DEVELOPMENT OF CANCER BIOMARKER ASSAYS FROM DNA IN VARIOUS BODILY FLUIDS

by
Simeon U. Springer

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
May, 2017

Abstract

DNA from cancer cells can enter surrounding fluid after apoptosis or necrosis. This DNA can be identified by sampling and sequencing the fluid looking for the mutations that gave rise to the cancer, referred to as a liquid biopsy. We have developed several liquid biopsies for the purpose of diagnosis and treatment selection through the use of targeted PCR and high throughput sequencing. We used the SafeSeqS methodology, a molecular barcoding technology previously developed in this lab, to further reduce sequencing errors to aid us in finding these rare mutations. We developed assays that identified genetic mutations from saliva, blood plasma, and pancreatic cysts. We investigating head and neck squamous cell carcinoma (HNSCC) from the oral cavity, oropharynx, hypopharynx, and larynx in 93 patients using a combination of saliva and plasma. We first identified either the E7 gene of HPV 16 or a primary mutation in the tumor and then tried to detect their presence in saliva and plasma. We collected saliva from every patient, while we only had plasma samples from 43 patients. We found that saliva performed best in oral cavity cancers detecting 100% of those patients, while plasma performed similarly across all sites detecting 87% of HNSCC patients. We detected mutations in 96% of patients when both saliva and plasma were available. Pancreatic cyst fluid was used to aid clinicians in the classification of pancreatic cysts. We created an 8 gene panel to look for mutations as well as additional tests for loss of heterozygosity and aneuploidy. We used DNA from fluid captured through endoscopic aspiration as well as from surgically resected cysts. When combined with the typical clinical features we were able to accurately predict which cysts needed surgery with a sensitivity and specificity of 89% and 69% respectively. Each of these projects began as an attempt to aid clinicians in dealing with these various diseases. Our results illustrate that liquid biopsies can be developed into an effective tool for the fight against cancer.

Thesis Advisor: Bert Vogelstein

Thesis Readers: Bert Vogelstein and Nick Papadopoulos

Acknowledgements

I would like to thank my mentors in the Ludwig Cancer Center for their support and advice. I would specifically like to thank the members of the A-Team without whom this work would not be possible. Everyone in the lab have been friendly, caring, and helped create an exciting place to work and learn.

I would also like to thank my friends that encouraged me along the way to pursue my degree and provided a much needed relief from the ensuing stress that developed.

Thank you as well to my parents, who believed in me. My mom in particular deserves a shout out for always being my biggest champion and who never stopped pushing me to reach for the stars.

Finally, the people who deserve the most thanks of all, my loving wife and daughter. The constant presence of their love and faith in me helped propel me through my entire time here in Baltimore.

Table of Contents

List of Tables

List of Figures

Chapter 1:

Introduction

Cancer is the result of the accumulation of somatic mutations in oncogenes and tumor suppressor genes. (Vogelstein and Kinzler 2004). When cancer cells die, their DNA can be released into the blood stream or other fluids. As early as 1977 it was discovered that cell-free DNA levels are raised in the serum of patients with cancer (Leon et al., 1977). These DNA fragments tend to be just under 200bp in size (Giacona et al., 1998) and have a half life anywhere from 16 minutes to two and a half hours in blood (Wang et al., 2016). Recenty studies have shown that circulating tumor DNA was found in plasma from all 15 different tumor types studied (Bettegowda et al., 2014). The presence of circulating tumor DNA could be used for several purposes including cancer diagnosis, assessing residual disease and risk of recurrence, treatment selection, and treatment monitoring. This has sparked a new area of research looking to create a liquid biopsy for the early detection of cancer.

The current method of performing a needle biopsy to get genetic data on a tumor has several drawbacks. In the case of cancer originating from internal organs, obtaining a biopsy can have complications in up to one in six biopsies (Overman et al., 2013). There is no such concern for taking a blood sample or collecting saliva. Furthermore, failure rates for these types of biopsies range from <10% to roughly 30% (VanderLaan et al., 2014). Other invasive clinical tests for finding cancer can involve inserting a fiber optic camera into a body cavity as in cystoscopy or colonoscopy. These procedures are painful to the patient and are therefore not often repeated, and in the case of colonoscopy are usually avoided despite their recommendation annual screening (Senore et al., 2011). A liquid biopsy, consisting of a blood draw or other sample collection, on the other hand is completely noninvasive and can give you the same genetic information. This makes taking multiple samples possible for following a patient during treatment or remission to get a timeline of the presence of the cancer and see if it has recurred.

Genetic profiling from circulating tumor DNA has already been shown to help inform and direct treatment to cancer. Circulating tumor DNA (ctDNA) can be used to get a genetic profile of the tumor in order to find which driver genes have been affected (Lebofsky et al., 2015). Liquid

biopsy can even provide additional information that is missed from conventional biopsies due to tumor heterogeneity and metastasis (Murtaza et al., 2015). This is important when considering that certain tumor types, such as those containing the $EGFR^{T790M}$ mutation, confer resistance to certain chemotherapy drugs (Oxnard et al., 2014). An accurate representation of the tumor mutations via liquid biopsy can lead to a properly informed treatment course.

Recent advances in sequencing techniques have increased the availability and reliability of these technologies for the use ctDNA discovery. Previous discovered blood based protein markers such as PSA and CEA lack in specificity and can arise for a number of reasons (Duffy, 2006), whereas genetic alterations are unique for neoplastic cells. With the development of digital droplet PCR and high throughput sequencing we are now able to detect low frequencies of ctDNA. To further increase our level of detection and help quantify the mutant fraction we find we used the SafeSeqS error reduction methodology invented in the Vogelstein lab. SafeSeqS is a type of molecular barcoding that attaches a unique 14 base sequence (UID) at the beginning of every template DNA molecule. This allows us to filter out artifacts introduced during PCR and sequencing as every read with the same UID should have the same sequence. This allows detection of mutations with frequencies as little as 0.01% (Kinde et al., 2011).

My thesis work focused on the development of molecular biomarker based assays of cell free DNA from various bodily fluids. The goal was to create a test that would help solve problems that clinicians currently face. These studies are the first step in bringing genetic based assays into a hospital setting where they can guide diagnoses and treatment decisions. Chapter 2 is a study that looked at DNA in fluid from four neoplastic cysts. These four neoplastic cysts (serous cystadenomas (SCA), solid-pseudopapillary neoplasm (SPN), mucinous cystic neoplasm (MCN), and intraductal papillary mucinous neoplasm(IPMN)) all have different treatment guidelines. SPNs and MCNs are malignant and require resection, while IPMNs tend to be stratified and only the high-grade should be removed (Law et al., 2014; Tanaka et al., 2012; Lennon et al., 2014).

Alternatively, barring symptoms SCAs can be monitored with no direct need for surgery (Lennon et al., 2013). A recent study has suggested that over 20% of resected cysts were benign meaning patients unnecessarily underwent a complicated highly invasive procedure (Valsangkar et al., 2012). My study took fluid from the cyst to identify the genetic alterations in the effort to find differences between them based on their mutational profile. The goal was to try to differentiate between cysts recommended for surgery and those that were not.

There is a need for biomarker assays for cancers that do not currently have one available such as head and neck squamous cell carcinoma (HNSCC). Chapter 3 focuses on the proof of principal study to develop an assay to try to detect HNSCC from saliva and plasma. More than half a million patients are diagnosed with HNSCC every year, making it the 7th most common cancer worldwide. Alarmingly, the incidence of certain types of HNSCC is on the rise. The outcomes associated with HNSCC are poor with a 5-year overall survival of only ~50%. We did whole exome sequencing of tumors from patients with cancer in the oral cavity, oropharynx, larynx, and hypopharynx. We then tested to see if saliva or blood plasma were capable of detecting the same mutations found in the tumors and determine which was more sensitive for detecting mutations. We also looked at both early, stage I and II, and late, stage III and IV, cancers to examine is we were able to find both early and late stage cancers.

Chapter 2:

Identifying molecular markers to improve the classification of pancreatic cysts

INTRODUCTION AND RATIONALE

Pancreatic cysts have been reported as incidental findings in 3% to 13% of individuals undergoing computed tomography (CT) or magnetic resonance imaging (MRI) (Laffan et al., 2008; Lee et al., 2010). The four most common types of neoplastic cysts of the pancreas are serous cystadenomas (SCA), solid-pseudopapillary neoplasm (SPN), mucinous cystic neoplasm (MCN) and intraductal papillary mucinous neoplasms (IPMN). SCAs have a very small risk of malignant transformation, and surveillance is usually recommended for these cysts in asymptomatic patients (Lennon et al., 2013). SPNs are low-grade malignant neoplasms (Bosman, 2010), and should be surgically resected when possible (Law et al., 2014). MCNs have the potential to progress to malignancy (Yamao et al., 2011), and current guidelines recommend surgical resection if possible (Tanaka et al., 2012). IPMNs can progress from low, to intermediate, to high-grade dysplasia, and ultimately to invasive adenocarcinoma (Lennon et al., 2014). Ideally, it is recommended that IPMNs with high-grade dysplasia, or an associated invasive adenocarcinoma, should undergo resection, while IPMNs with low- or intermediate-grade dysplasia should undergo surveillance (Tanaka et al., 2012; Sohn et al., 2004; Sahora et al., 2013).

The clinical management of patients with pancreatic cysts is currently based on clinical presentation, imaging and cyst fluid analysis (Tanaka et al., 2012). However, this approach is imperfect. For example, an evaluation of surgically resected pancreatic cysts at a high-volume center found that over 20% of the cysts resected due to concerns about their malignant potential, were entirely benign on histopathologic examination; in hind-sight, these cysts could have been safely observed (Valsangkar et al., 2012). Similarly, over 75% of resected IPMNs harbor only low-, or intermediate-grade dysplasia, and these also could have been safely observed (Sahora et al., 2013). Thus, better diagnostic tools are required to determine which patients truly benefit from surgical resection and which patients can be safely observed.

We previously performed whole exome sequencing of a well-characterized series of SCAs, SPNs, MCNs and IPMNs, and identified a distinct mutational profile in each cyst type (Wu et al., 2011; Wu et al., 2011). For example, *VHL* alterations were characteristic of SCAs, β-catenin gene (*CTNNB1*) mutations were found in SPNs, *GNAS* mutations in IPMNs, and *KRAS* and *RNF43* alterations were observed in both IPMNs and MCNs.

We have previously used an algorithm for analyzing multi-parametric features (MOCA, for Multivariate Organization of Combinatorial Alterations) to identify composite clinical markers of pancreatic cyst type and grade from a 1026-patient cohort (Masica et al., 2017). These composite markers, which included only clinical features such as age, sex, symptoms, and radiologic appearance, identified the common cyst types with high but imperfect accuracy (84% to 92%), and also correctly identified which cysts needed surgical resection with 82% accuracy.

The aim of the current study was to determine whether the molecular genetic features of pancreatic cyst fluid could be used to classify cysts and identify those that require surgical resection. Furthermore, we wished to determine whether combining the molecular markers identified here with the clinical markers identified by Masica et al, would increase the accuracy of diagnosis over either one alone (Masica et al., 2017).

METHODS

PATIENTS

The study was approved by the Institutional Review Boards for Human Research (IRB numbers: 00001584, 00-032, 1011003217, MOD07030072-52), and complied with Health Insurance Portability and Accountability Act.

Patients with SCAs, SPNs, MCNs or IPMNs who had undergone surgical resection at the participating institutions between September 2004 and September 2013 were included in the study. General demographics, the presence of symptoms, CT, MRI, endoscopic ultrasound (EUS)

features, cytology and cyst fluid carcinoembryonic antigen (CEA) levels were documented. When available, cross sectional imaging studies were reviewed by a single, experienced abdominal CT/MRI radiologist (S.R.), and the EUS studies reviewed by an experienced pancreatic endosonographer (A.M.L.). Multiple pancreatic cysts were defined as the presence of more than one cyst within the pancreas, which were anatomically separate from each other. The pathology of the surgically resected lesions was reviewed by one of two pancreatic pathologists (Bosman, 2010). The decision to resect a pancreatic cyst is multifactorial, and includes not only an assessment of the risk of the presence of high-grade dysplasia or invasive cancer within a cyst, but also the presence of symptoms secondary to the cyst, the age of the patient, and patient co-morbidities. For this study, cysts were considered as appropriately resected if they were found on histopathologic examination to be SPNs, MCNs, or IPMNs that had high-grade dysplasia or were associated with adjacent invasive adenocarcinoma.

CYST FLUID COLLECTION

Pancreatic cyst fluid was collected at the time of EUS or from the resected specimen in the surgical pathology laboratory. In the majority of instances, the results presented are from the analyses of cyst fluid aspirated post-operatively from the resected specimen. In 24 cases, paired samples, one of which was obtained at the time of EUS and the second at the time of surgical resection, were available from the same patient for mutation assessment, loss of heterozygosity (LOH) and aneuploidy analyses.

DNA PURIFICATION

DNA was purified from cyst fluid (0.25 to 1.0 mL) by adding 3 ml of RLTM buffer (Qiagen) and then binding to an AllPrep column (Qiagen) according to the manufacturer's instructions. DNA amounts were assessed by qPCR using the primers conditions described in

(Rago et al., 2007). A subset of the cysts reported here had been previously analyzed for *KRAS* and *GNAS* mutations (Wu et al., 2011).

ASSESMENT OF MUTATIONS

Because of their tremendous throughput, massively parallel sequencing instruments are highly cost effective instruments for DNA mutation analysis. However, sample preparation and sequencing steps introduce artifactual mutations into analyses at a low but significant frequency of approximately 9.1×10^{-6} (Kinde et al., 2011). In most clinical settings, this is irrelevant, as mutations of interest occur in a high fraction of alleles: 50% for germline mutations and greater than 10% for most tumors. Mutations in cysts occur much less frequently, often at 1% or less, and conventional sequencing cannot detect such low frequency mutations in a confident fashion. To better discriminate genuine mutations from artifactual sequencing variants introduced during these processes, we used Safe-SeqS, a technique which decreases the error rate from 9.1×10^{-6} to 4.5×10^{-7}, and allows detection of mutations present in as few as 0.01% of alleles, depending on sequencing depth and position of the mutation (Kinde et al., 2011; Kinde et al . 2013).

Safe-SeqS amplification primers were designed to amplify 60-bp to 262-bp segments each containing a region of interest. These regions of interest were derived from the following genes; *BRAF, CDKN2A, CTNNB1, GNAS, KRAS, NRAS, PIK3CA, RNF43, SMAD4, TP53, VHL* with primer sequences described in Table 2.5. These primers were used to amplify DNA in 25-uL multiplex PCRs reactions (Kinde et al., 2011; Kinde et al., 2013). For each sample, eight multiplex PCRs were performed, with each multiplex PCR containing 9 to 28 primer pairs. Reactions were purified with AMPure XP beads (Beckman Coulter) and eluted in 100 uL of Buffer EB (Qiagen, cat. no. 19086). Five uL of purified PCR products were then amplified in a second round of PCR (Kinde et al., 2011; Kinde et al., 2013). The PCR products were purified with AMPure and used for sequencing on a MiSeq instrument.

High quality sequence reads were analyzed as previously described (Kinde et al., 2011). Briefly, we selected reads that contained high quality base calls in their first 14 cycles as assessed by the quality scores generated by the sequencing instrument, which indicate the probability that an individual base call was made in error (Ewing, B and Green, P 1998). Reads in which each of these 14 cycles had a quality score >=15 were retained for further analysis. The template-specific portion of the reads that contained the sequence of an expected amplification primer was matched to the reference sequences using Bowtie (Bowtie 0.12.8). The unique identifier sequences (UIDs) that were incorporated as molecular barcodes into each template were used to group reads from a common template (Kinde et al., 2011). Artifactual mutations introduced during the sample preparation or sequencing steps were reduced by requiring that >90% of reads sharing the same UID contained the identical mutation (a "supermutant"). Normal peripheral blood DNA was used as a control to identify potential false positive mutations. Only supermutant frequencies in cysts that far exceeded supermutant frequencies in the control DNA samples (i.e., > mean + 5 standard deviations) were scored as mutations.

LOH ANALYSIS

This was performed in a fashion similar to that described above for mutations, but different primer sets were used. The primer sets amplified genomic regions of ~120 bp that contained common single nucleotide polymorphisms (SNPs) that were within or closely surrounding (within 1 Mb) the tumor suppressor genes *CDKN2A, RNF43, SMAD4, TP53*, or *VHL*. Analogously to the mutation protocol, each DNA sample was used for six multiplex PCRs, each containing 11 to 32 primer pairs (Table 2.6). The analysis was also carried out similarly, with the goal of identifying independent template molecules, defined by their UIDs, that were informative for the analyzed SNPs. The primer pairs used in this analysis were chosen from a large number of amplicons in the same region after extensive experimentation. The fraction of template molecules containing either allele in each of the chosen SNP amplicons was found to be 50% +/− 2% (mean +/− 1 SD) in the

analyses of 20 samples of peripheral blood DNA. A sample was scored as having LOH if >80% of the informative SNPs at one of the five loci assessed (*CDKN2A, RNF43, SMAD4, TP53*, or *VHL*) had allelic fractions lower or higher than 45% or 55%, respectively (i.e., lower or higher than 2.5 standard deviations from the mean).

ASSESSMENT OF ANEUPLOIDY

Aneuploidy across all non-acrocentric autosomal arms was assessed with a technology called FastSeqS (Kinde et al., 2012). With this technology, a single PCR is used to amplify ~20,000 loci scattered throughout the genome. After massively parallel sequencing, the fractional representation of each chromosomal arm can be determined by summing the reads that correspond to the loci in each arm.

IDENTIFICATION OF INDIVIDUAL MOLECULAR FEATURES OF INTEREST AND COMPOSITE MOLECULAR MARKERS

The MOCA algorithm for analysis of multi-parametric data sets has previously been described (Masica and Karchin, 2011; Masica and Karchin, 2013). Briefly, MOCA first selects features of interest, and then selects collections of features using Boolean logic operations. The composite features (termed "composite markers") are compared to the phenotypes under consideration (i.e., cyst type or need for surgery), and the corresponding Fisher's exact two-tailed P-value, sensitivity, and specificity recorded. Leave-one-out cross validations are used to identify the composite molecular markers that perform best. As the algorithm progresses, an optimization strategy is implemented, resulting in algorithmic convergence on sets of composite markers with optimal performance for predicting the phenotype under consideration. For each crossvalidation calculation, P-values are corrected for multiple testing using the Benjamini and Hochberg false discovery rate (FDR); composite markers were only considered if they had an FDR-corrected P-

value of < 0.05. Furthermore, composite markers were only considered if they were selected in each of the cross validations for discriminating a particular type or grade of cyst.

MOCA was applied to an independent set of patients with pancreatic cysts, but without molecular data, to generate composite clinical markers (Masica et al., 2017). In the current study, we tested these composite clinical markers on a set of 130 patients in whom molecular data was obtained as described above. None of the data from the patients in the current study were used to develop the original composite markers described in that study (Masica et al., 2017).

Finally, we used MOCA to identify composite markers that incorporated both molecular and clinical features (composite molecular/clinical markers) in the same way as described above. Sensitivity and specificity were used to quantify the performances of the composite molecular, clinical, and clinical/molecular markers, with the post-operative diagnoses based on histopathologic criteria.

RESULTS

BASIC PATIENT AND CYST CHARACTERISTICS

The cohort consisted of 130 patients (12 with SCAs, 10 with SPNs, 12 with MCNs, 96 with IPMNs). Of the IPMNs, 30 were main duct, 10 were mixed duct, 55 were branch duct types, and one was an intraductal tubulopapillary neoplasm (ITPN). The general demographics, symptoms, cyst features and pathological diagnoses of each of the 130 patients are presented in Table 2.1. Preoperative imaging was not available in 6 (5%) patients (2 IPMNs, 2 MCNs, 2 SPNs). The presumed pre-operative diagnosis, and the presence of high-risk or concerning features, is presented in Table 2.7.

CYST FLUID

The minimum amount of cyst fluid analyzed was 0.25 mL. The median DNA concentration was 4.9 ng/µl (range 0.05–270 ng/µl).

MOLECULAR FEATURES

Three types of molecular genetic tests were applied to each cyst. The first involved a search for subtle mutations (e.g. missense mutations or small insertions or deletions) of the genes known to be altered in pancreatic cysts. The most frequently mutated regions of six oncogenes *BRAF, CTNNB1, GNAS, KRAS, NRAS, PIK3CA*, and the great majority of the coding regions of five tumor suppressor genes (*CDK2NA, RNF43, SMAD4, TP53*, and *VHL*) were analyzed.

The second test involved a search for LOH of the same five tumor suppressor genes. We designed a massively parallel sequencing-based test to evaluate LOH events in a quantitative fashion, presented for the first time in this manuscript, and applied it to the DNA of cyst fluids.

Finally, it has been observed that aneuploidy is associated with malignant progression of neoplastic lesions of the pancreas (Bosci et al., 1998). We suspected that an evaluation of aneuploidy might help to identify high-risk cysts as well as to discriminate cyst types, and implemented a previously described PCR-based method for this purpose (Kinde et al., 2012).

Table 2.2 summarizes the mutational, LOH and aneuploidy analyses of the cysts; a detailed description of the aneuploidy data is presented in Table 2.8. One or more intragenic mutations, LOH events, or aneuploid chromosomes was identified in 9 (75%) SCAs, 100 (100%) SPNs, 7 (58%) MCNs, 94 (98%) IPMNs. Overall, at least one molecular genetic alteration was detected in 92% of the cyst fluid samples.

There were distinct mutational profiles associated with each type of cyst. The *VHL* gene was mutated in the cyst fluid of five (42%) of the 12 SCAs; intragenic *VHL* mutations were not found in any of the other type of pancreatic cyst. In addition, 7 (64%) of the SCAs had LOH of

chromosome 3 at the *VHL* gene locus, with 1 (8%) of the SCAs revealing aneuploidy at chromosome 3p. This finding is consistent with the rationale for evaluating LOH and aneuploidy described above. Overall, 8 (67%) of the SCAs either had a mutation in *VHL*, LOH of chromosome 3 or aneuploidy of chromosome 3p. There were four SCAs in which no abnormality was detected. Two of these four cases had matching surgical samples which were analyzed and no mutation in *VHL*, LOH or aneuploidy was identified.

The cyst fluid from all ten (100%) SPNs harbored a mutation in *CTNNB1*. One of the ten (10%) also harbored a mutation in *TP53*. Six (60%) patients with SPN had aneuploidy, involving chromosomes 11p (n=2) or 16p (n=6).

Of the 12 cyst fluid samples from MCNs analyzed, *KRAS* was the most commonly mutated gene, with activating mutations found in six samples (50%). *RNF43* was mutated in one MCN cyst fluid sample (8%) and this sample also harbored a *KRAS* mutation. LOH of chromosome 18 at the *SMAD4* gene locus was identified in one of the 12 MCN cyst fluid samples. Two (17%) MCNs had aneuploidy, involving chromosome 5p (n=1) or chromosome 16p (n=1).

Cyst fluid samples from 94 of the 96 (98%) IPMNs contained at least one mutation, LOH or aneuplody. *KRAS* was the most prevalent altered gene in the cyst fluid samples from IPMNs (78%), with single base substitutions occurring at codons 12, 13 or 61. *GNAS* mutations were identified in 56 (58%) IPMNs, and *GNAS* mutations were not present in any other cyst type. All *GNAS* mutations were single base substitutions at codon 201, resulting in substitution of an arginine with histidine (R201H), cysteine (R201C), or serine (R201S). *GNAS* mutations occurred in 9 (82%) of the IPMNs with intestinal type histology. They were found at a lower prevalence in the gastric (n=39 (61%)) and pancreaticobiliary (n=3 (38%)) type IPMNs, while none of the 3 oncocytic type IPMNs harbored a G*NAS* mutation. Overall, 86 (91%) of the IPMNs had a mutation in *KRAS* or *GNAS*, and 45 (47%) had a mutation in both genes. Mutations in *RNF43, TP53, SMAD4* and *CDKN2A*, also occurred in IPMNs, but less commonly. Six (6%) IPMNs had a mutation in

CTNNB1. In contrast to SPNs, where *CTNNB1* mutations occurred in isolation, all of the six IPMNs with a *CTNNB1* mutation also had another characteristic mutation, LOH or aneuploidy. A summary of the mutational, LOH and aneuploidy analyses based on IPMN type is presented in Table 2.9.

COMPARISON BETWEEN CYST FLUIDS OBTAINED DURING EUS VERSUS THOSE OBTAINED AT SURGERY

For 24 of the 130 cyst fluid specimens collected at the time of surgery, matching cyst fluid collected preoperatively by EUS was also available for analysis. This group of matching samples included cyst fluid collected from 17 IPMNs, 3 MCNs, 2 SCAs, 1 SPN and 1 ITPN. The number of genetic alterations detected in the cyst fluid samples collected at EUS (21 of 24, 87.5%) and at the time of surgery (20 of 24, 83.3%) was similar. Of the 1266 possible genetic alteration results (mutations, LOH, or aneuploidy) for these 24 matched cyst fluid samples, 1198 of the 1266 (94.6%) were concordant.

CYST CLASSIFICATION VIA COMPOSITE MOLECULAR MARKERS

We then used MOCA to identify composite molecular markers based on the individual features described above. SCAs were identified with 100% sensitivity and 91% specificity by the absence of a *KRAS, GNAS, RNF43* mutation, or by the absence of aneuploidy in chromosome 5p or 8p (Table 2.3). The presence of a *VHL* mutation has previously been shown to be predominantly associated with SCAs (Wu et al., 2011; Wu et al., 2011). In addition, on examination of the new molecular data presented here, the presence of LOH in chromosome 3 in the absence of LOH in chromosomes 9, 17 or 18, was exclusively identified in patients with SCA. These two features were therefore added to form the SCA composite molecular marker in expectation that this would be a useful feature for future assessments of cyst type; these "manually" added genetic features had no effect on the performance of the composite marker.

SPNs were identified with 100% sensitivity and 100% specificity by the presence of a *CTNNB1* mutation and the absence of *KRAS, GNAS*, or *RNF43* mutations or chromosome 18 LOH (Table 2.3).

MCNs were identified with 100% sensitivity and 75% specificity by the *absence* of *CTNNB1* or *GNAS* mutations, chromosome 3 LOH, or aneuploidy in chromosome 1q or 22q.

Finally, IPMNs were identified with 76% sensitivity and 97% specificity by the presence of a mutation in *GNAS, RNF43*, LOH in chromosome 9, or aneuploidy in chromosome 1q or 8p (Table 2.3).

CYST CLASSIFICATION VIA COMPOSITE CLINICAL MARKERS

In an independent set of cysts, we separately identified composite clinical markers for each cyst type (Masica et al., 2017). The new, 130-patient cohort gave us the opportunity to validate these composite clinical markers in an independent cohort. When applied to the 130-patient cohort, the composite clinical markers had high sensitivity for SCAs, SPNs, and MCNs (100%, 89%, and 90%, respectively), and modest sensitivity for IPMNs (75%; Table 2.10). The specificities of the composite clinical markers ranged from 71% to 88% for SCAs, SPNS, MCNs and IPMNs. The sensitivities and specificities estimated by cross-validation in Masica et al 2017 (Table 2.10) were in general similar to those estimated by analysis of this new 130-patient cohort. Though there were some differences—such as a higher sensitivity for IPMNs in the Masica study—all the markers were highly significant when applied to the 130-patient study. Although some of the sensitivities and specificities were outside the 95% confidence intervals of the Masica et al study (Table 2.10), this is not unexpected given that there is almost an order-of-magnitude difference in sample sizes between the cohorts.

Intuitively, one would expect that the combination of two different sets of biomarkers could, at least in certain circumstances, provide higher accuracy than either alone. For this purpose, we used MOCA to identify a new composite marker set, called "Composite molecular/ clinical markers" that included the composite molecular markers noted above plus the clinical or radiologic features identified by Masica et al as useful for cyst classification (Masica et al., 2017).

Because the composite molecular marker was so sensitive for identifying SCAs, sensitivity was not increased by adding clinical or radiologic features (Table 2.3). However, the *absence* of main pancreatic duct (MPD) dilation, communication with the MPD, or abdominal pain, increased the specificity for identifying SCA from 91% to 98% without compromising the 100% sensitivity.

The sensitivity and specificity of the composite molecular marker for identifying SPNs were both 100%. The addition of the clinical or radiologic features to the molecular markers decreased the sensitivity by 11% and decreased the specificity by 8%, for identifying this cyst type.

MCNs were similar to SCAs in that the composite molecular markers alone had perfect sensitivity (100%) but imperfect specificity (75%). The *presence* of age <75 years, and the *absence* of all three clinical or radiologic features (male gender, multiple cysts, communication with the MPD) increased the specificity to 97%, with a slight decrease in sensitivity to 90%. In contrast, an increase in sensitivity was realized when any of the following features (age ≥ 85 years, abdominal pain, MDP dilation or communication with the MPD) were added to the composite molecular marker for IPMNs. This composite molecular/clinical marker panel increased the sensitivity for having an IPMN from 76% (composite molecular marker alone) to 94%, while slightly decreasing specificity (from 97% to 84%; Table 2.3).

IDENTIFICATION OF CYSTS THAT REQUIRE SURGICAL RESECTION

From a practical perspective, the most important question in the management of cyst patients is the decision to perform surgical resection. We wished to determine the accuracy of the various composite markers described here and in Masica et al 2017 for determining this need for resection. As noted in Materials and Methods, we considered IPMNs characterized post-operatively as containing high-grade dysplasia or associated with invasive carcinoma to have been the most appropriate IPMNs for surgical excision. We also considered cysts that were histopathologically (i.e., post-operatively) diagnosed as SPNs or MCNs to have required surgical excision. SCAs were considered to have not required surgical excision. IPMNs with low-grade or intermediate-grade dysplasia were also considered in retrospect to have not required surgical resection at that time.

PREDICTING WHICH CYSTS NEED SURGERY USING THE COMPOSITE MOLECULAR MARKERS

To further characterize the molecular characteristics of IPMNs with high-grade dysplasia or associated with invasive carcinoma we analyzed the results from the 96 patients with a resected IPMN. The features of IPMNs that best predicted high-grade dysplasia or associated with invasive carcinoma are presented in Table 2.11. These were the presence of a mutation in *SMAD4*, chromosome 17q LOH (the region containing *RNF43*), or aneuploidy in chromosome 5p, 8p, 13q or 18q. We manually added a mutation in *TP53*, or chromosome 17p LOH (the region containing *TP53*), as these features have previously been described to occur in IPMNs with high-grade dysplasia or associated with invasive carcinoma (Kanda et al., 2012). A composite marker for IPMNs with high-grade dysplasia and/or associated with an invasive carcinoma based on these features, together with the composite markers for SCAs, MCNs, and SPNs described in Table 2.3, were then used to analyze the entire set of 130 patients. These composite molecular markers correctly identified patients requiring surgery with a sensitivity of 75% and a specificity of 92% (Table 2.4).

PREDICTING WHICH CYSTS NEED SURGERY USUING COMPOSITE CLINICAL MARKERS, WITH OR WITHOUT COMPOSITE MOLECULAR MARKERS

The composite clinical markers (Masica et al., 2017), when applied alone to the set of 130 patients, were able to identify cysts that required surgery with a sensitivity of 77% and a specificity of 75%. This provides a validation of the clinical composite markers, which were predicted to have a sensitivity of 84% and a specificity of 81% based on the cross-validation among the separate, 1026 patients in the original study (Masica et al., 2017). When both the clinical and molecular features were combined into a composite molecular/clinical marker to predict which cysts required surgery and applied to the 130 patients, the sensitivity increased to 89%, but at the expense of specificity, which fell to 69% (Table 2.4). The composite molecular markers provide information about the risk of high-grade dysplasia or an associated invasive cancer in IPMNs at the point in time the analysis is performed, and do not predict the risk of developing high-grade dysplasia or an associated invasive cancer in the future.

POTENTIAL TO AVOID UNNECESSARY SURGERY

All cysts included in this study underwent surgical resection, however those patients with SCA, or IPMNs with low- or intermediate-grade dysplasia in retrospect, may not have required surgery at this point in time. In this study, had the molecular analysis been performed prior to surgery, many unnecessary surgeries could have potentially been prevented. For example, the composite molecular/clinical marker correctly identified all 12 SCAs. If this had been realized prior to surgery, surgical resection would likely have been avoided in most. Similarly, many IPMNs were resected because of concern for the presence of high-grade dysplasia or an associated invasive cancer; however 62 of these IPMNs had only low- or intermediate-grade dysplasia. Fifty-six (90%) of these 62 patients would have been correctly identified as not needing surgery at the time of their evaluation using the composite molecular marker. Of the 74 patients with SCAs or IPMNs that in retrospect did not meet the histopathological criteria for surgical resection, the composite molecular

marker correctly identified 67 of these cases, thus potentially decreasing the number of unnecessary operations by 91%.

Table 2.1 Cyst and patient characteristics

	All Samples N = 130	IPMN* N = 96	MCN N =12	SCA N =12	SPN N =10
Sex (n=130^)					
Female - no. (%)	83 (64)	52 (54)	12 (100)	9 (75)	10 (100)
Race (n=130^)					
African American - no. (%)	7 (5)	4 (4)	1 (8)	1 (8)	1 (10)
White - no. (%)	114 (88)	87 (91)	10 (83)	8 (67)	9 (90)
Other - no. (%)	9 (7)	5 (5)	1 (8)	3 (25)	0 (0)
Age at surgery (n=130^)					
Years - mean (SD)	62.3 (17)	69.2 (10)	45.5 (14)	54.9 (15)	24.1 (4)
Symptoms (n=130^)					
Abdominal pain - no. (%)	23 (18)	21 (22)	2 (17)	0 (0)	0 (0)
Pancreatitis - no. (%)	16 (12)	14 (15)	2 (17)	0 (0)	0 (0)
Jaundice - no. (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Weight loss - no. (%)	6 (5)	5 (5)	1 (8)	0 (0)	0 (0)
Diabetes - no. (%)	25 (19)	21 (22)	1 (8)	2 (17)	1 (10)
Cyst size (n=130^)					
cm - median (IQR)	3.4 (2)	3 (2)	5.2 (4)	4 (2)	4 (2)
Cyst location (n=130^)+					
Head or Uncinate – no. (%)	56 (45)	51 (55)	0 (0)	1 (8)	4 (40)
Neck – no. (%)	16 (13)	13 (14)	0 (0)	2 (17)	1 (10)
Body or Tail – no. (%)	65 (52)	42 (45)	12 (100)	9 (75)	5 (50)
Multiple cysts (n=124^)					
Yes - no. (%)	31 (25)	29 (31)	0 (0)	1 (8)	1 (11)
Communication with MPD (n=107^)					
Yes - no. (%)	39 (36)	38 (46)	1 (14)	0 (0)	0 (0)
Mural Nodule (n=123^)					
Yes - no. (%)	32 (26)	19 (20)	3 (30)	2 (17)	8 (100)
CEA > 192 ng/mL (n=51^)					
Yes - no. (%)	30 (59)	24 (60)	6 (86)	0 (0)	0 (0)
IPMN Histotype (n=95^)					
Gastric - no. (%)	-	64 (67)	-	-	-
Intestinal - no. (%)	-	11 (11)	-	-	-
Pancreatobiliary - no. (%)	-	9 (9)	-	-	-
Oncocytic - no. (%)	-	3 (3)	-	-	-
Mixed - no. (%)°	-	8 (8)	-	-	-
Grade of Dysplasia/Invasive Cancer in IPMNs, and MCNs (n=108^)					
Low - no. (%)	25 (23)	15 (16)	10 (83)	-	-
Intermediate - no. (%)	49 (45.5)	47 (49)	2 (17)	-	-
High - no. (%)	22 (20.5)	22 (23)	0 (0)	-	-
Invasive Cancer - no. (%)	12 (11)	12 (12)	0 (0)	-	-

*Includes one ITPN.

^Indicates the number of patients in whom data on this variable was available.

+All sites where a pancreatic cyst was located were documented. Since some of the cysts extended to more than one site, this resulted in a number of locations that was greater than the number of cysts.

°Indicates more than one histologic su btypes in the same lesion.

SD = standard deviation. IQR = Interquartile range.

Table 2.2 Pancreatic cysts genetic alterations

	IPMN* N = 96	MCN N = 12	SCA N = 12	SPN N = 10
KRAS - no. (%)	75 (78)	6 (50)	0 (0)	0 (0)
GNAS - no. (%)	56 (58)	0 (0)	0 (0)	0 (0)
RNF43 - no. (%)	36 (38)	1 (8)	0 (0)	0 (0)
CDKN2A - no. (%)	3 (3)	0 (0)	0 (0)	0 (0)
CTNNB1 - no. (%)	6 (6)	0 (0)	0 (0)	10 (100)
SMAD4 - no. (%)	5 (5)	0 (0)	0 (0)	0 (0)
TP53 - no. (%)	9 (9)	0 (0)	0 (0)	1 (10)
VHL - no. (%)	0 (0)	0 (0)	5 (42)	0 (0)
BRAF - no. (%)	1 (1)	0 (0)	0 (0)	0 (0)
NRAS - no. (%)	0 (0)	0 (0)	0 (0)	0 (0)
PIK3CA - no. (%)	0 (0)	0 (0)	0 (0)	2 (20)
LOH chr3 (*VHL*) - no. (%)	4 (4)	0 (0)	7 (64)	0 (0)
LOH chr9 (*CDKN2A*) - no. (%)	8 (8)	0 (0)	0 (0)	0 (0)
LOH chr17 (*RNF43*) - no. (%)	11 (11)	0 (0)	1 (9)	0 (0)
LOH chr17 (*TP53*) - no. (%)	5 (5)	0 (0)	0 (0)	0 (0)
LOH chr18 (*SMAD4*) - no. (%)	10 (10)	1 (8)	0 (0)	0 (0)
Aneuploidy^	48 (50)	2 (17)	6 (50)	6 (60)

* Includes one ITPN patient.

^ Aneuploidy of at least one chromosome observed. Details are provided in Supplemental Table 3.

LOH = Loss of heterozygosity. chr = chromosome.

Table 2.3 Identification of pancreatic cyst type

Type of Cyst	Composite Molecular Markers				Composite Molecular AND Clinical[1] Markers			
	Any of these *Present*	Any of these *Absent*	Senstivity (95% CI)	Specificity (95% CI)	Any of these *Present*	Any of these *Absent*	Senstivity (95% CI)	Specificity (95% CI)
SCA	*VHL*[#] chr3 LOH[*#]	*KRAS* *GNAS* *RNF43* chr5p aneu chr8p aneu	100% (74–100%)	91% (84–95%)	Age ≥25 years	Abdominal pain Communication with MPD MPD dilation	100% (74–100%)	98% (94–99.8%)
SPN	*CTNNB1*	*KRAS* *GNAS* *RNF43* chr18 LOH	100% (69–100%)	100% (97–100%)	Age <55 years	Jaundice Multiple cysts Weight loss	89% (52–99.7%)	92% (85–96%)
MCN	None	*CTNNB1* *GNAS* chr3 LOH chr1q aneu chr22q aneu	100% (74–100%)	75% (66–82%)	Age <75 years	Sex (male) Communication with MPD Multiple cysts	90% (56–99.8%)	97% (91–99%)
IPMN	*GNAS* *RNF43*[^] chr9 LOH chr1q aneu chr8p aneu	None	76% (66–84%)	97% (85–99.9%)	Age ≥85 years Communication with MPD MPD dilation Abdominal pain	None	94% (86–98%)	84% (64–95%)

[#] These features were not identified by MOCA but were manually added.

[*] LOH of chromosome 3 but no LOH of chromosomes 9, 17 or 18.

[^] Mutations in *RNF43* have also been reported in MCNs, and in a larger cohort this feature may not be helpful for identifying IPMNs.

[1] The composite clinical marker is described in Masica et al[14].

MPD = main pancreatic duct. chr = Chromosome. CI=confidence intervals. aneu = aneuploidy.

Table 2.4 Comparison of different classification methods

	Sensitivity	Specificity	PPV	NPV
Composite molecular marker	75%	92%	88%	83%
Composite clinical marker	77%	75%	70%	81%
Composite molecular/clinical marker	89%	69%	69%	88%

[*] For this study, cysts were considered as appropriately resected if they were found on histopathologic examination to be SPNs, MCNs, or IPMNs that had high-grade dysplasia or were associated with adjacent invasive adenocarcinoma. Composite molecular and clinical markers are defined in Supplemental Table 5.

PPV=Positive predictive value. NPV=Negative predictive value.

Table 2.5 Primer sequences used in SafeSeqS

Gene	Forward Primer	Reverse Primer	Chr	Hg19 Start-End
BRAF	TGTTTTCCTTTACTTACTACACCTCAGA	AACTGTTCAAACTGATGGGACC	7	140453097-140453220
CDKN2A	GGGGAGAGCAGGCAGC	CACCTCCTCTACCCGACCC	9	21974743-21974859
CDKN2A	AGCCTTCGGCTGACTGG	GGCCTCCGACCGTAACTATT	9	21974684-21974798
CDKN2A	GGGTCGGGTAGAGGAGGTG	CTCCCGCTGCAGACCCT	9	21974652-21974761
CDKN2A	TGGCTCTGACCATTCTGTTCT	GGGTCGGGTGAGAGTGG	9	21971115-21971237
CDKN2A	CCGAGTGGCGGAGCTG	CACCAGCGTGTCCAGGAAG	9	21971073-21971187
CDKN2A	GACCCCGCCACTCTCAC	GCTCCTCAGCCAGGTCCA	9	21970997-21971138
CDKN2A	CTGGACGTGCGCGATG	GCATGGTTACTGCCTCTGGT	9	21970930-21971048
CDKN2A	AGGAGCTGGGCCATCG	ACAAATTCTCAGATCATCAGTCCTC	9	21970874-21971002
CDKN2A	TGTGCCACACATCTTTGACC	CTGTAGGACCTTCGGTGACTG	9	21968163-21968300
CTNNB1	GCCATGGAACCAGACAGAAA	CCTCAGGATTGCCTTTACCA	3	41266040-41266163
GNAS	CCAGACCTTTGCTTTAGATTGG	GGTCTCAAAGATTCCAGAAGTCA	20	57484325-57484449
KRAS	TTTACCTCTATTGTTGGATCATATTCG	TGACTGAATATAAACTTGTGGTAGTTGG	12	25398203-25398317
KRAS	TTCTCCCTTCTCAGGATTCCTAC	TGTACTGGTCCCTCATTGCAC	12	25380244-25380360
KRAS	GGAAATAAATGTGATTTGCCTTCT	TTTCAGTGTTACTTACCTGTCTTGTCTT	12	25378532-25378655
NRAS	ACACCCCCAGGATTCTTACAG	CGCCTGTCCTCATGTATTGG	1	115256485-115256609
PIK3CA	TTATTCCAGACGCATTTCCAC	ACATTCACGTAGGTTGCACAAA	3	178921436-178921555
PIK3CA	AAATAATAAGCATCAGCATTTGACTTT	CATCTACAAAATCCCTTTGGGTTAT	3	178921483-178921606
PIK3CA	TTTGATGACATTGCATACATTCG	GATCCAATCCATTTTTGTTGTCCAG	3	178951991-178952119
PIK3CA	CAATGAATTAAGGGAAAATGACAAA	CTCCATTTTAGCACTTACCTGTGAC	3	178936018-178936140
RNF43	GGTACCCATACCAGCCCCTA	CCTGTGTGTGCCATCTGTCT	17	56438051-56438185
RNF43	GGTACCCATACCAGCCCCTA	TGTGTGCCATCTGTCTGGAG	17	56438051-56438181
RNF43	ATGGTTGAAGTGCATTGCTG	CAGTCCTGTGCGTCCAAAG	17	56492849-56492969
RNF43	CCTGGCTGCTGATGGCTAC	GTCCATTTTCAAGGGGATCA	17	56492771-56492898
RNF43	CAGCGGTGGAGTCTGAAAGA	TTTCAGCAACACCAGCAAAC	17	56492716-56492838
RNF43	GGACCCCACAGGAAAACTG	ACATATTTCAAACAGATGGAAAGTGA	17	56492639-56492774
RNF43	GACCCCACAGGAAAACTGAA	ACATATTTCAAACAGATGGAAAGTGA	17	56492639-56492773
RNF43	GACCCCACAGGAAAACTGAA	CATATTTCAAACAGATGGAAAGTGAA	17	56492640-56492773

RNF43	AGCTCACTCCTGCTGCCTAA	CTCTCCAGCTTGACGATGCT	17	56448315-56448448
RNF43	AATGCCAGTGATGACGACAA	TGTGTAGGGCGAAGTGTGAG	17	56448247-56448373
RNF43	CTCAGCCCAACCTCTACTGTG	CCTTGAACACGCAAATGTCC	17	56440859-56440993
RNF43	GCAGGGAGAAGTCACAGCA	CACAAACTCCATCAGCTTCTCA	17	56440690-56440812
RNF43	TGGTGTTGATCTGGGGTAATG	GGCAAGGTCTGGAGGTCTAGT	17	56440607-56440736
RNF43	CAGCTCATCCCGGAAGTATG	AGCACCGAAGCCAGGATG	17	56439939-56440057
RNF43	GTGGGCACCATCTTTGTGAT	TAACCACCCACCCACACAC	17	56439845-56439976
RNF43	ATCTGAGCCCCATTCCTCTT	CTCCCTGAGTCTGGCCACT	17	56438202-56438330
RNF43	GGCCACCAGGAGGTACCAG	CTTACCTGCCCCTCAGAGAAC	17	56438139-56438264
RNF43	GGTACCCATACCAGCCCCTA	GTGTGTGCCATCTGTCTGGA	17	56438051-56438182
RNF43	ACAGAAGCCTTTGGTTTGGA	GCAAGTCCGATGCTGATGTA	17	56437535-56437650
RNF43	CATTTCCTGCCTCCATGAGT	TGCCTACACAGAGGGGAGTC	17	56437473-56437601
RNF43	CATCTGTGCTCTTGGTTCTTTTT	GAGGTGGTAGTGGGCATGG	17	56436078-56436212
RNF43	AGGTCGAAGACTCCACCTCA	ATGGCAGGAAGGGACCAG	17	56436005-56436132
RNF43	CTCCCTGCTGCCTACCTGT	ATGTGCAGCTCTGGGGAAG	17	56435952-56436080
RNF43	GCTCCAGGAGAGCAGCAG	GTAGGGGCACTGGGCAAG	17	56435831-56435945
RNF43	ACCTCCAATCCACCTCACAG	CCATCTGCCAGGTACCCACT	17	56435755-56435878
RNF43	CTGACAGCAGTGGATCTGGA	CGTGCAGTTGACCACAGAGT	17	56435688-56435815
RNF43	CCTGTCATGGCTCTTCCAGT	TGCAGTACACTAGGGGGTCAA	17	56435603-56435728
RNF43	GCAGCTCCCTAAGCAGTGAC	ACCACCGAGTCCAAGGAAC	17	56435527-56435644
RNF43	GGACATGCAGCCTAGTGTGA	CCGCTTTTTGTAGTGGTGGT	17	56435454-56435577
RNF43	TTCCAGCCATGTCCACTACC	GCTGTGTCCGAGGAATAGGA	17	56435378-56435505
RNF43	CCAGCCATGTCCACTACCAC	GAGGAATAGGAGGCCTGGAC	17	56435387-56435503
RNF43	AACCGGAGTCCCCCAGTC	GGCACTGTGGGTTAGAGAGC	17	56435291-56435421
RNF43	AACTCAGCAGCCCCTTCG	TGGATTTTTGCAAGTTGAACAG	17	56435201-56435333
RNF43	CTGCCCCAGTACCAGCAG	GCAAGCTGGGTGCACAGT	17	56435103-56435241
RNF43	CGACACCCACAGAGGAAAAG	GGACCAAGGATATGCCACAC	17	56435058-56435189
RNF43	TGCCAGATTTTTCCCCATTA	GGCCTGGGGTTTCTGGTAG	17	56434988-56435105
RNF43	ACCCCTTGATCTGTGGACCT	CAGGTGGATGTGGTTCCAG	17	56434916-56435047
RNF43	CAGTGTGGTTGTGCCTGACT	CTGGCAGTGCGGATAAGG	17	56434848-56434966

RNF43	CTGGAACCACATCCACCTG	CACCCACTTCCCTCTGAAAA	17	56434804-56434934
RNF43	AGGTAGGGCCCAAACACATC	TTCCTCCCCGCTCTCTAAAT	17	56432247-56432371
SMAD4	TTCCAAAGGATCAAAATTGCTT	ATGCACAATGCTCAGACAGG	18	48573350-48573479
SMAD4	TTACGAATACACCAACAAGTAATGATG	TCTTTACCAAACTTTCAATTGCTCT	18	48573433-48573552
SMAD4	ACAAGGTGGAGAGAGTGAAACATT	CATTTACTAGGATGAGCTCCATTTG	18	48573497-48573628
SMAD4	AAGTTTGGTAAAGAAGCTGAAGGA	AGCCTCCCATCCAATGTTCT	18	48573539-48573661
SMAD4	GCTATAACTACAAATGGAGCTCATCC	AAATCTGCCACCATAGAGGGTA	18	48573594-48573711
SMAD4	GTGACACATGAATAAATGGTCGTT	TTTAGTTCATTTTTGTGAAGATCAGG	18	48575021-48575135
SMAD4	CTATGCCCGTCTCTGGAGGT	ACAACTCGTTCGTAGTGATATGGA	18	48575088-48575216
SMAD4	CAGTATGCGTTTGACTTAAAATGTG	TCCTTTTCCCTTTATGTTTCTTAGG	18	48575152-48575276
SMAD4	AAAAACAGAGAGGATAGGACAAAACA	CATTTGTTTTCCCCTTTAAACAATTA	18	48575637-48575751
SMAD4	TTTTTCTGGGAATAGAAGCTTATAAAAAT	TGTCCTTCAGTGGACAACGA	18	48581095-48581226
SMAD4	GTGAAGGATGAATATGTGCATGA	AGAGCTGGGGTGCTGTATGT	18	48581168-48581295
SMAD4	CAGCATCCACCAAGTAATCGT	GTGGAAGCCACAGGAATGTT	18	48581243-48581361
SMAD4	CCCCCAAGTGACTACACATAAATAA	CTACCAGCACTGCCAACTTTC	18	48581319-48581440
SMAD4	ATCCATTCTGCTGCTGTCCT	CCATGTTAATGTCTTCTTGTTCCTCT	18	48584467-48584584
SMAD4	AGGCTGCCTACTTTTTTCTCAAC	CAGATAGCATCAGGGCCTCAG	18	48584542-48584669
SMAD4	CGGGTGGTGCTGAAGATG	AATTAACCCATGTGGGCCTTA	18	48584681-48584798
SMAD4	GGAAGTAGGACTGCACCATACAC	GCCCTTACAACAAAAACAAGAGC	18	48584730-48584853
SMAD4	CACCACTAAATCAATCTAAATACAGGAA	AAAATGGAATTTTTGTTGTCTTTTCT	18	48586206-48586336
SMAD4	CGTTTCAATCACCACTAAATCAATC	AATGGAATTTTTGTTGTCTTTTCTTTAG	18	48586208-48586345
SMAD4	GGGAGGATGTTCTTTCCCATT	AACAATAGGGCAGCTTGAAGG	18	48591760-48591881
SMAD4	TCCTTCAAGCTGCCCTATTGT	CCTTGCTCTCTCAATGGCTTC	18	48591860-48591977
SMAD4	TGTACATGGGAAAACATAACCTTG	TTGGGTCAACTCTCCAATGTC	18	48591927-48592062
SMAD4	ACATGGGAAAACATAACCTTGAAT	TTGGGTCAACTCTCCAATGTC	18	48591927-48592059
SMAD4	TTAAGCATGCTATACAATCTGAACTAAAA	CTTCACCTTTACATTCCAACTGC	18	48593307-48593432
SMAD4	GGTTGCACATAGGCAAAGGT	ACGCCCAGCTTCTCTGTCT	18	48593388-48593509
SMAD4	TTTGGGTCAGGTGCCTTAGT	CCTTTATATATGCACTTGGGTAGATCTTAT	18	48593439-48593558
SMAD4	TCCTTCCACCCAGATTTCAA	TGCACCTGGAGATGCTGTTC	18	48593509-48593636
SMAD4	TCAGGCATTGGTTTTTAATGTATG	GCCTGCTGCTGCATCTGT	18	48602937-48603051

SMAD4	TCAGGCATTGGTTTTTAATGTATG	CGCCTGCTGCTGCATCT	18	48602937-48603052
SMAD4	GATTTGCGTCAGTGTCATCG	GCTGGAGCTATTCCACCTACTG	18	48603014-48603141
SMAD4	CAAAGAAACTAATCAACTGAGTAAATCAAG	GCAGGAAACATCCCTGGC	18	48603095-48603227
SMAD4	CACCCTGTCCCTCTGATGTC	CCCAGCCTTTCACAAAACTC	18	48604586-48604706
SMAD4	TGGAATTGGTGTTGATGACCT	TTCAATCCAGCAAGGTGTTTC	18	48604642-48604756
SMAD4	ACCGGATTACCCAAGACAGAG	AAGGTTGTGGGTCTGCAATC	18	48604708-48604829
SMAD4	AGCATCAAAGAAACACCTTGCT	GGGCCCCAACGGTAAAA	18	48604727-48604858
TP53	CCTTCCAATGGATCCACTCAC	ACTGCCTTCCGGGTCACT	17	7579818-7579933
TP53	AGCCCCCTAGCAGAGACCT	CAGCCCAACCCTTGTCCTT	17	7579679-7579789
TP53	TGACTGCTCTTTTCACCCATC	TCATCTGGACCTGGGTCTTC	17	7579502-7579616
TP53	GCAATGGATGATTTGATGCTG	CGGTGTAGGAGCTGCTGG	17	7579441-7579572
TP53	AGCTCCCAGAATGCCAGAG	TGGGAAGGGACAGAAGATGA	17	7579388-7579501
TP53	CTGCACCAGCAGCTCCTAC	CAGAATGCAAGAAGCCCAGA	17	7579338-7579463
TP53	GCATTGAAGTCTCATGGAAGC	CCCTTCCCAGAAAACCTACC	17	7579270-7579396
TP53	GCCCTGACTTTCAACTCTGTCT	GGGGGTGTGGAATCAACC	17	7578475-7578593
TP53	CTCCGTCATGTGCTGTGACT	CAACAAGATGTTTTGCCAACTG	17	7578417-7578540
TP53	GCCATGGCCATCTACAAGC	ACCAGCCCTGTCGTCTCTC	17	7578335-7578455
TP53	GTCCCCAGGCCTCTGATT	CGAAAAGTGTTTCTGTCATCCA	17	7578211-7578326
TP53	GTGGAAGGAAATTTGCGTGT	CTTAACCCCTCCTCCCAGAG	17	7578137-7578260
TP53	TGTGATGATGGTGAGGATGG	TCATCTTGGGCCTGTGTTATC	17	7577513-7577635
TP53	TGGCTCTGACTGTACCACCATC	GTGGCAAGTGGCTCCTGA	17	7577480-7577606
TP53	TGCCTCTTGCTTCTCTTTTCC	GCGGAGATTCTCTTCCTCTGT	17	7577068-7577188
TP53	CGTGTTTGTGCCTGTCCTG	GCTTCTTGTCCTGCTTGCTT	17	7576997-7577121
TP53	TTTTATCACCTTTCCTTGCCTCT	CAAGACTTAGTACCTGAAGGGTGAA	17	7576840-7576956
TP53	AAGAAGAAAACGGCATTTTGAG	CCAGCCAAAGAAGAAACCAC	17	7576782-7576898
TP53	CCCTGGCTCCTTCCCAG	CTTCTCCCCCTCCTCTGTTG	17	7573947-7574059
TP53	GTTCCGAGAGCTGAATGAGG	TAGGAAGGCAGGGGAGTAGG	17	7573881-7574007
TP53	ATGTCATCTCTCCTCCCTGCT	AGTCTGAGTCAGGCCCTTCTG	17	7572929-7573043
TP53	GCCACCTGAAGTCCAAAAAG	GAGGCTGTCAGTGGGGAAC	17	7572893-7573009
VHL	CCCGGGTGGTCTGGAT	CCGCCGTCTTCTTCAGG	3	10183506-10183620

VHL	AGGCAGGCGTCGAAGAGTA	CGGCCTCCATCTCCTCCT	3	10183581-10183700
VHL	CGGCCCGGAAGAGTCC	GGACTGCGATTGCAGAAGAT	3	10183645-10183773
VHL	GCTGCGCTCGGTGAACT	CGGCAGCGTTGGGTAG	3	10183717-10183837
VHL	CCGTATGGCTCAACTTCGAC	CGTGCTATCGTCCCTGCT	3	10183788-10183917
VHL	CCGGTGTGGCTCTTTAACAA	GCTGTCCGTCAACATTGAGA	3	10188159-10188293
VHL	TCTATCCTGTACTTACCACAACAACC	CCAAACTGAATTATTTGTGCCATC	3	10188250-10188376
VHL	CACCGGTGTGGCTCTTTAAC	CCTGACATCAGGCAAAAATTG	3	10188157-10188423
VHL	GATTTGGTTTTTGCCCTTCC	GGTGGTCTTCCAGATCTTCGT	3	10191449-10191581
VHL	CAGGAGACTGGACATCGTCA	TCAATCTCCCATCCGTTGAT	3	10191532-10191649
VHL	CCCAAATGTGCAGAAAGACC	TCATCAGTACCATCAAAAGCTGA	3	10191580-10191703

Chr = Chromosome; Hg19: Human genome 19

Table 2.6 Primer sequences used in SafeSeqS for loss of heterozygosity

SNP	Gene Region	Forward Primer	Reverse Primer	Chr	Hg19 Start-End
SNP rs12602891	*RNF43*	TCAGAACTCCACCCAAGAGG	TCTGCTCCCCTCTCACTTTG	17	56269990-56270144
SNP rs11868650	*RNF43*	CCAAAGCCCTGCTAATTCTG	GAACTCGGTTTTGTGTCTGGA	17	56330905-56331126
SNP rs2632516	*RNF43*	CCCTGGTGTGTTCAAGACCT	GTCTCCAGTCCACCCCAAC	17	56409069-56409262
SNP rs1985749	*RNF43*	TTGTCCCGATAGGGAACTTG	TTTGCAGGAAGAAGAGTGCAT	17	56416970-56417121
SNP rs1859401	*RNF43*	CAGGGAGCTGTGAACCAAGT	ACTGTGCCAAGTGGGCTCT	17	56421550-56421728
SNP rs757485	*RNF43*	TCCTCAGGTCCATTAGGGAGT	TGATGCAACTCCAAGAGAAGAG	17	56424684-56424840
SNP rs2680702	*RNF43*	TCCCTGAAGCTTTCCCACTA	TCATTCCATCCTTCCTCTGC	17	56431523-56431720
SNP rs9904993	*RNF43*	ATGGGGGTAGTGGTGGAGAG	TTGTGAAGCAGTGGGACAAA	17	56468536-56468696
SNP rs7216856	*RNF43*	AGGAGCTTTGTTTTCCTCTCG	TGCTGTTACTTTCAGGGTTGTG	17	56472252-56472402
SNP rs2301867	*RNF43*	CCACAAAGGAAAACCTCAGC	AAATTGTCTATTCAGGTCCCAGAG	17	56480049-56480219
SNP rs1476596	*RNF43*	AACCAAATTCTCAAAACAAATTTACC	TGGAGAGCAGAATTCATTTACTACAC	17	56485332-56485531
SNP rs2526370	*RNF43*	TGCGCACAAACATTCTTAGG	TCATTAAAAATATTCTGCAGCTTCC	17	56487602-56487753
SNP rs3744093	*RNF43*	ATGGTTGAAGTGCATTGCTG	GTCCATTTTCAAGGGGATCA	17	56492771-56492969
SNP rs7215531	*RNF43*	TTTCCAACTTTGCTTGTGATTG	TTCTTCACGCTCACCTTTGAT	17	56493228-56493398
SNP rs7207286	*RNF43*	GTGACTGGGGGCCTTTTAAT	TGTTCCATGCTATACCAGATGC	17	56514313-56514490
SNP rs2302189	*RNF43*	TGCATGTTCTGCCTTCTGAG	CATCCAGGATCAGCAGCTTT	17	56584164-56584325
SNP rs2302190	*RNF43*	ACCCCGTGTGTTGTCATCTT	CTCAGCACCGGGAATAATGA	17	56584376-56584532
SNP rs3744108	*RNF43*	ATTTTCTTTGGGAGGCTGCT	TGTGTGAGACTCTCCAGACGTT	17	56585845-56586028
SNP rs740605	*RNF43*	CAGTGAACCCTGGGAATTGA	GCAGGGACATGAGGGAAGTA	17	56604789-56604941
SNP rs11078682	*TP53*	TTAGAAGGGTGGTTGGCTTG	GCTGCATTGTCTGGTGAGAA	17	7396033-7396219
SNP rs9890920	*TP53*	AACAACCTGAGGGTGACGAG	CCCTTCCAAGGAGAGACAGG	17	7399895-7400084
SNP rs7211392	*TP53*	CGCATTCTCCCATGGTCTAC	CATCTCATCCCCGTCAAAGT	17	7402445-7402642
SNP rs2277638	*TP53*	CAGCTCCTGGATCTCTGCTC	GGGTGGCTCCTCAAGGTT	17	7402536-7402693
SNP rs11078685	*TP53*	GATGTGCTGCATACCAGGTG	CACAGGGCCAGGACTGTAAT	17	7411331-7411489
SNP rs9889368	*TP53*	CAACAAGAGCGAGATTCCATC	TCCATGACACAGGGAACTGA	17	7412108-7412245
SNP rs2071502	*TP53*	CAGGAGTGGATCCTGGAGAC	CCTGAGTGGACAGCACAGAA	17	7414800-7414986
SNP rs4968212	*TP53*	TTTTCAGGGCTCAGATTTGG	GAGGCCTCCAGACTCTCCTC	17	7468078-7468248

SNP rs4968214	*TP53*	AAGGGTCACCCCCTTATGTC	GTCAGGAGTGGCCTGCAT	17	7477344-7477506
SNP rs727428	*TP53*	AAAACCCCGACAGACAGACA	GCCAGTCCAAGGACCAGAT	17	7537642-7537824
SNP rs2955617	*TP53*	CCCTCTCCATTGCTGAAGAC	GTCCCCTTCTGCAGAGATCA	17	7538763-7538962
SNP rs1641544	*TP53*	TAGGATGGGGTGAGTGGGTA	TCGCTTTGACATCGACTTTG	17	7539841-7539994
SNP rs1641511	*TP53*	CCTGCATATCCCCTGAGAGTT	GCGACTGTGTAGACAGGGAGA	17	7559520-7559725
SNP rs1050541	*TP53*	CCGAAGGGAAGAAACACAGA	AGGCATTTTAACCCTTTGTCCT	17	7560798-7560948
SNP rs7141	*TP53*	CCTCAATCACCAGCCAAGAT	AAAACTACTGAGGTGACGGCATA	17	7614461-7614624
SNP rs1544724	*TP53*	TTGCTGGCCTAACAGAACG	GCTGGGCACTCTCTTTTCATT	17	7621640-7621798
SNP rs3744258	*TP53*	GAGGAGCCTGGTGGGTACTT	CTGGCCAGGAGGAAGACA	17	7623209-7623412
SNP rs839721	*TP53*	TGGGCATTAAGGGAATTGAG	AGGACTCCGCACTAAGTGGA	17	7639621-7639783
SNP rs9909288	*TP53*	AGGCTCCAGCTGTTCATCAT	CGATGATGGCCCTTATCCTT	17	7673778-7673951
SNP rs9889453	*TP53*	TCCATATCCCATGCTCTTAACC	CCCCCAGCTCTCCTAACATT	17	7674878-7675007
SNP rs8066124	*TP53*	AACCCCAGTCCTTGCAGATT	CCGTCTTCCTAGATTTCCCATAG	17	7676557-7676749
SNP rs4791329	*TP53*	GCCTGTGTCTGCTTTGCAT	CAGAGGGCACGTGACAGATA	17	7693322-7693487
SNP rs9916791	*TP53*	GGCCTCACCTCTGACCTGTA	CTTCAGGCCCTAGAGAGCAA	17	7721076-7721272
SNP rs1105813	*TP53*	ACCTCCTTATCGCCAATGC	CCTGTCCCTCCATTCAGACA	17	7721494-7721666
SNP rs4791840	*TP53*	TTTTGCTTGATGTTTTGCACAT	CCATGTTACCATCCCCAATG	17	7744072-7744232
SNP rs1565816	*TP53*	TTTAGTGCTTCCGGTGTGC	TGGGAGAACAGCCACTTGAT	17	7767438-7767614
SNP rs4940009	*SMAD4*	TTTTGCAATTCTGAAGTCATGTG	GGAATCTCAGGTGTTAGTTCTGCT	18	48263843-48264034
SNP rs2255059	*SMAD4*	GGTCCCCATCAACAAAACAA	ATCTCACCACCCACCCTGTA	18	48327879-48328036
SNP rs6508037	*SMAD4*	AGGCTCCTCATACAGATCCTCA	GGCAGGTTGAGCTAAGCAGA	18	48360391-48360550
SNP rs6508038	*SMAD4*	CTCCTAGCCCTGTCATCTGC	TTCATTTTCCAAAACCATTGC	18	48360457-48360627
SNP rs8099088	*SMAD4*	GCTTTGTTTCTATGATCTACATGGATT	CAAACAAACTTGATGTGGGAAA	18	48360833-48360987
SNP rs9958074	*SMAD4*	AGGCTGCTCCTGAACTCTTG	TGGTTACCATTTGGAATGAGAAT	18	48472050-48472245
SNP rs8088712	*SMAD4*	GCCCATTGAGATACACAGCA	CCTCCAGGTTCATAAGCTTCAC	18	48514231-48514419
SNP rs2276163	*SMAD4*	GGAAAAGGATCTCAATAGTGTTTCA	CAATACTCGGTTTTAGCAGTCAAA	18	48575268-48575431
SNP rs7229678	*SMAD4*	GATCCCCCTAGTGGTGGTTT	AAATGAAGGAACTGTCCCATTTT	18	48577670-48577824
SNP rs2298617	*SMAD4*	TGGAAAAGTGTCCAATTCTGAAA	TGAAGGAGAAAGGGGATGATT	18	48603365-48603534
SNP rs3819122	*SMAD4*	TCAATGACAAGCAGCCTTTG	AGCCTACCCTTGGATGAGTACA	18	48610800-48610974
SNP rs4939652	*SMAD4*	GTCCTCTGCCTGACCTGTTC	TCCAGTCCTTCCTTCAGTCTTAAC	18	48648939-48649132

SNP rs12954819	*SMAD4*	GAGGAGGAGGGGTTTGTGAT	TTCTTAGAAGGCAGGAAGAGAATTAG	18	48686150-48686333
SNP rs2282543	*SMAD4*	CACCTCCTACCTGGGTTTCA	TGACACATCTTCACAGCTTCG	18	48693561-48693780
SNP rs11663428	*SMAD4*	TGGGGTGACAAGAATGACAA	GCCGGCATTAAAAGTCCATA	18	48741247-48741369
SNP rs4940069	*SMAD4*	GGGCCTTCTGAACCTAGGAA	CCTCTGTTCAGGGGGAGTTC	18	48771489-48771657
SNP rs9807456	*SMAD4*	ACCCCCACAGCATAGTGAAC	AGGGAACTAGAGTGCGTTTCC	18	48813762-48813932
SNP rs2445468	*SMAD4*	AGAAAATGGGGCTAGGTGCT	TTAACCTTGGGCTGCCTTTA	18	48857105-48857323
SNP rs6807326	*VHL*	AACGCATCTGAATGGAACTTTT	GTCTTAGATGCCATTCTCCCTTT	3	10036684-10036843
SNP rs13066757	*VHL*	CAGGAAGCCAGTGTCATTCA	TGGGGTCATCTTATTCCTCCT	3	10044907-10045069
SNP rs7637888	*VHL*	TTAAGCACTTACTGCGTTGATGTT	TGCTAAATATGAACAGTGACTACATCC	3	10080578-10080756
SNP rs1642742	*VHL*	CCGCTACGGATGTAGAATGG	GGACAGCTTGTATGTAAGGAGGTT	3	10191918-10192047
SNP rs11465853	*VHL*	CCCCCAAGAAGTGACAGCA	TGCGCTCTGAGTCCAGAGAA	3	10206959-10207115
SNP rs708030	*VHL*	TGCCAGAATGAATGATGAGC	GACTGGGCGAGGAGCAGT	3	10212386-10212541
SNP rs2619508	*VHL*	TGTGTGGATGACCAATGCTT	AAAGCCTCCAGGAGAAGTGC	3	10215775-10215972
SNP rs6442159	*VHL*	AGGCGGTGAGAATGGAGATT	CCAAATCCTTTTGGACCACTT	3	10243267-10243416
SNP rs3844283	*VHL*	CGTGGGAAGGACAAATCAAG	CGTATCTGCCAGAGGATTTCA	3	10264445-10264571
SNP rs394558	*VHL*	CCAAGAACTCTAGGGGTGGAG	TGCTGCTCAGAAGGAGAAAGA	3	10302138-10302322
SNP rs451952	*VHL*	CCCTCAAGAGAAACCCAGTG	TGGAACACATTTCATCAACAGAA	3	10302240-10302413
SNP rs35684	*VHL*	CACCTCGTTCTTTGCCTTTC	CCAGTTCACAGCAGACACGA	3	10326535-10326719
SNP rs1629816	*VHL*	CTTGACCTTGCCTCTCATCC	ATCCAACACAACCCAACACA	3	10336154-10336333
SNP rs696220	*VHL*	AAGAACTCTTGCTTTAGGGACAAA	TGTGCATTTAGGCTTCTTGCT	3	10344007-10344156
SNP rs629067	*VHL*	GCCCTGAATGAAGCAGAGTC	CCCAGTCAAACCACCAAGAC	3	10359103-10359279
SNP rs4327369	*VHL*	GGAAACGTTGGTTTTCTCTCC	GACGTCGCTTTAGCTGAGGA	3	10370351-10370510
SNP rs26310	*VHL*	CCTCTGTCCTTTCACCTTGC	GAGCTCAGAGCAGCCACAGT	3	10372722-10372882
SNP rs775018	*VHL*	GACCCTGGAATCCATGTTTG	AAGGACACAGAACCCACAGG	3	10399978-10400176
SNP rs28113	*VHL*	CATGGGCTCAATCACCTTCT	GGTGGGGAGGTGGTCTCTA	3	10400519-10400684
SNP rs3774166	*VHL*	GTGGGGGTGGAGAGCTATG	GGGAGAGGAGTGAGGACTTTG	3	10418797-10418956
SNP rs279543	*VHL*	GTTTCCTGACCCGTGACCTA	GGTTGGCTAGAACTGGGTGA	3	9935241-9935448
SNP rs279588	*VHL*	AGGAAATGCGACTCACCAGT	CAAGGTCCTTTGTGCTGAGG	3	9944893-9945112
SNP rs279578	*VHL*	CTTGTCACCGGGAAGAACAG	GGATCCTCTCATTGGGGGTA	3	9949140-9949247
SNP rs172155	*VHL*	AGACCTAGAGCCACGAGCAA	GAACTATCCCCAGAAAGAAGGAA	3	9950497-9950671

SNP rs279572	*VHL*	TCACTGACTACAGCCAGCACA	CCTCACCCCATCTGACTCTC	3	9952579-9952730
SNP rs455863	*VHL*	CAGCTCCATGCTTCACTCAG	CCCAGCCTCCCTCTTACCT	3	9956156-9956322
SNP rs708567	*CDKN2A*	GACTTGGGCCTGGAGAGAG	CTGGGAAGAGCCTGAAGATG	3	9960023-9960203
SNP rs2197008	*CDKN2A*	AACCAGCCACTCCCAGAAC	CCACAACCTTTCCTCATTGC	9	21685360-21685570
SNP rs7023954	*CDKN2A*	ATTCCATGAGTCCTGTTGTGG	TCCAAAAAGCTTAAAATACCATACCT	9	21816622-21816795
SNP rs4478653	*CDKN2A*	TGGTTCAGGTGGCTGACATA	TCCCGCTGGTTCAATAATGT	9	21853072-21853267
SNP rs7039105	*CDKN2A*	AAAGTGCCTTTCTACAAGGTTTTG	CCTGAAAATAAGGCCCAAGTC	9	21854962-21855142
SNP rs7047648	*CDKN2A*	GGGTTTATTTTGGGAAGGTCA	ATGCAGACATGAAGCATCCA	9	21857124-21857290
SNP rs2811708	*CDKN2A*	TGTAGGATGCTTAAACCCATTTG	TGGATAGTTTTGACAATTTTTAATGG	9	21973259-21973451
SNP rs3731211	*CDKN2A*	GCTCCTCCTAATTGTTTTTGAAGA	AACAATGACTCCCCCAGTTG	9	21986689-21986890
SNP rs7036656	*CDKN2A*	TGGACATAGCGTGCATTTCA	GCCTTTGACACACGGTAACA	9	21990289-21990477
SNP rs2811710	*CDKN2A*	GGGCTATGGTTCACTTGGAA	TGTTTATAGCTACTTCAGAAGGCTCA	9	21991800-21991950
SNP rs2106119	*CDKN2A*	ACCCAACTGAATTCGGACTCT	TGGGGTGTCTGCTGTAACTG	9	22017415-22017598
SNP rs10965215	*CDKN2A*	CAATAGGTGTGGGCCTCAGT	TTTTTGGATGTTTTGCAGGACT	9	22029414-22029591
SNP rs10757270	*CDKN2A*	TCCCCAAAGACACAAAGTTAAGA	AAGCAAAATGCTAGGAGCACA	9	22072698-22072849
SNP rs1333040	*CDKN2A*	GCCATGGGAATCTTCATTCA	TCAAGAGAGACAGGAGGGTCA	9	22083371-22083541
SNP rs2383206	*CDKN2A*	TTCAAATTTATGCTGCATTACTGAC	TTCAGGATTCAGGCCATCTT	9	22114870-22115051
SNP rs1537375	*CDKN2A*	ACTTAGCCCTTGGGACCATT	CCAATGTCATTTGATCCAATTTT	9	22115915-22116104
SNP rs10811658	*CDKN2A*	ACAGGGAGGGGCATTAGTTT	TCCTTTCCATGGCATCTCTC	9	22128455-22128638
SNP rs10757284	*CDKN2A*	GGAGGGGAAAAGCCACTTAG	CCTTCACAATGCATGGAGTACTTA	9	22138335-22138484
SNP rs1333054	*CDKN2A*	TGTCTTCTGCTCACGATTCCT	TTTTTAGCTTTTGAGCACTCTGC	9	22165442-22165613
SNP rs954399	*CDKN2A*	CACAAATGACTTTCTCCTGCTG	TCTAGGAAAAGGACAATTCATTCAG	9	22170948-22171107

Chr = Chromosome. Hg19 = Human genome 19

Table 2.7 Comparison of preoperative assessment with final pathological diagnosis in the 130 surgically resected pancreatic cystic neoplasms

Patient ID	Presence of concerning features prior to surgery*	Preoperative diagnosis[#]	Final pathology	Grade of Dysplasia/Invasive Cancer in IPMNs, and MCNs
Cyst-1	No	Branch duct IPMN	IPMN	Low
Cyst-2	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-3	No^	Cyst type unclear	MCN	Low
Cyst-4	Yes	MCN	MCN	Low
Cyst-5	Yes	Mixed duct IPMN	IPMN	Invasive cancer
Cyst-6	Yes	MCN	MCN	Low
Cyst-7	Yes	MCN	SCA	Not applicable
Cyst-8	Yes	Branch duct IPMN	SCA	Not applicable
Cyst-9	Yes	MCN	MCN	Low
Cyst-10	Yes	MCN	SCA	Not applicable
Cyst-11	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-12	Yes	Branch duct IPMN	IPMN	Low
Cyst-13	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-14	No	Branch duct IPMN	IPMN	Low
Cyst-15	Yes	Branch duct IPMN	IPMN	High
Cyst-16	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-17	Yes	Mixed IPMN	IPMN	High
Cyst-18	Yes	Mixed IPMN	IPMN	High
Cyst-19	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-20	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-21	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-22	No	SCA	SCA	Not applicable
Cyst-23	Yes	MCN	SCA	Not applicable
Cyst-24	Yes	Branch duct IPMN	SCA	Not applicable
Cyst-25	Yes	SCA	SCA	Not applicable
Cyst-26	Yes	Branch duct IPMN	SCA	Not applicable
Cyst-27	Yes	SCA	SCA	Not applicable
Cyst-28	No	Branch duct IPMN	SCA	Not applicable
Cyst-29	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-30	No^	Cyst type unclear	MCN	Intermediate
Cyst-31	No	Branch duct IPMN	IPMN	Intermediate
Cyst-32	No	MCN	MCN	Low
Cyst-33	No	Branch duct IPMN	MCN	Low
Cyst-34	No	Branch duct IPMN	IPMN	Intermediate
Cyst-35	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-36	Yes	Branch duct IPMN	IPMN	Low
Cyst-37	Yes	Mixed IPMN	IPMN	Intermediate
Cyst-38	Yes	Mixed IPMN	IPMN	High
Cyst-39	No	Mixed IPMN	IPMN	Intermediate
Cyst-40	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-41	Yes	Mixed IPMN	IPMN	Intermediate
Cyst-42	Yes	MCN	MCN	Low
Cyst-43	Yes	MCN	MCN	Low
Cyst-44	Yes	Branch duct IPMN	IPMN	Intermediate

Cyst-45	No	Mixed IPMN	IPMN	High
Cyst-46	Yes^	Main duct IPMN	IPMN	Invasive cancer
Cyst-47	Yes	Branch duct IPMN	IPMN	High
Cyst-48	No	SCA	IPMN	Low
Cyst-49	Yes	Mixed IPMN	IPMN	High
Cyst-50	No^	Mixed IPMN	IPMN	Intermediate
Cyst-51	No	Branch duct IPMN	IPMN	Intermediate
Cyst-52	No	Branch duct IPMN	IPMN	Low
Cyst-53	Yes	SCA	SCA	Not applicable
Cyst-54	Yes	Cyst type unclear	SCA	Not applicable
Cyst-55	Yes	Branch duct IPMN	IPMN	High
Cyst-56	No	Branch duct IPMN	IPMN	Intermediate
Cyst-57	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-58	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-59	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-60	Yes	SPN	SPN	Not applicable
Cyst-61	Yes	Branch duct IPMN	IPMN	Low
Cyst-62	No	Branch duct IPMN	IPMN	Low
Cyst-63	Yes	Branch duct IPMN	ITPN	High
Cyst-64	No	Branch duct IPMN	IPMN	Low
Cyst-65	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-66	Yes	SCA	MCN	Low
Cyst-67	Yes	SPN	SPN	Not applicable
Cyst-68	Yes	SPN	SPN	Not applicable
Cyst-69	Yes	SPN	SPN	Not applicable
Cyst-70	Yes	SPN	SPN	Not applicable
Cyst-71	Yes	SPN	SPN	Not applicable
Cyst-72	No^	SPN	SPN	Not applicable
Cyst-73	No^	SPN	SPN	Not applicable
Cyst-74	Yes	SPN	SPN	Not applicable
Cyst-75	Yes	SPN	SPN	Not applicable
Cyst-76	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-77	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-78	No	Branch duct IPMN	IPMN	Intermediate
Cyst-79	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-80	Yes	MCN	MCN	Intermediate
Cyst-81	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-82	Yes	MCN	MCN	Low
Cyst-83	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-84	Yes	SCA	IPMN	Intermediate
Cyst-85	No	Main duct IPMN	IPMN	Low
Cyst-86	Yes	Branch duct IPMN	IPMN	Low
Cyst-87	Yes	Branch duct IPMN	IPMN	Low
Cyst-88	Yes	Mixed IPMN	IPMN	High
Cyst-89	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-90	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-91	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-92	Yes	Branch duct IPMN	IPMN	Invasive cancer
Cyst-93	No	Branch duct IPMN	IPMN	High
Cyst-94	Yes	Branch duct IPMN	IPMN	Invasive cancer

Cyst-95	Yes	MCN	IPMN	Intermediate
Cyst-96	Yes	Branch duct IPMN	IPMN	High
Cyst-97	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-98	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-99	Yes	Branch duct IPMN	IPMN	High
Cyst-100	Yes	Mixed IPMN	IPMN	High
Cyst-101	No	Branch duct IPMN	IPMN	Intermediate
Cyst-102	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-103	No	SCA	IPMN	Intermediate
Cyst-104	Yes	MCN	IPMN	Low
Cyst-105	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-106	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-107	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-108	Yes	Mixed IPMN	IPMN	Intermediate
Cyst-109	Yes	Mixed IPMN	IPMN	High
Cyst-110	Yes	Mixed IPMN	IPMN	High
Cyst-111	Yes	Branch duct IPMN	IPMN	High
Cyst-112	Yes	Mixed IPMN	IPMN	High
Cyst-113	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-114	Yes	Branch duct IPMN	IPMN	High
Cyst-115	No^	Cyst type unclear	IPMN	Intermediate
Cyst-116	Yes	Mixed IPMN	IPMN	Low
Cyst-117	Yes	Branch duct IPMN	IPMN	High
Cyst-118	Yes	Mixed IPMN	IPMN	High
Cyst-119	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-120	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-121	Yes	Mixed duct-IPMN	IPMN	Low
Cyst-122	No	Branch duct IPMN	IPMN	Intermediate
Cyst-123	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-124	Yes	Branch duct IPMN	IPMN	High
Cyst-125	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-126	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-127	Yes	Branch duct IPMN	IPMN	Intermediate
Cyst-128	Yes	Mixed IPMN	IPMN	Invasive cancer
Cyst-129	Yes	Mixed IPMN	IPMN	Intermediate
Cyst-130	Yes	Mixed IPMN	IPMN	Invasive cancer

*Concerning features were the presence of any of the following: jaundice, mural nodule or solid component, dilation of the main pancreatic duct ≥5 mm in size, cyst size ≥3 cm, or cytology with marked atypia or diagnosic of malignancy. ^No preoperative imaging available for review. #Preoperative diagnosis was based on physician's assessment based upon clinical data, imaging, and cyst fluid analysis when available.

Table 2.8 Frequencies of aneuploid chromosomes by cyst type

Chromosome arm	IPMN N=96*		MCN N=12		SCA N=12		SPN N=10	
	gain, N (%)	loss, N (%)	gain, N (%)	loss, N (%)	gain, N (%)	loss, N (%)	gain, N (%)	loss, N (%)
1p	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1q	13 (14)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2p	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2q	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
3p	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	1 (8)	0 (0)	0 (0)
3q	3 (3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
4p	0 (0)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
4q	0 (0)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
5p	3 (3)	1 (1)	0 (0)	1 (8)	0 (0)	0 (0)	0 (0)	0 (0)
5q	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
6p	2 (2)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
6q	1 (1)	5 (5)	0 (0)	0 (0)	0 (0)	1 (8)	0 (0)	0 (0)
7p	6 (6)	0 (0)	0 (0)	0 (0)	0 (0)	0(0)	0 (0)	0 (0)
7q	6 (6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
8p	2 (2)	7 (7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
8q	5 (5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
9p	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
9q	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
10p	0 (0)	3 (3)	0 (0)	0 (0)	0 (0)	1 (8)	0 (0)	0 (0)
10q	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
11p	3 (3)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (20)
11q	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
12p	2 (2)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
12q	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
13q	2 (2)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
14q	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

15q	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
16p	3 (3)	4 (4)	1 (8)	0 (0)	3 (25)	0 (0)	5 (5)	0 (0)
16q	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
17p	1 (1)	7 (7)	0 (0)	0 (0)	1 (8)	1 (8)	0 (0)	0 (0)
17q	1 (1)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
18p	6 (6)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
18q	2 (2)	7 (7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
19p	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
19q	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
20p	4 (4)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
20q	3 (3)	3 (3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
21q	0 (0)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
22q	1 (1)	17 (18)	0 (0)	0 (0)	0 (0)	2 (17)	0 (0)	0 (0)

Values are presented as absolute number of chromosome arms altered and percentages in brackets
* Includes one ITPN

Table 2.9 Frequency of distribution of molecular markers and aneuploidy in branch duct, mixed or main duct IPMNs

IPMN	**Branch Duct**	**Mixed or Main Duct**
	N = 56	**N = 40**
KRAS - no. (%)	48 (86)	27 (68)
GNAS - no. (%)	30 (54)	26 (65)
RNF43 - no. (%)	23 (41)	13 (33)
CDKN2A - no. (%)	1 (2)	2 (5)
CTNNB1 - no. (%)	2 (4)	4 (10)
SMAD4 - no. (%)	2 (4)	3 (8)
TP53 - no. (%)	4 (7)	5 (13)
VHL - no. (%)	0 (0)	0 (0)
BRAF - no. (%)	1 (2)	0 (0)
NRAS - no. (%)	0 (0)	0 (0)
PIK3CA - no. (%)	0 (0)	0 (0)
LOH chr3 (VHL) - no. (%)	2 (3.6)	2 (5)
LOH Chr9 (CDKN2A) - no. (%)	3 (5)	5 (13)
LOH chr17 (RNF43) - no. (%)	3 (5)	8 (20)
LOH chr17 (TP53) - no. (%)	1 (2)	4 (10)
LOH chr18 (SMAD4) - no. (%)	3 (5)	7 (18)
Aneuploidy - no. (%)	26 (46)	22 (55)

Table 2.10 Composite clinical markers: Comparison of results from original cross-validation study (Masica et al) to those in the current study

Cyst type	Any of these *Present*	All of these *Absent*	Masica et al[14] (N=1,026)		Current study (N=130)		
			Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity	Specificity	*p*-value
SCA	Age ≥25	Abdominal pain Communication with MPD MPD dilation	95% (92-97%)	83% (79-86%)	100%	71%	1.3×10^{-5}
SPN	Age <55	Jaundice Multifocal Weight loss	89% (75-96%)	86% (83-88%)	89%	85%	6.8×10^{-6}
MCN	Age <75	Male Communication with MPD Multifocal	91% (82-96%)	83% (80-85%)	90%	78%	3.0×10^{-5}
IPMN*	Age ≥85 Communication with MPD MPD dilation Abdominal pain	None	94% (92-96%)	90% (87-93%)	75%	88%	2.2×10^{-8}

* Analysis was performed in the ninety-five IPMN and one ITPN patients.

Table 2.11 Composite Molecular Marker for High-Grade Dysplasia or Associated Invasive Adenocarcinoma

Composite Molecular Markers		Sensitivity (95% CI)	Specificity (95% CI)
Any of these *Present*	Any of these *Absent*		
SMAD4 chr17 LOH (*RNF43*) chr5p aneu chr8p aneu chr13q aneu chr18q aneu *TP53*^ chr17 LOH (*TP53*)^	None	59% (41-75%)	90% (80-96%)

^ These features were not identified by MOCA but were manually added.
chr = chromosome. CI=confidence intervals. aneu = aneuploidy.

Chapter 3:

Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas

INTRODUCTION AND RATIONALE

Head and neck squamous cell carcinomas (HNSCCs) are the seventh most common cancer worldwide, occurring in more than half a million new patients each year and in >50,000 patients in the United States alone (Ferlay et al., 2010; Siegel et al., 2015). The incidence of certain types of HNSCC appears to be increasing, especially among young people, in part due to the increasing prevalence of human papilloma virus (HPV) (Simard et al., 2012; Patel et al., 2011; Li et al., 2014; Chaturvedi et al., 2011; Chaturvedi et al., 2008). HNSCCs are associated with a poor 5-year overall survival of only ~50% that has remained relatively unchanged, especially for patients with HPV-negative tumors (Leemans et al., 2011). Only a few targeted therapies for this disease are available, in part because of the paucity of activating mutations in oncogenes that contribute to tumor development; most genetic alterations in HNSCCs inactivate tumor suppressor genes (Agrawal et al., 2011; Vogelstein et al., 2013; Stransky et al., 2011; Cancer Genome Atlas, 2015). There are also no available biomarkers for HNSCC to measure disease burden or response to therapy, further limiting progress in mitigating the impact of this often morbid and potentially lethal disease on human health.

Although HNSCC tumors are usually classified on the basis of histology, their biomedical properties, including demographics, risks factors, and clinical behavior, differ by anatomic site (Figure 3.1) (Howlader et al., 2014; Chaturvedi et al., 2013). Anatomically, the tumors are categorized as squamous cell carcinomas (SCCs) of the oral cavity (including the oral tongue), oropharynx (including the base of the tongue), larynx, and hypopharynx. Oral cavity SCC, with the exception of those of the oral tongue, is declining in incidence in the United States because of the reduction in cigarette smoking (Patel et al., 2011). In contrast, there is an increasing incidence of oropharyngeal SCC involving the palatine and lingual (base of the tongue) tonsils, particularly in younger men. These tumors are often associated with HPV. The survival of these patients is better than for those whose tumors are un-associated with HPV (Chaturvedi et al., 2011; D'Sou za et al.,

2007). Laryngeal SCC is declining in incidence and, unlike HNSCC at other sites, is generally associated with limited regional metastasis due to anatomic barriers (Pfister et al., 2014). The hypopharynx is the least common site for HNSCC and has decreasing incidence but relatively poor prognosis (Gourin et al., 2004; Kuo et al., 2014).

The idea that the genetic alterations present in tumors can be used as biomarkers for cancer was proposed more than two decades ago (Sidransky, 1997; Sidransky et al., 1991; Sidransky et al., 1992; Boyel et al., 1994). The advantage of genetic alterations over conventional bio-markers such as carcinoembryonic antigen or prostate-specific antigen is that genetic changes are exquisitely specific for neoplastic cells. One challenge in exploiting genetic alterations for this purpose is that the concentration of mutant templates is often low in bodily fluids. Over the last several years, however, technological advances have made it possible to detect such mutations even when they are rare. These advances have facilitated the detection of altered DNA sequences in plasma, stool, Pap smear fluids, sputum, and urine (Sidransky et al., 1991; Sidransky et al., 1992; Bettegowda et al., 2014; Diehl et al., 2008; Kinde et al., 2013; Dawson et al., 2013; Newman et al., 2014; Martignetti et al., 2014; Ralla et al., 2014; Hubers et al., 2013).

In this proof-of-principle study, we determined whether genetically altered DNA could be detected in the saliva or plasma of HNSCC patients with tumors of various stages and anatomical sites. We chose these two bodily fluids for obvious reasons: plasma has been shown to harbor tumor DNA from many cancers, including HNSCC, though only a few HNSCCs, all of late stage, have been previously examined (Bettegowda et al., 2014; Kang et al., 2015; Ahn et al., 2014). Tumor DNA that is released from the basal side of HNSCC epithelial cells into the lymphatics or venous system should be detectable in this compartment. On the other hand, DNA that is released primarily on the apical side of HNSCC should be detectable in the saliva (Bettegowda et al., 2014; Kang et al., 2015; Ahn et al., 2014). The studies described below were performed to test these hypotheses.

MATERIALS AND METHODS

STUDY DESIGN

This was a retrospective study with sample collection performed prospectively from 93 HNSCC patients donating saliva, 47 of whom are also donating plasma. Data analysis was performed in a blinded fashion, and all patient samples were de-identified.

SAMPLES

All samples from the 93 patients in this study were collected using Institutional Review Board–approved protocols at Johns Hopkins University (JHU) and MD Anderson Medical Center. None of the patients in the current study were included in the previously published study from our groups, in which the genomic landscapes of HNSCC were described (Agrawal et al., 2011).

Saliva samples were collected before definitive treatment for primary HNSCC (n = 71, 76% of 93 patients) and before salvage treatment for recurrent HNSCC (n = 22, 24% of 93 patients). In a subset of these patients (n = 9), post treatment saliva was also collected for surveillance. Most patients (95% of the 93) underwent a biopsy of the primary tumor and/or metastatic lymph node, on average 44 days before the first sample collection (Table 3.4). For the 22 patients with recurrent disease, previous treatment including an iteration of surgery, radiation, and/or chemotherapy was completed an average of 2.9 years before sample collection (Table 3.4).

Whole blood was collected from 47 of the 93 patients before treatment. Four to 10 ml of plasma was used for DNA purification, with the average amount of plasma being 6 ml.

Saliva was collected using two different protocols. Under the JHU protocol, patients were asked to swish 15 to 20 ml of 0.9% sodium chloride in their mouths for 10 to 15 s before spitting into the collection tube. Under the MD Anderson protocol, patients were asked to allow saliva to collect in the floor of the mouth for 5 min without swallowing before spitting into the collection

vial. There was no significant difference in the amounts of DNA purified, the fraction of mutant DNA, or the amount of HPV sequences found with the two protocols. Saliva was frozen at −80°C until DNA purification, and the entire volume of saliva, without centrifugation of cells, was used for DNA purification. The amount of saliva used averaged 15 ml (range, 10 to 20 ml).

When fresh tumor tissue from a surgical specimen of invasive SCC was available, it was immediately frozen at −80°C. When frozen tissue was not available, formalin-fixed, paraffin-embedded (FFPE) tissues were used for DNA purification. In either case (freshfrozen or FFPE), tumors were macrodissected to ensure neoplastic cellularity exceeding 30%. DNA was purified from the white blood cell pellet (normal DNA), saliva, and tumor using an AllPrep Kit (Qiagen, catalog #80204), and from plasma using an QIAamp Circulating Nucleic Acid Kit (Qiagen, catalog #55114).

TUMOR MUTATION PROFILING

A tiered approach was used to identify a somatic mutation within each tumor. Initially, the presence of HPV16 and HPV18 was assessed using the primers specific for the E7 oncogene of these variants (HPV16: TGTGACTCTACGCTTCGGTTG and GCCCATTAACAGGTCTTCCA; HPV18: GCATGGACCTAAGGCAACAT and GAAGGTCAACCGGAATTTCAT). When no HPV was present, multiplex PCRs containing primers amplifying regions of interest in *TP53*, *PIK3CA*, *CDKN2A*, *FBXW7*, *HRAS*, and *NRAS* were used to identify driver mutations in the tumors (Table 3.2). If a mutation was not identified in the queried regions, paired-end libraries of DNA from the tumors and white blood cell pellets of each patient were prepared and used for low-coverage whole-genome sequencing or exomic sequencing as previously described (Agrawal et al., 2012). Massively parallel sequencing was carried out on an Illumina HiSeq instrument, either in the Goldman Sequencing Facility at Johns Hopkins Medical Institutions or at PGDx Inc. Point mutations were identified as previously described (Agrawal et al., 2011; Agrawal et al., 2012;

Bettegowda et al., 2011), using the following criteria: allele fraction >20%, >5 reads for point mutations, and >1 read for translocations. Genomic rearrangements were identified through an analysis of discordantly mapping paired-end reads. The discordantly mapping paired-end reads were grouped into 1-kb bins when at least five distinct tag pairs (with distinct start sites) spanned the same two 1-kb bins, and further annotated on the basis of the approximate breakpoint (Sausen et al., 2013). One selected mutation was confirmed through an independent PCR and sequencing reaction, and then used to query the saliva or plasma. All oligonucleotides used in this study were synthesized by TriLink Biotechnologies.

MUTATION DETECTION IN SALIVA AND PLASMA

The same primers used to detect HPV16 in tumor DNA via PCR were used to detect HPV16 sequences in the DNA from saliva or plasma. Each saliva DNA or plasma DNA sample was assessed in at least three independent PCR assays, and all three assays had to be positive for the sample to be counted as positive. As an additional control for specificity, the PCR products were sequenced to ensure that they represented HPV16 sequences. To quantify the amount of HPV16 sequences present in saliva or plasma, we used digital PCR with the same primers (Vogelstein and Kinzler, 1999). Digital PCR was also used to quantify the amount of sequences with translocation using primers spanning the breakpoints, as previously described (Leary et al., 2010). For evaluation of point mutations in saliva or plasma, we used Safe-SeqS, a PCR-based error reduction technology for detection of low-frequency mutations in reactions each containing up to 3 ng of input DNA (Bettegowda et al., 2014; Kinde et al., 2013). High-quality sequence reads were selected on the basis of quality scores, which were generated by the sequencing instrument to indicate the probability a base was called in error (Ewing et al., 1998). The template-specific portion of the reads was matched to reference sequences. Reads from a common template molecule were then grouped on the basis of the unique identifier sequences (UIDs) that were incorporated as molecular barcodes. Artifactual mutations introduced during the sample preparation or sequencing

steps were reduced by requiring a mutation to be present in >90% of reads in each UID family ("supermutant"). Each PCR assay for each plasma or saliva sample was independently repeated at least three times, with the mutant allele fractions defined as the total number of supermutants divided by the total number of UIDs in all experiments. DNA from normal individuals was used as control, using at least five independent assays per queried mutation. Only saliva or plasma samples in which the mutant allele fractions significantly exceeded their frequencies in control DNA ($P < 0.05$) were scored as positive (Table 3.6 and Table 3.7).

STATISTICAL ANALYSIS

Sensitivity for the detection of tumor-specific mutations in the blood and saliva was calculated by tumor site, stage, and among HPV-associated tumors. Ability to detect tumor DNA in saliva and/or plasma was tested using Fisher's exact tests, and Wilcoxon rank sum tests were used to compare amounts of tumor DNA in saliva versus plasma (Agresti, 2002). For the comparison of mutant fractions in patients versus control in Safe-SeqS assays, *P* values were calculated using a two-sided χ^2 test of equal proportions or Fisher's exact test when conditions of the χ^2 test are not met. The concordance between mutant fractions in saliva and plasma was calculated using Pearson's product-moment correlation coefficient, a standard measure of linear dependence between two variables. All statistical analyses were performed using the R statistical package version 3.1.2.

RESULTS

MUTATIONS IN PRIMARY TUMORS

Ninety-three patients with HNSCC were enrolled in this study. Their average age was 60 and the majority (83%) were male, as is typical of HNSCC patients (Table 3.3). Forty-six, 34, 10, and 3 samples were from the oral cavity, oropharynx, larynx, and hypopharynx, respectively.

Twenty patients (22%) had early (stage I or II) disease, and the remaining 73 patients (78%) had advanced (stage III or IV) disease.

To begin this study, we attempted to identify at least one genetic alteration in each tumor. We first searched for the presence of either HPV type 16 (HPV16) or HPV18 sequences in tumor DNA. HPV is a well-established etiologic agent for a growing subset of HNSCCs, specifically oropharyngeal SCC (Chaturvedi et al., 2011; D'Sou za et al., 2007). With polymerase chain reaction (PCR) primer pairs specific for the E7 gene of the high-risk HPV types responsible for the overwhelming majority of HPV-associated HNSCCs, we identified 30 patients (32%) whose tumors contained HPV16 DNA and no patients with HPV18. The preponderance of HPV16 is not surprising given prior epidemiologic studies of this tumor type (Kreimer et al., 2005). In the other 63 patients (all of those without HPV), we searched for somatic mutations in genes or gene regions commonly altered in HNSCC, including *TP53*, *PIK3CA*, *CDKN2A*, *FBXW7*, *HRAS*, and *NRAS*, using multiplex PCR and massively parallel sequencing (Table 3.2) (Agrawal et al., 2011; Stransky et al., 2011; Cancer Genome Atlas, 2015). This allowed us to identify a driver mutation in 58 of the 63 samples. In the remaining five samples, genome-wide sequencing was performed at low coverage with the goal of identifying one driver mutation or translocation as previously described (Sausen et al., 2013). Ultimately, we identified and validated one genetic alteration in each of these 63 samples (Table 3.5). The most commonly mutated gene was *TP53* (86% of 63 patients). We also searched for mutations in the tumors of 25 of the patients with HPV and found mutations in 12 of those samples (Table 3.5).

MUTATIONS IN SALIVA AND PLASMA

Important characteristics of screening tests are that samples can be easily collected without discomfort and that the collection process is standardized. To achieve these goals, we used oral rinses, plasma, and commercially available kits to prepare DNA for conventional genotyping purposes. For saliva, we used the entire contents of the collection tube (including cells and cell

debris) to prepare DNA. Of the 93 patients who donated saliva for this study before their surgery, 47 patients (51%) volunteered to donate plasma at the same time. DNA from plasma was purified as previously described (Bettegowda et al., 2014). Digital PCR was used to query HPV sequences and translocations (Kinde et al., 2012), whereas point mutations were assessed by Safe-SeqS (Safe-Sequencing System), a PCR-based technology for the detection of low-frequency mutations, as previously described (Bettegowda et al., 2014; Kinde et al., 2013; Sausen et al., 2013; Kinde et al., 2012; Vogelstein and Kinzler 1999).

Tumor DNA was identified in 76% (n = 93) and 87% (n = 47) of the saliva and plasma samples from these patients, respectively (Table 3.1, Table 3.6 and Table 3.7). In the subset of patients with both plasma and saliva samples, 96% (n = 47) of patients had a tumor-specific alteration identified in at least one bodily fluid. Twenty-one of the 47 patients had HPV-positive tumors. Eighteen of the 21 patients (86%) had detectable HPV DNA in their plasma and/or saliva (Table 3.1, Table 3.6 and Table 3.7). Because HPV16 is rarely found in oral specimens of healthy individuals, we analyzed 10 saliva or plasma samples from patients whose tumors were not HPV-positive as controls (Kreimer et al., 2010; Gillson et al., 2009-2010). As expected, no HPV was detected in any of these samples, confirming the specificity of the test. In all 26 patients without HPV-positive tumors, endogenous DNA mutations, mostly in the *TP53* gene (92%), were identified in plasma or saliva (Table 3.5, Table 3.6 and Table 3.7). Thus, somatic mutations and HPV sequences were both useful as biomarkers for malignancy. The sensitivity of these biomarkers for detecting cancer was greatly improved when both plasma and saliva were examined, compared to testing saliva or plasma alone. There was no significant correlation between the amounts of tumor DNA in saliva versus plasma in the patients in whom both sample types were available (correlation coefficient of 0.074, P = 0.74), However, the use of versus plasma, and HPV versus somatic mutations, differed considerably with respect to the site of disease, as noted below.

SITE

All (100%) of the 46 patients with oral cavity cancers harbored detectable tumor DNA in their saliva (Table 3.1). The sensitivities of detection in saliva of malignancy at sites not directly sampled by an oral rinse were lower: 47% (n = 34), 70% (n = 10), and 67% (n = 3) of patients with oropharyngeal cancers, laryngeal cancers, and hypopharyngeal cancers had detectable tumor DNA, respectively. The detection rate of tumor DNA in plasma varied less with site, as expected: 80% (n = 15), 91% (n = 22), 86% (n = 7), and 100% (n = 3) of tumors of the oral cavity, oropharynx, larynx, and hypopharynx, respectively, had detectable tumor DNA in plasma.

It is well known that HPV-associated tumors are most often found at specific sites, particularly the oropharynx. Twenty-nine of the 34 (85%) oropharyngeal cancers were HPV-positive. The remaining five oropharyngeal cancers were negative for HPV by PCR, were associated with tobacco use, and harbored *TP53* mutations. In contrast, all but 1 of 59 samples from the oral cavity, larynx, and hypopharynx were HPV-negative. The finding that only 1 of the 46 oral cavity cancers tested was HPV-positive is consistent with recent evidence about the low prevalence of HPV-related cancers in the oral cavity (Isayeva et al., 2012; Lingen et al., 2013). For the HPV-associated cancers, which represent 30 (32%) of the total HNSCCs in our study, the presence of HPV DNA in bodily fluids represents a very convenient marker: HPV was detected in 40% (n = 30) of saliva samples and 86% (n = 21) of available plasma samples with a single primer pair specific for the E7 gene of HPV16 (Table 3.1).

Collectively, these data indicate that plasma rather than saliva is the optimal fluid for detecting tumor DNA in tumors of the oropharynx, larynx, and hypopharynx. Of the 32 patients with tumors from these sites in which both plasma and saliva were available, mutant DNA was detected in more plasma samples than saliva samples (29 versus 18, respectively). The amount of detectable mutant DNA alleles, expressed as a fraction of the total alleles assessed, was ~10-fold higher in the plasma compared with the saliva of these patients (median, 0.146% versus 0.015%; P

= 0.005, Wilcoxon rank sum test; Table 4.6 and Table 4.7). The higher fraction of alleles considerably simplifies the task of identifying such mutations. This pattern was not observed in the oral cavity: the fraction of patients harboring mutant DNA, as well as the mutant allele fraction, was similar in the saliva and plasma (median, 0.65% versus 0.54%; $P = 0.14$, Wilcoxon rank sum test; Table 3.6 and Table 3.7).

STAGE

Most HNSCC patients have advanced disease (stage III or IV) at diagnosis (2). Accordingly, only 22% of the 93 patients in our cohort presented with early-stage disease (Table 3.3). Overall, tumor-specific DNA could be detected in the plasma or saliva of 100% ($n = 20$) and 86% ($n = 73$) of patients with early and advanced disease, respectively ($P = 0.116$, Fisher's exact test). Saliva provided a more sensitive predictor of early-stage disease than plasma: 100% (15 of 15 oral cavity cancers, 3 of 3 oropharyngeal cancers, and 2 of 2 laryngeal cancers) versus 70% (5 of 7 oral cavity cancers, 2 of 2 oropharyngeal cancers, and 0 of 1 laryngeal cancers), respectively ($P = 0.03$, Fisher's exact test; Table 3.1). Contributing to the high sensitivity in saliva was the fact that 75% (15 of 20) of the earlystage cancers in the study were from the oral cavity, which are most readily detectable in saliva and are preferentially treated with surgery, explaining their enrichment in our study. As expected, plasma provided a more sensitive predictor than saliva in patients with advanced-stage disease: 92% ($n = 37$) versus 70% ($n = 73$), respectively ($P = 0.008$, Fisher's exact test; Table 4.1). When segregated by nodal status, tumor-specific DNA could be detected in the plasma or saliva of 83% ($n = 59$) and 100% ($n = 34$) of patients with or without nodal metastasis, respectively. When both saliva and plasma were available, there was little difference between the detectability of cancers with respect to stage of disease (Table 3.1). An important caveat is that only five patients with early-stage disease of non-oral cavity sites were available; although tumor DNA was detectable in all of these patients (all had detectable tumor DNA in saliva; two of the three patients with available plasma also had detectable tumor DNA in their plasma), the amount

of tumor DNA was considerably lower than that of late-stage patients (median, 0.007% versus 0.06%; $P = 0.03$, Wilcoxon rank sum test).

HUMAN PAPILOMA VIRUS

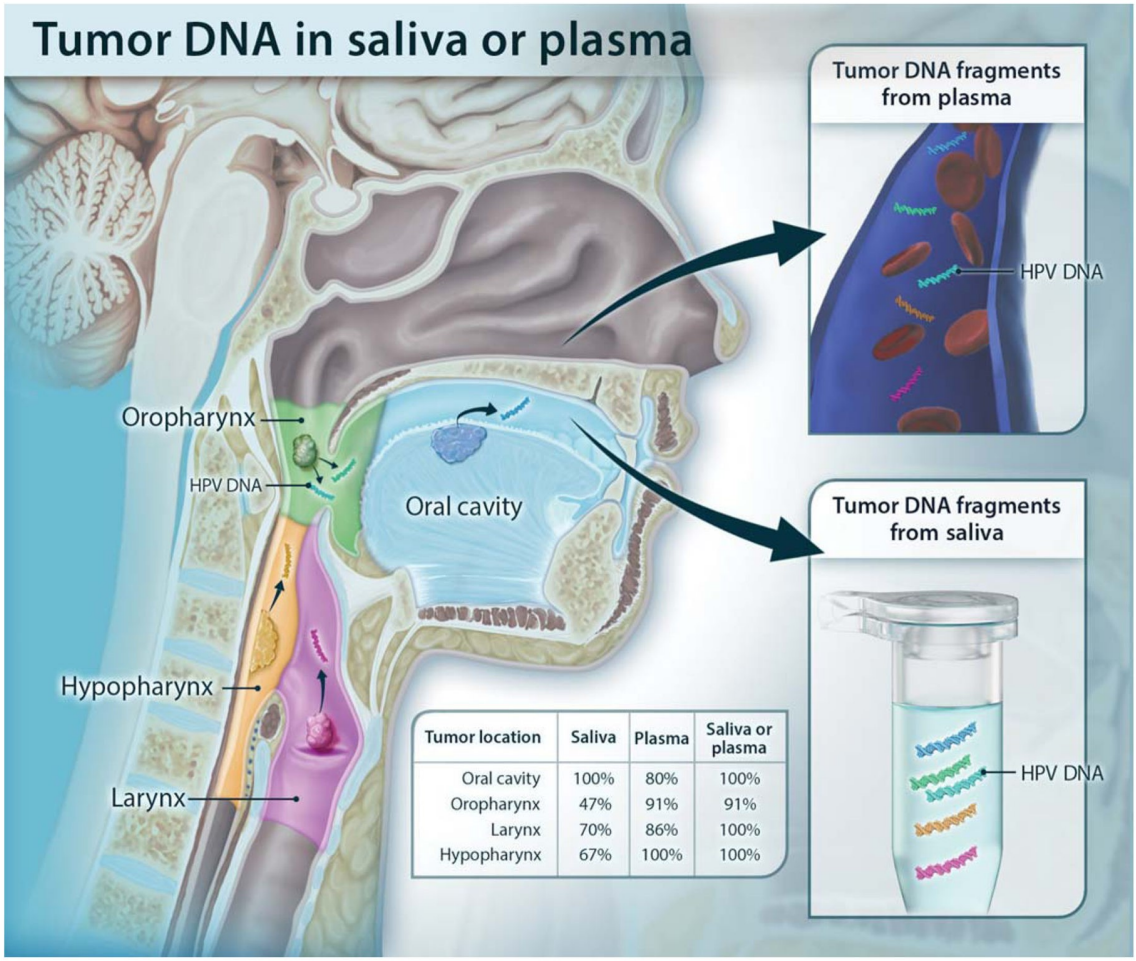
Thirty patients harbored HPV16 DNA in their tumors when assessed by PCR, and none had HPV18. Of these 30 tumors, 29 (97%) were thought to be HPV-associated upon clinical presentation on the basis of in situ hybridization with high-risk HPV sequences or immunohistochemistry with antibodies to p16; in one case, the HPV status had not been determined in the clinic. Additionally, there were no patients who were considered to have HPV-associated tumors in the clinic and did not have HPV16 DNA identified in their tumors by PCR. This supports the specificity and sensitivity of our assays. As expected from the literature, all except 1 of the 30 tumors containing HPV DNA were found in the oropharynx (D'Sou za et al., 2007; Isayeva et al., 2012). As expected, plasma from HPV-associated tumors was more informative than saliva; HPV DNA was detectable in the plasma of 86% ($n = 21$) of the patients but in only 40% ($n = 30$) of the saliva from these patients (Table 3.1).

SURVEILLANCE

Although not the primary purpose of this study, it was of interest to determine whether tumor DNA could be found in the saliva or plasma of patients after surgical removal of their tumors. "Follow-up" samples were available in nine patients in whom tumor DNA could be identified before therapy. Three of these patients were found to have tumor DNA in their saliva or plasma after surgery, but before clinical evidence of disease recurrence (Figure 3.2). For example, patient HN 399, with cancer of the oral cavity, was found to have tumor DNA in his saliva and plasma 4 months after surgery, whereas the recurrence only became evident clinically 19 months later (23.6 months after surgery). Similarly, tumor DNA was found in the saliva and plasma of patient HN 402, with cancer of the oral cavity, at 8 months after surgery, 9 months before the recurrence was

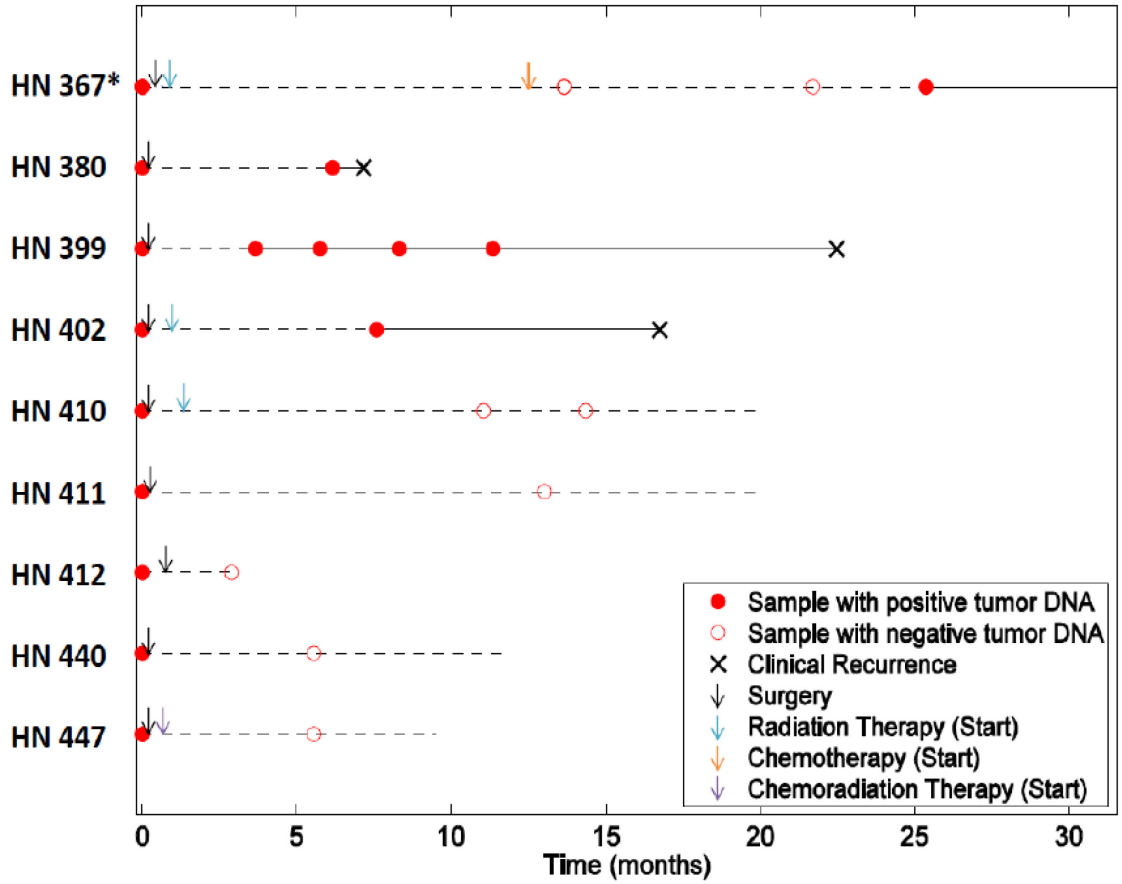
clinically evident. Patient HN 380 with cancer of the larynx was found to have tumor DNA in his saliva 7 months after surgery, before any clinical or radiologic evidence of disease recurrence; the patient died of recurrent disease soon thereafter. Tumor DNA was detectable in the saliva of patient HN 367 with cancer of the oropharynx 25 months after surgery; at the time of writing (36 months after surgery), no biopsy-proven disease is yet evident, but the clinical course has been complicated with suspicious imaging for locoregional and metastatic disease. No tumor DNA was detectable in the saliva and/or plasma of the other five patients in whom samples were available, all of whom have shown no clinical evidence of recurrence for a median follow-up of 12 months (Figure 3.3).

Figure. 3.1. Schematic showing the shedding of tumor DNA from head and neck cancers into the saliva or plasma



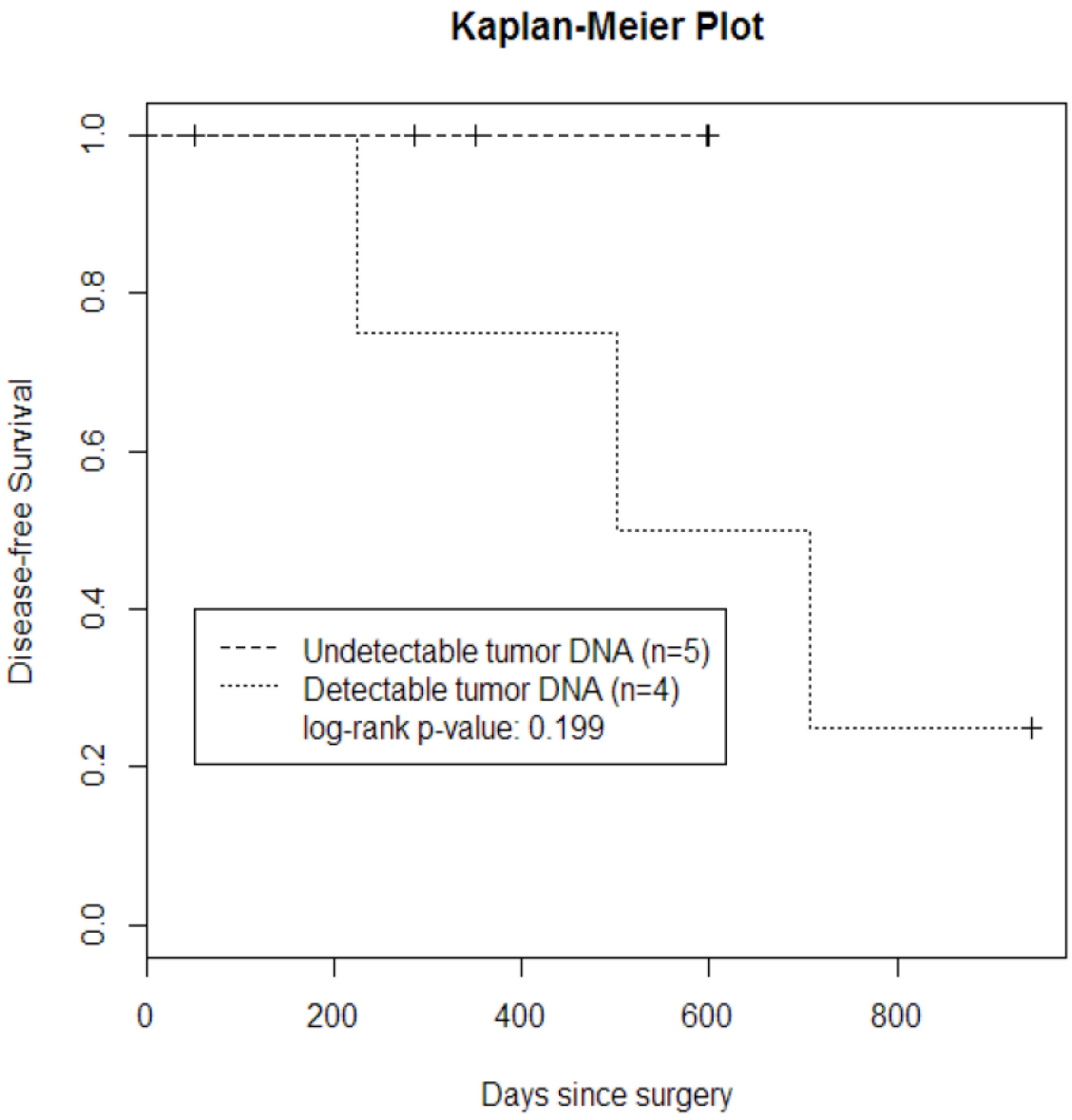
Tumors from various anatomic locations shed DNA fragments containing tumor-specific mutations and HPV DNA into the saliva or the circulation. The detectability of tumor DNA in the saliva varied with anatomic location of the tumor, with the highest sensitivity for oral cavity cancers. The detectability in plasma varied much less in regard to the tumor's anatomic location.

Figure 3.2 Timeline of patients following first positive tumor DNA sample.



Tumor DNA is Detectable in the saliva of patients before recurrence becomes clinically evident. Nine patients were followed-up for a median of 12 months after surgery. Dashed lines transition to solid lines wen tumor DNA was detected after surgery. *Twenty-five months prior to surgery, patient 367 also underwent chemoradiation therapy.

Figure 3.3 Kaplan-Meier plot of patients with detectable and undetectable tumor DNA.



Patients with undetectable tumor DNA after surgery have better disease-free survival.

Table 3.1 Detection of tumor-derived DNA in saliva and plasma

The percentages of patients whose tumors were detectable through the examination of saliva, plasma, or both are shown, grouped by tumor site, stage, and HPV status.

	Saliva, % with mutations (95% confidence intervals) (total number studied)	Plasma, % with mutations (95% confidence intervals) (total number studied)	Saliva or plasma, % with mutations (95% confidence intervals) (total number studied)*
Site			
Oral cavity	100 (92–100%) (46)	80 (52–96%) (15)	100 (78–100%) (15)
Oropharynx	47 (30–65%) (34)	91 (71–99%) (22)	91 (71–99%) (22)
Larynx	70 (35–93%) (10)	86 (42–99%) (7)	100 (59–100%) (7)
Hypopharynx	67 (9.4–99%) (3)	100 (29–100%) (3)	100 (29–100%) (3)
Stage			
Early (I and II)	100 (83–100%) (20)	70 (35–93%) (10)	100 (69–100%) (10)
Late (III and IV)	70 (58–80%) (73)	92 (78–98%) (37)	95 (82–99%) (37)
HPV			
Positive	40 (23–59%) (30)	86 (64–97%) (21)	86 (64–97%) (21)
Total	76 (66–85%) (93)	87 (74–95%) (47)	96 (85–99%) (47)

* Includes only patients from whom both saliva and plasma were available.

Table 3.2: Primer sequences used in multiplex assay for identification of driver mutations in tumors.

Gene	Genomic coordinates of amplified region*	Forward primer sequence	Reverse primer sequence	Amplicon length (bp)
FBXW7	chr4:153249328-153249457	GAAGTCCCAACCATGACAAGA	CGGACACTCAAAGTGTGGAA	130
FBXW7	chr4:153247295-153247409	TTGTTTTTCTGTTTCTCCCTCTG	CTGCAACATGACCCATCAAA	115
FBXW7	chr4:153247211-153247339	TTGAGACAGGCCAGTGTTTACAT	CAGTCTCTGGATCCCACACC	129
NRAS	chr1:115258719-115258840	GATGTGGCTCGCCAATTAAC	GATTGTCAGTGCGCTTTTCC	122
NRAS	chr1:115256485-115256609	ACACCCCCAGGATTCTTACAG	CGCCTGTCCTCATGTATTGG	125
PIK3CA	chr3:178916781-178916899	CCCCCTCCATCAACTTCTTC	GAAAAAGCCGAAGGTCACAA	119
PIK3CA	chr3:178921442-178921555	CAGACGCATTTCCACAGCTA	ACATTCACGTAGGTTGCACAAA	114
PIK3CA	chr3:178927879-178928007	AGAGATGATTGTTGAATTTTCCTTTT	AGTTTATATTTCCCCATGCCAAT	129
PIK3CA	chr3:178951924-178952055	GCATGCCAATCTCTTCATAAATC	TCCAAAGCCTCTTGCTCAGT	132
PIK3CA	chr3:178951991-178952119	TTTGATGACATTGCATACATTCG	GATCCAATCCATTTTTGTTGTCCAG	129
PIK3CA	chr3:178936018-178936140	CAATGAATTAAGGGAAAATGACAAA	CTCCATTTTAGCACTTACCTGTGAC	123
HRAS	chr11:534245-534371	GGCAGGAGACCCTGTAGGAG	GTTCTGGATCAGCTGGATGG	127
HRAS	chr11:533804-533926	GATGGCAAACACACACAGGA	GTGGTCATTGATGGGGAGAC	123
CDKN2A	chr9:21971073-21971187	CCGAGTGGCGGAGCTG	CACCAGCGTGTCCAGGAAG	115
CDKN2A	chr9:21970874-21971002	ACAAATTCTCAGATCATCAGTCCTC	AGGAGCTGGGCCATCG	129
CDKN2A	chr9:21970967-21971091	GCAGGTACCGTGCGACAT	CTTCCTGGACACGCTGGT	125
CDKN2A	chr9:21971115-21971237	GGGTCGGGTGAGAGTGG	TGGCTCTGACCATTCTGTTCT	123
CDKN2A	chr9:21970997-21971138	GCTCCTCAGCCAGGTCCA	GACCCCGCCACTCTCAC	142
CDKN2A	chr9:21974743-21974859	CACCTCCTCTACCCGACCC	GGGGAGAGCAGGCAGC	117
CDKN2A	chr9:21974684-21974798	GGCCTCCGACCGTAACTATT	AGCCTTCGGCTGACTGG	115
CDKN2A	chr9:21974652-21974761	CTCCCGCTGCAGACCCT	GGGTCGGGTAGAGGAGGTG	110
CDKN2A	chr9:21968163-21968300	CTGTAGGACCTTCGGTGACTG	TGTGCCACACATCTTTGACC	138
TP53	chr17:7579818-7579933	CCTTCCAATGGATCCACTCAC	ACTGCCTTCCGGGTCACT	116
TP53	chr17:7579679-7579789	AGCCCCCTAGCAGAGACCT	CAGCCCAACCCTTGTCCTT	111
TP53	chr17:7579502-7579616	TGACTGCTCTTTTCACCCATC	TCATCTGGACCTGGGTCTTC	115
TP53	chr17:7579441-7579572	GCAATGGATGATTTGATGCTG	CGGTGTAGGAGCTGCTGG	132
TP53	chr17:7579388-7579501	AGCTCCCAGAATGCCAGAG	TGGGAAGGGACAGAAGATGA	114

TP53	chr17:7579338-7579463	CTGCACCAGCAGCTCCTAC	CAGAATGCAAGAAGCCCAGA	126
TP53	chr17:7579270-7579396	GCATTGAAGTCTCATGGAAGC	CCCTTCCCAGAAAACCTACC	127
TP53	chr17:7578475-7578593	GCCCTGACTTTCAACTCTGTCT	GGGGGTGTGGAATCAACC	119
TP53	chr17:7578417-7578540	CTCCGTCATGTGCTGTGACT	CAACAAGATGTTTTGCCAACTG	124
TP53	chr17:7578335-7578455	GCCATGGCCATCTACAAGC	ACCAGCCCTGTCGTCTCTC	121
TP53	chr17:7578211-7578326	GTCCCCAGGCCTCTGATT	CGAAAAGTGTTTCTGTCATCCA	116
TP53	chr17:7578137-7578260	GTGGAAGGAAATTTGCGTGT	CTTAACCCCTCCTCCCAGAG	124
TP53	chr17:7577513-7577635	TGTGATGATGGTGAGGATGG	TCATCTTGGGCCTGTGTTATC	123
TP53	chr17:7577480-7577606	TGGCTCTGACTGTACCACCATC	GTGGCAAGTGGCTCCTGA	127
TP53	chr17:7577068-7577188	TGCCTCTTGCTTCTCTTTTCC	GCGGAGATTCTCTTCCTCTGT	121
TP53	chr17:7576997-7577121	CGTGTTTGTGCCTGTCCTG	GCTTCTTGTCCTGCTTGCTT	125
TP53	chr17:7576840-7576956	TTTTATCACCTTTCCTTGCCTCT	CAAGACTTAGTACCTGAAGGGTGAA	117
TP53	chr17:7576782-7576898	AAGAAGAAAACGGCATTTTGAG	CCAGCCAAAGAAGAAACCAC	117
TP53	chr17:7573947-7574059	CCCTGGCTCCTTCCCAG	CTTCTCCCCCTCCTCTGTTG	113
TP53	chr17:7573881-7574007	GTTCCGAGAGCTGAATGAGG	TAGGAAGGCAGGGGAGTAGG	127
TP53	chr17:7572929-7573043	AGTCTGAGTCAGGCCCTTCTG	ATGTCATCTCTCCTCCCTGCT	115
TP53	chr17:7572929-7573043	ATGTCATCTCTCCTCCCTGCT	AGTCTGAGTCAGGCCCTTCTG	115
TP53	chr17:7572893-7573009	GCCACCTGAAGTCCAAAAAG	GAGGCTGTCAGTGGGGAAC	117

*Coordinates refer to the human reference genome hg19 release (Genome Reference Consortium GRCh37, Feb 2009).

Table 3.3: Patient demographics

Patient ID	Age at diagnosis	Gender	Site	Subsite	Stage	T classification	N classification
HN 358	34	M	OC	Tongue	III	T2	N1
HN 359	67	M	L	Glottic	III	T3	N0
HN 361	68	M	OC	Tongue	I	T1	N0
HN 362	56	M	OP	Base of tongue	IV	T1	N2
HN 363	61	F	OC	Buccal mucosa	II	T2	N0
HN 364	61	M	OC	Tongue	IV	T4	N2
HN 365	57	M	OP	Soft palate	II	T2	N0
HN 366	70	M	OC	Alveolar ridge	IV	T4a	N0
HN 367	58	M	OP	Base of tongue	IV	T2	N2
HN 368	50	M	OC	Tongue	I	T1	N0
HN 369	74	M	L	Supraglottic	II	T2	N0
HN 370	54	M	OC	Buccal mucosa	IV	T2	N2
HN 371	77	F	OC	Tongue	I	T1	N0
HN 372	61	M	OP	Base of tongue	II	T2	N0
HN 373	69	M	L	Supraglottic	III	T3	N0
HN 375	54	M	OC	Alveolar ridge	I	T1	N0
HN 377	55	M	OP	Base of tongue	IV	T2	N2
HN 380	65	M	L	Transglottic	IV	T4a	N0
HN 381	45	M	OC	Tongue	II	T2	N0
HN 382	48	F	OC	Buccal mucosa	IV	T4a	N0
HN 383	61	M	OC	Tongue	I	T1	N0
HN 384	68	M	OP	Base of tongue	IV	T1	N2
HN 385	54	F	L	Supraglottic	IV	T4a	N1
HN 386	75	M	OC	Lip	IV	T4	N2
HN 389	59	M	OC	Hard palate	IV	T4b	N0
HN 390	61	M	OP	Tonsil	IV	T2	N3
HN 391	54	M	OP	Unknown	IV	Tx	N2
HN 392	65	M	OP	Base of tongue	IV	T2	N2
HN 393	83	M	OP	Base of tongue	IV	T1	N2
HN 394	51	M	L	Glottic	IV	T4a	N2
HN 395	55	M	OP	Tonsil	IV	T4a	N2
HN 396	58	M	OP	Base of tongue	IV	T1	N2
HN 397	55	M	OP	Tonsil	IV	T2	N3
HN 398	59	M	OC	Floor of mouth	III	T1	N1
HN 399	77	F	OC	Floor of mouth	I	T1	N0
HN 400	50	M	OP	Tonsil	IV	T1	N2
HN 401	52	M	OP	Base of tongue	IV	T2	N2
HN 402	38	M	OC	Tongue	III	T1	N1
HN 404	65	F	OP	Tonsil	IV	T2	N2
HN 405	65	M	H	Post cricoid	IV	T4b	N2
HN 406	65	M	OP	Base of tongue	III	T3	N0
HN 407	61	M	L	Transglottic	IV	T4a	N0
HN 408	61	M	OP	Tonsil	I	T1	N0
HN 409	68	M	OC	Tongue	IV	T4	N2
HN 410	58	M	OP	Tonsil	IV	T1	N2
HN 411	59	M	OP	Base of tongue	III	T1	N1
HN 412	58	M	OC	Tongue	III	T3	N0

HN 413	51	M	L	Supraglottic	IV	T4a	N2
HN 414	59	M	OP	Base of tongue	III	T2	N1
HN 415	56	M	OC	Tongue	II	T2	N0
HN 416	65	M	OC	Tongue	IV	T4	N2
HN 417	44	M	OC	Tongue	IV	T2	N2
HN 418	73	F	OC	Alveolar ridge	IV	T2	N2
HN 419	55	M	OC	Tongue	III	T3	N0
HN 420	40	M	OC	Tongue	IV	T2	N2
HN 421	62	M	OC	Floor of mouth	IV	T4	N2
HN 423	35	F	OC	Tongue	IV	T3	N2
HN 424	50	M	OC	Tongue	IV	T4a	N0
HN 425	68	M	OC	Floor of mouth	IV	T3	N2
HN 426	71	M	OC	Alveolar ridge	IV	T4	N2
HN 427	53	M	OC	Floor of mouth	IV	T4	N1
HN 428	58	F	OC	Tongue	II	T2	N0
HN 429	61	F	OC	Tongue	IV	T2	N2
HN 430	55	M	OC	Floor of mouth	II	T2	N0
HN 431	63	F	OC	Tongue	II	T2	N0
HN 432	71	F	OC	Tongue	II	T2	N0
HN 433	61	M	OC	Alveolar ridge	IV	T4	N2
HN 435	67	M	OC	Retromolar trigone	IV	T4a	N2
HN 436	62	M	OC	Tongue	IV	T2	N2
HN 438	56	F	OC	Tongue	I	T1	N0
HN 439	65	M	H	Piriform sinus	IV	T4a	N1
HN 440	53	M	OP	Base of tongue	III	T1	N1
HN 441	80	F	OP	Tonsil	III	T1	N1
HN 443	52	M	OC	Floor of mouth	IV	T4a	N2
HN 444	83	M	OP	Base of tongue	III	T3	N0
HN 445	74	F	OC	Tongue	I	T1	N0
HN 447	45	M	OP	Tonsil	IV	T2	N2
HN 449	52	M	OP	Tonsil	IV	T3	N2
HN 450	45	M	OP	Unknown	III	Tx	N1
HN 451	82	M	OC	Tongue	IV	T2	N2
HN 452	60	M	OP	Base of tongue	IV	T1	N2
HN 454	86	M	OC	Tongue	IV	T1	N3
HN 456	60	M	H	Piriform sinus	IV	T4a	N0
HN 457	67	M	OP	Base of tongue	III	T3	N1
HN 458	58	M	L	Glottic	IV	T4a	N0
HN 459	54	M	OC	Tongue	IV	T3	N2
HN 460	54	M	OP	Base of tongue	IV	T2	N2
HN 462	54	F	OC	Tongue	IV	T2	N2
HN 463	68	M	L	Glottic	I	T1a	N0
HN 464	47	M	OP	Base of tongue	IV	T1	N2
HN 469	58	M	OP	Tonsil	IV	T2	N2
HN 473	45	M	OP	Base of tongue	III	T1	N1
HN 474	39	M	OP	Base of tongue	IV	T2	N2

OC = oral cavity, OP = oropharynx, H = hypopharynx, L = larynx, and NA = not available

Table 3.4 Patient treatment information

Patient ID	Interval between Biopsy and Sample Collection (days)	Treatment Prior to Sample Collection for Patients with Recurrence (days prior)
HN 358	78	No prior treatment
HN 359	335	RT (449)
HN 361	29	No prior treatment
HN 362	66	No prior treatment
HN 363	24	Surgery (151)
HN 364	21	Surgery, RT (324)
HN 365	28	CRT (244)
HN 366	35	No prior treatment
HN 367	24	CRT (774)
HN 368	15	No prior treatment
HN 369	20	RT (190)
HN 370	34	No prior treatment
HN 371	40	No prior treatment
HN 372	75	Surgery, RT (3503)
HN 373	No prior biopsy	CRT (161)
HN 375	42	No prior treatment
HN 377	No prior biopsy	Surgery (709)
HN 380	27	CRT (432)
HN 381	32	CRT (194)
HN 382	14	No prior treatment
HN 383	26	No prior treatment
HN 384	49	No prior treatment
HN 385	93	No prior treatment
HN 386	84	Surgery (1249)
HN 389	35	No prior treatment
HN 390	33	No prior treatment
HN 391	48	No prior treatment
HN 392	9	No prior treatment
HN 393	28	No prior treatment
HN 394	27	RT (337)
HN 395	23	No prior treatment
HN 396	9	No prior treatment
HN 397	42	No prior treatment
HN 398	23	No prior treatment
HN 399	No prior biopsy	No prior treatment
HN 400	18	No prior treatment
HN 401	23	CRT (226)
HN 402	35	No prior treatment

HN 404	36	No prior treatment
HN 405	53	No prior treatment
HN 406	21	No prior treatment
HN 407	20	No prior treatment
HN 408	42	Surgery, RT (1783)
HN 409	28	No prior treatment
HN 410	25	No prior treatment
HN 411	78	No prior treatment
HN 412	20	No prior treatment
HN 413	109	No prior treatment
HN 414	22	Surgery, RT (3757)
HN 415	69	No prior treatment
HN 416	25	No prior treatment
HN 417	58	No prior treatment
HN 418	74	No prior treatment
HN 419	22	No prior treatment
HN 420	31	No prior treatment
HN 421	68	No prior treatment
HN 423	35	No prior treatment
HN 424	107	No prior treatment
HN 425	36	No prior treatment
HN 426	31	No prior treatment
HN 427	45	No prior treatment
HN 428	55	No prior treatment
HN 429	58	No prior treatment
HN 430	49	No prior treatment
HN 431	39	No prior treatment
HN 432	58	No prior treatment
HN 433	49	Surgery, CRT (1095)
HN 435	77	No prior treatment
HN 436	52	No prior treatment
HN 438	22	No prior treatment
HN 439	49	No prior treatment
HN 440	52	No prior treatment
HN 441	62	No prior treatment
HN 443	87	No prior treatment
HN 444	56	Surgery, RT (2279)
HN 445	25	No prior treatment
HN 447	8	No prior treatment
HN 449	29	No prior treatment
HN 450	48	No prior treatment
HN 451	No prior biopsy	CRT (157)
HN 452	90	No prior treatment

HN 454	23	No prior treatment
HN 456	26	RT (121)
HN 457	18	Surgery, RT (4729)
HN 458	40	No prior treatment
HN 459	22	Surgery, RT (69)
HN 460	37	No prior treatment
HN 462	41	No prior treatment
HN 463	No prior biopsy	No prior treatment
HN 464	28	No prior treatment
HN 469	58	No prior treatment
HN 473	11	No prior treatment
HN 474	44	No prior treatment

RT = Radiation therapy, CRT = Chemoradiation therapy

Table 3.5 Tumor mutation information

Patient ID	HPV status of Tumor	Gene mutated in Tumor	Type of Mutation	Genomic Position	cDNA Change	Amino Acid Change
HN 358	HPV negative	TP53	SBS	7578524	406C>T	Q136*
HN 359	HPV negative	TP53	SBS	7577082	856G>A	E286K
HN 361	HPV negative	TP53	SBS	7578268	581T>G	L194R
HN 362	HPV16 positive	PIK3CA	SBS	178936091	1633G>A	E545K
HN 363	HPV negative	TP53	SBS	7576851	c.993+2T>C	NA
HN 364	HPV negative	TP53	SBS	7576928	c.920-2A>G	NA
HN 365	HPV16 positive	HPV16	NA	NA	NA	NA
HN 366	HPV negative	TP53	SBS	7578553	377A>G	Y126C
HN 367	HPV16 positive	PIK3CA	SBS	178936082	1624G>A	E542K
HN 368	HPV negative	TP53	SBS	7578550	380C>A	S127Y
HN 369	HPV negative	TP53	INDEL	7577094	c.844_845insCTGTGCGCC	R282fs
HN 370	HPV negative	TP53	SBS	7577094	844C>T	R282W
HN 371	HPV negative	NOTCH1	SBS	139411837	1442G>T	G481V
HN 372	HPV negative	TP53	SBS	7577120	818G>T	R273L
HN 373	HPV negative	TP53	SBS	7574021	1006G>T	E336*
HN 375	HPV negative	TP53	SBS	7577574	707A>G	Y236C
HN 377	HPV16 positive	HPV16	NA	NA	NA	NA
HN 380	HPV negative	TP53	SBS	7577082	856G>A	E286K
HN 381	HPV negative	TP53	SBS	7577097	841G>A	D281N
HN 382	HPV negative	TP53	SBS	7577570	711G>A	M237I
HN 383	HPV negative	TP53	SBS	7577094	844C>T	R282W
HN 384	HPV16 positive	PIK3CA	SBS	178936091	1633G>A	E545K
HN 385	HPV negative	TP53	SBS	7579358	329G>T	R110L
HN 386	HPV negative	TP53	SBS	7578500	430C>T	Q144*
HN 389	HPV negative	PIK3CA	SBS	178952085	3140A>G	H1047R

HN 390	HPV16 positive	HPV16	NA	NA	NA	NA
HN 391	HPV16 positive	HPV16	NA	NA	NA	NA
HN 392	HPV16 positive	FBXW7	SBS	153249384	1394C>T	R465H
HN 393	HPV16 positive	HPV16	NA	NA	NA	NA
HN 394	HPV negative	TP53	SBS	7578179	670G>T	E224*
HN 395	HPV16 positive	HPV16	NA	NA	NA	NA
HN 396	HPV16 positive	PIK3CA	SBS	178916876	263G>A	R88Q
HN 397	HPV negative	TP53	SBS	7578478	452C>A	P151H
HN 398	HPV negative	NA	Translocation	chr11:69467879-chr11:69469070	NA	NA
HN 399	HPV negative	TP53	SBS	7577121	817C>T	R273C
HN 400	HPV16 positive	PIK3CA	SBS	178936091	1633G>A	E545K
HN 401	HPV16 positive	HPV16	NA	NA	NA	NA
HN 402	HPV negative	TP53	SBS	7577120	818G>A	R273H
HN 404	HPV16 positive	NRAS	SBS	115256530	181C>A	Q61K
HN 405	HPV negative	TP53	SBS	7577538	743G>A	R248Q
HN 406	HPV negative	TP53	SBS	7574000	1027G>T	E343*
HN 407	HPV negative	TP53	SBS	7578406	524G>A	R175H
HN 408	HPV negative	TP53	SBS	7577568	713G>T	C238F
HN 409	HPV negative	CDKN2A	INDEL	21971125-21971127	231delTC	T77fs
HN 410	HPV16 positive	HPV16	NA	NA	NA	NA
HN 411	HPV16 positive	HPV16	NA	NA	NA	NA
HN 412	HPV negative	NA	Translocation	chr18:45662870-chr11:69291050	NA	NA
HN 413	HPV negative	TP53	SBS	7577538	743G>T	R248L
HN 414	HPV16 positive	PIK3CA	SBS	178936082	1624G>A	E542K
HN 415	HPV16 positive	HPV16	NA	NA	NA	NA
HN 416	HPV negative	NOTCH1	INDEL	139396749	5359delC	L1787fs
HN 417	HPV negative	TP53	SBS	7578406	524G>A	R175H
HN 418	HPV negative	TP53	SBS	7577538	743G>A	R248Q

HN 419	HPV negative	TP53	SBS	7578271	578A>T	H193L
HN 420	HPV negative	TP53	SBS	7574003	1024C>T	R342*
HN 421	HPV negative	NOTCH1	INDEL	139418317	255_256insA	Y85fs
HN 423	HPV negative	TP53	SBS	7577081	857A>C	E286A
HN 424	HPV negative	TP53	SBS	7577046	892G>T	E298*
HN 425	HPV negative	TP53	INDEL	7577498	c.782+1_782+2insC	NA
HN 426	HPV negative	CDKN2A	SBS	21968242	c.458-1G>A	NA
HN 427	HPV negative	TP53	SBS	7578392	538G>T	E180*
HN 428	HPV negative	TP53	SBS	7577121	817C>T	R273C
HN 429	HPV negative	TP53	SBS	7577574	707A>G	Y236C
HN 430	HPV negative	TP53	SBS	7578212	637C>T	R213*
HN 431	HPV negative	TP53	SBS	7574017	1010G>T	R337L
HN 432	HPV negative	TP53	SBS	7577118	820G>T	V274F
HN 433	HPV negative	TP53	SBS	7578458	472C>G	R158G
HN 435	HPV negative	TP53	SBS	7578550	380C>A	S127Y
HN 436	HPV negative	TP53	SBS	7579366	321C>G	Y107*
HN 438	HPV negative	TP53	SBS	7577590	691A>G	T231A
HN 439	HPV negative	TP53	SBS	7578188	661G>T	E221*
HN 440	HPV16 positive	HPV16	NA	NA	NA	NA
HN 441	HPV16 positive	HPV16	NA	NA	NA	NA
HN 443	HPV negative	TP53	INDEL	7579590	97-2_97delAGT	S33fs
HN 444	HPV16 positive	PIK3CA	SBS	178936082	1624G>A	E542K
HN 445	HPV negative	PIK3CA	SBS	178952085	3140A>G	H1047R
HN 447	HPV16 positive	HPV16	NA	NA	NA	NA
HN 449	HPV16 positive	FBXW7	SBS	153249384	1394C>T	R465H
HN 450	HPV16 positive	HPV16	NA	NA	NA	NA
HN 451	HPV negative	TP53	SBS	7577094	844C>T	R282W
HN 452	HPV16 positive	FBXW7	SBS	153247289	1513C>G	R505G

HN 454	HPV negative	TP53	SBS	7578492	438G>A	W146*
HN 456	HPV negative	TP53	SBS	7574018	1009C>T	R337C
HN 457	HPV negative	TP53	SBS	7579358	329G>T	R110L
HN 458	HPV negative	TP53	SBS	7577121	817C>T	R273C
HN 459	HPV negative	TP53	SBS	7578271	578A>G	H193R
HN 460	HPV16 positive	PIK3CA	SBS	178936091	1633G>A	E545K
HN 462	HPV negative	TP53	SBS	7578492	438G>A	W146*
HN 463	HPV negative	TP53	SBS	7578208	641A>G	H214R
HN 464	HPV16 positive	HPV16	NA	NA	NA	NA
HN 469	HPV16 positive	HPV16	NA	NA	NA	NA
HN 473	HPV16 positive	HPV16	NA	NA	NA	NA
HN 474	HPV16 positive	HPV16	NA	NA	NA	NA

Table 3.6 Saliva mutation information

Patient ID	Somatic Mutation: Patient FRACTION of MUTANT reads	Somatic Mutation: Patient NUMBER of MUTANT reads	Somatic Mutation: Patient NUMBER of WILD-TYPE reads	Somatic Mutation: CONTROL FRACTION of MUTANT reads	Somatic Mutation: CONTROL NUMBER of MUTANT reads	Somatic Mutation: CONTROL NUMBER of WILD-TYPE reads	Somatic Mutation: P-value (Patient vs. Control)	HPV: templates/ng DNA
HN 358	0.541%	657	120798	0.000%	0	78462	3.00E-94	Not quantified
HN 359	0.009%	19	205715	0.000%	0	190784	7.25E-05	Not quantified
HN 361	0.245%	845	343819	0.004%	6	137370	5.81E-72	Not quantified
HN 362	0.000%	0	32958	0.000%	0	99828	NA	No HPV found
HN 363	5.443%	8843	153613	0.001%	3	244530	0.00E+00	Not quantified
HN 364	0.106%	189	177609	0.001%	2	285252	7.88E-66	Not quantified
HN 365	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	59.8
HN 366	3.296%	4905	143907	0.000%	0	78462	0.00E+00	Not quantified
HN 367	0.003%	28	862561	0.000%	0	165648	1.66E-02	No HPV found
HN 368	11.467%	22073	170413	0.002%	4	229130	0.00E+00	Not quantified
HN 369	0.001%	10	1190330	0.000%	0	3671899	7.73E-07	Not quantified
HN 370	2.390%	9233	377059	0.006%	6	94302	0.00E+00	Not quantified
HN 371	1.628%	9504	574434	0.000%	0	74781	2.77E-270	Not quantified
HN 372	0.007%	29	426931	0.000%	2	523939	1.40E-07	Not quantified
HN 373	0.000%	0	79917	0.000%	0	75699	NA	Not quantified
HN 375	0.040%	123	310509	0.000%	0	222409	1.49E-20	Not quantified
HN 377	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 380	0.026%	107	414013	0.000%	0	304190	1.81E-18	Not quantified
HN 381	0.282%	911	322360	0.000%	0	94302	1.37E-59	Not quantified
HN 382	6.120%	20130	308811	0.000%	0	73755	0.00E+00	Not quantified
HN 383	0.190%	394	207299	0.006%	6	94302	2.01E-37	Not quantified

HN 384	0.000%	0	33081	0.000%	0	99828	NA	No HPV found
HN 385	0.004%	29	717919	0.000%	0	297066	1.11E-03	Not quantified
HN 386	0.168%	166	98840	0.011%	4	36747	7.42E-13	Not quantified
HN 389	3.497%	5130	141558	0.000%	0	132561	0.00E+00	Not quantified
HN 390	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	0.6
HN 391	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 392	0.005%	8	165676	0.000%	0	113715	2.48E-02	0.3
HN 393	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 394	0.000%	0	111372	0.000%	0	8730	NA	Not quantified
HN 395	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	54.1
HN 396	0.000%	0	248376	0.000%	0	75132	NA	No HPV found
HN 397	0.000%	0	174729	0.000%	0	36747	NA	Not quantified
HN 398	43.000%	4301	5701	0.000%	0	9998	NA	Not quantified
HN 399	1.850%	5696	302191	0.014%	14	101826	0.00E+00	Not quantified
HN 400	0.000%	0	44937	0.000%	0	99828	NA	No HPV found
HN 401	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	243.5
HN 402	0.027%	90	336318	0.000%	0	101826	3.49E-07	No HPV found
HN 404	0.000%	0	435825	0.000%	0	96993	NA	No HPV found
HN 405	0.033%	57	175443	0.001%	1	73755	6.60E-06	Not quantified
HN 406	0.005%	22	446306	0.000%	0	151305	1.29E-02	Not quantified
HN 407	0.024%	94	398354	0.004%	8	189739	2.35E-07	Not quantified
HN 408	0.006%	18	294486	0.001%	2	222409	5.83E-03	Not quantified
HN 409	0.028%	102	370428	0.000%	0	62736	5.93E-05	Not quantified
HN 410	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	2.7
HN 411	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 412	41.000%	4099	5899	0.000%	0	11051	NA	Not quantified
HN 413	0.000%	0	139509	0.000%	0	73755	NA	Not quantified
HN 414	0.000%	0	118473	0.000%	0	99828	NA	0.3

HN 415	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	399.9
HN 416	0.060%	98	164341	0.000%	0	12387	1.17E-02	Not quantified
HN 417	0.560%	698	123940	0.011%	16	140283	2.58E-162	Not quantified
HN 418	6.097%	19333	297749	0.006%	12	215316	0.00E+00	Not quantified
HN 419	0.217%	302	139354	0.000%	0	243910	3.26E-116	Not quantified
HN 420	7.347%	32586	410934	0.004%	14	323181	0.00E+00	Not quantified
HN 421	3.469%	9920	276064	0.000%	0	52392	0.00E+00	No HPV found
HN 423	2.104%	3222	149904	0.000%	0	94302	0.00E+00	Not quantified
HN 424	0.653%	2935	446813	0.000%	0	94302	2.58E-136	Not quantified
HN 425	0.068%	253	374009	0.000%	0	213470	6.50E-33	Not quantified
HN 426	1.744%	842	47438	0.002%	2	103908	0.00E+00	Not quantified
HN 427	0.559%	390	69396	0.003%	4	140283	9.79E-169	Not quantified
HN 428	1.086%	2141	195085	0.009%	25	274420	0.00E+00	Not quantified
HN 429	2.072%	188	8878	0.000%	0	177032	0.00E+00	Not quantified
HN 430	0.759%	968	126514	0.012%	21	182440	2.16E-288	Not quantified
HN 431	1.263%	4979	389245	0.001%	3	323181	0.00E+00	Not quantified
HN 432	1.430%	1870	128915	0.000%	0	101826	0.00E+00	Not quantified
HN 433	0.204%	119	58051	0.000%	0	36747	9.36E-18	Not quantified
HN 435	2.200%	3592	159698	0.000%	0	78462	0.00E+00	Not quantified
HN 436	0.880%	2415	272049	0.000%	0	250810	0.00E+00	Not quantified
HN 438	0.062%	210	340200	0.001%	3	215316	1.04E-28	Not quantified
HN 439	0.000%	0	525321	0.000%	0	8730	NA	Not quantified
HN 440	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 441	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	0.3
HN 443	0.570%	3610	629240	0.000%	0	84825	1.60E-107	Not quantified
HN 444	0.020%	97	481403	0.000%	0	156321	3.93E-08	7.2
HN 445	0.151%	359	237184	0.000%	0	132561	3.49E-45	Not quantified
HN 447	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found

HN 449	0.019%	68	364936	0.002%	4	217514	4.92E-08	15.6
HN 450	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 451	0.419%	1506	358329	0.024%	24	101826	2.91E-83	Not quantified
HN 452	0.000%	0	466506	0.000%	0	247275	NA	No HPV found
HN 454	0.121%	113	92902	0.000%	0	36747	4.71E-11	Not quantified
HN 456	0.020%	69	351123	0.003%	2	75699	1.72E-03	Not quantified
HN 457	0.113%	2360	2086288	0.000%	0	297066	8.44E-75	Not quantified
HN 458	0.106%	32	30565	0.014%	14	101826	2.82E-13	Not quantified
HN 459	0.076%	248	327670	0.000%	0	137370	4.25E-24	Not quantified
HN 460	0.000%	0	94842	0.000%	0	99828	NA	No HPV found
HN 462	0.073%	44	60139	0.000%	0	36747	4.93E-07	Not quantified
HN 463	0.015%	61	404291	0.000%	0	354892	6.53E-13	No HPV found
HN 464	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 469	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	1.9
HN 473	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 474	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found

Patients in **BOLD** had both saliva and plasma available

Table 3.7 Plasma mutation information

Patient ID	Somatic Mutation: Patient FRACTION of MUTANT reads	Somatic Mutation: Patient NUMBER of MUTANT reads	Somatic Mutation: Patient NUMBER of WILD-TYPE reads	Somatic Mutation: CONTROL FRACTION of MUTANT reads	Somatic Mutation: CONTROL NUMBER of MUTANT reads	Somatic Mutation: CONTROL NUMBER of WILD-TYPE reads	Somatic Mutation: P-value (Patient vs. Control)	HPV: templates/ng DNA
HN 358	0.000%	0	69216	0.000%	0	81680	NA	Not quantified
HN 359	0.019%	54	282174	0.002%	3	128972	4.05E-05	Not quantified
HN 361	0.017%	21	120144	0.001%	3	292883	1.24E-09	Not quantified
HN 362	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 363	0.796%	1999	251160	0.000%	0	143400	3.89E-249	No HPV found
HN 364	0.901%	1831	203268	0.000%	0	162245	0.00E+00	No HPV found
HN 365	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	63
HN 366	0.582%	674	115812	0.000%	0	108590	1.19E-138	Not quantified
HN 367	0.423%	379	89622	0.000%	0	155200	8.88E-144	45.5
HN 368	0.867%	338	39024	0.000%	0	108590	1.95E-204	No HPV found
HN 369	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 370	0.498%	251	50454	0.010%	10	96060	6.68E-97	Not quantified
HN 371	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 372	0.082%	63	76758	0.000%	0	118580	2.16E-22	Not quantified
HN 373	0.092%	173	187566	0.000%	0	148060	3.80E-31	Not quantified
HN 375	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 377	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	6386.6
HN 380	1.379%	415	30132	0.002%	3	128972	0.00E+00	No HPV found
HN 381	1.852%	854	46104	0.000%	0	96070	0.00E+00	No HPV found
HN 382	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma

HN 383	0.031%	5	16158	0.004%	5	140730	1.87E-03	Not quantified
HN 384	0.022%	19	87564	0.000%	0	155200	2.64E-08	8.9
HN 385	0.089%	116	130764	0.000%	0	148650	4.82E-30	Not quantified
HN 386	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 389	0.022%	97	440478	0.000%	0	166610	2.79E-09	Not quantified
HN 390	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	4
HN 391	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 392	0.638%	770	120570	0.004%	5	126570	1.64E-173	33.1
HN 393	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 394	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 395	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	57.8
HN 396	0.344%	331	96294	0.011%	26	235469	7.38E-154	369.2
HN 397	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 398	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 399	0.000%	0	83514	0.000%	0	556028	NA	Not quantified
HN 400	0.279%	82	29580	0.000%	0	155200	5.36E-94	No HPV found
HN 401	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	126.2
HN 402	0.028%	13	44574	0.003%	3	118577	1.10E-05	Not quantified
HN 404	0.000%	0	20010	0.000%	0	136300	NA	No HPV found
HN 405	0.057%	40	70278	0.002%	3	162792	1.19E-18	Not quantified
HN 406	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 407	0.016%	5	30366	0.002%	2	105793	7.69E-03	Not quantified
HN 408	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 409	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 410	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	12.3
HN 411	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	1.1
HN 412	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 413	0.330%	455	137820	0.002%	4	162791	1.74E-115	Not quantified

HN 414	0.200%	247	123774	0.000%	0	155200	9.26E-69	20.7
HN 415	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 416	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 417	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 418	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 419	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 420	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 421	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 423	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 424	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 425	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 426	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 427	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 428	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 429	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 430	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 431	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 432	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 433	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 435	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 436	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 438	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 439	2.490%	2300	92358	0.000%	0	153345	0.00E+00	Not quantified
HN 440	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	3.4
HN 441	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 443	0.066%	27	41466	0.000%	0	123775	2.07E-18	Not quantified
HN 444	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 445	0.000%	0	119568	0.000%	0	166610	NA	Not quantified

HN 447	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	11.8
HN 449	0.084%	48	57852	0.004%	5	126570	6.40E-20	67.4
HN 450	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	7.9
HN 451	0.869%	1513	174054	0.010%	10	96060	3.07E-177	No HPV found
HN 452	0.069%	128	187146	0.001%	2	227918	6.51E-34	13.7
HN 454	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 456	2.858%	2224	77814	0.007%	10	148050	0.00E+00	No HPV found
HN 457	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 458	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 459	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 460	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 462	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 463	0.000%	0	44520	0.000%	0	153345	NA	Not quantified
HN 464	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	No HPV found
HN 469	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	NA	4.5
HN 473	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma
HN 474	No plasma	No plasma	No plasma	No plasma	No plasma	No plasma	NA	No plasma

Patients in **BOLD** had both saliva and plasma available

Chapter 4:

Discussion

DEVELOPMENT OF BIOMARKER ASSAYS

Liquid biopsies represent an exciting new field in cancer research. However, as with any biomarker type assay it is important to make sure that any new discoveries are not beset by any potential drawbacks. Over-diagnosis is always a potential issue with screening tests. Over-diagnosis can lead to direct harm to the patient. As with the case of PSA, poor specificity and many false positives can lead to mental anguish for the patient as well as expensive and invasive follow up tests. The easiest way to help minimize this problem is to focus on increasing the specificity of the test. A poor specificity will lead to a large number of people testing positive that do not even have cancer. Focusing on high specificity while retaining good sensitivity should be the aim of any biomarker assay that hopes to make a significant impact on cancer diagnosis and treatment.

IDENTIFYING MOLECULAR MARKERS TO IMPROVE THE CLASSIFICATION OF PANCREATIC CYSTS

The results from this study lead to several important conclusions. First, over half of the patients who underwent surgical resection were subsequently found to have a benign SCA, or an IPMN with low-, or intermediate-grade dysplasia, and could potentially have continued surveillance of their cyst rather than undergoing surgical resection. This highlights the difficulties of identifying those cysts that require surgery, versus those in whom surveillance is safe, using the modalities currently available to clinicians. These results are in accord with the numerous publications showing that current diagnostic criteria for managing cyst patients are inadequate (Sahora et al., 2013; Valsangkar et al., 2012).

Second, we show that the use of composite clinical or molecular markers could substantially increase diagnostic accuracy. When either the composite clinical marker or the composite molecular marker was used alone, the sensitivity for identifying cysts that required resection reached ~75%; in contrast, when used together, sensitivity increased to 92%. It is difficult

to estimate the number of cyst patients who would have surgery, and therefore would not develop PDAC, if these markers were widely applied in patient management. Similarly, we hesitate to calculate how many needless surgeries might be avoided if these new markers were broadly applied. However, the data strongly suggest that the combination of clinical and molecular features will be more accurate for assessing cyst type and need for surgical resection than either alone.

The clinical and radiologic findings incorporated into our composite clinical marker are the result of decades of careful study by clinicians (Lennon et al., 2013; Lennon et al., 2014). What was added in Masica et al 2017 was a rigorous and quantitative assessment of the most predictive features and combinations of features. Similarly, the molecular analysis we employed did not require the discovery of new genetic alterations present in pancreatic cysts. Guided by prior studies, we have developed assays employing massively parallel sequencing to robustly detect these genetic alterations, even when present in relatively low fractions of template molecules. We then used these data to identify the most predictive molecular features and combinations of molecular features in a rigorous fashion.

A balance between sensitivity and specificity nearly always must be made during the development of biomarkers for any disease. The current study was no exception. The most obvious example was in the determination of the need for surgery (Table 2.4). The composite molecular/clinical marker provided an excellent sensitivity (89%), considerably higher than either the composite molecular or composite clinical marker alone (75% or 77%, respectively). However, this increase in sensitivity with the composite molecular/clinical marker compromised specificity, reducing it from 92% with the composite molecular marker to 69%. Note that we purposefully designed these algorithms to reach maximum sensitivity, sacrificing specificity if necessary, as we considered it worse to "miss" a cyst that should be surgically excised than to unnecessarily perform surgery on a cyst that should have not been excised. However, this example illustrates that it is not

always possible to increase sensitivity without decreasing specificity, even with the aid of combinatorial approaches such as MOCA.

The summary data in Table 2.4, once validated independently, can be used in various ways, depending on the clinical situation. In a young and otherwise healthy patient, it would make sense to use the most sensitive method available to determine the need for surgery. The composite molecular/clinical marker might therefore be used to evaluate such a patient, as it is the most sensitive. In an elderly patient with significant comorbidities, however, it might make more sense to use the less sensitive, but more specific composite molecular marker to determine the need for surgery. Use of the composite molecular marker would largely avoid unnecessary surgery (specificity of 92%), while preserving a reasonable sensitivity for highrisk cysts (75%).

Our study has several limitations. The number of cyst fluid samples from some cyst types was relatively limited, thereby limiting confidence in our estimates of sensitivity and specificity (as indicated by the confidence intervals provided in all Tables). Another limitation is that our composite molecular marker was validated through cross-validation rather than through experimental validation of an independent cohort. Though crossvalidation is statistically sound, it is not as reliable as the evaluation of a distinct cohort. It is important to note that this limitation does not apply to the composite clinical marker; the 130 patients evaluated here were distinct from the patients used to define the composite clinical markers (Masica et al., 2017). It was therefore gratifying that the sensitivities and specificities estimated from the cross-validation in Masica et al 2017 were similar to those found in the current study (Table 2.8).

The possibility of false negative results is always a concern with any sequencing technique. This study was designed to detect all known mutations in oncogenes and most mutations in tumor suppressor genes that occur in cysts, based on genome wide sequencing (Wu et al., 2011). All missense mutations and small insertions or deletions present at allele frequencies of greater than 1% are easily detectable at the sequencing depth employed in this study (Kinde et al., 2011).

Moreover, the approach used in this study yields not just the presence or absence of a mutation; it reveals the precise fractional representation of the mutation in the DNA sample. Another source of false negative results could occur in tumor suppressor genes. Large deletions or insertions, as well as translocations, will not be detected upon sequencing. However, such changes are often associated with LOH or copy number changes, and these would be identified by the other assays employed in this study.

Obtaining large volumes of cyst fluid for analysis can be problematic and could potentially affect the ability to detect alterations in the molecular markers described in this study. However unlike analysis of cyst fluid CEA, which requires 0.5ml to 1ml of cyst fluid in most centers, we specifically chose to use methods that can be used on very small amounts of DNA and do not require library preparation. Each of the three methods is based on direct polymerase chain reaction, and 10 ng of DNA is adequate for all of the tests combined. Using 0.25 mL of either EUS or surgically obtained fluids has nearly always yielded sufficient DNA.

This was a retrospective study. While this has the advantage that all patients underwent surgical resection and therefore had defined pathology, it is possible that these markers will not perform as well in the general population of cyst patients. It was comforting in this respect that our analysis of paired fluids obtained from EUS-guided cyst aspiration and subsequent surgery revealed very similar molecular genetic alterations. In the future, the optimum study design will incorporate examination of cyst fluids taken at routine EUS sessions over time, then comparing the results to those obtained at surgery. In addition, the optimum role of the composite molecular markers, and which patients will be benefit from their use, should be addressed in these studies. It will be enticing in such research studies to use the composite molecular and clinical markers described in this work to help guide the decision about surgery. But because of the limitations of our study described above, any such guidance should be performed only in a research study. Use of our composite molecular or clinical markers in common practice, outside of a research study, is not yet warranted.

DETECTION OF SOMATIC MUTATIONS AND HPV IN THE SALIVA AND PLASMA OF PATIENTS WITH HEAD AND NECK SQUAMOUS CELL CARCINOMAS

Current diagnostic methods for HNSCC make the detection of early disease, assessment of response to treatment, and differentiation between the adverse effects of treatment versus persistent or recurrent disease challenging. These issues collectively compromise clinical decision-making and impair patient management. Although it is now abundantly clear that all cancers, including HNSCC, are the result of genetic alterations, this knowledge is just beginning to be applied to meet diagnostic challenges such as those described above (Vogelstein et al., 2013). In this proof-of-principle study, we show that tumor-derived DNA can be detected in the saliva of patients with HNSCC. We also show that the evaluation of plasma can complement that of saliva, together allowing detection of tumor-derived DNA in readily obtainable bodily fluids in >90% of the studied patients. Our findings lay the conceptual and practical foundation for clinical tests designed for the earlier detection of HNSCC, either for patients at high risk for the disease or for patients previously treated for HNSCC who are at risk for disease recurrence. Moreover, these results establish a paradigm for monitoring the response to treatment.

There were several notable findings in this study. The sensitivity for detection of tumorderived DNA in the saliva was site-dependent and most efficient for tumors in the oral cavity. Not only was tumor DNA detectable in every one of the 46 patients with cancers of the oral cavity, but also the fraction of mutant DNA in the saliva was particularly high (median, 0.65%; interquartile range, 0.17 to 2.2%; mean, 3.46%). Moreover, early-stage oral cavity cancers were highly detectable; 75% of the patients with oral cavity cancer were at an early stage (stage I or II), and all were detectable. The high fraction of tumor DNA in the saliva of patients with oral cancers makes anatomical sense and demonstrates the advantage of examining local bodily fluids for optimal sensitivity in this type of assay.

HNSCCs distal to the oral cavity (oropharynx, larynx, and hypopharynx) were still often detectable through the examination of saliva, but the frequency of their detection (47, 70, and 67%, respectively) and the fraction of mutant alleles (median, 0.015%) were considerably lower than found in the oral cavity (0.65%). Anatomical locations likely explain this difference. Gargling might increase the detectability of tumor DNA in these distal compartments, but to test this idea, a reproducible procedure for gargling would have to be devised.

One striking aspect of this study is the increased sensitivity demonstrated when assays of two compartments are combined. This increased sensitivity is possible only because of the exquisite specificity of mutant DNA as biomarker; because no false-positives are expected, any number of assays can be combined, increasing sensitivity without compromising specificity. The combination of saliva and plasma allowed detection of 96% of the cancers when both fluids were available, higher than obtained with either saliva or plasma alone.

We emphasize that our study establishes the proof of principle for the use of saliva and plasma to reveal the presence of HNSCCs, but does not comprise a clinical test. In each patient, we first evaluated the tumor, then used an alteration (either the presence of HPV or a somatic mutation) to query the saliva or plasma. In an actual diagnostic test, a panel of genes would have to be used to assess each case. Fortunately, technologies are available for finding mutations, even those present at low frequencies (Kinde et al., 2013; Kinde et al., 2012). On the basis of the results presented herein, as well as large studies of HNSCC genetics (Agrawal et al., 2011; Stransky et al., 2011; Cancer Genome Atlas, 2015), a panel including HPV16 DNA sequences, *TP53*, *PIK3CA*, *NOTCH1*, and *CDKN2A* would be able to detect >95% of invasive HNSCCs. Another limitation of our study is that the number of early-stage cancers beyond the oral cavity was small, in part reflecting the unfortunate fact that most of these cancers are detected only when they are late-stage. Future larger studies should be able to determine how often early-stage cancers of the oropharynx, larynx, and hypopharynx can be detected using the approach described here. The fact that at least

70% of the oropharyngeal cancers in the United States are associated with HPV simplifies this task (Chaturvedi et al., 2011; Cancer Genome Atlas, 2015).

One important application of the results described in this paper is in the diagnosis of clinically suspicious lesions. The often complex and highly specialized nature of current HNSCC diagnostic procedures can lead to delays in diagnosis and treatment, negatively impacting prognosis and survival (Kowalski and Carvalho, 2001; Guggenheimer et al., 1989; Wildt et al., 1995; Allison et al., 1998; Carvalho et al., 2002; Koivunen et al., 2001). These delays could be prevented in many patients through the examination of saliva and plasma for tumor DNA. Such a test could potentially be incorporated into routine examinations to complement current diagnostic modalities and inform clinical decision-making. Another obvious application is in disease monitoring and surveillance. In nine patients with positive pre-treatment saliva and/or plasma, samples were collected at various times after surgery. The fact that no mutations were identified after surgery in the five patients whose tumors did not recur highlights the specificity of the mutation-based assay. It was also encouraging that we identified tumor DNA in the saliva of patients whose tumors were found to recur at the clinical level only months later, suggesting that these tests could provide a clinically meaningful lead time. The results presented here provide the setting for a larger study to explore whether the presence of tumor DNA in either saliva or plasma can be used to help manage patients who appear free of disease after definitive treatment by clinical criteria.

REFERENCES

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Trevino L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011; 333:1154–1157.

Agrawal N, Jiao Y, Bettegowda C, Hutfless SM, Wang Y, David S, Cheng Y, Twaddell WS, Latt NL, Shin EJ, Wang LD, Wang L, Yang W, Velculescu VE, Vogelstein B, Papadopoulos N, Kinzler KW, Meltzer SJ. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. Cancer Discov. 2012; 2:899–905.

Agresti, A. Categorical Data Analysis. 2. John Wiley & Sons, Incorporated; New York: 2002.

Ahn SM, Chan JYK, Zhang Z, Wang H, Khan Z, Bishop JA, Westra W, Koch WM, Califano JA. Saliva and plasma quantitative polymerase chain reaction–based detection and surveillance of human papillomavirus–related head and neck cancer. JAMA Otolaryngol Head Neck Surg. 2014; 140:846–854.

Allison P, Franco E, Black M, Feine J. The role of professional diagnostic delays in the prognosis of upper aerodigestive tract carcinoma. Oral Oncol. 1998; 34:147–153.

Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH, Rodriguez FJ, Cahill DP, McLendon R, Riggins G, Velculescu VE, Oba-Shinjo SM, Marie SKN, Vogelstein B, Bigner D, Yan H, Papadopoulos N, Kinzler KW. Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma. Science. 2011; 333:1453–1455.

Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, Antonarakis ES, Azad NS, Bardelli A, Brem H, Cameron JL, Lee CC, Fecher LA,

Gallia GL, Gibbs P, Le D, Giuntoli RL, Goggins M, Hogarty MD, Holdhoff M, Hong SM, Jiao Y, Juhl HH, Kim JJ, Siravegna G, Laheru DA, Lauricella C, Lim M, Lipson EJ, Marie SKN, Netto GJ, Oliner KS, Olivi A, Olsson L, Riggins GJ, Sartore-Bianchi A, Schmidt K, Shih IM, Oba-Shinjo SM, Siena S, Theodorescu D, Tie J, Harkins TT, Veronese S, Wang TL, Weingart JD, Wolfgang CL, Wood LD, Xing D, Hruban RH, Wu J, Allen PJ, Schmidt CM, Choti MA, Velculescu VE, Kinzler KW, Vogelstein B, Papadopoulos N, Diaz LA Jr. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014; 6:224ra224.

Bocsi J, Berczi C, Balazs G, et al. Flow cytometric studies on the relationship between DNA content and clinicopathologic features of pancreatic cancers. Anticancer Res. 1998; 18:1839–1843.

Bosman, FT. World Health Organization. WHO classification of tumours of the digestive system. Lyon: International Agency for Research on Cancer; 2010. International Agency for Research on Cancer.

Bosman FT, Carneiro Ft, Hruban RH, et al. WHO classification of tumours of the digestive system: World Health Organization. 2010.

Boyle JO, Mao L, Brennan JA, Koch WM, Eisele DW, Saunders JR, Sidransky D. Gene mutations in saliva as molecular markers for head and neck squamous cell carcinomas. Am J Surg. 1994; 168:429–432.

Carvalho AL, Pintos J, Schlecht NF, Oliveira BV, Fava AS, Curado MP, Kowalski LP, Franco EL. Predictive factors for diagnosis of advanced-stage squamous cell carcinoma of the head and neck. Arch Otolaryngol Head Neck Surg. 2002; 128:313–318.

Cancer Genome Atlas. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; 517:576–582.

Chaturvedi AK, Engels EA, Anderson WF, Gillison ML. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. J Clin Oncol. 2008; 26:612–619.

Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruse S, Anderson WF, Rosenberg PS, Gillison ML. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011; 29:4294–4301.

Chaturvedi AK, Anderson WF, Lortet-Tieulent J, Curado MP, Ferlay J, Franceschi S, Rosenberg PS, Bray F, Gillison ML. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J Clin Oncol. 2013; 31:4550–4559.

D'Sou za G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, Gillison WHML. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med. 2007; 356:1944–1956.

Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013; 368:1199–1209.

Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, Kinzler KW, Vogelstein B. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. Gastroenterology. 2008; 135:489–498.

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res. 1998; 8:175–185.

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998; 8:186–194.

Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010; 127:2893–2917.

Giacona MB, et al. Cell-free DNA in human blood plasma; length measurments in patients with pancreatic cancer and healthy controls. Pancreas. 1998; 17:89-97.

Gillison ML, Broutian T, Pickard RKL, Tong ZY, Xiao W, Kahle L, Graubard BI, Chaturvedi AK. Prevalence of oral HPV infection in the United States, 2009–2010. JAMA. 2012; 307:693–703.

Gourin CG, Terris DJ. Carcinoma of the hypopharynx. Surg Oncol Clin N Am. 2004; 13:81–98.

Guggenheimer J, Verbin RS, Johnson JT, Horkowitz CA, Myers EN. Factors delaying the diagnosis of oral and oropharyngeal carcinomas. Cancer. 1989; 64:932–935.

Howlader, N.; Noone, AM.; Krapcho, M.; Garshell, J.; Miller, D.; Altekruse, SF.; Kosary, CL.; Yu, M.; Ruhl, J.; Tatalovich, Z.; Mariotto, A.; Lewis, DR.; Chen, HS.; Feuer, EJ.; Cronin, KA., editors. SEER Cancer Statistics Review, 1975–2011. National Cancer Institute; Bethesda, MD: 2014.

Hubers AJ, Prinsen CF, Sozzi G, Witte BI, Thunnissen E. Molecular sputum analysis for the diagnosis of lung cancer. Br J Cancer. 2013; 109:530–537.

Isayeva T, Li Y, Maswahu D, Brandwein-Gensler M. Human papillomavirus in non-oropharyngeal head and neck cancers: A systematic literature review. Head Neck Pathol. 2012; 6(Suppl 1):S104–S120.

Kanda M, Sadakari Y, Borges M, et al. Mutant TP53 in Duodenal Samples of Pancreatic Juice from Patients with Pancreatic Cancer or High-Grade Dysplasia. Clin Gastroenterol Hepatol. 2012.

Kang H, Kiess A, Chung CH. Emerging biomarkers in head and neck cancer in the era of genomics. Nat Rev Clin Oncol. 2015; 12:11–26.

Kinde I, Wu J, Papadopoulos N, et al. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A. 2011; 108:9530–9535.

Kinde I, Papadopoulos N, Kinzler KW, et al. FAST-SeqS: a simple and efficient method for the detection of aneuploidy by massively parallel sequencing. PLoS One. 2012; 7:e41162.

Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2012; 108:9530–9535.

Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih IM, Kurman R, Dao F, Levine DA, Giuntoli R, Roden R, Eshleman JR, Carvalho JP, Marie SK, Papadopoulos N, Kinzler KW, Vogelstein B, Diaz LA Jr. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med. 2013; 5:167ra164.

Koivunen P, Rantala N, Hyrynkangas K, Jokinen K, Alho OP. The impact of patient and professional diagnostic delays on survival in pharyngeal cancer. Cancer. 2001; 92:2885–2891.

Kowalski LP, Carvalho AL. Influence of time delay and clinical upstaging in the prognosis of head and neck cancer. Oral Oncol. 2001; 37:94–98.

Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: A systematic review. Cancer Epidemiol Biomarkers Prevent. 2005; 14:467–475.

Kreimer AR, Bhatia RK, Messeguer AL, González P, Herrero R, Giuliano AR. Oral human papillomavirus in healthy individuals: A systematic review of the literature. Sex Transm Dis. 2010; 37:386–391.

Kuo P, Chen MM, Decker RH, Yarbrough WG, Judson BL. Hypopharyngeal cancer incidence, treatment, and survival: Temporal trends in the United States. Laryngoscope. 2014; 124:2064–2069.

Laffan TA, Horton KM, Klein AP, et al. Prevalence of unsuspected pancreatic cysts on MDCT. AJR Am J Roentgenol. 2008; 191:802–807.

Law JK, Ahmed A, Singh VK, et al. A systematic review of solid-pseudopapillary neoplasms: are these rare lesions? Pancreas. 2014; 43:331–337.

Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, Antipova A, Lee C, McKernan K, De La Vega FM, Kinzler KW, Vogelstein B, Diaz LA Jr, Velculescu VE. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med. 2010; 2:20ra14.

Lee KS, Sekhar A, Rofsky NM, et al. Prevalence of incidental pancreatic cysts in the adult population on MR imaging. Am J Gastroenterol. 2010; 105:2079–2084.

Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. 2011; 11:9–22.

Lennon AM, Wolfgang C. Cystic neoplasms of the pancreas. J Gastrointest Surg. 2013; 17:645–653.

Lennon AM, Manos LL, Hruban RH, et al. Role of a Multidisciplinary Clinic in the Management of Patients with Pancreatic Cysts: A Single-Center Cohort Study. Ann Surg Oncol. 2014.

Lennon AM, Wolfgang CL, Canto MI, et al. The Early Detection of Pancreatic Cancer: What Will It Take to Diagnose and Treat Curable Pancreatic Neoplasia? Cancer Res. 2014; 74:3381–3389.

Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1997; 37:646-650.

Li R, Faden DL, Fakhry C, Langelier C, Jiao Y, Wang Y, Wilkerson MD, Pedamallu CS, Old M, Lang J, Loyo M, Ahn SM, Tan M, Gooi Z, Chan J, Richmon J, Wood LD, Hruban RH, Bishop J, Westra WH, Chung CH, Califano J, Gourin CG, Bettegowda C, Meyerson M, Papadopoulos N, Kinzler KW, Vogelstein B, DeRisi JL, Koch WM, Agrawal N. Clinical, genomic, and metagenomic characterization of oral tongue squamous cell carcinoma in patients who do not smoke. Head Neck. 201410.1002/hed.23807

Lingen MW, Xiao W, Schmitt A, Jiang B, Pickard R, Kreinbrink P, Perez-Ordonez B, Jordan RC, Gillison ML. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. Oral Oncol. 2013; 49:1–8.

Martignetti JA, Camacho-Vanegas O, Priedigkeit N, Camacho C, Pereira E, Lin L, Garnar-Wortzel L, Miller D, Losic B, Shah H, Liao J, Ma J, Lahiri P, Chee M, Schadt E, Dottino P. Personalized ovarian cancer disease surveillance and detection of candidate therapeutic drug target in circulating tumor DNA. Neoplasia. 2014; 16:97–103.

Masica DL, Karchin R. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. Cancer Res. 2011; 71:4550–4561.

Masica DL, Karchin R. Collections of simultaneously altered genes as biomarkers of cancer cell drug response. Cancer Res. 2013; 73:1699–1708.

Masica DL, Dal Molin M, Wolfgang CL, et al. A novel approach for selecting combination clinical markers of pathology applied to a large retrospective cohort of surgically resected pancreatic cysts. J Am Med Inform Assoc. 2017; 24:145-152.

Newman AM, Bratman SV, To J, Wynne JF, Eclov NCW, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrager JB, Loo BW Jr, Alizadeh AA, Diehn M. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014; 20:548–554.

Overman MJ, et al. Use of research biopsies in clinical trials: are risks and benefits adequately discussed? J Clin Oncol. 2013; 31:17-22.

Patel SC, Carpenter WR, Tyree S, Couch ME, Weissler M, Hackman T, Hayes DN, Shores C, Chera BS. Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years. J Clin Oncol. 2011; 29:1488–1494.

Pfister DG, Spencer S, Brizel DM, Burtness B, Busse PM, Caudell JJ, Cmelak AJ, Colevas AD, Dunphy F, Eisele DW, Gilbert J, Gillison ML, Haddad RI, Haughey BH, Hicks WL Jr, Hitchcock YJ, Jimeno A, Kies MS, Lydiatt WM, Maghami E, Martins R, McCaffrey T, Mell LK, Mittal BB, Pinto HA, Ridge JA, Rodriguez CP, Samant S, Schuller DE, Shah JP, Weber RS, Wolf GT, Worden F, Yom SS, McMillian NR, Hughes M. Head and neck cancers, version 2.2014. Clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2014; 12:1454–1487.

Rago C, Huso DL, Diehl F, et al. Serial assessment of human tumor burdens in mice by the analysis of circulating DNA. Cancer Res. 2007; 67:9364–9370.

Ralla B, Stephan C, Meller S, Dietrich D, Kristiansen G, Jung K. Nucleic acid-based biomarkers in body fluids of patients with urologic malignancies. Crit Rev Clin Lab Sci. 2014; 51:200–231.

Sahora K, Mino-Kenudson M, Brugge W, et al. Branch duct intraductal papillary mucinous neoplasms: does cyst size change the tip of the scale? A critical analysis of the revised international consensus guidelines in a large single-institutional series. Ann Surg. 2013; 258:466–475.

Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, Blackford A, Parmigiani G, Diaz LA Jr, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE, Hogarty MD. Integrated genomic analyses identify *ARID1A* and *ARID1B* alterations in the childhood cancer neuroblastoma. Nat Genet. 2013; 45:12–17.

Senore C, Ederle A, Fantin A, Andreoni B, Bisanti L, Grazzini G, Zappa M, Ferrero F, Marutti A, Guiliani O, Armaroli P, Segnan N. Acceptability and side-effects of colonoscopy and sigmoidoscopy in a screening setting. J Med Screen. 2011; 3:128-134.

Sidransky D, Von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, Paul M, Green P, Hamilton SR, Frost P, Volgelstein B. Identification of p53 gene mutations in bladder cancers and urine samples. Science. 1991; 252:706–709.

Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, Vogelstein B. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. Science. 1992; 256:102–105.

Sidransky D. Nucleic acid-based methods for the detection of cancer. Science. 1997; 278:1054–1059.

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015; 65:5–29.

Simard EP, Ward EM, Siegel R, Jemal A. Cancers with increasing incidence trends in the United States: 1999 through 2008. CA Cancer J Clin. 2012; 62:118–128.

Sohn TA, Yeo CJ, Cameron JL, et al. Intraductal papillary mucinous neoplasms of the pancreas: an updated experience. Ann Surg. 2004; 239:788–797. discussion 797–9.

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, Carter SL, Voet D, Cortes ML, Auclair D, Berger MF, Saksena G, Guiducci C, Onofrio RC, Parkin M, Romkes M,

Weissfeld JL, Seethala RR, Wang L, Rangel-Escareno C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Winckler W, Ardlie K, Gabriel SB, Meyerson M, Lander ES, Getz G, Golub TR, Garraway LA, Grandis JR. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011; 333:1157–1160.

Tanaka M, Fernandez-Del Castillo C, Adsay V, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology. 2012; 12:183–197.

Valsangkar NP, Morales-Oyarvide V, Thayer SP, et al. 851 resected cystic tumors of the pancreas: a 33-year experience at the Massachusetts General Hospital. Surgery. 2012; 152:S4–S12

VanderLaan PA, et al. Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. Lung Cancer. 2014; 84:39-44.

Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci USA. 1999; 96:9236–9241.

Vogelstein B, Kinzler, KW. Cancer genes and the pathways they control. Nat Med. 2004; 8:789-799.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013; 339:1546–1558.

Wang Y, Chen M, Xiao N, Liu H. Evaluation and comparison of in vitro degredation kinetics of DNA in serum, urine and saliva: a quantitative study. Gene. 2016; 590:142-148.

Wildt J, Bundgaard T, Bentzen SM. Delay in the diagnosis of oral squamous cell carcinoma. Clin Otolaryngol Allied Sci. 1995; 20:21–25.

Wu J, Jiao Y, Dal Molin M, et al. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. Proc Natl Acad Sci U S A. 2011; 108:21188–21193

Wu J, Matthaei H, Maitra A, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med. 2011; 3:92ra66.

Yamao K, Yanagisawa A, Takahashi K, et al. Clinicopathological features and prognosis of mucinous cystic neoplasm with ovarian-type stroma: a multiinstitutional study of the Japan pancreas society. Pancreas. 2011; 40:67–71.

Curriculum Vitae

Simeon Springer
The Johns Hopkins University · Ludwig Center for Cancer Genetics
Cancer Research Building One · Room 590
1650 Orleans St · Baltimore · MD 21205
(617) 797-0014 · Simeon.Springer@gmail.com

EDUCATION

Ph.D. in Human Genetics, August 2012 – May 2017
Johns Hopkins University, Baltimore, MD

Bachelor of Science, 2004-2008
Skidmore College, Saratoga Springs, NY

RESEARCH EXPERIENCE

Graduate Thesis, Johns Hopkins University, June 2013 – May 2017
Advisors: Bert Vogelstein, Nick Papadopoulos, Ken Kinzler
Thesis: Development of cancer biomarker assays from DNA in various bodily fluids

Research Technician, Massachusetts General Hospital, Cancer Center, 2008-2012
Mentored by: Dr. Daniel A. Haber, Dr. Shyamala Maheswaran, Dr. Brian V. Nahed
Project: Worked with a multidisciplinary team to develop microfluidic devices to isolate and characterize circulating tumor cells (CTCs) from human clinical blood samples

Senior Thesis, Skidmore College, 2007-2008
Advisor: Dr. Paul J. Arciero
Collaborators: Laura Henderson, Miriam Drace
Project: Investigated the effects of high and low glycemic index diets on metabolic factors after sprint interval exercise

Research Technician, Skidmore College, Summer Collaborative Research Project, Summer 2006
Advisor: Dr. T.H. Reynolds
Project: Investigated the role of mTORC1 in regulating the insulin signaling pathway in the skeletal muscle of ob/ob mice

TAUGHT COURSES

Advanced Topics in Human Genetics
Teaching Assistant
The Johns Hopkins University
2014

PUBLICATIONS

"Detection of upper tract urothelial carcinomas by genetic analysis of urinary cell DNA." In preparation, planned submission May 2017.

"Surveillance of bladder cancers from DNA in urine." In preparation, planned submission May 2017.

"Large cohort validation of molecular markers and clinical features for classification of pancreatic cysts." In preparation, planned submission June 2017.

Nguyen D, Taheri D, **Springer S**, Cowan M, Guner G, Mendoza Rodriguez MA, Wang Y, Kinde I, VandenBussche CJ, Olson MT, Ricardo BF, Cunha I, Fujita K, Ertoy D, Kinzler KW, Bivalacqua TJ, Papadopoulos N, Vogelstein B, Netto GJ. "High prevalence of TERT promoter mutations in micropapillary urothelial carcinoma." Vichows Arch. 2016 Oct; 469(4): 427-34.

Wang Y, Sundfeldt K, Mateoiu C, Shih leM, Kurman RJ, Scaefer J, Silliman N, Kinde I, **Springer S,** Foote M, Kristjansdottir B, James N, Kinzler KW, Papadopoulos N, Diaz LA, Vogelstein B. "Diagnostic potential of tumor DNA from ovarian cyst fluid." eLife 2016 Jul;5: e15175

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, Steel M, Tran B, Desai J, Jones I, Haydon A, Hayes T, Price TJ, Strausberg RL, Diaz LA Jr, Papadopoulos N, Kinzler KW, Vogelstein B, Gibbs P. "Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer." Sci Transl Med 2016 Jul 6;8(346):346ra92

Cowan ML, **Springer S**, Nguyen D, Taheri D, Guner G, Mendoza Rodriguez MA, Wang Y, Kinde I, Del Carmen Rodriguez Pena M, VandenBussche CJ, Olson MT, Cunha I, Fujita K, Ertoy D, Kinzler K, Bivalacqua T, Papadopoulos N, Vogelstein B, Netto GJ. "Detection of TERT promoter mutations in primary adenocarcinoma of the urinary bladder." Hum Pathol. 2016 Jul; 53:8-13

Masica DL, Dal Molin M, Wolfgang CL, Tomita T, Ostovaneh MR, Blackford A, Moran RA, Law JK, Barkley T, Goggins M, Irene Canto M, Pittman M, Eshleman JR, Ali SZ, Fishman EK, Kamel IR, Raman SP, Zaheer A, Ahuja N, Makary MA, Weiss MJ, Hirose K, Cameron JL, Rezaee N, He J, Joon Ahn Y, Wu W, Wang Y, **Springer S**, Diaz LL Jr, Papadopoulos N, Hruban RH, Kinzler KW, Vogelstein B, Karchin R, Marie Lennon A. "A novel approach for selecting combination clinical markers of pathology applied to a large retrospective cohort of surgically resected pancreatic cysts." J Am Med Inform Assoc. 2016 Jun;24(1):145-152

Cowan M, **Springer S**, Nguyen D, Taheri D, Guner G, Rodriguez MA, Wang Y, Kinde I, VandenBussche CJ, Olson MT, Cunha I, Fujita K, Ertoy D, Bivalacqua TJ, Kinzler K, Vogelstein B, Netto GJ, Papadopoulos N. "High prevalence of TERT promoter mutations in primary squamous cell carcinoma of the urinary bladder." Mod Pathol. 2016 May;29(5):511-5

Springer S[1], Wang Y[1], Dal Molin M[1], Masica DL[1], Jiao Y, Kinde I, Blackford A, Raman SP, Wolfgang CL, Tomita T, Niknafs N, Douville C, Ptak J, Dobbyn L, Allen PJ, Klimstra DS, Schattner MA, Schmidt CM, Yip-Schneider M, Cummings OW, Brand RE, Zeh HJ, Singhi AD, Scarpa A, Salvia R, Malleo G, Zamboni G, Falconi M, Jang JY, Kim SW, Kwon W, Hong SM, Song KB, Kim SC, Swan N, Murphy J, Geoghegan J, Brugge W, Fernandez-Del Castillo C, Mino-Kenudson M, Schulick R, Edil BH, Adsay V, Paulino J, van Hooft J, Yachida S, Nara S, Hiraoka N, Yamao K, Hijioka S, van der Merwe S, Goggins M, Canto MI, Ahuja N, Hirose K, Makary M, Weiss MJ, Cameron J, Pittman M, Eshleman JR, Diaz LA Jr, Papadopoulos N, Kinzler KW, Karchin R, Hruban RH, Vogelstein B, Lennon AM. "A combination of molecular markers and clinical features improve the classification of pancreatic cysts." Gastroenterology. 2015 Nov;149(6):1501-10

Wang Y[1], **Springer S**[1]**,** Zhang M[1], McMahon KW, Kinde I, Dobbyn L, Ptak J, Brem H, Chaichana K, Gallia GL, Gokaslan ZL, Groves ML, Jallo GI, Lim M, Olivi A, Quinones-Hinojosa A, Rigamonti D, Riggins GJ, Sciubba DM, Weingart JD, Wolinsky JP, Ye X, Oba-Shinjo SM, Marie SK, Holdhoff M, Agrawal N, Diaz LA Jr, Papadopoulos N, Kinzler KW, Vogelstein B, Bettegowda C. "Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord." Proc Natl Acad Sci USA 2015 Aug 4;112(31):9704-9

Wang Y[1], **Springer S**[1]**,** Mulvey CL, Silliman N, Schaefer J, Sausen M, James N, Rettig EM, Guo T, Pickering CR, Bishop JA, Chung CH, Califano JA, Eisele DW, Fakhry C, Gourin CG, Ha PK, Kang H, Kiess A, Koch WM, Myers JN, Quon H, Richmon JD, Sidransky D, Tufano RP, Westra WH, Bettegowda C, Diaz LA Jr, Papadopoulos N, Kinzler KW, Vogelstein B, Agrawal N. "Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas." Sci Transl Med. 2015 Jun 24;7(293):293ra104

Kinde I, Munari E, Faraj SF, Hruban RH, Schoenberg M, Bivalacqua T, Allaf M**, Springer S**, Wang Y, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N, Netto GJ. "TERT promoter

mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine." Cancer Res. 2013 Dec 15;73(24):7162-7

Springer S, Yi KH, Park J, Rajpurohit A, Price AJ, Lauring J. "Engineering targeted chromosomal amplifications in human breast epithelial cells." Breast Cancer Res Treat. 2015 Jul; 152(2):313-21

Sullivan JP, Nahed BV, Madden MW, Oliveira SM**, Springer S**, Bhere D, Chi AS, Wakimoto H, Rothenberg SM, Sequist LV, Kapur R, Shah K, Iafrate AJ, Curry WT, Loeffler JS, Batchelor TT, Louis DN, Toner M, Maheswaran S, Haber DA. "Brain tumor cells in circulation are enriched for mesenchymal gene expression." Cancer Discov. 2014 Nov;4(11):1299-309

Stott SL, Lee RJ, Nagrath S, Yu M, Miyamoto DT, Ulkus L, Inserra EJ, Ulman M, **Springer S**, Nakamura Z, Moore AL, Tsukrov DI, Kempner ME, Dahl DM, Wu CL, Iafrate AJ, Smith MR, Tompkins RG, Sequist LV, Toner M, Haber DA, Maheswaran S. "Isolation and characterization of circulating tumor cells from patients with localized and metastatic prostate cancer." Sci Transl Med. 2010 Mar 31;2(25):25ra23

Stott SL, Hsu CH, Tsukrov DI, Yu M, Miyamoto DT, Waltman BA, Rothenberg SM, Shah AM, Smas ME, Korir GK, Floyd FP Jr, Gilman AJ, Lord JB, Winokur D, **Springer S**, Irimia D, Nagrath S, Sequist LV, Lee RJ, Isselbacher KJ, Maheswaran S, Haber DA, Toner M. "Isolation of circulating tumor cells using a microvortexgenerating herringbone-chip." Proc Natl Acad Sci U S A. 2010 Oct 26;107(43):18392-7

[1] Denotes equal contribution to work