Published in final edited form as:

Cancer Epidemiol. 2012 February ; 36(1): e33–e39. doi:10.1016/j.canep.2011.06.003.

Analysis of microsatellite mutations in buccal cells from a case-control study for lung cancer

Jessica Baumann^{1,5}, Ming Li^{2,5}, Aslak Poulsen⁷, Nicholson S. Chadwick^{1,5}, Qiuyin Cai^{5,6}, Christine H. Chung^{1,3,5}, Yu Shyr^{2,5}, Jørgen H. Olsen⁷, Wei Zheng^{3,5,6}, and Robbert J.C. Slebos^{1,4,6,*}

¹Department of Cancer Biology, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

²Department of Biostatistics, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

³Department of Medicine, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

⁴Jim Ayers Institute for Precancer Detection and Diagnosis, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

⁵Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

⁶Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine; Nashville, TN 37232, USA

⁷Department of Biostatistics and Epidemiology, Institute of Cancer Epidemiology, The Danish Cancer Society, Copenhagen, Denmark

Abstract

Exposure to tobacco carcinogens is the major cause of human lung cancer, but even heavy smokers have only about a 10% life-time risk of developing lung cancer. Currently used screening processes, based largely on age and exposure status, have proven to be of limited clinical utility in predicting cancer risk. More precise methods of assessing an individual's risk of developing lung cancer are needed. Because of their sensitivity to DNA damage, microsatellites are potentially useful for the assessment of somatic mutational load in normal cells. We assessed mutational load using hypermutable microsatellites in buccal cells obtained from lung carcinoma cases and controls to test if such a measure could be used to estimate lung cancer risk. There was no significant association between smoking status and mutation frequency with any of the markers

© 2011 Elsevier Ltd. All rights reserved

* Author to whom correspondence should be addressed at: Department of Cancer Biology, Vanderbilt University School of Medicine, U1213A Medical Research Building III, 465 21st Avenue South, Nashville, TN 37232-8575; Phone 615 936-2806, FAX 615 343-8372; r.slebos@vanderbilt.edu.

<baumann003@gmail.com>

<ming.li@Vanderbilt.Edu>

<aslak@cancer.dk>

<nicholson.s.chadwick@gmail.com>

<qiyin.cai@Vanderbilt.Edu>

<cchung11@jhmi.edu>

<yu.shyr@Vanderbilt.Edu>

<jorgen@cancer.dk>

<wei.zheng@Vanderbilt.Edu>

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

tested. No significant association between case status and mutation frequency was observed. Age was significantly related to mutation frequency in the microsatellite marker D7S1482. These observations indicate that somatic mutational load, as measured using mutation frequency of microsatellites in buccal cells, increases with increasing age but that subjects who develop lung cancer have a similar mutational load as those who remain cancer free. This finding suggests that mutation frequency of microsatellite mutations in buccal cells may not be a promising biomarker for lung cancer risk.

Keywords

microsatellite repeat; lung carcinoma; small pool PCR

Introduction

Lung cancer accounts for approximately 15% of all new cancer diagnoses in the United States, and is a leading cause of world-wide cancer death in both men and women. Nearly 90% of lung cancer deaths are attributable to exposure to tobacco carcinogens, yet the lifetime risk of developing a tobacco-related cancer in a heavy smoker is less than 10% [1], indicating that within the population of long-term smokers there is variation in risk based on genetic background, environmental insults to genomic integrity and/or other factors. Established factors associated with an individual's risk for lung cancer besides tobacco exposure include age, family or personal history of lung cancer, occupational exposure to carcinogens, and diagnosis of prior precancerous lesions of the respiratory tract. These parameters, however, are of limited clinical utility in assessing the risk of cancer development. Polymorphisms of carcinogen-metabolizing enzymes or DNA repair genes may influence cancer risk at the molecular level, although the effects are generally subtle [2]. Studies of host DNA repair capacity as measured by degree of chromosomal damage [3–6] have provided compelling evidence for a link between defective DNA repair and cancer development. However, the technical skill, laboratory resources, and high degree of human error associated with these assays limits their applicability as screening tools in most clinical settings [7].

The development of small-pool PCR has allowed for the exploration of an alternative avenue for the quantification of genomic instability in carcinogen-exposed tissues: the measurement of DNA mutations in microsatellite repeats [8, 9]. Though typically neutral, microsatellite DNA repeats can result in slipped strand mispairing during DNA replication, and generally have higher mutation rates than non-repetitive DNA [10]. This characteristic of microsatellites makes them an especially attractive biomarker for evaluating genetic susceptibility to carcinogen exposure. Specific tetranucleotide repeats demonstrate high levels (average mutation rate $1-10 \times 10^{-3}$ per cell division) of instability in the germline [11], and thus may be sensitive to the effects of exogenous sources of host DNA damage. Particularly useful hypermutable tetranucleotide repeats are MycL1 [12], DXS981 [13, 14] and D7S1482 (UT5085/L17686) [15, 16]. When studied in a model system, the corresponding repetitive sequences demonstrated increased sensitivity to DNA damage caused by carcinogen exposure [17]. Furthermore, we, and others, have linked advancing age with increased levels of microsatellite mutations, either in normal human blood lymphocytes [9] or normal buccal cells [18]. Taken together, results from these studies suggest that microsatellites are exquisitely susceptible to damage, and that tetranucleotide repeats may be particularly sensitive to the tobacco-related mutagens implicated in most lung malignancies. If differences in mutational load could be established in lung cancer cases compared to cancer-free individuals with similar exposure status, the high-throughput

screening of microsatellite mutations could conceivably be developed into a practical, specific clinical assay for cancer risk.

The exposure of a large tissue area to mutagens is thought to result in the development of many distinct genetic abnormalities, each with malignant potential, a process known as *field cancerization*. This concept was originally proposed after the observance of multiple metaplastic and dysplastic lesions in the oral epithelium of smokers that seemed to arise independently from one another [19]. We, and others, later showed that the mutations characteristic of these precancerous lesions are also found in overt cancers [20–22]. This process explains the high incidence rates (3–5% per year) for the development of second primary cancers in patients who have been cured of a first lung- or head and neck carcinoma [23, 24]. Because microsatellite mutations caused by environmental insult can be screened for in a rapid, reproducible manner they are a sensitive measure for the somatic mutational load expected in a 'cancerized field' of normal human oral cells. Evidence of a unique mutagenic profile in carcinogen-exposed epithelium, combined with the ease of sampling such cells via oral rinses, made the use buccal cells for our assay an attractive means of testing the utility of screening for microsatellite mutations in a manner that would be clinically feasible.

In this study, we analyzed mutations in a select panel of tetranucleotide markers in buccal cells obtained as part of a case-control study for lung cancer and tested for an association between mutations in these markers and lung cancer status.

Methods

Biological materials

Samples for this study were obtained as part of a larger case-control lung cancer study in North Jutland County, Denmark, as part of the Vanderbilt SPORE in Lung Cancer. In a *prospective* sample collection, patients with suspected lung cancers were approached *prior* to the establishment of a cancer diagnosis and invited to provide a mouth rinse sample for DNA extraction. Of the patients who accepted, 50 case-control pairs were selected after establishment of a definite diagnosis and these were frequency-matched on sex, age (+ 2 years) and smoking status, classified as 1) never smoker, 2) former smoker (>5 years since quitting) and 3) current smoker (current or <5 year since quitting). There were 52 males and 48 females in the study: 6 never smokers, 14 former smokers and 80 current smokers (Table 1). DNA for small-pool polymerase chain reaction (SP-PCR) was isolated at the Survey and Biospecimen Shared Resource at the Vanderbilt-Ingram Cancer Center using the QIAmp DNA kit (Qiagen, Valencia CA) following the manufacturer's protocol and stored at –80°C in a solution of 5 ng/μl until use.

Detection of microsatellite mutations

The analysis strategy for the detection of microsatellite mutations was based on small-pool PCR whereby a known amount of DNA is diluted down to the single molecule level (approximately 9 pg/PCR, or 3 genome equivalents per well), from which specific amplicons are expanded using time-release PCR [18, 25]. In a single assay, hundreds of alleles are tested simultaneously and microsatellite mutations are visualized due to differences in retention time during capillary electrophoresis. In this study, we used the tetranucleotide markers MycL1, D7S1482, and DXS891, which accumulate high mutation levels following environmental insult and can be measured reliably by DNA fragment analysis. Detailed experimental procedures and reproducibility for the markers used in this study (MycL1, D7S1482 and DXS981) have been described previously [18, 26]. In summary, 3 serial dilutions of 1:100, 1:1000 and 1:10,000 from the 5 ng/μl solution were

each amplified in 48 wells using marker MycL1 and the number of blank wells was used to estimate the average number of molecules in each of the wells using a standard Poisson distribution. In the 1:1000 dilution, each well is expected to harbor the equivalent of approximately 2 genome equivalents (6 pg/well), but in practice this number varied substantially, presumably due to contamination of non-human DNA present in some of the mouth rinse samples. The estimate was used to create a solution containing an estimated 3 human genome equivalents per well and tested for each marker in 96 wells. Experiments that had an average of more than 4 molecules per well in a 96-well plate were excluded to retain sensitivity in the ability to separate mutated from normal alleles. Using information from this experiment, an improved estimate of the average genome equivalent per well could be made and this number was used to calculate the total number of molecules tested. If the number of marker molecules tested was below 100, an additional set of amplifications was started, adjusting for the improved estimate if necessary. This procedure was repeated until a minimum number of 300 genome equivalents were tested for the sum of all 3 markers (100 equivalents per marker), except for 3 cases where low DNA concentration limited the analysis to around 250 genome equivalents. The collection of 300 genome equivalents allows for the ability to detect mutation frequencies as low as 1%, and the use of three separate markers improves the likelihood that one or more markers may be able to uniquely characterize how mutation status relates to age, smoking status, and/or cancer status in cases versus controls. The median for all 100 samples was 483 tested genome equivalents with a range of 222 to 1303. Since DXS981 is an X-linked marker, this number was about half that in males. Scoring of microsatellite mutations was performed manually by examining of all of the capillary electrophoresis trace graphs using GeneMapper software, which uses peak height as a primary outcome measurement according to previously established criteria [18]. To distinguish PCR-induced slippage artifacts from true mutant molecules the signal from a given fragment length needed to be more than 50% that of the signal from the adjacent larger fragment (to the right) if present, and/or more than 10% of the signal from the adjacent smaller fragment (to the left) if present.

The primary endpoint for assessing mutational load was *overall mutation frequency*, defined as the number of observed abnormal DNA fragments across all markers as a fraction of the total number of genome equivalents tested. This total was estimated from the number of non-amplifying test tubes using the Poisson distribution. There are 2 possible explanations for abnormal DNA fragments in a given sample: either as an independent mutational event occurring in a single cell or as a linked mutational event arising from clonal outgrowth of a cell harboring the specific mutation due to field cancerization. The latter type of mutation is expected to be observed more than once in a given specimen and we can-not distinguish between independent and linked mutational events in this case. Thus, we also record the number of distinct DNA fragment sizes observed and, with overall genome equivalents, calculate our secondary endpoint as *unique mutation frequency*, since they can only be explained by single, independent mutational events.

Analysis and germline classification of microsatellite mutations

One essential part of determining mutant status of any allele is to first establish the normal (“wild-type”) germline fragment lengths for each of the markers for all subjects. A reliable way to accomplish this is to test larger quantities of DNA obtained from non-exposed tissues, such as lymphocyte DNA isolated from blood. This was not possible for this study since buccal cell DNA was the only biological material consistently available from all subjects. We therefore determined the normal germline pattern using larger quantities of buccal DNA (approximately 1 ng, or 350 genome equivalents) so that the expected fragment length pattern becomes apparent. The frequency of fragment sizes was consistent for MycL1 and DXS981 in all subjects, who had either one or two germline fragment sizes, but

fragment length patterns in D7S1482 were only consistent in 93 of 100 subjects. The remaining 7 subjects had similar intensities of three fragment sizes (4 cases and 3 controls) using D7S1482. The most likely explanation for this observation is that apart from the two allele sizes from the two chromosome 7 copies, buccal cell DNA from these subjects was dominated by the clonal outgrowth of a cell harboring a 3rd frequent fragment size. In such cases, it is impossible to determine which of the 3 fragment sizes represented the 2 normal germline sizes and which represented the fragment attributable to clonal outgrowth. In such cases, we arbitrarily assigned the least frequent of the three fragment sizes as the non-germline, mutant fragment from the cumulative measurements of all single-molecule analyses done for these subjects. Because of this uncertainty, all statistical tests were done either by including or excluding these 7 subjects in the analysis. However, excluding these 7 subjects did not result in any material change in test results for any of the statistical tests performed and thus only results from analyses on the full series of 100 patients are presented.

Statistical analyses

The statistical procedures for the analysis of microsatellite mutations were previously described in detail [18]. Our procedures are based on the assumption that the event of a microsatellite mutation is independent between subjects and that the rate of mutation is constant over time. The number of mutant alleles has a Poisson distribution given by $P(Y=y) = \mu^y * e^{(-\mu)} / y!$ where y is the number of mutant alleles observed and μ is the expected count of mutant alleles. This expected value, μ , is the product of the mutant rate and the total number of alleles. A generalized linear model (GLM) using Poisson regression was adopted to model the effects of age, smoking and other parameters. To compensate for varying amounts of human DNA in buccal samples, we used the number of blanks per 96-well plate to estimate the number of genome equivalents in each of the samples as described in detail previously [18]. With the estimated total number of alleles and the number of mutant alleles counted, we applied GLM to examine the effects of age, sex, cancer status, and smoking on mutation frequencies. Both, univariate and multivariate analyses were performed. In addition, we created logistical models to predict cancer or control status based on mutation rates and the other co-variates. We considered significance at the 0.05 level.

Results

To study the possible differences in somatic mutational load between cancer cases and controls, we obtained DNA isolated from buccal cells from a total of 50 lung cancer cases and 50 healthy controls. The subject pairs were frequency matched for age, sex and smoking history (Table 1). The buccal samples were obtained from patients suspected of lung carcinoma prior to the establishment of a definite diagnosis. All samples were tested for microsatellite mutations using SP-PCR for three complex tetranucleotide markers: D7S1482, DXS981 and MycL1.

Microsatellite mutation analysis

The median number of single molecules tested for each subject was 483 (range: 222–1303). An illustration of representative results of a small-pool diluted sample for the marker D7S1482 is shown in Figure 1. Two types of mutations were calculated: *overall mutations* where every single occurrence of a particular mutation was counted and *unique mutations* where only the number of different fragment sizes was counted, irrespective of the number of times a particular mutation was observed (see methods). Clonal expansion of a mutant cell can cause multiple occurrences of a mutant molecule and thus be counted multiple times as *overall mutants*, we also determined *unique mutants* which because of their size differences are most likely independent mutational events. Taken together, the highest

mutation frequencies for both overall and unique mutations were observed in D7S1482, followed by DXS981 and then MycL1 (Table 1).

Univariate analyses of microsatellite mutations

We performed univariate analyses of overall and unique mutation frequency in the three tetranucleotide markers for age, sex, smoking and case/control status. Overall mutation frequency in D7S1482 increased with age (see below), but the other two markers did not show this increase (Table 2). Unique mutation frequency did not appear to increase with age, although the power of this comparison was limited by the lower number of observed values. Univariate analysis for overall mutation frequency by sex was not significant for D7S1482 and MycL1 and borderline significant for DXS981 ($p=0.077$), while unique mutation frequency did reach significance in DXS981 ($p=0.045$). This result is complicated by the fact that different numbers of alleles are tested in males and females because of the difference in X-chromosome numbers. This aspect is further explored under multivariate analyses. As in our previous studies on head and neck carcinoma, there was no significant association between smoking status and mutation frequency with any of the markers tested [18].

Microsatellite mutations in D7S1482 are associated with aging

Previous work has suggested that microsatellite mutations increase in frequency with increasing age [9, 18]. In univariate analysis, the summed value of all mutations in all 3 markers was positively correlated with aging. This effect was driven mainly by results obtained with marker D7S1482 and univariate analysis with just this marker demonstrated a statistically significant correlation of overall mutation frequency with age ($p=0.027$). The point estimate for this increase was 0.023, or approximately a 2% increase in mutation frequency for each year of life. However, this estimate was less than 1% per year for unique mutation frequency ($p=0.298$), presumably because of the much lower number of possible mutation positions in comparison to overall mutation frequency for this marker. In a multivariate analysis that included case status, age, smoking history and sex, the effects of aging on overall D7S1482 mutation frequency remained significant ($p=0.044$). The same univariate and multivariate analyses for the other two markers, MycL1 and DXS981, did not reach statistical significance.

Multivariate analysis of microsatellite mutations and lung cancer status

The primary goal for this study was to determine if somatic mutational load as measured by microsatellite mutations was elevated in normal buccal cells obtained from lung cancer patients compared to matched healthy controls. The presence of such an association would indicate a possible use of this measurement as a means of predicting cancer risk. We first approached this question with a logistic regression model using mutation frequency, age, sex and smoking as predictor variables, and with a GLM as described above. None of the predictor variables, including overall and unique mutation frequency, was capable of predicting case/control status in the logistic regression model. We then expanded our univariate GLM analyses to multiple factors to determine if case/control status was associated with significant differences in mutation frequencies. It is necessary to incorporate multiple variables in the model because of the strong effect of age on mutation frequency in the marker with the highest number of mutations (D7S1482) and possible other confounding effects. In multivariate analyses, we noted no significant association in overall mutation frequency in MycL1 or D7S1482 between cases and matched controls but the marker DXS891 was significant at $p=0.021$. A similar finding was observed with unique mutation frequency: no association with MycL1 or D7S1482, but a significant association with DXS981 ($p=0.019$). In the same models, age was significantly related only to overall mutations in marker D7S1482 ($p=0.044$), but not with unique mutations, while sex was

significantly associated with overall and unique mutations in marker DXS981 ($p=0.041$ and $p=0.024$, respectively). Analysis of marker DXS981 is complicated by the fact that unequal numbers of alleles are present in males and females. We therefore performed a separate analysis for females only. In this model, DXS981 was not significantly associated with case status when tested for overall mutation frequency or unique mutation frequency. Thus, the observed association was probably biased because of differences in observed mutation frequencies between males and females.

Discussion

Lung cancers are among the most commonly diagnosed malignancies and they are associated with significant morbidity and mortality. The main risk factor, cigarette smoking, is associated with a process called “field cancerization” first proposed by Slaughter et al in 1953 [19]. In this process, normal airway epithelial cells become primed for malignant transformation through the accumulation of genetic damage, eg. “somatic mutational load”. We hypothesized somatic mutational load is detectable through the analysis of single-molecule mutation analysis and that the level of this load correlates with lung cancer risk. The hypothesis is that cancer cases have higher levels of mutations in normal cells than controls with similar demographics and exposure histories. Because of the challenges microsatellites pose to the DNA replication and repair machinery, they provide an attractive target for assessing somatic mutational load. When assessed in larger cell populations, these mutations are detectable as microsatellite instability (MSI), whereas our small-pool PCR approach is capable of characterizing mutations at the individual molecule level which allows precise quantitation of the mutational load. The ability to measure microsatellite mutations in cells obtained from buccal rinses [18], sputum samples [27], other noninvasive sampling techniques and from bronchial washings [28], allows the study of genomic alterations in tissues exposed to tobacco carcinogens. This study is the first of its kind to evaluate how mutations in the highly mutable tetranucleotide markers MycL1, D7S1482 and DXS981 relates to age, smoking status, and cancer status in DNA obtained from buccal rinses of patients suspected of lung carcinoma.

The main objective in this study was to determine if the frequency of microsatellite mutations is elevated in normal tissue obtained from lung cancer patients compared to matched controls from the identical patient group at high risk for smoking-associated disease. If such a difference in the pattern of microsatellite mutations existed, it would be invaluable in identifying those individuals who are most likely to develop lung cancer. Such individuals would then be candidates for aggressive monitoring of possible development of lung cancer.

There has been extensive examination of mutations in microsatellite markers that lie in close proximity to chromosomal regions associated with lung cancer, which seems to suggest a relationship between microsatellite instability and cancer status. Microsatellite instability in the 2p and 3p regions was first linked to poorer prognosis in NSCLC more than a decade ago [29, 30] and microsatellite instability (MSI) at other sites in these tumors was subsequently connected to an increased frequency of lymph node metastases [31]. A more recent NSCLC study using 5 markers in the 3p region found MSI at significantly higher rates in those with lung cancer compared to control cases [32]. A strong relationship between cancer status and MSI has also been noted in studies of DNA from laryngeal squamous cell carcinoma (LSCC) [33] and esophageal carcinoma [34], the latter of which indicated that a higher degree of MSI may correlate to more aggressive cancers. High levels of MSI in COPD, a disease that often results from similar carcinogen exposure as do malignancies of the lung, has also been linked to significantly higher overall exacerbation frequencies [35]. Although not strictly identical to MSI, our previous work suggested a

possible association between microsatellite mutations in buccal cells and head and neck cancer status [18].

Here we present an assessment of microsatellite mutations in a carefully selected case-control population for lung cancer. Our results do not provide evidence for higher microsatellite mutations in cancer patients versus controls, although our earlier observation of an increase with aging was confirmed. Logistical regression modeling demonstrated that none of the variables (mutational frequency, age, sex, and smoking status) was capable of predicting case/control status, while in multivariate analysis an apparent increase in DXS981 mutation frequency was presumably biased by male-female differences in X-chromosome copy numbers. The multivariate model was both valid and necessary given the strong effect of age and other confounding variables on mutational frequency. To assess the detection limits of our approach we used Poisson regression modeling of mutation counts and determined that a population of 100 individuals yields >80% power to detect a 2-fold difference in cancer risk between the case and control groups at 0.05 significance. Thus, it is unlikely that we would have missed any clinically relevant difference in lung cancer risk using microsatellite mutations in normal buccal cells.

Because sampling was conducted prior to any curative treatment that may have falsely elevated mutational frequency (i.e. chemotherapy or radiation) and tissue was obtained from an area outside that of the primary cancer, we can be confident that the elevated mutational load is an accurate reflection of somatic response to carcinogen exposure, i.e. the mechanisms that may have made cases more likely to develop the cancer compared to controls.

Our study differed from most others in that it only analyzed *normal* buccal epithelial cells versus the changes that are detectable in *cancer* cells. Moreover, many of the previous studies were performed using mono- and dinucleotide repeats, which can be problematic for the identification of mutant alleles. Mononucleotide slippage events differ from the normal allele by only one or two nucleotides, a small change that can be challenging to detect accurately. In contrast, tetranucleotide repeats separate well by capillary electrophoresis and can be scored individually using simple rules as outlined in the methods and reference [18]. We chose to study buccal cells because of the relative ease of collection for possible screening tests but it remains possible that somatic mutational load in bronchial epithelial cells more closely reflects cancer risk. However, the procedures to obtain such cells (bronchial washing during bronchoscopy) are not feasible as a general screening tool.

We also studied a possible direct correlation between the degree of carcinogen exposure and the degree of somatic mutational load. While the molecular pathogenesis of lung cancer in smokers has been extensively examined, the pathogenesis of lung cancer in nonsmokers has not received similar attention, and thus the identification of the differences in carcinogenic profile between the two is an area of intense interest. While there is some evidence that the frequency of microsatellite alterations is related to the degree of tobacco use [32], evidence arguing against a direct causal relationship between carcinogen exposure and MSI also exists. This includes a study on chronic obstructive pulmonary disease (COPD) that, while finding no mutations in healthy nonsmokers, also noted no statistically significant association between smoking status and MSI [35]. An investigation of squamous cell carcinoma of the lung concluded that MSI did not seem to be related to cumulative smoking exposure or smoking status [28]. Our study also failed to find any significant correlation between smoking status and mutation frequency in normal buccal epithelium. Thus, the level of microsatellite mutations and its likelihood to favor field cancerization and clonal outgrowth appears to be a reflection of factors independent of tobacco exposure. Such factors may include sex, genetic background, other exposures, age of the individual and

possibly other (unknown) factors. From the present data, it appears unlikely that microsatellite mutations as a measure of somatic mutational load is associated with the likelihood of developing lung carcinoma.

In our study, the tetranucleotide marker D7S1482 (which had both the highest overall and highest unique number of mutations of the three markers) demonstrated a statistically significant correlation between overall mutation frequency and age in both univariate analysis and a multivariate analysis that included case status, age, smoking history and sex. These results are in line with a similar association between age and mutation frequency in this tetranucleotide repeat that we noted previously in patients curatively treated for HNSCC [18], and thus confirms the significance of D7S1482 as a highly-mutable marker with increasing age. The association between microsatellite instability and increasing age has been established in studies of normal tissue and more recently in malignancies. Both CD4 and CD8 T cells from healthy older subjects have been observed to have significantly higher rates of MSI compared to younger individuals [9, 36] and it has been suggested that this increase in MSI with age may indicate overall genomic instability in the elderly [37]. MSI has been detected at a higher frequency in those of older age who develop gastric carcinomas [38–40] and lymphomas originating in gastric tissue [41], and has been linked to thyroid cancer in the elderly [42]. It is possible that increased mutation levels become fixed in the genome of the cancer cells and that age differences are thus revealed when analyzing tumor tissues. Our findings of an association between mutational frequency in the microsatellite marker D7S1482 and aging validates this marker as a highly sensitive measure of somatic mutational accumulation throughout the lifespan. The reason for a lack of a significant association between aging and our other markers MycL1 and DXS981 in multivariate analysis is unclear, but lower mutation rates in these markers may limit the sensitivity to detect age-related changes in mutation frequency. A previous study utilizing MycL1 in addition to 3 other microsatellite repeat sequences found it to have the lowest mutation rate in DNA mismatch repair (MMR)-deficient cells. MMR defects are caused by inactivation of genes involved in repair of single nucleotide errors that may occur during both normal DNA replication and in response to carcinogenic insult. However, in MMR-proficient cells, the relative rates were reversed, and MYCL1 was the least stable of the markers. The authors suggested that while MYCL1 and other tetranucleotide regions have lower overall mutation rates during replication but that errors in these repeats are corrected less efficiently than those in dinucleotide repeats [12]. Our observation of a significant correlation between age and mutation frequency in D7S1482 may indicate an increased mutability of this marker compared to MycL1 and DXS981 that results in significant mutation frequencies in the presumed presence of a functional MMR system.

Of final note, univariate analysis of mutation frequency by sex was borderline significantly different for overall mutations, and significantly different for unique mutations for the marker DXS981. In subsequent multivariate analysis, sex was significantly associated with both overall and unique mutations. The most likely explanation for this observation is that it reflects differences in the chance of finding a mutation if it is truly there between males and females. In males, we expect to see only a single DXS981 fragment size and hence, all of the possible mutant allele will be apparent by fragment size analysis. In females with two alleles of different sizes, a mutation that results in a new fragment size that is equal to the existing other allele will be missed. With the limited range of DXS981 fragment sizes compared to the other two markers, the proportion of missed mutant fragments might be sufficient to create an apparent difference in mutation frequencies in males and females, such as the apparent difference in DXS981 mutations by sex (Table 2). The fact that unique mutation frequency is more sensitive to this phenomenon also argues in favor of this explanation. In a separate analysis of only females, overall and unique mutation frequency in DXS981 was not significantly associated with case status.

In conclusion, we here report on the detection of microsatellite mutations by SP-PCR to buccal rinses of patients suspected of lung carcinoma prior to diagnosis. We reaffirm findings by us and others, indicating a significant correlation between increasing age and increasing overall mutational frequency in the marker D7S1482. However, our data do not support the hypothesis that microsatellite mutations as markers for somatic mutational load are associated with risk for lung cancer.

Acknowledgments

The research presented in this paper was supported by NIH Grants R03CA123571 and P50CA90949.

Literature

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin.* 2009; 59:225–49. [PubMed: 19474385]
2. Liu G, Zhou W, Christiani DC. Molecular epidemiology of non-small cell lung cancer. *Semin Respir Crit Care Med.* 2005; 26:265–72. [PubMed: 16052427]
3. Wu X, Gu J, Amos CI, Jiang H, Hong WK, Spitz MR. A parallel study of in vitro sensitivity to benzo[a]pyrene diol epoxide and bleomycin in lung carcinoma cases and controls. *Cancer.* 1998; 83:1118–27. [PubMed: 9740076]
4. Rajae-Behbahani N, Schmezer P, Risch A, Rittgen W, Kayser KW, Dienemann H, et al. Altered DNA repair capacity and bleomycin sensitivity as risk markers for non-small cell lung cancer. *Int J Cancer.* 2001; 95:86–91. [PubMed: 11241317]
5. Diem C, Runger TM. Processing of three different types of DNA damage in cell lines of a cutaneous squamous cell carcinoma progression model. *Carcinogenesis.* 1997; 18:657–62. [PubMed: 9111196]
6. Slebos RJ, Livanos E, Yim HW, Randell SH, Parsons AM, Detterbeck FC, et al. Chromosomal abnormalities in bronchial epithelium from smokers, nonsmokers, and lung cancer patients. *Cancer Genet Cytogenet.* 2005; 159:137–42. [PubMed: 15899385]
7. Berwick M, Vineis P. Markers of DNA Repair and Susceptibility to Cancer in Humans: an Epidemiologic Review. *J Natl Cancer Inst.* 2000; 92:874–897. [PubMed: 10841823]
8. Bacon AL, Farrington SM, Dunlop MG. Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue. *Oncogene.* 2001; 20:7464–71. [PubMed: 11709717]
9. Coolbaugh-Murphy MI, Xu J, Ramagli LS, Brown BW, Siciliano MJ. Microsatellite instability (MSI) increases with age in normal somatic cells. *Mech Ageing Dev.* 2005; 126:1051–9. [PubMed: 16098563]
10. Gordenin DA, Resnick MA. Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat Res.* 1998; 400:45–58. [PubMed: 9685581]
11. Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 2000; 16:551–8. [PubMed: 11102705]
12. Hatch SB, Farber RA. Mutation rates in the complex microsatellite MYCL1 and related simple repeats in cultured human cells. *Mutat Res.* 2004; 545:117–26. [PubMed: 14698421]
13. Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet.* 1992; 2:204–11. [PubMed: 1345170]
14. Mahtani MM, Willard HF. A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. *Hum Mol Genet.* 1993; 2:431–7. [PubMed: 8504304]
15. Temam S, Trassard M, Leroux G, Bosq J, Lubinski B, Lenoir G, et al. Cytology vs molecular analysis for the detection of head and neck squamous cell carcinoma in oesopharyngeal brush samples: a prospective study in 56 patients. *Br J Cancer.* 2003; 88:1740–5. [PubMed: 12771990]

16. Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J Natl Cancer Inst.* 1999; 91:332–9. [PubMed: 10050866]
17. Slebos RJ, Oh DS, Umbach DM, Taylor JA. Mutations in tetranucleotide repeats following DNA damage depend on repeat sequence and carcinogenic agent. *Cancer Res.* 2002; 62:6052–60. [PubMed: 12414628]
18. Slebos RJ, Li M, Vadivelu S, Burkey BB, Netterville JL, Sinard R, et al. Microsatellite mutations in buccal cells are associated with aging and head and neck carcinoma. *Br J Cancer.* 2008; 98:619–26. [PubMed: 18212747]
19. Slaughter D, Southwick H, Smejkal W. “Field cancerization” in oral stratified squamous epithelium: clinical implications of multicentric origin. *Cancer.* 1953; 6:953–968.
20. Westra WH, Baas IO, Hruban RH, Askin FB, Wilson K, Offerhaus GJ, et al. K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. *Cancer Res.* 1996; 56:2224–8. [PubMed: 8616876]
21. Tabor MP, Brakenhoff RH, van Houten VM, Kummer JA, Snel MH, Snijders PJ, et al. Persistence of genetically altered fields in head and neck cancer patients: biological and clinical implications. *Clin Cancer Res.* 2001; 7:1523–32. [PubMed: 11410486]
22. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, et al. Clonal genetic alterations in the lungs of current and former smokers. *J Natl Cancer Inst.* 1997; 89:857–62. [PubMed: 9196251]
23. Thomas P, Rubinstein L. Cancer recurrence after resection: T1 N0 non-small cell lung cancer. Lung Cancer Study Group. *Ann Thorac Surg.* 1990; 49:242–6. discussion 246–7. [PubMed: 2154958]
24. Deleyiannis FW, Thomas DB. Risk of lung cancer among patients with head and neck cancer. *Otolaryngol Head Neck Surg.* 1997; 116:630–6. [PubMed: 9215374]
25. Coolbaugh-Murphy M, Maleki A, Ramagli L, Frazier M, Lichtiger B, Monckton DG, et al. Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. *Genomics.* 2004; 84:419–30. [PubMed: 15234004]
26. Slebos RJ, Little RE, Umbach DM, Antipkin YG, T.D. Z, Mendel NA, et al. Mini- and microsatellite mutations in children from Chernobyl accident cleanup workers. *Mutat Res.* 2004; 559:143–151. [PubMed: 15066582]
27. Castagnaro A, Marangio E, Verduri A, Chetta A, D’Ippolito R, Del Donno M, et al. Microsatellite analysis of induced sputum DNA in patients with lung cancer in heavy smokers and in healthy subjects. *Exp Lung Res.* 2007; 33:289–301. [PubMed: 17694439]
28. Malhotra P, Behera D, Srinivasan R, Majumdar S, Wali A, Mir S, et al. Detection of microsatellite alterations in bronchial washings in squamous cell lung cancer: the first study from India. *Jpn J Clin Oncol.* 2004; 34:439–44. [PubMed: 15371460]
29. Pifarre A, Rosell R, Monzo M, De Anta JM, Moreno I, Sanchez JJ, et al. Prognostic value of replication errors on chromosomes 2p and 3p in non-small-cell lung cancer. *Br J Cancer.* 1997; 75:184–9. [PubMed: 9010024]
30. Rosell R, Pifarre A, Monzo M, Astudillo J, Lopez-Cabrerizo MP, Calvo R, et al. Reduced survival in patients with stage-I non-small-cell lung cancer associated with DNA-replication errors. *Int J Cancer.* 1997; 74:330–4. [PubMed: 9221814]
31. Woenckhaus M, Stoeckl R, Dietmaier W, Wild PJ, Zieglermeier U, Foerster J, et al. Microsatellite instability at chromosome 8p in non-small cell lung cancer is associated with lymph node metastasis and squamous differentiation. *Int J Oncol.* 2003; 23:1357–63. [PubMed: 14532977]
32. Carpagnano GE, Foschino-Barbaro MP, Spanevello A, Resta O, Carpagnano F, Mule G, et al. 3p microsatellite signature in exhaled breath condensate and tumor tissue of patients with lung cancer. *Am J Respir Crit Care Med.* 2008; 177:337–41. [PubMed: 17962633]
33. Zhigang H, Demin H, Han G, Erzhong F, Xiaohong C, Hongbo X. Significance and mechanism of microsatellite instability in laryngeal squamous cell carcinoma. *J Otolaryngol.* 2007; 36:168–74. [PubMed: 17711772]
34. Kubo N, Yashiro M, Ohira M, Hori T, Fujiwara I, Hirakawa K. Frequent microsatellite instability in primary esophageal carcinoma associated with extraesophageal primary carcinoma. *Int J Cancer.* 2005; 114:166–73. [PubMed: 15540218]

35. Makris D, Tzanakis N, Damianaki A, Ntaoukakis E, Neofytou E, Zervou M, et al. Microsatellite DNA instability and COPD exacerbations. *Eur Respir J*. 2008; 32:612–8. [PubMed: 18508815]
36. Krichevsky S, Pawelec G, Gural A, Effros RB, Globerson A, Yehuda DB, et al. Age related microsatellite instability in T cells from healthy individuals. *Exp Gerontol*. 2004; 39:507–15. [PubMed: 15050284]
37. Ben Yehuda A, Globerson A, Krichevsky S, On H. Bar, Kidron M, Friedlander Y, et al. Ageing and the mismatch repair system. *Mech Ageing Dev*. 2000; 121:173–9. [PubMed: 11164471]
38. Corso G, Pedrazzani C, Marrelli D, Pascale V, Pinto E, Roviello F. Correlation of microsatellite instability at multiple loci with long-term survival in advanced gastric carcinoma. *Arch Surg*. 2009; 144:722–7. [PubMed: 19687375]
39. Gu M, Kim D, Bae Y, Choi J, Kim S, Song S. Analysis of microsatellite instability, protein expression and methylation status of hMLH1 and hMSH2 genes in gastric carcinomas. *Hepatogastroenterology*. 2009; 56:899–904. [PubMed: 19621725]
40. Oki E, Kakeji Y, Zhao Y, Yoshida R, Ando K, Masuda T, et al. Chemosensitivity and survival in gastric cancer patients with microsatellite instability. *Ann Surg Oncol*. 2009; 16:2510–5. [PubMed: 19565284]
41. Starostik P, Greiner A, Schwarz S, Patzner J, Schultz A, Muller-Hermelink HK. The role of microsatellite instability in gastric low- and high-grade lymphoma development. *Am J Pathol*. 2000; 157:1129–36. [PubMed: 11021817]
42. Onda M, Nakamura I, Suzuki S, Takenoshita S, Brogren CH, Stampanoni S, et al. Microsatellite instability in thyroid cancer: hot spots, clinicopathological implications, and prognostic significance. *Clin Cancer Res*. 2001; 7:3444–9. [PubMed: 11705861]

Figure 1.

Examples of mutation detection in D7S1482. DNA from cells obtained by mouth rinse were analyzed by SP-PCR. Each of the 5 subpanels represents the results from one PCR tube. In panels A, D and E, only wild-type alleles were identified with sizes of 366 and 382 nucleotides, respectively (dotted line). Near-single molecule dilution is apparent from the absence of a signal for the larger allele in panel D and signal imbalance in panel A. Mutant DNA fragments are indicated by arrows in panels B and C.

Table 1

Characteristics of cases and controls

	Cases	Controls	Total
Age (median, range)	69 (46–79)	68.5 (47–79)	69 (46–79)
Sex (M/F)	26/24	25/25	51/49
Smoking ¹			
Never	3	3	6
Former	7	7	14
Current	40	40	80
Overall Mutant Frequency (median, range)			
MycL1	1 (0–18)	1 (0–22)	1 (0–22)
D7S1482	27 (8–115)	29 (0–126)	27 (0–126)
DXS981	1 (0–8)	2 (0–7)	2 (0–8)
Unique Mutant Frequency (median, range)			
MycL1	1 (0–5)	1 (0–6)	1 (0–6)
D7S1482	9 (3–20)	8 (0–29)	9 (0–29)
DXS981	1 (0–4)	2 (0–6)	1 (0–6)

¹ Former smokers quit less than 5 years prior to sample collection

Table 2

Univariate analysis of paired cases and controls

Mutation Frequency	Cancer status		Age		Smoking Status			Sex	
	Case	Control	<69	≥69	Never	Former	Current	Male	Female
Overall	2.32	2.33	2.34	2.31	4.67	1.07	2.37	2.80	1.84
MyelL1	38.54	34.45	30.26 ¹	42.17 ¹	27.83	29.79	38.37	41.64	31.29
DXS981	1.80 ²	2.59 ²	2.21	2.17	0.83	1.79	2.37	2.06	2.33
Unique	1.34	1.20	1.36	1.19	1.17	1.00	1.33	1.52	1.02
MyelL1	9.46	9.45	9.06	9.81	9.00	7.50	9.84	10.76	8.12
DXS981	1.28 ³	1.80 ³	1.57	1.50	0.67	1.36	1.63	1.46 ⁴	1.61 ⁴

¹ p=0.027, Quasi-likelihood model
² p=0.030, Quasi-likelihood model
³ p=0.026, Quasi-likelihood model
⁴ p=0.045, Quasi-likelihood model