POINT MUTATIONS IN NORMAL LUNGS OF SMOKERS AND NON-SMOKERS

by
Hiroko Sudo

B.S. Biology
University of Oregon, 1997

Submitted to the Division of Biological Engineering
In Partial Fulfillment of the Requirement for the Degree of

Doctor of Philosophy in Genetic Toxicology

at the
Massachusetts Institute of Technology

February, 2004

Signature of Author_____.
Division of Biological Engineering

Certified by _____.
Professor William G. Thilly
Thesis Advisor

Accepted by_____.
Professor Ram Sasisekharan
Chairman, Division Committee on Graduate Studies

Committee' Page

This doctoral thesis has been examined by a Committee of the Division of Biological Engineering as follows:

Professor Leona D. Samson, _____
Chairman

Professor William G. Thilly _____
Thesis Advisor

Professor Bevin P. Engelward _____

Professor John M. Essigmann _____

Dr. Thomas R. Skopek _____

Professor Alec A. Morley _____

POINT MUTATIONS IN NORMAL LUNGS OF SMOKERS AND NON-SMOKERS

by
Hiroko Sudo

Submitted to the Division of Biological Engineering
In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Genetic Toxicology

Abstract

It is a widely-held hypothesis that environmental mutagens play an essential role in human somatic and germinal cell mutagenesis. In particular, the finding of small amounts of chemical mutagens in cigarette smoke has led to the general hypothesis that mutagens in cigarette smoke induce oncomutations and thus account for the carcinogenic effect of cigarette smoking in human lungs. However, this hypothesis has not been tested by an assay of nuclear point mutations in lungs of smokers and nonsmokers.

Mismatch amplification mutation assay (MAMA), an effective form of allele-specific PCR, was applied for detection of point mutations in TP53 bp742, bp746 and bp747, K-ras bp35 and HPRT bp508 from a total of 291 tracheal-bronchial epithelial sectors from six smokers and nine non-smokers, yielding 949 individual mutational assays. The conditions of MAMA for each target point mutations were optimized such that the sensitivity of each was equal to or below 10^{-5}. Lung epithelial sectors of 2.3×10^6 cells in average contained 0-200 mutant cells in general, equivalent to mutant fractions (MFs) of 0 - 10^{-4} with an exception of rare sectors with MF larger than 4×10^{-4} (4.6%). Noticeably, the distributions of the MFs among sectors did not vary appreciably with the donor's smoking status. The mean MFs per lung were very similar between smokers and non-smokers for all five target mutations assayed ($p >> 0.05$). The mean MFs were slightly higher in females than males ($p = 0.015$). The mean MFs increased with age of the subjects although the correlation did not reach statistical significance due to large variances within the same age group. The distributions of MF among sectors of smokers and non-smokers did not differ significantly by Kolmogorov-Smirnov test for all target mutations but HPRT. By using hypothetical turnover unit sizes and Poisson distribution, the turnover unit size of human tracheal bronchial epithelium was estimated as 64 cells ($p = 0.05$).

These observations do not support the widely-held hypothesis that cigarette smoking causes lung cancer through its induction of point mutations in nuclear genes. The current findings demonstrate the necessity of investigation on alternative mechanisms for tobacco smoke in lung carcinogenesis.

Thesis Advisor: William G. Thilly
Title: Professor of Toxicology, Division of Biological Environmental Health

ACKNOWLEGMENTS

First of all, I would like to thank my thesis adviser Prof. William G. Thilly for his guidance during my graduate career. His constant inspiration and challenge kept me motivated and led me to achieve further. His teaching of "count the molecules" will never leave from my scientific career. I am amazed how much I have learned and academically improved since I came to the MIT; the credit goes to his thoughtful supervision and wrestling spirit.

I thank my thesis committee members, Prof. Leona D. Samson, Prof. John M. Essigmann, Prof. Bevin P. Engelward, Dr. Thomas R. Skopek and Prof. Alec A. Morley for their invaluable suggestions and comments. Their discussions always brought me new insights for the experiments and analyses.

I am also thankful to Prof. James C. Willey in the Medical College of Ohio for providing us with a sufficient amount of precious human lung samples, and to Prof. Emma E. Furth in the Department of Pathology and Laboratory Medicine at the University of Pennsylvania Medical School for providing us with human colon samples.

I would like to express my gratitude to Dr. Xiao Cheng Li-Sucholeiki and Dr. Luisa A. Marcelino, both former Thilly lab members who initiated this lung project. Particularly, Xiao Cheng patiently taught me how to perform PCR and CE in my first year, and even after her leave of MIT, had provided with scientific advices whenever I asked. I am very grateful for her kindness and generosity.

The works of Dr. Rita S. Cha and Dr. Hilary A. Coller, again both former Thilly lab members, were indispensable for my thesis. Dr. Cha has developed MAMA, a very sensitive and effective technique to detect mutations, on which my thesis was solely dependent. Dr. Coller has established the first valid argument with the mutagenic role for cigarette smoke in lung cancer in her mitochondrial analysis, which inspired my thesis hypothesis.

I would like to express a million thanks to Jacklene Goodluck-Griffith for her friendship and cheerful smiles in addition to her laboratory safety supervision. There were times when I faced to seemingly-inescapable dark and depressing phases along my research. She was always there listening to me and laughing out together. Without her enlightening aura and indispensable support, I would not have survived my graduate life. Particularly her occasional offerings of delicious hand-made lunch accompanied with joyful gossips were unforgetful.

I would like to express my appreciation to the current and past members of the Thilly lab: Amanda N. Gruhl, Dr. Elena V. Gostjeva, Weiming Zheng, Brindha P. Muniappan, Aoy Tomita-Mitchell, Pablo Herrero-Jimenez and Rita M. Demeo. Particularly Amanda was both my classmate and labmate, and went through all the academic ordeals together. I am grateful for her friendship, scientific exchanges and dance expertise.

I am also very grateful for the entire members of the Sherley lab for their friendly companionship, sharing joyful moments together.

Finally, I would like to thank my parents in Japan for their encouragement and everlasting faith on me throughout my life. Their unconditional love and mental support brought me here today. I dedicate this thesis, and all of my former and future accomplishments to them.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
APC	Adenomatous polyposis coli
ARMS	Amplification refractory mutation system
ASO	Allele-specific oligonucleotide
BaP	Benzo[a]pyrene
BG	Background
BP(s)	Biotin-labeled probe(s)
bp	Basepair
BPDE	Anti -7,8 -dihydroxy-9,10 -epoxy-7,8,9,10 tetrahydrobenzo[a]pyrene
BSA	Bovine serum albumin
°C	Celsius
CDCE	Constant denaturant capillary electrophoresis
DGGE	Denaturant gradient capillary electrophoresis
dNTP	Deoxynucleoside triphosphate
FA	Fluorescent anisotropy
GOI	Gain of imprinting
HAT	Hypoxanthine, Aminopterin and Thymine
hr	Hours
HPRT	Hypoxanthine-guanine phosphoribosyl transferase
IARC	The international agency for cancer on research
IS	Internal standard
LOH	Loss of heterozygosity
LOI	Loss of imprinting
MAMA	Mismatch amplifying mutation assay
min	Minutes
MF(s)	Mutant fraction(s)
MNU	N-Nitroso-N-methylurea
MT	Mutant
NS	Non-smoker
OLA	Oligonucleotide ligation assay

PAH(s)	Polycyclic aromatic hydrocarbon(s)
PASA	PCR amplification of specific alleles
PCR	Polymerase chain reaction
PIREMA	Primer-introduced restriction with enrichment for mutation alleles
RFLP	Restriction fragment length polymorphism
sec	Second
1SD(s)	One standard deviation(s)
2SD(s)	Two standard deviation(s)
SM	Smoker
SSCP	Single strand conformation polymorphism
Taq	*Thermus aquaticus*
WT	Wild-type

1. INTRODUCTION

Lung cancer is the single most common lethal cancer among men and women in the United States. The public health records and health questionnaire data for the U.S. population have established a clear cause and effect relationship between cigarette smoking and lung cancer. It has been known that cigarette smoke contains low levels of many kinds of mutagenic substances, such as nitrosamines, polycyclic aromatic hydrocarbons, reactive oxygen species, etc. It has been also established that DNA adducts of benzo[a]pyrene (BaP) metabolites are elevated in the tracheal bronchial epithelial cells of smokers (Phillips et al., 1990). An association between the pattern of BaP metabolite DNA adducts and certain G->T transversions in TP53 mutations in lung tumors has been further argued (Denissenko et al., 1996). These observations have led to a hypothesis that cigarette use causes lung cancers because mutagenic substances in cigarette smoke induce point mutations in tumor suppressor genes or protooncogenes, such mutations being necessary for tumor initiation and/or promotion. A natural expectation derived from this hypothesis is that the number of point mutations in lung epithelial cells would be greater in smokers than in non-smokers. This hypothesis has, however, remained untested, and several observations have led us to seriously consider the possibility that cigarette smoking causes lung cancer by pathways independent of a point mutation induction. Recent research into the mitochondrial sequences has demonstrated that the mutational spectra of all 17 mutational hotspots from smokers and non-smokers, including identical twins who were discordant in their smoking status, were indistinguishable (Coller et al., 1998). Indeed, the same set of mitochondrial point mutations arose spontaneously in cultured human cells as in cigarette smoker's lungs (Zheng, in preparation), pointing to endogenous causes of net mutations.

Knowledge of possible induced genetic difference at the nuclear DNA level remains scant due to a lack of technology capable of detecting nuclear DNA mutations from the lungs of cancer-free smokers and non-smokers. Our laboratory has previously developed Mismatch Amplification Mutation Assay (MAMA), a particularly effective method of measuring specified point mutations using allele-specific PCR. It was previously demonstrated that MAMA was capable of detecting ~10 mutant alleles in the

presence of a million wild-type alleles from human tissue sample. This assay enabled us to perform quantitative measurements of specific point mutations of genes in normal tracheal bronchial epithelial cells of smokers and non-smokers, and permitted this thesis to test directly the hypothesis that cigarette smoking causes lung cancer by inducing nuclear point mutations in lung epithelium. If the present observations invalidated this hypothesis it would have broad ramifications, challenging a primary premise underlying environmental genetic toxicology.

2. LITERATURE REVIEWS

2.1. Lung Cancer and Cigarette Smoking

2.1.1. History of Cigarette Smoking

In 1492, Christopher Columbus received tobacco as a gift from the Arawak people of Bahamas during his famous voyage to the America. From the late 16th to 18th century, cigarette was mostly used as a medicine. In the United States, John Rolfe introduced a successful experiment in tobacco cultivation in Virginia in 1613, and in 1730, the first tobacco factories "snuff mills" opened in Virginia. Until mid-19th century, pipe and chewing tobacco were the dominant forms of tobacco practice in the U.S. But after 1840, cigars were introduced, and after 1860, cigarette smoking was gradually adapted, a habit indirectly acquired though the British from their Turkish and French allies during the Crimean War (The Chemical Heritage Foundation, 2001). At that time, 13.9 million cigarettes were smoked annually in the U.S. and cigarette consumption was 0.36 / capita; through next 80 years, cigarette consumption in the United States reached 2,558 / capita (McGrew, 2003).

Today, tobacco smoking is practiced worldwide by over one billion people (IARC, 2002). In the United States, approximately 47 million adults smoke cigarettes, over 5 million use smokeless tobacco, over 3 million adolescents smoke cigarettes and over 750,000 use smokeless tobacco (Smith and Fiore, 1999). Harris, J.E. reconstructed trends in cigarette smoking among successive birth cohorts between 1880 and 1960 of men and women from smoking histories of respondents to the 1978-80 Health Interview Surveys (Figure 1. Harris, 1983). By 1920, over half of the young male population smoked cigarettes; whereas not until 1950 did more than a third of the young female population smoke cigarettes. During the past 30 years, however, the overall rates of smoking have been decreasing; the prevalence of current smoking in the U.S. adults decreased from 42.4% in 1965 to 24.7% in 1995. The proportion of men who ever smoked cigarettes progressively declined with each successive cohort born after 1920; by contrast, the proportion of women who ever smoked declined only among cohorts born after 1940. Maximum exposure to cigarette smoking probably occurred among men who are now in their seventh and eighth decades, whereas peak exposure to smoking probably occurred among women who are now only in their fifth and sixth decades.

In spite of vigorous campaigns held by the public health organizations, smoking cessation does not seem to progress among smokers. Enstrom and Heath (1999) investigated smoking cessation and mortality trends among 118,000 Californians between 1960 and 1997 and found no important decline in death rate despite of substantial degree of smoking cessation, suggesting the population impact of cessation appears to be less than currently believed. Moreover, the rates of teenage smoking have begun to increase; smoking rates in the U.S. high school students in 1997 were 32% higher compared with the rates in 1991. Overall, approximately 28% of Americans smoke in the United States today (Smith and Glynn, 2000).

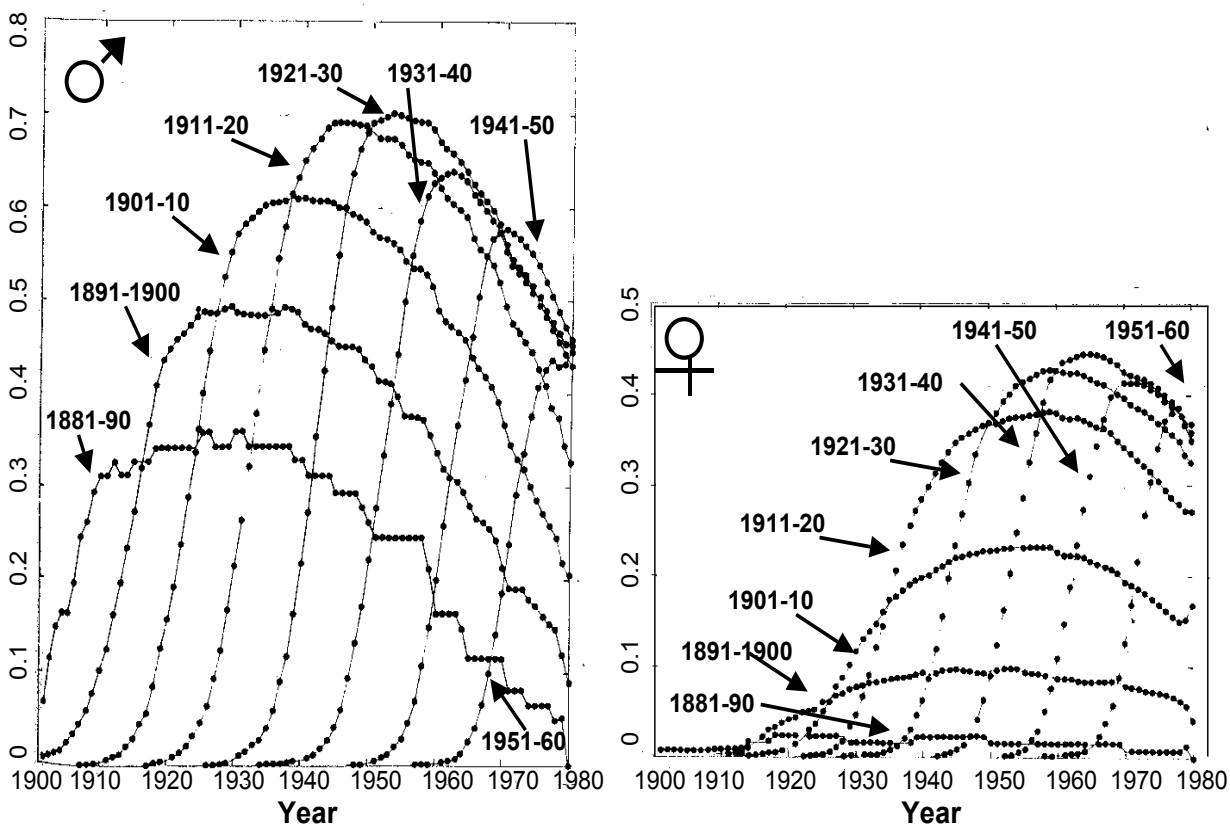


Figure 1. Age-specific prevalence of cigarette smoking (Harris. 1983). Trends in cigarette smoking among successive birth cohorts from 1880's to 1940's of men (left) and women (right) in1900-1980, derived from smoking histories of respondents to the 1978-80 Health Interview Surveys

2.1.2. Cigarette smoke constituents

The composition of tobacco smoke depends on the properties of 1) the leaf or tobacco blend, 2) fillers, 3) additives, 4) wrapper 5) filter and 6) smoking conditions. Inside a burning tobacco product, a large variety of chemical and physical processes occur in an oxygen-deficient, hydrogen-rich environment with a steep temperature gradient, generating a wide range of numerous chemicals (IARC Monographs, 1986). Over 4,800 chemicals are identified in the mainstream of tobacco smoke, among which 69 substances are classified as animal carcinogens and 48 of these are possibly also carcinogenic to human (IARC, 2000). The major toxic agents are nicotine, carbon monoxide, hydrogen cyanide, nitrogen oxides, some volatile aldehydes, some alkenes and some aromatic hydrocarbons (Table 1). The smoke particles that are inhaled are slightly charged (Kingdon, 1959). The pH of tobacco smoke is very important since it influences the proportion of basic components, such as nicotine, in the vapor phase and thus the inhalability of the mainstream smoke (Armitage and Turner, 1970).

The recent interventional attempts of risk reduction by changing tobacco composition and makeup of cigarettes led to the reduction of tar and nicotine yields; in 1955, the average tobacco contained 38mg of tar and 2.7 mg of nicotine per cigarette, whereas in 1994, the yield decreased to 13.5mg of tar and 1.0mg of nicotine per cigarette (Wynder and Hoffmann, 1994). Nevertheless, the anticipated health advantages of switching to lower yields may be offset by a major adjustment in smoking intensity and depth of inhalation by the habitual smokers (Hoffmann et al., 2000).

Table 1. Concentrations of biologically active agents in non-filter cigarette mainstream smoke.

Smoke constituent	Concentration / cigarette
Total particulate matter (“tar”)	15-40 mg
Carbon monoxide	10-23 mg
Nicotine	1.0-2.3 mg
Acetaldehyde	0.5-1.2 mg
Acetic acid	0.1-1.2 mg
Acetone	100-250 □g
Methanol	90-180 □g
Nitrogen oxides	100-600 □g
Formic acid	80-600 □g
Hydrogen cyanide	400-500 □g
Hydroquinone	110-300 □g
Catechol	100-360 □g
Ammonia	50-130 □g
Benzene	20-50 □g
Acrolein	60-100 □g
Phenol	60-140 □g
Croton aldehyde	10-20 □g
Formaldehyde	70-100 □g
Pyridine	16-40 □g
:	:
:	:
N-Nitrosonornicotine (NNN)	200-3000 ng
4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK)	80-770 ng
Vinyl chloride	1.3-16 ng
Benzo[a]pyrene (BaP)	20-40ng

* Adapted from IARC Monographs vol. 38 (1986)

2.1.3. Lung cancer epidemiology – mortality, gender and familial risk

Lung cancer is the leading cancer killer in the U.S. for both men and women of all ethnicities. Approximately 171,900 new cases of lung cancer will be diagnosed in 2003, and an estimated 157,200 Americans will die from lung cancer, accounting for 28% of all cancer deaths (Jemal et al., 2003). More Americans die each year from lung cancer than from breast, prostate and colorectal cancers all combined (Fiore, 1992). The median survival is 13 months due to lack of effective therapy and over 85% of people with lung cancer eventually succumb to this disease. Due to this low survival rate, the incidence rate of lung cancer closely reflects the mortality rate.

Previously, our laboratory has collected the population and lung cancer mortality data of the United Sates from the U.S. Department of Health and Human Services (Vital Statistics of the United States, 1937-1992) and the U.S. Bureau of the Census (Mortality Statistics, 1930-1936), and organized them as age-specific mortality rates for each birth decade from the 1820s forward (Figure 2. Herrero-Jimenez, 2001. http://epidemiology.mit.edu). Among American males, the lung cancer mortality has almost 25-fold increased since the end of nineteenth century, and it is still increasing. The mortality of other forms of cancer, such as pancreas, intestine and blood (lymphoma), also appears to have increased over history. Because our genetic materials are unlikely to change at such a dramatic speed within such a short time span, the rapid increase in these cancer mortalities can be attributed to solely environmental risk factors. However, none of such environmental risk factors have been clearly identified as the causalities of these cancers except cigarette smoking in lung cancer.

It has been reported that greater than 50% of newly diagnosed lung cancer occur among former smokers who had quit more than five years previously (BC Cancer Agency, 2003). Because there are approximately equal numbers of current smokers and former smokers in the U.S. at the presence, the incidence of the disease is not expected to drop significantly in the next several decades even if the public anti-tobacco campaigns succeed to reduce the smoking population.

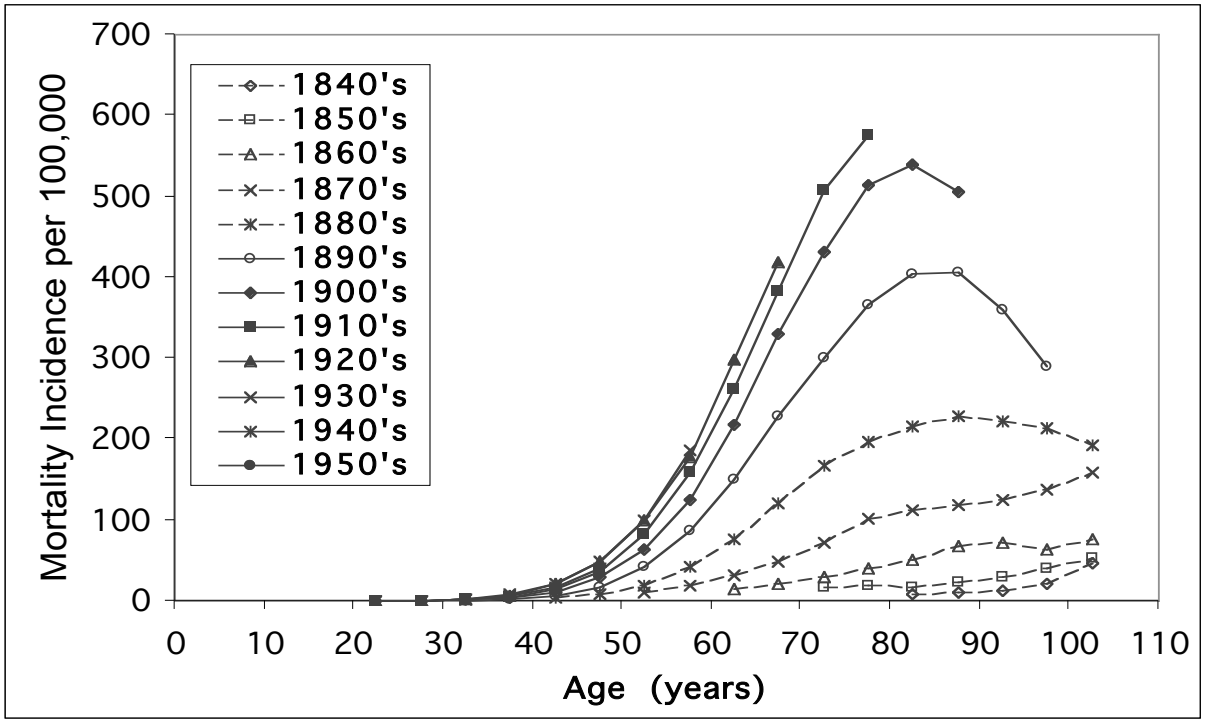


Figure 2. Age-specific lung cancer mortality for European American males from birth year cohorts 1840s-1950s (Herrero-Jimenez, 2001. http://epidemiology.mit.edu). The lung cancer mortality incidence has been dramatically increased since the beginning of the last century, and the increase has reached to almost 25-fold.

The observed historical shift of lung cancer mortality in females from males is one of the evidences supporting cigarette smoking as a causality of lung cancer. The apposite figure is the one showing an ~30 year "shift" of the lung cancer mortality between males and females at the birth year cohort 1880's and 1910's, respectively (Figure 3). This gender "shift" was seen only for lung cancer, but not other major cancers, such as colon, leukemia, lymphomas, pancreas, etc. This shift corresponds with the historical fact that men started the habit of smoking cigarette approximately 30 years earlier than women; over half of young men started to smoke by 1920, whereas not until 1950 were a third of young women smoking cigarettes. It is also noted that many women in the earlier cohorts began smoking after age 30, whereas nearly all men started smoking before age 25 (Harris, 1983). Thus, it appeared that the late starting historical year and the late initiation age for smoking among women attributes to the delayed increase of lung cancer mortality, almost 30 years behind that of men. When this gender-related latency of smoking habits is considered, there is little difference between men and women in the effect of smoking cigarettes on lung cancer mortality.

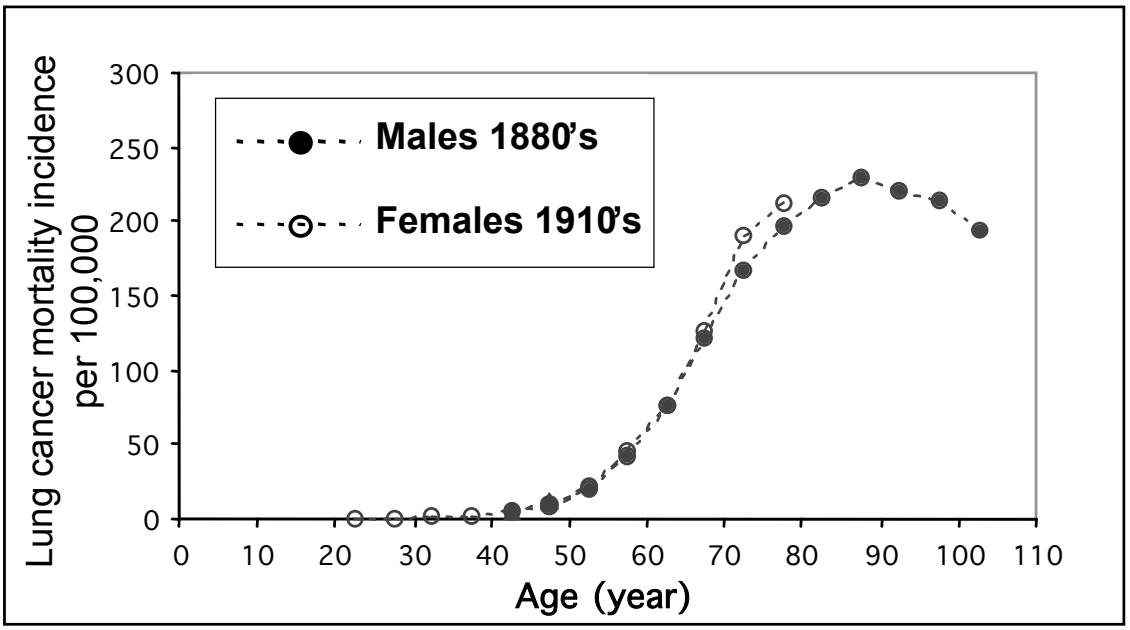


Figure 3. 30-year latency in the lung cancer mortality between European American males and females. The two mortality curves of European American males in birth cohort 1880s (closed circles) and of females in birth cohort 1910s (open circles) were super-imposable, suggesting the risk of lung cancer was equal in these two groups.

In addition to cigarette smoking, family history of lung cancer and personal history of lung diseases (e.g., asthma, chronic bronchitis, pneumonia or tuberculosis) also appear to be associated with an increased risk of lung cancer. It has been long argued whether or not inherited risks play a significant role in lung cancer. Numerous studies suggested a strong familial risk (Ooi et al., 1986; Wu et al, 1988; Shaw et al., 1991; Osann et al., 1991; Ambrosone et al., 1993; Yang et al., 1997; Brownson et al., 1997; Mayne et al., 1999). However, the problem which most of these studies faced is the difficulty in seperating heritable risks from environmental risks, since families often share both risk facors. Furthermore, smoking practice often confounds the results in such studies; as a result, the significance of inherited risk factor(s) still remains unclear. One interesting piece of work is a cohort study done by Braun et al. (1994) on 15,924 U.S. male twin pairs who served in World War II. It was found that the ratio of the observed to the expected concordance among monozygotic twins did not exceed that among dizygotic twins, suggesting little if any effect of inherited predisposition on the development of lung cancer.

2.1.4. Effect of cigarette smoking on lung carcinogenesis

The first association between tobacco and cancer was made by a London physician, John Hill, who reported six cases of "polypusses" related to excessive use of snuff in his Cautions Against the Immoderate Use of Snuff in 1761. Later in 1776, Dr. Percival Pott noted the high incidence of scrotal cancers in chimney sweeps, which was the first suggestion of the association of "substances" in the environment with the development of cancer in a particular occupational group. Samuel von Soemmering added pipe smoking to identified environmental carcinogens in 1795 (The Chemical Heritage Foundation, 2001). More recently, Peral, R. (1938) investigated family history records of 6,813 white males and reported that tobacco smoking was statistically associated with an impairment of life duration, and the amount or degree of this impairment increased as the habitual amount of smoking increased. Similarly, Ochsner and DeBakey (1939) discussed the possibility that smoking causes lung cancer by irritating the bronchial mucosa. More recently, Doll and Hill reconfirmed the association between smoking and carcinoma of the lung (1952). In 1957, as a first statement from

the Public Health Service, Dr. Burney, L addressed "the weight of evidence at present implicates smoking as the principal etiological factor in the increase incidence of lung cancer". Finally in 1964, for the first time in the United States, a country in which at the time 46% of adults smoked, the Surgeon General's Reports at last concluded that smoking causes cancer (Terry, et al).

Most clearly and concisely, Herrero-Jeminez demonstrated in his analysis of lung cancer mortality rates and cigarette use in the U.S. that the fraction at lifetime risk for all birth year cohorts analyzed was a simple linear function of reported cigarette use regardless of gender and ethnicity (Figure 4). Furthermore, he estimated that essentially 100% of smokers were at risk of death by lung cancer although less than 10% actually died of lung cancer (thesis, 2001). In addition to this clear indication, an enormous amount of epidemiological and clinical studies (Wynder and Graham, 1950; Stayner and Wegman, 1983) have heretofore suggested a strong link between cigarette smoking and lung cancer.

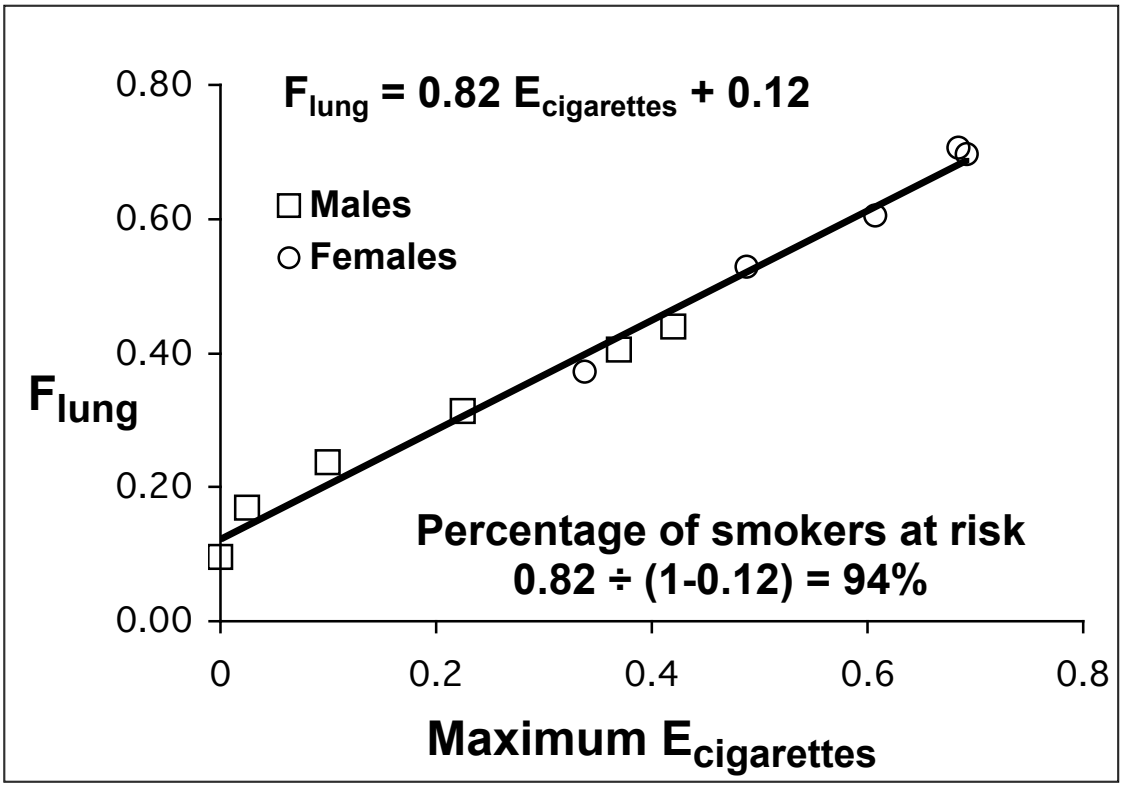


Figure 4. Fraction of the population at risk for lung cancer (F_{lung}) as a function of maximum smoking prevalence ($E_{cigarettes}$) for the birth decade cohorts of European American males (squares) and females (circles) who were born in the 1880s to 1920s (Herrero-Jimenez, 2001). Solving the implied algebraic relationship, the authors calculated a fraction of smokers at lifetime risk of lung cancer to be 0.94 at minimum.

However, the underlying mechanisms by which cigarette smoking is involved in lung carcinogenesis have remained poorly understood. This is partly due to the fact that cellular and molecular effects of cigarette smoking are extremely complex. Factors such as age, sex and nutritional status of experimental subjects, type of cigarette used, exposure conditions employed as well as specific tissue examined might significantly affect the results obtained.

2.1.4.1. Lung cancer types and susceptibility

More than 99% of malignant lung tumors arise from the respiratory epithelium and are termed bronchogenic carcinoma. Bronchogenic carcinoma is divided into two subgroups: small cell lung cancer and non-small cell lung cancer; the latter includes adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Both small cell carcinoma and squamous cell carcinoma usually arise in the proximal trachea and bronchi, whereas adenocarcinoma most arises peripherally from the smaller airways (bronchioles and alveoli).

It has been well proposed worldwide that small cell carcinoma and squamous cell carcinoma are strongly associated with lung cancer, and much less for adenocarcinoma, and not for large cell carcinoma (Sobue et al., 1988; Morabia and Wynder, 1991; Barbone et al., 1997; Makitaro et al., 1999). In the past, the most common histological type was squamous cell carcinoma; however, the rate of adenocarcinoma has dramatically surpassed that of squamous cell carcinoma since the mid-1980s (Travis et al., 1996). Adenocarcinoma has always been the most common form of lung cancers among non-smokers, female patients in general and early onset male patients (Kreuzer et al., 1999). This increase of adenocarcinoma, the histological type with assumingly less association with cigarette smoking, contradicts with the postwar expansion of popularity in cigarette smoking. By comparing the birth cohort effect and calendar period effect, Thun, MJ and his colleagues (1997) examined whether this increase of adenocarcinoma was due to an advance in diagnostic technology (i.e. enhanced ability to perform biopsies on distal airways), or changes in cigarette design or practice. They found that the increase of adenocarcinoma followed a clear birth cohort pattern, suggesting that it was

more consistent with changes in cigarette designs and smoking behaviors than with diagnostic advances.

Cigarette designs have been indeed changing over history (section 2.1.2.). This reduction of tar and nicotine yield in addition to the filtered cigarettes introduced in the 1960s could have been favored for the development of specific types of lung cancer (Hoffmann and Hoffmann, 1997). The smoke from unfiltered products was so irritating to inhale deeply that carcinogens were deposited on the epithelium at the large central bronchi, where squamous cell carcinoma preferentially arises. On contrary, filtered products remove larger particles in cigarette smoke; thus carcinogens move deep into the peripheral where adenocarcinoma arises. Also, smokers of low-yield filtered cigarettes have tended to compensate for the yield reduction by changing the number of cigarettes smoked per day, and by puffing more frequently or deeply to satisfy a craving for nicotine. Naturally, those practices may further help the development of adenocarcinoma. Moreover, blended reconstituted tobacco, introduced in the 1950s, releases higher concentration of nitrosamines from tobacco stems than tobacco leaves made predominantly before (Hoffmann et al., 1993). Because nitrosamines are smaller molecules and easier to get into narrower peripheral airways, the introduction of blended tobacco might have further accelerated adenocarcinoma more than squamous cell carcinoma.

These findings indicate that the historical changes of tobacco constituents and smoking practice might change tumor pathogenesis and led to development of different lung cancer types.

2.1.4.2. Histological and morphological changes in smokers' trachea and bronchus

During cigarette smoking, particles impinging upon lung tissues and the subsequent particle removal processes of the lung result in the introduction of a range of chemicals into the lung via a number of dissolution processes and biochemical mechanisms.

Histological and morphological changes in smokers' airways are remarkable even if the smokers retain otherwise healthy condition. Auerbach and Stout (1957, 1961) dissected the upper bronchial trees of cigarette smokers and non-smokers, rural and urban

dwellers, and subjected the multiple sections to microscopic examination. They noted that both ciliated and non-ciliated lesions with atypical nuclei were rare in sections from non-smokers but were common in smokers and increased monotonically with the number of cigarette smoked per day. They further extended their observations to ex-smokers and found that the number of lesions decreased some 40-fold and approached the low levels seen in non-smokers.

In addition to the loss of effective functionality of cilia or cilia itself, the most obvious transformation is displacement of normally ciliated pseudo-columnar cells by metaplastic stratified squamous and basal cells, causing thickening of surface epithelium (Macholda et al., 1970; Trevisani et al., 1992). Particularly, the proportion of goblet cells in healthy smokers is found as high as bronchitic patients (Thurlbeck et al., 1975; Spurzem et al., 1991). Nuclear atypia and increased mitosis also occur (atypical metaplasia or dysplasia). Importantly, cigarette smoking is known to contribute to inflammatory diseases of the respiratory tracts by promoting recruitment of inflammatory-immune cells, such as T-lymphocytes, macrophages and neutrophils, in peripheral airways (O'Shaughnessy et al., 1997; Saetta et al., 2000). If the exposure becomes chronic, the inflammatory response itself could also induce severe tissue damage (Maestrelli et al., 2001). It is also known that structural recovery occurs in bronchial epithelium in people who stop smoking for over two years (Bertram and Rogers, 1981); thus these histological changes induced by smoking is reversible. However, once carcinoma occurs, which is characterized by cellular disorganization and nuclear and mitotic atypia through a full thickness of the mucosa, the integrity of the basement membrane may be lost forever and frank infiltration of neoplastic cells into the underlying stroma may occur.

2.1.4.3. Molecular and genetic changes – BPDE adducts and TP53 mutations

Cigarette smoke contains low levels of approximately 3,800 different kinds of chemical substances including carcinogens such as nitrosamines, polycyclic aromatic hydrocarbons (PAHs), oxidative agents, etc. (Hetch, 1999). One of the most extensively studied markers of human exposure to cigarette smoke is, benzo[a]pyrene (BaP), a type of PAHs. Upon metabolic activation, BaP is transformed to a more harmful compound (±)

anti-7b,8a-dihydroxy-9a,10a-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene (BPDE), which generates predominantly covalent (+) trans adducts at the N2 position of guanine in DNA (Figure 5. Weinstein et al., 1976). BPDE has been known to induce base substitutions via G:C->TA transversions *in vitro* (Eisenstadt, 1982). It has been established that DNA adducts of PAHs are somewhat elevated in the normal tracheal bronchial epithelial cells of smokers (Phillips et al., 1990; Dunn et al., 1991; Izzotti et al., 1991; Routledge et al., 1992; Mustonen et al., 1993; Schoket et al., 1993; Schoket et al., 1995; Butkiewicz et al., 1999; Piipari et al., 2000), whereas others have argued that the adduct level is not always significantly higher in smokers' lungs than in non-smokers' lungs (Besarati et al., 2000; Cheng et al., 2000). A linear correlation between the number of years of smoking and the PAH-DNA adduct level was further found (Phillips et al., 1988; Ryberg et al., 1994). Moreover, it has been observed that the overall numbers of lung tumor samples containing TP53 mutations were higher among smokers (Ryberg, et al., 1994a; Kondo, et al., 1996; Vahakangas, et al, 2001), particularly G:C->T:A transversion mutations which is the same type of mutations that BPDE could create *in vitro* (Takeshima et al., 1993; Hussain et al., 2001). But again, others disagree that the difference was statistically significant (Husgafvel-Pursiainen and Kannio, 1996; Gao et al., 1997; Marchetti et al., 1998). Most importantly, an association between the pattern of BPDE DNA adducts and certain G->T transversions in TP53 mutations in lung tumors has been observed (Denissenko et al., 1996; Hernandez-Boussard and Hainaut, 1998; Gao et al., 2003).

These observations have led to a hypothesis that cigarette use causes a direct genetic damage to DNA sequences through the formation of bulky carcinogenic DNA adducts, such as BPDE, leading to mutations necessary for tumor initiation and/or promotion. A natural expectation based on this hypothesis, therefore, was that the number of point mutations in lung epithelial cells would be greater in smokers than in nonsmokers.

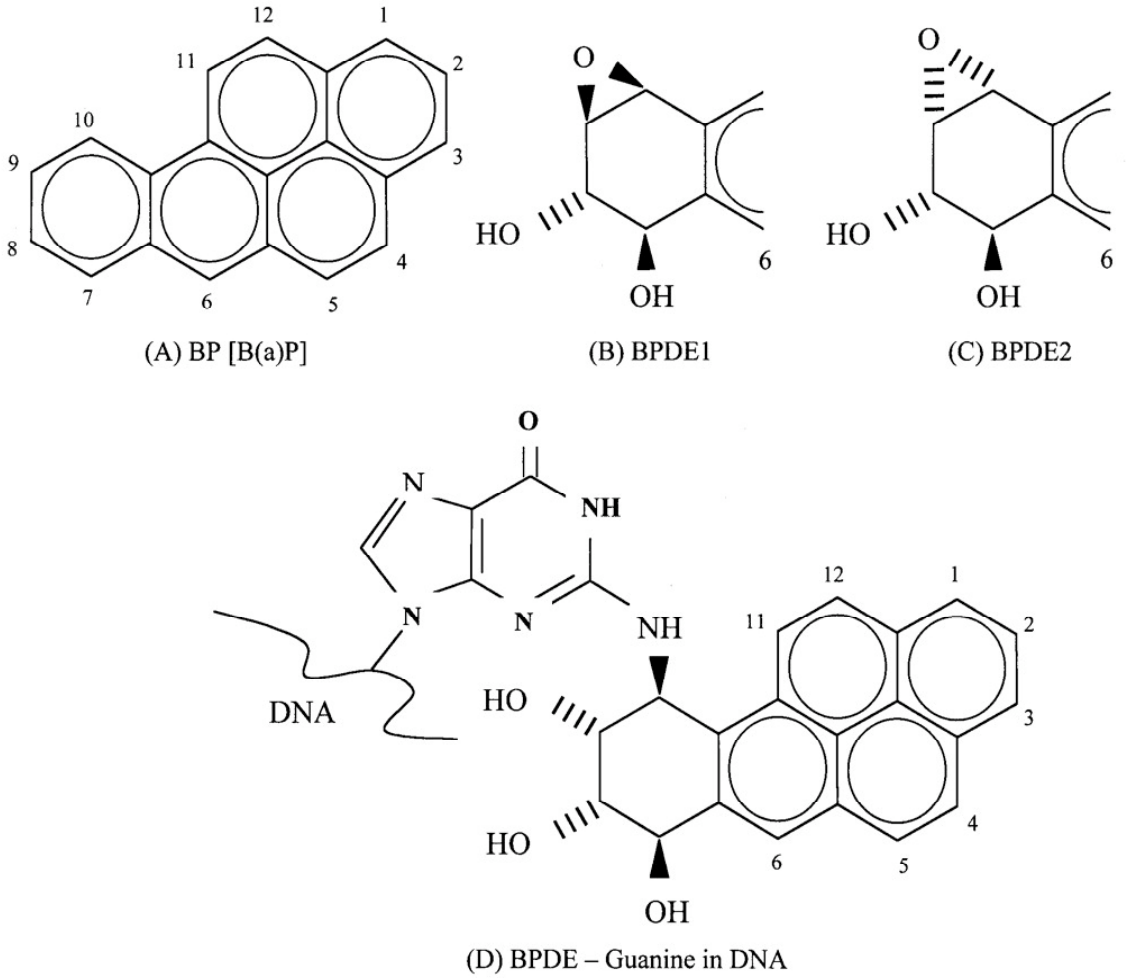


Figure 5. Structural formulae of BaP and its metabolites (Rubin, 2001a). (A) benzo[a]pyrene (BaP); (B) benzo[a]pyrene-diol-epoxide-1(BPDE-1); also called the *syn* or *cis* isomer, and diol-epoxide II; (C) benzo[a]diol-epoxide-2 (BPDE-2); also called the *anti* or *trans* isomer, and diol-epoxide I; (D) BPDE bound to the exocyclic 2-N of guanine in DNA. BPDE-2 is considered a carcinogen on the basis of binding to DNA, mutagenicity and extreme pulmonary carcinogenicity in newborn mice. BPDE-1 has a similar binding to DNA and mutagenicity, but it is not carcinogenic.

This hypothesis has, however, never been tested in nuclear DNA sequences of non-cancerous human lungs, and several observations have led me and my co-workers to seriously consider the possibility that cigarette smoking causes lung cancer by a pathway independent of point mutation induction.

First of all, the concentration of BaP in the tobacco smoke is only ~2% of the amount of pure BaP required to cause skin cancer in mice, which may not be sufficient to initiate carcinogenesis in smokers' lung. (Wynder et al., 1957; Roe et al., 1959; Rubin 2002). Furthermore, the mere fact that BaP-DNA adduct levels are measurably elevated in a range of cell types in smokers when compared to non-smokers posed of an enigma because the daily BaP intake levels for the general population are typically one or more magnitude higher than can be generated from moderate smoking. BaP intake level from a cigarette has been calculated to be 0.8-40 ng, varying widely among brands and generic types, e.g., tar levels and filtration (Smith and Hansch, 2000); in average, BaP concentrations in filtered mainstream cigarette smoke are estimated to be ~9 ng/cigarette (Chepiga et al, 2000), yielding an BaP intake of 180 ng/day for pack-a-day cigarette smokers. Compared with the long-term average daily intake of BaP by the general population of 2.2 μg per day (Hattemer-Frey and Travis, 1991), the intake of BaP from cigarette smoke, even for heavy smokers, is negligible. In other studies, the human exposure condition of BaP due to solely breathing ambient air was calculated. The concentration of BaP in the ambient airborne is largely dependent upon the local level of pollution; while the lowest indoor concentration of BaP is 0.1-0.15 ng/m^3, 2-4 ng/m^3 of BaP is present in typical suburban locations without extraordinary PAH emission sources (Butler and Crossley, 1979), up to 4.55 ng/m^3 for busy traffic areas in a city (Merlo et al., 1997) and reaches 8.1 ng/m^3 in Pillipsburg, NJ, a city containing a metal pipe foundry (Lioy et al., 1988). When a person inhales air 22 m^3/day, the BaP exposure to the lung ranges from 2.3 – 178 ng/day. This background exposure level is roughly equivalent to the amount of BaP intake from 1 to 20 cigarettes smoke per day. These studies concluded that cigarette smoking does not substantially increase human exposure to BaP relative to exposures to background levels of BaP present in the environment.

Secondly, noting an earlier paucity of data for TP53 mutations in non-smokers, Rodin and Rodin (2000) reevaluated the TP53 mutation data from more than 10,000 tumors compiled by the International Agency for Cancer on Research (IACR) and

available scientific literature. They analyzed the mutations as a function of its position in the TP53 gene sequence (mutational spectrum) and found that the mutational patterns were indistinguishable between smokers and non-smokers. They also found that the distribution of G:C to T:A transversions along the TP53 gene was comparable in lung tumors and tumors of organs unaffected by cigarette use. Most importantly, they found that smokers and non-smokers do not differ in the frequency of tumors that carry silent or unselected mutations, and that silent G:C to T:A transversions in lung cancer do not correlate with the spectrum of BPDE adducts formed at sites that would yield silent TP53 mutations. The authors reached a conclusion different from the prevailing hypothesis that is "targeted adduct formation rather than phenotypic selection appears to shape the TP53 mutational spectrum in lung cancer".

Similarly, Paschke (2000) argues that a large number of discrepancies existed in the classification of smoking status between successive releases of the IARC TP53 mutation database and that no statistically significant differences could be found in G:C to T:A transversion frequencies between smoking and non-smoking lung cancer patients. Pfeifer's and other groups have, nevertheless, continuously defended their central hypothesis from Rodin and Paschke's standpoints, and asserted that TP53 mutations in lung cancers can be attributed to direct DNA damage from cigarette smoke carcinogens rather than to selection of pre-existing endogenous mutations (Hainaut and Pfeifer, 2001; Hainaut et al., 2001; Cooper, 2002).

2.1.4.4. Smoking experiments on animal models

The pathogenesis of lung cancer remains highly elusive due to its aggressive biological nature and considerable heterogeneity. Rodent models of lung cancer have long been studied, but quite often the results seen with such animal models are dissimilate from what has been observed in humans. Lung tumors developed in mice or rats are usually measurable only late in their course, their metastatic pattern is not uniform, and their response to therapy is generally poor. For lung cancer, none of the available rodent models are optimal, in that none originated from an orthotopic (bronchial) primary site, and exhibit extensive extrathoracic metastasis (Howard et al., 1999). In fact, the transgenic mice with TP53 or RB mutations develop bronchial

hyperplasia, but die of other neoplasms, such as islet cell carcinoma, before progression of lung cancer can occur (Liu and Johnston, 2002). The models are not informative for an initiation step of lung carcinogenesis while it could be relatively useful for studies of tumor progression and metastasis,

Attempts to produce lung tumors in mice by inhalation of the smoke have either failed or could not be reproduced (Wynder et al., 1953; Davies, 1960). First of all, rodents are obligatory nose breathers and their nasal passages are more complex than those of humans, thereby affecting the dynamics of particle deposition in the respiratory tract. These obvious pharmacodynamic differences between rodents and human make the animal models difficult to reflect human carcinogenesis. Furthermore, even among rodents, there are considerable species-to-species variations in the degree and/or presence of these different abnormalities (Wright and Churg, 2002). For example, when BPDE was administered systemically, it causes lung tumors in mice, but not in rats (Culp et al., 1998); likewise, N-nitrosodiehylamine was found to be an effective pulmonary carcinogen in hamsters, but not in rats (Hecht, 1999). Although K-ras mutations are commonly observed in mouse lung tumors in general, they are rarely found in rat lung tumors; moreover, rodent lung tumors rarely contain mutated TP53 genes, which is the biggest genetic difference compared to human lung tumors (Hecht, 1998; Belinsky et al., 1997). Even mice, the best rodent model among others is far away from a perfect; when treated with carcinogen, they develop adenocarcinoma via visible benign adenomas which are very rare in human lungs.

The rodent model of lung cancer affected by tobacco carcinogens, therefore, has yet to optimally reflect some important biological properties of human cancer. Direct analyses of human tissue are therefore crucial for better understanding of lung cancer pathogenesis and carcinogenesis in humans.

2.1.4.5. Smoking experiments on human blood

One of the human tissues most easily available for research is blood. Many have investigated adduct formation and mutation frequency in circulating T-lymphocytes from smokers and non-smokers. The results are, however, somewhat inconsistent. The effect of smoking on the level of PAH-adducts in the T-cells is inconsistent (Phillips et al.,

1990; Savela and Hemminki, 1991; Gallagher et al., 1993; van Schooten et al., 1997; Godschalk et al., 1998; Cheng et al., 2000). Likewise, although the HPRT mutant frequency in T- cells was higher in smokers than in non-smokers in some observations (Cole et al. 1988; Jones et al., 1993; Huttner et al., 1995; Hou et al., 1999; Yang et al., 1999), it did not reach to the significant level in others (Davies et al., 1992; Branda et al., 1993; Robinson et al., 1994; Cheng et al., 1995). Importantly, many have observed no significant difference in the HPRT base substitution spectra of smokers and non-smokers (Vrieling et al., 1992; Burkhart-Schultz et al., 1996; Podlutsky et al., 1999; Hackman et al., 2000). Studies of the HPRT gene mutations in peripheral T-lymphocytes have, therefore, not *en toto* shown an increase in mutant fraction (MF) that could explain the high lung cancer risk in smokers compared to non-smokers (Cole and Skopek, 1994).

It has been in fact noted that T-lymphocytes are probably not a good biomarker for smoking because smokers are known to have more active immune system with higher turnover of T-cells, suggesting the higher mutant frequency observed in some studies can be mere reflection of more T-cells replications in smokers (Curry et al., 1990). Therefore, the effect of cigarette smoking on mutant frequency may not be accurately determined by studying T-cells. The direct assay of the target organ, lung, is therefore, essential to study smoking effect on mutant frequency.

2.1.4.6. Smoking experiments on mitochondria in human lung

Coller and her colleagues (1998) compared point mutational spectra in a 100bp mitochondrial sequence from bronchial brush biopsies of smokers and non-smokers. They found that there was no effect of cigarette smoking on the kinds or numbers of mitochondrial mutants in all seventeen hotspots assayed, even for identical twins who were discordant for smoking status. Since this set of hotspots included cryptic point mutations, they concluded that the rate of maintenance turnover death and division had not been affected by smoking. Work in progress by W.M. Zheng of our laboratory has demonstrated that twelve out of the seventeen hotspots, which accounts for 90% of the *in vivo* MF, are created when this same sequence is copied by the human mitochondrial DNA polymerase . An increase in mutant fraction was observed for all of the G:C->A:T transitions and the single G:C->T:A transversion when the DNA sample was subjected to

extended heat treatment. This observation suggests that unrepaired cytosine deaminations are the premutagenic lesions for the G:C->A:T transitions. The A:T->G:C transitions are not affected by DNA heat treatment and are apparently primary misincorporation errors of the DNA polymerase □. These findings suggest that endogenous factors, such as DNA replication errors by polymerase, could generate a significant level of spontaneous mutations, hence become significant contributors to human carcinogenesis.

2.2. Carcinogenesis Model and Lung Cancer Genes

2.2.1. Three-stage carcinogenesis model

Cancer is a disease caused by multiple genetic events, arising from accumulation of mutations that promote clonal selection of cells with increasingly aggressive behavior (Fialkow, 1976; Tomlinson et al., 1996; Fearon et al., 1997; Garcia et al., 1999). In order to understand the stepwise process of genetic alterations during tumor development, a mathematical model of three-stage carcinogenesis was established by collective contributions from Nordling (1953), Armitage and Doll (1954), Knudson (1971), Moolgavkar and Knudson (1981) and Herrero-Jimenez, et al. (2000).

The three-stage carcinogenesis model postulates that a normal cell is changed so that it becomes the founding cell of a preneoplastic lesion (initiation) which becomes the founding cell of a neoplastic lesion (promotion) which finally gives rise to lethal metastatic tumor (progression). The model proposes that the initiation requires "n" separate events in any normal cell, leading to growth of a preneoplastic colony. Each initiated cell has a finite probability of forming such a preneoplastic colony determined by the rates of cell division and death in that initiated cell and derived colony. Herrero-Jimenez's model predicted that the cellular net growth rate of preneoplastic colony is indeed enhanced in smokers (0.33 doubling / year) compared to non-smokers (0.17 doublings / year) (thesis, 2001). Also interestingly, the growth rate of preneoplastic colonies in non-smokers was noted to be very similar to the growth rate estimated for juveniles.

For several human organs, it appears that independent losses of both alleles of certain tumor suppressor genes constitute initiation, or n = 2. According to the

Knudson's "two-hit" model (Knudson, 1971), one germline hit and one somatic hit would be required in hereditary diseases, whereas two somatic hits would be required in non-hereditary diseases. The first inactivation event is mostly point mutations, frequently base pair substitutions, and the second event involves another independent point mutations, but also including chromosomal exchanges and recombination, resulting in loss of heterozygosity (LOH). The genes (and the organs) in which their loss appears to represent the only events of initiation are RB (retinoblastoma; Friend et al., 1987), APC (colorectum; Levy et al., 1994), VHL (kidney; Gnarra et al., 1994), P16 (melanoma; Kamb et al., 1994), PTCH (skin; Gailani et al., 1996) and NF1 and NF2 (central nervous system; Rouleau et al., 1993). Other genes involved in initiation of many cancer types are being pursued vigorously by the cancer genetic community. No gene has yet been found to play this role for lung cancer.

It is important to note that the first n-1 initiation mutation must occur in stem cells since a mutation first occurring in a non-stem cell would inexorably be lost during normal cell turnover, whereas a mutation in a stem cell can be passed down to the downstream turnover unit consisting of transition and terminal cells. Once the turnover unit consists of cells carrying n-1 of the initiation mutations, the n^{th} event could occur either in the stem cell or, as equally probable, in any newly divided cell including those that would otherwise differentiate into terminal cells.

Similar to initiation, the model proposes that promotion requires "m" separate events in a particular cell of the preneoplastic colony in order to completely transform into a neoplastic colony. However, there are no known genetic events associated with promotion in humans or experimental animals; thus, m might be zero.

Finally, progression, which is the events between the first surviving neoplastic colony and death of the organism, occurs rapidly at a cellular net growth rate of more than 20 doublings / year regardless of smoking status of the patients, and hence not a rate-limiting process. Many genetic events are recorded during progression, such as mutations and deletion in tumor suppressor genes, chromosomal recombination and aneuploidy.

After understanding the model of carcinogenesis, the original question came back; how might cigarette smoking cause lung cancer? During initiation, cigarette smoking

might 1) increase the point mutation rate in stem cells, which is indeed a prevailing hypothesis in the field of environmental carcinogenesis; 2) increase the LOH rate in stem cells, possibly as well as transition cells; 3) increase the number of or change the type of cells at risk; 4) increase the number of genes involved and/or 5) accelerate the onset age in which initiation occurs. During promotion, cigarette smoking might change rate(s) of any "m" events, most of which have not been elucidated. It might also enhance the growth rate of preneoplastic colonies, the outcome which was exactly predicted from the mathematical model. Overall, cigarette smoking could cause lung cancer by influencing any one or more of these events listed. In this study, I did not intend to discover a novel mechanism of how exactly cigarette smoking causes lung cancer; rather my goal was to eliminate the improbables from the list of all possible mechanisms. This strategy was taken because there is no single way to tackle this biologically highly-complex problem.

2.2.2. Lung cancer gatekeeper gene

Amongst the tumors suppressor genes and oncogenes identified to date, many have been candidates for the "gatekeeper genes" whose mutation, inactivation or loss is required to initiate lung cancer; none, however, has been conclusively proven to have sustained a mutation or loss in the initial stages of lung cancer development, a corollary to gatekeeping status. Considering the central premise of carcinogenesis, that is, neoplasm arise from a single cell of origin (Nowell, 1976), the initiation mutations that have occurred in the original cell must be present in all the descendent clones. No gene has yet showed such a high fraction of mutations in lung tumors. TP53 gene has been proposed as a lung gatekeeper gene in some observations; nevertheless, the fractions of lung tumors which contains the same mutations is within the range of 30% and 50% (Ryberg et al., 1994b; Kondo et al., 1996; Husgafvel-Pursiainen and Kannio, 1996; Lang et al., 2000; Vahakangas et al., 2001). Therefore, TP53 gene inactivation is not the first event in the lung carcinogenesis but rather a later event, probably during tumor progression.

Another candidate for lung gatekeeper gene is the fragile histidine triad (FITH) gene located in the short arm of chromosome 3 (3p14.2), containing 10 exons distributed in over 500kb and encoding a protein ApppA hydrolase which consists of 147 amino

acids. This gene has been reported to be lost in high frequency, particularly in small cell carcinomas (Whang-Peng et al., 1982; Naylor et al., 1987; Yokota et al., 1987; Mao et al., 1997; Kohno and Yokota, 1999). In most of these cases, the gene contains no point mutations but is homozygously deleted (Nelson et al., 1999); however, their LOH rates are not complete or universal in lung tumors (range 20-100%) (Marchetti et al, 1998; Hung et al., 1995; Kohno et al., 1999; Tseng et al, 1999; Geradts et al., 2000). Because by definition, the gatekeeper genes have to be inactivated in all cancer cells of a tumor colony with the same mutation, FHIT may not be the gatekeeper gene after all. It is still possible that the initiation and/or promotion require(s) not only one but multiple pathways; FHIT inactivation could be only one of them. Current data is too inconsistent to conclude that FHIT is involved in initiation or promotion events; further molecular analyses are necessary to discover gatekeeper gene(s) for lung cancer.

In the absence of a proven gatekeeper gene for lung cancer, the choice of suitable genes for the present study is based on their potential as mutation targets for lung cancer. In other words, genes proposed herein are any carcinogenic genes for studies of point mutagenesis rather than direct measurements of initiation mutations. Target genes were chosen based on their oncogenic role in lung cancer and arguable association with smoking-related mutations.

2.2.3. TP53 bp742, bp746 and bp747

TP53 is the most studied tumor suppressor gene. It appears to involve in many signaling pathways including cell cycle, growth, angiogenesis, DNA repair and apoptosis. TP53 mutations are found in about a half of sporadic cancers, and the germline mutations of this gene causes Li-Fraumeni syndrome that increases the risk of developing multiple primary cancers in the patients. In lung cancer, TP53 mutations appeared in a range from 33% in adenocarcinomas to 70% small cell carcinomas (Soussii, 2002).

TP53 G:C->T:A transversion mutations are genetic hallmarks for lung cancer; the prevalence of this type of mutation is observed more than 30% of lung tumors, which distinguishes lung cancer from other cancers such as colon (Nishisho et al., 1991), breast (Sasa et al., 1993; Meng et al., 1999) or bladder (Husgafvel-Pursiainen and Kannio, 1996, LaRue et al., 2000) in which the majority of mutations are G:C->A:T transitions.

Furthermore, G:C->T:A transversions are more frequently observed in lung tumors of smokers than that of non-smokers (Takeshima et al., 1993; Hernandez-Boussard and Hainaut, 1998; Marchetti et al. 1998; Bennett et al., 1999; Hussain et al., 2001; Soussi, 2002). Although it has not yet been demonstrated that BaP or cigarette smoke induces these specific G:C->T:A transversions at CpG sites of the TP53 gene in human cell *in vivo*, it has been demonstrated that the relevant BPDE preferentially reacts with guanine at these sites when DNA in human cells are methylated *in vitro* (Denissenko et al., 1996). Furthermore, treatment of human cells with cytotoxic concentrations of BaP or BPDE has resulted in the observation of induced G:C->T:A transversions although not under all conditions of exposure (Chen and Thilly, 1996). Based on these observations, it has been inferred that those TP53 mutations have been directly induced by components of cigarette smoke such as BaP.

Nearly 70% of total TP53 mutations are missense mutations, of which more than 95% are point mutations that are scattered over 250 codons. In this aspect, the TP53 gene differs from other tumor suppressor genes such as RB and APC that are frequently inactivated by deletions or nonsense mutations, and from oncogenes of the ras family which are activated by mutations at a small number of specific codons (Table 2). Moreover, the suitable target point mutations for the present study are the ones repeatedly reported in multiple tumors or persons so that we could enhance the chance of being able to make valid measurements of these mutations well above the detection limits of the system. As a consequence, it became crucial to identify TP53 mutation hotspots with respect to the type of cancer, particularly when different causative agents may induce tissue-specific hotspots. The hotspot is defined as one occurring at a higher relative frequency than expected from a random distribution (Poisson).

Based on the criteria that the target point mutations have an arguable association with chemicals in cigarette smoke and are hotspots in the human lung tumor mutational spectra, three TP53 gene point mutations were chosen from exon 7: C:G->T:A transition at bp742 (bp14069 in genomic sequence), G:T->T:A transversion at bp746 (bp14073) and G:C->T:A transversion at bp747 (bp14074).

Table 2. Types of mutations in TP53, APC and RB1 genes

	TP53	APC	RB1
Neucleotide substitutions	83 (86%)	191 (35%)	96 (46%)
missense	66 (68%)	18 (3%)	20 (10%)
nonsense	5 (5%)	139 (26%)	37 (18%)
splicing	12 (12%)	32 (6%)	36 (17%)
regulatory	0 (0%)	2 (0.4%)	3 (1%)
Small deletions	10 (10%)	232 (43%)	66 (32%)
Small insersions	2 (2%)	81 (15%)	29 (14%)
Small indels	0 (0%)	6 (1%)	7 (3%)
Gross deletions	2 (2%)	19 (4%)	8 (4%)
Gross insertions & duplications	0 (0%)	4 (1%)	1 (0.5%)
Complex rearrangements**	0 (0%)	9 (2%)	0 (0%)
Total	97 (100%)	542 (100%)	207 (100%)

* The numbers are the entries in the Human Gene Mutation Database (2003).

** Complex rearrangements includes inversions

Bp746 and bp747, both of which are located in codon 249, were chosen because the type of mutation, G:C->T:A transversions, is a hallmark of lung cancers observed in smokers. Codon 249 is the third most mutated codon in the TP53 gene followed by codon 273 and codon 248 in the human lung cancer mutational spectra. The chosen mutations cause amino acid changes at codon 249 from Arginine (A**GG)** to Methionine (A**T**G at bp746) or to Serine (AG**T** at bp747), which is a critical site of DNA binding domain of TP53 protein. These loci are also reported to be mutation hotspots for radon exposure in the lung (Hussain et al., 1997) and a fungi toxin Aflatoxin B1 exposure in the liver (Hsu et al., 1991; Bressac et al., 1991). Trans-4-Hydroxy-2-nonenal (4-HNE), a major product of lipid peroxidation, possibly produced by oxidants and free radicals present in cigarette smoke, also induces adduct at TP53 codon 249. The adduct could result in G:C to T:A transversions at bp747 in TK6 lymphoblastoid cells if not repaired (Hussain et al., 2000; Feng et al., 2003).

The C:T->T:A mutation at TP53 bp742 was chosen because it is a different type of mutation (transition) from the former two TP53 target choices (transversion). This mutation is reported to be caused by another potent tobacco carcinogen N-nitrosomethylbenzylamine *in vitro* through formation of O6-methylguanine adducts in rat esophageal papillomas (Wang et al., 1996). The mutation was also observed in human fibroblasts treated with a potent alkylating agent, N-ethyl-N-nitrosourea (Hussain et al., 1994). This mutation is not lung specific; many other cancers, including colon, exhibited hotpots at this locus, thus, useful for the study of tissue comparison.

In the IARC lung cancer database (IARC, 2002), these three mutations comprise 1.6% (bp742), 1.8% (bp746) and 1.6% (bp747) of all TP53 somatic tumor mutants (n = 1730). Figure 6 separates the TP53 mutant spectra of the lung and bronchial tumors into smokers and non-smokers. Their patterns of mutations are somewhat similar in these two groups; furthermore, the frequency of all three target point mutations; C:G->T:A at bp742, G:C->T:A at bp746 and G:C->T:A at bp747, are slightly higher in non-smokers (smoker vs. non-smoker: 1.8% vs. 2.6%, 2.6% vs. 3.0%, 1.5% vs. 1.7%, respectively), contradicting the proposed association of these mutations with cigarette smoking. It should be noted, however, that this discrepancy might originate from two flaws in the IARC database. Firstly, the IARC database is strongly biased towards mutations occurring in exons 5

through 8 that are the exons most frequently screened. The TP53 mutations might be distributed more sparsely outside of exon 5-8, hence result in slightly different hotspot frequency. Secondly, because the non-smoker group (n=230) is much smaller than the smoker group (n=723), the mutant spectra of non-smokers may be less robust. Therefore, more screening of the entire TP53 gene sequence and a larger number of non-smoker entries might reveal different patterns of distribution or more, or less, distinct difference from smokers.

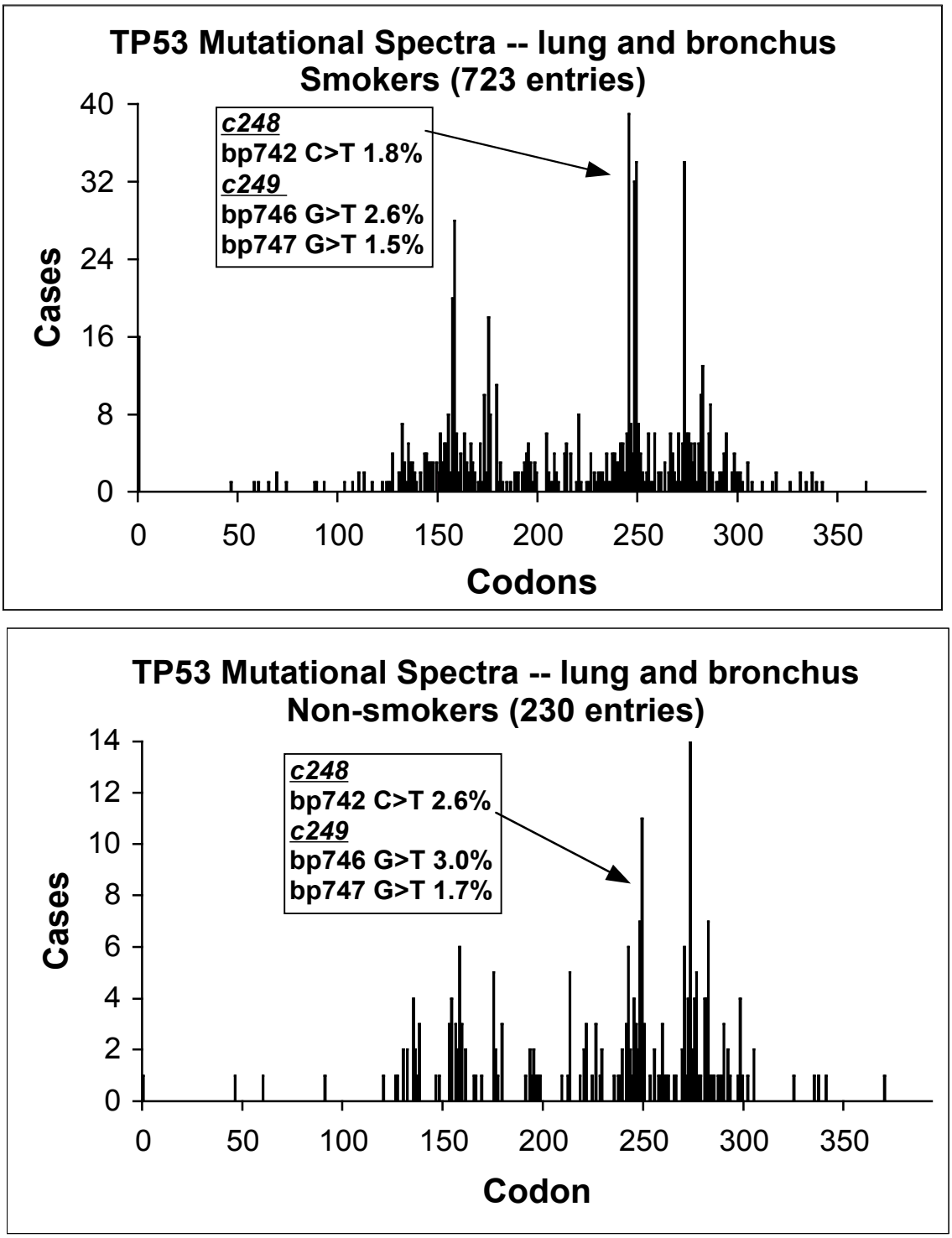


Figure 6. TP53 mutational spectra of lung and bronchial tumors of smokers (top) and non-smokers (bottom). The data was compiled from the IARC database. All three chosen hotspots: C:G->T:A at bp742, G:C->T:A at bp746 and bp747, are more common in non-smoker's lung tumors than smokers'. Codon 248 and 249 is the second and third most mutated codon in TP53 gene followed by codon 273 in the human lung cancer mutational spectra.

2.2.4. K-ras bp35

Ras is a membrane-bound guanosine triphosphate (GTP)/guanosine diphosphate (GDP)-binding G-protein that serves as a "molecular switch," converting signals from the cell membrane to the nucleus. These chemical signals lead to protein synthesis and regulation of cell survival, proliferation, and differentiation (Valencia et al., 1991; Boguski and McCormick, 1993). K-ras is the most frequently mutated gene among three in the ras family: K-ras, H-ras, and N-ras. The gene is located in 12p 12.1 and codes 189 amino acids.

Many tumors, including lung, pancreas and colon, contain K-ras mutations (Berrozpe et al., 1994; Vogelstein et al., 1988). Approximately 50% of lung tumors contain sectors with K-ras mutations (Mills et al., 1995), and many of such tumors are adenocarcinomas that represent the most common histological type of lung cancer among non-smokers.

It was previously argued that K-ras mutation is an early event in the lung carcinogenesis, particularly in adenocarcinomas (Westra et al., 1993; Li et al., 1994), but others disagreed (Sugio et al., 1994). K-ras does not seem to be a proto-oncogene whose mutation triggers oncogenic activity at the first step of the carcinogenesis because no lung tumors have been reported to carry a single K-ras mutant precursor cell throughout the tumor as would be expected for a mutation required for initiation or promotion.

The largest database on the frequency of the K-ras activating mutant sectors in human lung tumors has been compiled by Rodenhuis and Sleobos who have screened approximately 280 human lung tumors for such mutations (Rodenhuis et al., 1987; Rodenhuis and Slebos, 1992). Of the 280 tumors tested, 49 were found to carry sectors with an activating mutation in the ras gene. Importantly, among the tumors with an activated ras gene, 45 out of 49 were mutations in K-ras colon 12. Other reports have also confirmed that K-ras mutations involve only certain amino acids, predominantly in codon 12 (Kern et al., 1994), resulting in amino acid change from Glycine to Valine (G**G**T->G**T**T) which interferes with the GTPase function (Lerosey et al., 1991). Moreover, similar to TP53, it has been reported that this mutation at codon 12 of K-ras in the lung tumors is strongly associated with cigarette smoking (Rodenhuis and Slebos, 1992; Westra et al., 1993; Gealy et al., 1999; Ahrendt et al., 2001). G:C->T:A

transversion at bp35 in the 12th codon of K-ras gene, a mutation frequently observed in adenocarcinomas, was thus an appropriate choice for this study. This lung tumor type is the most frequent type of lung cancer among non-smokers and represents a rising percentage of lung tumors including those in smokers.

2.2.5. HPRT bp508

Hypoxanthine-guanine phosphoribosyltransferase (HPRT) protein salvages purine base products in the nucleotide metabolic pathway. HPRT enzyme recycles 6-oxopurine bases that are eventually converted back to ATP or GRP. HPRT gene is sex-linked and located in chromosome Xq26-27.2, thus hemizygous in males and functionally hemizygous in females due to X-inactivation. Germline mutations in the HPRT locus may lead to hyperuricemia, a debilitating neurological disorder, Lesch-Nyhan syndrome, or gouty arthritis, whereas somatic mutations are compatible with continual survivals of the subjects.

The HPRT gene is one of the most studied human genes. Besides its usual biochemical substrates: hypoxanthine and guanine, HPRT enzyme also phosphoribosylates purine analogue, 6-thioguanine, and the resistance of the analogue provides a highly efficient selection system for HPRT mutant cells, allowing them to grow while wild-type cells are killed (Szybalska and Szybalski, 1962; Albertini et al., 1982). The so-called HAT (hypoxanthine, aminopterin and thymine) reverse selection assay has enabled researches to establish both the frequency and spectrum of mutants in circulating T-lymphocytes and has offered a great amount of insights of molecular mechanism of mutagenesis *in vivo*. Using this method, many have measured HPRT mutant frequency in T-cells from smokers and non-smokers. The effect of smoking on HPRT mutations in T-cells is, however, inconsistent as described in section 2.1.3.5.

There is no evidence suggesting HPRT to be an important gene in lung carcinogenesis. Unlike TP53 and K-ras, HPRT is neither a tumor suppressor gene nor a proto-oncogene that might be selected against or for during the process of tumorigenesis. For this reason, HPRT was chosen as a "cancer-neutral" gene to serve as a control against oncogenic mutations which could be selected for during clonal expansion in the lung, thus bias numerical observations.

The C:G->T:A transition mutation at bp508 in codon 169, exon 7 of the HPRT gene is a nonsense, introducing stop codon (**C**GA->**T**GA). The mutation is a spontaneous hotspot with regard to both *in vivo* (somatic and germinal) and *in vitro* (TK6 cells with 1.9% frequency. Tomita-Mitchell et al., 2003); the mutability possibly accounts for methylation of the CpG sequence and its deamination. This mutation has been observed *in vitro* studies of human fibroblasts treated with BPDE (Yang et al., 1991) as well as T-lymphocytes treated with acetaldehyde, another tobacco smoke mutagens (Noori and Hou, 2001). The mutation is also reported in *in vivo* studies of T-lymphocytes of smokers and non-smokers (Vrieling et al., 1992; Burkhart-Schultz et al., 1996).

2.3. Technology for Detecting Rare Genetic Events From Human Tissues

2.3.1. Methods for mutational analysis

Detecting rare genetic events has become essential for human genetic studies, particularly after the complete decipherment of an entire human genome. A number of human genetic studies, such as early detection of cancer with known mutations, establishment of mutational spectra and mutagenesis induced by carcinogens and identification of oncogene mutations in normal tissues, require effective methods with high sensitivity in detecting rare mutants against the background of a vast excess wild-type sequences. Particularly, in order to distinguish the signals between pre-existing low-frequent mutations and induced or spontaneous mutations occurring during the treatment time period, the detection system must permit highly sensitive and rapid screening for mutations in a large number of genes.

In general, methods for mutant detection can be divided into two categories: phenotype-based and genotype-based. HAT reverse selection system is a classic phenotypic selection assay for HPRT mutants from circulating T-lymphocytes in human blood (Albertini et al. 1982). The method utilizes a purine analogue, 6-thioguanine, which is toxic for healthy cells, and only HPRT mutants can selectively survive. Although this method is widely used for determination of HPRT mutant frequencies and mutation spectra, the results are largely affected by cloning efficiency, and the application is mostly limited to T-lymphocytes in human study. Similar to HAT assay,

most phenotype-based assays are limited to selectable genes, and only applicable to cultured cells and to certain tissue types, the cells of which can be grown *in vitro*.

Genotypic mutant selection has a few major advantages over phenotypic mutant selections that inevitably requires clonal expansion of a particular mutant. First, genotypic selection methods enable to study early carcinogenesis in which some tumor suppressor genes or proto-oncogenes may be mutated but have yet manifested no recognizable phenotypic profiles. Second, genotypic selection allows us to directly analyze a larger pool of target genes and tissues of interest that may be responsible for specific diseases.

A number of amplification methods have been developed to determine single base mutations. These methods are mainly categorized into three groups. The first group takes an advantage of existing or introduces new restriction sites into PCR products at the mutated codons and selectively destroys the abundant wild-type alleles. This includes Restriction Fragment Length Polymorphism / Polymerase Chain Reaction (RFLP/PCR). RELP/PCR uses restriction enzymes to reduce the amount of wild-type sequence and PCR amplification with primers flanking the restriction site to amplify the uncut, mutant sequences (Kan and Dozy 1978; Felley-Bosco et al. 1991; Chen and Viola, 1991; Sandy et al., 1992). In this technique, the multiple cycles of restriction digestions and mutant enrichment by PCR are necessary to achieve sufficient sensitivity (Wilson et al. 1999). The sensitivity of this assay is limited by the efficiency of restriction digestion which is highly dependent upon sequence context. In addition, target sequences may not always contain an appropriate restriction site and are required to introduce a new restriction site by PCR. Overall, the range of suitable sequences for this assay is limited.

The second type of methods spatially separates the mutant and wild-type alleles based on conformational difference induced by denaturation. Single Strand Conformation Polymorphism (SSCP), Denaturing Gradient Gel Electrophoresis (DGGE) and Constant Denaturing Capillary Electrophoresis (CDCE) belong to this category. These methods employ denaturation of double strand DNA in which small variations in the target sequence may alter the conformation of single strand DNA (SSCP) or mutant / wild-type heteroduplexes (DGGE and CDCE), thus consequently its electrophoretic profile (Orita et al., 1989; Fischer and Lerman, 1983; Keohavong and Thilly, 1989;

Khrapko et al., 1994). The advantage of this approach is that mutational information is not required prior to the experiments; i.e., any unknown mutations in the target sequence of interest can be screened. CDCE carries the best sensitivity among the methods in this category; it can detect mutants at a fraction of as low as 10^{-6} (Khrapko et al., 1999; Muniappan and Thilly, 1999; Li-Sucholeiki and Thilly, 2000). However, this approach is only applicable to those target sequences juxtaposed to a 'GC clamp', a higher melting temperature sequence, in genomic DNA. Clamp ligation technique has been recently developed and made CDCE applicable to any sequences in genome. The current sensitivity achieved using this technique is $2x10^{-5}$ (Kim and Thilly, 2003).

The third group is so called allele-specific PCR, utilizing specifically designed oligonucleotide primers and probes that selectively amplify rare mutant alleles by PCR. Allele-specific PCR is a commonly applied method for speedy detection of known point mutations. Specificity is achieved by designing an oligonucleotide to match the desired allele but mismatch other alleles at its 3' end. The mismatch between the DNA and the oligonucleotide results in specific amplification of the desired allele, or mutants, and little or no amplification of the undesired alleles, or wild-type, by preventing elongation at the 3' end by DNA polymerase. Such methods includes: Allele-Specific Oligonucleotide / PCR (ASO/PCR) (Saiki et al., 1986), PCR coupled with Oligonucleotide Ligation Assay (PCR/OLA) (Delahunty et al., 1996), Amplification Refractory Mutation System (ARMS) (Newton et al., 1989), Mismatch Amplifying Mutation Assay (MAMA) (Cha et al., 1992), PCR Amplification of Specific Alleles (PASA) (Sommer et al., 1989), PCR-primer-introduced restriction with enrichment for mutation alleles (PCR-PIREMA) (Jacobson and Mills, 1994) and Blocker-PCR (Seyama et al., 1992).

Unfortunately, most of these methods are suitable for the detection of germ line mutations where the alteration occurs in at least one allele in each cell, or somatic mutations in tumor cells that give rise to populations of cells in which the ratio of mutant to normal alleles is high. Each of these techniques has some limitations. For example, although highly sensitive, the PCR-PIREMA requires 2 to 3 rounds of PCR involving a total of 80 to 100 amplification cycles, one step of gel purification and two rounds of restriction enzyme digestion. Furthermore, the technique was reported to be prone to false positives. As noticed by the authors, these false positives could have been produced

by DNA polymerase during the numerous amplification cycles or by the presence of uncut wild-type bands due to incomplete digestion by restriction enzymes. Likewise, although RFLP/PCR has been applied for detection of rare mutations from normal human tissue (Stork et al., 1991; Ouhtit et al. 1997), the method has several drawbacks: it is very labor intensive, the mutation must be within an appropriate restriction endonuclease sites and the cleavage efficiencies of many restriction endonucleases limit the sensitivity (the ratio of the minimum copies of the mutated template with detectable product to the maximum copies of the wild-type template with undetectable product) to a range of 10^3 to 10^5 (Persons and Heflich, 1997). In summary, most of available methods either do not achieve high selectivity or are not suitable for routine mass analysis. One of the exceptions is MAMA, a method with sufficient sensitivity to detect rare mutations in normal human lung tissue, as described in details in the next section.

2.3.2. Mismatch Amplification Mutation Assay (MAMA)

Our laboratory has developed a technique for measuring rare genetic events in the human genome. MAMA is a particularly effective form of allele-specific PCR and measures any specified point mutations from any human tissues. This technique is based on observations that correctly paired bases were much more readily extended by DNA polymerases than mispaired bases. By measuring the extension efficiencies of different mismatch primers hybridized, perfectly matched bases were 10^2-10^6 times more efficiently extended than mismatched bases, depending on DNA polymerases and the nature of mismatches (Petruska et al., 1988; Mendelman et al., 1990).

However, using G:C->A:T mutation in the 12th codon of the rat c-H-ras gene, Cha et al. (1992) demonstrated that the allele-specific primers carrying a single mismatch had no effect on the overall PCR yield, amplifying the target sequence with the efficiency of 70% per cycle during the exponential phase of DNA synthesis, whereas two or more consecutive mismatches at the 3' end failed to generate any detectable amplification product. Therefore, designing a primer which contains a single mismatch with the desired mutant but double mismatches with the wild-type sequence enables to extend and amplify only mutant DNA copies. Cha also demonstrated that MAMA could detect 30 copies of the transformed alleles present amongst $3x10^6$ copies of the wild-type alleles under the

optimum condition. The great advantage of this technique is its simplicity, speed, absence of expensive probe or radioactive labeling and sensitivity sufficient to be applicable for normal human tissue analysis. Furthermore, MAMA generates quantitative output, providing an opportunity to measure the actual fraction of the DNA sample carrying point mutations. This highly sensitive technology is now used in various applications (Cha et al, 1994; Zirnstein et al., 2000; Jinneman and Hills, 2001) and was employed in the present study to measure specific point mutations in the normal human lung epithelial tissue.

3. EXPERIMENTAL DESIGN AND METHODS

Figure 7 depicts the overall experimental procedure.

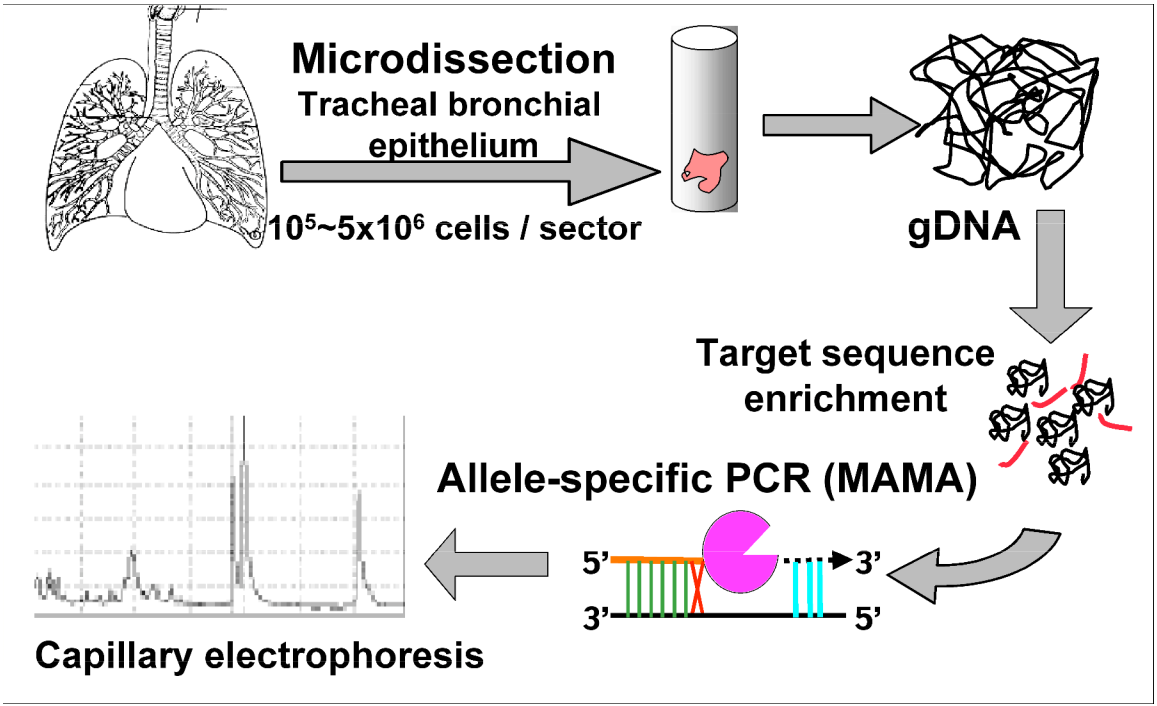


Figure 7. Strategy of the overall experimental procedure. The procedure consists of microdissection of normal human tracheal and upper bronchial epithelium, isolation of genomic DNA, restriction digestion, target sequence enrichments, MAMA and mutant detection by capillary electrophoresis (CE).

3.1. Cell Lines and Human Tissue Handling

3.1.1. Cell line controls

Human B-lymphoblastoid cell line TK6 was used as a negative control against human lung tissue. TK6 was originally derived from a 5 year-old male patient with hereditary spherocytosis (Levy et al., 1968) and has been commonly used for human cell mutation studies because it maintains its karyotype and remains pseudo-diploid in culture. TK6 has reported to contain at least 10-fold lower MF than human tissue in the mitochondrial sequences (Khrapko et al., 1997b). Furthermore, the presence of spontaneous or preexisting mutations in the TK6 DNA has previously reported the MF lower than 10^{-6} in the HPRT sequence (Tomita-Mitchell et al., 2003). TK6 was hence chosen as an ideal human DNA negative control in the present study. Before used, the cells were freshly recloned and frozen at its exponential growth phase by daily dilution of a 600ml culture at 10^6 cells/ml for more than 400 generations.

Positive controls were all derived from human tumor cell lines: rectal adenocarcinoma cell line SW 837 for C:G->T:A mutation in TP53 bp742; pancreatic adenocarcinoma cell line Hs700T for G:C->T:A mutation in TP53 bp746; hepatoma cell line PLC/PRF/5 for G:C->T:A mutation in TP53 bp747; and pulmonary adenocarcinoma cell line H441 for G:C->TA mutation in K-ras bp35. All mutant cell lines were purchased from American Type Culture Collection (Manassas, VA). Mutant alleles for C:G->T:A mutation in HPRT bp508 was created by PCR using one-mismatch (mutant) primer (5'- TTGCTGGTGAAAAGGACCCCA**T**GAAG –3', bold for mutation locus) and TK6 DNA as a template.

All reconstruction experiments were performed using these cell lines, and each assay was individually and extensively optimized before the use of human tissue.

3.1.2. Acquisition and dissection of human lung tissue

In collaboration with Professor James C. Willey in the Medical College of Ohio (Toledo, OH), the Willey laboratory transplant team collected epithelium from human tracheal and upper bronchial trees from six smoker and nine non-smoker donors and ensured the appropriate handling necessary for preservation of tissue architecture and viability. The protocol was approved by both the Medical College of Ohio Institutional

Review Board and the Committee on the Use of Human as Experimental Subjects at MIT. Informed consent was provided for each donor by his/her or their next of kin. The extensive medical records of all donors documented their smoking status, age, gender, occupation, previous history of illness, familial medical history and cause of death were also acquired (Table 3). In order to minimize the prejudice offered by pre-existing lung inflammation and diseases, only donors whose deaths were determined by physicians to be non-pulmonary related were chosen for the study. Their causes of death were mainly stroke (subarachnoid hemorrhage cerebrovascular) or accidents. Trachea and bronchus were obtained immediately postmortem and showed no sign of respiratory disease determined by medical records and physical examination. Following the Plopper technique of lung dissection (1983), the airway was cut into segments, starting from the first bifurcation of each lobe down to six or seven generations of airways. Tracheal/bronchial epithelial cells were then scraped off from the airways by a fresh clean scalpel. Each lung provided multiple, different number of sectors. All sectors were catalogued with both a description of its anatomical position and a binomial number as per the scheme devised by Phalen and Oldham (1983), e.g., LUL1 (left-upper-lobe) for the first section of the upper bronchus in the left lobe; RML3 (right-middle-lobe) for the third section of the middle bronchus of the right lobe and so on. The dissection of airways, anatomical excision of individual sectors, removal of the epithelium by scraping off the luminal surface and histological verification of normal epithelium under microscope were all performed by the Willey laboratory. Each tissue sector was placed in an eppendorf tube, then shipped to us in frozen with dry ice overnight. Most of the sector samples yielded about 10^5- $5x10^6$ cells.

As a control organ, colon epithelium was also acquired similarly to the lung. Colon epithelium was provided by Prof. Beth Furth in the Department of Pathology and Laboratory Medicine, University of Pennsylvania Medical School (Philadelphia, PA). The multiple colon epithelial sectors containing 10^7-$5x10^7$ cells per sector were collected from surgical discards from 58 year old male and 60 year old female.

Table 3. Clinical samples: tracheal bronchial epithelium

		Age	Gender	Smoking status	Sectors	Mutation assays performed
Smokers	I	41	female	1.5ppd x 7yr	9	43
	II	41	male	0.5 ppd	10	48
	III	47	female	1.0ppd x 25yr	67	134
	IV	55	male	1.5ppd x 26yr	23	94
	V	58	female	1.5ppd x 25yr	41	96
	VI	59	male	>2ppd	4	10
total					**154**	**425**
Non-smokers	I	38	female	Non-smoking	5	16
	II	40	female	"	9	19
	III	41	male	"	3	12
	IV	45	female	"	42	193
	V	50	male	"	50	117
	VI	59	female	"	52	109
	VII	67	male	"	3	5
	VIII	75	female	"	10	48
	IX	76	female	"	3	5
total					**177**	**524**

*ppd stands for package per day

3.1.3. Isolation of genomic DNA from cultured cells and human lung epithelial sectors

Genomic DNA isolation was processed using QIAGEN Blood & Cell Culture DNA Midi Kit (QIAGEN, Valencia, CA). In brief, cultured cell lines were pelleted by centrifugation, resuspended with PBS, lysed with sucrose and triton, and digested with proteinase K at 50C for one hour. Solid human lung tissue sectors were first cut into small pieces with clean surgical scissors, vigorously vortexed, then similarly lysed and digested with proteinase K and RNase for 2 hours or more until any visible particulates were dissolved. Digested samples both from cultured cells and tissue were applied to columns of anion-exchange resin which was washed and eluted with manufacturer's recommended buffers. DNA was isolated by isopropanol precipitation, washed with 70% ethanol and dissolved in 0.1xTE.

Concentration of the isolated genomic DNA per sector was assessed by two independent methods: UV spectrometer and quantitative PCR. In the former method, DNA was measured at 260nm wavelength and the concentration was calculated as:
A_{260} x 50□g/ml x dilution factor
Based on this method, the DNA content of anatomically distinct lung epithelial sectors ranged typically from 3to 90□g per sector, equivalent to 0.15 ~ 1.35 x 10^7 cells. In the latter methods, the internal standards whose copy numbers had been previously calculated were utilized for the assessment in quantitative PCR which is described in the next section. The concentrations measured by these two methods were compared and re-checked if they differed significantly. In general, the measurements by UV spectrometer and quantitative PCR corresponded with each other. Based on the concentration measured in this step, the total numbers of cells in each sector were calculated.

3.2. PCR-CE Setup – Quantitative PCR Using IS

This section describes "normal" PCR condition, distinguished from "MAMA" PCR. Normal PCR is mainly performed to check copy numbers of the target DNA molecules in samples.

3.2.1. Design primers

Table 4 summarizes the sequences of all primers and probes used in this study. The primer and probe sequences were designed using an oligo primer analysis software,

Oilgo 5.1 (Molecular Biology Insights, Inc, Cascade, CO), to search for the appropriate melting temperature (47-55°C), and to avoid palindromes and repeat sequences that contribute to hairpin loops and primer dimers. Primers and probes were synthesized and HPSF -purified by MWG Biotech, Inc (High Point, NC) or Synthetic Genetics (San Diego, CA, now Epoch Biosciences, Bothell, WA). All normal and MAMA primers were 20mer long; all IS primers were 35 mer long; and all probes were 25 or 30mer long.

Table 4. Sequences of normal, internal standard and MAMA primers and biotin labeled probes for TP53, K-ras and HPRT gene target.

* Bold nucleotides in MAMA primers indicate the critical mutation site.

[TP53 target sequence]	
Biotin labeled probe 1:	5'- CATGTGTAACAGTTCCTGCATGGGC -3'
Biotin labeled probe 2:	5'- TTCCAGTGTGATGATGGTGAGGATG -3'
P1 (upstream normal primer):	5'- ATCTCCTAGGTTGGCTCTGA -3'
P2 (downstream normal primer):	5'- GCAAGTGGCTCCTGACCTGG -3'
P1-IS (upstream IS, 10bp del):	5'- ATCTCCTAGGTTGGCTCTGAATCCACTACAACTAC -3'
P2-IS (downstream IS, 10bp del):	5'- GCAAGTGGCTCCTGACCTGGTGTGATGATGGTGAG-3'
bp742 MAMA primer (upstream):	5'- CTGCATGGGCGGCATGAA**AT** -3'
bp746 MAMA primer (downstream):	5'- GATGATGGTGAGGATGGG**TA** -3'
bp747 MAMA primer (downstream):	5'- TGATGATGGTGAGGATGG**TA** -3'
[Ras target sequence]	
Biotin labeled probe 1:	5'**-** CTGAATATAAACTTGTGGTAGTTGGAGCTG -3'
Biotin labeled probe 2:	5'**-** AATGATTCTGAATTAGCTGTATCGTCAAGG -3'
P1 (upstream normal primer):	5'- CCTGCTGAAAATGACTGAAT –3’
P2 (downstream normal primer):	5'- GATTCTGAATTAGCTGTATC -3'
P1-IS (upstream IS, 7bp del):	5'- CCTGCTGAAAATGACTGAATTGTGGTAGTTGGAGC –3’
bp35 MAMA primer (downstream):	5'- GGCACTCTTGCCTACGCC**GA** -3'
[HPRT target sequence]	
Biotin labeled probe 1:	5'- GTCCTTTTCACCAGCAAGCTGTTAATTACA -3'
Biotin labeled probe 2:	5'- GGATATAAGCCAGACTGTAAGTGAATTACT -3'
P1 (upstream normal primer):	5'- AACAGCTTGCTGGTGAAAAG -3'
P2 (downstream normal primer):	5'- AAAGATGGTTAAATGATTGA -3'
P2-IS (downstream IS, 8bp del):	5'- AAAGATGGTTAAATGATTGAGTAATTCACTTACAG -3'
bp508 MAMA primer (upstream):	5'- GCTGGTGAAAAGGACCCC**TT** -3'

3.2.2. Normal PCR condition

PCR was preformed inside of closed 10ul- or 50ul-scale glass capillaries using an Air Thermo-Cycler™ (Idaho Technology, Idaho Falls, ID). The PCR mixture contained 0.2uM each primer, 100uM each deoxinucleoside triphosphates (dNTP) (Amersham Pharmacia Biotech Inc., Piscataway, NJ), 100□g/ml bovine serum albumin (BSA) (New England Biolabs, Beverly, MA), 1 unit of *Thermus aquaticus (Taq)* DNA polymerase (New England Biolabs) and the specified polymerase buffer supplied by the manufacturer. Each PCR procedure consisted of 94°C for 2 min of polymerase activation, cycles of 94°C for 12 sec of template denaturation, 50°C or 48°C for HPRT sequence for 18 sec of template-primer annealing and 72°C for 18 sec of primer extension, followed by 72°C for 2 min of post-PCR extension and 45°C for 15 min of post-PCR incubation to remove interfering PCR byproducts. 10^3 – 10^6 copies of the templates were usually introduced and 36 or 38 cycles were performed depending on the amplification efficiency. PCR efficiency was estimated by measuring sequence amplification during the exponential phase of PCR and calculated using the following equation:

$N_{final} = N_{initial} (1+e)^C$

where $N_{initial}$ is the number of starting copies, e is efficiency, c is the number of PCR cycles and N_{final} is the number of copies measured after c PCR cycles. For most of PCR, 65% efficiency or above was usually achieved for all of the target sequences.

3.2.3. Mutant internal standards

For each target mutation, internal standards were created by using mutant templates and internal standard primers (Table 4) which are based on 20bp-long normal primers but skipped 7 - 10bp at the 3' end then continued another 15bp, resulting in amplification of shorter PCR products. Except the deletions, the internal standard sequences shared the same DNA sequence as the mutant and could be co-amplified by the same primers. The internal standard PCR products were over-amplified (~$6x10^{10}$ copies/ul) with 1.2 unit of *Taq* polymerase and extra PCR cycles, followed by 10-fold serial dilutions with 0.1 TE to produce the concentrations of 10^9, 10^8, 10^7, … 10^2 copies/ul. To check whether internal standard would be amplified with the same

efficiency as the fully-long target sequence by the same primer sets, internal standards were mixed with the TK6 DNA and amplified. The numbers of molecules were compared at different PCR cycle numbers (Figure 8). The results suggested that the deletion in the internal standard sequences did not affect PCR efficiency and both the internal standards and the target sequences were amplified with a remarkably similar efficiency.

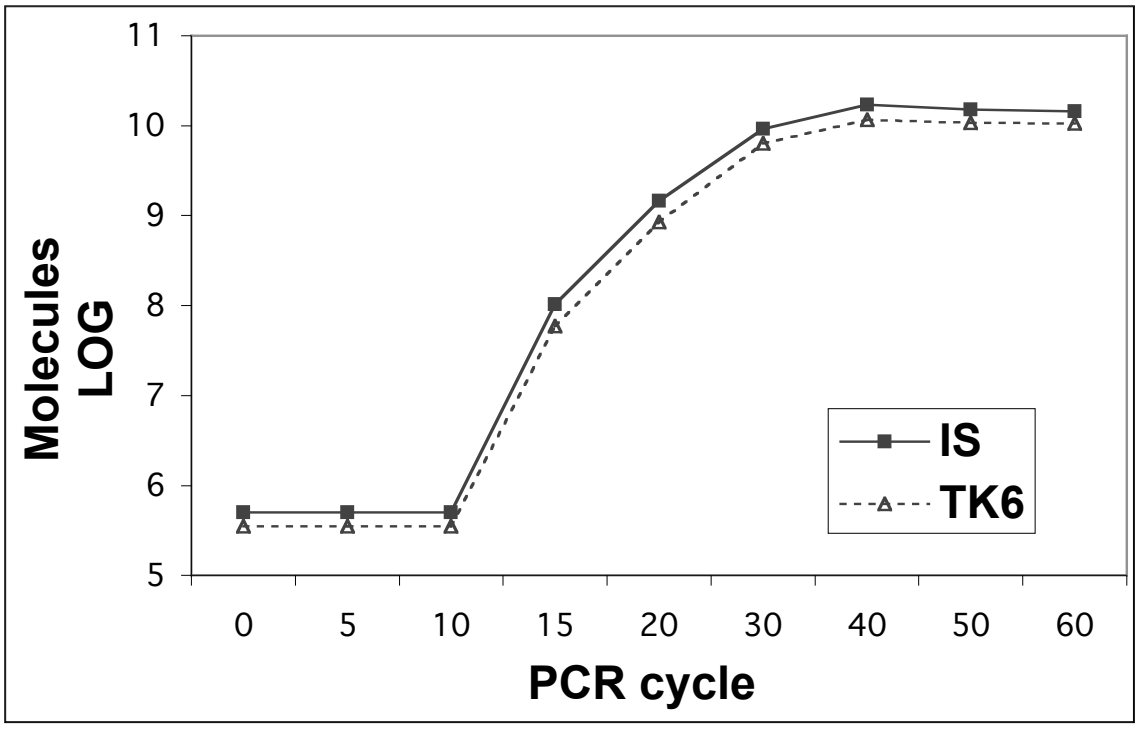


Figure 8. Competitive PCR between TK6 DNA and an internal standard in the TP53 sequence. The internal standard (IS) was previously created from TP53 mutant DNA by using P1 and P2-IS primers which produced a 10bp shorter product than the one created by P1 and P2 primers. Approximately $5x10^5$ copies of internal standard and TK6 template were mixed and amplified with P1 and P2 primers and the products were measured every 5 cycles. The PCR efficiency of the target and internal standard sequences were remarkably similar (66% and 67%, respectively).

3.2.4. Detection of PCR products by capillary electrophoresis (CE)

CE was performed similarly to that described previously (Khrapko et al., 1994). All CE was performed at room temperature and separated DNA molecules (primers, internal standards and target products) based on their lengths (Cohens et al., 1988). Electrophoresis was performed within a glass capillary with the inner diameter of 75μm and the outer diameter of 250μm (Polymicro Technologies, Phoenix, AZ). The capillary was pre-coated with 6% polymerized acrylamide to prevent electroosmotic flow. All coating materials were injected by a 100ul high-pressure syringe (SGE Inc., Austin, TX) of which needle end was inserted with a piece (~10cm) of Teflon tubing for reagent transmission. For the coating, capillary was treated with 1M NaOH for 1hr, washed with ddH_2O, filled with 1M HCl for 10 min, washed with 100% methanol and treated with γ-methacryloxypropyltrimehoxysilane overnight. After being washed with 100% methanol, the capillary was filled with a polymerizing solution consisting of 6% acrylamide in 1xTBE (89mM Tris, 89 mM borate, 1 mM EDTA, pH 8.4), 0.1% TEMED and 0.025% ammonium persulfate. The coated capillary was incubated at least one day before its use.

The coated capillary was cut 28cm in length with 0.8cm of an opened detection window at 18cm from the inlet. The gel matrix inside the capillary was replaced with 6% non-crosslinked polyacrylamide using the high pressure syringe before each CE run. A replaceable linear polyacrylamide matrix was prepared as described in previous studies (Khrapko et al., 1997a). To load a sample, the cathode was connected with a piece of platinum wire whose another end was submerged into a sample of 10-fold diluted PCR product which was also contacted with an inlet end of the capillary. PCR products were electro-injected onto the capillary column at 2μA for 20-30 sec. The electrophoresis was performed at 9μA for 15 min in average. Fluorescent molecules on the primers were excited by a 488nm argon laser beam (Ion Laser Technology, Salt Lake City, UT) and the emitted light was detected by a photomuliplier after passing through two filters of a 540nm and 530nm (Oriel, Stanford, CT). A computerized data acquisition system recorded A/D converted signals (Workbench Mac v.4.01, Strawberry tree, Sunnyvale, CA) that were analyzed by AcqKnowledge™ v.2.1.2 (BIOPAC System Inc., Goleta, CA). Figure 9 shows CE test separation of the internal standard from the target sequence in

TP53 gene. Based on the results of positive and negative control test runs, the relative positions to which the target sequences migrated through the gel could be identified.

To measure the copy number of the target sequences, a small aliquot of the sample was mixed with an internal standard whose copy number was previously calculated. The initial target copy number (Ni) was calculated based on measurement of the areas under the internal standard peak (A_{IS}) and the target peak (A_T):

$Ni = N_{IS} \times A_T / A_{IS}$

where N_{IS} is the internal standard copy number initially mixed with the sample (Figure 9).

The target copy number in each lung epithelial sector was individually measured twice during the whole experimental process: after genomic DNA isolation and after target sequence enrichment. The former was important to determine the total numbers of cells per sector, which should be comparative to the values measured by UV spectrometer. The latter is important to check the enrichment recovery, as well as to determine exact amounts of target molecules applied for MAMA.

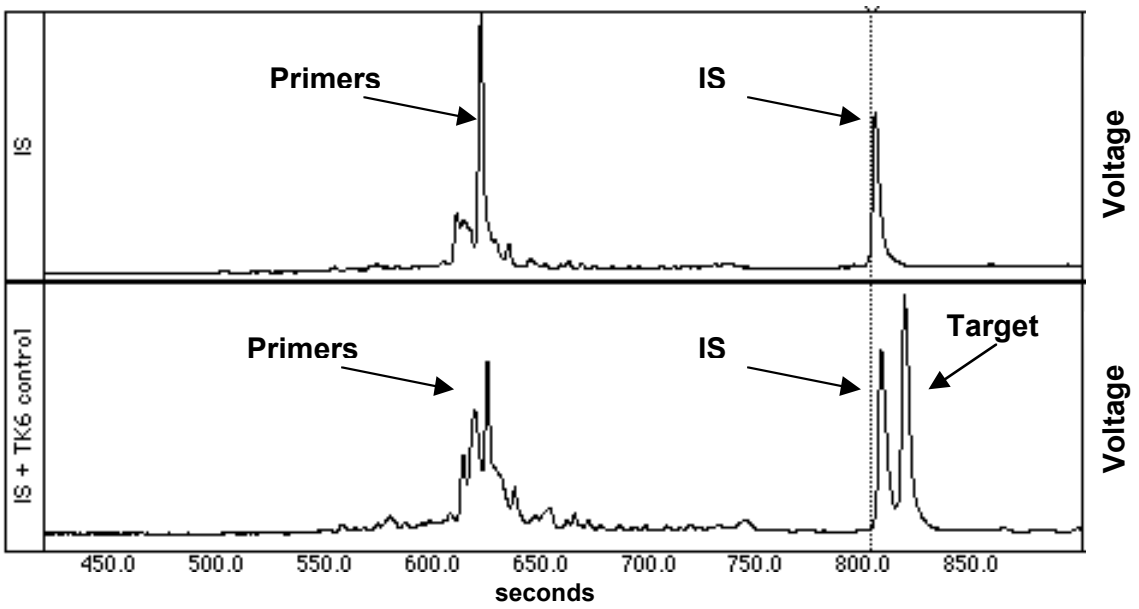


Figure 9. CE test runs for quantitative PCR. In the top chart, 10^4 copies of TP53 IS was amplified with P1 and P2 primers; at the bottom, a small aliquot of TK6 genomic DNA was mixed with 10^4 copies of the internal standard (IS) and amplified similarly. Because the internal standard product was 10bp shorter than the target product, its peak appeared earlier than the target peak. By taking a ratio of the areas under the peaks, the number of target TK6 molecules initially introduced to PCR was estimated to be 1.25 x 10^4 copies.

3.3. Target Sequence Enrichment

3.3.1. Restriction enzyme digestion

The major challenge for detecting rare genetic events in nuclear genes is that a large amount of genomic DNA is required to be assayed. However, the large input usually impedes performance of most of mutation assays. In this study, the quantity of DNA introduced to the assay was significantly reduced by employing a sequence-specific enrichment strategy to highly enrich the desired nuclear gene fragment. This strategy has been previously established by Li-Sucholeiki and Thilly (2000).

Isolated genomic DNA underwent double restriction digestions with BmsAI and SspI (New England Biolabs, Beverly, MA) to liberate the three specific sequences for the TP53, K-ras and HPRT target sequence enrichment. BmsAI was used to cut both sides in the exon 7 of TP53 gene (bp13960 and bp14502) to release the 543bp-long TP53 target sequence. BmsAI also cuts the upstream in the exon 7 of HPRT gene (bp39781), and SspI cuts the downstream (bp39950) to release the 382bp-long HPRT target. In the K-ras sequence, BmsAI cuts 364bp upstream of the start codon, and although the exact locus of the downstream cut was not clear due to a lack of intronic sequence information, the double digestion of BmsAI and SspI seemed to efficiently release the K-ras target sequence of which the length was estimated to be between 558bp and 1 kb.

The reaction mixture of restriction digestion contained 2U of restriction enzyme per □g of DNA, 0.1mg/ml of BSA and digestion buffer (NE buffer 2) supplied by the manufacturer. Genomic DNA was first incubated with the BmsAI digestion mixture at 55°C for 6 hours, then with the SspI mixture at 37°C for another 6 hours.

3.3.2. Target sequence enrichment

Enrichment of desired target sequences from genomic DNA was previously demonstrated by Li-Sucholeiki and Thilly (2000). The protocol is based on probe-target hybridization coupled with a biotin-streptavidin capture system. The probes (Table 4) were designed as to minimize intra-probe hairpins and inter-probe duplex formations and to share similar melting temperatures so that they can be simultaneously used in the same hybridization reaction. The enrichment procedure was performed twice; the first round was to enrich the TP53 and HPRT target sequences, and the second round was to enrich

the K-ras target sequence. In the first enrichment, 0.6mg of each streptavidin-coated glass paramagnetic beads (CPG, Lincoln Park, NJ) and non-magnetic colored beads (Bangs Laboratories, Inc., Fishers, IN) per 60μg of DNA were washed twice with 3xSSPE (0.56M NaCl, 30mM sodium phosphate pH7.4, 3mM EDTA) at a concentration of 10mg beads/ml. The washing was performed by centrifugation at 14,000rpm for 5-10 min. After resuspended with 3xSSPE, 60pmole of HPRT biotin-labeled probes (BPs) for both strands, BP1 and BP2, were added to magnetic beads per 60μg DNA; similarly, 60pmole of TP53 probes, BP1 and BP2, were added to non-magnetic colored beads. These mixtures were incubated in a microthermomixer at room temperature for 30min. The probe-bound beads were then washed three times each with 3xSSPE at a concentration of 10mg beads/ml. Meanwhile, digested genomic DNA in a 1.6ml eppendorf tube was denatured by boiling water bath for 2 min and immediately chilled in an ethanol-water ice bath for 10 min. Both magnetic and non-magnetic probe-bound beads were then added to single-stranded DNA samples. 20xSSPE was added to each sample to a final concentration of 6xSSPE and hybridized in a microthermomixer at 50°C for 2 hr, rotating at the maximum speed. After incubation, the target-bead hybrids were washed with 6xSSPE at a concentration of 20mg bead/ml at 50°C for 5 min, then with 6xSSPE and 3xSSPE at the same concentration at room temperature. All washed supernatant which contained a rest of genomic DNA was kept for other target sequence enrichment. With 1xSSPE, the beads were repeatedly rinsed and magnetically separated the two types of beads until no magnetic bead was seen in non-magnetic bead solution. Both beads were further washed with ice-chilled 1/3xSSPE and resuspended with 10-15ul of ddH_2O at 20mg beads/ml. The target DNA sequence was eluted from the probe-bound beads twice at 72°C for 2 min. The released targets were separated either magnetically from magnetic beads or by centrifugation from non-magnetic colored beads, and transferred to a new tube.

In the second round of K-ras target enrichment, the supernatant DNA from the first enrichment was similarly boiled and chilled. The K-ras probes were added to the DNA and incubated in a microthermomixer at 50°C for 2 hr. An appropriate amount of magnetic beads were washed, added to the target-probe hybrids and incubated in a

microthermomixer at 50°C for another 1 hr to allow the probe to hybrid with the beads. The remaining washing and elution steps were identical to the first enrichment process.

3.3.3. Estimating the enrichment recovery

From the copy numbers measured by quantitative PCR using internal standards, distribution of the target sequences to the three aliquots: the elute portion of the interest, by-eluted another target portion and the supernatant portion, can be calculated. Enrichment recovery percentage was estimated as the ratio of the target copy number in the target elute of the interest (N_T) to the sum of the target copy numbers in the elute, by-enriched non-target elute (N_{NT}) and the supernatant (N_S):

$R = N_T / (N_T+N_{NT}+N_S)$

The recovery efficiency depended upon the target sequence context and the choice of probes. In general, the enrichment mostly achieved more than 80% efficiency. If the recovery was lower than 65%, the enrichment procedure was repeated from the supernatant.

3.4. Mismatch Amplifying Mutation Assay (MAMA)

3.4.1. MAMA primer design and choices

In general, MAMA primer are designed such that they contain one mismatch with mutant allele but double mismatch with wild-type allele, leading to preferential amplification of only mutant alleles. Previously, I empirically demonstrated that the mismatch at 3'penultimate is more efficient for discrimination than at 3'ultimate; thus, primers were designed such that its 3'ultimate nucleotide is assigned at a critical locus of point mutation (a match with mutation), and the penultimate position was open for three choices of nucleotide to create a mismatch basepair. Since MAMA primers can be designed either forward or backward, total six variations of potential primers were of the choice for each target mutation (Figure 10). The best primer, which had a minimum background signal and highest specificity (described details in the section 3.4.2.), was empirically selected for each target mutation. All the MAMA primers (MWG Biotech, Inc., High Point, NC or Synthetic Genetics San Diego, CA) tested were 20-nucleotide long. The choices of MAMA primers were listed in Table 4.

Taq DNA polymerase was chosen for MAMA since it lacks 3'- to 5'- exonuclease activity (Innis, et al., 1988) and still amplifies single mismatched mutant templates but discriminately prohibits amplification of double mismatched wild-type templates.

The preference of 3' penultimate base of a mismatched primer mediated by DNA polymerases is sequence dependent. *Taq* polymerase has been shown to extend A-C and G-T mismatches more efficiently than T-C and T-T mismatches and the A-G, G-G and C-C mismatches have the lowest probability of extension (Mendelman et al., 1990). Furthermore, the base downstream of the mismatch in the template strand is believed to have great influence in the extension efficiency. By considering these criteria, a high level of specificity should be achievable for most sequences. The careful designing of MAMA primers and PCR conditions are very important to the clean and clear detection of mutants without false positive or negative signals.

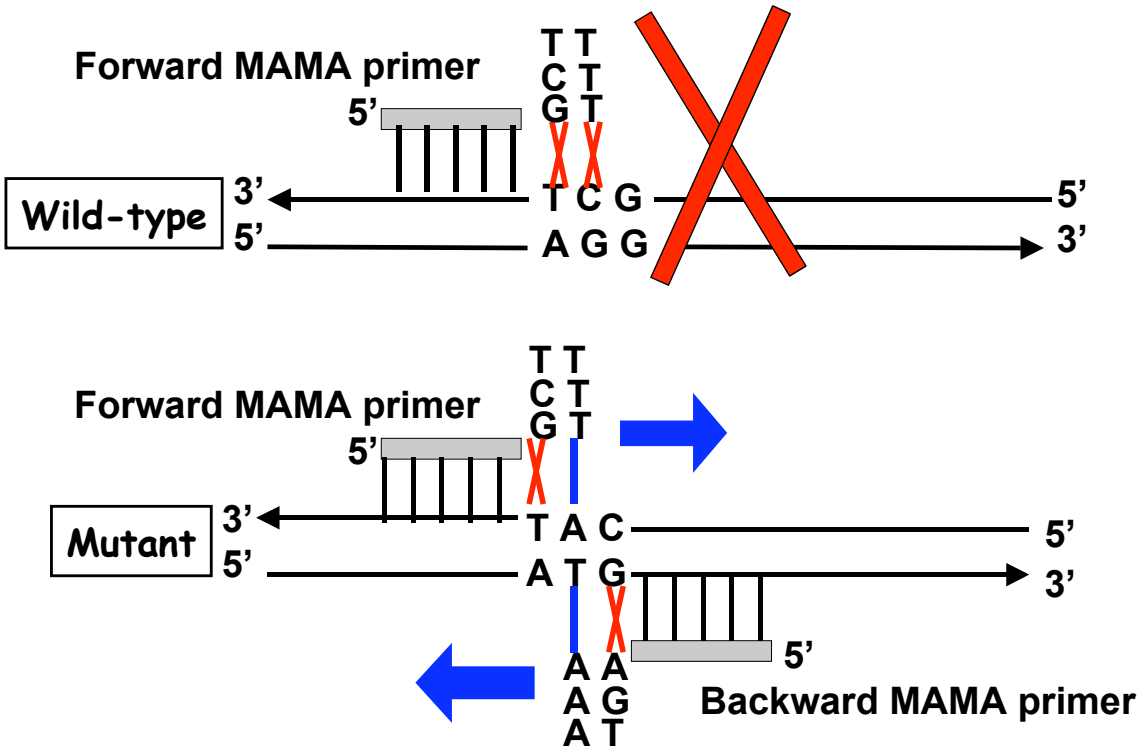


Figure 10. MAMA primer design of G:C->T:A mutation at bp746 in TP53. MAMA primers have double mismatches at the 3' end with the wild-type (top), impeding the primer extension by Taq polymerase, whereas they have a single mismatch at the3' penultimate end but a perfect match at the 3' ultimate end with the mutant, resulting efficient primer extension. Because the 3' penultimate nucleotide in the primer is open for the best choice, total six variations of MAMA primers, three each from forward and backward direction, have potentially suitable. Out of six, the best MAMA primer was chosen empirically based on its sensitivity and selectivity (section 4.1.).

3.4.2. MAMA PCR solute conditions

In addition to the primer modification, PCR conditions, particularly the reagent composition and temperature and length of PCR steps, had to be optimized to achieve the required high sensitivity and specificity of MAMA. The concentrations of magnesium ions (Mg^{2+}), primer and dNTP have strong influences on PCR fidelity and yield (Cha and Thilly, 1993). For each target, a series of different compositions of reaction mixture were tested to find the optimized solute conditions. These experiments showed that the concentration of Mg^{2+} and primers had a dramatic effect on PCR efficiency but relatively less on sensitivity, while the amount of dNTP affected both efficiency and yield. For instance, an excess of dNTP is more prone to generate unspecific amplification, whereas a very low concentration of dNTP compromises PCR yield. Based on these observations, the concentrations of 1.5mM Mg^{2+}, 0.2uM primers and 25uM each dNTP which is 4-fold lower concentration than the standard, were applied for the MAMA protocol.

Addition of glycerol into PCR reaction mixture has been also reported to improve the sensitivity by reducing the amplification of the wild-type sequence by MAMA primers (Cha et al., 1992). Reaction mixture with 5, 10 and 15% (volume/volume) glycerol were tested, resulting in the choice of the 10% solution for the best combination of sensitivity and yield, while the 5 and 15% solution compromised either PCR quality.

The sensitivity limit is determined by whether the mismatch primer is able not to amplify wild-type template in the first cycle or cycles of amplification. Reducing the number of intact wild-type alleles prior to amplification decreases the false positive signal generated from wild-type sequences, thereby increasing the sensitivity of detection. For mutations in some cancer genes which result in the loss of a restriction site, the number of wild-type copies can be significantly reduced by the use of restriction endonuclease prior to PCR (Wilson et al. 1999). Some of the target sequences chosen in this study contain such locus where restriction enzymes can specifically cut the wild-type sequence but not the mutant sequence. HaeIII for bp746 and bp747 in TP53 and DraIII for bp508 in HPRT were added for each sector after the target sequences were enriched and the copy numbers were determined. Both enzymes were purchased from New England Biolabs (Beverly, MA). The enriched target DNA sequences were restricted for 6 hours with the restriction endonuclease (1U/□g DNA) in the manufacturer's

recommended buffer and restriction conditions in a total volume of 100-150ul. Figure 11 shows the TP53 bp746 target sequence prepared with restriction, enrichment and PCR for MAMA.

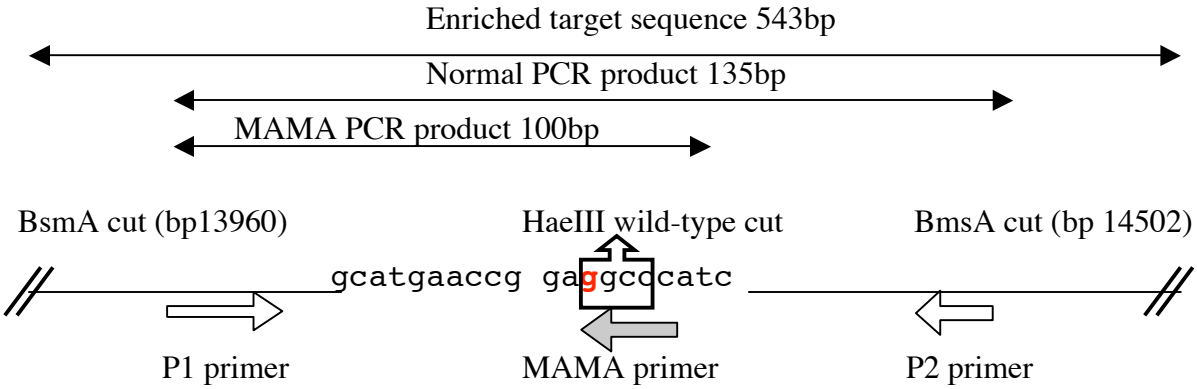


Figure 11. Diagram of the normal and MAMA primer position and the PCR product lengths for the TP53 bp746 target sequence. Genomic DNA was first cut with BmsA to release the target sequence for the enrichment process. Normal PCR using P1 and P2 primers was conducted for measurement of the copy number. After the enrichment, restriction digestion with HaeIII whose recognition site is GGCC destroyed the majority of the wild-type alleles, but saved the mutant allele (**T**GCC). MAMA was performed using P1 and MAMA primers, resulting preferential amplification of the mutants.

3.4.3. Temperature and time for MAMA PCR steps

Instead of a standard three-step PCR cycle consisting of a denaturation, annealing and extension step, two-step PCR cycles consisting of a denaturing step and a primer annealing / extension step were performed for MAMA. For all target sequences, the MAMA-PCR steps first started with DNA polymerase activation at 94°C for 2 minutes or 6 min for the HPRT target, proceeded to 42 cycles of varied condition specific for targets, and 72°C post-cycle extension for 2 minutes, followed by additional incubation at 45°C for 15 minutes to reduce PCR by-products (Table 5). Particularly, the temperature and time for the annealing / extending step was crucial for PCR sensitivity and yield. The cycle conditions were determined by optimization based on extensive reconstruction experiments. Normal *Taq* DNA polymerase (New England Biolabs, Beverly, MA) was used in MAMA for all but the HPRT target which had previously showed a difficulty in improving the sensitivity and yield. Because the background noise generated during MAMA is largely due to a low fidelity of the DNA polymerase, hot-start based Sure Start® *Taq* DNA polymerase (Stratagene, La Jolla, CA) was instead applied for the HPRT sequence, and optimized for the sufficient sensitivity and yield.

By combining the strategies described above, the assays had attained the required level of sensitivity (10^{-5}) to detect mutant alleles among normal lung epithelium (detail in section 4.1.).

Table 5. MAMA PCR cycle conditions for each target sequence.

Target	Polymerase activation	42 Cycles		Post-PCR Extension	Post-PCR Incubation
		Denaturing	Annealing / Extending		
TP53 bp 742	94°C for 2 min	94°C for 23 sec	60°C for 50 sec	72°C for 2 min	45°C for 15 min
TP53 bp 746		94°C for 23 sec	59°C for 50 sec		
TP53 bp 747		94°C for 21 sec	62°C for 45 sec		
K-Ras bp 35		94°C for 21 sec	62°C for 45 sec		
HPRT bp508	94°C for 6 min	94°C for 23 sec	62°C for 50 sec		

4. RESULTS

4.1. MAMA Reconstruction Experiments Using TK6 and Tumor Cell Lines

This thesis required quantitative measurements of five different mutations in normal human lungs epithelium. Because the key to detect accurate mutant signals lay in the sensitivity of the assays, the first goal of this study was to optimize the assay with regard to sensitivity sufficient for human tissue analysis.

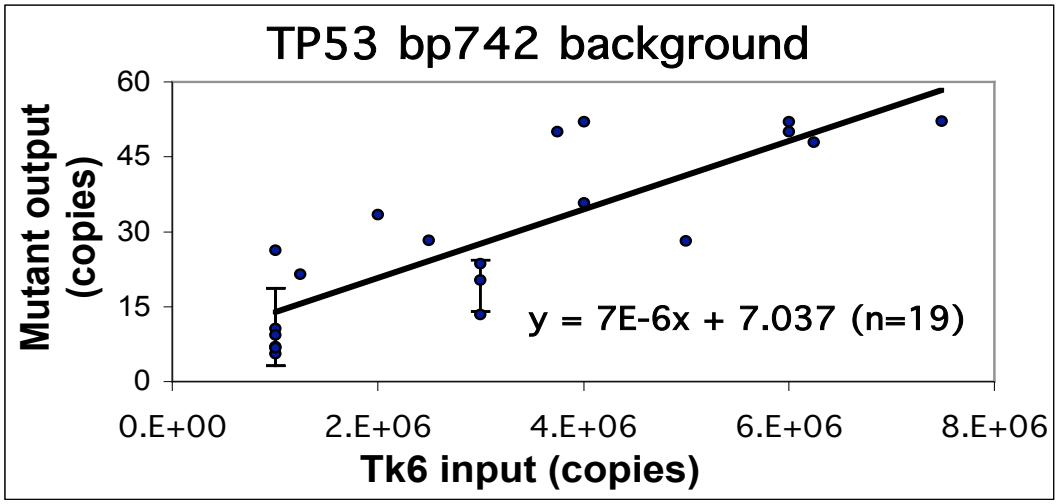
4.1.1. Assay sensitivity

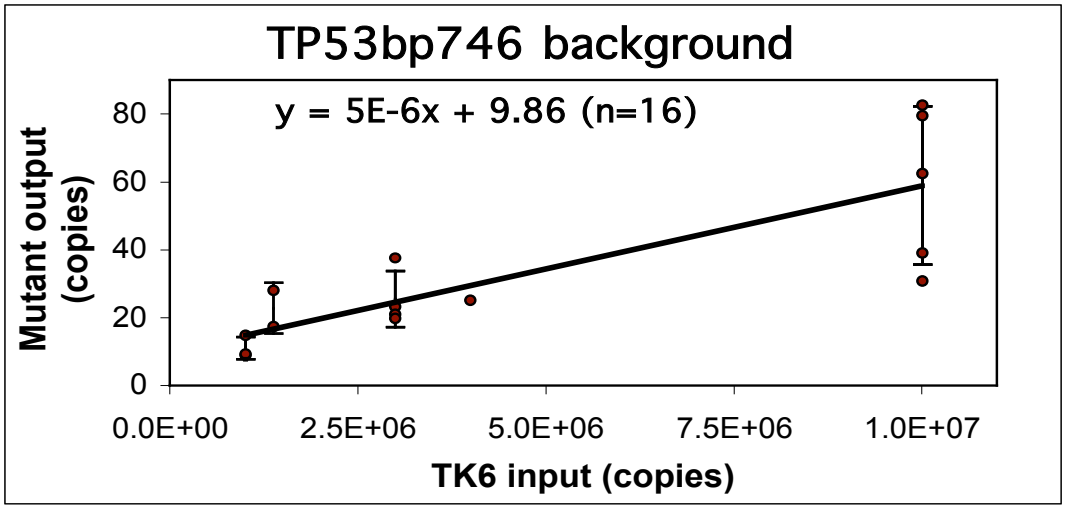
Before positive results from human specimens could be interpreted as detecting mutagenic or genotoxic events, or early steps in the carcinogenesis process, the background frequency of single base substitution mutations in oncogenic loci in human cell lines and human model tissues had to be determined. In the current study, the human lymphoblastoid cell line TK6 that has been reported to carry at least 10-fold lower MF than human tissue (Khrapko et al., 1997b; Tomita-Mitchell et al., 2003) was chosen as a model cell line for the reconstruction experiments. Using TK6 DNA as a negative control and human tumor cell lines with known mutations as positive mutant controls, optimal MAMA conditions were determined on a sequence-by-sequence basis. The reaction components and the thermocycling regimes were empirically optimized by: i) decreased concentration of dNTP (25uM each), ii) addition of 10% glycerol, iii) selection of the most favorable MAMA primers, iv) two-step PCR, v) higher annealing temperature, vi) shorter primer-template extension time and vii) the use of restriction endonucleases which selectively cut wild-type sequences. Particularly in iii), after finding the optimal condition, the most favorable MAMA primers were chosen out of six possible primer designs based on their minimum false positives and the highest sensitivity. In general, highly stringent conditions could generally bring high sensitivity but also sacrifices the yield; hence, the balance between sensitivity and yield was an important aspect of the optimization process.

First, in order to check whether the selected MAMA primers could create any products, the mutant alleles were amplified under MAMA conditions. As it has been reported (Cha et al., 1992), all single mismatch primers were capable of sufficiently replicating mutant templates; in the optimal MAMA conditions, the PCR efficiencies of

single mismatched mutant amplification were in average 68% for TP53, 56.8% for K-Ras and 55.4% for HPRT (per cycle at the exponential phase). These efficiencies for MAMA PCR were somewhat lower than the ones for normal PCR using perfectly matched primers that usually yielded more than 70% of efficiency. The slight decrease of PCR efficiency resulted from a single mismatch between primer and template and the regime of more stringent MAMA conditions. However, the number of product molecules amplified by MAMA reached to the equivalent number amplified by normal PCR after ~35 cycles. Therefore, MAMA single-mismatched primers amplify mutant templates with less efficiency than perfectly matched primers, but sufficiently enough to conduct MAMA.

Next, in order to detect the level of false positive signals, TK6 negative control was mixed with an appropriate amount of internal standards and subjected to MAMA. It should be noted that even though the same amount of input TK6 template copies were added repeatedly, the output (background) signals fluctuated over a non-negligible range. It was thus very important to run multiple controls and determine the mean and SD of the background. In general, the background increased with the input TK6 template copy numbers (Figure 12). The range of the background depended on target mutations. For instance, TP53 bp742 showed average 10 copies of the false positive when $5x10^5$ copies of TK6 templates were introduced to MAMA; the number increased to 52 copies when the number of input TK6 templates increased to $7.5x10^6$ copies. Averaging from 20 data points (n = 20), the mean background level of this target mutation was $1.01x10^{-5}$ with $5.16x10^{-6}$ of one standard deviation (1SD). This number was defined as a detection limit of this target mutation. Similarly, the mean background levels of TP53 bp746, bp747, K-ras bp35 and HPRT bp508 were determined as $9.32x10^{-6}$ (n = 16, 1SD = $4.45x10^{-6}$), $1.81x10^{-5}$ (n = 33, 1SD = $2.04x10^{-5}$), $6.01x10^{-6}$ (n = 21, 1SD = $5.99x10^{-6}$) and $3.01x10^{-5}$ (n = 17, 1SD = $1.80x10^{-5}$), respectively.





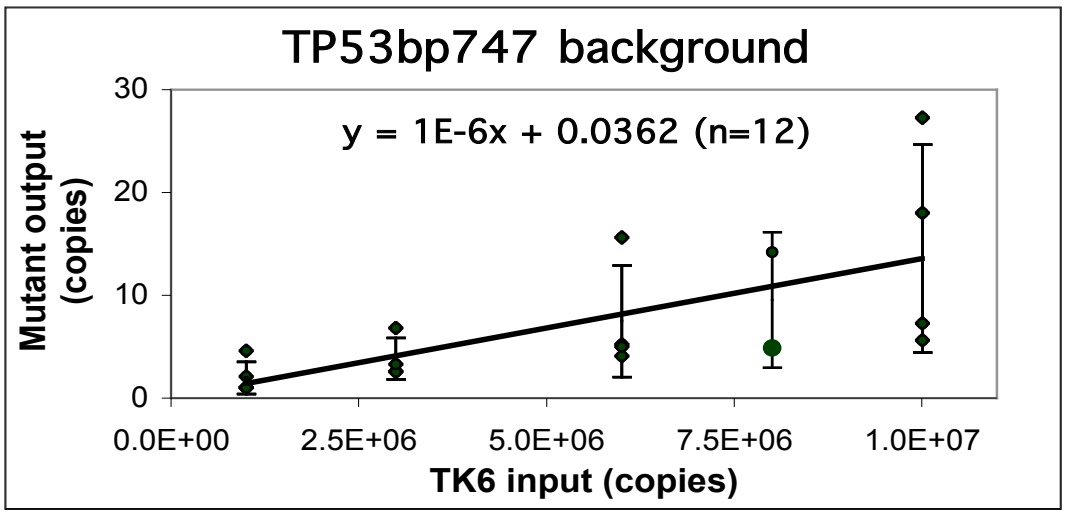


Figure 12. Linear relationship between an amount of input TK6 alleles and output positive signals (background) for TP53 bp742, bp746 and bp747 (from top to bottom). 10^6-10^7 copies of TK6 enriched target alleles were mixed with 50 copies of appropriate internal standards and subjected to MAMA. A linear dose-response was observed.

4.1.2. Assay selectivity

Next, the assay selectivity, which is determined by the ratio of minimum detectable copies of mutant alleles to the maximum copies of wild-type alleles, was examined. $3x10^6$ copies of enriched wild-type TK6 DNA was mixed with 0, 9, 30, 300 and 3,000 copies of mutant genomic DNA to artificially generate a series of MFs: 0, $3x10^{-6}$, 10^{-5}, 10^{-4} and 10^{-3}. Each mixture was further mixed with 200 copies of the internal standard and subjected to MAMA. At the same time, three controls: no template control, internal standard control and internal standard and mutant positive control, were assayed. The mutant copy number of the "0 MF" sample served as an assay background which was subtracted from the mutant copies observed in the remaining positive MF samples, in order to estimate the real mutant copy numbers. This "0 MF" corresponds to the background observed at the section 4.1.1. In the optimal condition, the mutant peaks became larger as the MF became larger compared with the internal standard which contained the constant copies for all reactions (Figure 13).

The selectivity was estimated by comparing the input and the output mutant copy numbers. For the most of the target mutations, the MFs between 10^{-3} and 10^{-4}, and between 10^{-4} and 10^{-5} showed approximately 10-fold difference in their mutant copy numbers measured based on the internal standard; however, at the level of $3x10^{-6}$, the most assays lost their selectivity and showed higher copy numbers than expected. Figure 14 compared the output and the input mutant copy numbers in a logarithmic scale. In the ideal scenario, a linear relationship with a slope of 1 (i.e. the input and the output are identical) and the y-intercept of 0 (i.e. zero background) would be expected. The reconstruction experiments were conducted multiple times (n) in order to verify the SD and its reproducibility. The observed log slopes were: 1.01, 0.96, 0.79, 0.85 and 0.90 for TP53 bp746, bp747 and bp742, K-ras bp35 and HPRT bp508, respectively (Figure 14). Particularly in the MF lower than 10^{-5}, the slope of less than 1 were more likely to be observed, which could lead to underestimation of the mutant copy numbers at low MF. Nevertheless, for cases in which the MFs were more than 10^{-5}, the slopes were close to 1 with relatively small 1SD; thus the assay selectivity of approximately 10^{-5}, which is sufficient for scanning mutational hotspots in DNA in human cells and tissue, was achieved for all target point mutations tested.

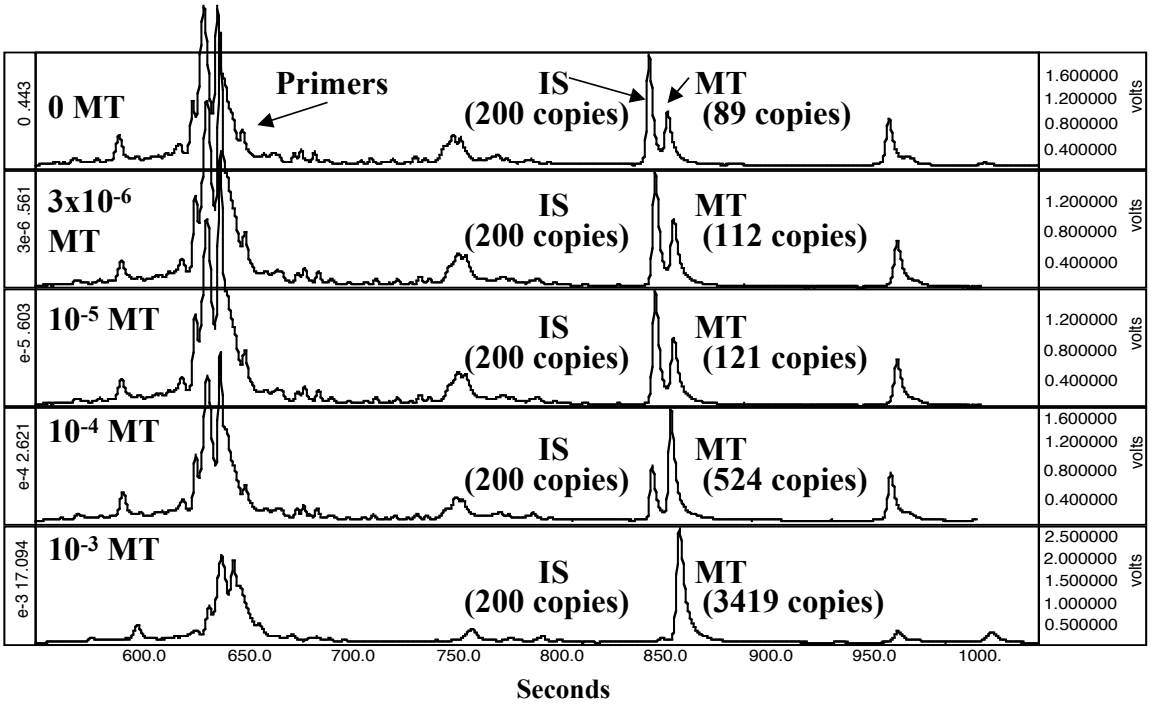


Figure 13. CE separations in the MAMA reconstruction experiments of TP53 bp746 G->T transversion mutation. Human lymphoblastoid cell line TK6 as the wild-type and human tumor Hs700T cell line as the mutant were used to make artificial MFs of 0, $3x10^{-6}$, 10^{-5}, 10^{-4} and 10^{-3} (from the top to the bottom) by mixing $3x10^{6}$ copies of the wild-type each with 0, 9, 30, 300 and 3,000 copies of the mutant, respectively. 200 copies of the internal standard were added to all the mixtures and subjected to the MAMA procedure. The MAMA-PCR products were detected in CE. The mutant (MT) peaks became larger compared to the internal standard (IS) peaks as the artificial MFs became larger (from top to bottom). The mutant copy number at the "0 MF" served as a background and subtracted from the mutants in the rest of all MF positive samples. The real output mutant copy numbers for the MFs of $3x10^{-6}$, 10^{-5}, 10^{-4} and 10^{-3} were thus estimated as 24, 32, 436 and 3330, respectively. These output mutant copy numbers were compared with the input copy numbers (9, 30, 300 and 3000). In this example, the observed (output) mutant copy numbers between the MFs of 10^{-3} and 10^{-4}, and between the MFs of 10^{-4} and 10^{-5} showed almost 10-fold difference, but the MF between 10^{-5} and $3x10^{-6}$ do not hold the expected ratio of 10 to 3, instead, 10 to 7.4. Thus the selectivity of this reconstruction experiment was determined as $7.4x10^{-6}$.

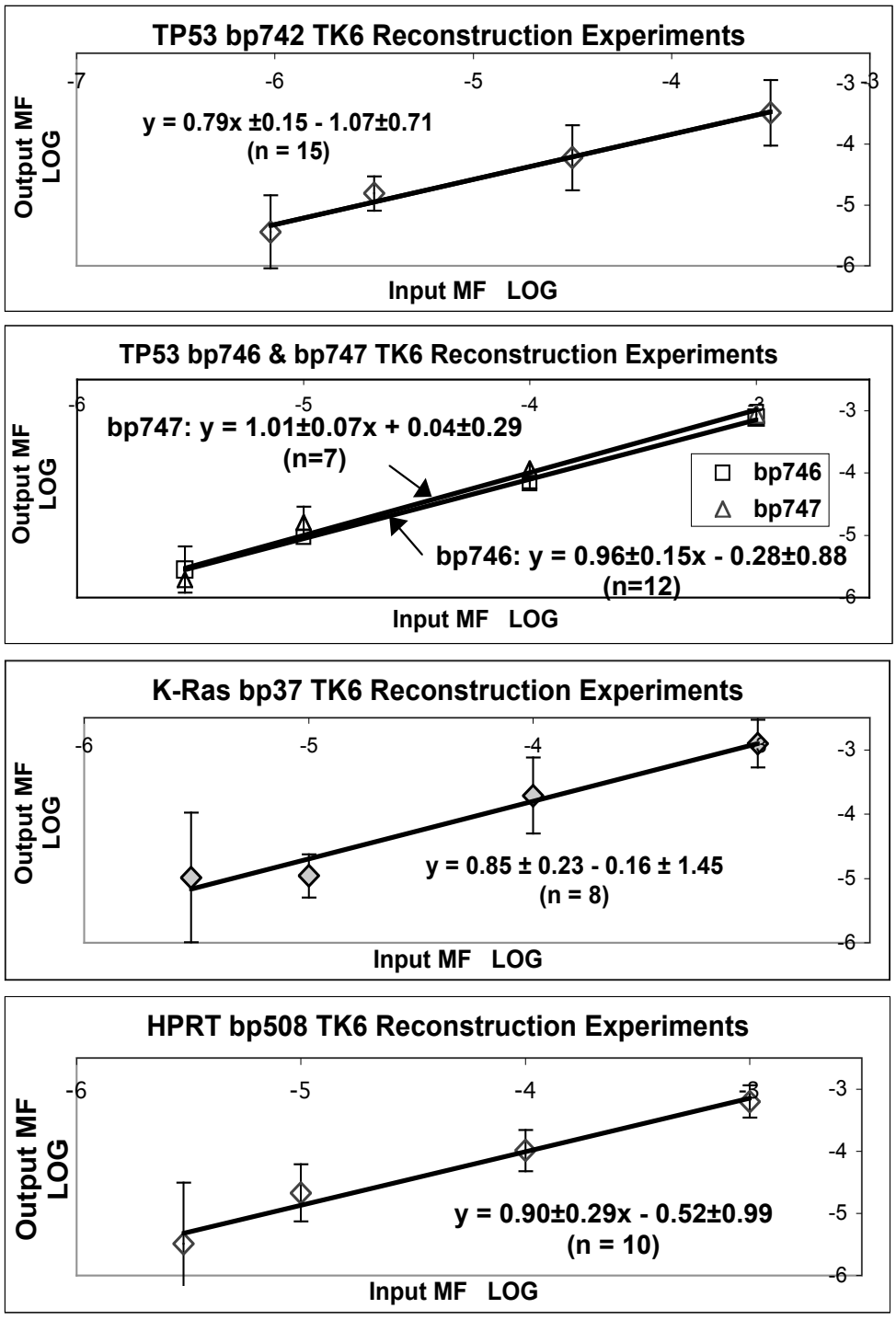


Figure 14. MAMA reconstruction experiments for TP53 bp742, bp746 and bp747, K-ras bp35 and Hprt bp508 using a series of artificial MFs: 10^{-3}, 10^{-4}, 10^{-5}, $3x10^{-6}$ and 0, created by TK6 and specific tumor cell lines. By comparing the input and the output mutant copy numbers, the sensitivity and selectivity of each MAMA were determined. The reconstruction experiments were conducted multiple tines (n) in order to verify the reproducibility and to determine the SD. For all the target mutations, approximately 10^{-5} of selectivity was achieved.

4.2. MAMA on Human Lung Epithelium

A total of 291 sectors from six smokers and nine non-smokers were analyzed for five different target mutations, or 949 individual mutational assays were performed. The target sequences from human lung sectors were prepared similarly to those from TK6 cell line. The amount of genomic DNA isolated from human lung tissue generally ranged from 5 to 10 g DNA or 0.5–5.0 $x10^6$ cells per sector given an average DNA yield of 70%. At every step of the target sequence preparations, the copy numbers of target sequences were measured for each sector by quantitative PCR (Appendix A).

Because every step of target sequence preparations could produce subtle but non-negligible systemic and random errors that consequently cause fluctuation of the background signals, the human lung tissue assays were always accompanied with TK6 negative control which was prepared in the same patch as the human tissue. This background level of TK6 control was re-determined every time with input copy numbers similar to the numbers of total copies in the tissue sector aliquot assayed.

4.2.1. Mutant fractions (MFs) of human lung epithelial cells

The majority of lung epithelial sectors, of which the average size was $2.3x10^6$ cells, provided positive signals larger than TK6 backgrounds and generally contained 0-200 mutant cells (Appendix A). In other words, the majority of lung sectors carried MFs lower than 10^{-4} (Figure 15, 16, 17). In general, the variance of MFs among sectors observed within the same lung was amazingly large, ranging from $5x10^{-6}$ to 10^{-3}. This large intra-individual variance in human tissue is consistent with previous report (Cole and Skopek, 1994).

The results from tissue samples detected in CE were very similar to the ones from cell lines, except that the tissue samples showed more non-specific PCR amplifications. Ideally, PCR products in CE should show only peaks of leftover primers, internal standard and the target mutant; as in figure 17, however, some CE runs from tissue samples showed multiple unidentified peaks along the time of electrophoresis. This is probably because tissue contains more heterogeneous biomolecular complexes, such as fats and proteins, and possibly more pseudogenes than in cell lines. These impurities in tissue other than target DNA sequences might interfere PCR reactions and show up as

non-specific amplifications. Peaks of an internal standard and the target mutants were mostly distinguishable from such non-specific peaks, identified by the relative positions to which the target sequences and positive control migrated through the gel. When the peaks migrated too close each other and hindered accurate measurement of the areas under the peaks, the PCR products were run repetitively until the peaks clearly separated from each other, or MAMA was redone as many times as quantity of the sample pool permitted for desirable results.

Figure 18 shows the distributions of TK6 background and lung sector mutant signals. Even thought the same copy number (10^6) of TK6 DNA was introduced to MAMA repetitively, the background signal fluctuated among multiple trials. More than two thirds of lung sectors assayed provided positive mutant signals clearly above the TK6 background, strongly suggesting the presence of mutations in human lung epithelium.

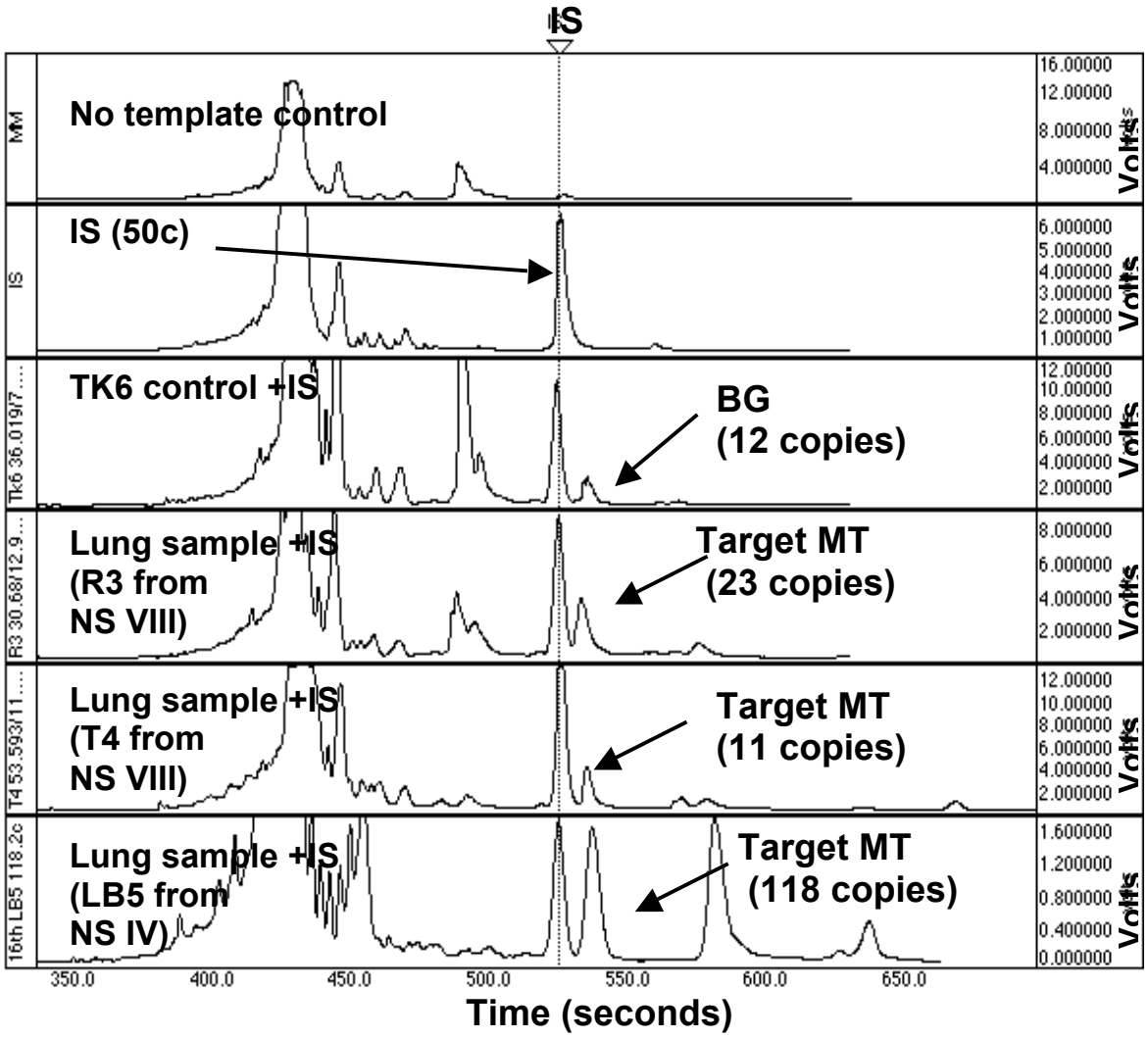


Figure 15. CE runs of MAMA products from lung tissue for K-ras bp35 G:C->T:A. From Top: no template control, 2nd: mutant internal standard (IS), 3rd: IS and TK6 control (background, or BG, is $1.2x10^{-5}$), 4th: IS and sector "R3" from non-smoker (NS) VIII, 5th: IS and sector "T4" from non-smoker VIII and the bottom: IS and sector "LB5" from non-smoker IV. An amount of the internal standard introduced was all 50 copies, and the amount of TK6 or lung samples introduced was all 10^6 copies, which resulted in the MFs of these lung sectors to be $1.1x10^{-5}$, 0 (negative) and $1.06x10^{-4}$, respectively.

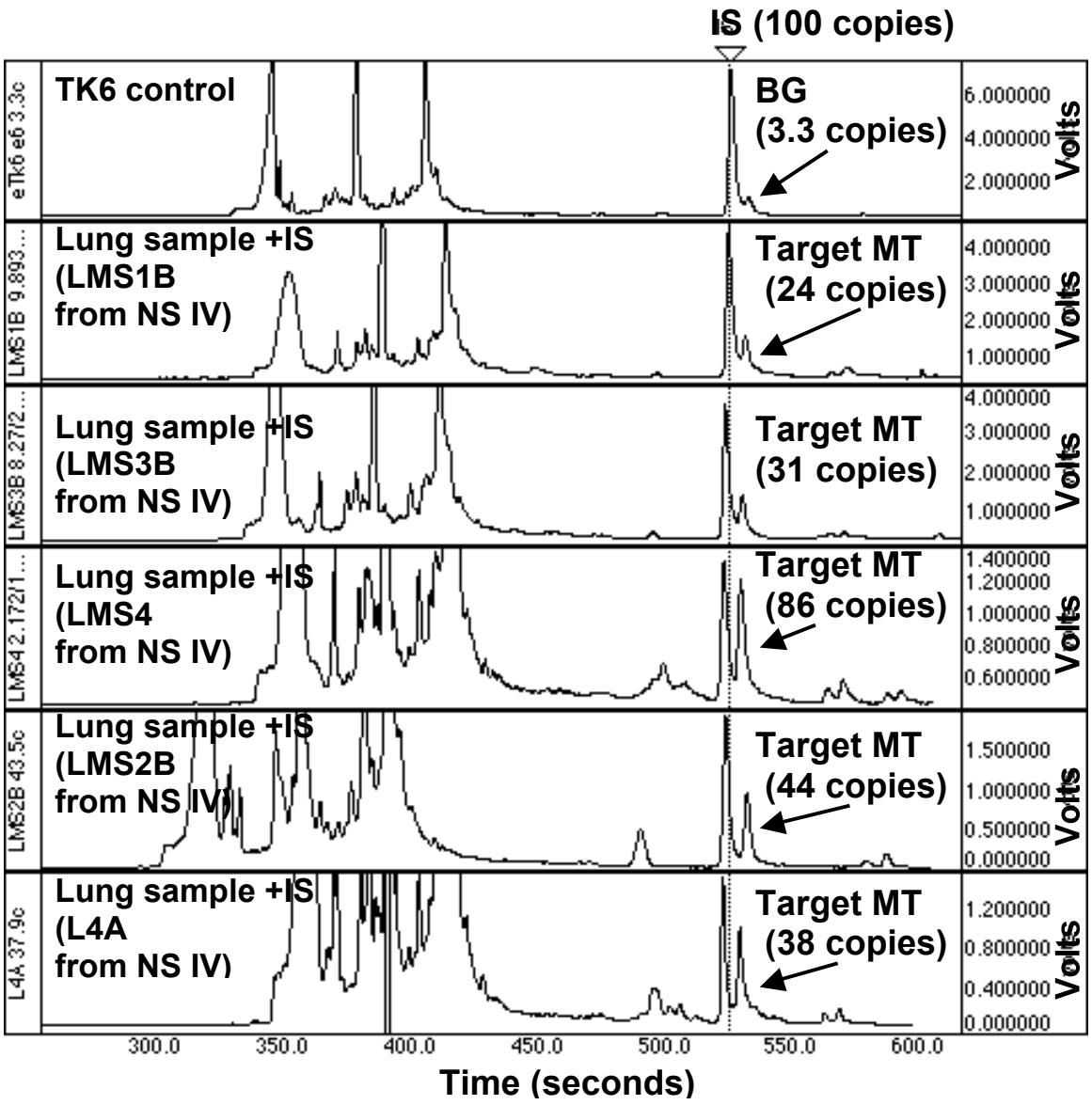


Figure 16. CE runs of MAMA products from lung tissue for Hprt bp508 C:G->T:A. From Top: internal standard (IS) and TK6 control (BG is $3.3x10^{-6}$), 2nd: IS and sector "LMS1B", 3rd: IS and sector "LMS3B", 4th: IS and sector "LMS4", 5th: IS and sector "LMS2B" and the bottom: IS and sector "L4A". All lung sectors were from non-smoker IV. An amount of the internal standard introduced was all 100 copies, and the amount of TK6 or lung samples introduced was all 10^6 copies, which resulted in the MFs of these lung sectors to be $2.1x10^{-5}$, $2.8x10^{-5}$, $8.3x10^{-5}$, $4.1x10^{-5}$ and $3.5x10^{-5}$, respectively.

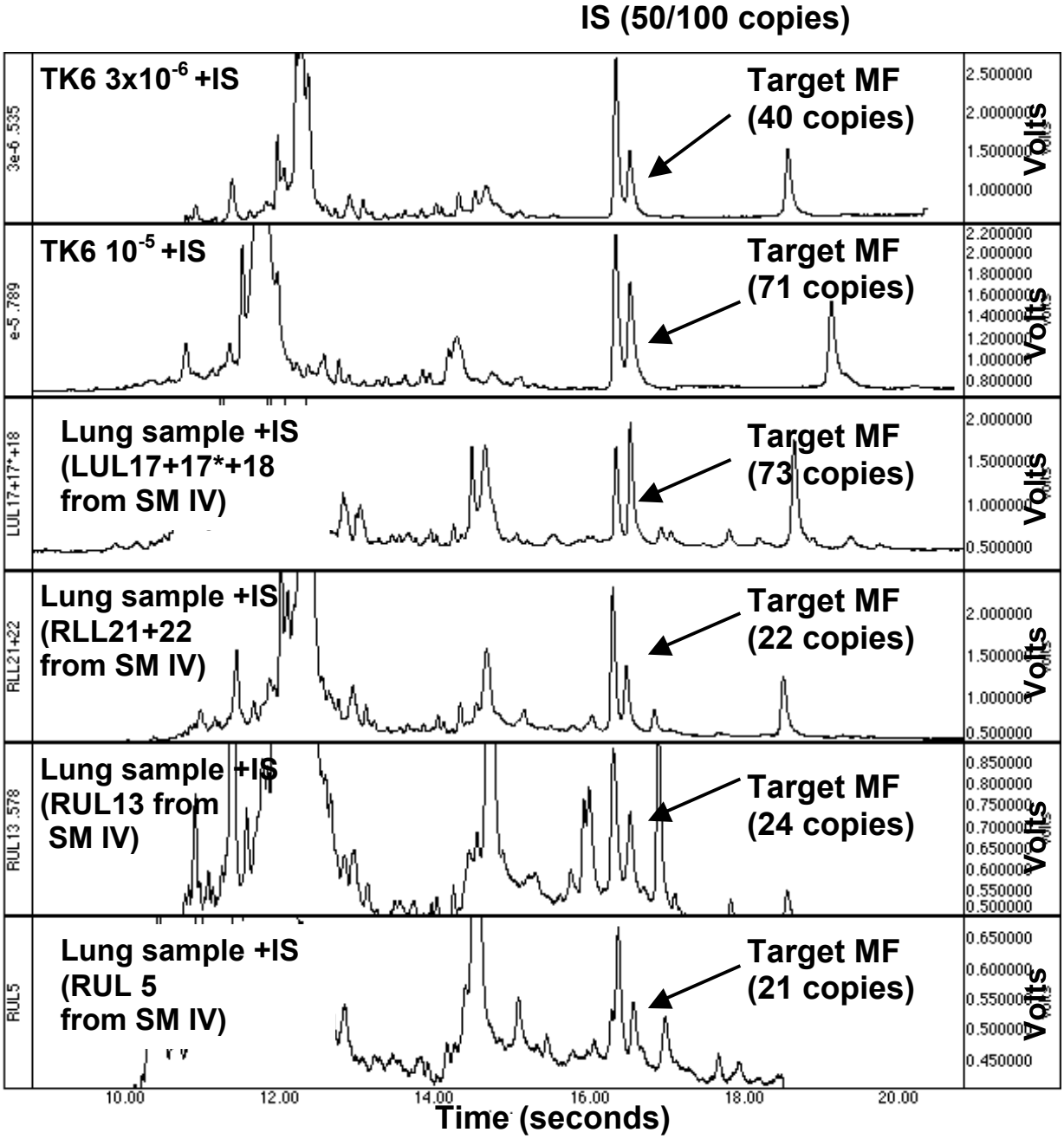


Figure 17. CE runs of MAMA products from lung tissue for TP53 bp746 G:C->T:A. From Top: internal standard (IS) and a mixture of TK6 and Hs700T whose MFs are $3x10^{-6}$ and (2^{nd}) 10^{-5}, 3^{rd}: IS and sector "LUL17+17*+18", 4^{th}: IS and sector "RLL21+22", 5^{th}: IS and sector "RUL13" and the bottom: IS and sector "RUL5". All lung sectors were from smoker (SM) IV. 100 copies of the internal standard and $3x10^{6}$ copies of the TK6 / Hs700T mixtures were introduced for control. 50 copies of the internal standard and 10^{6} copies of lung sectors were introduced into the lung sectors, which resulted in the MFs of these lung sectors to be $4.3x10^{-5}$ and 0 (negative) for the rest, respectively (BG was 25.3 copies for 10^{6} total copy input, not shown).

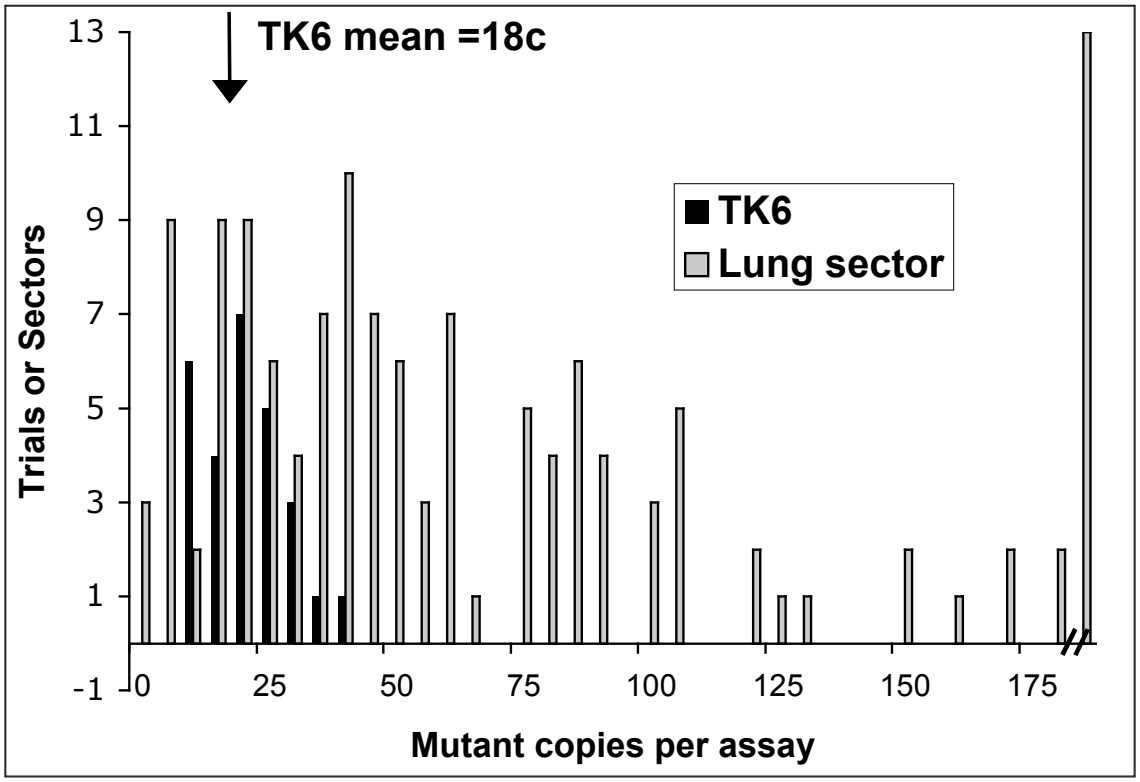
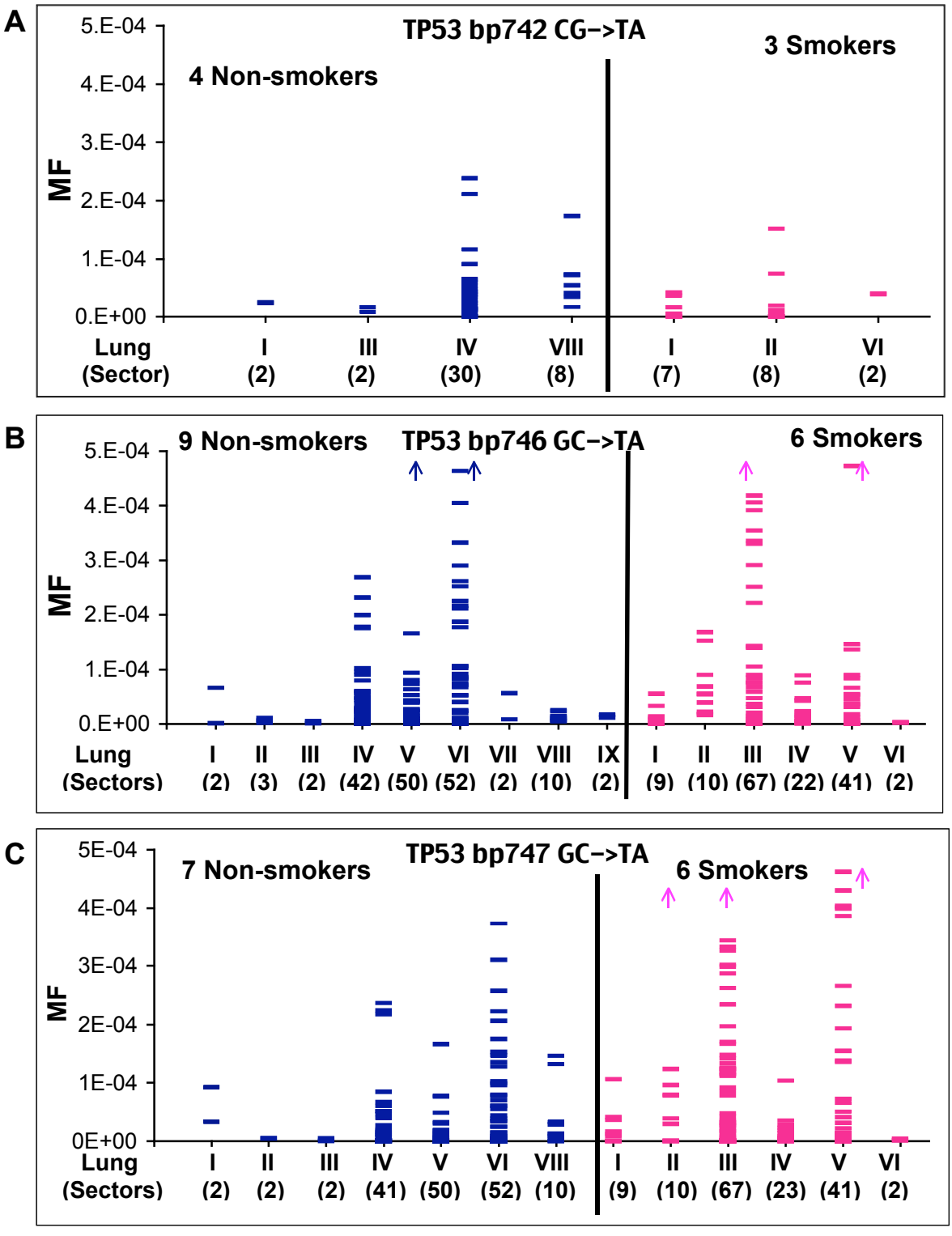


Figure 18. Distributions of sectors or trials as a function of observed mutant copies from TK6 negative controls (solid column) and lung sectors (grey column) for the TP53 bp746 mutation assay. 10^6 copies of templates were added to the all assays. The mean background defined by TK6 trials (n = 27) were 18 copies with 2.6 copies of 1SD. More than two third of the lung sectors provided positive signals clearly above the TK6 background level.

Figure 19 summarizes distributions of the MFs of all sectors from six smokers' and nine non-smokers' lungs for the five target mutations. Again, the variations of MFs within individuals were quite large, almost reaching to 100-fold (10^{-5} to 10^{-3}) regardless of the donor's smoking status. The variances of MFs were similar between the smoker and non-smoker group.

Notably, the MFs for the TP53 and K-Ras were found similar to the ones for a non-oncogenic control gene HPRT, implying that these mutations in TP53 and K-Ras genes were not induced by cigarette smoking. The observations also suggest that the TP53 and K-Ras genes, like HPRT, do not involve in the initial step of lung carcinogenesis, thus not gatekeeper genes for lung cancer.

With very rare occasions, sectors with MF larger than $4x10^{-4}$, or containing more than 1,000 mutant cells, were observed for some lungs regardless of the donor's smoking status. Such sectors comprised of 4.6% of total sectors (44 out of 949) and treated as outliers in the further statistical analyses to avoid skewness in the data (details in section 4.2.5.).


A
TP53 bp742 CG−>TA
4 Non-smokers
3 Smokers
MF
5.E-04
4.E-04
3.E-04
2.E-04
1.E-04
0.E+00
Lung (Sector)
I (2)
III (2)
IV (30)
VIII (8)
I (7)
II (8)
VI (2)
B
9 Non-smokers
TP53 bp746 GC−>TA
6 Smokers
MF
5.E-04
4.E-04
3.E-04
2.E-04
1.E-04
0.E+00
Lung (Sectors)
I (2)
II (3)
III (2)
IV (42)
V (50)
VI (52)
VII (2)
VIII (10)
IX (2)
I (9)
II (10)
III (67)
IV (22)
V (41)
VI (2)
C
7 Non-smokers
TP53 bp747 GC−>TA
6 Smokers
MF
5E-04
4E-04
3E-04
2E-04
1E-04
0E+00
Lung (Sectors)
I (2)
II (2)
III (2)
IV (41)
V (50)
VI (52)
VIII (10)
I (9)
II (10)
III (67)
IV (23)
V (41)
VI (2)

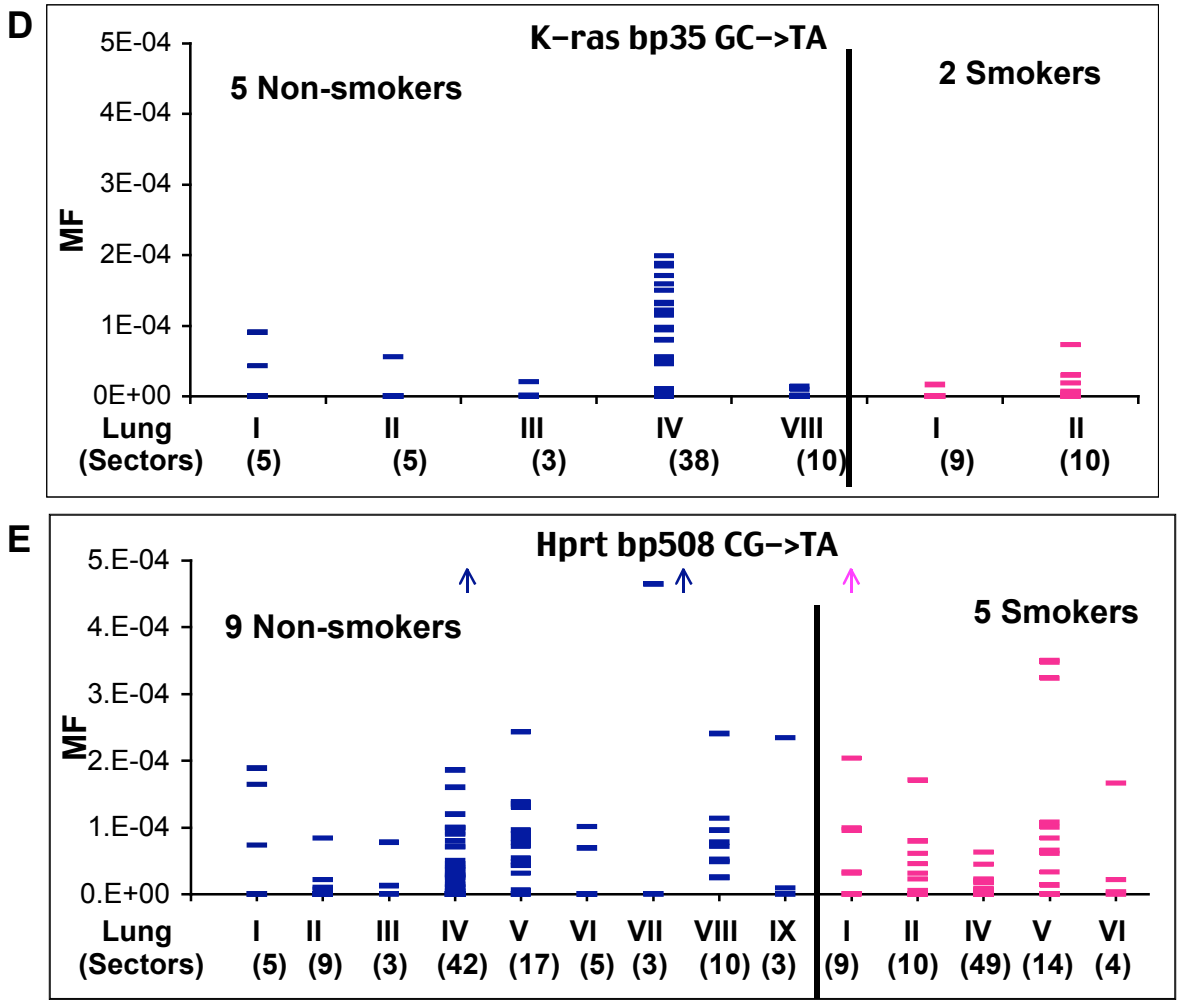


Figure 19. MFs of all lungs analyzed for TP53 bp742 (A), bp746 (B) and bp747 (C), K-ras bp35 (D) and Hprt bp508 (E) for non-smokers (black on left) and smokers (grey on right). Greek numbers correspond to the lungs labeled in Table 3, and each lung provided multiple sectors (parenthesized numbers). Each mark represents a MF of an individual sector. The range of variations of MF was similar in the smoker and non-smoker group. The arrows indicate lungs that contain outlying sectors with MF larger than $4x10^{-4}$ (4.6%).

In order to examine whether the smokers and non-smokers analyzed herein were significantly different in the MFs, two approaches were taken; 1) compare the mean MFs among sectors from an individual lung by two-sample t-test, assuming that every subject had a similar variance, 2) compare the MF distributions of all sectors combined for the smoker or the non-smoker group by the Kolmogorov-Smirnov test, assuming that the MF distributions of sectors were similar within the group.

4.2.2. Comparing the means of MFs for each lung from smokers and non-smokers

Table 6 summarizes the assay sensitivity measured by TK6, the mean MFs, their two standard deviations (2SD) and the number of sectors assayed for each lung from smokers and non-smokers. The mean MFs per lung ranged from 8.9×10^{-7} (K-ras for SM I) to 1.4×10^{-4} (HPRT for SM V). The grand mean MFs among lungs for TP53 bp742, bp746 and bp747, K-ras bp35 and HPRT bp508 were 2.66×10^{-5}, 2.16×10^{-5}, 3.60×10^{-5}, 7.07×10^{-6} and 5.79×10^{-5} for smokers and 3.28×10^{-5}, 2.12×10^{-5}, 2.48×10^{-5}, 2.39×10^{-5} and 4.78×10^{-5} for non-smokers, respectively (Table 6). These grand mean MFs among lungs were not statistically different between smokers and non-smokers ($p = 0.62$ by two-sample t-test).

Table 6. Summary of the assay sensitivity defined by TK6 reconstruction experiments (above top), mean MFs, two standard deviations (2SDs), number of sectors obtained from all lungs from 9 non-smokers (top) and 6 smokers (bottom) for TP53 bp742, bp746 and bp747, K-ras bp35 and Hprt bp508 (from the left to right).

		P53 bp742			P53 bp746			P53 bp747			K-ras bp37			HPRT bp508		
Sensitivity		1.02E-05	19	1.55E-05	1.42E-05	27	1.12E-05	1.76E-05	30	2.26E-05	7.40E-06	17	6.00E-06	2.44E-05	81	2.14E-05
	TK6	mean MF	sector	2SD	mean MF	sector	2SD	mean MF	sector	2SD	mean MF	sector	2SD	mean MF	sector	2SD
Non-smoker	I 38yo F	2.44E-05	2	2.35E-06	2.83E-05	2	4.62E-05	5.76E-05	2	4.27E-05	3.86E-05	5	4.03E-05	1.08E-04	5	8.88E-05
	II 40yo F				6.35E-06	3	4.38E-06	8.05E-06	2	1.39E-07	2.24E-05	5	2.50E-05	1.35E-05	9	2.73E-05
	III 41yo M	1.05E-05	2	3.02E-06	3.74E-06	2	3.02E-06	2.88E-06	2	3.19E-06	6.44E-06	3	1.14E-05	3.37E-05	3	4.16E-05
	IV 45yo F	3.49E-05	30	5.66E-05	4.18E-05	42	6.78E-05	2.63E-05	41	7.02E-05	4.90E-05	38	6.83E-05	3.53E-05	42	4.74E-05
	V 50yo M				1.95E-05	50	3.18E-05	8.36E-06	50	3.59E-05				7.45E-05	17	6.15E-05
	VI 59yo F				4.59E-05	52	9.24E-05	3.38E-05	52	8.70E-05				2.52E-05	5	4.79E-05
	VII 67yo M				1.89E-05	2	2.80E-05							3.93E-06	3	
	VIII 75yo F	6.12E-05	8	4.75E-05	1.26E-05	10	7.13E-06	3.68E-05	10	5.45E-05	2.84E-06	10	4.72E-06	6.64E-05	10	6.32E-05
	IX 76yo F				1.41E-05	2	8.83E-06							7.06E-05	3	1.33E-04
	Mean / Sum	3.28E-05	42	2.74E-05	2.12E-05	165	3.22E-05	2.48E-05	159	4.19E-05	2.39E-05	61	2.99E-05	4.78E-05	97	6.38E-05
Smoker	I 41yo F	1.24E-05	7	1.80E-05	1.11E-05	9	2.52E-05	2.22E-05	9	3.41E-05	8.49E-07	9	2.66E-06	8.59E-05	9	7.30E-05
	II 41yo M	2.79E-05	8	5.41E-05	6.23E-05	10	5.37E-05	4.48E-05	10	4.63E-05	1.33E-05	10	2.29E-05	2.65E-05	10	5.31E-05
	III 47yo F				3.34E-05	67	1.04E-04	8.86E-05	67	1.03E-04						
	IV 55yo M				1.01E-05	22	2.50E-05	3.23E-06	23	2.20E-05				7.17E-06	49	1.18E-05
	V 58yo F				1.00E-05	41	4.02E-05	5.53E-05	41	4.58E-05				1.39E-04	14	1.32E-04
	VI 59yo M	3.95E-05	2	4.82E-07	2.88E-06	2	1.03E-06	2.14E-06	2	1.29E-06				3.11E-05	4	7.94E-05
	Mean / Sum	2.66E-05	17	2.42E-05	2.16E-05	151	4.16E-05	3.60E-05	152	4.21E-05	7.07E-06	19	1.28E-05	5.79E-05	86	6.99E-05
Grand mean / Sum		3.13E-05	59	2.96E-05	2.20E-05	316	3.72E-05	3.15E-05	311	4.54E-05	1.60E-05	80	2.33E-05	5.11E-05	183	6.73E-05

* Greek numbers of lung labels correspond to Table 3; M and F stands for male and female, respectively, and yo stands for year-old.

** Outliers (sectors with MF larger than 4×10^{-4}, 4.6% of a total sectors) were excluded in calculating the mean MFs and 2SDs.

Figure 20 is a graphic representation of Table 6, depicting distributions of the mean MFs of the five target mutations assayed for all lungs analyzed. Again, the variances within the group were as large as the variances between the groups. The grand mean MFs were similar between smokers and non-smokers for all targets assayed. It should be also noted that the mean MFs were very similar across different target point mutations, including HPRT that is a non-oncogenic control gene. The analysis of variance (ANOVA) test in fact suggested that the mean MFs were similar across the five different target mutations for both smokers ($n = 22$, $p = 0.65$) and non-smokers ($n = 34$, $p = 0.43$).

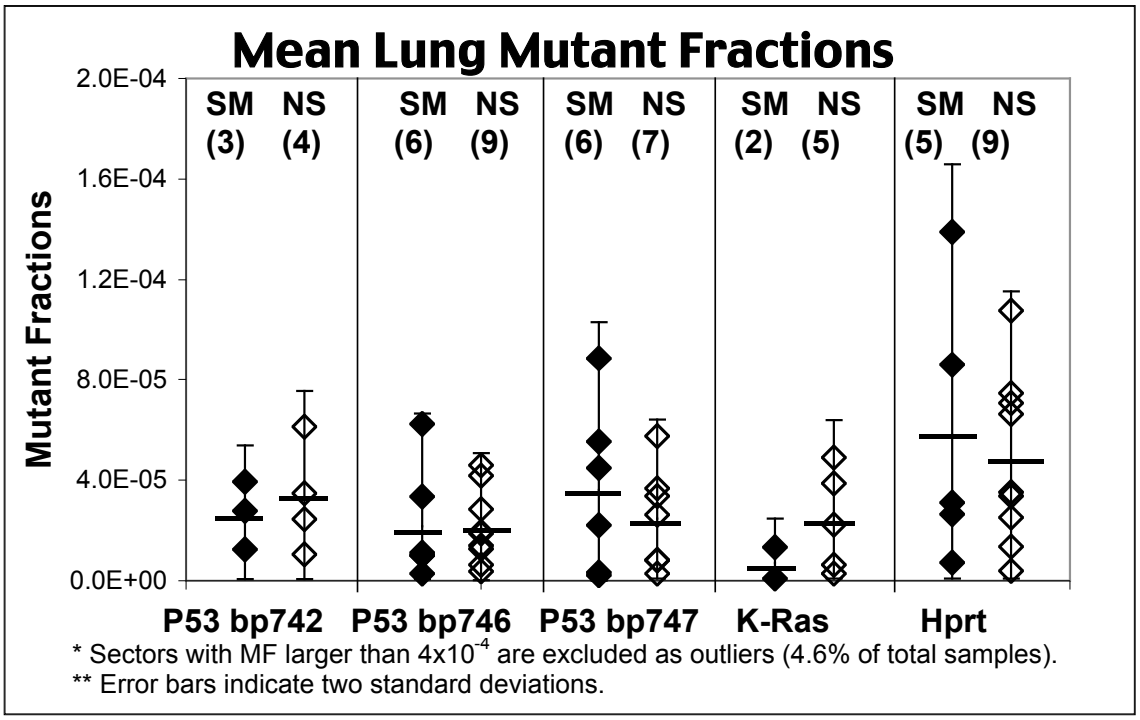


Figure 20. Distributions of the mean MFs among lungs of smokers (SM, solid diamonds on left) and non-smoker (NS, open diamonds on right) for TP53 bp742, bp746 and bp747, K-ras bp35 and Hprt bp508 (from left to right). Each diamond represents a mean MF of an individual lung, and the parenthesized numbers under SM and NS indicate the number of lungs analyzed for the specific target mutation. The short black bars indicate the grand mean in each target.

After proving that the distributions of mean MFs across five target mutations were statistically similar, the mean MFs were normalized for all target mutations assayed by taking the ratios of the mean MFs for each mutation, and recategorized into the smoker and non-smoker groups (Figure 21). The means of the relative MFs were strikingly similar in smokers (1.01, n = 22) and in non-smokers (0.99, n = 34). It should be noted herein that the dose of cigarettes smoked by these smoker donors would be expected with 4.5 to 9.7 higher risk of getting lung cancer (Agudo et al., 2000) and increase the lung cancer mortality rate almost 25-fold. The distributions of these two relative MFs were not significantly different by two-sample t-test ($p = 0.92$), assuming their normal distributions. The relative MF points mostly fell within 2SDs; out of 56 relative MF points, however, three points were out of the upper boundary of 2SD: (from lower to the highest) K-ras from non-smoker IV, TP53 bp746 from smoker II and TP53 bp747 from smoker III. It is important to note that lungs with the high relative MFs were observed both in the smoker and non-smoker group.

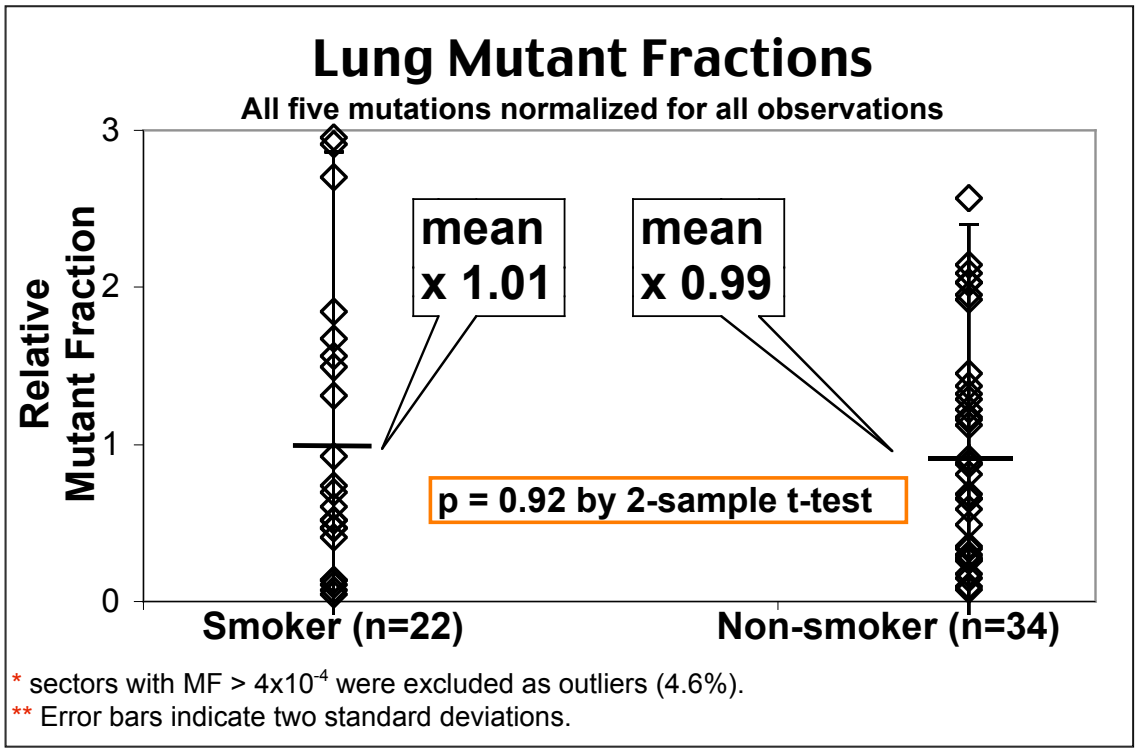


Figure 21. Distributions of the relative mean MFs among lungs from smokers (left) and non-smoker (right) for all five mutations normalized. Each diamond represents a relative mean MF of an individual lung for the specific target mutation. The short black bars indicate means of the relative mean MFs for smokers (1.01) and for non-smokers (0.99). Distributions of relative mean MFs did not differ significantly between smokers and non-smokers (p = 0.92, two-sample t-test).

4.2.2.1. Mean MFs and gender

It has been reported that women are more susceptible to tobacco-induced lung cancer. Both a higher average hydrophobic DNA adduct level and a higher frequency of G:C->T:A mutations have been reported in lung tumors of females than that of males despite of the fact that the level of exposure to carcinogens from cigarette smoke is lower among females (Ryberg et al., 1994a; Kure et al, 1996; Hernandez-Boussard and Hainaut, 1998). In this study, the lung donors comprised of three males and three females in the smoking group and three males and six females in the non-smoking group. When the mean MFs with all target mutations combined were compared by gender, the females carried significantly higher mean MFs (3.89×10^{-5}) than the males (2.06×10^{-5}) (Figure 22. $p = 1.5\%$, two-sample t-test). This observation is consistent with the previous understanding of higher genetic susceptibility among females.

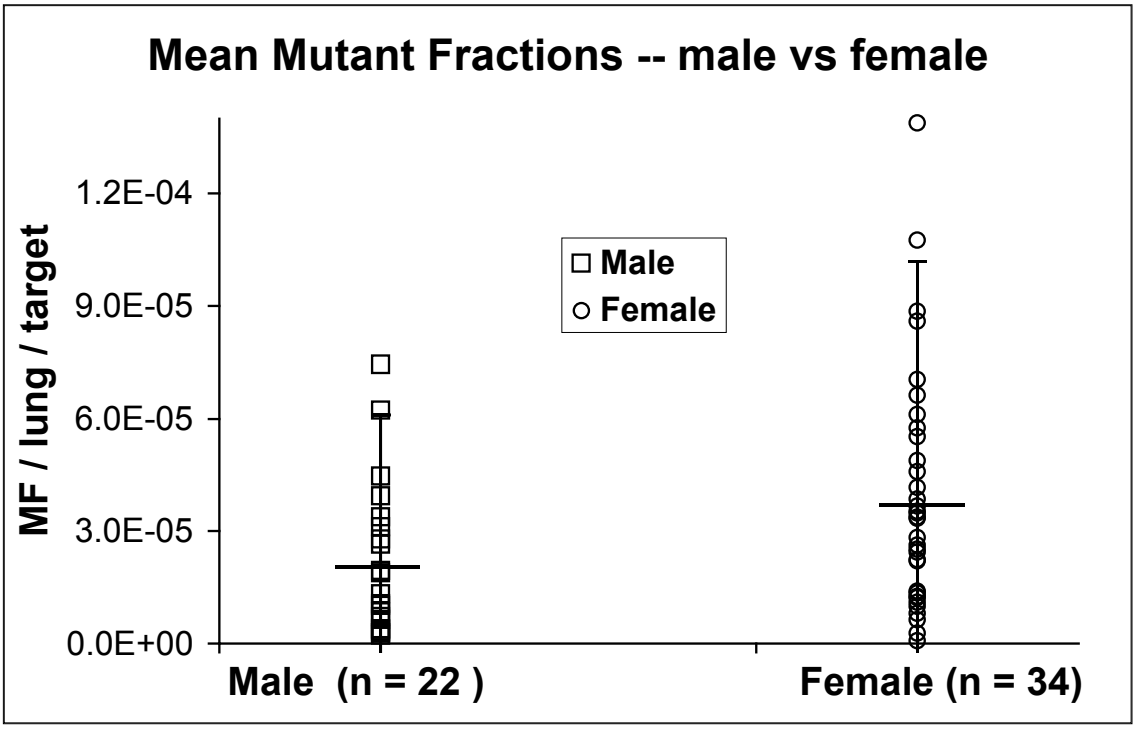


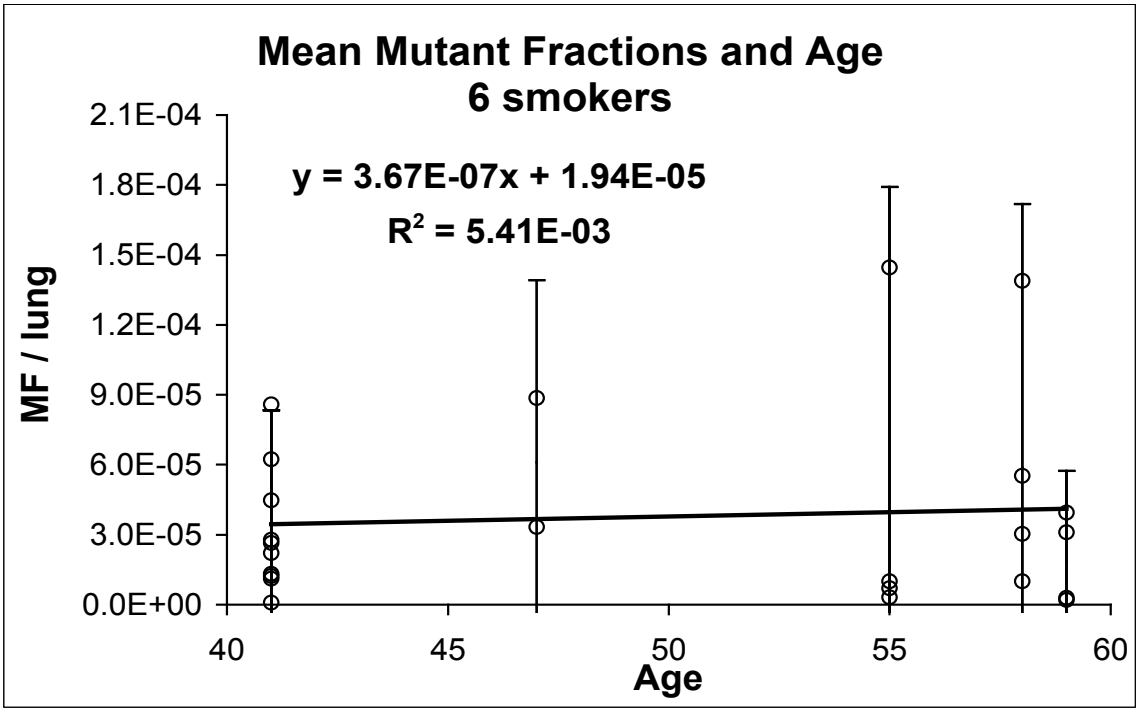
Figure 22. The male and female mean MFs among lungs for all target mutations combined. Twenty-two male mean MFs (squares on left) and thirty-four female mean MFs (circles on right) were compared. The short bars indicate the means (2.06×10^{-5} and 3.89×10^{-5}, respectively), and the error bars indicate 2SDs. The mean MFs were significantly higher in females than males (p = 1.5% by two-sample t-test). Sectors with MF larger than 4×10^{-4} were excluded as outliers (4.6% of the total sectors assayed).

4.2.2.2. Mean MFs and age

Aging is a complex mechanism that results from accumulation of mutations in somatic cells that results in failure of the cells either to survive, to proliferate or to function at complete efficiency (Morley, 1998). It has been reported that HPRT mutant frequency in peripheral T-lymphocytes increased significantly over age of the subject (Finette et al., 1994). The mutant frequency of the HLA-A locus in human T-lymphocytes was observed to increase at a rate of $7x10^{-7}$ / year as T-cells divide approximately once a year (Morley and Turner, 1999).

The present study enrolled fifteen subjects of age from 38 to 76. When the mean MFs were plotted over age of the subjects, the regression line resulted in positive incline for both smokers ($3.67x10^{-7}$ MF / year) and non-smokers ($2.69x10^{-7}$ MF / year), although the coefficient of determination (R^2) were very small (0.0133 and 0.0054, respectively) and the results did not reach to a statistical significance due to the large variances within the same age group (Figure 23).

It should be noted again that the mean MFs of the smoker and non-smoker groups were not affected by age since the ages of donors were similarly distributed in these two groups (the mean age per sector was 51 for the smoker group and 53 for the non-smoker group).



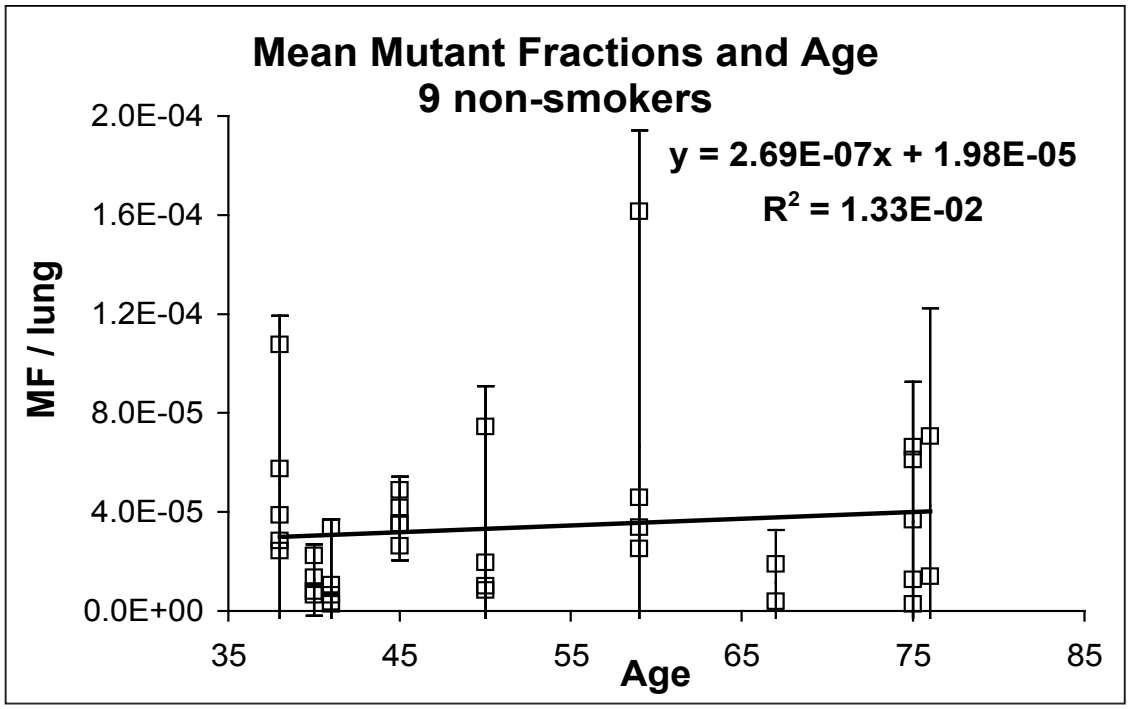


Figure 23. The mean MFs and the age of smokers (top) and non-smokers (bottom). The mean MF per lung were combined for all five target mutations. The mean MFs increased as the age of donor increased for both smokers and non-smokers (3.67×10^{-7} / year and 2.69×10^{-7}/ year, respectively), but the increases did not reach statistical significance (R^2 = 0.0541, 0.0133, respectively) due to the large variance within the same age group.

4.2.2.3. Mean MFs and cigarette smoke dose-response

The dose-response relationship of cigarette smoke with an incidence risk of lung cancer has been suggested by epidemiological and clinical studies (Doll and Peto, 1978; Zang and Wynder, 1996). The risk of getting lung cancer for smokers with less than 10 pack-year is 1.38, but increases to 3.70 for smokers with 10-19 pack-year and to 7.14 for smokers with more than 30 pack-year (Agudo et al., 2000). The smoker donors obtained in this study were mostly heavy smokers, the pack-year ranging from 10.5 to 60. No association was found between the mean MFs and the dose of cigarette smoke among smokers as well as non-smokers included as zero-dose smokers (Figure 24). The result is consistent with the previous observation in this study that the mean MFs were not affected by cigarette smoke.

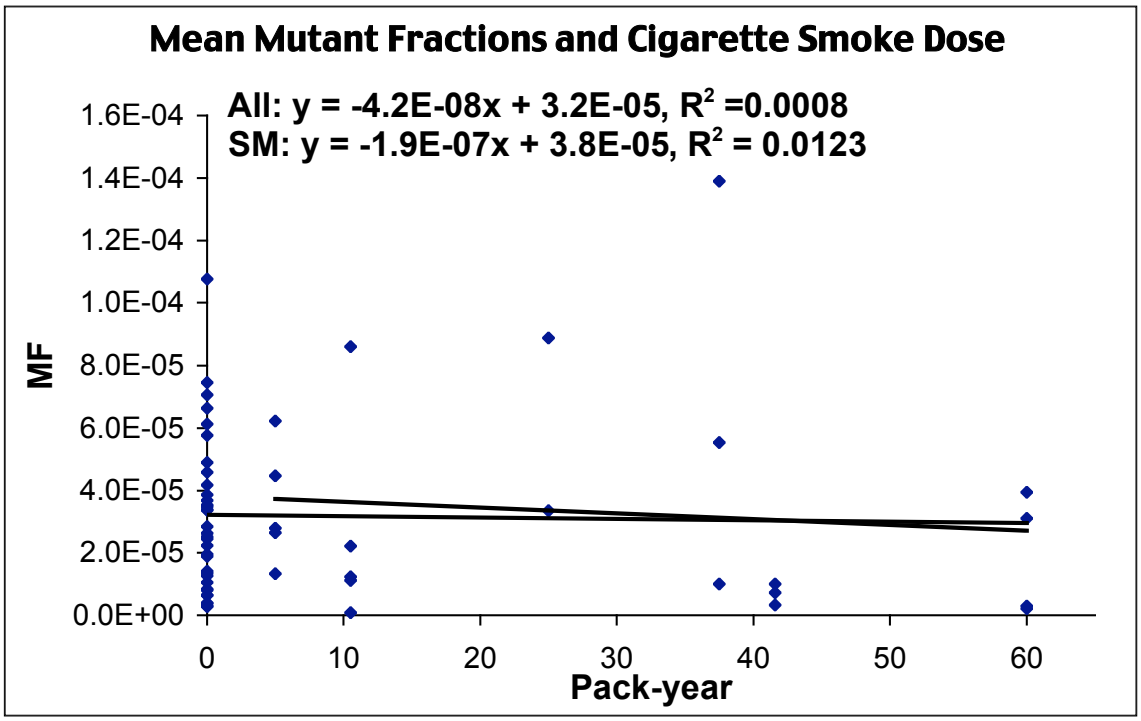
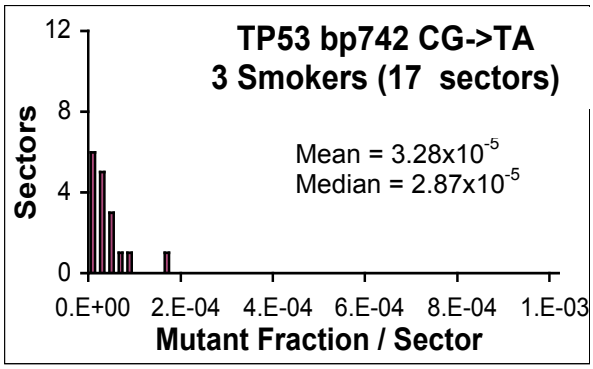


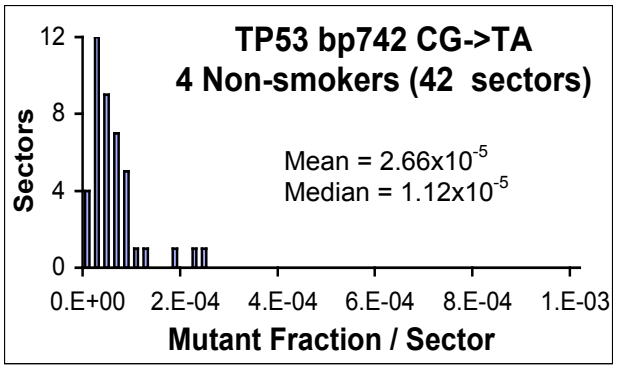
Figure 24. The mean MFs and cigarette smoke dose. The mean MFs from all target mutations were combined. The steeper regression line represents the mean MFs from all donors (**ALL**), whereas the less steeper regression line represents the mean MFs from smokers (**SM**). In both cases, a dose-response relationship was not observed.

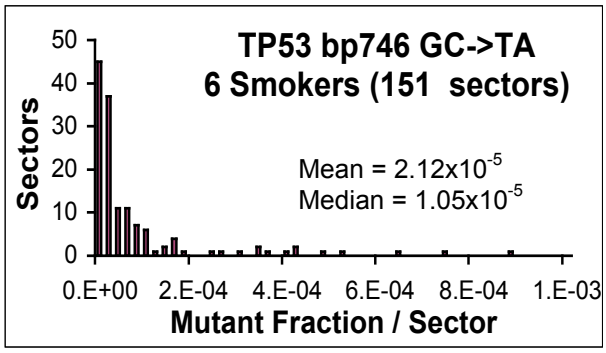
4.2.3. MF distributions among sectors

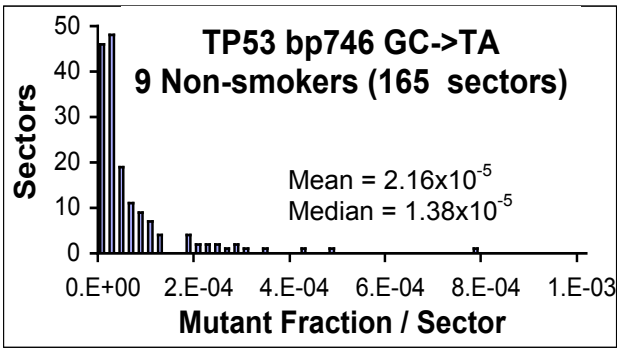
In the second approach, the MFs obtained from each sector assayed were combined for all smokers and for all non-smokers analyzed for each target mutations, under an assumption that the distributions of the MFs were similar within the same group. The distributions of sectors over MFs were similar in smokers and non-smokers in all target mutations (Figure 25). Because the mean does not represent itself correctly if the distribution is not normal, the median was also taken into accounts. After excluding outliers, the medians for TP53 bp742, bp746 and bp747, K-ras bp35 and HPRT bp508 were 1.12×10^{-5}, 1.38×10^{-5}, 1.80×10^{-5}, 0 and 2.34×10^{-5} for smokers and 2.87×10^{-5}, 1.05×10^{-5}, 4.51×10^{-5}, 9.81×10^{-7} and 3.17×10^{-5} for non-smokers, respectively. The means were appeared to be slightly higher than the medians due to a small number of outlying sectors. The medians of MF per sector were not statistically difference between smokers and non-smokers for all target mutations assayed ($p = 0.75$ by two-sample t-test). Likewise, the means of MF per sectors were not significantly different between these tow groups ($p=0.62$ by two-sample t-test).

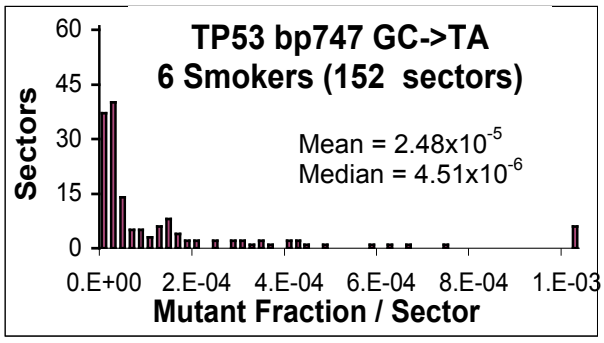
Figure25. Distributions of tracheal-bronchial sectors all combined for smokers (left) and non-smokers (right) as a function of MF at TP53 bp742, bp746 and bp747, K-ras bp35 and Hprt bp508 (from top to bottom). The means and medians were not significantly different between smokers and non-smokers for all five target mutations assayed ($p = 0.62$ and $p = 0.75$, respectively, by two-sample t-test). Most of sectors carried MF of less than $4x10^{-4}$. Sectors with MF larger than $4x10^{-4}$ were excluded as outliers (4.6% of total sectors) in calculating the means and medians.

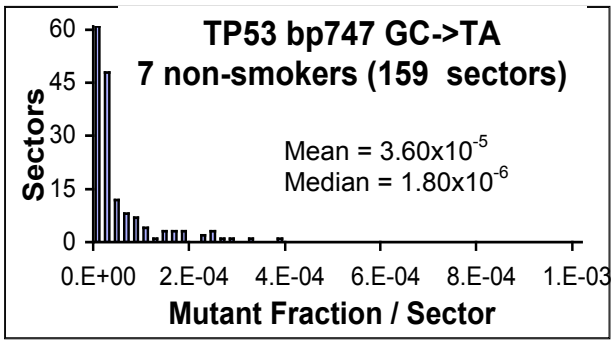

TP53 bp742 CG->TA
3 Smokers (17 sectors)
Mean = 3.28x10-5
Median = 2.87x10-5
Sectors
Mutant Fraction / Sector

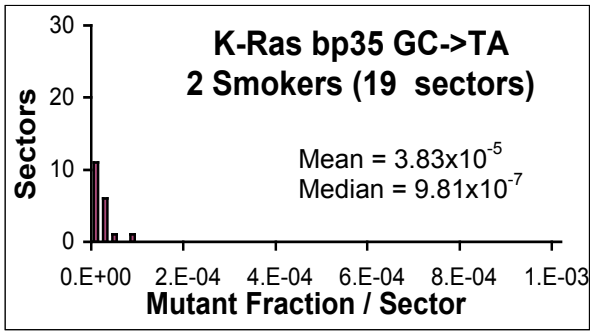

TP53 bp742 CG->TA
4 Non-smokers (42 sectors)
Mean = 2.66x10-5
Median = 1.12x10-5
Sectors
Mutant Fraction / Sector

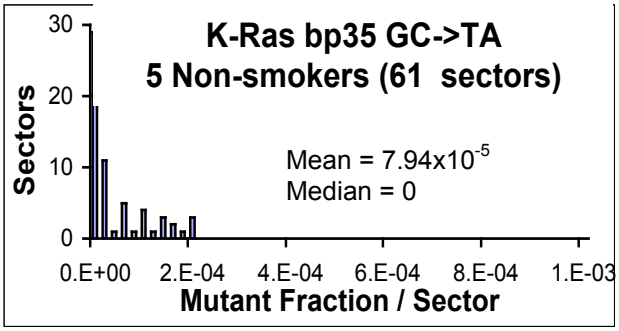

TP53 bp746 GC->TA
6 Smokers (151 sectors)
Mean = 2.12x10-5
Median = 1.05x10-5
Sectors
Mutant Fraction / Sector

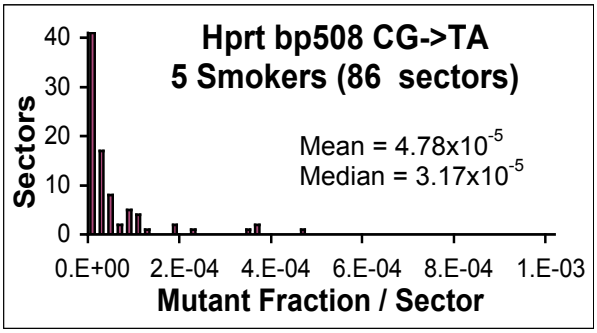

TP53 bp746 GC->TA
9 Non-smokers (165 sectors)
Mean = 2.16x10-5
Median = 1.38x10-5
Sectors
Mutant Fraction / Sector

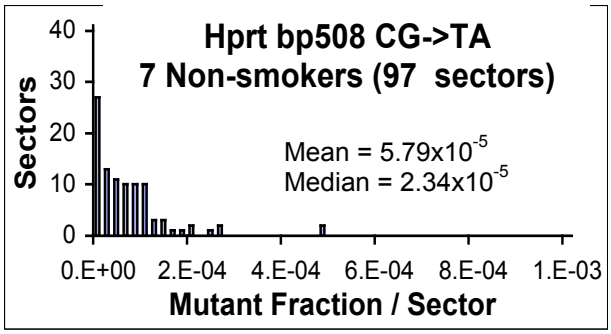

TP53 bp747 GC->TA
6 Smokers (152 sectors)
Mean = 2.48x10-5
Median = 4.51x10-6
Sectors
Mutant Fraction / Sector


TP53 bp747 GC->TA
7 non-smokers (159 sectors)
Mean = 3.60x10-5
Median = 1.80x10-6
Sectors
Mutant Fraction / Sector


K-Ras bp35 GC->TA
2 Smokers (19 sectors)
Mean = 3.83x10-5
Median = 9.81x10-7
Sectors
Mutant Fraction / Sector


K-Ras bp35 GC->TA
5 Non-smokers (61 sectors)
Mean = 7.94x10-5
Median = 0
Sectors
Mutant Fraction / Sector


Hprt bp508 CG->TA
5 Smokers (86 sectors)
Mean = 4.78x10-5
Median = 3.17x10-5
Sectors
Mutant Fraction / Sector


Hprt bp508 CG->TA
7 Non-smokers (97 sectors)
Mean = 5.79x10-5
Median = 2.34x10-5
Sectors
Mutant Fraction / Sector

Because the observed lung MFs among sectors were not normally distributed, the Kolmogorov-Smirnov test, which permits a comparison of two non-parametric distributions, was applied. The test uses statistics of the maximum vertical distance (M) between two empirical cumulative distribution functions as a measurement of how close together theses two distributions are (Figure 26). Table 7 shows the results of M and the critical value d with given degrees of freedom for each mutation. For the TP53 bp742, bp746 and K-ras bp35 mutations, the null hypothesis was accepted at the 95% confidence level, suggesting that the MF distributions among sectors were essentially the same for smokers and non-smokers. The TP53 bp747 target rejected the null hypothesis, but when the "outliers", sectors carrying MF larger than $4x10^{-4}$ and comprising 4.5% of all sectors assayed for this particular mutation, were excluded from both group, the null hypothesis was accepted at 95% confidence level. The HPRT bp508 data rejected the null hypothesis at the 95% confidence level, which was independent from exclusion of the outliers. This was unexpected first because smoker is the group supposedly insulted with chemical exposure of cigarette smoke and second because HPRT is a non-oncogenic control gene and had been expected no difference between the two groups. However, when it is recognized that five independent comparisons have been performed, the expectation that any one of the five could be rejected at the 95% confidence level for a single trial is 25%, and this recognition leads to the overall implication of no statistically significant differences between smokers and non-smokers.

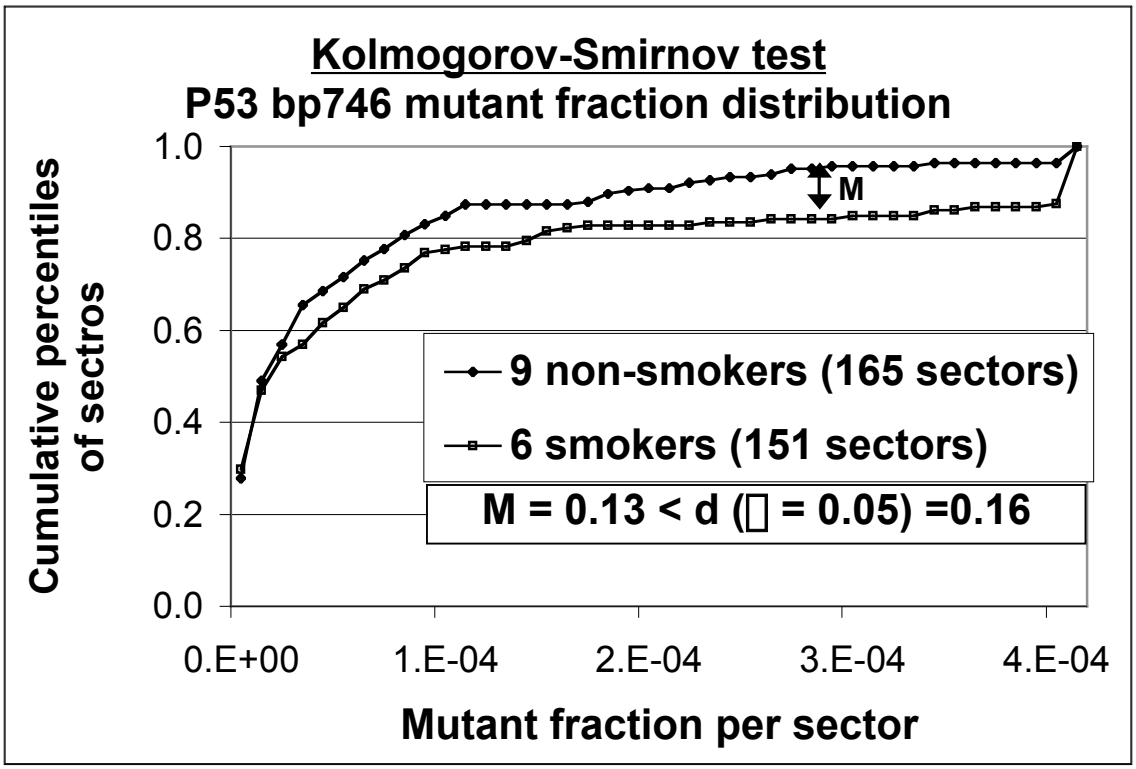


Figure 26. Empirical cumulative distribution functions of MFs from 151 sectors from six smokers (squares) and 165 sectors from nine non-smokers (diamonds) for the TP53 bp746 mutation. M indicates the maximum vertical distance of the two empirical cumulative distributions and d indicates the critical value at a significance level of 0.05 with given degree of freedom. For this target, the null hypothesis was accepted with 95% confidence limit (Kolmogorov-Smirnov test).

Table 7. Results of the Kolmogorov-Smirnov tests for the MF distributions among sectors of smokers and non-smokers.

	M	d (α = 0.05)	Null hypothesis
TP53 bp742	0.30	0.39	not rejected
TP53 bp746	0.12	0.15	not rejected
TP53 bp747	0.20 [0.13]	0.15 [0.16]	rejected [not rejected]
K-ras bp35	0.28	0.36	not rejected
HPRT bp508	0.27	0.20	rejected

* M indicates a maximum vertical distance between the two cumulative distribution functions.

** d indicates the critical value with a given degree of freedom, and α is a significance level.

*** The numbers and result within blankets in TP53 bp747 are the result after the outliers, sector with MF larger than 4×10^{-4} (4.5%), are excluded.

4.2.4. Estimating turnover unit size of human tracheal bronchial epithelium

Our data cannot be interpreted in terms of single mutants individually arising and dying within the sector assayed. How might they be distributed and why? It has been postulated that cells in tissue do not randomly distributed, but rather clustered as a turnover unit in which a multi-potent stem cell can asymmetrically give rise to a stem cell itself as well as a transition cell that can eventually give rise to fully differentiated terminal cells. The first genetic hit for initiating a mutant presumably happens only to a stem cell because other cells would be inexorably shed off from the unit and not keep the first mutation long enough for the subsequent genetic hits to be accumulated (Cairns, 1975). The mutation occurred in a stem cell thus eventually spreads to the rest of the unit which represents a mutant colony.

In order to test the idea that mutants were clustered as individual turnover units, the expected Poisson distribution was calculated as a random distribution of such mutant clusters over all sectors assayed but excluding the outlying sectors with more than 1,000 mutant cells. Using hypothetical cluster sizes of 2^n cells: 4, 8, 16, 32, 64, 128 and 256 cells, the numbers of observed mutant cells were then converted to the numbers of turnover units and the resulting cluster distributions were compared with the expected Poisson distribution. From Figure 27, it was clear that the expected Poisson and observed distributions of clusters per sector were poorly matched for the assumption of 4, 8, 16, 128 and 256 cells / turnover unit.

To obtain a better estimate of the turnover unit size, the Kolmogorov-Smirnov comparison was again applied in order to discover which hypothetical cluster sizes best fit in my observations. In contrast to a preliminary suggestion of 32 cells for the best fit, analyses of the sectors available from all lungs revealed that 64 cells was indeed the only hypothetical turnover unit size which accepted the hypothesis at the significance level of 5%, suggesting that these two distributions were similar (Figure 28).

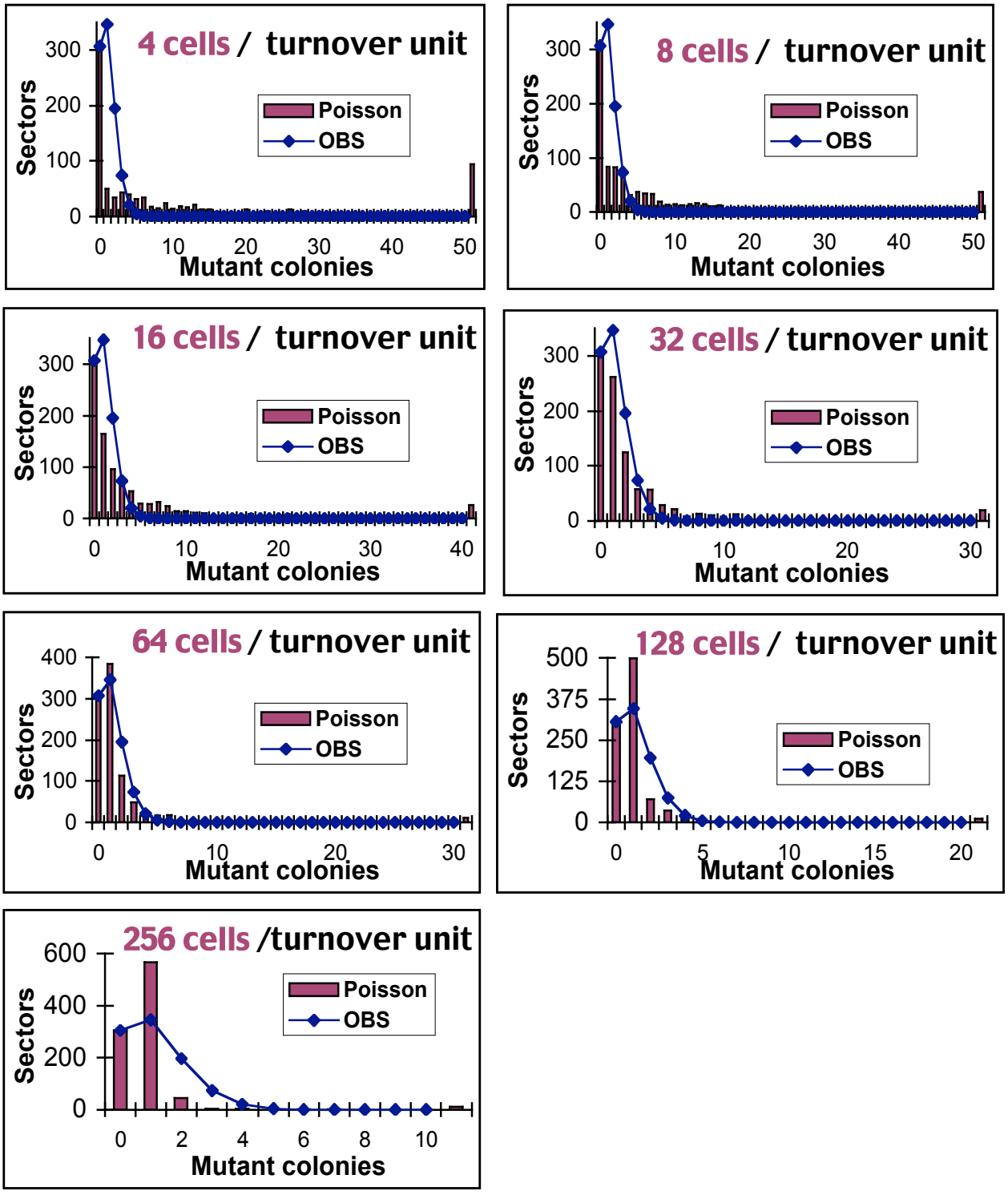


Figure 27. Estimating the turnover unit size of human lung epithelium using mutant cells obtained from all sectors analyzed except outliers. The observed mutant colonies per sectors (OBS) were estimated from mutant cells per sector divided by hypothetical turnover unit size of 2^n: 4, 8, 16, 32, 64, 128 and 256 cells, and compared with Poisson distribution (Poisson) calculated based on the observed number of "0" mutant cells per sector.

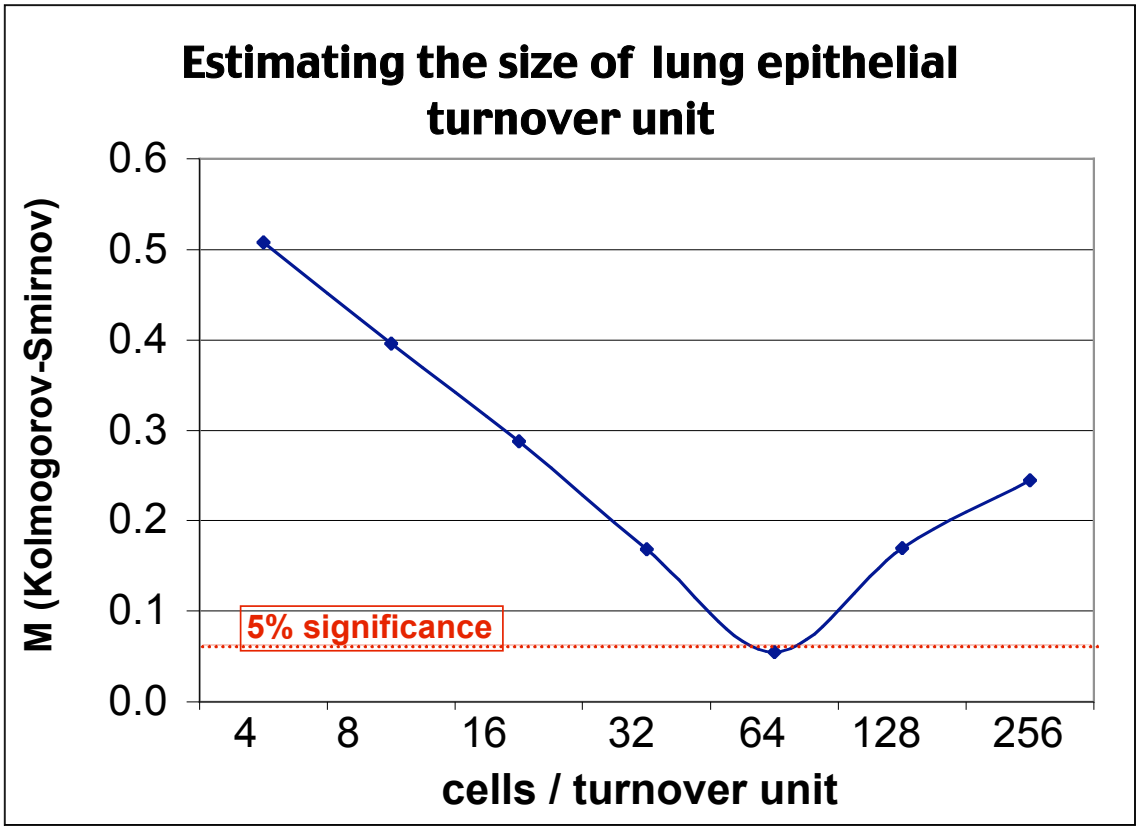


Figure 28. The hypothetical size of turnover unit for the human lung epithelium which reached the 5% significance level by Kolmogorov-Smirnov the test. The maximum vertical distance (M) between the observed and Poisson distributions obtained from figure 25 indicated how far these two distributions were apart. Only the hypothetical turnover unit size of 64 cells reached to the 5% significance level (the critical value d = 0.063 > M = 0.055).

The validity of this estimate depends essentially on the accuracy of determinations of total input copy numbers and observed mutant copy numbers in all sector assayed. Biases in either estimate would perforce be reflected in the accuracy of this estimate of cluster size. Provisionally, however, this value "64" can be interpreted to represent the previously unknown size, 2^6 cells, of overall maintenance turnover units in the human bronchial epithelium. It is biologically significant that this important tissue parameter has not been affected by years of heavy smoking despite the morphological changes created in the epithelium by cigarette smoke (Auerbach, 1957, 1961). These data and analysis led to a conclusion that the average turnover unit size for all mutants over all sectors (905 sectors without outlying values) is at or about 64 cells: one stem cell, 31 transition cells and 32 terminal cells. The turnover unit size of 64 cells is comparable to ~60 cells, the turnover unit size estimated for the human normal skin epithelium based on an observation of the number of cells stained for TP53 mutations (Jonason et al, 1996).

4.2.5. Outliers: sectors containing MF larger than $4x10^{-4}$

Most of the tracheal-bronchial epithelial sectors assayed herein contained 10^5-$5x10^6$ total cells and yielded mostly 0-200 mutant cells, or MF of up to 10^{-4}, whereas some rare sectors contained thousands to half a million mutants, yielding MF of 10^{-2} in a few extreme cases (Table 8). The sectors with MF larger than $4x10^{-4}$ which was outside of the upper 95% confidence limit and could not be distributed by the Poisson (section 4.2.4) were thus defined as outliers and excluded from the statistical tests performed above. 44 out of 949 (4.6%) sectors were categorized as outliers and present both in the smokers and non-smokers' lungs (Figure 23). The concordance of high mutant numbers in micro-anatomically adjacent sectors (grey highlights in Figure 23) suggests that these mutants were very large colonies extending over several sectors. Thus, the number of large mutant clusters with more than 1,000 mutant cells can be interpreted 12 not 19 such colonies were found among the 949 sectors per mutation assayed. Unfortunately, the numbers of such outlying sectors per each target mutation were too small to perform any valid statistical test, therefore, the association of the outliers with cigarette smoking cannot be clarified in the present study.

Table 8. Sectors containing more than 1,000 mutants

Non-Smoker				Smoker			
Lung	**Sector**	**MTcells/sector**	**MF**	**Lung**	**Sector**	**MTcells/sector**	**MF**
P53 bp742	0			0			
P53 bp746					RLL1-2	7067	7.01E-03
	LLL12	3198	5.89E-03	**SM III**	RLL15	1775	2.03E-03
NS VI	LLL14-3	1164	2.34E-03		L4	1426	2.34E-03
	LMS4	17436	5.02E-03				
	RMS1	4044	4.63E-04		RLL3	7866	2.37E-03
					RLL3*	76476	1.90E-02
					RLL4*	249126	5.59E-02
				SM V	RLL5	11556	2.04E-03
					RLL5*	43401	1.34E-02
					RUL10+11	1522	1.20E-03
P53 bp747	0			**SM II**	LB7	1818	2.49E-03
				SM III	LUL2-2	1567	7.25E-04
					RLL3*	1727	4.30E-04
				SM V	RLL4*	50687	1.14E-02
					RLL5	2282	4.03E-04
					RLL5*	28442	8.78E-03
K-ras	0			0			
HPRT	0			0			

* MT, SM and NS stand for mutant, smoker and non-smoker, respectively.

** Grey highlights indicate that these mutants might be very large cluster extending over several sectors.

*** Labeling of lungs corresponds with Table 3.

4.3. Human model tissue -- Colon

While this thesis focused on the human tracheal bronchial epithelium for the purpose of testing the specific hypothesis that cigarette smoking induces nuclear point mutations therein, the observations were extended to histologically normal colon as a tissue-control. It would be possible that some lung-specific bias existed, creating the appearance of high point MFs. To control for such a bias, normal colorectal epithelial sectors, each containing 10^7-$5x10^7$ cells, from two donors were assayed for TP53 bp742, bp746, bp747 and K-ras bp35 under the same protocol used for the lung assay. Two colon sectors assayed for TP53 bp746 and bp747 appeared to be negative; 9 - 14 mutant copies were observed per sector while TK6 control background was in average 30 copies when $7x10^6$ copies of the template were introduced for MAMA. The larger amount of template copies was used for colon because 10^6 copies of colon epithelial cells had previously shown negative results. This negative finding was indeed consistent with the mutational spectra of these target mutations that are 1.7% hotspots in lung tumors but have very low, if any, frequency in colon tumors. In contrast, TP53 bp742 and K-ras bp35 mutations have been reported in many colon tumors; the former is a 6.9% hotspot in the colon tumor and the latter were occasionally observed even in colon adenoma. Nevertheless, all 16 sectors assayed for TP53 bp742 resulted negative or the MF lower than $3x10^{-6}$; 13 – 25 mutant copies were observed per sector while TK6 background was in average 19 copies when $2x10^6$ copies of the templates were introduced. Likewise, all 24 sectors assayed for the K-ras bp35 mutation resulted in negative or MF lower than $2.5x10^{-6}$: 0 – 10 mutant copies were observed per sector while the TK6 background was 6 copies (n = 21) when 1.5 or 2 $x10^6$ copies of the templates were introduced for MAMA (Table 9). In summary, the observed colon MFs for all mutations assayed were very close to the TK6 background level, and none of the normal human colon sectors assayed showed a clear positive signal, or mutant colonies despite of the very high mutant frequency observed in the colon tumors (for TP53 bp742 and K-ras bp35). This result agrees with the analyses of colon cancer mortality data that suggest that the level of point mutations in the colon is at least ten-fold lower than apparently observed in the lung (Herrero-Jumanez, 2000). The tissue-specificity of the MF is thus very distinct in human

lung and colon. Herein, the colon served as a good negative control for random *in vivo* chemistry creating assay biases.

In the section 4.2.2., the mean MFs of lungs were very similar between smokers and non-smokers as well as across five different target point mutations, including HPRT which is a non-oncogenic control gene. Herein, the colon sectors as well as TK6 cell line showed MF lower than 10^{-5} for all or most of the target mutations; therefore, the similarly high level of MFs observed in both the smoker and non-smoker group as well as over five different target mutations described in the section 4.2.2. was not artificial errors but real.

Table 9. Total input copies, observed mutant copies and MFs per assay for colon A (A), colon B (B) and TK6 background (C).

A

Colon A (Male 58yo) 16 sectors

Sample	Copies/assay	MT#/assay I	II	III	mean MT#/assay	MT-BGc#/as	MF(-BG)
#1	1.5E+06	11.7	8.7		10.2	4.2	3.3E-06
#2	1.5E+06	0.4	0.1		0.3	0	0
#5	1.5E+06	0	0.3		0.2	0	0
#6	1.5E+06		0		0.0	0	0
#7	2.0E+06	9.1	11	11	10.4	4.4	2.54E-06
#10	1.5E+06	2.4	2.1		2.3	0	0
#11	2.0E+06	1	8.8	5.7	5.2	0	0
#15	1.5E+06	0	0.6		0.3	0	0
#23	1.5E+06	0	1.3		0.7	0	0
#25	2.0E+06	2.2	2.3		2.2	0	0
#29	2.0E+06	3	5.8	1.7	3.5	0	0
#30	1.5E+06	0	1		0.5	0	0
#31	1.5E+06	3.2	0.3		1.8	0	0
#37	2.0E+06	1	0.8		0.9	0	0
#38	1.5E+06	1	1.5		1.3	0	0
#39	2.0E+06	0.3	0.8		0.6	0	0
						mean MF	3.63E-07
						SD	1.00E-06

B

Colon B (Female 60yo) 8 sectors

Sample	Copies/assay	MT#/assay I	II	III	AveMT#/assay	MT-BGc#/as	MF(-BG)
#3	1.5E+06	1.9	0.6		1.3	0	0
#9	1.5E+06	6.5	1.6		4.1	0	0
#14	2.0E+06	1	4.3		2.7	0	0
#18	2.0E+06	3.3	8.3		5.8	0	0
#22	2.0E+06	2.1	6.2		4.2	0	0
#26	2.0E+06	3.2	0.5		1.9	0	0
#33	2.0E+06	7.7	10		8.9	2.8	1.78E-06
#36	2.0E+06	2.8	2.9		2.9	0	0
						mean MF	2.23E-07
						SD	6.29E-07

C

TK6 (n=21)	
Copy#/trial	BG copy#
1.5E+06	15.6
1.5E+06	8
1.5E+06	2.8
1.5E+06	2.9
1.5E+06	1.6
1.5E+06	1.4
1.5E+06	10.9
1.5E+06	17.1
1.5E+06	5.8
1.5E+06	6.7
1.5E+06	9.9
1.5E+06	14.5
1.5E+06	17.4
1.5E+06	5.9
1.5E+06	1
1.5E+06	1.6
1.5E+06	3.1
1.5E+06	0
1.5E+06	0
1.5E+06	0
1.5E+06	0
Mean	6.01
SD	5.99
BG MF	4.0E-06

*Some colon sectors were analyzed in duplicates or triplicates.

* Grey highlight indicates that the signal was lower than TK6 background (negative).

5. DISCUSSION

An overwhelming number of epidemiological and clinical evidences have indicated that cigarette smoking causes cancer. However, the mechanism of how exactly cigarette smoke cause lung cancer in a molecular and/or genetic level has remained poorly understood. The most prevailing hypothesis is based on mutagenic effects of chemicals in smoke and its consequent increase in the rate of DNA point mutations in tumor suppressor genes or oncogenic genes in smoker's lung. However, this hypothesis has never been directly tested in normal human lung epithelium of smokers and non-smokers due to a lack of sufficiently sensitive technology. The present work has demonstrated that the great sensitivity and specificity of the MAMA technique in the combination of target sequence enrichment, restriction of wild-type alleles and CE, enabled detection of rare point mutations in normal human tissues and allowed to test the mutagenic hypothesis of cigarette smoke in lung cancer. This powerful technique with the sensitivity of 10^{-5} was applied for detection of point mutations in TP53 bp742, bp746 and bp747, K-ras bp35 and HPRT bp508 from 291 individual tracheal-bronchial epithelial sectors from six smokers and nine non-smokers.

5.1. No Significant Difference in MFs of Smokers and Non-smokers

Lung epithelial sectors acquired contained $2.3x10^6$ cells in average and harbored 0-200 mutant cells in general, equivalent to MF of $0\text{-}10^{-4}$ with a rare exception of sectors containing giant mutant clusters. Noticeably, the variances of the MFs within the same lung reached almost 100-fold (10^{-5}-10^{-3}) for some individuals regardless of the donor's smoking status, which is consistent with previous observations that mutant frequency in human subjects varies as much as three orders of magnitude (Cole and Skopek, 1994). The distributions of the mean MFs per lung were very similar between smokers and non-smokers as well as across five different target mutations. When the mean MFs were normalized for all five target mutations, the distributions of the relative MFs of smokers and non-smokers were indistinguishable. Similarly, the distributions of MFs among sectors did not differ significantly between smokers and non-smokers by Kolmogorov-Smirnov test for all target mutations but HPRT. In addition, the mean MFs appeared slightly higher in females than males. Likewise, MF increased over age of the subjects

although the correlation did not reach to the statistical significance due to large variances within the same age group.

Our collaborator, Dr. L.C. Jing in the Zarbl laboratory at the Fred Hutchinson Cancer Research Center concurrently analyzed for the same K-ras bp35 G:C->T:A mutation on 209 sectors from four lungs: smokers V and VI and non-smokers IV and V, by using the same MAMA-based method. He has also observed large intra-individual variances of MFs ranging from ~$5x10^{-6}$ to ~10^{-3}, and very similar mean MFs among lungs between two smokers and two non-smokers: $1.0x10^{-5}$ (n=50) and $1.6x10^{-4}$ (n=51) vs. $1.5x10^{-4}$ (n=66) and $3.0x10^{-5}$ (n=42), respectively. This same outcomes brought by two independent researchers strongly suggest that cigarette smoking does not increase MFs in the chosen oncogenic genes.

The present observation is also consistent with H. Coller's mitochondrial study (1998) demonstrating that mutational spectra of all seventeen hotspots from bronchial epithelial cells of smokers and non-smokers were indistinguishable; even twins who were discordant in their smoking status resulted in no difference. These experimental outcomes led us to question the conventional link of chemically induced mutation to chemically induced cancers.

5.2. Potential Mechanisms of Causing Lung Cancer Without Inducing Mutations

If point mutation rates are not affected by cigarette use, how does cigarette smoking cause lung cancer?

5.2.1. Case examples – a mouse mammary model treated with MNU

To examine whether chemicals cause tumor by mutating cellular oncogenes, Zarbl et al (1985) characterized and compared the type of oncomutations found in N-Nitroso-N-methylurea- (MNU) induced rat mammary tumors to the known or suspected mutational specificity of the carcinogens. A specific G:C->A:T transition in codon 12 of the H-ras gene was found to be reproducibly associated with MNU-induced rat mammary tumors; nevertheless, they later found that the same type of mutants was already present in the mammary epithelial cells in rats before MNU treatment. Moreover, the mutants were clustered within organ sectors, consistent with their origin as mutational events during maintenance turnover divisions of stem cells in the mammary epithelium. Most

importantly, it was found that exposure of pubescent female rats to a single carcinogenic dose of MNU did not affect the number of H-ras mutants, the fraction of organ sectors containing mutant cells or the fraction of animals harboring mammary epithelial cells with H-ras mutations (Cha et al., 1994). Thus, despite a correlation between the mutagenic specificity of MNU and the type of oncogene mutation detected in MNU-induced tumors, MNU did not initiate mammary carcinogenesis by mutating H-ras gene; instead, the MNU-induced carcinomas arose from mammary epithelial cells that had already harbored activated H-ras proto-oncogenes prior to the carcinogen exposure. Subsequently, the Zarbl laboratory has discovered that MNU treatment changes tissue-specific hormonal regulation of DNA conformation within the H-ras1 promoter in rat mammary cells, suggesting a novel epigenetic mechanism of carcinogen action (Jin et al., 1996).

It has also been observed that carcinogen exposure physically disrupts the normal tissue architecture resulting in tumor development in mice mammary gland (DeOme et al., 1978; Medina et al., 1978). Breakdown of the tissue structure by carcinogen exposure might provide endogenously mutated clones with a selective and advantageous microenvironment for growth (Rubin, 2001b, 2002). In this mechanism, mutant colonies insulted by carcinogen exposure do not have to increase the mutation rate at the initiation of a neoplasm, which is agreeable to the current observation.

5.2.2. Mathematical modeling of lung carcinogenesis

As Moolgavkar and Knudson (1981) pointed out, the division and death rate of preneoplastic colonies could define the probability that a newly initiated preneoplastic cell will survive and eventually give rise to a tumor. Similarly, these kinetic rates governing the growth rate of a preneoplastic colony would affect the average age of individual mortality from lung cancer by given inception of the cigarette usage. Herrero-Jimenez's mathematical analyses of the age-specific lung cancer mortality of smokers and non-smokers concluded that the mutation rates of both tumor initiation and promotion remained constant while significantly faster growth of preneoplastic lesions of smokers (0.33 doublings / year) was predicted as compared with non-smokers (0.17 doublings / year) (thesis, 2001). The constant mutation rates during neoplastic growth

are concordant with the previous findings that carcinogen exposure may not "induce" mutations, but rather "select" previously existing mutants, and promote their growth to develop as preneoplastic colonies from which tumors would arise. Cigarette smoke might exert its effect by stimulating hyperplasia of some stem cell turnover units, allowing them to have more than its usual number of 64 cells.

5.3. Turnover Unit Size

The current results represent the first report of detecting TP53, K-ras and HPRT mutations in normal lung tissue of individuals free of lung cancer, but the important contribution also lies in the division of the airway into anatomically distinct sectors so that the distribution of mutants in clusters was discovered. Mutants were found to be non-randomly distributed among sectors. Most mutant copies were clustered as "colonies" presumably derived from a single mutant stem cell. The present observation suggested that the size of turnover unit for human tracheal and bronchial epithelium is approximately 64 cells, including 1 stem cell, 31 transition cells and 32 terminal cells, assuming a simple binomial expansion. Some rare sectors might contain smaller or larger sizes of mutant clones, but in general, the total number of mutational events seemed to be about equal in smokers and non-smokers.

Not many organs have revealed its turnover unit size. Jonason et al. observed that normal human skin contains clonal patches of TP53-mutated keratinocytes in size of 60-3000 cells. Within $1cm^2$ of the epidermis, they found $4.4x10^6$ keratinocytes and $7x10^4$ stem cells, leading to 63 cells per turnover units (1996). In colon crypt epithelium, 2000 cell per turnover unit was counted (Dr. Emma E. Furth, personal communication).

Multifocal areas of precancerous molecular changes with independent origin in histologically normal tissue, a phenomenon referred as "field cancerization", was first reported in oral caner by Slaughter et al. (1953). Since then, field cancerization has been repeatedly observed in normal bronchial epithelium of patients with lung cancer (Sozzi et al., 1995; Park et al. 1999; Wang et al., 1999; Park et al, 2000). It is plausible that such numerous small clonal patches of molecularly altered epithelium could be arisen from individual turnover units. The current observation supports the presence of a low level of mutant colonies even in the normal human tracheal bronchial epithelium. The growth of

these mutant colonies in the normal tissue might be enhanced by some triggers, possibly exposure to environmental carcinogens such as cigarette smoke, leading to a simultaneous development of multiple mutant subclonal patches with independent origins, observed as field cancerization. In fact, smokers were reported to exhibit more individual clonal patches of molecularly altered epithelium in the lungs (Barsky et al., 1998). The sizes of such mutant subclonal patch were measured in different organs: $1cm^2$ for bladder and gastric epithelium, 2mm diameter for skin and 10□m in diameter (~200 cells) for oral (Braakhuis et al., 2003). They could be the size of adenomas developed from individual turnover units.

5.4. Outliers

5.4.1. Rare sectors containing large numbers of mutants

In rare occasions, some sectors contained very large numbers of mutant cells, more than 1,000 mutant cells, or equivalently MF larger than $4x10^{-4}$, as opposed to the majority of sectors that harbors 0-200 mutant cell or MF smaller than 10^{-4}. Such sectors comprised of 4.6% of total sectors assayed and treated as outliers in most of the statistical tests performed herein. It appeared that smokers tend to harbor more outliers than non-smokers, particularly in TP53 bp746 and bp747 mutations. The number of outliers observed in this study was, nevertheless, so small that no valid statistical test could be performed. Therefore, the association of such outliers with cigarette smoking cannot be clarified in the current study.

The rare colonies containing large mutant clusters had been previously observed for mitochondrial DNA from both smokers' and non-smokers' bronchial epithelial cells (Coller et al., 1998). How and when mutants in these outliers were arisen cannot be answered in the present study. One potential source of these mutant clusters is mutational jackpots; mutations had occurred early in embryonic stage and undergone clonal expansion as the organism developed. However, the frequency of human jackpot mutation has not been known, and in general, there are too little studies to make the jackpot hypothesis convincing.

5.4.2. Hypotheses for how giant mutant clusters are arisen from normal tissue

How these giant mutant clusters in the rare sectors were arisen if the cigarette smoke is not inducing point mutations, is an interesting question, yet not answered by the present or previous studies. It is possible to conceive that they were arisen because i) the turnover rate of the stem cell unit was changed, or ii) the compartment size of the turnover unit was changed. In the former, a selected colony was endowed with growth advantage such that its mitotic rate is much higher than its apoptotic rate. In the latter, the compartment size of the turnover unit working as a physical confinement of cell growth was expanded. The latter hypothesis is consistent with the findings of Zhang and his colleagues (2001) who have also observed such mutant clusters in normal tissue of mouse epidermis, describing these rare clones as "imprisoned clones" because they continued to proliferate while not expanding their areas containing densely packed cells with very little cytoplasm. The imprisoned clones were only observed in the absence of UVB; with sustained UVB irradiation, the mutant clones were allowed to expand beyond their stem cell compartments. Thus, the rate-limiting step for mutant colony proliferation is not a UVB-induced proliferative (TP53) mutation but the expansion of the epidermal stem cell compartments. Analogically to lung cancer, it is plausible that chronic carcinogen exposure of tobacco smoke would evidently allows preexisting mutants to escape a barrier presented by the stem cell compartment arrangement. Clonal expansion of mutant epithelial cells emerges as an interplay between the obstacle presented by stem cell compartments and the driving force of physiological changes induced by sustained cigarette smoke exposure, which allows repeated breaching of this barrier. This hypothesis is consistent with Cairns' theory (1975) that stem cells pose a threat to multicellular organisms and must be constrained by stem cell compartments.

5.5. Endogenous Factors

How could the mutant be endogenously arisen if exogenous chemical carcinogens do not initiate mutation? Endogenous factors in fact play significant roles in human mutagenesis and carcinogenesis. A variety of mechanisms that causes spontaneous mutations and/or nullify gene functions have been postulated as well as experimentally demonstrated.

5.5.1. DNA polymerase errors

It has been known that DNA polymerase sometimes fruitlessly remove undamaged DNA bases, that would present the same opportunity for DNA polymerase error as removal from damaged DNA bases (Kolodner, 1996; Modrich and Lahue, 1996; Berdal et al., 1998). Assuming the repair rate *in vivo* is roughly equal to *in vitro*, Branum et al. (2001) estimated that $2x10^4$ nucleotides/day/human cell are subjected to the excinuclease action: the rate comparable with spontaneous DNA lesions. As resynthesis of DNA bases is invariably associated with mutations, gratuitous repair may be an important source of spontaneous mutations.

Khrapko et al. (1997b) studied mitochondrial mutation spectra in human cells and tissues and found the similarity of the hotspot sets in vivo and in vitro, leading to the conclusion that these mitochondrial mutations are primarily spontaneous in origin and arise either from DNA replication or reactions of DNA with endogenous metabolites. Moreover, W.M. Zheng of our laboratory has demonstrated that twelve out of the seventeen basepair substitution hotspots which accounts for 90% of the *in vivo* MF in both smokers and non-smokers' bronchial epithelium observed by Coller et al. (1998), are created when this same sequence is copied by the human mitochondrial DNA polymerase □. Similarly, B.P. Muniappan (2002) found significant concordance between *in vitro* replication errors of human DNA polymerase □ and *in vivo* point mutations of the adenomatous polyposis coli (APC) gene. In the tested 141bp DNA sequence in exon 15 of the APC gene, they observed that three out of seven hotspots created by polymerase □ were concordant with the APC hotspots detected in human colon cancer, which accounted for 54% of reported *in vivo* APC mutations. These observations suggest that endogenous mechanisms, such as excision of undamaged bases followed by DNA polymerase errors, might play significant roles as generating spontaneous mutations in human carcinogenesis.

5.5.2. Strand-biased repair and BPDE adduct

Mutation frequency depends not only on initial damage frequency, but also on the repair rate for each individual lesion; thus, repair efficiency may strongly contribute to the mutation spectrum in a cancer-associated gene. Repairing of damaged DNA does

not happen equally to the entire regions at risk, but strongly biased in transcribed strands because nucleotide excision repair is coupled with transcription. Preferential repair and strand-specific repair of BPDE adducts was demonstrated in the HPRT gene of diploid human fibroblasts (Chen et al. 1992; Wei et al; 1995). In addition, Denissenko and his colleagues (1998) reported that repair of BPDE adducts formed at mutational hotspots in TP53 gene is two to four times slower for sites in the non-transcribed strand than sites in transcribed strand, which might consequently lead to the strand-bias of G>T transversions in lung cancer.

After demonstrating that the types and patterns of mutational spectra of lung tumors obtained from the IARC were indistinguishable between smokers and non-smokers (2000), Rodin and Rodin focused on the strand-biased repair, rather than mutation induction, as the target mechanism affected by cigarette smoking (2002). They hypothesized that smoking may inhibit repair of G>T primary lesions on the non-transcribed strand. Because smokers and non-smokers were indistinguishable in the origin of primary lesions (2000), the difference emerges later, due to unbiased repair of non-transcribed strands. They further pointed out that cessation of smoking would be unlikely to result in the reported beneficial effect of reduced risk of lung cancer, if the major carcinogenic mechanism of smoking were in producing irreversible mutations. They concluded that smoking aggravates selection pressure, possibly via repair inhibition in non-transcribed strands, rather than inducing mutagenesis.

5.5.3. Loss of heterozygosity (LOH) and loss or gain of imprinting (LOI/GOI)

Cigarette smoking could directly or indirectly raise the rate of other potentially rate-limiting genetic changes in initiation and promotion of lung carcinogenesis without changing point mutation rates. Examples would be loss of heterozygosity (LOH) or loss or gain of imprinting (LOI/GOI) unrelated to point mutational events.

Because LOH events, including chromosomal exchanges and recombination that have a potential to trigger tumor development, could be affected by environmental factors, it is necessary to develop means to measure such changes in human tissues. For instance, the laboratory of J.D. Minna has found markedly elevated fractions of lung epithelium samples (600 - 800 cells) with detectable LOH events in approximately half of

the biopsies taken from current and former smokers, whereas no samples with measurable LOH were detected in non-smokers (Wistuba et al., 1997). Most interestingly, micro-sectors with elevated LOH were also found in the apparently normal epithelium of some smokers. Their findings lead the possibility that the ensemble of genetic changes leading to lung cancer could be accelerated through LOH-like events without accelerated rates of point mutations must now be considered seriously.

Similarly, changing DNA methylation patterns (imprinting) is an alternative mechanism for gene inactivation in cancers without changing mutation rates. Cytosine methylation is a post-replicative epigenetic modification of DNA that may lead to changes in gene expression without changing the sequence. Methylation of promoter regions represents a plausible pathway of "gene silencing" and thus loss of functional heterozygosity in tumor initiation or promotion. In fact, many tumor suppressor genes showing gains or losses of methylation status at the promoter regions in tumors as opposed to normal tissue have been discovered (Jones and Laird, 1999). Tracheal bronchial epithelium of smokers contained more aberrant methylation than that of never-smokers in multiple genes such as p16 and APC (Toyooka et al., 2003) as well as RARβ-2 (retinoic acid receptor β-2), RASSFIA (Ras association domain family I) and H-cadherin (Zochbauer-Muller et al., 2003). Furthermore, it is also suggested that hypomethylation status within exon 5-8 of TP53 gene from peripheral lymphocyte DNA of heavy smokers with lung tumors was associated with a 2-fold increased risk of lung cancer (Woodson et al., 2001). DNA methylation also could silence mismatch repair genes, leading to increased mutations by an epigenetic mechanism, as illustrated in hereditary non-polyposis colorectal cancer patients.

It is not clear if a part of a general change in methylation of genes is just concurrent with required changes in tumor initiation or promotion, or is an event selected during tumor progression. Further investigation is necessary to elucidate how LOH and LOI/GOI status changes could contribute to carcinogenesis.

5.5.4. Inflammation / hormonal response

It has been reported that asymptomatic smokers develop morphological changes in their surface airways such as displacement of ciliated pseudo-columnar cells by

unciliated stratified squamous cells causing bronchial wall thickening (Auerback et al., 1957, 1961), goblet cell hyperplasia (Thurlbeck et al., 1975; Spurzem et al., 1991) and recruitment of T-lymphocytes, neutrophils and macrophages (O'Shaughnessy et al., 1997; Saetta et al., 2000). Some, if not all, of these changes overlaps with the reactions caused by immune response. Indeed, using cDNA microarray and StaRT PCR technique, our collaborator, Dr. H. Zarbl at the Fred Hutchinson Cancer Research Center has observed that inflammation related genes were up-regulated in epithelium from smokers' bronchus and bronchioles (personal communication). In addition, his laboratory has previously demonstrated that hormonal regulation of DNA conformation induced by carcinogens promoted the growth of endogenous mutants (Jin et al., 1996), suggesting a novel epigenetic mechanism of carcinogenesis. It is thus conceivable that cigarette smoke is also pseudo-hormonal, lowering the continued existence of normally terminal cells. It is not necessary to overstate the potential meaning of these results for general understanding of chemically induced carcinogenesis. Many before us have already stressed the absence of direct data relating induced genetic changes to carcinogenesis and introduced the ideas of inflammation and hormone-mimicking and other selection parameters as alternatives to direct induction of requisite oncogenic changes. In order to pursue these ideas, careful investigation on how immune and/or hormonal reactions against an insult of cigarette smoke exposure at the airway are associated with the rise of lung carcinogenesis.

5.6. Mutations in Other Normal Tissues

For fifteen lungs analyzed for the five point mutations, the grand average MF of normal tracheal bronchial epithelium for both smokers and non-smokers was 3.3×10^{-5}. Considering the frequency of chosen hotspots, which is 1.7% in average, and the average age of the donor (~50 years old), the gene inactivation rate can be estimated as:
3.3×10^{-5} / 1.7% / age 50 = 1.9×10^{-5} gene inactivation event / year.
This rate is much higher than the rate estimated for other tissues.

5.6.1. Colon

Our collaborator, Prof. E.E. Furth of in the Department of Pathology at the University of Pennsylvania Medical School, has completed a long series of observations

enumerating the mitotic and apoptotic cells in colonic adenomas from many patients. Her observations led to an estimate that each crypt, which would represent a turnover unit of colon epithelium is comprised of ~2,000 cells.

In the current study, ~20 colon epithelial samples were assayed for TP53 and K-Ras mutations. The number of turnover units with the size of 2,000 cells in the average sector containing $2.5x10^7$ cells would be:

$2.5x10^7$ / 2,000 = $\sim 10^4$ turnover units / sector

Because none of the sectors assayed provided positive signals, or no mutant colony, the MF of colon epithelium is lower than 10^{-4} for 6.9% hotspot in TP53bp742. In order to determine the mutation rate of normal colon epithelium, further development of more sensitive and effective high-throughput technology is essential.

Prof. E.E. Furth also calculated the division and death rate of colon epithelium as about 9 per year. Combined the mathematical model which P. Herrero-Jimenez (2000) has developed, the adenoma growth rate was estimated as 0.16 doublings per year and the mutation rate of $2x10^{-7}$ per cell division in colon adenoma.

5.6.2. Skin

Skin cancer is the only cancer in which a link between its causality and the genetic mechanism: sunlight and its molecular footprint of CC->TT double-base change, has been known (Brash et al, 1991). Skin cancer caused by sunlight is quite an analogous to lung cancer caused by cigarette smoking. Jonason et al. (1996) observed that normal human skin contains multiple individual clonal patches of TP53-mutated keratinocytes, arising from the dermal-epidermal injunction. In sun-exposed skin, clones were both more frequent and larger (33 patches /cm^2) than in sun-shielded skin (3 patches /cm^2). These clones were 60-3000 cells in size and together involve as much as 4% of the epidermis. These numbers were comparative to the present observation of 4.6% of sectors with large mutant clusters in the normal lung epithelium. Furthermore, their patch size often varied 20-fold in the same individual, again consistent with the current observation of large intra-individual variances of MFs in the normal lung epithelium. They concluded that sunlight acts as a tumor promoter as well, endowing selective growth advantage to both endogenous and exogenous mutants. From the age

independence of TP53 clone frequency, substantial mutagenesis seems to occur during childhood and much less during adulthood, which might suggest that the mutagenicity is proportional to the net cellular growth rate. Interestingly, the lung cancer mathematical model suggested that the growth rate of normal juveniles (0.16 doubling / year) is very similar to the growth rate of preneoplastic lesions in smokers (0.17 doubling / year). Based on these findings, it can be speculated that cancers arisen in adults are initiated from a small fraction of adult cells was somehow maintained the positive cellular growth rate of juvenile, capable of developing to preneoplasias.

5.6.3. Blood

Mutations inactivating the HPRT gene and genetic changes causing LOH in HLA-A heterozygous have been measured in human peripheral T-cells in persons of varying age by observing colony formation in the presence of selective agents (Morley et al., 1982; Grist et al., 1992). The MFs show a great deal of variation among persons of the same age but over all appeared to increase lineally over age. A linear increase would be expected for mutations in stem cells if the stem cell number, turnover rate and mutation rate were reasonably constant throughout the life. It was estimated that an HPRT mutation rate of 2.5×10^{-7} per cell year and an HLA LOH rate of about 10×10^{-7} per cell year.

In the current study, the actual MFs in the lung epithelium were significantly higher than had been expected based on the published reports of peripheral T-lymphocytes in HPRT. The gene inactivation rate calculated in the lung epithelium is 1.9×10^{-5} / year, which is roughly 100 times higher than the rate for HPRT mutants in peripheral T-lymphocytes of persons in similar age, which is 2×10^{-7}. The coding region of TP53 is about 2.5 times longer than that of HPRT. Using this as an appropriated correction factor, it still appears that point mutations are occurring in stem cells of the tracheal bronchial epithelium at a rate of 100 / 2.5 or 40 times faster than estimated for stem cells of the erythro-leukopoietic system.

5.6.4. Kidney

The high MFs observed in the present study may not be restricted to lung epithelium. Martin et al., (1996) analyzed human epithelial cells from seventy-two kidneys and found that the MF was $5x10^{-5}$ for 10 year old, increasing to $2.5x10^{-4}$ for over the age 70 in the HPRT gene, concluding that somatic mutations are common in the kidney with the MF 10-fold higher than that has been described in human peripheral blood T-cells. This observation again emphasizes that the high lung MF observed in the present study is not due to an artifact, but rather to tissue-specificity.

Our findings did not reach to the reason why the lung carries such a high gene inactivation rate compared to other tissues, such as colon and blood. It is necessary to extend the analysis to other organs to investigate if this high MF is only tissue-specific and if not, to pursue the reason of what makes it different from other organs.

5.7. Potential Sources of Errors and assay background

The sensitivity of MAMA depends on how much amplification from excess wild-type alleles can be repressed while mutant alleles are accurately amplified. The improvement of the assay sensitivity would enable us analyses of a wider range of normal human tissues that might harbor lower MF than the lung. The followings are the possible sources of errors and the assay background in the present experimental procedure. These factors may be associated with sequence context such that the assay sensitivity is highly dependent on the nature of mutations chosen.

5.7.1. *Taq* polymerase error

Taq polymerase has relatively lower fidelity, $8x10^{-6}$/bp, which is a major factor to limit the assay sensitivity. For the HPRT bp508 mutation which originally had shown a high assay background ($\sim10^{-4}$), a hot-start based Sure Start® *Taq* DNA polymerase (Stratagene, La Jolla, CA) which carries higher fidelity was instead applied and successfully lowered the noise down to the level of $3x10^{-5}$. The only drawback of this polymerase is the cost; the Sure Start® *Taq* DNA polymerase is 650% more expensive

than the normal *Taq* DNA polymerase (New England Biolabs, Beverly, MA), hence not ideal for high throughput assays.

Other thermal stable polymerase enzymes (i.e., *Pfu, Pfu Turbo, Vent*, *Deep Vent*, etc.) that are reported to have higher fidelities than *Taq* polymerase would be useful for minimizing a risk of false positives. Due to the need to utilize MAMA primers containing one or more mismatches at the 3'- end, however, thermostable polymerase lacking the 3'- to 5'- exonuclease proofreading capabilities is favorable for the assay. Besides *Taq*, exo^{-} *Pfu* polymerase had first been tested in the current study, but its error rate of creating false positives was in fact higher than *Taq* polymerase, which is concordant to a previous report (Cline et al., 1996). Thus, *Taq* polymerase has remained the enzyme of choice for MAMA procedure. It should be noted, however, that the risk of false positives due to artifactual polymerase misincorporation or misligation can never been completely removed by enhancing the fidelity level of polymerase. It was thus critical to always analyzing tissue samples with similarly prepared negative controls (TK6) which provided the background signals.

The target mutations chosen in the present study are all G:C pair in wild-type (G:C->T:A or A:T) because this type of mutations is the most common in the lung cancer. The major deoxyguanosine mutation induced by *Taq* polymerase is G:C->A:T transition, and G:C->T:A transversion is least common. Therefore, the positive signals of three target mutations whose origins are controversially associated with cigarette smoke; TP53bp746, bp747 and K-Ras bp35, were less probable to attributed to *Taq* polymerase error.

5.7.2. Impurity in MAMA primers

Most of the MAMA primers purchased were purified by HPSF , a form of advanced HPLC, by the manufacturer (MWG Biotech, Inc., High Point, NC), providing with the purification level of <99.9% (<10^{-4}). This level of purification might not be sufficient when the assay pursues the sensitivity level of 10^{-5}. Such impurities might include even a small amount of contaminated primers with a single mismatch or perfect match with wild-type templates, capable of allowing primer-template annealing and extension, consequently increasing the false positive signals. The higher level of primer

purification system is therefore essential for minimizing the assay background, hence improving sensitivity of the assay.

5.7.3. Measurements of peak areas

The measurement of peak areas in CE is a factor that has an influence on the estimate of mutant cell numbers per sector, hence the MFs. PCR products that show distinguished peaks with little non-specific amplification in CE can be measured more accurately; in contrast, estimates of the areas of any closely migrated peaks are more difficult and can result in greater variations in the MF calculation. PCR products from tissue tended to contain more non-specific peaks than ones from cell culture. This is why replications of CE runs were performed as many times as the sample pool allowed particularly when non-specific amplification hinders the estimation of peak areas. In addition, thorough replacement of polyacrylamide gel matrix inside of a glass capillary or replacement of the capillary itself sometimes helped restore the resolution.

6. CONCLUSION

A proven technology for measuring rare genetic events in the human genome was used to test the hypothesis that cigarette smoking increases the number of point mutations in the lung epithelial cell population from which lung tumors arise. Contrary to the results predicted by the prevailing premise of chemical carcinogenesis, the present observation suggests that the quantitative distributions of all five target point mutations do not differ significantly between smokers and non-smokers. This finding in nuclear genes extends our previously reported observation that mitochondrial point mutations are not affected by cigarette smoking (Coller et al., 1998). Henceforth, we must seriously consider the alternative underlying mechanisms and possibly the proposition that the differences in the lung cancer risk between smokers and non-smokers are due to epigenetic or phenotypic changes resulting in "selection" of spontaneously arising mutant cells. Such selective mechanisms could be as subtle as stimulating a positive net growth rate in preneoplastic lesions. In humans, a small change in the net growth rate of preneoplastic colonies would be expected to have a dramatic effect on observed cancer incidence rates as a function of age (Herrero-Jimenez et al., 2000). The careful measurement of cell kinetics in lungs of smokers and non-smokers may provide additional information on the means by which exposure to the inhaled chemical mixture affects the age-specific lung cancer rate. It is also possible that pathways to lung cancer differ among subpopulstions of smokers and non-smokers. In any cases, it is going to take a substantial amount of work to sort out the key phenomena.

Many before us have stressed the absence of direct evidence linking induced mutations to carcinogenesis. It is an invalid argument to conclude that any genetic mutation observed in a tumor has been induced, selected or induced and selected by a prior experimental or environmental exposure. It became necessary to empirically demonstrate mutation induction in tissues for each example with appropriate controls rather than claiming it by association. Overall, the significance of this major shift in the paradigm of environmental carcinogenesis cannot be underestimated.

7. SUGGESTION FOR FUTURE STUDIES

7.1. Colon Cancer and APC Gene

In spite of vigorous efforts by the cancer genetic community, the gene(s) which initiate(s) the first mutant cells during lung carcinogenesis has not been discovered. In contrast, colon carcinogenesis is one of the most established human cancer model and its gatekeeper gene has been identified as APC of which mutational spectrum contains a few strong hotspots. Inactivation of this tumor suppressor gene is essential for >80% of human colon carcinomas and is specifically lost during initiation steps leading to adenomatous preneoplastic lesions. With an advantage of the substantial knowledge, we are currently planning to analyze APC mutations in human colon epithelium and studying the fundamental steps of human carcinogenesis with the combination of mathematical modeling.

7.2. Micro-arrays and Nano-cuvettes

MAMA is suitable for rare mutational analyses of any base sites in surgical, biopsy and other clinical specimens. However, the assay is highly labor-intensive and multiplexing, generally leading to a need of high-throughput automated systems that enable simultaneous analyses of multiple point mutations in hundred of samples.

In collaboration with the Hunter laboratory in the MIT BioInstrumentation Laboratory, we are developing an instrument that adapts MAMA to a high throughput micro-cuvette assay system which permits automatic distribution and mutational analyses of micro-anatomically distinct cell sectors. The instrument will enable to 1) computer-control micro-dissection of tubular organ specimens, 2) stack a tissue-containing microarrays with other microarrays containing digestion or PCR mixture, permitting direct robotic assays without any sample transfer, 3) proceed PCR with an internal heat transfer system and 4) measure the DNA amplification in a real-time manner with an attached detection device.

First, the instrument anatomically segregates mucosal sheets or thin slices of solid organs into thousands of micro-cuvettes in each of which a sample is independently analyzed for the presence of any of a set of desired specific point mutations. The BioInstrumentaion Laboratory has already fabricated microarrays containing 10,000 x

40nl micro-cuvettes in which each edge is 200μm wide and the depth is 1mm. The criterion for cutting precision is 500nm to permit automatic dissection based on local cell morphology.

A means to transfer liquids from one microarray to another has been already developed (Kanigan et al., 2000). Using this technique, the tissue-containing nano-cuvette array will be juxtaposed to another array containing the desired cell-digestion mixture including proteinase K and RNase. The "stacking" of these two arrays will mix the liquids in the two inline micro-cuvettes in a few seconds by passive diffusion followed by digested reaction, possibly with acoustic energy transfer to shred tissue samples. The third microarrays which contains necessary PCR mixture such as DNA polymerase and multiple pairs of primers for each intended assay, will be again stacked and proceeded to the MAMA process with an appropriate internal heat transfer system. The attached detection system will measure the amplification in a real-time manner. The necessary heat transfer steps for PCR have already perfected for micro-cuvette arrays (Hunter, unpublished), and we are currently testing the reproducibility of the system using DNA isolated from cell lines. We propose 10,000 micro-cuvettes each of which holds 100nl in volume.

7.3. TaqMAMA

Fortunately, the genetic analysis of the human tracheal bronchial epithelium was possible to perform manually due to its high MF therein. These fractions approach an estimated 2×10^{-3} inactivating mutations / allele for the TP53 gene. However, studies of HPRT MFs in human peripheral T cells have found inactivating MFs of about 2×10^{-5} (Grist et al., 1992). Calculations based on quantitative carcinogenesis models suggest the lower MFs to be expected for organs other than the lung. It is technically and physically challenging to analyze these organs with low MFs because the tissue segments have to be cut and divided into a large series of small sectors so that the sensitivity of the current assay would permit the detection of mutants within each sector. A conventional capillary electrophoresis approach in which each sample has to be run one at time is not well suited for such high sample throughputs.

Dr. Skopek at the Merck Research Laboratory (West Point, PA) has combined the highly sensitive MAMA procedure and TaqMan© assay to create TaqManMAMA, or TaqMAMA (Glaab and Skopek, 1999). The TaqMan© assay involves 5'-exonucleotic liberation of a fluorescent-labeled nucleotide probe that contains reporter and quenching molecules (Holland et al., 1991; Heid et al., 1996). In the TaqMAMA approach, an oligonucleotide (probe) is designed to hybridize to the parental DNA in the path of *Taq* polymerization prior to each PCR cycle. During its primer extension, *Taq* DNA polymerase "clears the tracks" by 5'-3' exonuclease activity and liberates the fluorescent chromophore now separated from the quenching moiety. Thus fluorescent signal is proportional to PCR procedure. A great advantage of the TaqMAMA approach is that there is no need for post-PCR separation of primers and products on the basis of sizes, which prevents potential PCR product carry-over contamination. In addition, because several dyes with different emission wavelength are available for the assay: FAM (512nm), SYBR (520nm), TET (538nm), VIC (552nm), and JOE (554nm), TaqMAMA permits simultaneous assays of multiple targets by using mutation-specific probes labeled with each dye. This real-time PCR will dramatically shorten the detection time compared to conventional systems.

I have applied the TK6 reconstruction experiments for TP53 bp746 and bp747, both G:C->T:A mutations, for TaqMAMA and have successfully achieved the sensitivity of 10^{-5} (Figure 29). My current result is 100-fold more sensitive than previously reported (Glaab and Skopek, 1999). This preliminary result is highly promising the validity of TaqMAMA as a multiplexed high throughput system that enables analysis of rare genetic events in human tissues. We are proposing to combine TaqMAMA technique to the micro-cuvettes robotic system described in section 7.2.

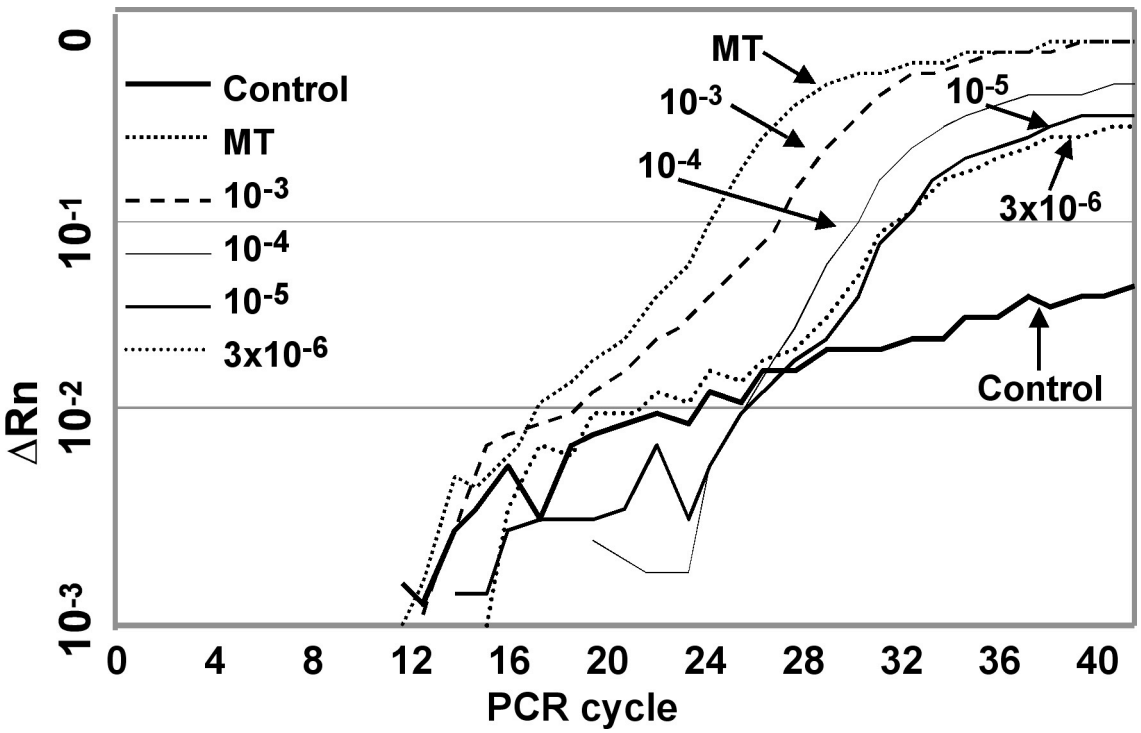


Figure 29. TaqMAMA TK6 reconstruction experiments on TP53 bp746 G:C->T:A transversion mutation. The series of MFs: 10^{-3}, 10^{-4}, 10^{-5}, $3x10^{-6}$ and 0, were artificially created from TK6 and human tumor cell line (Hs700T). $3x10^{6}$ copies of alleles were added as templates to each sample, including 3,000, 300, 30, 9, 0 copies of the mutant alleles, respectively. "MT" indicates mutant, or positive control, and "control" indicates no template negative control. ΔRn indicates the magnitude of the signal generated by the PCR. The amplification curves were clearly distinguishable between MT, 10^{-3}, 10^{-4} and 10^{-5}. The curve for $3x10^{-6}$ overlapped with that for 10^{-5}, showing the sensitivity did not reach to that level of $3x10^{-6}$. Similar results were obtained for TP53 bp747 G:C->T:A mutation as well (data not shown).

7.4. Fluorescent Anisotropy (FA)

In our first proposed embodiment, we envisaged adaptation of the TaqMAMA approach to the micro-cuvette platform. Although the method has proven its repeatability and accuracy, the challenges in designing appropriate probes and the cost of these reagents are limiting for high throughput application. In collaboration with the MIT BioInstrumentation Laboratory, we are developing an alternative real-time PCR detection technique based on fluorescent anisotropy (FA) that relied only on fluorescently labeled forward primers. This technique enables to distinguish molecules with different sizes without separation of probes, primers and products in real-time PCR. Its mechanism is based on the differential rotation angles of fluorophores, which is proportional to the size of the substance attached when excited by a plane-polarized light (Tsuruoka et al., 1996; Ye et al., 1998). The microarray optical scanning system already constructed was readily adapted to FA measurements. B. Crane in the Hunter laboratory has worked on the construction of an apparatus testing the specifications in which very low noise FA measurements have been achieved throughout the progress of PCR procedure.

Our ongoing project is the incorporation of cell digestion, thermal cycling and FA into a single machine to fully demonstrate the direct quantitative measurement of PCR amplification in a real-time manner. Our preliminary results showed that FA could successfully measure the real-time PCR amplification of genomic DNA extracted from cells as accurately as CE, and even of DNA not extracted from digested cells with a somewhat increased background noise (Figure 30). FA appeared to be prone to crowded impurities and cell debris such as proteins and fats that might overestimate the signals. We are trying to overcome this problem by setting the assay size very small (~200 cells/cuvette) so the cells are less concentrated in a cuvette, which is the same tactic as previously used and indeed improved the measurement in CE.

These promising systems will permit studies of a large number of samples from variety of organs.

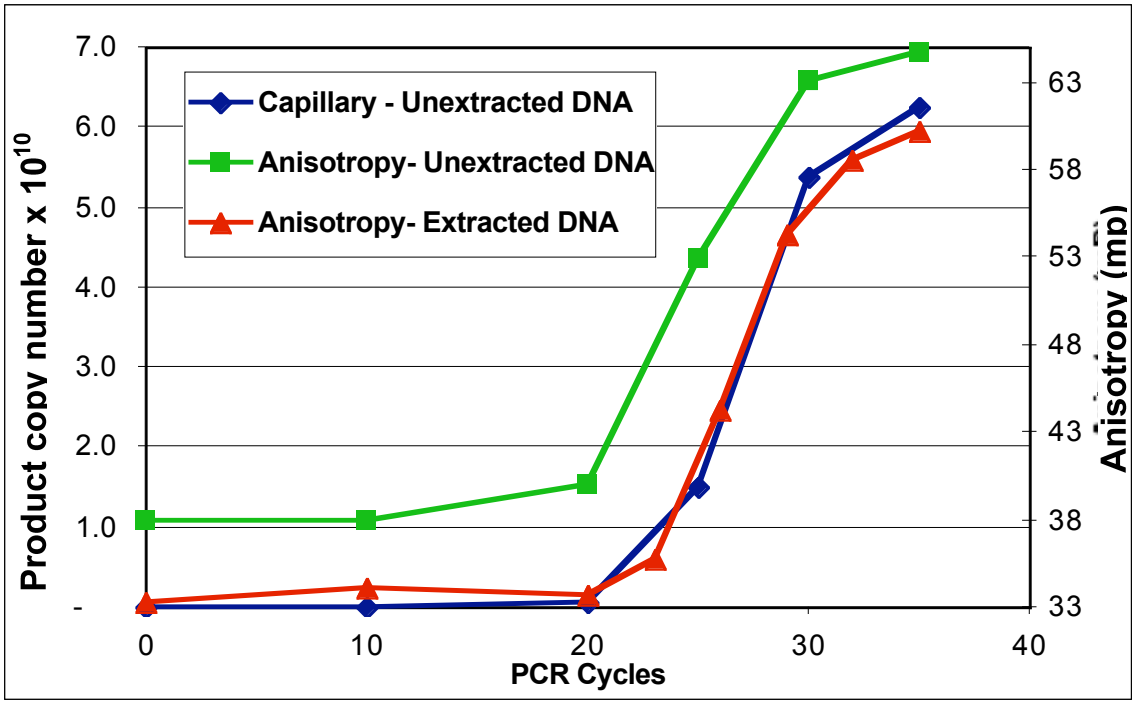


Figure 30. Real-time PCR detected by Fluorescent Anisotropy (FA) and CE. Two kinds of PCR templates: genomic TK6 DNA extracted from cells and digested cell (DNA not extracted) were used. FA (triangles) could measure DNA amplification as accurately as CE when DNA templates were extracted from cell previous to the PCR. When the template DNA was not extracted from cells, FA was still capable of measuring the amplification but with somewhat higher background noise, whereas CE showed no difference in measurements of extracted (not shown) and non-extracted DNA (diamonds).

8. REFERENCES

Agudo, A., Ahrens, W., Benhamou, E., Benhamou, S., Boffetta, P., Darby, S.C., Forastiere, F., Fortes, C., Gaborieau, V., Gonzalez, C.A., Jockel, K.H., Kreuzer, M., Merletti, F., Pohlabeln, H., Richiardi, L., Whitley, E., Wichmann, H.E., Zambon, P. and Simonato, L. (2000). Lung cancer and cigarette smoking in women: a multicenter case-control study in Europe. Int J Cancer. 88(5):820-7.

Ahrendt, S.A., Decker, P.A., Alawi, E.A., Zhu, Y.R., Sanchez-Cespedes, M., Yang, S.C., Haasler, G.B., Kajdacsy-Balla, A., Demeure, M.J. and Sidransky, D. (2001). Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. Cancer. 92(6):1525-30.

Albertini, R.J., Castle, K.L. and Borcherding, W.R. (1982). T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood. Proc Natl Acad Sci USA. 79(21):6617-21.

Ambrosone, C.B., Rao, U., Michalek, A.M., Cummings, K.M. and Mettlin, C.J. (1993). Lung cancer histologic types and family history of cancer. Analysis of histologic subtypes of 872 patients with primary lung cancer. Cancer. 72(4):1192-8.

Armitage, A.K. and Turner, D.M. (1970). Absorption of nicotine in cigarette and cigar smoke through the oral mucosa. Nature. 226(252):1231-2.

Armitage, P. and Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer. *8*:1-12.

Auerbach, O., Forman, J.B., Gere, J.B., Kassouny, D.Y., Muehsam, G.E., Petrick, T.G., Smolin, H.J. and Stout, A.P. (1957). Changes in the bronchial epithelium in relation to smoking and cancer of the lung; a report of progress. N Engl J Med. 256(3):97-104.

Auerbach, O., Stout, A.P., Hammond, E.C., Garfinkel, L.(1961). Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. N Engl J Med. 265:253-67.

Barbone, F., Bovenzi, M., Cavallieri, F. and Stanta, G. (1997). Cigarette smoking and histologic type of lung cancer in men. Chest. 112(6):1474-9.

Barsky, S.H., Roth, M.D., Kleerup, E.C., Simmons, M. and Tashkin, D.P. (1998). Histopathologic and molecular alterations in bronchial epithelium in habitual smokers of marijuana, cocaine, and/or tobacco. J Natl Cancer Inst. 90(16):1198-205.

BC Cancer Agency. 2003/02/17: Lung cancer prevention study screens ex-smokers for disease.
http://www.bccancer.bc.ca/ABCCA/NewsCentre/2003/LungCancerStudyScreensExSmokers.htm

Belinsky SA, Swafford DS, Finch GL, Mitchell CE, Kelly G, Hahn FF, Anderson MW, Nikula KJ. (1997). Alterations in the K-ras and p53 genes in rat lung tumors. Environ Health Perspect. 105 Suppl 4:901-6.

Bennett, W.P., Hussain, S.P., Vahakangas,K.H., Khan, M.A., Shields, P.G. and Harris, C.C. (1999). Molecular epidemiology of human cancer risk: gene-environment interactions and p53 mutation spectrum in human lung cancer. J Pathol *187*: 8-17

Berdal, K.G., Johansen, R.F. and Seeberg, E. (1998). Release of normal bases from intact DNA by a native DNA repair enzyme. EMBO J. 17(2):363-7.

Berrozpe, G., Schaeffer, J., Peinado, M.A., Real, F.X. and Perucho, M. (1994). Comparative analysis of mutations in the p53 and K-ras genes in pancreatic cancer. Int J Cancer. 58(2):185-91.

Bertram, J.F. and Rogers, A.W. (1981). Recovery of bronchial epithelium on stopping smoking. Br Med J (Clin Res Ed), 283(6306),1567-9.

Besarati Nia, A., Van Straaten, H.W., Kleinjans, J.C. and Van Schooten, F.J. (2000). Immunoperoxidase detection of 4-aminobiphenyl- and polycyclic aromatic hydrocarbons-DNA adducts in induced sputum of smokers and non-smokers. Mutat Res. 468(2):125-35.

Boguski, M.S. and McCormick, F. (1993). Proteins regulating Ras and its relatives. Nature 366:643–54.

Braakhuis, B.J., Tabor, M.P., Kummer, J.A., Leemans, C.R. and Brakenhoff, R.H. (2003). A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. Cancer Res. 63(8):1727-30.

Branda, R.F., Sullivan, L.M., O'Neill, J.P., Falta, M.T., Nicklas, J.A., Hirsch, B., Vacek, P.M. and Albertini, R.J. (1993). Measurement of HPRT mutant frequencies in T-lymphocytes from healthy human populations. Mutat Res. 285(2):267-79.

Branum, M.E., Reardon, J.T. and Sancar, A. (2001). DNA repair excision nuclease attacks undamaged DNA. A potential source of spontaneous mutations. J Biol Chem. 276(27):25421-6.

Brash, D.E., Rudolph, J.A., Simon, J.A., Lin, A., McKenna, G.J., Baden, H.P., Halperin, A.J. and Ponten, J. (1991). A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. Proc Natl Acad Sci U S A. 88(22):10124-8.

Braun, M.M., Caporaso, N.E., Page, W.F. and Hoover, R.N. (1994). Genetic component of lung cancer: cohort study of twins. Lancet. 344(8920):440-3.

Bressac, B., Kew, M., Wands, J. and Ozturk, M. (1991). Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature. 350(6317):429-31.

Brownson, R.C., Alavanja, M.C., Caporaso, N., Berger, E. and Chang, J.C. (1997). Family history of cancer and risk of lung cancer in lifetime non-smokers and long-term ex-smokers. Int J Epidemiol. 26(2):256-63.

Burkhart-Schultz, K.J., Thompson, C.L. and Jones, I.M. (1996). Spectrum of somatic mutation at the hypoxanthine phosphoribosyltransferase (hprt) gene of healthy people. Carcinogenesis. **17**(9):1871-83.

Butkiewicz, D., Cole, K.J., Phillips, D.H., Harris, C.C. and Chorazy, M. (1999). GSTM1, GSTP1, CYP1A1 and CYP2D6 polymorphisms in lung cancer patients from an environmentally polluted region of Poland: correlation with lung DNA adduct levels. Eur J Cancer Prev. 8(4):315-23.

Butler, J.D. and Crossley, P. (1979). An appraisal of relative airborne sub-urban concentrations of polycyclic aromatic hydrocarbons monitored indoors and outdoors. Sci Total Environ. 11(1):53-8.

Cairns, J. (1975). Mutation selection and the natural history of cancer. Nature. 255(5505):197-200.

Cha, R.S. and Thilly, W.G. (1993). Specificity, efficiency, and fidelity of PCR. PCR Methods Appl. 3(3):S18-29.

Cha, R.S., Thilly, W.G. and Zarbl, H. (1994). *N*-Nitroso-*N*-methylurea-induced rat mammary tumors arise from cells with preexisting oncogenic *Hras1* gene mutations. Proc. Natl. Acad. Sci. USA *91*:3794-53.

Cha, R.S., Zarbl, H., Keohavong, P. and Thilly, W.G. (1992). Mismatch amplification mutation assay (MAMA): Application to the c-H-ras gene. PCR Methods Appl. *2(1)*: 14-20.

The Chemical Heritage Foundation (2001). Magic Bullets: Chemistry vs. Cancer, Chemical Heritage Foundation. http://www.chemheritage.org/EducationalServices/pharm/chemo/readings/road.htm

Chepiga, T.A., Morton, M.J., Murphy, P.A., Avalos, J.T., Bombick, B.R., Doolittle, D.J., Borgerding, M.F. and Swauger, J.E. (2000). A comparison of the mainstream smoke chemistry and mutagenicity of a representative sample of the US cigarette market with two Kentucky reference cigarettes (K1R4F and K1R5F). Food Chem Toxicol. 38(10):949-62.

Chen, J. and Thilly, W.G. (1996). Mutational spectra vary with exposure conditions: benzo[a]pyrene in human cells. Mutat Res. 357(1-2):209-17.

Chen, J. and Viola, M.V. (1991). A method to detect ras point mutations in small subpopulations of cells. Anal Biochem. 195(1):51-6.

Cheng, T.J., Christiani, D.C., Liber, H.L., Wain, J.C., Xu, X., Wiencke, J.K. and Kelsey, K.T. (1995). Mutant frequency at the hprt locus in human lymphocytes in a case-control study of lung cancer. Mutation Res 332: 109-18.

Chen, R.H., Maher, V.M., Brouwer, J., van de Putte, P. and McCormick, J.J. (1992). Preferential repair and strand-specific repair of benzo[a]pyrene diol epoxide adducts in the HPRT gene of diploid human fibroblasts. Proc Natl Acad Sci USA. 89(12):5413-7.

Cheng, Y.W., Chen, C.Y., Lin, P., Huang, K.H., Lin, T.S., Wu, M.H. and Lee, H. (2000). DNA adduct level in lung tissue may act as a risk biomarker of lung cancer. Eur J Cancer. 36(11):1381-8.

Cline, J., Braman, J.C. and Hogrefe, H.H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. 24(18):3546-51.

Cohen, A.S., Najarian, D.R., Paulus, A., Guttman, A., Smith, J.A. and Karger, B.L. (1988). Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. Proc Natl Acad Sci USA. 85(24):9660-3.

Cole, J. and Skopek, T.R. (1994). Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. Mutation Res *304*: 33-105.

Cole, J., Green, M.H., James, S.E., Henderson, L. and Cole, H. (1988). A further assessment of factors influencing measurements of thioguanine-resistant mutant frequency in circulating T-lymphocytes. Mutat Res. 204(3):493-507.

Coller, A.H., Khrapko, K., Torres, A., Frampton, M.W., Utell, M.J. and Thilly, W.G. (1998). Mutational spectra of a 100-base pair mitochondrial DNA target sequence in bronchial epithelial cells: a comparison of smoking and nonsmoking twins. Cancer Res 58: 1268-77.

Cooper, C.S. (2002). Smoking, lung cancers and their TP53 mutations. Mutagenesis. 17(4):279-80.

Culp, S.J., Gaylor, D.W., Sheldon, W.G., Goldstein, L.S. and Beland, F.A. (1998). A comparison of the tumors induced by coal tar and benzo[a]pyrene in a 2-year bioassay. Carcinogenesis. 19:117-24.

Curry, J., Karnaoukhova, L., Guenette, G.C. and Glickman, B.W. (1999). Influence of sex, smoking and age on human hprt mutation frequencies and spectra. Genetics. 152(3):1065-77.

Davies, D.F. (1960). A review of the evidence on the relationship between smoking and lung cancer. J. Chronic Diseases, 11: 579-614.

Davies, M.J., Lovell, D.P. and Anderson, D. (1992). Thioguanine-resistant mutant frequency in T-lymphocytes from a healthy human population. Mutat Res. 265(2):165-71.

Delahunty, C., Ankener, W., Deng, Q., Eng, J. and Nickerson, D.A. (1996). Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay. Am J Hum Genet. 58(6):1239-46.

Denissenko, M.F., Pao, A., Pfeifer, G.P. and Tang, M. (1998). Slow repair of bulky DNA adducts along the nontranscribed strand of the human p53 gene may explain the strand bias of transversion mutations in cancers. Oncogene. 16(10):1241-7.

Denissenko, M.F., Pao, A., Tang, M. and Pfeifer, G.P. (1996). Preferential Formation of Benzo[a]pyrene adducts at lung cancer mutational hotspots in p53. Science *274*(5286):430.

DeOme, K.B., Miyamoto, M.J., Osborn, R.C., Guzman, R.C. and Lum, K. (1978). Detection of inapparent nodule-transformed cells in the mammary gland tissues of virgin female BALB/cfC3H mice. Cancer Res. 38(7):2103-11.

Doll, R. and Hill, A.B. (1952). A study of the aetiology of carcinoma of the lung. Br Med J 2:1471-86.

Doll, R. and Peto, R. (1978). Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health. **32**(4):303-13.

Dunn, B.P., Vedal, S., San, R.H., Kwan, W.F., Nelems, B., Enarson, D.A. and Stich, H.F. (1991). DNA adducts in bronchial biopsies. Int J Cancer. 48(4):485-92.

Eisenstadt, E., Warren, A.J., Porter, J., Atkins, D. and Miller, J.H. (1982). Carcinogenic epoxides of benzo[a]pyrene and cyclopenta[cd]pyrene induce base substitutions via specific transversions. Proc Natl Acad Sci USA. *79*(6):1945-9.

Enstrom, J.E. and Heath, C.W. Jr. (1999) Smoking cessation and mortality trends among 118,000 Californians, 1960-1997. Epidemiology. **10**(5):500-12.

Fearon. E.R. (1997). Human cancer syndromes: clues to the origin and nature of cancer. Science. 278(5340):1043-50.

Felley-Bosco, E., Pourzand, C., Zijlstra, J., Amstad, P. and Cerutti, P. (1991). A genotypic mutation system measuring mutations in restriction recognition sequences. Nucleic Acids Res. 19(11):2913-9.

Feng, Z., Hu, W., Amin, S. and Tang, M.S. (2003). Mutational spectrum and genotoxicity of the major lipid peroxidation product, trans-4-hydroxy-2-nonenal, induced DNA

adducts in nucleotide excision repair-proficient and -deficient human cells. Biochemistry. 42(25):7848-54.

Fialkow, P.J. (1976). Clonal origin of human tumors. Biochim Biophys Acta. *458*(3):283-321.

Finette, B.A., Sullivan, L.M., O'Neill, J.P., Nicklas, J.A., Vacek, P.M. and Albertini, R.J. (1994). Determination of hprt mutant frequencies in T-lymphocytes from a healthy pediatric population: statistical comparison between newborn, children and adult mutant frequencies, cloning efficiency and age. Mutat Res. 308(2):223-31.

Fiore, M.C. (1992). Trends in cigarette smoking in the United States: The epidemiology of tobacco use. Med Clin North Am. 76(2):289-303.

Fischer, S.G. and Lerman, L.S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. Proc Natl Acad Sci USA. *80(6):* 1579-83.

Friend, S.H., Horowitz, J.M., Gerber, M.R., Wang, X.F., Bogenmann, E., Li, F.P. and Weinberg, R.A. (1987). Deletions of a DNA sequence in retinoblastomas and mesenchymal tumors: organization of the sequence and its encoded protein. Proc Natl Acad Sci USA. 84(24):9059-63.

Gailani, M.R., Stahle-Backdahl, M., Leffell, D.J., Glynn, M., Zaphiropoulos, P.G., Pressman, C., Unden, A.B., Dean, M., Brash, D.E., Bale, A.E. and Toftgard, R. (1996). The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. Nat Genet. 14(1):78-81.

Gallagher, J., Mumford, J., Li, X., Shank, T., Manchester, D. and Lewtas, J. (1993). DNA adduct profiles and levels in placenta, blood and lung in relation to cigarette smoking and smoky coal emissions. IARC Sci Publ. (124):283-92.

Gao, H.G., Chen, J.K., Stewart, J., Song, B., Rayappa, C., Whong, W.Z. and Ong, T. (1997). Distribution of p53 and K-ras mutations in human lung cancer tissues. Carcinogenesis. 18(3):473-8.

Gao, W.M., Mady, H.H., Yu, G.Y., Siegfried, J.M., Luketich, J.D., Melhem, M.F. and Keohavong, P. (2003). Comparison of p53 mutations between adenocarcinoma and squamous cell carcinoma of the lung: unique spectra involving G to A transitions and G to T transversions in both histologic types. Lung Cancer. 40(2):141-50.

Garcia, S.B., Park, H.S., Novelli, M. and Wright, N.A. (1999). Field cancerization, clonality, and epithelial stem cells: the spread of mutated clones in epithelial sheets. J Pathol *187:* 61-81.

Gealy, R., Zhang, L., Siegfried, J.M., Luketich, J.D. and Keohavong, P. (1999). Comparison of mutations in the p53 and K-ras genes in lung carcinomas from smoking and nonsmoking women. Cancer Epidemiol Biomarkers Prev. 8(4 Pt 1):297-302.

Geradts, J., Fong, K.M., Zimmerman, P.V. and Minna, J.D. (2000). Loss of Fhit expression in non-small-cell lung cancer: correlation with molecular genetic abnormalities and clinicopathological features. Br J Cancer. 82(6):1191-7.

Glaab, W.E. and Skopek, T.R. (1999). A novel assay for allelic discriminatiuon that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan®) and mismatch amplification mutation assay. Mutation Res *430*: 1-12.

Gnarra, J.R., Tory, K., Weng, Y., Schmidt, L., Wei, M.H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F.M., et al. (1994). Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat Genet. 7(1):85-90.

Godschalk, R.W., Maas, L.M., Van Zandwijk, N., van 't Veer, L.J., Breedijk, A., Borm, P.J., Verhaert, J., Kleinjans, J.C. and van Schooten, F.J. (1998). Differences in aromatic-DNA adduct levels between alveolar macrophages and subpopulations of white blood cells from smokers. Carcinogenesis. 19(5):819-25.

Grist, S.A., McCarron, M., Kutlaca, A., Turner, D.R. and Morley, A.A. (1992). In vivo human somatic mutation: frequency and spectrum with age.
Mutation Res 266: 189-96.

Hackman, P., Hou, S.M., Nyberg, F., Pershagen, G. and Lambert, B. (2000). Mutational spectra at the HPRT locus in T-lymphocytes of nonsmoking and smoking lung cancer patients. Mutation Res *468:* 45-61.

Hainaut, P., Olivier, M., and Pfeifer, G.P. (2001). TP53 mutation spectrum in lung cancers and mutagenic signature of components of tobacco smoke: lessons from the IARC TP53 mutation database. Mutagenesis. 16(6):551-3; author reply 555-6.

Hainaut P. and Pfeifer, G.P. (2001). Patterns of P53 G->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis. 22(3), 367-74.

Harris JE. (1983). Cigarette smoking among successive birth cohorts of men and women in the united States during 1900-80. JNCI. 71:473-9.

Hattemer-Frey ,H.A. and Travis, C.C. (1991). Benzo-a-pyrene: environmental partitioning and human exposure. Toxicol Ind Health. 7(3):141-57.

Hecht, S.S. (1998). Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. Chem Res Toxicol. 11:559-603.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996). Real time quantitative PCR. Genome Res *10*: 986-94.

Hernandez-Boussard, T.M. and Hainaut, P. (1998). A specific spectrum of p53 mutations in lung cancer from smokers: review of mutations compiled in the IARC p53 database. Environ Health Perspect. *106*(7):385-91.

Herrero-Jimenez, P. (2001) Determination of the historical changes in primary and secondary risk factors for cancer using U.S. public health records. Thesis. Massachusetts Institute of Technology.

Herrero-Jimenez, P., Tomita-Mitchell, A., Furth, E.E., Morgenthaler, S. and Thilly, W.G. (2000). Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States. Mutat Res. 447(1):73-116.

Herrero-Jimenez, P. (2001). Mortality Data. http://epidemiology.mit.edu

Hetch, S.S. (1999). Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst *91*(14), 1194-210.

Hill, J. (1761). Cautions Against the Immoderate Use of Snuff and the Effects It Must Produce When This Way Taken into the Body. London: R. Baldwin & J. Jackson. [As cited in Redmond DE. (1970). Tobacco and cancer: the first clinical report, 1761. New Eng J Med. 282:21].

Hoffmann, D. and Hoffmann, I. (1997). The changing cigarette, 1950-1995. J Toxicol Environ Health. 50(4):307-64.

Hoffmann, D., Hoffmann, I. and El-Bayoumy, K. (2001). The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. Chem Res Toxicol. 14(7):767-90.

Hoffmann, D., Rivenson, A., Murphy, S.E., Chung, F.-L., Amin, S., Hecht, S.S. (1993). Cigarette smoking and adenocarcinoma of the lung: the relevance of nicotine-derived N-nitrosamines. J. Smoking Related Disorders, 4: 165-189.

Holland, P.M., van Loon, A.M., van der Avoort, H.G., Reimerink, J.H., Ras, A., Bestebroer, T.M., Drebot, M.A., Kew, O.M. and Koopmans, M.P. (1991). Detection of specific PCR product by utilizing the 5'->3' exonuclease activity of Thermus aquaticus DNA polymerase. PNAS USA *88*, 7276-80.

Hou, S.M., Yang, K., Nyberg, F., Hemminki, K., Pershagen, G. and Lambert, B. (1999). Hprt mutant frequency and aromatic DNA adduct level in non-smoking and smoking lung cancer patients and population controls. Carcinogenesis. 20(3):437-44.

Howard, R.B., Mullen, J.B., Pagura, M.E. and Johnston, M.R. (1999). Characterization of a highly metastatic, orthotopic lung cancer model in the nude rat. Clin Exp Metastasis. 17(2):157-62.

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J. and Harris, C.C. (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. Nature. 350(6317):427-8.

The Human Gene Mutation Database. Institute of Medical Genetics in Cardiff (2003). http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html

Hung, J., Kishimoto,Y., Sugio, K., Virmani, A., McIntire, D.D., Minna, J.D. and Gazdar, A.F. (1995). Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma. JAMA. 273(7):558-63.

Husgafvel-Pursiainen, K. and Kannio, A. (1996). Cigarette smoking and p53 mutations in lung cancer and bladder cancer. Environ Health Perspect. 104 Suppl 3:553-6.

Hussain, S.P., Aguilar, F. and Cerutti, P. (1994). Mutagenesis of codon 248 of the human p53 tumor suppressor gene by N-ethyl-N-nitrosourea. Oncogene. 9(1):13-8.

Hussain, S.P., Amstad, P., Raja, K., Sawyer, M., Hofseth, L., Shields, P.G., Hewer, A., Phillips, D.H., Ryberg, D., Haugen, A. and Harris, C.C. (2001a). Mutability of p53 hotspot codons to benzo(a)pyrene diol epoxide (BPDE) and the frequency of p53 mutations in nontumorous human lung. Cancer Res. *61*(17):6350-5.

Hussain, S.P., Kennedy, C.H., Amstad, P., Lui, H., Lechner, J.F. and Harris, C.C. (1997). Radon and lung carcinogenesis: mutability of p53 codons 249 and 250 to 238Pu alpha-particles in human bronchial epithelial cells. Carcinogenesis. 18(1):121-5.

Hussain, S.P., Raja, K., Amstad, P.A., Sawyer, M., Trudel, L.J., Wogan, G.N., Hofseth, L.J., Shields, P.G., Billiar, T.R., Trautwein, C., Hohler, T., Galle, P.R., Phillips, D.H., Markin, R., Marrogi, A.J. and Harris, C.C. (2000). Increased p53 mutation load in nontumorous human liver of wilson disease and hemochromatosis: oxyradical overload diseases. Proc Natl Acad Sci USA. 97(23):12770-5.

Huttner, E., Holzapfel, B. and Kropf, S. (1995). Frequency of HPRT mutant lymphocytes in a human control population as determined by the T-cell cloning procedure. Mutat Res. 1995 Oct;348(2):83-91.

Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A. (1988). DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc Natl Acad Sci U S A. 85(24):9436-40.

International Agency for Research on Cancer (IARC) Monographs. (1986). Chemistry and analysis of tobacco smoke. Vol. 38, 83-126.

International Agency for Research on Cancer (IARC) Monographs. (2000) The evaluation of carcinogenic risks of chemicals to humans, Vol. 1-77.

International Agency for Research on Cancer (IARC). p53 lung cancer database. Last up-dated September 23, 2002. http://www.iarc.fr/p53/index.html

Izzotti, A., Rossi, G.A., Bagnasco, M. and De Flora, S. (1991). Benzo[a]pyrene diolepoxide-DNA adducts in alveolar macrophages of smokers. Carcinogenesis. 12(7):1281-5.

Jacobson, D.R. and Mills, N.E. (1994). A highly sensitive assay for mutant ras genes and its application to the study of presentation and relapse genotypes in acute leukemia. Oncogene. 9(2):553-63.

Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E. and Thun, M.J. (2003). Cancer statistics, 2003. CA Cancer J Clin. 53(1):5-26.

Jin Z., Houle B., Mikheev, A.M., Cha, R.S. and Zarbl H. (1996). Alteration in *H-ras1* promoter conformation during *N*-Nitroso-*N*-methylurea-induced mammary carcinogenesis and pregnancy. Cancer Res *56*: 4927-35.

Jinneman, K.C. and Hill, W.E. (2001). *Listeria monocytogenes* lineage group classification by MAMA-PCR of the listeriolysin gene. Current Microbio *43*: 129-33.

Jonason, A.S., Kunala, S., Price, G.J., Restifo, R.J., Spinelli, H.M., Persing, J.A., Leffell, D.J., Tarone, R.E. and Brash, D.E. (1996). Frequent clones of p53-mutated keratinocytes in normal human skin. Proc Natl Acad Sci U S A. 93(24):14025-9.

Jones, I.M., Moore, D.H., Thomas, C.B., Thompson, C.L., Strout, C.L. and Burkhart-Schultz, K. (1993). Factors affecting HPRT mutant frequency in T-lymphocytes of smokers and nonsmokers. Cancer Epidemiol Biomarkers Prev. 2(3):249-60.

Jones, P.A. and Laird, P.W. (1999). Cancer epigenetics comes of age. Nat Genet. 21(2):163-7.

Kamb, A., Shattuck-Eidens, D., Eeles, R., Liu, Q., Gruis, N.A., Ding, W., Hussey, C., Tran, T., Miki, Y., Weaver-Feldhaus, J., et al. (1994). Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat Genet. 8(1):23-6.

Kan, Y.M. and Dozy, A.M. (1978). Antenatal diagnosis of sickle-cell anemia by DNA analysis of amniotic fluid cells. Lancet, *2*(8096):910-2.

Kanigan, T., Brenan, C., Lafontaine, S., Soswoski, L., Madden, P. and Hunter, I. (2000). Living Chips for drug discovery, SPIE Proceedings. 3926:172-180.

Keohavong, P. and Thilly, W.G. (1989). Fidelity of DNA polymerases in DNA

amplification. Proc Natl Acad Sci USA. 86(23):9253-7.

Kern, J.A., Slebos, R.J., Top, B., Rodenhuis, S., Lager, D., Robinson, R.A., Weiner, D. and Schwartz, D.A. (1994). C-erbB-2 expression and codon 12 K-ras mutations both predict shortened survival for patients with pulmonary adenocarcinomas. J Clin Invest *93*(2), 516-20.

Khrapko, K., Coller, H., Andre, P., Li, X.C., Foret, F., Belenky, A., Karger, B.L. and Thilly, W.G. (1997a). Mutational spectrometry without phenotypic selection: human mitochondrial DNA. Nucleic Acids Res. 25(4):685-93.

Khrapko, K., Coller, H.A., Andre, P.C., Li, XC., Henekamp, J.S. and Thilly, W.G. (1997b). Mitochondrial mutational spectra in human cells and tissues. Proc. Natl. Acad. Sci. USA *94*, 13798-803.

Khrapko, K., Hanekamp, J.S., Thilly, W.G., Belenkii, A., Foret, F. and Karger, B.L. (1994). Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. Nucleic Acids Res *22(*3), 364-9.

Kim, A.S. and Thilly, W.G. (2003). Ligation of high-melting-temperature 'clamp' sequence extends the scanning range of rare point-mutational analysis by constant denaturant capillary electrophoresis (CDCE) to most of the human genome. Nucleic Acids Res. 31(16):e97.

Kingdon, K.H. (1961). Possible biological effects of electrically charged particles in tobacco smoke. Nature. 189:180-2.

Klein-Szanto, A.J.P., Iizasa, T., Momiki, S., Garcia-Palazzo, I., Caamano, J., Metcalf, R., Welsh, J. and Harris, C.C. (1992). A tobacco-specific N-nitrosamine or cigarette smoke condensate (CSC) causes neoplastic transformation of xenotransplanted human bronchial epithelial cells. Proc. Natl. Acad. Sci. USA. 89, 6693-7

Knudson, A.G. (1971). Mutation and cancer: Statistical study of retinoblastoma. Proc Natl Acad Sci. 68(4):820-3.

Kohno, H., Hiroshima, K., Toyozaki, T., Fujisawa, T. and Ohwada, H. (1999). p53 mutation and allelic loss of chromosome 3p, 9p of preneoplastic lesions in patients with nonsmall cell lung carcinoma. Cancer. 85(2):341-7.

Kohno, T. and Yokota, J. (1999). How many tumor suppressor genes are involved in human lung carcinogenesis? Carcinogenesis. 20(8):1403-10.

Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev. 10(12):1433-42.

Kondo, K., Tsuzuki, H., Sasa, M., Sumitomo, M., Uyama, T. and Monden, Y. (1996). A dose-response relationship between the frequency of p53 mutations and tobacco consumption in lung cancer patients. J Surg Oncol. 61(1):20-6.

Kreuzer, M., Kreienbrock, L., Muller, K.M., Gerken, M. and Wichmann, E. (1999). Histologic types of lung carcinoma and age at onset. Cancer. 85(9):1958-65.

Kure, E.H., Ryberg, D., Hewer, A., Phillips, D.H., Skaug, V., Baera, R. and Haugen, A. (1996). p53 mutations in lung tumours: relationship to gender and lung DNA adduct levels. Carcinogenesis. 17(10):2201-5.

Lang, S.M., Stratakis, D.F., Freudling, A., Ebelt, K., Oduncu, F., Hautmann, H. and Huber, R.M. (2000). Detection of K-ras and p53 mutations in bronchoscopically obtained malignant and non-malignant tissue from patients with non-small cell lung cancer. Eur J Med Res. 5(8):341-6.

LaRue H, Allard P, Simoneau M, Normand C, Pfister C, Moore L, Meyer F, Tetu B, Fradet Y. (2000). P53 point mutations in initial superficial bladder cancer occur only in tumors from current or recent cigarette smokers. Carcinogenesis. 21(1):101-6.

Lerosey, I., Chardin, P., de Gunzburg, J. and Tavitian, A. (1991). The product of the rap2 gene, member of the ras superfamily. Biochemical characterization and site-directed mutagenesis. J Biol Chem. 266(7):4315-21.

Levy, D.B., Smith, K.J., Beazer-Barclay, Y., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. (1994). Inactivation of both APC alleles in human and mouse tumors. Cancer Res. 54(22):5953-8.

Levy, J., Virolainen, M. and Detendi, V. (1968). Human lymphoblast lines from lymph node and spleen. Cancer, **22**(3):517-24.

Li, Z.H., Zheng, J., Weiss, L.M., Shibata, D. (1994). c-k-ras and p53 mutations occur very early in adenocarcinoma of the lung. Am J Pathol. 144(2):303-9.

Lioy, P.L., Waldman, J.M., Greenberg, A., Harkov, R., and Pietarinen, C. (1988). The Total Human Environmental Exposure Study (THEES) to benzo(a)pyrene: comparison of the inhalation and food pathways. Arch Environ Health. 43(4):304-12.

Li-Sucholeiki, X.C. and Thilly, W. G. (2000). A sensitive scanning technology for low frequency nuclear point mutations in human genomic DNA. Nucleic Acids Res. 28(9):E44.

Liu, J. and Johnston, M.R. (2002). Animal models for studying lung cancer and evaluating novel intervention strategies. Surg Oncol. 11(4):217-27.

Macholda, F., Borek, Z. and Lhotka, J. (1970). Bronchogenic carcinoma. Anatomical and physiological conditions of its origin and evolution. Acta Univ Carol [Med] (Praha).

Suppl 41:1+.

Maestrelli, P., Saetta, M., Mapp, C.E. and Fabbri, L.M. (2001). Remodeling in response to infection and injury. Airway inflammation and hypersecretion of mucus in smoking subjects with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 164(10 Pt 2):S76-80.

Makitaro, R., Paakko, P., Huhti, E., Bloigu, R. and Kinnula, V.L. (1999). An epidemiological study of lung cancer: history and histological types in a general population in northern Finland. Eur Respir J. 13(2):436-40.

Mao, L., Lee, J.S., Kurie, J.M., Fan, Y.H., Lippman, S.M., Lee, J.J., Ro, J.Y., Broxson, A., Yu, R., Morice, R,C., Kemp, B.L., Khuri, F.R., Walsh, G.L., Hittelman, W.N. and Hong, W.K. (1997). Clonal genetic alterations in the lungs of current and former smokers. J Natl Cancer Inst. 89(12):857-62.

Marchetti, A., Pellegrini, S., Sozzi, G., Bertacca, G., Gaeta, P., Buttitta, F., Carnicelli, V., Griseri, P., Chella, A., Angeletti, C.A., Pierotti, M. and Bevilacqua, G. (1998). Genetic analysis of lung tumours of non-smoking subjects: p53 gene mutations are constantly associated with loss of heterozygosity at the FHIT locus. Br J Cancer. 78(1):73-8.

Martin, G.M., Ogburn, C.E., Colgin, L.M., Gown, A.M., Edland, S.D. and Monnat, R.J. Jr. (1996). Somatic mutations are frequent and increase with age in human kidney epithelial cells. Hum Mol Genet. 5(2):215-21.

Mayne, S.T., Buenconsejo, J. and Janerich, D.T. (1999). Familial cancer history and lung cancer risk in United States nonsmoking men and women. Cancer Epidemiol Biomarkers Prev. 8(12):1065-9.

McGrew, J.L. National Commission on Marihuana and Drug Abuse.
Schaffer Library of Drug Policy (2003).
http://www.druglibrary.org/schaffer/LIBRARY/studies/nc/nc2b.htm

Medina D, Shepherd F, Gropp T. (1978). Enhancement of the tumorigenicity of preneoplastic mammary nodule lines by enzymatic dissociation. J Natl Cancer Inst. 60(5):1121-6.

Mendelman, L.V., Petruska, J. and Goodman, M.F. (1990). Base mispair extension kinetics. Comparison of DNA polymerase alpha and reverse transcriptase. J Biol Chem. 265(4):2338-46.

Meng, L., Lin, L., Fresno, M., Morales, A.R. and Nadji, M. (1999). Frequency and pattern of p53 gene mutation in a cohort of Spanish women with node-negative breast cancer. Int J Oncol. 15(3):555-8.

Merlo, F., Bolognesi, C., Peluso, M., Valerio, F., Abbondandolo, A. and Puntoni, R. (1997). Airborne levels of polycyclic aromatic hydrocarbons: 32P-postlabeling DNA

adducts and micronuclei in white blood cells from traffic police workers and urban residents. J Environ Pathol Toxicol Oncol. 16(2-3):157-62.

Mills, N.E., Fishman, C.L., Rom ,W.N., Dubin, N. and Jacobson, D.R. (1995). Increased prevalence of K-ras oncogene mutations in lung adenocarcinoma. Cacner Res *55*, 1444-7.

Modrich, P. and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem.; 65:101-33.

Moolgavkar, S.H. and Knudson, A.G. Jr. (1981). Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst. 66(6):1037-52.

Morabia, A. and Wynder, E.L. (1991). Cigarette smoking and lung cancer cell types. Cacner. 68:2074-8.

Morley A.A. (1998). Somatic Mutation and Agiing. Ann N Y Acad Sci. 854, 20-2.

Morley, A.A., Cox, S. and Holliday, R. (1982). Human lymphocytes resistant to 6-thioguanine increase with age. Mech Ageing Dev. 19(1):21-6.

Morley, A.A. and Turner, D.R. (1999). The contribution of exogenous and endogenous mutagens to in vivo mutations. Mutation Res. 428:11-5.

U.S. Bureau of the Census, Mortality Statistics, (1900-1936), Special Reports, Washington Government Printing Office.

Muniappan, B.P. and Thilly, W.G. (1999). Application of constant denaturant capillary electrophoresis (CDCE) to mutation detection in humans. Genet Anal. 14(5-6):221-7.

Muniappan, B.P. and Thilly, W.G. (2002). The DNA polymerase beta replication error spectrum in the adenomatous polyposis coli gene contains human colon tumor mutational hotspots. Cancer Res. 62(11):3271-5.

Mustonen, R., Schoket, B. and Hemminki, K. (1993). Smoking-related DNA adducts: 32P-postlabeling analysis of 7-methylguanine in human bronchial and lymphocyte DNA. Carcinogenesis. 14(1):151-4.

Naylor, S.L., Johnson, B.E., Minna, J.D. and Sakaguchi, A.Y. (1987) Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. Nature. 329(6138):451-4.

Nelson, H.H., Wiencke, J.K., Gunn, L., Wain, J.C., Christiani, D.C. and Kelsey, K.T. (1998). Chromosome 3p14 alterations in lung cancer: evidence that FHIT exon deletion is a target of tobacco carcinogens and asbestos. Cancer Res. 58(9):1804-7.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. and Markham, A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 17(7):2503-16.

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., Hedge, P., Markham, A., Krush, A. J., Petersen, G., Hamilton, S. R., Nilbert, M. C., Levy, D. B., Bryan, T. M., Preisinger, A. C., Smith, K. J., Su, L.-K., Kinzler, K. W. and Vogelstein, B. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 253: 665-669.

Noori, P. and Hou, S.M. (2001). Mutational spectrum induced by acetaldehyde in the HPRT gene of human T lymphocytes resembles that in the p53 gene of esophageal cancers. Carcinogenesis. 22(11):1825-30.

Nordling, C.O. (1953). A new theory on the cancer-inducing mechanism. Br. J. Cancer. **7**:68-72,

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*(4260), 23-8.

Ochsner, M. and DeBakey, M. (1939). Symposium on cancer. Primary pulmonary pregnancy. Treatment by total pneumonectomy: analyses of 79 collected cases and presentation of 7 personal cases. Surg Gynecol Obstet. 68:435-51.

Ooi, W.L., Elston, R.C., Chen, V.W., Bailey-Wilson, J.E. and Rothschild, H. (1986). Increased familial risk for lung cancer. J Natl Cancer Inst. 76(2):217-22.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA. 86(8):2766-70.

Orita M., Suzuki Y., Sekiya, T. and Hayashi, K. (1989). Rapid and sensitive detection of point mutation and DNA polymorphisms using the polymerase chain reaction. Genomics *5(4),* 874-879.

Osann, K.E. (1991). Lung cancer in women: the importance of smoking, family history of cancer, and medical history of respiratory disease. Cancer Res. 51(18):4893-7.

O'Shaughnessy, T.C., Ansari, T.W., Barnes, N.C, and Jeffery, P.K. (1997). Inflammation in bronchial biopsies of subjects with chronic bronchitis: inverse relationship of CD8+ T lymphocytes with FEV1. Am J Respir Crit Care Med. 155(3):852-7.

Ouhtit, A., Ueda, M., Nakazawa, H., Ichihashi, M., Dumaz, N., Sarasin, A. and Yamasaki, H. (1997). Quantitative detection of ultraviolet-specific p53 mutations in normal skin from Japanese patients. Cancer Epidemiol Biomarkers Prev. 6(6):433-8.

Park, I.W., Wistuba, I.I., Maitra, A., Milchgrub, S., Virmani, A.K., Minna, J.D. and

Gazdar, A.F. (1999). Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. J Natl Cancer Inst. 91(21):1863-8.

Park, J.Y., Jeon, H.S., Park, S.H., Park, T.I., Son, J.W., Kim, C.H., Park, J.H., Kim, I.S., Jung, T.H. and Jun, S.H. (2000). Microsatellite alteration (MSA) in histologically normal lung tissue of patients with non-small cell lung cancer. Lung Cancer. 30(2):83-9.

Paschke, T. (2000). Analysis of different versions of the IARC p53 database with respect to G->T transversion mutation frequencies and mutation hotspots in lung cancer of smokers and non-smokers. Mutagenesis. 15(6), 457-8.

Persons, B.L. and Heflich, R.H. (1997). Genotypic selection methods for the direct analysis of point mutations. Mutat Res.;387(2):97-121.

Peral, R. (1938). Tobacco smoking and longevity. Science. 87(2253):216-7.

Petruska, J., Goodman, M.F., Boosalis, M.S., Sowers, L.C., Cheong, C. and Tinoco, I. Jr. (1988). Comparison between DNA melting thermodynamics and DNA polymerase fidelity. Proc Natl Acad Sci USA. 85(17):6252-6.

Phalen, R.F. and Oldham, M.J. (1983). Tracheobronchial airway structure as revealed by casting techniques. Am Rev Respir Dis. 128(2 Pt 2):S1-4.

Phillips, D.H., Hewer, A., Martin, C.N., Garner, R.C. and King, M.M. (1988). Correlation of DNA adduct levels in human lung with cigarette smoking. Nature. 336(6201):790-2.

Phillips, D.H., Schoket, B., Hewer, A., Bailey, E., Kostic, S. and Vincze, I. (1990). Influence of cigarette smoking on the levels of DNA adducts in human bronchial epithelium and white blood cells. Int J Cancer. 46(4):569-75.

Piipari, R., Savela, K., Nurminen, T., Hukkanen, J., Raunio, H., Hakkola, J., Mantyla, T., Beaune, P., Edwards, R.J., Boobis, A.R. and Anttila, S. (2000). Expression of CYP1A1, CYP1B1 and CYP3A, and polycyclic aromatic hydrocarbon-DNA adduct formation in bronchoalveolar macrophages of smokers and non-smokers. Int J Cancer. 86(5):610-6.

Plopper, C.G., Mariassy, A.T. and Lollini, L.O. (1983). Structure as revealed by airway dissection. A comparison of mammalian lungs. Am Rev Respir Dis. 128(2 Pt 2):S4-7.

Podlutsky, A., Hou, S.M., Nyberg, F., Pershagen, G. and Lambert B. (1999). Influence of smoking and donor age on the spectrum of in vivo mutation at the HPRT-locus in T lymphocytes of healthy adults. Mutat Res. 431(2):325-39.

Pott, P. (1775). Chirurgical Observations Relative to the Cataract, the Polypus of the Nose, the Cancer of the Scrotum, the Different Kinds of Ruptures, and the Mortification of the Toes and Feet. London.

Robinson, D.R., Goodall, K., Albertini, R.J., O'Neill, J.P., Finette, B., Sala-Trepat, M., Moustacchi, E., Tates, A.D., Beare, D.M., Green, M.H. and Cole, J. (1994). An analysis of in vivo hprt mutant frequency in circulating T-lymphocytes in the normal human population: a comparison of four datasets. Mutat Res. **313**(2-3):227-47.

Rodenhuis, S., van de Wetering, M.L., Mooi, W.J., Evers, S.G., van Zandwijk, N. and Bos, J.L. (1987). Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. N Engl J Med. 317(15):929-35.

Rodenhuis, S. and Slebos, R.J. (1992). Clinical significance of ras oncogene activation in human lung cancer. Cancer Res. 52(9 Suppl):2665-9.

Rodin, S.N. and Rodin, A.S. (2000). Human lung cancer and p53: The interplay between mutagenesis and selection. PNAS *97*(22), 12244-12249.

Rodin, S.N. and Rodin, A.S. (2002). On the origin of p53 G:C --> T:A transversions in lung cancers. Mutat Res. 508(1-2):1-19.

Roe, F.J.C., Salaman, M.H. and Cohen, J. (1959). Incomplete carcinogens in cigarette smoke condensate: tumour-promotion by a phenolic fraction. Br. J. Cancer, 13: 623-633.

Rouleau, G.A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B., et al. (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature. 363(6429):515-21.

Routledge, M.N., Garner, R.C., Jenkins, D. and Cuzick, J. (1992). 32P-postlabelling analysis of DNA from human tissues. Mutat Res. 282(3):139-45.

Rubin, H. (2001a). Synergistic mechanisms in carcinogenesis by polycyclic aromatic hydrocarbons and by tobacco smoke: a bio-historical perspective with updates. Carcinogenesis. 22(12):1903-30.

Rubin, H. (2001b). Selected cell and selective microenvironment in neoplastic development. Cancer Res. 61(3):799-807.

Rubin, H. (2002). Selective clonal expansion and microenvironmental permissiveness in tobacco carcinogenesis. Oncogene. 21(48):7392-411.

Ryberg, D., Hewer, A., Phillips, D.H. and Haugen, A. (1994a). Different susceptibility to smoking-induced DNA damage among male and female lung cancer patients. Cancer Res. 54(22):5801-3.

Ryberg, D., Kure, E., Lystad, S., Skaug, V., Stangeland, L., Mercy, I., Borresen, A.L. and Haugen, A. (1994b). p53 mutations in lung tumors: relationship to putative susceptibility markers for cancer. Cancer Res. 54(6):1551-5.

Saetta, M., Turato, G., Baraldo, S., Zanin, A., Braccioni, F., Mapp, C.E., Maestrelli, P., Cavallesco, G., Papi, A. and Fabbri, L.M. (2000). Goblet cell hyperplasia and epithelial inflammation in peripheral airways of smokers with both symptoms of chronic bronchitis and chronic airflow limitation. Am J Respir Crit Care Med. 161(3 Pt 1):1016-21.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn G.T., Mullis, K.B. and Erlich, H.A. (1988). Primer-directed enzymatic amplificationof DNA with a thermostable DNA plolymerase. Science *239*, 487-91.

Sandy, M.S., Chiocca, S.M. and Cerutti, P.A. (1992). Genotypic analysis of mutations in Taq I restriction recognition sites by restriction fragment length polymorphism/polymerase chain reaction. Proc Natl Acad Sci USA. 89(3):890-4.

Sasa, M., Kondo, K., Komaki, K., Uyama, T., Morimoto, T. and Monden, Y. (1993). Frequency of spontaneous p53 mutations (CpG site) in breast cancer in Japan. Breast Cancer Res Treat. 27(3):247-52.

Savela, K. and Hemminki, K. (1991). DNA adducts in lymphocytes and granulocytes of smokers and nonsmokers detected by the 32P-postlabelling assay. Carcinogenesis. 12(3):503-8.

Schoket, B., Kostic, S. and Vincze, I. (1993). Determination of smoking-related DNA adducts in lung-cancer and non-cancer patients. IARC Sci Publ. (124):315-9.

Seyama, T., Ito, T.,Hayashi, T., Mizuno, T., Nakamura, N. and Akiyama, M. (1992). A novel bloker-PCR method for detection of rare mutant alleles in the presence of an excess amount of normal DNA. Nucleic Acid Res *20*, 2493-6.

Shaw, G.L., Falk, R.T., Pickle, L.W., Mason, T.J. and Buffler, P.A. (1991). Lung cancer risk associated with cancer in relatives. J Clin Epidemiol. 44(4-5):429-37.

Sherley, J.L., Stadler, P.B. and Johnson, D.R. (1995). Expression of the wild-type p53 antioncogene induces guanine nucleotide-dependent stem cell division kinetics. Proc Natl Acad Sci USA. 92(1):136-40.

Slaughter, D.P., Southwick, H.W., Smejkal, W. (1953). Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 6(5):963-8.

Smith, C.J. and Hansch, C. (2000). The relative toxicity of compounds in mainstream cigarette smoke condensate. Food Chem Toxicol. 38(7):637-46

Smith, S.S. and Fiore, M.C. (1999). The epidemiology of tobacco use, dependence, and cessation in the Unites States. Prim care. 26(3):433-61.

Smith, R.A. and Glynn, T.J. (2000). Epidemiology of lung cancer. Radiol clin North Am. 38(3):453-70.

Sobue, T., Suzuki, T., Horai, T., Matsuda, M. and Fujimoto, I. (1988). Relationship between cigarette smoking and histologic type of lung cancer, with special reference to sex difference. Jpn J Clin Oncol. 18(1):3-13.

Sommer, S.S., Cassady, J.D., Sobell, J.L. and Bottema, C.D. (1989). A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. Mayo Clin Proc. 64(11):1361-72.

Soussii, T. (2002). The p53 web site at the Institut Curie. Laboratoire De genotoxicologie Des Tumeurs. Université P. & M. Curie / Institut Curie - Section Médicale. http://p53.curie.fr/

Sozzi, G., Miozzo, M., Pastorino, U., Pilotti, S., Donghi, R., Giarola, M., De Gregorio, L., Manenti, G., Radice, P., Minoletti, F., et al. (1995). Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. Cancer Res. 55(1):135-40.

Spurzem, J.R., Thompson, A.B., Daughton, D.M., Mueller, M., Linder, J. and Rennard, S.I. (1991). Chronic inflammation is associated with an increased proportion of goblet cells recovered by bronchial lavage. Chest. 100(2): 389-93.

Stayner, L.T. and Wegman, D.H. (1983). Smoking, occupation, and histopathology of lung cancer: a case-control study with the use of the Third National Cancer Survey. J Natl Cacner Inst 70, 421-6.

Stork, P., Loda, M., Bosari, S., Wiley, B., Poppenhusen, K. and Wolfe, H. (1991). Detection of K-ras mutations in pancreatic and hepatic neoplasms by non-isotopic mismatched polymerase chain reaction. Oncogene. 6(5):857-62.

Sugio, K., Kishimoto, Y., Virmani, A.K., Hung, J.Y. and Gazdar, A.F. (1994). K-ras mutations are a relatively late event in the pathogenesis of lung carcinomas. Cancer Res. 54(22):5811-5.

Szybalska, E.H. and Szybalski, W. (1962). Genetics of human cess line. IV. DNA-mediated heritable transformation of a biochemical trait. Proc Natl Acad Sci USA. 48:2026-34.

Takeshima, Y., Seyama, T., Bennett, W.P., Akiyama, M., Tokuoka, S., Inai, K., Mabuchi, K., Land, C.E. and Harris, C.C. (1993). p53 mutations in lung cancers from non-smoking atomic-bomb survivors. Lancet. 342(8886-8887):1520-1

Terry, L.L., et al. (1964). Smoking and Health: Report of the Advisory Committee of the Surgeon General of the Public Health Service. Centers for Disease Control and Prevention http://www.cdc.gov/tobacco/sgr/sgr_1964/sgr64.htm

Thun, M.J., Lally, C.A., Flannery, J.T., Calle, E.E., Flanders, W.D., Heath, C.W. Jr. (1997). Cigarette smoking and changes in the histopathology of lung cancer. J Natl Cancer Inst. 89(21):1580-6.

Thurlbeck, W.M., Malaka, D. and Murphy, K. (1975). Goblet cells in the peripheral airways in chronic bronchitis. Am Rev Respir Dis. 112(1):65-9.

Tomita-Mitchell, A., Ling, L.L., Glover, C.L., Goodluck-Griffith, J. and Thilly, W.G. (2003). The mutational spectrum of the HPRT gene from human T cells in vivo shares a significant concordant set of hot spots with MNNG-treated human cells. Cancer Res. 63(18):5793-8.

Tomlinson, I.P.M., Novelli, N.R. and Bodmer, W.F. (1996). The mutation rate and cancer.
PNAS. **93**:14800-3.

Toyooka, S., Maruyama, R., Toyooka, K.O., McLerran, D., Feng, Z., Fukuyama, Y., Virmani, A.K., Zochbauer-Muller, S., Tsukuda, K., Sugio, K., Shimizu, N., Shimizu, K., Lee, H., Chen, C.Y., Fong, K.M., Gilcrease, M., Roth, J.A., Minna, J.D. and Gazdar, A.F. (2003). Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer. 103(2):153-60.

Travis, W.D., Lubin, J., Ries, L. and Devesa, S. (1996). United States lung carcinoma incidence trends: declining for most histologic types among males, increasing among females. Cancer. 77(12):2464-70.

Trevisani, L., Sartori, S., Bovolenta, M.R., Mazzoni, M., Pazzi, P., Putinati, S. and Potena, A. (1992). Structural characterization of the bronchial epithelium of subjects with chronic bronchitis and in asymptomatic smokers. Respiration. 59(3):136-44.

Tseng, J.E., Kemp, B.L., Khuri, F.R., Kurie, J.M., Lee, J.S., Zhou, X., Liu, D., Hong, W.K. and Mao, L. (1999). Loss of Fhit is frequent in stage I non-small cell lung cancer and in the lungs of chronic smokers. Cancer Res. 59(19):4798-803.

Tsuruoka, M., Yano, K., Ikebukuro, K., Nakayama, H., Masuda, Y. and Karube, I. (1996). Rapid detection of complementary- and mismatched DNA sequences using fluorescence polarization. Analytical Letters *29*(10), 1741-9.

Vahakangas, K.H., Bennett, W.P., Castren, K., Welsh, J.A., Khan, M.A., Blomeke, B., Alavanja, M.C. and Harris, C.C. (2001). p53 and K-ras mutations in lung cancers from former and never-smoking women. Cancer Res. 61(11):4350-6.

Valencia, A., Chardin, P., Wittinghofer, A. and Sander, C. (1991). The ras protein family: evolutionary tree and role of conserved amino acids. Biochemistry 30:4637–48.

Van Schooten, F.J., Godschalk, R.W., Breedijk, A., Maas, L.M., Kriek, E., Sakai, H., Wigbout, G., Baas, P., Van't Veer, L. and Van Zandwijk, N. (1997). 32P-postlabelling of

aromatic DNA adducts in white blood cells and alveolar macrophages of smokers: saturation at high exposures. Mutat Res. 378(1-2):65-75.

U.S. Department of Health and human Services, Vital Statistics of the United States, (1937-1992), Vol.II -- Mortality Part A, U.S. Government Printing Office, Hyattsvile, MD.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y,, White, R,, Smits, A.M. and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. N Engl J Med *329*, 525-32.

Vrieling, H., Thijssen, J.C., Rossi, A.M., van Dam, F.J., Natarajan, A.T., Tates, A.D. and van Zeeland, A.A. (1992). Enhanced hprt mutant frequency but no significant difference in mutation spectrum between a smoking and a non-smoking human population. Carcinogenesis. **13**(9):1625-31.

Wang, D., Weghorst, C.M., Calvert, R.J. and Stoner, G.D. (1996). Mutation in the p53 tumor suppressor gene in rat esophageal papillomas induced by N-nitrosomethylbenzylamine. Carcinogenesis. 17(4):625-30.

Wang, X., Christiani, D.C., Mark, E.J., Nelson, H., Wiencke, J.K., Gunn, L., Wain, J.C. and Kelsey, K.T. (1999). Carcinogen exposure, p53 alteration, and K-ras mutation in synchronous multiple primary lung carcinoma. Cancer. 85(8):1734-9.

Wei, D., Maher, V.M. and McCormick, J.J. (1995). Site-specific rates of excision repair of benzo[a]pyrene diol epoxide adducts in the hypoxanthine phosphoribosyltransferase gene of human fibroblasts: correlation with mutation spectra. Proc Natl Acad Sci USA. 92(6):2204-8.

Weinstein, I.B., Jeffrey, A.M., Jennette, K.W., Blobstein, S.H., Harvey, R.G., Harris, C., Autrup, H., Kasai, H. and Nakanishi, K. (1976). Benzo(a)pyrene diol epoxides as intermediates in nucleic acid binding in vitro and in vivo. Science. 193(4253):592-5.

Westra, W.H., Slebos, R.J., Offerhaus, G.J., Goodman, S.N., Evers, S.G., Kensler, T.W., Askin, F.B., Rodenhuis, S. and Hruban, R.H. (1993). K-ras oncogene activation in lung adenocarcinomas from former smokers. Evidence that K-ras mutations are an early and irreversible event in the development of adenocarcinoma of the lung. Cancer. **72**(2):432-8.

Whang-Peng, J., Kao-Shan, C.S., Lee, E.C., Bunn, P.A., Carney, D.N., Gazdar, A.F. and Minna, J.D. (1982). Specific chromosome defect associated with human small-cell lung cancer; deletion 3p(14-23). Science. 215(4529):181-2.

Wilson, V.L., Wei, Q., Wade, K.R., Chisa, M., Bailey, D., Kanstrup, C.M., Yin, X., Jackson, C.M., Thompson, B. and Lee, W.R. (1999). Needle-in-a-haystack detection and identification of base substitution mutations in human tissues. Mutat Res. 406(2-4):79-100.

Wistuba, I.I., Lam, S., Behrens, C., Virmani, A.K., Fong, K.M., LeRiche, J., Samet, J.M., Srivastava, S., Minna, J.D. and Gazdar, A.F. (1997). Molecular damage in the bronchial epithelium of current and former smokers. J Natl Cancer Inst. 89(18):1366-73.

Woodson, K., Mason, J., Choi, S.W., Hartman, T., Tangrea, J., Virtamo, J., Taylor, P.R. and Albanes, D. (2001). Hypomethylation of p53 in peripheral blood DNA is associated with the development of lung cancer. Cancer Epidemiol Biomarkers Prev. 10(1):69-74.

Wright, J.L. and Churg, A. (2002). Animal models of cigarette smoke-induced COPD. Chest. 122(6 Suppl):301S-306S.

Wu, A.H., Yu, M.C., Thomas, D.C., Pike, M.C. and Henderson, B.E. (1988). Personal and family history of lung disease as risk factors for adenocarcinoma of the lung. Cancer Res. 48(24 Pt 1):7279-84.

Wynder, E.L., Fritz, L. and Furth, N. (1957). Effect of concentration of benzopyrene in skin carcinogenesis. J. Natl. Cancer Inst., 19: 361-370.

Wynder, E.L. and Graham, E.A. (1950). Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma: a study of six hundred and eighty-four proved cases. JAMA 143, 329-36.

Wynder, E.L., Graham, E.A. and Croninger, A.B. (1953). Experimental production of carcinoma with cigarette tar. Part 1. Cancer Res., 13: 855-864.

Wynder, E.L. and Hoffmann, D. (1994). Smoking and lung cancer: scientific challenges and opportunities. Cancer Res. 54(20):5284-95.

Yang, J.L., Chen, R.H., Maher, V.M. and McCormick, J.J. (1991). Kinds and location of mutations induced by (+/-)-7 beta,8 alpha-dihydroxy-9 alpha,10 alpha-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in diploid human fibroblasts. Carcinogenesis. 12(1):71-5.

Yang, P., Schwartz, A.G., McAllister, A.E., Aston, C.E. and Swanson, G.M. (1997). Genetic analysis of families with nonsmoking lung cancer probands. Genet Epidemiol. 14(2):181-97.

Yang, Q., Hergenhahn, M., Weninger, A. and Bartsch, H. (1999). Cigarette smoke induces direct DNA damage in the human B-lymphoid cell line Raji. Carcinogenesis. 20(9):1769-75.

Ye, B.C., Ikebukuro, K. and Karube, I. (1998). Quantitative analysis of polymerase chain reaction using anisotropy ratio and relative hydrodynamic volume of fluorescence polarization method. Nucleic Acid *26*(15), 3614-5.

Yokota, J., Wada, M., Shimosato, Y., Terada, M. and Sugimura, T. (1987). Loss of heterozygosity on chromosomes 3, 13, and 17 in small-cell carcinoma and on chromosome 3 in adenocarcinoma of the lung. Proc Natl Acad Sci USA. 84(24):9252-6.

Zarbl, H., Sukumar, S., Arthur, A.V., Martin-Zanca, D. and Barbacid, M. (1985). Direct mutagenesis of Ha-ras-1 oncogenes by N-nitroso-N-methylurea during initiation of mammary carcinogenesis in rats. Nature. 315(6018):382-5.

Zang, E.A. and Wynder, E.L. (1996). Differences in lung cancer risk between men and women: examination of the evidence. J Natl Cancer Inst. 88(3-4):183-92.

Zhang, W., Remenyik, E., Zelterman, D., Brash, D.E. and Wikonkal, N.M. (2001). Escaping the stem cell compartment: sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. Proc Natl Acad Sci USA. 98(24):13948-53.

Zheng, W.H., Khrapko, K., Coller, H., Thilly, W.G. and Copeland, W.C. Concordance of human in vivo mitochondrial point mutations with DNA polymerase errors. (manuscript in preparation).

Zirnstein, G., Helsel, L., Li, Y., Swaminathan, B. and Besser, J. (2000). Characterization of gyrA mutations associated with fluoroquinolone resistance in Campylobacter coli by DNA sequence analysis and MAMA PCR. FEMS Microbiol Lett. 190(1):1-7.

Zochbauer-Muller, S., Lam, S., Toyooka, S., Virmani, A.K., Toyooka, K.O., Seidl, S., Minna, J.D. and Gazdar, A.F. (2003). Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer. 107(4):612-6.

9. APPENDIX

9.1. Appendix A: Raw data on measurements and calculations of molecular copy numbers of DNA samples at each preparatory steps.

* The labeling of "sector" indicates the anatomical position (section 3.1.2.)

* "Total cells" per sector were calculated from quantitative PCR after genomic DNA isolation from tissue.

* The amount of "Cells / assay" were determined from the available target sequences of which the concentrations were measured by quantitative PCR after target sequence enrichment.

* "copy# / assay" were derived from "cells / assay" divided by two (two alleles / cell).

* "MT c(opies)# / assay" indicates the numbers of observed mutant copies per sector by MAMA.

* "MT c(opy)#-BG" indicates mutant copy numbers subtracted with TK6 background (bottom left corner).

* "MF (-BG)" indicates mutant fraction in which the assay background was already subtracted.

* "MT cell # / sector" was driven from:
"MT c# -BG" / 2 x "total cells" x 2 / "copy# / assay".
where the factor 2 comes from 2 alleles per cell.

* "MT colony # / sector" was driven from "MT cell # / sector" divided by 64 cells per turnover unit.

* Bottom: "Mean MF" indicates the mean of "MF (-BG)", whereas "Total MF" indicates the sum of "MT cell # / sector" divided by the sum of "total cells" from all sectors assayed for this assay and this lung. The latter was used as the mean MFs in section 4.2.2.

* In lungs which containes sectors with MF larger than $4x10^{-4}$, the numbers were calculated with and without such outliers

* Grey highlight indicates that the mutant signal was below the TK6 background (negative).

9.1.1. Appendix A.1. TP53 bp742 C:G->T:A in smokers

Smoker I, female 41yo (7 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/64cells
LMS1	2.90E+06	5.0E+05	1.0E+06	13.4	0	0	0	0
LMS2	3.70E+05	5.0E+05	1.0E+06	57	36.5	3.66E-05	13.5	0.2
RMS1	6.60E+05	1.5E+05	3.0E+05	1.6	0	0	0	0
RMS2	4.67E+06	5.0E+05	1.0E+06	25.1	4.6	4.64E-06	21.7	0.3
RMS3	1.30E+05	1.4E+05	2.7E+05	4.9	0	0	0	0
T1	2.16E+06	2.5E+05	5.0E+05	20.8	20.8	4.16E-05	89.8	1.4
T2	2.80E+06	5.0E+05	1.0E+06	36.3	15.8	1.58E-05	44.4	0.7
Sum	1.37E+07						169.3	
Mean	1.96E+06					1.41E-05		0.4

TK6	n=14
input (copies)	1E+06
BG	20.5
SD	10.8

Mean MF	1.41E-05
Total MF	1.24E-05
SD	1.80E-05

Smoker II, male 41yo (8 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/64cells
LB5	4.62E+06	4.0E+05	8.08E+05	23.6	3.1	3.89E-06	18.0	0.3
LB13	3.35E+06	4.1E+05	8.11E+05	20	0	0.0E+00	0	0
LB17	2.01E+06	4.2E+05	8.32E+05	29.8	9.3	1.12E-05	22.6	0.4
LB19	1.71E+06	4.0E+05	8.10E+05	36.1	15.6	1.93E-05	33.1	0.5
RB1	2.82E+06	4.0E+05	8.09E+05	15.4	0	0.0E+00	0	0
RB9	2.50E+06	4.1E+05	8.11E+05	143.1	122.6	1.51E-04	378.5	5.9
RB15	1.02E+06	3.4E+05	6.73E+05	5.9	0	0.0E+00	0	0
RB19	1.08E+06	3.2E+05	6.48E+05	68.5	48.0	7.42E-05	80.3	1.3
Sum	1.91E+07						532.5	
Mean	2.39E+06					3.25E-05	66.6	1.0

TK6	n=14
input (copies)	1E+06
BG	20.5
SD	10.8

Mean MF	3.25E-05
Total MF	2.79E-05
SD	5.41E-05

Smoker VI, M, 59yo (2 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/64cells
T2-1#1	3.04E+06	4.99E+05	9.98E+05	50	39.1	3.91E-05	118.9	1.9
T2-1#2	3.19E+06	4.98E+05	9.96E+05	50.6	39.7	3.98E-05	126.8	2.0
Sum	6.23E+06						245.7	
Mean	3.11E+06					3.95E-05		1.9

TK6	n=6
input (copies)	1E+06
BG	11.0
SD	7.7

Mean MF	3.95E-05
Total MF	3.95E-05
SD	4.82E-07

9.1.2. Appendix A.2. TP53 bp746 G:C->T:A in smokers

Smoker I, female 41yo (9 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcell#/sector	MTcolony#/sector
LMS1	2.90E+06	5.0E+05	1.0E+06	24.2	11.8	2.42E-05	34.1	0.5
LMS2	3.70E+05	1.0E+06	2.0E+06	19.3	6.9	9.65E-06	1.3	0.0
LMS3	2.00E+05	2.5E+05	5.0E+05	40.2	27.8	8.04E-05	11.1	0.2
RMS1	6.60E+05	5.0E+05	1.0E+06	45.3	32.9	4.53E-05	21.7	0.3
RMS2	4.67E+06	5.0E+05	1.0E+06	21.1	8.7	2.11E-05	40.5	0.6
RMS3	1.30E+05	5.0E+05	1.0E+06	4.3	0	0E+00	0.0	0.0
T1	2.16E+06	7.5E+05	1.5E+06	23.5	11.1	1.57E-05	15.9	0.2
T2	2.80E+06	5.0E+05	1.0E+06	26.2	13.8	2.62E-05	38.6	0.6
T3	7.80E+05	2.5E+05	5.0E+05	11	0	0E+00	0.0	0.0
SUM	1.5E+07						163.2	2.6
Mean	1.63E+06						18.1	0.3

Tk6	Input copy	1E+06
	Mean	12.4
	SD(n=14)	5.6

Mean MF	2.47E-05
Total MF	1.11E-05
SD	2.52E-05

Smoker II M 41yo (10 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LB5	1.42E+06	5.0E+05	1.0E+06	79.2	53.9	5.40E-05	76.7	1.2
LB7	2.19E+06	3.5E+05	7.0E+05	73.3	48.0	6.85E-05	150.1	2.3
LB13	2.65E+06	5.0E+05	1.0E+06	44.2	18.9	1.89E-05	50.0	0.8
LB17	1.80E+06	5.1E+05	1.0E+06	41.9	16.6	1.61E-05	29.1	0.5
LB19	1.21E+06	5.0E+05	1.0E+06	80.2	54.9	5.49E-05	66.4	1.0
RB1	2.23E+06	5.0E+05	1.0E+06	177.1	151.8	1.52E-04	338.9	5.3
RB9	9.35E+06	5.0E+05	1.0E+06	64.9	39.6	3.95E-05	369.8	5.8
RB15	2.53E+06	4.0E+05	8.0E+05	43.1	17.8	2.22E-05	56.3	0.9
RB17	2.66E+06	2.4E+05	4.7E+05	104.5	79.2	1.68E-04	448.1	7.0
RB19	1.38E+06	4.0E+05	8.0E+05	96.9	71.6	8.95E-05	123.5	1.9
Sum	2.74E+07						1708.8	
Mean	2.74E+06					6.84E-05	170.9	2.7

Tk6	input	1E+06
(copies)	BG	25.3
n=20	SD	27.5

Mean MF	6.84E-05
Total MF	6.23E-05
SD	5.37E-05

Smoker III F 47yo (67 sectors)

Sector	Total cells/s	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcell#/sector	MTcolony#/sector
LLL1-1	2.89E+06	7.22E+05	1.44E+06	6.3	0	0E+00	0	0
LLL1-2	1.07E+06	2.67E+05	5.34E+05	10.1	1.8	3.44E-06	3.7	0.1
LLL2-1	3.10E+05	7.76E+04	1.55E+05	5.4	0	0E+00	0	0
LLL2-2	2.64E+04	6.61E+03	1.32E+04	31.5	23.3	1.76E-03	46.5	0.7
LLL3-1	8.90E+04	2.23E+04	4.45E+04	2.2	0	0E+00	0	0
LLL3-2	2.43E+05	6.07E+04	1.21E+05	4.7	0	0E+00	0	0
LLL4-1	3.08E+06	7.71E+05	1.54E+06	21.1	12.8	8.31E-06	25.6	0.4
LLL4-2	1.14E+06	2.86E+05	5.72E+05	4.3	0	0E+00	0	0
LLL5	8.17E+04	1.36E+04	2.72E+04	1.1	0	0E+00	0	0
LLL7	8.64E+05	2.16E+05	4.32E+05	7.0	0	0E+00	0	0
LLL11	1.30E+06	3.26E+05	6.52E+05	6.5	0	0E+00	0	0
LLL15	3.06E+05	5.11E+04	1.02E+05	0.0	0	0E+00	0	0
LUL1	3.55E+06	8.87E+05	1.77E+06	8.1	0	0E+00	0	0
LUL1-2	5.73E+05	9.55E+04	1.91E+05	24.4	16.1	8.44E-05	48.4	0.8
LUL2-1	2.92E+06	7.31E+05	1.46E+06	14.0	5.8	3.97E-06	11.6	0.2
LUL2-2	2.16E+06	5.40E+05	1.08E+06	72.0	63.8	5.90E-05	127.5	2.0
LUL3	9.78E+05	2.45E+05	4.89E+05	52.4	44.1	9.02E-05	88.3	1.4
LUL4	6.14E+05	1.54E+05	3.07E+05	12.6	4.4	1.42E-05	8.7	0.1
LUL5	2.58E+05	6.45E+04	1.29E+05	2.3	0	0E+00	0	0
LUL7	5.23E+05	6.53E+04	1.31E+05	17.1	8.9	6.78E-05	35.4	0.6
LUL9	2.06E+05	5.15E+04	1.03E+05	2.2	0	0E+00	0	0
LUL10	7.15E+05	8.93E+04	1.79E+05	53.1	44.9	2.51E-04	179.4	2.8
LUL11-1	2.28E+04	3.80E+03	7.59E+03	6.1	0	0E+00	0	0
LUL11-2	1.23E+05	2.05E+04	4.09E+04	5.3	0	0E+00	0	0
LUL17	2.47E+05	6.19E+04	1.24E+05	17.7	9.5	7.64E-05	18.9	0.3
T1	3.06E+06	3.82E+05	7.64E+05	4.2	0	0E+00	0	0
T2	2.87E+06	3.59E+05	7.18E+05	7.9	0	0E+00	0	0
T3	1.45E+06	1.81E+05	3.62E+05	4.1	0	0E+00	0	0
T4	2.39E+05	3.98E+04	7.96E+04	0.0	0	0E+00	0	0
T5	1.48E+06	1.85E+05	3.71E+05	10.0	1.8	4.79E-06	7.1	0.1
T6	1.87E+06	4.66E+05	9.33E+05	0.8	0	0E+00	0	0
T8	2.84E+06	7.10E+05	1.42E+06	1.0	0	0E+00	0	0
T9	5.83E+05	9.72E+04	1.94E+05	9.1	0.9	4.51E-06	2.6	0.0
T10	3.81E+05	6.35E+04	1.27E+05	0.0	0	0E+00	0	0
T11	2.99E+06	7.47E+05	1.49E+06	0.3	0	0E+00	0	0
T12	1.69E+05	2.82E+04	5.64E+04	6.4	0	0E+00	0	0
T13	9.75E+05	2.44E+05	4.87E+05	14.3	6.1	1.25E-05	12.2	0.2
T15	2.42E+06	6.06E+05	1.21E+06	3.6	0	0E+00	0	0
T16	1.09E+05	2.73E+04	5.46E+04	26.6	18.3	3.35E-04	36.6	0.6
RUL1-1	9.12E+05	1.14E+05	2.28E+05	103.5	95.3	4.18E-04	381.0	6.0
RUL1-2	3.02E+05	3.77E+04	7.54E+04	18.9	10.7	1.41E-04	42.6	0.7
RUL2	4.82E+05	6.03E+04	1.21E+05	34.9	26.7	2.21E-04	106.7	1.7
RUL3	1.79E+05	2.98E+04	5.97E+04	6.7	0	0E+00	0	0
RUL4	1.68E+06	2.10E+05	4.19E+05	27.9	19.7	4.69E-05	78.6	1.2
RUL7	3.80E+05	9.50E+04	1.90E+05	75.6	67.4	3.54E-04	134.7	2.1
RMS1	1.24E+06	2.07E+05	4.14E+05	176.4	168.2	4.06E-04	504.5	7.9
RMS2	7.13E+05	1.19E+05	2.38E+05	42.0	33.8	1.42E-04	101.3	1.6
RLL1-1	2.63E+05	4.38E+04	8.77E+04	37.2	29.0	3.30E-04	86.9	1.4
RLL1-2	1.01E+06	1.68E+05	3.36E+05	2364.0	2355.8	7.01E-03	7067.3	110.4
RLL2	6.46E+05	1.08E+05	2.15E+05	92.4	84.2	3.91E-04	252.5	3.9
RLL3-1	1.25E+06	2.08E+05	4.16E+05	6.0	0	0E+00	0	0
RLL3-2	1.85E+05	3.09E+04	6.17E+04	53.5	45.3	7.33E-04	135.8	2.1
RLL4	1.49E+06	2.48E+05	4.95E+05	26.4	18.2	3.67E-05	54.5	0.9
RLL5-1	1.73E+06	2.88E+05	5.76E+05	88.8	80.6	1.40E-04	241.7	3.8
RLL5-2	1.30E+05	2.16E+04	4.32E+04	115.2	107.0	2.47E-03	320.9	5.0
RLL8	1.38E+05	2.29E+04	4.59E+04	69.6	61.4	1.34E-03	184.1	2.9
RLL9	8.84E+05	1.47E+05	2.95E+05	30.6	22.4	7.58E-05	67.1	1.0
RLL13	4.37E+05	7.28E+04	1.46E+05	12.8	4.6	3.16E-05	13.8	0.2
RLL14	4.26E+05	7.09E+04	1.42E+05	5.0	0	0E+00	0	0
RLL15	8.75E+05	1.46E+05	2.92E+05	600.0	591.8	2.03E-03	1775.3	27.7
RLL16	1.54E+05	2.56E+04	5.13E+04	0.6	0	0E+00	0	0
LI+C	2.46E+05	4.10E+04	8.21E+04	32.2	23.9	2.91E-04	71.8	1.1
L2	1.00E+06	1.67E+05	3.33E+05	15.0	6.8	2.03E-05	20.3	0.3
L3	5.87E+05	9.79E+04	1.96E+05	109.7	101.4	5.18E-04	304.3	4.8
L4	6.09E+05	1.01E+05	2.03E+05	483.6	475.4	2.34E-03	1426.1	22.3
L5	3.82E+05	6.36E+04	1.27E+05	21.6	13.4	1.05E-04	40.1	0.6
L6	7.68E+05	1.28E+05	2.56E+05	27.6	19.4	7.56E-05	58.1	0.9
Sum (n=67)	6.48E+07						1.41E+04	
Sum (n=57)	5.91E+07						1.98E+03	
Mean (n=67)	9.67E+05					3.35E-04		3.3
Mean (n=57)	1.04E+06					6.00E-05		0.5

TK6	BG (copies)
n=7	8.2
SD	1.8

n=67	Mean MF	3.35E-04
all sectors	Total MF	2.18E-04
	SD	9.86E-04
n=57	Mean MF	6.00E-05
MF<4E-4	Total MF	3.34E-05
	SD	1.04E-04

Smoker IV, male 55yo (22 sectors)

Sector	Total cells/s	Cells/assay	copy#/assay	MTcopy#/assay	MTcopy-BG	MF	MTcells/sector	MTcolony#/sector
LLL1+2+3	1.26E+07	2.52E+06	5.04E+06	38.6	22.9	4.54E-06	57.2	0.9
LLL4+4*+5	1.60E+07	3.20E+06	6.40E+06	21.9	6.2	9.72E-07	15.6	0.2
LLL6+7+7*	2.08E+07	3.12E+06	6.24E+06	22.6	6.9	1.11E-06	23.0	0.4
LUL3+4	4.53E+06	1.70E+06	3.40E+06	33.7	18.0	5.30E-06	24.0	0.4
LUL6+7+8	3.77E+06	1.23E+06	2.45E+06	45.7	30.0	1.22E-05	46.2	0.7
LUL8+12+9+9*	1.91E+07	3.82E+06	7.64E+06	179.7	164.0	2.15E-05	409.9	6.4
LUL11+11*+11**	6.34E+06	2.85E+06	5.71E+06	34.1	18.4	3.22E-06	20.4	0.3
LUL12+13+14	3.48E+06	1.74E+06	3.48E+06	35.7	20.0	5.74E-06	20.0	0.3
LUL17+17*+18	1.11E+07	2.22E+06	4.43E+06	36.1	20.4	4.60E-06	51.0	0.8
RML3+4	1.91E+06	6.69E+05	1.34E+06	46.2	30.5	2.28E-05	43.6	0.7
RLL15	1.63E+06	6.11E+05	1.22E+06	124.1	108.4	8.87E-05	144.5	2.3
RLL20	3.19E+06	1.20E+06	2.39E+06	18.1	2.4	1.02E-06	3.2	0.1
RLL21+22	8.20E+05	3.08E+05	6.15E+05	44.6	28.9	4.70E-05	38.5	0.6
RUL2	7.72E+06	2.90E+06	5.79E+06	102.5	86.8	1.50E-05	115.7	1.8
RUL5	3.85E+06	1.44E+06	2.89E+06	19.6	3.9	1.35E-06	5.2	0.1
RUL9	3.22E+06	1.21E+06	2.42E+06	118.5	102.8	4.26E-05	137.1	2.1
RUL10	1.67E+06	6.26E+05	1.25E+06	23.8	8.1	6.47E-06	10.8	0.2
RUL13	3.85E+06	1.44E+06	2.89E+06	57.8	42.1	1.46E-05	56.1	0.9
RUL14	3.44E+06	1.29E+06	2.58E+06	19.9	4.2	1.64E-06	5.7	0.1
RUL15	1.28E+06	4.80E+05	9.60E+05	56.7	41.0	4.27E-05	54.7	0.9
RUL16	1.47E+06	5.51E+05	1.10E+06	18.5	2.8	2.54E-06	3.7	0.1
RUL18	6.54E+05	2.45E+05	4.91E+05	52.9	37.2	7.58E-05	49.6	0.8
Sum	1.32E+08						1335.7	
Mean	6.02E+06					1.92E-05	60.7	0.9

TK6	BG (copies)
n=15	15.7
SD	9.0

Mean MF	**1.92E-05**
Total MF	**1.01E-05**
SD	**2.50E-05**

Smoker V F 58yo (41 sectors)

Sector	Total cells/s	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
LUL1T	3.49E+06	1.16E+06	2.33E+06	21.5	5.8	2.47E-06	8.6	0.1
LUL2	1.45E+06	4.85E+05	9.70E+05	2.2	0	0E+00	0	0
LUL3	2.26E+06	7.52E+05	1.50E+06	13.8	0	0E+00	0	0
LUL4	9.83E+05	3.28E+05	6.55E+05	27.4	11.7	1.79E-05	17.6	0.3
LUL5	2.27E+06	7.58E+05	1.52E+06	17.8	2.1	1.39E-06	3.2	0.0
LUL6	3.04E+06	1.01E+06	2.03E+06	91.0	75.3	3.72E-05	113.0	1.8
LUL7	2.16E+06	7.21E+05	1.44E+06	8.8	0	0E+00	0	0
LUL9	2.25E+06	7.51E+05	1.50E+06	2.9	0	0E+00	0	0
LUL10	6.16E+04	2.05E+04	4.11E+04	35.1	19.4	4.72E-04	29.1	0.5
LUL11	1.63E+05	5.42E+04	1.08E+05	22.8	7.1	6.55E-05	10.7	0.2
LUL15+15*+1	7.06E+04	2.35E+04	4.71E+04	45.0	29.3	6.23E-04	44.0	0.7
LLL1	1.13E+05	3.78E+04	7.56E+04	12.5	0	0E+00	0	0
LLL2+2C	7.19E+04	2.40E+04	4.79E+04	20.0	4.3	8.97E-05	6.5	0.1
LLL3	1.83E+05	6.10E+04	1.22E+05	20.0	4.3	3.52E-05	6.5	0.1
LLL4	1.46E+05	4.88E+04	9.76E+04	29.0	13.3	1.36E-04	20.0	0.3
LLL5	1.85E+05	6.16E+04	1.23E+05	26.0	10.3	8.36E-05	15.5	0.2
LLL6+7+8	2.91E+05	9.71E+04	1.94E+05	11.0	0	0E+00	0	0
LLL9+11	1.01E+05	3.38E+04	6.75E+04	4.0	0	0E+00	0	0
LLL15	1.58E+05	5.25E+04	1.05E+05	107.0	91.3	8.69E-04	137.0	2.1
RLL1	1.41E+05	4.70E+04	9.41E+04	11.0	0	0E+00	0	0
RLL3	3.31E+06	1.10E+06	2.21E+06	5260.0	5244.3	2.37E-03	7866.5	122.9
RLL3*	4.02E+06	1.34E+06	2.68E+06	51000.0	50984.3	1.90E-02	76476.5	1194.9
RLL4*	4.45E+06	1.48E+06	2.97E+06	166100.0	166084.3	5.59E-02	249126.5	3892.6
RLL5	5.66E+06	1.89E+06	3.77E+06	7720.0	7704.3	2.04E-03	11556.5	180.6
RLL5*	3.24E+06	1.08E+06	2.16E+06	28950.0	28934.3	1.34E-02	43401.5	678.1
RLL6	1.93E+06	6.44E+05	1.29E+06	24.0	8.3	6.45E-06	12.5	0.2
RLL7	1.25E+06	4.17E+05	8.35E+05	25.0	9.3	1.11E-05	14.0	0.2
RLL9	2.18E+06	7.26E+05	1.45E+06	6.0	0	0E+00	0	0
RLL13	2.78E+06	9.25E+05	1.85E+06	74.0	58.3	3.15E-05	87.5	1.4
RLL14	2.48E+06	8.27E+05	1.65E+06	91.0	75.3	4.55E-05	113.0	1.8
RUL1	7.42E+05	2.47E+05	4.95E+05	88.0	72.3	1.46E-04	108.5	1.7
RUL2	4.09E+06	1.36E+06	2.73E+06	12.0	0	0E+00	0	0
RUL3	1.83E+06	6.10E+05	1.22E+06	14.0	0	0E+00	0	0
RUL4	4.24E+05	1.41E+05	2.83E+05	31.0	15.3	5.41E-05	23.0	0.4
RUL5+7+8	2.25E+06	7.50E+05	1.50E+06	94.0	78.3	5.22E-05	117.5	1.8
RUL10+11	1.27E+06	4.24E+05	8.47E+05	1030.0	1014.3	1.20E-03	1521.5	23.8
RUL12	6.18E+05	2.06E+05	4.12E+05	11.0	0	0E+00	0	0
RUL13	4.15E+05	1.38E+05	2.77E+05	9.0	0	0E+00	0	0
Main Carina	2.25E+06	7.51E+05	1.50E+06	14.6	0	0E+00	0	0
Main Carina*	2.96E+07	9.86E+06	1.97E+07	50.0	34.3	1.74E-06	51.5	0.8
TRAC	4.87E+05	1.62E+05	3.24E+05	2.8	0	0E+00	0	0
Sum (n=41)	9.49E+07						390887	
Sum (n=32)	7.26E+07						728	
Mean (n=41)	2.31E+06					2.36E-03		149.0
Mean (n=32)	2.27E+06					2.56E-05		0.4

TK6	BG (copies)
n=15	15.7
SD	9.0

n=41	Mean MF	2.36E-03
All sectors	Total MF	4.12E-03
	SD	9.30E-03
n=32	mean MF	2.56E-05
MF>4e-4	total MF	1.00E-05
	SD	4.02E-05

Smoker VI, male, 59yo, (2 sectors)

Sector	Total cells/s	Cells/assay	copy#/assay	MTcopy#/assay	MTcopy-BG	MF	MTcells/sector	MTcolony#/sector
T2-1#1	3.04E+06	1.52E+06	3.04E+06	26.7	11.0	3.62E-06	11.0	0.2
T2-1#2	3.19E+06	1.59E+06	3.19E+06	22.6	6.9	2.17E-06	6.9	0.1
Sum	6.23E+06						18	
Mean	3.11E+06					2.89E-06	9	0.1

TK6	BG (copies)
n=15	15.7
SD	9.0

Mean MF	2.89E-06
Total MF	2.88E-06
SD	1.03E-06

9.1.3. Appendix A.3. TP53 bp747 G:C->T:A in smokers

Smoker I, female 41yo (9 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1	2.09E+06	5.0E+05	1.0E+06	55	15.7	1.57E-05	32.9	0.5
LMS2	1.50E+06	5.0E+05	1.0E+06	21.1	0.0	0.0	0.0	0.0
LMS3	6.56E+05	2.5E+05	5.0E+05	59.7	20.4	4.07E-05	26.7	0.4
RMS1	1.17E+06	5.0E+05	1.0E+06	145	105.7	1.06E-04	123.6	1.9
RMS2	2.89E+06	5.0E+05	1.0E+06	47.5	8.2	8.22E-06	23.8	0.4
RMS3	1.18E+06	5.0E+05	1.0E+06	31.1	0.0	0.0	0.0	0.0
T1	2.18E+06	7.5E+05	1.5E+06	92.4	53.1	3.54E-05	77.1	1.2
T2	1.15E+06	5.0E+05	1.0E+06	50.7	11.4	1.14E-05	13.1	0.2
T3	5.98E+05	2.5E+05	5.0E+05	0	0.0	0.0	0.0	0.0
Sum	1.34E+07						297.2	4.6
Mean	1.49E+06						33.0	0.5

Tk6	(n=5)
input	1E+06
BG (copies	39.3
SD	10.9

Mean MF	2.41E-05
Total MF	2.22E-05
SD	3.41E-05

Smoker II M 41yo (10 sectors)

Sample	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LB5	5.70E+06	5.0E+05	1.0E+06	26.5	0	0E+00	0	0
LB7	7.30E+05	3.5E+05	7.0E+05	1782.7	1743.4	2.49E-03	1818.0	28.4
LB13	4.14E+06	5.0E+05	1.0E+06	68.4	29.1	2.91E-05	120.3	1.9
LB17	2.48E+06	5.1E+05	1.0E+06	165.9	126.6	1.23E-04	305.8	4.8
LB19	2.11E+06	5.0E+05	1.0E+06	25.4	0	0E+00	0	0
RB1	3.48E+06	5.0E+05	1.0E+06	116.4	77.1	7.72E-05	268.7	4.2
RB9	3.09E+06	5.0E+05	1.0E+06	135.2	95.9	9.58E-05	296.0	4.6
RB15	1.26E+06	4.0E+05	8.0E+05	70.2	30.9	3.87E-05	48.7	0.8
RB17	4.73E+05	2.4E+05	4.7E+05	76.6	37.3	7.94E-05	37.5	0.6
RB19	1.34E+06	4.0E+05	8.0E+05	25.7	0	0.00E+00	0	0
Sum (n=10)	2.48E+07						2895.1	
Sum (n=9)	2.41E+07						1077.2	
Mean (n=10)	2.48E+06					2.93E-04		4.5
Mean (n=9)	2.67E+06					4.93E-05		1.9

Tk6	(n=5)
input	1E+06
BG (copies)	39.3
SD	10.9

n=10	Mean MF	2.93E-04
	Total MF	1.17E-04
	SD	7.73E-04
n=9	Mean MF	4.93E-05
MF<4E-4	Total MF	4.48E-05
	SD	4.63E-05

Smoker III F 47yo (67 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
LLL1-1	2.89E+06	7.22E+05	1.44E+06	59	51	3.51E-05	101	1.6
LLL1-2	1.07E+06	2.67E+05	5.34E+05	170	161	3.02E-04	323	5.0
LLL2-1	3.10E+05	7.76E+04	1.55E+05	59	51	3.26E-04	101	1.6
LLL2-2	2.64E+04	6.61E+03	1.32E+04	55	46	3.51E-03	93	1.5
LLL3-1	8.90E+04	2.23E+04	4.45E+04	177	168	3.78E-03	337	5.3
LLL3-2	2.43E+05	6.07E+04	1.21E+05	0	0	0E+00	0	0
LLL4-1	3.08E+06	7.71E+05	1.54E+06	227	218	1.41E-04	436	6.8
LLL4-2	1.14E+06	2.86E+05	5.72E+05	205	196	3.43E-04	393	6.1
LLL5	8.17E+04	1.36E+04	2.72E+04	5	0	0E+00	0	0
LLL7	8.64E+05	2.16E+05	4.32E+05	59	51	1.17E-04	101	1.6
LLL11	1.30E+06	3.26E+05	6.52E+05	40	31	4.76E-05	62	1.0
LLL15	3.06E+05	5.11E+04	1.02E+05	5	0	0E+00	0	0
LUL1	3.55E+06	8.87E+05	1.77E+06	171	162	9.14E-05	325	5.1
LUL1-2	5.73E+05	9.55E+04	1.91E+05	12	3	1.70E-05	10	0.2
LUL2-1	2.92E+06	7.31E+05	1.46E+06	5	0	0E+00	0	0
LUL2-2	2.16E+06	5.40E+05	1.08E+06	792	783	7.25E-04	1567	24.5
LUL3	9.78E+05	2.45E+05	4.89E+05	50	42	8.51E-05	83	1.3
LUL4	6.14E+05	1.54E+05	3.07E+05	22	13	4.18E-05	26	0.4
LUL5	2.58E+05	6.45E+04	1.29E+05	81	73	5.62E-04	145	2.3
LUL7	5.23E+05	6.53E+04	1.31E+05	9	0	1.94E-06	1	0.0
LUL9	2.06E+05	5.15E+04	1.03E+05	435	426	4.14E-03	853	13.3
LUL10	7.15E+05	8.93E+04	1.79E+05	62	53	2.99E-04	213	3.3
LUL11-1	2.28E+04	3.80E+03	7.59E+03	3	0	0E+00	0	0
LUL11-2	1.23E+05	2.05E+04	4.09E+04	0	0	0E+00	0	0
LUL17	2.47E+05	6.19E+04	1.24E+05	23	14	1.14E-04	28	0.4
T1	3.06E+06	3.82E+05	7.64E+05	28	19	2.51E-05	77	1.2
T2	2.87E+06	3.59E+05	7.18E+05	32	23	3.17E-05	91	1.4
T3	1.45E+06	1.81E+05	3.62E+05	18	9	2.56E-05	37	0.6
T4	2.39E+05	3.98E+04	7.96E+04	8	0	0E+00	0	0
T5	1.48E+06	1.85E+05	3.71E+05	13	4	1.04E-05	15	0.2
T6	1.87E+06	4.66E+05	9.33E+05	124	115	1.24E-04	231	3.6
T8	2.84E+06	7.10E+05	1.42E+06	288	279	1.97E-04	559	8.7
T9	5.83E+05	9.72E+04	1.94E+05	8	0	0E+00	0	0
T10	3.81E+05	6.35E+04	1.27E+05	10	1	8.62E-06	3	0.1
T11	2.99E+06	7.47E+05	1.49E+06	124	115	7.72E-05	231	3.6
T12	1.69E+05	2.82E+04	5.64E+04	2	0	0E+00	0	0
T13	9.75E+05	2.44E+05	4.87E+05	171	162	3.33E-04	325	5.1
T15	2.42E+06	6.06E+05	1.21E+06	211	202	1.67E-04	404	6.3
T16	1.09E+05	2.73E+04	5.46E+04	42	33	6.09E-04	67	1.0
RUL1-1	9.12E+05	1.14E+05	2.28E+05	62	53	2.34E-04	213	3.3
RUL1-2	3.02E+05	3.77E+04	7.54E+04	1	0	0E+00	0	0
RUL2	4.82E+05	6.03E+04	1.21E+05	11	2	1.70E-05	8	0.1
RUL3	1.79E+05	2.98E+04	5.97E+04	16	7	1.15E-04	21	0.3
RUL4	1.68E+06	2.10E+05	4.19E+05	3	0	0E+00	0	0
RUL7	3.80E+05	9.50E+04	1.90E+05	32	24	1.24E-04	47	0.7
RMS1	1.24E+06	2.07E+05	4.14E+05	4	0	0E+00	0	0
RMS2	7.13E+05	1.19E+05	2.38E+05	5	0	0E+00	0	0
RLL1-1	2.63E+05	4.38E+04	8.77E+04	20	12	1.33E-04	35	0.5
RLL1-2	1.01E+06	1.68E+05	3.36E+05	18	9	2.75E-05	28	0.4
RLL2	6.46E+05	1.08E+05	2.15E+05	1	0	0E+00	0	0
RLL3-1	1.25E+06	2.08E+05	4.16E+05	60	51	1.23E-04	154	2.4
RLL3-2	1.85E+05	3.09E+04	6.17E+04	49	40	6.55E-04	121	1.9
RLL4	1.49E+06	2.48E+05	4.95E+05	3	0	0E+00	0	0
RLL5-1	1.73E+06	2.88E+05	5.76E+05	10	1	1.48E-06	3	0.0
RLL5-2	1.30E+05	2.16E+04	4.32E+04	14	5	1.22E-04	16	0.2
RLL8	1.38E+05	2.29E+04	4.59E+04	10	1	1.86E-05	3	0.0
RLL9	8.84E+05	1.47E+05	2.95E+05	1	0	0E+00	0	0
RLL13	4.37E+05	7.28E+04	1.46E+05	10	1	5.86E-06	3	0.0
RLL14	4.26E+05	7.09E+04	1.42E+05	13	4	3.14E-05	13	0.2
RLL15	8.75E+05	1.46E+05	2.92E+05	52	43	1.47E-04	129	2.0
RLL16	1.54E+05	2.56E+04	5.13E+04	0	0	0E+00	0	0
LI+C	2.46E+05	4.10E+04	8.21E+04	6	0	0E+00	0	0
L2	1.00E+06	1.67E+05	3.33E+05	104	96	2.87E-04	287	4.5
L3	5.87E+05	9.79E+04	1.96E+05	60	51	2.62E-04	154	2.4
L4	6.09E+05	1.01E+05	2.03E+05	43	34	1.70E-04	103	1.6
L5	3.82E+05	6.36E+04	1.27E+05	8	0	0E+00	0	0
L6	7.68E+05	1.28E+05	2.56E+05	37	28	1.11E-04	85	1.3
Sum (n=67)	6.48E+07						8.66E+03	
Sum (n=60)	6.18E+07						5.48E+03	
Mean (n=67)	9.67E+05					2.82E-04		2.0
Mean (n=60)	1.03E+06					8.15E-05		1.4

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

n=67	Mean MF	2.82E-04
	Total MF	1.34E-04
	SD	7.89E-04
n=60	Mean MF	8.15E-05
MF<4E-4	Total MF	8.86E-05
	SD	1.03E-04

Smoker IV, male 55yo (23 sectors)

Sector	Total cells/se	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolonies/sector
LLL1+2+3	1.26E+07	5.04E+06	1.01E+07	18.0	9.3	9.20E-07	11.6	0.2
LLL4+4*+5	1.60E+07	6.40E+06	1.28E+07	12.1	3.4	2.63E-07	4.2	0.1
LLL6+7+7*	2.08E+07	6.24E+06	1.25E+07	10.2	1.5	1.19E-07	2.5	0.0
LUL3+4	4.53E+06	3.40E+06	6.80E+06	14.1	5.4	7.88E-07	3.6	0.1
LUL6+7+8	3.77E+06	2.45E+06	4.90E+06	19.5	10.8	2.20E-06	8.3	0.1
LUL8+12+9+9	1.91E+07	7.64E+06	1.53E+07	8.4	0	0E+00	0	0
LUL11+11*+11	6.34E+06	5.71E+06	1.14E+07	23.8	15.0	1.32E-06	8.3	0.1
LUL12+13+14	3.48E+06	3.48E+06	6.96E+06	42.3	33.6	4.82E-06	16.8	0.3
LUL17+17*+18	1.11E+07	4.43E+06	8.86E+06	15.5	6.8	7.65E-07	8.5	0.1
RML3+4	1.91E+06	1.34E+06	2.67E+06	24.1	15.3	5.73E-06	10.9	0.2
RLL15	1.26E+06	9.45E+05	1.89E+06	24.2	15.5	8.18E-06	10.3	0.2
RLL20	3.19E+06	1.36E+06	2.73E+06	24.2	15.4	5.65E-06	18.0	0.3
RLL21+22	8.20E+05	3.50E+05	7.01E+05	28.1	19.3	2.76E-05	22.6	0.4
RUL2	7.72E+06	3.09E+06	6.18E+06	8.1	0	0E+00	0	0
RUL5	3.85E+06	2.89E+06	5.78E+06	38.1	29.3	5.07E-06	19.5	0.3
RUL9	3.22E+06	2.42E+06	4.83E+06	32.7	24.0	4.96E-06	16.0	0.2
RUL8	2.44E+06	1.59E+06	3.17E+06	25.6	16.8	5.30E-06	12.9	0.2
RUL10	1.67E+06	1.09E+06	2.17E+06	83.7	75.0	3.45E-05	57.7	0.9
RUL13	3.85E+06	2.50E+06	5.01E+06	98.7	90.0	1.80E-05	69.2	1.1
RUL14	3.44E+06	1.20E+06	2.41E+06	20.9	12.2	5.06E-06	17.4	0.3
RUL15	1.28E+06	8.32E+05	1.66E+06	46.6	37.9	2.27E-05	29.1	0.5
RUL16	1.47E+06	1.03E+06	2.06E+06	36.8	28.1	1.36E-05	20.0	0.3
RUL18	6.54E+05	4.58E+05	9.16E+05	102.9	94.1	1.03E-04	67.2	1.1
Sum	1.34E+08						434.7	6.8
Mean	5.85E+06					1.18E-05	18.9	0.3

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

Mean MF	1.18E-05
Total MF	3.23E-06
SD	2.20E-05

Smoker V F 58yo (41 sectors)

35

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcell#/sector	MTcolony#/sector
LUL1T	3.49E+06	1.16E+06	2.33E+06	80	71.3	3.06E-05	106.9	1.7
LUL2	1.45E+06	4.85E+05	9.70E+05	30	21.3	2.19E-05	32	0.5
LUL3	2.26E+06	7.52E+05	1.50E+06	242	233.3	1.55E-04	349.9	5.5
LUL4	9.83E+05	3.28E+05	6.55E+05	6.2	0	0E+00	0	0
LUL5	2.27E+06	7.58E+05	1.52E+06	28.1	19.4	1.28E-05	29	0.5
LUL6	3.04E+06	1.01E+06	2.03E+06	36.3	27.6	1.36E-05	41	0.6
LUL7	2.16E+06	7.21E+05	1.44E+06	81.1	72.4	5.02E-05	108.5	1.7
LUL9	2.25E+06	7.51E+05	1.50E+06	22	13.3	8.82E-06	20	0.3
LUL10	6.16E+04	2.05E+04	4.11E+04	4.3	0	0E+00	0	0
LUL11	1.63E+05	5.42E+04	1.08E+05	3.9	0	0E+00	0	0
LUL15+15*+16	7.06E+04	2.35E+04	4.71E+04	7	0	0E+00	0	0
LLL1	1.13E+05	3.78E+04	7.56E+04	3.5	0	0E+00	0	0
LLL2+2C	7.19E+04	2.40E+04	4.79E+04	4	0	0E+00	0	0
LLL3	1.83E+05	6.10E+04	1.22E+05	10	1.3	1.03E-05	2	0.0
LLL4	1.46E+05	4.88E+04	9.76E+04	6.2	0	0E+00	0	0
LLL5	1.85E+05	6.16E+04	1.23E+05	8.1	0	0E+00	0	0
LLL6+7+8	2.91E+05	9.71E+04	1.94E+05	4	0	0E+00	0	0
LLL9+11	1.01E+05	3.38E+04	6.75E+04	18	9.3	1.37E-04	14	0.2
LLL15	1.58E+05	5.25E+04	1.05E+05	13	4.3	4.05E-05	6	0.1
RLL1	1.41E+05	4.70E+04	9.41E+04	3.5	0	0E+00	0	0
RLL3	3.31E+06	1.10E+06	2.21E+06	595	586.3	2.65E-04	879.4	13.7
RLL3*	4.02E+06	1.34E+06	2.68E+06	1160	1151.3	4.30E-04	1726.9	27.0
RLL4*	4.45E+06	1.48E+06	2.97E+06	33800	33791.3	1.14E-02	50686.9	792.0
RLL5	5.66E+06	1.89E+06	3.77E+06	1530	1521.3	4.03E-04	2281.9	35.7
RLL5*	3.24E+06	1.08E+06	2.16E+06	18970	18961.3	8.78E-03	28441.9	444.4
RLL6	1.93E+06	6.44E+05	1.29E+06	24	15.3	1.18E-05	23	0.4
RLL7	1.25E+06	4.17E+05	8.35E+05	62.6	53.9	6.45E-05	80.8	1.3
RLL9	2.18E+06	7.26E+05	1.45E+06	29.1	20.4	1.40E-05	31	0.5
RLL13	2.78E+06	9.25E+05	1.85E+06	260	251.3	1.36E-04	376.9	5.9
RLL14	2.48E+06	8.27E+05	1.65E+06	16	7.3	4.38E-06	11	0.2
RUL1	7.42E+05	2.47E+05	4.95E+05	85	76.3	1.54E-04	114.4	1.8
RUL2	4.09E+06	1.36E+06	2.73E+06	23	14.3	5.22E-06	21	0.3
RUL3	1.83E+06	6.10E+05	1.22E+06	18.2	9.5	7.75E-06	14	0.2
RUL4	4.24E+05	1.41E+05	2.83E+05	118	109.3	3.86E-04	163.9	2.6
RUL5+7+8	2.25E+06	7.50E+05	1.50E+06	610	601.3	4.01E-04	901.9	14.1
RUL10+11	1.27E+06	4.24E+05	8.47E+05	400	391.3	4.62E-04	586.9	9.2
RUL12	6.18E+05	2.06E+05	4.12E+05	104	95.3	2.31E-04	142.9	2.2
RUL13	4.15E+05	1.38E+05	2.77E+05	119	110.3	3.98E-04	165.4	2.6
Main Carina	2.25E+06	7.51E+05	1.50E+06	117	108.3	7.21E-05	162.4	2.5
Main Carina*	2.96E+07	9.86E+06	1.97E+07	225	216.3	1.10E-05	324.4	5.1
TRAC	4.87E+05	1.62E+05	3.24E+05	71.3	62.6	1.93E-04	93.8	1.5
Sum (n=41)	9.49E+07						8.79E+04	
Sum (n=35)	7.40E+07						3.31E+03	
Mean (n=41)	2.31E+06					5.93E-04		33.5
Mean (n=35)	2.11E+06					6.96E-05		1.5

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

n=41	Mean MF	5.93E-04
	Total MF	9.27E-04
	SD	2.20E-03
n=35	Mean MF	6.96E-05
MF<4E-5	Total MF	4.48E-05
	SD	1.09E-04

Smoker VI, male, 59yo, (2 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	Mtcolony#/sector
T2-1#1	3.04E+06	2.58E+06	5.17E+06	24.6	15.9	3.07E-06	9.3	0.1
T2-1#2	3.19E+06	3.19E+06	6.37E+06	16.7	7.9	1.25E-06	4.0	0.1
Sum	6.23E+06						13.3	
Mean	3.11E+06					2.16E-06	6.6	0.1

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

Mean MF	2.16E-06
Total MF	2.14E-06
SD	1.29E-06

9.1.4. Appendix A.4. K-ras bp35 G:C->T:A in smokers

Smoker I, female 41yo (9 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1	2.9E+06	2.5E+05	5.0E+05	0	0	0	0	0
LMS2	3.7E+05	2.5E+05	5.0E+05	0	0	0	0	0
LMS3	2.0E+05	2.5E+05	5.0E+05	2.9	0	0	0	0
RMS1	6.6E+05	2.5E+05	5.0E+05	1.4	0	0	0	0
RMS2	4.7E+06	1.5E+05	3.0E+05	1.3	0	0	0	0
RMS3	1.3E+05	1.0E+05	2.0E+05	0	0	0	0	0
T1	2.2E+06	4.0E+05	8.0E+05	4.1	0	0	0	0
T2	2.8E+06	2.5E+05	5.0E+05	1.5	0	0	0	0
T3	7.8E+05	3.0E+05	6.0E+05	10.8	4.8	8.0E-06	12.5	0.2
Sum	1.47E+07						12.5	0.2
Mean	1.63E+06					8.87E-07		

TK6	n=14
input (copies)	1E+06
BG	6.0
SD	6.0

Mean MF	8.87E-07
Total MF	8.49E-07
SD	2.66E-06

Smoker II, male 41yo (10 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LB5	3.86E+06	9.65E+05	1.93E+06	118.2	35.2	1.82E-05	70.4	1.1
LB7	7.48E+05	3.74E+05	7.48E+05	105.0	22.0	2.94E-05	22.0	0.3
LB13	2.12E+06	1.06E+06	2.12E+06	86.6	3.6	1.70E-06	3.6	0.1
LB17	8.51E+05	4.26E+05	8.51E+05	67.9	0	0E+00	0	0
LB19	3.29E+06	8.23E+05	1.65E+06	202.6	119.6	7.27E-05	239.2	3.7
RB1	3.71E+06	9.28E+05	1.86E+06	79.0	0	0E+00	0	0
RB9	3.81E+06	9.52E+05	1.90E+06	85.0	2.0	1.06E-06	4.0	0.1
RB15	3.14E+06	7.86E+05	1.57E+06	6.6	0	0E+00	0	0
RB17	3.04E+06	7.59E+05	1.52E+06	93.5	10.5	6.93E-06	21.0	0.3
RB19	4.04E+06	1.01E+06	2.02E+06	92.9	9.9	4.91E-06	19.8	0.3
Sum	2.86E+07						380.2	
Mean	2.86E+06					1.35E-05	38.0	0.6

TK6	n=22
input (copies)	1E+06
BG	83.0
SD	27.6

Mean MF	1.35E-05
Total MF	1.33E-05
SD	2.29E-05

9.1.5. Appendix A.5. HPRT bp508 C:G->T:A in smokers

Smoker I, female, 41yo, (9 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG/	MF (-\BG)	MTcells/sector	MTcolony#/sector
LMS1	1.75E+06	8.73E+05	1.75E+06	22.9	0	0E+00	0	0
LMS2	2.23E+05	2.23E+05	4.45E+05	227.8	203.6	4.57E-04	101.8	1.6
LMS3	1.20E+05	1.20E+05	2.41E+05	396.2	372.0	1.54E-03	186.0	2.9
RMS1	3.97E+05	3.97E+05	7.95E+05	56.7	26.5	3.34E-05	13.3	0.2
RMS2	2.81E+06	2.81E+06	5.62E+06	651.8	537.8	9.57E-05	268.9	4.2
RMS3	7.83E+04	7.83E+04	1.57E+05	23.3	0	0E+00	0	0
T1	1.30E+06	1.30E+06	2.60E+06	604.4	530.3	2.04E-04	265.2	4.1
T2	1.69E+06	1.69E+06	3.37E+06	408.9	334.8	9.93E-05	167.4	2.6
T3	4.70E+05	4.70E+05	9.39E+05	63.4	29.2	3.11E-05	14.6	0.2
Sum (n=9)	8.83E+06						1017.2	
Sum (n=7)	8.49E+06						729.4	
Mean (n=9)	9.81E+05					2.74E-04		1.8
Mean (n=7)	1.21E+06					6.62E-05		1.6

TK6 (n=17)		
input (copies)	BG	SD
5E+05	24.2	12.5
8E+05	30.2	
1E+06	**34.2**	18.2
3E+06	74.1	
5E+06	114.0	

n=9	Mean MF	2.74E-04
	Total MF	1.15E-04
	SD	4.98E-04
n=7	Mean MF	6.62E-05
MF<4E-4	Total MF	8.59E-05
	SD	7.30E-05

Smoker II, male, 41yo, (10 sectors)

Sample	Total cells	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG/assay	MF (-BG)	MTcells/sector	MTcolony#/sector
LB5	3.23E+05	3.23E+05	6.46E+05	70.8	39.1	6.05E-05	19.5	0.3
LB7	4.98E+05	4.98E+05	9.96E+05	44.9	0	0E+00	0	0
LB13	6.03E+05	6.03E+05	1.21E+06	101	54.6	4.53E-05	27.3	0.4
LB17	4.09E+05	4.09E+05	8.19E+05	65	25.9	3.17E-05	13.0	0.2
LB19	2.75E+05	2.75E+05	5.50E+05	125.5	93.8	1.70E-04	46.9	0.7
RB1	5.07E+05	5.07E+05	1.01E+06	126.7	80.3	7.92E-05	40.1	0.6
RB9	2.13E+06	2.13E+06	4.25E+06	68.9	22.5	5.29E-06	11.2	0.2
RB15	5.75E+05	5.75E+05	1.15E+06	22.3	0	0E+00	0	0
RB17	6.05E+05	6.05E+05	1.21E+06	43.9	0	0E+00	0	0
RB19	3.14E+05	3.14E+05	6.28E+05	46	14.3	2.27E-05	7.1	0.1
Sum	6.24E+06						165.2	
Mean	6.24E+05					4.15E-05	16.5	0.3

TK6 (n=17)		
input (copies)	BG	SD
5E+05	**28.1**	15.5
6E+05	31.7	23.4
7E+05	35.4	
8E+05	39.1	
1E+06	**46.4**	

Mean MF	4.15E-05
Total MF	2.65E-05
SD	5.31E-05

Smoker IV, male, 55yo, (49 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LLL2	4.51E+05	1.88E+05	3.76E+05	5.7	0	0	0	0
LLL3	4.73E+05	1.97E+05	3.95E+05	1.3	0	0	0	0
LLL4	5.62E+05	2.34E+05	4.68E+05	6.8	0	0	0	0
LLL4*	9.99E+05	3.33E+05	6.66E+05	14	2.9	4.32E-06	4.3	0.1
LLL6	6.17E+05	2.57E+05	5.15E+05	15.3	4.2	8.12E-06	5.0	0.1
LLL7	5.35E+05	2.23E+05	4.46E+05	14.5	3.4	7.57E-06	4.1	0.1
LLL7*	2.87E+06	6.71E+05	1.34E+06	95.4	84.3	6.28E-05	180.6	2.8
LLL9	2.26E+06	5.27E+05	1.05E+06	9.2	0	0	0	0
LLL11	3.36E+06	5.60E+05	1.12E+06	11.9	0.8	6.94E-07	2.3	0.0
LUL6	5.30E+05	2.65E+05	5.30E+05	4.9	0	0	0	0
LUL8+12	6.06E+05	2.52E+05	5.05E+05	1.4	0	0	0	0
LUL9*	3.03E+06	5.05E+05	1.01E+06	5.4	0	0	0	0
LUL11	1.35E+06	3.37E+05	6.74E+05	25.9	14.8	2.19E-05	29.6	0.5
LUL11*	3.53E+05	1.76E+05	3.53E+05	4.3	0	0	0	0
LUL12	4.40E+05	2.20E+05	4.40E+05	4.7	0	0	0	0
LUL17	6.04E+05	2.52E+05	5.03E+05	12.3	1.2	2.34E-06	1.4	0.0
LUL17*	9.47E+05	3.16E+05	6.31E+05	2	0	0	0	0
LUL18	7.73E+05	2.58E+05	5.15E+05	6.3	0	0	0	0
RLL3	2.04E+06	5.10E+05	1.02E+06	4.2	0	0	0	0
RLL4+5	2.06E+06	5.15E+05	1.03E+06	7.7	0	0	0	0
RLL7	1.64E+06	4.10E+05	8.19E+05	8.9	0	0	0	0
RLL8	1.37E+06	3.42E+05	6.84E+05	3.1	0	0	0	0
RLL9	2.56E+06	6.39E+05	1.28E+06	11.7	0.6	4.52E-07	1.2	0.0
RLL12	5.29E+05	1.76E+05	3.53E+05	13.4	2.3	6.46E-06	3.4	0.1
RLL13	1.37E+06	3.43E+05	6.86E+05	5.3	0	0	0	0
RLL14	1.46E+06	3.65E+05	7.31E+05	14.7	3.6	4.89E-06	7.2	0.1
RLL15	2.80E+06	7.00E+05	1.40E+06	16.8	5.7	4.05E-06	11.4	0.2
RLL20	1.14E+06	3.81E+05	7.62E+05	10.4	0	0	0	0
RLL21+22	5.22E+06	5.22E+05	1.04E+06	29	17.9	1.71E-05	89.4	1.4
RML1+2	1.57E+06	3.93E+05	7.86E+05	1	0	0	0	0
RML3	5.13E+05	2.14E+05	4.27E+05	13	1.9	4.40E-06	2.3	0.0
RML8,9,19,22	5.60E+05	2.80E+05	5.60E+05	3.9	0	0	0	0
RUL1*	1.47E+07	7.37E+05	1.47E+06	17	5.9	3.99E-06	58.8	0.9
RUL2	5.42E+06	5.42E+05	1.08E+06	33	21.9	2.02E-05	109.4	1.7
RUL4	1.15E+06	3.85E+05	7.69E+05	11.1	0	0	0	0
RUL4*	5.14E+05	1.71E+05	3.42E+05	9.1	0	0	0	0
RUL5	5.53E+05	1.84E+05	3.69E+05	27.7	16.6	4.50E-05	24.9	0.4
RUL7	5.69E+05	2.84E+05	5.69E+05	8.9	0	0	0	0
RUL8	5.22E+05	2.61E+05	5.22E+05	8.1	0	0	0	0
RUL9	7.51E+05	2.50E+05	5.00E+05	20.7	9.6	1.91E-05	14.4	0.2
RUL10	1.15E+06	3.83E+05	7.67E+05	2.3	0	0	0	0
RUL11	4.54E+05	2.27E+05	4.54E+05	12.9	1.8	3.92E-06	1.8	0.0
RUL12	4.88E+05	2.44E+05	4.88E+05	0.5	0	0	0	0
RUL13	6.30E+05	3.15E+05	6.30E+05	9.6	0	0	0	0
RUL14	1.28E+06	4.26E+05	8.53E+05	7	0	0	0	0
RUL15	5.99E+05	2.99E+05	5.99E+05	0.5	0	0	0	0
RUL16	6.82E+05	2.27E+05	4.54E+05	1.7	0	0	0	0
RUL17	6.15E+05	3.08E+05	6.15E+05	0.8	0	0	0	0
RUL19	1.20E+06	4.00E+05	8.00E+05	4.4	0	0	0	0
Sum	7.69E+07						551.2	
Mean	1.57E+06					4.84E-06		0.2

Tk6	n=9
input (copies)	1E+05
BG	11.1
SD	9.6

Mean MF	4.84E-06
Total MF	7.17E-06
SD	1.18E-05

Smoker V, female, 58yo, (14 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG/	MF (-BG)	MTcells/sector	MTcolony#/sector
LUL2	5.67E+05	2.13E+05	4.26E+05	15.9	0	0	0.0	0.0
LUL5	7.07E+05	1.59E+05	3.18E+05	134.4	110.5	3.47E-04	245.6	3.8
RLL1	8.63E+05	2.37E+05	4.74E+05	11.6	0	0	0.0	0.0
RLL2	8.00E+05	3.20E+05	6.40E+05	10.0	0	0	0.0	0.0
RLL9	4.31E+05	1.72E+05	3.45E+05	35.2	11.3	3.28E-05	14.1	0.2
RLL13	5.56E+05	2.22E+05	4.45E+05	171.3	143.8	3.23E-04	179.8	2.8
RUL2	4.62E+05	1.85E+05	3.70E+05	67.0	39.5	1.07E-04	49.4	0.8
RUL3	7.25E+05	2.90E+05	5.80E+05	42.4	7.8	1.35E-05	9.8	0.2
RUL4	8.24E+05	3.29E+05	6.59E+05	90.0	55.4	8.41E-05	69.3	1.1
RUL5,7,8	3.32E+05	1.66E+05	3.32E+05	9.3	0	0	0.0	0.0
RUL10+11	3.44E+05	1.72E+05	3.44E+05	58.2	34.3	9.98E-05	34.3	0.5
RUL15	2.97E+05	1.49E+05	2.97E+05	42.0	18.1	6.09E-05	18.1	0.3
Main carina	5.59E+05	2.24E+05	4.47E+05	56.8	29.3	6.56E-05	36.7	0.6
MC*	1.80E+06	1.80E+05	3.61E+05	153.6	126.1	3.50E-04	630.7	9.9
Sum	9.27E+06						1287.6	1.4
Mean	6.62E+05					1.06E-04	231.6	

TK6 (n=22)		
input (copies)	BG	SD
3E+05	23.9	
4E+05	27.5	
5E+05	31.0	
6E+05	34.6	
1E+06	**39.9**	24.7

Mean MF	1.06E-04
Total MF	1.39E-04
SD	1.32E-04

Smoker VI, male, 59yo, (4 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/assay	MF	MTcells/sector	MTcolony#/sector
T2-1#1	1.01E+06	7.56E+05	1.51E+06	37.1	5.8	3.84E-06	3.9	0.1
T2-1#2	6.68E+05	5.01E+05	1.00E+06	53	21.7	2.17E-05	14.5	0.2
T2-1#3	2.88E+05	2.16E+05	4.32E+05	94.1	71.8	1.66E-04	47.8	0.7
T2-1#4	1.65E+05	1.24E+05	2.48E+05	12.7	0.0	0	0.0	0
Sum	2.13E+06						66.2	1.0
Mean	5.32E+05					4.79E-05	16.5	0.3

Tk6	n=18	
input (co	BG	SD
5E+05	**22.3**	13.21
6E+05	24.8	
7E+05	26.6	
8E+05	28.4	
9E+05	30.2	
1E+06	**31.3**	18.04

Mean MF	4.79E-05
Total MF	3.11E-05
SD	7.94E-05

9.1.6. Appendix A.6. TP53 bp742 C:G->T:A in non-smokers

Non-smoker I, F, 38yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG/	MF	MTcells/sector	MTcolony#/sector
R9AB	2.12E+06	7.7E+05	3.85E+05	48	37.1	9.63E-05	50.9	0.8
3A	1.54E+06	7.6E+05	3.78E+05	48.6	37.7	9.96E-05	38.4	0.6
Sum	3.66E+06						89.3	
Mean	1.83E+06					9.80E-05		0.7

TK6	n=6
input (copies)	1E+06
BG	11.0
SD	7.7

Mean MF	9.80E-05
Total MF	2.44E-05
SD	2.35E-06

Non-smoker III, M, 41yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	meanMTc	MTc#-BG/	MF	MTcells/sector	MTcolony#/sector
L5A+L6C	4.04E+06	4.5E+05	9.0E+05	29	25.2	14.3	1.58E-05	64.0	1.0
		4.5E+05	9.0E+05	18.5					
		5.0E+05	1.0E+06	28.1					
L11B	7.15E+06	6.1E+05	1.0E+06	20.9	20.1	9.2	2.01E-05	54.1	0.8
		5.0E+05	1.0E+06	19.3					
		5.0E+05	1.2E+06	11.3					
Sum	1.12E+07							118.0	
Mean	5.60E+06						1.80E-05		0.9

TK6	n=6
input (copies)	1E+06
BG	11.0
SD	7.7

Mean MF	1.80E-05
Total MF	1.05E-05
SD	3.02E-06

Non-smoker IV, female, 45yo (30 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/32cells
LMS1A	2.99E+06	4.06E+05	8.11E+05	52	31.5	3.89E-05	116.4	1.8
LMS1B	9.38E+05	2.83E+05	5.67E+05	14.8	0	0.00E+00	0	0
LMS2B	6.91E+05	9.67E+04	1.93E+05	38	17.5	9.07E-05	62.7	1.0
LMS3A	3.65E+06	4.05E+05	8.09E+05	30.3	9.8	1.22E-05	44.3	0.7
LMS3B	1.40E+06	4.05E+05	8.10E+05	59	38.5	4.76E-05	66.5	1.0
LMS4	1.21E+06	4.05E+05	8.09E+05	31.2	10.7	1.33E-05	16.1	0.3
L4A	1.30E+06	4.05E+05	8.11E+05	35	14.5	1.79E-05	23.3	0.4
L4B	6.86E+05	7.55E+04	1.51E+05	38	17.5	1.16E-04	79.7	1.2
L5B	6.93E+05	1.04E+05	2.08E+05	70	49.5	2.38E-04	165.1	2.6
L7	8.55E+05	2.48E+05	4.96E+05	125	104.5	2.11E-04	180.2	2.8
L8	3.08E+06	4.05E+05	8.09E+05	48.1	27.6	3.42E-05	105.1	1.6
L9	1.63E+06	4.05E+05	8.11E+05	51	30.5	3.77E-05	61.3	1.0
L13A	1.19E+06	2.97E+05	5.93E+05	17.1	0	0.00E+00	0	0
L14A	9.81E+05	2.35E+05	4.71E+05	11.4	0	0.00E+00	0	0
L14B	1.49E+06	4.12E+05	8.24E+05	30.8	10.3	1.26E-05	18.7	0.3
L15B	1.05E+06	3.24E+05	6.48E+05	29.5	9.0	1.40E-05	14.6	0.2
T1	1.10E+06	3.02E+05	6.03E+05	52.8	32.3	5.36E-05	58.8	0.9
T4	9.58E+05	2.84E+05	5.67E+05	14.4	0	0.00E+00	0	0
T5	9.56E+05	2.83E+05	5.67E+05	45	24.5	4.33E-05	41.4	0.6
T6	9.18E+05	2.52E+05	5.05E+05	50.3	29.8	5.91E-05	54.3	0.8
T7	2.93E+06	4.04E+05	8.09E+05	40	19.5	2.42E-05	70.8	1.1
T10	8.10E+05	2.43E+05	4.86E+05	51.2	30.7	6.32E-05	51.2	0.8
R1	7.11E+05	7.11E+04	1.42E+05	21.3	0.8	5.93E-06	4.2	0.1
R2	1.98E+06	4.05E+05	8.10E+05	21.8	1.3	1.66E-06	3.3	0.1
R3	8.81E+05	3.17E+05	6.35E+05	27.4	6.9	1.09E-05	9.6	0.2
R4	1.80E+06	4.05E+05	8.10E+05	40	19.5	2.41E-05	43.3	0.7
R5	1.26E+06	4.05E+05	8.11E+05	43.7	23.2	2.87E-05	36.3	0.6
R6	1.79E+06	4.05E+05	8.10E+05	29.5	9.0	1.12E-05	20.0	0.3
R7	1.21E+06	4.05E+05	8.10E+05	72.9	52.4	6.47E-05	78.2	1.2
R8	8.10E+05	2.43E+05	4.86E+05	42	21.5	4.43E-05	35.9	0.6
Sum	4.19E+07						1461.4	
Mean	1.40E+06					4.40E-05	48.7	0.8

TK6	n=14
input (copies	1E+06
BG	20.5
SD	10.8

Mean MF	4.40E-05
Total MF	3.49E-05
SD	5.66E-05

Non-smoker VIII, female, 75yo (8 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcell#/sector	MTcolony#/32cells
R2	6.30E+06	5.00E+05	1.0E+06	61.4	40.9	4.09E-05	257.8	4.0
R3	3.80E+06	1.26E+05	2.5E+05	64.1	43.6	1.73E-04	657.2	10.3
T1	6.00E+06	5.01E+05	1.0E+06	54.3	33.8	3.38E-05	202.6	3.2
T2	3.30E+06	5.00E+05	1.0E+06	92.2	71.7	7.17E-05	236.5	3.7
T4	2.00E+06	1.62E+05	3.2E+05	44.2	23.7	7.34E-05	146.7	2.3
T5	3.20E+06	3.79E+05	7.6E+05	61.2	40.7	5.38E-05	172.1	2.7
T6	2.50E+06	4.05E+05	8.1E+05	79.5	59.0	7.29E-05	182.3	2.8
T7	4.45E+06	5.00E+05	1.0E+06	37.5	17.0	1.70E-05	75.8	1.2
Sum	3.16E+07						1931.1	30.2
Mean	3.94E+06						241.4	3.8

TK6	n=14
input (copies)	1E+06
BG	20.5
SD	10.8

Mean MF	6.71E-05
Total MF	6.12E-05
SD	4.75E-05

9.1.7. Appendix A.7. TP53 bp746 G:C->T:A in non-smokers

Non-smoker I, female, 38yo (2 sectors)

Sector	Total cells/s	Cells/assay	MTcopy#/assay	Mean MTcopy#	MTcopy-BG	MF	MTcells/sector	MTcolony#/sector
R9AB	2.12E+06	1.06E+06	22.9	17.4	1.7	8.04E-07	1.7	0.0
R9AB#2			11.9					
3A	1.54E+06	3.85E+05	66.7	66.7	51.0	6.62E-05	101.9	1.6
Sum	3.66E+06						103.6	
Mean	1.83E+06					3.35E-05	51.8	0.8

TK6	BG (copies)
n=15	15.7
SD	9.0

Mean MF	3.35E-05
Total MF	2.83E-05
SD	4.62E-05

Non-smoker II, female, 40yo (3 sectors)

Sector	Total cells/s	Cells/assay	MTcopy#/assay	MTcopy-BG/as	MF	MTcells/s	MTcolony#/sector
1D	2.67E+06	1.33E+06	33.0	17.3	6.49E-06	17.3	0.3
L6A	4.04E+06	2.02E+06	58.2	42.5	1.05E-05	42.5	0.7
R6C	3.75E+06	9.38E+05	19.0	3.3	1.77E-06	6.7	0.1
SUM	1.05E+07					66	
Mean	3.49E+06				6.26E-06	22.1	0.3

TK6	BG (copies)
n=15	15.7
SD	9.0

Mean MF	6.26E-06
Total MF	6.35E-06
SD	4.38E-06

Non-smoker III, male, 41yo (2 sectors)

Sector	Total cells/s	Cells/assay	MTcopy#/assay	Mean MTcopy#	MTcopy-BG	MF	MTcells/sector	MTcolony#/sector
L5A+L6C	4.04E+06	2.02E+06	16.5	19.8	4.1	1.00E-06	4.1	0.1
L5A+L6C#2			23.0					
L11B	7.15E+06	2.50E+06	30.0	42.1	26.4	5.28E-06	37.8	0.6
L11B#2			54.3					
Sum	1.12E+07						41.8	
Mean	5.60E+06					3.14E-06	20.9	0.3

TK6	BG (copies)
n=15	15.7
SD	9.0

Mean MF	3.14E-06
Total MF	3.74E-06
SD	3.02E-06

Non-smoker IV, F 45yo (42 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1A	3.70E+06	5.01E+05	1.00E+06	14.3	0	0	0	0
LMS1B	1.16E+06	3.50E+05	6.99E+05	187	161.7	2.3E-04	267.7	4.2
LMS2A	1.00E+06	5.00E+05	1.00E+06	83.2	57.9	5.8E-05	58.1	0.9
LMS2B	8.53E+05	4.00E+05	8.00E+05	165.6	140.3	1.8E-04	149.6	2.3
LMS3A	4.50E+06	5.00E+05	9.99E+05	49.3	24.0	2.4E-05	108.0	1.7
LMS3B	1.72E+06	5.00E+05	1.00E+06	104.5	79.2	7.9E-05	136.5	2.1
LMS4	1.50E+06	4.99E+05	9.99E+05	224.5	199.2	2.0E-04	298.2	4.7
L4B	8.47E+05	2.64E+05	5.28E+05	167	141.7	2.7E-04	227.4	3.6
L5A	4.71E+05	2.22E+05	4.45E+05	4.9	0	0	0	0
L5B	8.56E+05	4.00E+05	8.00E+05	101.4	76.1	9.5E-05	81.4	1.3
L7	1.06E+06	3.50E+05	7.01E+05	16.5	0	0	0	0
L8	3.80E+06	4.99E+05	9.99E+05	28	2.7	2.7E-06	10.2	0.2
L9	2.01E+06	5.00E+05	1.00E+06	115.1	89.8	9.0E-05	180.3	2.8
L10	1.07E+06	5.00E+05	1.00E+06	19.5	0	0	0	0
L12	7.37E+05	3.50E+05	7.00E+05	150	124.7	1.8E-04	131.2	2.1
L13A	1.47E+06	5.00E+05	1.00E+06	20.1	0	0	0	0
L13B	5.63E+05	2.50E+05	5.00E+05	14.4	0	0	0	0
L14A	1.21E+06	4.00E+05	8.00E+05	70.6	45.3	5.7E-05	68.5	1.1
L14B	1.84E+06	5.00E+05	1.00E+06	43.5	18.2	1.8E-05	33.4	0.5
L15A	5.29E+05	2.50E+05	4.99E+05	18.4	0	0	0	0
L15B	1.29E+06	4.00E+05	8.00E+05	73.6	48.3	6.0E-05	77.9	1.2
T1	1.35E+06	4.50E+05	9.00E+05	31.8	6.5	7.2E-06	9.7	0.2
T2	4.08E+05	2.04E+05	4.08E+05	10	0	0	0	0
T3	4.45E+05	2.23E+05	4.45E+05	12.5	0	0	0	0
T4	1.18E+06	3.50E+05	7.00E+05	56	30.7	4.4E-05	51.8	0.8
T5	1.18E+06	3.50E+05	7.00E+05	58.5	33.2	4.7E-05	56.0	0.9
T6	1.13E+06	3.50E+05	7.00E+05	96.4	71.1	1.0E-04	115.0	1.8
T7	3.62E+06	4.99E+05	9.99E+05	38.7	13.4	1.3E-05	48.5	0.8
T8	8.72E+05	4.00E+05	8.00E+05	55.6	30.3	3.8E-05	33.0	0.5
T9	6.22E+05	3.00E+05	6.00E+05	40.75	15.4	2.6E-05	16.0	0.2
T10	1.00E+06	3.00E+05	6.00E+05	38.8	13.5	2.2E-05	22.5	0.4
T11	7.26E+05	3.50E+05	7.00E+05	47.3	22.0	3.1E-05	22.8	0.4
R1	8.78E+05	4.00E+05	8.00E+05	29.2	3.9	4.8E-06	4.3	0.1
R2	2.45E+06	5.00E+05	1.00E+06	46.5	21.2	2.1E-05	51.8	0.8
R3	1.09E+06	3.50E+05	7.00E+05	13.3	0	0	0	0
R4	2.22E+06	5.00E+05	1.00E+06	32.6	7.3	7.3E-06	16.1	0.3
R5	1.56E+06	5.00E+05	1.00E+06	74.9	49.6	5.0E-05	77.3	1.2
R6	2.21E+06	5.00E+05	9.99E+05	35	9.7	9.7E-06	21.4	0.3
R7	1.49E+06	5.00E+05	1.00E+06	53.75	28.4	2.8E-05	42.4	0.7
R8	1.00E+06	3.00E+05	6.00E+05	12.8	0	0	0	0
R9	5.56E+05	2.50E+05	5.00E+05	38.6	13.3	2.7E-05	14.8	0.2
R11	6.76E+04	6.76E+04	1.35E+05	20	0	0	0	0
Sum	5.82E+07						2431.6	
Mean	1.39E+06					4.80E-05	57.9	0.9

Tk6	input	1E+06
(copies)	BG	25.3
n=20	SD	27.5

Mean MF	4.80E-05
Total MF	4.18E-05
SD	6.78E-05

Non-smoker V (50 sectors) 50 yo male

Sector	Total cells/s	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
LLL1	4.84E+06	1.61E+06	3.22E+06	2.8	0	0E+00	0	0
LLL2	2.38E+06	7.92E+05	1.58E+06	5.1	0	0E+00	0	0
LLL4	5.50E+06	1.83E+06	3.67E+06	0	0	0E+00	0	0
LLL5	2.38E+06	7.92E+05	1.58E+06	4.6	0	0E+00	0	0
LLL6+7	1.78E+06	5.92E+05	1.18E+06	3.3	0	0E+00	0	0
LLL8	4.65E+06	1.55E+06	3.10E+06	12.1	3.9	1.24E-06	5.8	0.1
LLL9+10	2.03E+06	6.75E+05	1.35E+06	78.6	70.4	5.21E-05	105.5	1.6
LLL11	2.98E+06	9.94E+05	1.99E+06	0	0	0E+00	0	0
LLL12	8.08E+05	2.69E+05	5.39E+05	2.2	0	0E+00	0	0
LLL13	2.62E+06	8.74E+05	1.75E+06	9.6	1.4	7.76E-07	2.0	0.0
LLL14+15	1.27E+06	4.24E+05	8.49E+05	42.1	33.9	3.99E-05	50.8	0.8
LUL1	3.55E+06	1.18E+06	2.37E+06	9.1	0.9	3.62E-07	1.3	0.0
LUL2	2.75E+06	9.17E+05	1.83E+06	14.4	6.2	3.36E-06	9.2	0.1
LUL3	3.29E+06	1.10E+06	2.19E+06	0.27	0	0E+00	0	0
LUL4	3.04E+06	1.01E+06	2.03E+06	0.81	0	0E+00	0	0
LUL5	5.10E+06	1.70E+06	3.40E+06	1.4	0	0E+00	0	0
LUL7	1.26E+06	4.21E+05	8.43E+05	0.13	0	0E+00	0	0
LUL8+9	3.33E+06	1.11E+06	2.22E+06	41.1	32.9	1.48E-05	49.3	0.8
LUL10	1.51E+06	5.04E+05	1.01E+06	9.2	1.0	9.49E-07	1.4	0.0
LUL11	2.83E+06	9.42E+05	1.88E+06	15.3	7.1	3.74E-06	10.6	0.2
LUL12	3.01E+06	1.00E+06	2.01E+06	9.5	1.3	6.27E-07	1.9	0.0
LUL15+16	4.38E+06	1.46E+06	2.92E+06	491	482.8	1.65E-04	724.1	11.3
LUL17+18	9.29E+06	3.10E+06	6.20E+06	588	579.8	9.36E-05	869.6	13.6
RLL1	8.80E+05	2.93E+05	5.86E+05	1.1	0	0E+00	0	0
RLL2	4.87E+06	1.62E+06	3.24E+06	13	4.8	1.47E-06	7.1	0.1
RLL3	2.82E+06	9.40E+05	1.88E+06	0	0	0E+00	0	0
RLL4	2.29E+06	7.64E+05	1.53E+06	6	0	0E+00	0	0
RLL5	7.57E+06	2.52E+06	5.05E+06	2.1	0	0E+00	0	0
RLL6	6.90E+06	2.30E+06	4.60E+06	10.9	2.7	5.77E-07	4.0	0.1
RLL7	1.51E+06	5.02E+05	1.00E+06	14	5.8	5.73E-06	8.6	0.1
RLL8	1.92E+06	6.40E+05	1.28E+06	19.6	11.4	8.88E-06	17.0	0.3
RLL9	1.69E+06	5.64E+05	1.13E+06	31	22.8	2.02E-05	34.1	0.5
RLL10+12	1.55E+06	5.15E+05	1.03E+06	73.5	65.3	6.33E-05	97.9	1.5
RLL13	1.11E+06	3.70E+05	7.40E+05	37	28.8	3.88E-05	43.1	0.7
RLL14	2.73E+06	9.11E+05	1.82E+06	28.8	20.6	1.13E-05	30.8	0.5
RLL18	3.61E+06	1.20E+06	2.41E+06	15.6	7.4	3.06E-06	11.0	0.2
RLL19	2.30E+06	7.65E+05	1.53E+06	8.2	0	0E+00	0	0
RLL22	1.07E+05	3.58E+04	7.15E+04	63.8	55.6	7.77E-04	83.3	1.3
RUL1	1.28E+06	4.27E+05	8.55E+05	70.6	62.4	7.30E-05	93.5	1.5
RUL2	1.58E+06	5.25E+05	1.05E+06	23.1	14.9	1.41E-05	22.3	0.3
RUL3	2.19E+06	7.31E+05	1.46E+06	20	11.8	8.04E-06	17.6	0.3
RUL4	9.49E+05	3.16E+05	6.33E+05	22.8	14.6	2.30E-05	21.8	0.3
RUL10	2.64E+06	8.81E+05	1.76E+06	149	140.8	7.99E-05	211.1	3.3
RUL11	2.07E+06	6.91E+05	1.38E+06	16.2	8.0	5.76E-06	11.9	0.2
RMS	8.26E+05	2.75E+05	5.51E+05	23	14.8	2.68E-05	22.1	0.3
RMS*	2.43E+06	8.11E+05	1.62E+06	76.2	68.0	4.19E-05	101.9	1.6
MC	1.19E+06	3.98E+05	7.96E+05	42.3	34.1	4.28E-05	51.1	0.8
LMS	7.67E+05	2.56E+05	5.12E+05	21.4	13.2	2.57E-05	19.7	0.3
TRAC	2.44E+06	8.14E+05	1.63E+06	17.1	8.9	5.44E-06	13.3	0.2
T*	2.13E+06	7.12E+05	1.42E+06	0.76	0	0E+00	0	0
Sum (n=50)	1.37E+08						2755	
Sum (n=49)	1.37E+08						2672	
Mean (n=50)	2.74E+06					3.31E-05		0.9
Mean (n=49)	2.79E+06					1.79E-05		0.9

TK6	BG (copies)
n=7	8.2
SD	1.8

n=50	Mean MF	3.31E-05
all sectors	Total MF	2.01E-05
	SD	1.12E-04
n=49	Mean MF	1.79E-05
MF<4e-4	Total MF	1.95E-05
	SD	3.18E-05

Non-smoker VI, F 59yo, 52 sectors

Sector	Total cells/s	Cells/assay	copy#/sector	MTcopy#/assay	MTc#-BG	MF	MTcell#/sector	MTcolony#/sector
LLL12	5.43E+05	1.81E+05	3.62E+05	2148	2132	5.89E-03	3198	50.0
LLL13	1.10E+06	5.49E+05	1.10E+06	0	0	0E+00	0	0
LLL14-1	9.69E+05	4.85E+05	9.69E+05	82	66	6.80E-05	66	1.0
LLL14-2	1.33E+05	4.44E+04	8.88E+04	0	0	0E+00	0	0
LLL14-3	4.98E+05	1.66E+05	3.32E+05	792	776	2.34E-03	1164	18.2
LLL15	6.69E+06	3.35E+06	6.69E+06	42	26	3.93E-06	26	0.4
LLL2	2.53E+05	8.43E+04	1.69E+05	34	18	1.06E-04	27	0.4
LLL3	9.33E+05	3.11E+05	6.22E+05	73	57	9.15E-05	85	1.3
LLL4	1.17E+05	3.92E+04	7.83E+04	2	0	0E+00	0	0
LLL5+6	1.37E+05	4.55E+04	9.10E+04	4	0	0E+00	0	0
LLL7	4.94E+05	1.65E+05	3.29E+05	87	71	2.16E-04	107	1.7
LMS1	4.36E+06	1.45E+06	2.91E+06	88	73	2.49E-05	109	1.7
LMS2	9.20E+05	3.07E+05	6.14E+05	146	130	2.12E-04	195	3.0
LMS3	1.76E+06	5.85E+05	1.17E+06	76	60	5.12E-05	90	1.4
LMS4	3.47E+06	1.16E+06	2.31E+06	11640	11624	5.02E-03	17436	272.4
LUL1+LLL2	8.11E+06	2.70E+06	5.40E+06	4	0	0E+00	0	0
LUL10	6.66E+05	2.22E+05	4.44E+05	1	0	0E+00	0	0
LUL11	2.33E+05	7.77E+04	1.55E+05	13	0	0E+00	0	0
LUL1-1	1.82E+06	6.06E+05	1.21E+06	28	13	1.03E-05	19	0.3
LUL3	2.97E+05	9.90E+04	1.98E+05	12	0	0E+00	0	0
LUL4+5	1.19E+05	3.96E+04	7.91E+04	1	0	0E+00	0	0
LUL6	4.34E+04	1.45E+04	2.89E+04	3	0	0E+00	0	0
LUL7	2.28E+05	7.59E+04	1.52E+05	16	0	0E+00	0	0
LUL9+14	3.46E+05	1.15E+05	2.30E+05	5	0	0E+00	0	0
RLL17	5.94E+06	2.97E+06	5.94E+06	60	44	7.46E-06	44	0.7
RLL18	8.19E+05	2.73E+05	5.46E+05	197	181	3.32E-04	272	4.2
RLL19+20	3.32E+05	1.11E+05	2.21E+05	80	64	2.90E-04	96	1.5
RLL2-1	7.30E+06	2.43E+06	4.87E+06	127	111	2.28E-05	166	2.6
RLL2-2	1.94E+06	6.47E+05	1.29E+06	258	242	1.87E-04	363	5.7
RLL3	3.13E+06	1.04E+06	2.09E+06	189	173	8.30E-05	260	4.1
RLL4	1.41E+06	4.69E+05	9.38E+05	82	67	7.09E-05	100	1.6
RLL6	9.41E+05	3.14E+05	6.27E+05	157	141	2.25E-04	211	3.3
RLL7	1.09E+05	3.64E+04	7.29E+04	0	0	0E+00	0	0
RLL8+9	1.82E+05	6.07E+04	1.21E+05	37	22	1.77E-04	32	0.5
RMS1	8.73E+06	2.91E+06	5.82E+06	2712	2696	4.63E-04	4044	63.2
RMS2+RLL1	3.84E+06	1.28E+06	2.56E+06	39	23	9.01E-06	35	0.5
RUL1	1.87E+06	6.22E+05	1.24E+06	40	24	1.92E-05	36	0.6
RUL10+11	1.25E+05	4.18E+04	8.36E+04	2	0	0E+00	0	0
RUL13	9.68E+05	3.23E+05	6.46E+05	185	169	2.62E-04	254	4.0
RUL2	6.74E+04	2.25E+04	4.49E+04	27	11	2.52E-04	17	0.3
RUL3	9.53E+04	3.18E+04	6.35E+04	22	7	1.02E-04	10	0.2
RUL4	1.76E+06	5.86E+05	1.17E+06	104	88	7.51E-05	132	2.1
RUL5+7	1.66E+05	5.54E+04	1.11E+05	13	0	0E+00	0	0
T1	1.09E+06	3.63E+05	7.27E+05	23	7	9.94E-06	11	0.2
T10	1.17E+06	3.91E+05	7.82E+05	57	41	5.28E-05	62	1.0
T11	3.09E+06	1.03E+06	2.06E+06	192	176	8.56E-05	264	4.1
T12	1.34E+06	4.45E+05	8.90E+05	90	74	8.34E-05	111	1.7
T13	3.16E+06	1.05E+06	2.10E+06	99	83	3.96E-05	125	2.0
T14	9.65E+05	3.22E+05	6.44E+05	84	68	1.06E-04	102	1.6
T2	2.99E+06	1.50E+06	2.99E+06	13	0	0E+00	0	0
T5	1.82E+05	6.06E+04	1.21E+05	2	0	0E+00	0	0
T9	9.65E+05	3.22E+05	6.43E+05	276	260	4.05E-04	390	6.1
Sum (n=52)	8.89E+07						2.97E+04	
Sum (n=47)	7.47E+07						3.43E+03	
Mean (n=52)	1.71E+06					3.35E-04		8.9
Mean (n=47)	1.59E+06					6.97E-05		1.1

TK6	BG (copies)
n=15	15.7
SD	9.0

n=52	Mean MF	3.35E-04
all sectors	Total MF	3.34E-04
	SD	1.09E-03
n=47	Mean MF	6.97E-05
MF<4E-4	Total MF	4.59E-05
	SD	9.24E-05

Non-smoker VII, male, 67yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MF mean	MTcells/sector	mean MTcells/	MTcolony#/sector
Bronchus	5.15E+06	1.72E+06	3.43E+06	235	209.7	6.11E-05	5.56E-05	314.5	286.1	4.5
Bronchus#2		5.00E+05	1.00E+06	75.4	50.1	5.01E-05		257.7		
Trachea	1.70E+07	2.83E+06	5.66E+06	80.3	55.0	9.71E-06	7.80E-06	164.9	132.4	2.1
Trachea#2		5.00E+05	1.00E+06	31.2	5.9	5.88E-06		99.8		
Sum	2.21E+07								418.5	6.5
Mean	1.11E+07					3.17E-05			209.2	3.3

Tk6	input	1E+06
(copies)	BG	25.3
n=20	SD	27.5

Mean MF	3.17E-05
Total MF	1.89E-05
SD	2.80E-05

Non-smoker VIII, female, 75yo (10 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG/	MF (-BG)	MTcell#/sector	MTcolony#/sector
R1	3.30E+06	5.00E+05	1.0E+06	37	24.6	2.46E-05	81.1	1.3
R2	6.30E+06	1.00E+06	2.0E+06	32.4	20.0	9.99E-06	62.9	1.0
R3	3.80E+06	1.00E+06	2.0E+06	34.8	22.4	1.12E-05	42.5	0.7
T1	6.00E+06	1.00E+06	2.0E+06	59	46.6	2.33E-05	139.7	2.2
T2	3.30E+06	1.00E+06	2.0E+06	26.4	14.0	6.99E-06	23.1	0.4
T3	2.93E+06	1.00E+06	2.0E+06	39.1	26.7	1.33E-05	39.1	0.6
T4	2.00E+06	1.00E+06	2.0E+06	21.6	9.2	4.59E-06	9.2	0.1
T5	3.20E+06	5.00E+05	1.0E+06	18.2	5.8	5.77E-06	18.5	0.3
T6	2.50E+06	5.00E+05	1.0E+06	26.8	14.4	1.44E-05	35.9	0.6
T7	4.45E+06	1.00E+06	2.0E+06	23.7	11.3	5.64E-06	25.1	0.4
Sum	3.78E+07						477.0	7.5
Mean	3.78E+06					1.20E-05	47.7	0.7

Tk6	Input copy	1E+06
	Mean	12.4
	SD(n=14)	5.6

Mean MF	1.20E-05
Total MF	1.26E-05
1SD	7.13E-06

Non-smoker IX, Female, 76yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MF mean	MTcells/sector	mean MTcells/	MTcolony#/sector
Trachea	7.81E+06	2.60E+06	5.21E+06	88.5	63.2	1.21E-05	1.20E-05	94.8	93.4	1.5
Trachea#2		5.00E+05	1.00E+06	37.1	11.8	1.18E-05		92.0		
Trachea-2	5.24E+06	1.75E+06	3.49E+06	50	24.7	7.07E-06	1.72E-05	37.0	90.2	1.4
Trachea-2#1		5.00E+05	1.00E+06	52.7	27.4	2.74E-05		143.3		
Sum	1.30E+07								183.5	2.9
Mean	6.52E+06					1.46E-05			91.8	1.4

Tk6	input	1E+06
(copies)	BG	25.3
n=20	SD	27.5

Mean MF	1.46E-05
Total MF	1.41E-05
SD	8.83E-06

9.1.8. Appendix A.8. TP53 bp747 G:C->T:A in non-smokers

Non-smoker I, female, 38yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTcopy#-BG	MF	MTcells/sector	MTcolony#/sector
R9AB	2.12E+06	2.12E+06	4.23E+06	144.8	136.0	3.22E-05	68.0	1.1
3A	1.54E+06	7.70E+05	1.54E+06	151.2	142.5	9.25E-05	142.5	2.2
Sum	3.66E+06						210.5	
Mean	1.83E+06					6.23E-05	105.2	1.6

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

Mean MF	6.23E-05
Total MF	5.76E-05
SD	4.27E-05

Non-smoker II, female, 40yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTcopy#-BG	MF	MTcells/sector	MTcolony#/sector
1D	2.67E+06	2.67E+06	5.33E+06	30.8	22.1	4.15E-06	11.0	0.2
L6A	4.04E+06	4.04E+06	8.08E+06	40.7	31.9	3.95E-06	16.0	0.2
Sum	6.71E+06						27.0	
Mean	3.35E+06					4.05E-06	13.5	0.2

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

Mean MF	4.05E-06
Total MF	8.05E-06
SD	1.39E-07

Non-smoker III, male, 41yo (2 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTcopy#-BG	MF	MTcells/sector	MTcolony#/sector
L5A+L6C	4.04E+06	4.04E+06	8.08E+06	0.0	0	0	0	0
L11B	7.15E+06	5.01E+06	1.00E+07	53.9	45.1	4.51E-06	32.2	0.5
Sum	1.12E+07						32.2	
Mean	5.60E+06					2.25E-06	16.1	0.3

Tk6	(n=5)
input	1E+06
BG (copies)	8.7
SD	7.3

Mean MF	2.25E-06
Total MF	2.88E-06
SD	3.19E-06

Non-smoker IV, F 45yo (41 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG/	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1A	3.70E+06	5.01E+05	1.0E+06	19.4	0	0E+00	0	0
LMS1B	1.16E+06	3.50E+05	7.0E+05	46.9	7.6	1.09E-05	12.6	0.2
LMS2A	1.00E+06	5.00E+05	1.0E+06	85	45.7	4.57E-05	45.9	0.7
LMS2B	8.53E+05	4.00E+05	8.0E+05	51.2	11.9	1.49E-05	12.7	0.2
LMS3A	4.50E+06	5.00E+05	1.0E+06	6.9	0	0E+00	0	0
LMS3B	1.72E+06	5.00E+05	1.0E+06	36.6	0	0E+00	0	0
LMS4	1.50E+06	4.99E+05	1.0E+06	45	5.7	5.73E-06	8.6	0.1
L4A	1.61E+06	5.00E+05	1.0E+06	20	0	0E+00	0	0
L4B	8.47E+05	4.00E+05	8.0E+05	28.4	0	0E+00	0	0
L5B	8.56E+05	4.00E+05	8.0E+05	86.9	47.6	5.95E-05	50.9	0.8
L7	1.06E+06	3.50E+05	7.0E+05	75	35.7	5.10E-05	53.8	0.8
L8	3.80E+06	4.99E+05	1.0E+06	24.6	0	0E+00	0	0
L9	2.01E+06	5.00E+05	1.0E+06	24.6	0	0E+00	0	0
L10	1.07E+06	5.00E+05	1.0E+06	25.4	0	0E+00	0	0
L12	7.37E+05	3.50E+05	7.0E+05	53	13.7	1.96E-05	14.4	0.2
L13A	1.47E+06	5.00E+05	1.0E+06	53	13.7	1.37E-05	20.1	0.3
L13B	5.63E+05	2.50E+05	5.0E+05	27.4	0	0E+00	0	0
L14A	1.21E+06	4.00E+05	8.0E+05	20	0	0E+00	0	0
L14B	1.84E+06	5.00E+05	1.0E+06	50	10.7	1.07E-05	19.7	0.3
L15A	5.29E+05	2.50E+05	5.0E+05	17.2	0	0E+00	0	0
L15B	1.29E+06	4.00E+05	8.0E+05	40.2	0.9	1.15E-06	1.5	0.0
T1	1.35E+06	4.50E+05	9.0E+05	63.4	24.1	2.68E-05	36.3	0.6
T2	4.08E+05	2.04E+05	4.1E+05	25.4	0	0E+00	0	0
T3	4.45E+05	2.23E+05	4.5E+05	157.2	117.9	2.65E-04	117.9	1.8
T4	1.18E+06	3.50E+05	7.0E+05	23.1	0	0E+00	0	0
T5	1.18E+06	3.50E+05	7.0E+05	14.4	0	0E+00	0	0
T6	1.13E+06	3.50E+05	7.0E+05	17.8	0	0E+00	0	0
T7	3.62E+06	4.99E+05	1.0E+06	23.2	0	0E+00	0	0
T8	8.72E+05	4.00E+05	8.0E+05	106.7	67.4	8.43E-05	73.5	1.1
T9	6.22E+05	3.00E+05	6.0E+05	50.2	10.9	1.82E-05	11.3	0.2
T10	1.00E+06	3.00E+05	6.0E+05	28.4	0	0E+00	0	0
T11	7.26E+05	3.50E+05	7.0E+05	191.4	152.1	2.17E-04	157.8	2.5
R1	8.78E+05	4.00E+05	8.0E+05	228.3	189.0	2.36E-04	207.3	3.2
R2	2.45E+06	5.00E+05	1.0E+06	90.3	51.0	5.10E-05	124.7	1.9
R3	1.09E+06	3.50E+05	7.0E+05	31.4	0	0E+00	0	0
R4	2.22E+06	5.00E+05	1.0E+06	105.6	66.3	6.63E-05	147.1	2.3
R5	1.56E+06	5.00E+05	1.0E+06	263.3	224.0	2.24E-04	349.5	5.5
R6	2.21E+06	5.00E+05	1.0E+06	78.3	39.0	3.90E-05	86.3	1.3
R7	1.49E+06	5.00E+05	1.0E+06	45.8	6.5	6.52E-06	9.7	0.2
R8	1.00E+06	3.00E+05	6.0E+05	31.6	0	0E+00	0	0
R9	5.56E+05	2.50E+05	5.0E+05	16.8	0	0E+00	0	0
Sum	5.93E+07						1562	
Mean	1.45E+06					3.58E-05	38.1	0.6

Tk6	(n=5)
input	1E+06
BG (copies)	39.3
SD	10.9

Mean MF	3.58E-05
Total MF	2.63E-05
SD	7.02E-05

Non-smoker V, 50yo, M (50 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
LLL1	4.84E+06	1.61E+06	3.22E+06	24.5	15.8	4.89E-06	23.6	0.4
LLL2	2.38E+06	7.92E+05	1.58E+06	35	26.3	1.66E-05	39.4	0.6
LLL4	5.50E+06	1.83E+06	3.67E+06	6.7	0	0E+00	0	0
LLL5	2.38E+06	7.92E+05	1.58E+06	23.4	14.7	9.25E-06	22.0	0.3
LLL6+7	1.78E+06	5.92E+05	1.18E+06	22	13.3	1.12E-05	19.9	0.3
LLL8	4.65E+06	1.55E+06	3.10E+06	24	15.3	4.92E-06	22.9	0.4
LLL9+10	2.03E+06	6.75E+05	1.35E+06	111	102.3	7.57E-05	153.4	2.4
LLL11	2.98E+06	9.94E+05	1.99E+06	8.1	0	0E+00	0	0
LLL12	8.08E+05	2.69E+05	5.39E+05	18	9.3	1.72E-05	13.9	0.2
LLL13	2.62E+06	8.74E+05	1.75E+06	9.6	0.9	4.88E-07	1.3	0.0
LLL14+15	1.27E+06	4.24E+05	8.49E+05	74.3	65.6	7.72E-05	98.3	1.5
LUL1	3.55E+06	1.18E+06	2.37E+06	23.7	15.0	6.32E-06	22.4	0.4
LUL2	2.75E+06	9.17E+05	1.83E+06	20	11.3	6.14E-06	16.9	0.3
LUL3	3.29E+06	1.10E+06	2.19E+06	0	0	0E+00	0	0
LUL4	3.04E+06	1.01E+06	2.03E+06	14	5.3	2.59E-06	7.9	0.1
LUL5	5.10E+06	1.70E+06	3.40E+06	6.3	0	0E+00	0	0
LUL7	1.26E+06	4.21E+05	8.43E+05	5.8	0	0E+00	0	0
LUL8+9	3.33E+06	1.11E+06	2.22E+06	76	67.3	3.03E-05	100.9	1.6
LUL10	1.51E+06	5.04E+05	1.01E+06	11.3	2.6	2.53E-06	3.8	0.1
LUL11	2.83E+06	9.42E+05	1.88E+06	29.9	21.2	1.12E-05	31.7	0.5
LUL12	3.01E+06	1.00E+06	2.01E+06	18.3	9.6	4.76E-06	14.3	0.2
LUL15+16	4.38E+06	1.46E+06	2.92E+06	16.3	7.6	2.58E-06	11.3	0.2
LUL17+18	9.29E+06	3.10E+06	6.20E+06	19.8	11.1	1.78E-06	16.6	0.3
RLL1	8.80E+05	2.93E+05	5.86E+05	2.1	0	0E+00	0	0
RLL2	4.87E+06	1.62E+06	3.24E+06	2.4	0	0E+00	0	0
RLL3	2.82E+06	9.40E+05	1.88E+06	0	0	0E+00	0	0
RLL4	2.29E+06	7.64E+05	1.53E+06	13.5	4.8	3.11E-06	7.1	0.1
RLL5	7.57E+06	2.52E+06	5.05E+06	10.5	1.8	3.47E-07	2.6	0.0
RLL6	6.90E+06	2.30E+06	4.60E+06	7.6	0	0E+00	0	0
RLL7	1.51E+06	5.02E+05	1.00E+06	12	3.3	3.24E-06	4.9	0.1
RLL8	1.92E+06	6.40E+05	1.28E+06	70	61.3	4.79E-05	91.9	1.4
RLL9	1.69E+06	5.64E+05	1.13E+06	9	0.3	2.25E-07	0.4	0.0
RLL10+12	1.55E+06	5.15E+05	1.03E+06	179	170.3	1.65E-04	255.4	4.0
RLL13	1.11E+06	3.70E+05	7.40E+05	2.5	0	0E+00	0	0
RLL14	2.73E+06	9.11E+05	1.82E+06	32.3	23.6	1.29E-05	35.3	0.6
RLL18	3.61E+06	1.20E+06	2.41E+06	3.7	0	0E+00	0	0
RLL19	2.30E+06	7.65E+05	1.53E+06	1.5	0	0E+00	0	0
RLL22	1.07E+05	3.58E+04	7.15E+04	21	12.3	1.71E-04	18.4	0.3
RUL1	1.28E+06	4.27E+05	8.55E+05	25.4	16.7	1.95E-05	25.0	0.4
RUL2	1.58E+06	5.25E+05	1.05E+06	8.3	0	0E+00	0	0
RUL3	2.19E+06	7.31E+05	1.46E+06	7.5	0	0E+00	0	0
RUL4	9.49E+05	3.16E+05	6.33E+05	1.4	0	0E+00	0	0
RUL10	2.64E+06	8.81E+05	1.76E+06	13	4.3	2.41E-06	6.4	0.1
RUL11	2.07E+06	6.91E+05	1.38E+06	1.3	0	0E+00	0	0
RMS	8.26E+05	2.75E+05	5.51E+05	4.5	0	0E+00	0	0
RMS*	2.43E+06	8.11E+05	1.62E+06	60	51.3	3.16E-05	76.9	1.2
MC	1.19E+06	3.98E+05	7.96E+05	7.2	0	0E+00	0	0
LMS	7.67E+05	2.56E+05	5.12E+05	1.3	0	0E+00	0	0
TRAC	2.44E+06	8.14E+05	1.63E+06	7.4	0	0E+00	0	0
T*	2.13E+06	7.12E+05	1.42E+06	0	0	0E+00	0	0
Sum	1.37E+08						1.14E+03	
Mean	2.74E+06					1.49E-05		0.4

Tk6	(n=5)
input	1E+06
BG (copies	8.7
SD	7.3

Mean MF	1.49E-05
Total MF	8.36E-06
SD	3.59E-05

Non-smoker VI, F 59yo (52 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcell#/sector	MTcolony#/sector
LLL12	5.43E+05	1.81E+05	3.62E+05	89.0	80.3	2.22E-04	120.4	1.9
LLL13	1.10E+06	5.49E+05	1.10E+06	117.6	108.9	9.92E-05	108.9	1.7
LLL14-1	9.69E+05	4.85E+05	9.69E+05	42.8	34.1	3.52E-05	34.1	0.5
LLL14-2	1.33E+05	6.66E+04	1.33E+05	36.1	27.4	2.06E-04	27	0.4
LLL14-3	4.98E+05	2.49E+05	4.98E+05	59.6	50.9	1.02E-04	50.9	0.8
LLL15	6.69E+06	3.35E+06	6.69E+06	1.1	0	0E+00	0	0
LLL2	2.53E+05	8.43E+04	1.69E+05	5.9	0	0E+00	0	0
LLL3	9.33E+05	3.11E+05	6.22E+05	92.4	83.7	1.34E-04	125.5	2.0
LLL4	1.17E+05	3.92E+04	7.83E+04	1.4	0	0E+00	0	0
LLL5+6	1.37E+05	4.55E+04	9.10E+04	10.9	2.2	2.39E-05	3	0.1
LLL7	4.94E+05	1.65E+05	3.29E+05	0.0	0	0E+00	0	0
LMS1	4.36E+06	1.45E+06	2.91E+06	11.8	3.0	1.04E-06	5	0.1
LMS2	9.20E+05	3.07E+05	6.14E+05	0.8	0	0E+00	0	0
LMS3	1.76E+06	5.85E+05	1.17E+06	12.6	3.9	3.29E-06	6	0.1
LMS4	3.47E+06	1.16E+06	2.31E+06	7.6	0	0E+00	0	0
LUL1+LLL2	8.11E+06	2.70E+06	5.40E+06	0.0	0	0E+00	0	0
LUL10	6.66E+05	2.22E+05	4.44E+05	0.0	0	0E+00	0	0
LUL11	2.33E+05	7.77E+04	1.55E+05	21.0	12.3	7.89E-05	18	0.3
LUL1-1	1.82E+06	6.06E+05	1.21E+06	37.8	29.1	2.40E-05	44	0.7
LUL3	2.97E+05	9.90E+04	1.98E+05	27.7	19.0	9.59E-05	28	0.4
LUL4+5	1.19E+05	3.96E+04	7.91E+04	14.3	5.5	6.99E-05	8	0.1
LUL6	4.34E+04	1.45E+04	2.89E+04	5.0	0	0E+00	0	0
LUL7	2.28E+05	7.59E+04	1.52E+05	5.9	0	0E+00	0	0
LUL9+14	3.46E+05	1.15E+05	2.30E+05	9.2	0.5	2.14E-06	1	0.0
RLL17	5.94E+06	1.98E+06	3.96E+06	0.3	0	0E+00	0	0
RLL18	8.19E+05	2.73E+05	5.46E+05	42.0	33.3	6.09E-05	49.9	0.8
RLL19+20	3.32E+05	1.11E+05	2.21E+05	37.0	28.2	1.27E-04	42	0.7
RLL2-1	7.30E+06	2.43E+06	4.87E+06	45.4	36.6	7.52E-06	54.9	0.9
RLL2-2	1.94E+06	6.47E+05	1.29E+06	205.8	197.1	1.52E-04	295.6	4.6
RLL3	3.13E+06	1.04E+06	2.09E+06	37.8	29.1	1.39E-05	44	0.7
RLL4	1.41E+06	4.69E+05	9.38E+05	12.6	3.9	4.11E-06	6	0.1
RLL6	9.41E+05	3.14E+05	6.27E+05	36.1	27.4	4.37E-05	41	0.6
RLL7	1.09E+05	3.64E+04	7.29E+04	2.5	0	0E+00	0	0
RLL8+9	1.82E+05	6.07E+04	1.21E+05	26.5	17.8	1.47E-04	27	0.4
RMS1	8.73E+06	2.91E+06	5.82E+06	0.0	0	0E+00	0	0
RMS2+RLL1	3.84E+06	1.28E+06	2.56E+06	30.2	21.5	8.39E-06	32	0.5
RUL1	1.87E+06	6.22E+05	1.24E+06	77.3	68.5	5.51E-05	102.8	1.6
RUL10+11	1.25E+05	4.18E+04	8.36E+04	0.8	0	0E+00	0	0
RUL13	9.68E+05	3.23E+05	6.46E+05	249.5	240.7	3.73E-04	361.1	5.6
RUL2	6.74E+04	2.25E+04	4.49E+04	22.7	13.9	3.10E-04	21	0.3
RUL3	9.53E+04	3.18E+04	6.35E+04	9.2	0.5	7.77E-06	1	0.0
RUL4	1.76E+06	5.86E+05	1.17E+06	79.0	70.2	5.99E-05	105.3	1.6
RUL5+7	1.66E+05	5.54E+04	1.11E+05	0.0	0	0E+00	0	0
T1	1.09E+06	3.63E+05	7.27E+05	19.3	10.6	1.46E-05	16	0.2
T10	1.17E+06	3.91E+05	7.82E+05	11.8	3.0	3.85E-06	5	0.1
T11	3.09E+06	1.03E+06	2.06E+06	537.6	528.9	2.57E-04	793.3	12.4
T12	1.34E+06	4.45E+05	8.90E+05	77.3	68.5	7.70E-05	102.8	1.6
T13	3.16E+06	1.05E+06	2.10E+06	89.9	81.1	3.86E-05	121.7	1.9
T14	9.65E+05	3.22E+05	6.44E+05	121.0	112.2	1.74E-04	168.3	2.6
T2	2.99E+06	9.98E+05	2.00E+06	2.5	0	0E+00	0	0
T5	1.82E+05	6.06E+04	1.21E+05	7.6	0	0E+00	0	0
T9	9.65E+05	3.22E+05	6.43E+05	30.2	21.5	3.34E-05	32	0.5
Sum	8.89E+07						3001.8	
Mean	1.71E+06					5.90E-05	57.7	0.9

Tk6	**(n=5)**
input	1E+06
BG (copies)	**8.7**
SD	7.3

n=52	**mean MF**	**5.90E-05**
	total MF	**3.38E-05**
	SD	**8.70E-05**

Non-smoker VIII, female, 75yo (10 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcell#/sector	MTcolony#/sector
R1	3.3E+06	5.0E+05	1.0E+06	51.0	11.7	1.17E-05	38.5	0.6
R2	6.3E+06	5.0E+05	1.0E+06	71.6	32.3	3.23E-05	203.6	3.2
R3	3.8E+06	5.0E+05	1.0E+06	171	131.7	1.32E-04	500.5	7.8
T1	6.0E+06	5.0E+05	1.0E+06	48	8.7	8.72E-06	52.3	0.8
T2	3.3E+06	5.0E+05	1.0E+06	35	0	0	0	0
T3	2.9E+06	5.0E+05	1.0E+06	44.1	4.8	4.82E-06	14.1	0.2
T4	2.0E+06	5.0E+05	1.0E+06	67.9	28.6	2.86E-05	57.2	0.9
T5	3.2E+06	5.0E+05	1.0E+06	185	145.7	1.46E-04	466.3	7.3
T6	2.5E+06	5.0E+05	1.0E+06	23.9	0	0.0	0.0	0.0
T7	4.5E+06	5.0E+05	1.0E+06	51.9	12.6	1.26E-05	56.2	0.9
Sum	3.8E+07						1388.8	21.7
Mean	3.8E+06							2.2

Tk6	(n=5)
input	1E+06
BG (cop	39.3
SD	10.9

Mean MF	3.76E-05
Total MF	3.68E-05
SD	5.45E-05

9.1.9. Appendix A.9. K-ras bp35 G:C->T:A in non-smokers

Non-smoker I, female, 38yo (5 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
1A	4.2E+05	2.8E+05	5.7E+05	19.8	0	0	0	0
3A	1.2E+06	7.4E+05	1.5E+06	30.2	0	0	0	0
L7A	2.3E+05	1.5E+05	3.1E+05	30.3	0	0	0	0
5C	7.1E+05	5.0E+05	1.0E+06	126	43.0	4.30E-05	30.4	0.5
R9AB	1.3E+06	7.5E+05	1.5E+06	219	136.0	9.07E-05	116.5	1.8
Sum	3.80E+06						146.9	
Mean	7.60E+05					2.67E-05	29.4	0.5

TK6	n=22
input (co	1E+06
BG	83.0
SD	27.6

Mean MF	2.67E-05
Total MF	3.86E-05
SD	4.03E-05

Non-smoker II, female, 40yo (5 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
1A	1.8E+05	1.1E+05	2.3E+05	33.1	0	0	0	0
3B	2.3E+05	1.5E+05	2.9E+05	41.5	0	0	0	0
L9E	8.2E+05	5.3E+05	1.1E+06	44.2	0	0	0	0
R5A	9.4E+05	6.0E+05	1.2E+06	150	67.0	5.58E-05	52.6	0.8
R7A	1.8E+05	1.1E+05	2.3E+05	90.3	7.3	3.20E-05	5.7	0.1
Sum	2.35E+06						58.3	
Mean	4.70E+05					1.76E-05	11.7	0.2

TK6	n=22
input (co	1E+06
BG	83.0
SD	27.6

Mean MF	1.76E-05
Total MF	2.48E-05
SD	2.55E-05

Non-smoker III, male, 41yo (3 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF	MTcells/sector	MTcolony#/sector
L5A	2.35E+06	1.0E+06	2.0E+06	85.5	2.5	1.26E-06	3.0	0.0
L11B	4.71E+06	1.5E+06	3.0E+06	45.8	0	0	0	0
L6C	3.06E+06	1.5E+06	3.0E+06	144	61.0	2.03E-05	62.2	1.0
Sum	1.01E+07						65.2	
Mean	3.37E+06					7.20E-06	21.7	0.3

TK6	n=22
input (co	1E+06
BG	83.0
SD	27.6

Mean MF	7.20E-06
Total MF	6.44E-06
SD	1.14E-05

Non-smoker IV, female, 45yo (38 sectors)

Sector	Cells/sector	Cells/assay	copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1A	1.35E+06	6.53E+05	1.31E+06	43.3	0	0E+00	0	0.0
LMS1B	1.35E+06	4.11E+05	8.22E+05	68.0	0	0E+00	0	0.0
LMS2A	1.36E+06	4.13E+05	8.26E+05	177.9	94.9	1.15E-04	156.1	2.4
LMS2B	1.34E+06	2.88E+05	5.76E+05	10.8	0	0E+00	0	0.0
LMS3A	1.09E+06	3.98E+05	7.96E+05	188.6	105.6	1.33E-04	144.9	2.3
LMS3B	8.12E+05	2.86E+05	5.71E+05	137.0	54.0	9.46E-05	76.8	1.2
LMS4	5.88E+05	5.30E+05	1.06E+06	66.0	0	0E+00	0	0.0
L4A	1.96E+06	4.74E+05	9.47E+05	12.9	0	0E+00	0	0.0
L4B	1.44E+06	2.82E+05	5.64E+05	18.4	0	0E+00	0	0.0
L5B	8.89E+05	3.56E+05	7.12E+05	169.8	86.8	1.22E-04	108.4	1.7
L7	8.54E+05	2.51E+05	5.02E+05	168.7	85.7	1.71E-04	145.8	2.3
L8	9.73E+05	5.25E+05	1.05E+06	4.8	0	0E+00	0	0.0
L9	1.09E+06	3.56E+05	7.12E+05	115.4	32.4	4.55E-05	49.4	0.8
L10	7.35E+05	4.60E+05	9.20E+05	1.0	0	0E+00	0	0.0
L12	7.42E+05	4.75E+05	9.50E+05	233.5	150.5	1.58E-04	117.6	1.8
L13A	9.59E+05	3.67E+05	7.35E+05	76.8	0	0E+00	0	0.0
L14A	6.79E+05	3.43E+05	6.85E+05	211.2	128.2	1.87E-04	127.1	2.0
L14B	2.44E+06	3.82E+05	7.63E+05	156.9	73.9	9.69E-05	236.7	3.7
L15B	1.48E+06	7.82E+05	1.56E+06	17.9	0	0E+00	0	0.0
T2	4.76E+05	2.05E+05	4.10E+05	50.7	0	0E+00	0	0.0
T3	9.80E+05	3.89E+05	7.77E+05	155.9	72.9	9.38E-05	92.0	1.4
T4	8.05E+05	5.50E+05	1.10E+06	60.6	0	0E+00	0	0.0
T5	4.13E+05	2.47E+05	4.93E+05	122.0	39.0	7.91E-05	32.7	0.5
T6	1.30E+06	4.33E+05	8.66E+05	85.0	2.0	2.33E-06	3.0	0.0
T7	1.20E+05	5.88E+05	1.18E+06	237.3	154.3	1.31E-04	15.7	0.2
T8	7.35E+05	3.88E+05	7.76E+05	52.4	0	0E+00	0	0.0
T9	1.10E+06	2.93E+05	5.85E+05	112.8	29.8	5.10E-05	56.0	0.9
T10	2.77E+05	3.32E+05	6.64E+05	214.5	131.5	1.98E-04	54.8	0.9
T11	1.08E+06	4.42E+05	8.83E+05	214.8	131.8	1.49E-04	160.9	2.5
R1	3.06E+06	8.26E+05	1.65E+06	175.0	92.0	5.57E-05	170.4	2.7
R2	8.68E+05	6.37E+05	1.27E+06	88.8	5.8	4.57E-06	4.0	0.1
R3	1.82E+06	2.60E+05	5.19E+05	88.2	5.2	1.01E-05	18.3	0.3
R4	1.64E+06	5.18E+05	1.04E+06	51.9	0	0E+00	0	0.0
R5	1.80E+06	1.82E+05	3.64E+05	150.0	67.0	1.84E-04	331.2	5.2
R6	1.27E+06	6.55E+05	1.31E+06	88.2	5.2	3.99E-06	5.1	0.1
R7	1.40E+06	4.05E+05	8.10E+05	91.5	8.5	1.05E-05	14.7	0.2
R8	1.18E+06	7.35E+05	1.47E+06	42.9	0	0E+00	0	0.0
R9	1.05E+06	6.60E+05	1.32E+06	93.3	10.3	7.82E-06	8.2	0.1
Sum	4.35E+07						2129.7	
Mean	1.14E+06					5.54E-05		0.9

TK6	n=22
input (copies)	1E+06
BG	83.0
SD	27.6

Mean MF	5.54E-05
Total MF	4.90E-05
SD	6.83E-05

Non-smoker VIII, female, 75yo (10 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcell#/sector	MTcolony#/sector
R1	3.3E+06	5.0E+05	1.0E+06	2.2	0	0	0	0
R2	6.3E+06	5.0E+05	1.0E+06	6.1	0.1	5.7E-08	0.4	0.0
R3	3.8E+06	5.0E+05	1.0E+06	19.1	13.1	1.3E-05	49.9	0.8
T1	6.0E+06	5.0E+05	1.0E+06	15	9.0	9.0E-06	53.9	0.8
T2	3.3E+06	2.5E+05	5.0E+05	6.5	0.5	9.8E-07	3.2	0.1
T3	2.9E+06	2.5E+05	5.0E+05	1	0	0	0	0
T4	2.0E+06	3.0E+05	6.0E+05	1	0	0	0	0
T5	3.2E+06	3.0E+05	6.0E+05	1	0	0	0	0
T6	2.5E+06	4.0E+05	8.0E+05	0	0	0	0	0
T7	4.5E+06	2.0E+05	4.0E+05	1	0	0	0	0
Sum	3.8E+07						107.5	1.7
Mean	3.8E+06					2.32E-06		0.2

TK6	n=14
input (c	1E+06
BG	6.0
SD	6.0

Mean MF	2.32E-06
Totlal MF	2.84E-06
SD	4.72E-06

9.1.10. Appendix A.9. HPRT bp508 G:C->T:A in non-smokers

Non-smoker I, female, 38yo (5 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
1A	1.58E+05	1.05E+05	2.10E+05	5.4	0.0	0.00E+00	0.0	0
3A	9.88E+04	6.59E+04	1.32E+05	44	21.7	1.64E-04	16.3	0.3
5C	1.48E+05	7.38E+04	1.48E+05	33.2	10.9	7.36E-05	10.9	0.2
L7A	3.05E+04	2.03E+04	4.06E+04	7.7	0.0	0.00E+00	0.0	0.0
R9AB	2.43E+05	1.22E+05	2.43E+05	68.1	45.8	1.88E-04	45.8	0.7
Sum	6.78E+05						72.9	1.1
Mean	1.36E+05					8.53E-05	14.6	0.2

Tk6	BG	SD
5e5 n=18	22.3	13.2
e6 n=18	31.3	18.0

Mean MF	8.53E-05
Total MF	1.08E-04
SD	8.88E-05

Non-smoker II, female, 40yo (9 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
1A	4.69E+04	4.69E+04	9.38E+04	6.9	0.0	0.00E+00	0.0	0
1D	5.52E+05	3.68E+05	7.35E+05	27.4	5.1	6.89E-06	3.8	0.1
3B	2.46E+04	1.64E+04	3.28E+04	5.6	0.0	0.00E+00	0.0	0
5D	1.44E+05	1.08E+05	2.16E+05	5.1	0.0	0.00E+00	0.0	0
L6A	5.05E+05	2.52E+05	5.05E+05	27.3	5.0	9.85E-06	5.0	0.1
L9E	2.15E+05	1.43E+05	2.87E+05	4	0.0	0.00E+00	0.0	0
R5A	2.63E+05	1.76E+05	3.51E+05	11.7	0.0	0.00E+00	0.0	0
R6C	1.84E+05	1.38E+05	2.77E+05	45.5	23.2	8.38E-05	15.4	0.2
R7A	2.39E+05	1.59E+05	3.18E+05	29.1	6.8	2.13E-05	5.1	0.1
Sum	2.17E+06						29.3	0.5
Mean	2.41E+05					1.35E-05	3.3	0.1

Tk6	BG	SD
5e5 n=18	22.3	13.2
e6 n=18	31.3	18.0

Mean MF	1.35E-05
Total MF	1.35E-05
SD	2.73E-05

Non-smoker III, male, 41yo (3 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
L5A	8.23E+05	5.49E+05	1.10E+06	19.1	0.0	0.00E+00	0.0	0
L6C	7.50E+05	3.75E+05	7.50E+05	80.4	58.1	7.75E-05	58.1	0.9
L11B	2.37E+05	1.19E+05	2.37E+05	25.3	3.0	1.25E-05	3.0	0.0
Sum	1.81E+06						61.0	1.0
Mean	6.03E+05					3.00E-05	20.3	0.3

Tk6	BG	SD
5e5 n=18	22.3	13.2
e6 n=18	31.3	18.0

Mean MF	3.00E-05
Total MF	3.37E-05
SD	4.16E-05

Non-smoker IV female, 45yo (42 sectors)

Sector	Total cells	Cells/assay	Copy#/assay	MTcopy#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS1A	1.16E+06	3.11E+05	6.22E+05	34.4	2.7	4.30E-06	5.0	0.1
LMS1B	1.16E+06	1.71E+05	3.41E+05	24	0	0.00E+00	0.0	0.0
LMS2A	1.16E+06	1.77E+05	3.54E+05	31.8	3.7	1.06E-05	12.3	0.2
LMS2B	1.15E+06	2.60E+05	5.20E+05	43.5	15.4	2.97E-05	34.0	0.5
LMS3A	9.36E+05	1.41E+05	2.82E+05	35.9	7.8	2.78E-05	26.0	0.4
LMS3B	6.96E+05	1.00E+05	2.00E+05	31	2.9	1.47E-05	10.2	0.2
LMS4	5.04E+05	1.57E+05	3.13E+05	86.3	58.2	1.86E-04	93.8	1.5
L4A	1.68E+06	6.37E+05	1.27E+06	37.9	0	0.00E+00	0.0	0.0
L4B	1.24E+06	4.68E+05	9.37E+05	25.7	0	0.00E+00	0.0	0.0
L5A	5.52E+04	2.09E+04	4.18E+04	0	0	0.00E+00	0.0	0.0
L5B	7.62E+05	2.89E+05	5.78E+05	34.4	2.7	4.63E-06	3.5	0.1
L7	7.32E+05	2.77E+05	5.55E+05	77.7	49.6	8.95E-05	65.5	1.0
L8	8.34E+05	3.16E+05	6.32E+05	51	19.3	3.05E-05	25.4	0.4
L9	9.30E+05	1.76E+05	3.52E+05	9.5	0	0.00E+00	0.0	0.0
L10	6.30E+05	2.39E+05	4.78E+05	14.2	0	0.00E+00	0.0	0.0
L11	2.70E+04	1.02E+04	2.05E+04	23.9	0	0.00E+00	0.0	0.0
L12	6.36E+05	2.41E+05	4.82E+05	7.7	0	0.00E+00	0.0	0.0
L13A	8.22E+05	3.12E+05	6.23E+05	106.2	74.5	1.20E-04	98.3	1.5
L13B	5.76E+05	2.18E+05	4.37E+05	97.9	69.8	1.60E-04	92.1	1.4
L14A	5.82E+05	2.21E+05	4.41E+05	44.8	16.7	3.80E-05	22.1	0.3
L14B	2.09E+06	7.94E+05	1.59E+06	113.6	67.2	4.23E-05	88.6	1.4
L15A	1.79E+06	6.80E+05	1.36E+06	80.4	34.0	2.50E-05	44.8	0.7
L15B	1.27E+06	2.00E+05	4.00E+05	67.95	39.9	9.97E-05	126.2	2.0
T1	9.60E+05	3.64E+05	7.28E+05	22.3	0	0.00E+00	0.0	0.0
T2	4.08E+05	1.55E+05	3.09E+05	49.9	21.8	7.06E-05	28.8	0.5
T3	8.40E+05	3.18E+05	6.37E+05	90	58.3	9.15E-05	76.9	1.2
T4	6.90E+05	2.62E+05	5.23E+05	21.6	0	0.00E+00	0.0	0.0
T5	3.54E+05	1.34E+05	2.68E+05	40.3	12.2	4.56E-05	16.2	0.3
T6	1.12E+06	4.23E+05	8.46E+05	122.7	83.6	9.89E-05	110.3	1.7
T7	1.03E+05	3.89E+04	7.78E+04	160.4	132.3	1.70E-03	174.6	2.7
T8	6.30E+05	2.39E+05	4.78E+05	62	33.9	7.11E-05	44.8	0.7
T9	9.42E+05	3.57E+05	7.14E+05	41.8	6.4	8.97E-06	8.4	0.1
T10	2.37E+05	8.98E+04	1.80E+05	32.2	4.1	2.31E-05	5.5	0.1
T11	9.24E+05	1.50E+05	2.99E+05	23.9	0	0.00E+00	0.0	0.0
R1	2.62E+06	4.89E+05	9.78E+05	39.1	0	0.00E+00	0.0	0.0
R2	7.44E+05	1.35E+05	2.69E+05	32.2	4.1	1.52E-05	11.3	0.2
R3	1.56E+06	5.91E+05	1.18E+06	42.2	0	0.00E+00	0.0	0.0
R4	1.40E+06	5.32E+05	1.06E+06	151.8	105.4	9.90E-05	139.0	2.2
R5	1.54E+06	5.84E+05	1.17E+06	105.2	58.8	5.03E-05	77.6	1.2
R6	1.09E+06	4.14E+05	8.28E+05	104.9	65.8	7.95E-05	86.8	1.4
R7	1.20E+06	4.55E+05	9.10E+05	80	33.6	3.69E-05	44.3	0.7
R8	1.01E+06	3.82E+05	7.64E+05	37.6	2.2	2.88E-06	2.9	0.0
Sum (n=42)	3.98E+07						1575.5	
Sum (n=41)	3.97E+07						1400.9	
Mean (n=42)	9.47E+05					7.80E-05		0.6
Mean (n=41)	9.68E+05					3.84E-05		

eTk6	input	BG	SD	
n=17	5E+05	**28.1**	15.5	
(copies)	6E+05	31.7	23.4	
	7E+05	35.4		
	8E+05	39.1		
n=12	1E+06	**46.4**		

n=42	Mean MF	7.80E-05
all sectors	Total MF	3.96E-05
	SD	2.61E-04
n=41	mean MF	3.84E-05
MF<4E-4	Total MF	3.53E-05
	SD	4.74E-05

Non-smoker V male, 50yo (17 sectors)

Sector	Total cells	Cells/assay	Copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LLL1	5.81E+05	1.94E+05	3.87E+05	121.5	94.0	2.43E-04	141.1	2.2
LLL8	6.06E+05	2.63E+05	5.25E+05	53.4	22.4	4.26E-05	25.8	0.4
LLL11	3.09E+05	1.23E+05	2.47E+05	58.0	34.1	1.38E-04	42.6	0.7
LLL13	4.01E+05	1.60E+05	3.21E+05	67.2	43.3	1.35E-04	54.1	0.8
LUL2	5.17E+05	1.72E+05	3.45E+05	55.0	31.1	9.02E-05	46.6	0.7
LUL5	5.75E+05	2.30E+05	4.60E+05	66.5	35.5	7.71E-05	44.3	0.7
LUL17+18	4.41E+05	1.76E+05	3.53E+05	29.8	2.3	6.62E-06	2.9	0.0
RLL1	3.35E+05	1.45E+05	2.90E+05	33.1	9.2	3.17E-05	10.6	0.2
RLL3	3.77E+05	1.63E+05	3.27E+05	47.2	23.3	7.13E-05	26.9	0.4
RLL4	5.76E+05	1.92E+05	3.84E+05	48.0	20.5	5.35E-05	30.8	0.5
RLL13	3.42E+05	1.14E+05	2.28E+05	45.7	21.8	9.56E-05	32.7	0.5
RLL19	4.80E+05	2.08E+05	4.16E+05	27.0	0	0	0.0	0.0
RLL22	7.40E+05	2.47E+05	4.93E+05	55.0	24.0	4.86E-05	36.0	0.6
RUL1	5.62E+05	2.44E+05	4.87E+05	52.3	21.3	4.37E-05	24.5	0.4
RUL3	3.46E+05	1.27E+05	2.53E+05	57.0	33.1	1.31E-04	45.1	0.7
RUL11	4.28E+05	1.57E+05	3.14E+05	5.2	0	0	0.0	0.0
MC	4.01E+05	1.34E+05	2.67E+05	46.0	22.1	8.27E-05	33.1	0.5
Sum	8.02E+06						597.3	18.7
Mean	4.72E+05					7.59E-05		1.6

Tk6	input	BG
(copies)	3E+05	23.9
	4E+05	27.5
	5E+05	31.0

Mean MF	7.59E-05
Total MF	7.45E-05
SD	6.15E-05

Non-smoker VI, female, 59yo (5 sectors)

Sector	Total cells	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG	MF (-BG)	MTcells/sector	MTcolony#/sector
LMS	1.06E+06	3.52E+05	7.04E+05	73.9	48.5	6.89E-05	72.7	1.1
LUL	7.64E+05	3.05E+05	6.11E+05	5.4	0	0	0	0
RLL	6.25E+05	2.66E+05	5.32E+05	76.0	53.7	1.01E-04	63.1	1.0
RUL	1.30E+06	5.20E+05	1.04E+06	6.9	0	0	0	0
T	1.65E+06	4.12E+05	8.24E+05	19.0	0	0	0	0
Sum	5.39E+06						135.8	
Mean	1.08E+06					3.40E-05	27.2	0.4

Tk6	input	BG (copies)	SD
n=20	5E+05	22.3	13.2
	7E+05	25.4	

Mean MF	3.40E-05
Total MF	2.52E-05
SD	4.79E-05

Non-smoker VII, male, 67yo (3 sectors)

Sample	Total cells/	Cells/assay	copies/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
Trachea-1	4.79E+04	3.20E+04	6.39E+04	52	29.7	4.64E-04	22.3	0.3
Trachea-2	6.86E+05	4.57E+05	9.14E+05	462	430.7	4.71E-04	323.0	5.0
Trachea-3	2.62E+05	1.75E+05	3.49E+05	23.7	1.4	3.93E-06	1.0	0.0
Sum	9.95E+05						346.3	5.4
Mean	3.32E+05					3.13E-04	115.4	1.8

Tk6	BG	STDEV
5e5 n=18	22.3	13.2
e6 n=18	31.3	18.0

n=3	Mean MF	3.13E-04
	Total MF	3.48E-04
	SD	2.68E-04
n=1		
MF<4E-4	MF	3.93E-06

Non-smoker VIII, female, 75yo (10 sectors)

Sector	Total cells/	Cells/assay	copy#/assay	MTc#/assay	MTc#-BG/	MF (BG)	MTcell#/sector	MTcolony#/sector
R1	3.30E+06	3.30E+06	6.60E+06	282	168	2.54E-05	83.9	1.3
R2	6.30E+06	2.48E+06	4.95E+06	240	126	2.54E-05	159.8	2.5
R3	3.80E+06	1.68E+06	3.36E+06	240	166	4.95E-05	187.9	2.9
T1	6.00E+06	3.00E+06	6.00E+06	406	292	4.87E-05	291.9	4.6
T2	3.30E+06	1.65E+06	3.30E+06	389	315	9.53E-05	314.5	4.9
T3	2.93E+06	1.47E+06	2.93E+06	407	333	1.14E-04	332.7	5.2
T4	2.00E+06	1.00E+06	2.00E+06	535	481	2.40E-04	481.0	7.5
T5	3.20E+06	1.60E+06	3.20E+06	322	248	7.76E-05	248.3	3.9
T6	2.50E+06	1.25E+06	2.50E+06	252	178	7.13E-05	178.2	2.8
T7	4.45E+06	2.23E+06	4.45E+06	343	229	5.14E-05	228.8	3.6
Sum	3.78E+07						2507.2	39.2
Mean	3.78E+06					8.0E-05	250.7	3.9

Tk6	input	BG
(copies)	2E+06	54
	3E+06	74
	5E+06	114

Mean MF	7.99E-05
Total MF	6.64E-05
SD	6.32E-05

Non-smoker IX, Female, 76yo (3 sectors)

Sample	Total cells/	Cells/assay	copies/assay	MTcopy#/assay	MTc-BG/a	MF	MTcells/sector	MTcolony#/sector
LB	2.17E+05	1.44E+05	2.89E+05	90	67.7	2.34E-04	50.8	0.8
RB	1.17E+05	7.77E+04	1.55E+05	23.7	1.4	8.81E-06	1.0	0.0
Tr	4.00E+05	2.67E+05	5.34E+05	15.4	0.0	0.0E+00	0.0	0
Sum	7.34E+05						51.8	0.8
Mean	2.45E+05					8.10E-05	17.3	0.3

Tk6	BG	STDEV
5e5 n=18	22.3	13.2
e6 n=18	31.3	18.0

Mean MF	8.10E-05
Total MF	7.06E-05
STDEV	1.33E-04