IDENTIFICATION OF NOVEL GENETIC ALTERATIONS IN NON-SMALL CELL LUNG CANCER PROGRESSION

By

AREK SIWOSKI

B.Sc., The University of British Columbia, 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Pathology
and Laboratory Medicine
We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Pathology

The University of British Columbia
Vancouver, Canada

Date Jan 11/01

Abstract

To improve the survival rate in lung cancer, novel molecular targets that facilitate early detection and intervention are required. This thesis aims at screening multiple loci from distinct stages of lung cancer using genome scanning methods to identify frequently occurring genetic alterations that could potentially reveal novel regions of chromosomal instability. Due to the tissue heterogeneity in biopsies, tissue microdissection is necessary to isolate specific populations of cells from formalin-fixed, paraffin-embedded tissues representing progression stages of lung cancer. Since some of the commonly used histological stains such as Hematoxylin & Eosin have been shown to degrade DNA due to their acidic nature, a panel of histological stains have been tested for minimal degradative effect on DNA. To work with the minute quantities of DNA recovered from premalignant lesions, we have adapted a Randomly Amplified Polymorphic DNA-Polymerase Chain Reaction (RAPD-PCR) assay to assess the quality of DNA extracted from the microdissected specimens, as well as a Southern hydridization assay to assess DNA quantity. DNA samples have been extracted from normal epithelial, hyperplastic, dysplastic, CIS and invasive carcinoma cells originated from 41 patients at various bronchial sites. DNA yield varied from 20ng to 2000ng depending on the amount of cells available in each lesion. Typical molecular techniques such as loss of heterozygosity (LOH) assay, DNA microarrays and fluorescent in situ hybridization (FISH) require large amounts of DNA to detect genomic changes. Using RAPD-PCR as a multi-loci fingerprinting technique, it is possible to screen for genetic alterations using minute DNA samples from microdissected specimens. RAPD-PCR involves simultaneous amplification of genomic DNA at multiple loci using short oligonucleotide primers.

Gains and losses of PCR signals that occur in multiple patients identify regions potentially harboring candidate tumour suppressor genes or oncogenes. Our results show that twelve patients showed a gain of PCR signal in carcinoma in situ (CIS) tissue compared to normal tissue. By cloning, sequencing, and mapping the fragments, the recurring gains in PCR signal in multiple patients corresponded to regions at 8q24.3 (6/17), 1q23.3 (2/18), 5q33.1 (2/16), and 7q36.2 (2/15). Microsatellite markers were chosen surrounding all four regions, however, due to minimal amounts of patient DNA, only regions 8q24.3 and 7q36.2 were verified for allelic imbalance in multiple patients using LOH. The groundwork has been laid for others to define the boundaries of these unstable regions as both regions contain many potential oncogenes and tumor suppressor genes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to my supervisor Dr. Wan Lam for his great ideas, scientific enthusiasm and guidance and members of the laboratory especially Krista Cleveland, Cathie Garnis and Chad Malloff for their technical advice and assistance. I would like to thank Dr. Juergen Vielkind for all his guidance. I would like to acknowledge Dr. Stephen Lam for providing us with lung biopsies, Chris Dawe for microdissection assistance, and Dr. Calum MacAulay for innovative ideas. Finally, I would like to thank my beautiful wife, Jenny for all her support, encouragement and laughter. I would also like to thank my family for their constant optimism and support throughout the years.

Chapter 1: Introduction

1.1 Background

As cancer is becoming the major cause of death in the civilized world, lung cancer mortality rate is increasing dramatically. In the United States, it was estimated that 171,600 new cases and 158,900 deaths were due to lung cancer in 1998 (Tseng, 1999). The most effective treatment for lung cancer is surgical resection however, present screening techniques result in 65% of patients presenting with advanced stage disease at time of diagnosis (Esteller, 1999). After surgical resection, many patients will succumb to recurrent lung cancer at distal sites. Despite some advances in treatment, the two-year survival rate of lung cancer patients without complete tumour resection was estimated at 10% (Wiest, 1997). As with other potential metastatic cancers, patients with early stage lung cancer, where the abnormal cells are localized and small in volume, show the best response to therapies and present the greatest survival rate compared to patients with the advanced-stage disease. Therefore, improvements in early detection techniques through the identification of prognostic markers in conjunction with the identification of a gene or group of genes linked to lung cancer predisposition are needed to reduce the mortality rate.

Lung cancer can be divided into two main types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC's account for ~25% of all lung cancer cases whereas the remainder of cases are adenocarcinoma, squamous cell carcinoma or large cell carcinoma, subtypes of NSCLC's. Adenocarcinomas account for 40% of all diagnosed lung tumours. Although there are four discrete histological types, it is well accepted that in many cases, an individual tumour can present features of more than one

type. It has been suggested that all histological types arise from similar bronchial epithelium differentiation pathway.

Embryologically, lung development begins with the out-pouching of epithelial cells from the foregut endoderm. Fetal lung cellular differentiation and branching morphogenesis gives rise to the characteristic respiratory epithelium. The pseudostratified epithelium of the bronchus is composed of basal, secretory and ciliated cells whereas the alveolar epithelium consists of type I and II alveolar cells (Burkitt, 1993). Presently, a pluripotent stem cell giving rise to the epithelial cells of the lung has not been identified.

Lung cancer development has been suggested to involve the accumulation of genetic mutations in the bronchial epithelium resulting in a progression through a series of morphologically distinct premalignant changes leading to invasive cancer. The progressive histological stages include basal cell hyperplasia, squamous cell metaplasia, mild, moderate and severe dysplasias, carcinoma in situ and invasive cancer. As this is a multistep process, it involves the mutation of several entities such as oncogenes, tumour suppressor genes and DNA repair genes. According to Knudson's two-hit theory of cancer progression, tumour suppressor genes are inactivated by a recessive mutation in one allele and loss of the other allele that retained wild-type function. Since both copies of the gene must be inactivated for a phenotypic change, the first copy is inactivated by point mutation, methylation or small deletion whereas the second copy is inactivated by large deletions. Regions of frequent chromosomal deletions represent potential tumour suppressor gene presence. Similarly, the mutagenic theory of cancer development, as introduced by Vogelstein and Fearon, encompasses four main ideas: 1) activation of

oncogenes coupled with the inactivation of tumour suppressor genes gives rise to tumours. 2) the formation of a malignant tumour requires mutations in at least four or five genes. 3) a tumour's biologic properties are representative of the total accumulation of genetic changes and 4) tumour suppressor genes that are mutated can cause a phenotypic effect in a heterozygous state. By identifying the early genetic mutations, potential prognostic markers may be discovered.

1.2. Genes Implicated In NSCLC

1.2.1 Potential oncogenes

Potential oncogenes at multiple chromosomal locations have been identified using cytogenetic and allelotyping assays. These oncogenes include the Ras gene, MYC gene and Bcl-2 gene.

The Ras gene family composed of K-Ras, H-Ras and N-Ras encode a Ras protein primarily involved in growth signal transduction to the cell nucleus. In its active signaling state, the Ras protein binds guanosine triphosphate (GTP) and in its inactive state, the Ras protein binds guanosine diphosphate (GDP). The Ras protein has an intrinsic GTPase activity causing hydrolysis of GTP to GDP. Mutations in the Ras gene cause Ras protein to lose its capability to hydrolyze GTP resulting in the permanent active state of the Ras protein and continued growth signal transduction. Common mutations in the Ras gene are single point mutations at codons 12, 13, or 61 occurring in 20-30% of lung adenocarcinomas and 15-20% of NSCLCs (Gazdar, 1994; Sugio, 1994; Sekido,1998). K-Ras mutations seem to account for 90% of Ras mutations in lung adenocarcinomas (Sekido,1998). Many studies report that K-Ras mutations predict a

poor prognosis in both early and late stage NSCLC and suggest its use as a clinical prognostic marker (Rosell,1993).

The MYC gene family composed of MYCL, MYCN, and MYC encode for a transcription factor activated by the RAS signal transduction cascade. The MYC proteins recognize a consensus sequence CACGTG resulting in activation of genes involved in normal cell growth and proliferation. It is implied that there is direct activation of genes involved in DNA synthesis, RNA metabolism and cell-cycle progression. Common aberrations of MYC genes include gene amplifications and transcriptional dysregulations resulting in overexpression of protein product. Previous studies show MYC amplifications in 18% of SCLC tumours and 8% of NSCLC tumours (Richardson, 1993). Comparative genomic hybridization (CGH) (described in 1.6.1) has detected gene amplifications ranging from 20 to 115 copies per cell. MYC genes localize to three distinct chromosomal regions: MYCL (1p32), MYCN (2p25) and MYC (8q24). Previous studies show that 31% and 20% of SCLC and NSCLC cell lines, respectively have MYC amplifications (Richardson, 1993). It is reported that MYC amplifications correlate with decreased survival as cell lines derived from metastatic lesions demonstrate increased frequencies of amplifications compared to primary tumours (Richardson, 1993).

Apoptotic pathways have been considered as possible areas of gene regulation involved in lung cancer development. Bcl-2, a protein involved in normal apoptosis, is believed to protect cells from the natural programmed cell death enabling tumour cells to escape apoptotic death. Several reports show that the Bcl-2 protein expression is greater in SCLCs than in NSCLCs. It has also been reported that 75-90% of SCLCs express increased levels of Bcl-2 protein (Jiang, 1995). Acting as an antagonist to Bcl-2, BAX, a

tumour suppressor gene candidate, encodes a protein involved in promoting apoptosis. The ratio of BAX:Bcl-2 is believed to indicate a cell's susceptibility to apoptosis.

1.2.2 Potential Tumour Suppressor Genes

Cytogenetic and allelotyping studies have elucidated potential tumour suppressor genes at multiple chromosomal locations. Mutated tumour suppressor genes such as $p16^{INK4}$, RB, p53 and FHIT have been characterized at 9p21, 13q14, 17p13 and 3p14, respectively.

Various mutations including deletions, insertions and splice mutations at chromosome 9p21 in many lung cancers indicated that there was a possible causative gene, $p16^{INK4}$. The $p16^{INK4}$ protein is responsible for cell cycle modulation of the RB pathway by inhibiting CDK4:cyclin D1 kinase activity. Abnormalities at $p16^{INK4}$ are frequent in NSCLC where homozygous deletions and point mutations have been reported in 10-40% of NSCLCs (Okami, 1997). Gene inactivation of $p16^{INK4}$ by hypermethylation of the 5' CpG island leading to disruption of the RB pathway has been reported in 30-70% of NSCLCs (Hamada, 2000; Sanchez-Cespedes, 2001). It has also been reported that $p16^{INK4}$ mutations can be correlated with tumour progression as the mutation frequency increases in metastatic lesions compared to primary lesions (Sekido, 1998).

Initially discovered in childhood retinoblastomas, the retinoblastoma gene, RB, was localized to chromosome 13q14. The RB gene encodes a nuclear phosphoprotein involved in the p16-cyclin D1-CDK4-RB pathway responsible for the transition of the cell cycle from G1 to S. In its inactive state, the RB protein binds the transcription factor E2F-1 responsible for the G1 to S transition, inhibiting S phase entry. When

phosphorylated by the cyclin dependent kinases, the RB protein loses its binding ability allowing S phase entry. The RB gene has also been linked to the repression of apoptosis and induction of cellular differentiation. Common aberrations of the RB gene include truncating deletions, nonsense mutations and splicing mutations. Studies report an abnormal RB protein in 90% of SCLCs and 15-30% of NSCLCs (Tamura, 1997). It was also shown that germline carriers of an RB mutation were 15 times more likely to die from lung cancer. It was also observed that lack of RB expression was correlated with poor prognosis in NSCLCs (Dosaka-Akita, 1997; Onuki, 1999).

The p53 gene localized at chromosome 17p13 encodes a transcription factor primarily involved in DNA damage response. The function of the p53 protein is to bind sites of DNA damage and activates a cascade of genes involved in arresting the cell cycle and apoptosis. It is believed that p53 function is critical in the death of genetically damaged cells preventing the evolution of cancer cells. Common alterations of the p53 gene include deletions, missense and nonsense mutations and splicing abnormalities of exons 5-8. Allelic loss of the p53 gene is frequent in both SCLC and NSCLC and mutational inactivation of the remaining allele occurs in 75-100% of SCLCs and 50% of NSCLC (Ahrendt, 2000; Onuki, 1999). Abnormal p53 protein expression has been shown in 40-70% of SCLCs and 40-60% of NSCLCs (Ahrendt, 1999). However, data has shown that the use of p53 expression as a prognostic marker for survival is not correlative (Dosaka-Akita, 1997).

Using cytogenetic and allelotyping analyses, it was determined that there is a deletion of one copy of the chromosome 3p in 90% of SCLCs and 80% of NSCLCs (Wistuba, 2000). This allelic loss occurred at both premalignant and malignant stages of

lung cancer. Using allelotyping, three distinct regions were identified at 3p25-26, 3p21.3-22 and 3p14-centomere suggesting that there are at least three tumour suppressor genes at 3p. Further, these regions were confirmed by homozygous deletions of these areas seen in several lung cancer cell lines. At 3p14.2, a potential tumour suppressor gene, FHIT, was proposed based on frequent loss of heterozygosity in NSCLC and homozygous deletions in several lung cancer cell lines. In 80% of smokers, the FHIT gene showed allelic imbalance. Studies on FHIT expression have shown that 49% of NSCLC specimens exhibited significant staining decrease for FHIT using immunohistochemistry assays.

Although many potential oncogenes and tumour suppressor genes have been identified, no particular gene has been supported as characteristic of NSCLC progression. There is a need to discover a gene(s) specific to an increased risk of NSCLC development and progression in order to develop therapeutic applications and impact the lung cancer mortality rate.

1.3 Chromosomal Regions Implicated in Premalignant Cases of NSCLC

Since cancer progression involves many accumulating genetic aberrations, it has become important to study the premalignant stages of tumourigenesis to identify abnormalities that act as gatekeepers to further progression. Studies report that premalignant lesions express genetic abnormalities found in cancer cells such as RAS upregulation, MYC overexpression, and allelic losses at the p53 locus. A previous allelotyping study of microdissected premalignant cells showed that the earliest change involved loss of the 3p allele followed by allele losses at 9p, 17p, 5q, and RAS mutations

suggesting that the earliest change at 3p may harbour a potential gate of lung cancer pathogenesis (Gazdar, 1994; Wiest, 1997). Another study suggests that KRAS mutations occurring in hyperplastic lesions may serve as potential gates for pathogenesis (Sugio, 1994). Interestingly, it has been shown that the allelic losses occurring in the spatially distinct premalignant lesions such as the loss of 3p, 9p and 17p, are identical to the allelic losses in the primary tumour. This evidence supports the field cancerization theory of cancer development whereby either identical clones spread throughout the lung or there exists an inherited predisposition of specific allelic loss. It is clear that mutations identified in premalignant stages have the most effective potentials as prognostic markers.

1.4 Epigentic Changes in Cancer Tumourigenesis

The accumulation of epigenetic changes such as DNA methylation is believed to be involved in cancer tumourigenesis. Gene expression is regulated at the 5' end of genes by the DNA methylation of the fifth carbon position of cytosine residues within the CpG islands. DNA methylation is a native process that is crucial in aging, embryonic development, and gene imprinting. Methylation of DNA is also critical in the regulation of housekeeping genes and tissue specific gene silencing. It is believed that the gene silencing is either caused by the inability of native transcription factors to bind to DNA sequences that have been methylated or the ability of repressive transcription factors to bind to methylated CpG islands or the altering of chromatin structure via methylation converting the DNA into an inactive form (Esteller, 1999). MYC expression has been shown to be affected by methylation as the MYC binding sites contain CpG islands and

several repressive transcription factors such as MeCP1 and MeCP2 have been identified (Momparler, 2000). Promoter hypermethylation of genes involved in cell cycle control, transcription factor regulation, and cell-cell interactions is believed to play a critical in tumourigenesis. Studies show that in NSCLCs, gene silencing of the tumour suppressor gene, $p16^{INK4}$, is caused by hypermethylation (Sekido, 1998). However, unlike the classic two hit hypothesis for disruption of a gene, methylation is more subtle as the loss of gene expression is related to methylation density. Since epigenetic modifications involve transcriptional inactivation of alleles without nucleotide sequence alterations, the changes are undetectable by genome-wide fingerprinting techniques such as RAPD-PCR or CGH.

1.5 Smoking and Lung Cancer

Unlike other malignancies, the majority of lung cancer prevalence is largely attributed to an environmental factor such as cigarette smoking. Tobacco smoke is known to contain thousands of substances including carcinogens and tumour promoters. Carcinogenesis of the lung involves the formation of covalent DNA adducts causing DNA mutation and misreplication. One potent carcinogen, benzopyrene diol epoxide has been shown to bind preferentially to regions of the p53 gene. Ahrendt et al report that 53% of 105 NSCLC patients who were smokers had p53 mutations. Cigarette smoking has also been associated with an increased incidence of K-ras mutations and chromosomal loss at the FHIT gene locus. It was reported that allelic losses and gains are present more often at chromosome arms 3p, 6q, 9p, 16p, 17p, and 19p in smokers than non-smokers (Sanchez-Cespedes, 2001). It is estimated that ~90% of all newly

diagnosed lung cancer cases will occur in patients with a prior history of smoking. When compared to non-smokers, heavy smokers have a 15 to 25 fold greater risk of dying from lung cancer (Ahrendt, 2000). There is some evidence of familial inheritance of lung cancer susceptibility genes in a Mendelian co-dominant fashion. It has been reported that there is a 6.1 fold increase in risk among relatives. Unfortunately, no gene has yet been identified.

1.6 Current Methods for Genome-wide Scanning of Alterations

1.6.1 Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) is a genome-wide scanning assay that identifies gene copy number differences at various chromosomal regions. The technique is based on the hybridization of two differently labeled whole genomic DNA probes, tumour DNA and normal DNA, to normal metaphase chromosome spreads. Probes are visualized by fluorochromes and signal intensites are measured and reported as either increased or decreased DNA content at a specific chromosomal region. CGH analysis has identified regions of chromosomal imbalances as gains, losses and amplifications of DNA in various solid tumours. Taguchi et al. reported DNA amplification detected by CGH at 7p12-13, 8q24, 9p24, and 10q22 in a NSCLC cell line. Michelland et al. (1999) reported observing more gains than losses in NSCLC as detected by CGH. The chromosomal gains included 1p, 2p, 7q, 8q, 15q, 17q, 19p, 19q, 20p, 21q, and 22q. Chromosomal losses were observed at 3p, 9p, 10p, 13q and 17p. Figure 1 shows a comprehensive study of 11 neuroendocrine (NE) tumours and 11 NSCLC tumours and the chromosomal imbalances seen in these patients. However, CGH has a limited capacity in detection. CGH is unable to provide a precise chromosomal location as

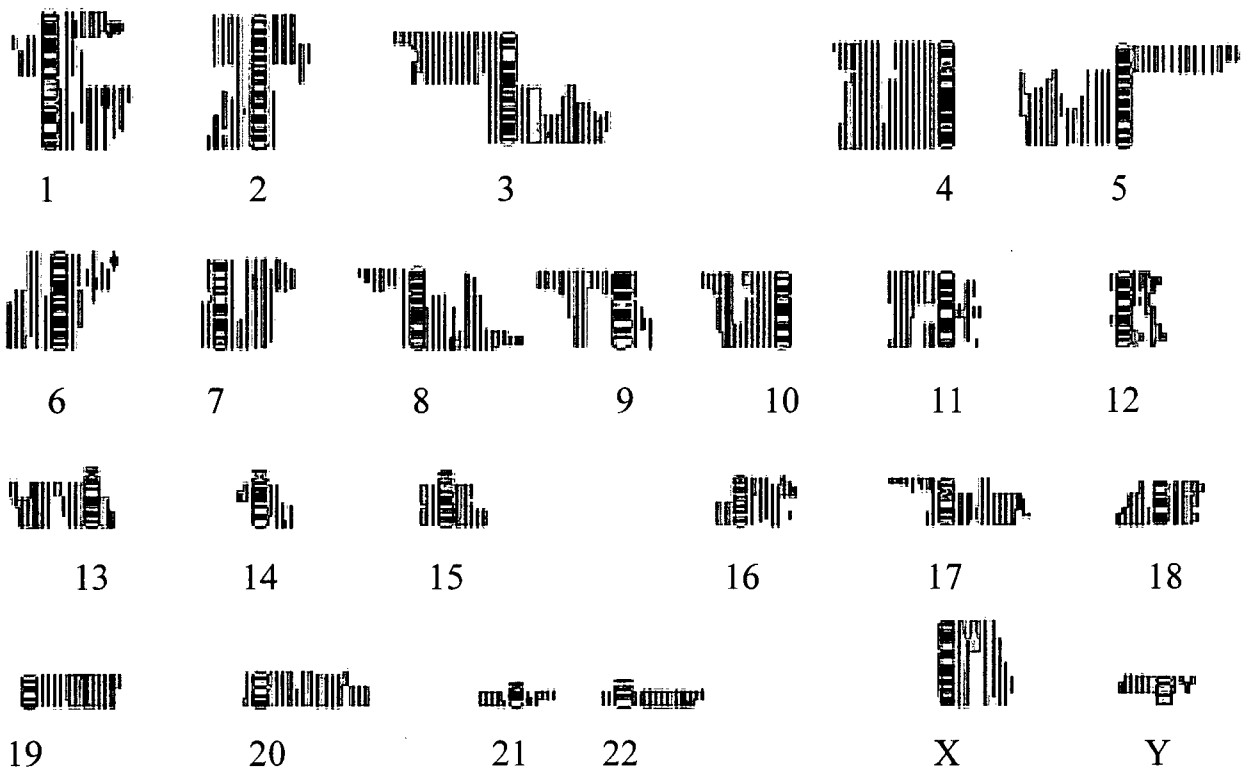


Figure 1. CGH Analysis of the Chromosomal Imbalances in 11 Neuroendocrine (thin lines) and 11 NSCLC (thick lines) Carcinomas. Vertical lines on the right side of each chromosome represent chromosomal material gains in the tumor; vertical lines on the left side correspond to losses. Each line illustrates a chromosomal aberration observed in one single tumor. Amplification sites are represented by squares. (Michelland et al., 1999)

resolution of genetic changes is limited to a chromosome band, at best, because the detection involves microscope-based visualization of chromosome spreads. Also, signal intensities involving a two-fold change representing a loss of a chromosomal region are difficult to discern using fluorophores.

1.6.2 Loss of Heterozygosity (LOH)

Loss of Heterozygosity (LOH) is a PCR based strategy for detecting regions of chromosomal instability using polymorphic microsatellites, nucleotide repeats, distributed throughout the genome. These polymorphic sites are generated by the differences in DNA sequence between maternal and paternal chromosomes at many genetic loci. The PCR reaction amplifies these polymorphic sites and they are represented as two maternal and paternal alleles on a polyacrylamide gel (Figure 2). The two alleles are different in their base pair composition attributed to the polymorphism at the loci. By comparing normal DNA versus tumour DNA, one can determine allelic imbalance at a specific loci. The imbalance can arise either as a deletion of the normal allele or a duplication of the mutant allele. In any event, the allelic imbalance is represented as a signal intensity ratio change between the two alleles in the normal DNA versus the ratio of the two alleles in the tumour DNA. Genetic alteration may be viewed as allelic imbalance. Nonrandom distribution of allelic imbalance is due to the presence of both oncogenes and tumour suppressor genes within chromosomal arms.

Microsatellites have been effective in allelotyping due to their abundance, hypervariability, high rate of polymorphism and genomic distribution. Girard et al. report

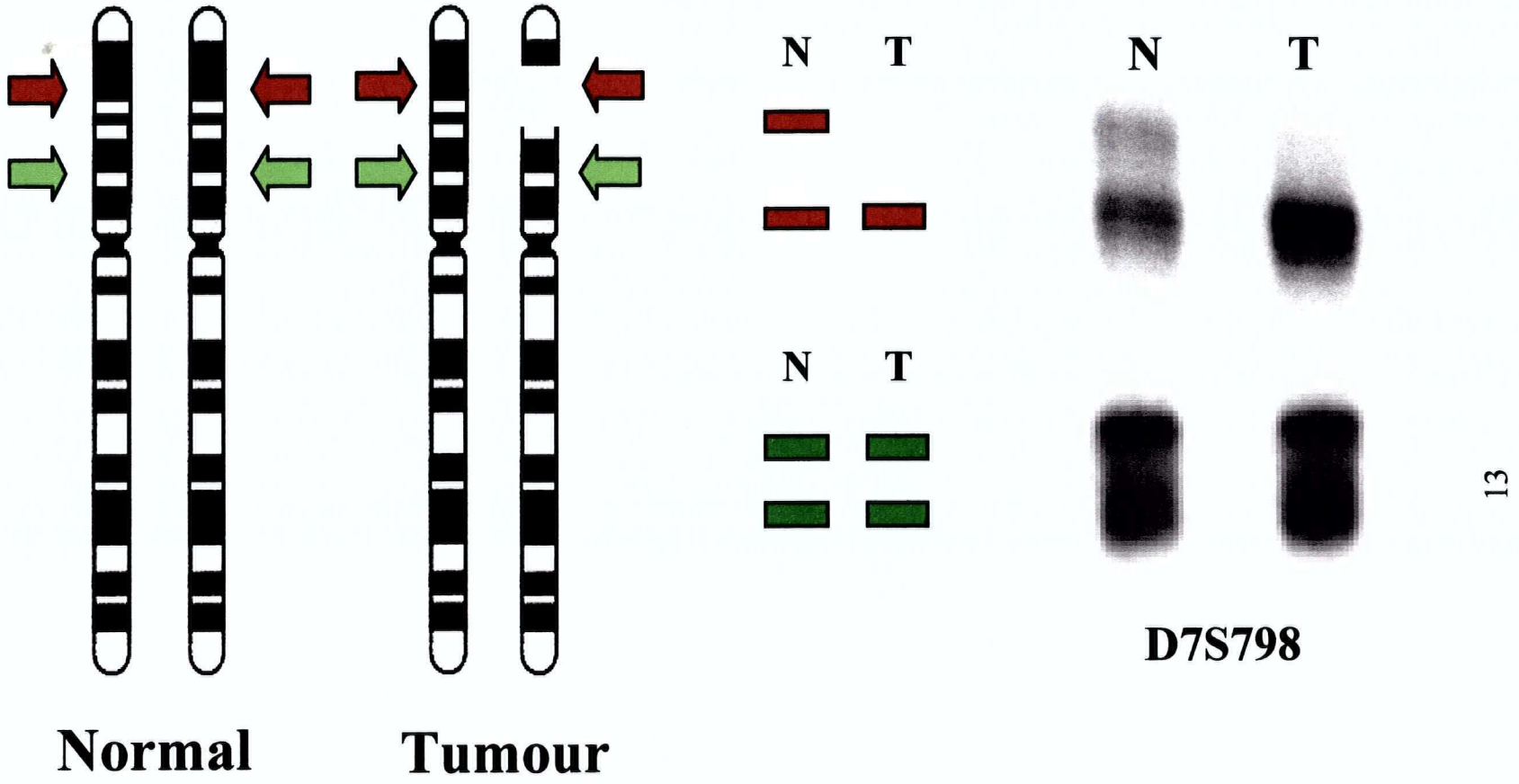


Figure 2: Loss of Heterozygosity (LOH). Polymorphic microsatellite markers are used to detect allelic imbalance at specific chromosomal regions. Allelic imbalance is caused by chromosomal gains, deletions and amplifications.

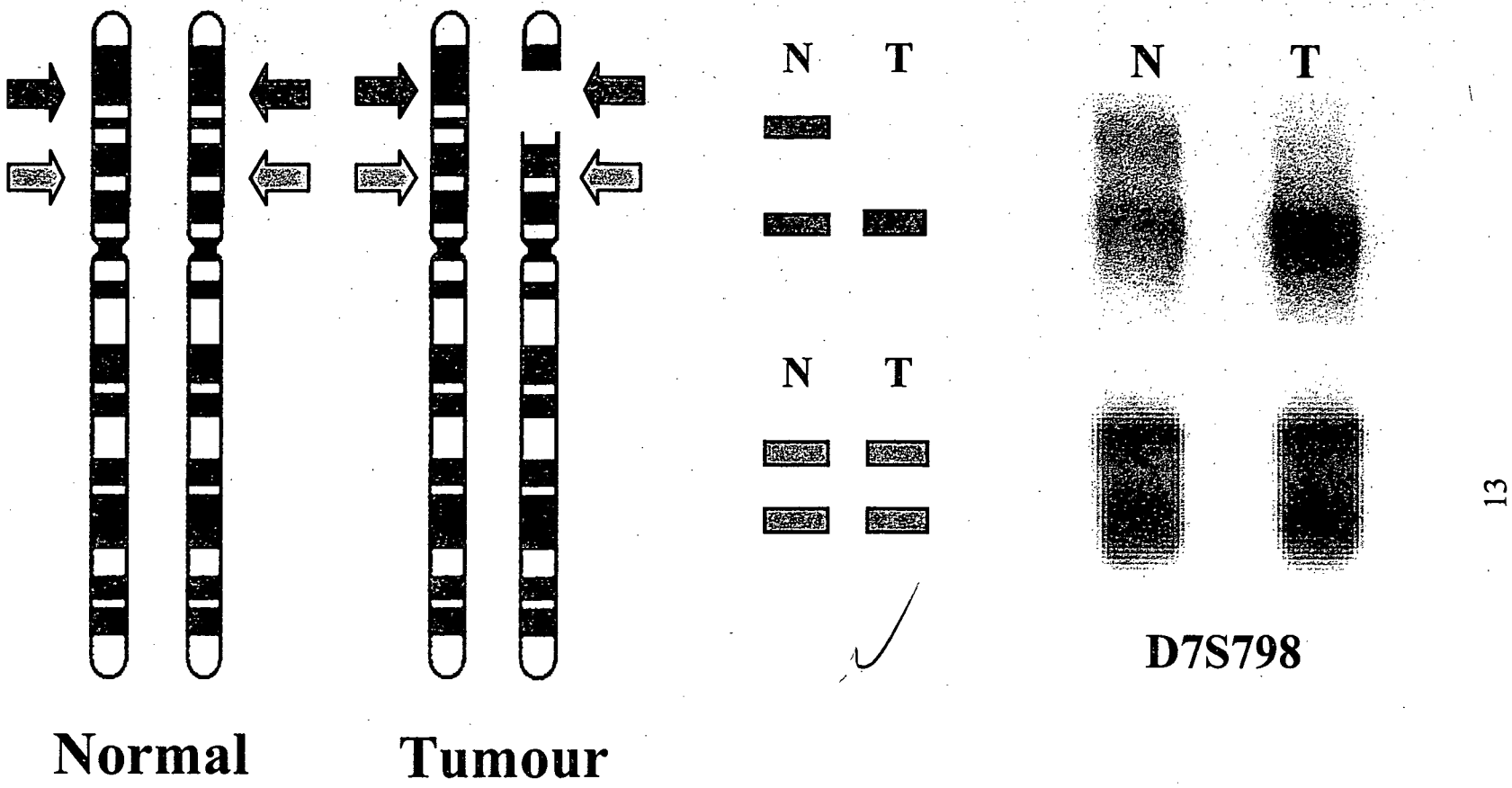


Figure 2: Loss of Heterozygosity (LOH). Polymorphic microsatellite markers are used to detect allelic imbalance at specific chromosomal regions. Allelic imbalance is caused by chromosomal gains, deletions and amplifications.

allelic imbalance at 1p, 3p, 4p, 4q, 5q, 6q, 8p, 9p, 10p, 10q, 13q, 15q, 17p, 18q and 19p (Shivapurkar, 1999; Wu, 1998; Wistuba, 1999; Nomoto, 2000). Zhou et al demonstrated that LOH could be used to determine tumour behaviour as increased allelic imbalance correlated with high frequency of tumour invasiveness and reduced patient survival.

Like CGH, LOH has many limitations that hamper its ability to screen the entire genome in search of novel regions of chromosomal instability. Firstly, each microsatellite marker is informative at only one locus making it inefficient in cases where limited microdissected DNA is available. Secondly, although microsatellite markers are distributed throughout the genome, some chromosomal regions do not contain markers preventing LOH analysis. Thirdly, using known polymorphic markers discovered in other cancers prevents discovery of novel regions of chromosomal instability.

1.7 Archival Tissue Use in Search of Novel Markers

Archived paraffin-embedded tissues (PET) are a vast resource of material that is being increasingly used for retrospective studies. Since cancers develop through the accumulation of somatic genetic changes that accompany histopathological progression from normal epithelium to invasive neoplasia, it is important to have tissue material from many patients and many histopathological grades to link an initial gene mutation critical to increased risk of progression thereby identifying prognostic markers.

Unfortunately, DNA quality from PET is influenced by several factors: fixative type, fixation time, and storage time. As it is believed that fixation is the primary source of DNA degradation in PET, Greer et al. (1994) evaluated a broad set of

fixatives and fixation time to discover conditions that are the least degradative. Acetone or 10% neutral-formalin were observed to be the best fixatives with minimum fixation times. Further, acetone was observed to be best fixative for long-term storage.

As many PET samples are degraded, determination of DNA quality and quantity is important. Researchers generally assess DNA quality using gene specific amplification of varied size fragments. Assays of this type tend to waste non-renewable DNA as multiple reactions are necessary to determine the extent of DNA degradation.

DNA quantitation of minute amounts of DNA is important in any study using archived paraffin-embedded tissues. Researchers use a variety of methods to measure nucleic acid concentrations including absorbance at 260nm (A_{260}) (Sambrook, 1989), fluorescent nucleic acid stains such as PicoGreen (Molecular Probes, Eugene, OR) (Serth, 2000) and Southern blot hybridization (Sambrook, 1989). Nucleic acid quantitation using A_{260} measurements is disadvantageous due to the assay's sensitivity limits, contamination interference by proteins and inability to distinguish DNA from RNA. PicoGreen quantitation of DNA is more sensitive and efficient but the technique is still influenced by contamination. Southern blot hybridization is efficient at quantitating DNA and is not influenced by contamination as the quantitation is dependent on DNA: DNA hybridization.

1.8 Tissue Heterogeneity and Microdissection

Pure cell populations represent the new focus of molecular analyses as native tissue environments have a high level of heterogeneity that affect cell specific studies. Unlike in the past when tissue studies involved crushing fresh tissue and assaying the

extracted molecules, isolating pure cell populations is more complicated as tissue is composed of a large number of various cell types organized in a three dimensional structure. For example, a lung biopsy of carcinoma may contain many cells including endothelial cells and fibroblasts in the stroma, normal epithelium, mild to severe dysplastic cells, premalignant carcinoma in situ lesions and clusters of invasive carcinoma cells. In any study, if the genetic focus is harvesting invasive carcinoma cells, this subpopulation may only represent a small fraction of the total tissue mass. Conducting genetic studies on one population of cells using tissue material that is contaminated by 80% with other cell subpopulations results in the masking of many observations. In the past, pure cell populations were studied by culturing specific cells from fresh tissue. Although this afforded the researcher with pure cells, it hampered native gene expression studies as cultured cell gene expression was influenced by the culture environment. In order to overcome the many problems surrounding tissue heterogeneity, microdissection was developed as a technique to separate individual cell populations. Microdissection involves the manual guidance of a needle to scrape cells of interest from a thin tissue section. With the advent of technology, there are many new devices such as the Laser Capture Microdissection (LCM) system that fuse the technique with a laser-assisted platform to make microdissection less operator dependent (Simone,1998).

1.9 Visualization of Morphology within Biopsies

The identification of benign, premalignant and malignant lesions is important in a progressive comparison of tumourigenesis. As most target lesions appear as small foci of cells, microdissection is important to overcome tissue heterogeneity.

Microdissection requires the precise identification of specific cellular features by histological stains such as haematoxylin and eosin (H&E) to be successful. Haematoxylin and eosin staining is the most common technique used in routine pathology. The basic dye, haematoxylin, stains acidic structures blue whereas the acidic dye, eosin, stains basic structures pink. Nuclei, ribosomes and rough endoplasmic reticulum have a high content of nucleic acids so the structures stain blue. Cytoplasmic proteins are basic therefore the cytoplasm stains pink. Although the stains distinguish morphology, there is a concern about the effects of the stains on tissue sample contents. Using an assay for DNA amplification by PCR, Murase et al report that several stains including hematoxylin have an adverse effect. It was proposed that the stains interfere with proteinase K digestion by binding DNA or influence the divalent cation concentration critical to *Taq* polymerase activity (Murase, 2000). A similar study confirmed that haematoxylin is detrimental to tissue for PCR amplification and that two stains, methyl green and nuclear fast red do not interfere with DNA amplification while maintaining their ability to distinguish chromatin within the nucleus (Burton, 1998). However, it is unclear if the integrity of DNA influences DNA amplifications.

1.10 Use of RAPD-PCR to Search for Novel Genetic Alterations

Although locus-by-locus allelotyping techniques such as LOH have been effective in discovering critical genetic events in cancer tumourigenesis, a technique that would screen multiple loci would be more appropriate for a genome scanning study. Developed by Williams et al., Randomly Amplified Polymorphic DNA (RAPD)-PCR (commonly

known as Arbitrarily Primed PCR (AP-PCR)) is a DNA fingerprinting technique that involves the PCR amplification of random fragments of genomic DNA with short oligonucleotides of arbitrary sequence (Micheli, 1994). The PCR amplification can be visualized as two short primers, generally 10 nucleotides in length, that anneal to many complementary sites under low stringency throughout the genome and providing that the primers are within a few hundred base pairs of each other and on opposite strands, amplification occurs (Figure 3). When amplification does not occur, the absence of product is either due to mutation, insertion or deletion of the DNA region. Polymorphisms between individuals are detected as DNA fragments that are amplified from one individual but not from another resulting in differences in the pattern of DNA fragments. A large number of polymorphic markers can be obtained by simply changing the primer sequence without previous sequence knowledge of any loci. RAPD-PCR is a very powerful technique used for a variety of purposes such as gene mapping, population and pedigree analysis, phylogenetic studies and the identification of bacterial strains. For example, RAPD was used to identify markers linked to the *Pseudomonas* resistance gene in the tomato (Martin, 1991), perform single-tree genetic linkage mapping in conifers (Tulsieram, 1992) and the estimation of outcrossing rates in *Dastica glomerata* (Fritsch, 1992).

RAPD-PCR simultaneously reveals about 50 arbitrarily sampled sequences per reaction when displayed on a denaturing polyacrylamide gel, revealing mutations that accumulated either somatically or differences in the relative abundances of corresponding sequences (Welsh, 1995). RAPD-PCR sensitivity is related to the types

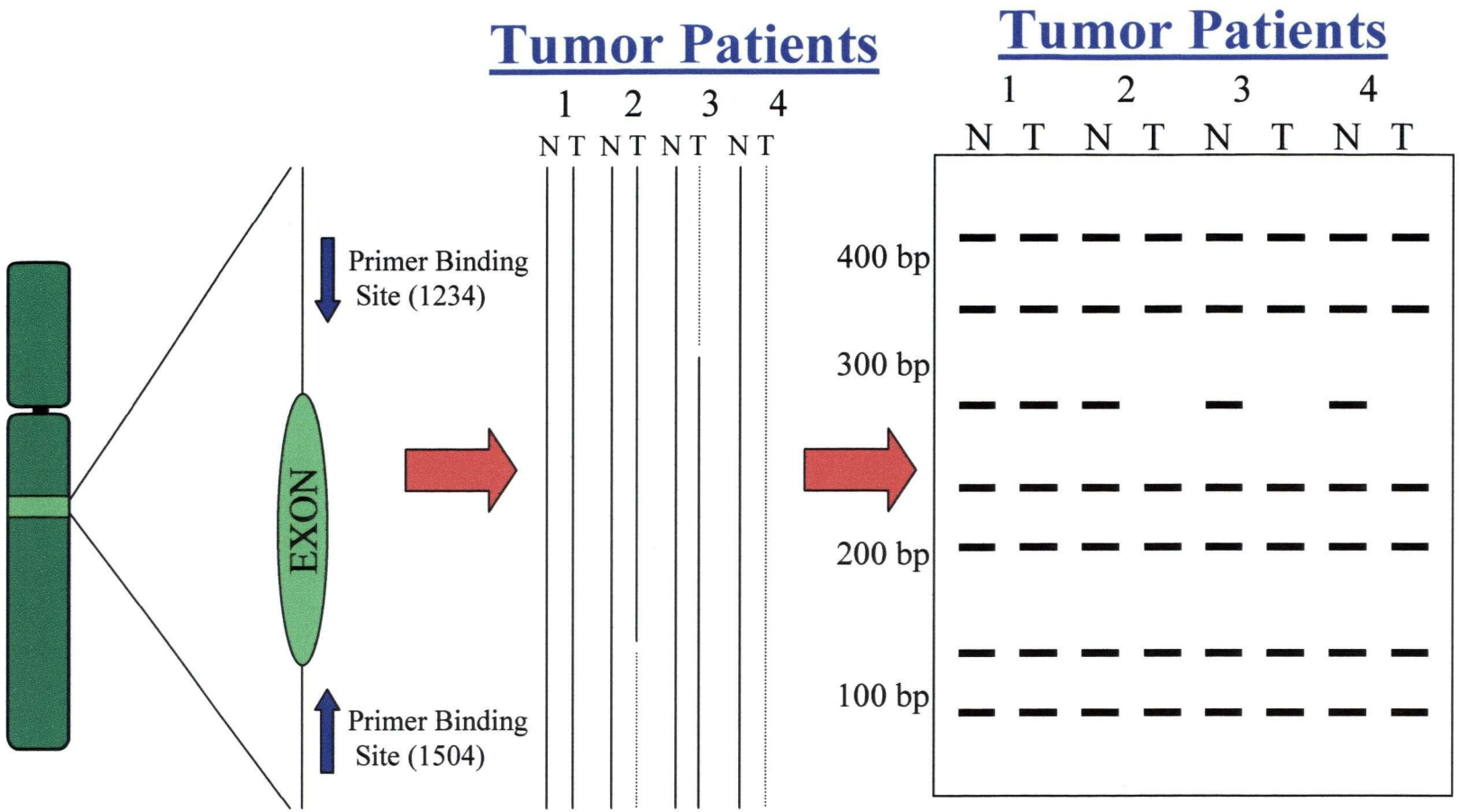


Figure 3: Randomly Amplified Polymorphic DNA (RAPD)-PCR. RAPD-PCR involves the simultaneous screening of multiple genetic loci via the high cycle amplification of genomic DNA using short 10 base pair primers. Minimum DNA quantity required is 0.5 nanograms.

of genetic alterations. There are three ways that mutations affect RAPD-PCR fingerprinting: altering the ability of the primer to anneal, altering the distance between the two primers and altering the relative amplification targets. Mutations that affect primer annealing result in the gain or loss of a band or a change in signal intensity. Single base mutations in the primer-binding site are more critical in shorter primers and in the 3' end of the primer. Deletions and insertions that alter the distance between primer binding sites result in the mobility shift of products or the loss of amplification, however, the detection of single base mutations is not possible (Williams, 1993). Ploidy is detectable as the relative band intensity is representative of corresponding template sequence. For example in heterozygotes, loss of a region would result in a loss of band as the only allele is lost. In homozygotes, a loss of a region would result in a 50% change in band intensity as one allele is lost. RAPD is a powerful tool for the detection of somatic genetic alterations during tumourigenesis. RAPD is applied to the detection of qualitative (structural) and quantitative (aneuploid) genetic alterations (Welsh, 1995).

RAPD has had applications in cancer research because of its semi-quantitative ability to detect copy number changes from chromosomal aberrations such as gains and losses of DNA sequence. The differences in RAPD band intensity between normal and tumour DNA provide an estimation of tumour aneuploidy. For example, Kohno et al. discovered a homozygous deletion on chromosome 2 in a lung cancer cell line. Losses and gains of DNA sequence can be identified because of the linkage to oncogenes and tumour suppressor genes. Ong et al. (1998) examined genomic instability in NSCLC and SCLC using several RAPD primers. It was found that genomic instability was

detected in 19/20 normal/tumour pairs. De Juan et al. (1999) analyzed 65 NSCLC patients in search of genomic alterations and found that 64% of all alterations were gains (increased band intensity) rather than losses (lowered or no band intensity). Kawakami et al.(1998) fingerprinted 44 NSCLCs using AP-PCR to find a loss of band intensity in 15 tumours that corresponded to the long arm of chromosome 10; the loss was confirmed with allelic imbalance analysis. De Juan et al. (1999) screened 65 NSCLC patients using AP-PCR to find that 45% of the sample set contained an amplified DNA fragment at 6p12. Anami et al. (2000) adapted the AP-PCR technique to fingerprint methanol-fixed lung cancer tissues and discovered DNA sequence losses at chromosome 7 and 22 in 41.7% of adenocarcinomas and 84.6% of small-cell carcinomas, respectively. Chromosome 22q13.3 losses were confirmed with microsatellite markers in 11 of 13 small cell carcinomas. They also reported DNA amplifications at chromosome 1, 8, 13 in 40% of adenocarcinomas and chromosome 2 in 63.3% of squamous-cell carcinomas. Yamada et al. (2000) used AP-PCR to detect chromosomal imbalances in 13 SCLCs and discovered gains and losses corresponding to chromosomes 1, 7, 8, 16 and chromosomes 2, 10 and 22, respectively. The most common band intensity gain corresponded to a DNA amplification at 8q24. Amplification of the C-MYC gene was confirmed in five of eight tumours analyzed, suggesting that another potential oncogene was amplified in the region. RAPD-PCR is an effective means of determining chromsomal alterations.

1.11 Objectives and Hypothesis

The first objective of this thesis was to deduce suitable conditions for the microdissection of formalin-fixed, paraffin-embedded archival lung tissue. The rationale was to identify a histological stain conducive to maintaining DNA integrity while enabling morphological distinctions within the biopsy. The second objective was to isolate DNA from progressive stages of lung carcinoma in order to conduct a stage-specific study. In collaboration with Dr. Stephen Lam at the BC Cancer Agency, a large cohort of biopsies from a large number of patients was obtained and microdissected. The third objective was to identify genetic alterations using a genome scanning technique known as RAPD-PCR. Hundreds of loci were analyzed using this technique. The fourth objective was to characterize genetic alterations by cloning and sequencing in hope of identifying critical genetic events in lung tumourigenesis. The fifth objective was to localize sequenced products to chromosomal regions and the sixth objective was to verify these regions for chromosomal instability.

These objectives were created in order to support the two-fold hypothesis: 1) RAPD-PCR will identify genetic alterations in NSCLC by simultaneously screening multiple loci from specific grades of archival, paraffin-embedded lung tissue and 2) evaluation of alterations will reveal chromosomal regions of instability.

Chapter 2: Materials and Methods

2.1 Tissue Microdissection and DNA Extraction

A total of 172 paraffin-embedded archived lung samples were provided by Dr. S. Lam. This set included 33 normal, 16 hyperplasia, 24 mild to severe dysplasias, 74 carcinoma in situ (CIS) and 25 invasive carcinomas. In all, 41 patients were represented. The patient cohort consisted of 27 men, 14 woman, 14 current smokers, 21 ex-smokers, and 4 non-smokers with a mean age of 69 and a mean smoking history of 58 pack years. Smoking history was not available for two patients. The rationale for the patient cohort was to include all available lung biopsies. All samples were histologically graded by Dr. J. LeRiche. Serial sections were cut from each paraffin block and stained with Methyl Green. Every seventh section was stained with hematoxylin and eosin and evaluated by the pathologist. Manual microdissection was performed by C. Dawe and myself. Tissue was digested in a digestion buffer (10mM Tris, 1mM EDTA pH 8.0, 0.5% sodium dodecyl sulfate (SDS) and 50mM NaCl) at 55°C for 72 hours with a fresh 20μg proteinase K spiking every 24hours. DNA was extracted twice with phenol/chloroform/isoamyl alcohol followed by precipitation with ethanol using 10μg glycogen as a DNA carrier. DNA material was extracted from blood in exactly the same manner once the cellular components were pelleted.

2.2 DNA Quantification

A 5% aliquot of unquantified DNA sample was run out on a 1% agarose gel along with known quantities of DNA, denatured and transferred to HyBond nylon membrane

(Amersham Pharmacia Biotech). The membrane was dried at 80°C for 30min and probed with 200 nanograms (ng) total genomic lung DNA. The probe DNA was labelled using a Random Priming Kit (Gibco BRL Inc.) as follows: 200ng denatured DNA, 14mM of each dCTP, dGTP and dTTP, 1X Random Priming Buffer (0.67M HEPES, 0.17M Tris-HCl, 17mM $MgCl_2$, 33mM 2-mercaptoethanol, 1.33 mg/ml BSA, 18 OD_{260} units/ml oligodeoxyribonucleotide primers, pH 6.8), 25 μCi $\alpha^{32}P$ ATP. The probed membrane was washed in 2X SSC (5mins), 2X SSC, 1% SDS (3 x 15 mins) and 0.2X SSC, 0.1% SDS (2 x 30mins). The membrane was exposed on Kodak X-OMAT AR autoradiography film and on a phosphoimager cassette for further analysis. The cassette was analyzed by the STORM phosphoimager and using ImageQuant 5.0 software (Molecular Dynamics), the DNA concentration was determined.

2.3 Gene-specific PCR

Two sizes of fragments were amplified from the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene (GenBank accession no. J04038) in 25 ul reaction mixtures containing: 1X PCR Buffer (50mM KCl, 10mM Tris-HCl pH 9, 0.1% Triton X-100), 2 mM $MgCl_2$, 0.2 mM deoxyribose trinucleotides (dNTPs), 20 pmoles of each GAPDH primer, 2.5 Units Taq DNA Polymerase and 2 ng of DNA. The PCR primers were: GAP1 (5'-AACCTGCCAAATATGATGACATCA-3'); GAP2 (5'-GTCGTTGAGGGCAATGCCA-3'); and GAP3 (GTCTTACTCCTTGGAGGCCATGA-3'). GAP1 and 2 produce a 163 bp fragment while GAP1 and 3 produce a 363 bp fragment (see Results, Fig.7A). PCR conditions were as follows: Denaturation at 94°C for 2 min, followed by 35 cycles of amplification (94°C for 45 sec, 60°C for 45 sec and

72°C for 2 min) and a final extension (72°C for 10 min) in a PTC-100 Thermal Cycler (MJ Research).

2.4 RAPD-PCR

Each RAPD-PCR fingerprint consisted of multiple patients of various histological grades analyzed with a favorable primer pair. All oligonucleotide primers were purchased from Alpha DNA Inc. In each reaction, 20 picomoles of each primer were end-labelled with 2μCi of adenosine 5' [γ-^{32} P] ATP (6000 Ci/mmol; Amersham Pharmacia Biotech), at 37°C for 1 hr and 65°C for 5 min. Each PCR reaction contained 2 ng of DNA, 200μM each dNTP, 10mM Tris-Cl, pH 8.3, 50 mM KCl, 2mM $MgCl_2$, 0.001% gelatin, and 2.5 units of Recombinant Taq DNA polymerase in a final volume of 10ul. All reactions were amplified for 45 cycles (94°C for 1min, 35°C for 1min, 72°C for 2min) in a thermal cycler (MJ Research PTC-100). All reaction products were run on 6% non-denaturing polyacrylamide gels at 800V for 3 hrs. Dried gels were exposed on Kodak X-OMAT AR autoradiography film. In each experiment, positive and negative controls consisted of DNA extracted from a frozen lung tissue and a reaction without DNA template, respectively. All noticeable PCR signal changes that recurred in many patients were further analyzed.

2.5 Cloning

Fragments of interest were cut out of the dried polyacrylamide gel after overlaying the exposed film against the dried gel with orientation markers. DNA was eluted into 20μl dH_2O by boiling for 10 minutes. Ten percent of eluted material was amplified using primers matching the original RAPD-PCR primer sequence but containing a Bam H1 or PST1 recognition sequence. As the RAPD-PCR fragments were

created using two possible primers, it was important to determine the terminal sequence of the products. Therefore, the reamplification reactions consisted of three possible primer combinations: primer 1 with itself, primer 1 and primer 2, and primer 2 with itself. Aliquots of PCR products were resolved on 1% agarose gels to determine which restriction enzyme was needed to prepare the product for insertion into a plasmid vector. PCR reactions contained 1X Mg-free PCR buffer (50mM KCl, 10mM Tris-HCl pH 9, 0.1% Triton X-100), 0.5mM dNTPs, 2 units recombinant Taq polymerase, 1.9 mM $MgCl_2$ and 10 picomoles of each primer. All reactions were performed using the MJ cycler. Three cycles of 94°C for 1min, 35°C for 1min and 72°C for 2min were first performed followed by 25 cycles of 94°C for 40 sec, 62°C for 40 sec and 72°C for 40 sec.

PCR products were digested by the appropriate enzyme (Bam H1 or Pst 1) and purified by excising from a 5% polyacrylamide gel. DNA was eluted from the excised gel piece using dH_2O at room temperature. The insert was quantified and combined with cut and dephosphorylated PSK^+ plasmid for subsequent ligation at 16°C overnight. Transformations involved combining ligation with heat shock competent DH5α E.coli. cells (Bergmans, 1981). Cells were plated on LB Amp^+ plates containing IPTG and X-Gal and grown at 37°C overnight.

Growing colonies containing an insert were white while colonies without an insert were blue. Ten white colonies (as a minimum) were subjected to colony PCR to confirm that the appropriate size insert was present in the plasmid. Colony PCR contained 1X Mg-free PCR buffer (buffer (50mM KCl, 10mM Tris-HCl pH 9, 0.1% Triton X-100), 0.25mM dNTPs, 2 units recombinant Taq polymerase, 5 picomoles each of M13 -47 and

-48 primers (New England Biolabs) and 2mM $MgCl_2$. Colony PCR reaction involves an initial denaturation at 94°C for 3 mins followed by 30 cycles at 94°C for 45 secs, 40°C for 45 secs and 72°C for 45 secs. Colonies that had appropriate size inserts were subjected to colony fingerprinting to determine that the colonies had the same sequence. Colony fingerprinting reaction contains 1X FP nucleotide stock (0.5mM dNTPs, 0.5mM ddGTP), 1X FP buffer (10mM Tris pH 8, 50mM KCl, 1.5mM $MgCl_2$), 0.5 pmol ^{32}P ATP labelled T7 primer, 2.5 units recombinant Taq polymerase, and 1ug RNase A. The T7 primer radiolabelling reaction contains 0.34 pmol $\gamma^{32}P$ dATP, 1 X PNK buffer, and 0.5 pmol T7 primer and involves incubation at 37°C for 1 hour followed by 65°C for 5mins. Colony fingerprinting reaction involved an initial denaturation at 94°C for 3 mins followed by 30 cycles at 94°C for 1 min, 40°C for 1 min and 72°C for 1 min. (Krishnan, 1991).

Plasmid preparation was performed on one representative colony (see Appendix for detailed protocol). The prepared plasmid was sequenced using the ABI Prism at the Nucleic Acids Protein Service (NAPS) Unit at the University of British Columbia (see Appendix for detailed protocol.

2.6 Localization of clones to a chromosomal position

Sequenced clones were matched to the partially sequenced human genome using NCBI standard nucleotide-nucleotide BLAST (blastn) at [www.ncbi.nlm.nih.gov]. Matched results confirmed chromosome locations for each clone.

2.7 Choosing Microsatellite Markers

In order to test a cloned region for allelic imbalance, microsatellite markers must be determined for each chromosomal region. Using the University of Southern

California genome database [http://genome.ucsc.org] to identify microsatellite markers surrounding the sequenced clone region and the Genome Database [http://www.gdb.org] to identify primer sequences and frequency of heterozygosity, microsatellite primers were chosen and ordered from Alpha DNA Inc.

2.8 Loss of Heterozygosity (LOH) Assay

One primer (2ng) from each LOH primer pair was end-labelled with 20μCi γ ^{32}P ATP using T4 polynucleotide kinase (PNK) at 37°C for 1 hour followed by 65°C for 5 mins. PCR amplification involved 4 ng genomic DNA, 2 ng labeled primer, 20 ng of each unlabelled primer, 0.5mM of each dATP, dCTP, dGTP, dTTP, 0.5 units of Taq DNA polymerase (Gibco BRL) and PCR buffer [16.6mM ammonium sulfate, 67mM Tris (pH 8.8), 6.7mM magnesium chloride, 10mM ß-mercaptoethanol, 6.7mM EDTA, and 0.9% dimethyl sulfoxide]. PCR amplification was performed for 35 cycles consisting of denaturation at 95°C for 30 sec, annealing at 50-60°C for 60 sec, and extension at 72°C for 60 sec. PCR products were separated on 6% urea-formamide-polyacrylamide gels run at constant power (50W) for 2.5 hours. The dried gels were exposed on Kodak X-OMAT AR autoradiography film.

The gels were analyzed by determining whether there was a signal intensity change for the two alleles seen in the normal DNA compared to tumour DNA. A signal intensity decrease of 50% was classified as an allelic loss. All samples showing allelic loss were subjected to repeat analysis after a second independent amplification.

Chapter 3: Results

3.1 Determining a Patient Pool

In order to determine prognostic markers for lung cancer, a disease that affects hundreds of thousands of individuals each year and develops over many decades, it is important to establish an unbiased, patient set. Ideally for a complete genetic study, the patient set should include all relevant information including age, sex, smoking history, recurrences and treatment.

The patient pool consisted of non-small cell lung carcinomas (NSCLCs) collected during fluorescence bronchoscopy by Dr. Stephen Lam (chair of the lung tumour group at the BC Cancer Agency) using Lung Imaging Fluorescence Endoscope (LIFE) system (Xillix Technologies Corp.). All biopsies were classified by a lung pathologist, Dr. Jean LeRiche, according to a histological grading system (Table 1). Biopsies were fixed in formalin, paraffin-embedded, and mounted on glass slides by Chris Dawe. Staining was performed as described in "Materials and Methods". (see Appendix for staining protocol).

Metaplasia is characterized as a maturation step where cells appear to be changing from one histological cell type to another. As seen in Table 1, the two possible grades for each early-stage pathological class are distinguished by the presence of absence of metaplasia (eg., 3.1 or 3.2, 4.1 or 4.2, 5.1 or 5.2 and 5.3 or 5.4). Cells that appear squamous are arranged in a flattened epithelial appearance whereas glandular appearance exhibits cells in a gland-like fashion. The two possible grades seen in the progressed

stages distinguish between squamous or glandular appearance (eg., 6.1 or 6.2, 7.1 or 7.2 and 8.1 or 8.2).

Table 1: Relevant Histological Grades in NSCLC

Pathological Class	Cellular Features	Grade
Normal	Pseudostratified ciliated columnar epithelium	1.0
Hyperplasia	Increase in the number of normal-appearing basal cells with normal ciliated or mucin secreting cells	3.1, 3.2
Mild Dysplasia	Stratified epithelium, hyperchromatism, slight increase in nuclear:cytoplasmic ratio, limited to lower 1/3 of epithelium	4.1, 4.2
Moderate Dysplasia	Stratified epithelium, hyperchromatism, larger increase in nuclear:cytoplasmic ratio, mitotic figures seen, limited to lower 2/3 of epithelium	5.1, 5.2
Severe Dysplasia	More pronounced changes than seen in moderate dysplasia	5.3, 5.4
Carcinoma *in situ*	Multi-layered epithelium with malignant cells but with an intact basement membrane	6.1, 6.2
Microinvasive	Cells have breached the basement membrane into the subepithelial layer; confined within layer above cartilage	7.1, 7.2
Invasive Carcinoma	Multi-layered epithelium with malignant like cells invading the underlying stroma	8.1, 8.2

3.2 Determining a Method for Obtaining Pure Cell Populations

Lung tissue is composed of various types of cells including epithelial, stromal, and inflammatory cell populations. Previous studies use gross tissue samples for genetic analyses risking the masking of specific genetic events by heterogenous tissue. In order to define stage specific markers for tumour progression, tissue heterogeneity had to be addressed. Microdissection was considered to be the only method for procuring specific cell populations. Two microdissection techniques were considered: manual microdissection and laser capture microdissection (LCM).

3.2.1 Manual Microdissection

Guided by many previously published studies using manual microdissection, an inverted Nikon microscope fitted with an X-Y-Z micromanipulator was made available by Dr. Calum MacAulay. Although many reports mentioned the use of a needle attached to a micromanipulator, we decided to practice the technique using our own manual dexterity. After several hours of practice and experimentation with various objects and needle sizes, it was determined that a 30 gauge needle was sufficient to dissect individual cells.

Manual microdissection was attempted using various solutions such as water, buffer and alcohol. The effect of the solution was to provide a moist medium to prevent the static repulsion of the tiny tissue flakes from the needle. It was determined that the 'wetting' of the slide with 80% EtOH prior to immediate dissection was most appropriate. In fact, not only did the alcohol provide moisture, as the alcohol dried on the slide it would cause the scraped tissue to adhere to the needle allowing safe transfer to the digestion vial. Figure 4 shows an example of two tissue biopsies before and after microdissection. In figure 4A, hyperplastic cells are present in three distinct areas in the epithelium. Figure 4C shows the biopsy after the removal of the hyperplastic area using manual microdissection. It is clear that the basement membrane is still intact confirming the lack of contamination by nonepithelial cells. In figure 4B, there is an invasive tumor within a deep layer of stroma. The CIS cells are microdissected and the biopsy post-dissection is seen in figure 4D.

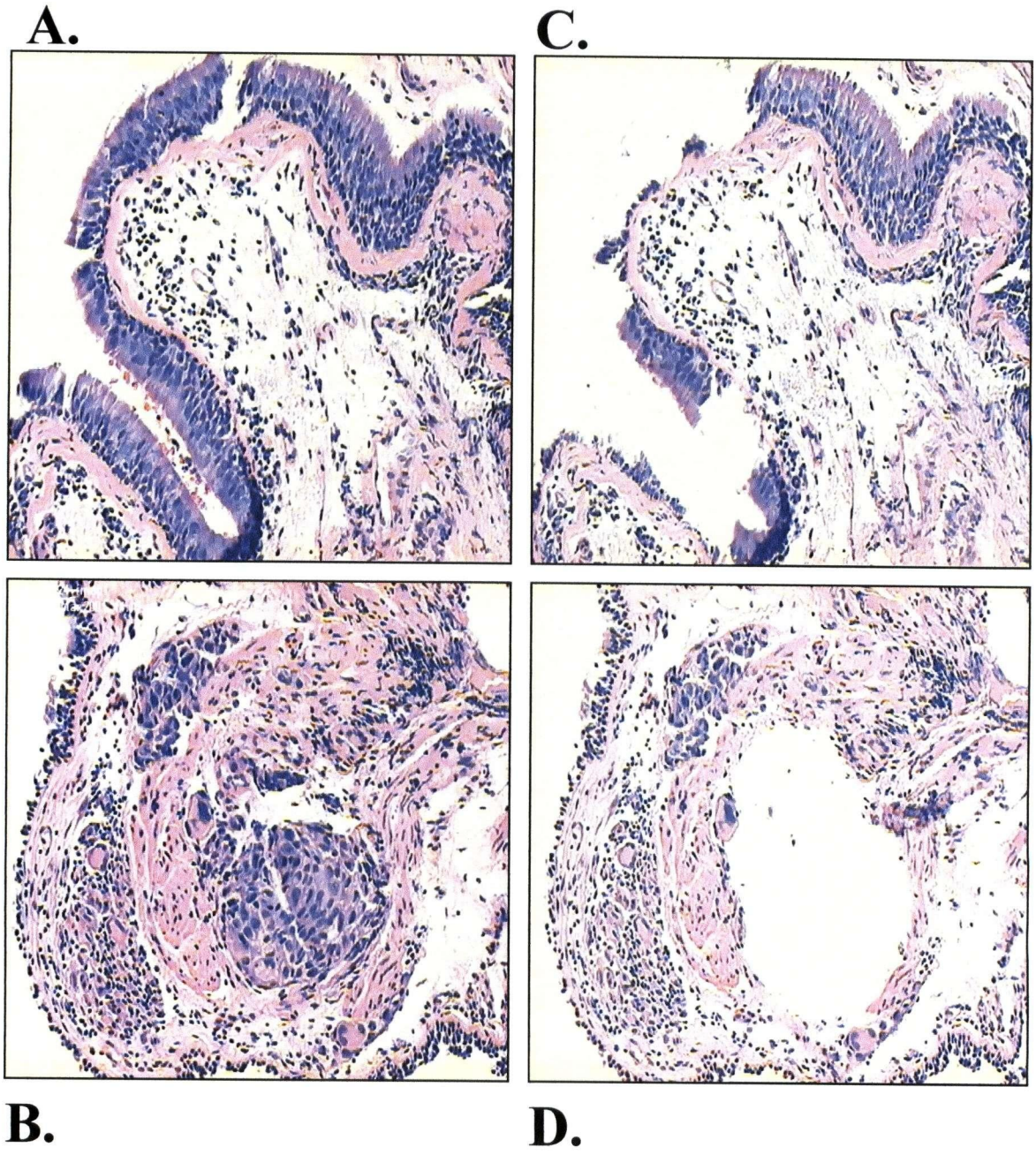


Figure 4: Manual Microdissection of Lung Biopsies.
A. Lung biopsy (before microdissection) with hyperplastic regions. B. Lung biopsy with invasive tumor. C., D. Lung tissue remaining after microdissection. Intact basement membrane in C. indicates no contamination by nonepithelial cells.

3.2.2 Laser Capture Microdissection (LCM)

Since LCM was promoted to be much more advantageous than manual microdissection, the use of PixCell II (Arcturus), kindly provided by Dr. Juergen Vielkind (Cancer Endocrinology, BCCA) was considered in this study. LCM is based on the fusion of cells of interest to a plastic polymer film via melting using an operator-guided laser. Several trial experiments were preformed to compare the efficiency of the technique versus manual microdissection specific to these small tissue biopsies. It was concluded that in order to maximize DNA quantity recovered from each biopsy, the tissue material was to be freely digested in a buffer as opposed to being digested while adhering to a polymer film. Experiments to optimize DNA yield were not performed since manual microdissection was adequate for obtaining pure cell populations.

3.3 The Effects of Histological Stains on DNA Quality

DNA quality remains an important issue in many genetic studies. Lung biopsies used in this study are classified as archival material because they have been formalin-fixed and paraffin-embedded immediately after excision from the patient. A common fixative, formalin has been known to cause the nicking of DNA and subsequent degradation. Despite these poor preservation techniques, the present concern is to maintain the level of degradation without further destruction of DNA. Although it is important to visualize cell morphology within these biopsies, common histological stains such as H&E have a low pH and could be detrimental to DNA quality.

To examine the effects of histological stains on DNA, a time trial was conducted where cultured Bronchial Epithelial Cells (BEC) were stained with Hematoxylin, a component of H&E, for various time lengths similar to common biopsy staining

practices. DNA was extracted and an aliquot was run on a 1% gel. (Figure 5A) The lengths of exposure to Hematoxylin stain at its natural pH 2.59 were 1, 10, 30, 90 and 270 seconds. It is clearly shown that with increased time exposure, there is a prominent degradation of the high molecular weight genomic band seen in the unstained BEC control with increased accumulation of smaller degraded DNA material.

A similar time trial was conducted using Methyl Green stain, pH 4.18, on BEC (Figure 5B1). The high molecular weight genomic band was present in the unstained BEC control. After staining the BEC cells for 1 sec, there is noticeable degradation of the DNA; DNA degradation accumulates with increasing staining lengths. Believing that the degradation was solely caused by the low pH, we decided to neutralize the Methyl Green stain to pH 6.0 and studied the effects on DNA quality (Figure 5B2). The previous experiment was conducted using the buffered stain and the DNA was analyzed for DNA quality. The modified stain did not affect DNA quality as all staining results showed the presence of the high molecular weight genomic band without accumulated degradation.

As most histological stains are used at an acidic pH, it was important to examine whether the acidity is necessary for proper stain binding by comparing histological stains at various pH's. Three stains, H&E, Nuclear Fast Red and Methyl Green at pH 2, 4, and 6 were used to stain serial sections of a lung biopsy (Figure 6). The stained sections were examined by Dr. LeRiche for the ability to distinguish varying morphological species. Staining the biopsy with H&E (pH 2.59), the nuclei appeared blue and the stroma appeared pink. With increasing pH, the H&E stain was unable to distinguish nuclei from stroma becoming a homogenous stain. Nuclear Fast Red (pH 2.17) staining of the biopsy allowed cells to be classified according to morphology but with increasing pH, the stain

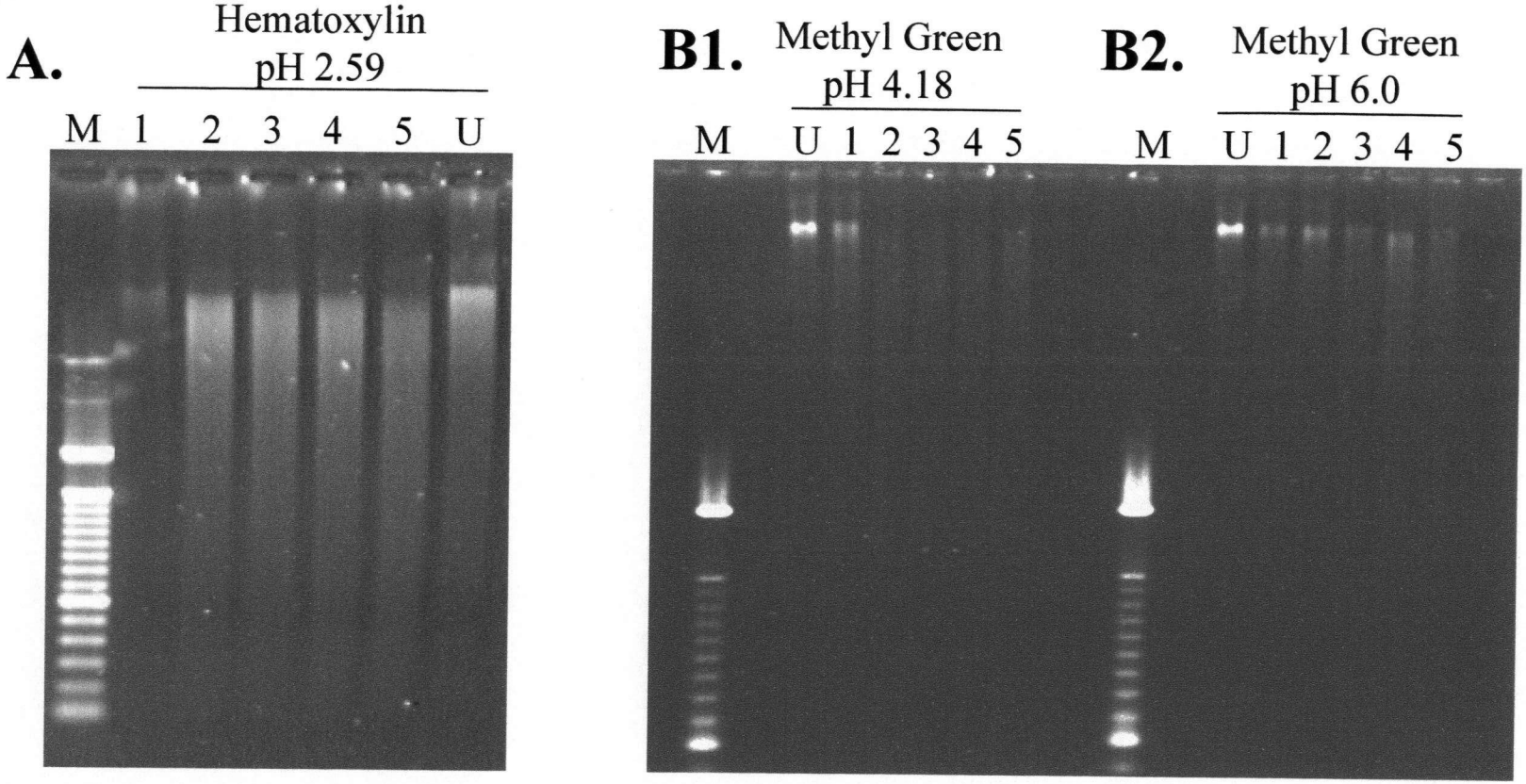


Figure 5: Assessing the Degradative Effect of Hematoxylin and Methyl Green on DNA Quality. Bronchial Epithelial Cells (BEC) slides were stained with A) Hematoxylin (pH 2.59), B1) Methyl Green (pH 4.18), and B2) Methyl Green (pH 6.0) for various time lengths including 1sec (1), 10 secs (2), 30 secs (3), 90 secs (4), and 270 secs (5). An unstained (U) slide was used as a negative control. DNA was extracted from each slide and run on 1% agarose. Methyl Green (pH 6.0) appeared to be optimal for maintaining DNA quality.

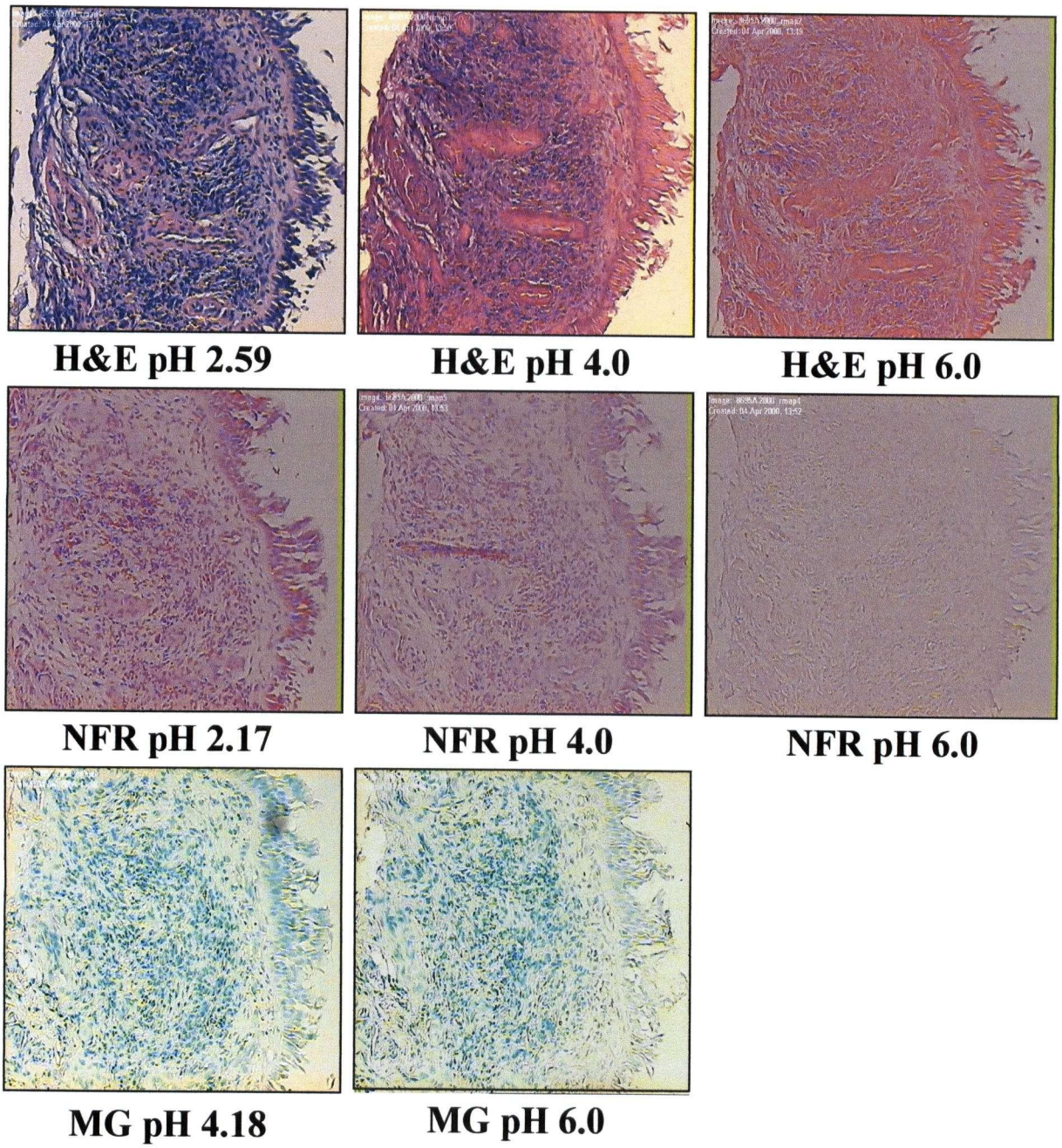


Figure 6: Determining Staining Ability of Histological Stains. Serial sections of lung biopsy stained with Hematoxylin & Eosin (H&E), Nuclear Fast Red (NFR) and Methyl Green (MG) at various pH's to assess effect on staining ability. MG pH 6.0 was the optimal stain.

became indiscriminate between cell types. Staining the biopsy with Methyl Green at its natural and buffered pH showed similar staining pattern allowing for morphological characteristics to be distinguished. Since Methyl Green at the buffered pH appeared to retain its staining ability and was not degradative to DNA, it was chosen for use with all microdissectable sections.

3.4 Developing an Assay for Assessing DNA Quality of Microdissected Archival Samples

Archival specimens vary considerably in DNA quality due to handling and storage techniques. Some DNA samples are highly fragmented or damaged by preservation procedures and require special handling in reactions, for example, the use of larger quantities of DNA in a PCR based assay, while other samples seem to be better preserved, readily yielding amplified products of larger size. Since specimen size is often a limiting factor in studying non-renewable, archival material, the ability to assess DNA quality, with a minimal amount of material, would allow the sorting of such specimens for specific analytical use. RAPD-PCR enables consistent genetic fingerprints from archival, microdissected specimens by the random amplification of multiple loci throughout the genome producing an unbiased representation of DNA quality

Using seven arbitrarily chosen oral squamous cell carcinomas provided by the Oral Biopsy Service of British Columbia, a trial experiment was designed to determine the applicability of RAPD-PCR as an alternative to gene specific PCR to determine the DNA quality of a sample prior to microdissection and further analyses.

3.4.1 Assessment of DNA Using Gene Specific PCR

DNA extracted from seven formalin-fixed, paraffin-embedded oral biopsy pairs (14 samples in total) was assessed by the amplification of two different size fragments, 163 bp and 363 bp, from the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene (Figure 7A). Figure 7B summarizes the results. Although all samples generated a 163 bp fragments only 8 yielded a 363 bp PCR product. The latter included samples from cases 402, 446, 448 and 451. In all cases, DNA extracted from normal cells in connective tissue (C) and tumour cells (T) from each specimen showed the same PCR response suggesting that the PCR limitation was a result of specimen preservation.

3.4.2 Assessment of DNA Quality Using RAPD-PCR

The same quantity of DNA was sufficient to generate RAPD profiles from the above samples (Fig. 7C). The size range of RAPD products varied between sample pairs, again with both members of the same sample pair giving a similar response. Three of the sample pairs (446, 448 and 451) had fingerprint patterns resembling the profile generated from high quality DNA extracted from cryopreserved tissue up to 500 bp. A fourth (402) gave reproducible PCR products up to 350 bp. All 4 of these sample pairs gave a 363 bp GAPDH fragment. In contrast, 405 gave a reliable profile up to 200 bp with less amplification of larger sized fragments (lighter bands) and variation between C and T. Specimens 414 and 416 gave even poorer profiles with fragments restricted to <150 bp. Increasing DNA concentration did not improve the quality of the profiles. This explains the absence of a GAPDH 363 bp fragment for these 3 samples as the DNA appears to be highly fragmented.

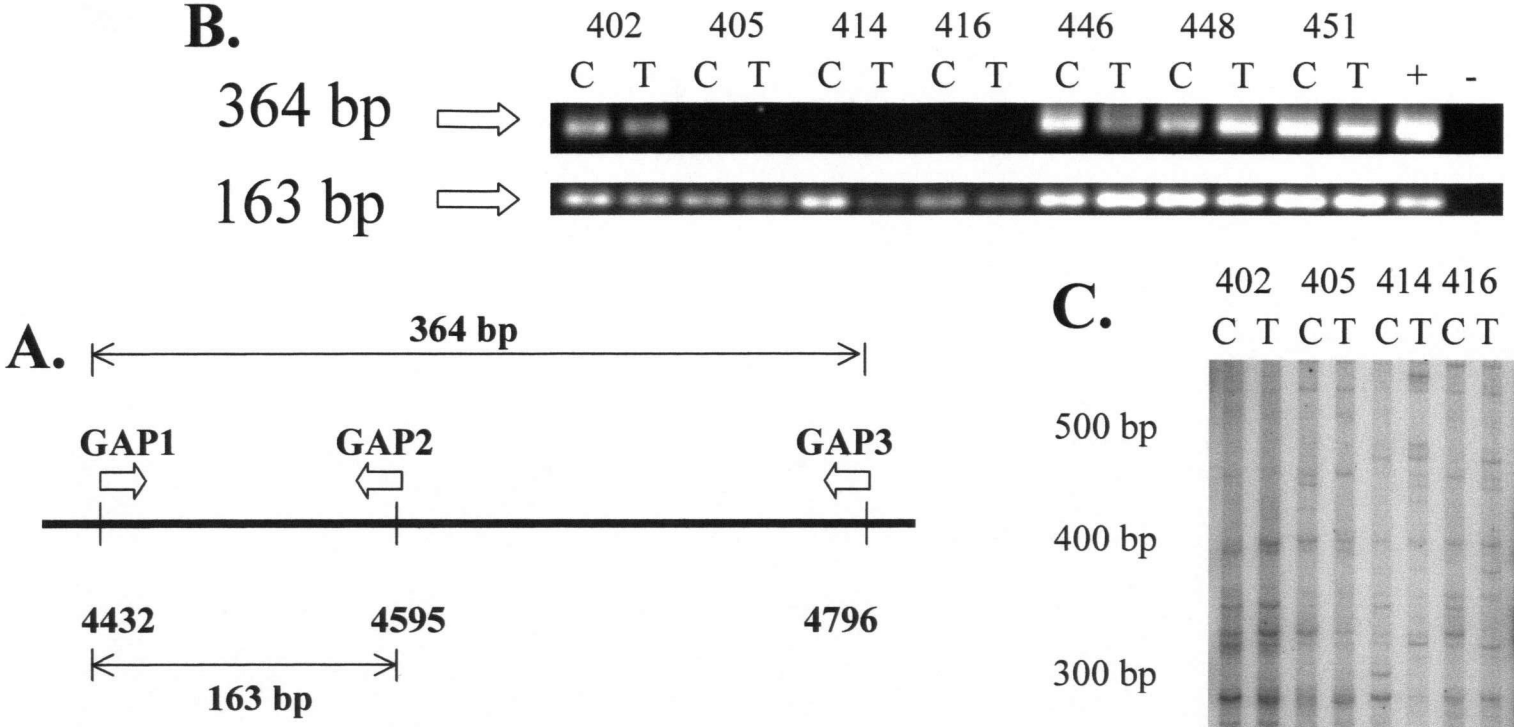


Figure 7: DNA Quality Assessment of Seven Microdissected Archival Oral Tumor Pairs. A) Primer positions on Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH) gene. B) PCR amplification (35 cycles) of a 163 bp and a 364 bp GAPDH gene fragment. C) Randomly Amplified Polymorphic DNA (RAPD)-PCR fingerprint using primer (5'-AATCGGGCTG) and (5'-GAAACGGGTG). Two nanograms was used in all reactions. +, frozen lung DNA control. -, negative control.

RAPD is an alternative manner of assessing DNA quality and it is advantageous compared to gene specific PCR assessment because it gives a clear representation of DNA quality in one reaction limiting the waste of precious microdissected DNA. As clearly shown, the same assessment requires a minimum of two reactions using gene specific PCR.

3.4.3 Screening Biopsies for DNA Degradation Using RAPD-PCR

Prior to microdissection, tissue material was scraped and digested from an unstained section of every biopsy to prescreen the DNA quality. DNA material was extracted and 1/20 of the DNA material was used in a RAPD-PCR reaction. Any biopsies unable to amplify DNA fragments greater than 200 bps were excluded from the study. All 180 microdissected cases used in this thesis consist of good quality DNA material as the early pre-selection excluded some biopsies.

3.5 DNA Quantification of Microdissected Archival Lung Biopsies Using Southern Blot Hybridization

Many techniques exist for the determination of DNA quantity in specimens. Quantitative DNA stains such as PicoGreen and spectrophotometer absorbance readings (A_{260}/A_{280}) tend to consume large amounts of DNA during quantitation and are influenced by residual nucleic acids and proteins. A Southern blot hybridization assay was proposed as a quantitation method because it was based on the hybridization of DNA to DNA on a membrane; it was more representative of DNA than absorbance readings.

A 5% aliquot of each microdissected sample was quantitated against known DNA quantities using a southern blot hybridization assay (Figure 8A). The probed membrane

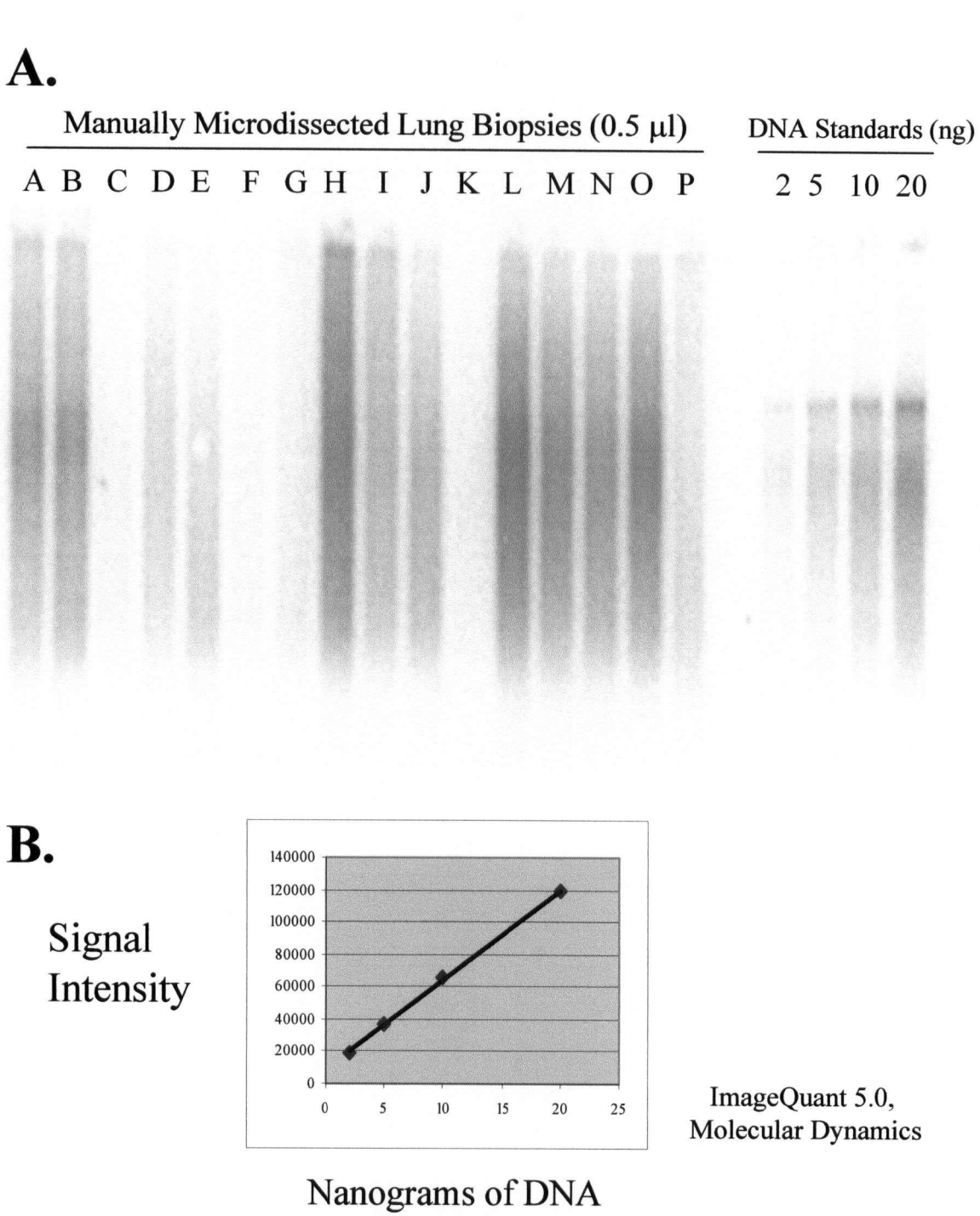


Figure 8: Southern Blot Quantitation Assay of Microdissected Lung Biopsies. A. Five percent of each microdissected sample was probed with DNA and quantitated by comparing signal intensities of unknown quantities of material with signal intensities of known quantities of DNA using ImageQuant. A-P represent microdissected samples, known DNA quantities are 2, 5, 10, 20ngs. B.Graph represents a linear relationship between signal intensity and known quantity of DNA.

was analyzed using ImageQuant 5.0 software (Molecular Dynamics) and DNA quantity was determined. The quantitation software expresses a numerical reading for radioactive signal present for each area of a lane. Numerical readings were also assigned to the signal emitted by the known quantities of DNA (2, 5, 10, 20ngs). To relate each numerical reading for an unknown quantity to the numerical reading of known quantity, the unknown quantity numerical reading was expressed as a factor of each known quantity numerical reading. The factor was multiplied by the known DNA quantity and an average DNA quantity was determined. In most cases, the relationship between numerical reading to known DNA quantities was linear (Figure 8B). DNA amount from all microdissected cases was quantitated in this manner.

Example:

$$\frac{\text{Unknown DNA Quantity Reading}}{\text{Known DNA Quantity Reading}} \times \text{Known DNA Quantity} = \text{Unknown DNA Quantity}$$

$$\frac{48671.81 \text{ (Unknown Reading)}}{53553.78 \text{ (2ng Reading)}} \times 2\text{ng} = 1.82\text{ ng}$$

3.6 Screening Oligonucleotides for Increased Amplification Density

Short oligonucleotides (10 base pairs) match variably throughout the human genome depending on DNA sequence. To achieve proper sampling of the human genome, it was expected that the primer set should amplify > 50 regions.

Sequence of thirty-six oligonucleotides (Table 2) was randomly designed with the criteria that each oligonucleotide must have a G:C content similar to the human genome, i.e. 60%.

Table 2: RAPD-PCR primers screened for increased scanning density

RAPD-PCR Primers	Sequence (5' to 3')	% GC Content
Alpha-1 (9mer)	CAGGCTATC	55.56%
Alpha-2 (9mer)	GAATCAGTG	44.44%
Alpha-3 (9mer)	AACTGGCAT	44.44%
Alpha-4 (9mer)	TCCAGCTAG	55.56%
Alpha-5 (9mer)	TTCAGTCAG	44.44%
Alpha-6 (9mer)	AACTGCATC	44.44%
Alpha-7 (8mer)	ACGTTGAC	50%
Alpha-8 (8mer)	AACTGCCA	50%
Alpha-9 (8mer)	TTCAGTCA	37.50%
Alpha-10 (8mer)	AACGTTCG	50%
Alpha-11 (9mer)	CGGCAGCAT	66%
Alpha-12 (9mer)	GGCTACCAG	66%
Alpha-13 (9mer)	CGCTTGGAC	66%
Alpha-14 (9mer)	CGCATCCAG	66%
Alpha-15 (9mer)	GCCATGGCA	66%
Alpha-16 (9mer)	CTCAGGCAC	66%
Alpha-17	GACGCCGCTT	70%
Alpha-18	ACGTCAGGCT	60%
Alpha-19	GGCTACTGCT	60%
Alpha-20	CGCCTAATGC	60%
Alpha-21	AGGCTCTCGT	60%
Alpha-22	CTAGGATCCG	60%
Alpha-23	CGCCTATGGT	60%
Alpha-24	GGCATACCGT	60%
Alpha-25	CACGGTAAGG	60%
Alpha-26	ATTCCGGACC	60%
Alpha 27	CGTATCGCCT	60%
Alpha 28	GCATACGGCA	60%
Alpha 29	ACTGCACGTC	60%
Alpha 30	TCGGATCCAG	60%
Alpha 31	CTAGCACGCT	60%
Alpha 32	ACTGACCGCT	60%
Alpha 33	CGTACGGCTG	70%
Alpha 34	CACGTGACGT	60%
Alpha 35	CATGGCTCAG	60%
Alpha 36	CCTAGACCGA	60%

Each primer was randomly combined with another in a RAPD-PCR reaction (see Materials & Methods) to screen for primer pairs that target many loci throughout the genome. The expectation was to find a primer pair that amplified > 50 bands in one reaction. Figure 9 is an example of a primer screen gel. Not all primer combinations succeeded in amplifying any PCR products. The primer combination Alpha 31/32 was chosen as the primary set for conducting RAPD-PCR on the lung biopsies.

3.7 Scanning Microdissected Archival Lung Biopsies for Genetic Alterations Using RAPD-PCR

In order to identify genetic alterations such as chromosome amplifications, translocations or deletions associated with the development of lung cancer, it is important to screen a patient set that is complete with both histologically normal tissue and progressive stages leading up to tumour development.

Using RAPD-PCR, 172 microdissected samples were analyzed. The samples were comprised of 41 different patient sets (see Appendix for patient details). Each patient set analyzed consisted of at least one histologically normal tissue (or blood equivalent) and one carcinoma in-situ (CIS). These two grades were the focus of the analysis. In many cases, there were multiple sites dissected from each patient corresponding to different sources of CIS or tumour. The resolved PCR products on polyacrylamide gels were analyzed in search of PCR products that did not exist throughout a patient set. Interesting alterations were identified and the gels were analyzed to determine if the same change occurred throughout other patient

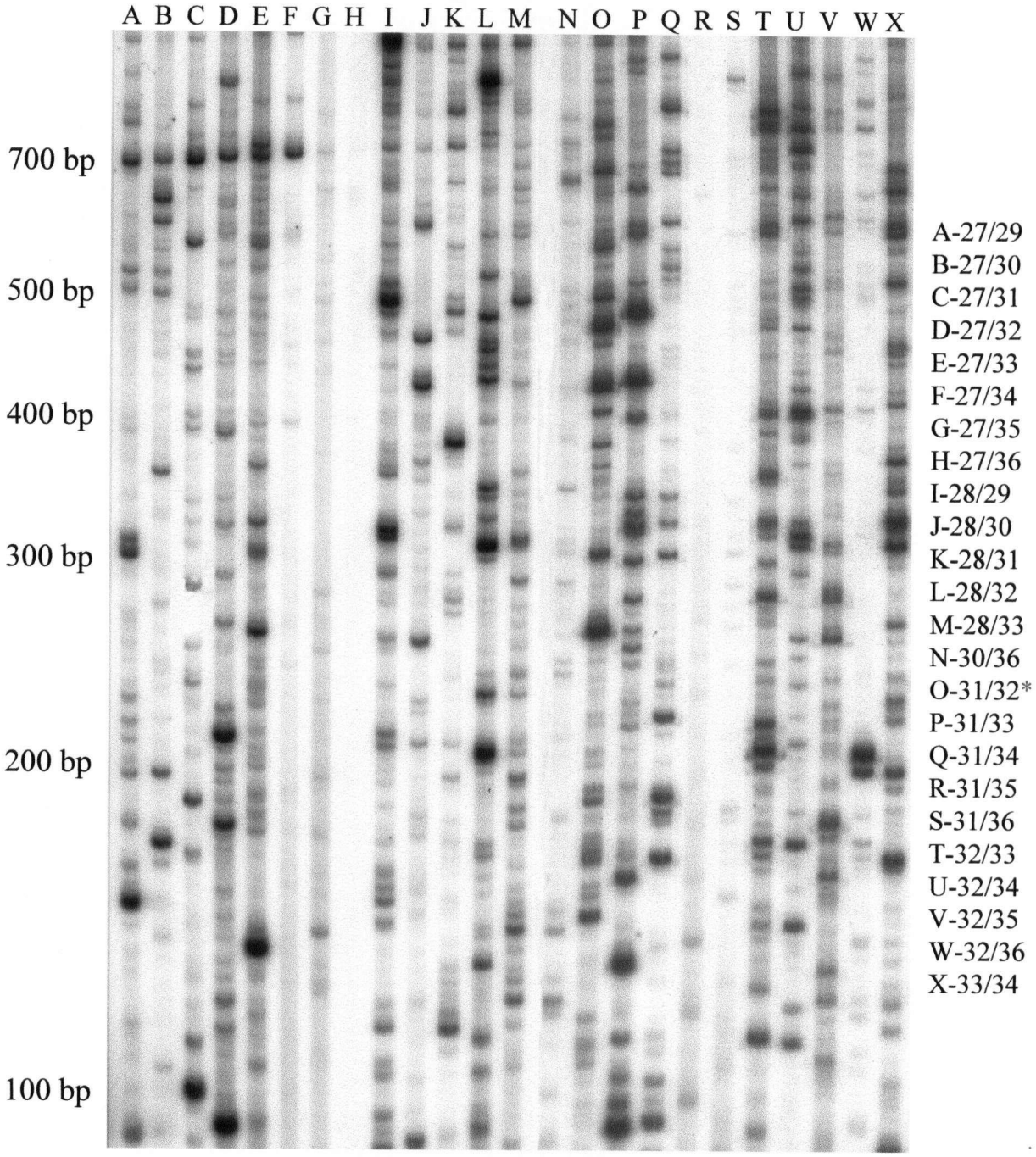


Figure 9: Screening Primer Pairs for Increased Scanning Density. Thirty-six randomly designed 10 bp primers were tested in random pairs to assess the number of SMAL-PCR fragments amplified. Pair 31/32 was chosen for the study because of the increased number of products.

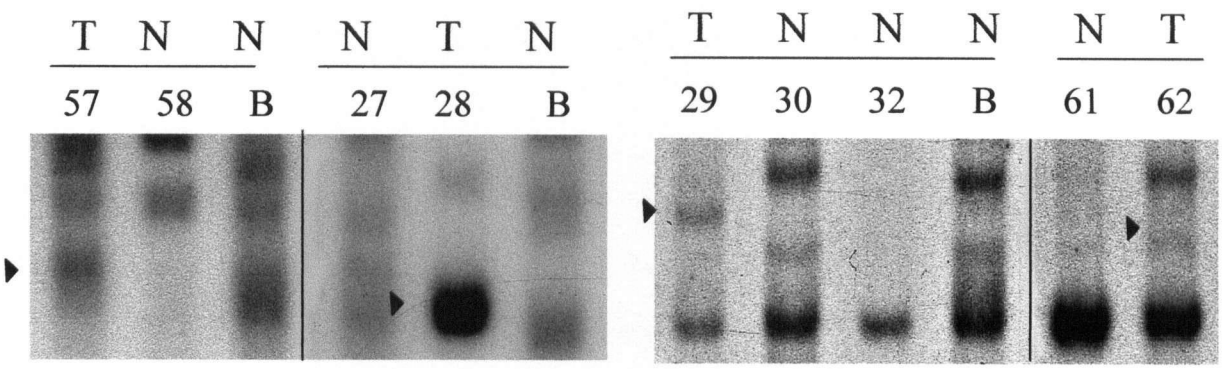
sets. As expected, more alterations were seen in the latter stages of progression. Fifteen different recurring band types (alterations) were identified from the entire patient pool. Table 3 categorizes all fifteen alterations according to size of alteration, number of recurrences in different patients, loss (PCR product present in the normal tissue only) or gain (PCR product present in the progressed stage), broad or tight band, faint or strong signal, singlet or doublet and chromosome localization. The number in parentheses denotes the number of patients exhibiting the polymorphism. Figure 10 shows four recurring alterations that mapped to distinct chromosomal regions. These alterations were identified on different gels therefore the band types do not appear perfectly matched across a gel. Alteration A is a 550bp gain in CIS that was seen in two of sixteen patients exhibiting the polymorphism. Alteration D is a 150bp gain in CIS that was seen in two of eighteen patients exhibiting the polymorphism. Alteration N is a 330bp gain in CIS that was seen in two of fifteen patients exhibiting the polymorphism. Alteration B/M is a 380bp gain in CIS that was seen in six of seventeen patients exhibiting the polymorphism. This alteration was also seen in biopsies from two different sites derived within one patient. Alteration B/M was assigned both letter designations because the alterations were initially identified as two separate band types that turned out to be the same alteration after further review.

3.8. Cloning of Genetic Alterations Potentially Associated with Lung Cancer Progression

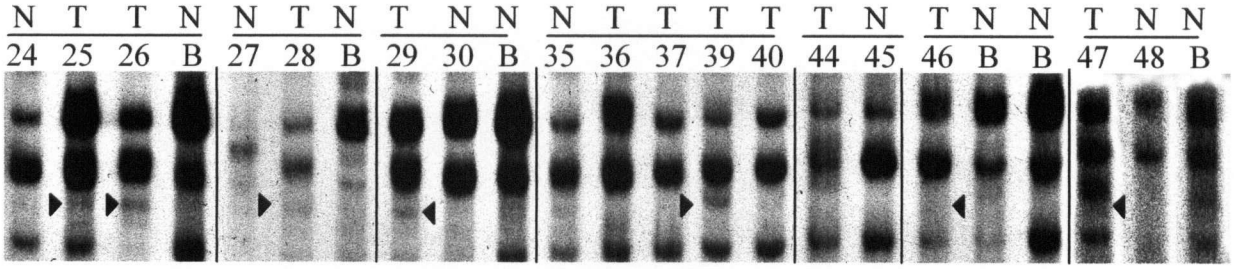
Cloning enables the identification of PCR products by the insertion of foreign DNA, by means of a vector, into bacterial cells allowing increased production of the desired DNA sequence.

Name	Size (kb)	# of Recurrences	Loss/Gain	Broad/Tight	Faint/Strong	Doublet/Singlet	Chromosome
A1	550	2(16)	loss	tight	faint	doublet	5q33.1
A2	550		loss	tight	faint	doublet	5q33.1
B1	380	4(17)	gain	broad	faint	singlet	8q24.3
B2	380		gain	broad	faint	singlet	8q24.3
B3	380		gain	broad	faint	singlet	8q24.3
B4	380		gain	broad	faint	singlet	8q24.3
C1	340	4	gain	tight	strong	singlet	Reamp Failed
C2	340		gain	tight	strong	singlet	Reamp Failed
C3	340		gain	tight	faint	singlet	Reamp Failed
C4	340		gain	tight	faint	singlet	Reamp Failed
D1	150	2(18)	gain	broad	strong	singlet	1q23.3
D2	150		gain	broad	strong	singlet	1q23.3
E1	140	2	gain	broad	faint	singlet	No Match
E2	140		gain	broad	faint	singlet	No Match
F1	260	3	gain	tight	strong	singlet	No Match
F2	260		gain	tight	strong	singlet	No Match
F3	260		gain	tight	strong	singlet	No Match
G1	240	3	gain	tight	faint	singlet	Ligation Failed
G2	240		gain	tight	strong	singlet	Ligation Failed
G3	240		gain	tight	strong	singlet	Ligation Failed
H1	220		loss	tight	strong	singlet	Reamp Failed
H2	220		loss	tight	strong	singlet	Reamp Failed
H3	220	3	loss	tight	faint	singlet	Reamp Failed
I1	650	2	loss	tight	strong	singlet	No Match
I2	650		loss	tight	strong	singlet	No Match
J1	340	2	loss	tight	strong	singlet	No Match
J2	340		loss	tight	faint	singlet	No Match
K1	410	3	gain	tight	faint	doublet	No Match
K2	410		gain	tight	faint	doublet	No Match
K3	410		gain	tight	faint	doublet	No Match
L1	550	2	loss	tight	faint	singlet	Ligation Failed
L2	550		loss	tight	faint	singlet	Ligation Failed
M1	380	3(17)	gain	tight	faint	singlet	8q24.3
M2	380		gain	tight	faint	singlet	8q24.3
M3	380		gain	tight	strong	singlet	8q24.3
N1	330	2(15)	gain	tight	strong	singlet	7q36.2
N2	330		gain	tight	strong	singlet	7q36.2
P1	130	2	loss	broad	strong	singlet	Repeat Element
P2	130		loss	broad	strong	singlet	Repeat Element

Table 3: Fifteen Different Band Type Alterations. All alterations were cloned; four alterations (A, B/M, D, N) mapped to distinct chromosomal regions, 5q33.1, 8q24.3, 1q23.3, 7q36.2, respectively. Number in parentheses indicates number of cases exhibiting polymorphism.



Alteration D (150 bp gain) **Alteration N** (330 bp gain)



Alterations B/M (380 bp gain)

Alteration A
(550 bp gain)

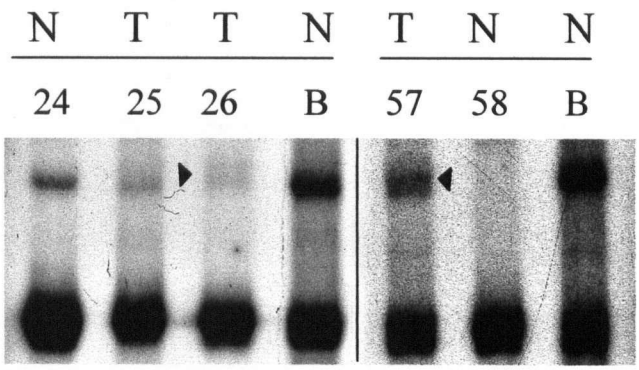


Figure 10: Recurring Alterations in Multiple Patients. Fifteen different band types were cloned; four band types mapped to four distinct chromosomal regions. Alteration A occurred in 2/16 patients, alteration D occurred in 2/18 patients, alteration N occurred in 2/15 patients and alteration B/M occurred in 7/17 patients. Numbers represent samples, B represents blood,

Fifteen band types were cloned as outlined in the Materials and Methods (See Table 3 for details). Alterations B/M are used in this cloning example to demonstrate the procedure. Figure 11A shows the reamplification of the original PCR products (B1, B2, B3, B4, M1, M2, M3) using modified SMAL-PCR primers with restriction sites within their sequence. The new PCR products contain BamH1 and PST1 sites. After cutting the PCR product with the enzymes, the product was precipitated and visualized on a polyacrylamide gel. Figure 11B shows the well-resolved PCR products on a polyacrylamide gel. Due to the increased resolution ability of the polyacrylamide, this gel shows many more PCR products that appeared as a single product on the agarose gel. The PCR product matching the appropriate size was excised, in this case, 380 bp. The product was inserted into a PSK^+ vector containing a β-galactosidase reporter gene and an ampicillin resistance gene. All colonies chosen for colony PCR were white because the inserted DNA disrupted the β-galactosidase gene. Figure 12A shows a colony PCR used to demonstrate that the colonies still contain the appropriate size insert. Colonies from plates containing the B and M clones were used in the reaction. In some cases, there is no product in the lane or the product size does not match the expected size. All colonies confirmed by colony PCR were tested by colony fingerprint (G-track sequencing) to determine if all contained the same sequence (See Appendix). Figure 12B shows a colony fingerprint of B and M clones. It is clear that all colonies contain the same sequence. One colony from each set of similar sequenced colonies was grown for plasmid purification (see Appendix). The purified plasmid was sequenced at Nucleic Acids Protein Service (NAPS) Unit. Figure 13 shows the sequence of the original B insert.

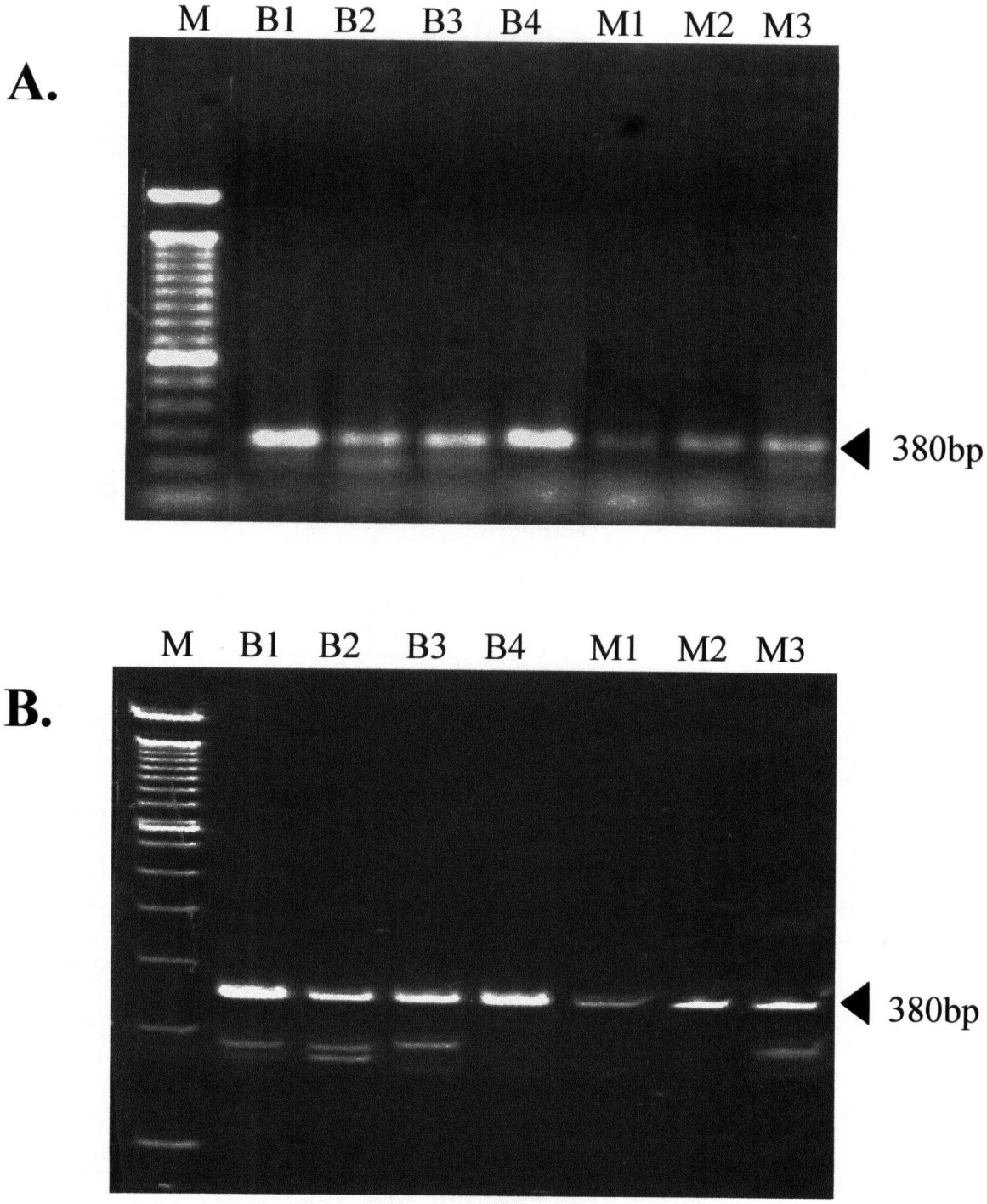


Figure 11: Reamplification and Purification in Cloning Procedure. A. Reamplification of eluted DNA representing alterations B and M. B. Acrylamide purification of cut inserts prior to ligation in vector. B1-4 represent four recurring alterations at 380 bp size; M1-3 represent three recurring alterations at 380 bp size.

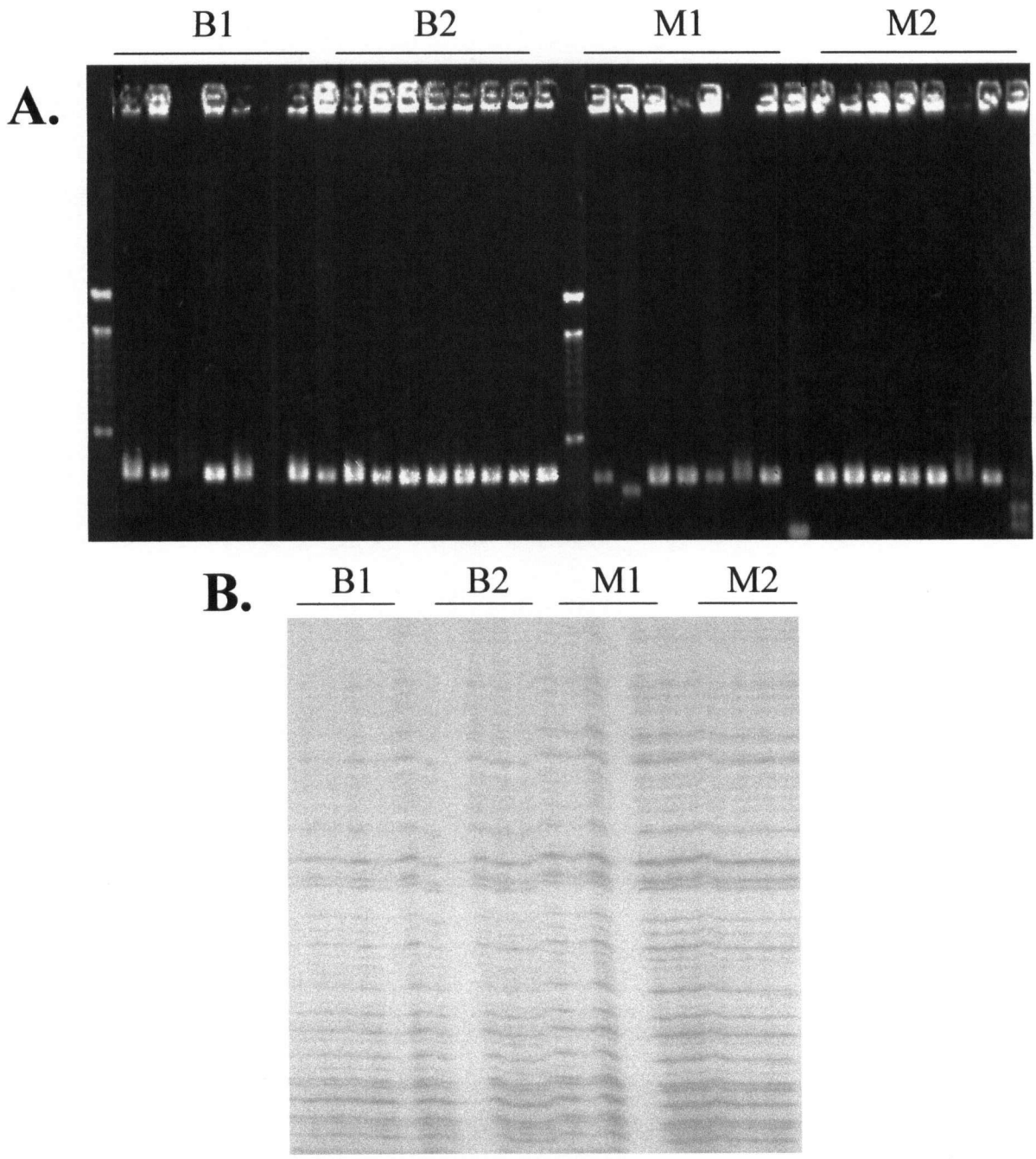


Figure 12: Colony PCR and Colony Fingerprint in Cloning Procedure. A. Colony PCR: Transformed colonies were verified to contain specific size insert. B. Colony Fingerprint: Transformed colonies containing same size inserts were verified to contain the same sequence by G-track sequencing. B1-2 and M1-2 represent transformed colonies with DNA insert B and M, respectively.

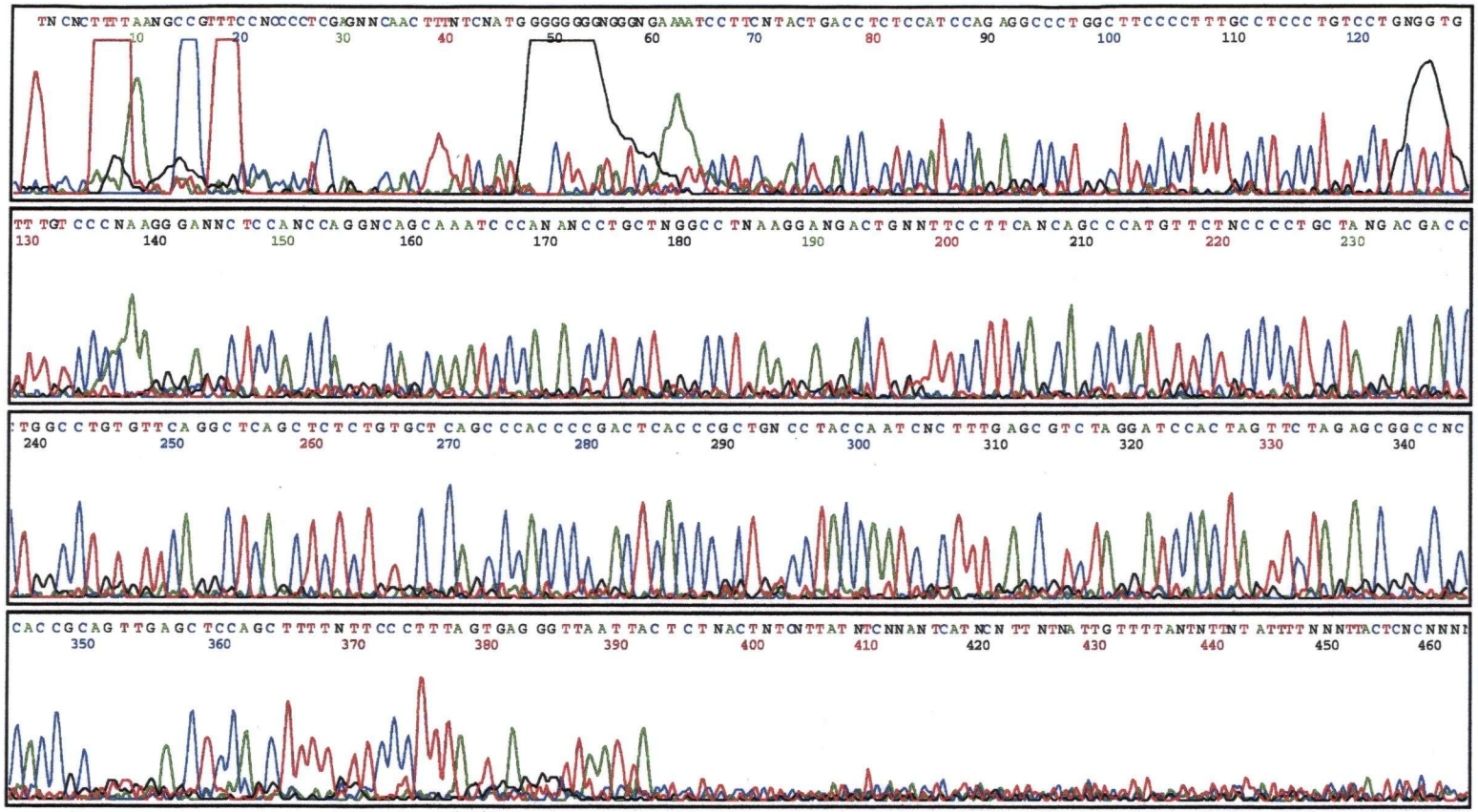


Figure 13: DNA Sequence of Clone B/M. Sequences for all clones determined by Nucleic Acids Protein Service (NAPS) Unit at the University of British Columbia. This sequence example is of clone B/M.

colors represent letters

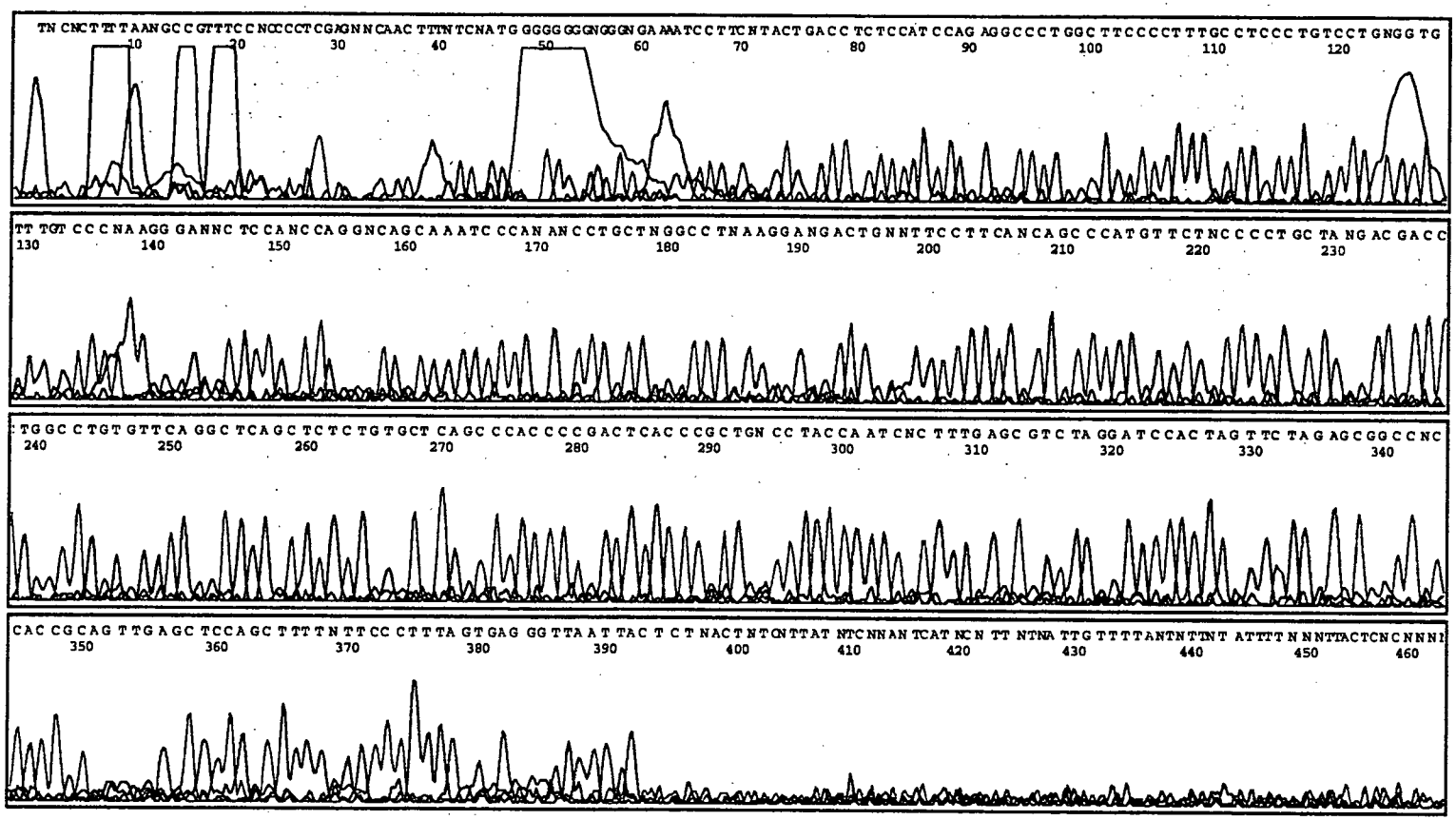


Figure 13: DNA Sequence of Clone B/M. Sequences for all clones determined by Nucleic Acids Protein Service (NAPS) Unit at the University of British Columbia. This sequence example is of clone B/M.

3.9. Localizing Sequenced Products to Chromosomal Locations

There are many chromosomal regions throughout the human genome proposed to be involved in lung cancer development. Many of these regions were identified by gross mapping using FISH or radiation hybrids (Yasuda, 1996).

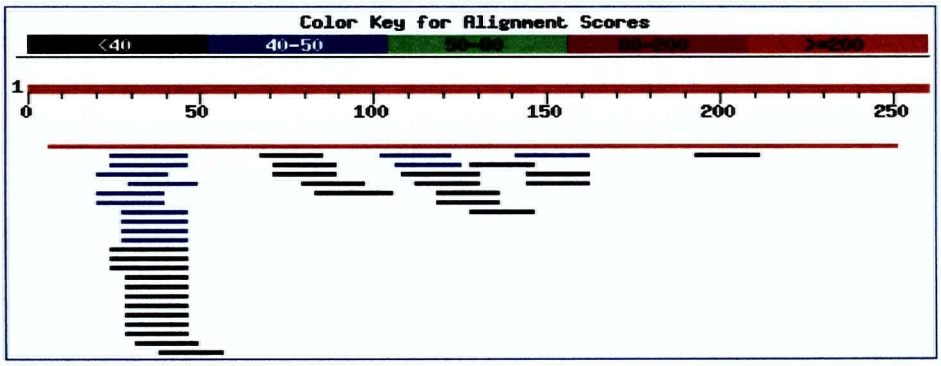
All cloned sequences (Table 3) were matched against the high throughput genome sequences (htgs) available through NCBI. Figure 14A shows alteration B matching to RP11-637F16 whereas figure 14B shows alteration N matching to RP11-26I06. Both alterations match their sequence with the bacterial artificial chromosomes (BACs) at high similarity. The same sequences were matched against the Human Genome Project Program at the University of Southern California (http://genome.ucsc.edu). Alteration B/M matches 8q24.3. Alterations N, D, and A match 7q36.2, 1q22.3, and 5q33.1, respectively.

3.10 Identification of LOH Markers Specific to Cloned Regions

There are many microsatellite markers scattered throughout the human genome. However, many of these di- and tri-nucleotide repeats are not polymorphic therefore their use is not informative in a LOH study.

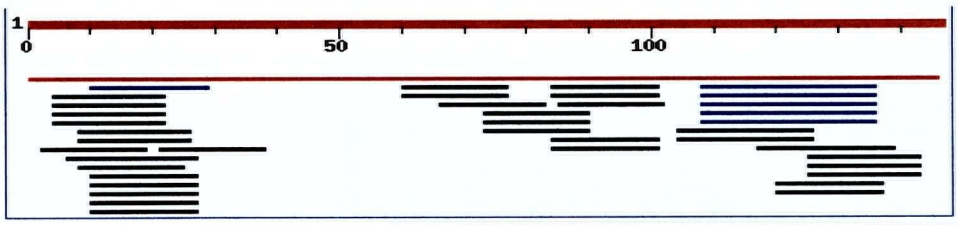
Using the human genome project at the University of Southern California (http://genome.ucsc.edu) potential LOH markers were chosen specific to all four sequenced regions. Figure 15A shows the sequenced region of 7q36.2 matched against

A.



Sequences producing significant alignments:		Score (bits)	E Value
gi|16506447|gb|AC021636.6|	Homo sapiens chromosome 8 clone ...	458	e-126
gi|9256419|gb|AC022418.4|AC022418	Homo sapiens chromosome 5...	46	0.019
gi|9256328|gb|AC012631.4|AC012631	Homo sapiens chromosome 5...	46	0.019
gi|6041753|gb|AC011805.1|AC011805	Homo sapiens chromosome 2...	44	0.074

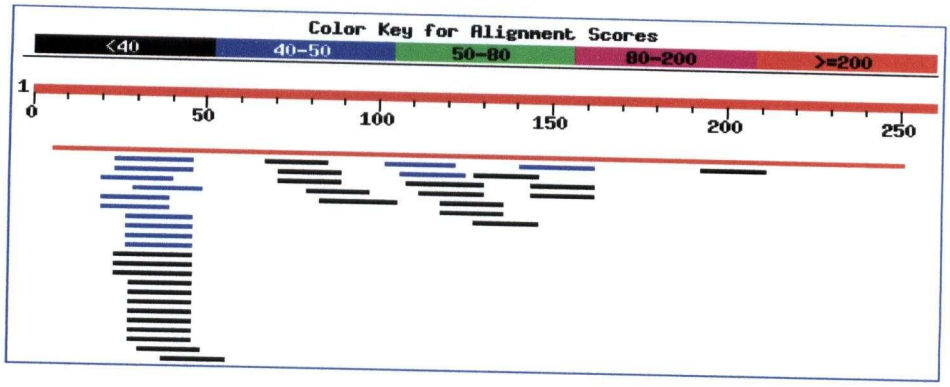
B.



Sequences producing significant alignments:		Score (bits)	E Value
gi|7770445|gb|AC016222.4|AC016222	Homo sapiens clone RP11-2...	210	3e-52
gi|15383794|gb|AC093505.1|	Homo sapiens chromosome 5 clone ...	42	0.15
gi|15208569|gb|AC093311.1|	Homo sapiens chromosome 5 clone ...	42	0.15
gi|15193401|gb|AC093267.1|	Homo sapiens chromosome 5 clone ...	42	0.15

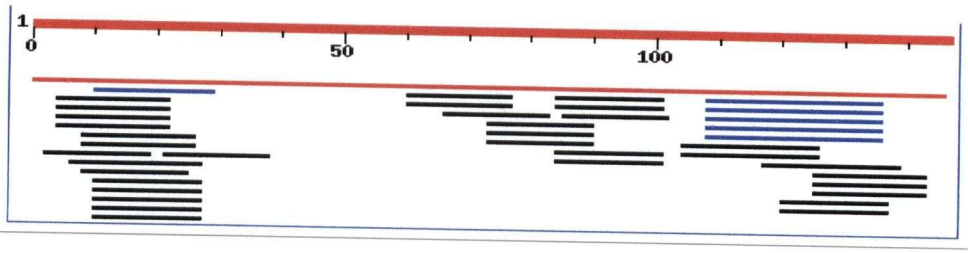
Figure 14. Localization of Sequenced Product to Chromosomes. Sequenced products were matched against the high throughput genome sequences database at NCBI. Matches were scored based on degree of similarity. A) Alteration B/M highly matched (redline) to a chromosome 8 clone RP11-637F16. B) Alteration N highly matched (red line) clone RP11-26I06. Both clones are sequenced BACs that were mapped using the Fingerprinting Contigs (FPC) program at Washington University.

A.



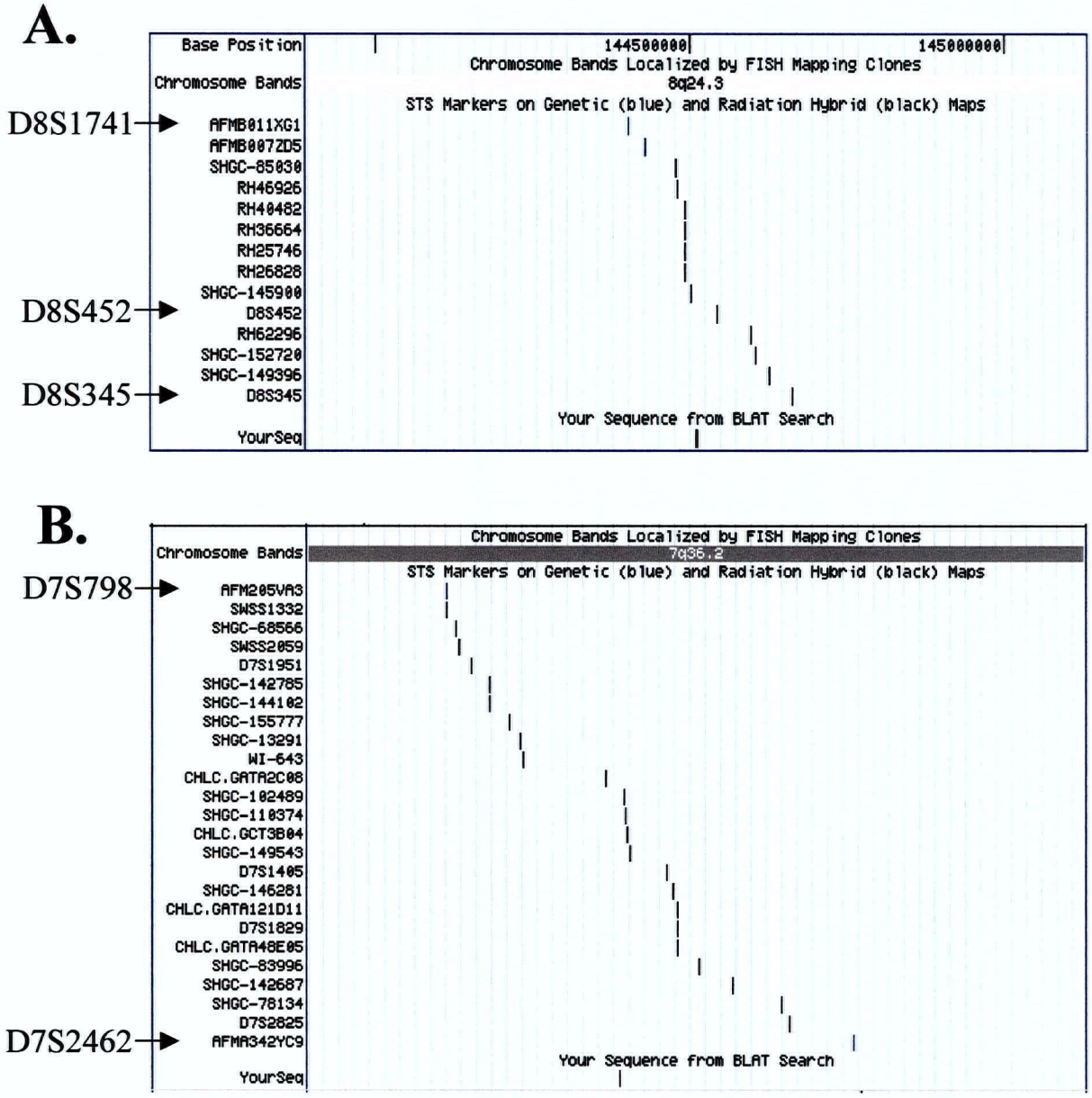
Sequences producing significant alignments:		Score (bits)	E Value
gi|16506447|gb|AC021636.6|	Homo sapiens chromosome 8 clone ...	458	e-126
gi|9256419|gb|AC022418.4|AC022418	Homo sapiens chromosome 5...	46	0.019
gi|9256328|gb|AC012631.4|AC012631	Homo sapiens chromosome 5...	46	0.019
gi|6041753|gb|AC011805.1|AC011805	Homo sapiens chromosome 2...	44	0.074

B.



Sequences producing significant alignments:		Score (bits)	E Value
gi|7770445|gb|AC016222.4|AC016222	Homo sapiens clone RP11-2...	210	3e-52
gi|15383794|gb|AC093505.1|	Homo sapiens chromosome 5 clone ...	42	0.15
gi|15208569|gb|AC093311.1|	Homo sapiens chromosome 5 clone ...	42	0.15
gi|15193401|gb|AC093267.1|	Homo sapiens chromosome 5 clone ...	42	0.15

Figure 14. Localization of Sequenced Product to Chromosomes. Sequenced products were matched against the high throughput genome sequences database at NCBI. Matches were scored based on degree of similarity. A) Alteration B/M highly matched (redline) to a chromosome 8 clone RP11-637F16. B) Alteration N highly matched (red line) clone RP11-26I06. Both clones are sequenced BACs that were mapped using the Fingerprinting Contigs (FPC) program at Washington University.



UCSC Genome Browser (http://genome.ucsc.edu)

Figure 15. Identification of LOH markers at chromsomal regions 8q24.3 and 7q36.2. Polymorphic microsatellite markers were picked from the Human Genome Project at UCSC. that mapped very close to the cloned sequenced regions.
A. Region 8q24.3 with three markers: D8S1741, D8S452 and D8S345. B. Region 7q36.2 with two markers: D7S798 and D7S2462. Vertical lines represent relative positions of sequence and markers.

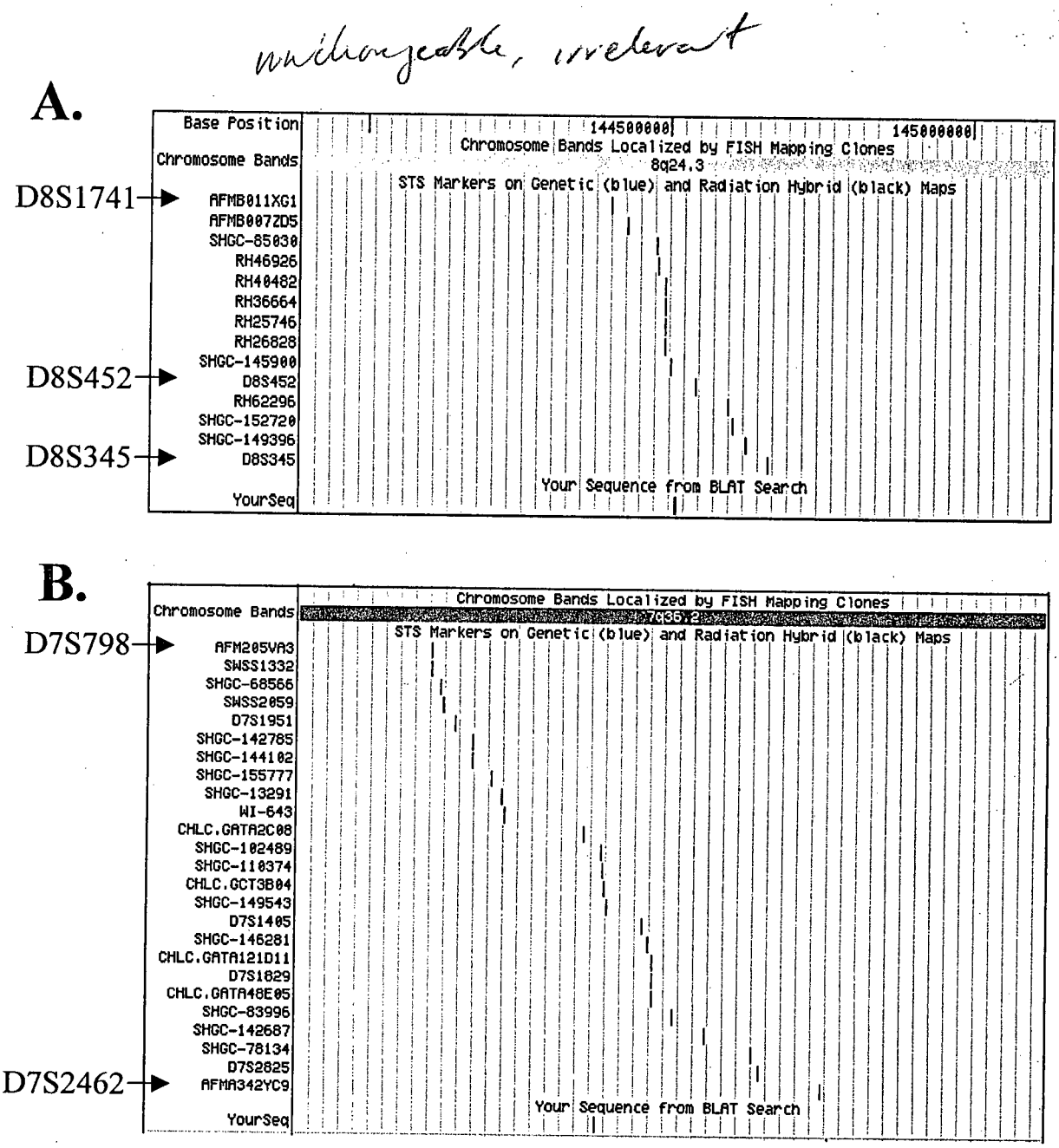


Figure 15. Identification of LOH markers at chromsomal regions 8q24.3 and 7q36.2. Polymorphic microsatellite markers were picked from the Human Genome Project at UCSC. that mapped very close to the cloned sequenced regions.
A. Region 8q24.3 with three markers: D8S1741, D8S452 and D8S345. B. Region 7q36.2 with two markers: D7S798 and D7S2462. Vertical lines represent relative positions of sequence and markers.

the UCSC data with the sequence and surrounding LOH markers. Figure 15B shows the sequenced region matching with 8q24.3 surrounded by LOH markers. Minimum product size and heterozygosity information for each marker was obtained at (www.gdb.org). Table 4 lists all microsatellite markers ordered from Alpha DNA Inc. and temperature tested. The table categorizes the markers according to chromosome regions targeted, size of product, optimal temperature, and level of heterozygosity within the human population.

Table 4: Microsatellite Markers Chosen for 8q24.3, 7q36.2, 5q33.1 and 1q23.3

Primer	Chromosome	Product Size(bp)	Annealing Temp. (°C)	Heterozygosity
D1S2768F	1q23.3	198	Only 50	61%
D1S2768R	1q23.3	198	Only 50	61%
D1S2844F	1q23.3	177	60>55>50	82%
D1S2844R	1q23.3	177	60>55>50	82%
D9S1864F	1q23.3	270	Only 50	69%
D9S1864R	1q23.3	270	Only 50	69%
D5S673F	5q33.1	260-288	60>55>50	82%
D5S673R	5q33.1	260-288	60>55>50	82%
D7S798F	7q36.2	200-218	60>55>50	79%
D7S798R	7q36.2	200-218	60>55>50	79%
D7S2462F	7q36.2	127	60>55>50	85%
D7S2462R	7q36.2	127	60>55>50	85%
D7S483F	7q36.2	166-188	60>55>50	83%
D7S483R	7q36.2	166-188	60>55>50	83%
D8S1727F	8q24.3	257	50>>55	76%
D8S1727R	8q24.3	257	50>>55	76%
D8S1741F	8q24.3	298	60>55>50	60%
D8S1741R	8q24.3	298	60>55>50	60%
D8S1743F	8q24.3	171	Failed	82%
D8S1743R	8q24.3	171	Failed	82%
D8S345F	8q24.3	291-296	50>>55	50%
D8S345R	8q24.3	291-296	50>>55	50%
D8S452F	8q24.3	218	60>55>50	50%

D8S452R	8q24.3	218	60>55>50	50%
D8S508F	8q24.3	230	60>55>50	55%
D8S508R	8q24.3	230	60>55>50	55%
D8S534F	8q24.3	176-210	55>50>60	80%
D8S534R	8q24.3	176-210	55>50>60	80%

LOH marker positions were verified by matching the sequence using NCBI and mapping the sequences using the Fingerprinting Contigs (FPC) Program at Washington University. Figure 16 shows two contigs of overlapping BACs with the sequence at 7q36.2 and 8q24.3 identified along with verified markers. Confirmation of precise marker positions was necessary using BAC contigs because marker positions using UCSC data are based on linkage studies.

3.11. Loss of Heterozygosity Analysis

Allelic imbalance is a characteristic method in assessing chromosomal instability on a locus-specific basis. Many studies have identified regions of chromosomal instability in lung cancer using microsatellite markers distributed throughout the human genome.

Due to DNA quantity constraints of microdissected lung samples, only two chromosomal regions, 8q24.3 and 7q36.2 were subjected to LOH analysis. Six patients were screened for allelic imbalance at 8q24.3 using D8S452, a microsatellite marker. Allelic imbalance is assessed as a difference in the PCR signal ratio of the two paternal and maternal alleles within the normal compared to the tumour DNA. Two patients showed allelic imbalance (AI), three patients were retention (R) and one patient was not informative (NI) (Figure 17A).

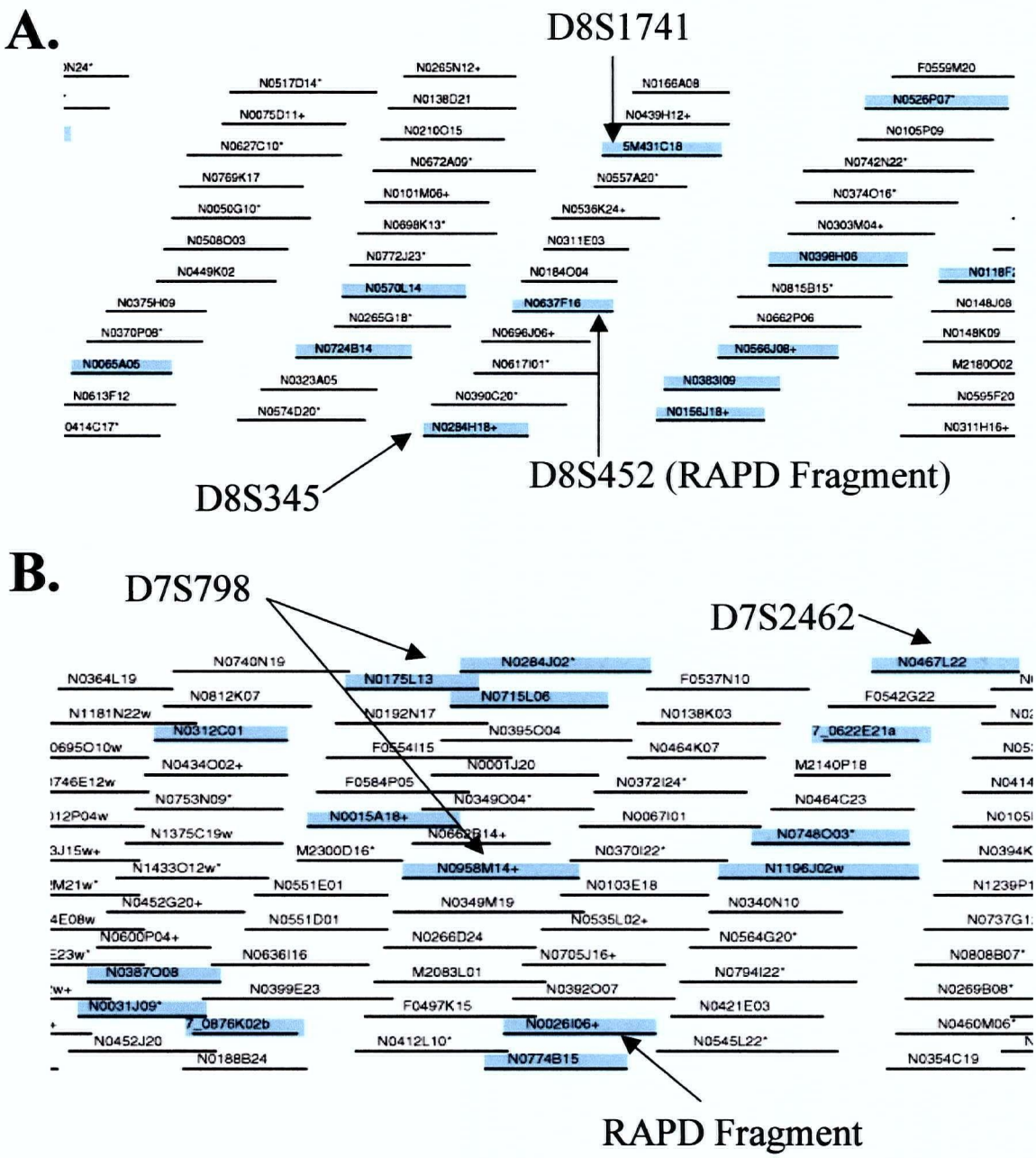


Figure 16. Contigs of Overlapping BACs. Microsatellite markers were mapped to overlapping BACs for both A) 8q24.3 and B) 7q36.2 using the Fingerprinting Contigs (FPC) Program at Washington University. Cloned region (RAPD fragment) was also localized to a BAC.

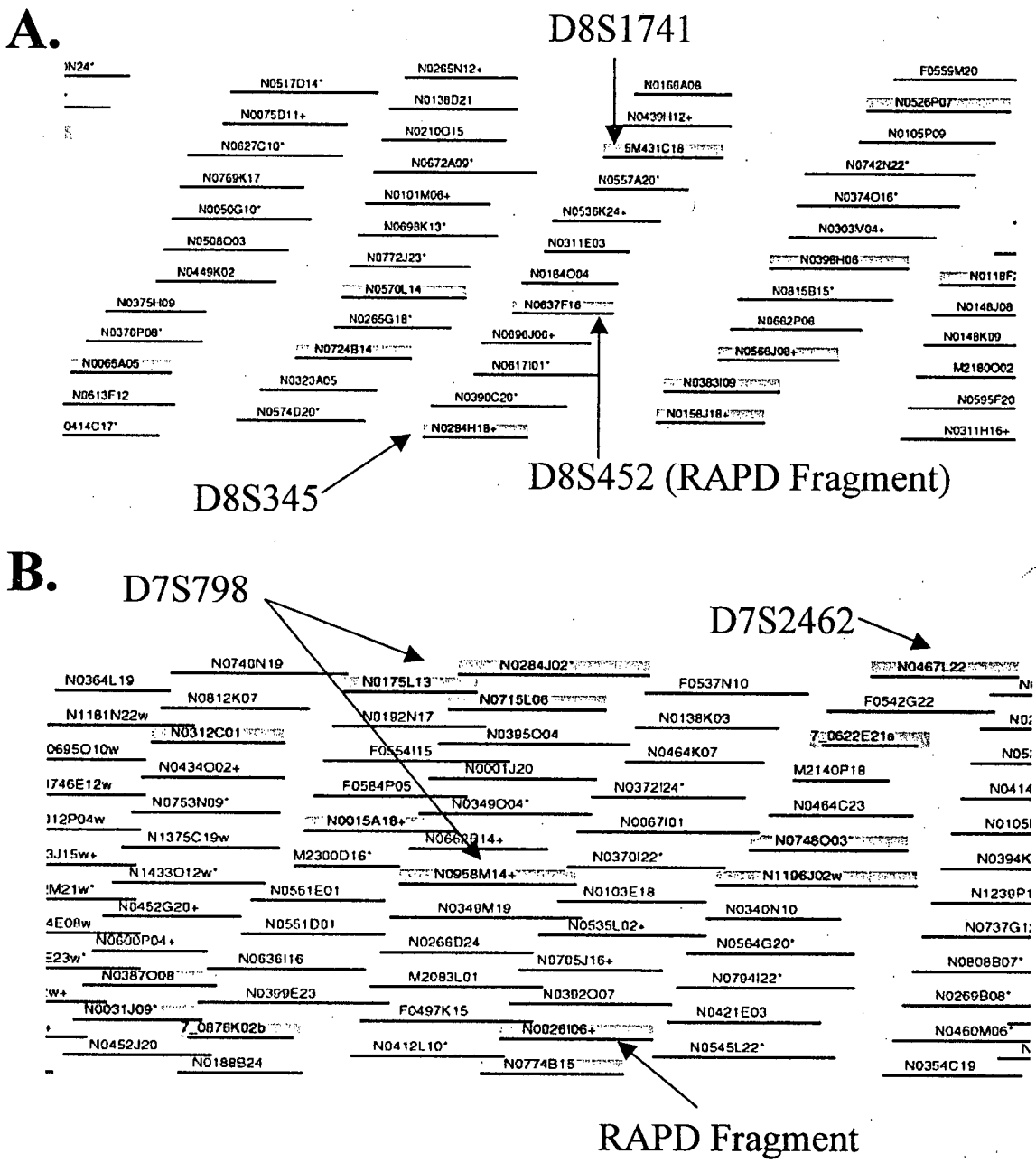


Figure 16. Contigs of Overlapping BACs. Microsatellite markers were mapped to overlapping BACs for both A) 8q24.3 and B) 7q36.2 using the Fingerprinting Contigs (FPC) Program at Washington University. Cloned region (RAPD fragment) was also localized to a BAC.

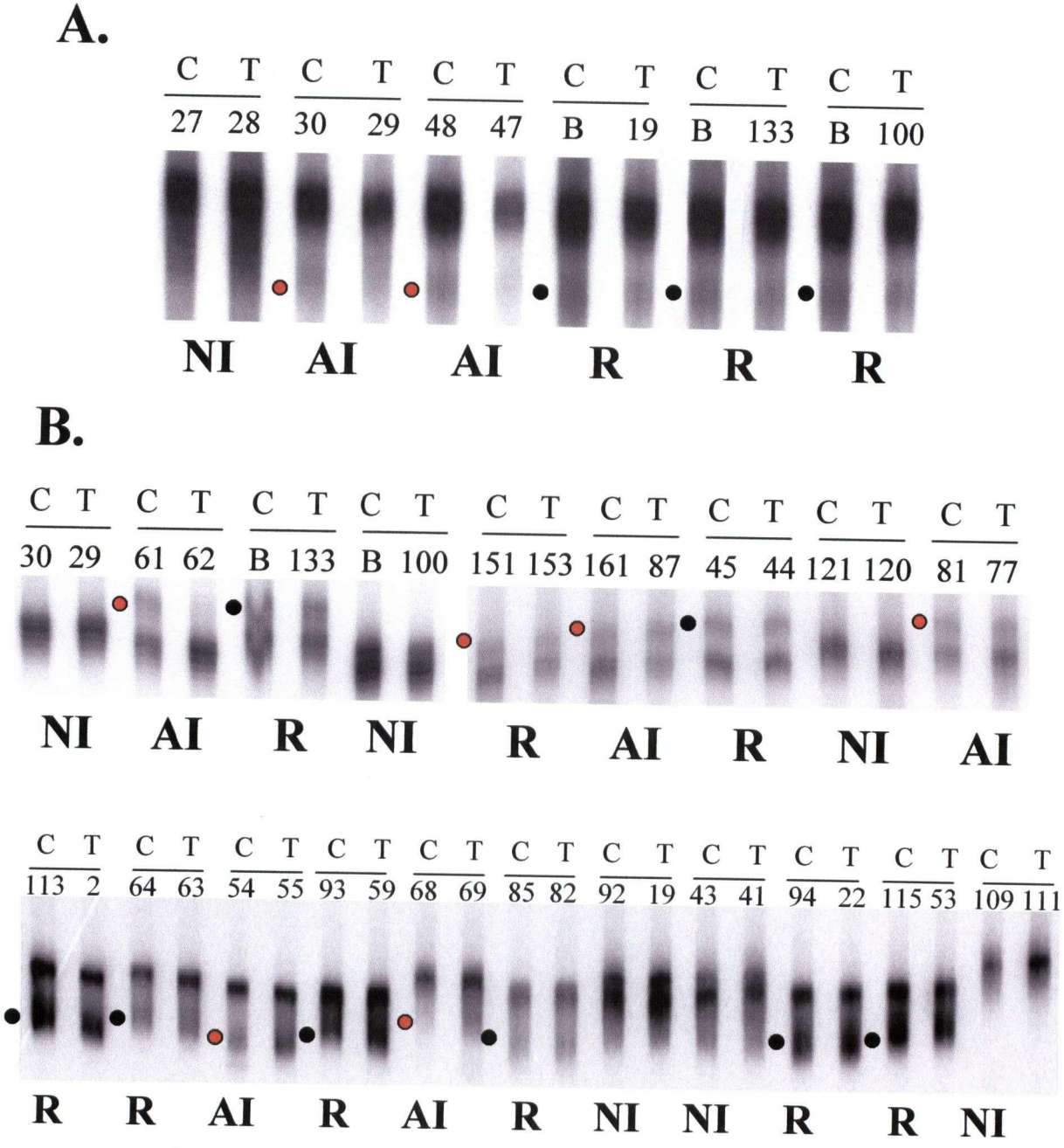


Figure 17: Loss of Heterozygosity (LOH) Analysis. A. Six patients were screened for allelic imbalance at 8q24.3 using D8S452. Two patients showed AI, three patients exhibited R and one patient was NI. B. Twenty NSCLC tumor/normal pairs were screened for allelic imbalance at 7q36.2 using D7S798. Five patients showed allelic imbalance (AI), nine patients exhibited retention(R) and six patients were non-informative(NI).

○ Allelic Imbalance ● Retention

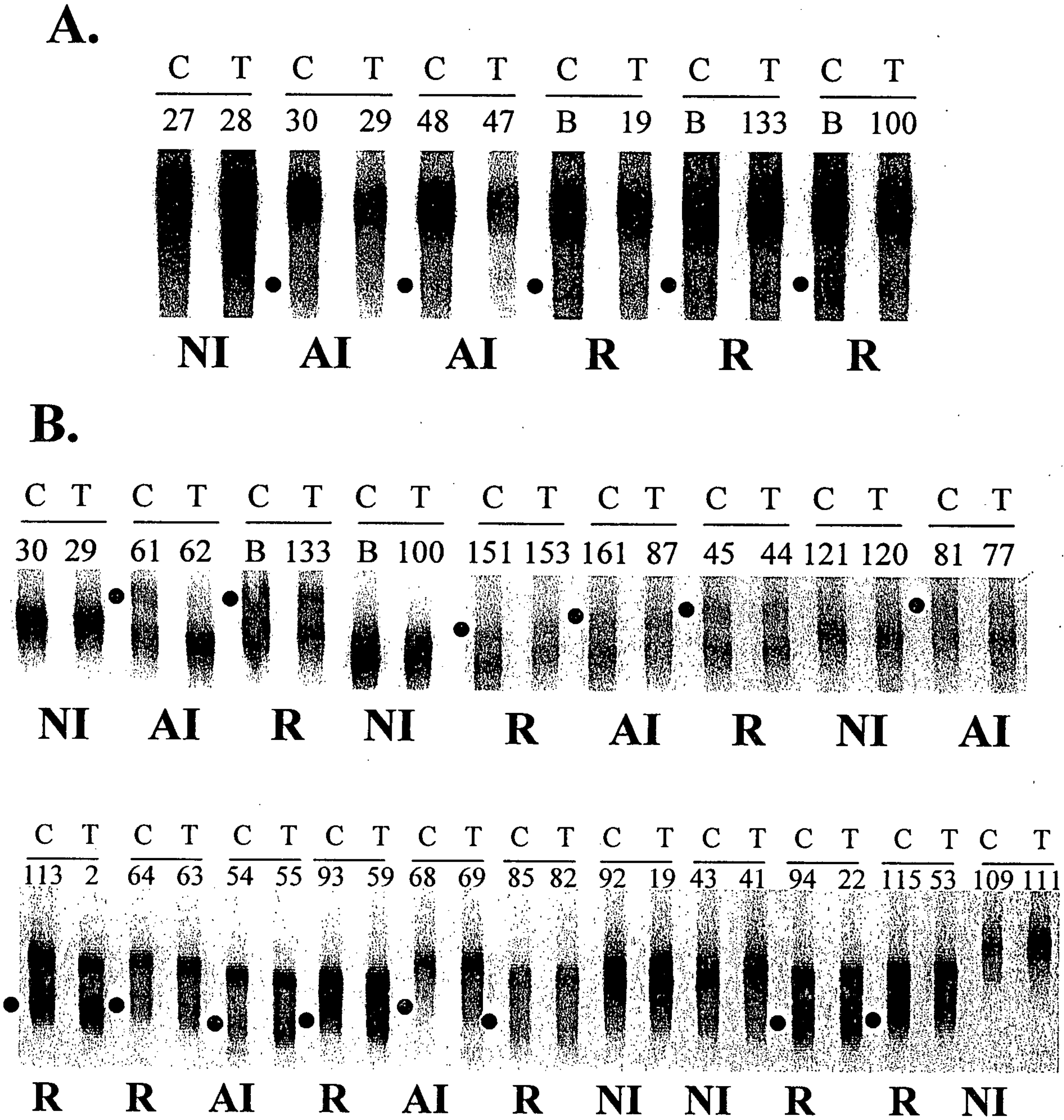


Figure 17: Loss of Heterozygosity (LOH) Analysis. A. Six patients were screened for allelic imbalance at 8q24.3 using D8S452. Two patients showed AI, three patients exhibited R and one patient was NI. B. Twenty NSCLC tumor/normal pairs were screened for allelic imbalance at 7q36.2 using D7S798. Five patients showed allelic imbalance (AI), nine patients exhibited retention(R) and six patients were non-informative(NI).

● Allelic Imbalance ● Retention

representated by letters

Twenty patients were screened for allelic imbalance at 7q36.2 using D7S798. Five patients showed AI, nine patients showed R and six patients were NI at the locus (Figure 17B).

To confirm the visual assessment of allelic imbalance, PCR signals were quantitated using a phosphorscreen and ImageQuant 5.0 software (Table 5). Quantitated signal was interpreted by calculating the ratio of lower allele signal to upper allele signal. The ratio was calculated and compared for both C and T. Ratio differences equal to or greater than 50% were considered as significant.

3.12 Previously Discovered Genes at 8q24.3 and 7q36.2

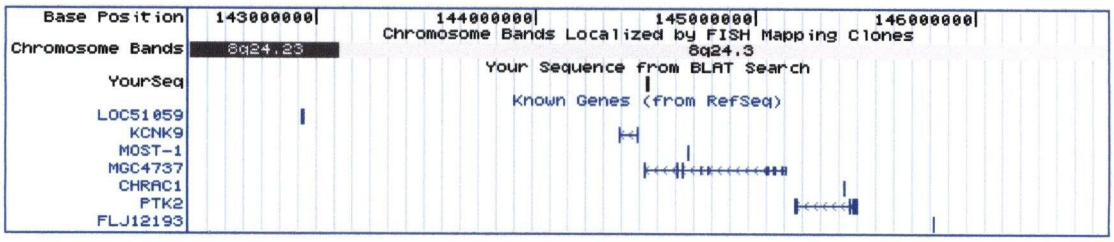
Many oncogenes such as ras and c-myc and tumour suppressor genes such as p53 and p16 have been implicated as potential markers of NSCLC; however, no one gene has been able to act as a great predictor of NSCLC. Novel genes need to be discovered as early indicators of NSCLC progression to assist in NSCLC disease management.

Using the UCSC database, known genes were identified within 2 Mb of each cloned region. Figure 18A shows seven genes in the 8q24.3 region including LOC51059, KCNK9, MOST-1, MGC4737, CHRAC1, PTK2 (FAK) and FLJ12193. Figure 18B shows four genes in the 7q36.2 region including FLJ21634, MLL3, XRCC2 and ARP3BETA. Unfortunately, most of these genes are not characterized.

Name	Signal Intensity	Ratio of Lower Allele/Upper Allele	LOH Designation
61L	1141	1.32	Allelic Imbalance
61U	866		
62L	1873	3.04	
62U	616		
BL	1487	1.40	Retention
BU	1064		
133L	1483	1.38	
133U	1073		
151L	251974	1.99	Retention
151U	126660		
153L	143799	1.48	
153U	97394		
161L	274999	2.07	Allelic Imbalance
161U	132823		
87L	135375	1.04	
87U	130032		
45L	258388	1.81	Retention
45U	143125		
44L	225255	1.84	
44U	122355		
81L	154866	1.31	Allelic Imbalance
81U	117912		
77L	246970	2.00	
77U	123510		
113L	271113	0.96	Retention
113U	283364		
2L	199117	0.89	
2U	222937		
64L	91254	0.55	Retention
64U	166443		
63L	106656	0.54	
63U	196431		
54L	88629	0.45	Allelic Imbalance
54U	195988		
55L	171238	0.74	
55U	229867		
93L	201621	0.66	Retention
93U	304562		
59L	255519	0.82	
59U	313312		
68L	42451	0.22	Allellic Imbalance
68U	194474		
69L	97030	0.38	
69U	254075		
85L	82725	0.45	Retention
85U	182550		
82L	102110	0.53	
82U	192157		
94L	268687	0.84	Retention
94U	320623		
22L	258564	0.87	
22U	297515		
115L	318397	0.87	Retention
115U	366043		
53L	258142	0.88	
53U	293722		

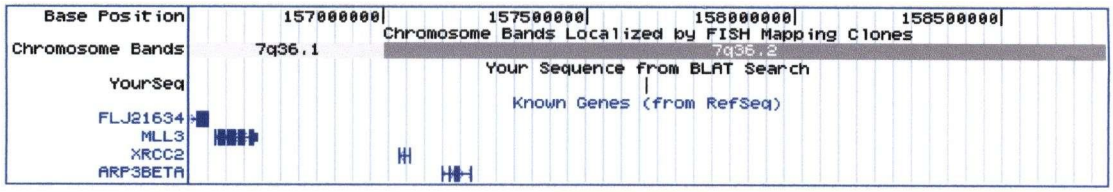
Table 5: Analysis of 7q36.2 LOH Results Using ImageQuant 5.0 Quantitation Software. Band signal intensities were determined and ratio of lower allele to upper allele were calculated. Ratios between tumor/normal within each patient were compared to designate either allelic imbalance or retention. A minimum of 50% change was required to designate allelic imbalance.

A.



LOC51059=hypothetical protein
KCNK9 = potassium channel
MOST-1= unknown protein
MGC4737 = hypothetical protein
CHRAC1= chromatin accessibility gene
PTK2 = protein tyrosine kinase 2 (FAK) gene
FLJ12193 = hypothetical protein

B.



FLJ21634= Unknown
MLL3=Myeloid/Lymphoid Leukemia
XRCC2 = X-Ray Damage Repair Gene
ARP3BETA=Actin Related Gene

Figure 18: Known Genes at 8q24.3 and 7q36.2. A) Seven genes are known within 2Mb radius of sequenced product. Some potential oncogenes/TSGs include CHRAC1 and PTK2 (FAK). B) Four known genes localize to within 2Mb radius of sequenced product. Potential oncogenes/TSGs are MLL3, XRCC2 and ARP3BETA. Many of these genes are not characterized and expression data is not available.

Chapter 4: Discussion

4.1 Summary of Results

The focus of this thesis was the identification of novel genetic alterations in microdissected NSCLC tumours and their precursors on a genome-wide scale. To discover grade specific alterations, it was necessary to isolate specific pure cell populations from surrounding cells using manual microdissection. Due to the small nature of lung biopsies, the quantity of DNA extracted was minimal as the biopsies contained few numbers of cells. As most other genome-wide analyses such as LOH require larger amounts of DNA, RAPD-PCR was used to perform high resolution scanning of the NSCLC sample set. Using one specific primer pair, fifteen different recurring alterations were identified among 41 patients. Four recurring alterations mapped to distinct chromosomal regions at 8q24.3, 7q36.2, 5q33.1 and 1q23.3. Due to the limited amount of microdissected patient DNA, verification of chromosomal instability was performed only at 8q24.3 and 7q36.2 using LOH analysis. A detailed physical map and a tiling set of human BAC clones were deduced for each region to facilitate future definition of minimal regions of alteration and gene identification.

4.2 Harvesting DNA from Multiple Grades of Tumourigenesis in Multiple Patients

Lung biopsies are composed of many different cell populations. To overcome the problem with tissue heterogeneity, 172 NSCLC biopsies composed of various stages including normal, hyperplasia, different degrees of dysplasia, CIS and invasive tumour were manually microdissected. Prior to the microdissection, Dr. Jean LeRiche classified sections according to progressive stages and identified cell populations representing

various grades to guarantee that the correct cells were extracted. Contamination by unwanted cells was limited to approximately 10% of the total pool of cells. This level of contamination was tolerable as only changes that were recurring were targeted for subsequent cloning, sequencing and chromosomal localization.

As very few cells were extracted, our concern was the DNA quantity. Total DNA yield from most grades was approximately 120 ng. As each RAPD reaction in this study only used 2ng of quantified DNA, most microdissected samples would support 60 separate analyses. Using RAPD-PCR, primer pair 31/32 generated about 50 fragments, targeting 50 separate genetic loci. Assuming that other appropriate primer pairs would target as many loci, the DNA material would support screening of approximately 3000 separate loci. Assuming that these loci are equally distributed throughout the human genome, the genome-wide analysis is sampling loci at less than 1 Mb intervals. Compared to LOH, that requires a minimum of 5ng per reaction to screen one locus, RAPD-PCR was an efficient method to use on this non-renewable resource.

Pathologists use various histological stains to discern morphologically distinct cells. However, reports suggest that certain stains are detrimental to DNA quality. Unfortunately, the cause of the degradation is not known. We have shown that increasing exposure of cultured epithelial cells to hematoxylin, a component of H&E, causes the degradation of DNA. It was also shown that methyl green, at its natural pH, degrades DNA. We believe that the low pH is the main cause of the increased degradation. We believe that an acidic environment would cause degradation by the hydrolysis of purine bases in the DNA structure. As an alternative, it was suggested that we localized specific cells on a stained section and microdissect from unstained serial sections. This

alternative would be possible when concentrating on large areas of tumour but in lung biopsies, cells of interest were usually present in a focal manner and there was a high level of cell diversity between serial sections. To overcome this problem, we decided to buffer the various histological stains to a neutral pH and to test their ability to distinguish cellular features. It was clear that the only functioning stain after buffering was methyl green.

4.3 RAPD-PCR, A Method of Genome-Wide Scanning

RAPD-PCR is a DNA fingerprinting technique that involves the PCR amplification of random fragments of genomic DNA with short oligonucleotides of arbitrary sequence. The main advantage of this high throughput technique is the ability to screen multiple loci simultaneously using minute quantities of DNA from archival specimens to detect gene copy number changes without the knowledge of the DNA sequence target.

Apart from the efficiency benefits due to the high-resolution capacity of RAPD-PCR, the discovery of novel genes is facilitated as RAPD-PCR promotes the quick identification of alterations through cloning. At best, other mapping techniques such as CGH resolve increased or decreased copy number changes to chromosomal bands whereas RAPD-PCR alterations are generally in the 100 to 800 bp size range. With the increased completion of the human genome sequence, simple database searches are now replacing experimental chromosomal localization through techniques such as fluorescence in situ hybridization (FISH) or human:rodent monochromosome cell hybrid hybridization (Yasuda, 1996).

A major concern of discovering alterations with RAPD-PCR is the inability of the technique to distinguish between chromosomal gains and losses. In the analysis of the RAPD-PCR fingerprinting gels, alterations are classified as either gains or losses of PCR fragments in the progressive disease stages compared to the normal stages. The technique does not predict the underlying mechanism of the alteration. A gain is characterized as a chromosomal amplification where a specific loci experiences an increased copy number in that region. However, a gain can also be characterized as a chromosomal deletion where new primer binding sites have been created via a deletion or a deletion has brought two primers binding sites close enough together to form a product. A RAPD-PCR loss is characterized as a deletion of a chromosomal region eliminating primer-binding sites. At the same time, a loss can be characterized as insertion of chromosomal material thereby separating two primer-binding sites far enough to prevent PCR amplification. Other infrequent chromosomal aberrations such as translocations or inversion are difficult to distinguish as these changes would appear as either gains or losses depending on the break point.

Another major concern of this genome-wide scanning technique is the potential missing of key genes as the technique is unable to detect single base mutations (i.e. point mutations). Point mutations in certain genes such as K-ras and p16 have been the cause of phenotypic changes. However, it is understood that gene disruption occurs by many different mechanisms in many different patients therefore the same mutational event is unlikely to occur in every patient. Although a gene may be disrupted in one patient by a point mutation, another patient's gene inactivation may be caused by a different mechanisms such as a large deletion. The hope is that the genes involved in signaling

tumour progression would be detected in a large cohort as clustering of locus-specific instability by various mechanisms including larger chromosomal amplifications or deletions.

4.4 Importance of Recurrences

Every individual is composed of a human DNA sequence that is based on the maternal and paternal chromosomes. Although the majority of the DNA sequences in a population is similar, there are regions where DNA sequence is varied. RAPD-PCR targets many different loci with each primer pair but the loci targeted are not always represented in every patient. All four cloned regions stemmed from recurring alterations. Region 8q24.3 was cloned from an alteration that was denoted as a gain of a PCR fragment in the latter stages of progression, specifically CIS. The gain was evident in 6 patients out of 17 patients that exhibited the polymorphism. Region 7q36.2 was cloned from an alteration that was a gain in CIS present in 2 of 15 patients that carried the polymorphism. Region 1q23.3 was cloned from an alteration that was a gain in CIS in 2 of 18 patients whereas 5q33.1 was derived from a gain seen in 2 of 16 patients.

Many mechanisms exist that cause the disruption of a gene's native function. These events include single base mutations, deletions, amplifications, translocations and transversions. Recurrences that occur at the frequencies seen in this thesis, 6/17, 2/15, 2/18, 2/16, are considered significant as the RAPD-PCR technique can only detect a fraction of the gene altering mechanisms. It has been suggested to use more primer pairs to screen for different loci. Sooner or later, by screening more primers, the regions of chromosomal instability will begin to cluster at specific regions. False positive results

are tolerated as only recurring alterations are further analyzed and false negative results are negligible as other primer pairs, in a comprehensive study, would detect the missed loci.

4.5 Two Candidate Regions at 8q24.3 and 7q36.2

After extensive LOH analysis, two novel candidate regions at 8q24.3 and 7q36.2 were identified that exhibited allelic imbalance. Previous reports have suggested these regions as areas of chromosomal instability. Michelland et al.(1999) report consistent patterns of chromosomal alterations in high-grade neuroendocrine (NE) carcinomas and NSCLC using CGH analysis. They observed amplifications in 6 of 11 NE carcinomas involving a minimal region of 8q23-telomere. The same group also reported amplifications in 2 of 11 NE carcinomas involving 7q31 – telomere and deletions in 2 of 11 NE carcinomas involving 7q21-telomere. Virmani et al.(1998) report allelic loss at 7q31.1-q31.2 in 2 of 19 lung cell lines, 1 of 6 SCLCs and 1 of 13 NSCLCs using LOH. Yamada et al.(2000) reported moderate gains of chromosomes 8q and 7q in 8 of 13 and 5 of 13 SCLCs, respectively.

At 8q24.3, high frequency of allelic imbalance was seen in 40% (2/5) of CIS. Similarly, allelic imbalance was seen in 36% (5/14) of CIS at 7q36.2. One can suggest that both these regions experience an early change in the development of NSCLC. To verify these regions as risk predictors, one would need to analyze dysplastic lesions for chromosomal instability at these regions. In comparison, it has been shown that certain regions of allelic imbalance have been effective as risk predictors of NSCLC progression. For example, allelic losses at 1p32-ter have been correlated with late stage NSCLC and

metastasis risk (Chizhikov, 2001). Zhou et al.(2000) report that microsatellite instability at either 3p14 or 10q24 was associated with shortened disease-specific survival. With verification, both 8q24.3 and 7q36.2 may act as a new region of predictive risk.

In close proximity to the cloned region at 8q24.3 exists a known oncogene, c-myc, located at 8q24.12, 12 Mb centromeric to the cloned region. C-myc has been associated with many different cancers including breast, colon and lung. Yamada et al. (2000) tested for c-myc amplification in the 8 SCLC tumours expressing gains in the 8q region and found that five tumours had associated c-myc amplification. These results are consistent with those shown by Little et al. (1983) and Okazaki et al.(1996) demonstrating a several fold increase in copy number of the c-myc gene in primary SCLC. However, three tumours did not show c-myc amplification suggesting that there were other genes in the regions over-represented in SCLC.

4.5.1 Genes Implicated at 8q24.3

Using data from the Human Genome Project at UCSC, we determined that there were seven known genes within a 2 Mb radius of the cloned sequence. These genes included LOC51059, KCNK9 (Kim, 2000), MOST-1, MGC4737, CHRAC1, PTK2 (Fiedorek, 1995) and FLJ12193. Some information is available on these genes but many have undetermined functions and tissue specificity. The KCNK9 gene encodes a protein that acts as a tandem pore domain acid-sensitive potassium channel while the CHRAC1 gene encodes for a chromatin accessibility complex. An interesting observation is that one of the genes, the PTK2 (FAK) gene encodes for a protein tyrosine kinase 2 believed to be an early step in intracellular signaling transduction pathways. The phosphorylation

of PTK2 is reported to be triggered by the interactions of integrins with various extracellular matrix adhesion molecules and by neuropeptide growth factors. PTK2 is expressed in most tissues types including T and B lymphocytes but has been shown to be most abundant in brain tissue. The potential role of PTK2 is believed to be oncogenic transformations resulting in increased kinase activity. Unfortunately, expression data is not available for these genes but possible experiments are described in the future directions

4.5.2 Genes Implicated at 7q36.2

Using data from the Human Genome Project at UCSC, we determined that there are four previously discovered genes within a 2 Mb radius of the cloned sequence. These genes include MLL3 (Tan, 2001), XRCC2 (Thacker, 1995), FLJ21634 and ARP3BETA (Jay, 2000). The MLL3 (myeloid lymphoid leukemia 3) gene encodes a protein associated with leukemia and developmental defects. It is suggested that the protein is involved in transcriptional regulation and studies have shown that MLL3 deletions are associated with hematological neoplasias. The XRCC2 (X-ray repair cross complementing protein 2) gene is believed to be involved in the control of the rearrangement process leading to repair of DNA double strand breaks. Both FLJ21634 and ARP3BETA have functions not remotely relevant to tumourigenesis. Unfortunately, expression data is not available for these genes but possible experiments are described in the future directions.

4.6 Implications of 8q24.3 and 7q36.2 in Epithelial Carcinomas

As epithelial cancers are categorized based on similar progressive characteristics, it is of great interest to compare our candidate region findings with observations seen in other epithelial cancers. Weist et al. (1997) reported that allelic loss was seen at 3p and 9p in lung, head and neck and oral cavity carcinomas. Saha et al. (2001) report identification of a phosphatase gene, PRL-3, at 8q24.3 that is associated with colorectal cancer metastasis. Kurdistani et al. (1998) report the downregulation of the RTP/rit42 gene may be involved in the development of a malignant tumour phenotype. In a separate study in our laboratory, RAPD-PCR was used to identify a region at 7q22 to be linked to oral epithelial carcinogenesis (personal communication). Multiple loci at 7q have been implicated in many cancers such as oral cancer (Wang, 1998), breast cancer (Zeng, 1999) and myeloid neoplasms (Tosi, 1999). Recently, Zenklusen et al (2001) identified ST7 as a highly conserved tumour-suppressor gene on chromosome 7q31. It is clear that both these regions are potential hot-spots of chromosomal instability that need to be further investigated.

4.7 Future Directions

Four distinct chromosomal regions were identified by the analysis of RAPD-PCR fingerprints of progressive stages of NSCLC. These regions include 8q24.3, 7q36.2, 5q33.1, and 1q23.3. Verification was only performed on 8q24.3 and 7q36.2 due to DNA quantity constraints. However, with increased microdissected DNA sources, verification of the other two regions will be performed. Increasing DNA quantity will also allow

fine-mapping of all four regions by defining boundaries for chromosomal instability and the discovery of the oncogenes or tumour suppressor genes in these regions.

As potential oncogenes or tumour suppressor genes are discovered, there are various ways to perform expression studies on these novel genes in NSCLC. In order to analyze expression, one needs to study RNA levels. Northern blot hybridization is an obvious choice for RNA analysis but is only appropriate when high RNA concentrations are extracted from frozen biopsies. An alternative is Real-Time PCR, a highly sensitive quantitative assay designed for analyzing gene expression from minute amounts of RNA. Also, RT-PCR has capabilities to screen a large number of genes and to process a large number of samples. Its sensitive and high-throughput approach makes it optimal for the rapid expression analysis of genes discovered by RAPD-PCR. Expression analysis using RT-PCR is optimal but it requires the availability of fresh frozen material. In many cases, tissue material is only available as formalin-fixed archival biopsies, which does not contain intact RNA due to fixation and lengthy storage. An alternative, in cases where RNA is not available, immunohistochemistry allows the analysis of gene expression using antibodies. The only drawback is that antibodies are not produced for every discovered gene, especially novel genes and all antibodies are not able to hybridize to archival biopsy sections.

With 8q24.3 and 7q36.2 regions verified, the next goal would involve the definition of boundaries of the two alterations. It is understood that the identified region could be altered by many different mutational mechanisms therefore it would be necessary to identify the region directly involved in these events. The easiest way to set boundaries would involve the picking of microsatellite markers on either side of the

cloned region at increasing distances. Boundaries would be set when experimental LOH analysis would show retention in a certain number of patients on both sides of the cloned region using distal markers. However, some disadvantages of this fine-mapping technique exist. Firstly, since the microdissected material is non-renewable, the sample material would not be able to support the hundreds of analyses required to define the boundaries. Secondly, polymorphic markers do not exist at regular intervals throughout the human genome preventing the ability to move at small intervals away from the region of interest. Finally, the definition would be very time-consuming and labor intensive. An alternative to fine-mapping using LOH analysis is the use of a new technology consisting of an array of overlapping bacterial artificial chromosomes (BACs) (Snijders, 2001). The BAC array is composed of sequence verified BACs containing approximately 100 kilobases each, organized in an overlapping fashion from one chromosome end to the other, fixed on a glass support. The technology involves the comparison of fluorescent signal representing hybridization of normal DNA to the signal representing hybridization of tumour DNA to the BAC array. Differences in signal intensities represent gene copy number differences. The rationale is that by using a tiled BAC contig for a specific region, one can identify a small region of a chromosome that has a gene copy number change represented by an altered signal on a BAC. More importantly, the identification of the region would only require small quantities of DNA, as little as 10ng and represent information equivalent to hundreds of LOH analyses. Fine-mapping of the four regions would be more efficient using the BAC array.

Conclusion

In conducting a genome-wide search of novel genetic alterations in archival grade progressive NSCLC biopsies using RAPD-PCR, many experimental goals were identified and achieved. As lung biopsies are composed of many varied cell populations, a method for separating specific cells representing different histological grades had to be developed. Manual microdissection was optimized for the extraction of cells from surrounding cell populations. Although archival tissues present a vast resource of material, the DNA quality is varied between biopsies and is, in some cases, degraded. Suitable staining conditions had to be developed to allow for the visualization of morphologically distinct cells while maintaining DNA integrity. After experimenting with many histological stains including H&E, a stain known as Methyl Green, at a buffered pH 6.0, was chosen as an optimal stain. To prevent the microdissection of archival tissues with initially degraded DNA, unstained biopsies were screened for good DNA integrity using RAPD-PCR. Good quality DNA material extracted from these small NSCLC biopsies was very limiting. A method had to be developed to accurately measure DNA concentrations of extracted material while minimizing loss of DNA used to quantitate. Southern blot hybridization was chosen as an efficient method for assessing DNA quantities. After the microdissection of 180 biopsies representing 41 different patients, RAPD-PCR was conducted to identify novel alterations recurring in multiple patients. RAPD-PCR was able to screen multiple loci from minute archival specimens in an unbiased fashion. Fulfilling the first hypothesis, RAPD-PCR was able to identify fifteen recurring alterations and four alterations, specific to CIS, were mapped to distinct chromsomal regions: 8q24.3, 7q36.2, 5q33.1 and 1q23.3. Due to patient DNA

quantity constraints, two regions 8q24.3 and 7q36.2 were chosen for further investigations. As polymorphic markers are not available in many regions, only a few markers were temperature-tested for verification of chromsomal instability in these regions. Fulfilling the second hypothesis, two of five patients showed allelic imbalance at 8q24.3 and five of twelve showed allelic imbalance at 7q36.2. Many possible genes exist in these two regions that could act as potential oncogenes or tumour suppressor genes however, further studies are necessary to define the expression patterns of these genes.

References

Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, Herman, JG, Wu L, Decker PA, Jen J, Sidransky D. (1999). Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. Journal of the National Cancer Institute 91(4): 332-339.

Ahrendt SA, Decker PA, Doffek K, Wang B, Xu L, Demeure MJ, Jen J, Sidransky D. (2000). Microsatellite instability at selected tetranucleotide repeats is associated with p53 mutations in non-small cell lung cancer. Cancer Research 60: 2488-2491.

Ahrendt SA, Chow JT, Yang SC, Wu L, Zhang MJ, Jen J, Sidransky D. (2000). Alcohol consumption and cigarette smoking increase the frequency of p53 mutation in non-small cell lung cancer. Cancer Research 60: 3155-3159.

Anami Y, Takeuchi T, Mase K, Yasuda J, Hiroshashi S, Perucho M, Noguchi M. (2000). Amplotyping of microdissected, methanol-fixed lung carcinoma by arbitrarily primed polymerase chain reaction. International Journal of Cancer 89: 19-25.

Bergmans HE, van Die IM, Hoekstra WP. (1981). Transformation in Escherichia coli: stages in the process. Journal of Bacteriology 146(2):564-70.

Burkitt, H.G., B. Young and J.W. Heath. (1993). Wheater's Functional Histology. Churchill Livingstone, New York.

Burton MP, Schneider BG, Brown R, Escamilla-Ponce N, Gulley ML. (1998). Comparison of histologic stains for use in PCR analysis of microdissected, paraffin-embedded tissues. Biotechniques 24:86-92.

Chizhikov V, Zborovskaya I, Laktionov K, Delektorskaya V, Polotskii B, Tatosyan A, Gasparian A. (2001). Two consistently deleted regions within chromosome 1p32-pter in human non-small cell lung cancer. Molecular Carcinogenesis 30: 151-158.

De Juan C, Iniesta P, Cruces J, Sanchez A, Massa MJ, Gonzalez-Quevedo R, Torres AJ, Balibrea JL, Benito M. (1999). DNA amplification on chromosome 6p12 in non small cell lung cancer detected by arbitrarily primed polymerase chain reaction. International Journal of Cancer 84: 334-349.

De Juan C, Iniesta P, Vega FJ, Peinado MA, Fernandez C, Caldes T, Massa MJ, Lopez JA, Sanchez A, Torres AJ, Balibrea JL, Benito M. (1998). Prognostic value of genomic damage in non-small-cell lung cancer. British Journal of Cancer 77 (11): 1971-1977.

Dosaka-Akita H, Hu S, Fujino M, Harada M, Kinoshita I, Xu H, Kuzumaki N, Kawakami Y, Benedict W. (1997). Altered retinoblastoma protein expression in nonsmall cell lung cancer: its synergistic effects with altered ras and p53 protein status on prognosis. Cancer 79:1329-1337.

Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. (1999). Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Research 59: 67-70.

Fearon ER, Vogelstein B. (1990). A genetic model for colorectal tumorigenesis. Cell 61: 759-767.

Fiedorek FT Jr, Kay ES. (1995). Mapping of the focal adhesion kinase (Fadk) gene to mouse chromosome 15 and human chromosome 8. Mammalian Genome 6:123-6.

Fritsch P, Riesenberg LH. (1992). High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Dastica glomerata*. Nature 359:633-636.

Gazdar AF, Bader S, Hung J, Kishimoto Y, Sekido Y, Sugio K, Virmani A, Fleming J, Carbone DP, Minna JD. (1994). Molecular genetic changes found in human lung cancer and its precursor lesions. Cold Spring Harbor Symposia on Quantitative Biology 59: 565-571.

Girard L, Zochbauer-Muller S, Virmani AK, Gazdar AF, Minna JD. (2000). Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. Cancer Research 60: 4894-4906.

Greer C, Wheeler C, Michele Manos M. (1994). Sample Preparation and PCR Amplification from Paraffin-Embedded Tissues. PCR Methods and Applications: S113-122.

Hamada K, Kohno T, Takahashi M, Yamazaki M, Tashiro H, Sugawara C, Ohwada S, Sekido Y, Minna JD, Yokota J. (2000). Two regions of homozygous deletion clusters at chromosome band 9p21 in human lung cancer. Genes Chromosomes Cancer 27: 308-318.

Jay P, Berge-Lefranc JL, Massacrier A, Roessler E, Wallis D, Muenke M, Gastaldi M, Taviaux S, Cau P, Berta P. (2000). ARP3beta, the gene encoding a new human actin-related protein, is alternatively spliced and predominantly expressed in brain neuronal cells. European Journal of Biochemistry 267:2921-8.

Jiang S, Sato Y., Kuwao S, Kameya T. (1995). Expression of bcl-2 oncogene protein is prevalent in small cell lung carcinoma. Journal of Pathology 177:135-138.

Kawakami K, Yasuda J, Shiraishi M, Kayama T, Doi K, Perucho M, Sekiya T. (1998). Detection of DNA abnormalities by arbitrarily primed PCR fingerprinting: allelic losses in chromosome 10q in lung cancers. Biochemical and Biophysical Research Communications 251: 153-157.

Kim Y, Bang H, Kim D. (2000). TASK-3, a new member of the tandem pore K(+) channel family. Journal of Biological Chemistry 275:9340-7.

Kohno T, Kawanishi M, Matsuda S, Ichikawa H, Takada M, Ohki M, Yamamoto T, Yokota J. (1998). Homozygous deletion and frequent allelic loss of the 21q11.1-q21.2 region including the ANA gene in human lung carcinoma. Genes Chromosomes Cancer 21: 236-243.

Krishnan BR, Blakesley RW, Berg DE (1991). Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies. Nucleic Acids Research 19(5): 1153.

Kurdistani SK, Arizti P, Reimer CL, Sugrue MM, Aaronson SA, Lee SW. (1998). Inhibition of tumor cell growth by RTP/rit42 and its responsiveness to p53 and DNA damage. Cancer Research 58(19):4439-44.

Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD. (1983). Amplification and expression of the c-myc oncogene in human lung cancer cell lines. Nature 306 (5939):194-6.

Martin GB, Williams JG, Tanksley SD. (1991). Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines. Proceedings of the National Academy of Sciences (U S A). 15;88:2336-40.

Michelland S, Gazzeri S, Brambilla E, Robert-Nicoud M. (1999). Comparison of chromosomal imbalances in neuroendocrine and non-small-cell lung carcinomas. Cancer Genetic Cytogenetic 114: 22-30.

Micheli MR, Bova R, Calissano P, D'Ambrosio E. (1993). Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers. BioTechniques 15(3):388-389.

Momparler R, Bovenzi V. (2000). DNA methylation and cancer. Journal of Cellular Physiology 183: 145-154.

Murase T, Inagaki H, Eimoto T. (2000). Influence of Histochemical and Immunohistochemical stains on polymerase chain reaction. Modern Pathology 13(2): 147-151.

Nomoto S, Haruki N, Tatematsu Y, Konishi H, Mitsudomi T, Takahashi T, Takahashi T. (2000). Frequent allelic imbalance suggests involvement of a tumor suppressor gene at 1p36 in the pathogenesis of human lung cancers. Genes Chromosomes Cancer 28: 342-346.

Okami K, Cairns P, Westra WH, Linn JF, Ahrendt SA, Wu L, Sidransky D, Jen J. (1997). Detailed deletion mapping at chromosome 9p21 in non-small cell lung cancer by microsatellite analysis and fluorescence *in situ* hybridization. International Journal of Cancer 74: 588-592.

Okazaki T, Takita J, Kohno T, Handa H, Yokota J. (1996). Detection of amplified genomic sequences in human small-cell lung carcinoma cells by arbitrarily primed-PCR genomic fingerprinting. Human Genetics 98(3):253-8.

Ong, TM, Song B, Qian HW, Wu ZL, Whong WZ. (1998). Detection of genomic instability in lung cancer tissues by random amplified polymorphic DNA analysis. Carcinogenesis 19(1): 233-235.

Onuki N, Wistuba II, Travis WD, Virmani AK, Yashima K, Brambilla E, Hasleton P, Gazdar AF. (1999). Genetic changes in the spectrum of neuroendocrine lung tumors. Cancer 85(3): 600-607.

Richardson G, Johnson B. (1993). The biology of lung cancer. Seminars in Oncology 20:105-127.

Rosell R, Li S, Skacel Z, Mate JL, Maestre J, Canela M, Tolosa E, Armengol P, Barnadas A, Ariza A. (1993). Prognostic impact of mutated K-ras gene in surgically resected non-small cell lung cancer patients. Oncogene 8:2407-2412.

Saha S, Bardelli A, Buckhaults P, Velculescu VE, Rago C, Croix BS, Romans KE, Choti MA, Lengauer C, Kinzler KW, Vogelstein B. (2001). A phosphatase associated with metastasis of colorectal cancer. Science 294(5545):1343-6.

Sambrook J, Fritsch EF, Maniatis T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Sanchez-Cespedes M, Ahrendt SA, Piantadosi S, Rosell R, Monzo M, Wu L, Westra WH, Yang SC, Jen J, Sidransky D. (2001). Chromosomal alterations in lung adenocarinoma from smokers and nonsmokers. Cancer Research 61: 1309-1313.

Sanchez-Cespedes M, Decker PA, Doffek KM, Esteller M, Westra WH, Alawi WA, Herman JG, Demeure MJ, Sidransky D, Ahrendt SA. (2001). Increased loss of chromosome 9p21 but not p16 inactivation in primary non-small cell lung cancer from smokers. Cancer Research 61: 2092-2096.

Sekido Y, Fong KM, Minna JD. (1998). Progress in understanding the molecular pathogenesis of human lung cancer. Biochimica et Biophysica Acta 1378: F21-F59.

Serth J, Kuczyk MA, Paeslack U, Lichtinghagen R, Jonas U. (2000). Quantitation of DNA extracted after micropreparation of cells from frozen and formalin-fixed tissue sections. American Journal of Pathology 156:1189-96.

Simone NL, Bonner RF, Gillespie JW, Emmert-Buck MR, Liotta LA. (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Reviews 14(7): 272-276.

Shivapurkar N, Virmani AK, Wistuba II, Milchgrub S, Mackay B, Minna JD, Gazdar AF. (1999). Deletions of chromosome 4 at multiple sites are frequent in malignant mesothelioma and small cell lung carcinoma. Clinical Cancer Research 5: 17-23.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. Nature Genetics 29(3):263-4.

Sugio K, Kishimoto Y, Virmani AK, Hung JY, Gazdar AF. (1994). K-*ras* mutations are a relatively late event in the pathogenesis of lung carcinomas. Cancer Research 54: 5811-5815.

Taguchi T, Cheng GZ, Bell DW, Balsara B, Liu Z, Siegfried JM, Testa JR. (1997). Combined chromosome microdissection and comparative genomic hybridization detect multiple sites of amplification DNA in a human lung carcinoma cell line. Genes Chromosomes Cancer 20(2): 208-12.

Tamura K, Zhang X, Murakami Y, Hirohashi S, Xu HJ, Hu SX, Benedict WF, Sekiya T. (1997). Deletion of three distinct regions on chromosome 13q in human non-small cell lung cancer. International Journal of Cancer 74: 45-49.

Tan YC, Chow VT. (2001). Novel human HALR (MLL3) gene encodes a protein homologous to ALR and to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated with leukemia and developmental defects. Cancer Detection Prevention 25:454-69.

Thacker J, Tambini CE, Simpson PJ, Tsui LC, Scherer SW. (1995). Localization to chromosome 7q36.1 of the human XRCC2 gene, determining sensitivity to DNA-damaging agents. Human Molecular Genetics 4:113-20.

Tosi S, Scherer SW, Giudici G, Czepulkowski B, Biondi A, Kearney L. (1999). Delineation of multiple deleted regions in 7q in myeloid disorders Genes Chromosomes Cancer 25(4):384-92.

Tseng JE, Kemp BL, Khuri FR, Kurie JM, Lee JS, Zhou X, Liu D, Hong WK, Mao Li. (1999). Loss of FHIT is frequent in Stage 1 non-small cell lung cancer and in the lungs of chronic smokers. Cancer Research 59: 4798-4803.

Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE. (1992). Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Biotechnology 10:686-90.

Virmani AK, Fong KM, Kodagoda D, McIntire D, Hung J, Tonk V, Minna JD, Gazdar AF. (1998). Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. Genes, Chromosomes & Cancer 21: 308-319.

Wang XL, Uzawa K, Miyakawa A, Shiiba M, Watanabe T, Sato T, Miya T, Yokoe H, Tanzawa H. (1998). Localization of a tumour-suppressor gene associated with human oral cancer on 7q31.1. International Journal of Cancer 75(5):671-4.

Weist JS, Franklin WA, Drabkin H, Gemmill R, Sidransky D, Anderson MW. (1997). Genetic markers for early detection of lung cancer and outcome measures for response to chemoprevention. Journal of Cellular Biochemistry Supplements 28/29: 64-73.

Welsh J, Rampino N, McClelland M, Perucho M. (1995). Nucleic acid fingerprinting by PCR-based methods: applications to problems in aging and mutagenesis. Mutation Research 338: 215-229.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18(22): 6531-6535.

Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon III JW, Garner HR, McKay B, Latif F, Lerman MI, Lam S, Gazdar AF, Minna JD. (2000). High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/ preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Research 60: 1949-1960.

Wistuba II, Behrens C, Virmani AK, Milchgrub S, Syed S, Lam S, Mackay B, Minna JD, Gazdar AF. (1999). Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer, Cancer Research 59: 1973-1979.

Wu X, Zhao Y, Kemp BL, Amos CI, Siciliano MJ, Spitz MR. (1998). Chromosome 5 aberrations and genetic predisposition to lung cancer. . International Journal of Cancer 79: 490-493.

Yamada T, Kohno T, Navvaro JM, Ohwada S, Perucho M, Yokota J. (2000). Frequent chromosome 8q gains in human small cell lung carcinoma detected by arbitrarily primed-PCR genomic fingerprinting. Cancer Genetics Cytogenetic 120: 11-17.

Yasuda J, Navarro JM, Malkhosyan S, Velazquez A, Arribas R, Sekiya T, Perucho M. (1996). Chromosomal assigment of human DNA fingerprint sequences by simultaneous hybridixation to arbitrarily primed PCR products from human/rodent monochromosome cell hybrids. Genomics 34: 1-8.

Zeng WR, Watson P, Lin J, Jothy S, Lidereau R, Park M, Nepveu A. (1999) Refined mapping of the region of loss of heterozygosity on the long arm of chromosome 7 in human breast cancer defines the location of a second tumor suppressor gene at 7q22 in the region of the CUTL1 gene. Oncogene 18(11):2015-21.

Zenklusen JC, Conti CJ, Green ED. (2001). Mutational and functional analyses reveal that ST7 is a highly conserved tumor-suppressor gene on human chromosome 7q31. Nature Genetics 27(4):392-8.

Zhou X, Kemp BL, Khuri FR, Liu D, Lee JJ, Wu W, Hong WK, Mao L. (2000). Prognostic implication of microsatellite alteration profiles in early-stage non-small cell lung cancer. Clinical Cancer Research 6:559-565.

APPENDIX 1.

MICRODISSECTION PROTOCOL

1. Using outlined borders of lesion or area, all H&E slides are photographed and used as guides to corresponding areas on Methyl Green slides.

2. Methyl Green slides are mounted on a dissection microscope, the diagnostic areas identified, hand scraped using a 30G needle and collected into a buffer solution.

3. "Before" and "after" photos are taken on one section per slide as a record.

STAINING PROTOCOL

Hematoxylin and Eosin (H&E)

1. Xylene soak for 10 minutes to remove wax from sections

2. Hydrate to water through graded Ethanol

3. Rinse in water

4. Stain in Hematoxylin for 5 minutes

5. Rinse in water

6. Differentiate in 1% Acid Alcohol, 10 dips

7. Rinse in water

8. Blue in Lithium Carbonate (Saturated) 3 minutes

9. Rinse in water

10. Stain in Eosin 10 dips

11. Dehydrate through graded Ethanol

12. Clear in Xylene and coverslip

Methyl Green

Staining Reagent Preparation:

-Methyl Green Stock Solution (2%)

-2.0 grams of Methyl Green dissolved in 100 ml d H_2O, mixed with 100 ml Chloroform. Shake and discard Chloroform layer. Repeat until violet (Crystal Violet) contamination is gone. Adjust pH to 6.0

Staining Procedure:

1. Dewax slides in Xylene
2. Hydrate through graded Ethanols to water
3. Rinse in dH_2O
4. 10 seconds in 0.2% Methyl Green, pH 6.0
5. Rinse in dH_2O
6. Rinse in 2 changes of 100% Ethanol and airdry

APPENDIX 2.

PATIENT PROFILES

Patient	SEX	BIRTHDATE	SUB GROUP	PACK YEARS
1	F	8/17/28	CURRENT-SMOKER	42
2	F	3/18/28	CURRENT-SMOKER	96
3	F	7/19/38	CURRENT-SMOKER	23
4	F	10/18/32	CURRENT-SMOKER	36
5	F	1/18/37	CURRENT-SMOKER	50
6	M	1/24/38	CURRENT-SMOKER	66
7	M	1/14/23	CURRENT-SMOKER	67.5
8	M	7/10/40	CURRENT-SMOKER	35
9	M	7/17/13	CURRENT-SMOKER	64
10	M	1/10/23	CURRENT-SMOKER	61
11	M	8/21/30	CURRENT-SMOKER	70.83
12	M	5/21/27	CURRENT-SMOKER	70.5
13	M	11/7/46	CURRENT-SMOKER	34
14	M	10/7/38	CURRENT-SMOKER	31.25
15	F	2/5/38	EX-SMOKER	50
16	F	6/23/28	EX-SMOKER	45
17	F	10/6/18	EX-SMOKER	108
18	F	9/9/22	EX-SMOKER	75
19	M	10/14/18	EX-SMOKER	32
20	M	6/11/44	EX-SMOKER	47
21	M	1/23/24	EX-SMOKER	100
22	M	1/19/37	EX-SMOKER	80
23	M	5/16/08	EX-SMOKER	10
24	M	7/12/31	EX-SMOKER	48
25	M	3/19/19	EX-SMOKER	75
26	M	6/15/29	EX-SMOKER	123
27	M	11/14/32	EX-SMOKER	64
28	M	7/13/38	EX-SMOKER	107.3
29	M	3/26/15	EX-SMOKER	43
30	M	10/20/43	EX-SMOKER	30

31	M	7/8/21	EX-SMOKER	47.25
32	M	2/11/20	EX-SMOKER	50
33	M	10/6/46	EX-SMOKER	72
34	M	11/2/32	EX-SMOKER	33
35	M	5/13/34	EX-SMOKER	45
36	F	8/3/52	NON-SMOKER	0
37	F	8/27/30	NON-SMOKER	0
38	F	1/29/30	NON-SMOKER	0
39	F	10/22/58	NON-SMOKER	0
40	F	6/7/60	?	?
41	M	?	?	?

MICRODISSECTED CASES

Patient Number	Microdissected Case	Histological Grade	DNA Concentration (ng/ul)
1	45	3.1	173.7
	44	8.2	38.9
2	113	1	6.5
	B	1	3.5
	1	6.1	52.5
	3	6.1	43.4
	4	6.1	82
	5	6.1	46.1
	6	6.1	42.2
	7	6.1	41.6
	8	6.1	44.3
	9	6.1	6.4
	10	6.1	19.3
	2	7.1	40.9
3	B	1	2.6
	24	1	4.8
	154	5.4	14.5
	25	6.1	42.5
	26	6.1	17.1
	89	6.1	36
	90	6.1	16.8
4	B	1	8.9
	27	2.2	10
	28	6.1	17.3
	88	6.2	40.7
	163	9.4	188.2
5	138	3.2	4
	136	3.2	4.1
	137	4.2	2.6
	135	8.1	27.6
6	114	1	5.6
	46	8.1	256.5
7	B	1	3.1
	48	3.1	22.9

	47	8.1	107.2
8	151	1	54.9
	150	5.2	12.4
	149	6.1	10.4
	152	8.1	21.7
	153	8.1	106.5
9	86	1	14.6
	B	1	30.1
	161	1	47.3
	87	6.1	10.8
10	B	1	110.7
	11	1	13.5
	13	6.1	9.1
	12	6.2	5.1
11	50	1	19.7
	49	8.5	40.5
	51	8.5	42.6
12	115	1	7
	52	8.2	100.1
	53	8.2	260.8
13	B	1	1.4
	64	3.1	27.8
	63	8.1	39.6
14	B	1	35.4
	54	3.1	21.1
	55	8.3	117.7
15	B	1	112.8
	35	1	8.2
	36	6.1	9.2
	37	6.1	10.4
	38	6.1	13.9
	39	6.1	6.9
	40	6.1	7.7
16	116	1	6.6
	56	8.5	101.2
17	B	1	12.2
	58	1	12
	57	8.1	220
18	B	1	3.2
	93	3.1	41.7
	59	8.4	135.7
	60	8.4	X
19	B	1	8.6
	17	1	8.4
	16	3.1	30
	14	6.1	78
	15	6.1	33.2
	18	6.1	80.8
20	61	1	21.1
	62	8.1	190.9
21	B	1	3.8

	68	1	9.3
	160	3.1	12.8
	69	6.1	30.1
	71	6.1	78.2
	70	7.1	30.6
22	80	3.1	X
	79	6.2	X
23	B	1	2.4
	85	3.1	15.1
	82	6.1	46.2
	84	6.1	68.8
	83	8.1	48
24	105	1	1.8
	B	1	15.2
	102	6.1	14.2
	103	6.1	5
	104	6.1	1.8
25	B	1	14
	98	3.1	7
	95	6.1	11.2
	99	6.1	29.4
	100	6.1	51.4
	101	6.1	37.1
	96	7.1	11.9
	97	8.1	37.1
26	B	1	120
	155	5.2	26.5
	156	6.1	21.1
	157	6.1	18.9
	158	6.1	49.9
27	132	1	30.5
	126	6.1	10.2
	127	6.1	9.9
	128	6.1	5.8
	129	6.1	7
	130	6.1	6.5
	131	6.1	24.7
28	109	1	3.6
	139	4.2	2.6
	141	4.2	3
	140	5.2	2.4
	142	5.2	5.4
	107	5.4	6.3
	106	6.1	3.8
	108	6.1	9.9
	110	6.1	10.2
	112	6.1	21.9
	111	8.1	24.1
29	B	1	13.3
	134	3.1	112.7
	133	7.1	48.5

30	124	1	2.3
	122	6	4.4
	123	6	3
	125	8.3	52.5
31	B	1	4.6
	119	1	5.9
	118	6.1	25.5
32	B	1	35.8
	43	1	7.4
	171	4.1	5.2
	173	4.2	7.4
	166	4.2	30.4
	170	4.2	18.5
	174	4.2	10.7
	175	4.2	10.9
	176	4.2	20.3
	169	5.1	11.5
	172	5.1	14.8
	167	5.2	15
	165	5.4	38.5
	164	5.4	16.4
	168	5.4	12.2
	177	6.1	31.9
	178	6.1	75
	179	6.1	53.1
	180	6.1	20.9
	41	7.1	10.3
	42	7.1	20.9
33	B	1	11.6
	81	1	14.4
	74	6.1	19.9
	76	6.1	17.4
	77	6.1	52.6
	78	6.1	X
	75	8.1	47.6
34	147	1	20.2
	B	1	2.5
	148	5.4	4.2
	143	6.1	29.3
	144	6.1	29.3
	145	6.1	66.7
	146	6.1	14.1
35	B	1	7.8
	30	1	8.5
	32	1	13.7
	29	8.2	20.1
	31	8.2	X
36	B	1	159.7
	92	3.2	83.7
	162	3.2	178.8
	19	6.1	56.5
	20	6.1	48.1

	21	6.1	62.4
37	B	1	4.9
	23	1	12.6
	22	6.1	15.3
	94	9.4	21
38	73	1	9.2
	117	1	3.6
	72	6.1	28.3
39	121	1	12.4
	120	8.1	32.2
40	66	3.1	9
	67	5.4	10.8
	65	6.1	13
41	B	1	13.5
	34	6.1	16.4
	33	7.1	10.5
	91	9.4	24.6

APPENDIX 3.

CLONING PROCEDURES

Preparation of Competent E.coli Cells by $CaCl_2$ Treatment.

Preparation

1. Pre-cool centrifuge (the one in Ling lab) to 4 degrees Celsius
2. Pre-cool 0.1M CaCl2 solution to 4 degrees Celsius

Growing cells

3. Grow cells to 0.5 units at OD600
4. Divide into 50 ml aliquots

Harvesting Cells

5. Chill cells on ice (10 min)
6. Spin cells at 3K rpm for 10 min, keep cold!
7. Pour off medium and drain (on ice for 30 sec), pipet off medium

Washing and resuspending cells

8. Add 0.8 ml of ice cold 0.1M CaCl2 and resuspend cells by pipetting (use filtered tip)
9. Add 15 ml of cold 0.1M CaCl2, invert to mix
10. Spin cells at 3K rpm for 10 min, keep cold!

11. Pour off medium and drain (on ice for 30 sec), pipet off medium

12. Add 0.8 ml of ice cold 0.1M CaCl2 and resuspend cells by pipetting (use filtered tip)

13. Add 1.2 ml of 0.1M CaCl2 and store cells in cold room

14. Cells are useful for two days

G-Track Sequencing

Similar to standard sequencing except it involves the use of only one dideoxynucleotide, ddGTP. (see Colony Fingerprinting for protocol).

Plasmid Preparation Protocol

1. From the plate of bacteria take one colony and inoculate 2mL LB broth containing the required antibiotic
2. swirl flask and place 1.5mL culture into 1.7mL microcentrifuge tube
3. spin 1min. 14000rpm
4. remove supernatant by dumping back into culture stock (which we put bleach into to kill): remove the last remaining bit with a pipett
5. resuspend fully in 100uL TE (10mM tris, 1mM EDTA)
6. Add 200uL alkaline lysis buffer (1%SDS, 0.2M NaOH)
7. Allow the samples to sit for 10min

FROM HERE EVERYTHING MUST BE ON ICE

8. add 150uL KOAc (3M, pH 5.5)
9. shake
10. place on ice for 15 min
11. spin samples in cold room (4 degrees) 14000rpm for 10min
12. place the supernatant into new tubes using 200uL pipett (toss the old ones)
13. add 2.5X volume ethanol (100%)
14. place samples in –20 for 30 min
15. spin samples in 4 degrees for 15min 14000 rpm
16. remove ethanol and airdry

17. add 100uL TE (make sure to resuspend pellet completely: do not want to see flakes- vortex and spin down)

18. add 1uL Rnase

19. sit for 30 min

20. add equal volumes of phenol/ shake/let sit for 5min/ spin for 5min to separate

21. take off aq. (top) layer (cut pipette tip): keep phenol tube for disposal

22. "guess and check" vol. with pipette (approx. 80uL)

23. add 1/10 vol NaOAc and 2.5X vol. Ethanol

24. place in –20 for 15 min.

25. spin in 4 degrees for 15min 14000

26. remove supernatant (if smells of phenol wash with 70% ethanol)
    -get EtOH from freezer
    -put in slowly 200uL as to not disrupt pellet
    -let sit for 5-10min
    -spin down and remove supernatant
    -resuspend in TE for freezing
27. freeze the sample in TE

TE: dissolves DNA
KOAc: ppt protein and cell debris
Phenol: protein goes into phenol and DNA goes into aqueous layer (top)

Sequencing Protocol

1. Big Dye reaction:

Big Dye Terminator	4μl
Primer 1.6 μM (T7)	2μl (final conc. = 3.2 pmol)
Plasmid	200ng
ddH_20	adjust
	10μl

2. PCR cycling :

94°C	2 min	
96°C	10 sec	)
50°C	5 sec	| 28 cycles
60°C	3 min	)
4°C		

3. Pelleting:

a. Add all of the Big Dye Rxn (10μl) to a tube with:
   2μl 3M sodium acetate, pH 4.6
   50μl 95% EtOH
b. Vortex. Incubate at room temperature, 15 min for long products, 1 hr for short products, to precipitate the extension products.
c. Centrifuge for 20 min at max speed
d. Remove the supernatant with pipet and discard.

e. Rinse the pellet with 250µl of 70% EtOH. Vortex.
f. Spin for 5 min. Remove the supernatant and discard.

Air-dry the pellet to evaporate all of EtOH. There should be a tiny pink pellet.