Proof-of-Principle of a Clinically-Compatible Assay for 17p LOH Detection and TP53 Mutational Analysis Utilizing LATE-PCR and its Associated Technologies

Kaiane Habeshian

April 28, 2008

Mentors: J. Aquiles Sanchez, Ph.D. and Lawrence J. Wangh, Ph.D.
ABSTRACT

Background: The most commonly altered tumor suppressor gene in human cancers is TP53, which is inactivated by mutations and/or allelic loss (loss of heterozygosity, LOH). Inactivation of TP53 is highly predictive of cancer development in the pre-malignant condition Barrett's Esophagus, the model system for assay construction in this thesis. Current methods for detecting both of these genetic alterations in any given sample are not clinically compatible because they require highly trained technicians capable of carrying out multi-step, time-consuming protocols on several pieces of expensive equipment. Linear-After-The-Exponential (LATE)-PCR provides a practical means to combine both LOH detection and mutational sequence analysis in a single-tube assay.

Results: The TP53 gene inactivation assay described here combines the rs4233018b SNP assay for 17p(TP53) LOH detection with amplification of TP53 exons 7/8 in a single LATE-PCR reaction. This duplex reaction successfully detects 17p LOH in a cell line derived from a Barrett's Esophagus biopsy at PCR end-point. The TP53 amplicon generated in the same reaction was subjected to Dilute'N'Go dideoxy-sequencing, a protocol for sequencing that can be performed directly on LATE-PCR amplification products. Dilute'N'Go dideoxy-sequencing confirmed the presence of a single point mutation in TP53 exon 7 known to be present in this particular Barrett's esophagus cell line.

Conclusions: This thesis provides proof-of-principle for the construction of a clinically compatible diagnostic method for the detection of TP53 tumor suppressor gene inactivation based on LATE-PCR. This assay should have broad applicability since TP53 gene inactivation is a key event in the progression of many pre-malignant conditions towards cancer. The first application of the LATE-PCR TP53 inactivation assay will be to Barrett's Esophagus, a pre-malignant condition where genetic alterations of TP53 are known predictors of future cancer risk.
INTRODUCTION

The problem of cancer and the importance of early detection

Cancer is still the second greatest cause of mortality in the United States. There are 1,437,180 new cases of cancer and 565,650 cancer deaths estimated for 2008 in the United States [1]. Nevertheless, the five-year survival rate has increased from 51% in 1975-1977 to 66% in 1996-2002 [Figure 1a, b]. This decline in cancer death rates is due to improved early detection, as well as more-effective treatments [2]. Screening for morphological changes at the site of cancer can detect cancers of the breast, colon, rectum, cervix, prostate, oral cavity, and skin at early stages. An example of a successful tool for early detection is mammography, which detects 80-90% of breast cancer in women without symptoms by screening for lumps. Partially due to the increase in the use of mammography in women over 40 from 50% in 1991 to 64% in 2002, death from breast cancer decreased 24% from 1990 to 2002 [1]. Unfortunately, such reliable methods for detecting early morphological changes do not exist for other cancers, including lung cancer, the leading cause of death from cancer in men since 1955 and in women since 1987. Unfortunately, too many cases go undetected, even among cancers with screening tools. It is for this reason that substantial research over the past twenty year has been aimed at defining reliable molecular biomarkers for many cancers.

The molecular basis of cancer

In order to understand the method of cancer detection described in this thesis, a basic understanding of the mechanisms of cancer development is required. Cancer is essentially uncontrolled cell proliferation triggered by gradually accumulated genetic alterations. These alterations can affect genes that normally promote cell growth (oncogenes), as well as genes that
normally inhibit cell growth (tumor suppressor genes). Amazingly, just one cell containing these growth-promoting alterations can expand through a process called clonal expansion and can thereby form an entire tumor.

The alterations that produce cancer occur in part in the control of the cell growth cycle. Several checkpoints exist between and within the phases of the cell cycle, M (mitosis), G1 (growth), S (synthesis), and G2, to prevent damaged cells from progressing. After each round of cell division, a cell must decide whether it will divide again or enter a non-growing state (Go). To divide again, the cell must pass through a G1 restriction point (R-point), which requires the presence of extracellular mitogenic growth factors. The production of growth factors is halted under external conditions that may cause cellular damage, resulting in entrance to Go. Thus the R-point serves to prevent cells that may potentially become damaged from proliferating. Once a cell passes through R point, growth no longer depends on external factors and the cell continues to progress through G1, S, G2 and M unless internal damage is detected at further checkpoints. The checkpoint between G1 and S phase prevents cells with damaged DNA from initiating genome replication. Also, a checkpoint operating within S phase slows down or pauses DNA replication if it detects errors in replication. Another checkpoint exists between G2 and M so that only cells that have completed replication can pass into M and also within M to ensure that each daughter cell receives a full set of chromosomes [3].

The molecular machinery that operates at these checkpoints consists of cyclin-dependent kinases (CDKs) that must associate with regulatory subunits called cyclins to fully activate their kinase activity. A unique CDK/cyclin complex operates at each checkpoint. The level of each cyclin fluctuates "naturally" in the absence of cellular damage, causing the activity of the associated CDK to increase or decrease accordingly. Active CDK/cyclin complexes
phosphorylate several downstream targets, leading to mRNA expression for the proteins requisite in the proceeding stage, including the expression of other cyclins. Damage at any point in the cell-cycle will cause the expression of tumor-suppressor genes (TSGs), the products of which can prevent cell-cycle progression, repair DNA damage, or induce cell-cycle arrest and apoptosis if damage is not reversible. TSGs are thus important for preventing cancerous cells from proliferating. Since checkpoint controls usually block cells with mutations, including those that confer a growth advantage, most cancer cells must inactivate one or more of their checkpoints to expand more rapidly.

**The inactivation of TSGs in human cancers: loss of heterozygosity (LOH)**

Oncogenes and tumor suppressor genes are the best-studied genes involved in cancer. Oncogenes act in a dominant manner to drive proliferation. Overexpression of just one copy of these genes is sufficient to confer a growth advantage to the cell. In contrast, both copies of a tumor suppressor gene must be either lost or mutated to inactivate their tumor suppressing function. Thus, the state of both copies must be analyzed when assessing the chances of progression to cancer. Knudson’s two-hit hypothesis [4] states that two independent “hits” must occur to lose both wildtype genes and thus to inactivate the TSG. The first hit to TSGs is generally a point mutation or small deletion, which is inherited in some instances. The first hit causes genomic instability that can lead to the second hit, which may be physical loss of chromosomes or chromosome fragments, promoter hypermethylation or replacement of the remaining wildtype allele with the mutated allele. These events are categorized as loss of heterozygosity (LOH), which refers to the effective abolishment of the heterozygous state of the
TSG, leaving only the mutated allele present either in the homozygous or hemizygous state depending on the mechanism of LOH. Thus LOH results in irreversible allele loss [5].

**The complexities of LOH events**

The commonality of TSG inactivation by LOH in many human cancers led to widespread investigations to discover new TSGs by detecting regions of chromosomes loss that are common to several types of tumors. But, identification of TSGs by LOH is not a simple process because of the biological complexities of LOH events in tumors. “Loss” of an allele refers not only to the physical loss of the respective segment of chromosome but also to the replacement of the original wild-type allele with an altered copy. Thus, a change in copy number of a gene is not necessarily informative in itself, as there may be an LOH event, such as mitotic recombination, gene conversion, or non-disjunction with reduplication, that does not change copy number [6]. Polymorphic markers that are prevalent in the genome, including restriction fragment length polymorphisms (RFLPs), microsatellite polymorphisms [5] and single nucleotide polymorphisms (SNPs), can be used to distinguish two TSG alleles and to identify the loss of one allele rather than directly assessing the TSG. But the cellular complexity of a tumor, which can contain both normal and altered cells, can make it difficult to detect the loss of one allele. LOH events themselves are heterogeneous because some of cells will have accumulated different LOH events compared to others. This variation is due to the clonal nature of tumor development so that only early LOH events are shared among all cancer cells. Therefore, it can be difficult to identify the critical LOH event responsible for loss of growth control.

Another complexity of detecting TSGs is the oversimplification of Knudson’s two-hit hypothesis, which states that both copies of a TSG must be lost to inactivate it. Many TSGs
actually display haploinsufficiency to varying degrees, meaning that one copy of the gene is not sufficient for normal, controlled proliferation. Therefore, loss of one allele of a TSG may be permissive to tumor formation. Furthermore, there are mechanisms in addition to LOH which can contribute to carcinogenesis dominant negative mutations that compete with the normal allele of the TSG, and the acquisition of novel functions by mutated alleles [5].

**The urgent need for clinically-compatible LOH detection in Barrett's Esophagus**

Despite their complexity, LOH events are predictive of cancer risk and can thus serve as clinical biomarkers in pre-malignant tissues associated with various cancers including oral leukoplakia (a precursor condition for oral cancer) and Barrett's Esophagus (a precursor condition for esophageal cancer) [7]. This thesis will focus on LOH and mutational events in Barrett's Esophagus. Barrett's Esophagus (BE) is the best characterized system for analyzing biomarkers predictive of cancer risk, as prospective studies have not been extensively performed for oral leukoplakia. It is the only known precursor to esophageal adenocarcinoma (EA), the incidence of which has increased over the past three decades at an alarming rate of 300% and 350% among male and female Caucasians in the Western world, respectively. Patients with BE have a 30- to 40-fold increase in the risk of developing EA. Approximately 0.5%-1.0% of BE patients per year develop the cancer [8].

Although the rate of EA progression in BE is very low, the cancer is almost always lethal unless it is detected early. BE patients whose cancer was detected early by surveillance have a five year survival rate of 73% compared to 0% among those whose cancers were detected by advanced clinical symptoms [9]. Because of the success of early detection and the fatality of advanced EA, most BE patients undergo endoscopies every six months to three years [10] to
categorize the morphological stage of BE progression. However, because only a small percentage of BE patients develop EA, repeated screening is costly to both the individual and society, indicating the need for a more convenient surveillance method. Furthermore, early detection based on morphological changes is now considered unreliable due to subjectivity of observer classification and heterogeneity of morphological stages of BE [6], indicating the need for objective molecular biomarkers.

Thanks to the efforts of Brian Reid and his collaborators at the Fred Hutchinson Cancer Institute over the last 20 years, three distinct genetic alterations in BE have been found to be extremely predictive of the development of esophageal adenocarcinoma. In their 2007 publication, Reid and his coworkers clearly demonstrated that the following three alterations are significant biomarkers for progression to EA: a) LOH of chromosome 17p affecting the tumor suppressor gene TP53, b) LOH of chromosome 9p affecting the tumor suppressor gene CDKN2A, and c) abnormal DNA content (aneuploidy, tetraploidy). Among patients with all three biomarkers at the start of the study, 79% developed EA over a 10 year period, while only 12% of patients with no baseline abnormality developed EA [7]. Although the combination of the three biomarkers is highly predictive of cancer risk, each biomarker alone also has predictive value and the additional presence of a second biomarker increases the likelihood of developing cancer. Given the extremely predictive value of 9p{TP53} LOH and 17p{CDKN2A} LOH, a clinically compatible assay for detecting these events would eliminate the current dependence on unreliable morphological biomarkers for assessing cancer risk and necessary degree of surveillance of Barrett's Esophagus patients.

This thesis aims to provide proof-of-principle on the development of a clinically-compatible assay for detection of LOH events involving the TP53 tumor suppressor gene in
Barrett’s Esophagus. This subject is very relevant because current methods for detecting LOH events involving TSGs in carcinogenesis are complicated and impractical for routine large-scale implementation due to the complicated nature of LOH events and the lack of a simple method for identifying polymorphisms.

**Alterations of the TP53 tumor-suppressor gene involving LOH and mutations in cancer**

A clinically-compatible method for the detection of 17p{TP53} LOH is crucial not only for monitoring BE status but also because 17pLOH and TP53 mutations occur together as the most common alterations in human cancers. Since its discovery in 1979, the TP53 gene and its protein have been the most extensively studied molecules in cancer biology. The gene is altered in over 50% of human cancers [4, 5, 11], including in over 40% of ovarian and colorectal cancers and over 35% of lung cancers [3]. Studies have indicated that alterations in TP53 are among the most predictive factors for disease development in many cancers, including nasopharyngeal [11], esophageal [12], and breast [13] cancers.

As stated earlier, both copies of the TP53 gene must be lost for its complete inactivation. The first inactivating mechanism is generally a point mutation. The second mechanism of inactivation is generally LOH, as in 92% of cases of esophageal adenocarcinoma (EA) [14]. LOH can involve copy number changes due to DNA deletion as well as copy number-independent events due to mitotic recombination, gene conversion, and nondisjunction with reduplication [6]. Considering the nature of the genetic alterations affecting TP53 during cancer progression, a diagnostic assay for complete TP53 inactivation must include two components. First, the assay must analyze the pre-malignant or cancerous tissue for 17p LOH. Second, the assay must examine the sequence of the remaining TP53 allele for mutational events.
Characteristics of the TP53 gene and its product

The TP53 protein is a transcription factor whose tumor-suppressor activity is due to the expression of the various proteins it regulates. TP53 integrates a wide range of internal and external stress signals and, in response, activates a particular downstream pathway to appropriately handle the damage. Specifically, the TP53 protein integrates signals from genotoxic and non-genotoxic damage. Genotoxic damage includes gamma or UV radiation, alkylation of bases, DNA crosslinking, depurination of DNA, alteration of the deoxyribose sugar moiety, and reaction with oxidative free radicals [19]. Non-genotoxic signals such as hypoxia, activated oncogenes [5,15], nitric oxide, nutrition deprivation, and heat/cold shock can also activate TP53’s various responses [15]. The three major outputs of TP53 activation include cell cycle arrest, senescence, or apoptosis, depending on the particular agents causing the damage and the tissue containing the alterations [15]. As suggested by its ability to integrate many upstream signals to produce several possible outcomes, the complete inactivation of TP53 by sequence mutations and LOH events results in overall genomic instability, which favors tumor progression by promoting the accumulation of further genetic alterations.

Although the inactivation of TP53 clearly compromises the integrity of the genome, mutations in the gene do not generally produce null TP53 proteins, and LOH is necessary for the complete loss of normal TP53 function. These characteristics of the TSG are due to the nature of the mutation events affecting it. The central domain of the TP53 protein (encoded by ~1Kb of DNA) contains 95% of the mutations found in the TP53 gene [3]. These TP53 mutations are unique compared to those in other TSGs because the majority of the alterations are missense mutations (single nucleotide change that produces a single amino acid change) rather than
frameshift mutations (insertion or deletion that disrupts the three-codon reading frame) or nonsense mutations (single nucleotide change that produces a premature stop codon). Unlike frameshift or nonsense mutations, missense mutations lead to a stable protein. Thus, altered TP53 protein often accumulates in the nucleus in human cancers [16].

Partially due to this accumulation of the protein in cancer cells, TP53 was originally classified as an oncogene. It was labeled as such after early experiments in which TP53 cDNA injected into cells interacted with several viral oncogene products to transform otherwise normal cells into tumor cells. However, later studies revealed that the cDNA capable of transforming normal cells actually carried a point mutation in its reading frame, appropriate as the cDNA had been extracted from tumor cells. Eventually, TP53 was correctly classified as a tumor suppressor gene. However, unlike the inactivation of typical TSGs, one mutated copy of TP53 was sufficient to confer a growth advantage to cells containing that mutated copy, a phenomenon called haploinsufficiency. TP53’s haploinsufficiency contradicts Knudson’s two hit hypothesis, which states that both alleles of a TSG must be lost for inactivation of the tumor suppressor protein.

This important observation about TP53 was later reconciled by findings from biochemical analysis of the protein, which led to the discovery that the TP53 protein exists as a homo-tetramer in-vivo [3]. Each subunit of the tetramer consists of five domains: (1) a transcription-activation domain, (2) a Proline-rich domain (involved in apoptotic function), (3) a DNA-binding domain, (4) an oligomerization domain (necessary to form tetramers), and (5) a domain involved in down-regulating DNA binding of the central domain [17]. TP53 protein subunits encoded for by the altered TP53 allele incorporate into tetramers with normal TP53 subunits and thus interfere with normal TP53 function, conferring a growth advantage to the cell.
Because of this dominant negative quality of mutated TP53 protein, it is necessary to assess whether a mutated copy of the TP53 gene is present in order to assess clinical status. However, it is important to emphasize that some normal TP53 activity is retained unless an LOH event follows the mutational event, in which case absolutely no normal TP53 protein remains within that cell.

In addition to the dominant negative phenotype of some TP53 mutants, mutations in TP53 protein often confer oncogenic properties to the protein [15, 18], as demonstrated by the early experiments characterizing the TP53 protein. These mutated proteins act directly as oncoproteins and bestow an additional growth advantage to damaged cells. There are yet other varieties of TP53 mutants. Interestingly, because the DNA-binding site recognized by the TP53 protein is highly degenerate with variable affinity for different sites [19], TP53 protein mutants often retain partial activity on a gene or a subset of its normal genes. Additionally, altered proteins often do not respond to normal regulation mechanisms such as negative regulation by the ubiquitin ligase MDM2 and accumulate in the nucleus [18].

Since different TP53 mutants act in different ways, knowledge of the specific TP53 mutation could be helpful for both characterization of a cancer and diagnosis and treatment of the particular cancer. Furthermore, the complete inactivation of TP53 by an LOH event that follows a mutation event leads to complete genomic instability due to TP53’s wide range of downstream tumor-suppressing effects. Thus, a clinical assay for analyzing mutational status and LOH status of the TP53 gene would provide powerful information for cancer risk assessment.
Inadequacy of the current methods for LOH detection and mutational analysis for widespread clinical application

Current methods of genomic analysis are not compatible for widespread clinical assays because they require multiple laborious, time-consuming steps between biopsy obtainment and subsequent analysis of DNA for LOH and mutations. Most biopsies of pre-malignant tissue contain both normal and neoplastic cells, so biopsy samples must be enriched for pre-malignant cells [6]. Still, because some normal cells remain, LOH assays involve detecting allele imbalances (a change in the amount of one allele compared to the other) rather than the complete absence of one allele. Fluorescent in-situ hybridization (FISH) is a method for detecting copy number loss directly from biopsy material. However, many LOH events, including that of the TP53 gene in BE, are not always due to decreased copy number but can also be due to replacement of the wildtype allele with the mutated allele [6], so FISH is not adequate to detect all LOH events. Other potential methods for LOH analysis require prior gene amplification. After amplification, LOH analysis can be performed in a separate assay through various methods such as gel electrophoresis, MALDI-TOF, CGH microarrays, STR analysis, dideoxy-sequencing, and pyrosequencing. However, all of these methods of LOH detection require significant technical expertise and expensive equipment and are thus impractical for frequent widespread clinical use.

The most direct methods for identifying LOH after gene amplification by PCR are various forms of genomic sequencing (mainly dideoxy-sequencing and pyrosequencing). Dideoxy-sequencing [20] and pyrosequencing [21-23] are also the most reliable and widely used methods for identifying point mutations and small deletions, necessary to understand the full risk of cancer development. Sequencing is still cumbersome, however, because all of its forms
consist of multi-step protocols after PCR. Since traditional asymmetric PCR is highly inefficient, gene amplification through symmetric PCR is necessary to generate a sufficient amount of DNA for sequencing. The shortcoming of combining symmetric PCR with sequencing is that it requires either a separate clean-up of reaction products, such as gel electrophoresis to isolate the desired DNA product, or careful titration of primers and dNTPs to ensure their depletion by the end of PCR. These limitations of sequencing have implications for high-throughput analysis required in clinical settings.

The ideal LOH assay would be performed directly on PCR products, but current PCR methods are not reliable for end-point LOH analysis. Symmetric PCR exits exponential amplification stochastically, and thus small differences in the time it takes to reach the plateau phase of the reaction produce highly variable amounts of total amplification products, regardless of starting amount. Therefore, accurate quantitative end-point analysis is not possible since the amount of accumulated amplicon does not reflect the initial amount of that allele and also because a great amount of variation exists among replicates [24]. This problem of variability could be solved using two probes of different colors to determine the relative amounts of each allele within a replicate. However, this method reduces the multiplexing capacity of the assay, since thermocyclers are capable of reading only a limited number of probes. Use of asymmetric PCR is another strategy to minimize variations in PCR product yield, since amplicons accumulate arithmetically. However, conventional asymmetric PCR is inefficient because it uses two primers of equal T_m (concentration-dependent melting point, at which 50% of the primer is bound to its target) at different concentrations. The use of limiting amounts of one of the primers leads to a decrease in its apparent T_m relative to the excess primer, dropping its T_m below the PCR annealing temperature and hindering the participation of this primer in the PCR reaction.
Lowering annealing temperature to compensate for the drop in limiting primer $T_m$ increases the likelihood of nonspecific amplification, since the affinity of primers for imperfect matches increases at lower temperatures. There is a need for a method of LOH detection that can be done directly on PCR amplified products, that is not based on copy number analysis, and that uses a single hybridization probe.

**A new strategy for clinically-compatible LOH detection**

Linear-After-The-Exponential (LATE)-PCR is an improved, reliable version of asymmetric PCR that permits end-point analysis, regardless of the starting amount of DNA. The key to LATE-PCR is the use of specific concentration ratios of specially designed primers that allow both primers to actively participate in the reaction, generating a sufficient amount of template to produce robust amounts of single-stranded DNA once the limiting primer has been completely incorporated into double-stranded DNA. LATE-PCR is different from conventional asymmetric PCR in the design of its limiting primer and excess primer, with $T_m^L$ and $T_m^X$, respectively, which follow the rule $T_m^L - T_m^X > 0$ (ideally 5°C). Traditional asymmetric PCR limiting primers, which have the same $T_m$s as the excess primer at equal concentrations, do not efficiently participate in the reaction, since decreasing the concentration of the limiting primer causes a drop in its apparent $T_m$ relative to excess primer $T_m$. To prevent this apparent drop in $T_m^L$, LATE-PCR limiting primers are designed to be longer than the excess primers. The increase in number of base pairs raises $T_m^L$ above $T_m^X$ so that apparent $T_m^L$ does not fall below annealing temperature during PCR. The full participation of the limiting primer renders LATE-PCR PCR as efficient as symmetric PCR [24].
The abundant amount of single-stranded DNA produced by LATE-PCR allows the
detection of two distinct alleles with one mismatch-tolerant probe in an end-point melt.
Mismatch-tolerant probes bind preferentially to their perfect complement at high temperatures
but bind equally to both the perfect complement and a mismatched target (the complement with a
single nucleotide change) at low temperatures. The percentage of each allele present in the
reaction can be determined by ratios of fluorescence at high to low temperature, a process which
normalizes the amount of single-stranded DNA products among replicates. Mismatch tolerant
probes can be targeted to single nucleotide polymorphisms (SNPs) to distinguish two alleles,
rendering LATE-PCR an effective method for genotyping and for LOH detection. In
genotyping, a heterozygous sample would yield 50% of the fluorescence of the low temperature
at the high temperature, while a homozygous sample would yield either 0% or 100% of the
fluorescence of the low temperature at the high temperature, depending whether the mismatched
or perfectly matched allele were present, respectively. Similarly, by comparing the percentage of
an allele in normal heterozygous cells to pre-malignant cells with LOH, one can readily define
the later, in which there would be either 0% or 100% of the probed allele depending on which
allele was lost [25].

Detection of 17p(TP53)LOH based on SNP sites with subsequent mutational analysis of
the TP53 gene requires multiplexing of amplicons containing the SNP site and the critical TP53
exon regions typically targeted by mutations. This thesis aims to provide proof-of-principle on
the multiplexing capacity of LATE-PCR SNP assays with TP53 amplification. Successful PCR
multiplexing depends on avoidance and suppression of mis-priming. Mis-priming refers to non-
specific primer interactions, which include primer-dimer formation, mis-priming of unintended
targets at low temperature before the start of the reaction and similar mis-priming events during
the early phase of amplification when target genomes are denatured and primers are in vast excess. In addition, the 3' ends of single-stranded amplicons generated in asymmetric reactions can prime on other single-stranded molecules accumulating in the reaction, particularly when the temperature is lowered. To enable multiplexing, the Wangh Laboratory developed a new class of molecules known as PrimeSafe (Smiths Detection, Edgewood, MD, USA), which can be added to both symmetric and asymmetric reactions to suppress all forms of mis-priming throughout amplification. One of the clear benefits of PrimeSafe is that it can be adjusted in terms of precise composition and concentration to minimally impact the efficiency of each monoplex or multiplex reaction. As a result, multiplex LATE-PCR assays can conveniently be readily constructed without the need to perfect primers to avoid possible non-specific interactions [26].

The abundant amount of single-stranded DNA produced by LATE-PCR allows for the streamlining of the process of LOH detection and sequencing through Dilute-'N'-Go dideoxy-sequencing, developed at the Wangh laboratory. Dilute-'N'-Go dideoxy-sequencing is more convenient, faster, and less expensive than sequencing double-stranded amplicons generated via conventional symmetric PCR. The preparation of LATE-PCR products for Dilute-'N'-Go sequencing simply requires diluting the PCR reaction to render concentrations of substances besides the desired single-stranded DNA negligible. Depending on which strand is to be sequenced, either the limiting primer is added back to the reaction or excess primer is allowed to prevail. The reaction is then sent off to a commercial dideoxy-sequencing company [26].

In this thesis, LATE-PCR LOH analysis followed by Dilute-'N'-Go sequencing are combined into a convenient multiplexed assay capable of detecting 17p{TP53} LOH and TP53 point mutations in BE cell line DNA. Specifically, the rs4233018b SNP assay was multiplexed
with primers for an amplicon containing exons 7 and 8 of the TP53 gene. This multiplex assay was capable of detecting LOH in a BE cell line known to contain TP53 LOH. Furthermore, Dilute-'N'-Go sequencing performed on the same samples was able to identify a known point mutation in exon 7 of the TP53 gene in that cell line. Based on these results, this thesis contributes to construction of a clinically useful multiplexed LATE-PCR assay for detection of LOH followed by immediate analysis of genetic mutations in the remaining copy of the gene.
Tables and Figures

Figure 1a: Cancer death rates for men in the US from 1930 to 2003 (American Cancer Society)

Figure 1b: Cancer death rates for women in the US from 1930 to 2003
MATERIALS AND METHODS

SNP Sites

Two heterozygous single nucleotide polymorphisms (SNP sites) in a normal genome are needed to reliably establish a diagnosis of LOH in the genome of a clinical sample from the same individual. The formula \( n = 2x (\log (0.02)/\log (1-HI)) \) was used to determine the number of SNP sites (n) around the TP53 gene required to obtain at least two heterozygous SNP sites with a given heterozygosity index (HI), in 98% of patients [27]. HI is the proportion of people in a population who are heterozygous for a given SNP site. The average HI of the SNP sites meeting the chosen criteria (discussed below) was calculated to be 0.495 using the average of the lower estimate of HI for these SNP sites, omitting the cases where margin of error is above 0.1. According to the above formula, \( n = 18.8 \). This number was rounded to 20 SNP sites to account for the SNP sites with a larger margin of error.

The following criteria was used to identify SNP sites in the vicinity of the TP53 gene for which to develop LATE-PCR assays used to screen for heterozygosity. The Pubmed dsSNP database [28] was searched to locate SNP sites in the vicinity of the TP53 gene that (1) have at least 40% heterozygosity index, (2) are located within 200 kBP upstream or downstream of the TP53 gene, (3) are found within a unique sequence complex, and (4) have DNA samples of known genotype available for assay validation. From the 20 selected SNP sites for which LATE-PCR assays are either completed or under construction (Table 1), assays for the following six SNP sites were validated to use in LOH analysis of BE cell line DNA: rs4233018b, rs8066124, rs8073498, rs858521, rs2543540, and rs4511593. The arrangement of these SNP sites around the TP53 gene is shown in Figure 2.
DNA Samples:

Cell line DNA derived from BE patient biopsy samples and corresponding normal DNA samples from the same patients were obtained from the Fred Hutchinson Cancer Research Center (Seattle, WA, USA) and stored at -20°C before use in PCR. 60 ng DNA from each sample at 3 ng/ul were received from three cell lines CP-B, CP-C, and CP-D from three Barrett’s patients, 13725, 15249, and 12881, respectively. Normal DNA from each patient was received for genotyping SNP sites. The normal DNA from cell lines CP-B, CP-C, and CP-D was 7031N, 310N, and 330N, respectively. Cell lines CP-C and CP-D each included DNA derived directly from Barrett’s epithelial cells, whereas cell line CP-B only included normal DNA. Table 2 characterizes the BE cell lines [12].

Probes:

Previously designed double-stranded hybridizations probes (DsPs) were used for the detection of the TP53 amplicons (Jesse Salk, Senior Honors Thesis, Brandeis University), with two modifications (Table 3). Exons 5-8 were chosen for amplification because 95% of TP53 point mutations found in human cancers occur in this region [3] (Figure 3). Fluorescent probes were specially designed and utilized for the SNP assays in this thesis. These probes consist of a linear oligonucleotide carrying a 3’ FAM fluorophore and a 5’ Black Hole Quencher (BHQ-1). An exception to this rule for probe design was the original rs4233018b probe, which contained a HEX fluorescent moiety at the 3’ end. This probe was subsequently modified to contain 5’ FAM and 3’ BHQ-1 before use on cell lines.

Probes and primer sequences for LOH assays were designed with the help of Visual OMP software (DNA Software, Inc., Ann Arbor, MI). Design specifications for fluorescently-labeled,
low T<sub>m</sub> mismatch-tolerant probes are as follows. The probe sequence comes from the limiting primer strand sequence, since the probe is complementary to the excess primer strand that accumulates as single strands in LATE-PCR. The probe will bind to and detect relative amounts both a matched and mismatched target with a single base-pair change. Therefore, to allow the greatest distinction between the matched and mismatched target, the mismatched target should exhibit the larger probe-target hybrid instability out of the two possible SNP allele choices. The relative stability of mismatched dideoxynucleotide bases is CG>AT>GG>GT>GA-AA>TC>AC [29]. Since LATE-PCR probes are low-T<sub>m</sub> (i.e., they bind below the PCR annealing temperature), their T<sub>m</sub> is constrained by the T<sub>m</sub> of LATE-PCR primers and amplification conditions. See T<sub>m</sub> specifications for probes in the primer section below.

**Data Analysis:**

In this thesis, the six SNP assays for LOH detection in BE cell lines were chosen on the basis that they fit the following binding criteria both in silico (VOMP, Figure 4) and on synthetic test target melts without Taq polymerase (Figure 5): (1) the probe bound the majority of the matched target before it began binding to the mismatched target at some middle temperature and (2) the probe bound 100% of alleles by 20°C. The experiments to matched and mismatched targets consisted of four replicates of each matched DNA, mismatched DNA, and no template controls (NTCs) that contained probe but no target. NTCs serve as a control to measure background fluorescence at each temperature, accounting for possible slight differences in the amount of probe among replicates. To test whether the probes maintained these binding criteria in the presence of Taq polymerase and other PCR conditions, their binding to short known homozygous and heterozygous DNA targets in PCR was tested (Figure 6). These experiments
included four replicates of each known heterozygous DNA, each homozygous allele, and NTCs. The SNP assays on short known targets were then repeated with 24 replicates to provide sufficient data for three-temperature analysis. In three-temperature analysis, fluorescence at three temperatures below the annealing temperature is measured at end-point to determine the relative amounts of each allele. The upper temperature is chosen to be too high for the probe to hybridize to either the perfectly complementary (matched) allele or the partially complementary (mismatched) allele and therefore provides a measure of background fluorescence. The lower temperature is chosen as the temperature at which the probe binds equally to the mismatched allele and matched allele. The middle temperature is the temperature at which there is the greatest difference in binding of the probe to its matched allele and its mismatched allele. The signals at the low and high temperature are then used to normalize the fluorescence at the middle temperature, which provides a basis to distinguish heterozygous and non-heterozygous DNA.

The following paragraph outlines the steps taken to normalize the fluorescent signals produced by LATE-PCR using Microsoft Excel (Figure 7a-f). The average fluorescent values of the NTCs at each temperature were calculated and subtracted from the raw data, and these values were plotted (Figure 7a,b). The first derivative of the raw fluorescence data minus fluorescent values of NTCs was calculated and plotted to determine the high, middle, and low temperatures for three-temperature analysis (Figure 7c). A plot of the first derivate produces two peaks for heterozygous DNA and one peak for each allele of homozygous DNA. The high temperature was chosen as the temperature 1°C inside the second peak on the first derivative graph. The middle temperature was chosen as the point between the two peaks at which fluorescence is zero. The low temperature was chosen as the temperature 1°C into the first peak. The raw fluorescent values were then divided by the values at the upper temperature and plotted (Figure 7d).
Average fluorescent values of NTCs were subtracted from the data at each temperature (Figure 7c). The upper temperatures chosen for normalization were slightly adjusted empirically for each SNP assay to produce the greatest distinction in probe binding to two alleles. The above steps were taken to determine whether a probe met binding criteria in experiments with matched and mismatched targets and of short known heterozygous and homozygous targets in PCR. The following additional steps were taken to test the ability of the SNP assays to distinguish known homozygous and heterozygous DNA. The data was normalized at the lower temperature, which accounts for variation in yield among replicate samples (Figure 7f). Then, the normalized fluorescent signals at middle temperature were compared between known heterozygous DNA and known homozygous DNA (Figure 7g). Boxes of +/- 3 standard deviations encompassing the known heterozygous samples were created and illustrate the 99.7% confidence interval for the fluorescent signals corresponding to the heterozygous genotype. Signals from known homozygous samples should fall outside this box if the SNP assay is adequate for distinguishing heterozygous and non-heterozygous DNA. The three temperatures chosen for analysis of the six SNP assays (Table 4) remained consistent in further experiments.

**Primers:**

Asymmetric PCR primers for amplification of the TP53 exons 5-6 (amplicon 5/6) and exons 7-8 (amplicon 7/8) were previously designed and validated at the Wangh Laboratory. SNP primers were designed according to the following criteria. The excess primer strand was chosen as the DNA strand with lesser secondary structure for the mismatch allele at 30°C, 70 nM monovalent cations, 3mM MgCl2 within 100 nucleotides surrounding the SNP as determined *in silico* by Visual OMP Software (VOMP). For LATE-PCR LOH assays, the following $T_m$ criteria
should apply to primers and probes:

1) **Limiting primer** $T_m = 71^\circ C$-$72^\circ C$ @ 50 nM and **Excess primer** $T_m = 66^\circ C$-$67^\circ C$ @ 1000 nMs, according to LATE-PCR primer design criteria ($T_m^L-T_m^X-5^\circ C$)

2) **Annealing temperature** = 63$^\circ$C-$64^\circ$C ($-2^\circ$C-$3^\circ$C below the excess primer $T_m$)

3) **Probe $T_m$ for the matched target** $\leq 62^\circ$C-$63^\circ$C @ 500 nM and 150 nM probe and target concentrations, respectively ($-9^\circ$C-$10^\circ$C below the limiting primer $T_m$ to prevent the probe from interfering with limiting primer extension).

4) **Probe $T_m$ for the mismatched target** $\sim 52^\circ$C-$54^\circ$C but ideally no lower than $45^\circ$C @ 500nM probe and 150 nM target concentrations ($-8^\circ$C-$10^\circ$C below the $T_m$ of the probe-matched target hybrid to allow the probe to bind to completion to the matched allele before binding to the mismatched allele).

Table 5 contains the $T_m$s of the probes, primers and amplicons for the six chosen SNP assays.

**PCR Conditions:**

Amplification and fluorescence detection were carried out using two PCR thermal cyclers: the ABI Prism 7700 Sequence Detector (PE Applied Biosystems, Foster City, CA, USA) and the Bio-Rad iQ-5 Multicolor Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) to show that the assays can be used across multiple platforms. Reactions of 25 μl included 1.25 units of Platinum Taq (Invitrogen Corp., Carlsbad, CA, USA) in 1x Platinum Taq Buffer, 3 mM MgCl₂, 250 nM dATP, dGTP, dCTP, dTTP (Promega, Madison, WI, USA), 25 nM PrimeSafe-060 (Smiths Detection, Edgewood, MD, USA), 300 nM PrimeSafe-001 (Smiths Detection, Edgewood, MD, USA), 50 nM limiting primer, 1000 nM excess primer, and 500nM probe (in the ratio of 300nM fluor:250 nM quencher for dsPs). Melts for match and mismatched targets without Taq polymerase consisted of the following steps: 95.0
C/2:00 min; 20.0°C/20:00 min; 1.0C increases from 20.0°C every 1:30 min to generate a melt profile. Thermocycling profiles of TP53 exon amplicons consisted of a 95°C/3 minute denaturation stage followed by 25 cycles of 95°C/10 sec, 65°C/30 sec and 75°C/30 sec, then 35 cycles of 95°C/10 sec, 65°C/30 sec, 75°C/30 sec and 45°C/20 sec. Thermocycling profile of amplification of SNP sequences consisted of a modified TP53 amplification protocol without the 45°C drop for reading. Multiplexing of TP53 amplicons and SNP sites was optimized by utilizing a modified TP53 amplification protocol without the 45°C drop for reading and also by modifying annealing temperature (see results for further details). Fluorescence data from the melt after thermocycling was used for end-point data analysis.

**Dilute' N'Go Dideoxy-Sequencing:**

PCR reactions containing the TP53 amplicon were diluted 1:20 with 10X Tris HCl pH 8.3 to prepare the samples for sequencing. 1 µl, 2 µl, 4 µl or 8 µl of diluted reaction mixture and 1 µl of limiting primer were added to 23 µl, 22 µl, 20 µl or 16 µl distilled H2O, respectively, for a final volume of 25 µl. Two 12 µl aliquots from each 25 µl mixture were sent to GENEWIZ Inc. (South Plainfield, NJ, USA) for dideoxy-sequencing. Results were obtained via the website [www.genewiz.com](http://www.genewiz.com) and read with Sequence Scanner 1.0 (Applied Biosystems, Foster City, CA, USA).
### Tables and Figures

<table>
<thead>
<tr>
<th>SNP rs#</th>
<th>Gene(s)</th>
<th>Alleles</th>
<th>Distance from previous SNP</th>
<th>Hetero. Index in Caucasians</th>
<th>Hetero. Index in all population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.45 +/- 0.15</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.41 +/- 0.20</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39 +/- 0.21</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.48 +/- 0.15</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.36 +/- 0.22</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Characterization of 20 SNP sites around the TP53 gene for which assays have been completed (green) or for which assays must still be completed (blue).

<table>
<thead>
<tr>
<th>CP-52731</th>
<th>5q</th>
<th>13q</th>
<th>18q</th>
<th>9p</th>
<th>CDK4/RB1</th>
<th>17p</th>
<th>P53</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CP-B)</td>
<td>2</td>
<td>N1</td>
<td>1</td>
<td>1</td>
<td>Leu-&gt;Gln 97</td>
<td>1</td>
<td>Arg-&gt;His 175</td>
</tr>
<tr>
<td>CP-94251</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>Wild-type</td>
<td>1</td>
<td>Arg-&gt;Trp 248</td>
</tr>
<tr>
<td>(CP-C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP-10821</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>Del(C5.1) (4N,8N)</td>
<td>1</td>
<td>Frameshift 302</td>
</tr>
<tr>
<td>(CP-D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Characterization of three Barrett's cultures that contain 17p LOH. Normal DNA corresponding to all three cell lines was available. DNA from BE cell line CP-B was not available [12].
Figure 2: Arrangement of SNP sites around the TP53 gene on chromosome 17p (HapMap.org)

Figure 3: Mutation frequency distribution in TP53 exons 4-10 (IARC Database 2003). The majority of TP53 point mutations found in human cancer occur in exons 5-8.
Primer Sequences

5/6 Excess 5' CTTTCTTTGCTGCGGTGTTTC 3'
5/6 Limiting 5' CCCAGAGACCCACGGTCTAACGAC 3'
7/8 Excess 5' CTCTGACTGTACCACCATCCACTAC 3'
7/8 Limiting 5' CCACCGCTTCTTGACCTGCTTGCTTACC 3'
rs2543540 Excess 5' ATGTTGAAAACCTTAAAAAGGCTGTGG 3'
rs2543540 Limiting 5' TTACAGGGCCATGCAGAATCATCATTCCC 3'
rs4233018b Excess 5' GTAGAGTACAGTGCAAATGCTATATT 3'
rs4233018b Limiting 5' CCGTGCTGGCCAAACAGATATTTAAAAAACA 3'
rs4511593 Excess 5' CAAGGAAATAAAGTTGGAGATTATGT 3'
rs4511593 Limiting 5' GCTAGTCTGACAATGTCCAGCAGCATCCAT 3'
rs8066124 Excess 5' TICCTAGATTTTTCCATAGGCCAAAAAG 3'
rs8066124 Limiting 5' TAATCTCTACTTTTACAGATCTTGGTGGGCGAG 3'
rs8073498 Excess 5' TTGGAGATTTTGAGTTTCAGATCTCATT 3'
rs8073498 Limiting 5' AGCTGTCTCTAGGCCCTAGAAAACCTATG 3'
rs858521 Excess 5' CAATCCCTTGACCTGTGGTGG 3'
rs858521 Limiting 5' GCCCAGCCGGGTGCATTCTCTGATCC 3'

Double Strand Probes Sequences

5/6 Fluor 5' [TET] AGACAGAGTGGAAAAATGCG [P] 3'
5/6 Quencher 5' GACTTTACAATTCTGTCT [DAB] 3'
7/8 Fluor 5' [FAM] GGAGTGATGATGAGTGTGA [P] 3'
7/8 Quencher 5' TCACACTACTACCAC [DAB] 3'

Original probes designed by Jesse Salk:
5/6 Fluor 5' [TET] AGACAGAGTGGAAAAATGCG [P] 3'
5/6 Quencher 5' GACTTTACAATTCTGTCT [BHQ-1] 3'
7/8 Fluor 5' [FAM] GGAGTGATGATGAGTGTGA [P] 3'
7/8 Quencher 5' TCACACTACTACCAC [BHQ-1] 3'

Low-Tm, Mismatch-Tolerant Exo-R Probe Sequences

rs2543540 5' [BHQ1] TTCCAATACGAGCAAA [FAM] 3'
rs4233018b 5' [HEX] TACCTTTAGGCTCAATA [BHQ1] 3'
rs4233018b REV 5' [BHQ1] TACCTTTAGGCTCAATA [FAM] 3'
rs4511593 5' [BHQ1] ATATTTAGCCACTGTTAT [FAM] 3'
rs8066124 5' [BHQ1] CACCCCGAGCCCTG [FAM] 3'
rs8073498 5' [BHQ1] GGCAGACATGGTCCCTG [FAM] 3'
rs858521 5' [BHQ1] CTCTTAGCTTCGAACATAG [FAM] 3'
**Amplicon Sequences**

rs2543540 (192 BP):
AATCTGGGTGGTGCTACAGGTGTCATCCTGGAATTTTCTTGGAACTTTTCTGTATA
TATCTTTGGAAAAATTTTGGAACAAAAATTTGCAGTGTGTTCTTGGC
TCGTATTTGGCCGTTACATATAACAAATGAGGAAAGGATGAGGATTGCCATTGGCC
TATGAAACAAATGACGCAAATTT

rs4233018b (150 BP):
GTAGAGTACAGTGCAAGCCATATTGGAAACCTAAGGGGAAAGGAGAAATCAGTAAT
CCCCCATCTGCTTTATTTATAATTTGTTGTGGTTGCTCTCATGAAATATTTTGCAATTC
TTTTGTCTTCTAAATCTGTGTGTGGCCAGGCACGG

rs4511593 (125 BP):
CAAGGAAATAAAATAGGTTGGGAAGTATGATAGATAATTATTAACGTTGAGAAAGTCTA
AAATAAAACAAAGTGGAATAAATGCAAATAAAATGATAAAATTCAGAAATGGATGGCTGGAG
ATTGTCGACTAGC

rs8066124 (137 BP):
CACTGTATAAACACACCCACCTAAGAACTTGGGCGTCCGTCTTCCTAGATTTTCCCAT
GGAAAAAGAAAGAAGGGCTGGGGGAAAGAAGCTTGGGGGATCCACTGGCCACCC
AACAGCCTCTGAAAGGATAGGATTA

rs8073498 (102 BP):
TTGGGAGATTTCGATTCATATCAATTAGGATAAAGGCCTGGGTCTCTGGAGTGAGG
GCTAGGGAAATGTCATGGCCTATTCCAGGGGCTAGAACAGCT

rs858521 (87 BP):
CAATCCCTTTGGACATTGGTGGAGAGAAAGGAGAAAAATAGCTATTGTTGAGCTG
GACGAGGGATCAGAAAAATGACACCGGCCTGGGC

---

**Table 3**: Probe, primer and amplicon sequences
Figure 4: rs2543540 probe binding profile to matched (green) and mismatched (blue) targets in VOMP demonstrates probe binding criteria: (1) at some middle temperature, the majority of probe binding is to the matched target and (2) the probe is bound to the totality of the targets by 20°C.

Figure 5: rs2543540 probe binding profile to matched and mismatched targets. Raw fluorescence data was normalized at the upper temperature (59°C). rs2543540 probe binding profile to matched target (red), mismatched target (green) and NTC (black) demonstrates probe binding criteria described in Figure 4.
Figure 6: rs858521 probe binding profile to short known heterozygous (blue), homozygous 1 (red) and homozygous 2 (green) targets amplified in PCR. Raw fluorescence data was normalized at the upper temperature (58 °C). Demonstrates that probe maintained binding criteria under PCR conditions.

Figure 7a: Plot of raw data for demonstration of normalization procedure.
Figure 7b: Raw data minus the average of NTCs

Figure 7c: First derivative of raw data minus average NTCs to determine three temperatures for normalization: 27°C (lower temperature), 46°C (middle temperature), and 65°C (upper temperature)
Figure 7d: Raw data normalized to the upper temperature (65°)

Figure 7e: Data normalized at upper temperature (65°) minus average of NTCs
Figure 7f: Data normalized to upper temperature (65°C) and lower temperature (27°C)

Figure 7g: Data normalized to all three temperatures (middle temperature = 46°C)
<table>
<thead>
<tr>
<th>SNP</th>
<th>Lower Temperature</th>
<th>Middle Temperature</th>
<th>Higher Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2543540</td>
<td>21</td>
<td>38</td>
<td>52</td>
</tr>
<tr>
<td>rs4233018b</td>
<td>33</td>
<td>47</td>
<td>61</td>
</tr>
<tr>
<td>rs4511593</td>
<td>35</td>
<td>42</td>
<td>52</td>
</tr>
<tr>
<td>rs8066124</td>
<td>20</td>
<td>50</td>
<td>62</td>
</tr>
<tr>
<td>rs8073498</td>
<td>33</td>
<td>47</td>
<td>61</td>
</tr>
<tr>
<td>rs858521</td>
<td>35</td>
<td>46</td>
<td>56</td>
</tr>
</tbody>
</table>

Table 4: Three temperatures (°C) chosen for normalization from probe binding to short known homozygous and heterozygous samples in PCR

<table>
<thead>
<tr>
<th>SNP</th>
<th>Variation</th>
<th>LP length</th>
<th>LP Tm</th>
<th>EP length</th>
<th>EP Tm</th>
<th>Amplicon length</th>
<th>Amplicon Tm</th>
<th>Probe Length</th>
<th>M Tm</th>
<th>MM Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2543540</td>
<td>A/G</td>
<td>30</td>
<td>71.8</td>
<td>30</td>
<td>67.1</td>
<td>95</td>
<td>83.6</td>
<td>16</td>
<td>52.3</td>
<td>37.1</td>
</tr>
<tr>
<td>rs4233018b</td>
<td>A/G</td>
<td>32</td>
<td>71.0</td>
<td>25</td>
<td>66.0</td>
<td>150</td>
<td>83.0</td>
<td>17</td>
<td>57.3</td>
<td>45.2</td>
</tr>
<tr>
<td>rs4511593</td>
<td>C/T</td>
<td>27</td>
<td>71.0</td>
<td>28</td>
<td>66.0</td>
<td>125</td>
<td>78.7</td>
<td>20</td>
<td>50.4</td>
<td>40.3</td>
</tr>
<tr>
<td>rs8066124</td>
<td>G/A</td>
<td>36</td>
<td>72.2</td>
<td>27</td>
<td>66.1</td>
<td>126</td>
<td>85.8</td>
<td>14</td>
<td>58.4</td>
<td>40.0</td>
</tr>
<tr>
<td>rs8073498</td>
<td>A/C</td>
<td>27</td>
<td>70.0</td>
<td>27</td>
<td>66.0</td>
<td>102</td>
<td>85.7</td>
<td>18</td>
<td>59.1</td>
<td>50.3</td>
</tr>
<tr>
<td>rs858521</td>
<td>C/G</td>
<td>26</td>
<td>72.0</td>
<td>20</td>
<td>65.0</td>
<td>87</td>
<td>83.6</td>
<td>20</td>
<td>54.4</td>
<td>42.9</td>
</tr>
</tbody>
</table>

Table 5: Primer, probe, and target Tₘₐₛ (°C) and length (bp) obtained through VOMP
RESULTS

Each LATE-PCR assay for loss of heterozygosity (LOH) measures the fluorescent ratio of a probe to two allelic target sequences at two different temperatures to establish whether or not either of those sequences has been lost in pre-malignant cells. Prior analysis of non-malignant cells from the same individual are used to establish in advance that the chosen target sequence is heterozygous in the intact genome.

Each probe is designed to hybridize to one allele of the SNP at an upper temperature of ~55°C and to both alleles of the SNP at a lower temperature of ~45°C. The fluorescent signal at three temperatures is used for ratio analysis that reveals whether or not LOH has taken place. All three temperatures are below the annealing temperature of the reaction. Thus, fluorescent readings are taken at end-point in the LATE-PCR assays. The upper temperature is chosen to be too high for the probe to hybridize to either the perfectly complementary allele or the partially complementary allele and therefore provides a measure of background fluorescence. The lower temperature is chosen as the temperature at which the probe binds equally to both alleles. The middle temperature is the temperature at which there is the greatest difference in binding of the probe to its perfectly complementary allele and partially complementary allele. The signals at the low and high temperature are then used to normalize the fluorescence at the middle temperature. When compared to the signal from known heterozygous DNA (non-LOH genomes), the normalized signal at the middle temperature reveals whether one or both alleles of the SNP site are present in pre-malignant DNA and thus whether LOH occurred. Boxes encompassing heterozygous samples correspond to three standard deviations and illustrate the 99.7% confidence interval for the fluorescent signals corresponding to the heterozygous
genotype. A non-heterozygous signal from a sample of pre-malignant DNA for a SNP site that was heterozygous in normal DNA indicates LOH has occurred in the vicinity of that SNP site.

In this thesis, LATE-PCR LOH assays were used to analyze BE cell lines with known 17p LOH events to demonstrate that the assay is capable of detecting LOH in Barrett’s DNA. The normal DNA corresponding to BE cell lines was first genotyped for SNP sites that are heterozygous and thus are “informative-SNP sites” for LOH analysis. Informative SNP sites were used for LOH detection in BE cell lines. The optimum SNP assay was then used to construct a duplex with one TP53 amplicon containing a known mutation to provide proof-of-principle of the ability to streamline LATE-PCR LOH analysis with mutational analysis through Dilute’N’Go sequencing on Barrett’s cell line DNA.

**Genotyping normal DNA to determine heterozygous SNP sites**

Before LOH analysis, it is necessary to establish whether an individual is heterozygous for a given SNP site in normal DNA and thus whether that SNP sites is useful for establishing LOH in pre-malignant DNA. Regarding the samples used in this thesis, BE epithelial biopsies were taken from three individual BE patients. The biopsies were enriched for pre-malignant cells, and cells from each patient containing 17p LOH were immortalized to create three separate BE cell lines CP-B, CP-C and CP-D. Normal samples were also obtained from each individual patient, and DNA from these cells was used to establish informative SNP sites in normal DNA corresponding to each cell line.

In genotyping experiments used to establish informative SNP sites, duplicates of the three normal DNA samples corresponding to the three BE cell lines CP-B, CP-C, and CP-D, as well as 20 control DNA samples heterozygous for each of the six above SNP sites (Coriell Institute
for Medical Research, Camden, NJ, USA) were subjected to end-point LATE-PCR. Each sample consisted of 3 ng genomic DNA (500 genomes/equivalent). The control heterozygous samples were used to establish the mean value of the fluorescence ration plus-or-minus three standard deviations. Thus, 99.7% of all heterozygous samples are expected to have fluorescence rations equal to the mean (+/-) three standard deviations. All SNP sites heterozygous in normal DNA have fluorescent ratio values within these boundaries. All homozygous SNP sites having two copies of either allele in normal DNA have fluorescent ratio values that are either greater than or less than the heterozygous mean value +/- three standard deviations (Figure 8).

The fluorescent ratio analysis for the normal DNA corresponding to cell line CP-B demonstrates that SNP sites rs2543540, rs4511593, rs8073498, and rs858521 are all heterozygous in the normal DNA and are thus informative for LOH detection in cell line CP-B. Normal DNA corresponding to cell line CP-C is heterozygous at rs4233018b (Figure 8), rs4511593, rs8073498, and rs858521. Normal DNA corresponding to cell line CP-D is heterozygous at rs8066124 only. The genotype of each SNP site in each normal DNA is presented in Table 6.

**LOH Detection in BE cell line DNA at SNP sites heterozygous in normal DNA**

After determining the informative SNP sites for each cell line, it was possible to analyze each of these SNP sites for the occurrence of LOH. In these experiments, for those SNP sites that proved heterozygous in normal DNA in the respective cell line, duplicate samples from cell lines CP-C and CP-D as well as 20 control DNA samples heterozygous for each SNP site were subjected to the same assays described above to determine the presence of 17p LOH in these cell lines. Cell line CP-B was not analyzed beyond normal DNA because although it contained
informative SNP sites, no BE DNA was available for this cell line. CP-C cell line DNA samples were tested for 17p LOH at the informative SNP sites rs4233018b, rs4511593, rs8073498 and, rs858521. Cell line CP-D DNA was tested for 17p LOH at the informative SNP site rs8066124. Five out of the six SNP assays previously described were utilized on either cell line CP-C or CP-D. The rs2543540 SNP assay was not utilized for LOH analysis because it was not informative in either available cell line. Each LATE-PCR LOH assay that was performed on cell lines CP-C and CP-D produced a fluorescent signal outside the range defined by control heterozygous samples, indicating that each SNP assay was able to detect 17p LOH (Table 7).

Despite the final consensus confirming 17p LOH in both the CP-C and CP-D BE cell lines, unexpected results were first produced from LOH analysis of cell line CP-C using the original rs4233018b HEX probe. Each LOH assay for informative SNPs rs4511593, rs8073498 and rs858521 demonstrated that cell line CP-C underwent 17p LOH. Contrary to this result, one rs4233018b replicate produced a heterozygous signal for the 17p allele. To show that the result of the original rs4233018b LOH analysis was not a reflection of the true genotype of the samples but instead was due to improper probe binding, the same assay was repeated with a modified probe. The rs4233018b probe originally contained a 5’HEX and 3’BHQ-1 and was redesigned to contain 5’BHQ-1 and 3’FAM, the format of all the other SNP probes used in the context of this thesis. Repeating LOH analysis with the redesigned probe produced signals that fell outside the heterozygous range defined by control samples (Figure 9), demonstrating that the original results may have been due to degradation of the probe. Thus, LATE-PCR LOH assays for SNP sites rs4233018b, rs4511593, rs8066124 rs8073498 and rs858521 were all able to detect known LOH events in Barrett’s cell line DNA.
Construction of duplex containing rs4233018b SNP amplicon and TP53 7/8 amplicon

If one copy of the TP53 gene is lost within a cell, it is important to establish whether the sequence of the remaining copy of the gene is normal or has already mutated, thereby eliminating both functional copies of this vital gene. 95% of the point mutations that inactivate the normal function of the TP53 gene in human cancers are found in exons 5-8. Thus, after confirming the ability of the five indicated LATE-PCR SNP assays to detect I.OH in the Barrett’s cell lines, the primers for one of these SNP assays was duplexed with a pair of primers which amplified exons 7/8 within the TP53 gene itself. The rs4233018b SNP assay with a modified probe was chosen for this proof-of-principle assay for the following reasons: (a) the SNP site was informative for cell line CP-C, which possessed three out of six possible informative SNP sites. Cell line CP-D, in contrast, contained just one of the six possible informative SNP sites; (b) the rs4233018b probe was highly reliable. Exons 5-8 of the TP53 gene are 1099 bp long and are thus too long to amplify with one set of primers. Jesse Salk, working in the Wangh Laboratory in 2004, designed two pairs of LATE-PCR primers to amplify exons 5/6 and 7/8 separately. Each of these amplicons was detected using its own double stranded probe (dsP). For the experiments described here the fluorophores and quenchers on the dpPs were slightly modified (see Materials and Methods). Successful amplification of TP53 amplicons using previously designed primers and identical PCR conditions was confirmed in real-time analysis with the modified dsPs, as well as by post-PCR Dilute’N’Go sequencing of the same samples.

Once the two monoplex TP53 assays were shown to successfully amplify the TP53 amplicons, construction of a duplex of the TP53 amplicon 7/8 with the rs4233018b SNP assay was undertaken. The TP53 amplicon 7/8 was chosen for this proof-of-principle experiment
because cell line CP-C is known to contain a single Arg to Trp amino acid change at position 248 [28], which is located in exon 7 of the TP53 gene (Figure 10). The rs4233018b FAM probe was used to detect the SNP allele after PCR. Dilute\textsuperscript{\textregistered}N\textsuperscript{\textregistered}Go sequencing was used to confirm amplification of exons 7/8 because the dsP for this amplicon also contained a FAM fluorophore and could not be used for detection with another FAM probe. PCR thermocycler conditions were initially kept almost identical to TP53 monoplex conditions (indicated in Materials and methods), with the elimination of the 45°C drop-down reading step. The initial attempt to duplex the rs4233018b SNP amplicon and the TP53 7/8 amplicon in a single PCR reaction did not produce successful amplification of the SNP allele (Figure 11). The absence of abundant single-stranded rs4233018b SNP amplicon DNA was indicated by the binding profile of the rs4233018b probe, which was identical to that of the no template controls (NTC). Poor amplification was not completely unexpected because the presence of multiple primers in a single PCR reaction increases the likelihood of mis-priming events, even in the presence of PrimeSafe (Smiths Detection, Edgewood, MD, USA). Mis-priming interferes with DNA amplification because the primers do not fully participate in the reaction.

The occurrence of mis-priming decreases as annealing temperature increases. Thus, in order to reduce mis-priming in the beginning of PCR, annealing temperature was increased from 64°C to 66°C for the first 20 amplification cycles. Once a sufficient number of amplicons had been generated to ensure specificity of primer binding, the annealing temperature was decreased to 60°C for the remaining 40 amplification cycles to increase primer efficiency. Four replicates of both the rs4233018b monoplex and the rs4233018b duplex with the 7/8 amplicon and four NTCs in both the monoplex and duplex were subjected to LATE-PCR in the BioRad thermocycler. In this experiment, the duplex had similar amplification as the monoplex as
indicated by its probe binding profile (Figure 12). Also, the TP53 amplicon generated by the LATE-PCR duplex was able to be sequenced by Dilute‘N’Go sequencing. Thus, by modifying annealing temperature, the LATE-PCR duplex assay was able to amplify both the rs4233018b SNP amplicon and the TP53 7/8 amplicon in a single PCR reaction on normal DNA corresponding to cell line CP-C.

**LOH detection in Barrett’s DNA by rs4233018b duplex SNP assay**

Once it was shown that both the SNP allele and TP53 amplicon could be successfully amplified in a LATE-PCR duplex, the same assay was used on BE cell line CP-C for LOH detection and subsequent Dilute‘N’Go sequencing of the TP53 amplicon in each sample. Twelve known heterozygous control replicates, four NTCs, and duplicates of both normal DNA and cell line CP-C Barrett’s DNA were amplified both in a monoplex with the rs4233018b SNP primers alone and in a duplex with the TP53 7/8 primers. Both the monoplex and the duplex successfully amplified the rs4233018b SNP amplicon as indicated by the probe binding profile. Furthermore, the duplex assay was able to detect the LOH event that occurred in CP-C Barrett’s DNA (Figure 13). In both the monoplex and duplex samples, the signal produced by BE DNA fell outside the box defined by heterozygous controls, while the normal DNA fell inside this box.

**Mutational analysis of TP53 exons 7/8 by Dilute‘N’Go Sequencing after LATE-PCR duplex**

The duplex samples on normal DNA and cell line CP-C DNA from the experiment discussed in the previous section were subjected to Dilute‘N’Go sequencing. TP53 7/8 amplicon from CP-C DNA was successfully amplified in both the monoplex and the duplex. Furthermore, the comparison of the sequence of normal DNA (Figure 14) and BE cell line DNA (Figure 15)
revealed a single G to A nucleotide change in codon 264 of the TP53 gene in BE cell line DNA. A CCG to CCA point mutation in codon 264 leads to an Arg to Trp amino acid change. This amino acid change is in agreement with the previous characterization of the CP-C cell line and demonstrates that Dilute’N’Go sequencing performed after amplification in a duplex is useful for mutational analysis of Barrett’s DNA. Thus, this proof-of-principle duplex LATE-PCR assay demonstrated that it is possible in a single reaction to detect both LOH of one copy of a gene and mutation within the remaining copy of the gene.
### Tables and Figures

<table>
<thead>
<tr>
<th>CP-B</th>
<th>CP-C</th>
<th>CP-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2543540</td>
<td>rs4233018b</td>
<td>rs4511593</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Table 7:** SNP sites heterozygous in normal DNA corresponding to indicated cell lines. Heterozygous SNP sites are indicated with + symbol and homozygous SNP sites with – symbol.

<table>
<thead>
<tr>
<th>rs2543540 (G/A)</th>
<th>rs4233018b (G/A)</th>
<th>rs4511593 (T/C)</th>
<th>rs8066124 (G/A)</th>
<th>rs8073498 (T/C)</th>
<th>rs858521 (G/C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-C</td>
<td>G/-</td>
<td>T/-</td>
<td>–/A</td>
<td>T/-</td>
<td>–/C</td>
</tr>
<tr>
<td>CP-D</td>
<td>–/-</td>
<td>–/-</td>
<td>–/A</td>
<td>–/A</td>
<td>–/A</td>
</tr>
</tbody>
</table>

**Table 8:** Genotype of BE cells lines at informative SNP sites. The cell lines underwent LOH at each informative SNP site. The symbol —— indicates that no LOH analysis was performed at the given SNP site in the indicated cell line because the SNP site was not informative in normal DNA. Cell line CP-B is not included in the table because BE DNA was not available LOH analysis of this cell line.
Figure 8: Genotyping at rs4233018b using the modified 3'FAM probe shows that normal DNA corresponding to cell line CP-C is heterozygous for rs4233018b and thus informative for LOH analysis. The SNP site is homozygous in normal DNA of cell lines CP-B and CP-D and is thus not informative for LOH analysis in these cell lines.

Figure 9: LOH analysis at rs4233018b shows that 17p LOH occurred in cell line CP-C because the signal falls outside the range defined by heterozygous control samples.
**Figure 10:** Normal DNA sequence and amino acid sequence of TP53 exons 5-8 [1]. Barrett’s cell line CP-C contains a single amino acid change at position 248 (location indicated in the yellow box) (Adapted from HapMag.org).
Figure 11: First derivative data of rs4233018b FAM probe fluorescence in monoplex of rs4233018b SNP assay (red) and duplex of rs4233018b SNP assay and TP53 7/8 amplicon (green), annealing temperature 64°C. The probe binding profile of the duplex indicates that it failed to amplify rs4233018b SNP amplicon because fluorescence values matches those of the NTCs (black and gray).
Figure 12: First derivative data of rs4233018b FAM probe fluorescence in monoplex of rs4233018b SNP assay (red) and duplex of rs4233018b SNP assay and TP53 7/8 amplicon (green), annealing temperature raised to 66°C for first 20 cycles and dropped to 60°C for remaining 40 cycles. Duplex amplification of DNA known to be heterozygous for rs4233018b (green) was successful with modified thermocycler conditions, as indicating probe binding profile that matches that of the monoplex SNP amplicon (red).
Figure 13: Normalized fluorescence signals at middle temperature of monoplex (left column) and duplex (right column). Known heterozygous DNA (neon green and dark green) were used to define boxes containing 99.7% of all heterozygous samples. Normal DNA (pink and red) and cell line CP-C Barrett's DNA (light blue and dark blue) were genotyped in this experiment. Normal DNA was correctly genotyped as heterozygous. LOH was detected in CP-C DNA by both the monoplex and duplex, as indicated by the generation of fluorescent signals that fall outside the box defining heterozygous samples.
Figure 14: Sequence of TP53 amplicon 7/8 from normal DNA generated by Dilute 'N' Go sequencing after LATE-PCR.
Figure 15: Sequence of TP53 amplicon 7/8 in cell line C-C generated by Dilute’N’Go sequencing after LATE-PCR. The assay was able to detect a known G to A point mutation in codon 264, located at position 515 of this sequence (pink box).
DISCUSSION

Cancer is essentially a disease of the genome where one alteration can lead to the eventual formation of a tumor. Early detection is critical for reducing deaths from cancer, the second greatest cause of mortality in the United States. Therefore, detecting changes on the molecular level is helpful for assessing cancer risk. One mechanism of tumor formation is the inactivation of tumor suppressor genes (TSGs), which normally prevent excess cell growth. The most commonly altered TSG in human cancers is the TP53 gene, which is inactivated by loss of heterozygosity (LOH), point mutations, and small deletions. Current methods for analyzing these genetic changes require highly trained technicians capable of carrying out multi-step, time-consuming protocols on several pieces of expensive equipment. It is therefore not surprising that biopsy samples of Barrett's Esophagus and many other cancers are currently only analyzed for LOH and gene mutations at a few research-oriented hospitals in the United States.

This thesis provides one step forward toward the construction of a convenient, affordable diagnostic method based on LATE-PCR that is suitable for widespread clinical use. The method is currently being developed for detection of LOH and gene mutation in samples of Barrett's Esophagus (BE), the pre-malignant condition to esophageal adenocarcinoma (EA). The incidence of EA has increased 300-350% in the past three decades, and the cancer has a 0.5-1.0% 5-year survival rate unless it is detected early. BE is appropriate for assay development because rigorous longitudinal studies carried out by Brian Reid and his colleagues at the Fred Hutchinson Cancer Institute have already established that 17p {TP53} LOH is the most predictive risk factor of EA development [7]. BE is also convenient for studying mutational analysis because 95% of TP53 point mutations found in human cancers are located within 1 Kb of the 19 Kb gene. The complete LATE-PCR assay for molecular analysis of BE will also detect

53
LOH of CDKN2a in chromosome 9p and will thereby facilitate early detection and management of BE patients at risk for developing cancer. In addition, the complete LATE-PCR assay will be combined with Dilute‘N’Go sequencing to quickly and conveniently perform mutational analysis, making this assay applicable to a wide range of human cancers. It is anticipated that all of the molecular information described above will be generated in a single tube. The combined assay will involve construction of a multiplex LATE-PCR assay. Proof-of-principle LATE-PCR multiplex SNP assays containing up to 15 pairs of primers have been described [26].

**Proof-of-principle of LATE-PCR LOH detection combined with mutational analysis**

This thesis demonstrates that it is possible to construct a duplex LATE-PCR for detection of LOH for one copy of a gene and mutation within the remaining copy of the gene. One amplicon for a heterozygous SNP site adjacent to TP53 is analyzed by fluorescent probe-target end-point analysis to determine whether LOH has occurred in the DNA of the pre-malignant cells. A second amplicon to a portion of the TP53 gene itself is generated in the same reaction, but need not be detected directly with a fluorescent probe. This “silent” amplicon is analyzed via Dilute-‘N’-Go sequencing, once the amplification reaction and end-point analysis are complete. This approach is doubly cost effective because it requires fewer amplification reactions with fewer steps per reaction, and because it guarantees that only those samples exhibiting LOH are processed for sequencing.

The duplex assay described here combined the rs4233018b SNP assay for 17p LOH detection with a TP53 amplicon containing exon7 and exon 8. Before creating the duplex, five monoplex SNP assays were used to detect LOH events in BE epithelial cell lines. Before LOH analysis was performed, the SNP assays were used to locate heterozygous SNP sites in normal
DNA corresponding to BE cell lines that were informative for LOH analysis. After genotyping the cell lines with monoplex assays, it was then possible to test the ability of the LATE-PCR duplex assay to detect those same LOH events after its construction. The rs4233018b SNP assay, which was informative for cell line CP-C, was readily duplexed with the primers for amplicon 7/8 of the TP53 gene. This duplex was capable of correctly genotyping normal DNA, as well as detecting LOH in Barrett’s cell line CP-C. Then the duplex assay was successfully used in conjunction with Dilute’N’Go dideoxy-sequencing to detect a single known G to A point mutation in exon 7 of the TP53 gene in the CP-C cell line. Based on these results, this thesis contributes to construction of a clinically useful multiplexed LATE-PCR assay for detection of LOH followed by immediate analysis of genetic mutations in the remaining copy of the gene.

**The importance of multiplexing for clinical application of LATE-PCR assays**

Multiplexing LATE-PCR assays is critical for the clinical application of this technology. As previously stated, two heterozygous SNP sites per patient are necessary to reliably diagnose LOH in biopsy material from that patient. A panel of 20 SNP sites around the TP53 gene allows the establishment of at least two heterozygous SNP sites in 98% of patients. Since a large number of SNP sites must be assessed in each patient, a multiplex of the full panel of SNP sites greatly facilitates this preliminary step. After determining informative SNP sites in normal DNA, the biopsy samples from that patient would be analyzed for TP53 LOH utilizing a multiplex of the two informative LATE-PCR SNP assays combined with the amplification of the two TP53 amplicons containing the critical exons 5-8. The biopsy material would first be analyzed by the LATE-PCR LOH assay while producing enough single-stranded DNA to easily
sequence the same DNA sample for TP53 point mutations through Dilute'N'Go dideoxysequencing.

An essential component to multiplexing LATE-PCR SNP assays is suppression of mispriming in the presence of several pairs of primers and amplicons. The PCR reagent PrimeSafe can be used for this purpose. Preliminary experiments show the feasibility of multiplexing up to 15 SNP assays utilizing PrimeSafe [26], but the concentrations of the reagent have yet to be optimized. Attempts to multiplex two SNP assays for this thesis suggest that it is necessary to obtain an intermediate condition between optimal PrimeSafe concentrations of two individual SNP assays to successfully amplify both strands (data not shown), a process which requires several careful titrations. Given the time constraint of this thesis, the duplex presented here was optimized by modifying annealing temperature rather than PrimeSafe concentrations. Future experiments will involve titrating PrimeSafe concentrations for this duplex. Ultimately, to simplify multiplex assay construction, it may be possible to replace current LATE-PCR primers with primers that contain 5’ universal tails, which incorporate into the amplicon sequence in the initial stages of PCR and can then be replaced with the universal primer. Twenty-seven genetic loci have been simultaneously amplified in this way [30].

Applying LATE-PCR multiplexes to heterogeneous DNA samples

Although the duplex described in this thesis demonstrates the utility of LATE-PCR for LOH and mutational analysis, this assay must be further tested under clinically-relevant conditions. This duplex was able to detect loss of 17p in BE cell lines containing 100% LOH. In the clinic, biopsics taken from pre-malignant tissue in BE patients contain mixtures of normal cells and pre-malignant LOH cells. It will therefore be helpful to apply a protocol for
enrichment of pre-malignant cells. Several such protocols are currently under development in the Reid Laboratory at Fred Hutchinson Cancer Institute. It still can be anticipated that clinical samples will likely contain mixtures of normal genomes and LOH genomes, even after being enriched for pre-malignant cells. Thus, if 17p LOH has occurred in the pre-malignant cells, the enriched biopsy sample is still heterogeneous for 17p LOH due to the presence of normal DNA. Proper diagnosis of LOH requires a highly sensitive assay that can detect small allele imbalances of a given TSG, the standard of which is 10% LOH to 90% no LOH (personal communication with Galipeau PC). Preliminary results in the Wangh Laboratory show that the LATE-PCR SNP assay is capable of detecting LOH in mixed samples of 90% normal DNA and 10% BE DNA containing 9p LOH (personal communication with Dr. J. Aquiles Sanchez). The LATE-PCR duplex assay presented in this thesis must be tested under similar conditions to further demonstrate its robustness.

A possible obstacle of utilizing multiplexed LATE-PCR assays on mixed DNA samples is that multiplexed assays produce more variability among replicates than monoplexes, see Figure 13. Therefore, the box defined by heterozygous controls that contains 99.7% of heterozygous samples covers a wider range of values in a multiplex. A concern is that signals produced by mixed samples of pre-malignant DNA containing LOH and normal DNA may fall into this larger box, compromising the ability of the LATE-PCR assay to detect LOH. In order to better distinguish pure heterozygous and mixed or non-heterozygous DNA, the scatter among multiplexed replicates could be reduced by optimizing the concentrations of PrimeSafe. The need for external heterozygous controls to define LOH events is itself an obstacle to clinical application of LATE-PCR assays because it reduces the number of assays that can be performed in one thermocycler. To eliminate the need for external controls and to thus facilitate clinical
application of these assays, the development of internal controls for each SNP assay is currently underway at the Wangh Laboratory.

Another method for detecting infrequent LOH events in heterogeneous samples using the LATE-PCR multiplex assay involves reducing the biopsy sample size analyzed in each assay, thereby increasing the sensitivity of the assay by decreasing complexity of the sample. The LATE-PCR assay can accurately genotype down to the single-cell level. Differences in the amount of starting material do not affect the result of analysis [25]. Therefore, it is possible to take small biopsies and perform multiple analyses on each sample using LATE-PCR. However, reducing sample size requires analysis of a greater number of samples in order to assess the entire field of the esophagus containing pre-malignant tissue, especially because early LOH events at the beginning stages of clonal expansion are found in highly localized regions. Fortunately, increasing the number of samples may not be an obstacle to clinical application because the ease of setting up the LATE-PCR assay makes it suitable for high throughput automation.

**Limitations of the LATE-PCR assay for LOH detection and for mutational analysis**

Limitations of the LATE-PCR SNP assay for detection of allele loss are the following. First, an assumption of utilizing SNP sites for detecting TSG LOH is that loss of a given TSG will result in loss of the investigated SNP site because LOH events cover wide stretches of chromosome. However, this assumption may not be true for all LOH events. Given the density of all SNP sites in the genome and the probability that those SNP sites are heterozygous for a given patient, losses smaller than 50 kb have less than 95% chance of detection [27]. Another limitation of utilizing SNP sites is that a patient may not have two informative SNP sites.
However, the panel of SNP assays for 17p LOH that is currently being developed at the Wangh Laboratory will ensure two informative SNP sites in 98% of patients. Additionally, it is possible that genes besides the alleged TSG that are lost in LOH may be involved in cancer progression. However, in the BE model system for this thesis, wherever there is a 17p LOH event, the remaining copy of the TP53 gene is found to be mutated, which suggests that TP53 is the target of the LOH event [8].

A limitation of the LATE-PCR multiplex assay for detecting inactivation of TSGs in general is that the assay cannot detect LOH by promoter hypermethylation. However, this is not an issue in the current model system because methylation of the TP53 promoter is not found in Barrett's esophagus [7] and furthermore is not common in human tumors [18]. Therefore, an assay for 17p LOH and TP53 mutations is complete in terms of analyzing loss of TP53.

The major limitation of the LATE-PCR assay for sequencing is the error introduced by Taq polymerase. PCR error has been estimated to be 1.18 +/- 0.34 x 10^{-4} mutations/bp [31]. The Taq polymerase does not have repair mechanisms in PCR as DNA polymerases do in vivo, so mistakes remain in the sequence and are reproduced in subsequent PCR cycles that use the mutated strand as a template. Mistakes in early cycles, when there are fewer than 1000 starting copies of the template generated, produce enough of the altered strand so that commercial sequencing of the material yields a sequence different from that of the starting material. A method for overcoming this limitation is to sequence both the limiting primer strand and the excess primer strand, instead of just the excess primer strand as was done in thesis, because the probability of Taq polymerase making the same error in two independent assays is very small (1.4 x 10^{-5}). It is possible to sequence the limiting primer strand through the Dilute'N'Go method even though it is not present as single-stranded DNA. The protocol is modified to dilute the
reaction mixture to a lesser degree in distilled water and uses the excess primer already present in the mixture [26].

**Application of the LATE-PCR multiplexed assay to a broad range of human cancers**

The first application of the LATE-PCR multiplex will be to Barrett’s esophagus, where 17p LOH is a known biomarker. In addition, this assay has potential for application to several other human cancers. LOH has been implicated as a biomarker in cancer of the colon/rectum, cervix (3p, 4p, 4q, 11q), lung (3p, 8p, 9p), skin (non-melanoma), breast and prostate (8p, 9p, 16q). Both LOH and TP53 mutations have been implicated in head and neck cancer (3p and 9p LOH) and bladder cancer (3p, 9p, 11p, 17p, and 18p LOH) [32]. Indeed, 17p LOH and point mutations in TP53 are together the most common alterations in human cancers. The frequency of mutant TP53 alleles in human cancer is illustrated in Figure 16. Although the presence of these alterations in human cancers is known, little is understood about their ability to predict cancer. Deeper understanding of these events is necessary for utilizing these possible biomarkers for early cancer detection and intervention. The LATE-PCR assay for combined LOH detection and mutational analysis has potential to accelerate research into the nature and temporal order of molecular alterations in various cancers. It may also be used to localize common regions of deletions to aid in the discovery of new tumor suppressor genes. This LATE-PCR assay will serve as a powerful tool in both basic science laboratories and clinics to help ameliorate the problem of cancer.
**Figure 16:** Frequency of mutant p53 alleles in human tumor cell genomes (Weinberg, *The Biology of Cancer*, 2007).
REFERENCES


