Detection of Microsatellite Instability by Fluorescence Multiplex Polymerase Chain Reaction

Karin D. Berg,* Cynthia L. Glaser,*
Richard E. Thompson,[†] Stanley R. Hamilton,*[†]
Constance A. Griffin,*[†] and James R. Eshleman*[†]

From the Departments of Pathology,* Oncology,[†] and
Biostatistics,[‡] The Johns Hopkins University School of Medicine,
Baltimore, Maryland

We have created a clinical molecular diagnostic assay to test for microsatellite instability (MSI) at multiple loci simultaneously in paraffin-embedded surgical pathology colon resection specimens. This fluorescent multiplex polymerase chain reaction (PCR) assay analyzes the five primary microsatellite loci recommended at the 1997 National Cancer Institute-sponsored conference on MSI for the identification of MSI or replication errors in colorectal cancer: Bat-25, Bat-26, D2S123, D5S346, and D17S250. Amplicon detection is accomplished by capillary electrophoresis using the ABI 310 Genetic Analyzer. Assay validation compared 18 specimens previously assessed by radioactive PCR and polyacrylamide gel electrophoresis detection to results generated by the reported assay. Germline and tumor DNA samples were amplified in separate multiplex PCR reactions, sized in separate capillary electrophoresis runs, and compared directly to identify novel length alleles in tumor tissue. A concordance of 100% between the two modalities was achieved. The multiplex assay routinely detected a subpopulation of 10% tumor alleles in the presence of 90% normal alleles. A novel statistical model was generated that corroborates the validity of using results generated by analysis of five independent microsatellites to achieve a single overall MSI diagnosis. The assay presented is superior to standard radioactive monoplex PCR, polyacrylamide gel electrophoretic analysis, primarily due to the multiplex PCR format. (*J Mol Diag* 2000, 2:20–28)

Microsatellite instability (MSI), or replication error, is a manifestation of genomic instability arising in a variety of human neoplasms where tumor cells have a decreased overall ability to faithfully replicate DNA. This phenomenon has been shown to be a frequent, if not obligatory, surrogate marker of underlying functional inactivation of one of the human DNA mismatch repair (MMR) genes.^{1–6} Functional loss of a MMR gene occurs due to biallelic inactivation via some combination of coding region mu-

tation, loss of heterozygosity (LOH), and/or promoter methylation.^{7,8} Germline mutation of a MMR gene has been shown to be the autosomal dominant genetic defect in most hereditary nonpolyposis colon cancer (HNPCC) kindreds.^{9,10} A second hit incurred by tumor cells in HNPCC individuals results in biallelic inactivation of the specific MMR gene, causing loss of faithful replication of microsatellite DNA in tumor.¹¹ MSI is thus a marker of an underlying DNA mismatch repair defect and, additionally, is associated with enhanced mutation rates in coding DNA.^{12,13} This mutator phenotype, which results from the MMR defect, causes both coding region base substitutions and frameshift mutations at direct repeats, each occurring at equal frequencies,¹⁴ in addition to resulting in MSI. Generation of MMR defects and the resultant mutator phenotype is thought to be an early event in tumorigenesis¹⁵ and has been suggested to occur as early as the aberrant crypt focus stage.¹⁶

Although implicating a germline defect in HNPCC families, MSI is also found in 15 to 20% of sporadic colorectal cancers,¹⁷ where the finding also reflects an overall increase in genomic instability. Several reports have associated the finding of MSI defects in tumors with a better prognosis in stage-for-stage matched tumors.^{18,19} Thus, it may become important clinically to identify tumors with MSI not only to implicate germline MMR defects (HNPCC families), but also for prognostic stratification. While clinical (Bethesda guidelines²⁰) and histopathological features²¹ may raise the suspicion that a colorectal tumor is microsatellite-unstable and perhaps has arisen in an HNPCC family, clinicopathological features are insufficient to diagnose the presence of MSI; thus, direct molecular testing has importance in documenting the MSI status of a clinically suspicious tumor.²²

Many different microsatellite markers or loci have been used by different investigators to identify MSI in tumors. Controversy has existed among experts as to which and how many loci should be analyzed to diagnose MSI. In an attempt to gain the most information from research trials and to provide uniformity in clinical diagnosis, the Na-

Supported by National Cancer Institute grants K08 CA66628 and R01 CA81439 (to J. R. E.).

Accepted for publication November 1, 1999.

S. Hamilton's current address: M.D. Anderson Cancer Center, Houston, Texas.

Address reprint requests to James R. Eshleman M.D., Ph.D., Johns Hopkins Medical Institutions, Department of Pathology, Division of GI Pathology, Ross Building, Room 632, 720 Rutland Avenue, Baltimore, MD 21205. E-mail: jeshlema@welchlink.welch.jhu.edu.

tional Cancer Institute (NCI) held an international meeting of MSI investigators in 1997 to create more consistent parameters for defining MSI in colorectal cancer (CRC) and to recommend microsatellite markers for use in CRC MSI testing in both clinical and research settings.²² Performing MSI assays using the five primary loci identified at the NCI MSI conference by standard radioactive monoplex polymerase chain reaction (PCR) and polyacrylamide gel electrophoresis (RMo-PAGE) requires a total of 10 separate PCR reactions (1 tumor and 1 normal DNA sample for each case, 5 loci assessed for each DNA sample), in addition to labor-intensive sequencing gel detection.

The logistic difficulty of RMo-PAGE analysis led us to create a fluorescent multiplex PCR-capillary electrophoresis (FM-CE) assay using the five primary loci identified by investigators at the NCI-sponsored consensus conference.²² This format permits simultaneous amplification of all five loci in a single PCR reaction, with subsequent detection of amplicon length alterations using the automated ABI 310 Genetic Analyzer. We have also examined the validity of this overall approach to clinical MSI diagnosis by creating and applying a novel statistical model to the problem. Using this model, we have examined the number of loci and degree of microsatellite informativity required to achieve sufficient MSI diagnostic sensitivity. Here, we report on the development of assay parameters and the theoretical and practical validation of the assay for clinical use.

Materials and Methods

Cells and Human Tissues

Lymphoid normal and tumor matched cell lines were generously provided by Dr. James Willson at Case Western Reserve University (Cleveland, OH). Vaco 670 is a colon cancer cell line known to have MSI due to a defect in the MMR gene, *hMSH2*. L670 is an Epstein-Barr virus-transformed lymphocyte cell line from the same patient as Vaco 670. Tumor and normal tissue blocks for the validation study were obtained from patients' colon resection specimens obtained from the files of Johns Hopkins Hospital Department of Pathology. These specimens had been previously established to possess or lack MSI using the traditional radioactively labeled PCR primer, monoplex PCR, polyacrylamide sequencing gel electrophoresis detection, autoradiography approach²³ (Hamilton SR, manuscript in preparation). The presence or absence of MSI by RMo-PAGE was blinded during FM-CE analysis.

Sectioning and Microdissection

Serial 5- μ m histological sections of formalin-fixed, paraffin-embedded tissue blocks of normal (N) and tumor (T) were prepared using DNA histology precautions. The first and fifth levels were stained with hematoxylin and eosin (H&E); the sandwiched tissue levels 2, 3, and 4 were mounted unstained on slides. Histological diagnoses

were verified by light microscopy for each block using levels 1 and 5, and the areas to be microdissected were marked. Normal or tumor tissue was microdissected from unstained slides for each case by overlaying the unstained slide onto the H&E-stained slide. Dissection of unstained slides was performed in a laminar flow tissue culture hood after UV irradiation.

DNA Isolation

Cell line DNA was isolated by the standard sodium dodecyl sulfate/proteinase K digestion, organic extraction, and ethanol/salt precipitation technique.²⁴ Tumor and normal DNA from microdissected paraffin embedded tissue was crudely isolated using xylene/ethanol deparaffinization followed by Proteinase K digestion and heat inactivation at 95°C for 10 minutes. Isolated genomic DNA was kept in a dedicated room which was used only to assemble PCR. No PCR products or equipment used in post-PCR analysis ever entered this room.

PCR and Microsatellite Analysis

The 16- μ l PCR fluorescent multiplex reaction mix pool for each reaction was composed of the following final constituents: 1 $\times$ AmpliTaq Gold PCR buffer with 1.5 mmol/L MgCl₂ (Perkin Elmer, Foster City, CA), 0.002 mmol/L dNTPs (Perkin Elmer), 0.3 units AmpliTaq Gold DNA polymerase (Perkin Elmer), and primer sets in the final molar amounts listed in Table 1, with phosphoramidite fluorescent labels as indicated. A 9- μ l aliquot of the reaction pool mix was used for each individual PCR reaction, with 1 μ l of normal or tumor sample DNA. Control reactions included a water control for exclusion of contamination and a mix of 90% L670 (N) DNA and 10% Vaco 670 (T) DNA. The 90%/10% cell line mixture served as a combined positive/negative and limit of detection control.

PCR was performed using either a PE 9600 or PE 9700 (Perkin Elmer) thermocycler under the following cycling conditions: initial denaturation 95°C for 9 minutes, followed by 35 cycles of: 94°C for 45 seconds, 55°C for 45 seconds and 72°C for 1 minute, with a final 45 minute, 60°C extension to aid nontemplated adenine addition. Fluorescently labeled PCR products were detected using the ABI 310 Genetic Analyzer and GeneScan Collection software. The following CE run parameters were used: GeneScan Short Tandem Repeat Performance Optimized Polymer (GS STR POP) 4 (1 ml) C module, GS POP 4 polymer, 5- to 10-second injection time at a voltage of 15.0 kV, a 15-kV electrophoresis voltage with a resultant 7- to 9-ampere current at a column temperature of 60°C, and a 20-minute electrophoresis time. GeneScan data analysis parameters were: matrix C virtual filter, 2800–6000 analysis range, baselining, multicomponent data processing, light peak smoothing, a minimum peak detection limit of 50 relative fluorescent units (RFU), size call range of 50 to 350 bp using the local Southern size calling method, Tamara labeled GS 500 size standard, no split peak correction, and a minimum peak half-width of 3 points.

Table 1. Primers and Characteristics of Microsatellite Loci Analyzed

Locus (phosphoramidite label): Primer sequence	GDB/Genbank no.	Product size	Molar concentration	Informativity
D2S123 (Tet) Forward: 5'-AAACAGGATGCCTGCCTTTA-3' Reverse: 5'-GGACTTTCCACCTATGGGAC-3'	187953	197-227	4.0 pmol/L 56.0 pmol/L	0.67
D17S250 (Fam) Forward: 5'-GGAAGAATCAAAATAGACAAT-3' Reverse: 5'-GCTGGCCATATATATTTAAACC-3'	177030 X54562	151-169	8.0 pmol/L 112 pmol/L	0.56
D5S346 (Fam) Forward: 5'-ACTCACTCTAGTGATAAATCGGG-3' Reverse: 5'-AGCAGATAAGACAGTATTACTAGTT-3'	181171 M73547	96-122	0.80 pmol/L 11.2 pmol/L	0.67
Bat-25 (Hex) Forward: 5'-TCGCCTCCAAGAATGTAAGT-3' Reverse: 5'-TCTGCATTTAACTATGGCTC-3'	9834508 U63834	120	2.0 pmol/L 28.0 pmol/L	0.73
Bat-26 (Tet) Forward: 5'-TGACTACTTTTGACTTCAGCC-3' Reverse: 5'-AACCATTCAACATTTTAAACCC-3'	9834505 L47575	116	3.06 pmol/L 42.3 pmol/L	1.00

For the radioactive monoplex PCR/PAGE assay, the loci analyzed were D18S55, D18S58, D18S61, D18S64, and D18S69. Forward and reverse primers (Research Genetics, Huntsville, AL) were used at a 1:1 molar ratio after end-labeling of the forward primer with γ - 32 P-dATP and T4 polynucleotide kinase (New England Biolabs, Beverly, MA). Amplicon detection used standard 7% polyacrylamide sequencing gel electrophoresis and autoradiography as previously described.²⁵ The size standard was generated using Phi-X 174 digested with *Hinf*I (Gibco/BRL, Rockville, MD), α - 32 P-dATP, and Klenow (Gibco/BRL) per manufacturer's instructions.

Statistical Model for Interpretation of Multiple Independent Assays

A statistical model was constructed to analyze the diagnostic situation in which multiple independent assays are combined to establish an overall diagnosis, in this case, MSI-High, MSI-Low, or MSS. The following equations were applied to evaluate a population of MMR-defective tumor cells: MSS = $(1 - I)^n$, MSI-Low = $I \times (1 - I)^{(n-1)} \times n$, and MSI-High = $1.00 - (MSS + \text{MSI-Low})$, where I is the constant informativity of each locus and n is the number of loci analyzed. These equations represent a special case of the more generalized binomial distribution equation:

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}.$$

Assay Interpretation

Identification of normal or tumor allele amplicon sizes was accomplished by examining the appropriate electropherogram (N or T) and determining the predominant amplicon size(s) for each allele at each locus (greatest peak height or heights). More than two peaks can be seen at each allele in tumor samples if a shift is present. Each true microsatellite peak should be surrounded by stutter peaks of lower intensity with the appropriate delta (1 base for mononucleotide repeats, 2 bases for dinucleotide repeats) from the dominant peak. Stutter peak sizes

can be larger or smaller than the predominant peak, but are more frequently smaller. Those peaks without stutter are considered to be nonspecific amplicons and are disregarded. Stutter peaks are thus considered a signature of amplification of a repetitive element lying within a forward and reverse primer set. Stutter peaks from a shifted dinucleotide locus allele can overlap with the predominant amplicon peak height of the second allele at that locus, causing difficulty in separating the stutter pattern of the first allele from the true predominant amplicon sizing of the second allele. The mononucleotide repeat loci, Bat-25 and Bat-26, are also subject to stutter and, as recently reported for Bat-26²⁶ and observed by the authors for Bat-25 (unpublished data), may also have a low occurrence of germline allele polymorphisms.

The amplicon length(s) for each locus are recorded, and a direct comparison of the amplicon lengths between normal and tumor is made. If the tumor specimen has novel amplicon lengths in relation to the patient's germline (normal) amplicon length(s) at a specific locus, this is considered a locus positive for MSI. Despite microdissection, some small degree of stromal tissue contamination of the tumor sample is essentially unavoidable; thus, both tumor and germline amplicons may (in fact, should) be present in the tumor specimen. Sufficient amplification needs to occur at each locus within the tumor sample to ensure that low level instability would be detected if present.

The parameters used for the diagnosis of MSI using the loci recommended by the NCI have been previously described.²² In brief, MSI-High (MSI-H) requires shifting of two of the five NCI recommended loci (≥ 30 to 40% of loci tested if more than five loci are analyzed). MSI-Low (MSI-L) is identified by shifting of one of the five NCI-recommended loci (≤ 30 to 40% of loci tested if more than five loci are analyzed). The diagnosis of microsatellite stable (MSS) requires that none of the loci analyzed be shifted.

Quality Control

Accurate data interpretation requires meeting specific run criteria to ensure validity. The water control must be

free of extraneous peaks. The 90% normal/10% tumor positive control must have detectable normal allele amplicons for all five loci in the peak height range of 2000 to 8000 RFU, and the tumor alleles should be shifted relative to germline in all five loci and have detectable peak heights in the range of 200-2000 RFU.

Results

Optimal Molar Ratios of Multiplexed PCR Primers Permit Simultaneous Amplification of All Five Recommended Primary Microsatellite Loci

Assay optimization involved identification of parameters important to the maximization of diagnostic accuracy, sensitivity, informativeness, and precision. Fluorescent labels were chosen such that the potential amplicons for each locus were sufficiently separated in size to prevent any potential overlap between PCR products (shifted or germline amplicons). Changes in fluorochrome assignments were made during optimization to enhance amplicon detection. For example, D2S123 was initially labeled in yellow (Hex), but was later moved to the green fluorochrome (Tet) to enhance its relative detectability.

Initial analysis verified that all five loci were amplified and detectable by fluorescent monoplex PCR and CE detection. The specificity of amplification was optimized by varying forward to reverse primer concentrations (presented below). After maximization of amplification specificity, the primer pair for Bat-26 was systematically mixed with additional loci, beginning with duplex PCR reactions and proceeding to full multiplex of all five loci. Finally, the relative molar amounts of the various locus primer sets were adjusted to produce simultaneous amplification where peak heights of the predominant amplicons varied by no more than threefold for all loci (Table 1). Electropherogram data generated by the described assay, as applied to the control cell lines, can be seen in Figure 1. In Figure 1A, we compare the radioactive format (left panel) to a fluorescent monoplex format (right panel), using normal and tumor cell lines assayed at the Bat-25 locus. In Figure 1B, the full multiplex electropherogram data are shown as applied to cell line samples Vaco670 and L670. Figure 1C shows the full multiplex assay result on tissue samples normal and tumor from validation sample 10. Each of the five loci produced amplicons within the appropriate size range, surrounded by stutter peaks of appropriate size, diminishing in intensity with increasing distance from the predominant peak (see Materials and Methods). Repeat analyses of the same control and validation specimens produced consistent amplicon sizing.

A Cross-Validation Study Establishes the Validity of the Fluorescent Multiplex Approach

Eighteen cases of colon cancer previously characterized by RMo-PAGE²³ (SR Hamilton, submitted) for the presence (14 cases) or absence (4 cases) of MSI were analyzed using the technique reported herein. Of the 14

cases diagnosed as MSI by RMo-PAGE, all 14 were diagnosed as MSI-H by FM-CE. Of the 4 cases shown to be MSS by RMo-PAGE, all four were MSS by FM-CE. Thus, of the 18 cases analyzed, 100% showed overall diagnostic concordance between RMo-PAGE analysis and the FM-CE method ($P = 0.0003$, Fisher's exact test, Table 2).

Varying the Forward Labeled to Reverse Unlabeled Molar Ratios Increases the Specificity of Amplification without Significant Loss of Amplicon Signal

Serial titration of labeled forward primer to unlabeled reverse primer was performed to determine whether non-specific amplification could be eliminated while maintaining specific amplification (Parsons R, personal communication). Serial dilutions revealed that a forward to reverse primer dilution of 1:15 was optimal in reducing nonspecific amplicons while maintaining adequate specific amplicon signal (Figure 2). At a forward to reverse primer ratio of 1:1 (Figure 2A), amplification was suboptimal due to detection of multiple nonspecific products, whereas at a forward to reverse ratio of 1:100 (Figure 2D) there was not enough forward primer present to allow generation of detectable PCR product. The ratio of 1:15 forward labeled to reverse primer was chosen as this primer ratio yielded the greatest specific amplicon product without loss of signal.

The Limit of Detection of FM-CE Is 2 to 10%

The limit of detection of tumor DNA mixed with normal DNA was tested. Tumor cell line DNA (Vaco 670) was serially diluted into germline DNA (L670). Using the criteria that amplicons from all five normal and tumor loci must be identified in the mixed sample, the consistent lower limit of detection for this assay is 10% tumor (shifted) alleles in 90% germline (unshifted) alleles (Figure 3). The lower limit of detection for the multiplex assay is improved to 2 to 5% tumor detectable mixed into normal, when fewer than five loci are required to be detected. Sensitivity can be further enhanced as clinically indicated by using monoplex fluorescent PCR reactions.

Statistical Modeling of Multicomponent Testing

We have developed a statistical model to examine some of the interpretive considerations when multiple independent assays are combined to produce a single overall assay result (see Materials and Methods). This model is applicable to the problem of MSI testing where the results of instability assessment at several microsatellite loci are pooled together to establish an overall categorization (MSI-High, MSI-Low, or MSS) using the standard definitions.²² We assumed that 100 tumors were analyzed,

Figure 1. A: Comparison of results generated by analysis of the Bat-25 locus in cell lines L670 (normal) and Vaco 670 (tumor) using the radioactive PAGE format (left panel, autoradiograph) and the fluorescent monoplex format (right panel, electropherograms). Arrowheads indicate the predominant amplicon size bands (peaks) at 121 bases for Normal (L670), and arrows indicate the predominant amplicon size bands (peaks) at 112 bases for Tumor (Vaco 670). Note the stutter bands surrounding the predominant peak and the absence of germline allele peaks in the tumor cell line. The asterisk to the right of the autoradiograph is the Phi-X174 *Hinf*I-digested band at 118 bases. **B:** The complete fluorescent multiplex electropherograms of cell lines L670 (normal) and Vaco 670 (tumor) are shown. The red tracing is the GS 500 size standard labeled with Tamara (100, 139, 150, 160, 200, and approximately 245 bases). In the top panel, normal germline alleles are seen where the alleles from one locus (D2S123) are indicated (219 and 225 bases) labeled with Tet (green). The blue tracing represents the D5S346 (2 alleles, 100 and 107 bases) and D17S250 (1 allele, 147 bases) loci labeled with 6-Fam, the green tracing represents Bat-26 (1 allele, 115 bases) labeled with Tet, and the black tracing Bat-25 (1 allele, 121 bases) labeled with Hex. The bottom panel is the fluorescent multiplex electropherogram for the matched tumor cell line Vaco 670. Again designated are the shifted alleles at D2S123 (211 and 213 bases). There are also shifts at all of the other four loci: D5S346 (102 and 109 bases), D17S250 (139 and 147 bases), Bat-26 (103 bases), and Bat-25 (112 bases). **C:** The complete fluorescent multiplex electropherograms of a representative normal tumor pair. The alleles for D2S123 are designated where the patient's normal sample reveals homozygosity at 208 bases. The tumor shows the germline allele in addition to novel length alleles at 198 with its accompanying stutter bands. Novel length alleles can be seen at each of the other four loci.

each of which has a defect which should result in MSI at all loci tested. The concept of informativity was created. We define informativity as the frequency that a given marker is unstable divided by the total number of MSI cases, and is therefore essentially the diagnostic sensitivity of a given microsatellite marker to detect the MSI phenotype when present. The model then assumes that the informativity of each locus is less than 100%, but that all loci are equally informative, and that each locus is an independent reporter of the phenotype. We initially set the informativity of each locus to 70% (a reasonable assumption based on our data of informativities ranging between 56 and 100% with an average of 73%; see Table 1). The number of loci analyzed varied from 3 to 7, as shown in Figure 4A. If only 3 loci are analyzed, the probability of misclassifying a tumor is almost 25%. They would be misclassified as either MSI-Low (19%) or MSS

(2.7%). When five loci are analyzed, 97% of tumors are correctly diagnosed as MSI-High and only 3% are misclassified as MSI-Low. Next, we kept the number of loci constant at five and varied the informativity of the microsatellites from 50 to 100% (Figure 4B). This has a fairly large influence, where an unacceptable number of tumors are misdiagnosed when the informativity of the loci drops to 50% (3% of cases erroneously called MSS, and 16% called MSI-Low) or an informativity of 60% (8% of cases called MSI-Low). Seventy percent informativity is probably adequate where the percentage of cases misdiagnosed as MSI-Low drops to only 3%. Thus, despite some oversimplification, the model provides a means for examining quantitatively the number of loci that should be analyzed, given defined informativity parameters, to provide adequate sensitivity and specificity for clinical testing.

Table 2. Comparison of Radioactive *versus* Fluorescence Assays

Sample number	RMo-PAGE result	FM-CE result
1N, 1T	MSI-H	MSI-H
2N, 2T	MSI-H	MSI-H
3N, 3T	MSI-H	MSI-H
4N, 4T	MSI-H	MSI-H
5N, 5T	MSS	MSS
7N, 7T	MSI-H	MSI-H
8N, 8T	MSI-H	MSI-H
9N, 9T	MSS	MSS
10N, 10T	MSI-H	MSI-H
11N, 11T	MSI-H	MSI-H
12N, 12T	MSI-H	MSI-H
13N, 13T	MSI-H	MSI-H
14N, 14T	MSI-H	MSI-H
15N, 15T	MSI-H	MSI-H
16N, 16T	MSI-H	MSI-H
17N, 17T	MSI-H	MSI-H
18N, 18T	MSS	MSS
19N, 19T	MSS	MSS
Total MSI-H	14	14
Total MSI-L	0	0
Total MSS	4	4

Figure 3. Determination of the limit of detection of the assay as demonstrated by detection of 10% Vaco 670 (tumor) DNA mixed into 90% L670 (normal) DNA at the Bat-25 locus. The predominant normal (germline) allele is indicated by the **bold arrowhead**, while the shifted (tumor) allele is indicated by the **bold arrow**. The Tamara labeled size standard is indicated by **broken arrows**, sizes as in Figure 1.

Discussion

The reported FM-CE assay is an advance in the field of MSI analysis primarily due to the fivefold reduction in cost and time required for its performance, in comparison with traditional formats. This multiplex PCR assay permits significant technical simplification, in conjunction with automated analysis, allowing relatively high-throughput analysis of tumor MSI status, limited primarily by the need for microdissection. CE technology also clearly provides ad-

A. 1:1 forward to reverse primer ratio

B. 1:15 forward to reverse primer ratio

C. 1:40 Forward to reverse primer ratio.

D. 1:100 Forward to reverse primer ratio.

Figure 2. Titration of labeled forward Bat-26 primer into the reverse, unlabeled primer. **A:** The 1:1 ratio of labeled forward primer to reverse primer produces a significant number of false amplification products indicated by **arrows**, and appropriate size amplicon (115 bases); **arrowhead**. **B:** The 1:15 forward to reverse primer dilution gives the best overall specific product to false amplicon ratio. Here, minimal background contaminant peaks are identified (**arrows**), and the peak height of the specific 115 base product (**arrowhead**) is still robust. **C:** The 1:40 primer dilution is entirely adequate for the Bat-26 locus, with essentially no false product detected, and specific amplicon is well within the linear range of the instrument. Other loci with less robust amplification cannot be visualized well (insufficient amplification) at this level of forward primer dilution (data not shown). **D:** The 1:100 labeled forward primer dilution produces no detectable amplicons. The other peaks represent internal size standard as above.

Figure 4. Statistical modeling of multiple independent tests to establish a diagnosis. **A:** Varying the number of loci analyzed determines the accuracy of MSI test results. The informativity of each of the microsatellites is kept constant at 70%, while the number of loci analyzed is varied from 3 to 7. **Dotted bars,** MSI-High; **striped bars,** MSI-Low; **solid bars,** MSS. The justification for a constant informativity is based on the overlapping 95% confidence intervals for the informativity of each individual locus (Bat 25: mean 0.73, CI = 0.45–0.92; Bat 26: mean 1.00, CI = 0.78–1.00; D2S123: mean 0.67, CI = 0.30–0.93; D5S346: mean 0.67, CI = 0.38–0.88; D17S250: mean 0.56, CI = 0.21–0.86). **B:** Varying the informativity determines the accuracy of MSI test results. The number of microsatellite loci is kept constant at five, while the informativity of each of the microsatellite markers analyzed is varied from 50 to 100%. Bar designations are as in **A**.

vantages in both analysis and management of data. Among these is the ability to generate quantitative information about the relative amounts of amplicons present in a specimen based on peak height data. This ability allows for more accurate assessment of phenomena such as stutter, or allele shift patterns,¹⁸ and may aid in study of the patterns of MSI phenotypes and their potential correlation with different underlying MMR defects, such as seen in *GTBP/hMSH6* altered tumors²⁷ (see below). The limit of detection of the assay can theoretically be manipulated as a function of amplicon loading adjustments and altered run parameters. From a practical standpoint, data storage, retrieval, and processing is entirely electronic, a significant improvement over data storage using autoradiographs. The reported assay should facilitate both clinical testing and accrual of research data by increasing assay throughput and permitting data analysis on a large scale. The clinical utility of such an assay is well recognized for screening of suspicious tumors to initiate the process of identifying HNPCC families.^{20,22} The assay may also be useful as a prognostic marker^{18,19} or potentially in the elucidation of the pharmacogenetic properties of a tumor (SR Hamilton, submitted).

The limit of detection of this multiplex assay for low numbers of alleles shifted due to MSI is comparable to, although somewhat less than, the lower limits of detection previously reported for RMo-PAGE.²⁸ Our multiplex assay, which has a limit of detection of 2 to 10%, has proven adequate to detect allele shifts in all of the microsatellite-unstable cases in our CRC validation study. It would be an uncommon clinical situation in which microdissection

could not enhance the tumor cell content of a specimen sufficiently to permit identification of microsatellite shifts at the limit of detection we have attained. In addition, in the event that increased sensitivity is required, any of the loci tested by the multiplex assay can be separated into a monoplex analysis, lowering the potential limit of detection to 1 to 5 cells in 100.

The statistical model presented is useful in understanding several facets of MSI testing, which can be extrapolated to other clinical testing situations. The clinical validation data presented confirm that FM-CE compares favorably with RMo-PAGE analysis; we appropriately diagnosed no cases as MSI-Low. When cases are interpreted as MSI-L (1/5 shifted loci) the question remains as to how such lesions should be classified clinically. Several theories have been proposed to explain the MSI-L diagnosis.

It is possible that the shifting phenotype seen in MSI-L reflects spontaneous background mutations, which occur at increased rates in microsatellites¹⁵ in comparison with coding regions.²⁹ It is highly unlikely that two microsatellite regions tested by a small panel of markers would coincidentally be mutated as part of the spontaneous background rate possessed by microsatellite regions. The unlikely chance occurrence of random mutation at more than one assayed marker supports the analysis of additional markers in MSI-L cases, to confirm or refute that the tumor has only a single shifted marker. It is also possible that an MSI-L phenotype could relate to unique MSI phenotypes such as that reportedly seen in association with defects of *GTBP/hMSH6*. Because *GTBP* abnormalities are associated with mononucleotide instabil-

ity,³⁰ they would be predicted to manifest as a muted MSI phenotype.^{27,31} Analysis of this type of lesion with the microsatellite panel described here may reduce the number of loci functionally analyzed from 5 to 2, by effectively eliminating the dinucleotide markers. Finally, the statistical model predicts that the finding of MSI-Low will in some instances be due to insensitivity of the assay as a whole, reflecting a combination of locus informativities which are less than 100% in conjunction with testing of a limited number of loci. One can use the model to predict how and to what degree the assay will lose sensitivity, given known locus informativities and the number of tests to be combined to render a single diagnosis.

Thus, one may use the model to predict either the locus informativity required to attain a certain level of assay sensitivity, given a certain number of tests being performed, or to determine the number of loci that must be tested to attain a level of sensitivity, given certain informativities. In a situation where the overall testing sensitivity may be limited, as is may be the case in MSI-Low phenotypes, the model justifies the use of additional testing loci. We used five to six additional loci from a list of backup microsatellites proposed at the NCI conference to further test cases with an initial diagnosis of MSI-Low. The model is also helpful in justifying the initial analysis of five loci to diagnose MSI, given a certain average level of locus informativity. It also emphasizes the importance of finding sets of microsatellite loci which are optimally informative for each tumor type, because the overall sensitivity of the test drops off substantially as the average of individual component informativity drops below 70%. In this regard, the independence of marker informativities is critically important because, according to the model, if markers are always shifted in exactly the same tumors, the redundancy effectively eliminates a number of loci analyzed.

The model described here should find useful applications in a variety of pathology testing situations, where multiple independent assays are used together to establish a single combined diagnostic result. It provides a means of determining how many independent component assays, with specific informativities, will be required to achieve an acceptable overall diagnostic sensitivity and allows one to predict for assay failures, identify problem situations, and preemptively create backup plans for the management of these situations.

The fluorescence multiplex assay presented has many advantages over conventional MSI analysis. This assay is logistically and financially more feasible in a clinical molecular diagnostic setting, and the results generated compare favorably to standard analysis techniques. The multicomponent testing statistical model validates the utility of analyzing five loci and also allows for prediction of and planning for assay failure. Lastly, the assay has great utility in MSI research applications, as it allows for relatively high throughput, analysis, storage, and management of MSI data. These capabilities may greatly streamline the generation, maintenance and comparability of MSI databases in the future.

Acknowledgments

We thank Dr. James Willson for generously providing the Vaco 670 and L670 cell lines, Ms. Patty Longo for technical expertise, and Drs. Sanford Markowitz, Bert Vogelstein, Wink Baldwin, Ralph Hruban, and Ramon Parsons for helpful discussions.

References

1. Parsons R, Li G, Longley M, Fang W, Papadopoulos N, Jen J, de la Chapelle A, Kinzler K, Vogelstein B, Modrich P: Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* 1993, 75:1227-1236
2. Leach F, Nicolaides N, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen L, Nystrom-Lahti M, Guan XY, Zhang J, Meltzer P, Yu J, Kao F, Chen D, Cerosaletti K, Fournier R, Todd S, Lewis T, Leach R, Naylor S, Weissenbach J, Mecklin J, Jarvinen H, Petersen G, Hamilton S, Green J, Jass J, Watson P, Lynch H, Trent J, de la Chapelle A, Kinzler K, Vogelstein B: Mutations of a MutS homolog in hereditary non-polyposis colorectal cancer. *Cell* 1993, 75:1215-1225
3. Fishel R, Lescoe MK, Rao MRS, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R: The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 1993, 75:1027-1038
4. Papadopoulos N, Nicolaides N, Wei Y, Ruben S, Carter K, Rosen C, Haseltine W, Fleischmann R, Fraser C, Adams M, Venter J, Hamilton S, Petersen G, Watson P, Lynch H, Peltomaki P, Mecklin J, de la Chapelle A, Kinzler K, Vogelstein B: Mutations of a mutL homolog in hereditary colon cancer. *Science* 1994, 263:1625-1629
5. Bronner C, Baker S, Morrison P, Warren G, Smith L, Lescoe M, Kane M, Earabino C, Lipford J, Lindblom A, Tannergard P, Bollag R, Godwin A, Ward D, Nordenskjold M, Fishel R, Kolodner R, Liskay R: Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 1994, 368:258-261
6. Nicolaides N, Papadopoulos N, Liu B, Wei Y, Carter K, Ruben S, Rosen C, Haseltine W, Fleischmann R, Fraser C, Adams M, Venter J, Dunlop M, Hamilton S, Petersen G, de la Chapelle A, Vogelstein B, Kinzler K: Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 1994, 371:75-80
7. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB: Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci USA* 1998, 95:6870-6875
8. Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, Li GM, Drummond J, Modrich PL, Sedwick WD, Markowitz SD: Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci USA* 1998, 95:8698-8702
9. Marra G, Boland CR: Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* 1995, 87:1114-1125
10. Eshleman JR, Markowitz SD: Microsatellite instability in inherited and sporadic neoplasms. *Curr Opin Oncol* 1995, 7:83-89
11. Hemminki A, Peltomaki P, Mecklin JP, Jarvinen H, Salovaara R, Nystrom-Lahti M, de la Chapelle A, Aaltonen LA: Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet* 1994, 8:405-410
12. Eshleman JR, Lang EZ, Bowerfind GK, Parsons R, Vogelstein B, Willson JKV, Veigl ML, Sedwick WD, Markowitz SD: Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. *Oncogene* 1995, 10:33-37
13. Bhattacharyya NP, Skandalis A, Ganesh A, Groden J, Meuth M: Mutator phenotypes in human colorectal carcinoma cell lines. *Proc Natl Acad Sci USA* 1994, 91:6319-6323
14. Eshleman JR, Markowitz SD, Donover PS, Lang EZ, Lutterbaugh JD, Li GM, Longley M, Modrich P, Veigl ML, Sedwick WD: Diverse hyper-

- mutability of multiple expressed sequence motifs present in a cancer with microsatellite instability. *Oncogene* 1996, 12:1425–1432
15. Shibata D, Peinado MA, Ionov Y, Malkhosyan S, Perucho M: Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet* 1994, 6:273–281
16. Augenlicht LH, Richards C, Corner G, Pretlow TP: Evidence for genomic instability in human colonic aberrant crypt foci. *Oncogene* 1996, 12:1767–1772
17. Aaltonen L, Peltomäki P, Leach F, Sistonen P, Pylkkanen L, Mecklin J, Jarvinen H, Powell S, Jen J, Hamilton S, Petersen G, Kinzler K, Vogelstein B, de la Chapelle A: Clues to the pathogenesis of familial colorectal cancer. *Science* 1993, 260:812–816
18. Thibodeau S, Bren G, Schaid D: Microsatellite instability in cancer of the proximal colon. *Science* 1993, 260:816–819
19. Sankila R, Aaltonen LA, Jarvinen HJ, Mecklin JP: Better survival rates in patients with MLH1-associated hereditary colorectal cancer. *Gastroenterology* 1996, 110:682–687
20. Rodriguez-Bigas M, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S: A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst* 1997, 89:1758–1762
21. Kim H, Jen J, Vogelstein B, Hamilton SR: Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol* 1994, 145:148–156
22. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Fodde R, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S: A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 1998, 58:5248–5257
23. Fujiwara TSJ, Watanabe T, Rashid A, Longo P, Eshleman JR, Booker S, Lynch HT, Jass JR, Green JS, Kim H, Jen J, Vogelstein B, Hamilton SR: Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. *Am J Pathol* 1998, 153:1063–1078
24. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press, 1989, pp 9.16–9.19
25. Jen J, Kim H, Piantadosi S, Liu Z-F, Levitt RC, Sistonen P, Kinzler KW, Vogelstein B, Hamilton SR: Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N Engl J Med* 1994, 331:213–221
26. Samowitz WS, Slattery ML, Potter JD, Leppert MF: BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am J Pathol* 1999, 154:1637–1641
27. Papadopoulos N, Nicolaides NC, Liu B, Parsons R, Lengauer C, Palombo F, D'Arrigo A, Markowitz S, Willson JKV, Kinzler KW, Jiricny J, Vogelstein B: Mutations of GTBP in genetically unstable cells. *Science* 1995, 268:1915–1917
28. Sidransky D: Nucleic acid-based methods for the detection of cancer. *Science* 1997, 278:1054–1058
29. Loeb LA: Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* 1991, 51:3075–3079
30. Drummond JT, Li GM, Longley MJ, Modrich P: Isolation of an hMSH2–p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 1995, 268:1909–1912
31. Verma L, Kane MF, Brassett C, Schmeits J, Evans DGR, Kolodner RD, Maher ER: Mononucleotide microsatellite instability and germline MSH6 mutation analysis in early onset colorectal cancer. *J Med Genet* 1999, 36:678–682