Journal of Pathology
J Pathol 2008; **215**: 204–210
Published online 18 March 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/path.2351

Original Paper

Most low-level microsatellite instability in colorectal cancers can be explained without an elevated slippage rate

T Graham,[1–3]* S Halford,[2] KM Page[1,3] and IPM Tomlinson[2,4]

[1] *Centre of Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, Gower St, London, WC1E 6BT, UK*
[2] *Molecular and Population Genetics, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK*
[3] *Department of Mathematics, University College London, Gower St, London, WC1E 6BT, UK*
[4] *Institute of Cancer, Barts and The London, Queen Mary's School of Medicine and Dentistry, Charterhouse Square, London, EC1M 6BQ, UK*

**Correspondence to:*
T Graham, CoMPLEX, University College London, Gower St, London, WC1E 6BT, UK.
E-mail: trevor.graham@ucl.ac.uk

No conflicts of interest were declared.

Received: 12 December 2007
Revised: 13 February 2008
Accepted: 9 March 2008

Abstract

Many cancers show a low level of microsatellite slippage and are labelled MSI-L (microsatellite instability — low). However, it is unclear whether this slippage can be attributed to some underlying genetic change that results in a mutator phenotype, analogous to mismatch repair deficiency in MSI-H cancers, or whether the apparent instability is the result of relatively frequent normal somatic slippage. Here, we have used a mathematical model of microsatellite slippage during cancer growth to estimate the degree of microsatellite slippage expected in a cancer due to normal somatic slippage. We compared the model to the slippage observed in 42 non-MSI-H cancers that were macro-dissected into four distinct regions and genotyped at $N = 9$ microsatellite loci. When the slippage rate was set at $\mu = 10^{-5}$ per locus per division, ten cancers showed a level of slippage in at least one region that was too severe to be expected from normal somatic slippage alone, suggesting that these cancers had acquired MSI-L. Only one of these ten cancers had putative MSI-L in all four regions. When we considered a slightly higher slippage rate of $\mu = 5 \times 10^{-5}$, none of the cancers showed a degree of slippage that could not be reasonably explained by normal somatic slippage. Counting the number of 'unstable' loci was a poor indicator of putative MSI-L status. We conclude that most low-level microsatellite instability in colorectal cancers can be explained without requiring an elevated slippage rate during neoplastic development, and hence there is little evidence for a discrete MSI-L group of cancers. Putative MSI-L status is indicated by the presence of at least one locus that has multiple alleles that differ by at least five motif repeats from the germline. If an underlying genetic change does cause MSI-L, it appears to be a relatively uncommon event that occurs late in oncogenesis.

Keywords: microsatellite instability; MSI-H; MSI-L; MSS; colorectal cancer; mathematical model

Introduction

The human genome contains many highly repetitive regions, in which a short motif of bases is repeated several times. These regions are termed microsatellites (MSs). Many cancers have one or more MS loci that show changes in the number of motif repeats when compared with the germline [1]. Cancers which have many loci that show a high degree of variability in motif repeats tend to have germline mutations in the mismatch repair (MMR) genes [2–4], or to have methylation of the MLH1 promoter, which is associated with protein loss [5,6]. These (epi)genetic changes impair the DNA repair machinery of the cell, causing an increase in the rate of MS slippage, so slippages accumulate at many loci. Consequently, MMR-deficient cancers are termed microsatellite instability — high (MSI-H) cancers.

It has been suggested that some cancers with a lower level of MS slippage (slippage at only a few loci), and without MMR deficiency, form a distinct pathological group [7]. These cancers are labelled low-level MS-unstable (MSI-L). Analogous to MMR deficiency in MSI-H cancers, putative MSI-L cancers are thought to have an underlying defect, possibly transient, that increases their MS slippage rate. If this group is to have pathological or clinical relevance, a reliable definition of the group is required [8]. Using this definition, it must be possible to reliably distinguish a MSI-L cancer from a MSI-H or MS-stable (MSS) cancer.

MSI-L cancers have been classified by having one, two or three MS loci that show slippage compared with the germline or, alternatively, by the presence of slippage at at least one but less than 30% of the loci examined (reviewed in ref 9). However, these relatively arbitrary definitions are contentious and some authors have questioned whether a distinct group of MSI-L cancers does exist [10]. The main reason for contention is that the rate of slippage errors at most MS loci is relatively high in normal cells — between 10^{-9} and 10^{-3} per locus per division, and typically towards the upper end of this range [11–14], whereas normal somatic DNA replication fidelity is between 10^{-10} and 10^{-6} per base per division [15–18]. Therefore, due to the large number of cell divisions that occur during neoplastic growth, a low-level of MS slippage in cancers is expected. Accordingly, Laiho *et al* showed that if enough loci in a cancer are examined, an 'unstable' locus will nearly always be found [19].

Here, we have proposed a null hypothesis: that the degree of MSI observed in a putative MSI-L cancer can be adequately explained by a low rate of normal 'background' MS slippage. We used a mathematical model of MS slippage during cancer growth to predict the level of MSI expected from background slippage and then compared this prediction with data from a group of non-MSI-H cancers. If the degree of MSI observed in a cancer was more severe than the prediction, then we concluded that the cancer had an elevated slippage rate and so was MSI-L. In contrast to previous work, for example ref 10, we considered both the number of MS loci showing slippage and the degree of heterogeneity present at each locus, to assess MSI status.

Materials and methods

Genotyping

Forty-seven paraffin-embedded colorectal cancers and paired normal samples were genotyped at $N = 9$ MS loci (D10S197, D13S175, D15S144, D15S659, D17S250, D18S58, D2S123, D5S346, and D8S87) that had been previously identified as being prone to slippage [20,21]. Permission to perform molecular analysis of these cancers was granted by the NHS Oxfordshire Research Ethics Committee B. Each cancer was macro-dissected into four distinct regions. DNA was extracted by incubation in proteinase K digestion buffer. PCR was performed using standard reaction conditions. PCR products were analysed with the ABI377 prism sequencer (Perkin-Elmer) and results were interpreted using Genotyper software (Perkin-Elmer). Additionally, the MSI-H status of each cancer was determined bythe presence of slippage at the BAT-26 locus [22].

Descriptive statistics

A collection of allelic lengths could be present in each PCR product. To describe the alleles, we calculated the average minimum mean-squared distance from the germline (A). The molecular methods that we used are unable to determine which germline allele(s) each length derived from. Hence the statistic A assumes 'maximum parsimony'; that is, that an allele of length l derives from the germline allele that is the most similar in size. Analysis of the PCR product indicates the relative concentration of each allele in the PCR product. However, these measurements are unreliable as an indicator of the true proportions of each allele in the cancer, due to possible contaminating normal tissue and slippage that occurs during PCR [23]. Hence we considered only whether an allele of a particular size was present or absent in the PCR product.

For each region of each cancer, we denoted the lengths of alleles present at the locus i by $l_{i1}, l_{i2}, \ldots, l_{in}$. Then the average minimum mean-squared distance from the germline at locus i (a_i) was defined as

$$a_i = \left(\frac{1}{n} \sum_{j=1}^{n} \min((l_{ij} - g_{i1})^2, (l_{ij} - g_{i2})^2) \right)^{1/2} \quad (1)$$

Then the data from all N loci were summarized as

$$A = \frac{1}{N} \sum_{i=1}^{N} a_i \quad (2)$$

Microsatellite slippage model

The mathematical model of MS slippage during cancer growth was based on that of Tsao *et al* [24] and is summarized in Figure 1. Our model described slippage of a MS before and during cancer growth. We assumed that a cancer grew from a single rogue cell. We simulated the lineage of this cell from birth until a cancer of approximately 10^9 cells, which corresponds to a cancer of about 1 cm^3 in volume, was formed. Exponential cancer growth began after T_G divisions, where T_G was chosen according to the patient's age at biopsy so that cancers had been growing for about 10 years before being removed from the body. We assumed that cells divided at a rate of $C = 182$ divisions per year, which corresponds to a cell division approximately every 2 days. We assumed that the cell division rate remained constant during cancer growth.

Pre-cancer cells were assumed to be diploid, so they contained two allelic copies of a MS locus. The initial germline lengths of the locus were g_1 and g_2, respectively, where, without loss of generality, $g_1 = 0$ and $g_2 \geq g_1$. Each time the cell divided, slippage of ± 1 motif repeat occurred at each allele of the MS locus with probability μ. To examine the effect of cell ploidy, we assumed that cancer cells were either diploid or tetraploid. In the tetraploid case, we assumed that cancer cells became tetraploid due to abnormal chromosome duplication at the initiation of cancer growth that resulted in two copies of each of

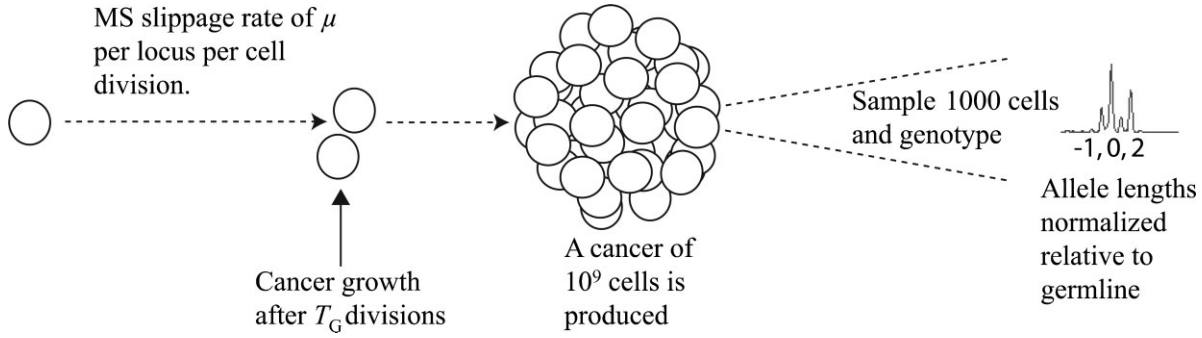

Figure 1. Illustration of a mathematical model of cancer growth. The model simulates the progeny of a single rogue cell, which initiates cancer growth after T_G divisions. We assume that the cancer grows exponentially, and T_G is chosen so that the rogue cell expands to form a cancer of approximately 10^9 cells after 10 years. Each cell has a number of MS loci. A slippage of ± 1 motif repeat occurs at each locus at a rate μ per locus per division. When the cancer has grown, 1000 cells are sampled and genotyped. An allele is considered present in the sample if it constitutes at least X% of the total number of alleles in the sample, where X is the PCR sensitivity. The lengths of the alleles present in the sample are normalized relative to the germline allele lengths, as indicated in the figure

the two original alleles. We assumed that no further abnormal chromosome duplication occurred. Slippage at each of these four alleles then occurred according to the stepwise mutation model. MS slippage in cancers is relatively poorly characterized; so this simple stepwise model was chosen as an adequate first approximation, which also minimized the degree of heterogeneity expected under the null hypothesis, since slippages of more than one motif repeat were not permitted. We examined the effect of two relatively low slippage rates: $\mu = 5 \times 10^{-5}$ and $\mu = 10^{-5}$ per cell per division [11,12].

Rather than growing exponentially, cancers may expand much more slowly (linearly), especially if there is a high rate of cell death within the cancer. This alternative, slower, growth would mean that there would be more cell divisions separating the cells in the cancer from their most recent common ancestor, and hence more variation would be expected at each MS locus. Hence our model produced a conservative estimate of the degree of MSI expected in a cancer.

When cancer growth had been simulated, in the model we randomly sampled a number of cells ($S = 1000$) from the cancer. Due to the inherent variability of the PCR process, an allele present in a sample will be reliably identified only if it constitutes a significant proportion of the total alleles in the sample [23]. Accordingly, in the model we set the PCR detection threshold at X%, so that an allele of length k was only considered present in the sample if at least X% of the total number of alleles in the sample had length k. We simulated the model for $X = 10\%$ and $X = 20\%$.

To compare the expected versus the observed MS slippage, we used the statistic A, described above. For each patient, we simulated the model for each of the N MS loci and described these N simulations as a single *instance* of the model. By repeatedly simulating the model, 10^5 instances of the model for each cancer, and recording each value of A, we produced an approximation to the underlying distribution of A (labelled $\hat{A}$) that was expected under our null hypothesis. The distribution $\hat{A}$ depended on the germline allele lengths of each MS locus and the age of the patient at biopsy, and so had to be simulated for each cancer in turn.

To calculate the p value, the probability that MS slippage at least as extreme as the slippage observed in each region of the cancer would occur if the null hypothesis were true, we calculated

$$p = P(A_{\text{sim}} \geq A_{\text{obs}}) \tag{3}$$

where A_{sim} is the simulated value of A and A_{obs} is the value of A observed in the cancer. For each sample, the null hypothesis was rejected if $p < 0.05$ and the cancer was labelled MSI-L.

Results

The heart of our approach was a mathematical model of normal 'background' MS slippage during cancer growth. By comparing the degree of slippage predicted by the model with the degree of slippage observed in a group of non-MSI-H cancers, we tested our null hypothesis that the MS slippage observed in a cancer was due only to a low rate of background slippage.

Rejection of null hypothesis for MSI-H cancers

First we tested whether our method would successfully reject the null hypothesis for MSI-H cancers. We initially examined five MSI-H cancers, which were identified by slippage at the BAT-26 locus [22]. All five cancers had significant MS slippage at most of the loci examined. We set the slippage rate of $\mu = 10^{-5}$ per locus per cell division and assumed that cancer cells were diploid and that the PCR detection threshold was 10%. Then in regions of the five cancers, the degree of variation exhibited was very unlikely to occur under the null hypothesis (largest p value: $p < 10^{-5}$). Similar results were obtained when the

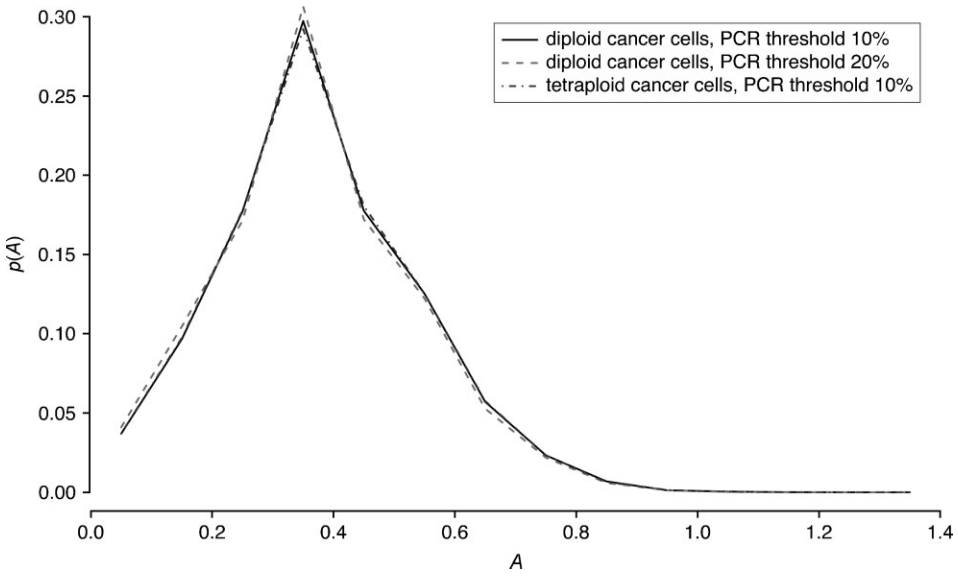

Figure 2. Effect of the PCR detection threshold or cancer cell ploidy. The model was simulated for $N = 9$ loci, with $\mu = 5 \times 10^{-5}$ and initial germline lengths $g_{i1} = 0$, $g_{i2} = 1$, for a patient aged 80 years at biopsy. Neither the PCR detection threshold nor the ploidy of the cancer cells significantly affects the distribution of A expected under our null hypothesis

slippage rate was increased to $\mu = 5 \times 10^{-5}$ (largest p value: $p < 0.026$). Cancer cell ploidy and the PCR detection threshold did not significantly alter the p values (see Figure 2). We therefore confirmed that the MS slippage in the five MSI-H cancers must have resulted from an elevated slippage rate.

Most MSI in non-MSI-H cancers can be explained by background slippage

Next we examined whether the null hypothesis could explain the slippage observed in the 42 non-MSI-H cancers. We initially assumed a slippage rate of $\mu = 10^{-5}$ per locus per cell division; that cancer cells were diploid; and that the PCR detection threshold was 10%. Using these assumptions, ten out of the 42 non-MSI-H cancers (24%) contained at least one region where the degree of MSI was too severe to be explained by the null hypothesis (largest p value: $p < 0.047$), suggesting that these ten cancers were MSI-L (see Table 1). Borderline levels of MSI ($p < 0.1$) were observed in regions of two cancers, but both of these cancers contained another region where the amount of slippage was indicative of MSI-L.

We examined the degree of MSI in these ten putative MSI-L cancers (see Table 1). Four of these ten cancers (cancers 9–11 and 38) had a putative MSI-L region that had only one unstable locus, which always had multiple extra alleles that showed a maximum deviation from the germline of at least five motif repeats. Five cancers had MSI-L regions with two unstable loci (cancers 4, 18, 20, 29, and 36) and only one cancer had a MSI-L region with three unstable loci (cancer 35); these cancers had multiple extra sizes of alleles at at least two loci or had at least two unstable loci, one of which having an allele that differed by at least three motif repeats from the germline. In comparison, non-MSI-L cancers always had less than three unstable regions, which exhibited less severe MSI than MSI-L cancers. Typically, a value of $A > 0.25$ was sufficient to produce a p value less than 0.05. Cancer 36 had A values of 0.37, 0.37, and 0.34 in regions 2–4, respectively, but these values did not produce a significant p value. This is because only five loci were successfully genotyped in cancer 36 and the variance of A is greater when fewer loci are examined (data not shown).

We compared the Dukes' stage, location in the colorectum, the patient's sex, and age at biopsy between the putative MSS and MSI-L cancers. No significant associations were found.

Level of MSI differs between regions of a cancer

Of the ten cancers identified as putative MSI-L, only one (cancer 29) appeared to have MSI-L in all four regions. Cancer 29 had two unstable loci that were both unstable in all four regions of the cancer and had a maximum change of at least three repeats from the germline. Of the remaining nine putative MSI-L cancers, five cancers had MSI-L in only one region; one cancer had MSI-L in two regions; and a further two cancers had MSI-L in three regions. Therefore, if these nine cancers were truly MSI-L, then the increase in slippage rate would likely have occurred late in the development of the cancer. If MSI-L occurred early, then all the regions of a cancer would show similarly high degrees of MSI.

No cancers appear MSI-L if the background slippage rate is increased

Next we assumed a slightly higher slippage rate of $\mu = 5 \times 10^{-5}$ per locus per cell per division and again assumed that cancer cells were diploid and that the

Table 1. Summary of the MS slippage present in each section of each cancer. Values are magnitudes of the statistic *A*, the average minimum mean-squared distance from the germline. Values in bold produced a significant *p* value ($p < 0.05$) when the background slippage rate was set at $\mu = 10^{-5}$ per locus per division, suggesting that the slippage observed in that region was due to an elevated slippage rate

							Value of *A* in cancer section			
ID	**Age at biopsy (years)**	**Sex**	**Location**	**Dukes' stage**	**MSI-H**	***N***	**1**	**2**	**3**	**4**
H1	63	F	Ascending	B	yes	7	**2.02**	**2.01**	**1.97**	**1.83**
H2	67	F	Sigmoid	B	yes	7	**1.49**	**1.49**	**1.49**	**1.55**
H3	69	F	sigmoid	C	yes	9	**0.82**	**0.82**	**0.82**	**1.20**
H4	75	M	sigmoid	C	yes	8	**0.91**	**0.91**	**0.81**	**0.79**
H5	103	F	ascending	C	yes	9	**2.14**	**1.93**	**1.70**	**1.91**
1	23	F	rectosigmoid	B	no	8	0.07	0.07	0.07	0.07
2	50	F	rectum	C	no	9	0.00	0.00	0.00	0.00
3	53	F	sigmoid	B	no	9	0.00	0.00	0.00	0.00
4	54	M	rectum	—	no	7	**0.56**	**0.32**	**0.25**	0.14
5	55	F	rectum	B	no	9	0.00	0.00	0.00	0.00
6	56	M	rectum	—	no	9	0.00	0.00	0.19	0.00
7	57	M	rectum	B	no	9	0.00	0.00	0.00	0.00
8	58	M	rectum	B	no	9	0.00	0.00	0.20	0.00
9	58	F	ascending	C	no	9	0.00	0.00	0.00	**0.44**
10	58	F	rectum	A	no	9	**0.36**	0.00	0.13	0.00
11	59	F	rectum	B	no	8	0.00	**0.45**	0.00	0.00
12	60	F	sigmoid	B	no	9	0.06	0.00	0.00	0.00
13	63	F	Hepatic flexure	A	no	9	0.00	0.00	0.00	0.00
14	64	F	rectum	—	no	9	0.00	0.00	0.00	0.00
15	64	M	rectum	B	no	9	0.06	0.06	0.06	0.00
16	64	M	sigmoid	C	no	9	0.00	0.00	0.00	0.00
17	64	M	sigmoid	B	no	8	0.00	0.00	0.00	0.00
18	64	M	rectum	A	no	8	**0.36**	0.00	0.07	**0.42**
19	65	M	rectum	—	no	9	0.13	0.19	0.00	0.06
20	65	F	sigmoid	A	no	6	**0.30**	**0.38**	0.19	**0.31**
21	65	M	rectum	C	no	9	0.00	0.12	0.12	0.00
22	65	M	sigmoid	C	no	9	0.15	0.09	0.15	0.12
23	66	F	sigmoid	B	no	9	0.00	0.00	0.00	0.00
24	67	M	ascending	C	no	9	0.00	0.00	0.00	0.00
25	67	M	descending	C	no	9	0.00	0.00	0.00	0.00
26	70	M	sigmoid	—	no	9	0.00	0.00	0.00	0.00
27	70	F	sigmoid	B	no	9	0.06	0.06	0.06	0.06
28	71	M	rectum	C	no	6	0.22	0.12	0.23	0.00
29	71	M	sigmoid	A	no	6	**0.54**	**0.77**	**0.49**	**0.58**
30	72	M	descending	C	no	9	0.12	0.12	0.12	0.12
31	73	F	sigmoid	C	no	9	0.00	0.00	0.00	0.00
32	75	M	rectum	—	no	9	0.00	0.00	0.00	0.00
33	75	M	rectum	B	no	9	0.00	0.00	0.00	0.00
34	78	F	rectum	B	no	9	0.00	0.00	0.14	0.00
35	78	F	rectum	C	no	9	0.00	0.06	**0.32**	0.00
36	79	M	ascending	C	no	5	**0.49**	0.37	0.37	0.34
37	80	M	ascending	B	no	9	0.00	0.00	0.00	0.00
38	80	F	sigmoid	C	no	9	0.00	**0.41**	0.31	0.00
39	81	M	rectum	A	no	9	0.00	0.00	0.00	0.00
40	83	F	rectum	C	no	9	0.12	0.14	0.12	0.22
41	84	F	rectum	B	no	9	0.23	0.00	0.23	0.00
42	86	M	rectum	B	no	9	0.00	0.00	0.00	0.00

PCR detection threshold was 10%. With this slightly higher slippage rate, the slippage observed in all the non-MSI-H cancers could be adequately explained by the null hypothesis (smallest p value: $p < 0.094$).

p values do not depend strongly on cell ploidy or the PCR detection threshold

Assuming that cancer cells were tetraploid or increasing the PCR detection threshold to 20% did not significantly alter the p value assigned to each cancer region (see Figure 2).

Number of unstable loci is a poor indicator of MSI-L status

We considered how reliable the number of 'unstable' loci was as a predictor of MSI-L status. We assumed that our model with $\mu = 10^{-5}$ correctly identified regions of a cancer that were MSI-L and computed the sensitivity and specificity of identifying these regions by their having at least one, two or three loci that showed slippage (see Table 2). Classifying a cancer as MSI-L by the presence of slippage at a single locus had

Table 2. Sensitivity and specificity of defining MSI-L status by counting the number of loci that show slippage. We assumed that our model, with $\mu = 10^{-5}$, could reliably identify MSI-L in a region of a cancer and then examined the success of a diagnostic test for MSI-L which considered only the number of loci that showed slippage in each region of a cancer

Minimum number of unstable loci required for MSI-L classification	Sensitivity	Specificity
1	1 (18/18)	0.69 (102/150)
2	0.72 (13/18)	0.97 (145/150)
3	0.06 (1/18)	1 (150/150)

poor specificity, whereas requiring slippage at multiple loci for MSI-L classification had poor sensitivity.

Discussion

If we assume that normal somatic slippage at MS loci is a rare event (slippage rate of $\mu = 10^{-5}$ per locus per cell per division), then a minority of non-MSI-H cancers (24%) contain a region that has slippage that is too severe to be expected under the hypothesis that there is no increase in the MS slippage rate during neoplastic development. If we consider the proportion of the regions with severe slippage, then the frequency falls to 18/168 (11%). If MSI-L is a real phenomenon, then it is relatively rare. Furthermore, MSI-L cannot be a frequent early event in tumourigenesis since only a single cancer was observed that had putative MSI-L in all four regions that were genotyped. Therefore we conclude that acquisition of MSI-L is not a frequent early change in oncogenesis.

We urge caution when labelling a cancer as MSI-L. We have shown that simply counting the number of loci that show slippage in a cancer is a poor indicator of MSI-L status. Our model suggests that cancers that have at least one locus that shows significant slippage (a change of at least five motif repeats from the germline and multiple extra alleles), or multiple unstable loci, at least one of which is highly unstable, are the most likely to have acquired a mutator phenotype. An equivalent indicator is a value of $A > 0.25$ when sufficient loci ($N \geq 9$) are examined.

If we assumed a slight increase in the normal MS slippage rate (to $\mu = 5 \times 10^{-5}$ per locus per cell per division), all of the slippage observed in our cohort of cancers could be adequately explained by background slippage. Estimates of MS slippage rates range from 10^{-8} to 10^{-3} per locus per division [11,12] and are typically towards the upper end of this scale [13,14], so an average background rate of $\mu = 5 \times 10^{-5}$ appears reasonable. Moreover, the nine MS loci examined were selected as they were sensitive for detection of putative MSI-L cancers [20,21], so they are likely to have a relatively high slippage rate. Although we cannot be sure of the precise background MS slippage rate, our model illustrates that it is reasonable to expect a non-negligible degree of slippage in a cancer. Furthermore, our simple model minimizes the degree of genetic instability expected by chance, so is conservative with respect to the existence of MSI-L. Therefore, we conclude that an elevated slippage rate is not required to explain most MS slippage observed in non-MMR-deficient cancers.

Is it reasonable, then, to conclude that MSI-L exists even in a minority of non-MSI-H cancers? During PCR, MS loci can slip [23], so MSI will be occasionally observed in cancers that are actually MSS. Rather than an intrinsic genetic defect, MSI-L could be the result of external mutagenic factors. Previous work has looked for an association between MSI-L and clinicopathological variables, such as K-ras mutations or loss of heterozygosity events at APC (reviewed in ref 9). The findings have been inconsistent, with most studies finding no significant association between MSI-L and any clinico-pathological or molecular variable, and occasional studies finding an association between MSI-L and mutations in K-ras [25], or methylation of the DNA repair enzyme O^6-methylguanine methyltransferase [26]. Potentially, the reason for the inconsistency could be the difficulty in identifying MSI-L cancers. Our definition of MSI-L status may prove useful to study these associations.

Acknowledgements

We are grateful for the use of UCL Research Computing facilities (C^3) and to Lisa Willis for her useful comments on the study. TG is supported by an EPSRC LSI DTC studentship.

References

1. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 1993;**363:**558–561.
2. de la Chapelle A. Genetic predisposition to colorectal cancer. *Nature Rev Cancer* 2004;**4:**769–780.
3. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature* 2001;**411:**366–374.
4. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996;**87:**159–170.
5. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, *et al*. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* 1997;**57:**808–811.
6. Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, *et al*. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci U S A* 1998;**95:**8698–8702.
7. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, *et al*. 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.
8. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, *et al*. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004;**96:**261–268.

9. Tomlinson I, Halford S, Aaltonen L, Hawkins N, Ward R. Does MSI-low exist? *J Pathol* 2002;**197:**6–13.
10. Gonzalez-Garcia I, Moreno V, Navarro M, Marti-Rague J, Marcuello E, Benasco C, *et al*. Standardized approach for microsatellite instability detection in colorectal carcinomas. *J Natl Cancer Inst* 2000;**92:**544–549.
11. Bhattacharyya NP, Skandalis A, Ganesh A, Groden J, Meuth M. Mutator phenotypes in human colorectal carcinoma cell lines. *Proc Natl Acad Sci U S A* 1994;**91:**6319–6323.
12. Hanford MG, Rushton BC, Gowen LC, Farber RA. Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair. *Oncogene* 1998;**16:**2389–2393.
13. Kruglyak S, Durrett RT, Schug MD, Aquadro CF. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci U S A* 1998;**95:**10774–10778.
14. Lai Y, Sun F. The relationship between microsatellite slippage mutation rate and the number of repeat units. *Mol Biol Evol* 2003;**20:**2123–2131.
15. Elmore E, Kakunaga T, Barrett JC. Comparison of spontaneous mutation rates of normal and chemically transformed human skin fibroblasts. *Cancer Res* 1983;**43:**1650–1655.
16. Loeb LA. A mutator phenotype in cancer. *Cancer Res* 2001;**61:**3230–3239.
17. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, *et al*. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;**314:**268–274.
18. Wang TL, Rago C, Silliman N, Ptak J, Markowitz S, Willson JK, *et al*. Prevalence of somatic alterations in the colorectal cancer cell genome. *Proc Natl Acad Sci U S A* 2002;**99:**3076–3080.
19. Laiho P, Launonen V, Lahermo P, Esteller M, Guo M, Herman JG, *et al*. Low-level microsatellite instability in most colorectal carcinomas. *Cancer Res* 2002;**62:**1166–1170.
20. Halford S, Sasieni P, Rowan A, Wasan H, Bodmer W, Talbot I, *et al*. Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait. *Cancer Res* 2002;**62:**53–57.
21. Halford SE, Sawyer EJ, Lambros MB, Gorman P, Macdonald ND, Talbot IC, *et al*. MSI-low, a real phenomenon which varies in frequency among cancer types. *J Pathol* 2003;**201:**389–394.
22. Stone JG, Tomlinson IP, Houlston RS. Optimising methods for determining RER status in colorectal cancers. *Cancer Lett* 2000;**149:**15–20.
23. Sieben NL, ter Haar NT, Cornelisse CJ, Fleuren GJ, Cleton-Jansen AM. PCR artifacts in LOH and MSI analysis of microdissected tumor cells. *Hum Pathol* 2000;**31:**1414–1419.
24. Tsao JL, Yatabe Y, Salovaara R, Jarvinen HJ, Mecklin JP, Aaltonen LA, *et al*. Genetic reconstruction of individual colorectal tumor histories. *Proc Natl Acad Sci U S A* 2000;**97:**1236–1241.
25. Jass JR, Biden KG, Cummings MC, Simms LA, Walsh M, Schoch E, *et al*. Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *J Clin Pathol* 1999;**52:**455–460.
26. Whitehall VL, Walsh MD, Young J, Leggett BA, Jass JR. Methylation of *O*-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. *Cancer Res* 2001;**61:**827–830.