Low-Level Microsatellite Instability Occurs in Most Colorectal Cancers and Is a Nonrandomly Distributed Quantitative Trait

Sarah Halford, Peter Sasieni, Andrew Rowan, Harpreet Wasan, Walter Bodmer, Ian Talbot, Nicholas Hawkins, Robyn Ward, and Ian Tomlinson

Molecular and Population Genetics Laboratory [S. H., A. R., I. To.] and Mathematics, Statistics and Epidemiology [P. S.], Imperial Cancer Research Fund, London W2CA 3PX. United Kingdom; Cancer and Immunogenetics Laboratory, Imperial Cancer Research Fund, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX1 9DZ, United Kingdom [S. H., W. B.]; Department of Medical Oncology, Hammersmith Hospital, London W12 OHS, United Kingdom [H. W.]; Academic Department of Pathology, St Mark’s and Northwick Park Hospitals, Harrow HA1 3UJ, United Kingdom [I. Ta.]; School of Pathology, University of Sydney, New South Wales, 2052, Australia [N. H.]; Department of Medical Oncology, St Vincent’s Hospital, Victoria St, Darlinghurst, NSW, 2010, Australia [R. W.]

Abstract

About 10–15% of colorectal cancers show high-level microsatellite instability. The characteristics and very existence of low-level instability (MSI-L) are unclear, although some studies have found associations between MSI-L and molecular characteristics, notably more frequent K-ras mutations and a low level of allele loss near APC. We have attempted to define a MSI-L group of tumors by analyzing 107 sporadic colorectal carcinomas at 44 microsatellites. Ten (9.7%) MSI-H cancers were identified, but there was no evidence for a discrete MSI-L group. However, the 97 non-MSI-H cancers showed greater variation in the frequency of MSI than was expected by chance. Most cancers (68%) in the non-MSI-H group showed some MSI and could therefore be classed as nominally MSI-L. No association was found between MSI-L (or the level of MSI) and any clinicopathological or molecular variable, including K-ras mutation and loss of heterozygosity at APC. The causes of variation in level of the MSI in non-MSI-H colorectal cancers are unknown, but the differences are quantitative and probably reflect the evolutionary histories of the cancers rather than qualitatively different genetic pathways of tumorigenesis.

Introduction

Slippage of repeat number at microsatellite loci is a stochastic process occurring with a certain probability in all cells; it is for this reason that many microsatellites are highly polymorphic. For several years, the phenomenon of MSI was ignored by workers who were using highly polymorphic markers to study colorectal cancers. The microsatellite bandshifts observed in DNAs from a subset of cancers were ascribed to a combination of random, background events and artifact derived from Taq polymerase slippage. However, since the importance of MSI and its relationship to DNA mismatch repair were first recognized (1–3), the situation could hardly be more different, because MSI was ascribed to a combination of random, background events and error in DNA repair. MSI is now known to be a consequence of mutations in genes that code for enzymes involved in DNA repair, and the phenomenon of MSI has become a reliable and popular method for identifying cancers with mutations in these genes. The causes of variation in level of MSI in colorectal cancers are unknown, but the differences are quantitative and probably reflect the evolutionary histories of the cancers rather than qualitative differences in the mutational pathways of tumorigenesis.

In most colorectal cancers, DNA mismatch repair is impaired, but it has been shown that some cancers have a lower level of MSI than others (4, 5). Instability resulting from defects in mismatch repair involves a high proportion of microsatellites in each tumor, typically affects mononucleotide repeats such as microsatellites in each tumor, typically affects mononucleotide repeats such as microsatellites from MSS lesions is unproven. It is likely, moreover, that because microsatellite repeat slippage occurs about every 1,000 to 10,000 divisions in normal cells, the clonal nature of cancers would make this slippage detectable and might account for at least some MSI-L cancers. We therefore analyzed a set of sporadic colorectal cancers to determine the following: (a) would (almost) all cancers be MSI-L if sufficient microsatellite markers were typed; (b) would cancers defined as MSI-L and MSS by one set of markers also be defined in the same categories when a second set of microsatellites was typed; (c) does the distribution of MSI fit a bimodal (MSS and MSI-H) or trimodal (MSS, MSI-L, and MSI-H) distribution; and (d) are there any clinicopathological or molecular features specific to MSI-L cancers?

Materials and Methods

Frozen specimens of 107 colorectal carcinomas from an unselected series and paired normal bowel tissue were taken from colectomy specimens from St. Mark’s Hospital, London. Histological review showed that all H&E sections taken contemporaneously from the same tumor consisted predominantly of carcinoma. Cases with inflammatory bowel disease or a known Mendelian cancer syndrome were excluded. For each tumor, the following clinicopathological data were collected: patient age; Duke’s stage; grade; and anatomical site in colon. DNA was extracted from each specimen using standard methods.

Fifty-four microsatellite markers (details available from authors, but including D2S123, BAT26, BAT25, MYCL, D5S346, D18S474, D8S1771,
We initially derived a histogram showing for each cancer the proportion of (successfully amplified) markers that were unstable (Fig. 1). No attempt was made to distinguish between different markers at this stage. By simple inspection, there appeared to be a non-unimodal distribution of instability. The mean level of instability was 5.0% at any marker. Ten cancers (9.4%), corresponding to MSI-H cancers, had instability at the marker BAT26. An additional 2 cancers were MSI+ at BAT26 but had only ~11% of unstable markers overall and were therefore not classified as MSI-H. The MSI-H cancers tended to have clinicopathological features known to typify these lesions, such as location in the cecum, ascending and transverse colon (Fisher’s exact test, \( P < 0.00035 \)), low overall frequency of LOH (\( t = 4.52, \) degrees of freedom \( = 105, \) \( P < 0.001 \)), and low frequency of \( K^{ras} \) mutation (Fisher’s exact test, \( P = 0.05 \)). One MSI-H cancer did, however, have LOH at 25% of markers, including those close to \( APC \) and \( p53 \), suggesting that this might be an example of a rare MSI-H tumor with chromosomal instability.

The 97 non-MSI-H cancers had between 0 and 7 unstable microsatellites, with the proportion of unstable markers being no more than 25% in this group of 97 (Fig. 1). We investigated whether there was any evidence that the 97 non-MSI-H cancers came from a mixture of populations. There were 24 tumors that were stable at all microsatellites tested (mean number of MSI+ markers successfully amplified, 37; range, 24–44). The variation in the proportion of the tumors that tested MSI+ for each marker was much greater than would be expected if all of the markers had an equal probability of displaying instability in these tumors. In fact, the variance of the proportions was approximately three times the expected variance under the assumption of equal underlying probabilities. We therefore estimated different underlying probabilities of instability for each marker using the following. Let \( X_i = r/n_i \) be the proportion of \( n_i \) tumors with MSI+ for the \( i \)th marker. Then, \( \text{var}(X_i) = \mathbb{E} \{ \text{var}(X_i, n_i) \} + \text{var}(\mathbb{E}(X_i, n_i)) \). Assuming a common probability of instability \( p \), \( \text{var}(X_i, n_i) = p(1-p)/n_i \) and \( \mathbb{E}(X_i, n_i) = p \). Hence, \( \text{var}(X_i) = p(1-p) \times \text{mean}(1/n_i) \). Using our data gave a value of 0.00053, whereas the actual (sample) variance of \( X \) was 0.00154. Even if all of the markers did not have a common probability, we assumed that the mean of the probabilities was \( p \) (i.e., \( \mathbb{E}(p_i) = p \)), and then \( \text{var}(\mathbb{E}(X_i, n_i)) = \text{var}(p_i) \). Hence, \( \text{var}(p_i) \) was estimated by \( 0.00154 - 0.00053 = 0.00101 = p(1-p)/44 \), where \( p = 0.0477 \) is the (sample) mean of the \( p_i \)s. Thus, shrinkage estimates were performed using the formula: (no. tumors with instability + 2.147)/(total no. of tumors studied + 44), where 2.147/44 corresponded to the proportion of samples positive overall, and 44 was chosen to match the expected variance of a sample from the shrunk estimates of the observed variance of the raw proportions. Given the shrunk estimates for each of the 44 markers, we simulated 10,000 data sets corresponding to testing 97 tumors with the 44 markers. The simulations took into account the exact list of markers that were successfully used for each tumor.

The results of the simulations were analyzed in two ways:

(a) We looked at the number of samples with a given number of unstable sites (Table 1). There were two tumors with seven (or more) unstable sites compared with only 0.28 expected (one-sided \( P = 0.032 \)) and six with five or more compared with 3.94 expected.
that were unstable. There were 2 tumors with /H11005 P D18S474 and classed as MSS.

were unstable at one or more markers and were therefore classed as a nominal subdivision of the 97 non-MSI-H tumors into MSI-L and which instability was homogeneous throughout the non-MSI-H cancers.

definite excess of samples with between 10 and 25% microsatellites Thus, based on the proportion of unstable markers, there was a /H11022 10% of markers MSI microsatellites (/H11022 /H11001 markers MSI microsatellites). In the other tail, there were 24 tumors with no instability compared with 15.6 expected (P = 0.016).

(b) We examined the results obtained on the proportions of markers that were unstable. There were 2 tumors with >17.5% of markers showing instability compared with only 0.17 expected (P = 0.015), and in fact both these tumors were unstable at >20% of markers (P = 0.005). Furthermore, there were five tumors with >12.5% markers MSI+ (P = 0.053), and all were unstable at >15% of microsatellites (P = 0.0005). In all, there were 12 samples with >10% of markers MSI+ compared with 6.2 expected (P = 0.021). Thus, based on the proportion of unstable markers, there was a definite excess of samples with between 10 and 25% microsatellites unstable and an excess with no instability, compared with a model in which instability was homogeneous throughout the non-MSI-H cancers.

On the basis of previous classifications (7, 27), we decided to make a nominal subdivision of the 97 non-MSI-H tumors into MSI-L and MSS categories for the purpose of further testing of the MSI-L hypothesis. Seventy-three non-MSI-H cancers (68% of all studied) were unstable at one or more markers and were therefore classed as MSI-L. Twenty-four cancers (22% of total) were stable at all markers and classed as MSS.

The frequencies of allelic loss at APC (D5S346) and SMAD4 (D18S474) were not significantly different in the nominally MSS and MSI-L cancers (32% versus 27%, P > 0.70; and 33% versus 42%, P > 0.50). K-ras mutations were not significantly more common (Fisher’s exact test, P = 0.08) in the MSI-L cancers (20 mutant versus 42 wild-type) than the MSS cancers (2 mutant versus 17). Because, however, there were more K-ras mutations in the MSI-L group and an association between MSI-L and K-ras had been shown previously (27) using a smaller set of markers, we extended the MSS group to include cancers with a single unstable marker (and reduced the MSI-L group accordingly), in case occasional slippage in otherwise stable cancers had obscured a true association. Using this reclassification, however, no association was found (Fisher’s exact test, P > 0.30; 16 mutant versus 34 wild-type in MSI-L, 6 versus 25 in MSS). LOH at 1p32 (MYCL), 8p12 (D8S1771), and 8p21 (D8S887) was not associated with MSI-L status (Fisher’s exact test, P > 0.1 in all cases). In case our cutoff for nominal MSI-L was incorrect, we performed an alternative analysis of the same data by comparing the frequencies of MSI by mutation or LOH subgroup; we asked, for example, whether K-ras mutant versus K-ras wild-type non-MSI-H cancers differed in their levels of MSI. However, for K-ras mutation and each of the above sites of LOH, there was no significant tendency for any group to have higher or lower levels of MSI (t test, P > 0.3 in all cases).

We then tested whether the same cancers were consistently classed as nominally MSI-L and MSS by microsatellites from sets 1 and 2 (i.e., one or more markers unstable in a set of 22 markers caused a cancer to be categorized as MSI-L for that set). If MSI-L were a purely stochastic event, then there should be no general tendency to microsatellite slippage and hence no association between classification as MSI-L or MSS in the two sets of markers. We found no significant tendency (Table 2) for a cancer classed as MSI-L by the first 22 markers also to be classed as MSI-L by the second 22 markers (Fisher’s exact test, P > 0.14).

We compared the overall frequency of allelic loss in the nominally MSI-L and MSS groups. A mean of 21.8% (SD, 16.0; median, 20.7%; range, 0–63.9%) of informative markers showed LOH in the MSI-L group, compared with 23.1% (SD, 16.2; median, 20.2%; range, 0–61.5%) in the MSS group. This difference was nonsignificant (t test, P > 0.4). Using Fisher’s exact, x², and t tests (details not shown), we found no significant differences between MSI-L and MSS cancers with regard to clinicopathological variables [age, location in colorectum, Dukes’ stage, and grade (differentiation)].

Finally, additional dinucleotide markers (details available from authors) were typed in 16 of the 24 MSS cancers to determine whether a truly MSS set of cancers existed. Between 2 and 12 microsatellites successfully amplified for each tumor, and 6 of the tumors were found to be MSI+ at one of the additional markers (at least). The 6 positive results occurred in a total of 149 tests (4.0%), lower than, but not significantly different from, the overall mean frequency of instability of 5.0% (x² test, P > 0.4). Notably, when these additional data were used to reclassify the 6 MSS cancers as MSI-L, there was still no significant association between MSI-L status and either K-ras mutation (Fisher’s exact test, P > 0.3) or LOH at any of the sites described above.

In conclusion, we used a set of 44 microsatellite markers to assess the frequency and characteristics of microsatellite instability in 107 colorectal carcinomas. Overall, we found that MSI-H cancers could be distinguished with some confidence from other cancers but found no good evidence for qualitatively different groups within the remaining non-MSI-H cancers. Nominally MSI-L cancers, defined as showing MSI at any of 44 markers tested, were the most frequent group, making up ~68% of all tumors. Nominally MSS cancers were correspondingly less frequent (~22%). The proportion of MSI-L cancers in our set was much higher than in previous studies (4, 5, 8–26), undoubtedly because of the larger number of microsatellites that we used. Our data show that the more markers used, the greater the chance of MSI-L; presumably, random microsatellite slippage (or PCR artifact) can occur in any cell, and the fact that tumors are clonal means that MSI is detectable, when it would not be in normal (polycional) tissue or blood. We found no clinicopathological or molecular differences between nominally MSI-L and MSS cancers.

MSI-H cancers, for which the cutoff lay between 25 and 35% (Fig. 1), were at a frequency (9.4%) similar to that reported by many other studies. Markers provided 40–100% sensitivity to detect MSI-H cancers. DSS1771 and D17S939 were 100% sensitive and also had high specificities (99 and 98%, respectively). BAT26 and MYCL, both in

<table>
<thead>
<tr>
<th>Table 1 Results of simulations of MSI frequency in non-MSI-H cancers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of cancers observed</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9+</td>
</tr>
</tbody>
</table>

Proportion of markers unstable

| ≤0.01 | 24 | 15.61 |
| 0.05  | 36 | 46.64 |
| 0.1   | 25 | 28.54 |
| 0.125 | 7  | 4.20  |
| 0.15  | 0  | 1.36  |
| 0.175 | 3  | 0.48  |
| 0.2   | 0  | 0.13  |
| 0.25  | 1  | 0.03  |
| 0.25  | 1  | 0.03  |
| >0.25 | 0  | 0.01  |


In conclusion, we used a set of 44 microsatellite markers to assess the frequency and characteristics of microsatellite instability in 107 colorectal carcinomas. Overall, we found that MSI-H cancers could be distinguished with some confidence from other cancers but found no good evidence for qualitatively different groups within the remaining non-MSI-H cancers. Nominally MSI-L cancers, defined as showing MSI at any of 44 markers tested, were the most frequent group, making up ~68% of all tumors. Nominally MSS cancers were correspondingly less frequent (~22%). The proportion of MSI-L cancers in our set was much higher than in previous studies (4, 5, 8–26), undoubtedly because of the larger number of microsatellites that we used. Our data show that the more markers used, the greater the chance of MSI-L; presumably, random microsatellite slippage (or PCR artifact) can occur in any cell, and the fact that tumors are clonal means that MSI is detectable, when it would not be in normal (polycional) tissue or blood. We found no clinicopathological or molecular differences between nominally MSI-L and MSS cancers.

MSI-H cancers, for which the cutoff lay between 25 and 35% (Fig. 1), were at a frequency (9.4%) similar to that reported by many other studies. Markers provided 40–100% sensitivity to detect MSI-H cancers. DSS1771 and D17S939 were 100% sensitive and also had high specificities (99 and 98%, respectively). BAT26 and MYCL, both in

<table>
<thead>
<tr>
<th>Table 2 Concordance between first and second 22 markers typed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSI-L</strong></td>
</tr>
<tr>
<td>MSI-L 2nd</td>
</tr>
<tr>
<td>MSS 2nd</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

a MSI-L is defined as any marker with MSI.
common use, also had high sensitivity (both 90%) and specificity (98% and 97%). PPVs of D8S1771, D17S939, BAT26, and MYCL were 90, 83, 82, and 75%, respectively. D9S319 and BAT25 had sensitivities of 100 and 99%, respectively, high PPVs (100 and 88%) but lower specificities (40 and 78%). Of the other widely used markers, D2S123 had sensitivity/specitivity/PPV of 80/97/73%, and D5S346 had values of 60/98/75%. Thus, it might be possible to substitute some of the above markers for D2S123 and D5S346 when undertaking MSI analysis.

In addition to the 10 tumors with MSI-H, our analysis suggested that there is some heterogeneity in terms of MSI among the non-MSI-H colorectal cancers. Although we emphasize that the great majority of non-MSI-H cancers fit within a “normal” category, with a low but definite probability of MSI at any marker, our simulations showed that perhaps as many as 8 tumors were “super-stable” and that 6 of the 12 samples with 10–25% MSI+ markers might come from a “true MSI-L” population. If these conclusions are correct, it is nevertheless almost impossible to assign tumors with any certainty to any of these three non-MSI-H groups. Even if all 44 markers were used, one would expect ~12% of “normal” colorectal cancers to show no instability and >5% to show instability in at least five markers. Evidently, distinguishing “super-stable” or “true-MSI-L” cancers would require typing many hundreds or thousands of microsatellite markers per tumor, which is currently impractical in most diagnostic settings.

In our opinion, it is most likely that our evidence for a small number of “super-stable” and “true MSI-L” cancers results from quantitative variation in the tendency for microsatellite slippage to occur in non-MSI-H colorectal cancers. This variation is likely to be relatively subtle, as evidenced by our failure consistently to define nominally MSI-L and MSS cancers in our first and second sets of microsatellite markers. (It is also likely, although difficult to test without a very much larger sample, that the level of MSI-H varies more than expected.) One possible explanation for the greater than expected variation in MSI-L is that the natural history of tumors varies, such that some have undergone patterns of replication that are associated with greater tendency to microsatellite slippage or greater opportunity to acquire and accumulate such slippages. These differences might be environmental or genetic. For example, if a tumor had grown along a pathway that involved few cell divisions, fewer microsatellite slippages would have occurred, although there would be no qualitative difference between this cancer and others as regards the tendency for microsatellite slippage to occur in each cell division. Alternative explanations for the excess variation in the non-MSI-H cancers include germ-line polymorphisms in DNA mismatch repair genes or somatic mutations in genes that have very subtle effects on the efficiency of the mismatch repair system. Finally, we cannot exclude the possibility that the apparent differences between the non-MSI-H cancers are artifactual, resulting from excess normal tissue (where the part of the cancer assessed by histology was not representative of the whole sample), or from poor quality DNA in some samples (28), or even from misclassification of MSI-H cancers as non-MSI-H.

We conclude that MSI-H in colorectal cancers is a well-established phenomenon with a characterized underlying cause. There is no readily identifiable group of MSI-L cancers, and we found no association between nominal MSI-L or the level of MSI in non-MSI-H cancers and clinicopathological or molecular variables. Nevertheless, there is greater variability than expected in the level of MSI found in non-MSI-H cancers. The reasons for this variation deserve investigation but probably temporally and spatially reflect the “evolutionary history” of a tumor, perhaps including features such as somatic mutational load, germ-line genotype, and microenvironment. Innovative work has shown how MSI-H can be used to trace tumor histories in a minority of colorectal cancers (29). It seems likely that, despite greater technical challenges, MSI at a lower level could be used to trace the evolution of almost every cancer of any site.

Acknowledgments

We are very grateful to the Equipment Park, Imperial Cancer Research Fund for genotyping assistance.

References

18. Gioia, R., Maestri, I., Matteuzzo, M., Santini, A., Ferretti, S., Cavazzoni, L., and Lanza, G. Sporadic colorectal adenocarcinomas with high-frequency microsatellite instabil-


Low-Level Microsatellite Instability Occurs in Most Colorectal Cancers and Is a Nonrandomly Distributed Quantitative Trait

Sarah Halford, Peter Sasieni, Andrew Rowan, et al.


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/62/1/53

Cited articles  This article cites 29 articles, 12 of which you can access for free at: http://cancerres.aacrjournals.org/content/62/1/53.full#ref-list-1

Citing articles  This article has been cited by 18 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/62/1/53.full#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/62/1/53. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.