Epidermal growth factor receptor intron 1 polymorphism and microsatellite instability in sporadic colorectal cancer

SONJA MARINOVIĆ1*, KRISTINA VUKOVIĆ1*, ANITA ŠKRTIĆ2, MIRKO POLJAK3, SARA PETEK1, LARA PETEK1 and SANJA KAPITANOVIĆ1

1Laboratory for Personalized Medicine, Division of Molecular Medicine, Ruder Boskovic Institute; Departments of 2Pathology and 3Surgery, Clinical Hospital Merkur, 10000 Zagreb, Croatia

Received June 18, 2020; Accepted November 9, 2020

DOI: 10.3892/ol.2020.12392

Abstract. Epidermal growth factor receptor (EGFR) expression is commonly upregulated in sporadic colorectal cancer (CRC) and its high expression is associated with poor prognosis in patients with CRC. CA-SSR1 is a dinucleotide CA repeat of the EGFR gene that can modulate EGFR transcription and is a potential target of the mismatch repair machinery in tumours with microsatellite instability (MSI). In the present study, 160 sporadic colon cancer samples were analysed for EGFR CA-SSR1 polymorphism and MSI status. Additionally, EGFR mRNA and protein expression levels in the tumour centre and in the invasive tumour front, compared with those in adjacent normal tissue samples, were evaluated in 80 tumour samples. An inverse association was identified between EGFR mRNA levels and the sum of repeats in both alleles of the CA-SSR1 polymorphism in normal tissues. Changes in CA-SSR1 were detected in the tumour centre as well as in the invasive tumour front and metastases in all MSI high (MSI-H) tumours. Analysis of EGFR expression at the mRNA and protein levels according to MSI status revealed lower EGFR mRNA and protein expression in MSI-H tumours than microsatellite-stable (MSS) tumours. Furthermore, higher EGFR levels in the invasive tumour front compared with in the tumour centre in MSS tumours were identified, suggesting a role of EGFR in tumour progression and higher invasive potential of MSS than MSI-H tumours.

Introduction

Sporadic colorectal cancer (CRC) is the second most common cause of cancer death in developed countries with an estimated 1.8 million new cases and 862,000 deaths in 2018 (1,2). Despite the overall advances in diagnosis and therapy, survival rates for colorectal cancer remain disappointing at approximately 65% depending on the stage. Hence, finding molecular markers that can improve patient diagnosis and treatment selection is necessary.

Epidermal growth factor receptor (EGFR) is a membrane receptor of the receptor tyrosine kinase (ErbBs) family that plays an important role in cell proliferation, survival, differentiation and invasion in sporadic colon cancer (3,4). Several studies have shown that EGFR is overexpressed in approximately 50% of colon tumours (5-7) and there is an increased level of EGFR protein at the invasive tumour front (ITF) in comparison to the tumour centre (8,9). This increased EGFR expression has been related to the presence of tumour budding leading to tumour invasion and other aggressive pathohistological features (10-13). Overexpression of EGFR protein in CRC can be rarely attributed to gene amplification and is more often attributed to polymorphisms in the EGFR gene (14).

Polymorphic regions are excellent candidates as possible prognostic biomarkers for cancer patients. Their main advantage is that they can be easily assessed from blood and normal and tumour tissue and determined by straightforward and well-established methods. One of these regions is the CA dinucleotide repeat polymorphism in intron 1 of the EGFR gene. CA-SSR1 is important due to its close location to the second enhancer (15) which endows it with the ability to influence the expression of the EGFR gene (16-21). Given that CA-SSR1 could modulate EGFR transcription, changes in its sequence could also alter the levels of EGFR protein. Different studies investigated the role of CA repeats and its prognostic implication in various types of cancer and some of them have brought into connection an increasing number of repeats and decreasing levels of both EGFR mRNA and protein expression (22,23). However, the results and its potential predictive impact in CRC remain contrasting and inconclusive (24,25).

Sporadic CRCs can be divided into microsatellite-stable (MSS) tumours and tumours having microsatellite instability (MSI) as a result of failure in the DNA mismatch repair (MMR) system (26). This failure results in changes in the length of microsatellite sequences, potentially also affecting the EGFR CA-SSR1 polymorphism. Even though it is recognized that MSI can affect repeat elements of the EGFR gene,

Correspondence to: Professor Sanja Kapitanović, Laboratory for Personalized Medicine, Division of Molecular Medicine, Ruder Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia
E-mail: kapitan@irb.hr

*Contributed equally

Key words: sporadic colorectal cancer, epidermal growth factor receptor, CA-SSR1 polymorphism, microsatellite instability
and subsequent EGFR expression (27,28), clinical and pathological significance is still not extensively studied.

Therefore, we decided to investigate the effect of the EGFR CA-SSR1 polymorphism on mRNA and protein expression by considering microsatellite status in CRC tumours. Furthermore, we wanted to examine the difference in EGFR mRNA and protein expression between the tumour centre and invasive tumour front in accordance to MSI status to clarify the role of EGFR in CRC tumour progression. Additionally, we performed a correlation analysis between the EGFR expression and CRC clinicopathological characteristics in MSS and MSI-H tumours.

Materials and methods

Study subjects and DNA isolation. Our study included tumour samples from 160 patients diagnosed with sporadic colon adenocarcinoma obtained from the Croatian Tumour Bank (29). Tissue samples were collected from 2013 to 2019 during routine surgery performed in Merkur Clinical Hospital. Fresh tumour samples were stored at -80°C until DNA and RNA extraction. From each patient two samples of tumour were obtained: One corresponding to the tumour centre (T1) and the other corresponding to the invasive tumour front (T2) as well as adjacent normal colon tissue. If metastasis (M) was present, samples were retrieved for the analysis. In MSS tumours 15 and in MSI tumours 2 metastasis samples were obtained. Before use in the study, each specimen was verified by a pathologist (A.Š.).

DNA was extracted from the blood and tumour tissues as well as corresponding normal tissue samples located 15 cm from the tumour edge. DNA extraction was performed using proteinase K digestion and phenol-chloroform extraction (30).

PCR. For MSI analysis paired normal and tumour DNA was analysed for changes in five loci each, using previously published Bethesda panel (26).

The primer sequences used were as follows: D2S123: forward, 5'-AAACAGGATGCTGCTTATT-3' and reverse, 5'-GGCGTTTCCCAATGGGAC-3'; D5S346: forward, 5'-ACTCATTCTATGATAATCGGG-3' and reverse, 5'-AGCACATAAGACAGTATTACCTTT-3'; D17S250: forward, 5'-GAAGAATCAAATAGACAAT-3' and reverse, 5'-GCTGGCCATATATATTTAAAACC-3'; BAT-25a: forward, 5'-TGCTCGCTCAAGAATGTAAGT-3' and reverse, 5'-TCCTTTTTATACTGGTCTC-3'; BAT-26: forward, 5'-CTCGGGTAATCAAGTTTTTAG-3' and reverse, 5'-AACATTCAACATTTTACCACTCC-3'. Samples were considered MSI low (MSI-L) if only one marker was changed and MSI high (MSI-H) if at least two out of five markers showed instability.

MSI was defined when staining intensity was either intermediate or strong; in MSS, samples were considered MSI low category. EGFR expression was defined low when there was no staining (Score 0) or there was a weak positive (light brown) staining (Score 1), and high EGFR expression was defined when staining intensity was either intermediate (Score 2) or strong (Score 3).
Statistical analysis. Statistical analyses were performed using the GraphPad Prism statistical package (GraphPad Software, Inc.). Correlations between the \textit{EGFR} CA-SSR1 genotype and \textit{EGFR} mRNA expression were analysed using Spearman's correlation coefficient and linear regression analysis. The relationship between the sum of repeats in both alleles in the \textit{EGFR} CA-SSR1 polymorphism and \textit{EGFR} protein levels was examined by two-tailed unpaired Student’s t-test and the Mann-Whitney test. Two-way analysis of variance (ANOVA) with Bonferroni correction was used to compare \textit{EGFR} mRNA expression between the groups. For \textit{EGFR} IHC analysis and further correlation with clinicopathological characteristics, contingency table with Fisher’s exact test was used to calculate statistical significance. Overall survival rate was determined by Kaplan-Meier, and statistical differences between groups were calculated with the log-rank test. Data are presented as the mean ± SEM. Values of *P<0.05, **P<0.01 and ***P<0.001 were considered statistically significant.

Results

CA-SSR1 polymorphism of the \textit{EGFR} gene and microsatellite instability. The \textit{EGFR} intron 1 CA repeat polymorphism (CA-SSR1) was genotyped in 160 blood samples from patients with sporadic colon cancer. The number of CA repeats ranged in length from 15 to 21 and homozygous 16/16 CA repeats were the most frequent genotype (24.5%), followed by heterozygous 16/18 (16.6%) and 16/20 (14.4%) CA repeats. The frequencies of 10 most common genotypes (present in 140 patients; 87%) are shown in Fig. 1A.

Furthermore, tumour centre (T1), invasive tumour front (T2) and, if present, corresponding metastasis (M) samples in comparison to corresponding normal tissues (N) were analysed for MSI status and the \textit{EGFR} CA-SSR1 polymorphism. Analysis showed that MSI was present in 33 (20.6%) tumour samples, 13 (8.1%) tumours were MSI-L and 20 of 160 analysed tumours (12.5%) were MSI-H. The remaining 127 (79.4%) tumours were classified as microsatellite stable (MSS) (Table 1).

Table I. MSI status of tumours.

<table>
<thead>
<tr>
<th>Number of markers exhibiting instability</th>
<th>Number (%)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>20 (12.5)</td>
<td>MSI-H</td>
</tr>
<tr>
<td>1</td>
<td>13 (8.1)</td>
<td>MSI-L</td>
</tr>
<tr>
<td>0</td>
<td>127 (79.4)</td>
<td>MSS</td>
</tr>
</tbody>
</table>

MSI, microsatellite instability; MSI-H, microsatellite instability high; MSI-L, microsatellite instability low; MSS, microsatellite stable.

Association of CA-SSR1 with \textit{EGFR} mRNA and protein expression. To determine whether the number of CA-SSR1 CA repeats is associated with changes in \textit{EGFR} transcription and protein expression levels, 80 specimens with the most frequent CA-SSR1 genotypes were analysed for \textit{EGFR} mRNA and protein expression with regards to MSI status.

When we calculated the sum of CA repeats from both alleles per patient, the median sum of all samples present in study was 34 (range 31-42), with 52.2% patients having less than 34 repeats. Therefore, we classified patients as having...
either low (≤34) or high (>34) numbers of CA repeats in both alleles. An inverse correlation was found between a higher sum of CA repeats in the EGFR CA-SSR1 polymorphism and lower EGFR mRNA expression in both MSS (P=0.0047) and MSI-H (P=0.6335; P=0.0036) tumours. (B) Correlation between the sum of CA repeats from both alleles per patient (≤34 or >34) of the CA-SSR1 polymorphism and EGFR immunohistochemical score. Statistical significance was analysed using (A) Spearman’s correlation coefficient and linear regression analysis and (B) two-tailed unpaired Student’s t-test. (C) Representative immunohistochemical images of epidermal growth factor receptor staining. Images represent the tumour staining scores between 0 and 3 (magnification, x100). EGFR, epidermal growth factor receptor; MSS, microsatellite-stable; MSI-H, microsatellite instability high.

**EGFR expression in the tumour centre and at the invasive margin.** Since it is known that EGFR can promote cell migration, invasion, and metastatic dissemination we also evaluated EGFR mRNA and IHC protein levels in the tumour centre (T1) and invasive tumour front (T2) samples.

The analysis showed that EGFR mRNA levels were significantly higher in both T1 (P=0.009) and T2 (P=0.024) samples of MSS tumours than in normal tissues. However, in MSI-H tumours, EGFR mRNA levels were not significantly increased in either T1 or T2 samples in comparison to adjacent normal tissue samples (P>0.999). Both T1 and T2 of MSI-H tumours showed decreased EGFR mRNA levels in comparison to those in MSS tumour, nevertheless, this was not statistically significant. There was no difference in the expression of EGFR mRNA between normal tissues adjacent to MSS and normal tissues adjacent to MSI-H tumours (Fig. 3).

Immunohistochemical analysis further confirmed these results. EGFR protein expression was detected in 6 (35.5%) adjacent normal tissues, 50 (92.6%) T1 tissues and in 39 (97.5%) T2 tissues of MSS tumours as well as in 2 (28.5%) adjacent normal tissues, 12 (66.7%) T1 tissues and 6 (54.5%) T2 tissues of MSI-H tumours.

EGFR protein expression was significantly increased in tumour tissues in comparison to adjacent normal tissues in MSS tumours (P<0.0001). Moreover, the analysis showed a difference in EGFR protein expression between T1 and T2 samples of MSS tumours (P=0.001) (Table II and Fig. 4).
There was also significantly higher EGFR expression in MSS tumours than in MSI-H tumours in both T1 and T2 samples (P=0.040, P=0.001, respectively) (Table II and Fig. 4). However, there was no difference in the expression of EGFR protein in tumour tissues in comparison to adjacent normal tissues in MSI-H tumours or between normal tissues adjacent to MSS and normal tissues adjacent to MSI-H tumours (Table II and Figs. 4, S1 and S2). Immunohistochemical staining showed similar EGFR protein expression in T2 and liver or lymph node metastasis samples (Figs. 4, S1 and S2).

There was no correlation between EGFR protein expression and tumour size, histological grade, Dukes' stage of tumours (P>0.05) (Table III) in either MSS or MSI-H tumours. The survival of the two MSS tumour subgroups, based on EGFR immunohistochemical score, had no statistically significant difference (P=0.717) (Table IV).

Discussion

EGFR receptor overexpression is found in a wide range of cancers, including CRC, and it is associated with an aggressive tumour phenotype and poor prognosis (33). Nonetheless, the mechanisms regulating the levels of EGFR expression in cancer have not been fully characterized. The results of our study show that the CA-SSR1 of the \( \text{EGFR} \) gene is altered in MSI-H sporadic colorectal tumours and that it has an effect on EGFR expression at the mRNA level but not at the protein level in both MSS and MSI-H tumours. Moreover, we demonstrated that MSI-H tumours have lower \( \text{EGFR} \) mRNA and protein levels in the tumour centre and invasive tumour front than MSS tumours. Additionally, we confirmed that in MSS tumours EGFR expression is higher in the invasive tumour front than in the tumour centre.

In recent years, \( \text{EGFR} \) gene intron 1 length has been considered a factor affecting expression through modification of \( \text{EGFR} \) transcription. In this regard, it has been suggested that CA-SSR1 has an effect on gene transcription. This hypothesis was tested in several cell lines (20,22,23) as well as in head and neck, lung, pancreas, colon and mammary tumours (17-20,34,35), but the results were inconsistent and somewhat contradictory, most likely due to the limited number of analysed samples. We found the same distribution of CA-SSR1 alleles in normal tissues as previously reported (18,28,36), and the most common genotype was 16/16.

The distribution of CA-SSR1 alleles and MSI status, were analysed in normal and tumour tissues (tumour centre and invasive tumour front) as well as in corresponding metastasis samples when available. Our results showed for the first time that all MSI-H tumours showed instability in the CA-SSR1 polymorphism. However, the genotype varied from shortening to elongation, with presence of one additional allele to several of them, regardless of the length of the alleles present in normal tissues. CA-SSR1 instability was present in the tumour centre as well as in invasive tumour front and metastasis samples. To further characterize the possible effect of the CA-SSR1 polymorphism of the \( \text{EGFR} \) gene, and its changes in MSI-H tumours, we measured \( \text{EGFR} \) mRNA and protein levels. Since there is a lack of consensus regarding cut off values defining shorter versus longer CA repeats (37-40), we decided to take the median sum of the CA repeats from both alleles.

<table>
<thead>
<tr>
<th>MSI status</th>
<th>MSS N</th>
<th>MSS T1</th>
<th>MSS T2</th>
<th>MSI-H N</th>
<th>MSI-H T1</th>
<th>MSI-H T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>11 (64.7)</td>
<td>4 (7.4)</td>
<td>5 (29.4)</td>
<td>11 (64.7)</td>
<td>4 (7.4)</td>
<td>5 (29.4)</td>
</tr>
<tr>
<td>T1</td>
<td>2 (5.0)</td>
<td>3 (5.0)</td>
<td>0 (0)</td>
<td>2 (5.0)</td>
<td>3 (5.0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>T2</td>
<td>0 (0)</td>
<td>1 (2.5)</td>
<td>2 (4.5)</td>
<td>0 (0)</td>
<td>1 (2.5)</td>
<td>2 (4.5)</td>
</tr>
</tbody>
</table>

Letters indicate the difference between MSS N and MSS T1, MSS N and MSS T2, MSI-H N and MSI-H T1, MSI-H N and MSI-H T2, MSS N and MSI-H N and MSI-H T1, MSS N and MSI-H T2, MSI-H N and MSI-H T1, MSI-H N and MSI-H T2, MSS N and MSI-H N and MSI-H T1, MSS N and MSI-H T2, MSI-H N and MSI-H T1, MSI-H N and MSI-H T2. P-values were obtained by Fisher's exact test. N, adjacent normal tissue; T1 tumour centre; T2 invasive tumour front; \( \text{EGFR} \), epidermal growth factor receptor; MSI, microsatellite instability; MSS, microsatellite stable; MSI-H, microsatellite instability-high.
Table III. Clinicopathological characteristics of 147 patients with sporadic adenocarcinoma stratified by MSI status.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>MSS, n=127 (86.4%)</th>
<th>MSI-H, n=20 (13.6%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low EGFR expression, n (%)</td>
<td>High EGFR expression, n (%)</td>
<td>Low EGFR expression, n (%)</td>
</tr>
<tr>
<td>Tumor size, cm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>11 (42.3)</td>
<td>15 (57.7)</td>
<td>3 (50.0)</td>
</tr>
<tr>
<td>≥5</td>
<td>7 (26.9)</td>
<td>19 (73.1)</td>
<td>8 (61.5)</td>
</tr>
<tr>
<td>Histological grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well (1)</td>
<td>8 (40.0)</td>
<td>12 (60.0)</td>
<td>2 (50.0)</td>
</tr>
<tr>
<td>Moderate and poor (2 and 3)</td>
<td>9 (74.3)</td>
<td>26 (25.7)</td>
<td>8 (53.3)</td>
</tr>
<tr>
<td>Dukes' stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A+B</td>
<td>5 (21.7)</td>
<td>18 (78.3)</td>
<td>7 (70.0)</td>
</tr>
<tr>
<td>C+D</td>
<td>9 (30.0)</td>
<td>21 (70.0)</td>
<td>4 (44.4)</td>
</tr>
</tbody>
</table>

Low EGFR, immunohistochemical score 0 and 1; high EGFR, immunohistochemical score 2 and 3. MSI, microsatellite instability; MSS, microsatellite stable; MSI-H, microsatellite instability-high; EGFR, epidermal growth factor receptor.
per patient from all samples present in our study. The results showed that \textit{EGFR} mRNA expression levels declined with an increasing sum of CA-SSR1 alleles which is in line with other papers (22,23), possibly due to changes in DNA secondary
structure. However, this effect was absent at the protein level, an effect seen also by McKay et al (41) and Buisine et al (28). This could be explained by posttranscriptional regulation via miRNAs (42,43) and possibly other regulatory mechanisms on protein levels like EGFR dimerization, internalization, degradation or recycling (44–46). Interestingly, MSI tumours had lower levels of EGFR on both mRNA and protein levels in comparison to MSS tumours, however, our results show that this is not mediated by EGFR CA-SSR1 polymorphism but via some other mechanism that should be further investigated.

To highlight the potential relationship between EGFR overexpression and tumour invasion, we analysed tumour centre and invasive tumour front samples from each patient. Our results showed for the first time that EGFR mRNA and protein levels were lower in both the tumour centre and invasive tumour front of MSI-H tumours than in such samples from MSS tumours. This is in accordance with several studies that showed a smaller number of metastatic lymph nodes in MSI-H patients than MSS patients (9,47), and it could also clarify why MSI-H metastatic CRC is rare (48). In addition to increased host immunity (49,50), a decrease in EGFR expression in MSI-H CRC could partially explain why these cancers are less aggressive and have a more promising prognosis. Additionally, in MSS tumours, there was a higher expression of EGFR at the invasive tumour front in comparison to the tumour centre which confirms the putative role of EGFR in tumour invasive-ness and the development of metastasis in CRC. This indicates that even though changes in non-coding regions are usually background effects, intron 1 polymorphism could play a role since it is located near the region that has a regulatory function in the EGFR gene (51).

Correlation between EGFR expression and colon cancer staging along with histological grade and tumour size is still at issue (8). In our study, we found no association between tumour size, histological grade, or Dukes’ stage with a level of EGFR expression either in MSS or in MSI-H tumours. Several studies addressed the possible relationship between EGFR overexpression and tumour stage and/or histological grade. McKay et al (41) reported a significant association between the histological grade and EGFR, however, they showed no correlation between EGFR expression and the Dukes’ stage. On the other hand, Theodoropoulos et al (52) reported an association between advanced tumour stage and high EGFR expression, excluding correlation with tumour grade. Whereas, Del Carmen et al (42) showed no correlation with either tumour size, TNM stage or tumour differentiation altogether leading to the conclusion that EGFR remains a controversial prognostic factor. Even though survival analysis for MSS tumours showed no statistical significance, it should be taken in consideration that we had a very small dataset for MSS tumours and a complete lack of survival data for MSI-H tumours. Therefore, this is our research should be expanded in the future.

In conclusion, the present study identified the CA-SSR polymorphism in intron 1 of the EGFR gene as a potential new CRC marker that is exclusively altered in MSI-H colorectal tumours. Additionally, we showed higher EGFR expression in MSS than in MSI-H tumours. The highest EGFR expression found in invasive tumour front of MSS tumours suggests more important role of EGFR in MSS tumours progression than in MSI-H tumours where MSI is a dominant molecular genetic change. Furthermore, instability in EGFR CA-SSR1 polymorphism that correlates with decreased EGFR expression suggests that it could serve as an indicator of sporadic MSI-H CRC progression.

Acknowledgements

Not applicable.

Funding

The present study was supported by the Croatian Science Foundation (grant no. HRZZ-IP-2016-06-1430).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors’ contributions

SK is the PI on the project HRZZ-IP-2016-06-1430 and designed the study. SM, KV, SP and LP performed the experiments and the statistical analyses. AS and MP collected the samples and analysed the immunohistological data. SK and SM wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Written informed consent was obtained from all patients included in the present study. The present study was approved by the Ethics Committee of Merkur Clinical Hospital, Zagreb (Zagreb, Croatia) and Medical School, University of Zagreb (Zagreb, Croatia) and was performed in accordance with the ethical standards of the Helsinki Declaration.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


49. Xiao Y and Freeman GF: The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy. Cancer Discov 5: 16-18, 2015.


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.