3'-UTR poly(T/U) repeat of *EWSR1* is altered in microsatellite unstable colorectal cancer with nearly perfect sensitivity

Johanna Kondelin^{1,2}, Sari Tuupanen^{1,2}, Alexandra E. Gylfe^{1,2}, Mervi Aavikko^{1,2}, Laura Renkonen-Sinisalo³, Heikki Järvinen³, Jan Böhm⁴, Jukka-Pekka Mecklin⁵, Claus L. Andersen⁶, Pia Vahteristo^{1,2}, Esa Pitkänen^{1,2}, Lauri A. Aaltonen^{1,2}

1.Medicum/Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

2. Genome-Scale Biology Research Program, Research Programs Unit, University of Helsinki, Haartmaninkatu 8, PO Box 63, 00014 Helsinki, Finland

3.Department of Surgery, Helsinki University Central Hospital, Haartmaninkatu 4, 00029 Helsinki, Finland

4.Department of Pathology, Jyväskylä Central Hospital, University of Eastern Finland, Keskussairaalantie 19, 40620 Jyväskylä, Finland

5.Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Keskussairaalantie 19, 40620 Jyväskylä, Finland

6.Department for Molecular Medicine (MOMA), Aarhus University Hospital, Brendstrupgårdsvej 21, 8200 Aarhus, Denmark

Abstract

Approximately 15% of colorectal cancers exhibit instability of short nucleotide repeat regions, microsatellites. These tumors display a unique clinicopathologic profile and the microsatellite instability status is increasingly used to guide clinical management as it is known to predict better prognosis as well as resistance to certain chemotherapeutics. A panel of five repeats determined by the National Cancer Institute, the Bethesda panel, is currently the standard for determining the microsatellite instability status in colorectal cancer. Recently, a quasimonomorphic mononucleotide repeat 16T/U at the 3' untranslated region of the *Ewing* sarcoma breakpoint region 1 gene was reported to show perfect sensitivity and specificity in detecting mismatch repair deficient colorectal, endometrial, and gastric cancers in two independent populations. To confirm this finding, we replicated the analysis in 213 microsatellite unstable colorectal cancers from two independent populations, 148 microsatellite stable colorectal cancers, and the respective normal samples by PCR and fragment analysis. The repeat showed nearly perfect sensitivity for microsatellite unstable colorectal cancer as it was altered in 212 of the 213 microsatellite unstable (99.5 %) and none of the microsatellite stable colorectal tumors. This repeat thus represents the first potential single marker for detecting microsatellite instability.

Keywords

Colorectal cancer Microstellite unstable colorectal cancer Microsatellite instability marker *EWSR1*

Introduction

Colorectal cancer (CRC) is the third most common cancer in Western countries with a mortality rate of nearly 50 % [1]. Approximately 15 % of CRCs display microsatellite instability (MSI) as a result of defected mismatch repair (MMR) machinery, which leads to excessive accumulation of small insertions and deletions [2]. These most often target short repeat regions of the genome, microsatellites.

Over 90 % of MSI CRCs are sporadic [3]. Lynch syndrome, a hereditary cancer predisposition syndrome arising due to germline mutations in DNA MMR genes (*MLH1, MSH2, MSH6,* and *PMS2*) accounts for approximately 2-7 % of MSI CRCs [4]. MSI is also observed in 10–20 % of sporadic endometrial and gastric cancers [5].

MSI CRCs are known to arise through a distinct genetic pathway as compared to microsatellite stable (MSS) CRCs [2]. They also display a unique clinicopathologic profile, and thus the MSI status is increasingly used to guide clinical management. A large number of genetic loci recurrently affected by MSI have been reported [6]. Currently, the analysis of five microsatellite markers, the Bethesda panel, is the standard method for determining MSI in CRC according to the instructions of the National Cancer Institute Workshop in 1997 [7].

Recently, a quasimonomorphic mononucleotide T/U16 (EWST16) repeat at the 3' untranslated region (UTR) of *Ewing sarcoma breakpoint region 1 (EWSR1*) was reported by Kishore et al. to show perfect sensitivity and specificity in detecting MMR deficient cancers in two independent populations [8]. The first sample set consisted of 85 Swiss Lynch syndrome-related cancers and 113 sporadic CRCs, and the second set of eight Finnish Lynch syndrome-related cancers, 50 sporadic CRCs as well as 64 sporadic gastric cancers. The locus does not affect the protein coding sequence of *EWSR1* but was shown to associate with changes in the expression and subcellular localization of the encoded protein.

Here, we analyzed the EWS16T locus in 213 MSI (157 Finnish and 56 Danish) 148 MSS (Finnish) CRCs, and corresponding normal samples. The repeat showed nearly perfect sensitivity for MSI CRC: it was mutated in 212 out of 213 (99.5 %) MSI tumors while all the 148 MSS tumors were wild type. Our results are similar to those of Kishore and co-workers and support the utility of the repeat as an effective marker for detecting MSI status.

Materials and methods

Study subjects

A series of 157 MSI CRCs and 148 MSS CRCs were derived from a previously characterized population-based series of 1042 Finnish CRCs and an additional series of approximately 1000 unselected CRCs [9–11]. All tumor DNAs were extracted from histologically evaluated fresh-frozen tissue. Thirty one of the tumors were from Lynch syndrome patients. Normal tissue samples were available from all patients and DNA was extracted either from blood or normal colonic epithelium distant from the site of the tumor. Additional 56 Danish MSI CRCs and their corresponding normal samples were also available for analysis. The tumor DNA of these samples was extracted from gross dissected cancer tissue with the Genetra Puregene DNA purification kit (Qiagen, Limburg, The Netherlands).

MSI status determination

For the samples derived from the set of 1042 CRCs, the MSI status was determined by radioactive labeling techniques, fluorescence-based PCR methods or fragment analysis in previous studies between 1993 and 1999 [9,10].

In the radioactive labeling techniques, seven markers (D5S404, D17S787, D5S346, D1S216, D11S904, D10S197, and TP53) were analyzed by two reviewers. A sample was called MSI if 2/7 markers showed instability. If none of the markers showed instability, the sample was called MSS, as long as at least 5/7 markers were successfully analyzed. If 1/7 markers showed instability, more markers (DCC, D13S175, D7S519, D2OS100, D15S120, D2S136, and D14S79) were analyzed so that in total at least ten markers were reviewed. If at least one of the extra markers showed instability, the sample was called MSI. If none of the markers exhibited instability, the sample was called MSS.

In cases where a fluorescence-based PCR method was used, 16 markers (D8S254, MYC, NM23, D5S346, TP53, D1S228, D8S261, D7S496, D8S137, DCC, D7S501, MCC, D5S318, D1S507, D19S394, and RB1) were tested for. If at least 30 % of the alleles exhibited instability, the sample was called MSI. Later, two markers (BAT26 and TGFBRII) were used. Both markers were evaluated by two independent reviewers. If BAT26 showed deletions, the result was compared to that of the normal sample to ensure the change was somatic in origin.

For the samples derived from the latter set of 1000 CRCs, the Bethesda panel of five markers (BAT25, BAT26, D5S346, D17S250, and D2S123) was analyzed by fragment analysis [7]. If at least 2/5 markers showed instability, the sample was called MSI.

For the 56 Danish MSI CRCs and corresponding normal samples, MSI status was determined by pentaplex polymerase chain reaction with five quasimonomorphic mononucleotide repeats (BAT25, BAT26, and three poly(T) tract segments of the following genes: *SLC7A8* (NR-21, 21T), *transmembrane precursor protein B5* (NR-22, 22T), and *zinc finger-2* (*NR-24, 24T*) [12]. If at least two markers showed instability, the sample was called MSI.

In previous studies, if a sample derived from the set of 1042 CRCs showed MSI, Sanger sequencing of *MLH1* and *MSH2* was performed [9,10]. For the latter sample set, MMR gene mutation data was obtained from diagnostic laboratories [11].

Fragment analysis

In this study, the 3'-UTR poly T tract in EWSR1, (g.2608 2624 containing a 16T repeat, GRCh38, www.ensembl.org) amplified by PCR with the following primers: 5'-AATGTTCATGGTTGTGATGT-3' (forward FAM-labeled) and 5'-GAAGGATGACTCTTTATAA-3' (reverse) [8]. The PCR contained mix 1.5 µl of 1 × PCR-buffer (Applied Biosystems, Brachburg, NJ, USA), 0.2 µl of deoxynucleotide triphosphate (dNTP) (FINNZYMES, Espoo, Finland), ® 0.6 µl of forward and reverse primer (SIGMA, GENOSYS, Helsinki, Finland), 0.15 µl of DNA polymerase enzyme AmpliTaqGold[®] (Applied Biosystems, Brachburg, NJ, USA), 9.95 μl of water, and

enzyme AmpliTaqGold (Applied Biosystems, Brachburg, NJ, USA), 9.95 µl of water, and 25 ng of DNA extracted from fresh-frozen colorectal tumor. The PCR program consisted of five steps: (1) denaturing at 95 °C for 10 min, (2) denaturing at 95 °C for 30 s, (3) annealing at

55 °C for 1 min 15 s, (4) extension at 72 °C for 1 min, (5) extension at 72 °C for 30 min. Steps 2–4 were cycled through 35 times.

Fragment analysis was performed at the Institute for Molecular Medicine Finland (FIMM) Technology Centre. The fragments were run in an ABI3730xl electrophoresis run with a GeneScanTM-500 LIZ^{R} size standard (Applied Biosystems). All methods were performed according to manufacturers' instructions.

The fragment analysis graphs were analyzed by GeneMarker (SoftGenetics, State College, PA, version 1.4) software. Each tumor graph was compared to its corresponding normal graph to ensure the somatic origin of the observed changes. A length difference of one or more nucleotides between the tumor and normal graphs was considered significant.

Ethics approval

The study was approved by the Ethics Committee of the Hospital district of Helsinki and Uusimaa. All samples were derived either after an informed consent signed by the patient or authorization by the National Supervisory Authority for Welfare and Health.

Results

In this study, we analyzed the quasimonomorphic EWS16T locus in 157 MSI and 148 MSS Finnish CRCs, in 56 Danish MSI CRCs, and in their corresponding normal samples. Assessment of the EWS16T repeat length with fragment analysis revealed that from the initial set of 157 MSI CRCs, 156 tumors (99.4 %) showed shortening of the repeat length and only one tumor was wild type for the repeat (Table 1). Of the 157 tumors, 31 were from patients with Lynch syndrome. All the 148 MSS tumors were wild type for the locus.

In the additional set of 56 Danish MSI CRCs, all tumors showed alterations in the repeat length; one tumor exhibited an insertion of one base whereas the other 55 tumors showed contraction of the repeat. Therefore, similar to the previous study and as is typical for MSI, instability was driven mostly by deletion rather than insertion [8].

For the one MSI CRC that displayed two wild type alleles for the EWS16T tract, the Bethesda panel was redone to ensure the MSI status of the sample. The tumor exhibited instability in four of the five markers: BAT25, BAT26, D17S250, and D2S123. Exome sequencing data was available for this tumor, and it was scrutinized for mutations elsewhere in *EWSR1* with negative results.

Discussion

EWS16T, a mononucleotide tract consisting of 16 thymines in the 3'-UTR of the *EWSR1* gene, was recently reported to represent a novel, quasimonomorphic MSI target gene locus, which identifies both hereditary and sporadic MSI colorectal, gastric, and endometrial cancers with perfect sensitivity and specificity [8]. In our study, the repeat was analyzed by fragment analysis in altogether 213 MSI (157 Finnish and 56 Danish) and 148 MSS samples. The results showed nearly perfect sensitivity for MSI CRC both in the Finnish and Danish samples; the repeat was mutated in 156 out of the 157 Finnish and in all the 56 Danish MSI CRCs. All the 148 MSS tumors tested were wild type for the locus. All the 156 mutations seen in the Finnish

MSI CRCs were deletions and of the 56 mutated Danish MSI CRCs all but one showed a deletion. Our results thus support the previously reported finding [8].

The MSI status of the Finnish tumor that was wild type for the EWS16T locus was confirmed as it showed instability in four of the five Bethesda panel markers. According to the pathologist's evaluation, the tumor was a small adenocarcinoma classified as T2N0M0. One could speculate that perhaps the tumor was still in such early stage that EWS16T had simply not acquired mutations yet. On the other hand, instability in four out of the five markers of the Bethesda panel suggests that the tumor was already rather unstable. Exome sequencing data was available for this tumor, and it was scrutinized for mutations elsewhere in EWSR1 with negative results.

A large number of genetic loci have been shown to be affected by MSI but the possible selective advantage of most such mutations is unclear [6]. EWSR1 is located at chromosome 22q12 and it encodes a nuclear protein consisting of 656 amino acids [13]. It is thought to be involved in a variety of cellular functions such as mitotic spindle formation, microtubule stabilization as well as DNA repair. The gene is also known to be involved in the development of Ewing sarcoma, the second most common bone malignancy in adolescents and young adults [14]. In 85 % of Ewing sarcomas, EWSR1 has fused with FLI1 to form EWSR1-FLI1 that results from a translocation between chromosomes 22 and 11. Later, fusions of EWSR1 have been reported in many other sarcoma types such as myxoid liposarcomas and extraskeletal myxoid chondrosarcomas [13]. The deletions and insertions observed in the 16T/U-repeat occur in the 3'-UTR of *EWSR1* and therefore do not alter the coding sequence of the gene. The gene does, however, undergo alternative polyadenylation resulting in two transcripts of different length [15]. In the study by Kishore et al. higher expression for the longer isoform was observed in MSI CRC cell lines [8]. They also reported significant down regulation of protein levels associated with EWS16T tract deletions. In an immunohistochemical examination, Lynch syndrome patients were, interestingly, shown to exhibit diffuse cytoplasmic expression of the protein, whereas in sporadic MSI cancers the protein was expressed exclusively in the nucleus. Therefore, this protein was thought to play a role in the tumorigenesis of MSI CRC.

Detection of MSI has important clinical implications as it can be used in guiding prognosis and therapeutic choices. MSI CRC has been linked to better prognosis as well as resistance to certain chemotherapeutics [2]. Also diagnosis of MSI is critical for identifying patients with Lynch syndrome that should potentially receive genetic counseling as well as clinical monitoring. Since 1992 when MSI was first described, a variety of methods have been established for determining the MSI status of tumors [2]. Several studies have shown different methods to be equally precise in identifying MSI [16,17]. Immunohistochemistry detects absence of protein expression and has been the first widely-used method for identifying MSI in tumors as MSI CRCs are known to show absence of protein expression in proteins of the MMR machinery such as MLH1 and MSH2 [17,18]. Hundreds of primer sequences have been published for amplification of different mono-, di-, tri-, and tetranucleotide repeats by PCR and several different methods have been used for their visualization [12,16]. In 1997, the National Cancer Institute hosted "The International Workshop on MSI and RER Phenotypes in Cancer Detection and Familial Predisposition" to review the field, and a panel of five markers, the so called Bethesda panel, was determined [7]. The panel consists of two mononucleotide markers (BAT25 and BAT26), and three dinucleotide markers (D5S346, D17S250, and D2S123) and is currently the most used method for DNA-based MMR determination in clinical use [19].

Next generation sequencing (NGS) technologies are increasingly used in cancer genetics studies as well as in clinical diagnostics. Several recent studies have involved successful sequencing of tumors exhibiting MSI, and NGS has been reported to have high accuracy in detection of mutations targeting microsatellites [20–23]. Two studies have evaluated NGS to have high accuracy in MSI status determination [24,25]. NGS methods, however, are, costly and therefore not suitable for routine diagnostics at present. Due to the high accuracy of the EWS16T marker in determining MSI in both our study and in the study by Kishore et al. it seems that the marker might have the potential to replace the Bethesda panel in MSI diagnostics in routine practice [8]. Utilizing only one marker instead of a panel of five markers would reduce the costs of MSI diagnostics as well as manual work in the laboratory. A subsequent study comparing the sensitivity and specificity of the Beteshda panel and EWS16T in an extensive, uncharacterized set of CRCs would finally confirm the accuracy of this marker before adopting it in routine clinical practice.

Acknowledgments

The authors thank Sini Nieminen, Sirpa Soisalo, Marjo Rajalaakso, Mairi Kuris, Inga-Lill Svedberg, Iina Vuoristo, and Alison Ollikainen for excellent technical assistance. This study was supported by Grants from the Academy of Finland (Finnish Center of Excellence Program 2012–2017), The Finnish Cancer Society, and the Sigrid Juselius Foundation, and SYSCOL (an EU FP7 Collaborative Project). The Doctoral Programme in Biomedicine in the Doctoral School of Health Sciences at University of Helsinki is thanked for funding.

Conflicts of interest The authors have no conflicts of interest to disclose.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clin 61:69–90. doi:10.3322/caac.20107

2. Boland CR, Goel A (2010) Microsatellite instability in colorectal cancer. Gastroenterology 138:2073–2087. doi:10.1053/j.gastro.2009.12.064

3. Bogaert J, Prenen H (2014) Molecular genetics of colorectal cancer. Ann Gastroenterol 27:9–14

4. Abdel-Rahman WM, Mecklin JP, Peltomaki P (2006) The genetics of HNPCC: application to diagnosis and screening. Crit Rev Oncol Hematol 58:208-220. doi: S1040-8428(05)00225-8

5. Hamelin R, Chalastanis A, Colas C, El Bchiri J, Mercier D, Schreurs AS, Simon V, Svrcek M, Zaanan A, Borie C, Buhard O, Capel E, Zouali H, Praz F, Muleris M, Flejou JF, Duval A (2008) Clinical and molecular consequences of microsatellite instability in human cancers. Bull Cancer 95:121–132. doi:10.1684/bdc.2008.0571

6. Alhopuro P, Sammalkorpi H, Niittymaki I, Bistrom M, Raitila A, Saharinen J, Nousiainen K, Lehtonen HJ, Heliovaara E, Puhakka J, Tuupanen S, Sousa S, Seruca R, Ferreira AM, Hofstra RM, Mecklin JP, Jarvinen H, Ristimaki A, Orntoft TF, Hautaniemi S, Arango D, Karhu A, Aaltonen LA (2012) Candidate driver genes in microsatellite-unstable colorectal cancer. Int J Cancer 130:1558–1566. doi:10.1002/ijc.26167

7. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S (1998) 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 58:5248–5257

8. Kishore S, Piscuoglio S, Kovac MB, Gylling A, Wenzel F, Trapani F, Altermatt HJ, Mele V, Marra G, Peltomaki P, Terracciano L, Zavolan M, Heinimann K (2014) 3'-UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair-deficient cancers. Cancer Res 74:224–234. doi:10.1158/0008-5472.CAN-13-2100

9. Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, Chadwick RB, Kaariainen H, Eskelinen M, Jarvinen H, Mecklin JP, de la Chapelle A (1998) Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 338:1481–1487. doi:10.1056/NEJM199805213382101

10. Salovaara R, Loukola A, Kristo P, Kaariainen H, Ahtola H, Eskelinen M, Harkonen N, Julkunen R, Kangas E, Ojala S, Tulikoura J, Valkamo E, Jarvinen H, Mecklin JP, Aaltonen LA, de la Chapelle A (2000) Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol 18:2193–2200

11. Tanskanen T, Gylfe AE, Katainen R, Taipale M, Renkonen-Sinisalo L, Mecklin JP, Jarvinen H, Tuupanen S, Kilpivaara O, Vahteristo P, Aaltonen LA (2013) Exome sequencing in diagnostic evaluation of colorectal cancer predisposition in young patients. Scand J Gastroenterol 48:672–678. doi:10.3109/00365521.2013.783102

12. Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, Seruca R, Iacopetta B, Hamelin R (2002) Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology 123:1804–1811. doi:10.1053/gast.2002.37070

13. Fisher C (2014) The diversity of soft tissue tumours with EWSR1 gene rearrangements: a review. Histopathology 64:134–150. doi:10.1111/his.12269

14. Ludwig JA (2008) Ewing sarcoma: historical perspectives, current state-of-the-art, and opportunities for targeted therapy in the future. Curr Opin Oncol 20:412–418. doi:10.1097/CCO.0b013e328303ba1d

15. Martin G, Gruber AR, Keller W, Zavolan M (2012) Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. Cell Rep 1:753–763. doi:10.1016/j.celrep.2012.05.003

16. Bocker T, Diermann J, Friedl W, Gebert J, Holinski-Feder E, Karner- Hanusch J, von Knebel-Doeberitz M, Koelble K, Moeslein G, Schackert HK, Wirtz HC, Fishel R, Ruschoff J (1997) Microsatellite instability analysis: a multicenter study for reliability and quality control. Cancer Res 57:4739–4743

17. Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R, Ruschoff J (1997) Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. Cancer Res 57:4749–4756

18. Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, McDonnell SK, Schaid DJ, Vockley CW, Michels VV, Farr GH Jr, O'Connell MJ (1998) Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res 58:1713–1718

19. de la Chapelle A, Hampel H (2010) Clinical relevance of microsatellite instability in colorectal cancer. J Clin Oncol 28:3380–3387. doi:10.1200/JCO.2009.27.0652

20. McIver LJ, Fonville NC, Karunasena E, Garner HR (2014) Microsatellite genotyping reveals a signature in breast cancer exomes. Breast Cancer Res Treat 145:791–798. doi:10.1007/s10549-014-2908-8

21. Kim TM, Laird PW, Park PJ (2013) The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell 155:858–868. doi:10.1016/j.cell.2013.10.015

22. Yoon K, Lee S, Han TS, Moon SY, Yun SM, Kong SH, Jho S, Choe J, Yu J, Lee HJ, Park JH, Kim HM, Lee SY, Park J, Kim WH, Bhak J, Yang HK, Kim SJ (2013) Comprehensive genome- and transcriptome-wide analyses of mutations associated with microsatellite instability in Korean gastric cancers. Genome Res 23:1109–1117. doi:10.1101/gr.145706.112

23. Gan C, Love C, Beshay V, Macrae F, Fox S, Waring P, Taylor G (2015) Applicability of next generation sequencing technology in microsatellite instability testing. Genes (Basel) 6:46–59. doi:10.3390/genes6010046

24. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC (2014) Microsatellite instability detection by next generation sequencing. Clin Chem 60:1192–1199. doi:10.1373/clinchem.2014.223677

25. Lu Y, Soong TD, Elemento O (2013) A novel approach for characterizing microsatellite instability in cancer cells. PLoS ONE 8:e63056. doi:10.1371/journal.pone.0063056

			EWES16T tract status		
Samples analyzed	Number	MSI status	wt n (%)	Contraction n (%)	Expansion n (%)
MSI colorectal cancers	213	MSI-high	1 (0.5)	212 (99.5)	1 (0.5)
Finnish	157	MSI-high	1 (0.6)	157 (100)	0 (0)
Danish	56	MSI-high	0 (0)	55 (98.2)	1 (1.7)
MSS colorectal cancers					
Finnish	148	MSS	148	0	0

Table 1. EWS16T instability in 213 microsatellite unstable colorectal cancers and 148 microsatellite stable colorectal cancers