Intrinsic structural variation of the complex microsatellite marker MYCL1 in Finnish and Somali populations and its relevance to gastrointestinal tumors

Hanna Vauhkonen and Antti Sajantila

Department of Forensic Medicine, University of Helsinki, Helsinki, Finland.

Abstract

The structurally complex MYCL1 microsatellite marker is often used to determine microsatellite instability in colorectal cancers but the allelic variation of this marker has remained largely uncharacterized in both populations and in cancers. Our study describes the allelic distributions of MYCL1 in Finnish (n = 117) and Somali population samples (n = 61) of non-related individuals and compares this distribution with the instability pattern obtained from 61 gastrointestinal tumors.

Key words: colorectal cancers, complex microsatellite, gastrointestinal tumors, microsatellite instability, Finnish and Somali populations.

Received: September 20, 2005; Accepted: February 14, 2006.

The MYCL1 microsatellite is a polymorphic Alu variable poly (A) marker located on chromosome 1p34.4 was previously known as HY-TM1 and used in association studies on infantile neuronal ceroid lipofuscinosis (INCL) (Mäkelä et al. 1992, Vesa et al. 1993) but the instability of this marker in microsatellite unstable colorectal cancers was subsequently discovered and it was re-named MYCL1 (Dietmaier et al. 1997, Boland et al. 1998). Microsatellite instability (MSI) is characterized by the emergence of new alleles not present in normal tissue and, depending on the proportion of affected loci, can be phenotypically categorized as high frequency MSI (MSI-H) or low frequency MSI (MSI-L) (Boland et al. 1998). Although it is known that MSI-H is due to malfunction of the mismatch repair (MMR) system (Boland et al. 1998, Umar et al. 2004) the mechanism of MSI-L and whether or not this type of allele should be separated from the microsatellite stable (MSS) phenotype is still unclear (Boland et al. 1998, Jass et al. 2001, Laiho et al. 2002, Halford et al. 2003, Umar et al. 2004). The mechanism producing germ-line microsatellite mutations (i.e. DNA polymerase slippage during replication) has been suggested as being responsible for the production of new alleles in both MSI-H and MSI-L tumours (Di Rienzo et al. 1998, Sturzeneker et al. 2000, Bacon et al. 2000).

Five microsatellites have been suggested for categorizing MSI, with five additional markers (including MYCL1) being recommended for confirmation of ambiguous results (Boland et al. 1998, Umar et al. 2004). However, due to structural differences different alleles might show differential susceptibility for alteration (Bacon et al. 2000) and thus a genotype with stable alleles may give a false MSS result. In addition, MYCL1 has also been reported to serve as a sensitive indicator of MSI-L (Dietmeier et al. 1997, Lino et al. 1999, Jass et al. 2001).

The MYCL1 marker has been used to study tumor-related MSI since the discovery of its usefulness in the detection of MSI colorectal cancers (Young et al. 1995, Dietmaier et al. 1997, Lino et al. 1999, Perinchery et al. 2000, Jass et al. 2001, Umar et al. 2004) and loss of heterozygosity at the MYCL1 locus has also been reported as being clinically relevant in colorectal cancer (Kambara et al. 2004). Although MYCL1 is widely used as a marker, very little is known about its characteristics and population genetics. Previous reports describe this marker as a structurally complex microsatellite consisting of mono-, tetra- and pentanucleotide repeats (Mäkelä et al. 1992, Dietmeier et al. 1997, Jass et al. 2001, Hatch and Farber 2004). However, although the PCR fragment sizes of 16 MYCL1 alleles have been described in families collected by CEPH (Centre d’Etude du Polymorphisme Humain) and Finnish INCL families (Mäkelä et al. 1992, Vesa et al. 1993) no thorough description of the level of polymorphism, allelic distribution or of the unstable regions of MYCL1 have been reported. Our previous studies have indicated that the tumor-related instability of a microsatellite marker is associated with its diversity in a population (Vauhkonen et al. 2004a, Vauhkonen et al. 2004b). Based on our previous
findings, the study described in the present paper aimed to reveal the variability pattern of MYCL1 in tumors and between two distant populations (Finnish and Somali).

We obtained DNA from lymphocyte samples from 14 healthy Finnish volunteers and non-cancerous biopsy tissue samples from 103 Finnish patients (n = 117 for the Finnish population; 48 males, 69 females) and lymphocyte samples from 61 Somalis (30 males, 31 females) seeking family reunion in Finland. All individuals were unrelated. Tissue samples from 61 primary gastrointestinal cancers (34 gastric and 27 colorectal) and their cancer-free adjacent areas from 61 Finnish patients (25 female, 36 male, age range 49-85) were collected as described previously (Vauhkonen et al. 2004a). This study was evaluated and approved by the local ethics committee.

Lymphocyte DNA was extracted using proteinase K, phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitation, while DNA from tissues was extracted using proteinase K and Qiaquick columns (Qiagen, Hilden, Germany) according to the method of Vauhkonen et al. (2004a). Microsatellite instability was determined using 15 autosomal loci and the AmpFISTR® SGM Plus (containing the primers for D3S1358, vWA, FGA, TH01, D16S539, D2S1338, D8S119, D21S11, D18S51 and D19S433) and AmpFISTR® Profiler (containing the primers for D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317 and D7S820) kits (Applied Biosystems, Foster City, CA, USA). The analysis was carried out according to the manufacturer’s instructions using the PTC-225 DNA Engine Tetrad (MJ Research, Boston, MA, USA) and ABI Prism CE310 capillary electrophoresis (Applied Biosystems) (Vauhkonen et al. 2004a). The guidelines of Boland et al. (1998) were used for tumor categorization, MSI-H tumors being classified as those having ≥ 5 unstable loci, MSI-L tumors as having 1 to 4 unstable loci and MSS tumors as those showing no microsatellite instability. We confirmed MSI-H using BAT-26 (Hoang et al. 1997) and MYCL1 analysis was performed using PCR primers (Mäkelä et al. 1992) with a FAM-label ed forward primer (TAGC, Copenhagen, Denmark). For DNA amplification we used a 20-μL reaction volume containing 1 ng of sample DNA, 1 unit of AmpliTaqGold (Applied Biosystems), 1 x AmpliTaqGold PCR buffer with 1.5 mM MgCl2, 0.2 mM of each dNTP, and 0.5 pmol of primer and the touch-down protocol described by Vauhkonen et al. (2004b). Each allele was sequenced from one to five individuals with the BigDye Terminator Cycle Sequencing kit (Applied Biosystems). Homozygote genotypes were sequenced without allele separation, whereas heterozygote alleles were separated in 3% Metaphor agarose gels (FMC BioProducts, Rockland, Maine, USA) and the bands excised, soaked overnight in 50 μL of water and re-amplified prior to sequencing. The nomenclature of the alleles was based on the nucleotide sequence, combining the tetranucleotide repeat number (gaaa), with the mono- and pentanucleotide structure (gaaa)n-gaa-(gaaa)-[(gaa)n]z-(gaaa)n-z, which were designated as the PENTA (i.e. pentanucleotide) structure which were numbered 1-13 (Table 1). Polymorphic information content (PIC) and heterozygosity index (HET) were calculated using the PowerStatsV12 program.

In the population samples we found a total of 40 MYCL1 alleles, of which 28 were present in the Finnish and 29 in the Somali population sample (Table 2). Only 16 (40%) of the alleles were shared by both populations. The basic structure (gaaa)-(gaaa)-(gaaa)-(gaaa)n-z, (gaaa)n-z (gaaa)n-z was found in all alleles, of which the tetranucleotide region (gaaa)n, with z ranging from 3 to 17 repeats) and pentanucleotide regions (gaaa)n, with z ranging from 1 to 5 and (gaaa)n, with z ranging from 1 to 6) showed length variation. The low-level variation in the pentanucleotide and mononucleotide regions enabled the categorization of the alleles into 13 distinct PENTA structures (Table 1). Only four (31%) of the PENTA structures were present in both populations, with PENTA1 being present in 74% of the Finnish population and 56% of the Somali population, PENTA2 in 13% of the Finnish and 4% of the Somali population, PENTA3 in 5% of the Finnish and 3% of the Somali population and PENTA4 in 2% of the Finnish and 27% of the Somali population. Only two alleles in the Finnish population (13-PENTA1 and 14-PENTA1, i.e. PENTA1 structures with 14 and 13 tetranucleotide repeats respectively) and four alleles in the Somali population (14-PENTA1, 13-PENTA1, 12-PENTA1 and PENTA4, i.e. PENTA structures with 13, 12 and 12 repeats respectively) reached a frequency above 10% (Table 2). The tetranucleotide region showed the highest repeat number and variability, and thus produced an allelic distribution for each PENTA structure (Figure 2). We found that PENTA4 was more than ten-times as common in the Somali as in the Finnish population (27% vs. 2%) and the 12-PENTA4 allele was found only among Somalis, where it was the second most frequent

---

**Table 1 - PENTA structures.**

<table>
<thead>
<tr>
<th>PENTA</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>PENTA1</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA2</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA3</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA4</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA5</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA6</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA7</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA8</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA9</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA10</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA11</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA12</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
<tr>
<td>PENTA13</td>
<td>(gaaa)n-gaa-(gaaa)</td>
</tr>
</tbody>
</table>
allele (Figure 2). Mäkelä et al. (1993) reported a PIC value of 0.87 and a HET value of 0.87 (both of which indicate the presence of polymorphism) in CEPH families. Our results calculated from the Finnish population data (PIC 0.90 and HET 0.87) support these results.

In the set of 61 cancers from Finnish patients we found 10 MSI-H, 13 MSI-L and 38 MSS tumors. We also found that MYCL1 was unstable in 18% (11/61) of all tumors and in 80% of MSI-H tumors and 7.9% of MSS tumors (Table 3). The observed overall instability level of 11% for MYCL1 in colorectal tumors was in line with previous reports (Young et al. 1995, Dietmaier et al. 1997) while in gastric cancers the corresponding instability level was 24%. The new alleles were produced mainly by single-step (94%) contractions (86%) at the tetranucleotide repeat, while the penta- and mononucleotide regions remained intact (Table 3). Only the alleles with a structure that was highly variable in the population were altered in tumor samples. A recent study by Hatch and Farber (2004) assessed the mutability of MYCL1 in MMR-deficient and MMR-proficient cultured human cell lines using a transfection assay in which cells were transfected with a vector containing a 14-PENTA1 allele (shown in Figure 1) up-

Figure 1 - The basic structure of the 14-PENTA1 ((gaaa)-, (gaaa)k, (gaaa)(gaaa)z, (ga-1,2,3,4,5-(gaaa)z) MYCL1 allele which has 14 tetranucleotide repeats ((gaaa)k, (gaaa)z, (ga-1,2,3,4,5-(gaaa)z) mono- and pentanucleotide structure. The tetranucleotide and pentanucleotide repeats showing length variation are underlined and the mono- and pentanucleotide region forming the PENTA structure is boxed.
stream of an antibiotic resistance gene but colonies were produced only if mutations restoring the reading frame occurred in MYCL1. Tetranucleotide repeat mutations were present in both MMR-deficient and MMR-proficient cells but the MMR-proficient cells also produced novel allelic structures. However, since only one reading frame was used in the construct only one-third of the possible MYCL1 mutations were detected.

In our present study, comparison of the allelic distributions of the structurally complex MYCL1 microsatellite in two distant populations has provided further insight into the dynamics of the different MYCL1 structures. Independent distribution of alleles belonging to the same PENTA group indicated that changes in the tetranucleotide repeats produced the observed distributions (Figure 2). Our analysis of MYCL1 in MSI tumors confirmed the preferential mutability of the tetranucleotide repeat of MYCL1, which further suggests that this marker should be considered primarily as a tetranucleotide marker. The alleles which were mutated in tumors presented the major PENTA structures with the highest number of alleles in the populations examined. This might have been due either to stochastic effects or allele structure, as, for example, has been previously been reported for the D2S123 allele by Bacon et al. (2000). The consequences of differential allelic mutability may be significant when using the marker for MSI categorization. In addition, MYCL1 with regions of both high and low mutability may prove to be a useful marker in studies of microsatellite dynamics or population genetics.

Acknowledgments

We thank Professor Pentti Sipponen for the tissue samples, Minttu Hedman for the Somali DNA samples and Matti Vauhkonen M.D., Ph.D. for his contribution in the preparation of the manuscript.

References


Hoang JM, Cottu PH, Thuille B, Salmon RJ, Thomas G and Hamelin R (1997) BAT-26, an indicator of the replication error

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Tumor origin</th>
<th>Phenotype$^1$</th>
<th>PENTA</th>
<th>(gaaa), mutation</th>
<th>Change (+ or -) in repeat number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Colorectal cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>14 → 13</td>
<td>-1</td>
</tr>
<tr>
<td>4A</td>
<td>Colorectal cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>16 → 15 → 14</td>
<td>-1, -2</td>
</tr>
<tr>
<td>4C</td>
<td>Gastric cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>13 → 12</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>Colorectal cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>9 → 8</td>
<td>-1</td>
</tr>
<tr>
<td>55</td>
<td>Gastric cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>13 → 14</td>
<td>+1</td>
</tr>
<tr>
<td>91</td>
<td>Gastric cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>14 → 13 → 12</td>
<td>-1, -2</td>
</tr>
<tr>
<td>154</td>
<td>Gastric cancer</td>
<td>MSS</td>
<td>PENTA1</td>
<td>13 → 12</td>
<td>-1</td>
</tr>
<tr>
<td>161</td>
<td>Gastric cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>17 → 16</td>
<td>-1</td>
</tr>
<tr>
<td>174</td>
<td>Gastric cancer</td>
<td>MSS</td>
<td>PENTA1</td>
<td>14 → 15</td>
<td>+1</td>
</tr>
<tr>
<td>176</td>
<td>Gastric cancer</td>
<td>MSI-H</td>
<td>PENTA1</td>
<td>12 → 11</td>
<td>-1</td>
</tr>
<tr>
<td>184</td>
<td>Gastric cancer</td>
<td>MSS</td>
<td>PENTA1</td>
<td>14 → 13</td>
<td>-1</td>
</tr>
</tbody>
</table>

$^1$MSI-H = microsatellite instability-high; MSS = microsatellite stable.

Table 3 - Mutations observed in tumor samples from Finnish patients. In tumors 4C, 10 and 91, mutations were observed at both alleles. See Table 2 for PENTA structure.
phenotype in colorectal cancers and cell lines. Cancer Res 57:300-303.


Associate Editor: Emmanuel Dias Neto