



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 (Dietmaier *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, Iino *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, Dietmaier *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 non-related. 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 AmpF/STR® SGM Plus (containing the primers for D3S1358, vWA, FGA, TH01, D16S539, D2S1338, D8S119, D21S11, D18S51 and D19S433) and AmpF/STR® 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-labeled 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)_x with the mono- and pentanucleotide struc-

ture (gaaa)_y-ga₈-(gaaa)_z-[(ga₆)₁₋₂-(gaaa)₂]₀₋₁, 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 (gaa)_x-(gaaa)_y-ga₈-(gaaa)_z-[(ga₆)₁₋₂-(gaaa)₂]₀₋₁ (Figure 1) was found in all alleles, of which the tetranucleotide region ((gaa)_x, with *x* ranging from 3 to 17 repeats) and pentanucleotide regions ((gaaa)_y, with *y* ranging from 1 to 5 and (gaaa)_z, with *z* ranging from 1 to 6) showed length variation. The low-level variation in the penta- 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 12-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.

PENTA	Structure
PENTA1	(gaaa) ₂ -ga ₈ -(gaaa) ₃ -ga ₆ -(gaaa) ₂
PENTA2	(gaaa) ₄ -ga ₈ -(gaaa) ₃ -ga ₆ -(gaaa) ₂
PENTA3	(gaaa) ₄ -ga ₈ -(gaaa) ₃ -ga ₆ -(gaaa) ₂
PENTA4	(gaaa) ₂ -ga ₈ -(gaaa) ₄ -ga ₆ -(gaaa) ₂
PENTA5	(gaaa) ₂ -ga ₈ -(gaaa) ₃
PENTA6	(gaaa) ₃ -ga ₈ -(gaaa) ₄ -ga ₆ -(gaaa) ₂
PENTA7	(gaaa) ₃ -ga ₈ -gaaa-ga ₆ -(gaaa) ₂
PENTA8	(gaaa) ₂ -ga ₈ -(gaaa) ₃ -ga ₆ -ga ₆ -(gaaa) ₂
PENTA9	(gaaa) ₄ -ga ₈ -(gaaa) ₅ -ga ₆ -(gaaa) ₂
PENTA10	(gaaa) ₅ -ga ₈ -(gaaa) ₄ -ga ₆ -(gaaa) ₂
PENTA11	(gaaa) ₂ -ga ₈ -(gaaa) ₆ -ga ₆ -(gaaa) ₂
PENTA12	(gaaa) ₄ -ga ₈ -(gaaa) ₄ -ga ₆ -(gaaa) ₂
PENTA13	gaaa-ga ₈ -(gaaa) ₄ -ga ₆ -(gaaa) ₂

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.

Tumor	Tumor origin	Phenotype ¹	PENTA	(gaaa) _x mutation	Change (+ or -) in repeat number
1	Colorectal cancer	MSI-H	PENTA1	14 → 13	-1
4A	Colorectal cancer	MSI-H	PENTA1	16 → 15 → 14	-1, -2
4C	Gastric cancer.	MSI-H	PENTA1	13 → 12	-1
4C	Gastric cancer.	MSI-H	PENTA1	16 → 15 → 14	-1, -2
10	Colorectal cancer	MSI-H	PENTA1	9 → 8	-1
10	Colorectal cancer	MSI-H	PENTA1	14 → 13	-1
55	Gastric cancer	MSI-H	PENTA1	13 → 14	+1
91	Gastric cancer	MSI-H	PENTA2	12 → 9	-3
91	Gastric cancer	MSI-H	PENTA1	14 → 13 → 12	-1, -2
154	Gastric cancer	MSS	PENTA1	13 → 12	-1
161	Gastric cancer	MSI-H	PENTA1	17 → 16	-1
174	Gastric cancer	MSS	PENTA1	14 → 15	+1
176	Gastric cancer.	MSI-H	PENTA1	12 → 11	-1
184	Gastric cancer	MSS	PENTA1	14 → 13	-1

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

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 the 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

- Bacon AL, Farrington SM and Dunlop MG (2000) Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells. *Hum Mol Gen* 9:2707-2713.
- Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN and Srivastava S (1998) A National Cancer Institute workshop on microsatellite instability for cancer detection on family predisposition; development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 58:5248-5257.
- Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R and Rüschoff J (1997) Diagnostic microsatellite instability: Definition and correlation with mismatch repair protein expression. *Cancer Res* 57:4749-4756.
- Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK and Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci and implications for human demographic histories. *Genetics* 148:1269-1284.
- Halford SER, Sawyer EJ, Lambros MB, Gorman P, Macdonald ND, Talbot IC, Foulkes WD, Gillett CE, Barnes DM, Akslan LA, Lee K, Jacobs IJ, Hanby AM, Ganesan TS, Salvesen HB, Bodmer WF, Tomlinson IP and Royle RR (2003) MSI-low, a real phenomenon which varies in frequency among cancer types. *J Pathol* 201:389-394.
- Hatch SB and Farber RA (2004) Mutation rates in the complex microsatellite MYCL1 and related simple repeats in cultured human cells. *Mutat Res* 545:117-126.
- Hoang JM, Cottu PH, Thuille B, Salmon RJ, Thomas G and Hamelin R (1997) BAT-26, an indicator of the replication error

- phenotype in colorectal cancers and cell lines. *Cancer Res* 57:300-303.
- Iino H, Jass JR, Simms LA Young J, Leggett B, Ajioka Y and Watanabe H (1999) DNA microsatellite instability in hyperplastic polyps, serrated adenomas, and mixed polyps: A mild mutator pathway for colorectal cancer? *J Clin Pathol* 52:5-9.
- Jass JR, Young J and Leggett BA (2001) Correspondence: Biological significance of microsatellite instability-low (MSI-L) status in colorectal tumours. *Am J Pathol* 158:779-780.
- Kambara T, Sharp GB, Nagasaka T, Takeda M, Sasamoto H, Nakagawa H, Isozaki H, MacPhee DG, Jass JR, Tanaka N and Matsubara N (2004) Allelic loss of a common microsatellite marker MYCL1: A useful prognostic factor of poor outcomes in colorectal cancer. *Clin Cancer Res* 10:1758-1763.
- Laiho P, Launonen V, Lahermo P, Esteller M, Guo M, Herman JG, Mecklin JP, Jarvinen H, Sistonen P, Kim KM, Shibata D, Houlston RS and Aaltonen LA (2002) Low-level microsatellite instability on most colorectal carcinomas. *Cancer Res* 62:1166-1170.
- Mäkelä TP, Hellsten E, Vesa J, Alitalo K and Peltonen L (1992) An Alu variable poly A repeat polymorphism upstream of L-myc at 1p32. *Hum Mol Genet* 1:217.
- Perinchery G, Nojima D, Goharderakhshan R, Tanaka Y, Alonso J and Dahiya R (2000) Microsatellite instability of dinucleotide tandem repeat sequences is higher than trinucleotide, tetranucleotide and pncucleotide repeat sequences in prostate cancer. *Int J Oncol* 16:1203-1209.
- Sturzeneker R, Bevilacqua R, Haddad L, Simpson A and Pena S (2000) Microsatellite instability in tumors as a model to study the process of microsatellite mutations. *Hum Mol Genet* 9:347-352.
- Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN and Srivastava S (2004) Revised Bethesda guidelines for hereditary colorectal cancer (Lynch Syndrome) and microsatellite instability. *J Natl Cancer Inst* 96:261-268.
- Vauhkonen H, Hedman M, Vauhkonen M, Kataja M, Sipponen P and Sajantila A (2004a) Evaluation of gastrointestinal cancer tissues as a source of genetic information for forensic investigations by using STRs. *Forensic Sci Int* 139:159-167.
- Vauhkonen H, Vauhkonen M, Sipponen P and Sajantila A (2004b) Comparison of the allelic distribution of X-chromosomal STRs in the normal population and in phenotypically different tumour tissues. A study using four X-STR loci (DXS7423, DXS8377, ARA, DXS101). *Ann Hum Gen* 68:55-562.
- Vesa J, Hellsten E, Mäkelä TP, Järvelä I, Airaksinen T, Santavuori P and Peltonen L (1993) A single PCR marker in strong allelic association with the infantile form of neuronal ceroid lipofuscinosis facilitates reliable prenatal diagnostics and disease carrier identification. *Eur J Hum Genet* 1:125-132.
- Young J, Searle J, Buttenshaw R, Thomas L, Ward M, Chenevix-Trench G and Leggett B (1995) An Alu VpA marker on chromosome 1 demonstrates that replication errors manifest at the adenoma-carcinoma transition in sporadic colorectal tumors. *Genes, Chromosomes & Cancer* 12:251-254.

Associate Editor: Emmanuel Dias Neto