

ARTICLE

Optimization of a pentaplex panel for MSI analysis without control DNA in a Brazilian population: correlation with ancestry markers

Nathália C Campanella^{1,5}, Gustavo N Berardinelli^{1,5}, Cristovam Scapulatempo-Neto¹, Danilo Viana²,
Edenir I Palmero¹, Rui Pereira³ and Rui M Reis^{1,4,*}

Microsatellite instability (MSI) testing has been advocated for all newly diagnosed colorectal cancer patients. One of the most common tests is composed by a pentaplex panel of mononucleotides markers (NR-27, NR-21, NR-24, BAT-25, and BAT-26), which allows the analysis of MSI in tumors without the need of reference DNA. For that, it is fundamental to establish a quasi-monomorphic variation range (QMVR) for each marker. Herein, we aimed to establish the QMVR in a Brazilian healthy population, to evaluate the feasibility of MSI determination of tumors, without the matching normal DNA. Furthermore, we intend to assess their ancestry using specific ancestry-informative markers (AIMs) and correlate with QMVR. The QMVR was assessed in 214 individuals, through a pentaplex PCR followed by fragment analysis. The ancestry analysis was done by 46 AIMs in a single multiplex PCR followed by capillary electrophoresis. Following QMVR establishment, we observed 23 individuals with alleles outside the QMVR. Importantly, none of them exhibited more than one marker outside the range. Therefore, individuals with instability at ≥ 2 markers would be accurately classified as MSI. The European ancestry proportion was the most frequent (67.5%), followed by the African (19.6%). The comparison of the individuals with alleles within ($n = 191$) and outside ($n = 23$) the QMVR showed statistical difference in the proportions of European and African alleles, confirming the higher polymorphic nature of African ancestry. In conclusion, the present study reports an accurate methodology to assess MSI status without matched-normal DNA and independently of the ethnicity, even in the highly admixed population of Brazil.

European Journal of Human Genetics (2014) **22**, 875–880; doi:10.1038/ejhg.2013.256; published online 6 November 2013

Keywords: microsatellite instability; quasi-monomorphic variation range; ancestry

INTRODUCTION

The microsatellite instability (MSI) arises in short repetitive DNA sequences (or microsatellites) due to defects and loss of function of the MMR family genes (*MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *PMS1*, and *PMS2*), a system that recognizes and repairs errors that occur during DNA replication, as well as some forms of DNA damage.¹ The MSI is a change in length due to either insertion or deletion of repeating units of microsatellite in tumor DNA that is not seen in the corresponding normal DNA.^{2,3}

The MSI phenotype was first described and constitutes a hallmark of tumors associated with hereditary non-polyposis colorectal cancer (HNPCC) or Lynch's syndrome that is a hereditary cancer predisposition syndrome caused by inactivating germline mutations in the MMR gene family.^{2,4–8} These patients develop tumors at early ages and frequently have multiple tumors in colon and rectum, and less frequently have endometrial, ovarian, and stomach cancers.^{5–7,9} In sporadic colorectal cancer (CRC), the MSI is also observed, but at a lower rate in ~ 10 –15%.^{2,3,10} Tumors with the MSI phenotype appear to be associated with particular molecular, histopathological, and

clinical features, including mutation profile, specific location, tumors poorly differentiated, high tumor lymphocyte infiltration, low frequency of distant metastasis, and good prognosis.^{11–13}

At a molecular level, studies have showed that oncogenic alterations found in MSI tumors are somatic mutational events that affect coding repeated sequences in numerous target genes.^{13–15} Several MSI-target genes have been identified in MSI-positive tumors, which are involved in important diverse cellular functions and pathways such as DNA repair (eg, *MRE11*, *MSH6*, *BRCA1*, and *BRCA2*), cell growth (eg, *TGF β RII* and *EGFR*), and apoptosis (eg, *IGFR* and *BAX*).^{14,16,17} In sporadic tumors, the MSI is strongly associated with the presence of *BRAF* oncogene mutations,¹⁸ a lack of *KRAS* mutation, and the inactivation of *PTEN* tumor suppressor gene.^{19,20} Thus, the mutation profile in MSI tumors is fundamentally different from other CRCs.¹⁴

From a histopathological and clinical point of view, CRCs with MSI have distinct features, such as location in the proximal colon and the major prevalence in stage II ($\sim 20\%$) compared with stage III ($\sim 12\%$). In stage IV or metastatic CRC, the MSI is relatively uncommon ($\sim 4\%$).^{13,21,22} It is well accepted that CRC patients

¹Molecular Oncology Research Center, Barretos Cancer Hospital, São Paulo, Brazil; ²Oncogenetic Department, Barretos Cancer Hospital, São Paulo, Brazil; ³IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; ⁴Life and Health Sciences Research Institute, University of Minho, and ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal

⁵These authors contributed equally to this work.

*Correspondence: Dr RM Reis, Molecular Oncology Research Center, Barretos Cancer Hospital, Rua Antenor Duarte Vilela, 1331, CEP 14784 400, Barretos, São Paulo, Brazil. Tel: +55 17 3321 6600; E-mail: ruireis.hcb@gmail.com

Received 17 June 2013; revised 19 September 2013; accepted 27 September 2013; published online 6 November 2013

with MSI have an overall better survival than MSS patients.^{23–25} Importantly, MSI is being considered as a predictive biomarker of 5-fluorouracil (5-FU), irinotecan, and other chemotherapeutic agents response.¹⁴ Following an initial study of Elsaleh *et al*²³ where MSI tumors were associated with better response to a 5-FU regimen, many others report tried to validate these findings; however, its predictive value remains controversial. Large retrospective studies confirmed that the effect of 5-FU is restricted to stage II MSI-positive cases²⁴ and is not applicable to stage III MSI-positive ones.²⁵ Yet, a recent meta-analysis interrogated this issue and did not confirm the predictive value of MSI for 5-FU therapy at any stage.²⁶ Interestingly, a recent study of Dorard *et al*¹⁵ identified the *HSP110* as a novel MSI-target gene in colorectal tumors. The authors reported that the overexpression of the mutant *HSP110* caused the sensitization of cells to oxaliplatin and 5-FU, and raised the hypothesis that the MSI-target genes rather than the MSI phenotype can be predictive of chemotherapy regimens.¹⁵

Due to the need of a better understanding of the clinical and histologic manifestations of HNPCC, the National Cancer Institute (NCI) hosted in 1997 an international workshop on HNPCC, which culminated with the development of a panel of genetic markers that allows the identification of MSI-positive individuals at-risk for HNPCC. This panel is known as Bethesda panel²⁷ and includes five markers: two mononucleotide (BAT-25 and BAT-26) and three dinucleotide (D5S346, D2S123, and D17S250) repeats. Tumors with instability at two or more of these markers were defined as MSI-high (MSI-H), whereas those with instability in one repeat or showing no instability were defined as MSI-low (MSI-L) and microsatellite stable (MSS), respectively.²⁸ Following this landmark panel, several others were reported using tetra, di, and mononucleotide repeats, for MSI screening/diagnosis. However, all of these assays required standardization and the need to include matched-normal DNA as a reference.^{2,8,28}

Posteriorly, in 2002 another HNPCC workshop was held at NCI for re-evaluation and improvement of the Bethesda Guidelines.^{27,28} This workshop concluded the existence of limitations on the application of the above-mentioned markers, and it was recommended that the dinucleotide repeats should be substituted by mononucleotide repeats.²⁸ Consequently, an optimized assay of five mononucleotide markers (BAT-25, BAT-26, NR-21, NR-24, and NR-27) was established for MSI screening avoiding the use of paired normal DNA.^{8,29} Thereafter, this panel of markers has been used by some authors in a pentaplex PCR assay to evaluate MSI status without the need of matched-reference DNA.^{8,28–31} Using this panel, the criterion used to classify a tumor as MSI is not consensual. Several authors suggest that the minimum number of unstable markers required are 2 out of 5,^{29,32} whereas other studies suggest 3 out of 5 markers to determine an MSI-H phenotype.^{27,28}

These five mononucleotide markers show a quasi-monomorphic variation in accordance with different ethnicities: Sub-Saharan Africans showed variant alleles above or approaching 10% for NR-27, BAT-25, and BAT-26, and the other populations (European, East Asian, and Native American) did not reach a frequency of 10% variant alleles for any of the markers.²⁸ In Europe, >95% of individuals do not show any variant alleles in the five markers, whereas in the remaining 5%, individuals present variant alleles in only one marker.²⁸ In the current admixed Brazilian population, these markers were not evaluated and the quasi-monomorphic variation range (QMVR) is still undetermined.

One accurate estimative of ancestry proportions at the population level is possible by making use of ancestry-informative markers (AIMs).^{33–36} Some of these studies are focused on insertion

deletion polymorphisms (INDELs) that are likely to be important factors underlying inherited traits in humans.³³ Recently, using a set of AIM-INDELs, we estimated the ancestry proportions of individuals from Amazonas, Belém, and Rio de Janeiro populations in Brazil.^{33,37} We concluded that this panel has accuracy and is suitable to estimate genetic ancestry in highly admixed populations such as the Brazilian.^{33,37}

Herein, we aimed to establish the QMVR of BAT-25, BAT-26, NR-21, NR-24, and NR-27 markers, in a Brazilian healthy population, to evaluate the feasibility of MSI status determination of tumors without the need of matching normal DNA. Furthermore, we intend to determine the ancestry of Brazilian individuals using specific AIM-INDELs and correlate with their QMVR.

MATERIALS AND METHODS

Samples

Blood normal DNA was obtained from 214 healthy individuals provided by the BioBank of the Barretos Cancer Hospital, Barretos, São Paulo, Brazil. This study was approved by the local ethic commit (600/2012). The average age of the individuals was 33 years old, 52.3% were male, and 90% of the individuals came from southeast region of Brazil (São Paulo and Minas Gerais states), whereas others came from Paraná, Rio Grande do Sul, Bahia, Mato Grosso, Mato Grosso do Sul, Paraíba, Pernambuco, and Rondônia regions. Blood DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. In addition, cancer cell lines were used as controls. The CRC cell lines Co115, DLD1, and LoVo were used as an MSI-positive control, and CaCo2, DIFI, HT29, SW480, and SW620 were used as an MSI-negative control. The DNA from cancer cell lines was extracted using Trizol reagent (Life Technologies, Gaithersburg, MD, USA) following the manufacturer's protocol.

MSI analysis

The MSI evaluation was performed using a multiplex PCR comprising five quasi-monomorphic mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24, and NR-27).^{28,38} Primer sequences were described elsewhere.²⁸ Each antisense primer was end labeled with a fluorescent dye: FAM (6-carboxyfluorescein) for BAT-26 and NR-21; VIC (2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein) for BAT-25 and NR-27; and NED (2,7,8-benzo-5-fluoro-2,4,7-trichloro-5-carboxyfluorescein) for NR-24. PCR was performed using the Qiagen Multiplex PCR Kit (Qiagen), with 0.5 μ l of DNA at 50 ng/ μ l and the following thermocycling conditions: 15 min at 95 °C; 40 cycles of 95 °C for 30 s, 55 °C for 90 s and 72 °C for 30 s; and a final extension at 72 °C for 60 min. PCR products were then submitted to capillary electrophoresis on an ABI 3500 xL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions and the results were analyzed using GeneMapper v4.1 (Applied Biosystems) software.

The QMVR of each marker was established from an average of the alleles' size with a range of plus or minus three nucleotides in accordance with the literature.^{29,30}

Ancestry analysis

The ancestry analysis was performed using a set of 46 AIMs among the most informative INDELs for each population group as previously described.³³ Primer sequences and PCR conditions were according to Pereira *et al*.³³ A 46plex PCR was performed and the amplified products were further submitted to fragment analysis on an ABI 3500 xL Genetic Analyzer (Applied Biosystems), according to the manufacturer's instructions. The electropherograms were analyzed and genotypes were automatically assigned with GeneMapper v4.1 (Applied Biosystems).

Ancestry proportions were then assessed using the Structure v2.3.3 software,^{39,40} considering the four major population groups of Native American, European, African, and East Asian as possible contributors to the current genetic makeup of Brazilians. Using the data available for the HGDP-CEPH panel as a reference for the ancestral populations, a supervised analysis was performed to estimate ancestry membership proportions of the individuals

involved in the study. Structure runs considering $K=4$ consisted of 100 000 burnin steps followed by 100 000 Markov Chain Monte Carlo iterations. The option 'Use population Information to test for migrants' was used with the Admixture model, considering allele frequencies correlated, and updating allele frequencies using only individuals with POPFLAG = 1.

The SPSS 19.0 software (IBM Corp, Armonk, NY, USA) was used for all statistics analysis. The P -value established for the statistics significance was <0.05 .

RESULTS

Determination of QMVR using the 5-marker panel

All the 214 samples were successfully amplified for the five markers, generating 428 alleles for each marker. The allele's sizes and the absolute and relative frequencies for each marker are showed in Table 1 and Figure 1. All raw data are shown in Supplementary Table 1.

NR-27 demonstrated a stable profile with a QMVR of 82–88 bp, with only one allele (81 bp) outside the QMVR. For NR-21, the size of the quasi-monomorphic alleles ranged from 101 to 107 bp, and five alleles of 100 bp were outside the established QMVR. NR-24 showed a QMVR of 119–125 bp and no allele outside the range was found. Both BAT-25 and BAT-26 exhibited a bimodal distribution: the determined QMVR of BAT-25 was 142–148 bp and 6 individuals (1.38%) harbored variant alleles outside it, while BAT-26 had an established QMVR of 174–180 bp and 14 individuals (3.23%) had alleles outside the QMVR range (Table 1; Figure 1).

Overall, we identified a total of 23 individuals that showed alleles outside the QMVR. Importantly, none of them showed more than one marker outside the range (Table 2), and independently of the 2/5 or 3/5 cutoff used for MSI-H definition, none of the healthy individuals would be erroneously classified as MSI. As expected, all colorectal cell lines known to be MSI (HCT15, DLD1, and LoVo) exhibited a MSI-H profile, and the known MSS cell lines (CaCo2, DIFI, HT-29, SW480, and SW620), were correctly classified (Supplementary Table 2).

Ancestry analysis

The results obtained for the complete AIM-INDEL panel allowed us to estimate the AFR, EUR, EAS, and NAM ancestral proportions of the 214 healthy individuals (all raw data are shown in Supplementary

Table 1). The average ancestry proportions for all individuals were 67.5% for EUR, 19.6% for AFR, 6.7% for NAM, and 6.2% for EAS (Table 3; Figure 2).

We further compared the ancestry estimates of individuals exhibiting alleles outside and within the QMVR. We observed that the 23 individuals with alleles outside the QMVR showed mean ancestry proportions of 49.6% for EUR, 37.2% of AFR, 7.6% of EAS, and 5.6% of NAM, whereas the 191 individuals within the QMVR showed 69.7% for EUR, 17.5% of AFR, 6.8% of NAM, and 6% of EAS (Table 4). These differences were statistically significant for the AFR ($P<0.001$) and EUR ($P=0.001$) ancestral contributions (Table 4). In fact, when we consider individuals with essentially European ancestry (EUR membership proportion $>90\%$) only one allele outside the QMVR was observed.

DISCUSSION

The determination of MSI status is of extreme importance for at-risk identification of HNPCC individuals and molecular characterization

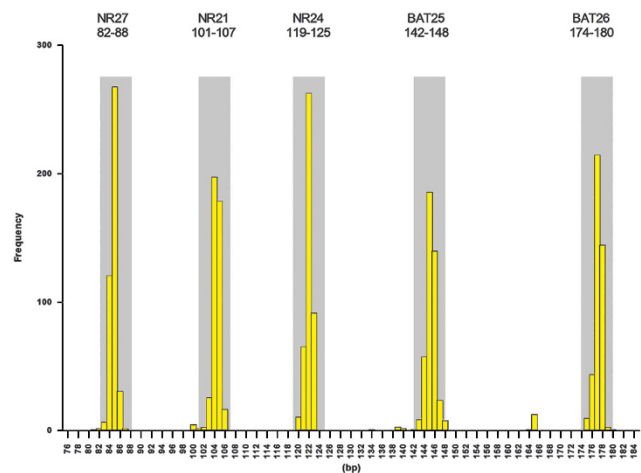


Figure 1 Frequency of allele size distribution (in base pairs) for the five markers from 214 normal DNA specimens. For each marker, the gray shading indicates the QMVR established.

Table 1 Sizes of the alleles and the QMVR for each marker

NR-27			NR-21			NR-24			BAT-25			BAT-26		
Size (bp)	N	%	Size (bp)	N	%	Size (bp)	N	%	Size (bp)	N	%	Size (bp)	N	%
80			99			118			133			163		
81	1	0.23	100	4	0.93	119			134	1	0.23	164	1	0.23
82	2	0.46	101	2	0.46	120	11	2.57	135			165	13	3
83	7	1.63	102	3	0.7	121	65	15.18	139	3	0.69	174		
84	119	27.8	103	26	6.07	122	263	61.4	140	2	0.46	175	11	2.57
85	265	61.9	104	195	45.56	123	89	20.79	142	1	0.23	176	45	10.51
86	32	7.48	105	182	42.52	124			143	9	2.1	177	212	49.53
87	2	0.46	106	16	3.73	125			144	55	12.85	178	142	33.17
88			107			126			145	185	43.22	179	3	0.7
89			108						146	140	32.7	180	1	0.23
									147	24	5.6	181		
									148	8	1.87	182		
									149					
									150					

The dark-gray shading represents the mean and the light-gray shading shows the established QMVR for each marker.

Table 2 Detail of all markers of the individuals with alleles outside the QMVR

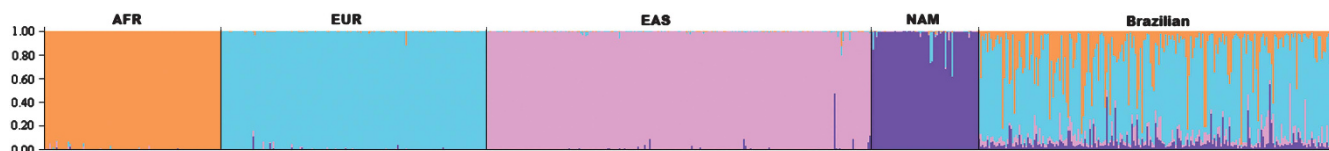
ID	Age	Gender	NR-27		NR-21		NR-24		BAT-25		BAT-26	
			82–88		101–107		119–125		142–148		174–180	
C1	23	F	85	85	100	105	122	122	145	146	177	178
C15	23	M	84	85	104	105	122	123	145	146	165	177
C3	29	M	85	85	104	105	122	123	134	146	177	178
2	18	M	85	85	103	105	120	120	139	146	177	178
13	18	F	84	85	100	104	122	123	145	146	175	176
16	31	M	84	84	104	105	121	122	144	146	165	177
22	41	M	85	85	104	104	122	123	145	145	165	176
34	18	M	85	85	105	105	120	120	139	140	177	179
36	31	F	81	85	104	105	122	122	145	146	177	178
45	41	F	85	85	105	105	122	122	145	146	165	176
49	18	M	84	85	104	105	122	123	139	146	177	178
53	31	M	84	85	103	105	122	122	144	144	165	177
56	41	M	84	85	104	105	122	122	145	146	165	176
63	51	M	85	85	104	104	121	122	144	146	165	177
75	18	M	84	85	100	105	122	122	146	146	176	177
91	41	F	85	85	100	105	122	122	144	145	177	178
97	41	M	84	85	104	105	121	122	145	145	165	177
108	51	F	82	86	105	105	121	122	145	146	165	178
125	51	M	86	86	104	104	122	122	145	146	165	177
135	18	F	85	85	102	103	122	122	144	145	165	177
136	51	M	85	85	102	103	122	122	144	145	165	177
151	18	M	85	85	105	105	121	121	145	146	165	178
162	41	M	85	86	104	105	122	122	144	145	164	177

Abbreviations: F, Female; M, Male; light gray, alleles outside the QMVR.

Table 3 Ancestral membership proportions (average) for the HGDP-CEPH Diversity Panel reference samples and the studied Brazilian population

	AFR	EUR	EAS	NAM
HGDP-CEPH AFR origin	99.5%	0.2%	0.2%	0.1%
HGDP-CEPH EUR origin	0.2%	99.3%	0.2%	0.3%
HGDP-CEPH EAS origin	0.1%	0.3%	99.0%	0.6%
HGDP-CEPH NAM origin	0.1%	2.3%	0.4%	97.2%
Present Brazilian population (SD)	19.6% (21.9)	67.6% (23.9)	6.1% (6.2)	6.7% (7.3)

Abbreviations: AFR, African; EUR, European; EAS, East Asian; NAM, Native American; SD, standard deviation.

**Figure 2** Individual ancestry estimates for the testing Brazilian population using the HGDP-CEPH diversity panel genetic data as a training set.

of sporadic CRC.^{2,7} In fact, the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group recommends MSI testing to all newly diagnosed CRC patients.⁴¹ Of the several panel of genetic markers available, the use of pentaplex quasi-monomorphic markers (NR-27, NR-21, NR-24, BAT-25, and BAT-26) is one of the most disseminated worldwide, and are the base of the Promega MSI kit (Promega, Madison, WI, USA).^{2,7,30} The optimization of QMVR for each quasi-monomorphic marker is necessary because the allelic size estimation can be influenced by several factors, such as the use of specific reagents and sensibility of

the sequencer apparatus utilized, as well as population ethnicity. In the current admixed Brazilian population, the variation of the pentaplex quasi-monomorphic markers was not evaluated and the QMVR was undetermined. In the present study, we analyzed the allelic size variation of the five mononucleotide markers in a series of germline DNA from 214 healthy Brazilian individuals, determined their QMVR and correlated with individuals ancestry determined by AIMs.

The QMVRs established from each marker were 82–88bp for NR-27, 101–107 for NR-21, 119–125 for NR-24, 142–148 for BAT-25,

Table 4 Ancestral membership proportions (average) considering the subsets of individuals within and outside the QMVR

	Individuals with alleles within QMVR	Individuals with alleles outside QMVR	P-value (Mann–Whitney test)
AFR (SD)	17.5% (20.1)	37.2% (27.7)	<0.001
EUR (SD)	69.7% (22.4)	49.6% (28.6)	0.001
EAS (SD)	6% (6.2)	7.6% (6.5)	0.277
NAM (SD)	6.8% (7.6)	5.6% (5.2)	0.786

Abbreviations: AFR, African; EUR, European; EAS, East Asian; NAM, Native American; SD, standard deviation.

and 174–180 for BAT-26. The frequency of variant alleles for each marker was variable: BAT-26 marker showed the major proportion of variant alleles, with 14 alleles outside the QMVR. BAT-25 presented six alleles outside the QMVR, followed by NR-21 with four alleles and NR-27 with one allele. For NR-24, no alleles were observed out of the determined QMVR. The QMVR for all markers in our individuals differed by few base pairs in relation to what had been reported previously.^{28,29} When considering the five markers, the analyses showed 191 individuals (89.25%) within the QMVR, and 23 individuals (10.75%) exhibited one allele outside the QMVR. Buhard *et al*²⁸ had established a QMVR analyzing 1206 individuals from 55 different worldwide populations including the HGDP-CEPH panel and found that 87.5% of cases showed alleles within the QMVR,²⁸ a similar result as observed in our study. Importantly, the authors assessed 45 individuals from 2 native Brazilian populations (Karitiana and Surui) and none of them exhibited a marker outside the QMVR.²⁸ Therefore, analyzing the current Brazilian population, which results from the miscegenation of distinct ethnic groups, our study is of great relevance for MSI determination for clinical application.

To correlate markers allele size with individual's ancestry, we performed the analysis of an AIM-INDEL panel that could discriminate a four-group contribution. This four-group analysis is due to the historical migration of populations, where the Native Americans suffered admixture with Europeans, followed by African, and more recently with a significant East Asian community, mainly in São Paulo state.

Using this AIM-INDEL panel, we showed that the average ancestry proportions of the 214 individuals were 67.5% of European, 19.6% of African, 6.7% of Native Americans, and 6.2% of East Asian. Using the same panel, Pereira *et al*³³ analyzed a Brazilian population from Belém (northeastern region of Brazil), and identified distinct frequency of ancestry: 53.5% EUR; 14.8% AFR; 22.9% NAM, and 8.8% EAS,³³ most probably representing the higher native American ancestry of the region.

Furthermore, Manta *et al*³⁷ analyzed a Brazilian population from Rio de Janeiro, and using the same AIMs panel observed 55.2% of EUR, 31.1% of AFR, and 13.7% of NAM. Using a different panel of the markers, Pena *et al*⁴² performed a wider ancestry analysis of Brazilian population from distinct geographical regions (North, Northeast, Southeast, and South). The authors described for the southeast region similar frequencies to the ones obtained in our studied population, namely 74.2% of European, 17.3% of African, and 7.3% of Amerindian ancestry.⁴² The majority of the alleles found in our population outside the QMVR are on the BAT-25 and BAT-26 markers. These loci are described as rather polymorphic in African populations.³⁰ In consonance, we showed that the Brazilian individuals harboring those variants showed a higher African

ancestry. Nevertheless, we observed that none of the individuals with polymorphism exhibited more than one marker outside the established QMVR.

An important issue with the use of the present pentaplex panel for MSI testing without the need of reference DNA is the minimum number of unstable markers needed to categorize a tumor as MSI. The literature is ambiguous, varying from 2 to 3 out of 5 markers. In the present study of an admixed population from the Southeast of Brazil, none of the individuals showed more than one unstable marker. Therefore, we suggest classifying as MSI-H, individuals with instability at two or more markers. Since the presence of instability in one marker can be due to polymorphic variants, we proposed that these subjects should be further analyzed by MMR immunohistochemistry, or a PCR using paired normal DNA, for accurately determine the MSI-L status of patients. Therefore, our results demonstrated that this methodology allows characterizing the MSI status without the need of the matched blood for each patient. This constitute a very important advantage since molecular diagnostic laboratories often received only the formalin-fixed paraffin-embedded tumor tissue, not having access to matched blood of the patient.

In conclusion, this is the first study that reports the optimization and establishment of a suitable QMVR for a pentaplex system of MSI markers and characterizes their genetic variation in the Brazilian population for molecular diagnosis purposes. We found a noticeable number of alleles outside the QMVR, which appear directly associated with an important African ancestral component present in the current Brazilian population, as depicted by the use of AIMs. Despite the higher frequency of variant alleles in individuals with African ancestry, no individuals showed more than one allele outside the established QMVR in our study, and therefore our results corroborate that this methodology may be used with confidence to assess MSI status without matched-normal DNA and independently of the ethnicity, even in the highly admixed population of Brazil.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The present study was partially supported by a MCT/FINEP/CT-INFRA-PROINFRA 02/2010 grant and a CNPq Universal grant (475358/2011-2). Nathália Cristina Campanella is a recipient of an FAPESP Master Grant (2012/01732-2). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by the Portuguese Foundation for Science and Technology (FCT). Rui Pereira is a recipient of a postdoctoral fellowship from FCT (SFRH/BPD/81986/2011).

- 1 Karran P: Microsatellite instability and DNA mismatch repair in human cancer. *Semin Cancer Biol* 1996; **7**: 15–24.
- 2 Vilar E, Gruber SB: Microsatellite instability in colorectal cancer—the stable evidence. *Nat Rev Clin Oncol* 2010; **7**: 153–162.
- 3 Boland CR, Thibodeau SN, Hamilton SR *et al*: 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* 1998; **58**: 5248–5257.
- 4 Viana-Pereira M, Lee A, Popov S *et al*: Microsatellite instability in pediatric high grade glioma is associated with genomic profile and differential target gene inactivation. *PLoS One* 2011; **6**: e20588.
- 5 Boland CR, Goel A: Microsatellite instability in colorectal cancer. *Gastroenterology* 2010; **138**: 2073–2087, e2073.
- 6 Thibodeau SN, Bren G, Schaid D: Microsatellite instability in cancer of the proximal colon. *Science* 1993; **260**: 816–819.
- 7 Vilkki S, Launonen V, Karhu A, Sistonen P, Vastrik I, Aaltonen LA: Screening for microsatellite instability target genes in colorectal cancers. *J Med Genet* 2002; **39**: 785–789.

- 8 Suraweera N, Duval A, Reperant M *et al*: Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology* 2002; **123**: 1804–1811.
- 9 Fearon ER: Molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011; **6**: 479–507.
- 10 Calin GA, Gafa R, Tibiletti MG *et al*: Genetic progression in microsatellite instability high (MSI-H) colon cancers correlates with clinico-pathological parameters: a study of the TGRbetaR11, BAX, hMSH3, hMSH6, IGF1R and BLM genes. *Int J Cancer* 2000; **89**: 230–235.
- 11 Jass JR, Do KA, Simms LA *et al*: Morphology of sporadic colorectal cancer with DNA replication errors. *Gut* 1998; **42**: 673–679.
- 12 Ribic CM, Sargent DJ, Moore MJ *et al*: Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* 2003; **349**: 247–257.
- 13 Sinicrope FA, Sargent DJ: Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. *Clin Cancer Res* 2012; **18**: 1506–1512.
- 14 Vilar E, Tabernero J: Molecular dissection of microsatellite unstable colorectal cancer. *Cancer Discov* 2013; **3**: 502–511.
- 15 Dorard C, de Thonel A, Collura A *et al*: Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis. *Nat Med* 2011; **17**: 1283–1289.
- 16 Alhopuro P, Sammalkorpi H, Niittymäki I *et al*: Candidate driver genes in microsatellite-unstable colorectal cancer. *Int J Cancer* 2012; **130**: 1558–1566.
- 17 Mori Y, Yin J, Rashid A *et al*: Instability typing: comprehensive identification of frameshift mutations caused by coding region microsatellite instability. *Cancer Res* 2001; **61**: 6046–6049.
- 18 Oliveira C, Pinto M, Duval A *et al*: BRAF mutations characterize colon but not gastric cancer with mismatch repair deficiency. *Oncogene* 2003; **22**: 9192–9196.
- 19 Parsons DW, Wang TL, Samuels Y *et al*: Colorectal cancer: mutations in a signalling pathway. *Nature* 2005; **436**: 792.
- 20 Goel A, Arnold CN, Niedzwiecki D *et al*: Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res* 2004; **64**: 3014–3021.
- 21 Koopman M, Kortman GA, Mekenkamp L *et al*: Deficient mismatch repair system in patients with sporadic advanced colorectal cancer. *Br J Cancer* 2009; **100**: 266–273.
- 22 Roth AD, Tejpar S, Delorenzi M *et al*: Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J Clin Oncol* 2010; **28**: 466–474.
- 23 Elsaleh H, Joseph D, Griew F, Zeps N, Spry N, Iacopetta B: Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet* 2000; **355**: 1745–1750.
- 24 Sargent DJ, Marsoni S, Monges G *et al*: Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol* 2010; **28**: 3219–3226.
- 25 Sinicrope FA, Foster NR, Thibodeau SN *et al*: DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy. *J Natl Cancer Inst* 2011; **103**: 863–875.
- 26 Des Guetz G, Schischmanoff O, Nicolas P, Perret GY, Morere JF, Uzzan B: Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. *Eur J Cancer* 2009; **45**: 1890–1896.
- 27 Umar A, Boland CR, Terdiman JP *et al*: Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004; **96**: 261–268.
- 28 Buhard O, Cattaneo F, Wong YF *et al*: Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. *J Clin Oncol* 2006; **24**: 241–251.
- 29 Goel A, Nagasaka T, Hamelin R, Boland CR: An optimized pentaplex PCR for detecting DNA mismatch repair-deficient colorectal cancers. *PLoS ONE* 2010; **5**: e9393.
- 30 Patil DT, Bronner MP, Portier BP, Fraser CR, Plesec TP, Liu X: A five-marker panel in a multiplex PCR accurately detects microsatellite instability-high colorectal tumors without control DNA. *Diagn Mol Pathol* 2012; **21**: 127–133.
- 31 Viana-Pereira M, Almeida I, Sousa S *et al*: Analysis of microsatellite instability in medulloblastoma. *Neuro Oncol* 2009; **11**: 458–467.
- 32 Xicola RM, Llor X, Pons E *et al*: Performance of different microsatellite marker panels for detection of mismatch repair-deficient colorectal tumors. *J Natl Cancer Inst* 2007; **99**: 244–252.
- 33 Pereira R, Phillips C, Pinto N *et al*: Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One* 2012; **7**: e29684.
- 34 Parra EJ, Marcini A, Akey J *et al*: Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 1998; **63**: 1839–1851.
- 35 Kosoy R, Nassir R, Tian C *et al*: Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat* 2009; **30**: 69–78.
- 36 Tian C, Gregersen PK, Seldin MF: Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 2008; **17**: R143–R150.
- 37 Manta FS, Pereira R, Caiafa A, Silva DA, Gusmão L, Carvalho EF: Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann Hum Biol* 2013; **40**: 94–98.
- 38 Buhard O, Suraweera N, Lactard A, Duval A, Hamelin R: Quasimonomorphic mononucleotide repeats for high-level microsatellite instability analysis. *Dis Markers* 2004; **20**: 251–257.
- 39 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.
- 40 Falush D, Stephens M, Pritchard JK: Inference of population structure using multi-locus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; **164**: 1567–1587.
- 41 Berg AO, Armstrong K, Botkin J *et al*: Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet Med* 2009; **11**: 35–41.
- 42 Pena SD, Di Pietro G, Fuchshuber-Moraes M *et al*: The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 2011; **6**: e17063.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)