8-1-2009

Intra-tumor heterogeneity of MLH1 promoter methylation revealed by deep single molecule bisulfite sequencing.

Katherine E Varley
Department of Genetics, Center for Genome Sciences, Washington University School of Medicine

David G Mutch
Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Washington University School of Medicine

Tina B Edmonston
Thomas Jefferson University, Department of Pathology, Anatomy and Cell Biology

Paul J Goodfellow
Department of Surgery, Washington University School of Medicine

Robi D Mitra
Department of Genetics, Center for Genome Sciences, Washington University School of Medicine

Published online 3 June 2009

Nucleic Acids Research, 2009, Vol. 37, No. 14 ***4603–4612***
doi:10.1093/nar/gkp457

Intra-tumor heterogeneity of *MLH1* promoter methylation revealed by deep single molecule bisulfite sequencing

Katherine E. Varley[1], David G. Mutch[2], Tina B. Edmonston[3], Paul J. Goodfellow[4] and Robi D. Mitra[1,*]

[1]Department of Genetics, Center for Genome Sciences, Washington University School of Medicine, [2]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Washington University School of Medicine, St Louis, Missouri, [3]Thomas Jefferson University, Department of Pathology, Anatomy and Cell Biology, Philadelphia, PA and [4]Department of Surgery, Washington University School of Medicine, St Louis, MO, USA

Received November 13, 2008; Revised May 11, 2009; Accepted May 14, 2009

ABSTRACT

A single tumor may contain cells with different somatic mutations. By characterizing this genetic heterogeneity within tumors, advances have been made in the prognosis, treatment and understanding of tumorigenesis. In contrast, the extent of epigenetic intra-tumor heterogeneity and how it influences tumor biology is under-explored. We have characterized epigenetic heterogeneity within individual tumors using next-generation sequencing. We used deep single molecule bisulfite sequencing and sample-specific DNA barcodes to determine the spectrum of *MLH1* promoter methylation across an average of 1000 molecules in each of 33 individual samples in parallel, including endometrial cancer, matched blood and normal endometrium. This first glimpse, deep into each tumor, revealed unexpectedly heterogeneous patterns of methylation at the *MLH1* promoter within a subset of endometrial tumors. This high-resolution analysis allowed us to measure the clonality of methylation in individual tumors and gain insight into the accumulation of aberrant promoter methylation on both alleles during tumorigenesis.

INTRODUCTION

Tumors are often genetically heterogeneous; a phenomenon in which cells from the same tumor contain different sets of somatic mutations (1–3). The study of genetic heterogeneity can provide insight into the dynamics of tumor development and the order in which mutations occur during tumorigenesis (4,5). Tumor heterogeneity accounts for a variety of clinically defined phenotypes, including outcome. In lung cancer and chronic myeloid leukemia, resistance to kinase inhibitors is associated with expansion of rare populations of cells that carry a drug resistant second-site mutation in addition to the original activating mutation (6–9). Significantly higher risk for progression to cancer is associated with greater mutation diversity in Barrett's esophagus, the premalignant precursor esophageal adenocarcinoma (10).

Epigenetic heterogeneity is likely to play a similarly important role in tumor development and response to therapy. Epigenetic mutations (epimutations) can phenocopy genetic mutations. Aberrant promoter methylation and associated silencing of tumor suppressor genes can provide a selective advantage to neoplastic cells (11). In recent years, the significance of epigenetic defects in cancer biology has become evident. Aberrant DNA methylation has been observed in all types of cancer cells thus far examined and is frequently associated with inappropriate transcriptional gene silencing (11). However, intra-tumor heterogeneity of promoter methylation has only rarely been examined (12–15), in part due to a lack of adequate technology. Therefore, we sought a method that could sensitively characterize the heterogeneity of DNA methylation in many individual tumors in parallel. Deep single molecule bisulfite sequencing using next-generation machines has recently been applied to sequence promoter methylation in cancer (14,16). Taylor and colleagues' seminal work demonstrated that disease-specific tags and 454 sequencing can be used to identify methylation patterns that differ between types of leukemia and lymphoma (16). Korshunova *et al.* (14) incorporated sample-specific barcodes with 454 sequencing and found complex methylation in breast cancer and sera DNA at biomarker loci. We sought to adapt these methods to obtain a high-resolution profile of intra-tumor heterogeneity deep within individual tumors to begin to discern

*To whom correspondence should be addressed. Tel: +314 362 2751; Fax: +314 362 2157; Email: rmitra@genetics.wustl.edu

how aberrant DNA methylation accumulates at a causative locus.

We focused our efforts on characterizing *MLH1* promoter methylation heterogeneity in endometrial tumors. Germline mutations in *MLH1*, a DNA mismatch repair gene, result in hereditary colorectal and endometrial cancers with microsatellite instability (MSI) (17–19). Sporadic endometrial cancers that have lost DNA mismatch repair frequently exhibit promoter hypermethylation and concomitant silencing of *MLH1*, leading to a mutator phenotype, referred to as MSI in these tumors. Over 70% of MSI positive endometrial cancers have hypermethylation of the *MLH1* promoter (20). *MLH1* methylation is thought to be an early event in sporadic microsatellite unstable endometrial cancer that contributes to clonal expansion (21,22). *MLH1* promoter methylation is clearly a key event in the development of many endometrial cancers; however, the spectrum of *MLH1* promoter methylation within individual tumors has never been examined. Characterizing the heterogeneity of *MLH1* promoter methylation for thousands of single molecules in individual endometrial tumors will provide more information about the timing and variability of this event in tumor development.

MATERIALS AND METHODS

Sample acquisition, DNA extraction, bisulfite treatment

Endometrial tissue specimens and blood were obtained at the time of surgery, snap frozen and stored at −70°C (IRB approval 93-0828). Tumors were histologically evaluated to ensure high neoplastic cellularity for the tissues used for DNA preparations. DNA was prepared using proteinase K and phenol extraction or with the DNeasy Tissue Kit (Qiagen Inc, Valencia, CA). DNA was extracted from matched peripheral blood leukocytes as previously described (23,24) or using DNeasy Tissue Kits (Qiagen Inc, Valencia, CA). The quality of the genomic DNA was assessed by measuring absorbance, and samples were required to have an A260/280 ratio of 1.77–1.85 to be included in the study. Genomic DNA was also analyzed on a 1% agarose gel to ensure that it was present in high molecular weight fragments (>5 kb). Genomic DNA (250 ng) from each of 33 samples was sodium bisulfite treated in parallel using the EZ DNA Methylation-Gold Kit (Zymo Research Corp, Orange, CA).

Amplicon design, PCR, FLX sequencing

Primers were designed to PCR amplify the 700-bp promoter upstream of the *MLH1* transcription start site (Refseq NM_000249 UCSC Human Genome March 2006 Assembly http://genome.ucsc.edu/). PCR amplification was performed from each sodium bisulfite treated DNA sample in two separate reactions, a Distal PCR and a Proximal PCR. The Distal PCR primer sequences were 5′ AGTAGTTTTTTTTTAGGAGTGAAGGAG GTTA 3′ and 5′ CTTCTCAAACTCCTCCTCTCCCCT TA 3′. The Proximal PCR primer sequences were 5′ TA AGGGGAGAGGAGGAGTTTGAGAAG3′ and 5′ AA AATACCTTCAACCAATCACCTCAATACCT 3′. The PCR was performed with primers specific to the locus that were also tailed with a sample-specific DNA barcode sequence and a 454 Life Sciences machine specific primer (Supplementary Data 1). There are 1024 possible 5-bp DNA sequences, and we selected 33 sample-specific barcodes, one for each sample, that did not contain homopolymers and had the least sequence similarity to each other. The PCR for each locus in each sample was performed in a total volume of 50 µl. The reaction contained 1× PCR Buffer $MgCl_2$ (Invitrogen, Carlsbad, CA), 10 units Platinum Taq Polymerase (Invitrogen Carlsbad, CA), 0.5 mM each dNTP, 1M Betaine, 0.5 µM Forward Primer, 0.5 µM Reverse Primer and 125 ng bisulfite treated genomic DNA. This reaction was incubated at 93°C for 2 min, followed by (93°C for 2 min, 55°C for 6 min) × 30 cycles, and held at 4°C. One-fifth of the PCR reaction for each of 66 reactions was quantified by electrophoresis on a 2% agarose gel. This was achieved by quantitatively comparing the intensity of the gel band containing the PCR product to the intensity of the similar size band in the Low Molecular Weight DNA Ladder (NEB). The quantity of PCR product is computed by dividing the PCR product intensity by the scaling factor, which is the ladder band intensity divided by the ladder band molecular weight. Equimolar quantities of each PCR product were then pooled into a single tube, purified on a Qiaquick column (Qiagen Inc, Valencia, CA) and submitted to Cogenics Inc. (www.cogenics.com) for sequencing on the 454 Life Sciences FLX machine.

Cloning and Sanger sequencing

We cloned and sequenced 45 molecules from the distal and proximal amplicons from both the endometrial cancer sample and the matched normal blood from a single patient for comparison between sequencing methods. To clone the PCR products we ligated them into the pGEM-T Easy Vector using Rapid Ligation Buffer according to the manufacturer's instructions (Promega, Madison, WI). We then transformed the ligated vector into GC10 Competent Cells (Gene Choice) and grew them overnight on LB-agar (Luria-Broth) plates containing standard concentrations of carbenicillin, X-gal and IPTG. After overnight growth, colonies were picked from the plates and added to 50 µl colony PCR reactions containing 1× PCR Reaction Buffer (Sigma, St Louis, MO), 1.25 units Jumpstart Taq Polymerase (Sigma), 0.2 mM each dNTP, 0.5 µM M13 Forward Primer (5′ CGCCAGGGTTTTCC CAGTCACGAC 3′), 0.5 µM M13 Reverse Primer (5′ TC ACACAGGAAACAGCTATGAC 3′) and 0.01% Tween. The reaction was incubated at 94°C for 10 min, followed by (94°C for 1 min 30 s, 55°C for 1 min, 72°C for 1 min) × 35 cycles and held at 4°C. These reactions were then treated with 10 µl ExoSAP to degrade the remaining primers and nucleotides by adding 0.2 units Exonuclease I (USB, Cleveland, OH) and 0.2 units Shrimp Alkaline Phosphatase (SAP) (Promega, Madison, WI) in 1× SAP buffer (Promega, Madison, WI), incubating at 37°C for 30 min, then heat inactivate by incubating at 80°C for 30 min. The Sanger sequencing/cycle sequencing reactions were 20 µl and contained 1.5 µl

Exo-SAP Treated colony PCR, 1 µl Big Dye Terminator v3.1 RR-100 Mix (Applied Biosystems, Foster City, CA), 2 mM MgCl2 and 0.16 µM M13 Forward Primer. They were incubated at 96°C for 1 min, followed by (96°C for 10 s, 50°C for 5 s, 60°C for 4 min) × 24 cycles, and held at 4°C. The reactions were ethanol precipitated with sodium acetate and submitted to the Washington University Genome Sequencing Center to load on the ABI 3730 (Applied Biosystems, Foster City, CA).

Sequence analysis

We obtained 180 reads from Sanger sequencing from the distal and proximal regions in both the tumor sample and normal blood sample of a single patient. To calculate the correlation coefficient between the results from the Sanger sequencing and the FLX sequencing of the same samples we first calculated the percent of methylated reads at each CG position in each sample and stored these values in a vector. We then calculated the Pearson correlation coefficient between the vectors from the Sanger sequencing results and the vectors from the FLX sequencing.

We obtained more than 33 000 reads from FLX sequencing that had perfect matches to either the distal or proximal *MLH1* promoter PCR primers. We determined which reads came from each sample based on exact matches to a sample-specific DNA barcode that was the first five bases in each read. We created a completely methylated bisulfite treated reference sequence by substituting a T for all C's in the sequence that were not in CG dinucleotides. We aligned all the reads from a particular sample and locus to the reference sequence using CLUSTALW (25). We computationally extracted each column in the multiple sequence alignment that had a C in the reference sequence. We substituted a 1 for C, a 0 for T and a 2 for any other base or a gap. We visualized this matrix as an image using the Matlab (The Mathworks, Natick, MA) software package.

MSI analysis, COBRA and mismatch repair protein immunohistochemistry

MSI analysis was performed using five National Cancer Institute consensus microsatellite markers (*BAT25, BAT26, D2S123, D5S346 and D17S250*) as previously described (26,27). Tumors were classified as microsatellite stable if there was no evidence of MSI in any marker. Tumors were designated as having high-level MSI (MSI+) if aberrant marker alleles were identified in at least two of the five consensus panel markers.

The COBRA method was used to evaluate methylation of the *MLH1* promoter (28). Briefly, tumor DNA was bisulfite converted using commercially available kits (CpGenome DNA Modification Kit, Intergen Company, Purchase, NY and EZ DNA Methylation Gold Kit, Zymo Research, Orange, CA) following manufacturers' recommended protocols. Target CpG-rich sequences in the *MLH1* promoter region were amplified using two rounds of PCR. Restriction digest of PCR products was then undertaken using enzymes that recognize sequences potentially altered by methylation. PCR primers and conditions, as well as restriction enzymes for this assay have been previously published by our group (29).

Immunohistochemistry for MLH1, MSH2, MSH6 and PMS2 was performed using formalin-fixed, paraffin-embedded tissue sections. Tissues were stained with antibodies to hMLH1 [clone G168 728; 1 mg/ml (Pharmingen)], hMSH2 [clone FE11; 0.5 mg/ml (Oncogene Science)], hMSH6 [clone 44, 0.5 µg/ml (Transduction Laboratories)] and hPMS2 (clone A16-4, BD Pharmingen) as previously described by our group (30). The expression status for each tumor was based on nuclear staining in tumor cells. Tumors that lacked nuclear staining but had normal positive staining in surrounding non-neoplastic cells were classified as negative for expression of the respective proteins.

RESULTS

Experimental design to measure intra-tumor heterogeneity of *MLH1* promoter methylation in endometrial cancer

To characterize the epigenetic heterogeneity at the *MLH1* locus in endometrial tumors, we bisulfite treated genomic DNA from 14 endometrial cancers, 14 matched bloods and five normal endometrial tissue. To determine if deep single molecule bisulfite sequencing could detect methylation that the COBRA (28) method could not, we sequenced six tumors that were unmethylated according to COBRA and eight tumors that were methylated according to COBRA. We also included normal endometrial tissue from healthy control patients in this study to determine the background level of *MLH1* promoter methylation by deep sequencing. To determine if we could detect methylated tumor DNA in blood of individuals with endometrial cancer we included matched blood samples.

Sodium bisulfite treatment converts unmethylated cytosine into uracil while leaving methylated cytosine intact, allowing us to determine the pattern of cytosine methylation by sequencing (Figure 1A) (31,32). We performed PCR to amplify the desired loci from each sample. Because the average read length on the 454 Life Sciences Inc. FLX is ~350 bp, we separately amplified the distal and proximal regions of the 700 bp promoter upstream of the *MLH1* transcription start site (RefSeq # NM_000249). It has previously been shown that methylation within this region of the promoter is correlated with gene expression; a 70 bp section in the proximal promoter at −209 to −139 bp relative to the transcriptional start site (Refseq NM_000249 UCSC Human Genome March 2006 Assembly) was particularly well correlated (33). The locus-specific PCR primers were tailed at the 5′ end with sample-specific DNA barcodes and 454 sequencing primers (Figure 1B). Equimolar quantities of each PCR product were pooled in a single tube ($N = 66$, two amplicons for each of 33 starting templates) and sequenced from both ends using 454 FLX sequencer (Figure1 C). More than 33 000 reads were obtained. The first five bases of each sequence are sample-specific barcodes that indicate the sample of origin and the remaining bases reveal the

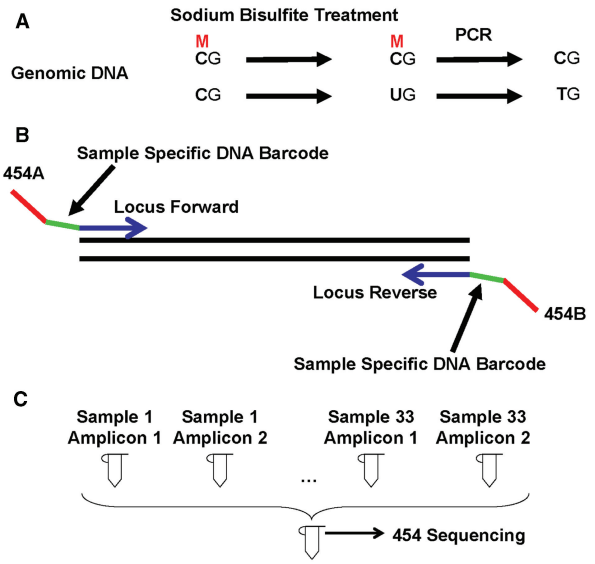

Figure 1. Experimental design for deep bisulfite sequencing in individual samples. (**A**) Detection of DNA methylation by sodium bisulfite treatment and PCR. (**B**) PCR amplification with sample specific DNA barcodes. (**C**) Pool PCR products for 454 sequencing.

methylation status of the DNA molecule, as well as any SNPs in the sequence.

Comparison to conventional Sanger sequencing

The standard method for determining methylation patterns on single molecules is bisulfite treatment followed by cloning and Sanger sequencing. Single molecule PCR methods (34,35) have been recently introduced as simplified workflow and to mitigate cloning bias. We compared the methylation patterns obtained using the standard method to those obtained using deep single molecule sequencing on second-generation sequencing machines. We cloned bisulfite-treated DNA amplified from the tumor and matched blood of a single patient (Patient #1684) and Sanger sequenced 45 clones from both the distal and proximal region of the promoter (Figure 2A). Results for conventional sequencing analysis were compared to the 454 FLX sequencing of the same samples, comprising 496 reads from the distal region of the *MLH1* promoter and 480 reads from the proximal region (Figure 2B). The 454 FLX sequencing and Sanger sequencing results were nearly identical. To quantify the similarity between the results from the two methods we calculated the percent of cytosines methylated at each CG in each sample and determined the correlation coefficient for the two analytic approaches. The methods produce highly similar methylation patterns ($R^2 = 0.96$). Both methods revealed the majority of tumor DNA molecules from the *MLH1* promoter were densely methylated. A small percentage of molecules in the tumor were unmethylated. These unmethylated sequences are likely derived from normal non-neoplastic cells in the tumor. The neoplastic cellularity (NPC) of tumor 1684 was estimated to be 70% based on histological assessment of the tissue used for DNA preparation. Both 454 FLX and Sanger sequencing revealed dense methylation of the proximal promoter

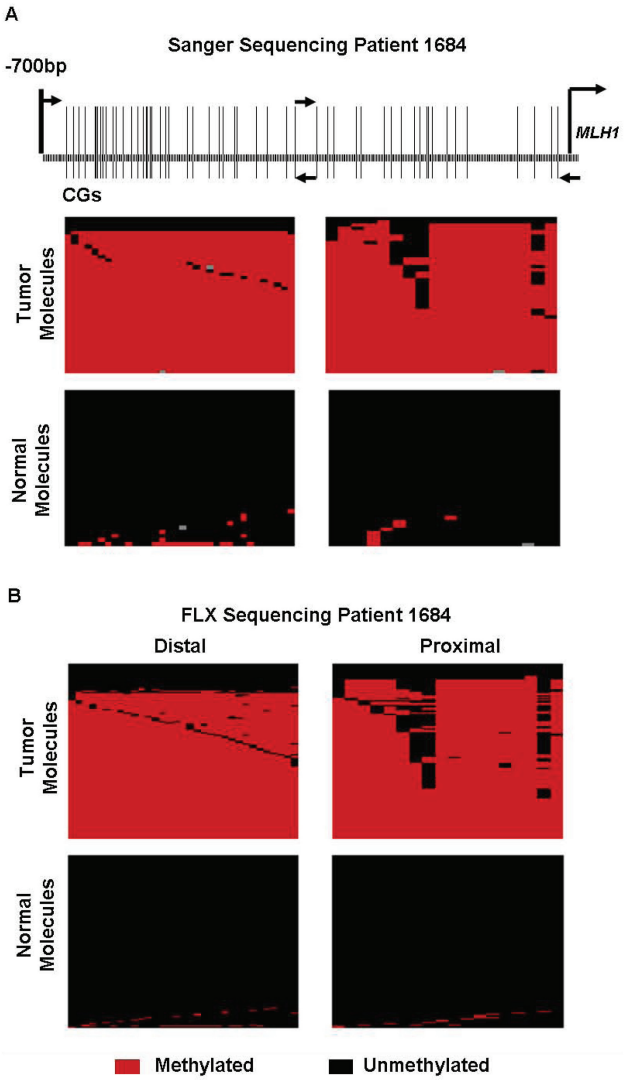

Figure 2. Comparison of sequencing results from conventional cloning and Sanger sequencing and 454 Life Sciences FLX sequencing of *MLH1* promoter PCR products from bisulfite treated DNA of an endometrial cancer and normal blood from Patient #1684. (**A**) Schematic of the *MLH1* promoter is presented with arrows indicating the location of PCR primers and vertical lines representing the position of CG dinucleotides. Below the schematic are the results from cloning and bisulfite sequencing 45 molecules from the distal (left) and proximal (right) promoter in each sample (tumor and matched normal blood). Each column represents a CG dinucleotide in the sequence, and corresponds to each vertical line in the promoter schematic. Each row represents a single molecule. The color of the boxes represents the methylation state of each cytosine. Red, methylated; Black, unmethylated. (**B**) The results from the FLX single molecule sequencing of the same samples. The distance of each CpG from the transcription start site (UCSC Human Genome March 2006) is listed from distal to proximal. The Distal Amplicon: −671, −662, −654, −648, −634, −632, −630, −626, −623, −619, −609, −605, −596, −584, −576, −569, −566, −564, −560, −558, −548, −540, −537, −512, −505, −483, −470, −465, −449, −446, −421, −405, −380, −368. The Proximal Amplicon: −341, −325, −318, −286, −280, −249, −240, −226, −209, −202, −192, −190, −184, −165, −154, −139, −70, −47, −23, −15. As an additional point of reference, the translation start site (ATG) is currently annotated at +60 downstream of the transcription start site (position 0).

Table 1. Comparison of FLX bisulfite sequencing to traditional measurements in endometrial cancer tumor samples

Tumor sample no.	Percent of methylated molecules in proximal promoter	COBRA	MSI	MLH1 IHC	NPC (%)	Age	Stage	Grade	Heterogenous methylation patterns
1499	0	u	–	+	90	77.53	IA	1	–
1472	0	u	–	+	85	62.02	IB	1	–
1487	0	u	–	+	100	81.12	IB	1	–
1556	0	u	–	+	70	82.55	IVB	1	–
1673	0	u	–	+	75	74.5	IVB	1	–
1758	0	u	+	–	60	59.47	IB	3	–
1645	15	m	+	–	80	61.47	IC	2	Y
1789	48	m	+	+	90	61.52	IVB	2	N
1727	49	m	+	–	75	75.58	IC	1	N
1576	54	m	+	–	70	55.91	IC	2	Y
1495	59	m	+	–	70	49.96	IIA	1	Y
1669	65	m	+	–	80	90.73	IVB	3	N
1569	77	m	+	–	85	81.05	IIIA	1	N
1684	86	m	+	–	>70	72.32	IB	1	Y

FLX bisulfite sequencing was summarized by calculating the fraction of molecules from the proximal promoter with >50% of CpGs methylated. COBRA is reported as u, unmethylated; m, methylated; MSI, Microsatellite Instability is reported as –, stable; +, unstable; MLH1 IHC, MLH1 protein immunohistochemistry; NPC (%), estimated neoplastic cellularity by microscopy. The age of the patient at diagnosis as well as the stage and grade of the tumor are listed in columns 7–9. Shading distinguishes classes of tumors; Grey, unmethylated, active MLH1; White, unmethylated, inactivated MLH1; Red, methylated. The presence of heterogeneous patterns of methylation within the tumor is indicated by a Y in the last column, absence of heterogeneity is indicated by an N.

with heterogeneous methylation of five CpG positions. The normal blood from Patient #1648 was unmethylated in the MLH1 promoter as assessed by both 454 FLX and Sanger sequencing (Figure 2A). We found that >99% of non-CpG cytosines were converted in each sample, indicating the sodium bisulfite conversion was successful. The similarity between the Sanger and FLX sequencing for this patient's samples confirmed that the FLX sequencing strategy can be used to bisulfite sequence single molecules in individual samples in high-throughput.

MLH1 promoter methylation in MLH1 deficient endometrial tumors is heterogeneous

Tumors lacking MLH1 mismatch repair activity have either loss-of-function mutations or epigenetic silencing of *MLH1*(17,18,20–22). We studied nine MSI-positive primary endometrial cancers that lacked MLH1 expression based on immunohistochemical evaluation. Eight of nine tumors had methylation of the proximal *MLH1* promoter based on COBRA (28,36) (Table 1). An accepted model of endometrial tumorigenesis is that a single somatic cell acquires dense promoter methylation with the inactivation of *MLH1*. The inactivation confers a selective advantage to the cell. The cell then undergoes a clonal expansion, propagating the aberrant *MLH1* promoter methylation through division (37,38). Therefore we expected to observe a homogeneous population of densely methylated *MLH1* promoter molecules in these tumors.

In four of eight of the *MLH1* methylated tumors analyzed, we observed the expected dense homogeneous methylation (Table 1). For example, the tumor from patient #1569 is densely methylated at all CpGs in both the distal and proximal regions of the *MLH1* promoter (Figure 3A). In tumors 1569, 1669, 1727 and 1789, dense *MLH1* promoter methylation appears to have been an early event propagated throughout the tumor during the subsequent clonal expansion.

In four of eight of the *MLH1* methylated tumors analyzed, we observed heterogeneity in the pattern of methylation across the molecules within individual tumors (Figure 3C, D, E and F). For example, we observed two distinct patterns of methylation in the distal region of the *MLH1* promoter in tumor of Patient #1495 (Figure 3C). Although the distal promoter is heavily methylated, ~20% of the methylated molecules exhibit a distinct alternating (checkerboard) pattern of unmethylated CpGs. The distinct pattern of methylation for a subpopulation of sequences indicates that the tumor is not clonal for methylation. Because demethylation is believed to be a rare event (39,40), the pattern of alternating methylated and un-methylated CpGs likely resembles the methylation state in the initiating tumor cell. The observed heterogeneity could be explained by expansion of this initial tumor cell followed by methylation of additional CpGs along certain lineages. Interestingly, the four tumors that were identified as having heterogeneous patterns of methylation were from younger women with earlier stage disease compared to the four cases with homogeneous patterns of MLH1 methylation (Table 1).

Each tumor that displayed epigenetic heterogeneity at the MLH1 locus had a unique signature of methylation. Heterogeneity occurred in either the distal region (Figure 3C and E) or in the proximal promoter region (Figure 3D and F). All four tumors with heterogeneous methylation do not express MLH1 as assessed by IHC (Table 1). Although the same set of CpGs are not methylated in every tumor, all of the tumors have >50% of CpG positions methylated at this locus. These data support the model, proposed by others, that it is the degree of methylation across the molecule rather than site-specific methylation that is associated with gene silencing (41–44).

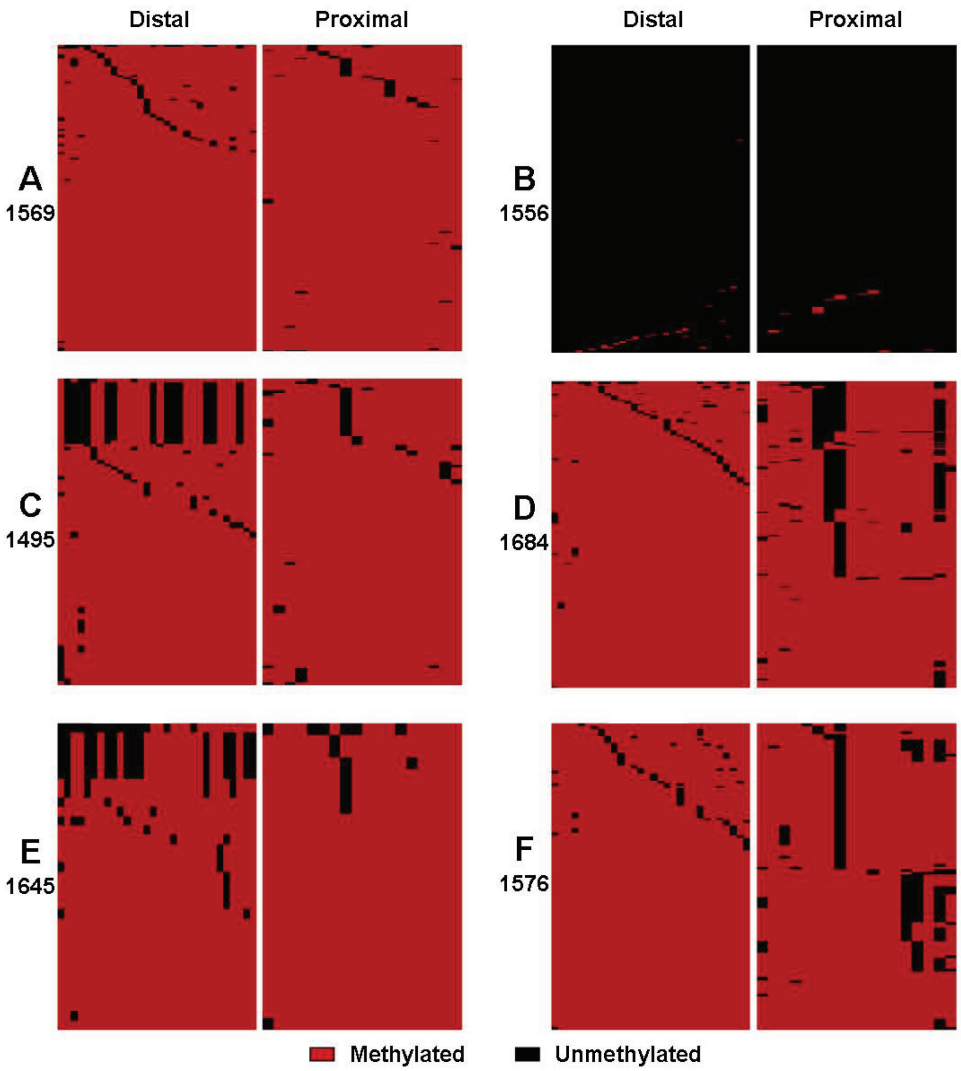

Figure 3. Representation of the different types of methylation observed in endometrial cancer specimens. The numerical patient identifier is followed by the bisulfite sequencing results for the distal and proximal *MLH1* promoter in the tumor. (**A**) Methylated molecules from a tumor with dense homogenous promoter methylation. (**B**) Homogenous tumor with no promoter methylation. (**C**) and (**E**), Methylated molecules from tumors with distinct heterogeneous patterns of unmethylated cytosines in distal promoter. (**D**) and (**F**), Methylated molecules from tumors with distinct heterogeneous patterns of unmethylated cytosines in proximal promoter. Completely unmethylated molecules from the tumors in A, C, D, E and F were not included to allow for better resolution of the methylation patterns. Each column represents a CG dinucleotide. Each row represents a single molecule. The color of the boxes represents the methylation state of each cytosine. Red, methylated; Black, unmethylated.

Lack of *MLH1* promoter methylation in endometrial cancer patient blood, healthy control endometrial tissue and MLH1 expressing tumors

Since deep single molecule bisulfite sequencing provides the opportunity to detect rare methylated molecules and methylation at sparse CpGs across a locus, we examined if we could detect methylation in 14 matched endometrial cancer patient blood samples, five healthy control endometrial tissues and six MLH1-expressing tumors. Even with deep bisulfite sequencing, we did not detect DNA methylation in any of these samples.

Allelic promoter methylation

We next asked if different alleles of the *MLH1* promoter displayed different patterns of methylation. Four of the methylated tumor samples were heterozygous for a SNP (rs1800734) in the proximal promoter region of *MLH1*, allowing us to identify the pattern of methylation for each allele (Figure 4). All four tumors have dense methylation on both alleles in near equal fractions, even when only a subclone of the tumor is methylated (Supplementary Table 1). The tumor sample from patient #1576 has a heterogeneous pattern of methylation in the proximal promoter, with distinct patterns of methylation evident for the two alleles. If the heterogeneous pattern of methylation arose randomly, both MLH1 alleles would be expected to show the heterogeneous pattern of methylation. Alternatively, if the patterns start in an initiating tumor cell and methylation accumulates as the tumor divides and expands, then the alleles should maintain their distinct patterns. Our data indicate that

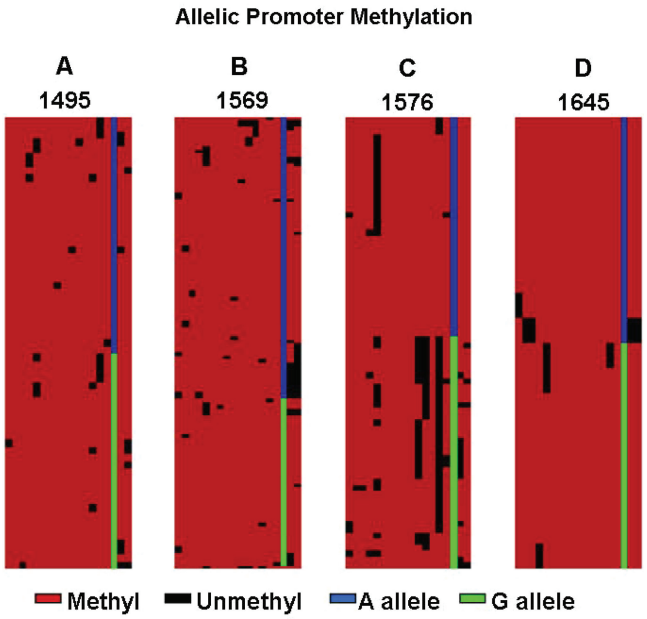

Figure 4. Patterns of methylation for different MLH1 alleles. Methylated sequences from the proximal *MLH1* promoter in four tumors heterozygous at the rs1800734 SNP are presented. Both alleles are methylated in all four patient's tumors. (**A**) Patient 1495. (**B**) Patient 1569. (**C**) Patient 1576, this tumor displays heterogeneous patterns of methylation that are distinct on the two alleles, suggesting the allelic patterns are inherited as the tumor divides. (**D**) Patient 1645. Completely unmethylated molecules from these tumors are excluded from these graphs for better resolution of methylation patterns. Each column represents a CG dinucleotide or the SNP. Each row represents a single molecule. The color of the boxes represents the methylation state of each cytosine. Red, methylated; Black, unmethylated. The position of the SNP in the sequencing reads is indicated by the green and blue column. The color of the box indicates the base at that position, Blue, A allele; Green, G allele.

heterogeneous patterns differ between alleles suggesting the allelic methylation patterns are stably inherited (Figure 4C).

Comparison of 454 sequencing to standard assays

Traditional bulk assays of the molecular characteristics of tumors have enabled the classification of tumors into subtypes. We compared our high resolution methylation measurements to the results from the traditional COBRA method for these tumors. We found that *MLH1* promoter methylation existed in two states, unmethylated or densely methylated, so we summarized our data by calculating the fraction of molecules from the proximal promoter with greater than 50% of CpGs methylated. As seen in Table 1, our metric correlates with the COBRA methylation status, MLH1 expression and MSI. A notable exception is patient 1758 whose tumor is unmethylated (according to both assays), but is not expressing MLH1 and is MSI+. This finding suggests another mechanism of MLH1 inactivation, possibly mutation. None of the patients in this study had recognized hereditary non-polyposis colorectal cancer (HNPCC), so further investigation is needed to identify the somatic mutations responsible for this loss of MLH1.

Since biallelic *MLH1* promoter methylation is thought to be an early event during tumorigenesis, all tumor cells would be expected to contain methylated molecules. Therefore, the fraction of methylated molecules is likely to represent the DNA in the sample that is from neoplastic cells. Unmethylated molecules, on the other hand, are likely derived from adjacent or infiltrating normal cells. We compared the fraction of molecules that were densely methylated with the neoplastic cellularity estimated by light microscopy of tumor tissues. We found one notable discrepancy between the measures (Table 1). In the tumor from Patient #1645, only 15% of the sequenced molecules are densely methylated, but the neoplastic cellularity was estimated to be 80%. In this case the 15% of molecules that are methylated may represent a distinct subclone in the tumor that has acquired aberrant promoter methylation. This methylation appears on both alleles in near equal fractions (18% of the molecules from the A allele and 11% of the molecules from the G allele) indicating both alleles were methylated in the subclone of the tumor. Under this scenario, it is assumed that a different tumor initiating event caused MLH1 to be silenced, resulting in an MSI+ tumor. This patient's family history of cancer included a sister with endometrial cancer and other most distantly related family members with cancers. Although her history does not fulfill the established clinical criteria for HNPCC, it is likely that she had an inherited cancer susceptibility and that her tumor was caused by mutation of MLH1. The observed methylation in her tumor is consistent with recent reports in colorectal cancer that rare methylation events can be found in cancers caused by germline mutations (45). The sensitive resolution of the deep single molecule bisulfite sequencing made it possible to discover this exceptional case, a tumor that is polyclonal for *MLH1* promoter methylation.

Evaluation of PCR Bias

The biased amplification of certain molecules during PCR could result in a small number of template molecules generating the bulk of the sequencing reads. The presence of PCR bias would prevent the accurate assessment of intratumor heterogeneity. There are two complimentary methods to evaluate whether sequencing reads were generated from a biased PCR amplification of a limited number of template molecules or whether sequencing reads were generated from diverse template molecules and are representative of the sample. First, for samples that are heterozygous for a SNP, if one template molecule is amplified preferentially then the ratio of reads from each allele would deviate from the expected 50%. Four of the patients we evaluated were heterozygous for a SNP in the proximal promoter. For each patient we counted the number of reads from each allele. In each sample, the fraction of reads from the A allele is: 1495 47%, 1569 56%, 1576 50%, 1645 50%. These results indicate an absence of PCR bias, since each alleles is represented at the expected frequency (50%). The second way to determine if particular template molecules are overrepresented in the sequencing reads is to examine the patterns of cytosines in the sequence. Since the sodium bisulfite conversion

of the samples in this study was ~99% efficient, some cytosines in individual template molecules were not converted, giving some molecules unique patterns of unconverted cytosines. If PCR bias occurs then the patterns of cytosines in the sequencing reads would be clonal. In contrast, if the sequencing reads originate from diverse molecules, then we expect to see diverse patterns of unconverted cytosines. To analyze the diversity of cytosine patterns in the sequencing reads, we determined the fraction of reads that contained at least one cytosine. These are the reads that we can evaluate for diversity or clonality. We next calculated the fraction of reads with cytosines that have a unique pattern (Supplementary Table 2). We found that, in each sample, greater than 93% of the reads with cytosines had distinct patterns. Thus, we concluded that our sequencing reads were generated from a diverse pool of starting molecules.

DISCUSSION

We used deep single molecule bisulfite sequencing and sample-specific DNA barcodes to reveal the spectrum of *MLH1* promoter methylation across an average of 1000 molecules in each of 33 individual samples in parallel, including endometrial cancer, matched blood and normal endometrium. This high-resolution analysis allowed us to measure the clonality of methylation in tumors and gain insight into the accumulation of aberrant promoter methylation during tumorigenesis.

MLH1 promoter methylation is a tumor initiating event in sporadic MSI+ endometrial cancers. As such, the *MLH1* promoter methylation is expected to be uniform and dense throughout these tumors. Deep single molecule bisulfite sequencing revealed unexpected heterogeneous patterns of *MLH1* promoter methylation within individual tumors. This intra-tumor heterogeneity was found in half (four out of eight) of the tumors that were methylated at the *MLH1* promoter. We observed distinct patterns of methylation in which promoter molecules were densely methylated, but with different sets of CpGs methylated. The patterns of methylation found in each heterogeneous tumor allow us to hypothesize as to how DNA methylation accumulated. Because DNA demethylation is thought to be rare (39,40), heterogeneous patterns of methylation are likely attributable to gains in methylation rather than losses. The promoter sequences with the least (fewest) CpG methylation events likely represent the methylation pattern most similar to the one present in the tumor initiating cell. As the tumor divides, the pattern of methylation would be propagated and certain lineages in the tumor would accumulate additional methylation. Each of the four tumors with epigenetic heterogeneity displayed distinct patterns of methylation, suggesting that each tumor was initiated with a different pattern of methylation in the *MLH1* promoter. Four *MLH1* methylated endometrial tumors were heterozygous for a SNP (rs1800734) in the *MLH1* promoter, allowing us to examine allele-specific methylation. We observed dense methylation on both alleles in each of the four samples. The methylation was present on both alleles in equal fractions of the reads, even when the methylation was a late event in Patient 1645 (Supplementary Table 1). This suggests that both alleles are methylated at the same time and the methylation is propagated in the subclone of the tumor. One tumor exhibited distinct patterns of methylation on each allele. This pattern supports the model that once the alleles are methylated, the methylation pattern can be stably inherited throughout the expansion of the tumor. Further development of the methods described herein to enable high-throughput bisulfite sequencing from microdissected tumor cells would allow one to test this model.

Using next-generation sequencing to assess intra-tumor epigenetic heterogeneity we identified a new molecular subclass of MSI+ endometrial cancers: tumors with heterogeneous *MLH1* promoter methylation. Follow-up studies are needed to determine if classification of tumors based on their epigenetic heterogeneity can be used to stratify disease subtypes with distinct prognosis or responses to treatment. It will be important to have high-throughput methods, such as the one described here, to identify tumor subclasses that are defined by distinct epigenetic defects.

High-throughput deep sampling of methylation in individual tumors affords new opportunities for modeling tumorigenesis. Analysis of multiple loci from a single tumor, including neutral loci, will provide the opportunity to apply the mathematical framework of population genetics to analyze tumor development and evolution, as in the pioneering work of Shibata, Nowak and colleagues (46–50). This will be useful for determining the frequency, timing and order of aberrant methylation events during tumorigenesis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Amy Schmidt and Mary Ann Mallon for assistance in specimen preparation, MSI analysis and COBRA for the samples studied here. We thank Jason Gertz for suggestions, discussion and critical reading of the manuscript. We thank Lee Tessler, Todd Druley, German LeParc, Michael Brooks and Yue Yun for discussion. In memory of Suzanne K. Varley.

FUNDING

Siteman Cancer Center Endometrial Cancer Working Group Research Development Award; Genome Analysis Training Program [T32 HG000045]; Center for Excellence in Genome Sciences grant from National Human Genome Research Institute [5P50HG003170-03]; Epigenetics Roadmap grant [1R01DA025744-01]. Funding for open access charge: National Institutes of Health grant# 1R01DA025744.

Conflict of interest statement. None declared.

REFERENCES

1. Aubele,M., Mattis,A., Zitzelsberger,H., Walch,A., Kremer,M., Hutzler,P., Hofler,H. and Werner,M. (1999) Intratumoral heterogeneity in breast carcinoma revealed by laser-microdissection and comparative genomic hybridization. *Cancer Genet. Cytogenet.*, **110**, 94–102.
2. Baisse,B., Bouzourene,H., Saraga,E.P., Bosman,F.T. and Benhattar,J. (2001) Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int. J. Cancer*, **93**, 346–352.
3. Gonzalez-Garcia,I., Sole,R.V. and Costa,J. (2002) Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA*, **99**, 13085–13089.
4. Shibata,D. (2006) When does MMR loss occur during HNPCC progression? *Cancer Biomark.*, **2**, 29–35.
5. Shibata,D. (2008) Stem cells as common ancestors in a colorectal cancer ancestral tree. *Curr. Opin. Gastroenterol.*, **24**, 59–63.
6. Branford,S., Rudzki,Z., Walsh,S., Grigg,A., Arthur,C., Taylor,K., Herrmann,R., Lynch,K.P. and Hughes,T.P. (2002) High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. *Blood*, **99**, 3472–3475.
7. Gorre,M.E., Mohammed,M., Ellwood,K., Hsu,N., Paquette,R., Rao,P.N. and Sawyers,C.L. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, **293**, 876–880.
8. Pao,W., Miller,V.A., Politi,K.A., Riely,G.J., Somwar,R., Zakowski,M.F., Kris,M.G. and Varmus,H. (2005) Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.*, **2**, e73.
9. Shah,N.P., Nicoll,J.M., Nagar,B., Gorre,M.E., Paquette,R.L., Kuriyan,J. and Sawyers,C.L. (2002) Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*, **2**, 117–125.
10. Maley,C.C., Galipeau,P.C., Finley,J.C., Wongsurawat,V.J., Li,X., Sanchez,C.A., Paulson,T.G., Blount,P.L., Risques,R.A., Rabinovitch,P.S. *et al.* (2006) Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.*, **38**, 468–473.
11. Lyko,F. and Brown,R. (2005) DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. *J. Natl Cancer Inst.*, **97**, 1498–1506.
12. Aggerholm,A., Guldberg,P., Hokland,M. and Hokland,P. (1999) Extensive intra- and interindividual heterogeneity of p15INK4B methylation in acute myeloid leukemia. *Cancer Res.*, **59**, 436–441.
13. Bhawal,U.K., Tsukinoki,K., Sasahira,T., Sato,F., Mori,Y., Muto,N., Sugiyama,M. and Kuniyasu,H. (2007) Methylation and intratumoural heterogeneity of 14-3-3 sigma in oral cancer. *Oncology Rep.*, **18**, 817–824.
14. Korshunova,Y., Maloney,R.K., Lakey,N., Citek,R.W., Bacher,B., Budiman,A., Ordway,J.M., McCombie,W.R., Leon,J., Jeddeloh,J.A. *et al.* (2008) Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res.*, **18**, 19–29.
15. Rastetter,M., Schagdarsurengin,U., Lahtz,C., Fiedler,E., Marsch,W., Dammann,R. and Helmbold,P. (2007) Frequent intra-tumoural heterogeneity of promoter hypermethylation in malignant melanoma. *Histol. Histopathol.*, **22**, 1005–1015.
16. Taylor,K.H., Kramer,R.S., Davis,J.W., Guo,J., Duff,D.J., Xu,D., Caldwell,C.W. and Shi,H. (2007) Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res.*, **67**, 8511–8518.
17. Bronner,C.E., Baker,S.M., Morrison,P.T., Warren,G., Smith,L.G., Lescoe,M.K., Kane,M., Earabino,C., Lipford,J., Lindblom,A. *et al.* (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature*, **368**, 258–261.
18. Papadopoulos,N., Nicolaides,N.C., Wei,Y.F., Ruben,S.M., Carter,K.C., Rosen,C.A., Haseltine,W.A., Fleischmann,R.D., Fraser,C.M., Adams,M.D. *et al.* (1994) Mutation of a mutL homolog in hereditary colon cancer. *Science*, **263**, 1625–1629.
19. Lynch,H.T. and Lynch,J.F. (1994) 25 years of HNPCC. *Anticancer Res.*, **14**, 1617–1624.
20. Simpkins,S.B., Bocker,T., Swisher,E.M., Mutch,D.G., Gersell,D.J., Kovatich,A.J., Palazzo,J.P., Fishel,R. and Goodfellow,P.J. (1999) MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum. Mol. Genet.*, **8**, 661–666.
21. Esteller,M., Catasus,L., Matias-Guiu,X., Mutter,G.L., Prat,J., Baylin,S.B. and Herman,J.G. (1999) hMLH1 promoter hypermethylation is an early event in human endometrial tumorigenesis. *Am. J. Pathol.*, **155**, 1767–1772.
22. Veigl,M.L., Kasturi,L., Olechnowicz,J., Ma,A.H., Lutterbaugh,J.D., Periyasamy,S., Li,G.M., Drummond,J., Modrich,P.L., Sedwick,W.D. *et al.* (1998) Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc. Natl Acad. Sci. USA*, **95**, 8698–8702.
23. Lahiri,D.K. and Nurnberger,J.I. Jr. (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.*, **19**, 5444.
24. Miller,S.A., Dykes,D.D. and Polesky,H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.*, **16**, 1215.
25. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
26. Boland,C.R., Thibodeau,S.N., Hamilton,S.R., Sidransky,D., Eshleman,J.R., Burt,R.W., Meltzer,S.J., Rodriguez-Bigas,M.A., Fodde,R., Ranzani,G.N. *et al.* (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.
27. Kowalski,L.D., Mutch,D.G., Herzog,T.J., Rader,J.S. and Goodfellow,P.J. (1997) Mutational analysis of MLH1 and MSH2 in 25 prospectively-acquired RER+ endometrial cancers. *Genes Chromosomes Cancer*, **18**, 219–227.
28. Xiong,Z. and Laird,P.W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.*, **25**, 2532–2534.
29. Whitcomb,B.P., Mutch,D.G., Herzog,T.J., Rader,J.S., Gibb,R.K. and Goodfellow,P.J. (2003) Frequent HOXA11 and THBS2 promoter methylation, and a methylator phenotype in endometrial adenocarcinoma. *Clin. Cancer Res.*, **9**, 2277–2287.
30. Case,A.S., Zighelboim,I., Mutch,D.G., Babb,S.A., Schmidt,A.P., Whelan,A.J., Thibodeau,S.N. and Goodfellow,P.J. (2008) Clustering of Lynch syndrome malignancies with no evidence for a role of DNA mismatch repair. *Gynecol. Oncol.*, **108**, 438–444.
31. Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
32. Clark,S.J., Harrison,J., Paul,C.L. and Frommer,M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.*, **22**, 2990–2997.
33. Deng,G., Chen,A., Hong,J., Chae,H.S. and Kim,Y.S. (1999) Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. *Cancer Res.*, **59**, 2029–2033.
34. Chhibber,A. and Schroeder,B.G. (2008) Single-molecule polymerase chain reaction reduces bias: application to DNA methylation analysis by bisulfite sequencing. *Anal. Biochem.*, **377**, 46–54.
35. Weisenberger,D.J., Trinh,B.N., Campan,M., Sharma,S., Long,T.I., Ananthnarayan,S., Liang,G., Esteva,F.J., Hortobagyi,G.N., McCormick,F. *et al.* (2008) DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res.*, **36**, 4689–4698.
36. Whelan,A.J., Babb,S., Mutch,D.G., Rader,J., Herzog,T.J., Todd,C., Ivanovich,J.L. and Goodfellow,P.J. (2002) MSI in endometrial carcinoma: absence of MLH1 promoter methylation is associated with increased familial risk for cancers. *Int. J. Cancer*, **99**, 697–704.

37. Nowell,P.C. (1976) The clonal evolution of tumor cell populations. *Science*, **194**, 23–28.
38. Issa,J.P. (2000) The epigenetics of colorectal cancer. *Ann. NY Acad. Sci.*, **910**, 140–153; discussion 153–145.
39. Pfeifer,G.P., Steigerwald,S.D., Hansen,R.S., Gartler,S.M. and Riggs,A.D. (1990) Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl Acad. Sci. USA*, **87**, 8252–8256.
40. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
41. Cameron,E.E., Baylin,S.B. and Herman,J.G. (1999) p15(INK4B) CpG island methylation in primary acute leukemia is heterogeneous and suggests density as a critical factor for transcriptional silencing. *Blood*, **94**, 2445–2451.
42. Helmle,K.E., Otto,C.J., Constantinescu,G., Honore,L.H. and Andrew,S.E. (2005) Variable MLH1 promoter methylation patterns in endometrial carcinomas of endometrioid subtype lacking DNA mismatch repair. *Int. J. Gynecol. Cancer*, **15**, 1089–1096.
43. Kanaya,T., Kyo,S., Maida,Y., Yatabe,N., Tanaka,M., Nakamura,M. and Inoue,M. (2003) Frequent hypermethylation of MLH1 promoter in normal endometrium of patients with endometrial cancers. *Oncogene*, **22**, 2352–2360.
44. Strazzullo,M., Cossu,A., Baldinu,P., Colombino,M., Satta,M.P., Tanda,F., De Bonis,M.L., Cerase,A., D'Urso,M., D'Esposito,M. *et al.* (2003) High-resolution methylation analysis of the hMLH1 promoter in sporadic endometrial and colorectal carcinomas. *Cancer*, **98**, 1540–1546.
45. Bettstetter,M., Dechant,S., Ruemmele,P., Grabowski,M., Keller,G., Holinski-Feder,E., Hartmann,A., Hofstaedter,F. and Dietmaier,W. (2007) Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin. Cancer Res.*, **13**, 3221–3228.
46. Beerenwinkel,N., Antal,T., Dingli,D., Traulsen,A., Kinzler,K.W., Velculescu,V.E., Vogelstein,B. and Nowak,M.A. (2007) Genetic progression and the waiting time to cancer. *PLoS Comput. Biol.*, **3**, e225.
47. Iwasa,Y., Michor,F., Komarova,N.L. and Nowak,M.A. (2005) Population genetics of tumor suppressor genes. *J. Theoret. Biol.*, **233**, 15–23.
48. Kim,J.Y., Tavare,S. and Shibata,D. (2005) Counting human somatic cell replications: methylation mirrors endometrial stem cell divisions. *Proc. Natl Acad. Sci. USA*, **102**, 17739–17744.
49. Nicolas,P., Kim,K.M., Shibata,D. and Tavare,S. (2007) The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput. Biol.*, **3**, e28.
50. Yatabe,Y., Tavare,S. and Shibata,D. (2001) Investigating stem cells in human colon by using methylation patterns. *Proc. Natl Acad. Sci. USA*, **98**, 10839–10844.