Gene 448 (2009) 151–167

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements

Sebastian Szpakowski [a,b], Xueguang Sun [b], José M. Lage [b], Andrew Dyer [b], Jill Rubinstein [a,b,c], Diane Kowalski [b], Clarence Sasaki [d], Jose Costa [b], Paul M. Lizardi [a,b,*]

[a] Interdepartmental Program in Computational, Biology and Bioinformatics, Yale University School of Medicine, Room LH-208, 310 Cedar Street, New Haven, CT 06520, USA
[b] Department of Pathology, Yale University School of Medicine, Room LH-208, 310 Cedar Street, New Haven, CT 06520, USA
[c] M.D.-Ph.D. Program, Yale University School of Medicine, Room LH-208, 310 Cedar Street, New Haven, CT 06520, USA
[d] Department of Surgery, Yale University School of Medicine, Room LH-208, 310 Cedar Street, New Haven, CT 06520, USA

A R T I C L E I N F O

Article history:
Received 30 April 2009
Received in revised form 19 July 2009
Accepted 6 August 2009
Available online 21 August 2009

Received by D. Mager

Keywords:
Mobile DNA
Natural selection
Methylation
Epigenetic control
Disease progression
DNA repair

A B S T R A C T

Close to 50% of the human genome harbors repetitive sequences originally derived from mobile DNA elements, and in normal cells, this sequence compartment is tightly regulated by epigenetic silencing mechanisms involving chromatin-mediated repression. In cancer cells, repetitive DNA elements suffer abnormal demethylation, with potential loss of silencing. We used a genome-wide microarray approach to measure DNA methylation changes in cancers of the head and neck and to compare these changes to alterations found in adjacent non-tumor tissues. We observed specific alterations at thousands of small clusters of CpG dinucleotides associated with DNA repeats. Among the 257,599 repetitive elements probed, 5% to 8% showed disease-related DNA methylation alterations. In dysplasia, a large number of local events of loss of methylation appear in apparently stochastic fashion. Loss of DNA methylation is most pronounced for certain members of the SVA, HERV, LINE-1P, AluY, and MaLR families. The methylation levels of retrotransposons are discretely stratified, with younger elements being highly methylated in healthy tissues, while in tumors, these young elements suffer the most dramatic loss of methylation. Wilcoxon test statistics reveals that a subset of primate LINE-1 elements is demethylated preferentially in tumors, as compared to non-tumoral adjacent tissue. Sequence analysis of these strongly demethylated elements reveals genomic loci harboring full length, as opposed to truncated elements, while possible enrichment for functional LINE-1 ORFs is weaker. Our analysis suggests that, in non-tumor adjacent tissues, there is generalized and highly variable disruption of epigenetic control across the repetitive DNA compartment, while in tumor cells, a specific subset of LINE-1 retrotransposons that arose during primate evolution suffers the most dramatic DNA methylation alterations.

1. Introduction

The DNA of most tumors has a reduced content of methylated cytosine residues. This so-called global "hypomethylation" affects primarily DNA sequences that belong to interspersed DNA repeats. In normal human tissues, DNA repeats are predominantly methylated, consistent with the requirement to maintain genomic stability by transcriptional silencing of retroelements whose potential deleterious functions include DNA mobilization as well as the facilitation of recombination events in somatic cells. There have been a considerable number of reports of transcriptional activation of retrotransposons in the context of loss of DNA methylation. Expression of human endogenous retroviruses (HERVs) has been detected in breast cancer (Wang-Johanning et al., 2001), ovarian cancer (Menendez et al., 2004; Wang-Johanning et al., 2007), leukemia cell lines (Patzke et al., 2002), and urothelial and renal cell carcinomas (Florl et al., 1999). Increased transcriptional expression of HERV-K has been reported in teratocarcinoma (Löwer et al., 1984; Herbst et al., 1998), breast cancer cells and adjacent tissues (Wang-Johanning et al., 2003; Golan et al., 2008), and in melanoma (Muster et al., 2003; Büscher et al., 2006; Serafino et al., 2009). Stauffer et al. (2004) used massively parallel signature sequencing (MPSS) to define the number and type of transcripts of endogenous retroviruses of the LTR family in various cancers. This study reported that HERV-H, a relatively young retrotransposon, was expressed in cancers of the intestine, bone marrow, bladder, and cervix and was more highly expressed than the other families in cancers of the stomach, colon, and prostate. Recently Alves et al.

Abbreviations: DSB, double-strand break; EIDR, endonuclease-independent DNA repair; ERV, endogenous retrovirus; HERV, human endogenous retrovirus; HNSCC, head and neck squamous cell carcinoma; LINE, long interspersed nuclear element; LTR, long terminal repeat; MaLR, mammalian apparent LTR retrotransposons; MDA, multiple displacement amplification; MPSS, massively parallel sequencing; ORF, open reading frame; SINE, short interspersed nuclear element; SVA, SINE-VNTR-Alu; VNTR, variable-number tandem repeat.

* Corresponding author. Interdepartmental Program in Computational, Biology and Bioinformatics, Yale University School of Medicine, Room LH-208, 310 Cedar Street, New Haven, CT 06520, USA. Tel.: +1 203 785 5107; fax: +1 203 785 3583.
E-mail address: Paul.Lizardi@yale.edu (P.M. Lizardi).

0378-1119/$ – see front matter
doi:10.1016/j.gene.2009.08.006

(2008) have reported that a specific HERV-H element present in the X chromosome is selectively transcribed in 60% of colon cancers and in a high proportion of metastatic colon cancers. There is evidence for context-specific induction of LINE-1 transcription during oxidative stress (Teneng et al., 2007). In a relatively large study of squamous head and neck carcinomas, Smith et al. (2007) reported that the DNA methylation level of LINE-1 elements was significantly reduced and correlated with environmental insults such as alcohol use and smoking, as well as tumor stage.

Here we report a systematic study of DNA methylation changes occurring in the repetitive DNA compartment of squamous carcinomas of the head and neck. In contrast to previous studies, we use a novel microarray-based approach to obtain discrete DNA methylation data at hundreds of thousands of individual repetitive DNA loci in the human genome. We then use extensive annotation resources for different subfamilies of repeats to evaluate possible relationships between loss of epigenetic silencing in the context of natural history of cancer and the evolutionary history of repetitive element sub-compartments in the human genome.

2. Materials and methods

A microarray analysis method developed in our laboratory permits genome-wide assessment of DNA methylation status using restriction endonucleases (see Supplementary Materials Section 1). Among the 339,314 probes in the microarray, 257,599 are dedicated to the measurement of the methylation levels of individual members of interspersed DNA repeat families.

2.1. Principle of the DNA methylation analysis method

Multiple displacement amplification (MDA; Dean et al., 2002; Lage et al., 2003; Lage and Lizardi, 2005) is an isothermal amplification method based on random priming and DNA hyper-branching, catalyzed by a strand-displacing DNA polymerase. The yield of the MDA reaction is strongly influenced by the size of the DNA used as template (Lage and Lizardi, 2005). We have systematically studied the dependence of amplification yield using DNA templates of different size and also built a computational model of the reaction that fits the experimental data. The results of this analysis (data not shown) indicate that the yield of DNA derived from any sequence segment depends on template size and additionally on the distance of the sequence segment from the nearest DNA terminus on the template molecule. We reasoned that a specific cleavage event in a genomic DNA molecule could be detected by measuring DNA amplification yield using a DNA microarray and that a probe in the microarray would be able to measure a local reduction in sequence representation due to cleavage, even if that cleavage event occurred as far as 1200 bases upstream or downstream from the location of the probe. This property enables the use of probe designs that measure cleavage events not only in unique DNA sequences overlapping a probe but also within repetitive DNA sequences that contain CpG dinucleotides, located in the vicinity of a probe of unique sequence, within a window of approximately 2400 bases surrounding the probe. Below we will present experimental data that help to define the approximate size of the window that enables probing-at-a-distance.

2.2. Microarray probe design

DNA probes of unique sequence (uniqueness assessed using merEngine, Healy et al., 2003) were designed to map as closely as possible to every CpG island in the human genome. We examined the DNA sequences located within a window of plus or minus 4 kb from loci coding for microRNAs and noticed that many of these regions contain small clusters of CpG residues. We then created a relatively lax "CpG islet" specification, requiring that a region in the genome contain a minimum of 7 CpG residues, that the ratio of the CG count to the GC content be larger than 0.53, and that the region be no shorter

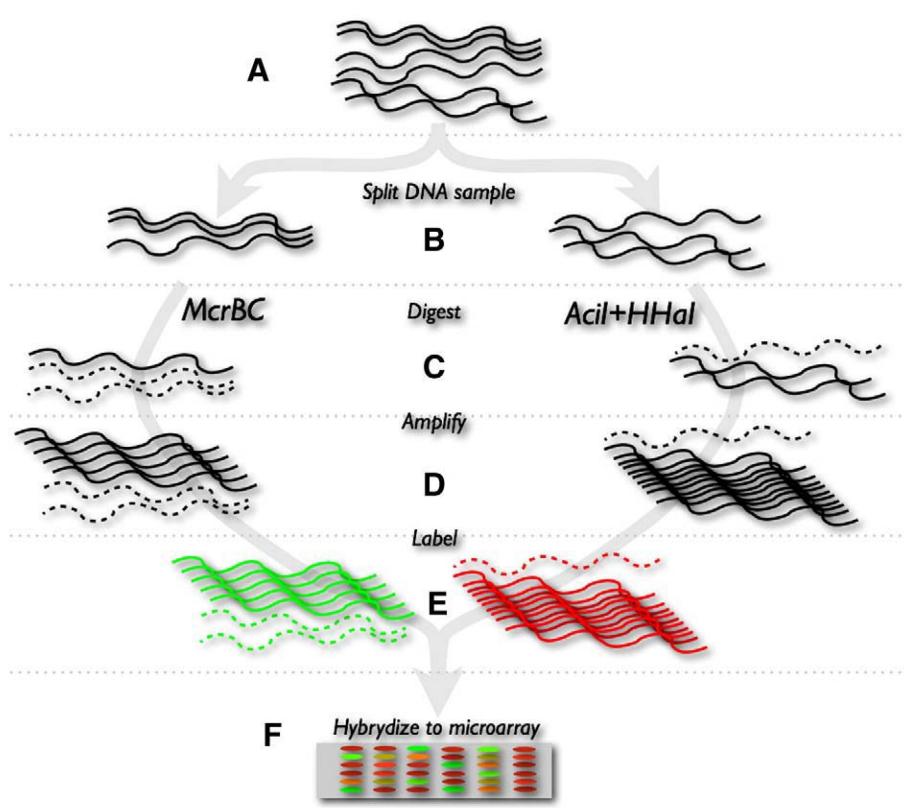

Fig. 1. A simplified diagram summarizing the steps discussed in the Materials and methods section. (A) DNA is first acquired from a tissue material. (B) the DNA is split into two equal aliquots. (C) Each of them is then digested with methylation-sensitive or -dependent enzymes. (D) the DNA is then amplified, (E) labeled, and (F) hybridized to a microarray.