Available online at www.sciencedirect.com
ScienceDirect
www.sciencedirect.com/science/journal/16720229
GENOMICS PROTEOMICS & BIOINFORMATICS

Article

The Most Redundant Sequences in Human CpG Island Library Are Derived from Mitochondrial Genome

Ximiao He[1,2#], Shu Tao[1,2#], Jing Jin[3], Songnian Hu[1*], and Jun Yu[1*]

[1]*CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*
[2]*Graduate University of Chinese Academy of Sciences, Beijing 100049, China;*
[3]*Department of Biology, University of Science and Technology of China, Hefei 230027, China.*

Genomics Proteomics Bioinformatics 2010 Jun; 8(2): 81-91. DOI: 10.1016/S1672-0229(10)60009-5

Abstract

An altered pattern of epigenetic modifications, such as DNA methylation and histone modification, is critical to many common human diseases, including cancer. Recently, mitochondrial DNA (mtDNA) was reported to be associated with tumorigenesis through epigenetic regulation of methylation patterns. One of the promising approaches to study DNA methylation and CpG islands (CGIs) is sequencing and analysis of clones derived from the physical library generated by methyl-CpG-binding domain proteins and restriction enzyme MseI. In this study, we observed that the most redundant sequences of 349 clones in a human CGI library were all generated from the human mitochondrial genome. Further analysis indicated that there was a 5,845-bp DNA transfer from mtDNA to chromosome 1, and all the clones should be the products of a 510-bp MseI fragment, which contained a putative CGI of 270 bp. The 510-bp fragment was annotated as part of cytochrome c oxidase subunit II (COXII), and phylogenetic analysis of homologous sequences containing COXII showed three DNA transfer events from mtDNA to nuclear genome, one of which underwent secondary transfer events between different chromosomes. These results may further our understanding of how the mtDNA regulates DNA methylation in the nucleus.

Key words: human, DNA methylation, CpG islands, nuclear mitochondrial DNA, molecular phylogeny

Introduction

DNA methylation is a critical biochemical modification of eukaryotic DNA involved in various biological processes including gene silencing, chromosomal structure stabilization, X-chromosome inactivation, imprinting, and cell differentiation (*1-9*). In mammals, DNA methylation mainly occurs at the fifth carbon position of the cytosine in a CpG context, and this biochemical process is owing to the activity of DNA methyltransferases (DNMTs) (*10, 11*). The dinucleotide CpG is remarkably under-represented in the human genome with only about 20% of the expected frequency. However, there are many genomic regions that contain a much high frequency (about 10 times higher than the average of genome) of CpG dinucleotides, and these regions are called CpG islands (CGIs) (*12*). The usual formal defination of a CpG island is a region at least 200 bp in length, with a GC percentage greater than 50% and an observed/expected CpG ratio greater than 0.6 (*13*).

# Equal contribution.
*Corresponding authors.
E-mail: husn@big.ac.cn; junyu@big.ac.cn

Recently, a stricter rule (>500 bp, GC content >55%, and CpG ratio >0.65) of CpG island prediction was suggested in order to exclude other GC-rich genomic sequences such as Alu repeats (*14*).

The CGIs are traditionally thought to be unmethylated, even some of them were found to be hypermethylated in the imprinted genes (*15*). Considering this property of CGIs, and that unmethylated CpG sites can be methylated *in vitro*, Cross *et al* introduced an approach to purify CGIs using a methylated DNA binding (MDB) column (*16*). Heisler *et al* analyzed the CGI clones derived from a physical library generated by MDB and MseI, and suggested that the clones are representatives of CGIs annotated on the human genome, and there may be value in continuing to isolate clones from the library (*17*). The CGIs typically occur at or near the transcription start site of genes, particularly housekeeping genes, in vertebrates. The methylation of CGIs is involved in the regulation of gene expression, which plays an important role in disease development including tumorigenesis (*18, 19*). To better understand the interplay of CGI methylation and cancer, He *et al* continued to isolate and sequence the clones derived from the same CGI library, and the sequences were deposited in the database of MethyCancer (*20*). Based on the comprehensive analysis of these CGI clones, we observed that the most redundant sequences of 349 clones were generated from the human mitochondrial genome (mtDNA).

The nuclear mitochondrial DNA, first denoted as "NUMT" in cat by Lopez *et al* (*21*), refers to DNA segment that has been transferred from mitochondrial genome to nuclear genome. This phenomenon has been observed in diverse eukaryotes such as human, mouse, rat, rice, *Arabidopsis* and insects, with a large NUMT number and size variation across species (*22*). Recently, mtDNA was found to be associated with tumorigenesis through epigenetic regulation of methylation patterns. Xie *et al* (*23*) studied the effect of mtDNA depletion on cancer progression and found that mtDNA depletion promotes cancer progression through activating hypermethylation pattern of cancer-associated genes' promoter CpG islands, and this activation was achieved through the induction of DNMT1. Similarly, Smiraglia *et al* (*24*) investigated whether mtDNA copy number variation, a feature of many human tumors, can affect methylation changes in the nucleus, and found that methylation pattern is reversible for a number of genes following depletion and restoration of mtDNA.

In this study, we reported a DNA transfer of about 6k mtDNA (namely $NUMT^{ND\text{-}COX}$) to chromosome 1, which covered all of the most redundant 349 clones. Further analysis of $NUMT^{ND\text{-}COX}$ indicates that all the clones should be the products of an MseI fragment with the length of 510 bp ($NUMT^{ND\text{-}COX}$: 3,841-4,110), which contained a putative CGI of 270 bp. Phylogenetic analysis of homologous sequences containing cytochrome c oxidase subunit II (COXII) allows us to detect three DNA transfer events, and one of these events underwent possible secondary interchromosomal transfer events. This observation may provide us a clue for further understanding of the roles of mtDNA in regulation of DNA methylation in the nucleus.

Results

Genomic mapping of CGI library clones

Heisler *et al* (*17*) analyzed and compared the CGI clones of 12k set (12,192 clones) deposited at the Wellcome Trust Sanger Institute, and the 9k set (8,554 clones) isolated from the same CGI library (*25*) by the Huang laboratory, and found that there was only a small degree of overlap between the two sets, with only 753 common genomic loci of the total 9,595. While in 17,606 sequences obtained from MethyCancer, 17,068 sequences were aligned to the human genome using BLAT/BLAST (*26*). Clones sharing the overlapped genome locations were clustered into 10,648 distinct genomic loci, with the redundancy of 37.61%.

We combined the three datasets of BIG18K (the 17,606 clones sequenced by Beijing Institute of Genomics, CAS), Sanger12K (the 12,192 clones deposited at Sanger) and UHN8K (the 8,373 clones downloaded from UHN). Overall, the 35,602 clones mapped to the genome were clustered into 18,240 genomic loci. The number of distinct genomic loci of BIG18K was 7,932 (74.49%), and the number of common loci of the three sets was only 913 (**Figure 1A**), which implies that the majority of loci of clones