論文の内容の要旨

STUDY ON THE SEQUENCE-DEPENDENCY OF DNA METHYLATION IN MEDAKA EMBRYOS

（メダカ胚における DNA methylation の配列依存性の研究）

チョウ　コク　メイ
張　　國　　鳴
氏名：CHEUNG KWOK MING

Introduction

The heavily methylated vertebrate genomes are punctuated by stretches of poorly methylated DNA sequences (i.e. hypomethylated domains, referred to as 'HypoMDs') that usually mark gene regulatory regions. It has been proven that hypermethylated regulatory regions confer transcriptional silencing to downstream genes, whereas hypomethylation is associated with active or poised transcription. Given its governance on the transcriptome, genome-wide DNA methylation pattern is directly linked to cellular functions and identity, and is tightly regulated. However, how the methylation pattern is determined *in vivo* remains enigmatic.

Recent *in vitro* studies proposed that the methylation state is autonomously determined by local DNA sequence. It is presumed that the genome is methylated by default, while specific loci are recognized and hypomethylated, leading to the formation of HypoMDs. Bioinformatics analyses revealed that high local CpG density in HypoMDs may serve as an

evolutionary conserved signature that repels *de novo* methylation. *In vitro* studies further evidenced that transcription factor binding site-like DNA motifs isolated from HypoMDs could confer the hypomethylated state to their nearby sequences. These facts advocate that DNA sequence is encoded with information which dictates its own methylation state. However, this notion has never been carefully validated *in vivo*.

This is the first *in vivo* study, with the use of medaka (*Oryzias latipes*) embryos as model, in which the sequence dependency of DNA methylation was rigorously tested.

Results

Methylation state correlates with sequence pattern

Statistical association between medaka genomic sequences and local methylation states was modelled by machine learning (support vector machine; kmer-SVM). HypoMDs, as well as hypermethylated regions (HyperMDs) with similar GC-content, were subjected to model training with 10-fold cross-validation. The resultant models were highly precise and sensitive (area under precision-recall curve ≥ 0.83 vs. 0.08 for random classifier), indicating that HypoMDs and their hypermethylated counterparts carry distinct sequence patterns. In addition, consecutive CpGs (especially 5'-CGCGCG-3') and CGnull-repeats (i.e. AGCTAG, GCTAGC, TAGCTA) are found highly enriched in HypoMDs, suggesting a relationship between the presence of these motifs and the local hypomethylation state. These results are in concordance with the postulated link between methylation states and nucleotide sequences.

Highly enriched DNA motifs are not the primary determinants of local hypomethylation

To test if the above enriched DNA motifs are related to the hypomethylated state of HypoMDs. The $(CG)_3$ motif and CGnull-repeats were disrupted *in vivo* in 6 HypoMDs using CRISPR-Cas9. However, successful motif disruption did not result in any change in the hypomethylated state of the targeted HypoMDs in both F0 and F1 medaka embryos.

Methylation states are not autonomously determined by nucleotide sequences *in vivo*

To verify the dependency between methylome and genomic sequence, genomic regions with high CpG density were captured, PCR-amplified and reintegrated into genome at random positions. Despite the strong statistical association between nucleotide sequences and methylation states, the integrated genomic fragments could not recapitulate their

endogenous methylation status. Correlation between the methylation rate at endogenous loci and that at ectopic positions: Spearman's $\rho = 0.08$, Kendall's $\tau = 0.07$. The lack of sequence dependency was further illustrated by a drastically different ectopic methylation pattern when the genomic fragments were artificially methylated prior to injection and genome integration.

To exclude possible position effect due to random integration, six HyperMDs were subcloned and integrated into a gene desert region of the chromosome 18 using the PhiC31 medaka transgenic strain via PhiC31 integrase-mediated site-specific recombination. In parallel, eleven HypoMDs were similarly processed, but with artificial methylation prior to integration. Regardless of the endogenous state of the cloned and integrated sequence, the artificially conferred methylation state (i.e. hypomethylation in HyperMDs and hypermethylated in HypoMDs) of the all these sequences was faithfully maintained from 1-cell stage (i.e. injection and integration) to blastula stage. This observation indicates that there is no direct dependency between DNA sequence and methylation state.

HypoMDs could not define their methylation state at their endogenous sites

The observed absence of sequence-dependency in the above experiment could be due to the fact that the assayed sequences were interrogated at non-native, ectopic genome locations. To verify this, the methylation state of two HypoMDs were edited *in situ* via CRISPR-Cas9-triggered homology-directed repair with artificially methylated repair templates. However, consistent with the above observations, the loci were rendered hypermethylated after *in situ* editing, in spite of their original hypomethylated state. Moreover, these loci remained hypermethylated in the edited embryos at 3 and 7 (i.e. hatching stage) day-post-fertilization. This finding highlights that genomic sequence and its methylation state were not coupled even at the endogenous positions.

Summary

The above results clearly demonstrated that there is no direct connection between DNA methylation and local sequence *in vivo*, defying the notion of the governance on DNA methylation by nucleotide sequence, but instead suggest the involvement of other epigenetic factors in defining and maintaining the DNA methylation landscape of vertebrate genomes.