Methods 187 (2021) 13–27

Contents lists available at ScienceDirect

Methods

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

Targeted bisulfite sequencing for biomarker discovery

Marco Morselli [a,b,c,*], Colin Farrell [a], Liudmilla Rubbi [a], Heather L. Fehling [d], Rebecca Henkhaus [d], Matteo Pellegrini [a,b,c,*]

[a] Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA 90095, United States
[b] UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095, United States
[c] Institute for Quantitative and Computational Biosciences – The Collaboratory, University of California Los Angeles, Los Angeles, CA 90095, United States
[d] Clinical Reference Laboratory, Inc., Lenexa, KS 66215, United States

ARTICLE INFO

Keywords:
Next-generation sequencing
Target bisulfite-seq
DNA methylation
Biomarker discovery
Epigenetic clock

ABSTRACT

Cytosine methylation is one of the best studied epigenetic modifications. In mammals, DNA methylation patterns vary among cells and is mainly found in the CpG context. DNA methylation is involved in important processes during development and differentiation and its dysregulation can lead to or is associated with diseases, such as cancer, loss-of-imprinting syndromes and neurological disorders. It has been also shown that DNA methylation at the cellular, tissue and organism level varies with age. To overcome the costs of Whole-Genome Bisulfite Sequencing, the gold standard method to detect 5-methylcytosines at a single base resolution, DNA methylation arrays have been developed and extensively used. This method allows one to assess the status of a fraction of the CpG sites present in the genome of an organism. In order to combine the relatively low cost of Methylation Arrays and digital signals of bisulfite sequencing, we developed a Targeted Bisulfite Sequencing method that can be applied to biomarker discovery for virtually any phenotype. Here we describe a comprehensive step-by-step protocol to build a DNA methylation-based epigenetic clock.

1. Introduction

The methylation of the 5th carbon of cytosines is a covalent modification found in many organisms, such as prokaryotes, fungi, algae, plants and animals [1–4]. In mammals the reaction is catalyzed by the activity of three enzymes called DNA Methyltransferases (DNMT3a, DNMT3b and DNMT1) and it is found predominantly in the CpG dinucleotide context [5,6]. For each cell type, DNA methylation patterns are established during development and differentiation and faithfully maintained through cell division [7]. DNA methylation is involved in numerous processes such as genomic imprinting, X-chromosome inactivation, genome stability and transcriptional regulation [8]. Aberrant DNA methylation patterns have been observed in loss-of-imprinting syndromes, many cancer types, autoimmune diseases, and metabolic, neurological and psychological disorders [9]. DNA methylation changes have not been observed only during differentiation or in diseases, but to estimate the age of an individual, known as epigenetic clocks [10,15–17], and other traits such as BMI [18], smoking [18,19] and type-2 diabetes [20]. The difference between the epigenetic and the chronological age can inform on the biological or physiological age of an individual [21]. Other biomarkers have been utilized to predict biological age, but DNA methylation is generally more accurate than other approaches [22].

Several methods have been described to detect DNA methylation and they can be classified into four major groups:

1) methylation-specific restriction endonucleases;
2) Immunoprecipitation using anti-$5^{me}C$ antibodies or affinity purification by methyl-DNA binding proteins;
3) Sodium bisulfite treatment;
4) Sequencing of the native DNA molecule using third-generation sequencing technologies (Pacific Biosciences and Oxford Nanopore).