

Chapter 28

Prediction of Epigenetic Target Sites by Using Genomic DNA Sequence

Guo-Cheng Yuan

Harvard School of Public Health, USA & Dana-Farber Cancer Institute, USA

ABSTRACT

Epigenetic regulation provides an extra layer of gene control in addition to the genomic sequence and is critical for the maintenance of cell-type specific gene expression programs. Significant changes of epigenetic patterns have been linked to developmental stages, environmental exposure, ageing, and diet. However, the regulatory mechanisms for epigenetic recruitment, maintenance, and switch are still poorly understood. Computational biology provides tools to deeply uncover hidden connections and these tools have played a major role in shaping the current understanding of gene regulation, but its application in epigenetics is still in the infancy. This chapter reviews some recent developments of computational approaches to predict epigenetic target sites.

INTRODUCTION

Epigenetics refers to heritable changes of gene expression or genotypes without change of the DNA sequence (Waddington, 1942). In a multicellular organism, the DNA sequence is constant in all cell lineages, but the gene activities in different cell-types are highly variable. Such cell-type specific gene expression patterns are controlled by epigenetic mechanisms, including nucleosome

positioning, histone modifications, and DNA methylation. Together these mechanisms control the accessibility of the genomic DNA to regulatory proteins. Only a small part of the genomic blueprint is used in any cell type.

The rapid advance of microarray and DNA sequencing technologies (Barski et al., 2007; Mikkelsen et al., 2007; Ren et al., 2000) has allowed researchers to identify genome-wide epigenetic patterns in various species. Recent epigenomic studies have identified dramatic epigenetic differences between different cell-types (Barski et

DOI: 10.4018/978-1-4666-3604-0.ch028

al., 2007; Heintzman et al., 2009; Meissner et al., 2008; Mikkelsen et al., 2007; Mohn et al., 2008), between normal and disease tissues (Schlesinger et al., 2007; Seligson et al., 2005; TCGA, 2008), and between stimulated and resting cells (Saccani & Natoli, 2002; Wei et al., 2009). These differences are also highly correlated with gene expression level changes. Importantly, the epigenetic changes are not permanent but can be reversed. Strikingly, the entire epigenetic state in an adult cell can be reprogrammed to a pluripotent cell state (called iPS cell) that is highly similar to an embryonic stem (ES) cell (Okita et al., 2007; Takahashi & Yamanaka, 2006; Wernig et al., 2007; Yu et al., 2007). This reversibility makes epigenetic marks the ideal targets for therapeutic treatment (Sharma et al., 2010). Indeed, a number of drugs have been developed and currently used to treat a number of diseases including cancer (Yoo & Jones, 2006). However, a major challenge is to avoid off-target interactions, since our current understanding of the targeting mechanisms of epigenetic factors is still limited.

The epigenetic pattern is not randomly distributed across the genome (Bernstein et al., 2007). A fundamental question is how target specificity is achieved. The targeting mechanism is complex and involves many factors such as the genomic DNA sequence, chromatin modifiers, transcription factors (TFs), and non-coding RNAs. How these factors work together to regulate epigenetic patterns is still poorly understood. Among these factors, the association with DNA sequence has been most studied. Perhaps the most commonly studied factor is the DNA sequence. Here I review some of computational studies aimed at prediction of epigenetic patterns based on the DNA sequence. Additional information on certain specific aspects can be found in some excellent reviews (Bock & Lengauer, 2008; Kaplan et al. 2010). In addition, the readers are referred to some excellent reviews for biological background (Jiang & Pugh, 2009; Kouzarides, 2007; Rando & Chang, 2009; Hawkins et al. 2010; Zhou et al. 2011).

METHODS TO PREDICT EPIGENETIC TARGETS

Nucleosome Positioning

The eukaryotic DNA is packaged into chromatin. The fundamental unit of chromatin is the nucleosome, consisting of two copies each of four core histone proteins: H2A, H2B, H3, and H4 (Kornberg & Lorch, 1999). Each nucleosome wraps around 146 bp DNA in about 2 turns. Ever since the initial discovery of nucleosomes in the 70's (Kornberg, 1974), the regulatory mechanisms underlying nucleosome positioning have been intensely investigated. A potentially important role of DNA sequences was noticed decades ago. The investigators recognized that the structural properties of DNA are dependent upon the base pair composition; therefore specific DNA sequences might be favored for nucleosome binding. By extracting nucleosome core particles from chicken red blood cells and then analyzing the DNA sequences attached to these nucleosome particles, Satchwell et al. (1986) observed an approximately 10 bp periodic pattern of the frequency of the dinucleotide pair AA/TT. Similar results were found by several other groups (Ioshikhes et al., 1996; Widom, 2001). The 10 bp periodicity pattern agrees well with the high-resolution nucleosome structure (Luger et al., 1997; Richmond & Davey, 2003), where the histones interact with the DNA sequence approximately once every 10 base pair.

However, early studies were limited by the fact that only a handful of sequences were known to be nucleosome bound. During the past decade, rapid progress has been made thanks to the development of high throughput technologies such as microarrays and DNA sequencing. As a result, high-resolution, genome-scale maps of nucleosome positions have been identified in various organisms including human (Chodavarapu et al. 2010; Johnson et al., 2006; Lanterman et al. 2010; Lee et al., 2007; Mavrich, Ioshikhes et al., 2008; Mavrich, Jiang et al., 2008; Oszolak et al., 2007;

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/prediction-epigenetic-target-sites-using/76081

Related Content

Social Issues Related To Gene Patenting in Latin America: A Bioethical Reflection

Eduardo Rodriguez and Fernando Lolas (2011). *Genomics and Bioethics: Interdisciplinary Perspectives, Technologies and Advancements* (pp. 152-170).

www.irma-international.org/chapter/social-issues-related-gene-patenting/47299

Optimal Fuzzy Cluster Partitioning by Crow Search Meta-Heuristic for Biomedical Data Analysis

Janmenjoy Nayak, Bighnaraj Naik, Pandit Byomakesha Dash and Danilo Pelusi (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 1252-1269).

www.irma-international.org/chapter/optimal-fuzzy-cluster-partitioning-crow/342573

Classification of Tandem Repeats in the Human Genome

Yupu Liang, Dina Sokol, Sarah Zelikovitz and Sarah Ita Levitan (2012). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-21).

www.irma-international.org/article/classification-tandem-repeats-human-genome/77808

Role of Data Mining Techniques in Bioinformatics

Pushpa Singh and Narendra Singh (2021). *International Journal of Applied Research in Bioinformatics* (pp. 51-60).

www.irma-international.org/article/role-of-data-mining-techniques-in-bioinformatics/267825

Cross-Platform Microarray Data Integration Combining Meta-Analysis and Gene Set Enrichment Analysis

Jian Yu, Jun Wu, Miaoxin Li, Yajun Yi, Yu Shy and Lu Xie (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences* (pp. 12-27).

www.irma-international.org/chapter/cross-platform-microarray-data-integration/48361