

Chapter 6

Application of a CAGE Method to an Avian Development Study

Ruslan Deviatiiarov, Marina Lizio, and Oleg Gusev

Abstract

Cap analysis of gene expression (CAGE) is a convenient approach for genome-wide identification of promoter regions at single base-pair resolution level and accurate expression estimation of the corresponding transcripts. Depending on the initial biomaterial amount and sequencing technology, different computational pipelines for data processing are available, as well as variations of the CAGE protocol that improve sensitivity and accuracy. Therefore, this chapter elucidates the key steps of sample preparation, sequencing, and data analysis via an example of a promoter expression estimation study in chicken development. We also describe the applicability of this approach for studying other avian and reptilian species.

Key words Cap analysis gene expression (CAGE), Promoters, Chicken development, FANTOM

1 Introduction

During recent years, next-generation sequencing (NGS) technologies drastically accelerated the accumulation of genomic data. The Genomes OnLine Database (GOLD) lists more than 13.5 K sequenced eukaryotic genomes (<https://gold.jgi.doe.gov>), and the Genome 10K Project has already sequenced about 280 vertebrate genomes with a further goal of 10,000 [1, 2]. At the same time, different approaches for RNA signal detection have been developed, from methods like microarrays, applied on a relatively small set of genes, to whole-genome transcriptomics like RNA-Seq, CAGE, which is based on sequencing of short 5' ends of transcripts, DGE (digital gene expression) for accurate tag-based expression estimation, and RACE (rapid amplification of cDNA ends), designed for identification of full-length RNA transcripts, allowing researchers to estimate gene expression on a genome-wide scale and to understand the principles of gene regulation [3–6].

Since utilization of high-throughput sequencing technologies became more affordable and accessible, many research groups