Methods 48 (2009) 398–408

Contents lists available at ScienceDirect

Methods

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

Review Article

Using ChIP-chip and ChIP-seq to study the regulation of gene expression: Genome-wide localization studies reveal widespread regulation of transcription elongation

Daniel A. Gilchrist [a], David C. Fargo [b], Karen Adelman [a,*]

[a] Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA
[b] Library and Information Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

ARTICLE INFO

Article history:
Accepted 27 February 2009
Available online 9 March 2009

Keywords:
Transcription elongation
Gene expression
ChIP-chip
ChIP-seq

ABSTRACT

Transcription is a sophisticated multi-step process in which RNA polymerase II (Pol II) transcribes a DNA template into RNA in concert with a broad array of transcription initiation, elongation, capping, termination, and histone modifying factors. Recent global analyses of Pol II distribution have indicated that many genes are regulated during the elongation phase, shedding light on a previously underappreciated mechanism for controlling gene expression. Understanding how various factors regulate transcription elongation in living cells has been greatly aided by chromatin immunoprecipitation (ChIP) studies, which can provide spatial and temporal resolution of protein–DNA binding events. The coupling of ChIP with DNA microarray and high-throughput sequencing technologies (ChIP-chip and ChIP-seq) has significantly increased the scope of ChIP studies and genome-wide maps of Pol II or elongation factor binding sites can now be readily produced. However, while ChIP-chip/ChIP-seq data allow for high-resolution localization of protein–DNA binding sites, they are not sufficient to dissect protein function. Here we describe techniques for coupling ChIP-chip/ChIP-seq with genetic, chemical, and experimental manipulation to obtain mechanistic insight from genome-wide protein–DNA binding studies. We have employed these techniques to discern immature promoter-proximal Pol II from productively elongating Pol II, and infer a critical role for the transition between initiation and full elongation competence in regulating development and gene induction in response to environmental signals.

1. Introduction

1.1. Background

Synthesis of messenger RNA by RNA polymerase II (Pol II) is a carefully orchestrated process. Although the regulated recruitment of the transcription machinery to a gene promoter has been studied for decades [1], recent evidence suggests that regulation can occur at many steps in the transcription cycle, and may be particularly prevalent during transcription elongation [2–5]. Transcription begins with promoter recognition and binding by the pre-initiation complex (PIC) consisting of Pol II and general transcription factors including TFIID and TFIIH. The C-terminal domain (CTD) of the Rpb1 subunit of Pol II, consisting of multiple copies of the consensus sequence YSPTSPS, is largely unphosphorylated during initial promoter binding, which favors interactions between the CTD and activators such as the Mediator complex [6,7]. Unwinding of DNA by TFIIH allows Pol II to access the template DNA strand and begin incorporating nucleotides into a nascent RNA chain. As the nascent RNA is extended, TFIIH phosphorylates Serine-5 of the CTD, which is thought to positively influence the association of the mRNA capping machinery [8,9]. Factors such as the Negative Elongation Factor, or NELF, complex in collaboration with the heterodimeric DSIF complex (comprised of Spt4/Spt5) can impede elongation through the promoter-proximal region [10–14].

Recruitment of the P-TEFb kinase signals the transition to productive elongation, by phosphorylating the CTD at Serine-2 and helping to overcome NELF-dependent stalling of early elongation [15–17]. The Serine-2 phosphorylated, fully elongation-competent form of Pol II is then bound by RNA processing and termination factors as it transcribes in a highly processive manner toward the polyadenylation site, which signals for termination. In addition to this coordinated regulation of the phosphorylation status of the Pol II CTD and the association of transcription elongation and processing factors, histone remodeling and modifying factors are specifically recruited to facilitate efficient polymerase elongation [18].

Each aspect of transcription is governed by interactions between the largely proteinaceous transcription machinery and the DNA template. Chromatin immunoprecipitation (ChIP) is capable

* Corresponding author. Fax: +1 919 541 0146.
E-mail address: adelmank@niehs.nih.gov (K. Adelman).

1046-2023/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ymeth.2009.02.024

of providing high-resolution spatial and temporal information about the interactions between proteins and DNA in living cells. It is therefore well suited for dissecting phases of the transcription cycle by placing individual protein complexes at specific genomic locations at biologically significant times. For example, biochemical studies had identified the NELF complex as capable of inhibiting Pol II elongation, in a manner that can be reversed by the kinase activity of P-TEFb [10–13], leading to the hypothesis that NELF played a role in regulating the efficiency of transcript elongation by Pol II prior to the transition to full elongation competence. Subsequent ChIP studies have shown that the NELF complex is associated broadly with unphosphorylated or Serine-5 phosphorylated Pol II near transcription start sites. This association is not maintained in downstream regions where the polymerase is Serine-2-phosphorylated and productive elongation occurs, thereby confirming that the biochemically determined activity of the NELF complex is relevant and placing it in a global in vivo context [19,20].

The power of ChIP has been tremendously increased by its coupling with DNA microarray technology. In traditional ChIP assays, protein complexes are localized to genomic loci by querying immunoprecipitated DNA with quantitative or semi-quantitative PCRs using primer pairs designed to amplify specific regions of interest. In ChIP-chip, immunoprecipitated material is labeled with fluorescent dyes (with or without prior amplification) and hybridized to DNA microarrays containing several hundred thousand, to several million probes (Fig. 1). Performing ChIP coupled with DNA microarrays has several significant advantages over traditional ChIP. First, instead of querying a limited number of loci selected by researchers with inherent biases, large contiguous genomic regions are probed in a single experiment, eliminating bias and permitting discovery of unanticipated sites of protein-DNA binding, as well as regions where binding is unexpectedly absent. Second, localization of protein binding can be accomplished with optimized commercially available platforms, eliminating time spent designing and testing primer pairs and running expensive large-scale quantitative PCR assays. In addition, the use of the same platforms by different research groups facilitates direct comparison of binding data obtained for many individual proteins; groups such as the ENCODE consortium have used this to a great advantage (e.g. [21]). Third, the parallel analysis of thousands of genes allows one to parse the data into distinct classes of genes based on different binding distributions or behaviors, and permits statistical comparisons between classes.

Genomic distribution of Pol II and other transcription elongation factors has also been determined through a process referred to as ChIP-seq [22], which offers an appealing complementary or alternative method for mapping protein–DNA interactions. The strategy is similar to ChIP-chip but instead of labeling immunoprecipitated material and hybridizing it to a microarray, immunoprecipitated material is used to construct a library of millions of individual DNA fragments which are amplified and then sequenced in parallel (Fig. 1). Massively parallel sequencing technology, also referred to as deep, or high-throughput sequencing, is now widely available on a variety of platforms, each with distinct characteristics (see Section 2.6.2 below).

1.2. Genome-wide analyses illuminate novel aspects of transcription elongation

Global analyses of Pol II distribution have provided insight into mechanisms of regulation of transcription elongation that are unattainable with either traditional ChIP or biochemical techniques. In particular, ChIP-chip studies have detected a widespread

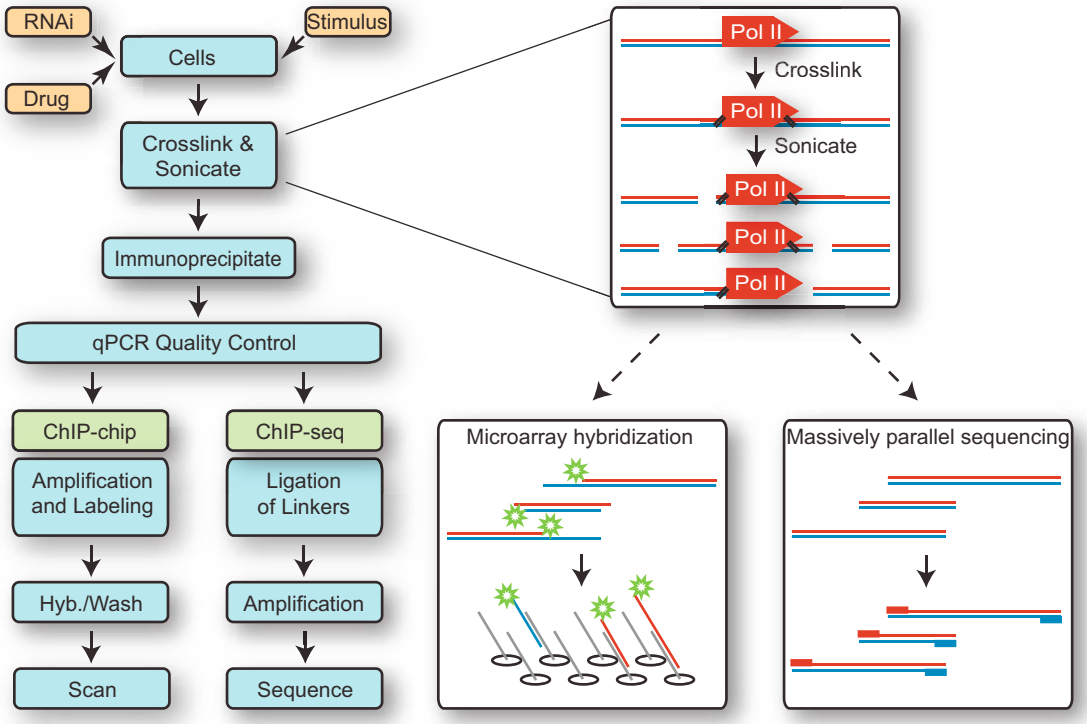

Fig. 1. Workflow for ChIP-chip and ChIP-seq experiments. Following experimental manipulation (yellow boxes), cells are crosslinked with formaldehyde, sonicated to fragment chromatin, and protein–DNA complexes immunoprecipitated with antibodies targeting the protein or modification of interest (here, Pol II). Following quality control qPCR to confirm expected ChIP signal at control regions, immunoprecipitated DNA is processed specifically for either ChIP-chip or ChIP-seq. ChIP-chip can provide information about all immunoprecipitated DNA sequences complementary to tiling array probes in a strand-insensitive manner. ChIP-seq provides information about all mappable sequences located at the 5′-ends of immunoprecipitated DNA (red and blue boxes).