BioSystems 102 (2010) 61–69

Contents lists available at ScienceDirect

BioSystems

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

Circadian transcription in liver

K. Bozek[a], A.L. Rosahl[b], S. Gaub[b], S. Lorenzen[c,*], H. Herzel[b]

[a] Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany
[b] Institute for Theoretical Biology, Invalidenstr. 43, 10557 Berlin, Germany
[c] AG Systems Biology, Charité Centrum für Grundlagenmedizin, Tucholskystr. 2, 10117 Berlin, Germany

ARTICLE INFO

Article history:
Received 4 May 2010
Accepted 15 July 2010

Keywords:
Circadian clock
Transcription factor binding site
DNA array
Clock controlled genes

ABSTRACT

Circadian rhythms regulate a wide range of cellular, physiological, metabolic and behavioral activities in mammals. The complexity of tissue- and day-time specific regulation of thousands of clock controlled genes (CCGs) suggests that many transcriptional regulators are involved. Our bioinformatic analysis is based on two published DNA-array studies from mouse liver. We search overrepresented transcription factor binding sites in promoter regions of CCGs using GC-matched controls. Analyzing a large set of CCG promoters, we find known motifs such as E-boxes, D-boxes and cAMP responsive elements. In addition, we find overrepresented GC-rich motifs (Sp1, ETF, Nrf1), AT-rich motifs (TBP, Fox04, MEF-2), Y-box motifs (NF-Y, C/EBP) and cell cycle regulators (E2F, Elk-1).

In a subset of system-driven genes, we find overrepresented motifs of the serum response factor SRF and the estrogen receptor ER. The analysis of published ChIP data reveals that some of our predicted regulators (C/EBP, E2F, HNF-1, Myc, MEF-2) target relatively many clock controlled genes. Our analysis of CCG promoters contributes to an understanding of the complex transcriptional regulation of circadian rhythms in liver.

1. Introduction

Circadian clocks are self-sustained oscillators that regulate the temporal organization of physiology, metabolism and behavior (Levi and Schibler, 2007). These rhythms allow organisms to adapt to the 24-h period of the solar day. Two small clusters of hypothalamic neurons called the suprachiasmatic nuclei (SCN) constitute the master pacemaker. Light signals detected by the eyes can entrain rhythms in the SCN through the retinohypothalamic tract. The SCN sends synchronization signals to other cells of the body by hormone secretion, sympathetic enervation and indirect cues such as body temperature, feeding time and activity rhythms. The cell-autonomous circadian oscillations in virtually all cells of central and peripheral organs are generated by interlocked transcriptional–translational feedback loops (Takahashi et al., 2008). The transcription factor heterodimer CLOCK:BMAL1 activates via E-boxes the expression of *Period* genes (*Per1*, *Per2* and *Per3*), nuclear receptors (*Rev-Erbα*, *Rorα*) and *Chryptochrome* genes (*Cry1* and *Cry2*). PER and CRY proteins form complexes and repress their own expression by interacting with the CLOCK:BMAL1 dimer. REV-ERBα and RORα regulate the transcription of *Bmal1* in separate feedback loops through ROR regulatory elements. Light input to the SCN and intercellular coupling between SCN neurons is mediated by CREB binding motifs. Several clock output genes are regulated through D-boxes (Mitsui et al., 2001) by the transcription factors DBP, HLF, TEF and E4BP4. Thus E-boxes, ROR elements (RREs), cAMP response elements (CREs) and D-boxes are core elements of the circadian gene regulatory network (Ueda et al., 2005).

Transcriptome profiling studies have revealed that in addition to the core clock genes hundreds of clock controlled genes (CCGs) show oscillatory transcription (Ueda et al., 2002; Akhtar et al., 2002; Panda et al., 2002). In particular, it has been uncovered that the sets of CCGs differ dramatically from tissue to tissue (Storch et al., 2002). The molecular details of the regulation of clock controlled genes in different tissues are still widely unknown. Genes might be regulated by the intrinsic circadian clock in individual cells or by systemic cues such as feeding–fasting cycles, hormones and body temperature (Kornmann et al., 2007a). An in-depth analysis of cis-regulatory elements and transcription factors can provide valuable information on the regulation of clock controlled genes in specific tissues.

Recently, first promoter analyses of clock controlled genes appeared (Bozek et al., 2007, 2009; Yan et al., 2008). In these papers, thousands of CCGs extracted from published microarray data were analyzed. It was found that, in addition to the known regulatory elements, many other transcription factor binding sites (TFBSs) are overrepresented in CCG promoters. It has been found in Bozek et al. (2009) that promoters of clock controlled genes are GC-rich. Consequently, the predicted transcription factors in Bozek et al. (2007) and Yan et al. (2008) exhibit a bias towards GC-rich binding sites.

* Corresponding author. Tel.: +49 30 450 528 222; fax: +49 30 450 528 972.
E-mail address: stephlo1976@gmail.com (S. Lorenzen).

0303-2647/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2010.07.010

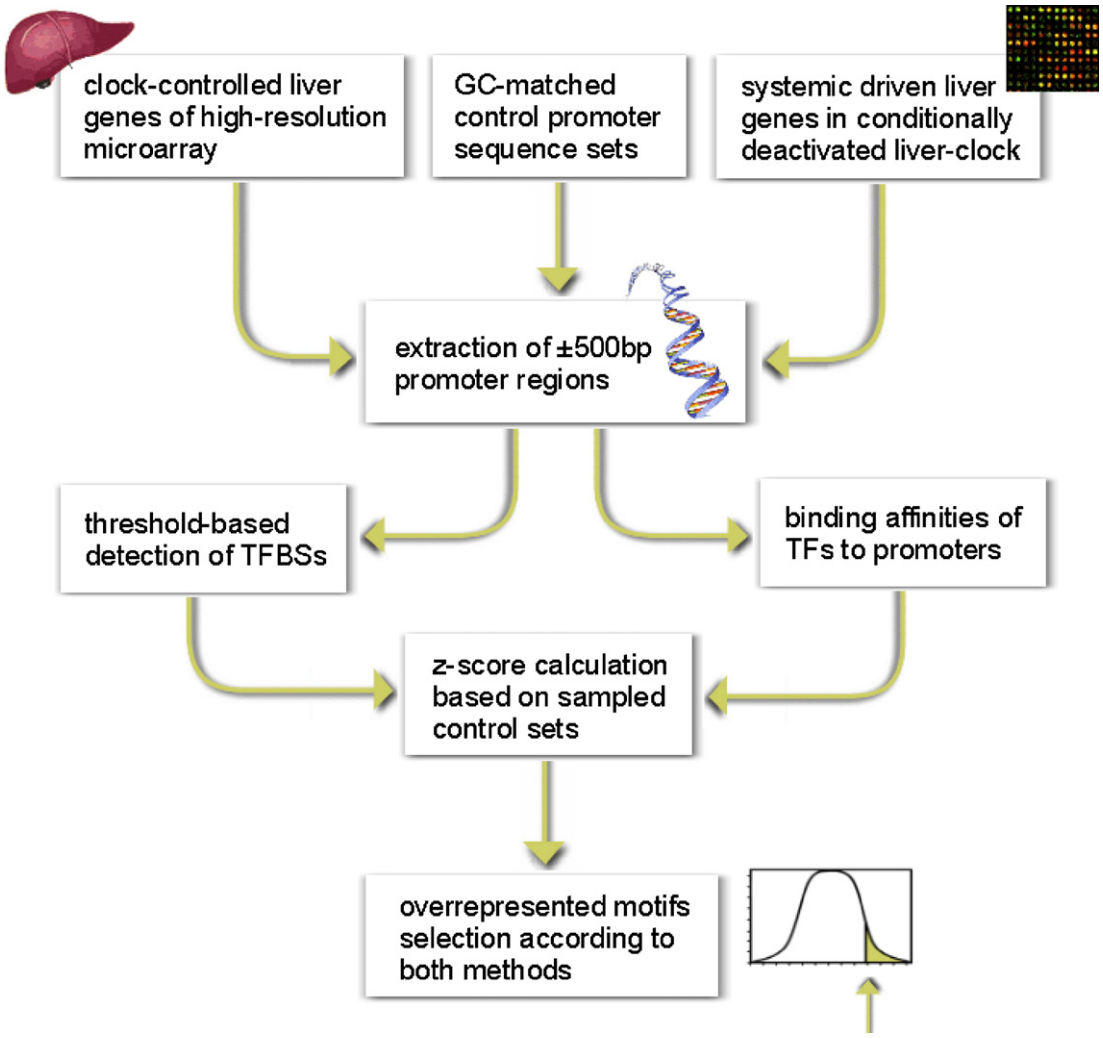

Fig. 1. Workflow of our analysis.

Furthermore, bioinformatic predictions of TFBSs suffer from a large amount of false positive predictions (Wasserman and Sandelin, 2004). Thus, independent confirmations of previous bioinformatic studies are required.

This paper exploits a recent high-resolution analysis of the liver transcriptome (Hughes et al., 2009). In contrast to previous studies with sampling in 4-h intervals, expression profiles are obtained every hour for 48 h. A rigorous statistical analysis including multiple testing corrections was applied. Over 3000 transcripts were found to oscillate with a false discovery rate (FDR) of 0.05. Here we search TFBSs around the transcriptional start site (TSS) of the corresponding genes using two different methods: (i) a threshold-based method with a FDR of 0.05 (Rahmann et al., 2003) and (ii) a novel method based on the estimation of the overall affinities of transcription factors to the promoter region (Roider et al., 2007). This approach takes also weaker TFBSs into account, comparable to a recent study of fly genes (Segal et al., 2008). Furthermore, we tested some of our predictions using recent data from chromatin immunoprecipitation (ChIP) studies. Our approach confirms that E-boxes and D-boxes play a major role in the regulation of CCGs. In addition, promising predictions of novel regulators such as SP-1, C/EBP, NF-Y, E2F, HNF-1, and MEF-2 are obtained.

2. Materials and methods

As illustrated in Fig. 1, our bioinformatic promoter analysis is based on published microarray data. Hughes et al. (2009) collected liver samples from 6-week-old male C57B2/6J mice housed in complete darkness. Over a time-span of 48 h, 3–5 mice were sacrificed every hour to excise liver samples. With Affymetrix Mouse Genome 430 2.0 arrays, the RNA expression levels were quantified. On average, 18,581 transcripts were detected per array (Hughes et al., 2009). Cycling genes were identified using COSOPT and Fisher's *G*-test (Hughes et al., 2007) with false discovery rates less than 0.05 in both tests. With this technique, over 3000 transcripts with circadian oscillations were identified (see supporting information in Hughes et al., 2009). Many of these clock controlled genes are driven by the cell-autonomous clock, e.g. via E-boxes and D-boxes. Another class of rhythmically expressed genes may be under the control of systemic signals, such as hormones and body temperature. In order to detect these system-driven genes, Kornmann et al. (2007a,b) established a mouse model in which local hepatocyte oscillators can be switched off. In short, inducible overexpression of REV-ERBα represses *Bmal1* transcription (see Kornmann et al., 2007a, for details). It turned out that only a relatively small number of 61 genes remained rhythmic with similar expression levels (Table 1 in Kornmann et al., 2007b). Many of these genes encode heat shock proteins or belong to the cholesterol metabolism.

In previous promoter analyses (Yan et al., 2008; Bozek et al., 2009) a meta-analysis of multiple array data from different organisms and tissues was performed. Moreover, due to uncertainties of the precise location of the transcription start sites (TSS), several thousand base pairs were scanned for transcription factor binding sites. In this paper, we focus on two recently published mouse-liver data sets with a so far unprecedented temporal resolution (Hughes et al., 2009) and with an interesting set of system-driven genes (Kornmann et al., 2007b). Moreover, we try to reduce the number of false positive predictions by analyzing only ±500 bp around the transcription start sites. Recent immunoprecipitation studies suggest (see e.g. Birney et al., 2007) that transcription factor binding sites are often symmetrically distributed around the TSS. Furthermore, we apply two different motif searching methods: a threshold-based weight matrix search (Rahmann et al., 2003) and a novel affinity-based approach (Roider et al., 2007). This Transcription Factor Affinity Prediction (TRAP) integrates over all contributions of individual sites using a thermodynamic framework.