OPEN ACCESS
CellPress

Cell Reports
Report

Molecular Architecture of Transcription Factor Hotspots in Early Adipogenesis

Rasmus Siersbæk,[1] Songjoon Baek,[2] Atefeh Rabiee,[1] Ronni Nielsen,[1] Sofie Traynor,[1] Nicholas Clark,[3] Albin Sandelin,[3] Ole N. Jensen,[1] Myong-Hee Sung,[2] Gordon L. Hager,[2] and Susanne Mandrup[1,*]

[1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark
[2]Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, USA
[3]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre, University of Copenhagen, 2200 Copenhagen N, Denmark
*Correspondence: s.mandrup@bmb.sdu.dk
http://dx.doi.org/10.1016/j.celrep.2014.04.043

SUMMARY

Transcription factors have recently been shown to colocalize in hotspot regions of the genome, which are further clustered into super-enhancers. However, the detailed molecular organization of transcription factors at hotspot regions is poorly defined. Here, we have used digital genomic footprinting to precisely define factor localization at a genome-wide level during the early phase of 3T3-L1 adipocyte differentiation, which allows us to obtain detailed molecular insight into how transcription factors target hotspots. We demonstrate the formation of ATF-C/EBP heterodimers at a composite motif on chromatin, and we suggest that this may be a general mechanism for integrating external signals on chromatin. Furthermore, we find evidence of extensive recruitment of transcription factors to hotspots through alternative mechanisms not involving their known motifs and demonstrate that these alternative binding events are functionally important for hotspot formation and activity. Taken together, these findings provide a framework for understanding transcription factor cooperativity in hotspots.

INTRODUCTION

The recent increase in the number of genome-wide maps of transcription factor binding has made it increasingly clear that transcription factors do not work alone but rather collaborate with other transcription factors at genomic target regions. Colocalization of two cooperating factors has been observed in several different systems, e.g., activating protein 1 (AP1) has been shown to regulate the chromatin accessibility at glucocorticoid receptor (GR)-binding sites in mammary cells (Biddie et al., 2011), FoxA1 regulates the estrogen receptor (ER)-activated gene program by directly targeting ER-binding sites (Carroll et al., 2005; Hurtado et al., 2011), PU.1 cooperates with several different factors at specific target sites including CCAAT/enhancer-binding protein β (C/EBPβ) (Heinz et al., 2010), and peroxisome proliferator-activated receptor γ and C/EBPα-binding profiles overlap extensively in mature adipocytes (Lefterova et al., 2008; Nielsen et al., 2008) and cooperate in gaining access to chromatin (Madsen et al., 2014). In addition to cooperation between transcription factor pairs, we (Boergesen et al., 2012; Siersbæk et al., 2011; 2014, this issue of *Cell Reports*) and others (Chen et al., 2008; Gerstein et al., 2012; He et al., 2011) have demonstrated that multiple transcription factors can target the same genomic regions that we refer to as transcription factor hotspots (Siersbæk et al., 2011). Cooperative binding of transcription factors is also the basis for modeling of *cis*-regulatory modules determining cell specificity (Frith et al., 2003; Krivan and Wasserman, 2001). However, chromatin immunoprecipitation sequencing (ChIP-seq), which is the most widely used method to map transcription factor binding, has a relatively low resolution and does not discriminate between direct protein binding to DNA and indirect recruitment to chromatin through protein-protein interactions. The molecular architecture (i.e., the organization) of transcription factors at hotspots is therefore poorly understood.

Here, we have used digital genomic footprinting to precisely define protein localization for several adipogenic transcription factors at a genome-wide level. In combination with ChIP-seq data, these analyses reveal molecular insight into the organization of transcription factors at hotspot regions, which provides a framework for understanding transcription factor cooperativity on chromatin.

RESULTS

Digital Genomic Footprinting Reveals Precise Protein Footprints at a Genome-wide Level

To begin to understand how transcription factors communicate in hotspots, it is essential to know how the factors are organized at these regions, including which factors are engaged in direct DNA interactions. We therefore employed high-resolution digital genomic footprinting (Figure 1A), a recently developed method to identify protein footprints (i.e., areas of restricted nuclease access) within DNase I hypersensitive (DHS) regions based on ultradeep sequencing (>100 M sequence tags) of DHS-seq

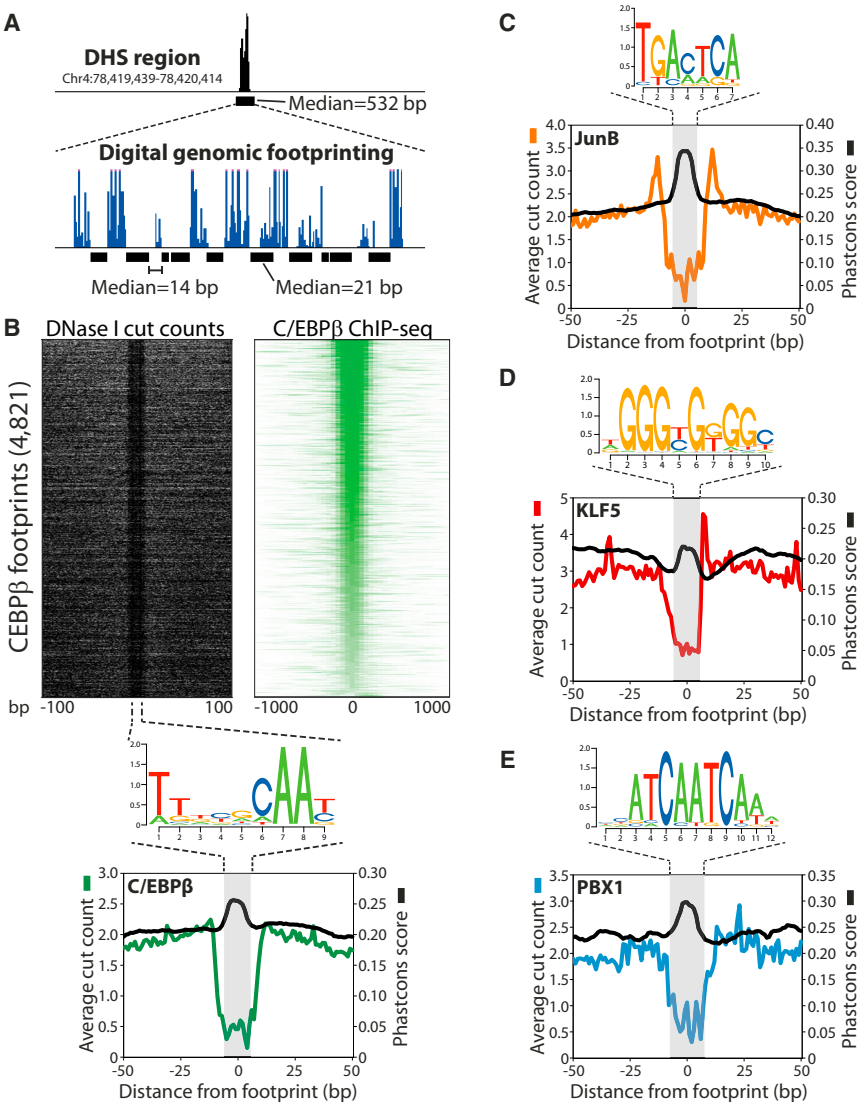

Figure 1. Digital Genomic Footprinting Reveals Transcription Factor Footprints at a Genome-wide Level

(A) Schematic overview of digital genomic footprinting. Extensive sequencing reveals small protected areas of 8–30 bp within overall DNase I hypersensitive regions corresponding to protein footprints. The median size of DHS regions and footprints as well as the median distance from one footprint to the nearest neighbor obtained from Figure S1 are indicated.

(B) DNase I cut counts (left) and C/EBPβ ChIP-seq signal (right) in the vicinity of C/EBPβ footprints (top). These regions were defined as footprints containing a C/EBP predicted site that overlap a ChIP-seq peak for C/EBPβ. Note the different scales used for visualization of DNase I cut counts and ChIP-seq data. Average DNase I cut counts and phastCons score in the vicinity of C/EBPβ footprints are shown at the bottom.

(C–E) Average DNase I cut counts as in (B) for JunB (C), KLF5 (D), and PBX1 (E).

See also Figure S1.

libraries (Boyle et al., 2011; Hesselberth et al., 2009; Neph et al., 2012). We used an algorithm that we have recently developed to efficiently identify protein footprints at a genome-wide level with high sensitivity and specificity (M.-H.S., M. Guertin, S.B., and G.L.H., unpublished data). Based on analysis of 129,028 DHS regions detected 4 hr after induction of 3T3-L1 adipogenesis, we identify ~330,000 protein footprints (false discovery rate = 1%) using this algorithm. These footprints have a median size of 21 bp but are mostly relatively short (8–12 bp) and long (25–30 bp) (Figure S1B), and they are located with a median distance of 14 bp from each other within DHS regions (Figure S1C). The small sizes of the identified footprints clearly demonstrate the increased precision with which footprinting analysis can map protein binding to DNA compared to regular DHS-seq analyses, which identify accessible regions with a median size of 532 bp (Figure S1A).

To assign the identified footprints to specific transcription factors, we combined motif analyses with ChIP-seq data for several transcription factors involved in the early phase of adipocyte differentiation (Siersbæk et al., 2011, 2014). Footprints were assigned to specific factors based on the presence of a ChIP-seq peak for that factor as well as a predicted binding site using the known position weight matrix for the same factor (unless otherwise noted, predicted sites discussed below will always refer to predictions made by the corresponding canonical motifs in this way—see Supplemental Experimental Procedures for more details). As illustrated in Figure 1, this approach revealed strong footprints for several factors. The localized decrease in DNase I digestion at footprints is highly correlated with a colocalized peak in sequence conservation (Figures 1B–1E), indicating that generally these footprints occur at functionally important sites.

C/EBP-ATF Footprints at a Composite Motif Demonstrate Heterodimer Formation on Chromatin

The high resolution of genomic footprinting allows us to obtain information about protein localization within ChIP-seq peak regions (see Figure 1B for an example of the high resolution of footprinting analyses compared to ChIP-seq. Note the different scales used for ChIP-seq and DNase I cut count visualization). Detailed investigation of the distances between footprints for different factors revealed that ATF footprints overlap specifically with C/EBP footprints, whereas minimal overlap was observed between ATF footprints and footprints for KLFs and AP1 (Figure 2A). Interestingly, the exact same composite motif containing a half site from the ATF motif in the 5′ end and a half site from the C/EBP motif in the 3′ end was found at almost all these regions,