HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy.
HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy

Joëlle Michaud, Viviane Praz, Nicole James Faresse, et al.

Genome Res. 2013 23: 907-916 originally published online March 28, 2013
Access the most recent version at doi:10.1101/gr.150078.112
HCFCI is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy

Joëlle Michaud, Viviane Praz, Nicole James Faresse, Courtney K. JnBaptiste, Shweta Tyagi, Frédéric Schütz, and Winship Herr

In human transcriptional regulation, DNA-sequence-specific factors can associate with intermediaries that orchestrate interactions with a diverse set of chromatin-modifying enzymes. One such intermediary is HCFC1 (also known as HCF-1). HCFC1, first identified in herpes simplex virus transcription, has a poorly defined role in cellular transcriptional regulation. We show here that, in HeLa cells, HCFC1 is observed bound to 5400 generally active CpG-island promoters. Examination of the DNA sequences underlying the HCFC1-binding sites revealed three sequence motifs associated with the binding of (1) ZNF143 and THAP11 (also known as Ronin), (2) GABP, and (3) YY1 sequence-specific transcription factors. Subsequent analysis revealed colocalization of HCFC1 with these four transcription factors at ~90% of the 5400 HCFC1-bound promoters. These studies suggest that a relatively small number of transcription factors play a major role in HeLa-cell transcriptional regulation in association with HCFC1.

[Supplemental material is available for this article.]

In eukaryotes, DNA-sequence-specific transcription factors and chromatin-modifying activities work together to regulate the initiation of transcription at promoters by core-promoter-binding factors and RNA polymerases. There exists also a more limited class of transcriptional regulators whose members coordinate the interaction of the DNA-binding transcription factors and chromatin-modifying activities. One of these factors is the host-cell factor HCFC1 (also known as HCF-1), which was discovered in studies of herpes simplex virus (HSV) transcription (for reviews, see Wysocka and Herr 2003; Kristie et al. 2010) and for which a mechanistic understanding of its cellular role has remained relatively enigmatic, largely because it does not display DNA-binding activity.

HCFC1 is synthesized as a 2035-amino-acid precursor that is cleaved by O-GlcNAc transferase (OGT) to generate a heterodimeric complex of amino-terminal HCFC1N and carboxy-terminal HCFC1C subunits (Capotosti et al. 2011) that regulate different aspects of the cell-division cycle (Julien and Herr 2003).

Although HCFC1 does not display direct DNA-binding activity, it associates with chromatin via a Kelch-repeat domain within the HCFC1N subunit (Wysocka et al. 2001). The Kelch-repeat domain is predicted to form a β-propeller structure that binds to a short sequence motif, D/EHxY, called the HCFC1-binding motif (HBM) (Freiman and Herr 1997; Lu et al. 1998), which is found in several HCFC1-associated DNA-binding transcription factors (for review, see Zargar and Tyagi 2012). HCFC1 likewise associates with a constellation of chromatin-modifying activities. These latter activities include the histone H3 lysine 4 (H3K4) methyltransferases SETD1A and mixed lineage leukemia 1 (MLL), histone demethylases KDM1A and PHF8, histone acetyltransferase (HAT) KAT8, histone deacetylase (HDAC) SINE3A, glycosyl transferase OGT, ubiquitin hydroxase RNF2 (BAP-1), and the phosphatase PPA1 (for references, see Zargar and Tyagi 2012). Both these DNA-binding transcription factors and chromatin-modifying activities can be associated with either or both activation and repression of transcription.

The aforementioned indicates that human HCFC1 could play intimate and complex roles in the regulation of gene transcription, and yet to date, a vision of the roles of HCFC1 at specific sites genomewide has been limited to a study of HCFC1 with the mouse embryonal stem (ES)-cell proliferation factor THAP11 (also known as Ronin) (Dejosez et al. 2010). Here, we have examined HCFC1-genome association through high-throughput chromatin immunoprecipitation (ChIP-seq) analysis of proliferating human HeLa cells. The results show that HCFC1 can associate with the large majority of active promoters in cells.

Results

HCFC1 is a major promoter-binding factor

To identify genomic sites bound by HCFC1, we performed ChIP-seq with asynchronously cycling HeLa cells using either HCFC1C- or HCFC1N-subunit-specific antibodies. Figure 1A shows a representative selection of peaks from HCFC1C and HCFC1N ChIP-seq.
analyses in an annotated 500-kb region of the human X chromosome. In this display, HCFC1 appears to be associated with the 5′-proximal regions of some but not all genes. Genome-wide analysis identified 8097 peaks in the HCFC1 C ChIP-seq peaks that can be categorized with respect to transcription units as shown in Figure 1B and Table 1 (for details, see Methods and Supplemental Tables 1 and 2); the majority of HCFC1 C-binding sites are within or near the 22,048 RefSeq-annotated transcription units (83%), and fully 67% (5400) are located near (−1000 to +500 bp) a transcription-start site (TSS). This strong preference for HCFC1 binding near TSSs can also be observed in a cumulative plot of all HCFC1 peaks (Fig. 1C). Of the 17% of HCFC1 C-binding sites located far from transcription units (“Intergenic” in Fig. 1B), many are also likely to be associated with transcription units (Table 1). The large number of HCFC1 C-binding sites is consistent with HCFC1 being an abundant chromatin-associated factor (Wysocka et al. 2001); its presence near many TSSs (at least 5400 out of 22,048, or a quarter) suggests the potential for a broad role in transcriptional regulation.

Because HCFC1 is a heterodimer of HCFC1N and HCFC1C subunits, we also performed ChIP-seq analysis for the HCFC1 N subunit. This analysis revealed a similar number of HCFC1 N (8235) and HCFC1 C (8097) binding sites (Table 1), but they did not always correspond to one another (Fig. 1A). A supplementary quantitative PCR analysis of selected HCFC1N and HCFC1C-binding sites (Supplemental Fig. 2) revealed that TSS-associated binding sites identified in just one of the two ChIP-seq analyses are, in fact, occupied by both subunits. In contrast, a parallel analysis of intergenic binding sites indicated that only those appearing in both ChIP-seq analyses are true binding sites (Supplemental Fig. 2D). These results suggest that HCFC1 asso- ciates with chromatin as an HCFC1N–HCFC1C heterodimer, consistent with the recent finding that the two HCFC1 subunits associate via the formation of a stable hybrid Fibronectin type III–repeat structure (Park et al. 2012). Consistent with this hypothesis, cumulative mapping of HCFC1 subunit binding sites (Fig. 1E) indicates that on average both subunits bind 40 bp upstream of the TSS.

Because the HCFC1C analysis appeared more robust and the HCFC1N analysis revealed few additional HCFC1-binding sites (347 or 6%) (Fig. 1D), we chose to refer to the HCFC1C analysis for the studies described below.
The presence of HCFC1 at TSSs correlates with transcriptional activity

Given the association of HCFC1 with a quarter of annotated TSSs in proliferating HeLa cells, we asked whether HCFC1-bound and -unbound TSSs might differ in associated transcriptional activity (Fig. 2). We thus performed ChIP-seq analyses for trimethylated H3K4 (H3K4Me3) and TSS-associated RNA polymerase II (Pol II), markers for active promoters, and trimethylated H3K36 (H3K36Me3), a marker for gene transcription (Bannister et al. 2005). Figure 2A shows that HCFC1 N- and HCFC1 C-subunit binding is highly correlated with the H3K4Me3 and Pol II active promoter marks. Cumulative mapping shows that the two HCFC1 subunits are generally positioned just upstream of the Rpb2 Pol II subunit and in close proximity to the TSS region (Fig. 1E).

To investigate further the relationship of HCFC1 TSS association and transcriptional activity, we divided the 22,048 RefSeq-annotated TSSs into four different groups (absent, low, medium, and high HCFC1 occupancy) according to the density of HCFC1 sequence tags over the –250- to +250-bp TSS region. Figure 2B (panel a) shows the distribution of HCFC1-binding densities of the TSS regions for each selected group. In parallel, we plotted for the same four TSS groups the corresponding (1) TSS-associated Pol II (panel b) and H3K4Me3 (panel c) density distributions, and (2) transcription unit–associated H3K36Me3 (panel d) density distribution. Comparison shows an excellent correlation between HCFC1 and Pol II, H3K4Me3, and H3K36Me3 occupancy, suggesting that HCFC1 TSS presence correlates with transcriptional activity. Consistent with this hypothesis, HCFC1 TSS occupancy also correlates with mRNA transcript levels as determined by Affymetrix microarray analysis (Fig. 2B, panel c), a measure, albeit indirect, of transcriptional activity.

Depletion of HCFC1 results in changes in mRNA level for a large number of genes

Given the association of HCFC1 with active promoters, we asked whether genes are generally misregulated in the absence of HCFC1. We compared in triplicate mRNA levels in cells treated with siRNA against either HCFC1 or, as a negative control, luciferase (Supplemental Fig. 3A). We observed that for 19% of the transcription units analyzed, the relative level of mRNA changed either up or down—with a multiple testing corrected P-value of ≤0.05—upon depletion of HCFC1. This large number of changes in mRNA level upon HCFC1 depletion suggests a broad role for HCFC1 in the regulation of gene transcription. Consistent with this conclusion, although the genes affected differ, a broad effect on gene expression has also been observed using another strategy of HCFC1 inactivation called cytoplasmic sequestration (Khurana and Kirstie 2004).

For up-regulated genes (i.e., associated with HCFC1 repression), HCFC1-bound and -unbound TSSs revealed little difference in the number of genes affected (Fig. 3, bars 1 and 2) or in the level of response (Supplemental Fig. 3B, plots 1 and 3). For down-regulated genes (i.e., associated with HCFC1 activation), however, there was a large difference (7% vs. 16%) in the number (Fig. 3, bars 3 and 4) as well as in the level of response (Supplemental Fig. 3B, plots 2 and 4) between HCFC1-unbound and -bound TSSs. These observations suggest that HCFC1 directly and indirectly regulates gene transcription with more often a direct role in transcriptional activation.

Indeed, the 74% of HCFC1-bound genes for which we do not detect a change in gene expression upon depletion of HCFC1 could in some cases represent genes that are regulated by HCFC1 at a particular stage of the cell cycle and whose differential expression might therefore be missed in an analysis of a heterogeneous population of asynchronized cells. To test this hypothesis, we synchronized cells in S phase following a double-thymidine block and performed both ChIP-seq of the HCFC1 N subunit and comparative gene-expression microarray analysis of HCFC1-depleted and nondepleted cells. The large majority (89%) of the HCFC1-bound TSSs that are associated with genes showing no expression difference upon HCFC1 depletion in asynchronized cells are still HCFC1-bound in S phase—interestingly, while not differentially expressed in asynchronized HCFC1-depleted cells, 264 (8%) of these genes significantly differ in expression in S phase (Supplemental Table 3). Consistent with an HCFC1-specific role in the regulation of gene transcription, the ChIP-seq analysis shows that, among these 264 S-phase-specific genes, 188 (71%) show a TSS-associated increase in HCFC1 occupancy as shown for two examples in Supplemental Figure 5. These results indicate that HCFC1 plays a broad role in cell-cycle-specific regulation of gene transcription.

Potential regulatory pathways controlled by HCFC1

The HCFC1-depletion microarray gene-expression analysis provides a tool to identify HCFC1 controlled regulatory pathways by Gene Ontology analysis (Supplemental Table 4). Such an analysis revealed many genes involved in cellular metabolism, with genes activated by HCFC1 being mostly involved in the ubiquitin cycle, DNA replication, cell division, and spindle formation. Perturbation of any one or all of these processes could play a role in the cell-cycle

---

Table 1. Location of HCFC1 peaks

<table>
<thead>
<tr>
<th>Feature</th>
<th>HCFC1C ChIP-seq peaks</th>
<th>% of total HCFC1C ChIP-seq peaks</th>
<th>HCFC1N ChIP-seq peaks</th>
<th>% of total HCFC1N ChIP-seq peaks</th>
<th>HCFC1C peaks positive for HCFC1N peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features (total)</td>
<td>8097</td>
<td>100</td>
<td>8235</td>
<td>100</td>
<td>3866</td>
</tr>
<tr>
<td>RefSeq transcription unit*</td>
<td>6742</td>
<td>83</td>
<td>6315</td>
<td>77</td>
<td>288</td>
</tr>
<tr>
<td>TSS PROX region of transcribed unit</td>
<td>5460 (5400)</td>
<td>67</td>
<td>3666 (4153)</td>
<td>45</td>
<td>3806</td>
</tr>
<tr>
<td>Intergenic*</td>
<td>1347</td>
<td>17</td>
<td>1920</td>
<td>23</td>
<td>288</td>
</tr>
<tr>
<td>Alternative transcription unit*</td>
<td>618 (180)</td>
<td>7.6</td>
<td>682</td>
<td>8.3</td>
<td>192</td>
</tr>
<tr>
<td>Small noncoding RNAs</td>
<td>39</td>
<td>0.5</td>
<td>22</td>
<td>0.3</td>
<td>7</td>
</tr>
<tr>
<td>Repeat elements</td>
<td>395</td>
<td>4.8</td>
<td>528</td>
<td>6.4</td>
<td>22</td>
</tr>
<tr>
<td>No feature</td>
<td>295</td>
<td>3.6</td>
<td>688</td>
<td>8.4</td>
<td>67</td>
</tr>
</tbody>
</table>

*See Methods for details.
*Two HCFC1C peaks can correspond to the same HCFC1N peak.
*Peak number falling within TSS-PROX regions of these transcription units.
defects (G1/S arrest and M phase) observed in the absence of HCFC1. The genes more likely to be direct targets of HCFC1 regulation (i.e., both change expression and are bound by HCFC1) affect broad processes such as gene expression and RNA processing. These results are consistent with a picture of HCFC1 as a broad regulator of gene expression.

TSS-associated HCFC1-binding sites lie within CpG islands

Approximately two-thirds of the RefSeq transcription unit promoters possess relatively high concentrations of the dinucleotide sequence CpG (Illingworth et al. 2010), referred to as "CpG islands." Comparison of the list of HCFC1-associated TSSs with a list of ex-
enriched in HCFC1-bound promoters, albeit less than the HCFC1 analysis revealed that the E2F1-binding-site motif is, indeed, E2F transcription factors (Knez et al. 2006; Tyagi et al. 2007). Direct transcription factors known to associate with HCFC1, such as those for complemental Fig. 4B).

The MEME analysis did not identify motifs for some transcription factors known to associate with HCFC1, such as those for E2F transcription factors (Knez et al. 2006; Tyagi et al. 2007). Direct analysis revealed that the E2F1-binding-site motif is, indeed, enriched in HCFC1-bound promoters, albeit less than the HCFC1 MEME Motifs 1–3. These results indicate that the HCFC1 MEME Motifs 1–3 are likely the predominant HCFC1-associated motifs but that more restricted HCFC1-binding-site-associated motifs are also involved in HCFC1 function.

To identify DNA-binding transcription factor targets of HCFC1, we compared the three MEME-motif sequences to motifs for DNA-binding transcription factors in the TRANSFAC database (Matys et al. 2006). The 3’ half of HCFC1 MEME Motif 1 showed strong similarity with the binding site for the human zinc-finger transcription factor ZNF143 (also known as Stat). This motif also matches a sequence identified in mouse embryonic stem (ES) cell studies of binding sites for mouse HCFC1 and the DNA-binding transcription factor THAP11 (Dejosez et al. 2010). HCFC1 MEME Motifs 2 and 3 correlate with the binding sites for transcription factors GABP and YY1 (Fig. 5A, panels b and c, respectively). Interestingly, of the four transcription factors associated with these three motif sequences, the human proteins GABP and YY1 and the mouse protein THAP11 have been shown to bind HCFC1 (Vogel and Kristie 2000; Dejosez et al. 2008; Yu et al. 2010). Although Sp1 has been shown to associate with HCFC1 (see Wysocka et al. 2003), the SP1-binding site motif was not enriched at HCFC1-binding sites. Although the reason is unclear, we note that the SP1-binding site is frequent in CpG promoters, which may make an enrichment more difficult to discern.

**Figure 3.** Effect of HCFC1 depletion on mRNA levels. Percentages of HCFC1 unbound (1 and 3) or bound (2 and 4) TSSs that display a significant differential expression upon depletion of HCFC1 for both the up-regulated (1 and 2) and the down-regulated genes (3 and 4). The number of TSSs used here is 20,571 because 7% of TSSs identified by ChIP-seq analyses were not present on the Affymetrix chip.

Because HCFC1 is not known to bind directly to DNA but instead to site-specific proteins that tether it to the DNA (see the introduction), we searched for recurrent motifs in the DNA sequences associated with the HCFC1 peaks to identify candidate proteins that recruit HCFC1 to its target promoters. We searched for motifs from 6 to 20 bp long within the 200-bp sequence (±100 bp) centered on each TSS-associated HCFC1 peak using the sequence-analysis tool MEME (Bailey and Elkan 1994). This analysis revealed three motifs referred to as HCFC1 MEME Motifs 1 (20 bp long), 2 (9 bp), and 3 (12 bp) (Fig. 5A, panels a–c). All three motifs are highly enriched in HCFC1-bound (15%–20%) versus HCFC1-unbound (2%–3%) promoters (Fig. 5B) and are positioned within −10 bp of the HCFC1 peak itself (Supplemental Fig. 4A), suggesting a binding relationship between HCFC1 and the three HCFC1-associated MEME motifs. In contrast, other motifs identified by the MEME analysis were not enriched in HCFC1-bound promoters (see Supplemental Fig. 4B).

The MEME analysis did not identify motifs for some transcription factors known to associate with HCFC1, such as those for E2F transcription factors (Knez et al. 2006; Tyagi et al. 2007). Direct analysis revealed that the E2F1-binding-site motif is, indeed, enriched in HCFC1-bound promoters, albeit less than the HCFC1

**Table 2.** HCFC1 associates with CpG island promoters

<table>
<thead>
<tr>
<th>Feature</th>
<th>Total number</th>
<th>Number falling within CpG islands</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human promoters</td>
<td>22,048</td>
<td>14,558</td>
<td>66%</td>
</tr>
<tr>
<td>HCFC1 peaks in promoters</td>
<td>5747</td>
<td>5559</td>
<td>97%</td>
</tr>
</tbody>
</table>
transcription factors. GABP being an α/β heterodimer, we analyzed the DNA-binding GABPA subunit. Of these four DNA-binding transcription factors, three displayed, like HCFC1, a strong binding-site bias for TSSs: Thus, 49% of the 3572 ZNF143-binding sites, 61% of the 8042 GABPA-binding sites, and 36% of the 7757 YY1-binding sites identified were bound within -1000 to +500 bp of an annotated TSS. In contrast, for THAP11, only 10% of the 8657 identified binding sites were similarly associated with TSSs—40% were found within transcription units. Only the THAP11 peaks associated with TSSs, which possessed high scores, were analyzed here.

Our ChIP-seq analyses for each of these transcription factors showed that the TRANSFAC motifs accurately predict binding by these transcription factors with >90% of the identified motif sequences bound by the corresponding transcription factor in the case of ZNF143, YY1, and GABP (Supplemental Fig. 4C). In contrast, we observed more peaks than motif sequences, suggesting that these transcription factors bind directly or indirectly other target motifs as well.

We examined the coassociation of the four transcription factors with HCFC1 at TSSs both in detail (Fig. 7A) and globally (Fig. 5C). Figure 7A shows the gene-rich X-chromosome region from the Immunoglobulin-binding protein 1 (IGBP1) to the OGT gene together with peaks for HCFC1, ZNF143, THAP11, GABPA, and YY1. For comparison, CpG islands (Illingworth et al. 2010) and the regions enriched for histone H3K36Me3-modified nucleosomes (i.e., transcriptionally active) are shown.

Figure 4. HCFC1 and CpG island correlation. (A) Boxplot of the distribution of the expression level in HeLa cells of genes lacking (1) or containing (2) a CpG island (Illingworth et al. 2010) and genes containing a CpG island with a peak for HCFC1 (3) or not (4).

Figure 5. Identification of transcription factors associated with HCFC1-bound TSSs. (A) The HCFC1 MEME Motif logo identified in HCFC1-bound TSS-associated sequences is displayed above the most similar TRANSFAC motif(s). For HCFC1 MEME Motif 1, the motif corresponding to the experimentally described sequence for mouse Thap11 (Ronin) is also shown (Dejosez et al. 2010). (a) ZNF143 and THAP11; (b) GABP; (c) YY1. (B) Percentage of HCFC1-bound (+) or -unbound (−) sequences that contain the HCFC1 MEME motif. (C) Venn diagrams showing the overlap between HCFC1-bound TSS regions and ZNF143, THAP11, GABPA, and YY1-bound TSS regions as identified by ChIP-seq. The percentages of TSS regions that contain the transcription factor-binding sites and are HCFC1 positive are given. (*) Percentage of ZNF143 binding sites that are HCFC1 positive; (+) percentage of THAP11 binding sites that are HCFC1 positive. (D) Cumulative mapping of sequence tags around -250 to +250 of TSSs (determined as in Fig. 1E) shared between HCFC1 and the corresponding transcription factor. The most enriched position is indicated as the distance in base pairs from the TSS, which is indicated as a dashed vertical line.
contrast, HCFC1 itself associates with sets of promoters that are distinguished by a different DNA-binding protein(s) with which it associates, creating distinct subsets of promoters depending on the associated DNA-binding protein as shown in Figure 7. It seems likely that this manner of HCFC1 association with promoters permits it to regulate distinct groups of genes according to the cellular context.

**Discussion**

For many years, the analysis of human transcriptional regulation rested on the analysis of individual viral or cellular promoters (for review, see Lemon and Tjian 2000). These studies accentuated the specificity of transcriptional regulation because different promoters often displayed dependence on different sets of site-specific transcription factors. The availability of ChIP-seq allows a simultaneous genome-wide analysis of gene regulation, as presented here. This analysis has accentuated a high degree of commonality in the regulation of gene transcription. Thus, HCFC1, originally discovered in a highly specific role in viral transcription, when examined in proliferating human HeLa cells is found to be broadly associated with actively transcribed genes.

To date, few transcriptional "coregulators" that, like HCFC1, associate directly with sequence-specific DNA-binding transcription factors but are not known to bind to DNA directly themselves, have been analyzed for genome-wide occupancy. One, p300, does not display the TSS specificity of HCFC1, because it is often associated with enhancers as well as promoters (Visel et al. 2009; Ramos et al. 2010). In addition to showing TSS specificity, HCFC1 also displays a high degree of specificity for CpG-island promoters. This may reflect the association of HCFC1 with chromatin-modifying activities that can directly recognize the CpG dinucleotide (see below).

Although HCFC1 lacks known DNA-binding activity, the sequences underlying HCFC1 TSS-binding sites revealed sequence motifs for the binding of ZNF143 and THAP11 together, and GABP and YY1 individually. Interestingly, these four factors have all been associated with cell proliferation, consistent with the role of HCFC1 in cell-cycle progression. For example, GABP regulates S-phase entry (Yang et al. 2007), and in mice, THAP11 sustains ES-cell proliferation (Dejosez et al. 2008).

ZNF143 and THAP11 have been previously characterized independently. In HeLa cells, ZNF143 regulates both noncoding small RNA and protein-encoding gene transcription units (Anno et al. 2011). ZNF143 and mouse THAP11 both possess transcriptional regulatory roles in mouse ES cells: ZNF143 regulates Nanog gene expression (Chen et al. 2008), which maintains the undifferentiated status of ES cells, and THAP11 maintains ES-cell proliferation (Dejosez et al. 2008) and is often associated with HCFC1 on ES-cell promoters (Dejosez et al. 2010). Shown here, these two transcription factors often co-occupy promoters, and of these co-occupied promoters, 98% also bind HCFC1 (see Fig. 5), perhaps because both ZNF143 and THAP11 can bind HCFC1 independently. These results suggest that HCFC1, ZNF143, and THAP11 often partner to regulate gene expression in different cell types including ES and cancer cells.

**HCFC1: A frequent link between DNA-binding transcription factors and chromatin modifiers**

The studies presented here reveal that HCFC1 is a common component of the transcriptional machinery found at active promoters in proliferating human cells. Figure 8 shows how we imagine HCFC1 playing a role at these promoters: HCFC1 associates at transcrip-
tionally active TSSs with both sequence-specific DNA-binding transcription factors (TFs) and a chromatin-modifying activity, of which HCFC1 interacts with many (see the introduction); here we show HCFC1 with an H3K4 histone methyltransferase (H3K4 HMT) as histone H3K4 methylation is associated with active promoters. Consistent with this model, (1) HCFC1 can bind many different DNA-bound transcription factors often via the short and degenerate D/EHxY HBM sequence; (2) HCFC1 associates with both the MLL and SETD1A H3K4 histone methyltransferases on human promoters (Tyagi et al. 2007); and (3) HCFC1-bound promoters are highly H3K4 methylated (see Fig. 2), which forms binding sites for components of the histone H3K4 methyltransferases (i.e., WDR5) (Wysocka et al. 2005). In addition to these transcription factor characteristics, nearly all promoter-bound HCFC1 is associated with CpG islands. We note with interest that a component of H3K4 histone methyltransferases—the CXXC domain of the MLL protein (Allen et al. 2006) and the CXXC finger protein 1 component of the SETD1A H3K4 histone methyltransferase (Lee and Skalnik 2005) (shown in green)—binds to the unmethylated CpG dinucleotides associated with active promoters.

The model illustrates the multiplicity of contacts via which HCFC1 can associate—indirectly—with many active promoters. We imagine HCFC1 association with promoters being dynamic and thus, for example, initiated and maintained by different mechanisms (e.g., association with DNA-binding transcription factors to initiate promoter association and with histone methyltransferases and CpG

---

**Figure 7.** Correlation between HCFC1 and transcription-factor binding patterns. (A) Genome distribution of peaks for HCFC1c and ZNF143, THAP11, GABP, and YY1 transcription factors. CpG islands (Illingworth et al. 2010) are indicated. Actively transcribed transcription units are indicated using the H3K36Me3 distribution. Genes bound by HCFC1 and actively transcribed are indicated at the bottom of the figure along with their direction of transcription. (B) Summary of the distribution of the transcription-factor binding sites, CpG islands, and H3K36Me3 status for HCFC1-bound TSSs. The global percentage of HCFC1c-bound TSSs containing each of the described features is indicated at the bottom.

**Figure 8.** Schematic representation of an HCFC1-containing transcription initiation complex accentuating the multiple potential HCFC1 contact points at an active CpG-island promoter in human cells (see text for details).
DNA sequences to maintain promoter association). Whichever the case, a multiplicity of interaction points for HCFC1 on the large number of cellular promoters described here likely provides a rich source of targets to regulate human gene transcription.

Methods

Cell culture

HeLa-S cells were grown in suspension in Joklik’s modified Eagle’s medium (JMEM) with 5% of fetal calf serum. Adherent HeLa cells were grown in Dulbecco’s modified Eagle’s medium (DMEM) with 10% of fetal bovine serum. HeLa-S cells were synchronized as described (Tyagi et al. 2007).

Chromatin immunoprecipitation (ChIP)

HeLa-S cells were cross-linked for 8 min using 1% formaldehyde. DNA was isolated and sonicated to 100–300 bp using a Bioruptor (Diagenode) and 30-sec pulses on and off at maximum power. Sonicated DNA was immunoprecipitated, washed, and eluted as described (Tyagi et al. 2007). Two × 10⁷ to 8 × 10⁷ cells were used per ChIP-seq. The following antibodies were used: polyclonal anti-HCFC1 C (H12) (Wilson et al. 1993), polyclonal anti-HCFC1 N (961-1011) (Machida et al. 2009), polyclonal anti-YY1 (sc-1703), and polyclonal anti-THAP11 (G4275) ZNF143 (Yuan et al. 2007), polyclonal anti-GABPA (sc-22810), polyclonal H3K36Me3 (Abcam ab-9050), polyclonal anti-STAT1 (Ruppert et al. 1996) and mixed protein-Sepharose A/G (GE Healthcare). Immunoprecipitated material was immunoblotted using the Flag beads (M2, Sigma-Aldrich a-2220) or with TBP antibody (SL30). Recombinant proteins and GST-pull-down assays were performed using the TNT T7 Quick transcription/translation system (Promega) as recommended by the manufacturer. Equal amounts of GST-ZNF143 and GST-ZNF143/DBD (amino acids 220–428) were used: polyclonal anti-HA (12CA5) or polyclonal anti-ZNF143 (Yuan et al. 2007) antibodies.

Ultra-high-throughput sequencing and analysis

Five to 10 ng of ChIP-DNA was transformed into libraries using the ChIP-seq DNA Sample Prep Kit (Illumina) and sequenced on an Illumina Genome Analyzer 2. Total input DNA was also sequenced. Thirty-eight-base-pair sequence tags were then aligned to the genome (NCBI36/hg18) using Eland and fetchGWI (Iseli et al. 2007). All tags, regardless of the sequencing score, were used for mapping. Only the sequence tags with a unique and perfect match were retained. Moreover, to reduce some potential PCR artifacts, a maximum of two identical sequence tags was kept for further peak identification. Peak identification was performed using SISRs (Jothi et al. 2008) with the program’s default options (FDR = 0.001 as compared to a random background model based on Poisson probabilities), except for the following ones: The fragment length of the library was as determined experimentally before the sequence analysis; overlapping peaks were clustered together. The relatively few peaks (5% on average) common between input and ChIP samples were excluded from subsequent analysis.

Transcription unit annotation

The standard list of protein-encoding genes used in this study was generated from the UCSC Browser gene list filtered using their RefSeq status (NM_ accession numbers for transcripts; priority to validated status, release 37). Alternative 5′ TSSs were kept as distinct transcripts. The transcripts with the longest 3′ ends were used if the TSSs were similar. If no validated RefSeq was available, provisional RefSeq annotations were also used. Alternative lists were generated with noncurated transcription units and other genomic features. The peaks away from any RefSeq sequences (Intergenic) were then compared with these alternative lists in the following order: (1) small noncoding RNAs (from UCSC RNA Genes Table and NR_ accession numbers from RefSeq), (2) Ensembl genes (Hubbard et al. 2002), (3) SIB predicted genes, and (4) repeats.

Quantitative PCR

Quantitative PCR on ChIP samples was performed in duplicate using MESA Blue qPCR Mastermix Plus for SYBR Assay from Eurogentec and a Rotorgene RG300A sequence detector (Corbett Research). The ChIP samples were normalized with the total input DNA amount using the ΔCt method.

Gene expression microarrays

Cells were treated with siRNA against HCFC1 or luciferase, as previously described (Julien and Herr 2003), twice 12 h apart using Oligofectamine (Invitrogen). RNA samples were collected 48 h following the first transfection, labeled using the MessageAmp II-Biotin Enhanced kit (Ambion 1791), and hybridized to the GeneChip Human Genome U133 Plus 2.0 Array. Three biological replicates were used for each condition. The analysis was done using the R statistical software; the intensities of each probe set were normalized and summarized using the Robust Multi-array Analysis algorithm (RMA) (Irizarry et al. 2003), and differential expression was assessed using the LIMMA package (Smyth 2004).

MEME motif search

One hundred base pairs on each side of the center of HCFC1 peaks was used for a de novo search for 6–20 bp motifs using the sequence-analysis tool MEME (Bailey and Elkan 1994) with a combined P-value < 0.01 as threshold. The matrices generated by MEME were then compared with the TRANSFAC database using the DNA-binding motif similarity tool STAMP (Mahony and Benos 2007). The MEME motifs were used to search in other sequence files using the motif alignment and search tool MAST (Bailey and Gribskov 1998).

Coimmunoprecipitation

Nuclear extracts of HeLa Flp-In cells (Tighe et al. 2004) or stable HeLa Flp-In cells synthesizing a Flag- and HA-tagged HCFC1 N1011 fragment were prepared as described (Tyagi et al. 2007). The extracts were precleared and then incubated overnight with anti-Flag beads (M2, Sigma-Aldrich a-2220) or with TBP antibody (SL30) (Ruppert et al. 1996) and mixed protein-Sepharose A/G (GE Healthcare). Coimmunoprecipitated material was immunoblotted using the monoclonal anti-HA (12CA5) or polyclonal anti-ZNF143 (Yuan et al. 2007) antibodies.

Recombinant proteins and GST-pull-down assays

The GST-ZNF143 and GST-ZNF143/DBD proteins were synthesized in bacteria as described (Yuan et al. 2007). Sequences encoding GFP were cloned in the pSBet (Schenk et al. 1995) vector containing a 5′-GST tag and synthesized in bacteria as for ZNF143. In vitro transcription/translation and [35S]methionine labeling of HCFC1 fragments were performed using the TNT T7 Quick transcription/translation system (Promega) as recommended by the manufacturer. Equal amounts of recombinant GST fusion proteins were incubated overnight with glutathione-agarose beads (Sigma-Aldrich G4510) and the indicated HCFC1 fragment.

Data access

Gene expression and sequencing data have been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE31419.
Acknowledgments

We thank the Lausanne Genomic Facility and Affymetrix microarray hybridizations; Philippe LHôte for cell culture; Thomas Zwaika for the HeLa-Flp cell line. This research was supported by the University of Lausanne and SNSF grants CRSII3_125230 and 31003A_130829 to W.H. N.J.F. was supported by the research funding of N. Hernandez.

References


