MIT Libraries | DSpace@MIT

MIT Open Access Articles

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

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Michaud, J., V. Praz, N. James Faresse, C. K. JnBaptiste, S. Tyagi, F. Schutz, and W. Herr. "HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy." Genome Research 23, no. 6 (June 1, 2013): 907-916. © 2013, Published by Cold Spring Harbor Laboratory Press

As Published: http://dx.doi.org/10.1101/gr.150078.112

Publisher: Cold Spring Harbor Laboratory Press

Persistent URL: http://hdl.handle.net/1721.1/82917

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Massachusetts Institute of Technology

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

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.

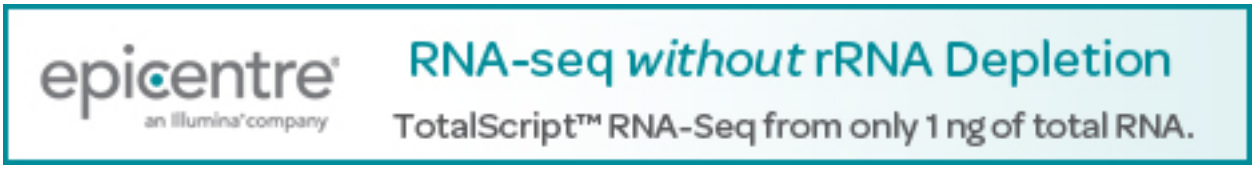
Genome Res. 2013 23: 907-916 originally published online March 28, 2013
Access the most recent version at doi:10.1101/gr.150078.112

Supplemental Material http://genome.cshlp.org/content/suppl/2013/04/02/gr.150078.112.DC1.html

References This article cites 45 articles, 23 of which can be accessed free at: http://genome.cshlp.org/content/23/6/907.full.html#ref-list-1

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

© 2013, Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

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

Joëlle Michaud,[1,3] Viviane Praz,[1] Nicole James Faresse,[1] Courtney K. JnBaptiste,[1,4] Shweta Tyagi,[1,5] Frédéric Schütz,[2] and Winship Herr[1,6]

[1]*Center for Integrative Genomics, University of Lausanne, Génopode, 1015 Lausanne, Switzerland;* [2]*Bioinformatics Core Facility, SIB Swiss Institute of Bioinformatics, Génopode, 1015 Lausanne, Switzerland*

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 $HCFC1_N$ and carboxy-terminal $HCFC1_C$ 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 $HCFC1_N$ subunit (Wysocka et al. 2001). The Kelch-repeat domain is predicted to form a β-propeller structure that binds to a short sequence motif, $^{D}/_{E}$HxY, 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) SIN3A, glycosyl transferase OGT, ubiquitin hydrolase 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 genome-wide 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.

Present addresses: [3]Gene Predictis SA, Site EPFL, PSE-B, CP 128, 1015 Lausanne, Switzerland; [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; [5]Laboratory of Cell Cycle Regulation, Centre for DNA Fingerprinting and Diagnostics (CDFD), Nampally, Hyderabad 500001, India.
[6]Corresponding author
E-mail winship.herr@unil.ch
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.150078.112.

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 $HCFC1_C$- or $HCFC1_N$-subunit-specific antibodies. Figure 1A shows a representative selection of peaks from $HCFC1_C$ and $HCFC1_N$ ChIP-seq

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

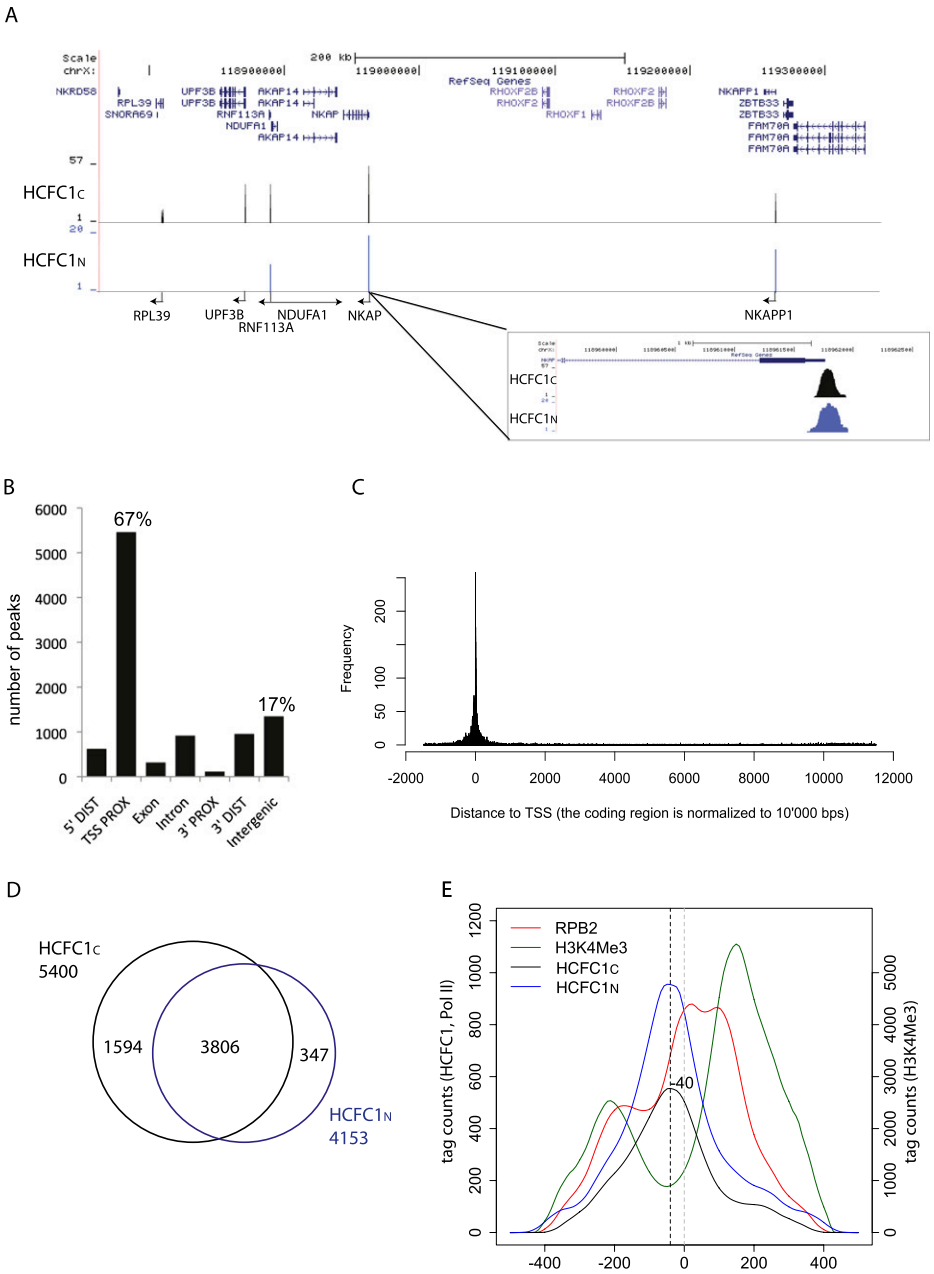


Figure 1. Genomic localization of human HCFC1. (*A*) Typical profile of HCFC1 peaks throughout the genome (UCSC Genome Browser) and magnification of one HCFC1 peak. Both $HCFC1_C$ and $HCFC1_N$ data are shown. (*B*) Distribution of the $HCFC1_C$ peaks according to their genomic localization. (5′ DIST) −5000 to −1000 of TSSs; (TSS-PROX) −1000 to +500 of TSSs; (3′ PROX) −500 to +1000 of end of transcript; (3′ DIST) +1000 to +5000 of end of transcript; exon and intron within the coding sequence only; (Intergenic) >5000 bp distal of annotated transcript regions. Percentages of the total number of peaks are indicated for the TSS-PROX and intergenic regions. (*C*) Enrichment of HCFC1 peaks within and surrounding annotated transcript regions. All transcribed regions are normalized to 10,000 bp. (*D*) Venn diagram showing the number of shared and distinct RefSeq TSSs bound by either $HCFC1_C$ or $HCFC1_N$. (*E*) Cumulative mapping of $HCFC1_N$ (blue), $HCFC1_C$ (black), Pol II (RPB2; red), and H3K4Me3 (green) binding sites within −250 to +250 of TSSs. Binding sites were mapped as the center of ChIP-seq fragments artificially extended to the experimentally determined average fragment length. (*Left* axis) Corresponds to $HCFC1_N$, $HCFC1_C$, and Pol II distribution; (*right* axis) corresponds to H3K4Me3 distribution. The most enriched position for $HCFC1_C$ is indicated as the base pair distance from the TSS and by the dashed black vertical line. TSS is indicated by the dashed gray vertical line.

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 large 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 $HCFC1_N$ and $HCFC1_C$ 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 $HCFC1_N$- and $HCFC1_C$-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 associates with chromatin as an $HCFC1_N$–$HCFC1_C$ 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 $HCFC1_C$ analysis appeared more robust and the $HCFC1_N$ analysis revealed few additional HCFC1-binding sites (347 or 6%) (Fig. 1D), we chose to refer to the $HCFC1_C$ analysis for the studies described below.

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

Table 1. Location of HCFC1 peaks

Feature	$HCFC1_C$ ChIP-seq peaks	% of total $HCFC1_C$ ChIP-seq peaks	$HCFC1_N$ ChIP-seq peaks	% of total $HCFC1_N$ ChIP-seq peaks	$HCFC1_C$ peaks positive for $HCFC1_N$ peaks
All features (total)	8097	100	8235	100	
RefSeq transcription unit[a]	6742	83	6315	77	3866
TSS PROX region of transcribed unit	5460 (5400[b])	67	3666 (4153[b])	45	3806[c]
Intergenic[a]	1347	17	1920	23	288
Alternative transcription unit[a]	618 (180[d])	7.6	682	8.3	192
Small noncoding RNAs	39	0.5	22	0.3	7
Repeat elements	395	4.8	528	6.4	22
No feature	295	3.6	688	8.4	67

[a]See Methods for details.
[b]Number of TSS-RefSeq transcription units bound by the subunit.
[c]Two $HCFC1_C$ peaks can correspond to the same $HCFC1_N$ peak.
[d]Peak number falling within TSS-PROX regions of these transcription units.

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 between H3K4Me3-modified histones (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 e), 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 Kristie 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 $HCFC1_C$ 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

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

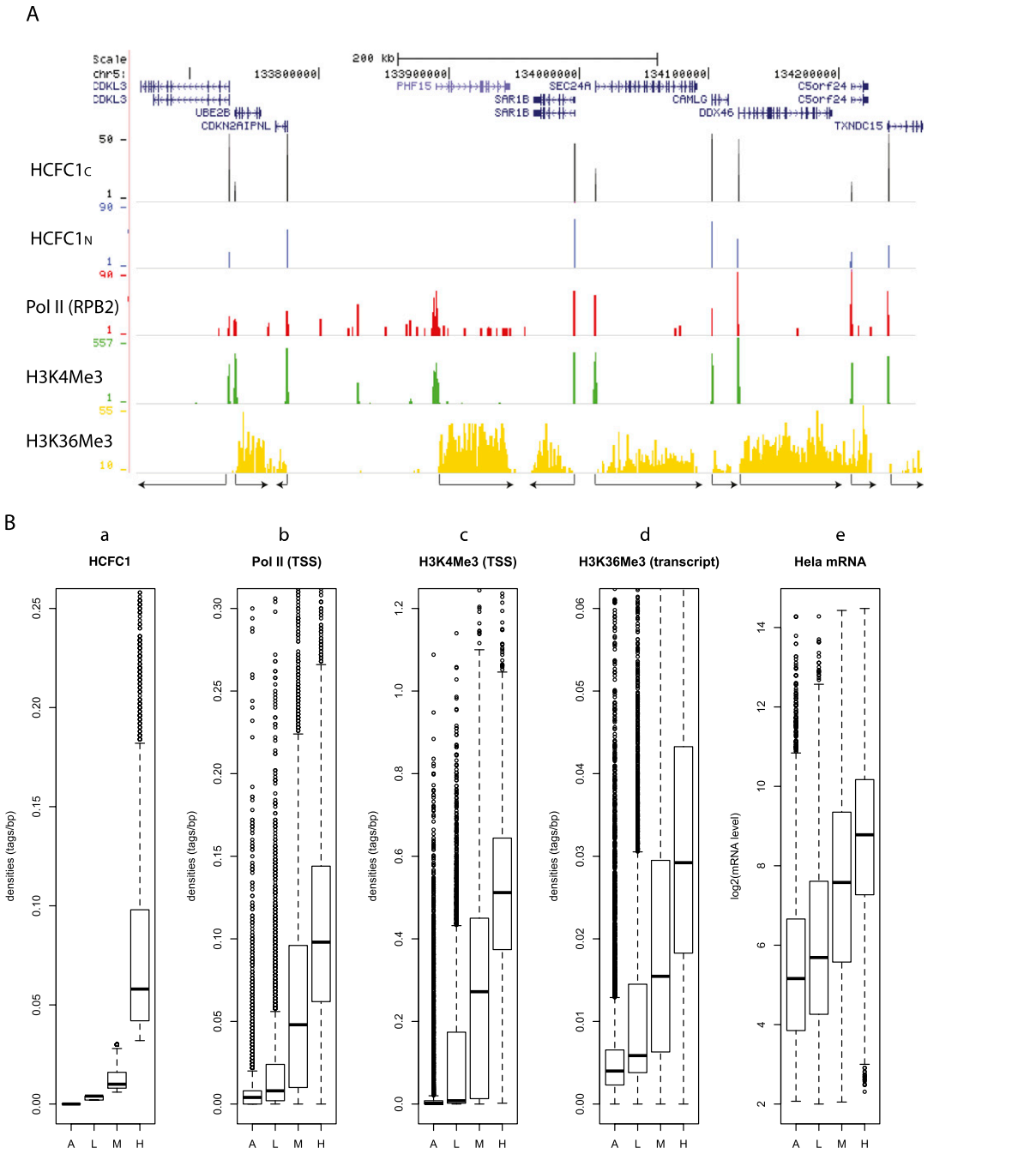


Figure 2. HCFC1 peaks correlate with sites of active promoters. (*A*) UCSC Genome Browser shot showing a typical distribution of $HCFC1_C$ and $HCFC1_N$ binding sites together with Pol II, and H3K4Me3 and H3K36Me3 modified histones. (*B*) TSS regions (−250 to +250 bp from TSSs) were split into four categories according to the density of sequence tags (number of tags centered to the fragment length as in Fig. 1E divided by the number of base pairs, i.e., 500) for $HCFC1_C$ (panel *a*). (A) Absent (density = 0, 6270 transcription units); (L) low density (density > 0 and ≤0.004, 4806 transcription units); (M) medium density (density > 0.004 and ≤0.03, 7902 transcription units); (H) high density (density > 0.03, 3417 transcription units). The sequence-tag densities for Pol II (panel *b*) and H3K4Me3 (panel *c*) at promoter regions and H3K36Me3 (panel *d*) within transcribed regions are shown for each category. The distribution of transcript abundance in $\log_2$ of transcripts in HeLa cells is shown for each category (panel *e*). The results are displayed as boxplots.

defects (G_1/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 HCFCI-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-

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

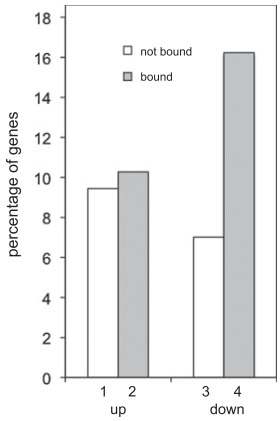


Figure 3. Effect of HCFC1 depletion on mRNA levels. Percentages of HCFC1 unbound (1 and 3) or bound (2 and 4) TSSs that display a significant differential expression upon depletion of HCFC1 for both the up-regulated (1 and 2) and the down-regulated genes (3 and 4). The number of TSSs used here is 20,571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

perimentally defined human CpG-island promoters (Illingworth et al. 2010) showed that there is a strong bias for HCFC1 association with CpG-island promoters: 97% of TSS-associated HCFC1 peaks lie within the CpG island of CpG-island promoters (Table 2). These results suggest that there is a strong link between CpG island–containing promoters and the recruitment of HCFC1.

CpG-island promoters are often more active than non-CpG-island promoters (for review, see Antequera and Bird 1999). We observe the same effect for the HeLa-cell mRNA analysis of Figure 2B (panel e), as shown in Figure 4 (cf. boxplots 1 and 2). Among the genes with CpG-island promoters, those with promoters associated with HCFC1 are even more highly expressed (cf. boxplots 2 and 3 vs. 2 and 4), establishing association of HCFC1 with a large number of active CpG-island promoters.

DNA sequence motifs corresponding to the human transcription factors ZNF143, THAP11, GABP, and YY1 are enriched in HCFC1-binding sites

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 $HCFC1_C$ 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 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 Staf). 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 HCF-1-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.

HCFC1 binds ZNF143

The fourth of the HCFC1 MEME motif–associated transcription factors, human ZNF143, has been primarily characterized in the context of regulation of *RNU6* small nuclear RNA (snRNA) gene transcription (Schaub et al. 1999; Yuan et al. 2007). HCFC1 was previously found in a proteomic screen for ZNF143-associated proteins, but the nature of the association was not defined (Yuan et al. 2007). We therefore defined the association. In an extract from cells containing an epitope-tagged $HCFC1_N$ subunit, epitope-tag-specific immunoprecipitation recovered the endogenous ZNF143 protein (Fig. 6B). A pull-down experiment (Fig. 6C) with a purified GST–ZNF143 fusion protein and HCFC1 fragments synthesized by in vitro translation (Fig. 6A) showed that, in contrast to a control GST–GFP protein (lanes 1,3), the GST–ZNF143 protein (lanes 2,4) recovered all of the amino-terminal $HCFC1_N$ fragments but not the $HCFC1_C$ subunit (cf. lanes 3 and 4). These results suggest that ZNF143 binds the HCFC1 Kelch domain. Within ZNF143, the DNA-binding domain (DBD) is sufficient to bind to the HCFC1 N380 Kelch domain (Fig. 6D). Thus, all four transcription factors identified in the HCFC1 binding-site analysis bind HCFC1.

Nearly 90% of all HCFC1-bound promoters are also bound by one or more of ZNF143, THAP11, GABP, or YY1 transcription factors

To examine the genomic relationship between HCFC1 and the four human ZNF143, THAP11, GABP, or YY1 transcription factors, we performed ChIP-seq analysis for each of these four DNA-binding

Table 2. HCFC1 associates with CpG island promoters

Feature	Total number	Number falling within CpG islands	Percentage
Human promoters	22,048	14,558	66%
HCFC1 peaks in promoters	5747	5559	97%

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

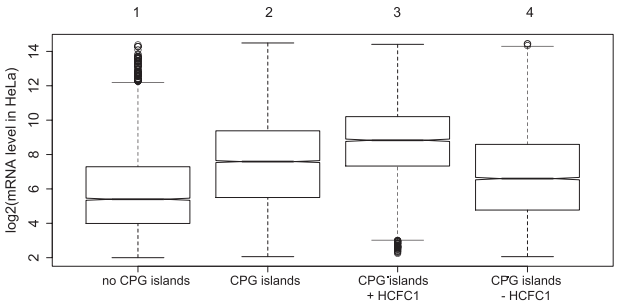


Figure 4. HCFC1 and CpG island correlation. (*A*) Boxplot of the distribution of the expression level in HeLa cells of genes lacking (1) or containing (2) a CpG island (Illingworth et al. 2010) and genes containing a CpG island with a peak for HCFC1 (3) or not (4).

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 $HCFC1_C$, 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.

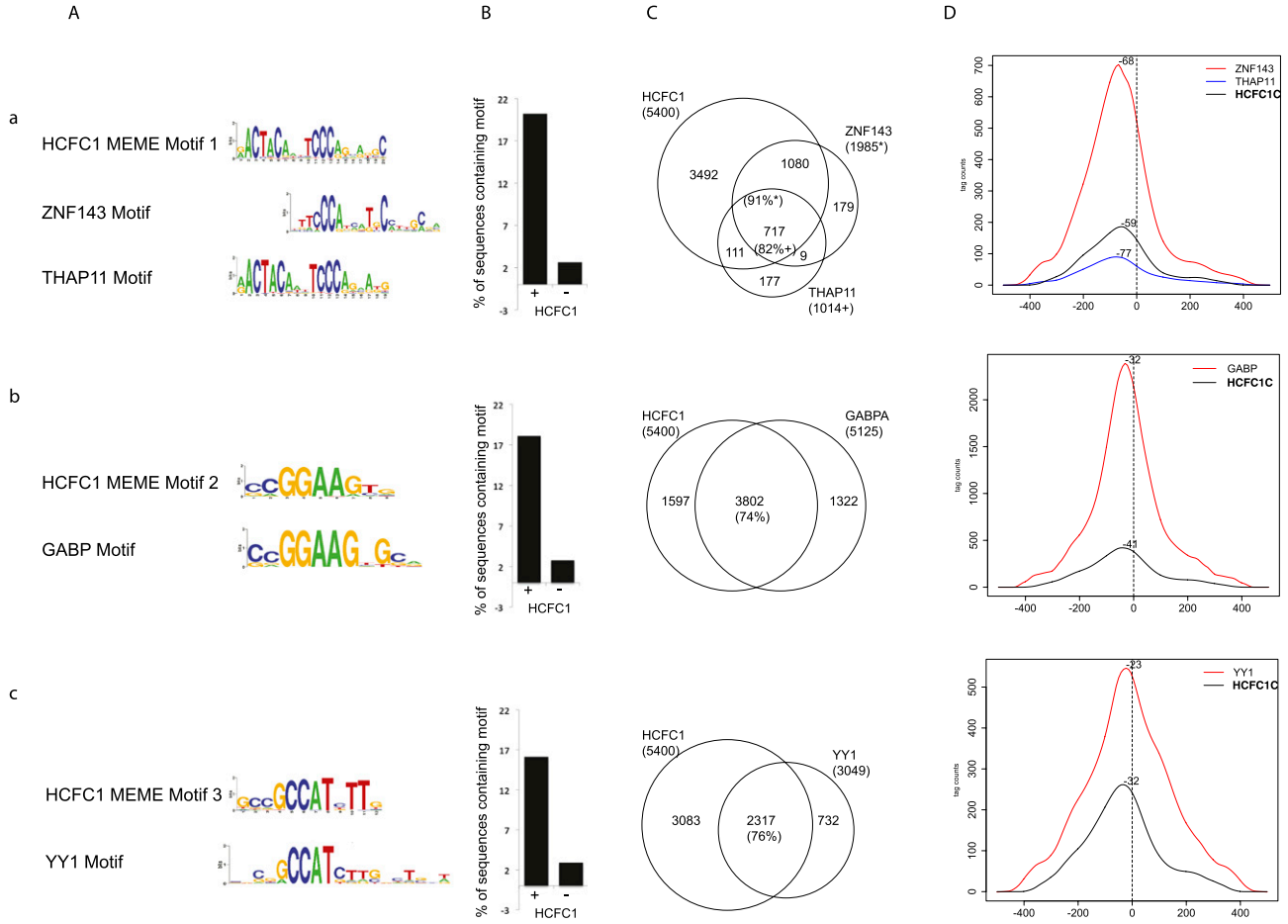


Figure 5. Identification of transcription factors associated with HCFC1-bound TSSs. (*A*) The HCFC1 MEME Motif logo identified in HCFC1-bound TSS-associated sequences is displayed *above* the most similar TRANSFAC motif(s). For HCFC1 MEME Motif 1, the motif corresponding to the experimentally described 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 $HCFC1_C$ 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.

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

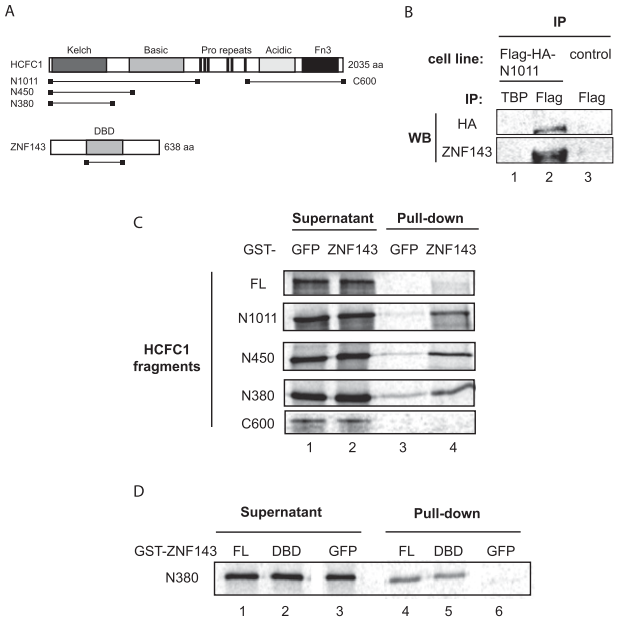


Figure 6. HCFC1 binds ZNF143. (*A*) Schematic representation of the HCFC1 and ZNF143 proteins and fragments used in the analysis. Features are indicated *above* the proteins, and HCFC1 fragment names are indicated *below*. (*B*) Immunoprecipitation (IP) of extracts from control (lane *3*) or Flag-HA–HCFC1$_{N1011}$ positive cells (lanes *1*,*2*) using Flag (lanes *2*,*3*) or a naive (TBP) antibody (lane *1*). The proteins were immunoblotted (WB) using HA or ZNF143 antibodies. (*C*) GST-pull-down assay using either GFP as negative control (lanes *1*,*3*) or ZNF143 recombinant protein (lanes *2*,*4*) and different in vitro–translated full-length or fragments of HCFC1 as indicated to the *left*. (*D*) GST-pull-down assay using different fragments of ZNF143 (full-length [lanes *1*,*4*]; DNA-binding domain [lanes *2*,*5*]) or GFP (lanes *3*,*6*), and the Kelch domain (N380) of HCFC1. (Supernatant) Unbound material after pull-down; (pull-down) material recovered on the glutathione-agarose beads.

Seven, all TSS-associated, HCFC1$_C$ peaks (labeled 1–7) were identified of which six are CpG-island associated and five associated with histone H3K36Me3 positive transcription units, consistent with the aforementioned HCFC1 association with active transcription units (see Fig. 2B). Interestingly, all seven HCFC1$_C$-bound TSSs are also associated with at least one of the four transcription factors (Fig. 7, cf. A and B). This typical region illustrates how the nature of specific transcription factors that colocalize with HCFC1 can be highly variable.

Genome-wide, 88% of the 5400 HeLa-cell HCFC1-bound TSSs are also bound by one or more of the four transcription factors analyzed. Among the four transcription factors, we observed most prominent coassociation on promoters between THAP11 and ZNF143 (71% of THAP11 TSS-associated sites also display ZNF143 binding), consistent with the overlap between the ZNF143 and THAP11 DNA-binding-site motifs in HCFC1 MEME Motif 1 (Fig. 5A, panel a), as well as the adjacent factor binding (see cumulative mapping in Fig. 5D, panel a). Interestingly, whereas 71% of THAP11-bound TSSs also display ZNF143 binding, only 37% (726 out of 1985) of ZNF143-binding sites are associated with THAP11-binding sites.

A high percentage of the promoters bound by each of the four DNA-binding transcription factors was also bound by HCFC1 (ZNF143, 91%; THAP11, 82%; GABPA, 74%; YY1, 76%) (Fig. 5C). Furthermore, the colocalization of HCFC1 and these four DNA-binding transcription factors on promoters revealed by the cumulative mapping shown in Figure 5D suggests that they are often associated with HCFC1 when bound to their target promoters. In 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).

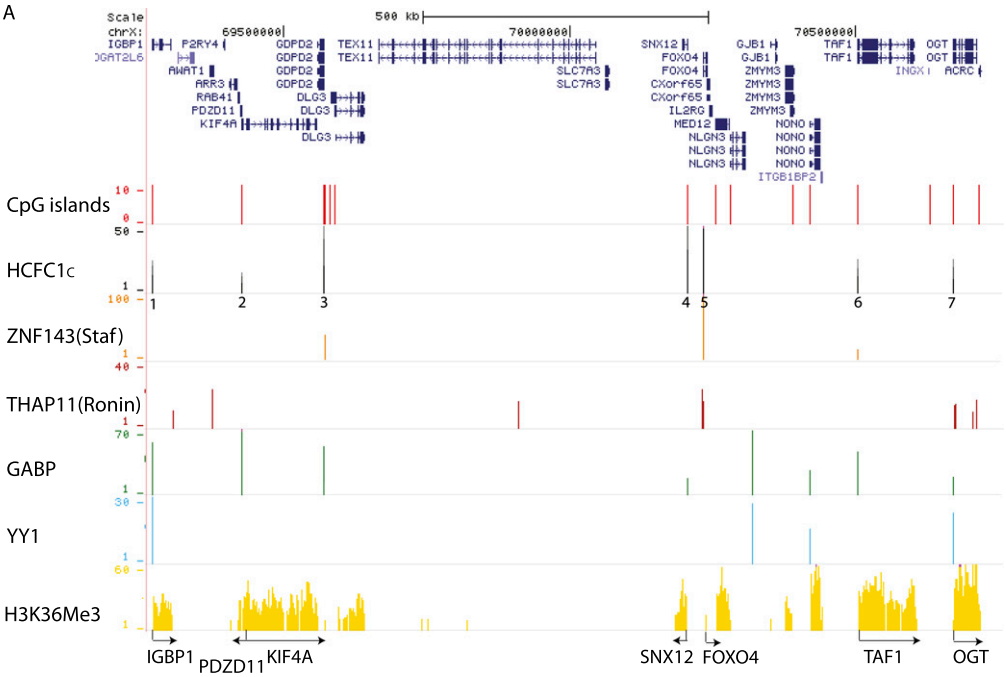
ZNF143 and THAP11 have been previously characterized independently. In HeLa cells, ZNF143 regulates both noncoding small RNA and protein-encoding gene transcription units (Anno et al. 2011). ZNF143 and mouse THAP11 both possess transcriptional regulatory roles in mouse ES cells: ZNF143 regulates *Nanog* 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-

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

A



B

	Gene name	Nucleotide composition	Transcription factors				Histone modification
HCFC1C peak		CpG island	ZNF143	THAP11	GABP	YY1	H3K36Me3
1	*IGBP1*	+	-	-	+	+	+
2	*PDZD11/KIFA4*	+	-	-	+	-	+
3	-	+	+	-	+	-	-
4	*SNX12*	+	-	-	+	-	+
5	*FOXO4*	-	+	+	-	-	-
6	*TAF1*	+	+	-	+	-	+
7	*OGT*	+	-	+	+	+	+
5400 total		96%	32%	15%	68%	42%	

Figure 7. Correlation between HCFC1 and transcription-factor binding patterns. (*A*) Genome distribution of peaks for $HCFC1_C$ 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 $HCFC1_C$-bound TSSs containing each of the described features is indicated at the *bottom*.

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 $^{D}/_{E}$HxY 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

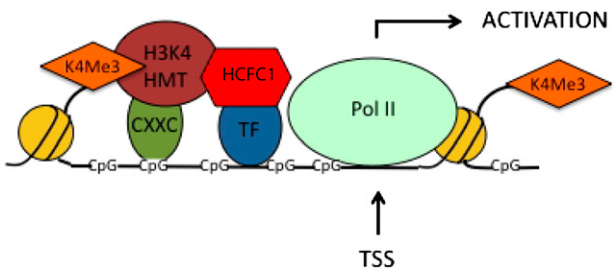


Figure 8. Schematic representation of an HCFC1-containing transcription initiation complex accentuating the multiple potential HCFC1 contact points at an active CpG-island promoter in human cells (see text for details).

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

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 (Bioruptor UCD-200; 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 $\times$ 10^7 to 8 $\times$ 10^7 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-Pol II (POLR2B, sc-67318), polyclonal H3K4Me3 (Abcam ab-8580), polyclonal H3K36Me3 (Abcam ab-9050), polyclonal anti-ZNF143 (Yuan et al. 2007), polyclonal anti-GABPA (sc-22810), polyclonal anti-YY1 (sc-1703), and polyclonal anti-THAP11 (G4275) (Dejosez et al. 2008).

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 SISSRs (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 probeset 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- to 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). Immunoprecipitated 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 (amino acids 220–428) proteins were synthesized in bacteria as described (Yuan et al. 2007). Sequences encoding GFP were cloned in the T7 expression system from a pSBet (Schenk et al. 1995) vector containing a 5′-GST tag and synthesized in bacteria as for ZNF143. In vitro transcription/translation and [^{35}S]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.

Downloaded from genome.cshlp.org on September 11, 2013 - Published by Cold Spring Harbor Laboratory Press

Acknowledgments

We thank the Lausanne Genomic Technologies Facility for performing the ultra-high-throughput sequencing and Affymetrix microarray hybridizations; Philippe L'Hôte for cell culture; Thomas Zwaka for the THAP11/Ronin antibody; Nouria Hernandez for discussions and critical reading of the manuscript; and S.S. Taylor for the HeLa-Flp cell line. This research was supported by the University of Lausanne and SNSF grants CRSI33_125230 and 31003A_130829 to W.H. N.J.F. was supported by the research funding of N. Hernandez.

References

Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, Johnson CM, Freund SM, Bycroft M, Warren AJ. 2006. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J* **25:** 4503–4512.

Anno YN, Myslinski E, Ngondo-Mbongo RP, Krol A, Poch O, Lecompte O, Carbon P. 2011. Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription. *Nucleic Acids Res* **39:** 3116–3127.

Antequera F, Bird A. 1999. CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr Biol* **9:** R661–R667.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28–36.

Bailey TL, Gribskov M. 1998. Combining evidence using *P*-values: Application to sequence homology searches. *Bioinformatics* **14:** 48–54.

Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280:** 17732–17736.

Capotosti F, Guernier S, Lammers F, Waridel P, Cai Y, Jin J, Conaway JW, Conaway RC, Herr W. 2011. *O*-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144:** 376–388.

Chen X, Fang F, Liou YC, Ng HH. 2008. Zfp143 regulates *Nanog* through modulation of Oct4 binding. *Stem Cells* **26:** 2759–2767.

Dejosez M, Krumenacker JS, Zitur LJ, Passeri M, Chu LF, Songyang Z, Thomson JA, Zwaka TP. 2008. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* **133:** 1162–1174.

Dejosez M, Levine SS, Frampton GM, Whyte WA, Stratton SA, Barton MC, Gunaratne PH, Young RA, Zwaka TP. 2010. Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev* **24:** 1479–1484.

Freiman RN, Herr W. 1997. Viral mimicry: Common mode of association with HCF by VP16 and the cellular protein LZIP. *Genes Dev* **11:** 3122–3127.

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, et al. 2002. The Ensembl genome database project. *Nucleic Acids Res* **30:** 38–41.

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6:** e1001134.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31:** e15.

Iseli C, Ambrosini G, Bucher P, Jongeneel CV. 2007. Indexing strategies for rapid searches of short words in genome sequences. *PLoS ONE* **2:** e579.

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. 2008. Genome-wide identification of in vivo protein–DNA binding sites from ChIP-seq data. *Nucleic Acids Res* **36:** 5221–5231.

Julien E, Herr W. 2003. Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *EMBO J* **22:** 2360–2369.

Khurana B, Kristie TM. 2004. A protein sequestering system reveals control of cellular programs by the transcriptional coactivator HCF-1. *J Biol Chem* **279:** 33673–33683.

Knez J, Piluso D, Bilan P, Capone JP. 2006. Host cell factor-1 and E2F4 interact via multiple determinants in each protein. *Mol Cell Biochem* **288:** 79–90.

Kristie TM, Liang Y, Vogel JL. 2010. Control of α-herpesvirus IE gene expression by HCF-1 coupled chromatin modification activities. *Biochim Biophys Acta* **1799:** 257–265.

Lee JH, Skalnik DG. 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys[4] methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* **280:** 41725–41731.

Lemon B, Tjian R. 2000. Orchestrated response: A symphony of transcription factors for gene control. *Genes Dev* **14:** 2551–2569.

Lu R, Yang P, Padmakumar S, Misra V. 1998. The herpesvirus transactivator VP16 mimics a human basic domain leucine zipper protein, luman, in its interaction with HCF. *J Virol* **72:** 6291–6297.

Machida YJ, Machida Y, Vashisht AA, Wohlschlegel JA, Dutta A. 2009. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J Biol Chem* **284:** 34179–34188.

Mahony S, Benos PV. 2007. STAMP: A web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res* **35:** W253–W258.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. 2006. TRANSFAC and its module TRANSCompel: Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34:** D108–D110.

Park J, Lammers F, Herr W, Song JJ. 2012. HCF-1 self-association via an interdigitated Fn3 structure facilitates transcriptional regulatory complex formation. *Proc Natl Acad Sci* **109:** 17430–17435.

Ramos YF, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, van Galen M, van Dam H, van Ommen GJ, den Dunnen JT, Zantema A, et al. 2010. Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. *Nucleic Acids Res* **38:** 5396–5408.

Ruppert SM, McCulloch V, Meyer M, Bautista C, Falkowski M, Stunnenberg HG, Hernandez N. 1996. Monoclonal antibodies directed against the amino-terminal domain of human TBP cross-react with TBP from other species. *Hybridoma* **15:** 55–68.

Schaub M, Krol A, Carbon P. 1999. Flexible zinc finger requirement for binding of the transcriptional activator Staf to U6 small nuclear RNA and tRNA^{Sec} promoters. *J Biol Chem* **274:** 24241–24249.

Schenk PM, Baumann S, Mattes R, Steinbiss HH. 1995. Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare ^{Arg}tRNAs. *BioTechniques* **19:** 196–200.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3.

Tighe A, Johnson VL, Taylor SS. 2004. Truncating APC mutations have dominant effects on proliferation, spindle checkpoint control, survival and chromosome stability. *J Cell Sci* **117:** 6339–6353.

Tyagi S, Chabes AL, Wysocka J, Herr W. 2007. E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* **27:** 107–119.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854–858.

Vogel JL, Kristie TM. 2000. The novel coactivator C1 (HCF) coordinates multiprotein enhancer formation and mediates transcription activation by GABP. *EMBO J* **19:** 683–690.

Wilson AC, LaMarco K, Peterson MG, Herr W. 1993. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74:** 115–125.

Wysocka J, Herr W. 2003. The herpes simplex virus VP16-induced complex: The makings of a regulatory switch. *Trends Biochem Sci* **28:** 294–304.

Wysocka J, Reilly PT, Herr W. 2001. Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol Cell Biol* **21:** 3820–3829.

Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W. 2003. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* **17:** 896–911.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD. 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121:** 859–872.

Yang ZF, Mott S, Rosmarin AG. 2007. The Ets transcription factor GABP is required for cell-cycle progression. *Nat Cell Biol* **9:** 339–346.

Yu H, Mashtalir N, Daou S, Hammond-Martel I, Ross J, Sui G, Hart GW, Rauscher FJ III, Drobetsky E, Milot E, et al. 2010. The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* **30:** 5071–5085.

Yuan CC, Zhao X, Florens L, Swanson SK, Washburn MP, Hernandez N. 2007. CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription. *Mol Cell Biol* **27:** 8729–8738.

Zargar Z, Tyagi S. 2012. Role of Host Cell Factor-1 in cell cycle regulation. *Transcription* **3:** 187–192.

Received September 28, 2012; accepted in revised form March 27, 2013.