Global mapping of the regulatory interactions of histone residues

CrossMark

Inkyung Jung [a], Jungmin Seo [b], Heun-Sik Lee [b], Lawrence W. Stanton [c], Dongsup Kim [a], Jung Kyoon Choi [a,c,*]

[a] Department of Bio and Brain Engineering, KAIST, Daejeon 34141, Republic of Korea
[b] Research Institute of Bioinformatics, Omicsis, Inc., BVC, KRIBB, Daejeon 305-333, Republic of Korea
[c] Genome Institute of Singapore, Singapore 138672, Republic of Singapore

ABSTRACT

Histone residues can serve as platforms for specific regulatory function. Here we constructed a map of regulatory associations between histone residues and a wide spectrum of chromatin regulation factors based on gene expression changes by histone point mutations in *Saccharomyces cerevisiae*. Detailed analyses of this map revealed novel associations. Regarding the modulation of H3K4 and K36 methylation by Set1, Set2, or Jhd2, we proposed a role for H4K91 acetylation in early Pol II elongation, and for H4K16 deacetylation in late elongation and crosstalk with H3K4 demethylation for gene silencing. The association of H3K56 with nucleosome positioning suggested that this lysine residue and its acetylation might contribute to nucleosome mobility for transcription activation. Further insights into chromatin regulation are expected from this approach.

© 2015 The Authors. Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

As the structural units of chromatin, the core histones (H3, H4, H2A, and H2B) play a critical role in the epigenetic regulation of DNA. The biological significance and universality of their function are reflected by the remarkably high evolutionary conservation of the protein sequences from yeast to humans. The amino acid residues of histones can be partitioned into four major geographical domains: buried, disk (protein surface that does not contact DNA), lateral (protein surface that contacts DNA), and tail (protruding unstructured region).

To explore the functional role of each residue across the different histone domains, systematic mutant libraries were generated

and screened for phenotypic changes in yeast [1,2]. The library of synthetic histone H3 and H4 mutants [1] was created by substituting each non-alanine residue with alanine to avoid deletion effects, while mutating alanine to serine. The mutant strains were tested for phenotypic changes in viability, transcriptional silencing, transcriptional elongation, response to DNA damage, response to microtubule disruption, and response to temperature shock. A database of the phenotypes of histone mutants collected from different studies, named HistoneHits, has been developed [3].

This library also included replacements of modifiable residues with amino acids mimicking modified and unmodified states. Post-translational modification (PTM) of histones includes acetylation, methylation, ubiquitination, phosphorylation, and sumoylation of lysine, arginine, serine, and threonine [4]. Although most of the modifications are observed in the histone tail, the globular domain also contains modifiable amino acids, including acetylated H3K56 and methylated H3K79 [5–9]. The histone code hypothesis [10] proposes that PTMs, alone or in combination, serve as selective binding platforms for regulatory proteins such as chromatin modifiers (CMs) and transcription factors (TFs).

However, modifiable residues cannot account for all histone functions. The modification of nucleosomes is preceded by the formation of nucleosomes through the interactions of histones with wrapping DNA. The nucleosome code hypothesis [11,12] predicts that DNA dictates its own physical packaging into the chromatin

structure by modulating its binding with histone proteins. The importance of the DNA-histone interaction was illustrated by a viability test of the histone mutants. In the two mutation studies [1,2], most of the lethal substitution mutations were commonly mapped to the nucleosome lateral surface near the dyad axis or at the DNA entry/exit site, where histones make contact with DNA. The H2A and H2B essential residues were also located on the surface of nucleosomes [2].

Transcriptional regulation should be the primary mechanism by which each residue of histones contributes to phenotypic consequences. How nucleosome assembly and histone PTMs can influence gene transcription has been studied extensively, thanks to genomic technologies such as DNA microarray and next-generation sequencing. Various PTMs and nucleosome patterns near gene promoters have been intensively examined at the genome-wide level [13–20]. For example, the canonical promoter contains a nucleosome-free region (NFR) upstream of the transcription start site (TSS) and a +1 nucleosome downstream of the TSS. Several PTMs, including H3 or H4 acetylation and H3K4 methylation, are associated with active gene transcription.

In this work, we employed genomics approaches to study the function of individual residues of histone H3 and H4 by utilizing a public histone mutant library [1]. Gene expression microarrays were employed to measure the influence of each residue on transcription by comparing gene expression patterns in the histone mutants and wild type. We selected 123 mutants with the highest phenotypic effects for gene expression profiling. To determine whether certain mutants of high transcriptional importance alter nucleosome positioning, the mutated nucleosomes were purified and profiled by deep sequencing.

2. Materials and methods

2.1. Selection of histone mutants

Our mutant selection procedure was based on the HistoneHits database [3]. In essence, we used the alanine substitution mutants of H3 and H4, while discarding lethal mutants [1] and including the strains where the mutation was targeted to a known modifiable residue. The information of modifiability was obtained from the HistoneHits database except for H3K37, whose methylation was recently reported [44]. The degree of phenotypic change for each mutant was obtained from the same database, which collected phenotype scores observed in multiple experiments that belong to one of nine different categories. The nine categories and their readouts were: ribosomal silencing (growth on plates), telomeric silencing (colony color as gain of telomeric silencing), mating efficiency (enzyme assay and growth on plates), growth rate (growth on plates), DNA damage (growth on plates), Spt-phenotype (growth on plates), transcription elongation defect (growth on plates), K56 hyperacetylation suppression (growth on plates), and mating cassette silencing (growth on plates). The score for each experiment was represented as an integer ranging from −2 to 2 with a high absolute value indicating a higher degree of phenotypic change. The average for each category was calculated and then the average of the nine scores was obtained as the final measure for the degree of phenotype changes by each mutation. We first selected the mutants with the final phenotype score greater than 1 and then tried to filter out adjacent mutants that had similar response across the nine experimental conditions. To do so, we first identified the two adjacent mutants that had the same direction of response (the same sign of the phenotype score) in more than seven out of the nine conditions and then removed the one having the lower final phenotype score. H3Q5A was later added because H3Q5 was reported to crosstalk with H3K4me3 [2].

To investigate the effects of different acetylation states, we included H3K56R, H3K56Q, H4K16R, and H4K16Q. Arginine (R) substitution mimics unacetylated lysine and glutamine (Q) mimics acetylated lysine. Overall, a total of 123 mutants were profiled and compared against four replicates of the wild type by microarray experiments, totaling 127 microarray datasets.

2.2. Yeast strains and cell culture

We obtained the yeast histone mutant library from Open Biosystems (catalog number: YSC5106, none essential histone H3 & H4 mutant collection). The 2 ml glycerol stocks of selected mutants and the wild types were cultured in SD-ura medium for 22 h at 25 °C. After streaking on YPD plate at 25 °C, single colony was cultured in 2 ml SD-ura medium for 22 h at 25 °C, 500 μl out of which was subcultured in 10 ml SD-ura medium for 22 h at 25 °C.

2.3. Microarray hybridization

RNA was prepared using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen). After DNase treatment (TaKaRa Recombinant DNaseI), the first-strand complementary DNA was synthesized from 1 μg total RNA at 42 °C for 2 h, which was followed by second-strand cDNA synthesis at 16 °C for 2 h. The resulting double-stranded cDNA was purified. Fluorescence-labeled RNA was generated by carrying out an in vitro transcription reaction (double-stranded cDNA in nuclease-free water 16 μl, T7 rNTP mix 16 μl, T7 10× Reaction Buffer 4 μl, and T7 Enzyme Mix) at 37 °C for 16 h. The labeled RNA was subsequently purified and chemically fragmented at 70 °C for 15 min in fragmentation buffer (Ambion fragmentation reagent). The fragmented, labeled complementary RNA was hybridized to NimbleGen 12 × 135 K oligonucleotide microarrays at 42 °C for 16–20 h according to the provided instructions. The spotted microarrays of 60-mer oligonucleotide probes that represent 5777 yeast ORFs contained three to eight probes per gene with three replicates for each probe. The microarrays were washed in three consecutive steps by using the provided kit and the readouts were scanned using NimbleScan 2.5.26.

2.4. Microarray data processing

The raw microarray data was normalized by using VSN (Variance Stabilization and Normalization) algorithm [45]. This method utilized variance stabilizing transformation based on the parametric form $h(x) = \mathrm{arcsinh}(a + bx)$, which is derived from a model of the variance-versus-mean dependence for microarray intensity data. For large intensities, h coincides with the log transformation, and Δh with the log ratio. Following the preprocessing by VSN, microarray batch effects were removed by means of the Combat algorithm [46]. We applied the non-parametric empirical Bayes frameworks to the VSN-normalized data. After removing batch effects, we obtained the relative gene expression changes as the log2 ratios between the mutants and wild type. The expression levels for the wild type were calculated as the mean of the replicates. A very high reproducibility ($R = 0.98$) was observed between the wild-type replicates. The genes that were up-regulated or down-regulated >1.5 fold were selected for each histone mutation. We performed gene-set enrichment analysis for these changed genes. To obtain a sufficient statistical power, we only considered gene cohorts that included more than 50 genes, resulting in 91 up-regulated gene cohorts and 71 down-regulated gene cohorts. Afterward, we conducted gene sent enrichment analysis for Gene Ontology (GO) terms by using DAVID (https://david.ncifcrf.gov/) and then performed hierarchical clustering for $-\log_{10}$

(FDR-corrected P value). We only considered GO terms with an FDR-corrected P value < 0.05. In the heatmap, each column indicates up- or down-regulated gene cohort and each row indicates GO biological function. Any GO terms detected less than 5 times were excluded from downstream analyses.

2.5. Calculation of association scores

The genes that were up-regulated or down-regulated >1.5 fold were selected for each histone mutation, resulting in three cohorts of genes: activated, repressed, and non-responsive. As detailed in Supplementary Table 2, the datasets for CMs, TFs, PTMs, and nucleosomes were collected from different sources. For ChIP-chip- or ChIP-seq-based location profiles (except TF binding), binding levels at the promoter (−500–100 bp of the transcription start site) and those at the coding region were separately treated. Only promoter binding levels were used for the TF dataset. For the deletion datasets (ΔCM and ΔTF), the absolute change of gene expression was considered to reflect the binding affinity of the deleted regulator to the given gene. For each histone residue, the difference in CM/TF binding affinity or PTM/nucleosome density between the activated and non-responsive cohort and that between the repressed and non-responsive cohort were calculated based on the Kolmogorov–Smirnov (KS) statistic (see Fig. 2B for a schematic representation). The KS test tries to determine if two distributions differ significantly (e.g., green curve vs. blue curve or red curve vs. blue curve in Fig. 2B). The KS-test has the advantage of making no assumption about the distribution of data. This non-parametric statistic proved to be useful for extracting comparable summary statistics across heterogeneous datasets from different sources [47,48]. For the association score, $-\log_{10}$ (P value of the KS test) was taken. If the association (KS) score for the activated cohort (KS_{rep} in Fig. 1B) is greater than that for the repressed cohort (KS_{act} in Fig. 1B), this means the mutation (or the original residue) tends to activate (or repress) the genes via interaction with the given regulation factor, namely, CM, TF, PTM, or nucleosome. Therefore, we assigned a negative association (KS) score for this association ($KS_{rep} > KS_{act}$ in Fig. 1B) to indicate that the original residue (prior to the mutation) represses gene transcription in concert with the given factor. For the constitutive acetylation mutation (glutamine substitution), a negative score (gene activation) was interpreted to mean that the acquired mutation (not the original residue) activates the given cohort of genes. Concerning the HR-CM association map and the HR-TF association map used in the main text, we collapsed the columns (representing the regulation factors) by taking the greater absolute association (KS) score between the promoter and coding-region profiles. For identified modules, we manually checked if the CM profiles were from the promoter or coding region.

2.6. HR-CM heatmap and HR interaction network

HR-CM associations were represented as a heatmap of the association scores obtained for the CM and ΔCM dataset. HR-CM association modules were identified based on hierarchical clustering of the association scores. We ran Cluster and TreeView (http://www.eisenlab.org/eisen/?page_id=42) with default options after column-wise and row-wise centering and normalization. Red indicates a positive KS score ($KS_{act} > KS_{rep}$) and green represents a negative KS score ($KS_{act} > KS_{rep}$). We also constructed an HR interaction network based on correlation coefficients of the HR-CM association scores between HRs. We defined an interaction between HRs as having a correlation >0.7. As a result, a total of 151 interactions among 53 HRs were identified. We identified several subnetwork modules by using MCODE (http://baderlab.org/Software/MCODE).

2.7. Mononucleosome purification and sequencing

The MNase-mediated purification of mononucleosomes was carried out as previously described [12,49–51]. Yeast cells were grown in YPD to an A_{600} OD of 1.5 at 200 rpm. Cells were fixed in 1% formaldehyde, and incubated for 30 minutes at room temperature, shaking, at 200 rpm. 2.5 M glycine was added to a final concentration of 125 mM to quench the formaldehyde. Cell pellet was resuspended in 10 ml β-ME buffer (20 mM EDTA, 0.7 M β-mercaptoethanol). After 30 min of incubation at 30 °C, cell pellets were precipitated by 5 min of centrifugation at 3000 rpm. Spheroplast preparation was conducted at 30 °C for 30 minutes in 10 ml lyticase lysis buffer (1 M sorbitol. 50 mM Tris, pH 7.8; 5 mM β-ME) using 10,000 units of Lyticase (Sigma). After digesting cell walls, spheroplasts were spun down, washed twice with 1 M sorbitol. DNA digestion was conducted at 37 °C for 20 min in 1.6 ml of digestion buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris–HCl at pH 7.5, 5 mM MgCl2, 1 mM CaCl2, 1 mM β-mercaptoethanol, 0.5 mM spermidine and 0.075% v/v NP40). 400 μL aliquots of spheroplasts were digested with 3.5 units of MNase (Sigma). Reaction was stopped by the addition of MNase stop buffer (5% SDS, 250 mM EDTA). Samples were treated with 5 μL of proteinase K (20 mg/mL) at 37 °C for 5 h, then phenol:chloroform extracted, ethanol precipitated, RNase A treated, and then ethanol precipitated. The digested DNA was purified and separated on a 1.5% agarose gel, and the mononucleosomal DNA (∼147 bp) was cut out of the gel. The mononucleosomal DNA fragments were sequenced by Illumina Genome Analyzer, subjected to 36 cycles of single-read sequencing.

2.8. Sequencing data processing

Sequence reads were mapped to the yeast genome (NCBI June 2008) by means of the Illumina sequencing pipeline. The sequencing tags were extended to the average size of fragments in the library (150 bp) and the number of overlapping sequence reads was obtained at 200-bp intervals across the genome. The ratio of (target read count/200 bp)/(total read count/genome size) was obtained and log2 transformed [52]. This normalized read count was used as an estimate of nucleosome level at the given genomic locus.

2.9. Computation of nucleosome fuzziness

Nucleosome fuzziness is a measurement of how delocalized or spread out a nucleosome position is. We used Genetrack software [53], which first identifies nucleosome positions and then calculates the standard deviation of all read coordinates that contributes to identify mononucleosome locations [54]. The standard deviation served as the fuzziness measure for each positioned nucleosome.

3. Results

We profiled global gene expression changes in the strains of the comprehensive H3 and H4 mutant library [1], where all available HRs were substituted by alanine (A) to avoid unwanted deletion effects. We selected all known modifiable HRs and the HRs with significant phenotypic changes [3] (see Methods), resulting in a gene expression compendium of 123 histone mutations (Supplementary Table S1). To first characterize cis-level features of HR regulation, we investigated overall patterns of gene expression changes and enriched biological functions of the genes affected by each HR mutation. First, we observed varying numbers of up-regulated and down-regulated genes across different histone mutations (Fig. 1A). The strongest mutation was H4L97A. In terms of structural classification, mutations on modifiable histone
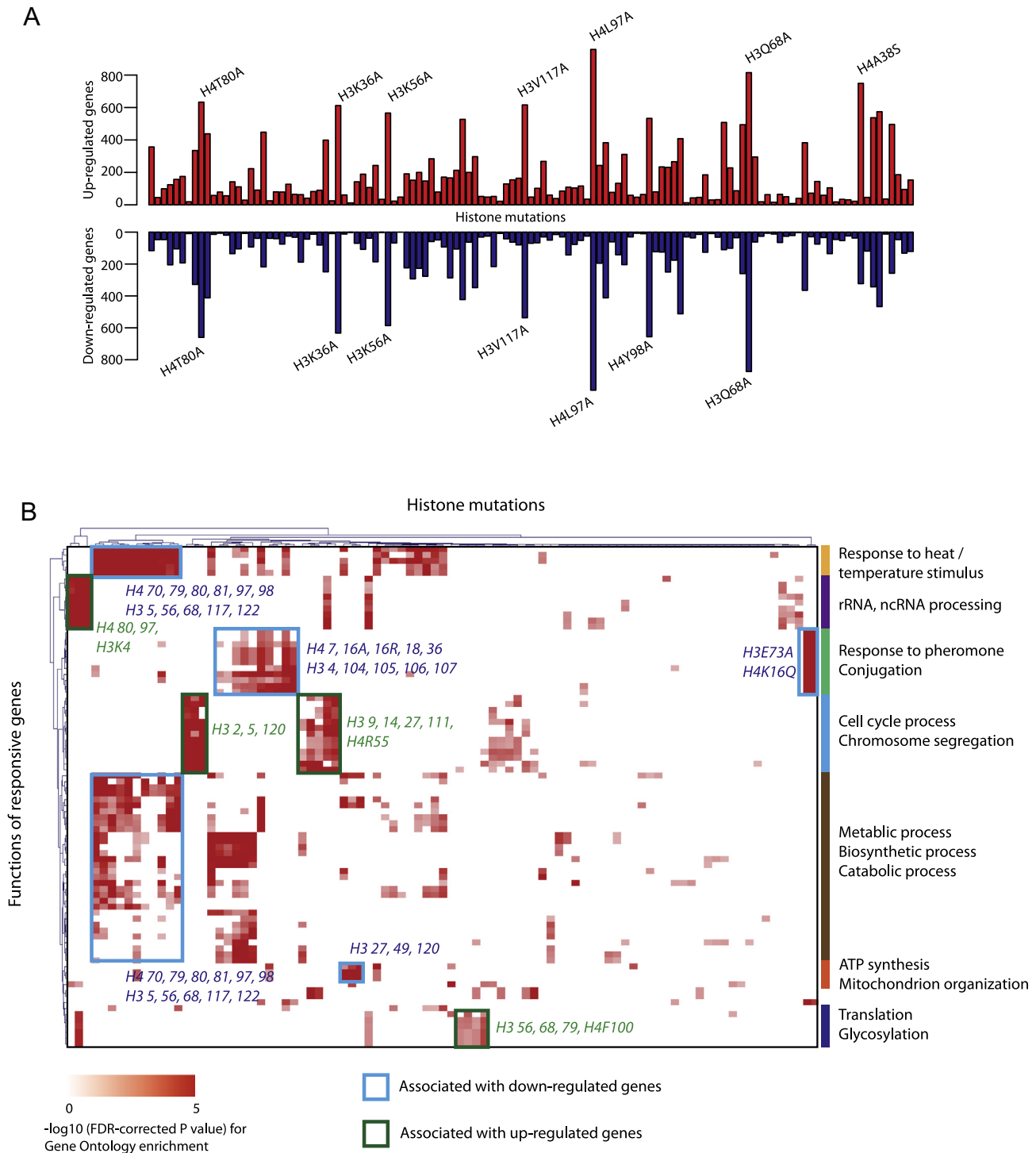
Fig. 1. Characterization of genes affected by histone mutations. (A) The number of activated and repressed genes in response to each histone mutation. (B) Functional map of genes affected by histone mutations. The statistical significance of functional enrichment and the direction of gene expression changes were color-coded as depicted below the map. Enriched Gene Ontology terms are described on the right side of the map. For example, genes down-regulated by H3E73A and H4K16Q are enriched for response to pheromone and conjugation.

residues tended to have stronger effects than other histone residues (Supplementary Fig. S1, left). The mutations on the DNA-contacting surface regions also exhibit stronger mutation effect than mutations on other geographical domains (Supplementary Fig. S1, right).

We performed gene-set enrichment analysis for the genes affected by each HR (Methods) to identify several modules associating mutated HRs and biological functions of the affected genes (Fig. 1B). For example, down-regulated genes responding to histone H4 7/16/18/36 and H3E73 mutations are strongly enriched by mating-related biological processes. Interestingly, H4H18A, H4K16Q, and H3E73A mutations were reported to phenotypically affect mating efficiency [3]. Although it is intriguing that particular HRs can be involved in the regulation of genes with particular
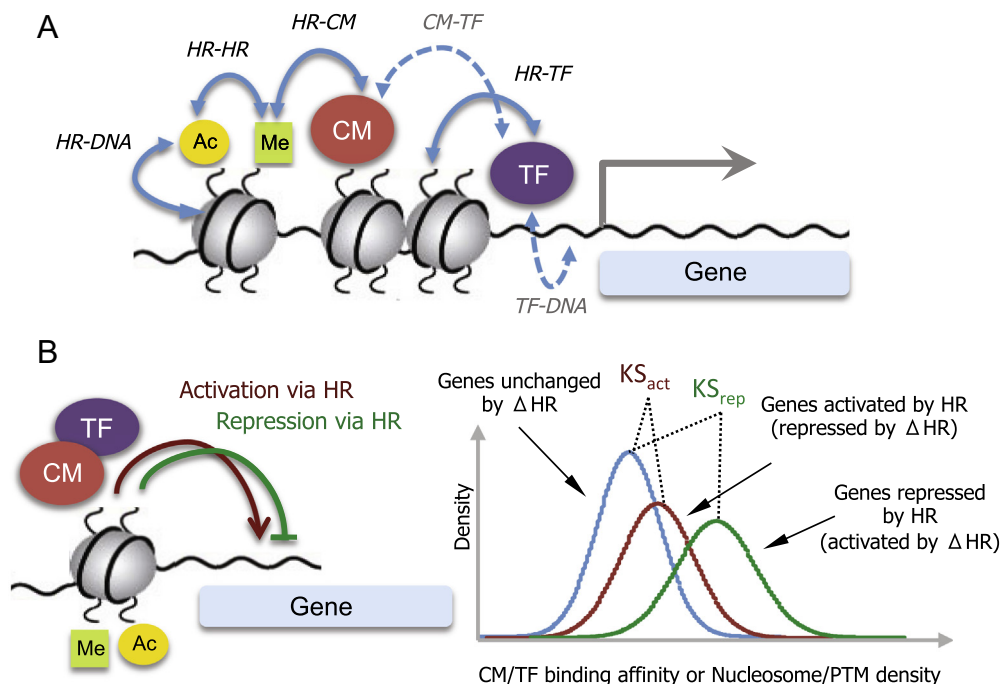
Fig. 2. Histones as the platform for gene regulation (A) and our data analysis scheme (B). (A) Interactions of histone residues (HRs) with DNA (HR-DNA), with chromatin modifiers (HR-CM), and with transcription factors (HR-TF), and crosstalk between histone residues (HR-HR) play a critical role in transcription regulation. (B) For each HR, the difference in CM/TF binding affinity or PTM/nucleosome density between the changed genes and unchanged genes by an HR mutation is measured by the KS statistic. In this example, the KS score for the activated cohort (KS_{rep}) is greater than that for the repressed cohort (KS_{act}), meaning that the original HR tends to interact with the regulatory factor (CM, TF, PTM, or nucleosome) for gene repression. Therefore, we take the negative of KS_{rep} for this association to indicate repression. When KS_{act} is greater than KS_{rep}, the raw KS_{act} score (positive) is used.

function, this type of specific association is unlikely achieved by HRs themselves because histones are globally distributed across the whole genome. This highlights a role for trans-regulatory factors that bind HRs or their modification in the cis-regulatory regions of particular genes. Therefore, we proceeded to investigate the regulatory relationships among HRs and trans-regulators (Fig. 2A).

In our data analysis scheme (Fig. 2B), a positive association score was assigned when the genes associated with a regulation factor (e.g., bound by a CM) are repressed by an HR loss-of-function mutation, an indication that this HR normally activates the genes in correspondence with the regulation factor. Exceptionally, a negative score for gain-of-function mutations such as H4K16Q and H3K56Q (constitutive acetylation) should be interpreted as an indication that the mutation, not the original residue, activates gene expression. When interrogated against the CM, TF, PTM, and NUC profiles (Fig. 3A and Supplementary Table S2), the HR compendium exhibited higher associations with CMs than with TFs, which was predictable, and with nucleosome occupancy than with PTM density (Fig. 3B). It is likely that the interactions of the modified residues with CMs are more important than the direct influence of the PTMs themselves. On the other hand, histone-DNA interaction emerged as one of the key aspects to consider. These findings encouraged us to focus on the interactions of the modifiable HRs with CMs and those of the surface HRs with DNA.

From the HR-CM association map (Fig. 3C), we identified three modules that seemed related to the function of H3K4me3 and H3K36me3. A summary of our analyses of this map is provided in Supplementary Table S3. Module I integrated separate findings on the interplay of Set1 with Gcn5 and Esa1 on the 5′ portion of coding regions [21–23]. Notably, the CMs in this module were identified based on their open reading frame (ORF)-binding profiles. The four identified HRs turned out to have the highest associations with the PTM levels (Fig. 3B). The two modifiable residues

among them, H4K91 and H4R92, lie in the interface between histone H3/H4 tetramers and H2A/H2B dimers [24]. Notably, H4K91 acetylation was found to destabilize the histone octamer by weakening H3/H4 – H2A/H2B binding [25], which is exactly in line with the function of the FACT complex, that is, nucleosome destabilization by H2A/H2B dimer removal during RNA polymerase II (Pol II) passage [26]. The SWR1 complex performs ATP-driven exchange of H2AZ variant for H2A [27,28]. Surprisingly, H4K91 and H4R92 were the top two HRs most strongly associated with ORF H2AZ occupancy (Fig. 4A). The function of H4R92 methylation is currently unknown. Our data propose a possible crosstalk between these two adjacent modification sites. The association of Rad6 suggests a role for H2B ubiquitylation in H2B dissociation during Pol II elongation (described later).
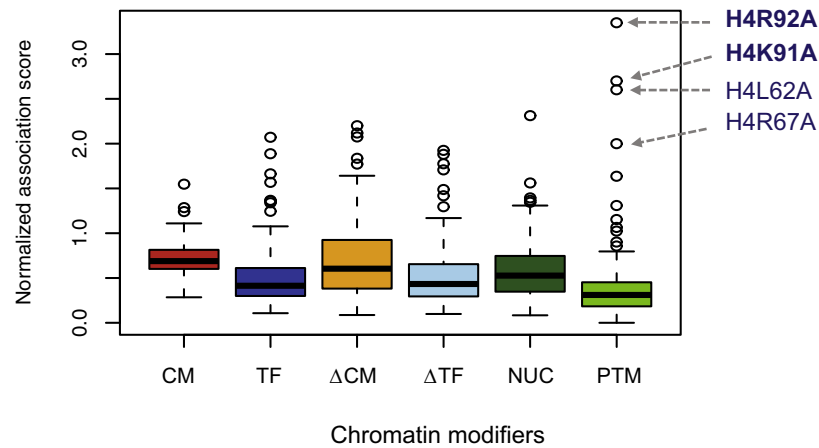
We examined which PTMs are involved in Module I function. Although various histone acetylations as well as H3K4me3 were found at Module I genes (Fig. 4B), our HR mutation experiments showed that the loss of the H3K27 residue only had transcriptional effects (Fig. 3C). The other PTMs may play a secondary role. Especially, even with the involvement of Set1 and H3K4me3 in this module, the mutation of this residue (H3K4A) did not change the expression of the relevant genes. This reinforces the hypothesis that H3K4me3 serves only as a molecular memory of recent Pol II elongation initiation, which is based on the observation that H3K4me3 persists for a considerable time even after Set1 dissociation and transcriptional inactivation [23,29].

Apparently, H4K16Ac is not involved in the Set1 module (Fig. 4B) but in Module II and III. Module II is represented by JmjC domain-containing enzymes, namely the H3K4 demethylase Jhd2 and two other histone demethylases, Jhd1 and Ecm5 [30]. The crosstalk between H3K4 demethylation by Lsd1 and H4K16 deacetylation by Sir2 was recently discovered in *Drosophila* [31]. It seems that this crosstalk is present in yeast and mediated by Jhd2 and Sir2-related histone deacetylases (HDACs), namely Hst1
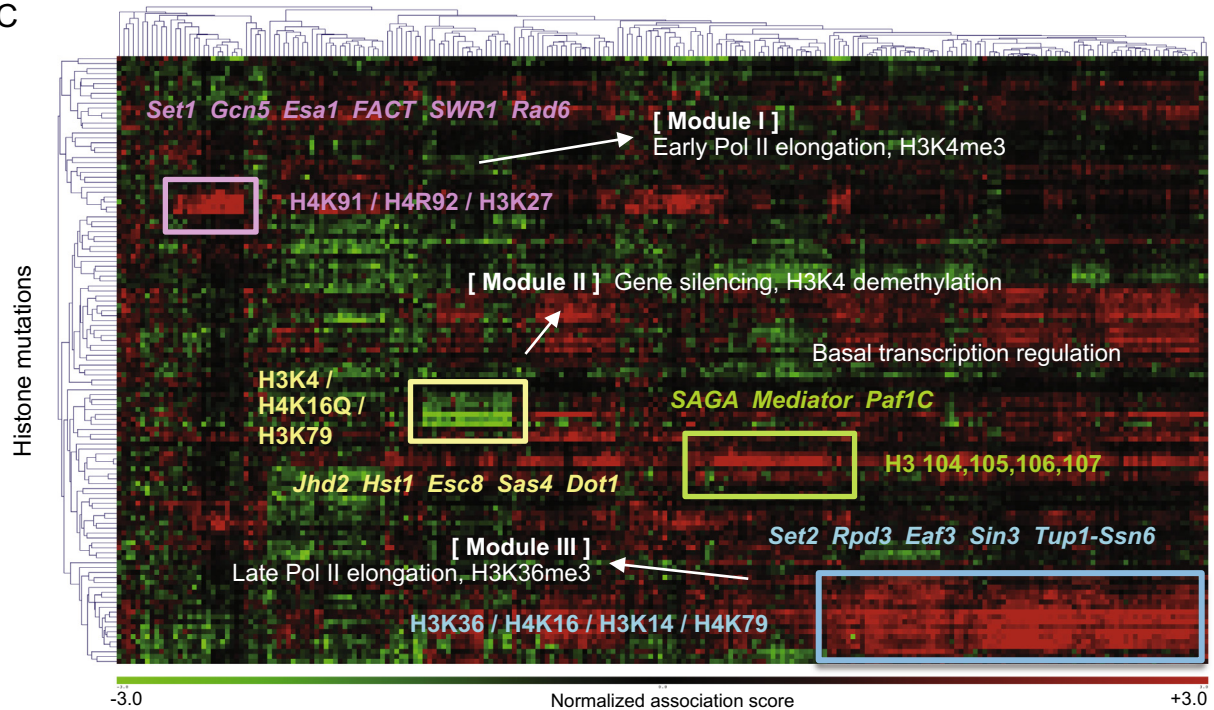
Fig. 3. Overview of HR associations. (A) The genome-wide datasets for chromatin regulation factors or elements that were used in this study. More details are provided in Supplementary Table S2. (B) The average association score for each HR was obtained across all the factors or elements of each dataset. The association scores were scaled to unit standard deviation beforehand for comparison between different datasets. (C) HR-CM associations represented as a heatmap of the association scores obtained for the CM and ΔCM dataset. HR-CM association modules could be identified based on hierarchical clustering of the association scores. For example, Module I, II, and III were discovered with a focus on the function of H3K4me3 and H3K36me3. Spt16 and Pob3 are pertaining to FACT, and Swr1 and Swc2 to SWR1. The individual components of SAGA, Mediator, and Paf1C are detailed in Supplementary Fig. S3.
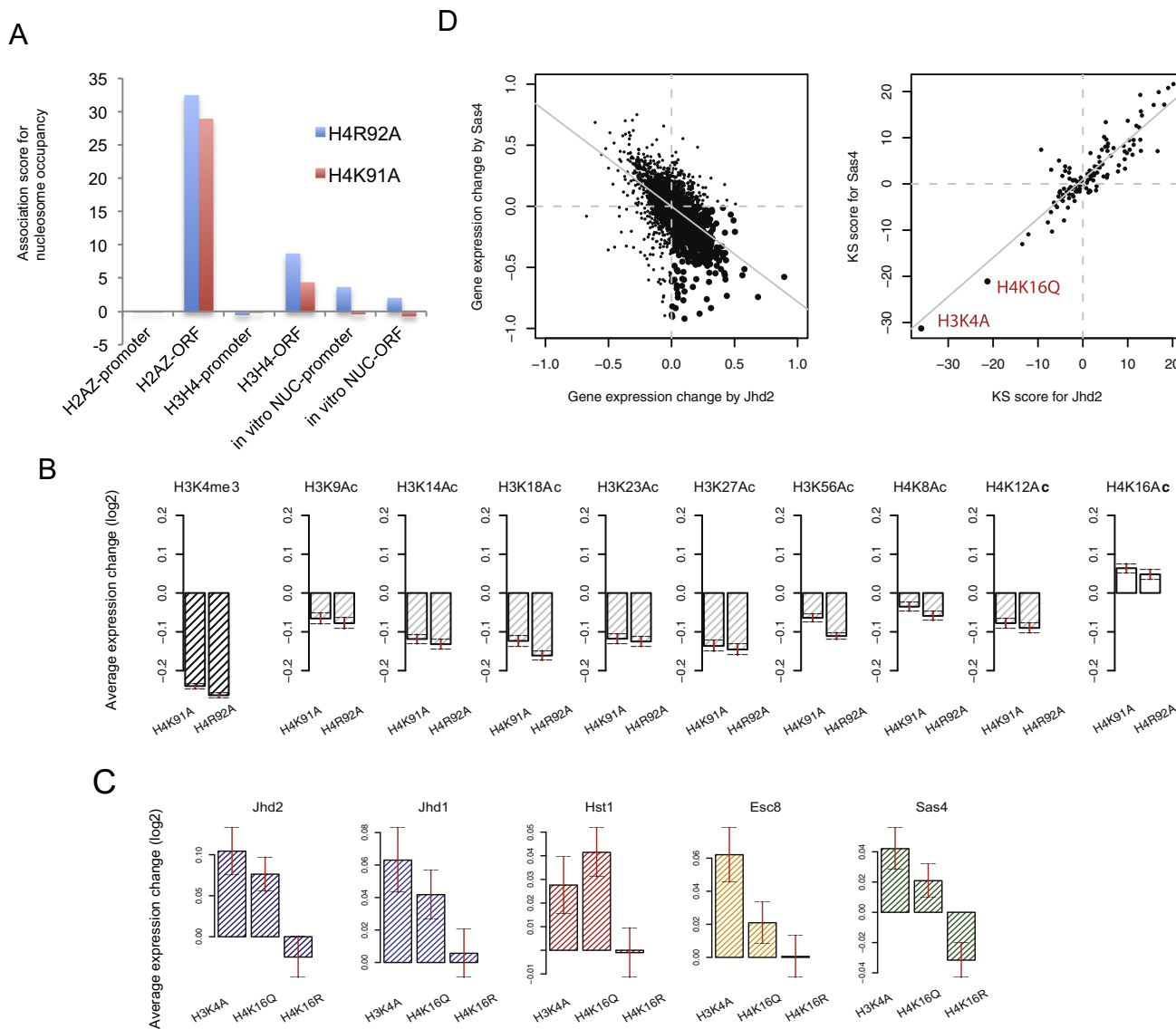
Fig. 4. Analysis of HR-CM associations in Module I and II. (A) HR-NUC associations between H4K91/H4R92 with promoter H2AZ, ORF H2AZ, promoter H3/H4, ORF H3/H4, promoter in vitro nucleosome, and ORF in vitro nucleosome occupancy (from left to right). (B) The expression changes by the mutation of H4K91 and H4K92 from Module I were negatively correlated with the PTM levels measured for coding regions, except for H4K16Ac. (C) The prevention of gene deactivation by the H3K4 mutation and H4K16 acetylation mutation in Module II. (D) Gene expression changes by ΔJhd2 and ΔSas4 (left) and association scores of Jhd2 and Sas4 with HRs (right).

(Sir2 homolog) and Esc8 (Sir2-interacting protein). In contrast to the case for the Set1 module, the H3K4 deletion caused transcriptional changes here (Fig. 3). H4K16A did not cause transcriptional changes, however, indicating that the recognition of the H4K16 residue is not required in this process. H4K16Q (acetylated lysine) as well as H3K4A prevented deactivation by Jhd1 and the HDACs (Fig. 4C). Taken together, H3K4me3 recognition and demethylation by Jhd2 may precede H4K16 deacetylation, which results in actual gene repression. Interestingly, Sas4, which acetylates H4K16, was also associated in this module (Fig. 3) showing the same patterns except with H4K16R (unacetylated lysine) (Fig. 4C). The involvement of Sas4 in yeast chromatin silencing was previously suggested [32]. These findings can be explained by a recent model proposed in human cells, where inactive genes primed by H3K4 methylation are repressed by HDACs that remove acetyl groups added by transient binding of histone acetyltransferases (HATs) [33]. Constitutive acetylation modelled by H4K16Q prevents repression by HDACs and thereby this mutation activates gene expression as indicated by the green color in the heatmap (Fig. 3C).

Our data suggests the key role of Jhd2 and Sas4 in this mechanism. The repressor Jhd2 and the activator Sas4 tend to affect the same genes in the opposite direction with similar histone-interaction mechanisms (left panel of Fig. 4D), especially mediating the crosstalk between H3K4 methylation and H4K16 acetylation (right panel of Fig. 4D).

Module III delineates the relatively well-characterized interaction of Set2 with Rpd3 and Eaf3 as a mechanism to antagonize inappropriate histone acetylation during late Pol II elongation and suppress spurious intragenic transcription [34–36]. A possible link between Set2 and Sin3 was suggested in human [37]. The counteracting crosstalk between H3K36me3 and H4K16Ac observed in *Drosophila* [38] may be also present in yeast. H4K16R and H4K16A were highly correlated with the H3K36 mutation in terms of association with Set2 and the HDACs (Fig. 3C). Therefore, it seems that H4K16 deacetylation plays a dual role depending on its crosstalk: with H3K4me3 in gene silencing and with H3K36me3 in Pol II elongation. The interaction of the Tup1-Ssn6 complex with hypoacetylated histones during Pol II elongation is suggested (Fig. 3C).

These three HR clusters were consistently found in the HR-TF association map (Supplementary Fig. S2), providing the capability to predict CM-HR-TF interactions. Interestingly, the HRs in Module III were positively associated with the Set2-associated HDACs but negatively with a host of TFs, a pattern that nicely explains the different mechanisms by which histone deacetylation affects transcription initiation and elongation. In other words, histone deacetylation inhibits transcription initiation but promotes transcription elongation at the 3′ end of the coding region. Another HR-CM module (green box; Fig. 3C), which was not found in the HR-TF map, turned out to be enriched for multiprotein complexes that associate with Pol II and the general transcription factors (Supplementary Fig. S3). A series of adjacent non-modifiable HRs on the H3 globular domain were involved. Presumably, they are required to modulate chromatin structure for the action of the Pol II initiation complex.
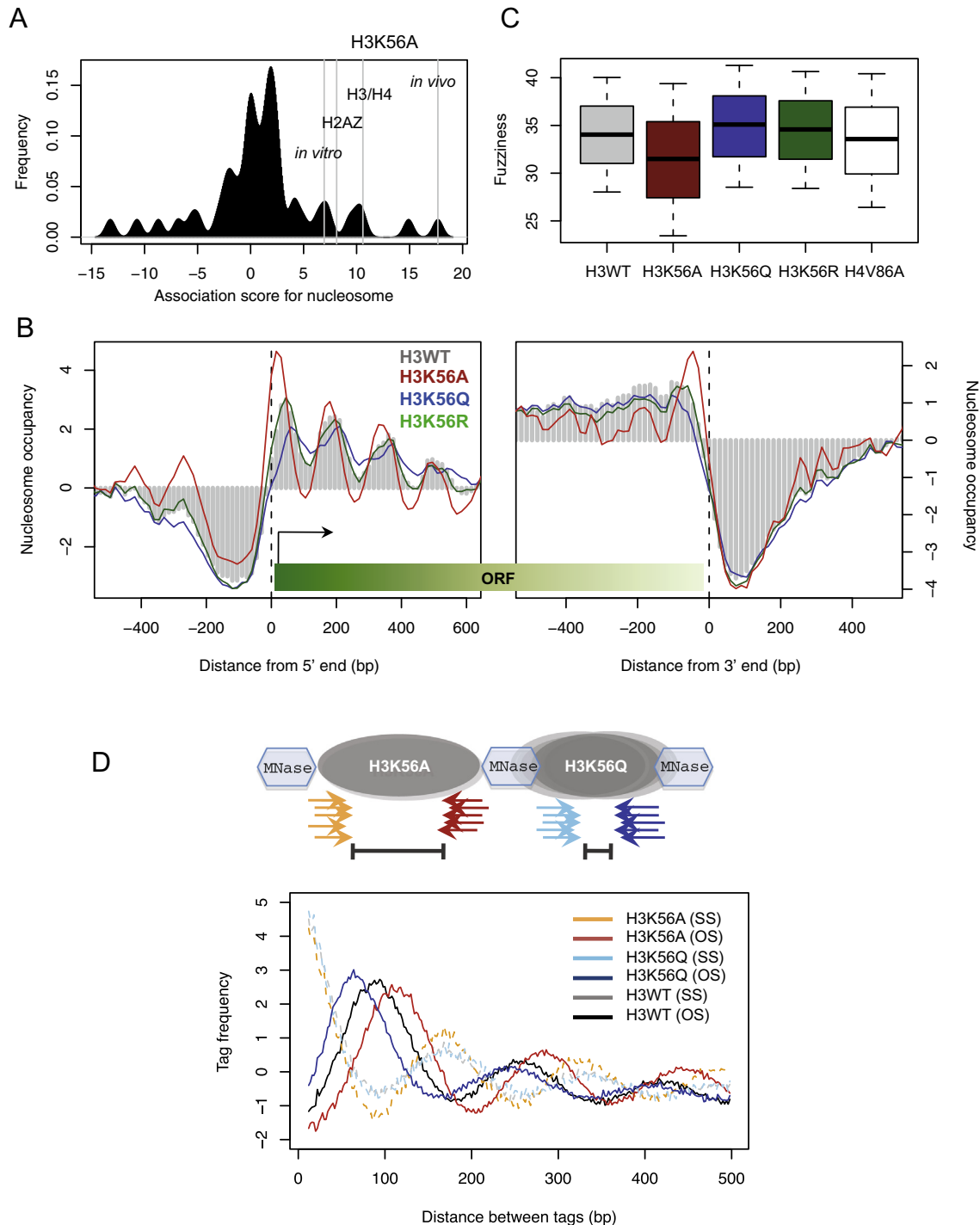

Fig. 5. Nucleosomal changes by H3K56 mutations. (A) Associations of the modifiable HRs on the DNA-contacting surface with nucleosome positioning. The KS scores for H3K56A and the NUC datasets were depicted by the vertical lines. (B) Nucleosome positioning surrounding the coding region for the wild type and H3K56 mutants (C) The fuzziness of the wild type and mutant nucleosomes. H4V86A was arbitrarily included for comparison. (D) Distances between sequencing tags on the same strand (SS) and between those on the opposite strand (OS). The illustration above the panel shows that the average distance to the OS tags indicates the length of the MNase-inaccessible region.

Regarding the regulatory associations with nucleosome occupancy, the modifiable HRs on the DNA-contacting surface of histones stood out (Supplementary Fig. S4) with the maximum positive score found for H3K56A and one of the NUC datasets (Fig. 5A). Other NUC measures also produced a positive score, the interpretation being that H3K56 contributes to transcription activation probably by promoting nucleosome eviction or relocation. Residing at the DNA entry-exit points, H3K56 is in a position to make contact with both ends of the wrapping DNA. To understand the detailed mechanism, we sequenced the genome-wide positions of mononucleosomes in the wild type and three H3K56 mutants (see Methods).

Indeed, there was a striking change in nucleosome positioning, including the increased occupancy of the −1 and +1 nucleosomes and the 3′-end nucleosome, and a sharpened nucleosome phasing with higher peaks and deeper troughs (red curve; Fig. 5B), all reflecting more stable nucleosome positioning. In fact, H3K56A increased the number of positioned nucleosomes by >2000 and reduced the fuzziness [39] of the nucleosomes (Fig. 5C). H3K56Q acted in the opposite direction, resulting in the disrupted nucleosome phasing and reduced occupancy of the −1, +1, and 3′-end nucleosomes (blue curve; Fig. 5B). H3K56Ac was shown in vitro to increases the transient unwrapping (breathing) of the ends of the DNA [40].

DNA breathing will cause further invasions of MNase into the histone-protected region. We calculated the distance between the tags on the same and opposite strand. The distance to the closest peak on the opposite strand (OS lines; Fig. 5D) corresponds to the length of the MNase-inaccessible region. The change of this distance was accurately reflecting the expected change of DNA breathing in the H3K56A and H3K56Q nucleosome. The distance to the closest peak on the same strand (SS lines; Fig. 5D), which reflects nucleosome spacing, did not change.

One thing to note is that H3K56R (unacetylated lysine) did not induce considerable changes. This proposes that the lysine residue itself is structurally important for nucleosome mobility; H3K56 itself makes a water-mediated contact with the DNA phosphate backbone at the DNA entry-exit gates [41]. It is possible that H3K56Ac is not constitutively abundant in vivo so its loss does not trigger global changes. Intriguingly, H3K56Q had the maximum negative association with H2AZ (KS < −20), but no significant association with H3/H4. H3K56Ac may cause transcriptional changes only in concert with H2AZ through the actions of CMs, in contrast to the constitutive function of the residue itself.

4. Discussion

In this work, we developed a systematic approach to investigating the regulatory function of individual histone residues. Although the role of histone modification has been the focus of interest, our results emphasize the associations of HRs with CMs and NUC. For example, associations between NUC and HRs in the lateral surface highlight the importance of HR-NUC interactions in transcription control. Specifically, H3K56 was found to play a key role in modulating nucleosome relocation. The ability of a single residue to globally affect nucleosome stability and occupancy is surprising. Given the high impact of this residue on gene expression, we propose that nucleosome fuzziness and nucleosome positioning are key factors to be considered in the study of gene transcription [18]. Research into how K56 and its acetylation function in the dynamic control of nucleosome positioning by modulating histone structure and DNA interaction will also be interesting.

More prominent associations were observed with CMs. Primarily, three modules in HR-CM associations suggested unknown HR function (Supplementary Table S3). Here we focused on H3 and H4, which was fruitful in identifying novel function and associations. A more comprehensive map that includes the other histones [2] will provide further insights into chromatin

regulation. For example, the association scores between Rad6/Bre1 and Set1/Dot1 (Supplementary Fig. S5) support the crosstalk between H2B ubiquitylation and H3K4 and K36 methylation [42]. The association of Rad6 in Module I (Fig. 3C) and with FACT (Supplementary Fig. S5) reflect the cooperation of H2B ubiquitylation and FACT for Pol II elongation [43]. Are there other residues or CMs involved in this crosstalk between H2B and H3/H4? What is the counterpart of H4K91Ac in H2B that prevents salt bridge formation between H4 and H2B [25]? What is the role of H2B ubiquitylation in salt bridge prevention? How does H3K56Ac interact with H2AZ for chromatin remodeling resulting in transcriptional changes? These are only a few of the questions that it will help us to answer.

As shown in the HR-CM association map (Fig. 3C), some HRs have similar association profiles in terms of their interaction with CMs. If multiple HRs interact with a similar set of CMs, they can be also regarded as communicating with one another. As another type of visualization, we employed an HR network that represents similarity of HRs in their association with CMs (Supplementary Fig. S6). Interestingly, we were not able to detect any interacting partners for modifiable tail residues, maybe reflecting that these HRs have a direct and unique regulatory relationship with CMs. In contrast, HRs located at the lateral DNA-contacting surface tended to interact with multiple HRs. In this pilot analysis, we used a high correlation threshold of 0.7 for easy visualization. By lowering the threshold, a more complex network with a larger number of nodes and edges will be obtained. It would be interesting to investigate whether HRs in the network have a redundant or synergistic regulatory function with associated CMs.

In conclusion, our novel method that leveraged genome-wide measurements of gene expression changes by the mutations of single histone residues revealed potential interplay between histone residues and a range of regulatory factors. Genome-wide profiles for >120 histone mutations and >400 *trans*-regulators were associated by means of statistical tests. Based on the data, especially associations with chromatin modifiers, we were able to reinforce previous hypotheses or suggest novel hypotheses on the role of histone residues or their modifications. However, our analysis is primarily based on association and limited to static relationships. Association may not necessarily imply causality. Novel but putative relationships derived from our approach need to be tested by further experiments for causal, mechanistic, or dynamic interactions of histone residues.

Competing interests

The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2015.11.016.

References

[1] Dai, J., Hyland, E.M., Yuan, D.S., Huang, H., Bader, J.S. and Boeke, J.D. (2008) Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell 134, 1066–1078.

[2] Nakanishi, S., Sanderson, B.W., Delventhal, K.M., Bradford, W.D., Staehling-Hampton, K. and Shilatifard, A. (2008) A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. Nat. Struct. Mol. Biol. 15, 881–888.

[3] Huang, H., Maertens, A.M., Hyland, E.M., Dai, J., Norris, A., Boeke, J.D. and Bader, J.S. (2009) HistoneHits: a database for histone mutations and their phenotypes. Genome Res. 19, 674–681.

[4] Kouzarides, T. (2007) Chromatin modifications and their function. Cell 128, 693–705.

[5] Xu, F., Zhang, K. and Grunstein, M. (2005) Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell 121, 375–385.

[6] Masumoto, H., Hawke, D., Kobayashi, R. and Verreault, A. (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature 436, 294–298.

[7] Schneider, J., Bajwa, P., Johnson, F.C., Bhaumik, S.R. and Shilatifard, A. (2006) Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. J. Biol. Chem. 281, 37270–37274.

[8] Van Leeuwen, F., Gafken, P.R. and Gottschling, D.E. (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 109, 745–756.

[9] Ng, H.H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y. and Struhl, K. (2002) Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev. 16, 1518–1527.

[10] Jenuwein, T. and Allison, C.D. (2001) Translating the histone code. Science 293 (80), 1074–1080.

[11] Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z. and Widom, J. (2006) A genomic code for nucleosome positioning. Nature 442, 772–778.

[12] Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J. and Segal, E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458, 362–366.

[13] Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K. and Young, R.A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517–527.

[14] Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. and Zhao, K. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. 40, 897–903.

[15] Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887–898.

[16] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.

[17] Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R. and Nislow, C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. 39, 1235–1244.

[18] Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C. and Pugh, B.F. (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature 446, 572–576.

[19] Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J. and Rando, O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science 309 (80), 626–630.

[20] Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N. and Rando, O.J. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. PLoS Biol. 3, e328.

[21] Ginsburg, D.S., Govind, C.K. and Hinnebusch, A.G. (2009) NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. Mol. Cell. Biol. 29, 6473–6487.

[22] Govind, C.K., Zhang, F., Qiu, H., Hofmeyer, K. and Hinnebusch, A.G. (2007) Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. Mol. Cell 25, 31–42.

[23] Ng, H.H., Robert, F., Young, R.A. and Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell 11, 709–719.

[24] Hyland, E.M., Cosgrove, M.S., Molina, H., Wang, D., Pandey, A., Cottee, R.J. and Boeke, J.D. (2005) Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 25, 10060–10070.

[25] Ye, J., Ai, X., Eugeni, E.E., Zhang, L., Carpenter, L.R., Jelinek, M.A., Freitas, M.A. and Parthun, M.R. (2005) Histone H4 lysine 91 acetylation. Mol. Cell 18, 123–130.

[26] Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M. and Reinberg, D. (2003) FACT facilitates transcription-dependent nucleosome alteration. Science 301, 1090–1093.

[27] Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S. and Wu, C. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science 303, 343–348.

[28] Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G., Wei, D. and Wu, C. (2005) Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. Nat. Struct. Mol. Biol. 12, 1064–1071.

[29] Gerber, M. and Shilatifard, A. (2003) Transcriptional elongation by RNA polymerase II and histone methylation. J. Biol. Chem. 278, 26303–26306.

[30] Liang, G., Klose, R.J., Gardner, K.E. and Zhang, Y. (2007) Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase. Nat. Struct. Mol. Biol. 14, 243–245.

[31] Mulligan, P., Yang, F., Di, Stefano L., Ji, J.-Y., Ouyang, J., Nishikawa, J.L., Toiber, D., Kulkarni, M., Najafi-Shoushtari, S.H., Mostoslavsky, R., Gygi, S.P., Gill, G., Dyson, N.J. and Näär, A.M. (2011) A SIRT1-LSD1 corepressor complex regulates notch target gene expression and development. Mol. Cell 42, 689–699.

[32] Xu, E.Y., Kim, S. and Rivier, D.H. (1999) SAS4 and SAS5 are locus-specific regulators of silencing in *Saccharomyces cerevisiae*. Genetics 153, 25–33.

[33] Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W. and Zhao, K. (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 138, 1019–1031.

[34] Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P. and Workman, J.L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123, 581–592.

[35] Joshi, A.A. and Struhl, K. (2005) Eaf3 chromodomain interaction with methylated H3–K36 links histone deacetylation to Pol II elongation. Mol. Cell 20, 971–978.

[36] Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., Boone, C., Emili, A., Weissman, J.S., Hughes, T.R., Strahl, B.D., Grunstein, M., Greenblatt, J.F., Buratowski, S. and Krogan, N.J. (2005) Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123, 593–605.

[37] Brown, M.A., Sims, R.J., Gottlieb, P.D. and Tucker, P.W. (2006) Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. Mol. Cancer 5, 26.

[38] Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N.D., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Imhof, A., Garza, D., Peters, A.H.F.M. and Schübeler, D. (2007) Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. EMBO J. 26, 4974–4984.

[39] Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., Gilmour, D.S., Albert, I. and Pugh, B.F. (2008) Nucleosome organization in the *Drosophila* genome. Nature 453, 358–362.

[40] Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. and Chin, J.W. (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol. Cell 36, 153–163.

[41] Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W. and Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. 319, 1097–1113.

[42] Kim, J., Guermah, M., McGinty, R.K., Lee, J.-S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W. and Roeder, R.G. (2009) RAD6-mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell 137, 459–471.

[43] Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A. and Reinberg, D. (2006) Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell 125, 703–717.

[44] Unnikrishnan, A., Gafken, P.R. and Tsukiyama, T. (2010) Dynamic changes in histone acetylation regulate origins of DNA replication. Nat. Struct. Mol. Biol. 17, 430–437.

[45] Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18, S96–S104.

[46] Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127.

[47] Steinfeld, I., Shamir, R. and Kupiec, M. (2007) A genome-wide analysis in *Saccharomyces cerevisiae* demonstrates the influence of chromatin modifiers on transcription. Nat. Genet. 39, 303–309.

[48] Choi, J.K. and Kim, Y.-J. (2008) Epigenetic regulation and the variability of gene expression. Nat. Genet. 40, 141–147.

[49] Gupta, S., Dennis, J., Thurman, R.E., Kingston, R., Stamatoyannopoulos, J.A. and Noble, W.S. (2008) Predicting human nucleosome occupancy from primary sequence. PLoS Comput. Biol. 4, e1000134.

[50] Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O. and Schreiber, S.L. (2004) Global nucleosome occupancy in yeast. Genome Biol. 5, R62.

[51] Ponts, N., Harris, E.Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G.R., Hardiman, G., Lonardi, S. and Le Roch, K.G. (2010) Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res. 20, 228–238.

[52] Choi, J.K. (2010) Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol. 11, R70.

[53] Albert, I., Wachi, S., Jiang, C. and Pugh, B.F. (2008) GeneTrack – a genomic data processing and visualization framework. Bioinformatics 24, 1305–1306.

[54] Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I. and Pugh, B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. 18, 1073–1083.