Nucleic Acids Research, 2024, 1–16
https://doi.org/10.1093/nar/gkae361
Computational Biology

asteRIa enables robust interaction modeling between chromatin modifications and epigenetic readers

Mara Stadler[1,2,*], **Saulius Lukauskas**[3], **Till Bartke**[3,†] and **Christian L. Müller**[1,2,4,†]

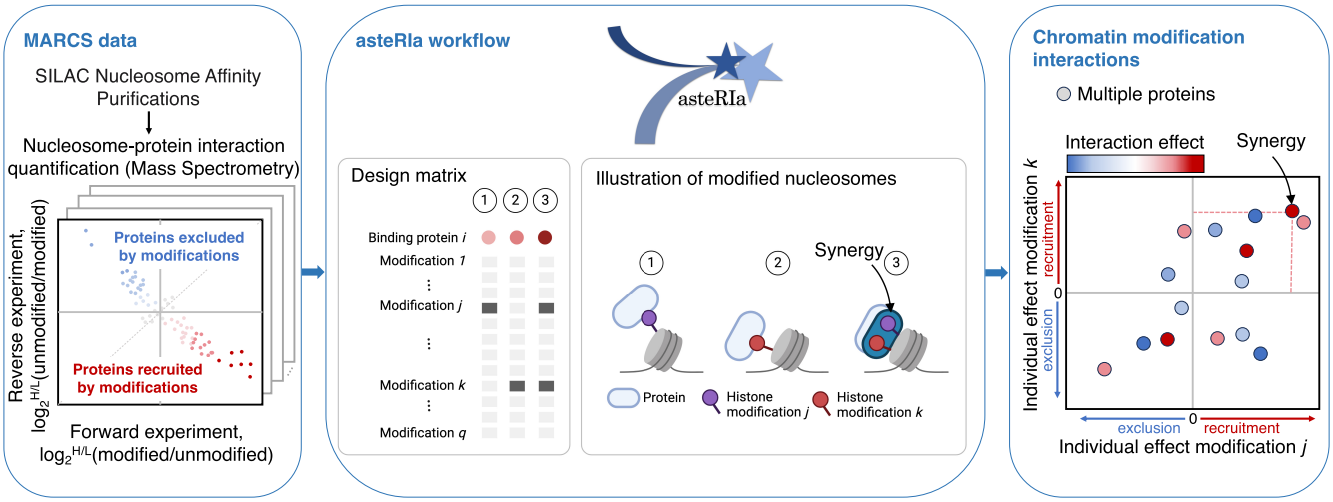
[1]Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[2]Department of Statistics, Ludwig-Maximilians-University Munich, 80539 Munich, Germany
[3]Institute of Functional Epigenetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[4]Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, USA

[*]To whom correspondence should be addressed. Tel: +49 8921803466; Email: mara.stadler@stat.uni-muenchen.de
[†]Christian L. Müller and Till Bartke co-supervised the project.

Abstract

Chromatin, the nucleoprotein complex consisting of DNA and histone proteins, plays a crucial role in regulating gene expression by controlling access to DNA. Chromatin modifications are key players in this regulation, as they help to orchestrate DNA transcription, replication, and repair. These modifications recruit epigenetic 'reader' proteins, which mediate downstream events. Most modifications occur in distinctive combinations within a nucleosome, suggesting that epigenetic information can be encoded in combinatorial chromatin modifications. A detailed understanding of how multiple modifications cooperate in recruiting such proteins has, however, remained largely elusive. Here, we integrate nucleosome affinity purification data with high-throughput quantitative proteomics and hierarchical interaction modeling to estimate combinatorial effects of chromatin modifications on protein recruitment. This is facilitated by the computational workflow `asteRIa` which combines hierarchical interaction modeling, stability-based model selection, and replicate-consistency checks for **a st**able **e**stimation of **R**obust **I**nter**a**ctions among chromatin modifications. `asteRIa` identifies several epigenetic reader candidates responding to specific *interactions* between chromatin modifications. For the polycomb protein CBX8, we independently validate our results using genome-wide ChIP-Seq and bisulphite sequencing datasets. We provide the first quantitative framework for identifying cooperative effects of chromatin modifications on protein binding.

Graphical abstract



Introduction

Eukaryotic cells store the genetic material in the nucleus where it is packaged into chromatin, a nucleo-protein complex made up primarily of DNA and histone proteins. Both DNA and histones carry chemical modifications that can either directly affect chromatin structure or recruit so-called epigenetic reader proteins that mediate downstream events. As these modifications are involved in the regulation of all DNA-templated processes, such as transcription, DNA replication, or DNA repair, they play central roles in controlling chromatin function (1). The basic repeating unit of chromatin is the nucleosome, which coordinates 147 bp of DNA wrapped around an octamer consisting of two copies each of the core histones H2A, H2B, H3 and H4 (2). Nucleosomes are folded into higher-order structures to form chromatin. Since DNA and histone modifications show extensive overlap in the genome (3) and

Received: January 17, 2024. **Revised:** March 15, 2024. **Editorial Decision:** April 20, 2024. **Accepted:** April 24, 2024

decorate histones and nucleosomes in specific combinations (4–10), it is likely that these modifications act in a concerted manner. This is supported by the observation that most chromatin regulators contain multiple modification binding domains or are part of multi-subunit complexes harbouring multiple such domains, and are therefore likely to read out multiple chromatin modifications (11). Indeed, the idea that combinations of histone modifications may form a 'histone code' that together with DNA modifications could store epigenetic information in the chromatin template, thereby expanding the genetic information encoded in the DNA sequence, has been around for over two decades (12–14).

To date, the functions and readers of a host of *individual* chromatin modifications have been described (see, e.g., (15,16) and references therein for an overview). Moreover, epigenetic regulators that read the modification status of more than one epigenetic mark on histones or the DNA have been described using functional and structural studies (17–23). Several DNA repair factors were also found to recognize dual histone modification signatures, ranging from individual interactions (24–29) to combinatorial ones (30,31). One prime example is the ubiquitin ligase UHRF1, an essential player in DNA methylation maintenance, that recognizes a triple modification signature on histone H3 (32–34) and the DNA (35–37).

The gap in knowledge about the combinatorial nature of factors that read multiple DNA and histone modifications can be partially attributed to the fact that one of the most prevailing high-throughput technology to study histone modifications and their readers is chromatin immunoprecipitation followed by deep sequencing (ChIP-seq). Here, antibodies are used to detect the localization of specific modifications *or* chromatin-binding proteins at a genome-wide scale (38). Despite its groundbreaking influence on our understanding of the histone code through community efforts such as the NIH Roadmap Epigenomics Mapping Consortium (39), ENCODE (40), and ChIP-Atlas (41), ChIP-seq alone can only probe a single modification or reader protein in each experiment, thus making it difficult to assess combinatorial synergies or antagonistic effects on epigenetic readers. However, careful integration of multiple genome-wide ChIP-seq experiments of individual modifications enabled the application of *multivariate* statistical analysis techniques to uncover chromatin states and interactions. For example, using hidden Markov modeling techniques, the `ChromHMM` method (42,43) revealed cell-type specific discrete chromatin states that characterize the combinatorial presence or absence of modifications on the genome. Alternatively, sparse partial correlation estimation techniques were proposed to learn multivariate association networks between histone modifications (44). The latter framework was extended in (45,46) to include both histone modifications and a small set of chromatin modifiers. Using linear regression and sparse partial correlation estimation, the studies derived *de novo* high-confidence backbones of 'chromatin signaling networks' from ChIP-Seq data. There, the inferred network edges are to be interpreted as additive (or main) effects between histone modifications on chromatin modifiers and vice versa. The analysis of the derived chromatin signaling networks revealed both histone-protein interactions known from literature and several novel hypothetical interactions. To show the power of the network approach, the authors were also able to experimentally verify the statistically hypothesized interactions between H4K20me1 and members of the polycomb repressive complexes 1 and 2 (PRC1 and PRC2, respectively) (46). Nevertheless, none of these ChIP-Seq-based computational approaches allow the statistical estimation of how *multiple* histone modifications co-operate in recruiting epigenetic regulators.

In this contribution, we present a statistical interaction modeling approach, termed `asteRIa`, that tackles this challenge. Rather than considering genome-wide ChIP-Seq data, `asteRIa` uses novel nucleosome affinity purification data with high-throughput quantitative proteomics, as provided in the Modification Atlas of Regulation by Chromatin States (MARCS), to make robust and reproducible predictions of combinatorial effects of chromatin modifications on chromatin-interacting proteins. The MARCS data, available at https://marcs.helmholtz-munich.de comprises a collection of Stable Isotope Labeling with Amino acids in Cell culture (SILAC) nucleosome affinity purification (SNAP) experiments (47) that probe the binding of proteins from HeLa S3 nuclear extracts to a library of semi-synthetic di-nucleosomes (referred to as nucleosomes throughout the manuscript) incorporating biologically meaningful combinations of chromatin modifications representing promoter, enhancer and heterochromatin modification states. Each affinity purification measures the relative abundances of nuclear proteins on a modified nucleosome in relation to an unmodified control nucleosome using the SILAC labelling and quantitative proteomics as a read out. This allows the high-throughput identification of proteins that are either recruited or excluded by the modification(s), and also indicates the relative extent of the recruitment or exclusion. Collectively, the MARCS data set catalogs the binding responses of 1915 nuclear proteins to nucleosomes carrying 55 different modification signatures. The constructive nature of these data, paired with an appropriate statistical model, thus enables the direct analysis of combinatorial effects of different modification features on the nucleosome binding of the measured proteins. At its core, `asteRIa` uses a linear regression model with pairwise (or 'two-way') interactions among chromatin modifications to predict the binding affinities of each protein. Regression models with pairwise interactions have a long tradition in statistics and experimental design (48–50) but are notoriously difficult to estimate in the presence of noisy, scarce data and/or incomplete experimental designs, and are prone to misinterpretation (51,52). As we will show, the `asteRIa` framework incorporates several model and design principles that (i) guard against common pitfalls and (ii) take the properties of the MARCS data (and biological data in general) into account. Firstly, we posit that our framework should work in the underdetermined regime, i.e. the number of features q (here the chromatin modifications) and pairwise interactions exceeds the number of measurements n. We achieve this by including sparsity-inducing penalization of the model coefficients (53–55). Secondly, we assume that the underlying interaction model obeys the so-called 'strong hierarchy' principle (50,53,56), i.e. interactions among features are only included in the model if both features are present as main effects. Thirdly, we embrace the principle of statistical 'stability' (57–59) for model selection, implying that interactions are only included when they are reproducibly identified across subsets of the data. To respect the ubiquitous measurement variability of biological systems, we also require replicate consistency (60) of our combinatorial models. This means that models with interactions need to be (at least partially) consistent across available technical

or biological replicates, further ensuring the general robustness and validity of the resulting models. While these design principles and the underlying computational workflow, available at https://github.com/marastadler/asteRIa.git, are general, we illustrate the framework to detect novel combinatorial interactions between chromatin modifications on epigenetic reader recruitment.

On the MARCS data, we show that considering interaction effects between chromatin modifications can consistently improve the predictive performance of the binding profiles of a subset of proteins. `asteRIa` not only recovers known binding patterns, such as, e.g. the well-known H3K27me3-CBX8 pairing, but also identifies novel interaction effects between chromatin modifications on the binding behavior of proteins not yet implicated as epigenetic readers (e.g. ACTL8). Our analysis also allows to define and quantify the extent of distinct modes of apparent chromatin modification interactions, ranging from synergistic and antagonistic to competitive effects. Our post-hoc model analysis shows that proteins belonging to the same protein complexes do read combinatorial chromatin modification signatures in a similar fashion, thus allowing the delineation of a protein complex - chromatin modification interaction network.

Independent confirmation of the identified combinatorial interactions is challenging due to the uniqueness of the MARCS data and the accompanying statistical analysis. Nevertheless, we provide a validation workflow on ENCODE ChIP-Seq, ChIP-Atlas ChIP-Seq and WGBS (Whole Genome Bisulfite Sequencing) data that demonstrates that our findings are not limited to a specific cell type or experimental setup. Specifically, we show that one of the found combinatorial interactions for CBX8 are consistent with these orthogonal datasets. The latter analysis also illustrates how to validate other interactions found in this study, thus inviting the generation of new ChIP-Seq data collections for previously understudied proteins.

Materials and methods

The Modification Atlas of Regulation by Chromatin States dataset

The Modification Atlas of Regulation by Chromatin States (MARCS), as introduced in (61), builds on two experimental components: (i) a designed library of engineered dinucleosomes (referred to as nucleosomes throughout the manuscript) comprising combinatorial chromatin modifications and (ii) nucleosome affinity purifications coupled to high-throughput quantitative proteomics measurements employing SILAC labeling (SNAP) (47). The modified nucleosomes were assembled from a biotinylated DNA containing two 601 nucleosome positioning sequences (62) and histone octamers containing semi-synthetic site-specifically modified histones H3.1 and H4 prepared by native chemical ligation (63). Some nucleosomes were also assembled using CpG-methylated DNA (5mC) or the histone variant H2A.Z. The complete library design matrix comprises $n_{\text{total}} = 55$ modified nucleosomes with thirteen possible chromatin modifications (see left panel of Figure 1 for a conceptual picture). The available modifications include six lysine residues on the tails of histone H3 (K4, K9, K14, K18, K23 and K27) and five on histone H4 (K5, K8, K12, K16 and K20) as well as the variant histone H2A.Z and CpG methylated (5mC) DNA on both DNA strands (symmetric methylation), respectively. The lysines are modified with acetylation (ac) or mono-, di-, or trimethylation (me1, me2, me3). H3-5ac denotes that multiple acetylations (K9, K14, K18, K23, K27) on the tails of histone H3 are present. H4-4ac denotes that multiple acetylations (K8, K5, K12, K16) on the tails of histone H4 are present. For our computational analysis, we do not consider engineered nucleosomes that include subsets of acetylations (namely, not all five acetylations on H3 or not all four acetylations on H4) since building their mathematical products would result in perfectly collinear (thus fully redundant, and therefore not distinguishable) pair-wise interaction features (see **Interaction-modeling strategy** for further clarification). Our analysis thus excludes 22 nucleosomes from the initial nucleosome library and considers a subset of $n = 33$ nucleosomes with $q = 12$ different chromatin modifications, resulting in the design matrix $L \in \{0, 1\}^{33 \times 12}$. The (transposed) design matrix L with the available combinatorial modifications is shown in the top panel ((Step 1) of Figure 2). Note that the design pattern in L does not follow any particular statistical experimental design guideline (50) but is driven by biological expertise about common modification co-occurrences.

For each modified nucleosome in MARCS, SNAP experiments are provided in two experimental 'label-swap' replicates of the nucleosome affinity purification process, a 'forward' (F) and 'reverse' (R) nucleosome pull-down. Nucleosomes are immobilized on streptavidin beads and incubated with nuclear extracts from HeLa S3 cells cultured either in isotopically light or heavy-labelled SILAC media. In the 'forward' experiments the heavy extracts are incubated with the modified and the light extracts with the unmodified nucleosome, in the 'reverse' experiments the extracts are exchanged. Bound proteins are eluted from the beads and identified and quantified by mass spectrometry. For each SNAP experiment the relative abundance of a given protein on the modified nucleosome is determined in relation to the unmodified nucleosome by measuring the ratios between the heavy and the light peptides (H/L ratios) identified for that particular protein (47). The H/L ratios indicate binding preferences to the modified or the unmodified nucleosomes and allow the unbiased identification of proteins that are either recruited or excluded by the modification(s) present on the modified nucleosomes. In addition, the SILAC enrichment ratios also indicate a relative 'strength' of the recruitment or exclusion of a given protein by the modifications. In total, the MARCS dataset comprises the binding behavior of $p = 1915$ proteins in the forward (F) and reverse (R) experiments. For our analysis, we consider the protein measurement matrices $P^F, P^R \in \mathbb{R}^{33 \times 1915}$ that correspond to the subset of $n = 33$ nucleosomes, described above.

Interaction modeling strategy

We aim at predicting the binding profile of each protein captured in MARCS $(P_i)_{1 \leq i \leq 1915}$ (either from the forward or reverse experiment) from the combinations of nucleosome modifications $(L_j)_{1 \leq j \leq 12}$. Given the binary design matrix L, the baseline model of uncovering (joint) additive effects of the modifications on a binding profile $Y = P_i \in \mathbb{R}^n$, $i = 1, \ldots, p$, is the linear model

$$Y = \beta_0 + \sum_{j=1}^{q} \beta_j L_j + \epsilon, \tag{1}$$

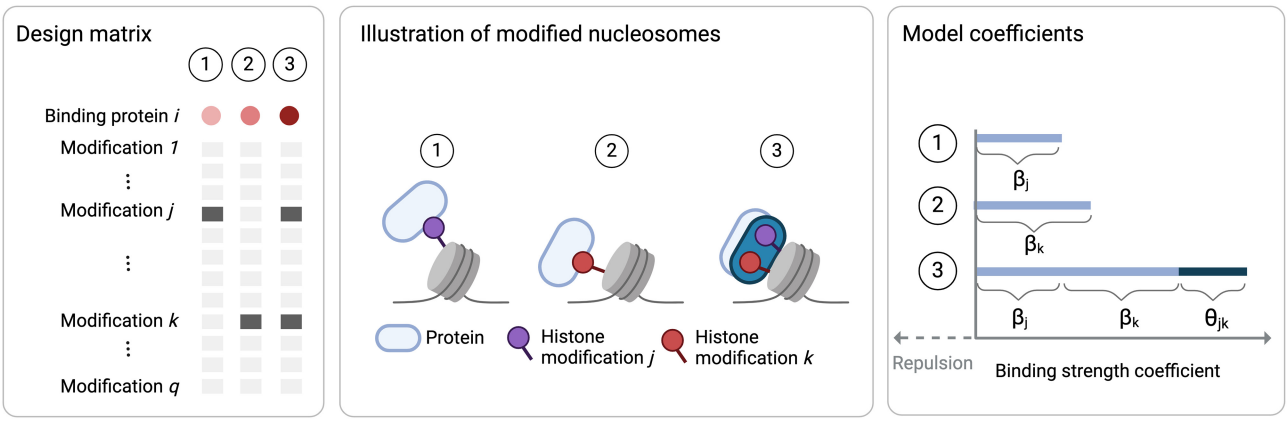


Figure 1. Left: Three exemplary columns of the design matrix L. Dark gray boxes indicate that a modification has been installed on the respective nucleosome. Above the design, the binding behavior of an exemplary protein to the modified nucleosomes is shown by color. The shade of red indicates the strength of the binding effect. Center: Illustration of two *individual* binding effects of chromatin modifications j and k on a protein P_i (1 and 2). Synergistic *combinatorial* effect of the modifications j and k on protein P_i (dark blue) compared to expected binding effect under independence of modification j and k (light blue) (3). Right: Model coefficients/estimated binding strength of protein P_i for the three scenarios. Light blue bar in scenario 3 shows the binding strength under independence of modification j and k, $\beta_j + \beta_k$. Dark blue shows the additional combinatorial effect θ_{jk} that goes beyond additive combinatorial effects (created with BioRender.com).

where $\beta_0 \in \mathbb{R}^n$ is a protein-specific (constant) intercept, β_j is the effect of modification j on the binding profile $Y = P_i$ of protein i, and ϵ models the technical and biological noise component. In (61), a simplified version of this baseline model was investigated through 'feature effect estimates' via pairwise comparisons of the enrichments of individual proteins on nucleosomes differing by a single modification feature. This, however, only allowed robust prediction of the effects of individual modifications or blocks of modifications and did not provide any information on combinatorial effects. Here, we extend the baseline model by including all pairwise interactions between modifications. For each protein binding profile $Y = P_i$, $i = 1, ..., p$, the core model in `asteRIa` thus reads

$$Y = \beta_0 + \sum_{j=1}^{q} \beta_j L_j + \frac{1}{2} \sum_{j=1}^{q} \sum_{k=1}^{q} \Theta_{jk} L_j L_k + \epsilon\,, \qquad (2)$$

where Θ_{jk} models interaction effects between epigenetic readers that cannot be captured by linear additive effects. Robustly and reproducibly estimating non-zero entries in the interaction matrix Θ from replicated data is at the heart of the `asteRIa` workflow. The sign of the interaction coefficients also allows a characterization of epigenetic reader interplay. For example, when $\hat{\Theta}_{jk} > 0$ we interpret the two modifications j and k to have a synergistic binding effect if both $\beta_j > 0$ and $\beta_k > 0$ (see Figure 1 for illustration).

To guarantee identifiability and interpretability of individual interaction models, we first need to ensure that the interaction design matrix $L_j L_k$ has no co-linear columns. In the concrete example of the MARCS data, we group modifications of the complete design matrix to a set of $n = 33$ non-redundant nucleosomes (see top panel (Step 1) of Figure 2). Secondly, to enable estimation in the present underdetermined regime ($q(q + 1)/2 > n$) with $q(q + 1)/2 = 78$, we perform regularized maximum-likelihood estimation with ℓ_1-norm (lasso) penalization (64) on the linear and interaction coefficients, respectively. Given the log-likelihood function of the model $l(\beta_0, \beta, \Theta) = \left\| Y - \beta_0 - L\beta - \frac{1}{2} L\Theta L^T \right\|_2^2$, the (all-pairs) lasso problem reads

$$\min_{\beta_0, \beta, \Theta} l(\beta_0, \beta, \Theta) + \lambda \|\beta\|_1 + \tfrac{\lambda}{2} \|\Theta\|_1\,, \qquad (3)$$

where $\lambda > 0$ is a tuning parameter and controls the sparsity levels of the coefficients β and Θ, respectively. To further ease model interpretability, we follow the statistical principle of hierarchy (also known as marginality or heredity) and allow the presence of an interaction in the model *only if* the associated linear (main) effects are in the model as well (see (53), and references therein). In mathematical terms, this so-called strong hierarchy principle can be expressed as

$$\hat{\Theta}_{jk} \neq 0 \Rightarrow \hat{\beta}_j \neq 0 \text{ and } \hat{\beta}_k \neq 0,$$

implying that interaction effects are only present if both linear effects enter the model. This hierarchy can be achieved by adding a constraint on the interaction effects $\Theta_j \in \mathbb{R}^q$ and a symmetry constraint on Θ. The corresponding optimization problem with hierarchical interactions thus reads

$$\min_{\beta, \Theta} l(\beta_0, \beta, \Theta) + \lambda \|\beta\|_1 + \tfrac{\lambda}{2} \|\Theta\|_1$$
$$\text{s.t. } \Theta = \Theta^T, \qquad \|\Theta_j\|_1 \leq |\beta_j|\,. \qquad (4)$$

To solve the non-convex optimization problem in (4), we follow Bien et al. (53) who proposed a convex relaxation of the problem and provide an efficient implementation in the corresponding R package `hierNet` (65) (v1.9). In `asteRIa`, we use `hierNet` to model each protein binding profile $Y = P_i$, $i = 1, ..., p$ with hierarchical interactions. Apart from reducing the number of spurious interaction effects, a major advantage of the strong hierarchy constraint is the so called 'practical sparsity'. The strong hierarchy constraint favors models that 'reuse' measured variables. In the context of the MARCS data, this becomes important when generating hypotheses for follow-up functional analysis (where experiments are complex and costly). Concretely, our models assumes that a protein or protein complex must have a domain capable of recognizing a particular chromatin modification. Thus, if there exists a response of a protein to an interaction effect between two modifications, a (possibly small) linear effect to both modifications is expected.

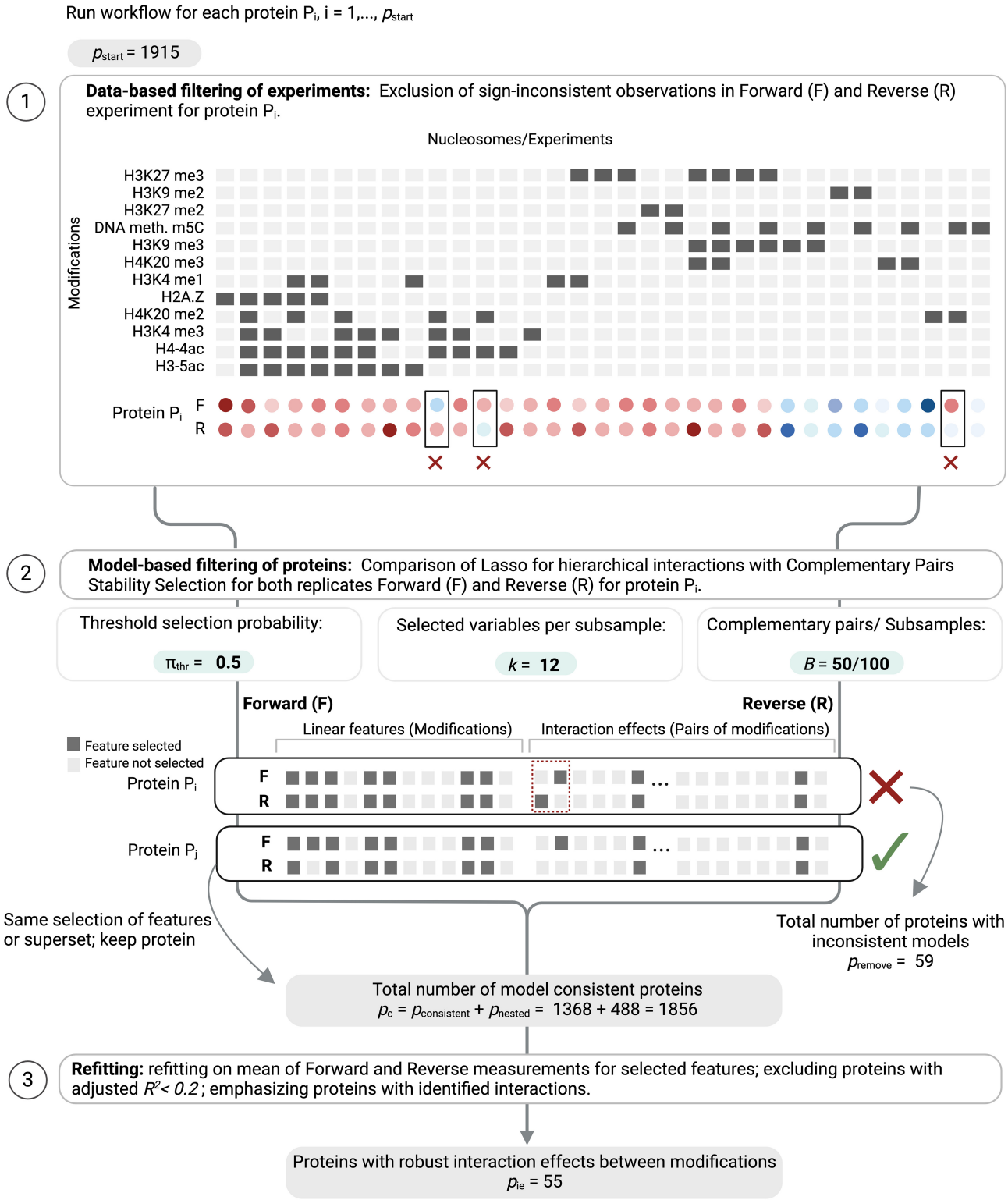


Figure 2. Graphical representation of the `asteRIa` workflow for the robust detection of hierarchical interactions (created with BioRender.com). Step 1: Design matrix and measured binding behavior for a protein (two replicates F and R). Removal of observations with different signs in the replicates (sign consistency). Step 2: Hierarchical interaction modeling with default complementary pairs stability selection (CPSS) parameters. Comparison of selected features for each replicate (nested model consistency): The first example shows a protein prediction model that gets filtered out since the selected features from the forward and reverse replicate are neither identical nor nested. The second example shows a 'consistent' protein model where the selected features learned from the reverse replicate is a nested subset of the features learned from the forward replicate. Step 3: Least-squares refitting on averaged replicate data for final prediction model building. The intersection of two selected feature sets is used for refitting. Models with adjusted $R^2 < 0.2$ are discarded.

Stability-based model selection for hierarchical interactions

One of the core challenges in high-dimensional penalized regression is determining a suitable regularization parameter λ that trades off sparsity (i.e. interpretability) of the model coefficients and out-of-sample predictive performance of the model (66,67). Standard procedures for (hierarchical) interaction models include cross-validation (53) and Information Criteria, including the Akaike (AIC) and the extended Bayesian Information Criterion (BIC) (55). However, it has been observed that, both in simulation and practice, cross-validation and Information criteria tend to select more predictors (and interactions) than necessary (55).

To address this shortcoming, we follow the principle of stability (57) in `asteRIa` and introduce stability selection (58) for the identification of a reproducible set of predictive features *and* interactions. Stability selection has been proven useful across several scientific applications, ranging from network learning (68,69) to data-driven partial differential equation identification (70,71). In the regression context, stability selection repeatedly learns sparse regression models from subsamples of the data of fixed size (e.g. $n_s = \lfloor n/2 \rfloor$), records the frequency of all selected predictors across the models, and selects the most frequent predictors to fit the final regression model. Here, we use a variant of stability selection, the so-called complementary pairs stability selection (CPSS) (59) which draws B subsamples as complementary pairs $\{(A_{2b-1}, A_{2b}): b = 1, ..., B\}$, with $A_{2b-1} \cap A_{2b} = \emptyset$ of samples $\{1, ...n\}$ of size $\lfloor n/2 \rfloor$. Drawing complementary pairs is particularly beneficial when dealing with unbalanced experimental designs, as the resulting random splits ensure that individual subsamples are independent of each other. After applying a variable selection procedure S (e.g. using the first k predictors that enter the penalized model), each feature j in the model gets an individual estimated selection probability $\hat{\pi}(j)$, given by

$$\hat{\pi}(j) = \frac{1}{2B} \sum_{b=1}^{2B} \mathbb{1}_{\{j \in \hat{S}(A_b)\}}, \tag{5}$$

and the final selection set is given by $\hat{S}^{\text{CPSS}} = \{j : \hat{\pi}(j) \geq \pi_{\text{thr}}\}$, for a threshold π_{thr} defining the minimum selection frequency. In our workflow we use the corresponding R package `stabs` (72) (v0.6-4) that provides an efficient implementation of CPSS. The CPSS procedure includes the following hyperparameters: The set of regularization parameters Λ, a threshold $\pi_{\text{thr}} \in [0, 1]$, the number of predictors k that first enter the sparse model, and the number of complementary splits B. In `asteRIa`, we set as default parameters Λ to be the internal λ-path in Bien and Tibshirani (65), $\pi_{\text{thr}} = 0.5$, $k = 12$ and $B = 50$, resulting in 100 subsamples. For the MARCS data, this means that chromatin modifications (as main or interaction effects) are part of the pairwise interaction model 2 for protein binding profile i, $Y = P_i$, if it is among the $k = 12$ selected modifications in at least 50 subsamples. While these default values may need to be tuned in other scenarios, we verified in a realistic semi-synthetic simulation scenario (see Supplementary information and Supplementary Figures S1 and S2 for details) that hierarchical interaction modeling with stability selection greatly outperforms cross-validation, particularly in terms of false positive rate.

Replicate consistency

Biological datasets typically include replicated measurements (replicates) to probe different sources of variability in the underlying experimental procedure or study object (73). The MARCS dataset, for example, comprises two technical replicates of the SILAC-based protein binding affinities. Replicate consistency, i.e. assessing how consistent two or multiple replicated measurements are in terms of direction or size, is an important property to evaluate experimental protocols and downstream analysis quality (see, e.g. (74) for a discussion in the context of RNA sequencing data).

In `asteRIa`, we propose and include two replicate-consistency mechanisms: (i) data sign-consistency and (ii) nested model consistency. While there are alternative ways of performing filtering, data sign-consistency can be considered as a data filtering step that ensures that replicated measurements agree on the direction, i.e., the sign of the measured unit, and removes experiments where sign consistency does not hold. In MARCS, we perform data sign consistency for each protein P_i separately using the forward and reverse replicates (see Figure 2, Step 1) and remove nucleosomes (experiments) where measured protein binding affinities disagree in sign. Although this reduction in sample size (for each protein $n_i \leq n$ samples are available) decreases the power for subsequent hierarchical interaction modeling, the filtering increases the chance of estimating pairs of consistent interaction models. In a second post-hoc step, nested model consistency further ensures that only pairs of consistent interaction models are considered for downstream analysis. Nested model consistency deems estimated interaction models valid only if they comprise the same set of features (main and interaction coefficients) across replicates *or* one model comprises a nested subset of main and interaction effects of the other model (see Figure 2, Step 2, for illustration).

The `asteRIa` workflow

The `asteRIa` workflow incorporates the described model and design principles as illustrated in Figure 2 on the MARCS data. `asteRIa` comprises three main steps: Step (1) uses sign consistency to filter pairs of forward and reverse experiments for each protein $(P_i)_{1 \leq i \leq p = 1915}$. Step (2) comprises model estimation using the hierarchical interaction model, CPSS-based model selection, and the post-hoc nested model consistency filter. Step (3) performs least-squares 'refitting' to estimate main and interaction effect sizes on the selected model coefficients from averaged replicate data. The resulting signed model coefficients are then used for functional categorization and downstream analysis.

On the MARCS data, the experiment filtering step (1) removes on average 11 experiments across all proteins. In step (2), using the internal λ-path in Bien and Tibshirani (65), and CPSS parameters $\pi_{\text{thr}} = 0.5$, $k = 12$, and $B = 50$, `asteRIa` learns $p_{\text{consistent}} = 1368$ fully consistent regression models across forward and reverse replicates, as well as $p_{nested} = 488$ models that obey the nested model consistency criterion. Only $p_{\text{remove}} = 59$ models are inconsistent across replicates. Among all $p_c = 1856$ consistent models, `asteRIa` identifies 58 models that include robust interaction coefficients. The refitting estimation process in step (3) uses the averaged binding affinities as outcome and performs least-squares refitting on the *intersection* of the per-replicate selected features. The refit coefficients are the final effect sizes. For downstream analysis,

asteRIa removes poorly-performing prediction models with adjusted R^2 below 0.2 (three out of 58).

Results

Enhanced predictive performance of protein binding through chromatin modification interaction

We first quantify the overall predictive performance of asteRIa models for all proteins included in the MARCS dataset and then assess the degree to which hierarchical interaction modeling improves overall predictive performance of protein binding affinities. For a majority of the p=1915 protein binding profiles, asteRIa deems main effects models (i.e., the baseline linear model in 1) to be sufficient for robust prediction. For more than 200 proteins, main effects models achieve adjusted $R^2 > 0.8$, and for more than 500 proteins, main effects models achieve adjusted $R^2 > 0.5$ (see Supplementary Figure S4 for a list of top protein binding models and associated coefficients). The top-six protein binding models achieve near-perfect predictive performance and include the protein ING5, a dimeric, bivalent reader of histone H3K4 me3 (75), with an $R^2 = 0.99$, the methyl–lysine histone-binding protein L3MBTL3 ($R^2 = 0.99$), SMARCC2 ($R^2 = 0.99$) which is part of the chromatin remodeling complex SNF/SWI, the histone acetyltransferase KAT7 ($R^2 = 0.98$), the YAF2 protein ($R^2 = 0.98$), and the histone lysine demethylase KDM2B ($R^2 = 0.98$).

However, asteRIa also identifies a set of $p_{ie} = 55$ models that comprise stable interaction effects among modifications with enhanced predictive performance. This provides statistical evidence that cooperative effects between chromatin modifications may play a crucial role in the binding of specific reader proteins and thus in controlling chromatin function. Figure 3A shows the modification design matrix (left panel) and binding profiles (both the 'forward' and the 'reverse' experiments) of the 55 proteins explained by interaction models. The proteins are sorted by data density (i.e., in terms of number of experiments removed due to sign consistency filtering step (1) in asteRIa, Figure 3A, gray boxes). Figure 3C shows the corresponding predictive performance of the models in terms of adjusted R^2 both for main effects (light blue) and interaction models (dark blue), respectively. While the light blue segment denotes the proportion of variance explained by all selected main effects combined, the dark blue portion represents the additional explained variance attributed solely to one interaction. We observe that the inclusion of robust interaction among modifications can boost the performance of up to 0.5 (e.g., for proteins CDKAL1 and PEX11B). For others, such as, e.g., RFC3, the binding behavior can only be sufficiently described by taking into account interaction effects. While the improvement is less dramatic for proteins with well-performing main effects models, asteRIa still provides evidence for stable interactions among modifications. Figure 3B illustrates the stabilities (inclusion probabilities) $\hat{\pi}$ of all model coefficients for the protein CBX8. On both forward and reverse experimental data, asteRIa estimates a high selection probability ($\approx$0.7) of an interaction effect between DNA methylation m5C and H3K9me3 while all other interaction effects emit a low inclusion probability. For detailed model inspection, we provide similar stability plots for all other proteins in the Supplementary Material. To illustrate the improvement in binding prediction, Figure 3C (right panel) shows predicted vs. observed binding profiles for the protein RNF2. Comparison of the fits of both the main effect (light gray) and interaction model (dark blue) visually and quantitatively ($R^2 = 0.76$ versus $R^2 = 0.9$) confirm the enhanced predictive performance of the interaction model.

Modes of chromatin modification interactions

To categorize the interaction effects uncovered in asteRIa, we establish potential modes of chromatin modification interactions. This is achieved by contrasting the effects of individual chromatin modifications (modification j and k) on the binding behavior of specific proteins, represented by the linear model coefficients $\hat{\beta}_j$ and $\hat{\beta}_k$ with the combinatorial effects identified during our analysis, represented by $\hat{\Theta}_{j,k}$ for the corresponding pair (see Figure 4A and B). We define three major modes: synergistic combinatorial behavior, antagonistic combinatorial behavior, and conflicting combinatorial behavior. We further divide these into two sub-modes each of which describes the direction of the combinatorial effect, either towards binding (b, $\Theta_{j,k} > 0$) or towards repulsion (r, $\Theta_{j,k} < 0$). The direction and strength of the combinatorial effect is color-coded in Figure 4B.

The 'Synergy b+b+b' category (shown in blue in Figure 4) includes proteins that bind to two modifications individually and exhibit particularly strong binding, i.e., stronger than the sum of the two individual effects when both modifications are present. For example, we uncover that UHRF1 (Figure 4B, 1st quadrant) responds in a synergistic way to an interaction effect between DNA methylation m5C and H3K9me3. UHRF1 is a RING-type E3 ubiquitin ligase that plays an essential role in DNA methylation by mediating the recruitment of the maintenance DNA methyltransferase DNMT1 (76). UHRF1 is known to bind to H3K9me3 via a tandem tudor domain and to recognize hemi-methylated DNA via a SRA domain. Our analysis therefore validates previously known binding behaviors and, additionally, unveils that there is a true synergistic effect between H3K9me3 and DNA methylation in the recruitment of UHRF1.

For the maintenance DNA methyltransferase DNMT1 (77), we identify an individual binding effect to H3K9me3 and a modest individual binding to DNA methylation. Furthermore, we also identify an interaction effect between DNA methylation m5C and H3K9me3. In this case, however, the addition of DNA methylation m5C leads to a reduction in binding of DNMT1 to H3K9me3. We define this behavior as 'Antagonism b+b+r' or preferential binding (pink category in Figure 4). UHRF1 and DNMT1 were found to interact with each other (see references in (76)), and binding of DNMT1 to H3K9me3 is likely mediated through UHRF1 (see above). Both UHRF1 and DNMT1 are flexible multi-domain proteins, that consist of several different domains and can change their shape or structure. They are involved in a complex network of interactions, both within themselves (intra-molecular) and with each other (inter-molecular). This network helps control their function through allosteric regulation events involving conformational rearrangements of autoinhibitory domains (changes in the structure of certain domains within the proteins) in both molecules (76,78). The antagonistic effect of DNA methylation on the recruitment of DNMT1 to H3K9me3 indicates that while symmetric DNA methylation stimulates binding of UHRF1 to the doubly modified nucleosomes (see above), it disrupts the interaction with DNMT1.

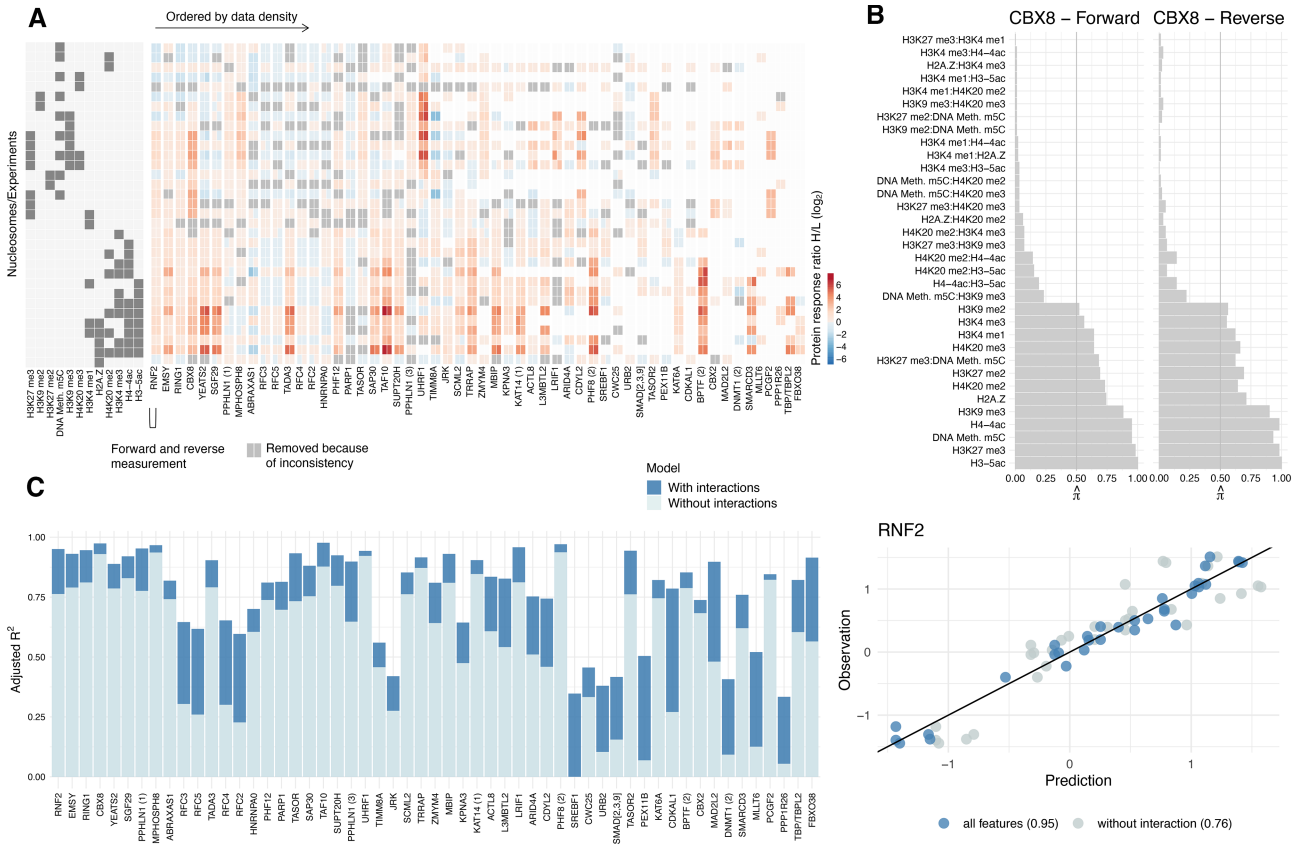


Figure 3. (**A**) Observed protein binding profiles (forward and reverse experiment) for the $p_{ie} = 55$ proteins for which interactions between modifications have been detected. Proteins are arranged from left to right based on data density (number of non-zero measurements), with proteins with the highest data density being on the left. (**B**) Stability plots for CBX8 of the hierarchical interaction model with complementary pairs stability selection (CPSS). Vertical lines show the threshold for the selection probability threshold $\pi_{thr} = 0.5$. Stability plots for all proteins are provided in Extended (B). (**C**) Adjusted R^2 for all $p_{ie} = 55$ proteins of the main effect (light blue) and interaction model (dark blue) (left panel). Scatter plot of observed vs. predicted values for the protein RNF2 (right panel). Scatter plots for all proteins are provided in Extended (C).

This suggests a mechanism within DNMT1 that senses symmetrically methylated DNA (the end product of the DNA methylation reaction) and triggers the release from chromatin upon completion of its enzymatic reaction. Apart from the catalytic domain of DNMT1, which is responsible for the main activity of the protein, this observed behavior could involve a CXXC domain that has a special ability to bind to certain DNA sequences, specifically sequences that contain unmethylated CpG nucleotides, and could contribute to sensing the DNA methylation status.

Two proteins, MAD2L2 and ACTL8, exhibit a similar behavior with respect to DNA methylation m5C and H3K9me3. However, for these proteins, DNA methylation m5C exhibits a slight repulsive effect on its own. These proteins belong to the category 'Conflict, dominated by repulsion b+r+r' (grey category in Figure 4).

Proteins in the 'Conflict, dominated by binding b+r+b' category (yellow category in Figure 4) are repelled by one modification and bind to another modification if they are considered individually. In combination, these modifications show a stronger binding effect on the protein than expected under additivity. The chromodomain-containing protein CBX8, which is a component of the polycomb repressive complex 1 (PRC1) (79), also falls into this category. Our analysis reveals that DNA methylation m5C enhances the binding of CBX8 to H3K27me3, while DNA methylation m5C itself exhibits a slight repulsive effect on CBX8. The association of CBX8 with both DNA and H3K27me3 has been investigated in Connelly et al. (80). Here, the authors identified a dual interaction mechanism for the CBX8 chromodomain, where the engagement of both DNA and H3K27me3 mediates the association of CBX8 with chromatin. Similar binding behaviors are observed for the PRC1 subunits RNF2 and RING1. However, in contrast to CBX8, RNF2, and RING1 are shared among multiple complexes, including the canonical polycomb repressive complex 1 (PCR1) and various non-canonical versions of the complex (ncPRC) (79). The nucleosome binding profiles of these shared subunits reflect a superposition of the binding profiles of all the complexes they are associated with. This introduces additional complexity to the interpretation of combinatorial effects.

Chromatin modification interaction in the recruitment of proteins and complexes

Our analysis suggests that proteins within the same protein complex tend to exhibit similar binding patterns not only to individual chromatin modifications, but also with regards to interaction effects of modifications.

Our analysis reveals seven distinct combinations of chromatin modifications demonstrating a robust combinatorial effect on the shortlisted 55 proteins (see Figure 5A). While six of the discovered interactions affect multiple proteins, H2A.Z incorporation appears to interact solely with H4K20me2,

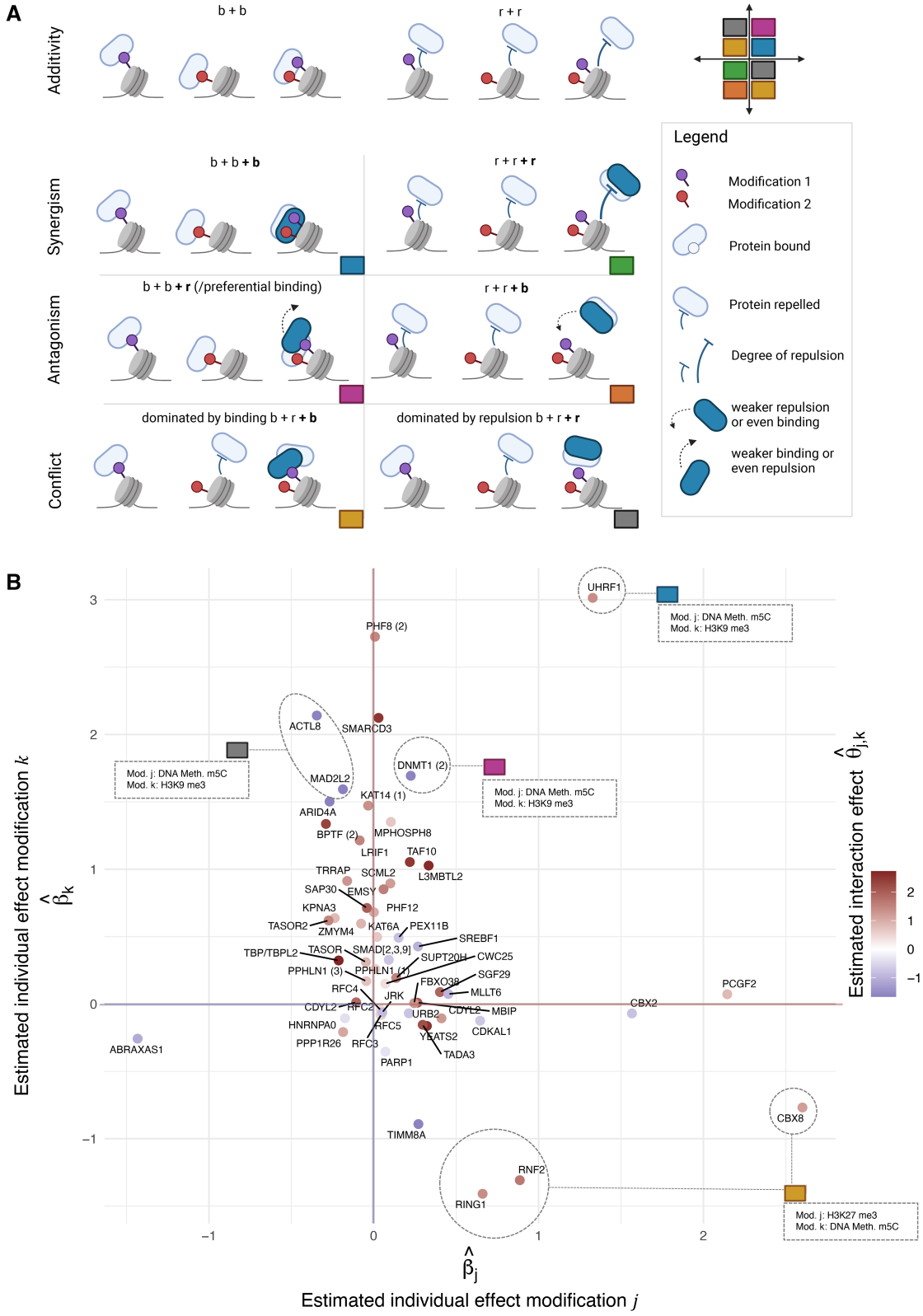


Figure 4. (**A**) Six combinations of combinatorial interaction effects of two modifications on the binding behavior of chromatin-associated proteins (created with BioRender.com). The top row illustrates what would be expected under a purely additive dependence of the effects in a scenario where a protein shows *individual* binding effects to two distinct chromatin modifications (overall *additive* effect is the sum of both individual binding effects, b + b) (left) and where a protein is repelled by two distinct chromatin modifications (right). The rows below shows different modes of deviations due to (directional) interaction effects. (**B**) Scatter plot of protein binding effects with unspecific linear effects β_j and β_k on the *x*- and *y*-axis and corresponding interaction effect $\Theta_{j,k}$ represented by color. For some example proteins detailed information is provided.

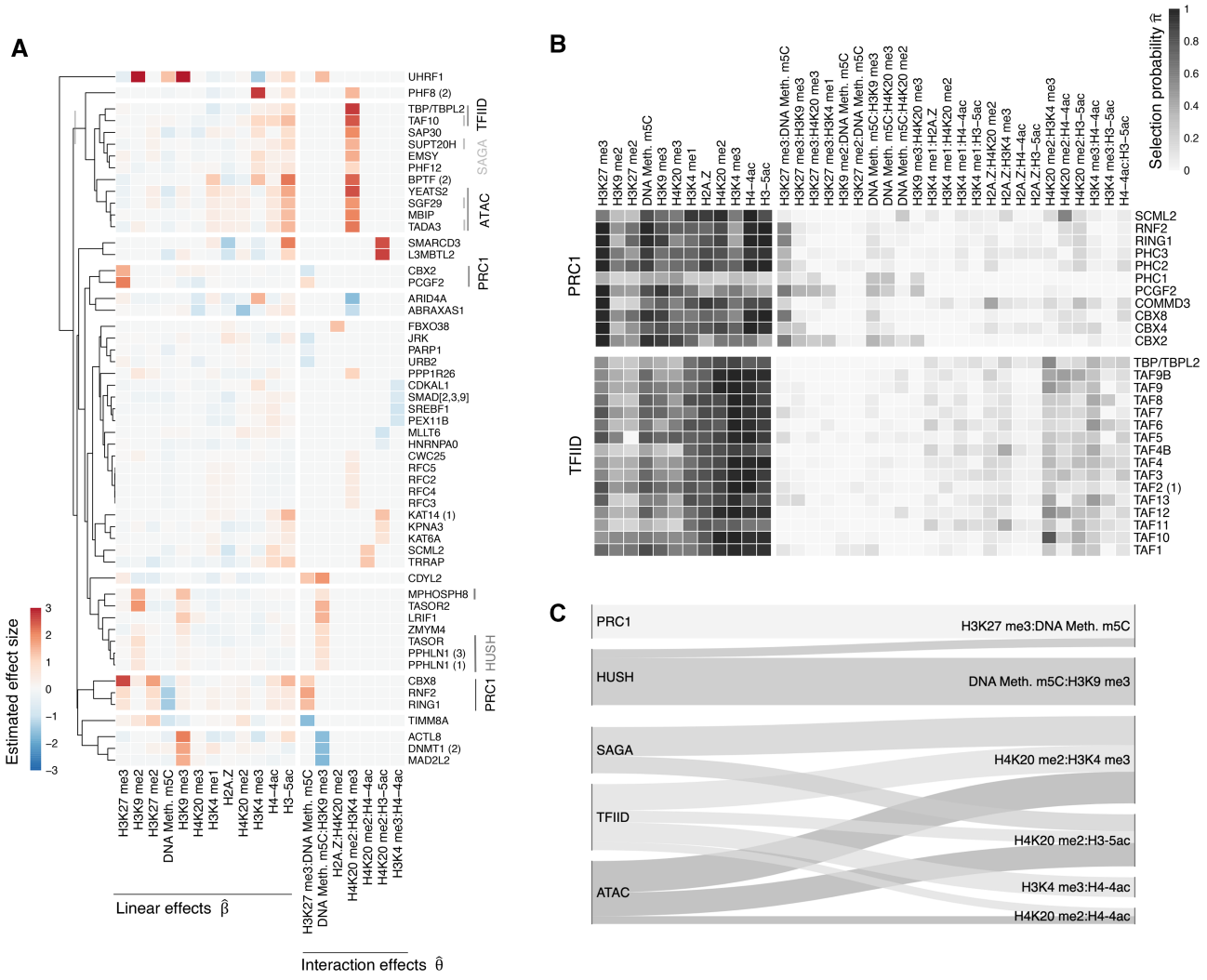


Figure 5. (**A**) Clustered representation of robustly estimated linear and interaction coefficients for $p_I = 55$ proteins for which interaction coefficients have been identified. Proteins belonging to notable protein complexes are highlighted. (**B**) Selection probabilities $\hat{\pi}$ for all proteins in the TFIID and PRC1 complexes. Selection probability plots for all protein complexes are provided in Extended (B). (**C**) Sankey diagram of mean selection probabilities (>0.2) for interaction effects for proteins within a complex.

influencing only the protein FBXO38. Given FBXO38's notably low data density (see Figure 3A, last column), we did not investigate this interaction further. Notably, TAF10 and TBP, which are both part of the Transcription Factor II D (TFIID) complex, respond similarly to the combination of H4K20me2 and H3K4me3. Similarly, members of the PRC1 complex, such as CBX8, RNF2, RING1, CBX2, and PCGF2, are found to respond to the combination of H3K27me3 and DNA methylation m5C (Figure 5A). In addition to examining the effect sizes obtained from `asteRIa`, our approach allows for the interpretation of protein-specific selection probabilities for each individual chromatin modification and each interaction between chromatin modification combinations. The selection probability indicates how stable a feature is in predicting a proteins binding profile across subsamples.

We observe that proteins belonging to the same complex show similar modification selection probability patterns (see Figure 5B for an illustration using the TFIID and PRC1 complex, respectively). These similarities in selection probability patterns justify the exploration of mean selection probabilities over proteins within the same complexes, leading to a more general analysis of how entire complexes respond to interaction effects between chromatin modifications (see Figure 5C).

One major discovery is that the Ada-Two-A-containing (ATAC), Spt-Ada-Gcn5 acetyltransferase (SAGA), and TFIID complexes exhibit multiple combinations of co-operative chromatin modifications that stimulate their binding (Figure 5C). Notably, our analysis reveals several interactions where H4K20me2 is involved, particularly in conjunction with H3K4me3, H3-5ac and H4-4ac.

SAGA is a highly conserved transcriptional co-activator with four distinct functional modules. Its enzymatic functions, including histone acetylation and deubiquitination modules, play crucial roles in chromatin structure and gene expression (81). The ATAC complex, which shares subunits with SAGA, also exhibits histone acetyltransferase activity (81). TFIID, another essential transcription factor, is also a histone acetyltransferase, but additionally recognizes core promoter sequences, recruits the transcription pre-initiation complex, and interacts with SAGA subunits. TFIID contributes to transcription initiation and gene expression by collaborating with cofactors, gene-specific regulators, and chromatin modifica-

tions associated with active genomic regions (82). As such ATAC, SAGA and TFIID are all protein complexes that possess activities that are intricately involved in the process of transcription initiation and that thereby contribute to the regulation of chromatin structure and gene expression.

H4K20me2 is a pervasive modification found on 80% of all histone H4 proteins, marking nearly every nucleosome throughout the genome. Since newly incorporated histone H4 is unmodified at K20 (H4K20me0), the H4K20me2 modification serves as a marker of not yet replicated 'old' chromatin, while H4K20me0 marks newly replicated chromatin during the cell cycle. This modification is used by the DNA repair machinery to determine between different DNA repair pathways in different cell cycle phases (83). The synergistic effect between H4K20me2 and active modifications in recruiting protein complexes associated with transcriptional initiation is therefore surprising and hints to a so far unknown possible function of this modification in the context of promoter regulation.

In contrast, members of the repressive PRC1 and HUSH complexes show a response to an interaction effect between H3K27me3 and DNA methylation m5C and an interaction effect between DNA methylation m5C and H3K9me3, respectively.

The human silencing hub (HUSH) complex is well-established for its role in transcriptionally repressing long interspersed element-1 retrotransposons (L1s) and retroviruses through the modification of histone H3 lysine 9 trimethylation (H3K9me3) (84). Our analysis not only confirms H3K9me3 to be an important binding determinant, in line with previous findings, but it also reveals the involvement of DNA methylation m5C in this regulatory process. Furthermore, our analysis uncovers a previously unreported synergistic interaction between these two modifications, indicating a more complex interplay between H3K9me3 and DNA methylation m5C than previously known.

As a last example, we find that for several members of the PRC1 complex, there is an increased likelihood of responding to an interaction between H3K27me3 and DNA methylation m5C, as previously discussed for CBX8 and the subunits RNF2 and RING1. The PRC1 complex is known to be capable of recognizing H3K27me3 and facilitating transcriptional repression (79), while there are no known associations between the PRC1 complex and methylated DNA. Our results suggest a distinct behavior of DNA methylation and H3K27me3 on regulating the recruitment of the PRC1 complex, with DNA methylation m5C having minimal or even a slightly repulsive effect and H3K27me3 having a binding effect on their own. However, in combination, our analysis reveals an interaction between these two modifications that enhances binding.

Validation of the effects of H3K27me3 and DNA methylation on the binding of CBX8 with ChIP-seq and WGBS data

To validate and compare our findings with orthogonal data sources, we leverage publicly accessible ChIP-seq and WGBS (Whole Genome Bisulfite Sequencing) datasets from the ENCODE project (https://www.encodeproject.org) (40,85–87) and ChIP-Atlas (https://chip-atlas.org) (88–90). Specifically, we design a validation workflow that compares partial correlations from modification co-occurrence patterns with `asteRIa`'s linear and interaction coefficients.

Given the unique design of the MARCS data, our ability to independently validate our discoveries hinges on the availability of ChIP-seq/WGBS experiments that encompass chromatin modifications for which we have identified interaction effects *and* are available in the same cell type. After a comprehensive search, we have identified only the trio of H3K27me3 (ChIP-seq), methylated DNA (WGBS), and the CBX8 protein (ChIP-seq) as the only adequate data set.

As previously described, `asteRIa` reveals a modest interaction effect between H3K27me3 and methylated DNA concerning the binding of CBX8 in the nucleosome binding data. This interaction effect is categorized as 'conflict, dominated by binding b+r+b' (see Figure 4A). We detect a slight repulsive effect of methylated DNA on CBX8 and a recruitment to H3K27me3. Notably, we identify an additional positive interaction effect on CBX8 binding when methylated DNA and H3K27me3 co-occur. Consequently, our results indicate a subtle enhancing effect on CBX8 binding when methylated DNA co-occurs with H3K27me3 (see Figure 4B, lower right corner), resulting in improved predictive accuracy (see Figure 3C).

For this combination, we found matching ChIP-seq and WGBS experiments in A549 (human lung carcinoma epithelial cells), K562 (human myelogenous leukemia cells), and H1 cells (human embryonic stem cells) on ENCODE. Additionally, we use mES cell (mouse embryonic stem cells) data from ChIP-Atlas. For these four cell types, we perform the following analysis workflow: (i) We calculate averages of WGBS data and averages of fold-change values to a reference genome in the ChIP-seq data within consecutive genome bins of 1000 base pairs (bp) with no spacing between bins. We accomplish this by utilizing the 'bins' mode within deeptools on the Galaxy web platform (91), and we ensure the exclusion of blacklisted regions (hg38 for A549, K562 and H1 cells and mm9 for mES cells) during these calculations. (ii) We then conduct a genome-wide analysis of the behavior of H3K27me3 and methylated DNA in CBX8 peak regions. We observe increased H3K27me3 fold-changes and simultaneously decreased DNA methylation values (decreased in K562, A549 and mES; unaffected in H1) in CBX8-bound regions across all cell types under investigation (see Figure 6A and Supplementary Figure S3). This substantiates the (linear) dependencies identified in the `asteRIa` workflow. (iii) We compute Kendall's partial correlations (92) (package version v1.1) of the genome-wide co-occurrence patterns between CBX8, methylated DNA, H3K27me3, and the 'interaction' between methylated DNA and H3K27me3 (i.e. the product of WGBS and H3K23me3 ChIP-seq values, denoted by H3K27me3:WGBS). We use this rank-based correlation coefficient to account for the fact that WGBS and ChIP-seq data are measured and interpreted on different scales. The resulting partial correlations patterns are shown in Figure 6B. The interpretation of the partial correlation coefficients aligns with the coefficients in `asteRIa`'s interaction model. Specifically, the first column of each partial correlation matrix (CBX8) can be understood as follows. The partial correlation between CBX8 and H3K27me3, as well as between CBX8 and the WGBS abundances, reflects the individual (linear) effects of these modifications on CBX8 binding (after conditioning on all other effects). We observe that they are (moderately) positive for CBX8 and H3K27me3 across all cell types, and nega-

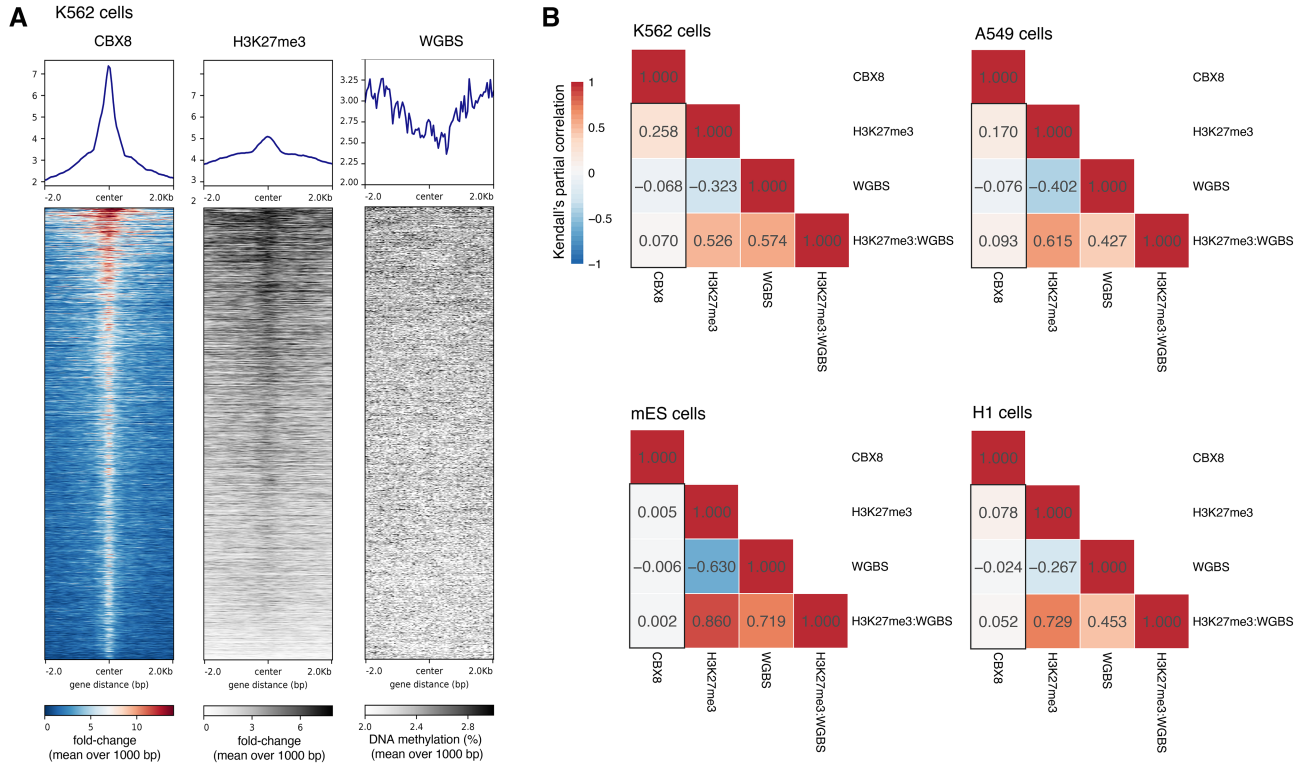


Figure 6. (**A**) Heatmap for score distributions across CBX8 IDR (Irreproducible Discovery Rate) thresholded peaks in K562 cells created with deeptools on the Galaxy web platform. (**B**) Kendall's partial correlation between CBX8, H3K27me3, WGBS data and the product between H3K27me3 and WGBS for A549, K562, H1 and mES cells. The first column in each partial correlation plot recapitulates the main and interaction effects, derived by `asteRIa`. For ENCODE and ChIP-Atlas identifier see caption of Supplementary Figure S3.

tive for CBX8 and WGBS (b+r pattern). Furthermore, the partial correlation between CBX8 and the product of WGBS and H3K27me3 ChIP-seq values represents the additional combinatorial interaction effect, complementing the individual effects. This partial correlation is positive across all cell types, leading to the b+r+b pattern observed in `asteRIa`, Furthermore, it tends to be larger in magnitude than the negative partial correlation between the CBX8 and WGBS data, which also aligns with the `asteRIa` results on the CBX8 nucleosome binding data.

In summary, this analysis provides evidence that `asteRIa`'s estimated main and interaction effects can be recapitulated using other high-throughput experimental data. Moreover, this type of analysis provides a recipe for further validation and invites to perform new ChIP-Seq experiments for other candidate proteins that show evidence of combinatorial interaction effects.

Discussion

While many functions and readers of individual chromatin modifications have been described (15,16), the understanding of how multiple modifications cooperate in recruiting epigenetic regulators has remained largely elusive. To gain insights into these cooperative effects, we have introduced `asteRIa`, a workflow for the robust statistical detection of interaction effects, and applied the workflow to the recently published MARCS nucleosome binding dataset. The MARCS data comprise a library of semi-synthetic di-nucleosomes followed by nucleosome affinity purification with high-throughput quantitative proteomics measurements. Despite MARCS' unique approach to probe the binding behavior of proteins to combinatorial chromatin modifications at a large scale, the imbalanced design matrix and the low sample size pose considerable challenges for consistent statistical interaction estimation. `asteRIa` presents a first step toward identifying robust combinatorial effects between chromatin modifications and is tailored specifically to address these challenges. At its core, `asteRIa` combines the lasso for hierarchical interactions (53) with the complementary pairs stability selection (CPSS) concept (59), and incorporates replicate consistency mechanisms to minimize the identification of spurious interaction effects. We also confirm in a realistic synthetic simulation scenario that combining the interaction model with CPSS reduces the number of spurious effects considerably and leads to more robust results compared to the standard cross-validation procedure (see Supplementary information and Supplementary Figures S1 and S2).

By employing `asteRIa` in conjunction with the MARCS dataset, our study provides the first quantitative framework for the identification of cooperative effects of chromatin modifications on protein binding. We identify a list of 55 epigenetic reader candidates that likely respond to combinatorial modification effects. For the set of 55 proteins we confirmed that interactions enhance predictive performance of protein binding.

To evaluate the validity of `asteRIa`'s data consistency checks, we performed a sensitivity analysis, comparing `asteRIa`'s sign-consistency checks to distance-based consistency filtering and no data filtering. Our analysis demonstrates that requiring data sign-consistency results in the largest number of replicate consistent models and gives the largest set of ro-

bustly identified proteins responding to chromatin modification interactions (see Supplementary Figure S5).

For the 55 proteins identified, we observed consistent responses to these combinations across multiple proteins within the same protein complex, further substantiating the robustness of our findings. The derived candidate set also allowed for a quantitative categorization of different modes of potential chromatin modification interactions.

While our analysis is naturally limited to combinations of chromatin modifications that co-occur in at least one MARCS experiment, we were able to both recapitulate established effects of chromatin modifications on protein binding behavior and discover novel interaction effects between chromatin modifications, potentially promising candidates for future functional analyses. An intriguing finding of our analysis is the discovery of several combinations of cooperative chromatin modifications that elicit responses of the ATAC, SAGA and TFIID complexes. In particular, we identified several interactions involving the H4K20me2 modification, especially in combination with H3K4me3, H3-5ac, and H4-4ac. Another intriguing finding from our analysis is the similar binding profile observed for the proteins DNMT1, MAD2L2 and ACTL8 - all exhibiting a repulsive combinatorial effect in response to DNA methylation m5c and H3K9me3. The function of ACTL8 has not been extensively studied. However, its analogous behavior to MAD2L2 and especially DNMT1 provides an initial hint to a potential function of ACTL8.

We demonstrated the generalizability of our findings beyond a specific cell type or experimental setup by comparing the interaction effect of H3K27me3 and methylated DNA on CBX8, as identified by `asteRIa`, using publicly available ChIP-seq and WGBS data from K562, A549, H1 and mES cells sourced from ENCODE and ChIP-Atlas. Our analysis revealed that, even with the modest improvement in predictive accuracy observed for CBX8 when considering the identified interaction effect between H3K27me3 and methylated DNA, similar patterns are consistently observed in ChIP-seq and WGBS experiments across these diverse cell types.

However, it is important to note that the majority of combinatorial chromatin modification interaction effects identified by `asteRIa`, particularly those characterized by strong interaction effect sizes, are not present in publicly available ChIP-seq datasets. Consequently, we posit that our study serves as a first unbiased attempt to identify chromatin regulators that respond to more than one modification and thereby act as a hypothesis generator, suggesting specific combinations of proteins and chromatin modifications worthy of further investigation in future biological experiments. In particular, we recommend focusing on proteins that exhibit relatively poor predictive accuracy when considering individual chromatin modification effects alone. For instance, proteins like RFC2, RFC3, RFC4 and RFC5 show a substantial enhancement in predictive accuracy when considering the identified interaction effect between H4K20me2 and H3K4me3.

Moreover, `asteRIa` functions as a versatile tool that can be readily updated whenever new nucleosome affinity purification experiments become available. As tools are developed to conduct a greater number of experiments with additional combinations of modifications, our workflow can be conveniently extended to explore more and higher-order interaction effects between chromatin modifications, allowing a more comprehensive understanding of the combinatorial complexity of chromatin modifications.

Even though our statistical workflow has been specifically designed and optimized for the MARCS dataset, its methodology and approach can be broadly applied in scenarios where robust assessment of hierarchical interactions is required, particularly in data-scarce regimes with high levels of noise.

In conclusion, our study provides compelling evidence that large-scale SILAC nucleosome affinity purification data, when combined with `asteRIa`, is a potent resource for generating hypotheses related to epigenetic reader candidates.

Data availability

The `asteRIa` workflow, the processed data, and the code for reproducing all figures and results are available at https://figshare.com/articles/software/asteRIa/25003103 and (partly, without large files) at https://github.com/marastadler/asteRIa.git. The MARCS data is available at https://marcs.helmholtz-munich.de. Mass spectrometry data for MARCS was submitted to the PRIDE database (https://www.ebi.ac.uk/pride/) (accession number: PXD018966).

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We greatly acknowledge Roberto Olayo Alarcon for valuable discussion on the data validation. We acknowledge the ENCODE Consortium and the ENCODE production laboratorys generating the particular datasets used in this study.

ENCODE K562 identifier: ENCFF405HIO, ENCFF687ZGN, ENCFF522HZT, ENCFF459XNY; ENCODE A549 identifier: ENCFF702IOJ, ENCFF081CPV, ENCFF723WVM, ENCFF552VXR; ENCODE H1 identifier: ENCFF345VHG, ENCFF284JDC, ENCFF975NYJ, ENCFF483UZG; ChIP-Atlas mES identifier: SRX426373, SRX006968, DRX001152, SRX5090173.05.

Author contributions: M.S. developed asteRIa, conducted the analysis on the MARCS data, ChIP-Seq and bisulfite data and conducted the analysis on synthetic data. C.L.M. and T.B. supervised the work. M.S. and C.L.M. conceived the statistical workflow. S.L. and T.B. analyzed the results and provided feedback. M.S., C.L.M. and T.B. wrote the manuscript. All authors read and approved the final manuscript.

Funding

M.S. is supported by the Helmholtz Association under the joint research school 'Munich School for Data Science - MUDS'; T.B. is supported by the Helmholtz Association. Funding for open access charge: Helmholtz Gemeinschaft.

Conflict of interest statement

None declared.

References

1. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.

2. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
3. Roadmap Epigenomics Consortium, Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J., *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
4. Garcia,B.A., Pesavento,J.J., Mizzen,C.A. and Kelleher,N.L. (2007) Pervasive combinatorial modification of histone H3 in human cells. *Nat. Methods*, **4**, 487–489.
5. Pesavento,J.J., Bullock,C.R., LeDuc,R.D., Mizzen,C.A. and Kelleher,N.L. (2008) Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J. Biol. Chem.*, **283**, 14927–14937.
6. Shema,E., Jones,D., Shoresh,N., Donohue,L., Ram,O. and Bernstein,B.E. (2016) Single-molecule decoding of combinatorially modified nucleosomes. *Science*, **352**, 717–721.
7. Tvardovskiy,A., Schwämmle,V., Kempf,S.J., Rogowska-Wrzesinska,A. and Jensen,O.N. (2017) Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nucleic Acids Res.*, **45**, 9272–9289.
8. Voigt,P., LeRoy,G., Drury,W.J. III, Zee,B.M., Son,J., Beck,D.B., Young,N.L., Garcia,B.A. and Reinberg,D. (2012) Asymmetrically modified nucleosomes. *Cell*, **151**, 181–193.
9. Young,N.L., DiMaggio,P.A., Plazas-Mayorca,M.D., Baliban,R.C., Floudas,C.A. and Garcia,B.A. (2009) High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics*, **8**, 2266–2284.
10. Li,S., Peng,Y. and Panchenko,A.R. (2022) DNA methylation: Precise modulation of chromatin structure and dynamics. *Curr. Opin. Struct, Biol.*, **75**, 102430.
11. Ruthenburg,A.J., Li,H., Patel,D.J. and Allis,C.D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.*, **8**, 983–994.
12. Turner,B.M. (1993) Decoding the nucleosome. *Cell*, **75**, 5–8.
13. Strahl,B.D. and Allis,C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.
14. Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
15. Greenberg,M. V.C. and Bourc'his,D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.*, **20**, 590–607.
16. Bannister,A.J. and Kouzarides,T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.
17. Li,B., Gogol,M., Carey,M., Lee,D., Seidel,C. and Workman,J.L. (2007) Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*, **316**, 1050–1054.
18. Tsai,W.-W., Wang,Z., Yiu,T.T., Akdemir,K.C., Xia,W., Winter,S., Tsai,C.-Y., Shi,X., Schwarzer,D., Plunkett,W., *et al.* (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature*, **468**, 927–932.
19. Eustermann,S., Yang,J.-C., Law,M.J., Amos,R., Chapman,L.M., Jelinska,C., Garrick,D., Clynes,D., Gibbons,R.J., Rhodes,D., *et al.* (2011) Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat. Struct. Mol. Biol.*, **18**, 777–782.
20. Ruthenburg,A.J., Li,H., Milne,T.A., Dewell,S., McGinty,R.K., Yuen,M., Ueberheide,B., Dou,Y., Muir,T.W., Patel,D.J., *et al.* (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*, **145**, 692–706.
21. Su,W.-P., Hsu,S.-H., Chia,L.-C., Lin,J.-Y., Chang,S.-B., Jiang,Z.-d., Lin,Y.-J., Shih,M.-Y., Chen,Y.-C., Chang,M.-S., *et al.* (2016) Combined interactions of plant homeodomain and chromodomain regulate NuA4 activity at DNA double-strand breaks. *Genetics*, **202**, 77–92.
22. Borgel,J., Tyl,M., Schiller,K., Pusztai,Z., Dooley,C.M., Deng,W., Wooding,C., White,R.J., Warnecke,T., Leonhardt,H., *et al.* (2017) KDM2A integrates DNA and histone modification signals through a CXXC/PHD module and direct interaction with HP1. *Nucleic Acids Res.*, **45**, 1114–1129.
23. Jurkowska,R.Z., Qin,S., Kungulovski,G., Tempel,W., Liu,Y., Bashtrykov,P., Stiefelmaier,J., Jurkowski,T.P., Kudithipudi,S., Weirich,S., *et al.* (2017) H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. *Nat. Commun.*, **8**, 2057.
24. Botuyan,M.V., Lee,J., Ward,I.M., Kim,J.-E., Thompson,J.R., Chen,J. and Mer,G. (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell*, **127**, 1361–1373.
25. Fradet-Turcotte,A., Canny,M.D., Escribano-Díaz,C., Orthwein,A., Leung,C.C.Y., Huang,H., Landry,M.-C., Kitevski-LeBlanc,J., Noordermeer,S.M., Sicheri,F., *et al.* (2013) 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature*, **499**, 50–54.
26. Nakamura,K., Saredi,G., Becker,J.R., Foster,B.M., Nguyen,N.V., Beyer,T.E., Cesa,L.C., Faull,P.A., Lukauskas,S., Frimurer,T., *et al.* (2019) H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat. Cell Biol.*, **21**, 311–318.
27. Sobhian,B., Shao,G., Lilli,D.R., Culhane,A.C., Moreau,L.A., Xia,B., Livingston,D.M. and Greenberg,R.A. (2007) RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science*, **316**, 1198–1202.
28. Kim,H., Chen,J. and Yu,X. (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science*, **316**, 1202–1205.
29. Yan,J., Kim,Y.-S., Yang,X.-P., Li,L.-P., Liao,G., Xia,F. and Jetten,A.M. (2007) The ubiquitin-interacting motif containing protein RAP80 interacts with BRCA1 and functions in DNA damage repair response. *Cancer Res.*, **67**, 6647–6656.
30. Wilson,M.D., Benlekbir,S., Fradet-Turcotte,A., Sherker,A., Julien,J.-P., McEwan,A., Noordermeer,S.M., Sicheri,F., Rubinstein,J.L. and Durocher,D. (2016) The structural basis of modified nucleosome recognition by 53BP1. *Nature*, **536**, 100–103.
31. Hu,Q., Botuyan,M.V., Zhao,D., Cui,G., Mer,E. and Mer,G. (2021) Mechanisms of BRCA1-BARD1 nucleosome recognition and ubiquitylation. *Nature*, **596**, 438–443.
32. Rajakumara,E., Wang,Z., Ma,H., Hu,L., Chen,H., Lin,Y., Guo,R., Wu,F., Li,H., Lan,F., *et al.* (2011) PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell*, **43**, 275–284.
33. Arita,K., Isogai,S., Oda,T., Unoki,M., Sugita,K., Sekiyama,N., Kuwata,K., Hamamoto,R., Tochio,H., Sato,M., *et al.* (2012) Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12950–12955.
34. Rothbart,S.B., Krajewski,K., Nady,N., Tempel,W., Xue,S., Badeaux,A.I., Barsyte-Lovejoy,D., Martinez,J.Y., Bedford,M.T., Fuchs,S.M. and et,al. (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.*, **19**, 1155–1160.
35. Arita,K., Ariyoshi,M., Tochio,H., Nakamura,Y. and Shirakawa,M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
36. Avvakumov,G.V., Walker,J.R., Xue,S., Li,Y., Duan,S., Bronner,C., Arrowsmith,C.H. and Dhe-Paganon,S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
37. Hashimoto,H., Horton,J.R., Zhang,X., Bostick,M., Jacobsen,S.E. and Cheng,X. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, **455**, 826–829.
38. Park,P.J. (2009) ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
39. Bernstein,B.E., Stamatoyannopoulos,J.A., Costello,J.F., Ren,B., Milosavljevic,A., Meissner,A., Kellis,M., Marra,M.A.,

Beaudet,A.L., Ecker,J.R., *et al.* (2010) The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.*, **28**, 1045–1048.
40. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
41. Oki,S., Ohta,T., Shioi,G., Hatanaka,H., Ogasawara,O., Okuda,Y., Kawaji,H., Nakaki,R., Sese,J. and Meno,C. (2018) ChIP-Atlas: a data-mining suite powered by full integration of public Ch IP-seq data. *EMBO Reports*, **19**, e46255.
42. Ernst,J. and Kellis,M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods*, **9**, 215–216.
43. Ernst,J. and Kellis,M. (2017) Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.*, **12**, 2478–2492.
44. Lasserre,J., Chung,H.-R. and Vingron,M. (2013) Finding associations among histone modifications using sparse partial correlation networks. *PLoS Comput. Biol.*, **9**, e1003168.
45. Perner,J. (2015) Bioinformatic approaches for understanding chromatin regulation. PhD thesis.
46. Perner,J., Lasserre,J., Kinkley,S., Vingron,M. and Chung,H.-R. (2014) Inference of interactions between chromatin modifiers and histone modifications: from ChIP-Seq data to chromatin-signaling. *Nucleic Acids Res.*, **42**, 13689–13695.
47. Bartke,T., Vermeulen,M., Xhemalce,B., Robson,S.C., Mann,M. and Kouzarides,T. (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, **143**, 470–484.
48. Nelder,J. (1977) A reformulation of linear models. *J. R. Stat. Soc. Ser. A: Stat. Soc.*, **140**, 48–63.
49. Aiken,L.S., West,S.G. and Reno,R.R. (1991) Multiple Regression: Testing and Interpreting Interactions. Sage.
50. Hamada,M. and Wu,C.J. (1992) Analysis of designed experiments with complex aliasing. *J. Qual. Technol.*, **24**, 130–137.
51. Duncan,R.P. and Kefford,B.J. (2021) Interactions in statistical models: three things to know. *Methods Ecol. Evol.*, **12**, 2287–2297.
52. Simonsohn,U. (2022) Interacting with curves: How to validly test and probe interactions in the real (nonlinear) world.
53. Bien,J., Taylor,J. and Tibshirani,R. (2013) A lasso for hierarchical interactions. *Ann. Stat.*, **41**, 1111.
54. Lim,M. and Hastie,T. (2015) Learning Interactions via Hierarchical Group-Lasso Regularization. *J. Comput. Graph. Stat.*, **24**, 627–654.
55. Hao,N., Feng,Y. and Zhang,H.H. (2018) Model selection for high-dimensional quadratic regression via regularization. *J. Am. Stat. Assoc.*, **113**, 615–625.
56. Peixoto,J.L. (1987) Hierarchical variable selection in polynomial regression models. *Am. Stat.*, **41**, 311–313.
57. Yu,B. (2013) Stability. *Bernoulli*, **19**, 1484–1500.
58. Meinshausen,N. and Bühlmann,P. (2010) Stability Selection. *J. R. Stat. Soc. Ser. B*, **72**, 417–473.
59. Shah,R.D. and Samworth,R.J. (2013) Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, **75**, 55–80.
60. Capraz,T. and Huber,W. (2023) Feature selection by replicate reproducibility and non-redundancy. bioRxiv doi: https://doi.org/10.1101/2023.07.04.547623, 04 July 2023, preprint: not peer reviewed.
61. Lukauskas,S., Tvardovskiy,A., Nguyen,N.V., Stadler,M., Faull,P., Ravnsborg,T., Özdemir Aygenli,B., Dornauer,S., Flynn,H., Lindeboom,R.G., *et al.* (2024) Decoding chromatin states by proteomic profiling of nucleosome readers. *Nature*, **627**, 671–679.
62. Lowary,P. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
63. Muir,T.W. (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.*, **72**, 249–289.
64. Tibshirani,R. (1996) Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, **58**, 267–288.
65. Bien,J. and Tibshirani,R. (2020) hierNet: A Lasso for Hierarchical Interactions, R package version 1.9.
66. Lederer,J. and Müller,C. (2015) Don't fall for tuning parameters: Tuning-free variable selection in high dimensions with the TREX. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. **29**.
67. Wu,Y. and Wang,L. (2020) A survey of tuning parameter selection for high-dimensional regression. *Annu. Rev. Stat. Appl.*, **7**, 209–226.
68. Liu,H., Roeder,K. and Wasserman,L. (2010) Stability approach to regularization selection (stars) for high dimensional graphical models. *Adv. Neu. Inf. Proc. Syst.*, **24**, 1432–1440.
69. Bodinier,B., Filippi,S., Nøst,T.H., Chiquet,J. and Chadeau-Hyam,M. (2023) Automated calibration for stability selection in penalised regression and graphical models. *J. R. Stat. Soc. Ser. C: Appl. Stat.*, **72**, 1375–1393.
70. Maddu,S., Cheeseman,B.L., Sbalzarini,I.F. and Müller,C.L. (2022) Stability selection enables robust learning of differential equations from limited noisy data. *Proc. R. Soc. A*, **478**, 20210916.
71. Fasel,U., Kutz,J.N., Brunton,B.W. and Brunton,S.L. (2022) Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. R. Soc. A*, **478**, 20210904.
72. Hofner,B. and Hothorn,T. (2021) stabs: Stability Selection with Error Control, R package version 0.6-4.
73. Blainey,P., Krzywinski,M. and Altman,N. (2014) Replication: quality is often more important than quantity. *Nat. Methods*, **11**, 879–881.
74. Teng,M., Love,M.I., Davis,C.A., Djebali,S., Dobin,A., Graveley,B.R., Li,S., Mason,C.E., Olson,S., Pervouchine,D., *et al.* (2016) A benchmark for RNA-seq quantification pipelines. *Genome Biol.*, **17**, 74.
75. Ormaza,G., Rodríguez,J.A., de Opakua,A.I., Merino,N., Villate,M., Gorroño,I., Rábano,M., Palmero,I., Vilaseca,M., Kypta,R., *et al.* (2019) The tumor suppressor ING5 is a dimeric, bivalent recognition molecule of the histone H3K4me3 mark. *J. Mol. Biol.*, **431**, 2298–2319.
76. Xie,S. and Qian,C. (2018) The growing complexity of UHRF1-mediated maintenance DNA methylation. *Genes (Basel)*, **9**, 600.
77. Petryk,N., Bultmann,S., Bartke,T. and Defossez,P.A. (2021) Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. *Nucleic Acids Res.*, **49**, 3020–3032.
78. Jeltsch,A. and Jurkowska,R.Z. (2016) Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res.*, **44**, 8556–8575.
79. Geng,Z. and Gao,Z. (2020) Mammalian PRC1 Complexes: Compositional Complexity and Diverse Molecular Mechanisms. *Int. J. Mol. Sci.*, **21**, 8594.
80. Connelly,K.E., Weaver,T.M., Alpsoy,A., Gu,B.X., Musselman,C.A. and Dykhuizen,E.C. (2019) Engagement of DNA and H3K27me3 by the CBX8 chromodomain drives chromatin association. *Nucleic Acids Res.*, **47**, 2289–2305.
81. Cheon,Y., Kim,H., Park,K., *et al.* (2020) Dynamic modules of the coactivator SAGA in eukaryotic transcription. *Experimental & Molecular Medicine*, **52**, 991–1003.
82. Timmers,H. T.M. (2021) SAGA and TFIID: Friends of TBP drifting apart. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.*, **1864**, 194604.
83. Chen,B.-R. and Sleckman,B.P. (2022) The Regulation of DNA End Resection by Chromatin Response to DNA Double Strand Breaks. *Front. Cell Dev. Biol.*, **10**, 932633.
84. Seczynska,M., Bloor,S., Cuesta,S.M., *et al.* (2022) Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature*, **601**, 440–445.
85. Luo,Y., Hitz,B.C., Gabdank,I., Hilton,J.A., Kagda,M.S., Lam,B., Myers,Z., Sud,P., Jou,J., Lin,K., *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.*, **48**, D882–D889.

86. Kagda,M.S., Lam,B., Litton,C., Small,C., Sloan,C.A., Spragins,E., Tanaka,F., Whaling,I., Gabdank,I., Youngworth,I., *et al.* (2023) Data navigation on the ENCODE portal. arXiv doi: https://arxiv.org/abs/2305.00006, 04 May 2023, preprint: not peer reviewed.
87. Hitz,B.C., Lee,J.-W., Jolanki,O., Kagda,M.S., Graham,K., Sud,P., Gabdank,I., Strattan,J.S., Sloan,C.A., Dreszer,T., *et al.* (2023) The ENCODE Uniform Analysis Pipelines. bioRxiv doi: https://doi.org/10.1101/2023.04.04.535623, 06 April 2023, preprint: not peer reviewed.
88. Oki,S. and Ohta,T. (2015) ChIP-Atlas. https://chip-atlas.org.
89. Zou,Z., Ohta,T., Miura,F. and Oki,S. (2022) ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res*, **50**, W175–W182.
90. Oki,S., Ohta,T., Shioi,G., Hatanaka,H., Ogasawara,O., Okuda,Y., Kawaji,H., Nakaki,R., Sese,J. and Meno,C. (2018) ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep*, **19**, e46255.
91. Community,T.G. (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*., **50**, W345–W351.
92. Kim,S. (2015) ppcor: Partial and Semi-Partial (Part) Correlation, R package version 1.1.

Received: January 17, 2024. **Revised:** March 15, 2024. **Editorial Decision:** April 20, 2024. **Accepted:** April 24, 2024