Published online 20 November 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 22* ***13689–13695***
doi: 10.1093/nar/gku1234

Inference of interactions between chromatin modifiers and histone modifications: from ChIP-Seq data to chromatin-signaling

Juliane Perner[1], Julia Lasserre[1], Sarah Kinkley[2], Martin Vingron[1] and Ho-Ryun Chung[2,*]

[1]Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany and [2]Otto-Warburg-Laboratories Epigenomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany

Received January 24, 2014; Revised October 08, 2014; Accepted November 07, 2014

ABSTRACT

Chromatin modifiers and histone modifications are components of a chromatin-signaling network involved in transcription and its regulation. The interactions between chromatin modifiers and histone modifications are often unknown, are based on the analysis of few genes or are studied *in vitro*. Here, we apply computational methods to recover interactions between chromatin modifiers and histone modifications from genome-wide ChIP-Seq data. These interactions provide a high-confidence backbone of the chromatin-signaling network. Many recovered interactions have literature support; others provide hypotheses about yet unknown interactions. We experimentally verified two of these predicted interactions, leading to a link between H4K20me1 and members of the Polycomb Repressive Complexes 1 and 2. Our results suggest that our computationally derived interactions are likely to lead to novel biological insights required to establish the connectivity of the chromatin-signaling network involved in transcription and its regulation.

INTRODUCTION

Transcription and its regulation are facilitated by a complex interplay between various molecular players, such as transcription factors, chromatin modifiers (CMs), histone modifications (HMs) and RNA polymerase II (Pol II). Together these components form a chromatin-signaling network (1) whose signaling activity affects the transcriptional and the chromatin state of a particular genomic region. Thus, it is not surprising that the presence of certain HMs at the promoter or the gene body coincides with the transcriptional status of the corresponding gene (2,3). This close link is further substantiated by the finding that there is even a quantitative relationship between HM levels and the steady-state level of mRNAs (4–6).

HMs are closely linked to the transcriptional process but their functional role in transcription remains largely unknown. On one hand, HMs may modulate the stability of nucleosomes or the chromatin conformation (7) and thereby directly interfere with Pol II recruitment or processivity. On the other hand, HMs may play an indirect role by recruiting CMs to well-defined regions of the genome. Thus, because histones are firmly bound to DNA, HMs may restrict the signaling activity to certain genomic features, such as enhancers and promoters.

The activity of the chromatin-signaling network leads to co-localization of HMs and CMs on the genome, which can be determined by chromatin immunoprecipitation followed by sequencing (ChIP-Seq; (8–10)). Accordingly, clustering HM and CM ChIP-Seq data identifies patterns of co-localized HMs and CMs, which can be associated with genomic features like enhancers and promoters (11). The co-localization pattern specific to, e.g. promoters, unravels those CMs and HMs that constitute the building blocks of the underlying chromatin-signaling network. However, such an analysis is unlikely to identify the specific interactions between CMs and HMs.

Recently, two approaches, one based on Bayesian Network inference (12) and the other on a maximum entropy framework (13), have been proposed to infer chromatin-signaling networks in *Drosophila melanogaster*. Both approaches require discrete data. This, however, involves difficult decisions on optimal decision thresholds. To circumvent these problems we use the ChIP-Seq levels directly and infer a human chromatin-signaling network. We construct this network drawing on two complementary philosophies. We model each HM level as a weighted linear combination of the CM levels and select those CMs that have the most consistent quantitative information about the HM level using Elastic Nets (14). This approach accounts for interactions induced by correlations between CMs, but is not able to remove interactions induced by correlations be-

[*]To whom correspondence should be addressed. Tel: +49 30 8413 1122; Fax: +49 30 8413 1960; Email: chung@molgen.mpg.de

tween HMs. Consequently, we prune the so-derived candidate chromatin-signaling network by computing sparse partial correlation networks (SPCN) (15), which is aimed to identify direct interactions between HMs and CMs accounting for correlations between CMs and HMs.

MATERIALS AND METHODS

ChIP-Seq and gene expression data

The raw HM and CM ChIP-Seq reads were obtained from the SRA Archive (GSE29611 and GSE32509). We merged multiple replicates and mapped uniquely mapping reads to the hg19 genome using Bowtie (16). We counted the number of reads falling into a ±2000 bp window centered at the Transcription Start Sites (TSSs) of all known RefSeq genes (accessed: 19 October 2012). Only promoter regions with at least one sample having a read count larger than the input control were used. The expression data from Cap Analysis of Gene Expression (CAGE) was obtained from the UCSC genome browser (accessed: 14 November 2012; K562CellPapAlnRep1/2.bam and H1hescCellPapAlnRep1/2.bam). The CAGE-counts were averaged over the available replicates.

Read count normalization

We normalized the HM and CM read counts by the following procedure: We estimated the slope of the correlation between the read counts of the sample (S) versus the read counts of the input control (C) (adding a pseudo-count of 1) by the median ($m = \text{median}\,((S+1)/(C+1))$) of the ratio between the two over all promoters. The read counts were then replaced by the enrichment of the sample over the input normalized by the median ($S_{\text{norm}} = (S+1)/(C+1) * 1/m$). This procedure shrinks all the read counts that are highly correlated with the input toward zero. The normalized read counts and average CAGE-counts were log-transformed and scaled to have mean zero and standard deviation one.

Linear regression and regularization using Elastic Nets

We use a combination of computational methods to decipher the chromatin-signaling network as described in the Result section. First, we would like to uncover direct interactions between each HM and the CMs taking into account all other CMs at hand. This can be done by predicting each HM from the CMs using linear regression. Linear regression has been applied in various problems for outcome prediction. Here, apart from achieving good prediction accuracy, we are interested in determining the subset of variables (CMs) that is most useful for the prediction. The latter can be obtained with regularized linear regression methods, which, in contrast to simple linear regression models, impose soft constraints on the number of non-zero coefficients. Moreover, it would be desirable that correlated variables, i.e. equally good predictors, have similar weights. This is especially useful for our case, where we might have sets of CMs that interact with an HM only when being in a complex. For this reasons, we used Elastic Nets (14) as implemented in the *glmnet*-package (17) for R (18). The objective function of Elastic Net (as for simple linear regression) is the Residual Sum of Squares (RSS) criterion: $\text{RSS} = \sum_{j=1}^{N} \left(y_j - \beta_0 - \sum_{i=1}^{p} X_{ij}\beta_i\right)^2$, which is the sum of squared errors that should be minimized. In the Elastic Net this objective function is subjected to the constraint: $(1-\alpha)||\beta||_1 + \alpha||\beta||_2 \leq t$, where $||\beta||_1 = \sum_{i=1}^{p} |\beta_i|$ and $||\beta||_2^2 = \sum_{i=1}^{p} \beta_i^2$, for $\alpha \in [0, 1]$ and some t. The first constraint is based on the L1-norm and forces the coefficients to shrink to 0, thereby favoring sparsity (LASSO-type). The second constraint is based on the L2-norm and favors similar values for the coefficients (Ridge-type), thereby avoiding picking one variable over another when both are redundant. The α-parameter specifies the contribution of each constraint. Throughout the paper we first choose α between 0.01 and 0.99 using 10-fold cross-validation (CV) on each cross-fold. The best α is selected such that the average RSS of the selected α lies within standard deviation of the α having the minimal average RSS. Once α is fixed, the t-parameter is then automatically optimized by the *cv.glmnet*- function in a similar fashion.

We estimated the importance of a CM in predicting a specific HM using Elastic Nets and 10-fold CV. Due to the large number of promoters and due to the smoothing operated by the L2-norm, we expect all coefficients to be non-zero, as the prediction accuracy will increase more with one coefficient than the penalty. However, the L1-norm will enhance the contrast between useful and unuseful variables, and will make the selection for the network representation easier. For the graphical representation of the important CMs we select only those CMs that have an average coefficient that deviates from the average of all coefficients by at least one standard deviation (Supplementary Figure S4).

Partial correlations and the SPCN

We combine the Elastic Net approach described above with SPCN (15), which take into account both HMs and CMs. The SPCN approach is based on the partial correlation coefficient $P(X, Y|Z)$ that gives the correlation coefficient between X and Y after they are controlled for Z. In other words, X and Y are both regressed against the control set Z, and the correlation between their respective residuals $r(X)$ and $r(Y)$ is computed. This allows us to focus on associations that are as direct as possible within the data set at hand. For a data set D, the pairwise partial correlations $P(X, Y|D\backslash\{X, Y\})$ between every pair X and Y, where all other variables $D\backslash\{X, Y\}$ are in the control set, can be efficiently computed by inverting and normalizing the covariance matrix of a data set D.

We build the SPCN on all CMs and HMs (15). In short, we compute the pairwise partial correlation between the ranked ChIP-Seq levels of a CM and an HM conditioned on all other variables (Supplementary Figure S5). Only those edges having a significant, non-zero partial correlation coefficient are retained. Sparseness is introduced in a 10-fold CV scheme which, at the same time, is designed to maintain high accuracy of the resulting (15). For the graphical representation we select only those links from the full SPCN that are between HMs and CMs.

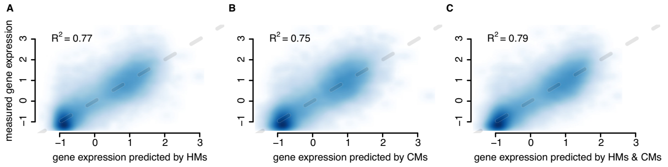

Figure 1. HMs and CMs hold redundant information about gene expression. Scatterplots with the predicted gene expression by HMs (A), CMs (B) and both (C) on the *x*-axis and the measured gene expression (CAGE-tags) on the *y*-axis. The blue color indicates the densities of points, the darker the denser. The gray dashed line indicates identity. In the left upper corner of each plot the coefficient of determination (R^2), i.e. the variance in the gene expression measure explained by the model, is indicated.

Cell lysis and immunoprecipitation (IP)

K562 cells (3×10^6) were lysed in 350 μl cytoskeletal lysis buffer (10 mM PIPES, 100 mM NaCl, 300 mM Sucrose, 3mM $MgCl_{2,}$ 0.1% NP40) for 10 min on ice. The lysate was then centrifuged at 5000 × *g* for 5 min and the supernatant discarded. The pellet was resuspended in 350 μl of chromatin lysis buffer (300 nM NaCl, 50 mM Hepes pH 7.4, 0.5% Igpal, 2.5 mM $MgCl_{2,}$ 5 U Benzonase from Novagene, 1× protease inhibitor cocktail from Roche) for 30 min on ice, with periodic mixing. The lysate was centrifuged at 13 000 × *g* for 10 min and the supernatant collected.

Note that 2 μg of a mouse immunoglobulin G (IgG) control antibody (Diagenode C15200007) or 2 μg of a monoclonal mouse H4K20me1 antibody (Diagenode C15200147) were incubated with 10 μl of magnetic protein G beads (Dynabeads Life Technologies) for 2 h under rotation at 4°C and then washed several times in the IP buffer. A total of 150 μl of nuclease digested chromatin lysate was diluted with dilution buffer (100 nM NaCl, 50 mM Hepes) to 500 μl and incubated with the antibody coated beads for 4 h under rotation at 4°C. The beads were then washed 3× with IP buffer and resuspended in 50 μl of chromatin lysis buffer supplemented with 10 μl of 5× Lammeali buffer. The input and IPs were then heated to 99°C for 10 min prior to loading on a 4–12% gradient gel (Invitrogen). The immunoblot was detected with specific antibodies against H4K420me1 (Abcam ab9057), EZH2 (Epitomics 1940-7) and CBX2 (Abcam ab18968 and Bethyl A302-524A).

RESULTS

HMs and CMs hold redundant information about gene expression

As both, HMs and CMs, are components of the chromatin-signaling network involved in transcription and its regulation, both should contain information about gene expression. To test this idea we used linear regression models to predict gene expression values from HM or CM levels at promoters in the human K562 cell line. The HM levels explain 76% of the variance in gene expression (Figure 1A), which is similar to the results from earlier work (4–6). The CMs capture 75% of the variance in gene expression (Figure 1B). The good predictive performance confirms that both HMs and CMs contain extensive information about gene expression.

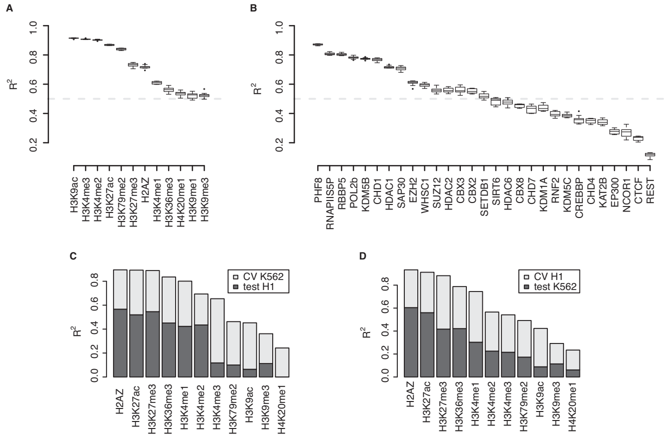

Figure 2. CM levels predict HM levels and *vice versa*. (A and B) Boxplots showing the range of coefficients of determination (R^2) obtained by 10-fold CV using CMs to predict HMs (A) and HMs to predict CMs (B). The boxes indicate the range of R^2 values between the first and third quartile, the horizontal thick line indicates the median and the whiskers extend the range to 1.5-fold the range from the median to the lower and upper hinge of the box. R^2 values outside this range are depicted as points. The dashed gray line indicates an R^2 of 0.5. (C and D) Barplots showing the coefficients of determination (R^2) obtained by training the model with data from the K562 cell line and testing it in the H1 cell line (C) and by training in H1 and testing in K562 (D). The total height of the bar indicates the average R^2 obtained by 10-fold CV in the training cell line, while the darker part indicates the R^2 obtained by testing in the other.

If HMs and CMs reflect the same chromatin-signaling network, both should contain redundant information about gene expression such that combining them should yield only a marginal increase in the predictive power. Indeed, using both, CMs and HMs, improves the explained variance in gene expression only by 3% (4%) compared to using only HMs (CMs) at the expense of a higher model complexity (Figure 1C). Thus, these findings support that CMs and HMs jointly constitute a chromatin-signaling network involved in transcription and its regulation.

CM levels predict HM levels and *vice versa*

Given that CMs and HMs are coupled together by the chromatin-signaling network, the levels of CMs should contain information about the HM levels and *vice versa*. To test this idea we separated the HMs from the CMs and modeled each group of variables using the other. For each HM we built simple linear regression models using 10-fold CV and predicted the HM level based on a weighted combination of the CM levels. For all HMs the models account for at least 50% of the variance in the HM or CM level (Figure 2A). For H3K9ac, H3K4me3, H3K4me2 and H3K27ac the model explains even more than 85% of the variance, which is close to the agreement between biological replicates (Supplementary Figure S1). The high explanatory power of CMs for these four HMs suggests that many CMs interact with these HMs. Indeed, roughly half of the CMs are known to interact with modifications of the H3K4, the H3K9 or the H3K27 residue (Supplementary Table S1).

We repeated this analysis by predicting CM levels from a linear combination of HM levels. For about half of the CMs the models account for over 50% of the variance (Figure 2B). Thus, for those well-predicted CMs the HMs in the data set cover the bulk of the recruitment mechanisms and enzymatic targets.

Under the assumption that the chromatin-signaling network is a common mechanism underlying transcription and its regulation, we expect that the contribution of a CM to the prediction of an HM in one cell type is similar in another cell type. Thus, given the regression model trained on the data from K562 cells we should be able to predict the HM levels in another cell type. We tested this using ChIP-Seq data for 14 CMs and 11 HMs in human embryonic stem cells (H1) that were also measured in the K562 cells. Indeed, the regression models learned from the data available for both cell types show good agreement (Figure 2C and D). The lower performance of the models when tested on the data from a different cell type is expected due to biological variation, e.g. different expression levels of the CMs. Thus, the quantitative effects of the interactions within the chromatin-signaling network are preserved suggesting a cell-type independent chromatin-signaling network involved in transcription and its regulation.

From co-localization to interactions

We have shown that CM levels accurately predict HM levels and *vice versa*. We argued that the prediction accuracy depends on the expression and biochemical activities of the available CMs toward the HMs. To identify CM-HM pairs that are likely to interact with each other, we selected those CMs that contributed most to the prediction of an HM level. The most straightforward approach is to select those CM-HM pairs that show the highest pair-wise correlation. This has been done in recent work by clustering HMs and CMs into correlated subgroups based on their cooccupancy patterns (11).

There are groups of HMs and CMs that exhibit very high pairwise correlation (Supplementary Figure S2), suggesting that they are functionally related. However, within these groups no internal structure is visible, rendering an identification of interactions between the group members difficult. As CMs and HMs constitute a chromatin-signaling network, this high correlation is expected due to direct interactions between its components. However, high correlations could also be induced by other factors connecting the respective CM and HM. In general, the identity of these additional factors is not known, but we can account for those factors that are present in the data set. Thus, we want to recover interactions between CMs and HMs that cannot be 'explained away' by other variables in the data set.

We recovered these interactions by applying a two-step procedure (see Materials and Methods). First, we used a regularized regression technique called 'Elastic Net', where the CMs are used to predict HMs, to select only CMs that are informative for the prediction of a HM. Moreover, in case of groups of strongly correlated CMs the members of these groups tend to remain all in the model or are removed together (14). This approach accounts for possible interactions induced by correlations within the CMs but does not take into account correlations between the HMs. This indicates that highly correlated HMs might be predicted by similar sets of CMs, while only certain CMs actually interact with specific HMs. Second, to remedy this situation we used a technique called 'SPCN' (15), where the pairwise rank correlation between a CM and a HM is conditioned on all other variables in the data set. This method takes into account the correlation structure of both, CMs and HMs, and is conservative in proposing interactions. As a consequence, in groups of strongly correlated CMs and/or HMs, interactions may be explained away by individual members of the group (15). Thus, in the SPCN framework an identified interaction is likely to represent a direct interaction in the sense that it cannot be explained by other variables in the data set. However, the failure to recover an interaction does not imply the absence of a biologically meaningful interaction. Within the SPCN framework some interactions between CMs and HMs arise from logical dependencies induced by sharing a common target. Thus, to recover interactions, we establish first the necessary condition that a CM is consistently highly predictive for an HM level by the Elastic Net approach and in a second step we prune those interactions that may be induced by correlations between the HMs using the SPCN approach. Thus, we focus only on interactions that are recovered by both methods. These interactions may originate from a direct function of the CM in setting, erasing or binding the HM but also from indirect interactions via unobserved CMs.

Distinct sets of CMs associate with each HM

In the Elastic Net network each HM is linked to a different set of CMs indicating the different specificities of the CMs toward the individual HMs (Supplementary Figure S3A). The densest part of the network connects several CMs to the HMs H3K4me3, H3K9ac, H3K27ac and H3K79me2. The effect of the SPCN framework becomes most apparent on this dense cluster (compare Supplementary Figure S3A and B) where most of the interactions are resolved. It is important to note that the lack of a predicted interaction by the SPCN is not sufficient evidence to prove the absence of a biological relevant interaction. However, an interaction recovered by both approaches is likely to represent a true interaction between the CM and the HM.

The chromatin-signaling network recovers biologically meaningful interactions

Many of the interactions identified by both Elastic Net and SPCN (Figure 3) are supported by published experimental evidence (Supplementary Note S1 and Supplementary Table S2), strengthening our confidence in the recovered interactions. For example, H3K27me3 has a positive interaction with members of the Polycomb Repressive Complex (PRC) 1 (CBX2 and CBX8; (19,20)) and members of the PRC2 (EZH2 and SUZ12 (20)), as well as a negative interaction with Pol II phosphorylated at serine 5 (RNAPIIS5P).

The interaction between H3K27me3 and EZH2 is direct, because EZH2 sets H3K27me3 (21–24). The interaction between H3K27me3 and SUZ12 may be direct, because it cannot be 'explained away' by EZH2. However, EED which forms a trimeric complex together with SUZ12 and EZH2 binds H3K27me3 directly (25), and most likely explains the interaction between H3K27me3 and SUZ12 (26). The interaction between the PRC1 components CBX2 and CBX8

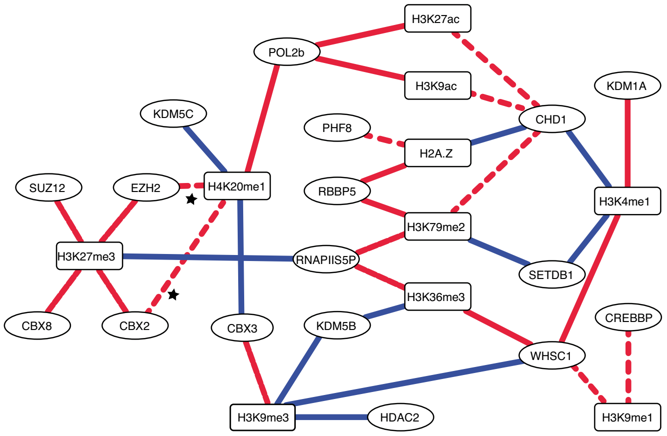

Figure 3. Chromatin-signalling network. Graphical representation of the interactions between CMs (circles) and HMs (squares). Shown are the interactions recovered both by the Elastic Net and SPCN approach. Red lines indicate positive and blue lines negative interactions. The continuous lines indicate interactions with supporting evidence in the literature, while the dashed lines indicate interactions without supporting evidence. The stars indicate the two interactions confirmed in this study.

and H3K27me3 is direct, because CBX2 and CBX8 bind to H3K27me3 (23).

A negative interaction connects RNAPIIS5P and H3K27me3 in our network. The serine 5 phosphorylation of Pol II is mediated by the pre-initiation complex factor TFIIH (27–30) and is present in the initiating and the elongating form of Pol II (31). A role of H3K27me3 is to repress transcription, which is accompanied by low levels of initiating and/or elongating Pol II marked by serine 5 phosphorylation, explaining the negative interaction with RNAPIIS5P in our network.

Using H3K27me3 as an example, these results show that our approach identifies biological meaningful interactions between the members of PRCs and H3K27me3. If we did not have any prior information about the interactions between H3K27me3 and PRC, we would conclude that members of the PRCs are involved in setting and/or reading H3K27me3 and that high levels of H3K27me3 are incompatible with high levels of Pol II phosphorylated at serine 5.

In summary, 19 (58%) of the 33 identified interactions are supported by experimental evidence as collected from the literature, showing a direct interaction or involving only one unobserved, additional protein (Supplementary Note S1 and Supplementary Table S2). Our predictions complement the experimental evidence obtained either *in vitro* or by using one or few genes as model system. In addition, as we used ChIP-Seq data the inferred interactions between CMs and HMs provide evidence for the interactions *in vivo* and genome-wide. Finally, we provide testable hypotheses regarding novel interactions, which may be instrumental to define chromatin signaling and its impact on transcription.

Verification of two predicted interactions links H4K20me1 to Polycomb-mediated repression

Two predicted interactions involve the HM H4K20me1 and CBX2 and EZH2, which are components of PRC1 and 2, respectively. In both cases the interaction is positive suggesting that CBX2 and EZH2 are involved in setting, stabilizing and/or reading H4K20me1.

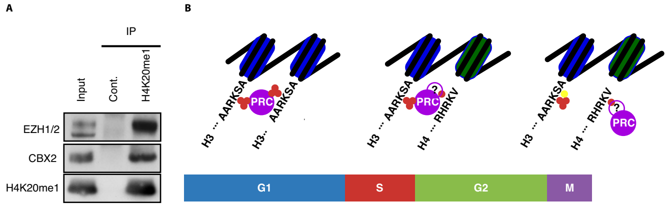

Figure 4. Verification of two predicted interactions links H4K20me1 to Polycomb-mediated repression. (A) H4K20me1 co-IP: K562 cells were immunoprecipitated with a control IgG and an H4K20me1 specific antibody and analyzed by immunoblot for the co-precipitation of EZH2, CBX2 and H4K20me1. 10% input was loaded. (B) Model of the role of H4K20me1 in the maintenance of Polycomb-mediated repression through the cell-cycle. During the G1-phase PRCs bind to H3K27me3 (indicated by three red circles) on two adjacent nucleosomes. During S-phase one of the two nucleosomes is replaced by a new one, which acquires H4K20me1 (indicated by a single red circle). After replication PRCs bind to H3K27me3 on the old nucleosome (in blue) and H4K20me1 on the new (in green), possibly via a yet unknown factor (indicated by the violet circle with the question mark). In M-phase, serine 28 gets phosphorylated (indicated by a yellow circle), which prevents PRCs from binding. PRCs are maintained on chromatin by their interaction with H4K20me1.

Given the biochemical properties of CBX2 and EZH2, a role in setting or stabilizing H4K20me1 seems unlikely. However, CBX2 and EZH2 may directly or indirectly bind to H4K20me1. To test the latter possibility, we performed an IP against H4K20me1 and probed for the presence of CBX2 and EZH2 (Figure 4A). The presence of a positive signal of CBX2 and EZH2 in the H4K20me1 IP and the absence in the control IgG IP suggests that both proteins interact with H4K20me1. Our results are in line with the idea that H4K20me1 is linked to Polycomb-mediated repression by interacting with PRCs 1 and 2.

DISCUSSION

Taken together, we propose a novel computational approach to enrich for potential direct interactions linking CMs and HMs within a chromatin-signaling network. We have applied this approach to the most comprehensive set of CMs and HMs in human cells and identified interactions between the CMs and HMs. Furthermore, we have demonstrated that at least two of the predicted but yet unknown interactions can be verified by experimental means. These verified interactions provide an unexplored link between Polycomb-mediated repression and H4K20me1.

Analyzing the pairwise correlation patterns between the levels of CMs and HMs identifies groups of CMs and HMs, which are likely to constitute the building blocks of a chromatin-signaling network. However, unraveling specific interactions between the group members by focusing only on the pairwise correlations is difficult. This difficulty arises from the propagation of correlations along the direct interactions of the network components. For example, H3K27me3 is set by EZH2, which is in a complex with SUZ12 and EED, which itself binds to H3K27me3 (Figure 5A). Thus, the H3K27me3 ChIP-Seq levels correlate with those of EZH2, EED and SUZ12. However, only in the case

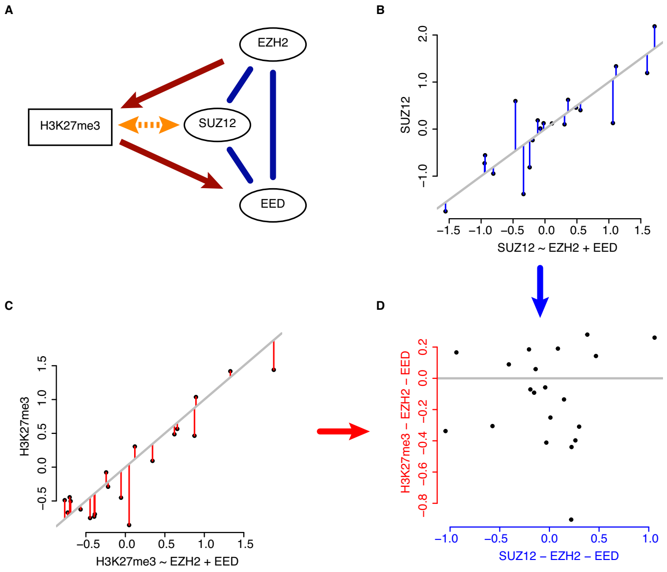

Figure 5. From correlation to direct interactions. (A) Model of the interaction of the PRC2 trimeric complex (EZH2, SUZ12 and EED) with H3K27me3. The blue lines indicate protein-protein interactions. The red arrows indicate direct causal interactions, with EZH2 setting H3K27me3 and EED reading H3K27me3. The orange double-headed arrow indicates a correlation between SUZ12 and H3K27me3 induced by either EZH2 and/or EED. (B–D) Toy example of the de-correlation action of multivariate regression. Modeling of, e.g. H3K27me3 levels by a linear combination of EZH2, EED and SUZ12 leads to an estimate of the influence of SUZ12 independent on the influence of EZH2 and EED. This is achieved by modeling SUZ12 (B) and H3K27me3 levels (C) by a linear combination of EZH2 and EED. The predictions of these models are subtracted from the actual SUZ12 and H3K27me3 levels (residuals, depicted by blue (SUZ12) and red (H3K27me3) vertical lines). The residuals of SUZ12 after incorporating the information of EZH2 and EED are used to predict the corresponding residuals of H3K27me3 (D), which in this case fails because there is no information of SUZ12 on H3K27me3 left after considering EZH2 and EED levels.

of EZH2 and EED this is due to a direct interaction with H3K27me3. To remedy such a situation, in our example we need to ask how much more information SUZ12 provides on H3K27me3 given the information provided already by EZH2 and EED. We achieve this by modeling H3K27me3 levels as a weighted linear combination of EZH2, EED and SUZ12 levels. Here, the correlations between EZH2, EED and SUZ12 are taken into account, such that we obtain a weight for SUZ12, which corresponds to the remaining information that SUZ12 has on H3K27me3 after the information of EZH2 and EED on SUZ12 (Figure 5B) and H3K27me3 (Figure 5C) has been subtracted (Figure 5D).

We use this mathematical framework to explain away indirect interactions and thus to obtain the most direct interactions given the data. This implies that the uncovered interactions may change if additional information is added. For example, we had only data for H3K27me3, EZH2 and SUZ12, but lacked data for EED. Our analysis uncovers an interaction between H3K27me3 and EZH2, which has been shown to set H3K27me3 (21–24). We also identified an interaction between H3K27me3 and SUZ12. The latter interaction is independent of EZH2, but may be dependent on the unobserved EED, such that the addition of EED to the data set will remove the indirect interaction between SUZ12 and H3K27me3.

Within this mathematical framework, we have shown that HM levels are accurately predicted by CM levels and *vice versa* (Figure 2), suggesting a close relationship between CMs and HMs. Given the high predictive power, we are confident to take the weights of the Elastic Net as evidence for an interaction between a HM and a CM. By combining Elastic Net and SPCN we further eliminated indirect interactions moving closer toward a mechanistic understanding of the interactions between HMs and CMs (Figure 3).

These interactions should not be confused with causal interactions. Inference of causality from data requires perturbation experiments as discussed extensively in the literature (32). In our setting such experiments are notoriously difficult to perform, because perturbations of CMs usually either lead to pleiotropic effects, including cell death (33), or are buffered by redundant mechanisms (34,35). Additionally, manipulation of the histones, i.e. single amino acid substitutions, is not feasible in most organisms except for yeast (36) and Drosophila (37,38).

Our analysis predicted many interactions between CMs and HMs, of which many are supported by the literature (Supplementary Note S1 and Supplementary Table S2). Others provide novel hypotheses about yet unknown interactions between CMs and HMs, which are amenable to experimental verification. To demonstrate this, we validated two interactions involving the HM H4K20me1 and the CMs EZH2 and CBX2 by co-IP (Figure 4A). These results link H4K20me1 to Polycomb-mediated repression by PRCs 1 and 2, which may form a mechanistic basis for the maintenance of Polycomb-repression through the cell cycle.

The progression of cells through the cell cycle constitutes two challenges for the maintenance of Polycomb-mediated repression: (i) During DNA replication old and newly synthesized nucleosomes are randomly distributed to the daughter strands (39). This leads to an effective dilution of H3K27me3-bearing nucleosomes by half. (ii) During mitosis HMs, chromatin composition and structure change dramatically, rendering the proper transmission of H3K27me3 difficult.

H4K20me1 is tightly regulated during the cell cycle. It starts accumulating during S-phase and attains high levels during mitosis (40). Given this pattern, H4K20me1 may play an important role in maintaining PRCs at their target sites throughout the replicative and mitotic challenges by recruiting PRCs 1 and 2 to regions with old H3K27me3- and new H4K20me1-bearing nucleosomes (Figure 4B).

Taken together, we provide a chromatin-signaling network in K562 cells that links CMs to specific HMs. Our approach aims at high specificity and sacrifices sensitivity leading to high-confidence interactions. We verified two yet unknown interactions, which gives rise to novel biological insights about the interplay between Polycomb-mediated repression and H4K20me1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

EU-FP7 'BLUEPRINT' [282510]; Bundesministerium für Bildung und Forschung 'Deutsches Epigenom Programm (DEEP)' [01KU1216C]; Deutsche Forschungsgemeinschaft [SFB618]. Funding for open access charge: Max Planck Society.

Conflict of interest statement. None declared.

REFERENCES

1. Schreiber,S.L. and Bernstein,B.E. (2002) Signaling network model of chromatin. *Cell*, **111**, 771–778.
2. Zhou,V.W., Goren,A. and Bernstein,B.E. (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.*, **12**, 7–18.
3. Li,B., Carey,M. and Workman,J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.
4. Kumar,V., Muratani,M., Rayan,N.A., Kraus,P., Lufkin,T., Ng,H.H. and Prabhakar,S. (2013) Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.*, **31**, 615–622.
5. Dong,X., Greven,M.C., Kundaje,A., Djebali,S., Brown,J.B., Cheng,C., Gingeras,T.R., Gerstein,M., Guigó,R., Birney,E. *et al.* (2012) Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.*, **13**, R53.
6. Karlić,R., Chung,H.-R., Lasserre,J., Vlahoviček,K. and Vingron,M. (2010) Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2926–2931.
7. Shogren-Knaak,M., Ishii,H., Sun,J.-M., Pazin,M.J., Davie,J.R. and Peterson,C.L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, **311**, 844–847.
8. Robertson,G., Hirst,M., Bainbridge,M., Bilenky,M., Zhao,Y., Zeng,T., Euskirchen,G., Bernier,B., Varhol,R., Delaney,A. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*, **4**, 651–657.
9. Johnson,D.S., Mortazavi,A., Myers,R.M. and Wold,B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497–1502.
10. 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.
11. Ram,O., Goren,A., Amit,I., Shoresh,N., Yosef,N., Ernst,J., Kellis,M., Gymrek,M., Issner,R., Coyne,M. *et al.* (2011) Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, **147**, 1628–1639.
12. van Bemmel,J.G., Filion,G.J., Rosado,A., Talhout,W., de Haas,M., van Welsem,T., van Leeuwen,F. and van Steensel,B. (2013) A network model of the molecular organization of chromatin in Drosophila. *Mol. Cell*, **49**, 759–771.
13. Zhou,J. and Troyanskaya,O.G. (2014) Global quantitative modeling of chromatin factor interactions. *PLoS Comput. Biol.*, **10**, e1003525.
14. Zou,H. and Hastie,T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 768–768.
15. Lasserre,J., Chung,H.-R. and Vingron,M. (2013) Finding associations among histone modifications using sparse partial correlation networks. *PLoS Comput. Biol.*, **9**, e1003168.
16. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
17. Friedman,J., Hastie,T. and Tibshirani,R. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, **33**, 1–22.
18. R Core Team. (2014) R:A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.
19. Simon,J.A. and Kingston,R.E. (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*, **10**, 697–708.
20. Margueron,R. and Reinberg,D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343–349.
21. Müller,J., Hart,C.M., Francis,N.J., Vargas,M.L., Sengupta,A., Wild,B., Miller,E.L., O'Connor,M.B., Kingston,R.E. and Simon,J.A. (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, **111**, 197–208.
22. Kuzmichev,A., Nishioka,K., Erdjument-Bromage,H., Tempst,P. and Reinberg,D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.*, **16**, 2893–2905.
23. Cao,R., Wang,L., Wang,H., Xia,L., Erdjument-Bromage,H., Tempst,P., Jones,R.S. and Zhang,Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039–1043.
24. Czermin,B., Melfi,R., McCabe,D., Seitz,V., Imhof,A. and Pirrotta,V. (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, **111**, 185–196.
25. Hansen,K.H., Bracken,A.P., Pasini,D., Dietrich,N., Gehani,S.S., Monrad,A., Rappsilber,J., Lerdrup,M. and Helin,K. (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.*, **10**, 1291–1300.
26. Margueron,R., Justin,N., Ohno,K., Sharpe,M.L., Son,J., Drury,W.J., Voigt,P., Martin,S.R., Taylor,W.R., De Marco,V. *et al.* (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, **461**, 762–767.
27. Helenius,K., Yang,Y., Tselykh,T.V., Pessa,H.K.J., Frilander,M.J. and Mäkelä,T.P. (2011) Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. *Nucleic Acids Res.*, **39**, 5025–5035.
28. Serizawa,H., Mäkelä,T.P., Conaway,J.W., Conaway,R.C., Weinberg,R.A. and Young,R.A. (1995) Association of Cdk-activating kinase subunits with transcription factor TFIIH. *Nature*, **374**, 280–282.
29. Shiekhattar,R., Mermelstein,F., Fisher,R.P., Drapkin,R., Dynlacht,B., Wessling,H.C., Morgan,D.O. and Reinberg,D. (1995) Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, **374**, 283–287.
30. Roy,R., Adamczewski,J.P., Seroz,T., Vermeulen,W., Tassan,J.P., Schaeffer,L., Nigg,E.A., Hoeijmakers,J.H. and Egly,J.M. (1994) The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell*, **79**, 1093–1101.
31. Bataille,A.R., Jeronimo,C., Jacques,P.-É., Laramée,L., Fortin,M.-È., Forest,A., Bergeron,M., Hanes,S.D. and Robert,F. (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol. Cell*, **45**, 158–170.
32. Koller,D. and Friedman,N. (2009) *Probabilistic Graphical Models.* MIT Press, Cambridge.
33. O'Carroll,D., Erhardt,S., Pagani,M., Barton,S.C., Surani,M.A. and Jenuwein,T. (2001) The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell. Biol.*, **21**, 4330–4336.
34. Montgomery,R.L., Davis,C.A., Potthoff,M.J., Haberland,M., Fielitz,J., Qi,X., Hill,J.A., Richardson,J.A. and Olson,E.N. (2007) Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.*, **21**, 1790–1802.
35. Jurkin,J., Zupkovitz,G., Lagger,S., Grausenburger,R., Hagelkruys,A., Kenner,L. and Seiser,C. (2011) Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis. *Cell Cycle*, **10**, 406–412.
36. Park,E.C. and Szostak,J.W. (1990) Point mutations in the yeast histone H4 gene prevent silencing of the silent mating type locus HML. *Mol. Cell. Biol.*, **10**, 4932–4934.
37. Pengelly,A.R., Copur,Ö., Jäckle,H., Herzig,A. and Müller,J. (2013) A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*, **339**, 698–699.
38. Günesdogan,U., Jäckle,H. and Herzig,A. (2010) A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. *EMBO Rep.*, **11**, 772–776.
39. Jackson,V. and Chalkley,R. (1985) Histone segregation on replicating chromatin. *Biochemistry*, **24**, 6930–6938.
40. Pesavento,J.J., Yang,H., Kelleher,N.L. and Mizzen,C.A. (2008) Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol. Cell. Biol.*, **28**, 468–486.