The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset

Christian Feller[1], Matthias Prestel[1], Holger Hartmann[2], Tobias Straub[1], Johannes Söding[2] and Peter B. Becker[1],*

[1]Adolf-Butenandt-Institute and Center for Integrated Protein Science of the Ludwig-Maximilians-University, Schillerstraße 44, 80336 München, Germany and [2]Gene Center Munich, Department of Chemistry and Biochemistry, Ludwig-Maximilians-University, Feodor-Lynen-Straße 25, 81377 Munich, Germany

ABSTRACT

The MOF (males absent on the first)-containing NSL (non-specific lethal) complex binds to a subset of active promoters in *Drosophila melanogaster* and is thought to contribute to proper gene expression. The determinants that target NSL to specific promoters and the circumstances in which the complex engages in regulating transcription are currently unknown. Here, we show that the NSL complex primarily targets active promoters and in particular housekeeping genes, at which it colocalizes with the chromatin remodeler NURF (nucleosome remodeling factor) and the histone methyltransferase Trithorax. However, only a subset of housekeeping genes associated with NSL are actually activated by it. Our analyses reveal that these NSL-activated promoters are depleted of certain insulator binding proteins and are enriched for the core promoter motif 'Ohler 5'. Based on these results, it is possible to predict whether the NSL complex is likely to regulate a particular promoter. We conclude that the regulatory capacity of the NSL complex is highly context-dependent. Activation by the NSL complex requires a particular promoter architecture defined by combinations of chromatin regulators and core promoter motifs.

INTRODUCTION

Eukaryotic organisms consist of a diversified set of highly specialized cells. Their individual identities are determined by the appropriate expression of cell-specific genes while a battery of genes that are expressed in all cells maintain general ('housekeeping') functions. Gene expression at the transcriptional level is governed by an intricate interplay between transcription regulators and local chromatin organization. In general, the packaging of genomes into chromatin brings about a default state of repression, as nucleosome assembly constantly competes with transcription factors for promoter binding sites. Overcoming this repression requires a concerted action of various chromatin-modifying principles. These include ATP-dependent nucleosome remodeling factors, which are targeted to specific loci by DNA-bound proteins and post-translational histone marks where they reorganize nucleosomes to facilitate transcription (1). An example for such an activity in *Drosophila melanogaster* is NURF (nucleosome remodeling factor), whose large regulatory subunit, NURF301, interacts with a diversity of transcription factors and methyl marks on lysine 4 of histone H3 (H3K4me3) (2,3) (and references therein). NURF has also been reported to bind to acetylated lysine 16 of histone H4 (H4K16ac) (2), a nucleosome modification that prevents nucleosome–nucleosome interactions that promote the folding of the nucleosomal fiber into more compact structures. The acetyltransferase MOF (males absent on the first) is a major enzyme responsible for this modification in both, *Drosophila* and mammalian cells (4,5).

MOF is best known for its key role in the *Drosophila* dosage compensation process. It is a subunit of the dosage compensation complex [DCC, also known as male-specific lethal (MSL) complex], which brings about the 2-fold transcriptional activation of genes on the single male X chromosome to equalize expression with the corresponding genes transcribed from the two female X chromosomes (6). The DCC is constituted only in male flies and the five protein components, MSL1, MSL2, MSL3, maleless (MLE) and MOF, as well as the non-coding *roX* RNAs are essential for male viability. According to the current model, the DCC recruits MOF to the transcribed regions of X-chromosomal genes. Subsequent acetylation of

---

H4K16 renders chromatin more accessible and potentially facilitates transcriptional elongation (7,8).

With the exception of MSL2, all DCC protein subunits are also expressed in female flies, and therefore also serve more general, yet barely understood functions (9). For example, the acetyltransferase MOF appears to be involved in more global transcription regulation as it has recently been found in an alternative complex together with MCRS2, the WD40-repeat protein WDS (will-die-slowly), NSL1, NSL2, NSL3 and the plant homeo domain (PHD) protein MBD-R2 (10–12). With reference to the dosage compensation 'MSL complex', this alternative MOF-containing assembly was termed 'NSL complex' (for 'non-specific lethal'), as its subunits are essential in both sexes (10). The incorporation of MOF into either the DCC or the NSL complex is determined by association of MOF with the PEHE domains of the respective MSL1 or NSL1 subunits (10). Genome-wide mapping by chromatin immunoprecipitation (ChIP) coupled to DNA microarrays (ChIP-chip) identified MOF binding sites at many, but not all active promoters in male and female cells (13). Subsequent studies revealed that MBD-R2 colocalizes with MOF at many active promoters in both sexes, suggesting that the NSL complex recruits MOF to these sites (12). This is compatible with a recent ChIP-Seq study (ChIP DNA analyzed by massive parallel sequencing), which found MCRS2 and NSL1 peaks at promoters in mixed-sex 3rd instar larval salivary glands (11).

In male cells the association of MOF with NSL subunits is in competition with its incorporation into the DCC, which redirects it to the transcribed regions of X chromosomal genes (12). However, key aspects of MOF's targeting in the context of the NSL complex are unclear. What determines the binding of the NSL complex to only a subset of the active promoters? The available data also are ambiguous when it comes to the role of the NSL complex; does it activate or repress target genes, or perhaps both? Ablating the NSL subunit MBD-R2 in male embryonic cells resulted in a reduced expression of many MBD-R2 target genes (12). In contrast, a similar fraction of genes was found up- and downregulated when MBD-R2 and NSL3 were depleted in 3rd instar salivary glands (11).

In this study, we created novel data sets and analyzed existing ones to compare functional interactions of NSL subunits in different developmental tissues to better define the targets of the NSL complex. We systematically explored the common properties of the NSL target genes, searching for colocalizing chromatin factors and prevalent sequence motifs in target promoters. We traced the NSL complex through monitoring the NSL1 subunit and found that it preferentially binds to promoters of housekeeping genes, which are also approached by the chromatin remodeler NURF and the methyltransferase Trithorax. There, NSL1 binding correlates best with the core promoter element DNA replication-related element (DRE). However, only a defined fraction of NSL1-bound genes are actually regulated by the complex. Those promoters are depleted for insulator proteins and are enriched for the E-box-derived promoter motif 'Ohler 5'. Our analysis provides a functional classification of housekeeping genes according to their NSL coregulator requirements.

MATERIALS AND METHODS

Generation of the NSL1 antibody

A cDNA fragment corresponding to NSL1 amino acids 1271–1550 was Polymerase Chain Reaction (PCR) amplified from cDNA clone #LP09056 (*Drosophila* Genomics Resource Center; see Table 1) and cloned into the pGEX2TKN. The N-terminally glutathion-S-transferase (GST)-tagged NSL1 fragment was expressed in *Escherichia coli* BL21, purified on glutathione beads and used to raise antibodies in rabbit by a commercial supplier.

RNA interference in S2 cells, immunoblotting and indirect immunofluorescence

Male *Drosophila* S2 cell cultivation and RNA interference (RNAi) were carried out as described before (12). Briefly, $1.5 \times 10 \mathrm{e}^6$ cells were incubated with $10\,\mu\mathrm{g}$ dsRNA targeted against NSL1 or GST as a control. Primer sequences used for dsRNA production are listed in Table 1. Cells were harvested after 6 or 7 days and processed for RNA (see below) and protein. For every $10\,\mathrm{e}^6$ cells, cells were lysed for $10\,\mathrm{min}$ in $100\,\mu\mathrm{l}$ of N-buffer [$15\,\mathrm{mM}$ (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 7.5, $60\,\mathrm{mM}$ KCl, $15\,\mathrm{mM}$ NaCl, $0.5\,\mathrm{mM}$ ethylene glycol tetraacetic acid pH 8, 0.25% Triton-X, $10\,\mathrm{mM}$ sodium butyrate, $1\,\mathrm{mM}$ phenylmethanesulfonylfluoride, $0.1\,\mathrm{mM}$ Dithiothreitol protease inhibitor cocktail (Roche)] on ice and the chromatin fraction was pelleted by centrifugation. RNA for Affymetrix expression profiling was prepared as described (12). RNA labeling and cDNA hybridization to a *Drosophila*

Table 1. Primer table

Construct	Forward primer sequence	Reverse primer sequence
NSL1 RNAi amplicon 1	TTAATACGACTCACTATAGGGA GCGTC CGAGCTCAAC CTTC	TTAATACGACTCACTATAGGGA CACATGGGTGTGTTCATTAGTC
NSL1 RNAi amplicon 2	TTAATACGACTCACTATAGGGA GATGTCGCATCAAAGTCAGAGG	TTAATACGACTCACTATAGGGA GACTCGAGAAGAGCTCGCTGAT
GST RNAi amplicon	TTAATACGACTCACTATAGGGAG AATGTCCCCTATACTAG GTTA	TTAATACGACTCACTATAGGGAGA ACGCAT CCAGGCACATTG
NSL1 antibody cloning	CGCTCCATGGCTTTCATT AAGTTCCCCTGGAGCACC	ATTTCTAGATTAGATGC GTCTGCTGCGAACACCCTC

Genome GeneChip 2.0 was performed at the Gene Center Affymetrix Microarray Platform (Munich, Germany). Immunoblot analysis and immunofluorescence microscopy (IFM) analysis was performed as described previously (14). The lamin antibody was obtained from H. Saumweber (Berlin) and the MSL1 antibody was described previously (15).

Reporter gene ChIP assay and luciferase reporter assay

The reporter gene ChIP assay and luciferase reporter assay have been described before (12).

Chromatin extraction and immunoprecipitation

Chromatin extraction and immunoprecipitation were previously described (12). Briefly, chromatin extracts from sex-sorted adult flies were prepared and the DNA concentration of the extract was determined. DNA (7.5–15 µg) were used for a single ChIP experiment. Five microliters of anti-NSL1 serum was used in a single IP reaction. After the precipitation and extensive washing, DNA was extracted with phenol/chloroform, ethanol precipitated and further cleaned using the GenElute PCR clean-up kit (SIGMA). DNA was amplified using the whole-genome amplification kit (WGA, SIGMA). Labeling, hybridization to customized high-resolution NimbleGen tiling arrays (comprising the euchromatic part of the entire X chromosome, 5 Mb of 2 L, 2 R and 3 L, respectively, as well as 10 Mb of 3 R) (12), scanning and feature extraction was performed by imaGenes (Berlin).

ChIP-chip data processing

ChIP-chip data analysis was performed using R/Bioconductor (www.r-project.org; www.bioconductor.org). Raw signals of the NimbleGen NSL1 ChIP-chip were normalized and log2-transformed using the 'vsn' package (16). IP/input ratios of the modENCODE data were scaled to a mean of zero and a standard deviation of one. Promoter enrichments were calculated by summarizing the probe level signals in a window of 600 bp centered at the transcriptional start site (TSS) (FlyBase release 5.22). Promoter binding was classified based on the bimodal distribution of binding values, where genes within the population of lower values were considered 'unbound' and genes within the population of higher values were considered 'bound'. Alternatively, 'bound' were selected based on the fdr values from the 'locfdr' package applied on the promoter binding values with a fdr cutoff of <0.2. The results are robust to several normalization methods and promoter window definitions.

Genes were classified 'active' when (i) their Affymetrix expression value exceeded four (see below) and (ii) RNA polymerase II [modENCODE profile (17)] was classified as 'bound' on their promoters. A similar result was obtained using genes which are 'bound' (modeled on the bimodal distribution of the averaged binding along the transcribed region) by the elongating polymerase [serine 2 phosphorylated RNA polymerase II, data from (18)].

Promoters were classified as 'peaked', 'broad' and 'weak peak' promoters according to Hoskins et al. (19) and Ni et al. (20). Hierarchical cluster analysis of the promoter binding pattern was carried out using the R package 'hclust' and the 'complete' or 'ward' clustering approach as indicated in the figure legends.

All available modENCODE chromatin ChIP-chip data sets were screened for factors, which are enriched at promoter locations (by March 2011). After initial data quality assessment probe level binding was assessed for promoter probes (broad: ±300 bp centered at TSS; narrow: ±100 bp centered at TSS; upstream-biased: −300 - TSS- +100 bp), transcriptional termination (TT) sites (broad: ±300 bp centered at TT; narrow: ±100 bp centered at TT; downstream-biased: -100 – TT+ 300 bp), gene probes (probes corresponding to annotated genes without promoter and termination probes) and intergenic probes (defined as probes not found in previous groups). Only ChIP-chip data sets with a clear enrichment for promoter probes relative to gene, intergenic and termination probes were selected for this study.

Transcriptome data analysis

Transcriptome data analysis was conducted as described previously (12). Briefly, raw signals were normalized, summarized and log2-transformed using the 'gcrma' package. Significant change of gene expression was calculated applying the 'locfdr' package on a 'sam' statistics using a cutoff of fdr <0.35. Alternatively, genes with log2 (NSL1 RNAi–GST RNAi) < (−1) were considered 'down-regulated'. The results are robust to various parameters in data analysis, as assessed by choosing varying thresholds. All expression data set values are log2-normalized with a theoretical dynamic range of $2\exp16$ (Affymetrix.com).

Housekeeping gene definition

Affymetrix expression data sets of 40 different *Drosophila* tissues [GSE7763, (21)] were processed as described above for the NSL1 transcriptome data set. For every gene, the standard deviation was calculated across all 40 samples (gene variation index). Filtering for active genes, the distribution of standard deviations resulted in two major populations with the best discrimination at a standard deviation of ∼1.5 (Supplementary Figure S9A and B). Consequently, genes with a gene variation index <1.5 were considered housekeeping genes and genes with a gene variation index >1.5 were considered differentially regulated genes. The results are robust to different applied thresholds. In an alternative analysis (presented in Supplementary Figure S2E), we took the more stringent call for housekeeping gene function according to the classification of Weber and Hurst (22). Here, active genes which belong to either the 'tau' class or to the 'breadth' class were considered housekeeping genes.

ChIP-Seq data analysis

NSL1 ChIP-Seq and corresponding input data sets (11) were obtained from the ArrayExpress repository (E-MTAB-214). Sequence reads were mapped to the *Drosophila melanogaster* genome (dm3) using bowtie (23). Uninformative reads and read anomalies were filtered out using the R package 'SPP' (24), resulting in 7840131

unique NSL1 ChIP reads and 6094163 unique input reads. Peaks were identified using SPP with the following parameters: 'tag-wtd' method, fdr = 0.01, minimal distance between detected peaks = 100 bp. The input data was used to determine statistical significance of NSL1 peaks, resulting in the 'peak score'.

Core promoter motif analysis

We used the 10 promoter motifs described by Ohler and colleagues (25) to analyze promoter motif occurrences. For every motif a log-odds weight matrix description P of the binding sites is given, which was used to calculate a motif score for a specific sequence. It ranges between zero and one and measures how similar a binding site is to the consensus. In a first step, the log-odds score for the consensus site L_C is determined by

$$L_C = \sum_{i=1}^{w} \max\{P_{ib} : b \in \{A,C,G,T\}\},$$

where w is the motif length. The motif score given a specific binding site starting at position k in sequence X is calculated by

$$\text{motif score } (X,k) = \frac{1}{L_C} \sum_{i=1}^{w} P_{iX_{k+i-1}}$$

The motif score is the ratio of the log-odds score of the site at position k to the log-odds score of the consensus site. The motif score for the entire sequence X is given by the highest motif score in the sequence:

$$\text{motif score } (X) = \max_k\{\text{motif score } (X,k)\}$$

For the analyses, we used a threshold of motif score >0.3 to consider a binding site as functional. The *de novo* sequence analysis algorithm will be reported elsewhere (Hartmann and Soeding, manuscript in preparation).

RESULTS

NSL1 colocalizes with MBD-R2 at many active promoters

The genomic interaction profile of MOF differs in adult male and female flies, reflecting its incorporation into the male-specific DCC and the general NSL complex (11,12). We previously monitored the MBD-R2 distribution in adult male and female flies but could not detect any significant difference (12). Since MBD-R2 is the only NSL complex protein which may interact with DCC members (10) we sought to compare the genome-wide binding pattern of the NSL complex with the potential core subunit of the complex, NSL1. In order to compare the NSL1 interactions in the genomes of adult male and female flies, an antibody was raised against NSL1 and

its specificity confirmed combining RNAi with subsequent detection by indirect immunofluorescence microscopy (IFM) and immunoblotting (Supplementary Figure S1 and see below). The antibody was then used for ChIP-chip experiments, where NSL1 was precipitated from chromatin preparations from hand-sorted adult male and female flies and the associated DNA was amplified and hybridized to high-resolution DNA tiling microarrays representing the X chromosome and an equivalent amount of the autosomes. The binding profile in male and female flies did not show any significant differences (Supplementary Figure S2A). In addition, NSL1 was found at the same loci as MBD-R2 (Supplementary Figure S2B), in agreement with the results of the biochemical definition of both proteins as 'NSL' complex subunits (10–12). The ChIP-chip profiling suggested that NSL1 globally binds target loci independent of the fly sex, confirming previous ChIP-qPCR analyses at selected loci (11).

Re-examination of the previously published NSL1 ChIP-Seq profiles, which had been generated from salivary glands of mixed-sex third instar larvae (11), revealed a systematic enrichment of NSL1 peaks at RNA polymerase II—promoters relative to genes transcribed by RNA polymerases I and III (Table 2). Applying a superior peak calling algorithm (24) to these data identified the majority of NSL1 binding events within a window of 200 base pairs (bp) around the annotated TSS (Figure 1A), implicating the NSL complex in transcriptional initiation.

In order to avoid the heterogeneous salivary gland tissue, which impedes a comparison of NSL binding with the transcriptional activity and with other known promoter binding factors, an NSL1 ChIP-chip profile was generated from *Drosophila* S2 cells. These cells are commonly used in the chromatin community because they provide a homogeneous biological material, a fact that allows comparing our data to other published genomic data sets, such as the comprehensive collection of chromatin factors and histone modifications generated by the modENCODE consortium with a similar ChIP-chip strategy (17).

The newly generated NSL1 ChIP-chip profile correlated well with our previously published MBD-R2 profile (12) as well as with the MBD-R2 profile generated by the modENCODE consortium using a different antibody (Supplementary Figure S2C). Therefore, in the following we subsume the individual NSL1 and MBD-R2 profiles as the 'NSL complex' binding, unless stated otherwise. We related the NSL complex binding at promoters with the transcriptional activity of the corresponding genes, using the ChIP-chip profile of the elongating polymerase as a direct readout for active transcription (18). The NSL complex binds active genes with high preference, but only a subset of ~60–70% (depending on the threshold) (Figure 1B, left). A similar result was obtained when

Table 2. NSL1 ChIP-Seq peaks mapped to transcript type

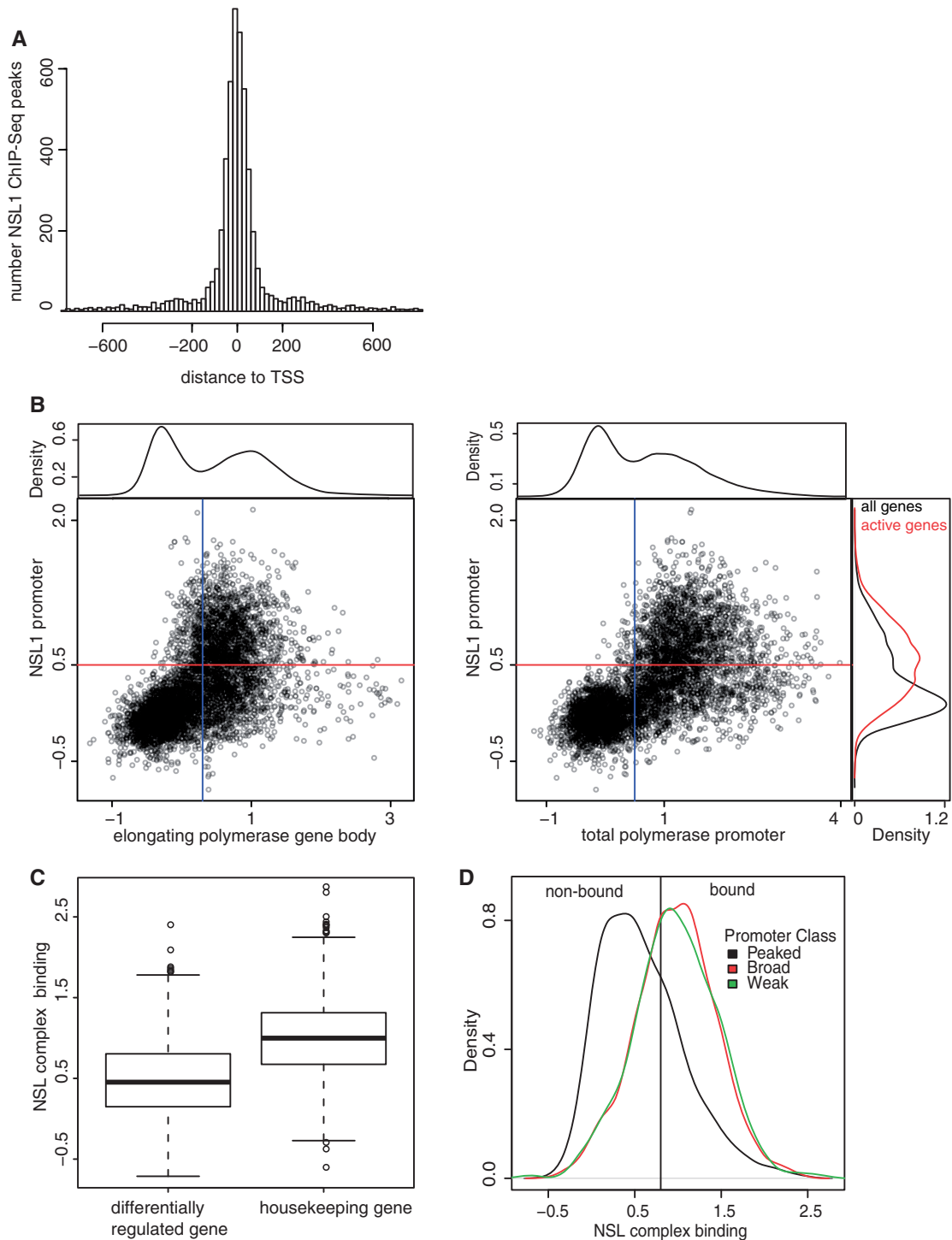| Transcript type	MiRNA	mRNA	ncRNA	rRNA	snoRNA	snRNA	tRNA
Number annotated transcripts	194	22 765	189	160	249	47	314
Number NSL1 peaks mapped to transcript TSS	0	4302	14	0	1	1	5
Fraction NSL1-bound transcripts rel. to all transcripts	0	18.9	7.4	0	0.4	2.1	1.6

Figure 1. The MOF-containing NSL complex is enriched at promoters of most housekeeping genes. (**A**) NSL1 peaks map close to the TSS. The histogram displays the distance between the summit of NSL1 ChIP-Seq (11) peak and the closest TSS. Refinement of NSL1 positions relative to the TSS compared to (11) was achieved by using an improved peak calling algorithm [SPP (24) and more precisely mapped TSSs (56)]. (**B**) NSL1 prevalently binds active gene promoters. Scatter plots of NSL1 promoter binding versus (i) elongating polymerase at all genes (left) (18), (ii) total polymerase promoter occupancy at all genes (right) [modENCODE (17)]. Significant binding cutoffs of NSL1 (red) and polymerase (blue) are indicated. The density plots on the top of each histogram depict the signal distribution of the elongating polymerase (left) and the total polymerase (right). The density plot to the right indicates the NSL1 promoter signal distribution for all genes (black) and active genes [red, based on the elongating polymerase II gene signal (18)], respectively. (**C**) The NSL complex preferentially associates with housekeeping genes (Welch two sample *t*-test, $P < 2.2\,\mathrm{e}^{-16}$). The boxplot depicts NSL1 binding at differentially regulated and housekeeping genes (for categorization, see 'Materials and Methods' section). A similar result is gained using the available MBD-R2 ChIP-chip data sets (12). An alternative, more stringent categorization for housekeeping genes after Hurst and colleagues (22) is shown in Supplementary Figure S2. (**D**) The NSL complex prevalently binds to dispersed promoters ('broad with peak' and 'weak peak' promoters) over peaked promoters (Welch two sample *t*-test, $P < 2.2\,\mathrm{e}^{-16}$). Density plot of NSL1 binding at genes, which were grouped according their transcriptional start site usage in 'peaked' promoters, 'broad with peaked' promoters and 'weak peak' promoters using the data of (19). The window is split for NSL1-bound (right) and -unbound (left) promoters.

displaying gene activity as a function of polymerase promoter binding (Figure 1B, right) or Affymetrix RNA expression profiling (data not shown), in agreement with previous studies examining other markers of the NSL complex (11,12).

The NSL complex specifically binds promoters of most housekeeping genes

As noted above, the NSL1 profile is very similar in nuclei of different sex and developmental stage despite significant expression differences (Supplementary Figure S2A and D). This indicates that the NSL complex may associate with 'housekeeping' genes, which are equally expressed in these diverse tissues. To test this hypothesis, we classified genes as 'housekeeping' or 'differentially regulated' according to their expression variation index, i.e. the standard deviation of expression, when compared between several *Drosophila* tissues (21). According to this classification the NSL complex showed a significant preference for 'housekeeping' over 'differentially regulated' genes (Figure 1C). The same conclusion was reached when 'housekeeping' genes were classified according to the more exclusive definition of Hurst and colleagues (22) (Supplementary Figure S2E). This conclusion is further illustrated by a gene ontology (GO) analysis of bound and unbound genes, which revealed that active NSL-bound genes are enriched in housekeeping functions such as 'cofactor biosynthetic processes', 'microtubule-based processes', 'protein complex biogenesis' (Supplementary Figure S3), whereas active genes which are not bound by the NSL complex are enriched in categories such as 'sensory perception', 'cell adhesion' and 'tissue developmental genes' (Supplementary Figure S4).

Recent improvements in high-throughput RNA profiling techniques facilitated quantitative mapping of TSSs at base pair resolution (19,20). Whereas some promoters possess well-defined TSS, where transcription reliably initiates within a few base pairs ('focused' or 'peaked' promoters), many promoters show a dispersed zone of transcription initiation of up to a few hundred base pairs, which may be dominated by a major TSS ('broad promoters') or not ('weak peak promoter') (20). Notably, differentially regulated genes tend to have peaked promoters whereas housekeeping genes are enriched for broad or weak promoters (19). Concordantly, we found that the NSL complex is strongly overrepresented at promoters of the latter classes (Figure 1D).

The NSL complex activates only a specific subset of bound genes

It has remained controversial whether NSL target genes are activated or repressed after RNAi ablation of NSL complex components (11–13). Akthar and coworkers observed that similar fractions of NSL target genes were up- or downregulated following RNAi against MOF, NSL3 and MBD-R2 and subsequent microarray-based transcriptome profiling (11,13). By contrast, we found that the transcription of genes that had the NSL subunit MBD-R2 bound was mostly reduced when MBD-R2 levels were lowered (12). However, since MBD-R2 is the only NSL complex subunit that was suggested to interact with components of the DCC (10), it was necessary to exclude indirect effects. We therefore examined the expression of NSL target genes after depletion of the core subunit of the NSL complex, NSL1.

RNAi against NSL1 in S2 cells efficiently depleted the protein as examined by immunoblotting and IFM (Supplementary Figure S1). Genome-wide transcriptome profiling of the NSL1-depleted cells led to the down-regulation of a considerable fraction of genes (Figure 2A), most of which had been classified as 'NSL-bound' before (Figure 2B). This is consistent with reporter gene assays where the transcription brought about by tethering MOF to a model promoter was diminished upon NSL1 depletion (for details, see Supplementary Figure S5). Importantly, the expression of the majority of NSL1 target genes was unchanged (Figure 2A), such that only 20–30% of them (depending on the threshold) required
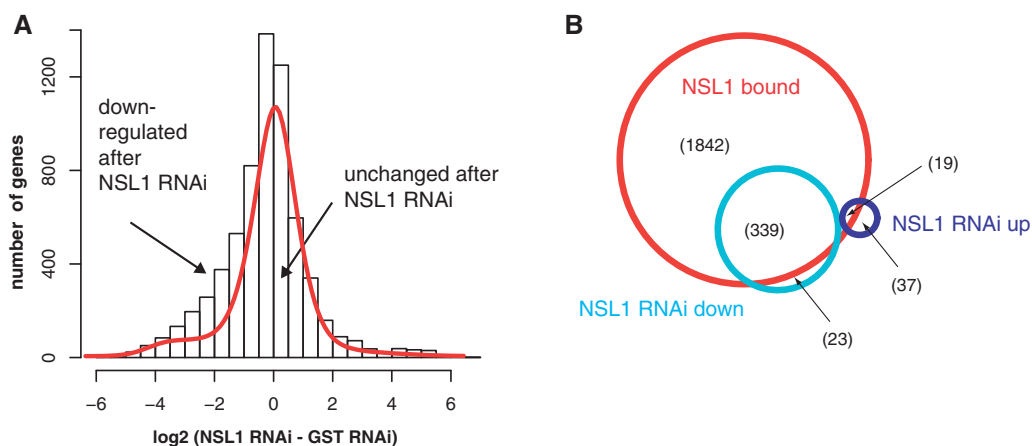

Figure 2. The NSL complex activates only a subset of its target promoters. (**A**) The histogram depicts transcriptome changes upon NSL1 depletion in S2 cells. The overlayed modeled normal distribution (red) reveals a skew toward the population of down-regulated genes. The gene set was filtered for active genes based on the transcriptome of control cells. (**B**) Proportional Venn diagram depicts genes 2-fold up- or down-regulated after RNAi against NSL1, respectively, and NSL1-bound genes (NSL1 ChIP-chip in S2 cells). Numbers in parentheses indicate the size of the respective genes sets. Only active genes (total polymerase promoter occupancy determined by the modENCODE consortium) represented on our custom-tailored microarray tiling array are shown.

NSL1 for proper expression. The same trend had been observed earlier in the context with MBD-R2 (12) (and data not shown). The MBD-R2 ChIP-chip profile and the MBD-R2 RNAi transcriptome data are indeed very similar to the NSL1 data (Supplementary Figures S2C and S5C), arguing that they form a functional complex bound to chromatin.

We next asked whether the genes that were activated by the NSL complex coded for related housekeeping functions. The GO classification revealed that the genes whose expression was diminished upon NSL1 depletion were enriched in genes involved in nucleic acid metabolism, such as genes involved in transcription, RNA processing, translation, DNA replication and DNA repair (Supplementary Figure S6). Evidently, the NSL complex only activates a specific subset of the many housekeeping genes. In order to explore whether the promoters of NSL-responsive genes could be recognized by a combination of *cis*-elements and *trans*-factors, we set out to identify chromatin proteins with genome binding profiles related to the NSL complex and to investigate whether the promoters regulated by NSL shared particular core promoter motifs.

The NSL complex co-occupies target promoters together with the chromatin remodeler NURF and the histone methyltransferase Trithorax

The NSL1 ChIP-chip profile in S2 cells allowed a direct comparison with the chromatin profiles recorded by the modENCODE consortium (17), which used the same cell line and the same profiling technique. We mined the modENCODE data for profiles of general chromatin factors (excluding sequence-specific transcription factors) and histone modifications, which are preferentially enriched at promoters (see 'Materials and Methods' section for a detailed discussion on selection algorithm). We created a pairwise correlation matrix for 23 selected protein and histone modification profiles and performed an unsupervised hierarchical clustering to reveal the extent of correlation with the NSL complex. We found the profiles of the interband protein Chromator, the WD40-repeat protein WDS, the NURF complex subunit NURF301 and the methyltransferase Trithorax highly correlated with the NSL complex profile (Figure 3A and B; Supplementary Figure S7). Chromator had been found in an early NSL complex purification (10) but could not been recovered in more recent experiments (11,12), possibly due to more transient or indirect interaction. Notably, 5–15% of promoters which contain NSL1, MBD-R2, WDS, NURF301 and Trithorax lack Chromator. The WD40-repeat protein WDS consistently copurifies with NSL complex members (10–12) and other chromatin complexes including the *Drosophila* ATAC acetyltransferase complex (26) and mammalian MLL methyltransferase complexes (27,28). NURF301 is the diagnostic marker subunit of the Imitation Switch (ISWI)-containing nucleosome remodeling factor NURF, which stimulates transcription by remodeling promoter nucleosomes (29,30). Trithorax was originally described to counteract the repression of homeotic genes

by the polycomb group proteins (31–33). More recently, genome-wide ChIP-chip studies have indicated a widespread binding of Trithorax to many promoters (34,35).

The pairwise relationships between the tested factors are further illustrated by the scatter plots depicted in Figure 3C, which emphasize that the NSL complex, WDS, Chromator, NURF301 and Trithorax co-occupy target promoters at linearly proportional levels (Figure 3C). Promoters which are strongly bound by the NSL complex are also highly enriched for NURF301, Chromator and Trithorax. The same strong correlation can be seen in an unbiased analysis using all microarray probe signals, confirming the promoter-focused analysis described above (Supplementary Figure S7B).

Searching for factors enriched at promoters we found the heterochromatin protein 1c (HP1c) and, consistent with previous results (36), the insulator proteins BEAF32 and CP190 (37) enriched at housekeeping promoters (Supplementary Figure S8). These factors localize to minor subsets of the NSL/Chromator/NURF301/ Trithorax target promoters (Figures 3 and 5; Supplementary Figures S7B and 11). Importantly, the presence of BEAF32, CP190 and HP1c determines whether the bound NSL complex functions as an activator or not (see below).

Quantitative NSL1 binding correlates best with the DNA replication-related element

Conceivably, the association of the NSL complex and its colocalized chromatin modifiers may be determined by a particular core promoter architecture. Different promoters are characterized by the presence and combination of a range of sequence motifs that provide contact surfaces for general transcription factors and, therefore, modulate the formation of the transcription pre-initiation complex (38–40). The core promoter sequence motifs can be classified as canonical core promoter motifs which have fixed positions with regard to the TSS, such as the TATA box, the MTE (motif ten element), the DPE (downstream core promoter element) and the INR (initiator), or as motifs with weaker positional information (Ohler 1, Ohler 5, Ohler 6, Ohler 7, Ohler 8 and DRE) (25,41). Canonical core promoter motifs are enriched in peaked promoters, whereas weakly positioned motifs are characteristic of dispersed promoters. The mechanisms of action of most dispersed elements are unknown [with the exception of the DRE (39)].

Since NSL1 peaks within the core promoters of genes with dispersed transcriptional start sites (Figure 1A and D) we investigated whether the NSL complex is associated with a specific set of core promoter motifs. We first characterized the core promoter motifs with regard to their distribution at active housekeeping and differentially regulated genes (Supplementary Figure S9). As the motifs deviate from their defined consensus sequences in many cases, a similarity score (motif score) was calculated for each promoter reflecting the similarity of the sequence to any of the ten promoter consensus motifs described by Ohler and colleagues (25) (see 'Materials and Methods' section). We found that over 70% of all active promoters
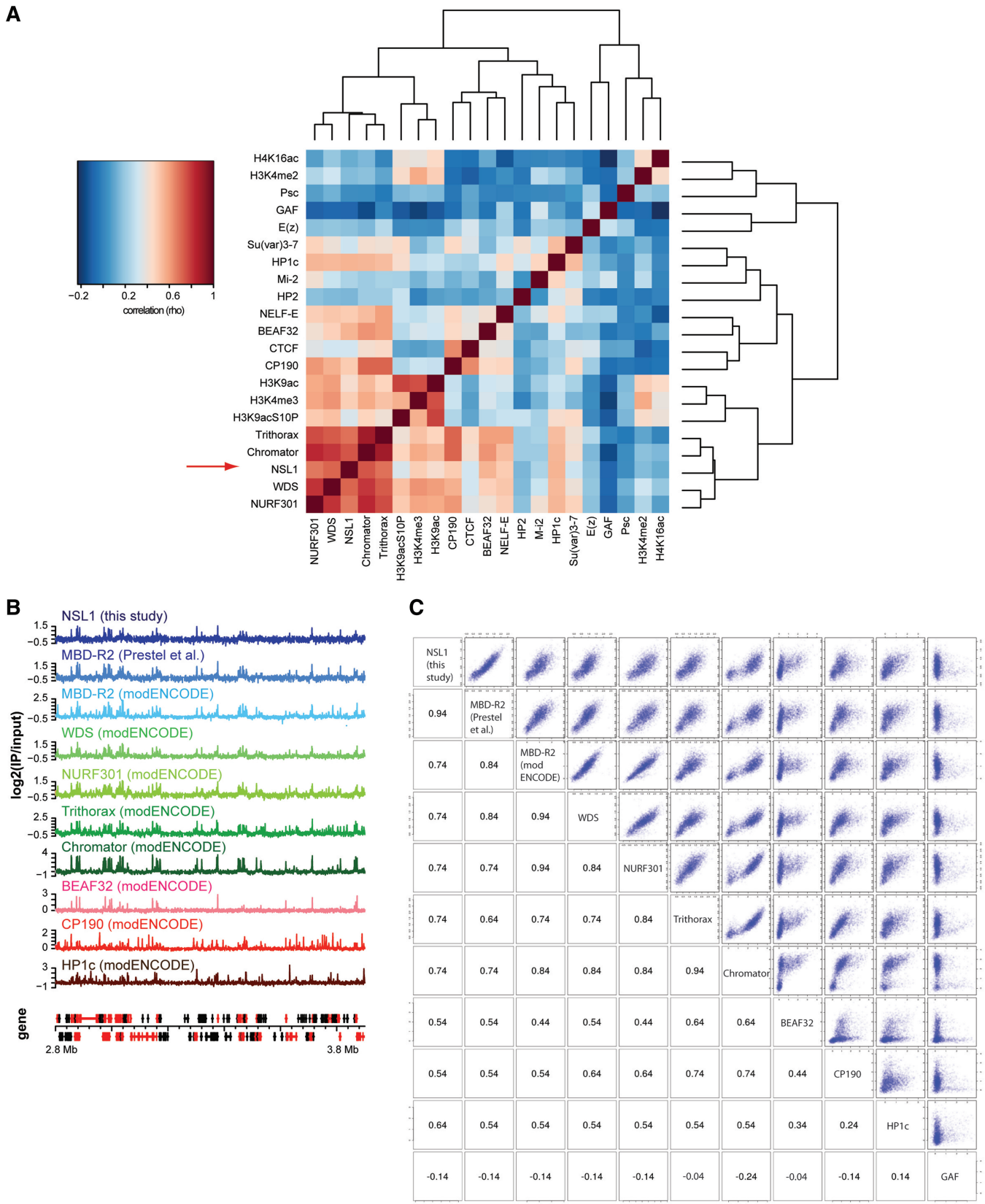
Figure 3. The NSL complex cooccupies target promoters together with the chromatin remodeler NURF and the histone methyltransferase Trithorax. (**A**) Heat map visualized correlation matrix of promoter-enriched chromatin factors and histone modifications at active genes. Pairwise Spearman correlations were calculated using only active autosomal genes. The dendrogram shows the hierarchical clustering after which the matrix was sorted. (**B**) ChIP-chip profile of the indicated proteins along a representative region of the chromosome arm 2 L. The gene structure is indicated below (active genes are red). (**C**) Pairwise scatter plot of promoter binding for each indicated factor using only active autosomal genes. Spearman non-parametric correlation coefficients are provided for each pair.

can be described by these ten motifs, indicating that our analysis is representative (Supplementary Figure S9). In agreement with previous analyses (41–43), the promoter motifs INR, MTE and DPE were clearly overrepresented in differentially regulated genes, which fits their enrichment at peaked promoters (Supplementary Figures S8 and S9). Accordingly, housekeeping genes are enriched for the motifs DRE, Ohler1, Ohler 5, Ohler 6 and Ohler 7.

Being able to characterize the core promoter motifs allowed us to examine whether there is a differential association of NSL with any of them. Active genes were categorized either as NSL targets or as non-targets based on their NSL complex promoter occupancy. As expected, the NSL1 target genes are enriched for the housekeeping promoter motifs DRE, Ohler 1, Ohler 5, Ohler 6 and Ohler 7 and depleted for TATA, INR, MTE and DPE. Consistently, when we performed *de novo* motif analysis of the sequences covered by the NSL1 ChIP-Seq peaks (11), we again obtained the same motifs (Supplementary Figure S10). This confirms that the NSL1 peaks at core promoter motifs are diagnostic for housekeeping genes.

Is there any correlation between the 'strength' of NSL1 binding and how well an underlying motif matches its consensus sequence? In order to address this question we used the NSL1 ChIP-Seq data set (11), which due to its good dynamic range allowed to categorize the ChIP-Seq peak score as a surrogate for binding 'strength'. We binned the ChIP-Seq peaks in equally sized groups according to their peak score [determined by SPP, (24)] and displayed the fraction of promoters bound by a specific group at a given motif score (Figure 4A, left). Among the ten tested core promoter motifs the DRE motif, and to a lesser extent motif Ohler 7, are the only motifs with scores that correlate with the NSL1 ChIP-Seq peak score. This suggests that DRE-containing promoters (and those containing the less abundant Ohler 7 motif) primarily contain NSL complex targeting clues (Figure 4B).

The combination of chromatin factors and core promoter motifs enhance the prediction of NSL-regulated promoters

Whether or not a promoter-bound transcription factor engages in active regulation often depends on the context of close-by *cis* elements and interacting factors (44). This appears to be the case for the NSL complex, as we showed that the complex only activates a subset of the promoters it associates with. NSL binds with high preference to a set of housekeeping promoter motifs and its binding 'strength' correlates best with the presence of the DRE motif. Can the subset of these NSL targets whose transcription is diminished after depletion of NSL (i.e. those promoters at which the complex is functional as an activator) be distinguished at the sequence level? We grouped active genes according to their core promoter motif class (see 'Materials and Methods' section) and monitored the transcriptome changes after NSL depletion for each group. Strikingly, only promoters containing the core promoter motif 'Ohler 5' were strongly enriched for NSL complex functional sites (Figure 5A). We note that 'Ohler 5'-containing promoters do not show the strongest correlation to NSL binding strength (Figure 4B)

suggesting that quantitative differences in factor binding are not directly translated into a functional output.

We had observed that most promoters bound by HP1c, BEAF32 and CP190 are among those also occupied by the NSL complex (Figure 3 and Supplementary Figure S11). Most of the HP1c, BEAF32 and CP190 binding occur at distinct subsets as the three factors only colocalize at a minority of sites (Supplementary Figure S11). Intriguingly, promoters bound by any of the three factors HP1c, BEAF32 or CP190 are obviously underrepresented among the genes, whose transcription is activated by the NSL complex (Figure 5B and Supplementary Figure S11).

In summary, the data suggest that the functionality of a promoter-associated NSL complex is modulated by positive effectors (e.g. unidentified interactors of the 'Ohler 5' element) and negative regulators (HP1c and the insulator proteins BEAF32 and CP190).

DISCUSSION

In this study, we show that the NSL complex is a potential coactivator, which binds to many active genes, but regulates only a specific subset of them. In our efforts to describe the circumstances that define complex association and function, we considered the contributions of two major parameters: the diverse DNA sequences around the core promoters, which are characterized by combinations of recurring sequence motifs, and the association of chromatin regulators that have recently been mapped by the modENCODE consortium. Combining these diverse data sets, we were able to improve the prediction toward whether the transcription of an NSL-bound gene is modulated by the NSL complex. To our knowledge, this is the first systematic study demonstrating the usefulness of this type of data integration.

The NSL complex is a transcription cofactor dedicated to housekeeping genes

Following our observation that the NSL complex binds to only a subset of all active promoters, we discovered that the target genes were mostly housekeeping genes. This was surprising as to our knowledge so far no transcription coregulator dedicated to housekeeping genes is known. This may simply reflect the fact that historically the mechanisms underlying differential transcription regulation received more attention. Several lines of evidence support the conclusion that the NSL complex preferentially localizes to the majority of housekeeping promoters. (i) We do not detect significant differences in the global chromatin binding profile of NSL complex members in cells of different sex or developmental stage. (ii) Genes that have NSL bound at their promoters show little expression variation among different tissues as compared to active genes that lack the NSL complex. (iii) NSL-bound promoters are depleted of sequence motifs known to be enriched in genes differentially regulated during development and in tissue homeostasis (38). (iv) GO analysis of the active NSL-bound genes revealed an overrepresentation of categories for housekeeping functions, whereas the converse data set of active genes not bound by NSL
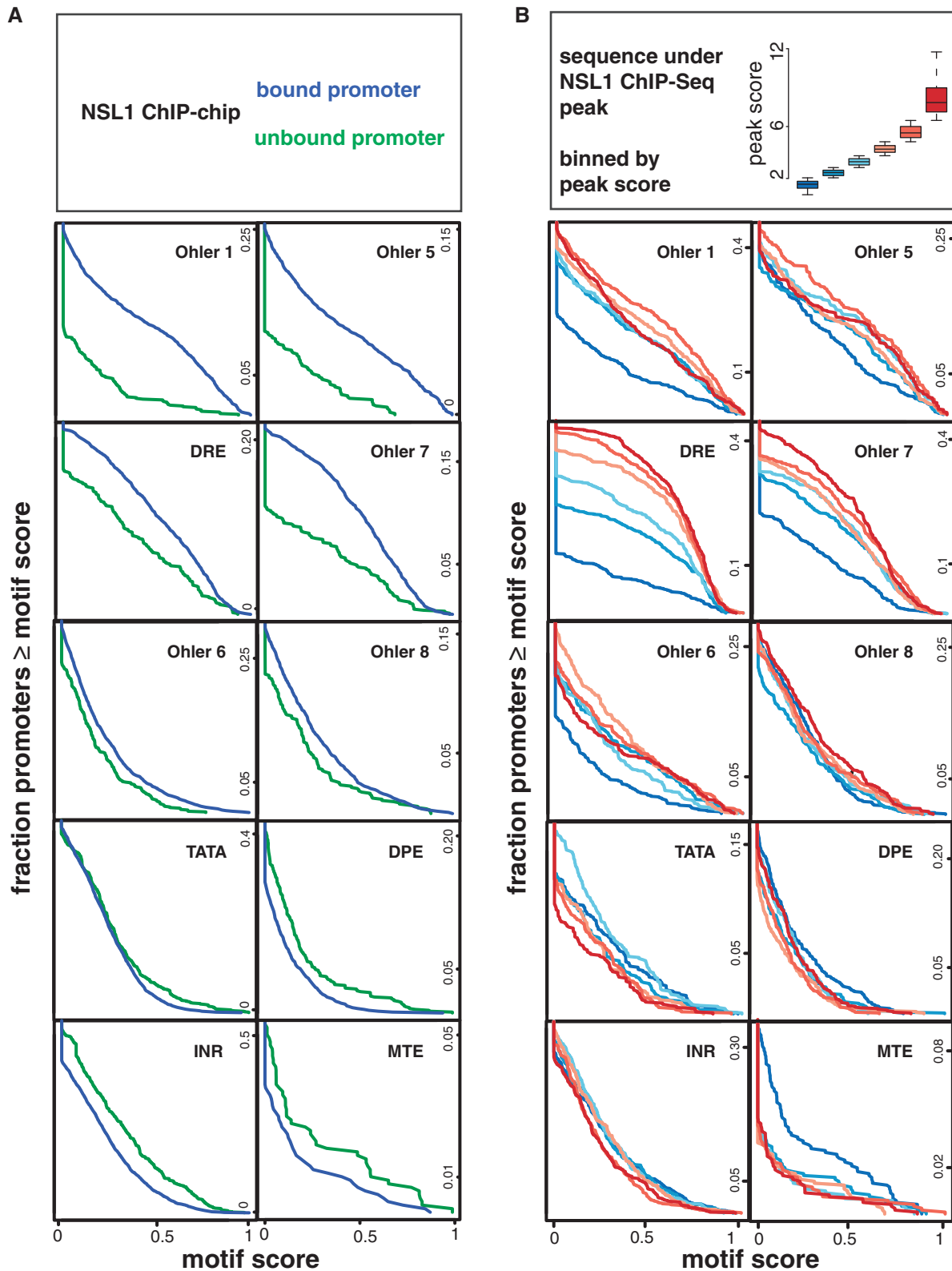
Figure 4. NSL1 binding correlates best with the DRE motif score. (**A**) Promoters containing motifs Ohler 1, Ohler 5, DRE and Ohler 7 are overrepresented among the NSL complex target genes. Active genes were grouped into NSL1-bound and NSL1-unbound genes, as determined by NSL1 ChIP-chip in S2 cells. The curves display the cumulative fraction of promoters, which exceed a given motif score. The motif score is derived from the information content of the position weight matrix, as taken from Ohler and colleagues (25), see 'Materials and Methods' section for details). (**B**) NSL1 binding 'strength' linearly scales to the DRE motif score. NSL1 ChIP-Seq peaks (11) are grouped in equal-sized bins according to their peak score. Similar to (A), the fraction of promoters is displayed relative to the motif score.
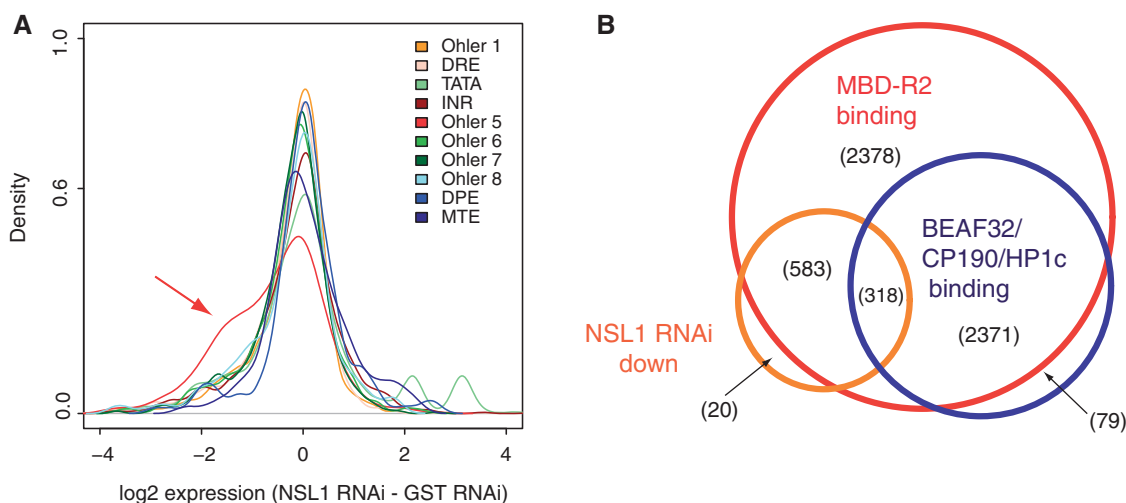
Figure 5. A combination of chromatin factors and core promoter motifs enhance the prediction of NSL transcriptional targets. (**A**) The core promoter motif Ohler 5 is overrepresented in genes, which are regulated by the NSL complex ($P < 2.2\,\mathrm{e}^{-16}$, one-sample *t*-test). The density plot of the transcriptome changes after RNAi against NSL1 for groups of genes enriched for the ten core promoter motifs defined by Ohler *et al.* (25). In the presented analysis only active NSL-bound genes were considered (using the modENCODE MBD-R2 data set (17), which allows the comparison to a larger transcriptome data set). The same conclusion was reached if all active genes (without filtering for NSL complex binding) or different 'NSL complex binding' criteria (such as NSL1 binding) were employed (data not shown). (**B**) Venn diagram representation of gene sets bound by the NSL complex (modENCODE MBD-R2 ChIP-chip profile), down-regulated after ablation of the NSL complex and bound by at least one of the three investigated proteins BEAF32, CP190 and HP1c (17). Numbers in brackets represent the number of genes in the respective group. The modENCODE MBD-R2 data set served as a surrogate for the NSL complex to minimize technical variation in the comparison with the BEAF32/CP190/HP1c data sets. Similar results were obtained using the NSL1 data set and the MBD-R2 (12) data set, respectively.

present diverse categories including 'developmental programs' and 'acute signaling'. Other chromatin constituents, like HP1c and the insulator proteins BEAF32 and CP190 also interact preferentially with housekeeping gene promoters, as previously shown by Ohler and colleagues (36), but these factors bind to a much more limited number of genes in this class. Our analysis supports the concept of global coregulation of functionally related gene classes by common cofactors.

The developmental regulators NURF and Trithorax colocalize with NSL at housekeeping promoters

The extensive colocalization of the NSL complex with the methyltransferase Trithorax and the chromatin remodeler NURF is puzzling since those factors are best known as regulators of transcription of very restricted sets of genes (developmental and highly inducible genes) (30,32), and only recently has their extensive genome-wide localization at many active gene promoters been noticed (34,35,45,46). Conceivably, these three complexes cooperate to regulate the transcription of housekeeping genes at the level of chromatin organization and/or transcription initiation. This hypothesis is supported by previous reports of biochemical or genetic interactions between components of the three factors. A genetic interaction between the *Xenopus* BPTF (the NURF301 homolog) and *Xenopus* WDR5 (a homolog of the NSL subunit WDS) has been reported (47). Furthermore, Dou *et al.* (27) described a 'supercomplex' containing the human NSL as well as the MLL1 complexes [MLL1 is homologous to *Drosophila* Trithorax].

At present it is not clear whether NURF and Trithorax-containing complexes contribute to the targeting of the NSL complex (or vice versa), or whether all three regulators are attracted by an additional common denominator of target promoters. None of the three complexes contains any specific DNA-binding subunit. NURF can be recruited to inducible genes via direct interactions between the large NURF301 subunit and transcription factors, such as the GAGA factor (29) or the ecdysone receptor (48). However, these interactions certainly do not explain the widespread targeting of NURF to housekeeping genes *in vivo* reported here. We noted a good quantitative correlation between the NSL1 binding levels and the DRE core promoter motif score, which opens the possibility that a DRE—recognizing factor may stimulate NSL recruitment. One candidate for such a factor is DREF, which has been isolated as a DRE binding factor (49). DREF may also contribute to the recruitment of NURF, since an association of DREF with NURF has been observed in a much larger complex based on the immunoaffinity purification of the TATA box binding protein (TBP)-related factor TRF2 (39).

In addition to direct recruitment by DNA-binding proteins, transcription cofactors may be tethered by specific local histone modifications through recognition domains (50). It is likely that this principle will also contribute to the observed colocalization of NSL, NURF and Trithorax complexes. Trithorax (the *Drosophila* MLL1 homolog) is an enzyme that methylates histone H3 at lysine 4 (H3K4me3), a mark that characterizes active promoters (46). Interestingly, WDS, which copurifies with NSL complexes from *Drosophila* and mammalian cells

(27) has been shown to preferentially interact with methylated H3K4 (28). The mammalian homolog of NURF301 (BPTF) also recognizes mononucleosomes marked with methylated H3K4 and acetylated H4K16ac through its PHD finger and bromodomain, respectively (51). Acetylation of H4K16 by MOF in the NSL complex may, therefore, contribute to the local enrichment of NURF at target promoters. Our study gives rise to numerous testable hypotheses as to the nature of the interaction network that leads to the observed selective targeting of the NSL complex.

The NSL complex only regulates a subset of target promoters

The detailed analysis of the transcriptional effects of the NSL complex revealed that the NSL complex regulates only a subset of bound genes. Such a situation is not without precedent as it has been shown for a number of transcription factors that many binding events appear to be non-functional (44). In fact, it is a major challenge to predict the functional sites from the interaction profiles of single factors as functionality is frequently determined by the local clustering of binding sites, synergism between colocalized proteins, and recently, chromatin accessibility (52,53). Accordingly, we favor the idea that a combination of chromatin factors and core promoter elements determines the activity of the NSL complex at any target promoter. An even more immediate influence of promoter DNA on interacting proteins may be imagined as a direct effect of a DNA sequence on the conformation and, therefore, the activity of a bound transcription factor has been described (54).

Alternatively, it is possible that the default state of every chromatin-bound NSL complex is functional, but that the realization of this potential is restricted by negative factors. We found that the presence of either one of the three proteins HP1c, BEAF32 or CP190 correlated with lack of NSL1 regulation. Insulator binding proteins like BEAF32 and CP190 are known to decrease enhancer-promoter interactions, which may lead to decreased transcriptional output. Interestingly, antagonistic roles for BEAF32 and DREF have been suggested for some overlapping *in vivo* binding sites (55). Resolving the mechanistic intricacies of complex promoter regulation remains a challenging task for future endeavors.

ACCESSION NUMBER

Microarray accession number: GSE30991.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–11 and Supplementary References [57].

ACKNOWLEDGMENTS

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Clapier,C.R. and Cairns,B.R. (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.*, **78**, 273–304.
2. Kwon,S.Y., Xiao,H., Wu,C. and Badenhorst,P. (2009) Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities. *PLoS Genet.*, **5**, e1000574.
3. Tsukiyama,T. and Wu,C. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*, **83**, 1011–1020.
4. Smith,E.R., Cayrou,C., Huang,R., Lane,W.S., Cote,J. and Lucchesi,J.C. (2005) A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol. Cell. Biol.*, **25**, 9175–9188.
5. Taipale,M., Rea,S., Richter,K., Vilar,A., Lichter,P., Imhof,A. and Akhtar,A. (2005) hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.*, **25**, 6798–6810.
6. Straub,T. and Becker,P.B. (2007) Dosage compensation: the beginning and end of generalization. *Nat. Rev.*, **8**, 47–57.
7. Gelbart,M.E., Larschan,E., Peng,S., Park,P.J. and Kuroda,M.I. (2009) Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.*, **16**, 825–832.
8. Larschan,E., Bishop,E.P., Kharchenko,P.V., Core,L.J., Lis,J.T., Park,P.J. and Kuroda,M.I. (2011) X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature*, **471**, 115–118.
9. Prestel,M., Feller,C. and Becker,P.B. (2010) Dosage compensation and the global re-balancing of aneuploid genomes. *Genome Biol.*, **11**, 216.
10. Mendjan,S., Taipale,M., Kind,J., Holz,H., Gebhardt,P., Schelder,M., Vermeulen,M., Buscaino,A., Duncan,K., Mueller,J. *et al*. (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol. Cell*, **21**, 811–823.
11. Raja,S.J., Charapitsa,I., Conrad,T., Vaquerizas,J.M., Gebhardt,P., Holz,H., Kadlec,J., Fraterman,S., Luscombe,N.M. and Akhtar,A.

(2010) The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol. Cell*, **38**, 827–841.

12. Prestel,M., Feller,C., Straub,T., Mitlohner,H. and Becker,P.B. (2010) The activation potential of MOF is constrained for dosage compensation. *Mol. Cell*, **38**, 815–826.

13. Kind,J., Vaquerizas,J.M., Gebhardt,P., Gentzel,M., Luscombe,N.M., Bertone,P. and Akhtar,A. (2008) Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. *Cell*, **133**, 813–828.

14. Straub,T., Neumann,M.F., Prestel,M., Kremmer,E., Kaether,C., Haass,C. and Becker,P.B. (2005) Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. *Chromosoma*, **114**, 352–364.

15. Gilfillan,G.D., Straub,T., de Wit,E., Greil,F., Lamm,R., van Steensel,B. and Becker,P.B. (2006) Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex. *Genes Dev.*, **20**, 858–870.

16. Huber,W., von Heydebreck,A., Sultmann,H., Poustka,A. and Vingron,M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18(Suppl. 1)**, S96–S104.

17. Celniker,S.E., Dillon,L.A., Gerstein,M.B., Gunsalus,K.C., Henikoff,S., Karpen,G.H., Kellis,M., Lai,E.C., Lieb,J.D., MacAlpine,D.M. *et al.* (2009) Unlocking the secrets of the genome. *Nature*, **459**, 927–930.

18. Regnard,C., Straub,T., Mitterweger,A., Dahlsveen,I.K., Fabian,V. and Becker,P.B. (2011) Global analysis of the relationship between JIL-1 kinase and transcription. *PLoS Genet.*, **7**, e1001327.

19. Hoskins,R.A., Landolin,J.M., Brown,J.B., Sandler,J.E., Takahashi,H., Lassmann,T., Yu,C., Booth,B.W., Zhang,D., Wan,K.H. *et al.* (2011) Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome Res.*, **21**, 182–192.

20. Ni,T., Corcoran,D.L., Rach,E.A., Song,S., Spana,E.P., Gao,Y., Ohler,U. and Zhu,J. (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods*, **7**, 521–527.

21. Chintapalli,V.R., Wang,J. and Dow,J.A. (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.*, **39**, 715–720.

22. Weber,C.C. and Hurst,L.D. (2011) Support for multiple classes of local expression clusters in Drosophila melanogaster, but no evidence for gene order conservation. *Genome Biol.*, **12**, R23.

23. 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.

24. Kharchenko,P.V., Tolstorukov,M.Y. and Park,P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.*, **26**, 1351–1359.

25. Ohler,U., Liao,G.C., Niemann,H. and Rubin,G.M. (2002) Computational analysis of core promoters in the Drosophila genome. *Genome Biol.*, **3**, RESEARCH0087.

26. Suganuma,T., Gutierrez,J.L., Li,B., Florens,L., Swanson,S.K., Washburn,M.P., Abmayr,S.M. and Workman,J.L. (2008) ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat. Struct. Mol. Biol.*, **15**, 364–372.

27. Dou,Y., Milne,T.A., Tackett,A.J., Smith,E.R., Fukuda,A., Wysocka,J., Allis,C.D., Chait,B.T., Hess,J.L. and Roeder,R.G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, **121**, 873–885.

28. Wysocka,J., Swigut,T., Milne,T.A., Dou,Y., Zhang,X., Burlingame,A.L., Roeder,R.G., Brivanlou,A.H. and Allis,C.D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, **121**, 859–872.

29. Xiao,H., Sandaltzopoulos,R., Wang,H.M., Hamiche,A., Ranallo,R., Lee,K.M., Fu,D. and Wu,C. (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell*, **8**, 531–543.

30. Badenhorst,P., Voas,M., Rebay,I. and Wu,C. (2002) Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.*, **16**, 3186–3198.

31. Simon,J.A. and Tamkun,J.W. (2002) Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr. Opin. Genet. Dev.*, **12**, 210–218.

32. Ringrose,L. and Paro,R. (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.*, **38**, 413–443.

33. Klymenko,T. and Muller,J. (2004) The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Reports*, **5**, 373–377.

34. Schuettengruber,B., Ganapathi,M., Leblanc,B., Portoso,M., Jaschek,R., Tolhuis,B., van Lohuizen,M., Tanay,A. and Cavalli,G. (2009) Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol.*, **7**, e13.

35. Enderle,D., Beisel,C., Stadler,M.B., Gerstung,M., Athri,P. and Paro,R. (2011) Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Res.*, **21**, 216–226.

36. Rach,E.A., Winter,D.R., Benjamin,A.M., Corcoran,D.L., Ni,T., Zhu,J. and Ohler,U. (2011) Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.*, **7**, e1001274.

37. Bushey,A.M., Dorman,E.R. and Corces,V.G. (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol. Cell*, **32**, 1–9.

38. Juven-Gershon,T. and Kadonaga,J.T. (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.*, **339**, 225–229.

39. Hochheimer,A., Zhou,S., Zheng,S., Holmes,M.C. and Tjian,R. (2002) TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature*, **420**, 439–445.

40. Goodrich,J.A. and Tjian,R. (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat. Rev.*, **11**, 549–558.

41. Ohler,U. (2006) Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction. *Nucleic Acids Res.*, **34**, 5943–5950.

42. FitzGerald,P.C., Sturgill,D., Shyakhtenko,A., Oliver,B. and Vinson,C. (2006) Comparative genomics of Drosophila and human core promoters. *Genome Biol.*, **7**, R53.

43. Rach,E.A., Yuan,H.Y., Majoros,W.H., Tomancak,P. and Ohler,U. (2009) Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biol.*, **10**, R73.

44. MacArthur,S., Li,X.Y., Li,J., Brown,J.B., Chu,H.C., Zeng,L., Grondona,B.P., Hechmer,A., Simirenko,L., Keranen,S.V. *et al.* (2009) Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.*, **10**, R80.

45. Schwartz,Y.B., Kahn,T.G., Stenberg,P., Ohno,K., Bourgon,R. and Pirrotta,V. (2010) Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet.*, **6**, e1000805.

46. Kharchenko,P.V., Alekseyenko,A.A., Schwartz,Y.B., Minoda,A., Riddle,N.C., Ernst,J., Sabo,P.J., Larschan,E., Gorchakov,A.A., Gu,T. *et al.* (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature*, **471**, 480–485.

47. Wysocka,J., Swigut,T., Xiao,H., Milne,T.A., Kwon,S.Y., Landry,J., Kauer,M., Tackett,A.J., Chait,B.T., Badenhorst,P. *et al.* (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, **442**, 86–90.

48. Badenhorst,P., Xiao,H., Cherbas,L., Kwon,S.Y., Voas,M., Rebay,I., Cherbas,P. and Wu,C. (2005) The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes Dev.*, **19**, 2540–2545.

49. Hirose,F., Yamaguchi,M., Handa,H., Inomata,Y. and Matsukage,A. (1993) Novel 8-base pair sequence (Drosophila DNA replication-related element) and specific binding factor involved in the expression of Drosophila genes for DNA polymerase alpha and proliferating cell nuclear antigen. *J. Biol. Chem.*, **268**, 2092–2099.

50. 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.

51. 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.

52. Li,X.Y., Thomas,S., Sabo,P.J., Eisen,M.B., Stamatoyannopoulos,J.A. and Biggin,M.D. (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol.*, **12**, R34.

53. Kaplan,T., Li,X.Y., Sabo,P.J., Thomas,S., Stamatoyannopoulos,J.A., Biggin,M.D. and Eisen,M.B. (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet.*, **7**, e1001290.

54. Meijsing,S.H., Pufall,M.A., So,A.Y., Bates,D.L., Chen,L. and Yamamoto,K.R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science*, **324**, 407–410.

55. Hart,C.M., Cuvier,O. and Laemmli,U.K. (1999) Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma*, **108**, 375–383.

56. Nechaev,S., Fargo,D.C., dos Santos,G., Liu,L., Gao,Y. and Adelman,K. (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science*, **327**, 335–338.

57. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.