Resource

Cell

Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers

Michiel Vermeulen,[1,6,7,*] H. Christian Eberl,[1,6] Filomena Matarese,[2,6] Hendrik Marks,[2] Sergei Denissov,[2] Falk Butter,[1] Kenneth K. Lee,[3] Jesper V. Olsen,[1,5] Anthony A. Hyman,[4] Henk G. Stunnenberg,[2,*] and Matthias Mann[1,*]

[1]Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, D-82152 Martinsried, Germany

[2]Department of Molecular Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen, Geert Grooteplein 26 Zuid, 6525 GA Nijmegen, The Netherlands

[3]Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA

[4]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

[5]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark

[6]These authors contributed equally to this work

[7]Present address: Department of Physiological Chemistry and Cancer Genomics Centre, University Medical Center Utrecht, Utrecht, The Netherlands

*Correspondence: m.vermeulen-3@umcutrecht.nl (M.V.), h.stunnenberg@ncmls.ru.nl (H.G.S.), mmann@biochem.mpg.de (M.M.)

DOI 10.1016/j.cell.2010.08.020

SUMMARY

Trimethyl-lysine (me3) modifications on histones are the most stable epigenetic marks and they control chromatin-mediated regulation of gene expression. Here, we determine proteins that bind these marks by high-accuracy, quantitative mass spectrometry. These chromatin "readers" are assigned to complexes by interaction proteomics of full-length BAC-GFP-tagged proteins. ChIP-Seq profiling identifies their genomic binding sites, revealing functional properties. Among the main findings, the human SAGA complex binds to H3K4me3 via a double Tudor-domain in the C terminus of Sgf29, and the PWWP domain is identified as a putative H3K36me3 binding motif. The ORC complex, including LRWD1, binds to the three most prominent transcriptional repressive lysine methylation sites. Our data reveal a highly adapted interplay between chromatin marks and their associated protein complexes. Reading specific trimethyl-lysine sites by specialized complexes appears to be a widespread mechanism to mediate gene expression.

INTRODUCTION

In the eukaryotic nucleus, DNA is wrapped around an octamer of histone proteins, which constitute the nucleosomes. Rather than merely serving as a means to store genetic material, nucleosomes play an active role in regulating processes such as transcription, DNA repair, and apoptosis. The N-terminal tails of the four core histones that protrude from the core structure of the nucleosome are subject to a variety of posttranslational modifications such as acetylation, methylation, and phosphorylation. One role of these modifications is the recruitment of regulatory proteins that in turn exert their function on chromatin (Jenuwein and Allis, 2001; Kouzarides, 2007).

The major lysine methylation sites on the N terminus of histone H3 and histone H4 with a clearly defined biological function are H3K4me3, H3K9me3, H3K27me3, H3K36me3, and H4K20me3, which are associated with different functional states of chromatin. H3K4me3 is almost exclusively found on promoter regions of actively transcribed genes while H3K36me3 is linked to transcription elongation. H3K9me3, H3K27me3, and H4K20me3 are generally found on silent heterochromatic regions of the genome. Part of the functional distinction between these methylation sites relates to the proteins interacting with them. A number of these "chromatin readers" for various histone methyl lysine sites have already been identified and characterized (Kouzarides, 2007; Shilatifard, 2006; Taverna et al., 2007), but this list is unlikely to be exhaustive. To obtain a comprehensive map of the histone methyl lysine interactome, unbiased screening methods are required.

Mass spectrometry (MS)-based proteomics is increasingly used in functional biological studies and has proved to be a powerful tool to characterize histone modifications (Garcia et al., 2007; Vermeulen and Selbach, 2009). For protein-protein interactions a quantitative format is desirable, as this enables to distinguish specific and background binders (Vermeulen et al., 2008). In particular, the technology of stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002) can be used to expose peptide baits bearing a posttranslational modification to "heavy" SILAC-labeled cell extracts, whereas the unmodified peptide is exposed to "light" labeled cell extract. Binders specific to the modified form of the peptide appear in mass spectra with a significant ratio between heavy and light form of the protein. Using this approach, we discovered that TFIID binds to H3K4me3, thereby providing a link between

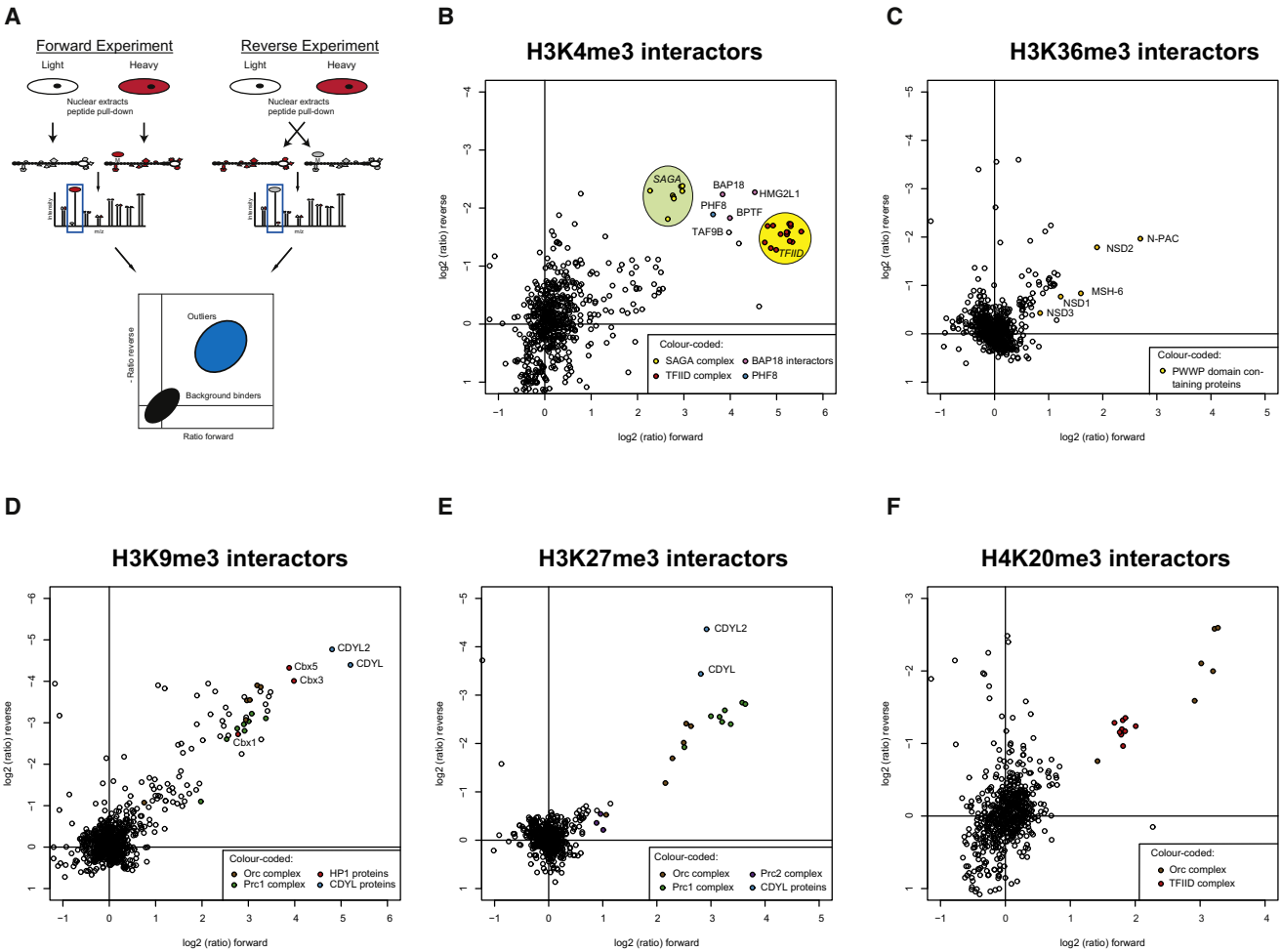

Figure 1. A Histone Peptide Pulldown Approach Using SILAC Technology

(A) Schematic representation of the experimental approach (M indicates trimethyl lysine).

(B) The H3K4me3 interactome. Proteins are plotted by their SILAC-ratios in the forward (x axis) and reverse (y axis) SILAC experiment. Specific interactors should lie close to the diagonal in the upper right quadrant. The two major transcriptional coactivator complexes that were found to interact with this mark (TFIID and SAGA) are encircled. TAF9b, which is localized between TFIID and SAGA in the figure, is a shared subunit between these two complexes.

(C) The H3K36me3 interactome. Proteins carrying a PWWP domain are colored yellow.

(D–F) The interactome of H3K9me3, H3K27me3 and H4K20me3, respectively. Note that the ORC complex, including LRWD1, binds to these three marks.

See also Figure S1 and Table S1.

this modification and activation of transcription (Vermeulen et al., 2007).

Here, we refine this technology and perform an unbiased interaction screen for the known activating and repressive trimethyl histone marks on H3 and H4. We apply the BAC-GFP transgeneOmics technology (Poser et al., 2008) to characterize chromatin readers and their complexes. Chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) with the same BAC-GFP lines identifies the in vivo target genes, which are found to overlap with the histone marks they interact with. This integrative approach provides not only an interactome of the studied histone marks, including many previously uncharacterized factors, but also mechanistic insights into epigenetic regulation of gene expression.

RESULTS

A Large-Scale Methyl Lysine Interactome

To characterize the interactome of trimethyl-lysine chromatin marks, we developed an interaction screen based on a recently described technology (Vermeulen et al., 2007). In brief, nuclear extracts derived from HeLaS3 cells grown in "light" or "heavy" medium were incubated with immobilized biotinylated histone peptides (Figure 1A). After incubation, beads from both pulldowns were pooled, run on a one-dimensional PAGE gel, and subjected to in-gel trypsin digestion. The resulting peptide mixtures were measured by high-resolution on-line electrospray MS on a hybrid linear ion trap, Orbitrap (see Experimental Procedures). Computational analysis was done with the MaxQuant

algorithms (Cox and Mann, 2008), which enabled sub parts-per-million mass assignment and accurate quantitation even for very low abundance SILAC pairs. Eluates from methylated and non-methylated peptides each contained hundreds of proteins and are visually indistinguishable on 1D gels (Figure S1A available online). Nevertheless, the SILAC-ratios reliably retrieved specific binders even when they were hundred-fold less abundant than background binders (Figure S1B). We determined the interactome of the two activating marks H3K4me3 and H3K36me3 and three repressive marks, H3K9me3, H3K27me3, and H4K20me3 (Table S1; Figures 1B–1F). Each measurement identified between 600 and 1200 proteins at a confidence level of 99%. Of these, between 10 and 60 had highly significant ratios indicating specific binding to the respective marks.

In our previous study, we identified interactions of members of the TFIID complex with H3K4me3. Here, we performed the interaction screen in the "forward" and "reverse" format to obtain higher discrimination between specific baits and background. The forward experiment consists of incubating the modified peptide with heavy labeled cell lysate and the nonmodified peptide with light labeled cell lysate, whereas in the reverse format the labels are switched. These two experiments also constitute a biological replicate. With a minimum of two quantification events, every significant interactor is supported by at least four quantitative measurements. Plotting interaction data for H3K4me3 in a two-dimensional space and inverting the SILAC-ratios of the reverse experiment places the true interactors into the top right quadrant (Figure 1A). Nonlabeled contaminants, such as keratin and proteins derived from the medium will not change the ratio in the reverse experiment and are located in other quadrants. Furthermore, a number of other proteins, such as polypyrimidine tract-binding protein 2, were automatically filtered out because they show significant ratios only in one of the labeling experiments, and are color coded accordingly in Table S1. In some cases, interactions may be biophysically correct but they may not occur in vivo because of compartmentalization in the cell (for example, mitochondrial hsp60 binding to H3K9me3). We noticed that the entire TFIID protein complex clustered together in the two-dimensional plot, indicating very similar SILAC ratios in the forward and reverse experiments (Figure 1B). This prompted us to inspect the interaction plots for other protein complexes binding to specific chromatin marks.

Sgf29 Links the Human SAGA Complex to H3K4me3

The measured H3K4me3 interactome contained eight subunits of the human SAGA complex, which tightly clustered together in the two-dimensional plot (green circle in Figure 1B). Inspection of the sequences of all known SAGA subunits revealed a double Tudor domain in the C terminus of Sgf29 (Figure 2A). Double Tudor domains are known to have affinity for H3K4me3 (Huang et al., 2006). We therefore speculated that Sgf29 could be the subunit within the SAGA complex that directly binds to H3K4me3. To address this question, we used RNAi to knock down Sgf29 in HeLa cells (Figure 2B). The nuclear extracts derived from these cells as well as nuclear extracts derived from cells transfected with control oligonucleotides were used for peptide pulldowns. Western blotting shows that the SAGA subunit GCN5 only binds to H3K4me3 and not to H3K4me0 (Figure 2B). This binding is abolished upon knockdown of Sgf29, while GCN5 levels in these cells are similar to those in the cells treated with mock siRNA. These experiments also imply that, at least in mammalian cells, Sgf29 is responsible for the observed interaction between H3K4me3 and SAGA, and not CHD1, as has been suggested in yeast (Pray-Grant et al., 2005).

To biophysically characterize this interaction, we expressed Sgf29 as a recombinant protein in *E. coli* and used the induced bacterial lysates for histone peptide pulldowns. As shown in Figure 2C, Sgf29 binds to histone H3 peptides, with a clear preference for H3K4me3. This binding is specific as no interaction with other histone lysine methylation sites such as H3K9me3 or H3K36me3 was observed. Sgf29 binds to both H3K4me2 and H3K4me3 with a slight preference for H3K4me3 (Figure 2D). Based on sequence alignments between yeast, *Drosophila* and human Sgf29 we selected conserved and nonconserved residues for mutational analyses (Figure 2A). Results of nine pulldown experiments revealed that conserved residues in the second Tudor domain of Sgf29 are particularly important for H3K4me3 binding. As expected, mutating nonconserved residues did not affect the binding (Figure 2E). We used isothermal calorimetry experiments to measure the affinity of the interaction between Sgf29 and H3K4me3 (Figure 2F). The binding constant of 4 μM is comparable to that of other trimethyl-lysine marks to their readers and in particular to the interaction constant of the Tudor domain of JMJD2A, which is 10 μM (Huang et al., 2006). No affinity between Sgf29 and the unmethylated histone H3 peptide could be observed. Together, these results demonstrate that the human SAGA complex binds to H3K4me3 and that the double Tudor domain in its subunit Sgf29 is both necessary and sufficient to mediate this interaction.

Functional Insights into Chromatin Readers Using BAC transgeneOmics

Our screening of the H3K4me3 and H3K36me3 interactome, two lysine methylations associated with actively transcribed genes, revealed a large number of chromatin readers of unknown function. To gain insight into the molecular mechanism of their interaction with the lysine methylation sites, we tagged a selection of these proteins with GFP using the recently developed BAC transgeneOmics technology (Poser et al., 2008). In this strategy, a GFP-tagged fusion of the protein of interest is stably integratedpreserving the endogenous genomic context—in HeLa cells by recombineering (Zhang et al., 1998). Fusion proteins are therefore expressed at near endogenous levels, as demonstrated previously (Poser et al., 2008). Furthermore, we tested expression levels of several of the GFP-tagged BAC lines and found very similar expression levels to the endogenous proteins (Figures S2E–S2H).

Quantitative SILAC-based GFP pulldowns employing wild-type parental cells as control were optimized such that protein complexes can be identified and visualized in a single two hour MS analysis without the need to separate proteins on an SDS PAGE gel (Hubner et al., 2010). As a proof of principle we applied this workflow to the K4me3 binding protein Sgf29, which is known to assemble into either the SAGA or the ATAC complex (Nagy et al., 2010). Both SAGA and ATAC complex subunits co-purified with GFP-Sgf29 demonstrating the applicability of single

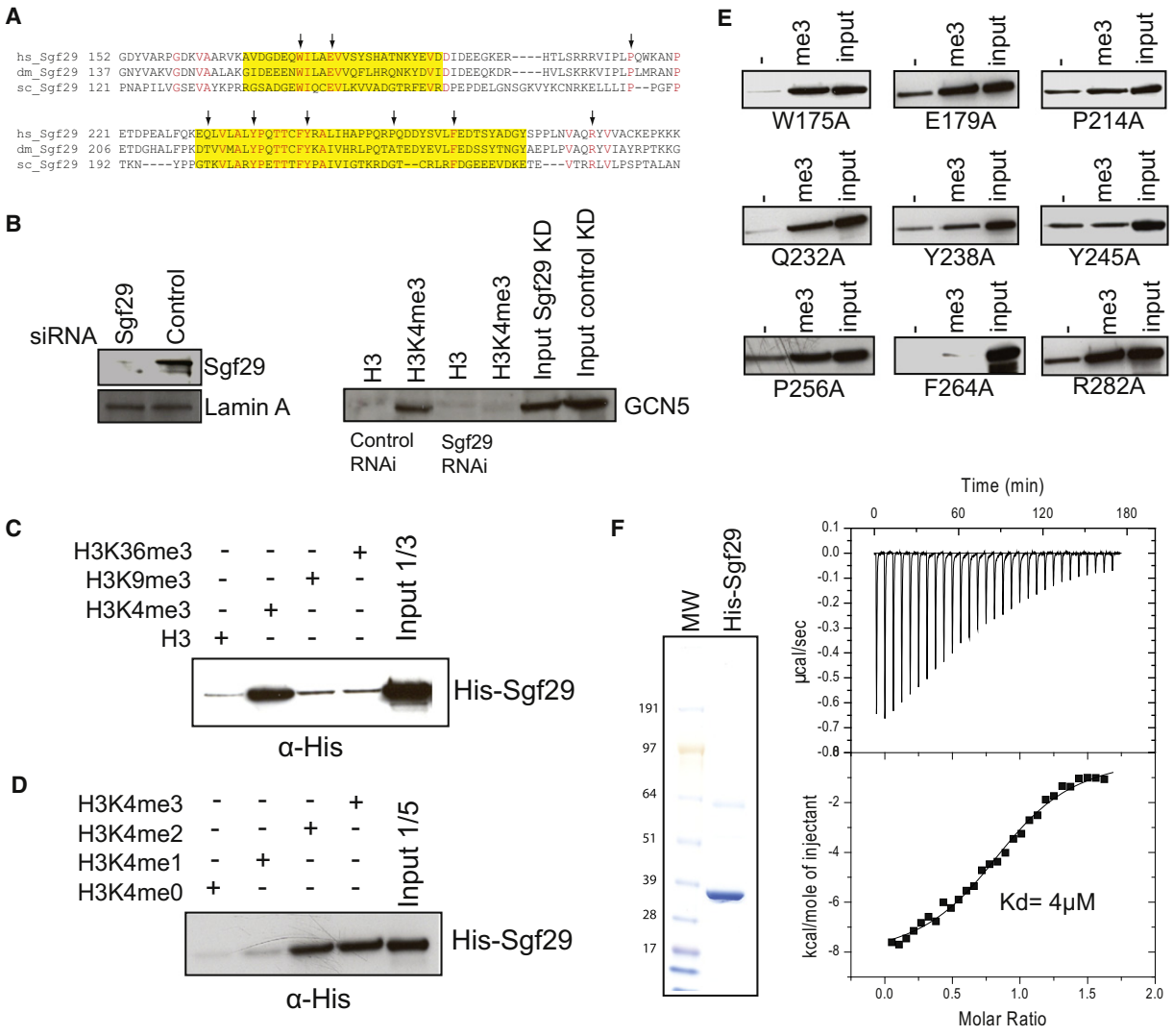

Figure 2. Sgf29 Links the SAGA Complex to H3K4me3

(A) Alignment of the C-terminal part of human, Drosophila, and yeast Sgf29. Tudor domains are indicated in yellow.

(B) siRNA experiments followed by peptide pulldowns show that Sgf29 links the SAGA complex to H3K4me3.

(C and D) Bacterial lysates expressing recombinant his-tagged Sgf29 were incubated with the indicated peptides. Following incubation and washes, the amount of bound Sgf29 protein was determined by western blotting using an anti-His antibody.

(E) Bacterial lysates expressing the indicated Sgf29 mutants were used for histone peptide pulldowns to determine their binding affinity for H3K4me3. The first lane represents peptides without the me3 modifications.

(F) Isothermal calorimetry experiment revealing the affinity of the full-length Sgf29 protein for H3K4me3.

step GFP affinity purification to identify protein-protein interactions for chromatin readers (Figure 3A; Table S2). We then applied this approach to the as-yet uncharacterized protein C17orf49, which we had found to interact with H3K4me3 (Figure 1B). C17orf49 is an 18 kDa protein that carries a SANT domain, which commonly occurs in chromatin associated proteins. Pulldown of the GFP fusion protein from stably transfected HeLa cells specifically copurified subunits of the human NuRF/BPTF complex (Figure 3B; Table S2). Strikingly, HMG2L1, another highly significant interactor of H3K4me3 (Figure 1B) is one of the most prominent interactors of C17orf49. Thus, this experiment established C17orf49 and HMG2L1 as subunits of the human NuRF/BPTF complex. Their association with H3K4me3 is explained by their interaction with the H3K4me3 reader BPTF. We name the uncharacterized open reading frame C17orf49 as "BPTF associated protein of 18 kDa" (BAP18).

GATA zinc finger domain containing 1 (GATAD1) is another protein of unknown function that was identified as a H3K4me3 interactor. Using the GFP pulldown approach, we identified subunits of the Sin3b/HDAC complex, the H3K4me3-specific lysine demethylase Jarid1A/RBBP2, and the breast cancer

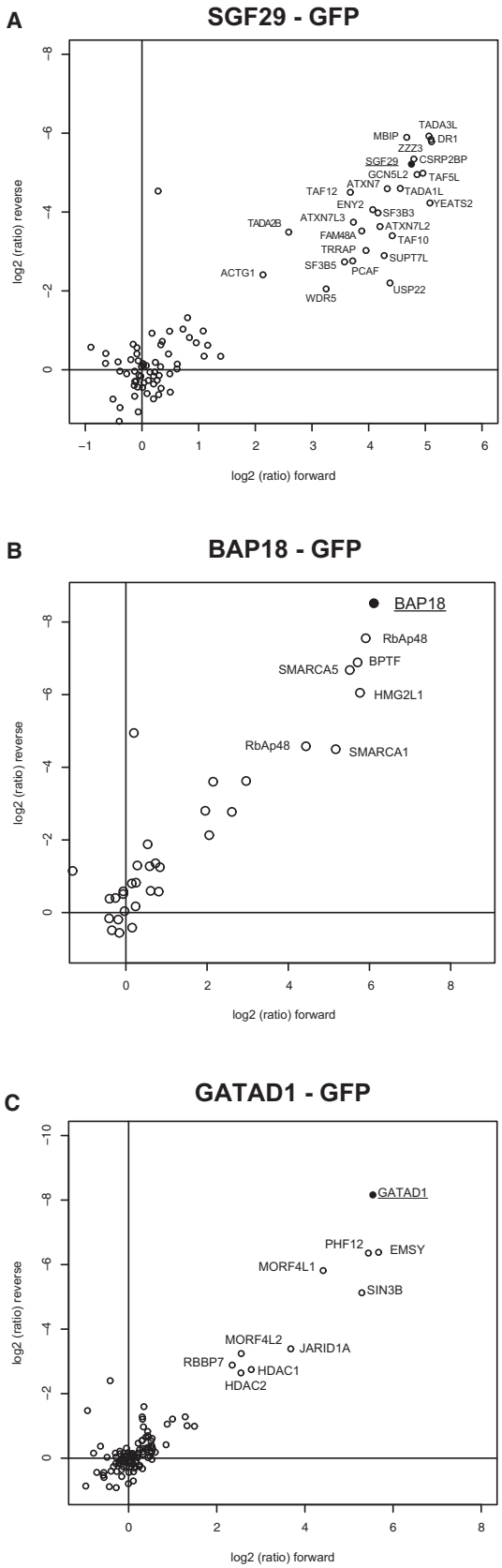

Figure 3. GFP Pulldowns for H3K4me3 Readers

HeLa Kyoto cells expressing GFP-Sgf29 (A), GFP-C17orf49/BAP18 (B), and GFP-GATAD1 (C) were SILAC-labeled and subjected to single-step affinity purifications in a "forward" and reverse" pulldown using GFP nanotrap beads. In each panel the ratio of the identified proteins in the forward and reverse pulldown is plotted. Proteins interacting with the baits are indicated.
See also Figure S2 and Table S2.

associated protein EMSY (Hughes-Davies et al., 2003) as interactors for GATAD1 (Figure 3C; Table S2). Because all of the subunits in this complex were identified as H3K4me3 readers with similar ratios, we hypothesized that they form an as-yet uncharacterized chromatin reading complex (Figures S2A–S2D). Jarid1a was recently reported to bind tightly to H3K4me3 with a Kd of 0.75 μM (Wang et al., 2009a) and therefore forms the direct link between the complex and the chromatin mark. Further evidence for our hypothesis comes from a subsequently published *Drosophila* Lid complex (Lee et al., 2009; Moshkin et al., 2009). Lid is the *Drosophila* homolog of the mammalian Jarid1 family of proteins, consisting of Jarid1a, Jarid1b and Jarid1c. The complex furthermore contains homologs of the Sin3 proteins, as well as an EMSY and GATAD1 homolog. In mammals, interactions between the Sin3/HDAC complex and Jarid1a have also been reported (van Oevelen et al., 2008). However, EMSY has not been tied to any of these proteins yet. EMSY is known to be a repressor of transcription (Hughes-Davies et al., 2003) but the mechanisms underlying this repressive activity are poorly understood. The identification of the above-described complex provides important clues as to how EMSY represses transcription. We hypothesize that gene repression involves histone deacetylation coupled with H3K4me3 demethylation.

Localizing the Chromatin Readers on the Genome

To further investigate the function of our proteins of interest in vivo, we performed ChIP-Seq profiling using an anti-GFP antibody on the BAC-GFP lines. Figure 4A shows a representative snapshot of the ChIP-Seq data. Profiling of GFP-tagged proteins interacting with H3K4me3 and H3K36me3 was performed on biological replicas and showed that the approach is highly reproducible (Pearson correlation >0.85; Figures S3F and S3G). In agreement with our peptide pulldown data, the identified H3K4me3 readers Sgf29, TRRAP, PHF8, GATAD1, and BAP18, are associated mainly with promoters (Figures S3A and S3B) and coincide with H3K4me3 marking (Figures 4B and 4C; Figure S3C). We also identified a small number of binding sites of H3K4me3 readers outside of annotated promoters (Figure S3A). As these are not associated with H3K4me3 (Figure S3B), the interactor proteins are apparently recruited to these loci by H3K4me3 independent mechanisms. Nevertheless, for each of these five proteins we observed a good genome-wide correlation with H3K4me3 (Pearson correlation BAP18: 0.71, GATAD1: 0.71, PHF8: 0.66, TRRAP: 0.66, SGF29: 0.55).

For Sgf29, TRRAP, and BAP18, it was expected that they would localize to promoters, as they are part of conserved complexes associated with active transcription– SAGA/ATAC, SAGA/NuA4, and BPTF/NuRF, respectively (Nagy et al., 2010; Wysocka et al., 2006). PHD finger protein 8 (PHF8) belongs to

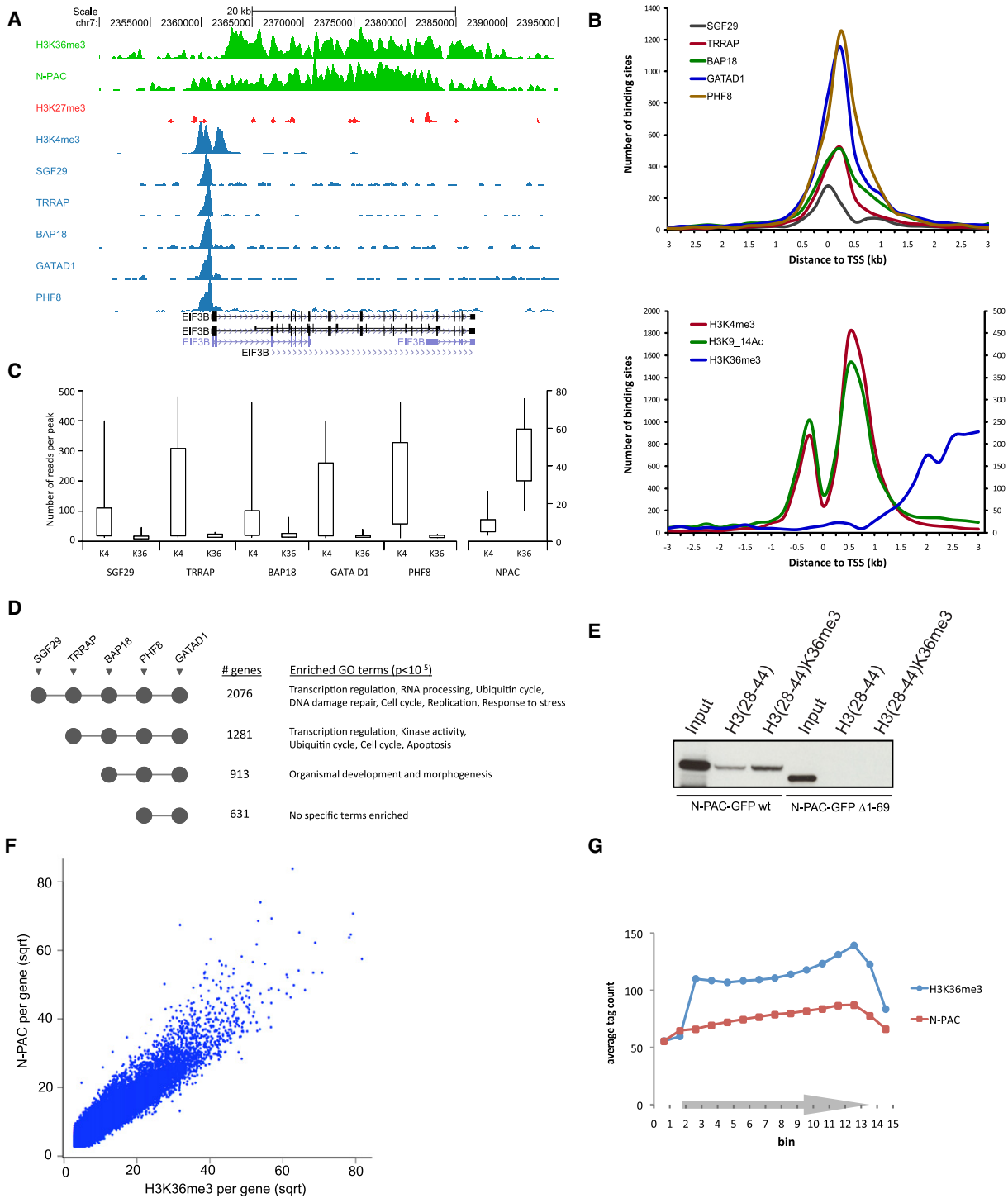

Figure 4. ChIP Sequencing of H3K4me3 and H3K36me3 Readers

(A) ChIP-Seq profiles of three histone modifications and the interactors across the Eif3B gene on human chromosome 7.

(B) Distance distribution of the binding sites for the H3K4me3 interactors and the three histone modifications relative to the closest transcription start site (TSS). x axis is in 1000 bp; on the y axis the number of binding sites is indicated. Values for H3K36me3 are plotted on a separate scale (right side).

(C) Number of reads for H3K4me3 and H3K36me3 (indicated with K4 and K36, respectively) within the binding sites for the H3K4me3 interacting proteins. The ends of the whiskers represent the 9th and 91st percentile, respectively. Values for SGF29, TRRAP, BAP18, PHF8, and GATAD1 are on the scale on the left side of the plot, while values for N-PAC are on a separate scale on the right.

(D) Promoters clustered by the binding sites for the H3K4me3 interacting proteins (Figure S3). Co-occurrence of binding sites is indicated with gray circles under

the JmJc domain-containing family of proteins that can remove methyl groups from arginine or lysine residues (Cloos et al., 2008). PHF8 can remove the repressive mark H3K9me2 (Horton et al., 2010), associating it with activation of transcription, which is in agreement with our ChIP-Seq analyses.

We found GATAD1 to interact with Jarid1a/EMSY/Sin3 (Figure 3C). Jarid1a is a JmJc domain-containing protein that demethylates H3K4me3 (Cloos et al., 2008). In addition, the GATAD1 purification enriched for components of the Sin3/HDAC transcriptional corepressor complex, including two histone deacetylases, HDAC1 and HDAC2. Despite the repressive enzymatic activities associated with GATAD1, our ChIP-Seq analysis reveals that this complex binds to promoters marked with H3K4me3. These data may be explained by invoking a mechanism of cyclical recruitment of ''writers'' and ''erasers'' to sites of active transcription (Wang et al., 2009c).

Interestingly, our ChIP-Seq analyses showed that many target genes can be occupied by each of the five H3K4me3 readers. Analysis of all identified target genes resulted in four discrete clusters (Figures S3D and S3E; Table S3). PHF8 and GATAD1 were the only factors found to be common to all clusters and therefore are likely to have a general role in transcription. The two largest clusters combined genes whose promoters were bound by Sgf29 and/or TRRAP, indicating that transcriptional regulation of these genes involves SAGA/NuA4-related complexes. Gene ontology (GO) annotation of the genes in these clusters revealed a number of highly enriched ($p < 10^{-5}$) functional terms that agree very well with the biological functions of these complexes (Figure 4D). For example, SAGA/ATAC and NuA4 complexes are crucial regulators of transcription, DNA repair, DNA replication, and the cell cycle (Squatrito et al., 2006). Distinct GCN5/PCAF-containing complexes function as coactivators and are involved in transcription factor and global histone acetylation (Nagy and Tora, 2007). SAGA was shown to regulate various stress-response genes (Huisinga and Pugh, 2004; Nagy et al., 2010), while TRRAP-containing complex NuA4 regulates apoptosis (Ikura et al., 2000; Tyteca et al., 2006). Thus, each functional category of the GO analysis corresponds to an established function of the SAGA and NuA4 complex, which independently validates the connection between the activating histone mark and its reader found in our experiments.

N-PAC, MSH-6, and NSD1 as well as NSD2 were identified as H3K36me3 interactors (Figure 1C; Table S2). Interestingly, these four proteins share a PWWP domain which is part of the Tudor domain ''Royal Family'' and includes the Tudor, chromo and MBT domains that can interact with methylated lysine residues. The PWWP domain of Set9 was recently identified as a reader for H4K20me1 (Wang et al., 2009b). Our peptide pulldown data suggest that this domain is also capable of recognizing H3K36me3, which is associated with elongation of transcription and peaks in coding regions of genes (Shilatifard, 2006). Very recently the PWWP domain of Brpf1 was shown to bind specifically to H3K36me3 (Vezzoli et al., 2010). Indeed, deletion analyses revealed that the PWWP domain of N-PAC is necessary for H3K36me3 binding (Figure 4E). This PWWP domain mediated K36me3 binding is most likely direct, since purification of N-PAC-GFP from a BAC line did not reveal protein-protein interactions (data not shown). To investigate the genomic binding pattern of N-PAC, we generated the corresponding BAC-GFP line and performed ChIP-Seq analysis. Consistent with our peptide pulldown data, N-PAC binds to coding regions of active genes correlating with the presence of H3K36me3 (Figures 4C and 4F). N-PAC and H3K36me3 increase toward the 3′ end (Figures 4A and 4G). Together our data establish the PWWP domain as a putative binder of H3K36me3. In addition to a PWWP domain, N-PAC also contains an AT-hook that is often found in proteins that are associated with elongation of transcription and an enzymatic domain of unknown function. Our ChIP-Seq analysis revealed that both H3K36me3 and N-PAC are present almost exclusively over gene bodies (data not shown), and that the vast majority of H3K36me3 marked regions are also bound by N-PAC, indicating a broad or universal function of this protein in transcriptional elongation.

The Interactome of the Repressive Histone Methyl Marks

We next investigated the chromatin readers of H3K9me3, H3K27me3 and H4K20me3, histone methyl marks associated with gene repression (Figures 1D–1F). H3K9me3 yielded the richest set of interactors, including all three HP1 isoforms (CBX1, CBX3, and CBX5). The chromodomain-containing HP1 proteins are classical readers of H3K9me3 (Jenuwein and Allis, 2001) and our analysis confirms that they are restricted to this repressive modification. Two chromodomain proteins, CDYL and CDYL2, were identified as binders for both H3K9me3 and H3K27me3 but not H4K20me3. These proteins are members of a family of three chromodomain proteins, the third one being chromodomain Y protein, whose gene is located on the Y chromosome and whose expression is testis specific. Recently, direct binding of CDYL and CDYL2 to H3K9me3 and H3K27me3 has been reported (Fischle et al., 2008; Franz et al., 2009). As expected, Polycomb group proteins represent the major readers for H3K27me3, but many of these proteins were also identified as specific interactors for H3K9me3. Given the high degree of sequence identity surrounding H3K9 and H3K27 (TARKST and AARKSA for K9 and K27, respectively), it is not surprising to find Polycomb group proteins as interactors

the corresponding interactor names. Four major groups of promoters were identified, for which the number of genes within each group and highly enriched GO terms (p value $< 10^{-5}$) are listed.

(E) Full-length N-PAC-GFP and Δ1-69 N-PAC-GFP were transfected into HeLa Kyoto cells. Extracts from these cells were subsequently used for K36/K36me3 peptide pulldowns. Unlike the wild-type protein, Δ1-69 N-PAC-GFP, that lacks most of the PWWP domain, does not bind to H3K36me3.

(F) Dotplot showing the correlation between H3K36me3 and N-PAC ($R^2 = 0.86$). Every dot represents the number of N-PAC or H3K36me3 ChIP-Seq tags per gene.

(G) All genes containing H3K36me3 (>5 kb) were each divided in 15 bins followed by counting and averaging of the H3K36me3 and N-PAC ChIP-Seq tags within each bin.

See also Figure S3 and Tables S3 and S4.

for H3K9me3. Literature evidence also supports the interaction of Polycomb group proteins with H3K9me3, although their affinity for H3K27me3 is higher (Fischle et al., 2003b; Ringrose et al., 2004). Finally, we identified the origin recognition complex (ORC) as an interacting complex for all three repressive sites.

We purified complexes associated with the HP1 family members to ascertain if the H3K9me3 readers physically interact with them using BAC-GFP constructs (Figures 5A–5C). Among the specifically interacting proteins, known HP1 interactors were identified, such as chromatin assembly factors CHAF1A/CHAF1B and ADNP (Lechner et al., 2005; Mandel et al., 2007). Two uncharacterized proteins, POGZ and Znf828, consistently interacted with high ratios with all HP1 family members. We confirmed the binding of POGZ to H3K9me3 by western blotting (Figure S1C). POGZ and Znf828 have an interesting domain structure and multiple zinc fingers, suggesting that these proteins may specifically bind DNA sequences. POGZ or POGO transposable element with a ZNF domain is a 1410 amino acid protein containing two domains that are also present in the centromeric protein B (CenPB). Next, we generated BAC-GFP constructs for these proteins. Pulldowns with POGZ and Znf828 reciprocally confirmed interaction with HP1 and, interestingly, with each other (Figures 5D and 5E). Additionally, POGZ interacted specifically with mitotic spindle checkpoint protein, Mad2l2. To substantiate this possible connection to a prominent cell cycle protein, we performed a GFP pulldown with a cell line of this protein, which clearly demonstrated reciprocal binding (Figure 5F). Thus, a combination of repressive mark interactors and full-length protein interactomes allows us to deconstruct the majority of protein interactions involved in the biology of the repressive marks.

We noticed that LRWD1 clusters together in the two-dimensional interaction plots with the ORC complex in the pulldowns of each of the repressive marks (Figures 1D–1F). LRWD1 has not been characterized but obtains its name from a leucine-rich repeat and a stretch of WD40 domains. To test if this protein is a subunit of the ORC complex, we generated the BAC-GFP cell line of Orc2L. Pulldown with this ORC subunit indeed demonstrated specific interaction with LRWD1 (Figure 5G). Furthermore, ChIP-Seq of the BAC LRWD1-GFP line revealed a strong enrichment on satellite repeats, correlating with high levels of H3K9me3 which is known to be enriched over satellites (Figure 5H) (Martens et al., 2005).

Triple SILAC Pulldowns Reveal Differential Fine-Tuning of Trimethyl Lysine Binding

The five trimethyl lysine marks that we screened for interactors are flanked by numerous residues that can also be subjected to posttranslational modifications. These modifications could, either agonistically or antagonistically, affect trimethyl lysine binding. To study such potential interplay between different posttranslational modifications (PTMs) occurring in close proximity on the histone H3 tail, we applied triple pulldown experiments involving a combination of methylation and other PTM marks, in this case acetylations or phosphorylations (Vermeulen et al., 2007). In this approach, cells are grown in three different SILAC media, each containing different stable isotopic versions of lysine and arginine. These extracts, which are distinguishable by MS, are each incubated with a differently modified histone peptide (triple pulldown). Peptides appear as triplets in the MS spectra and a significant ratio between the first two peaks indicates specific binding to the H3K4me3 mark. The highest mass peak in the triplet originates from the eluate of the combinatorially modified peptide and its intensity compared with the eluate from the singly modified peptide (middle peak) indicates either agonistic or antagonistic binding or no effect. On genes that are actively being transcribed, H3K4me3 often co-occurs with acetylation of H3K9 and H3K14. A number of readers for H3K4me3 carry both a domain that recognizes H3K4me3 as well as one or multiple bromodomains, which bind to acetylated lysine residues. We therefore wondered whether these acetylations would function agonistically with H3K4me3 to bind H3K4me3 readers to the histone H3 tail. Consistent with our previous findings (Vermeulen et al., 2007), TFIID and BPTF bound more strongly to the H3K4me3 mark when it was flanked by acetylation on H3K9 and H3K14 acetylation (Figures S4A and S4B). In addition, we also observed—by quantitative proteomics and by western blotting—agonistic binding to the methylated and acetylated peptide for the SAGA complex (Sgf29 in Figures 6A and 6C). In contrast, recombinant Sgf29 does not display preferential H3K9,14Ac binding (Figure 6D), indicating that the observed effects in the triple pulldown are due to the agonistic binding effects of the Sgf29 double Tudor domain and the GCN5 bromodomain. Finally, we also observed agonistic binding of PHD finger protein 8 (PHF8) to H3K4me3 and H3K9,14 Ac (Figure 6B). PHF8 carries an H3K4me3-binding PHD finger (Horton et al., 2010), but it does not contain a bromodomain. Therefore, we hypothesize that this protein either carries an unidentified acetyl lysine binding motif, or interacts with an as-yet unidentified bromodomain-containing protein. These results indicate that agonistic H3K4me3 and H3K9,14Ac recognition occurs in several chromatin readers. The mechanisms are diverse; for example, a PHD finger domain can be combined with a bromodomain in one protein (BPTF), or in different subunits of the same complex (TAF3 PHD finger and TAF1 bromodomains in the TFIID complex). Moreover, a different recognition domain combination can be used (Tudor domain of Sgf29 with the bromodomain of GCN5 in the SAGA complex). Clearly, these chromatin readers have each evolved the ability to target combinatorially marked nucleosomes allowing regulation of specific subsets of genes.

To study potential antagonistic histone PTM crosstalk, we decided to focus on phosphorylations on the histone H3 tail. Phosphorylation of histone H3S10 results in the release of HP1 from chromatin during mitosis even though levels of H3K9me3 remain unchanged (Fischle et al., 2005). H3K27me3 is also flanked by a serine residue that can be phosphorylated (Winter et al., 2008). To investigate if these trimethylations co-occur with the respective adjacent phosphorylations, we analyzed our recent large-scale study of the proteome and the phosphoproteome of the cell cycle (Olsen et al., 2010). Indeed, we found the corresponding doubly modified peptides. Moreover cell cycle data indicates that they are specific for mitotic cells (Figures S4G–S4J). As shown in Figure 6E, H3S10 phosphorylation does not appear to drastically affect the binding of HP1 to H3K9me3. These results are in agreement with data reporting

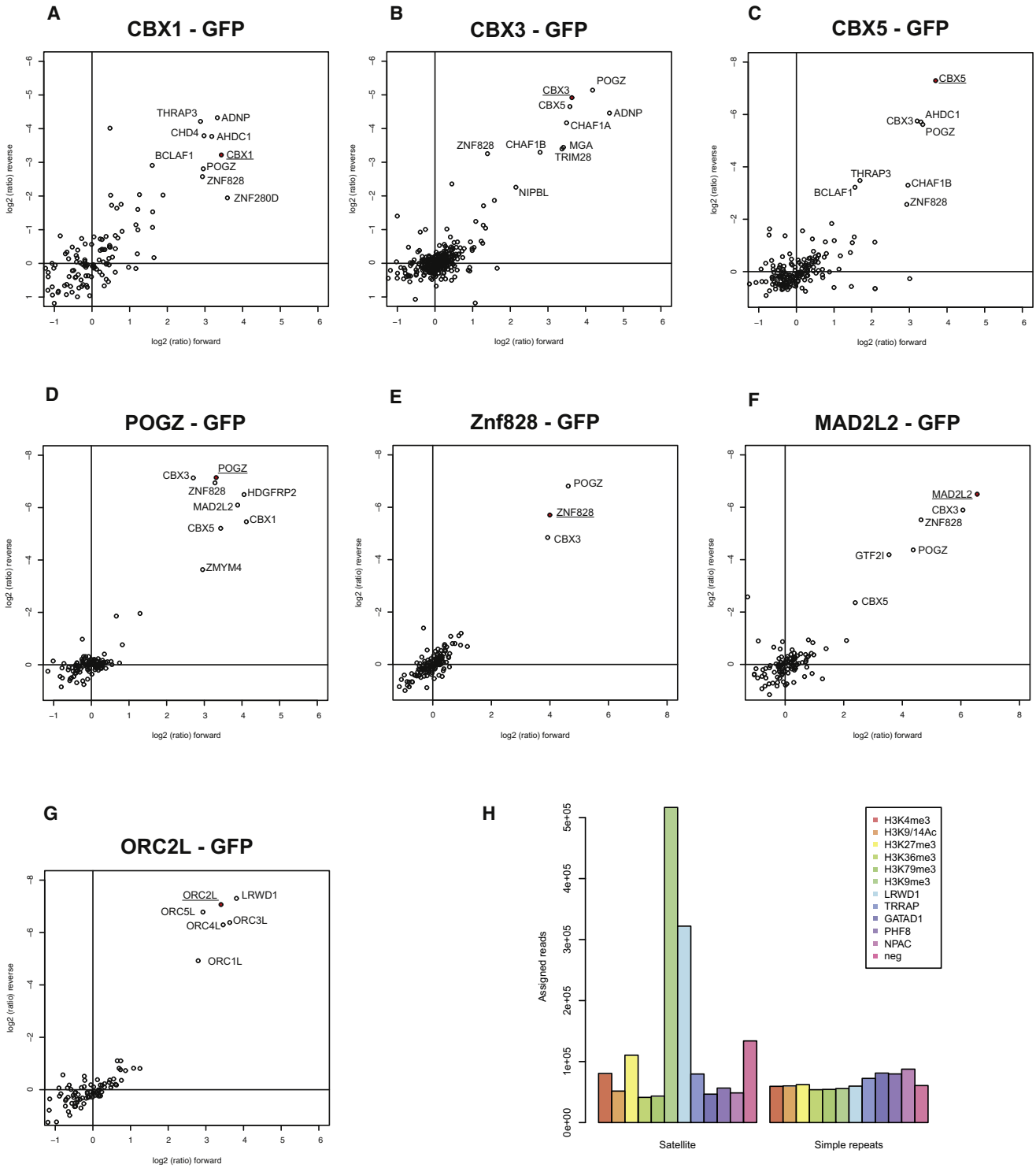

Figure 5. GFP Pulldowns for Readers of the Repressive Histone Marks

(A–G) GFP-fusions proteins expressed in SILAC-labeled HeLa cells were enriched on GFP-nanotrap beads. In each figure, the ratio of the identified proteins in the forward and reverse pulldown is plotted. Proteins interacting with the baits are indicated.

(H) The total number of ChIP-Seq reads present on either satellite repeats or simple repeats for the indicated proteins and histone marks is shown.

See also Table S2.

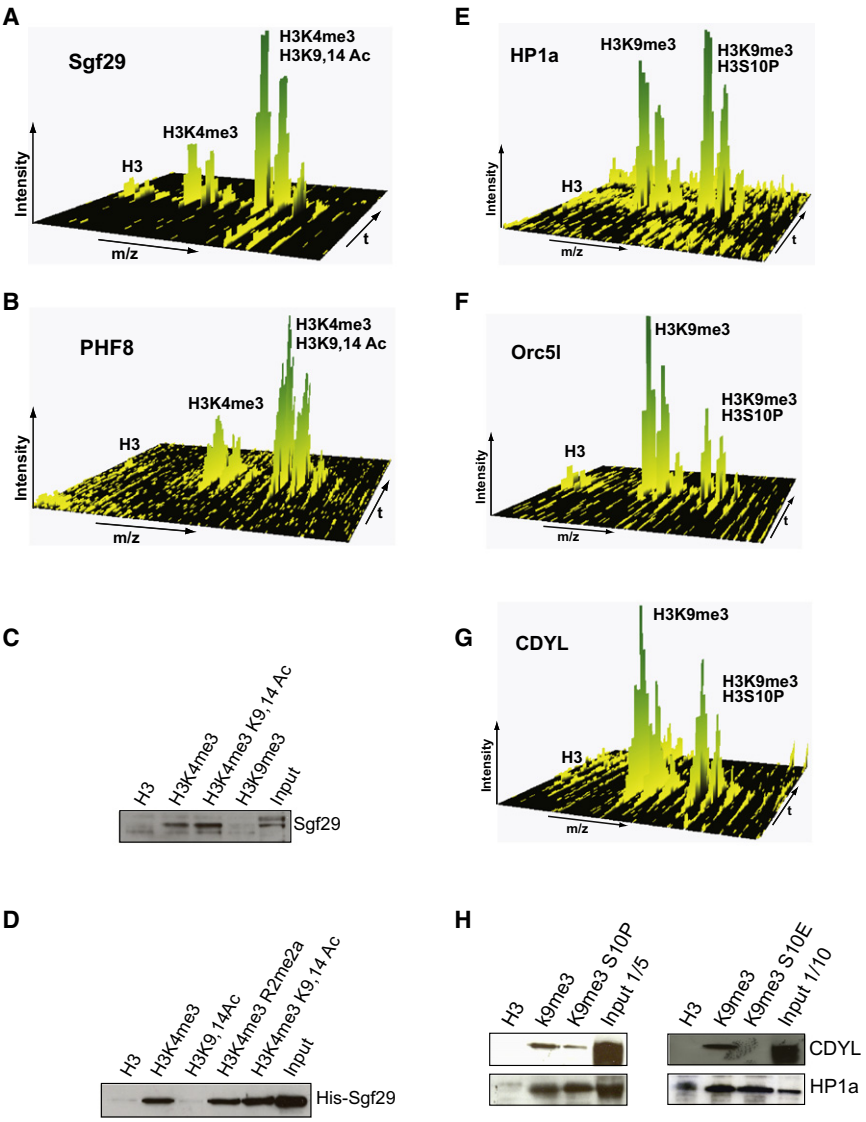

Figure 6. Triple SILAC Pulldowns Revealing Histone Modification Crosstalk

(A) Three-dimensional representation of the MS signal of an Sgf29 peptide identified in a triple pulldown SILAC experiment (the m/z scale is the x axis, the chromatographic retention time is the y axis, and the MS-signal is the z axis). Each group of signals represents the natural isotope pattern of the peptide. The relative intensities of the triplet peak of the Sgf29 peptide indicates the preference of binding to the modification states (unmethylated histone H3 peptide [left peak], H3K4me3 peptide [middle peak], and the double-modified H3K4me3/H3K9,14 Ac peptide [right peak]).

(B) Same as (A) for a PHF8 peptide identified in the same triple pulldown.

(C) Nuclear extracts derived from HeLa cells were incubated with the indicated histone peptides. The amount of Sgf29 protein bound to these peptides was determined by western blotting using an antibody against endogenous Sgf29.

(D) Bacterial lysates expressing recombinant His-tagged Sgf29 were incubated with the indicated peptides. Following incubation and washes, the amount of bound Sgf29 protein was determined by western blotting using an anti-His antibody. Note that Sgf29 does not bind to a peptide containing H3K9,14 acetylation and that the binding of Sgf29 to H3K4me3 is not affected by asymmetric dimethylation of H3R2.

(E–G) Three-dimensional representation of an HP1α (E), Orc5 (F), and CDYL (G) peptide identified in a triple pulldown SILAC experiment. The spectra show the MS-signal representing the relative binding of these peptides to the unmethylated histone H3 peptide (left peak), the H3K9me3 peptide (middle peak), and the double-modified H3K9me3/H3S10P peptide (right peak).

(H) Histone peptide pulldowns in HeLa nuclear extracts were performed with the indicated peptides. The amount of HP1α and CDYL binding to these peptides was determined by western blotting using an antibody against HP1α and CDYL.

See also Figure S4.

stabilization of HP1 binding by H3S10 phosphorylation (Mateescu et al., 2004). Indicating that our assay can indeed reveal antagonistic effects, we observed that CDYL as well as the ORC complex subunits do show reduced H3K9me3 binding in combination with H3S10 phosphorylation (Figures 6F and 6G). These experiments were further confirmed by western blotting, also making use of a phosphomimetic peptide where H3S10 was mutated to glutamic acid (Figure 6H). Similarly, H3S28 phosphorylation destabilizes the binding of CDYL and ORC complex subunits to H3K27me3, whereas this phosphorylation only mildly affects the binding of Polycomb group proteins (Figures S4C–S4F). Taken together, these results suggest that phosphorylations on the N-terminal tails of histones selectively affect the binding of proteins to adjacent modified lysines residues. Such so-called phospho-methyl switches are quite common on core histones (Fischle et al., 2003a). We have also identified H3S57 and H3T80 as phosphorylation sites on histone H3 (for H3S57P and H3T80, Figures S4K and S4L), both of which are adjacent to modified lysine residues. Thus, almost all of the modified lysine residues on histone H3 can be flanked by phosphorylated residues. An important function of these phosphorylation sites could be the differential regulation of protein binding to neighboring methylated or acetylated lysines in specific cellular situations and for specific genes.

DISCUSSION

Here, we have characterized the association of chromatin readers with histone trimethyl-lysine modifications by a combination of three technologies. The major findings from our integrated approach are visualized and summarized in Figure 7. High-accuracy, quantitative proteomics based on SILAC identified known and previously unknown binders to each of the chromatin marks. Plotting SILAC ratios from forward and reverse experiments grouped distinct protein clusters together, representing functional complexes. To investigate these complexes, we turned

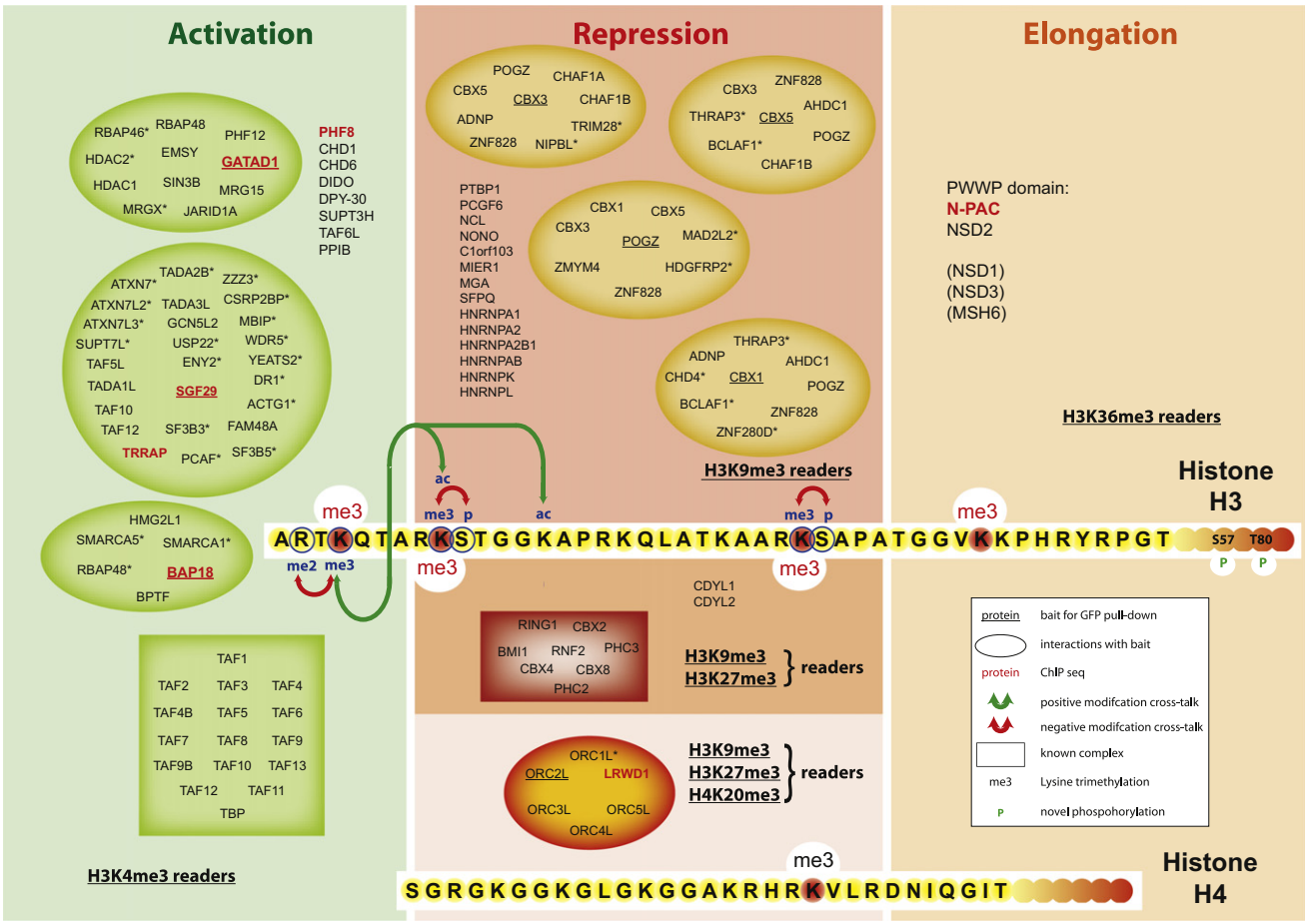

Figure 7. Visualization of the Histone Trimethyl-Lysine Interactome

Proteins interacting with the five trimethyl lysine marks are indicated. Encircled are proteins that were additionally identified in GFP pulldown experiments; baits in these pulldowns are underlined. Proteins in those circles marked with an asterisk were not identified as interactors in the peptide pulldowns. Proteins clustered in rectangles were identified in the peptide pulldowns and were previously shown to interact with each other (TFIID for H3K4me3 and PRC1 for H3K9me3 and H3K27me3). For proteins that are color coded red in vivo verification by ChIP-Seq is also provided. The arrows and associated labels indicate histone modification crosstalk investigated in this study. In the globular part of histone H3, two identified histone phosphorylations (H3S57P and H3T80), are indicated.

to the recently developed BAC-transgeneOmics technology (Poser et al., 2008), which allowed rapid generation of stable cell lines containing the entire gene of interest fused to GFP in its endogenous context. Therefore, this technology provides a generic "handle" for the members of chromatin reader complexes while maintaining endogenous control. We used these cell lines in a next round of SILAC-based quantitative interaction screens to establish physical interactions between the chromatin readers. Furthermore, the GFP-tag was utilized for chromatin immunoprecipitation followed by next generation DNA sequencing to localize the readers on the genome. The synergistic use of these three approaches enabled us to create data sets and reagents that provide a resource for researchers interested in epigenetic questions. While shown here for histone modifications, our approach can be extended to posttranslational modifications on other chromatin-associated proteins and to other cellular systems such as stem cells. Illustrating the usefulness of this resource, we were able to dissect several mechanisms of chromatin reader associations with their chromatin marks starting from basic interaction data.

One such example is the human SAGA complex, all identified members of which clustered tightly in the two-dimensional interaction plot (Figure 1B). SAGA is a highly conserved complex, which plays key roles in the activation of transcription of RNA polymerase II target genes. However, the mechanisms of activation are not completely understood. In yeast, it has been suggested that CHD1 links the complex to H3K4me3 (Pray-Grant et al., 2005). However, this association is controversial as it has been reported that yeast CHD1 does not bind to H3K4me3 (Sims et al., 2005). While we identified human CHD1 as a specific binder to this mark, it did not co-cluster with the SAGA subunits in our H3K4me3 peptide pulldowns. Furthermore, we were not able to identify CHD1 as an interactor of the SAGA subunit Sgf29 in a GFP pulldown. Instead, starting with the observation that Sgf29, which we identified as a H3K4me3 interactor, has a double Tudor domain (Lee and Workman, 2007) and given

the fact that Tudor domains can bind methylated lysines (Huang et al., 2006), we established by biochemical and biophysical means that the double Tudor domain of Sgf29 forms the direct molecular link between SAGA and H3K4me3. This binding mode is likely conserved down to yeast, which has a homolog of Sgf29 that also contains a double Tudor domain (Figure 2A). Such conservation is not universal as it is not the case for association of TFIID with H3K4me3. This interaction is mediated by the PHD-finger domain of human TAF3, but yeast TAF3 lacks the PHD-finger domain (Vermeulen et al., 2007).

Bioinformatic analysis of the interactors of the activating H3K36me3 mark revealed that four of the most prominent specific interactors shared the same domain. This PWWP domain is part of the Tudor domain "Royal family" of domains (Maurer-Stroh et al., 2003) and therefore almost certainly mediates direct binding to H3K36me3. In agreement with this, deletion analysis revealed that the PWWP domain of N-PAC is essential for its interaction with H3K36me3 (Figure 4E).

In the interactome of the repressive marks we identified, in addition to expected heterochromatin associated proteins, several other proteins. Interaction studies with BAC GFP-fusion proteins uncovered many interactions with members of the HP1 family. This HP1 family and associated proteins represent a large portion of the H3K9me3 interactome and establish the HP1 proteins as interaction hubs in mediating repressive gene functions. Interestingly, several HP1 interactors contain zinc finger domains (such as POGZ and Znf828), which may serve to recruit HP1 to specific sites in the genome.

The origin recognition complex (ORC) has a key function in replication firing. It is known to localize to heterochromatic regions (Prasanth et al., 2004) and it interacted with all three repressive marks. LRWD1 grouped with the ORC complex members in the two-dimensional interaction plots. Pulldowns with an Orc2L BAC-GFP cell line demonstrated that LRWD1 is indeed an ORC complex subunit and ChIP-Seq experiments established that it co-enriches with H3K9me3 on satellite repeats. The WD40 repeat domain of LRWD1 may mediate the interaction of the ORC complex with the repressive marks as it was recently shown that the WD40 repeats of the Polycomb group protein EED directly binds to H3K27me3 (Margueron et al., 2009).

A triple-encoding variant of the SILAC peptide pulldown allowed us to directly address the question of agonistic and antagonistic binding to combinatorial histone modifications. These experiments recapitulated several known combinatorial interactions, such as the agonistic effects between H3K4me3 and nearby acetylations. The general conclusion from these experiments is that the trimethyl marks constitute the major docking sites for chromatin readers and that other nearby modifications fine-tune these primary interactions by augmenting or destabilizing specific interactions. For example, our data show that H3S10 phosphorylation destabilizes the ORC complex and CDYL binding to H3K9me3, whereas HP1 binding does not appear to be affected. Consistent with this paradigm, we have not been able to determine specific interactors with peptides bearing only the ancillary modifications. This is unlikely to be an artifact due to pulldowns with synthetic peptides because similar results are obtained when performing pulldowns with entire immobilized nucleosomes carrying particular epigenetic marks (T. Bartke, M.V., M.M., and T. Kouzarides, unpublished data). In this context, mass spectrometry can also contribute by identifying and quantifying the combinatorially modified peptides in vivo, as shown for several examples here.

A striking finding that emerges from our integrative investigation into the nature of the relationship between histone marks and their readers is the degree of overlap between the known biological functions of the marks and the biological functions of their associated readers (Figure 7). Histone modifications are usually studied by techniques such as ChIP, ChIP-Seq, or immunofluorescence that associate them with particular genes or nuclear processes. The same holds true for transcription factors or other chromatin regulators. By its nature, our strategy combines investigation of chromatin marks and transcriptional regulators and is thereby uniquely suited as an integrative tool for the investigation of epigenetic regulation of gene expression.

EXPERIMENTAL PROCEDURES

Recombinant Protein Expression and ITC Calorimetry

Full-length Sgf29 constructs were expressed with an N-terminal His-tag and a maltose binding protein (MBP) domain using expression plasmid pETM44 (Novagen). For histone peptide pulldown experiments, crude induced bacterial lysates were used as described (Vermeulen et al., 2007). His-tag westerns were performed using a penta-His antibody (QIAGEN). For isothermal calorimetry (ITC) experiments the Sgf29 protein was enriched using Ni NTA beads after which the protein was further purified on a Superdex 200 column. ITC measurements were performed on a VP-ITC Microcal calorimeter (Microcal, Northampton, MA) at 25°C. During titration, 7 μl of H3K4me3 peptide (aa 1–17) at a concentration of 300 μM was injected into a solution of 25 μM Sgf29 protein.

GFP Pulldowns

Generation of the BACs with GFP-fusion constructs was done as described (Poser et al., 2008). Nuclear extracts from BAC-GFP-tagged or wild-type HeLa cells were SILAC labeled with heavy lysine (Isotec, Sigma). For CBX3, no BAC was available and SILAC-labeled HeLa cells were transfected with plasmid pBCHGN-CBX3 (Addgene). GFP nanotrap beads (Chromotek) were used to precipitate GFP-tagged proteins from these lysates. Approximately 500–1000 μg of nuclear extract was used per pulldown in a buffer containing 300 mM NaCl, 0.25% NP40, 0.5 mM DDT, 20 mM HEPES KOH (pH 7.9), and protease inhibitors. Following incubation and washes with the same buffer, beads from both pulldowns were combined, proteins were eluted with acidic glycine (0.1 M [pH 2.0]) and digested overnight with LysC (Wako Biochemicals, Japan) using the FASP protocol (Wisniewski et al., 2009) prior to LC/MS-MS analysis.

Mass Spectrometry of Proteins

Gel lanes representing each pulldown were cut into eight equally sized slices as described (Vermeulen et al., 2007). Peptide identification was performed on an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany) essentially as described (Olsen et al., 2004). Full-scan MS spectra were acquired with a resolution of 60,000 in the Orbitrap analyzer. For every full scan, the five most intense ions were fragmented in the linear ion trap. Raw data were processed and analyzed using the MaxQuant software (version 1.0.12.33) and searched with the Mascot search engine against a human IPI database 3.52 as described (Butter et al., 2010). Phosphopeptide enrichment of core histones and MS analysis of these were performed as described (Hurd et al., 2009).

Deposition of MS-Related Data

Mass spectrometric data for peptide pulldowns and GFP pulldowns, consisting of raw data files, unfiltered "proteingroups" tables, and identified

peptides, can be accessed at the TRANCHE repository (https://proteomecommons.org/) under the name "Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers."

Chromatin Immunoprecipitation and Deep Sequencing

ChIP experiments were performed using 3.3×10^6 cells per ChIP according to standard protocols (Denissov et al., 2007), with two minor modifications. Crosslinking of the cells was done on the culture plates for 20 min, while ChIP'ed DNA was purified by Qiaquick PCR purification Kit (QIAGEN cat. no. 28106). ChIP enrichment levels were analyzed by qPCR using specific primers (available upon request) for quality control. ChIP-Seq samples were prepared and analyzed according to the manufacturer (Illumina). Enriched regions were identified by FindPeaks (Fejes et al., 2008). Table S4 summarizes the ChIP-Seq output. For the repeat analysis of the H3K9me3 and LRWD1 ChIP-Seq profiles, mappings were performed by maq aligner (Li et al., 2008). For further information about the ChIP-Seq methods and data analysis see Extended Experimental Procedures. All ChIP-Seq data are present in the NCBI GEO SuperSeries GSE20303.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2010.08.020.

ACKNOWLEDGMENTS

Eva Janssen-Megens, Kees-Jan François, and Yan Tan helped with the sequencing, the MPI core facility expressed and purified recombinant proteins, and Dr. F. Tashiro kindly provided Sgf29 antibody. This work was supported by the Max-Planck Society for the Advancement of Science, HEROIC, an Integrated Project funded by the European Union under the 6th Framework Program (LSHG-CT-2005-018883) and by Interaction Proteome (LSHG-CT-2003-505520). M.V. received a fellowship of the Dutch Cancer Society (KWF/NKB) and a grant from the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

Received: February 17, 2010
Revised: June 9, 2010
Accepted: August 13, 2010
Published: September 16, 2010

REFERENCES

Butter, F., Kappei, D., Buchholz, F., Vermeulen, M., and Mann, M. (2010). A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO Rep. *11*, 305–311.

Cloos, P.A., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev. *22*, 1115–1140.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R., and Stunnenberg, H. (2007). Identification of novel functional TBP-binding sites and general factor repertoires. EMBO J. *26*, 944–954.

Fejes, A.P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., and Jones, S.J. (2008). FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics *24*, 1729–1730.

Fischle, W., Wang, Y., and Allis, C.D. (2003a). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475–479.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003b). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Fischle, W., Franz, H., Jacobs, S.A., Allis, C.D., and Khorasanizadeh, S. (2008). Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. J. Biol. Chem. *283*, 19626–19635.

Franz, H., Mosch, K., Soeroes, S., Urlaub, H., and Fischle, W. (2009). Multimerization and H3K9me3 binding are required for CDYL1b heterochromatin association. J. Biol. Chem. *284*, 35049–35059.

Garcia, B.A., Shabanowitz, J., and Hunt, D.F. (2007). Characterization of histones and their post-translational modifications by mass spectrometry. Curr. Opin. Chem. Biol. *11*, 66–73.

Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat. Struct. Mol. Biol. *17*, 38–43.

Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748–751.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. *189*, 739–754.

Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S.F., Milner, J., Brown, L.A., Hsu, F., Gilks, B., et al. (2003). EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell *115*, 523–535.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol. Cell *13*, 573–585.

Hurd, P.J., Bannister, A.J., Halls, K., Dawson, M.A., Vermeulen, M., Olsen, J.V., Ismail, H., Somers, J., Mann, M., Owen-Hughes, T., et al. (2009). Phosphorylation of histone H3 Thr-45 is linked to apoptosis. J. Biol. Chem. *284*, 16575–16583.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463–473.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

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

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem. Biophys. Res. Commun. *331*, 929–937.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.

Lee, N., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2009). The H3K4 Demethylase Lid Associates with and Inhibits Histone Deacetylase Rpd3. Mol. Cell. Biol. *29*, 1401–1410.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.

Mandel, S., Rechavi, G., and Gozes, I. (2007). Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. Dev. Biol. *303*, 814–824.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, 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.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. *24*, 800–812.

Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep. *5*, 490–496.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem. Sci. *28*, 69–74.

Moshkin, Y.M., Kan, T.W., Goodfellow, H., Bezstarosti, K., Maeda, R.K., Pilyugin, M., Karch, F., Bray, S.J., Demmers, J.A., and Verrijzer, C.P. (2009). Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RPD3 complexes during NOTCH silencing. Mol. Cell *35*, 782–793.

Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene *26*, 5341–5357.

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., and Tora, L. (2010). The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell. Mol. Life Sci. *67*, 611–628.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics *3*, 608–614.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics *1*, 376–386.

Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods *5*, 409–415.

Prasanth, S.G., Prasanth, K.V., Siddiqui, K., Spector, D.L., and Stillman, B. (2004). Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. EMBO J. *23*, 2651–2663.

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434–438.

Ringrose, L., Ehret, H., and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol. Cell *16*, 641–653.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu. Rev. Biochem. *75*, 243–269.

Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J. Biol. Chem. *280*, 41789–41792.

Squatrito, M., Gorrini, C., and Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. *16*, 433–442.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tyteca, S., Vandromme, M., Legube, G., Chevillard-Briet, M., and Trouche, D. (2006). Tip60 and p400 are both required for UV-induced apoptosis but play antagonistic roles in cell cycle progression. EMBO J. *25*, 1680–1689.

van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Mol. Cell *32*, 359–370.

Vermeulen, M., and Selbach, M. (2009). Quantitative proteomics: a tool to assess cell differentiation. Curr. Opin. Cell Biol. *21*, 761–766.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vermeulen, M., Hubner, N.C., and Mann, M. (2008). High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Curr. Opin. Biotechnol. *19*, 331–337.

Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Gottgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. Nat. Struct. Mol. Biol. *17*, 617–619.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009a). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847–851.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., III, and Jia, S. (2009b). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Mol. Cell *33*, 428–437.

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

Winter, S., Fischle, W., and Seiser, C. (2008). Modulation of 14-3-3 interaction with phosphorylated histone H3 by combinatorial modification patterns. Cell Cycle *7*, 1336–1342.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

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.

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. *20*, 123–128.

Cell

c-Myc Regulates Transcriptional Pause Release

Peter B. Rahl,[1] Charles Y. Lin,[1,2] Amy C. Seila,[3,4] Ryan A. Flynn,[3] Scott McCuine,[1] Christopher B. Burge,[2] Phillip A. Sharp,[2,3] and Richard A. Young[1,2,*]
[1]Whitehead Institute for Biomedical Research
[2]Department of Biology, Massachusetts Institute of Technology
[3]Koch Institute, Massachusetts Institute of Technology
Cambridge, MA 02142, USA
[4]Present address: Alnylam Pharmaceuticals, 300 Third Street, Cambridge, MA 02142, USA
*Correspondence: young@wi.mit.edu
DOI 10.1016/j.cell.2010.03.030

SUMMARY

Recruitment of the RNA polymerase II (Pol II) transcription initiation apparatus to promoters by specific DNA-binding transcription factors is well recognized as a key regulatory step in gene expression. We report here that promoter-proximal pausing is a general feature of transcription by Pol II in mammalian cells and thus an additional step where regulation of gene expression occurs. This suggests that some transcription factors recruit the transcription apparatus to promoters, whereas others effect promoter-proximal pause release. Indeed, we find that the transcription factor c-Myc, a key regulator of cellular proliferation, plays a major role in Pol II pause release rather than Pol II recruitment at its target genes. We discuss the implications of these results for the role of c-Myc amplification in human cancer.

INTRODUCTION

Regulation of transcription is fundamental to the control of cellular gene expression programs. Recruitment of the RNA polymerase II (Pol II) transcription initiation apparatus to promoters by specific DNA-binding transcription factors is generally recognized as a key regulatory step in selective transcription at most eukaryotic genes (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997; Roeder, 2005). Additional regulatory steps can occur subsequent to recruitment of the transcription apparatus, and these are known to play important roles in controlling the expression of a subset of genes (Core and Lis, 2008; Margaritis and Holstege, 2008; Peterlin and Price, 2006).

Promoter-proximal pausing of Pol II is a postinitiation regulatory event that has been well-studied at a small number of genes. Promoter-proximal pausing, for the purpose of discussion here, will be used to describe events including attenuation, stalling, poising, abortive elongation, and promoter-proximal termination. The *Drosophila Hsp70* gene is regulated through both recruitment of the initiation apparatus and promoter-proximal pausing prior to the transition to elongation (Gilmour and Lis, 1986; O'Brien and Lis, 1991; Rougvie and Lis, 1988). Paused Pol II molecules can also be detected in some human genes (Bentley and Groudine, 1986; Espinosa et al., 2003; Sawado et al., 2003). At genes regulated through promoter-proximal pausing, the pause factors DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF) generate a Pol II pause just downstream of the transcription start site (TSS) (Wada et al., 1998a; Yamaguchi et al., 1999). Certain sequence-specific transcription factors may recruit pause-release factors such as the positive transcription elongation factor b (P-TEFb) to these genes (Barboric et al., 2001; Core and Lis, 2008; Eberhardy and Farnham, 2001, 2002; Kanazawa et al., 2003; Peterlin and Price, 2006).

Recent reports suggest that postinitiation regulation is important for transcriptional control at a subset of metazoan protein-coding genes. In human embryonic stem cells, for example, approximately 30% of genes experience transcription initiation but show no evidence of further elongation (Guenther et al., 2007). These results indicate that a regulatory step subsequent to recruitment of the initiation apparatus is key for transcriptional control at these genes. Although the genes that experience transcription initiation but not elongation are a minority, the recent discovery that Pol II can initiate transcription in both the sense and antisense directions (Core et al., 2008; Seila et al., 2008) suggests that a postinitiation regulatory step may be required more generally at promoters, if only to prevent unregulated antisense transcription.

We report here evidence that promoter-proximal pausing does occur generally in embryonic stem (ES) cells, at genes that are fully transcribed as well as at genes that experience initiation but not elongation. At genes with detectable levels of Pol II, ChIP-Seq data revealed that most of the enzyme typically occupies DNA in the promoter-proximal region together with the pause factors DSIF and NELF. Inhibition of the pause-release factor P-TEFb caused Pol II to remain at these sites genome-wide. Because c-Myc plays key roles in ES cell self-renewal and proliferation (Cartwright et al., 2005) and can bind the pause-release factor P-TEFb in tumor cells (Eberhardy and Farnham, 2001, 2002; Gargano et al., 2007; Kanazawa et al., 2003),

we investigated whether c-Myc functions to regulate pause release in ES cells. Our results indicate that c-Myc plays a key role in pause release rather than Pol II recruitment at a substantial fraction of actively transcribed genes in ES cells.

RESULTS

Pol II Tends to Occupy Promoter Regions

We used chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) to determine how Pol II occupies the ES cell genome (Figure 1, Table S1 and Table S2 available online). An antibody that binds to the N terminus of the largest subunit of Pol II (N-20) was used, allowing us to monitor Pol II independent of the phosphorylation status of its C-terminal domain (CTD). We found that the bulk of Pol II occupied the promoter-proximal region of the vast majority of genes (Figure 1A). This tendency to occupy promoter-proximal regions was evident both for genes that are actively transcribed (with H3K4me3- and H3K79me2-modified nucleosomes) and for nonproductive genes that show evidence of initiation but not elongation (with H3K4me3-, but not H3K79me2-modified nucleosomes). At actively transcribed genes, low levels of Pol II signal were observed throughout the transcribed region up to the polyadenylation site, with higher signals observed downstream where transcription termination takes place. These data are consistent with more lengthy occupancy of promoter and terminator regions than the central body of actively transcribed genes.

The presence of high polymerase density at the promoter region relative to the gene body has previously been cited as evidence for promoter-proximal pausing or some form of postinitiation regulation in *E. coli*, *Drosophila*, and human cells (Fuda et al., 2009; Price, 2008; Wade and Struhl, 2008). The pattern of Pol II binding we observed suggests that promoter-proximal pausing occurs frequently in mouse embryonic stem (mES) cells. To further characterize Pol II occupancy in mES cells, we calculated the relative ratio of Pol II density in the promoter-proximal region and the gene body (Figure 1B), which has been termed the traveling ratio (TR) (Reppas et al., 2006) or the pausing index (Zeitlinger et al., 2007). At genes where the rate of promoter-proximal clearance is similar to the rate of initiation, the TR is close to 1 (Reppas et al., 2006). However, at genes where promoter-proximal clearance is lower than the initiation rate, the TR is greater than 1. Using this metric, we found that 91% of genes have a Pol II TR of more than 2, confirming that higher Pol II density is detected in the promoter-proximal region than in the gene body at the vast majority of genes (Figure 1C, Figure S1, Table S3). The presence of high polymerase density in the promoter regions of most active ES cell genes, where maximal signal occurs ~35 bp downstream of the transcriptional start site, suggests that these genes experience some form of postinitiation regulation.

The large subunit of Pol II contains a CTD that is modified at various stages of transcription; Pol II is recruited into the preinitiation complex with a hypophosphorylated CTD, and the CTD is phosphorylated on serine 5 (Ser5P) during initiation and then on serine 2 (Ser2P) during elongation (Fuda et al., 2009). To determine how these two phosphorylated forms of Pol II occupy ES cell genes, ChIP-Seq experiments were conducted with antibodies against these two phosphorylated forms of the CTD (Figure 1A). Ser5P Pol II was detected in the promoter region and the transcribed region of active genes, with the peak located in the promoter-proximal region. For genes that experience initiation but not elongation (nonproductive), Ser5P Pol II was detected only within the promoter region, as expected. Ser2P Pol II was detected predominantly downstream of the promoter region, with the peak in the region downstream of the polyadenylation site where termination likely takes place. These results are consistent with the idea that Pol II typically experiences a promoter-proximal, rate-limiting step after being recruited to promoters and after becoming Ser5 phosphorylated. Pol II may also experience a slow release from DNA in regions of transcription termination (Core et al., 2008; Glover-Cutter et al., 2008).

P-TEFb Inhibition Prevents Pause Release at Most Active Genes

The pattern of Pol II occupancy of genes suggests that a postinitiation regulatory step, such as pause release, may be important for transcriptional control of most genes. The *Drosophila Hsp70* gene is regulated subsequent to initiation by P-TEFb-dependent pause release (Lis et al., 2000). Active P-TEFb, a heterodimer consisting of the cyclin-dependent kinase Cdk9 and a cyclin component (CycT1, CycT2, or CycK), phosphorylates at least three targets important for transcriptional control: the Spt5 subunit of DSIF, the NelfE subunit of NELF, and Ser2 of the Pol II CTD (Kim and Sharp, 2001; Marshall et al., 1996; Marshall and Price, 1995; Wada et al., 1998b; Yamada et al., 2006). To assess the role of P-TEFb-dependent pause release in global transcriptional control, we repeated the ChIP-Seq experiment for total Pol II in mES cells treated with flavopiridol (FP), an inhibitor of Cdk9 kinase activity (Chao et al., 2000; Chao and Price, 2001). As expected, FP treatment caused reduced phosphorylation of Spt5 and Pol II Ser2 within 60 min, whereas Ser5 phosphorylation was not substantially affected (Figure 2A and Figure S2A). If Pol II pause release is required at transcribed genes, we would expect that in the presence of FP, Pol II molecules would remain associated with promoter-proximal pause sites but be depleted from DNA farther downstream. This change in the pattern of Pol II occupancy was observed at most actively transcribed genes (Figure 2B and Figure S2B). We analyzed TR to further evaluate changes in Pol II occupancy genome-wide. TR changes with FP treatment were generally observed at actively transcribed genes, where promoter-proximal Pol II signals were relatively unaffected but Pol II signals farther downstream were depleted (Figure 2C). We found that 75% of genes had a change in Pol II TR of at least 1.5 upon drug treatment. TRs were generally unchanged at genes that normally experience initiation but not elongation (Figure 2D). These results suggest that P-TEFb-dependent pause release is required for Pol II transcription of most actively transcribed genes in mES cells.

Promoter-Proximal Sites Are Co-occupied by Pol II, DSIF, and NELF

P-TEFb antagonizes the negative elongation activity of the pause factors DSIF and NELF (Cheng and Price, 2007; Kim and Sharp, 2001; Wada et al., 1998b). DSIF (Spt4 and Spt5) and NELF

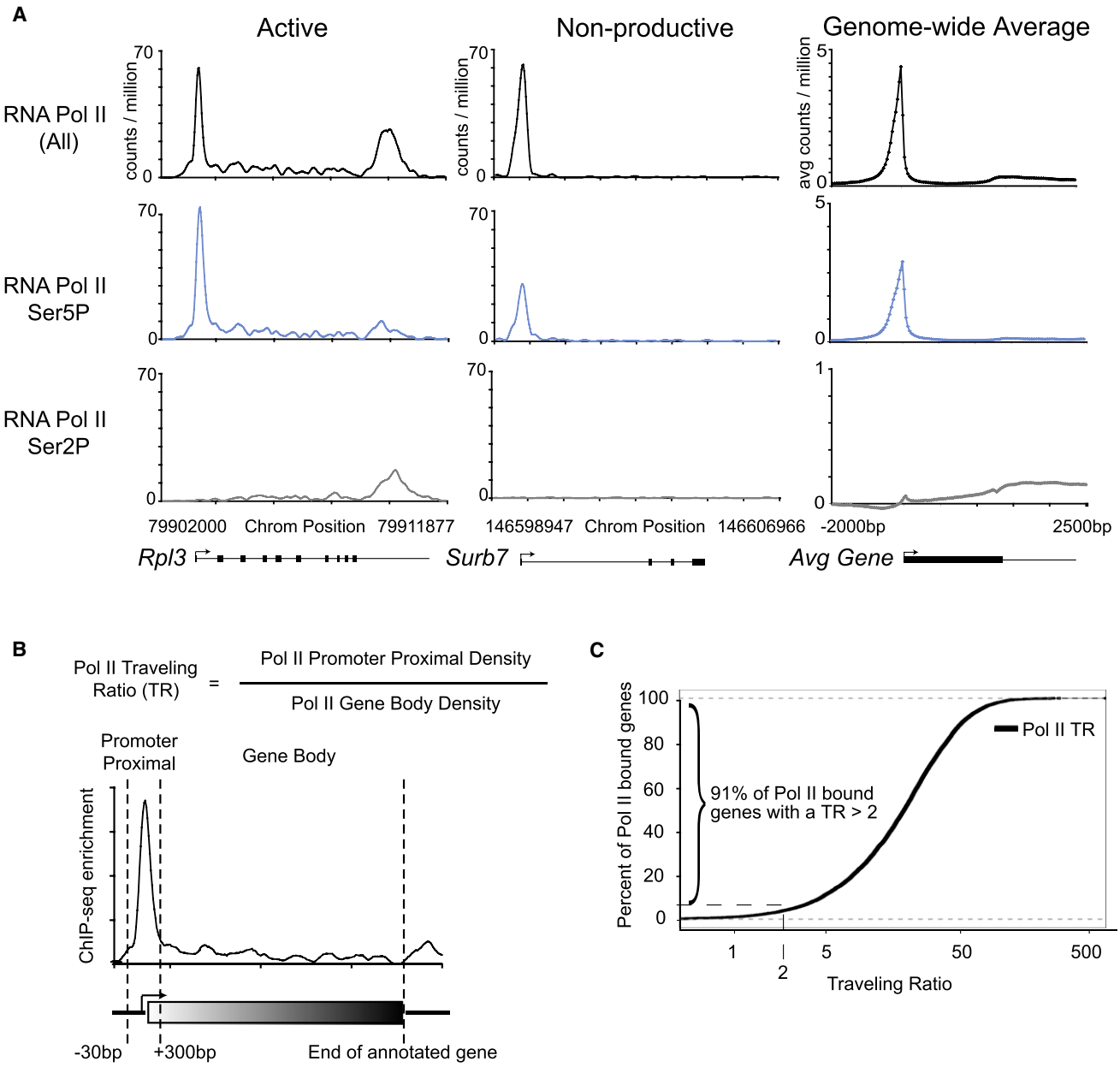

Figure 1. Genome-wide Occupancy of Pol II

(A) Occupancy of RNA Pol II (all), RNA Pol II Ser5P, and RNA Pol II Ser2P in mES cells, determined by ChIP-seq analysis. Enrichment at a representative active gene (*Rpl3*) and nonproductive gene (*Surb7*) is shown. Genome-wide binding averages (introns not depicted), in 50 bp bins, are shown for each Pol II form to display the general binding patterns along the transcription unit from 2 kb upstream of the transcriptional start site to 2.5 kb downstream of the end of each annotated gene.

(B) Schematic representation describing the calculation used to determine the traveling ratio (TR) at each Pol II-bound gene in mES cells. The promoter-proximal bin is defined using a fixed window from −30 bp to +300 bp around the annotated start site. The transcribed region (gene body) bin is from +300 bp to the annotated end. The TR is the ratio of Pol II density in the promoter-proximal bin to the Pol II density in the transcribed region bin.

(C) Distribution of the percent of Pol II-bound genes with a given TR. Approximately 91% of genes have a TR greater than 2, indicating that the majority of Pol II-bound genes have more Pol II in the promoter-proximal region compared to the downstream transcribed region.

See also Figure S1.

(NelfA, NelfB, NelfC/D, and NelfE) are both associated with promoter-proximal Pol II at genes regulated through pausing (Wada et al., 1998a; Yamaguchi et al., 1999). Following the transition to elongation, NELF dissociates and a form of DSIF remains associated with the elongation complex (Andrulis et al., 2000; Wu et al., 2003). If P-TEFb-dependent pause release

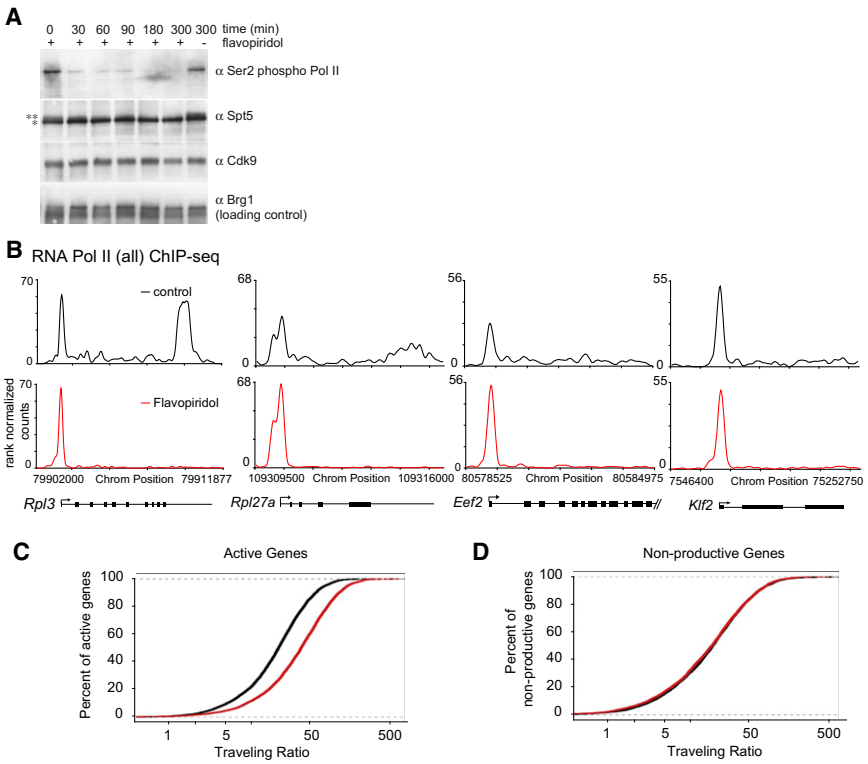

Figure 2. P-TEFb Inhibition Prevents Release of Promoter-Proximal Pol II

(A) mES cells were treated with 1 μM flavopiridol for the indicated time. Extracts were analyzed by western blot using antibodies against Pol II Ser2P, Spt5, Cdk9, and Brg1 (used as a loading control). ** indicates higher-molecular-weight Spt5 species, as reported in Yamada et al. (2006), that is flavopiridol sensitive. * indicates lower-molecular-weight Spt5 species. See also Figure S2A.

(B) RNA Pol II (all) ChIP-seq analysis in mES cells treated with control (DMSO for 60 min, black) or flavopiridol (1 μM for 60 min, red). This panel shows the changes in Pol II occupancy at four example actively transcribed genes following flavopiridol treatment. See also Figure S2B.

(C) Pol II traveling ratio distribution in flavopiridol-treated (red) and control-treated (black) mES cells for active genes (Pol II-bound with H3K79me2-modified nucleosomes). Higher TR values indicate a higher degree of pausing.

(D) Pol II traveling ratio distribution for nonproductive genes in mES cells (Pol II-bound but without H3K79me2-modified nucleosomes), demonstrating that the TR distribution remains relatively the same for nonproductive genes whether treated with control (black) or flavopiridol (red).

is generally required at genes transcribed by Pol II, DSIF and NELF should occupy the promoter-proximal regions of these genes together with Pol II.

We used ChIP-Seq to determine the genome-wide occupancy of NelfA (NELF) and Spt5 (DSIF) in murine ES cells (Figure 3). The results revealed that NelfA and Spt5 occupy precisely the same promoter-proximal sites as Pol II throughout the genome (Figure 3A). The co-occupancy of Pol II, NelfA, and Spt5 in promoter-proximal regions was evident at both actively transcribed genes and genes that experience transcription initiation but not elongation (nonproductive) (Figure 3A). Spt5 and NelfA occupancy positively correlates with Pol II occupancy (Figure S3). The largest NelfA and Spt5 peaks were detected in the promoter-proximal region, but only Spt5 was also enriched farther downstream in actively transcribed genes (Figure 3A). The Spt5 enrichment at the 3′ end of actively transcribed genes was similar to that of Ser2P Pol II, suggesting that it remains associated with Pol II until termination. The NelfA and Spt5 peaks overlapped with the promoter-proximal site of the Pol II peak, which is flanked by H3K4me3-modified nucleosomes (Figures 3B and 3C). These results demonstrate that the pause factors DSIF and NELF co-occupy the promoter-proximal regions of genes together with Pol II, consistent with the model that P-TEFb-dependent pause release is generally required at genes transcribed by Pol II.

Factors such as the PAF1 complex are involved in postinitiation events that are independent of promoter-proximal pausing. PAF1 is involved in elongation, mRNA processing events, and elongation-associated chromatin modifications (Saunders et al., 2006). To test if the Pol II promoter-proximal peak is specific for factors involved in promoter-proximal pausing, we conducted ChIP-Seq with the Ctr9 subunit of the PAF1 complex. Although a limited signal could be detected in the promoter-proximal region of some genes, Ctr9 occupancy did not generally overlap with the promoter-proximal Pol II peak (Figure 3A). Ctr9 was typically found within coding regions of active genes, just downstream of promoter-proximal Pol II, and extending to the 3′ end of transcribed genes. Ctr9 occupancy peaked at the 3′ end of actively transcribed genes, which is similar to the results obtained for Ser2P Pol II and Spt5, suggesting that it remains associated with Pol II until termination. The Ctr9 ChIP-seq data indicate that the PAF1 complex generally associates with the transcribed portion of most active genes, which is consistent with its proposed roles in elongation, mRNA processing, and chromatin modification (Adelman et al., 2006; Krogan et al., 2003; Pokholok et al., 2002; Zhu et al., 2005). These results support the view that the Pol II promoter-proximal peaks represent regions of postinitiation regulation and not simply an artifact of the ChIP-Seq method.

DSIF and NELF function prior to P-TEFb at genes regulated by pause release (Fuda et al., 2009; Peterlin and Price, 2006). This predicts that DSIF and NELF should be present at promoter-proximal sites with Pol II even without P-TEFb activity. We used ChIP-chip to determine if Spt5 and NelfA co-occupy promoter-proximal sites with Pol II following FP treatment. We find that Spt5 and NelfA continue to co-occupy promoter-proximal sites with Pol II following FP treatment (Figure S3B). Spt5 was depleted downstream of these promoter-proximal sites following FP treatment, supporting the model that Spt5 localization in the gene body is dependent on Pol II (Ni et al., 2004, 2008).

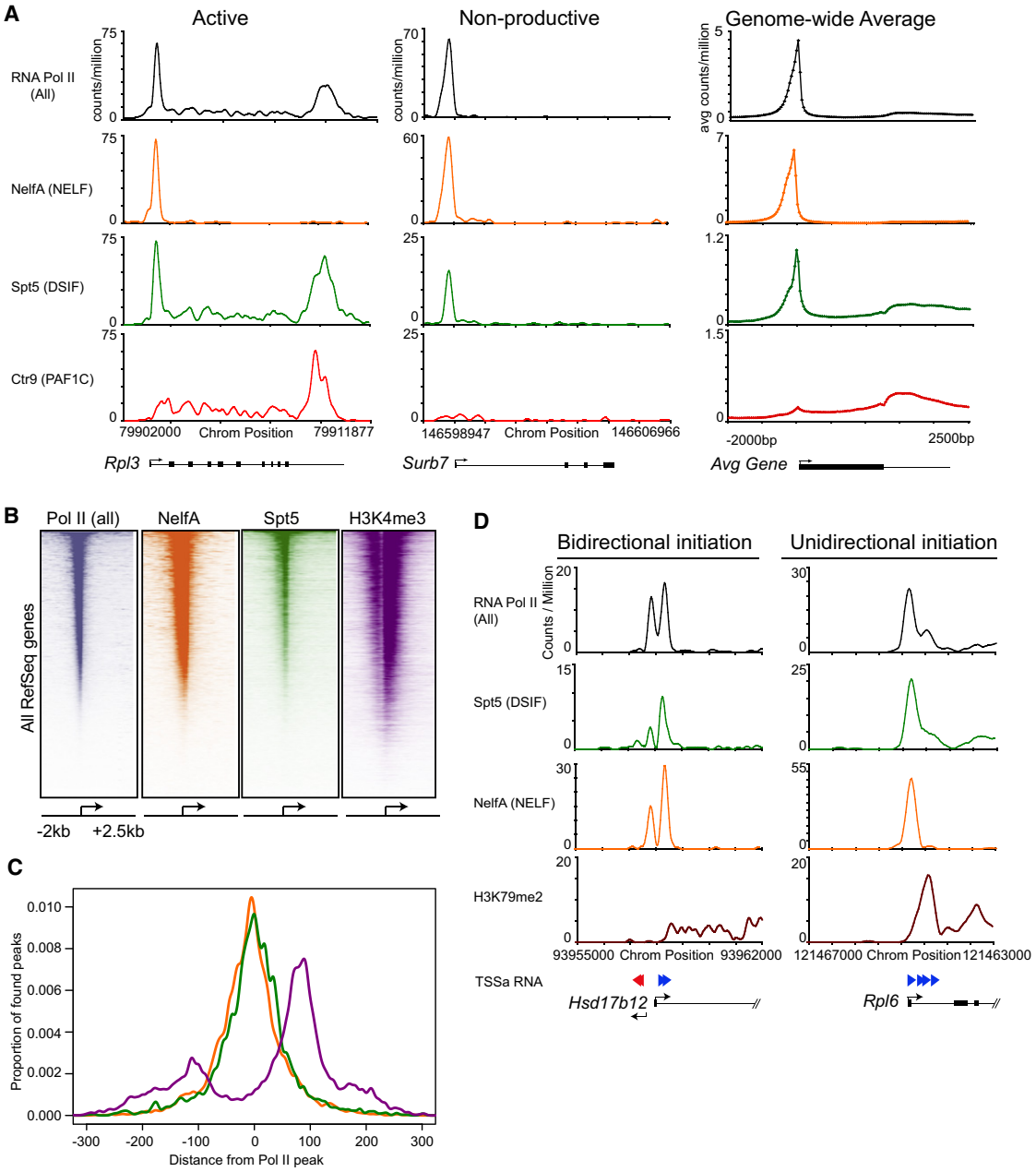

Figure 3. DSIF and NELF Co-occupy Most Genes with Pol II

(A) Binding of Pol II (all), NelfA (NELF subunit), Spt5 (DSIF subunit), and Ctr9 (PAF1 subunit) using ChIP-seq analysis at a representative active gene (*Rpl3*) and nonproductive gene (*Surb7*) in mES cells. Genome-wide binding averages (introns not depicted), in 50 bp bins, are shown for each factor to display the general binding patterns along the transcription unit of RefSeq genes, from 2 kb upstream of the transcriptional start site to 2.5 kb downstream of the end of each annotated gene.

(B) Heatmap representation of ChIP-seq binding for Pol II (all; blue), NelfA (orange), Spt5 (green), and H3K4me3 (purple) at all mouse RefSeq genes, rank ordered from most Pol II to lowest Pol II. Color means enrichment, white means no enrichment. See also Figure S3.

(C) Spatial distribution of the distance (base pairs) of Spt5 (green), NelfA (orange), and H3K4me3 (purple) peaks from the promoter-proximal Pol II peak at each enriched Pol II gene, demonstrating general overlaps with Spt5, NelfA, and Pol II peaks.

(D) ChIP-seq-binding plots showing Pol II (all), Spt5 (DSIF), NelfA (NELF), elongation-associated chromatin modification (H3K79me2), and TSSa-RNA reads that map to this genomic region at a bidirectional initiated gene (*Hsd17b12*) and unidirectional initiated gene (*Rpl6*). Red arrows represent TSSa-RNA species that map in the antisense direction to the gene, and blue arrows represent TSSa-RNA species that map in the sense direction to the gene.

These results indicate that DSIF and NELF co-occupy promoter-proximal sites with Pol II prior to P-TEFb function.

Bidirectional and Unidirectional Genes

It was recently reported that Pol II can initiate transcription in both the sense and antisense directions at many genes (Core et al., 2008; Seila et al., 2008). We separated genes into bidirectional and unidirectional classes based on evidence for sense and antisense *t*ranscription *s*tart site *a*ssociated RNAs (TSSa-RNAs) in ES cells (Seila et al., 2008). To determine how DSIF and NELF occupy the promoter-proximal regions of these two classes of genes, we re-examined the ChIP-seq data for Pol II, Spt5, NelfA, and H3K79me2 (a marker for elongation) at higher resolution (Figure 3D). Approximately 65% of active genes with TSSa-RNA reads fell into the bidirectional class, and at the promoters of these genes we found that the two sites occupied by Pol II were both co-occupied by NelfA and Spt5. Approximately 35% of active genes with TSSa-RNA reads fell into the unidirectional class, and at the promoters of these genes we found that the one site occupied by Pol II was co-occupied by NelfA and Spt5. These results demonstrate that DSIF and NELF generally co-occupy promoter-proximal regions wherever Pol II is found, whether initiation is occurring in one direction or two, further supporting the model that P-TEFb-dependent pause release is a general feature of transcription initiation by Pol II.

Pause-Factor Knockdown Alters Pol II Gene Occupancy

The pause factors NELF and DSIF co-occupy promoters with Pol II at most genes that experience transcription initiation. Previous studies have shown that loss of NELF causes a decrease in Pol II density at promoters, and thus a decrease in Pol II traveling ratio (or pausing index), at a small number of *Drosophila* genes (Muse et al., 2007). To determine how loss of vertebrate NELF or DSIF might influence Pol II occupancy, we used shRNA-mediated knockdown of NelfA and Spt5 followed by Pol II ChIP-seq analysis in mES cells (Figure 4).

The most significant change in Pol II density was found following Spt5 knockdown, where increases in Pol II density were frequently observed downstream of the promoter at actively transcribed genes (Figure 4B). At these active genes, depletion of a pausing factor appeared to result in increased transcription through the pause site, but because there was little effect on promoter-proximal Pol II, high rates of initiation maintained Pol II promoter levels. NelfA was found to continue to occupy the promoter-proximal regions following Spt5 knockdown (Figures S4A and S4B). The effects of Spt5 knockdown on Pol II density were quantified using the TR metric (Figure 4C and Figures S4E and S4F). There was a substantial shift in TR upon Spt5 knockdown, demonstrating that genes generally experience an increase in Pol II density in the transcribed region at active genes when the levels of DSIF are reduced. These results confirm that Spt5 function contributes to the control of promoter-proximal Pol II in mES cells.

NelfA knockdown had less impact on Pol II occupancy (Figures 4B and 4C), but a modest effect was observed at some genes and the pattern of change was similar to that observed for the Spt5 knockdown experiment at nonproductive genes, as evidenced by the change in TR (Figures S4E, S4F, and S4G). This result is similar to that observed previously in *Drosophila* embryos, where a fraction of genes showed a loss of Pol II density at the promoter (Muse et al., 2007). Spt5

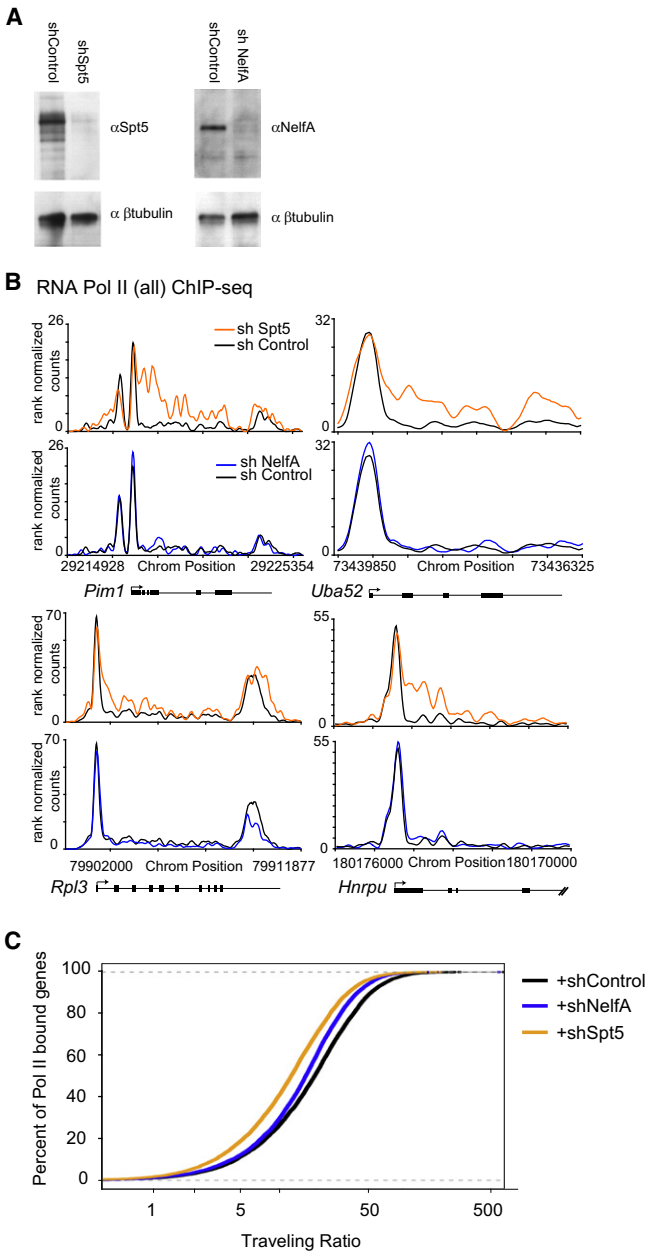

Figure 4. DSIF Knockdown Alters Pol II Occupancy at Many Genes

(A) Spt5 (left) and NelfA (right) protein levels after the indicated shRNA-mediated knockdown in mES cells as determined by western blot using Spt5 or NelfA antibodies. β-Tubulin protein is a loading control.

(B) RNA Pol II (all) ChIP-seq binding density in shControl (black), shSpt5 (orange), and shNelfA (blue) mES cells analysis at four active genes in mES cells.

(C) RNA Pol II TR calculations in shControl, shSpt5, and shNelfA mES cells, showing that many genes become less paused following Spt5 knockdown and a more subtle change following NelfA knockdown. Lower TR values indicate a lower degree of pausing.

See also Figure S4.

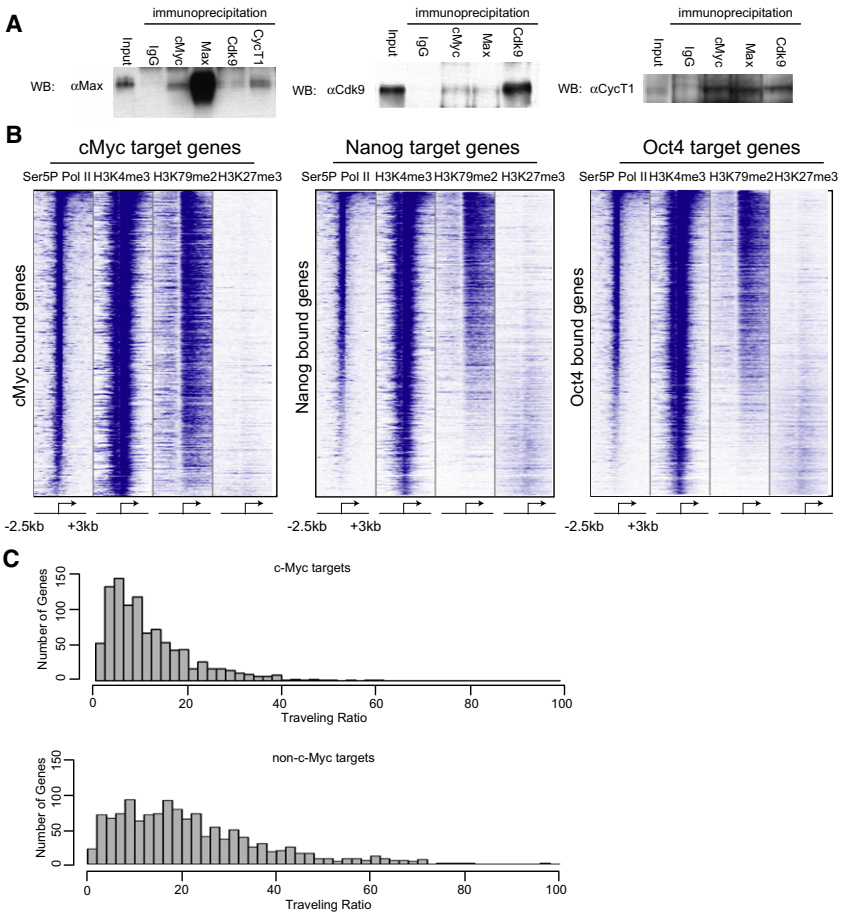

Figure 5. c-Myc Target Genes Are Enriched in Actively Transcribed Genes and c-Myc/Max Associates with P-TEFb in mES Cells

(A) Coimmunoprecipitation experiments in mES cells using antibodies against IgG (to measure background binding) or endogenous c-Myc, Max, Cdk9, and CycT1. Proteins were immunopreipicated from mES cell lysates and analyzed by western blot analysis by probing for Max, Cdk9, and CycT1. Ten percent input was loaded for the Max and Cdk9 blots, 1% input was loaded for the CycT1 blot.

(B) Heatmap representation illustrating the transcriptional state of c-Myc, Oct4, and Nanog target genes in mES cells, as determined by Pol II Ser5P, H3K4me3 (initiation-associated chromatin modification), H3K79me2 (elongation-associated chromatin modification), and H3K27me3 (repressive chromatin modification). Each target gene set was rank ordered based on the amount of Pol II bound at each gene, from the highest amount of Pol II to the lowest amount, and the enrichment of the indicated chromatin modification or Pol II is displayed from −2.5 kb to +3 kb surrounding each annotated transcription start site. Blue indicates enrichment and white indicates no enrichment.

(C) c-Myc target genes have lower TR values than nontarget genes. Histograms were made for the number of genes with given TR values for high-confidence c-Myc target genes and nontarget genes. Genes with lower TR values are less paused than genes with higher TR values.

See also Figure S5.

occupancy was largely unaffected in the promoter-proximal regions following NelfA knockdown (Figures S4C and S4D). In summary, we find that Spt5 knockdown, and to a more limited extent NelfA knockdown, can produce increased Pol II occupancy in transcribed regions relative to promoter-proximal regions, consistent with the proposed roles of these factors in controlling promoter-proximal pausing.

c-Myc Binds P-TEFb and Contributes to Pause Release in ES Cells

Certain DNA-binding transcription factors may be responsible for recruiting P-TEFb to release paused polymerase at active genes if P-TEFb-dependent pause release is a general feature of transcription by Pol II. Such a role has been proposed for c-Myc based on evidence that this transcription factor can bind P-TEFb and stimulate elongation at specific genes in tumor cells (Eberhardy and Farnham, 2001, 2002; Gargano et al., 2007; Kanazawa et al., 2003). Because c-Myc is a key ES cell transcription factor required for self-renewal and proliferation (Cartwright et al., 2005), which occupies a third of active genes (see below), we investigated whether c-Myc plays a role in P-TEFb-dependent pause release at the genes it occupies in ES cells.

If c-Myc contributes to P-TEFb-dependent pause release in ES cells, it might be expected to bind P-TEFb in these cells. To function as a transcription factor, c-Myc forms a heterodimer with Max (Eilers and Eisenman, 2008). We used coimmunoprecipitation analysis to determine if endogenous c-Myc/Max interacts with P-TEFb in ES cells. We found that immunoprecipitation of P-TEFb components Cdk9 and CycT1 coimmunoprecipitates Max, and similarly, immunoprecipitation of c-Myc and Max coimmunoprecipitates Cdk9 and CycT1 (Figure 5A). Therefore, c-Myc/Max can bind P-TEFb in ES cells.

If a predominant function of c-Myc is to contribute to pause release in ES cells, then we expect that it should be associated almost exclusively with actively transcribed genes, unlike other key ES cell regulators like Oct4 and Nanog, which are associated with both active and repressed genes. We examined published ChIP-Seq data to determine the fraction of genes bound by c-Myc, Oct4, and Nanog that were actively transcribed (Chen et al., 2008; Marson et al., 2008), as indicated by the presence of nucleosomes containing histones H3K4me3 and H3K79me2 (Figure 5B). Just over half of Oct4- and Nanog-occupied genes show evidence of transcription elongation (H3K79me2-modified nucleosomes). In contrast, almost all of the c-Myc-occupied genes have H3K79me2-modified nucleosomes, indicating that the majority of c-Myc targets in mES cells experience transcription elongation. Furthermore, c-Myc target genes have lower TR values compared to non-c-Myc targets (Figure 5C). We estimate that 33% of actively transcribed genes in ES cells are bound by c-Myc within 1 kb of the transcriptional start site (Figure S5).

The association of c-Myc with a substantial fraction of actively transcribed genes, coupled with evidence that it can bind P-TEFb, is consistent with the model that c-Myc contributes to P-TEFb-dependent pause release at a large portion of active genes in ES cells.

To more directly test whether c-Myc regulates pause release, we used a low-molecular-weight inhibitor of c-Myc/Max, 10058-F4, which inhibits c-Myc/Max heterodimerization both in vitro and in vivo (Hammoudeh et al., 2009; Wang et al., 2007; Yin et al., 2003). Max co-occupies c-Myc-binding sites as determined by ChIP, confirming that c-Myc and Max function together at target genes in ES cells (Figure S6A). Treatment of mES cells with 10058-F4 (50 μM for 6 hr) caused a decrease in the expression of most c-Myc target genes tested but did not affect the expression of two non-c-Myc target genes, indicating that c-Myc/Max function is inhibited by 10058-F4 under these conditions (Figure 6A). The magnitude of the decrease observed (~20%–40%) is consistent with the relatively short duration of inhibitor treatment relative to typical mRNA half-lives of ~7 hr in mES cells (Sharova et al., 2009).

If a key function of c-Myc is to contribute to pause release at the active genes it occupies in ES cells, then loss of c-Myc would be expected to cause a reduction in the levels of Ser2-phosphorylated Pol II (the form associated with elongation) but should not affect the levels of Ser5-phosphorylated Pol II (the form associated with initiation). We found that treatment of ES cells with 10058-F4 did indeed cause a significant reduction in the levels of Pol II Ser2P, whereas Ser5P remained unaffected (Figure 6B). The 1/3 of genes that are regulated by c-Myc are among the most highly transcribed genes in the cell, which likely explains why the reduction in total Ser2P Pol II levels is greater than 33%.

If c-Myc regulates pause release, then inhibition should have an effect on Pol II levels in promoter and gene bodies similar to that of FP, but only at c-Myc-occupied genes. We tested this idea by determining how 10058-F4 affects Pol II occupancy using ChIP-seq in mES cells. There was little effect on Pol II density at promoters, but there was a clear reduction in transcribed regions (Figure 6C). This effect on Pol II density was also observed following c-Myc shRNA knockdown (Figures S6B–S6D). The magnitude of the effect with 10058-F4 was somewhat milder than with FP, probably because the inhibition of c-Myc/Max heterodimerization is not complete (Hammoudeh et al., 2009; Wang et al., 2007; Yin et al., 2003). Treatment with 10058-F4 did not alter the protein levels of P-TEFb components Cdk9 or CycT1, indicating that this effect is not a result of reduced levels of P-TEFb (Figure S6E). Importantly, genes that lack evidence of c-Myc binding showed patterns of Pol II occupancy that were unaffected by treatment with 10058-F4 (Figure 6C). We confirmed that genes that are not targets of c-Myc do require P-TEFb function to release paused Pol II by showing that FP treatment causes a block in pause release (Figure 6C), which suggests that transcription factors other than c-Myc are involved in recruiting P-TEFb to stimulate pause release at these genes.

We carried out a more global analysis of the effect of 10058-F4 on Pol II occupancy of genes bound by c-Myc and compared these patterns to genes that are not bound by this factor but show evidence of elongation (Figure 6D and Figure S6G). The results show that high-confidence c-Myc target genes generally retained promoter-proximal Pol II but had reduced Pol II density in their transcribed regions, whereas Pol II occupancy does not change at genes that are not c-Myc targets (Figure 6D). Further analysis confirmed that there were statistically significant changes in the gene bodies ($p = 7.341e-06$) but not the promoter regions ($p = 0.4536$) of c-Myc targets. Additionally, following 10058-F4 treatment, a substantial increase in TR was observed at c-Myc target genes, but no such shift was observed at non-c-Myc targets (Figure 6E). A similar shift in TR at the c-Myc target genes was also detected following c-Myc shRNA knockdown, indicating that genes become more paused (Figure S6F). The observation that reduced c-Myc activity had little effect on the levels of promoter-proximal Pol II but caused a reduction in the levels of Pol II across transcribed portions of c-Myc target genes is consistent with the model that c-Myc/Max generally plays a role in Pol II pause release at its target genes in mES cells.

Loss of Oct4 and c-Myc Have Different Effects on Pol II Gene Occupancy

Loss of another key ES cell transcription factor, Oct4, leads to reduced transcription of many Oct4-bound active genes in ES cells (Hall et al., 2009; Matoba et al., 2006). To determine how loss of Oct4 affects Pol II levels at the promoters and transcribed regions of its target genes, we utilized a doxycycline-inducible Oct4 shutdown mES cell line (Niwa et al., 2000) and monitored Pol II levels by ChIP-Seq before and after Oct4 shutdown (Figure 7). Oct4 protein levels were substantially reduced within 12 hr and nearly eliminated at 24 hr after exposure to doxycycline (Figure 7A). At Oct4-occupied genes that experience reduced transcription, Pol II occupancy was generally reduced in both the promoter-proximal region and the gene body at 12 and 24 hr (Figure 7B). These effects were not observed at most genes that are not occupied by Oct4 (Figure 7B). The loss of Pol II in the promoter-proximal regions of Oct4 target genes, given the commensurate loss of Pol II in the gene body, is likely due to reduced recruitment of the transcription apparatus. For these Oct4 target genes, where Pol II is lost from both promoter-proximal and gene body regions, we would expect no change in TR, and a global analysis of such genes revealed that this is indeed the case (Figure 7C and Figure S6G). We conclude that the pattern of reduced Pol II density at Oct4 target genes that occurs upon loss of Oct4 differs from that at c-Myc target genes upon loss of c-Myc and suggest that this is due to differences in the stages at which the two transcription factors play their dominant regulatory roles.

DISCUSSION

Transcription factors bind to specific DNA sequences and regulate gene expression by recruiting the transcription initiation apparatus to promoters (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997; Roeder, 2005). Recent studies have shown that an additional level of regulation must occur subsequent to initiation at certain genes and have proposed that certain transcription factors regulate this step (Core and Lis, 2008; Margaritis and Holstege, 2008; Peterlin and Price, 2006). The evidence described here indicates that promoter-proximal pausing is

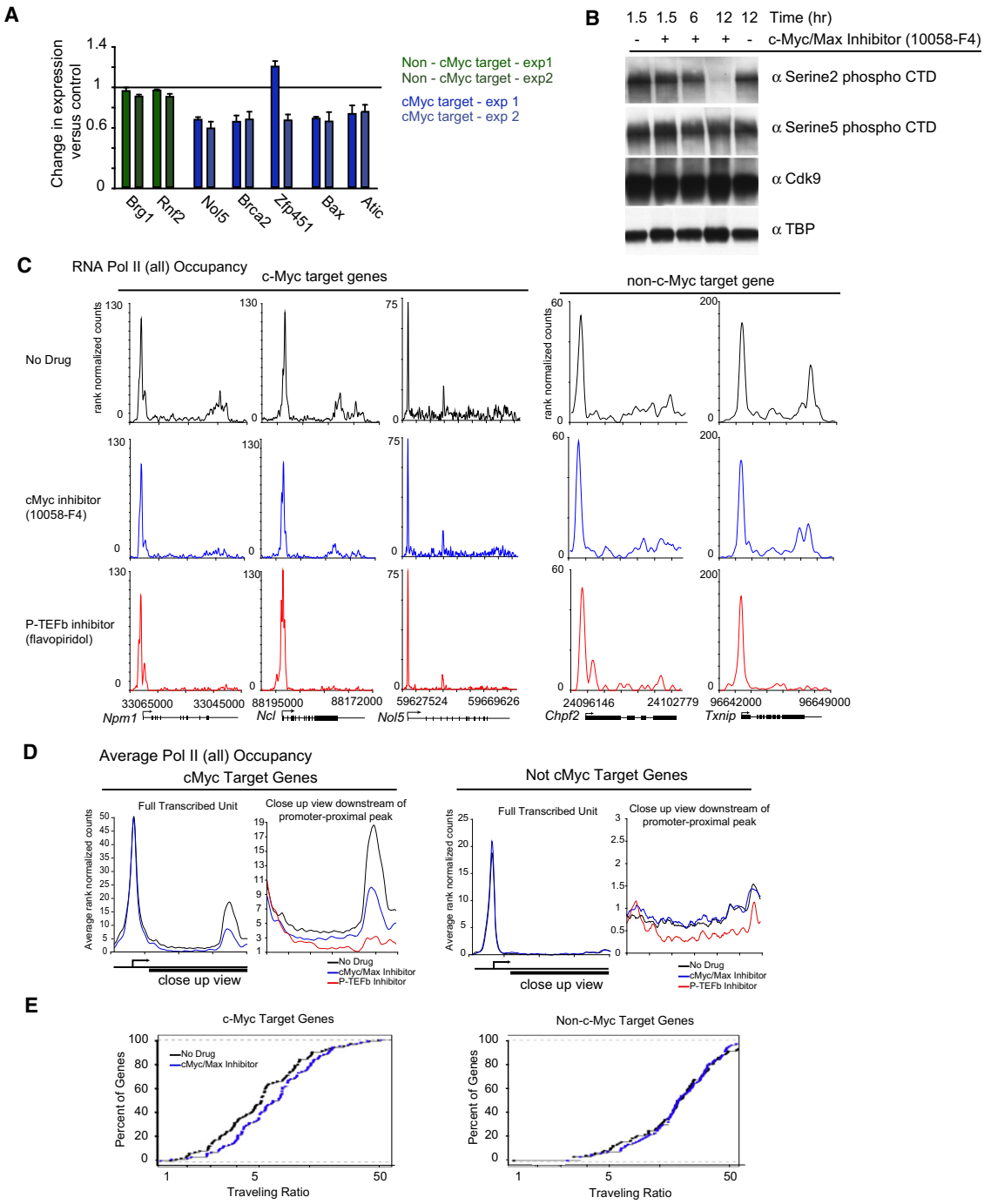

Figure 6. c-Myc Inhibition Affects Transcription at the Pause-Release Step

(A) RNA was extracted from mES cells treated with 10058-F4 or vehicle alone (DMSO) for 6 hr and used to generate cDNA using reverse transcription. Expression change was calculated for 10058-F4-treated cells compared to vehicle alone control for two non-c-Myc target genes (Brg1, Rnf2—green) and five c-Myc target genes (Bax, Nol5, Zfp451, Brca2, and Atic—blue) from two independent experiments. Error bars represent standard deviation (SD) from triplicate qPCR reactions.

(B) mES cells were treated with 10058-F4 for 1.5, 6, or 12 hr. Extracts were analyzed using western blot with antibodies against Pol II Ser2P CTD and Pol II Ser5P CTD to determine the levels of the modified forms of Pol II. TBP was used as a loading control.

(C) Pol II ChIP-seq binding profiles in mES cells treated with 10058-F4 (c-Myc/Max inhibitor; blue), DMSO alone (black), or flavopiridol (P-TEFb inhibitor; red). Pol II occupancy is shown for three c-Myc target genes (*Ncl*, *Npm1*, and *Nol5*) and two non-c-Myc target gene (*Txnip* and *Chpf2*). Cells were treated with 10058-F4 or DMSO for 6 hr. See also Figure S6.

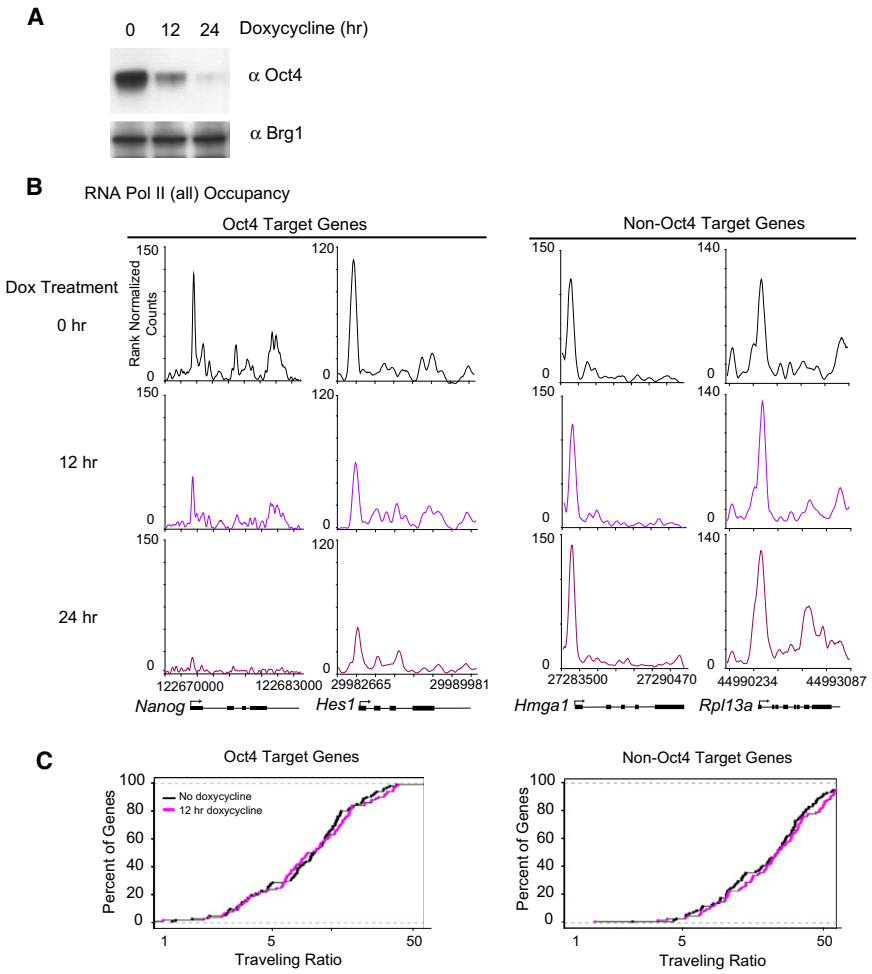

Figure 7. Oct4 Shutdown Reduces Pol II Initiation at Oct4-Dependent Genes

(A) Oct4 protein levels in doxycycline-inducible Oct4 knockdown mES cells following 0, 12, or 24 hr of doxycycline treatment. Extracts were probed with an antibody against Oct4. Brg1 was used as a loading control.

(B) Pol II ChIP-seq binding profiles at Oct4 target genes following the indicated time of doxycycline treatment, inducing Oct4 knockdown. Of note, the Oct4-bound genes change Pol II occupancy in both the promoter-proximal region and the transcribed region. The panel to the right shows Pol II ChIP-seq binding profiles at non-Oct4 target following the indicated time of doxycycline treatment, inducing Oct4 knockdown.

(C) Pol II traveling ratio (TR) as described in Figure 1C for the high-confidence Oct4-dependent genes and Oct4 nontarget genes after either 0 or 12 hr of doxycycline treatment. The left panel is the TR for the Oct4 targets and right panel is the TR for non-Oct4 targets.

a more general feature of transcription by Pol II in vertebrate cells and identifies c-Myc as playing a key role in pause release at a large population of actively transcribed genes in ES cells.

We describe several lines of evidence supporting the hypothesis that promoter-proximal pausing is a general feature of transcription by Pol II in ES cells. First, genome-wide analysis shows that the bulk of Pol II occupies the promoter-proximal region of genes, even when these genes are among the most actively transcribed in the cell. Second, the pause factors DSIF and NELF typically co-occupy these sites with Pol II, consistent with the idea that they generally bind to the enzyme during early steps of transcription elongation. Third, inhibition of the pause-release factor P-TEFb prevents release of promoter-proximal paused Pol II at essentially all genes.

A handful of genes have been identified that are regulated by P-TEFb-dependent pause control, such as *Hsp70* and *cad* (Eberhardy and Farnham, 2002; Lis et al., 2000). However, in retrospect, there were a number of observations that indicated that this step in transcription is more frequently regulated and might be general. Germ cells repress Pol II transcription globally by inhibiting P-TEFb function. In *C. elegans* germline blastomeres and in *Drosophila* primordial germ cells, PIE-1 and Pgc, respectively, repress global transcription by inhibiting P-TEFb function (Hanyu-Nakamura et al., 2008; Seydoux and Dunn, 1997; Zhang et al., 2003). The transcription factors of lentiviruses and retroviruses such as HIV and T-lymphotropic virus type I function by recruiting P-TEFb to their promoter regions, attenuating Pol II transcriptional pausing (Wei et al., 1998; Zhou et al., 2006). These viruses have thus generated their own means to overcome pause control.

The model that promoter-proximal pausing is general has several implications for transcriptional control. A step subsequent to recruitment of the transcription initiation apparatus can, in principle, be regulated at any gene. Promoter-proximal pausing may facilitate assembly of RNA-processing factors and has been proposed to couple transcription and mRNA-processing events (Glover-Cutter et al., 2008; Moore and Proudfoot, 2009). DSIF and Ser5P Pol II can bind capping enzyme and stimulate mRNA capping (Mandal et al., 2004; McCracken et al., 1997a, 1997b; Wen and Shatkin, 1999). Ser2 phosphorylation by P-TEFb leads to splicing factor and 3′ end processing factor recruitment and is required for proper processing (Ahn et al.,

(D) Average Pol II binding plots for the high-confidence c-Myc targets and non-c-Myc target genes in no drug (black) and 10058-F4 treatment (blue). The left panel shows the entire gene average. The right panel is a close-up of the transcribed region to show the difference in amounts of elongating Pol II density under the different conditions. Also included in the right panel for comparison is elongating Pol II density following flavopiridol treatment (red).

(E) Pol II traveling ratio (TR) for the high-confidence c-Myc target genes and non-c-Myc target genes following 10058-F4 treatment (blue) or no drug (black). The left panel is the TR for the c-Myc targets and right panel is the TR for non-c-Myc targets. Higher TR values indicate a higher degree of pausing.

2004; Komarnitsky et al., 2000; McCracken et al., 1997b; Ni et al., 2004). Promoter-proximal pausing also provides a mechanism to control transcription from bidirectional promoters (Core et al., 2008; Seila et al., 2008), perhaps facilitating the formation of nucleosome-depleted regions and thus providing improved access to regulators (Gilchrist et al., 2008).

Multiple lines of evidence support the contention that c-Myc/Max generally plays a role in Pol II pause release at its target genes in ES cells and does so through recruitment of P-TEFb. Loss of c-Myc reduces the levels of elongating Pol II but does not affect the levels of promoter-proximal Pol II. Inhibition of c-Myc/Max function leads to a substantial reduction in the levels of Ser2-phosphorylated Pol II in cells, which is the form associated with elongation, but does not affect the levels of Ser5-phosphorylated Pol II, which is the form associated with initiation. c-Myc binds P-TEFb, which is responsible for Ser2-phosphorylated Pol II. Consistent with a role in pause release, c-Myc is associated almost exclusively with genes that are actively transcribed, unlike other key ES cell regulators like Oct4 and Nanog, which occupy both active and repressed genes. Furthermore, c-Myc occupies promoter-proximal sites (Figure S5A), which are heavily enriched for the E-box core motif that it binds (Figure S5B), where c-Myc would be optimally positioned to recruit P-TEFb.

In ES cells, c-Myc occupies genes involved in cellular proliferation, as it does in other cell types (Chen et al., 2008; Kidder et al., 2008; Kim et al., 2008). Our results indicate that c-Myc plays a key role in pause release in ES cells and does so at perhaps 1/3 of all actively transcribed genes. These results explain how ectopic expression of c-Myc can substantially enhance the efficiency of reprogramming of fibroblasts to induced pluripotent stem cells (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). Since c-Myc is expressed in a broad spectrum of proliferating cell types (Eilers and Eisenman, 2008; Meyer and Penn, 2008), and has been shown to bind P-TEFb and stimulate elongation at a set of genes in tumor cells (Eberhardy and Farnham, 2001, 2002; Gargano et al., 2007; Kanazawa et al., 2003), we expect that c-Myc functions to effect pause release at this population of genes in most proliferating cells.

c-Myc amplification is the most frequent somatic copy-number amplification in tumor cells (Beroukhim et al., 2010). Our results suggest that tumor cells that overexpress c-Myc have enhanced expression of proliferation genes due to the role of c-Myc in recruiting P-TEFb to effect Pol II pause release at these genes. It is therefore possible that combinations of drugs that reduce the activity of both c-Myc and P-TEFb could be especially effective therapeutic agents in tumor cells that overexpress c-Myc.

EXPERIMENTAL PROCEDURES

mES Cell Culture

V6.5 (C57BL/6-129) murine ES cells were grown under typical mES conditions on irradiated mouse embryonic fibroblasts (MEFs). For location analysis, cells were grown for two passages off of MEFs. For location analysis on mES cells following treatment with small-molecule inhibitors, cells were grown two passages off feeders, and prior to formaldehyde crosslinking, the cells were treated with the indicated final concentration of flavopiridol (1 μM for 1 hr for ChIP-chip and ChIP-seq experiments, or the indicated concentration and time for western blot analysis) or c-Myc/Max inhibitor 10058-F4 (50 μM for 6 hr), both dissolved in DMSO, in the growth medium. As a control, DMSO alone was added at the same final volume as with drug. Small-molecule inhibitors used were as follows: flavopiridol (Sigma cat #F3055) and c-Myc inhibitor 10058-F4 (Sigma cat #F3680). For location analysis following shRNA knockdown (OpenBiosystems), viral media was collected 48 hr after cotransfection in 293T cells and the mES cells were directly infected with the viral media 24 hr after initial plating of mES cells. The infection media were 1:2 viral media:mES media with 2 mM polybrene. The efficiently infected cells were selected for 24 hr post-infection with mES media containing 2 μM puromycin. Cells were crosslinked 72 hr post-selection. For location analysis following Oct4 shutdown, ZHBTc4 mES cells (Niwa et al., 2000) were grown under standard mES cell culture conditions and expanded for two passages off MEF feeders. mES media with 2 μg/ml doxycycline were added to the cells for 0 hr, 12 hr, and 24 hr prior to formaldehyde crosslinking.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was done following the Agilent Mammalian ChIP-on-chip protocol. The antibodies and ChIP conditions used can be found in the Extended Experimental Procedures. For ChIP-chip analysis, Cy3- and Cy5-labeled ligation-mediated PCR products were hybridized to a 44,000 feature Agilent mouse microarray. For ChIP-seq analysis, Solexa/Illumina sequencing and analysis were done following the protocol described in Marson et al. (2008). Refer to the Extended Experimental Procedures for a detailed description of these methods.

Active and Nonproductive Gene Classes in mES Cells

The active and nonproductive genes were classified in mES cells using H3K4me3 (initiation-associated chromatin modification) and H3K79me2 (elongation-associated chromatin modification), as determined by ChIP-seq (Marson et al., 2008), as markers of transcriptional state. Active genes had both H3K4me3 and H3K79me2 chromatin modifications, nonproductive genes had only H3K4me3, and inactive genes did not have H3K4me3 or H3K79me2 (Guenther et al., 2007; Marson et al., 2008; Pokholok et al., 2005).

Traveling Ratio Calculation

Pol II levels peak in the 5′ region of many genes. To quantify this effect, we have developed a measure called traveling ratio (TR) that compares the ratio between Pol II density in the promoter region and the gene region. We defined the promoter region from −30 to +300 relative to the TSS and the gene body as the remaining length of the gene.

Heatmap Analysis of ChIP-seq Data

ChIP-seq enrichment for the indicated factor or modification was determined in 50 bp bins (enrichment in the bin as counts per million), centered on each transcriptional start site. Generally, the gene list for each representation was rank ordered based on the amount of Pol II (all) in mES cells, from most to least to correlate the enrichment of the given factor with the amount of Pol II at each gene. Cluster 3.0 and Java Treeview were used to visualize the data and generate figures shown in this manuscript.

Previously Published ChIP-seq Datasets Used in This Study

Previously published datasets are as follows: H3K4me3, H3K79me2, and Oct4 occupancy in mES cells (Marson et al., 2008); Nanog and c-Myc occupancy in mES cells (Chen et al., 2008); and H3K27me3 occupancy in mES cells (Mikkelsen et al., 2007).

ACCESSION NUMBERS

The ChIP-seq and microarray data are deposited in GEO under accession number GSE20485.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and three tables and can be found with this article online at doi:10.1016/j.cell.2010.03.030.

ACKNOWLEDGMENTS

We thank Garrett Frampton for help with Solexa data analysis; David Price for antibodies and discussions; Hiroshi Handa for antibodies; Tom Volkert, Jennifer Love, Jeong-Ah Kwon, and Sumeet Gupta at the Whitehead Institute Sequencing Core for help with Solexa sequencing; and John Lis and members of the Young, Sharp, and Burge laboratories for helpful discussions. This work was supported by grants RO1-HG002668 (R.A.Y.), RO1-GM34277 (P.A.S.), RO1-CA133404 (P.A.S.) from the National Institutes of Health, PO1-CA42063 from the National Cancer Institute (P.A.S.), and Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute (P.A.S.). A.C.S. was supported by a National Institutes of Health postdoctoral fellowship 5-F32-HD051190.

Received: October 14, 2009
Revised: January 14, 2010
Accepted: February 21, 2010
Published: April 29, 2010

REFERENCES

Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Mol. Cell. Biol. *26*, 250–260.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol. Cell *13*, 67–76.

Andrulis, E.D., Guzman, E., Doring, P., Werner, J., and Lis, J.T. (2000). High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes Dev. *14*, 2635–2649.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol. Cell *8*, 327–337.

Bentley, D.L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702–706.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885–896.

Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J. Biol. Chem. *276*, 31793–31799.

Chao, S.H., Fujinaga, K., Marion, J.E., Taube, R., Sausville, E.A., Senderowicz, A.M., Peterlin, B.M., and Price, D.H. (2000). Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. J. Biol. Chem. *275*, 28345–28348.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J. Biol. Chem. *282*, 21901–21912.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Eberhardy, S.R., and Farnham, P.J. (2001). c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. J. Biol. Chem. *276*, 48562–48571.

Eberhardy, S.R., and Farnham, P.J. (2002). Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. J. Biol. Chem. *277*, 40156–40162.

Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev. *22*, 2755–2766.

Espinosa, J.M., Verdun, R.E., and Emerson, B.M. (2003). p53 functions through stress- and promoter-specific recruitment of transcription initiation components before and after DNA damage. Mol. Cell *12*, 1015–1027.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Gargano, B., Amente, S., Majello, B., and Lania, L. (2007). P-TEFb is a crucial co-factor for Myc transactivation. Cell Cycle *6*, 2031–2037.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). Genes Dev. *22*, 1921–1933, NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. *22*, 1921–1933.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol. Cell. Biol. *6*, 3984–3989.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell *5*, 597–609.

Hammoudeh, D.I., Follis, A.V., Prochownik, E.V., and Metallo, S.J. (2009). Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J. Am. Chem. Soc. *131*, 7390–7401.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730–733.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. *17*, 1309–1320.

Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene *22*, 5707–5711.

Kidder, B.L., Yang, J., and Palmer, S. (2008). Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PLoS ONE *3*, e3932.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Kim, J.B., and Sharp, P.A. (2001). Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J. Biol. Chem. *276*, 12317–12323.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol. Cell *11*, 721–729.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes Dev. *14*, 792–803.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., and Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc. Natl. Acad. Sci. USA *101*, 7572–7577.

Margaritis, T., and Holstege, F.C. (2008). Poised RNA polymerase II gives pause for thought. Cell *133*, 581–584.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. *270*, 12335–12338.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J. Biol. Chem. *271*, 27176–27183.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A., and Ko, M.S. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. PLoS ONE *1*, e26.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997a). 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. *11*, 3306–3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. Nat. Rev. Cancer *8*, 976–990.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.R., and Lis, J.T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. Mol. Cell *13*, 55–65.

Ni, Z., Saunders, A., Fuda, N.J., Yao, J., Suarez, J.R., Webb, W.W., and Lis, J.T. (2008). P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. Mol. Cell. Biol. *28*, 1161–1170.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

O'Brien, T., and Lis, J.T. (1991). RNA polymerase II pauses at the 5′ end of the transcriptionally induced Drosophila hsp70 gene. Mol. Cell. Biol. *11*, 5285–5290.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol. Cell *9*, 799–809.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Price, D.H. (2008). Poised polymerases: On your mark…get set…go! Mol. Cell *30*, 7–10.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Reppas, N.B., Wade, J.T., Church, G.M., and Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol. Cell *24*, 747–757.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. *579*, 909–915.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795–804.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol. *7*, 557–567.

Sawado, T., Halow, J., Bender, M.A., and Groudine, M. (2003). The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. Genes Dev. *17*, 1009–1018.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. Development *124*, 2191–2201.

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. *16*, 45–58.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Wade, J.T., and Struhl, K. (2008). The transition from transcriptional initiation to elongation. Curr. Opin. Genet. Dev. *18*, 130–136.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998a). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998b). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. EMBO J. *17*, 7395–7403.

Wang, H., Hammoudeh, D.I., Follis, A.V., Reese, B.E., Lazo, J.S., Metallo, S.J., and Prochownik, E.V. (2007). Improved low molecular weight Myc-Max inhibitors. Mol. Cancer Ther. *6*, 2399–2408.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451–462.

Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev. *13*, 1774–1779.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev. *17*, 1402–1414.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol. Cell *21*, 227–237.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51.

Yin, X., Giap, C., Lazo, J.S., and Prochownik, E.V. (2003). Low molecular weight inhibitors of Myc-Max interaction and function. Oncogene *22*, 6151–6159.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516.

Zhang, F., Barboric, M., Blackwell, T.K., and Peterlin, B.M. (2003). A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. Genes Dev. *17*, 748–758.

Zhou, M., Lu, H., Park, H., Wilson-Chiru, J., Linton, R., and Brady, J.N. (2006). Tax interacts with P-TEFb in a novel manner to stimulate human T-lymphotropic virus type 1 transcription. J. Virol. *80*, 4781–4791.

Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev. *19*, 1668–1673.

Cell

EXTENDED EXPERIMENTAL PROCEDURES

mES Cell Culture

V6.5 (C57BL/6-129) murine ES cells were grown under typical mES conditions on irradiated mouse embryonic fibroblasts (MEFs). In summary, cells were grown on gelatinized tissue culture plates in Dulbecco's modified Eagle medium supplemented with 15% fetal bovine serum (characterized from Hyclone), 1000 U/ml leukemia inhibitory factor (LIF, Chemicon; ESGRO ESG1106), nonessential amino acids, L-glutamine, Penicillin/Streptomycin and β-mercaptoethanol. For location analysis, cells were grown for two passages off of MEFs, on gelatinized tissue-culture plates. Mouse embryonic fibroblasts were prepared and cultured from DR-4 strain mice.

For location analysis following Spt5 and NelfA knockdown, shRNA plasmids targeting the mouse Spt5 and NelfA mRNAs and an empty plasmid (control) (Open Biosystems, Huntsville, AL RMM4534-NM_013676, RMM4534-NM_011914, and RMM4534_NM_010849, RHS4080) were used. We purchased and tested each set of shRNA hairpins for ability of each hairpin to knockdown the mRNA of the factor of interest. We then selected the hairpin that performed the best for use in ChIP-seq analysis. For Spt5, hairpin TRCN0000092761 was used. For NelfA, hairpin TRCN0000124874 was used. For c-Myc, hairpin TRCN0000042516 was used. 293T cells were plated in 6-well dishes at 6 × 10∼5 cells/well. The shRNA plasmids and lentiviral components were cotransfected into 293T cells. V6.5 mES cells were plated in T-75 flasks at 2 × 10∼6 cells/flask. Viral media was collected 48 hr after cotransfection and the V6.5 mES cells were directly infected with the viral media 24 hr after initial plating of the mES cells. The infection media was 1:2 viral media:mES cell media with 2 mM polybrene. The efficiently infected cells were selected for 24 hr post infection with mES cell media containing 2 μM puromycin. V6.5 cells were crosslinked 72 hr post selection and frozen for ChIP-Seq experiments and western blotting. To assess mRNA knockdown a small fraction of cells were collected and mRNA was prepared (using a QIAGEN RNeasy Mini Kit, QIAGEN, Valencia, CA). RT-qPCR was used to determine relative gene expression between mock knockdown and shRNA knockdown samples. Taqman gene expression assays from Applied Biosystems, Foster City, CA were ordered for the Spt5, NelfA, and GAPDH (control) genes (Mm01217228_m1, Mm01170629_m1, Mm99999915_g1, Mm00487804_m1). Western blot analysis (described below) was done to assess knockdown at the protein level prior to ChIP-seq analysis. c-Myc knockdown was also assessed at the mRNA level (described below) using RT-PCR.

For location analysis on mES cells following treatment with small-molecule inhibitors, cells were grown two passages off feeders and prior to formaldehyde crosslinking, the cells were treated by addition of the indicated final concentration of flavopiridol (1 μM for 1 hr for ChIP-chip and ChIP-seq experiments, or the indicated concentration and time for western blot analysis), or c-Myc/Max inhibitor 10058-F4 (50 μM for 6 hr for ChIP-seq experiments or the indicated time for western blot analysis), both dissolved in DMSO, to the growth medium. 50 μM is within the concentration range commonly used for 10058-F4 in vivo to investigate c-Myc function (Arabi et al., 2005; Fang et al., 2009; Faumont et al., 2009; Follis et al., 2008; Hammoudeh et al., 2009; Khanna et al., 2009; Lee et al., 2009; Sampson et al., 2007; Wang et al., 2007). As a control, vehicle alone (DMSO) was added to the growth medium at the same final volume as with drug. Small-molecule inhibitors used were: Flavopiridol (Sigma cat #F3055), and c-Myc inhibitor 10058-F4 (Sigma cat #F3680).

For location analysis following Oct4 shutdown, ZHBTc4 mES cells (Niwa et al., 2000) were grown under standard mES cell culture conditions and expanded for two passages off MEF feeders. ES cell culture media with 2 μg/ml doxycycline was added to the cells for 0 hr, 12 hr and 24 hr prior to formaldehyde crosslinking. Loss of Oct4 was assessed using western blot analysis (see below). Oct4 protein was essentially depleted at the 24 hr time point and we noticed that the cells appeared morphologically different from mES cells and the 0 hr and 12 hr time points. It is well established that loss of Oct4 causes ES cell differentiation. Therefore, in order to minimize measuring secondary effects, we performed our TR analysis on Oct4 targets and non-targets on the earliest time points: 0 hr and 12 hr following doxycycline treatment.

Chromatin Immunoprecipitation

ChIP was done following the Agilent Mammalian ChIP-on-chip protocol (version 9.1, Nov 2006). In summary, mES cells were grown as described above and cross-linked for 15 min at room temperature by the addition of one-tenth of the volume of 11% formaldehyde solution (11% formaldehyde, 50mM HEPES pH 7.3, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH8.0) to the growth media followed by two washes with PBS. Cells were scraped and frozen in liquid nitrogen. 100 μl of Dynal magnetic beads (Sigma) were blocked with 0.5% BSA (w/v) in PBS. Magnetic beads were bound with 10 μg of the indicated antibody. Antibodies used are as follows: Pol II (all; Rpb1 N terminus): Santa Cruz sc-899; Ser5P Pol II: Abcam ab5131; Ser2P Pol II: Abcam (H5 clone) ab24758 with Upstate IgG-IgM linker antibody 12-488; Spt5: gift from Yuki Yamaguchi and Hiroshi Handa (Wada et al., 1998); NelfA: Santa Cruz (A-20) sc-23599;; Ctr9: Bethyl labs A301-395; and Max: Santa Cruz sc-197. For all of the experiments analyzing Pol II occupancy following shRNA-mediated knockdown, flavopiridol, or 10058-F4 treatment, the Pol II (all; Santa Cruz sc-899, Pol II N-20) antibody was used. Crosslinked cells were lysed with lysis buffer 1 (50 mM HEPES pH 7.3, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100) and washed with lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA pH 8.0 and 0.5 mM EGTA pH 8.0).

For Spt5 ChIPs, cells were resuspended and sonicated in lysis buffer 3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.1% Na-Deoxycholate and 0.5% N-lauroylsarcosine) for 8 cycles at 30 s each on ice (18 W) with 60 s on ice between cycles. Triton X-100 was added to a final concentration of 1% to the sonicated lysates. Sonicated lysates were cleared and

incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed four times with RIPA (50 mM HEPES pH 7.3, 500 mM LiCl, 1 mM EDTA, 1% NP-40 and 0.7% Na-Deoxycholate) and once with TE + 50 mM NaCl. Bound complexes were eluted in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) at 65°C for 15 min with occasional vortexing. Crosslinks were reversed overnight at 65°C. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation.

For Pol II Ser5P, Pol II (all), NelfA, Ctr9 and Max ChIPs, cells were resuspended and sonicated in sonication buffer (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS) for eight cycles at 30 s each on ice (18 W) with 60 s on ice between cycles. Sonicated lysates were cleared and incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed three times with sonication buffer, one time with sonication buffer with 500 mM NaCl, one time with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and one time with TE. DNA was eluted in elution buffer. Crosslinks were reversed overnight. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation.

For Pol II Ser2P ChIP, cells were resuspended and sonicated in sonication buffer II (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS) for eight cycles at 30 s each on ice, at 18 W with 60 s on ice between cycles. Sonicated lysates were cleared and incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by Pol II Ser2P. Beads were washed two times with sonication buffer II, one time with LiCl wash buffer (2 mM Tris pH 8.0, 0.02 mM EDTA, 50 mM LiCl, 0.1% NP-40, 0.1% Na-deoxycholate), and one wash with TE. DNA was eluted in elution buffer. This protocol is similar to that used in (Stock et al., 2007). Crosslinks were reversed overnight. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation.

ChIP-PCR Analysis

Max ChIP DNA was analyzed using SYBR Green real-time PCR analysis (Applied Biosystems). Fold enrichment was determined from replicate PCR reactions at seven c-Myc-binding sites (Prdx1, Josd3, Actb, Eef1g1, Nol5, Mat2a, Ybx1) and one non-c-Myc binding site (Gata1), as determined by previously published c-Myc ChIP-seq (Chen et al., 2008), over input DNA. The oligos used for this analysis are as follows:

Prdx1 fwd: ttagttcccggacctgttg
Prdx1 rev: acaaactcgtcccaccaag
Josd3 fwd: cctggagggcgtttttagt
Josd3 rev: acccttttcggaacgtaacc
Actb fwd: gatcactcagaacggacacc
Actb rev: acacgctaggcgtaaagttg
Eef1g1 fwd: CTGGGTCTCCATTGTCTGG
Eef1g1 rev: AGTTCCACCAACCTGCTCA
Nol5 fwd: GGCTCCGAAAAGATGTGAA
Nol5 rev: AGCAGAGGTCGCCCTAAAT
Mat2a fwd: GTCTCCGAAGGTCCCATCT
Mat2a rev: TGAAGGCTAAAGGGCATGT
Ybx1 fwd: AGATCCTGGACCGACTTCC
Ybx1 rev: GTTCCCAAAACCTTCGTTG
Gata1 fwd: agagcctaaaaggtcctcca
Gata1 rev: caccttctccctcctctttc

Hybridization to DNA Microarray

Purified immunoprecipitated DNA was amplified using two rounds of ligation mediated PCR (LM PCR), as described in (Lee et al., 2006b) and the Agilent Mammalian ChIP-on-chip protocol (version 9.1, Nov 2006). LM-PCR immunoprecipitated DNA was labeled with Cy5, LM-PCR input DNA was labeled with Cy3 using Invitrogen Bioprime random primer labeling kit. For microarray hybridization, the Agilent Mammalian ChIP-on-chip protocol (version 9.1, Nov 2006) was followed. In brief, mouse Cot1 DNA, Agilent blocking buffer (1X final conc.), Agilent hybridization buffer (1× final conc.), was added to Cy5- and Cy3-labeled DNA. The mixture was incubated at 95°C for 3 min, followed by 37°C for 30 min. Sample was centrifuged for 1 min and sample was hybridized to Agilent DNA microarray. For experiments testing effects of flavopiridol on NelfA and Spt5 occupancy, and NelfA or Spt5 occupancy following Spt5 or NelfA knockdown, respectively, ChIP samples were hybridized to Agilent arrays MTvB (44,000 features covering the promoter region, from approximately −6 kb to +2 kb, of approximately 10% of mouse genes – MTvB). DNA was hybridized to microarray for 40 hr at 65°C. Microarray was washed and scanned following the Agilent Mammalian ChIP-on-chip protocol. The Agilent DNA microarray scanner BA was used. PMT settings were set manually to normalize bulk signal in the Cy3 and Cy5 channel. Data were processed as described in Lee et al. (2006a) to calculate enrichment ratios and determine bound regions. Figures presented in this manuscript using ChIP-chip data display the chromosomal coordinates from mouse genome build mm6.

Solexa/Illumina Sequencing

All protocols used for Solexa/Illumina ChIP-seq analysis (sample preparation, polony generation on Solexa flow-cells, sequencing, and Solexa data analysis) are described in Guenther et al. (2008) and Marson et al. (2008). A summary of the protocol used is described below.

Sample Preparation

Purified immunoprecipitated DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented DNA was prepared by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:100 dilution (in water) of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. Following amplification, the library was size selected to a narrow range of fragment sizes by separation on a 2% agarose gel and a band between 150–300 bp (representing shear fragments between 50 and 200 nt in length and ~100 bp of primer sequence) was excised. The DNA was purified from the agarose and this DNA library was subsequently used for polony generation and sequencing.

Polony Generation and Sequencing

The DNA library (2–4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1000 strands in 1 μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing.

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 or 32 cycles.

Solexa Data Analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. Sequences were aligned using ELAND software to NCBI Build 36 (UCSC mm8) of the mouse genome. Only sequences that mapped uniquely to the genome with zero or one mismatch were used for further analysis. When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. Refer to Table S1 for a list of the total number of mapped reads used for analysis of each ChIP-seq dataset.

Analysis methods were derived from previously published methods (Johnson et al., 2007; Mikkelsen et al., 2007; Marson et al., 2008; Guenther et al., 2008). Sequence reads from multiple flow cell runs were combined for Pol II Ser2P ChIP-seq dataset. Each read was extended 100 bp, toward the interior of the sequenced fragment, based on the strand of the alignment. The number of ChIP-Seq reads across the genome, in 25 bp bins within a 1kb window surrounding each bin (±500 bp) was tabulated. The 25 bp genomic bins that contained statistically significant ChIP-Seq enrichment was identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a 1kb window can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins (40) into which each read maps, divided by the total number of bins available (we estimated 70%). Enriched bins within 1kb of one another were combined into regions. The complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on December 20, 2008. Genes with enriched regions within 1 kb to their transcription start site to annotated stop site were called bound.

The Poissonian background model assumes a random distribution of background reads, however we have observed significant deviations from this expectation. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples and can create many false positives in ChIP-Seq experiments. To remove these regions, we compared genomic bins and regions that meet the statistical threshold for enrichment to a set of reads obtained from Solexa sequencing of DNA from whole cell extract (WCE) in matched cell samples. We required that enriched bins and enriched regions have 5-fold greater ChIP-Seq density in the specific IP sample, compared with the control sample, normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-Seq reads. A summary of the bound regions and genes for each antibody is provided in Table S2.

For comparison of Pol II occupancy following either shRNA-mediated knockdown or small-molecule inhibition with a control dataset, rank normalization was used to normalize the datasets to be compared. This normalization method is described in (Bilodeau et al., 2009). Briefly, a quantile normalization method was used for analysis. For each dataset compared, the genomic bin with the greatest ChIP-Seq density was identified. The average of these values was calculated and the highest signal bin in each dataset

was assigned this average value. This was repeated for all genomic bins from the greatest signal to the least, assigning each the average ChIP-Seq signal for all bins of that rank across all datasets.

ChIP-seq Binding Tracks Available on GEO Database

ChIP-seq datasets generated in this study can be downloaded from the GEO datasets database (http://www.ncbi.nlm.nih.gov/gds) under the accession number GSE20485. Included under this accession number is ChIP-seq binding tracks that can uploaded onto the UCSC genome browser. Descriptions for the datasets contained in each of the WIG files are listed below.

RNAPolII_Phosphorylation_Tracks.WIG.gz—contains three ChIP-seq datasets mapped to mouse genome mm8 using a +200 bp extension model that can be uploaded to the UCSC genome browser to view binding events (http://genome.ucsc.edu/). The datasets include Pol II Ser5P in wild-type V6.5 mES cells, Pol II Ser2P in wild-type V6.5 mES cells, and Pol II (all) in wild-type V6.5 mES cells + shControl. Enrichment is shown as counts per million reads in 25 bp genomic bins and the track floor is at 1 count per million reads. Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

Factor_Tracks.WIG.gz—contains three ChIP-seq datasets mapped to mouse genome mm8 using a +200 bp extension model that can be uploaded to the UCSC genome browser to view binding events. The datasets include Spt5, NelfA, and Ctr9 in wild-type V6.5 mES cells. Enrichment is shown as counts per million reads in 25bp genomic bins and the track floor is at 1 count per million reads. Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

RNAPolII_PauseFactorKnockdown_Tracks.WIG.gz—contains three ChIP-seq datasets mapped to mouse genome mm8 using a +200bp extension model that can be uploaded to the UCSC genome browser to view binding events. The ChIP experiments were done using the N20 Pol II antibody (sc-899). The datasets include Pol II (all) mES V6.5 + shSpt5, Pol II (all) in mES V6.5 + shNelfA, and Pol II (all) mES V6.5 + shControl. Binding enrichment tracks for all datasets are shown as rank normalized counts in 25bp genomic bins and the track floor is at 2 count per million reads. Rank normalized data are indicated with the title of the track (title of the track is, for example, mES_Pol2_shControl_norm). Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

RNAPolII_FlavopiridolTreated_Tracks.WIG.gz—contains two ChIP-seq datasets mapped to mouse genome mm8 using a +200 bp extension model that can be uploaded to the UCSC genome browser to view binding events. The ChIP experiments were done using the N20 Pol II antibody (sc-899) to determine total Pol II occupancy. The datasets include Pol II (all) in mES V6.5 + DMSO control, and Pol II (all) in mES V6.5 + flavopiridol (P-TEFb inhibitor). In these experiments, cells were treated with either flavopiridol (1 μM) or DMSO alone for 60 min prior to crosslinking and ChIP with Pol II antibody. Binding enrichment tracks for all datasets are shown as rank normalized counts in 25 bp genomic bins and the track floor is at 2 count per million reads. Rank normalized data are indicated with the title of the track (title of the track is, for example, mES_Pol2_Flavo_norm). Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

RNAPolII_10058F4Treated_Tracks.WIG.gz—contains two ChIP-seq datasets mapped to mouse genome mm8 using a +200bp extension model that can be uploaded to the UCSC genome browser to view binding events. The ChIP experiments were done using the N20 Pol II antibody (sc-899) to determined total Pol II occupancy. The datasets include Pol II (all) in mES V6.5 + DMSO control, and Pol II (all) in mES V6.5 + 10058-F4 (c-Myc/Max inhibitor). In these experiments, cells were treated with either 10058-F4 (50 μM) or DMSO alone for 6 hr prior to crosslinking and ChIP with Pol II antibody. Binding enrichment tracks for all datasets are shown as rank normalized counts in 25 bp genomic bins and the track floor is at 2 count per million reads. Rank normalized data is indicated with the title of the track (title of the track is, for example, mES_DMSO_norm). Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

RNAPolII_cMycKnockdown_Tracks.WIG.gz—contains two ChIP-seq datasets mapped to mouse genome mm8 using a +200bp extension model that can be uploaded to the UCSC genome browser to view binding events. The ChIP experiments were done using the N20 Pol II antibody (sc-899) to determine total Pol II occupancy. The datasets include Pol II (all) mES V6.5 + shc-Myc and Pol II (all) mES V6.5 + shControl. Binding enrichment tracks for all datasets are shown as rank normalized counts in 25 bp genomic bins and the track floor is at 2 count per million reads. Rank normalized data are indicated with the title of the track (title of the track is, for example, mES_Pol2_shControl_norm). Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

RNAPolII_Oct4Shutdown_Tracks.WIG.gz—contains three ChIP-seq datasets mapped to mouse genome mm8 using a +200bp extension model that can be uploaded to the UCSC genome browser to view binding events. The ChIP experiments were done using the N20 Pol II antibody (sc-899) to determine total Pol II occupancy. The datasets include Pol II (all) mES ZHBTc4 + doxycycline 0 hr, Pol II (all) mES ZHBTc4 + doxycycline 12 hr, and Pol II (all) mES ZHBTc4 + doxycycline 24 hr. Binding enrichment tracks for all datasets are shown as rank normalized counts in 25 bp genomic bins and the track floor is at 2 count per million reads. Rank normalized data are indicated with the title of the track (title of the track is, for example, mES_Pol2_Dox0hr_norm). Additional tracks identify genomic regions that have been determined enriched above background at the given p value.

Active and Nonproductive Gene Classes in mES Cells

The active and nonproductive genes were classified in mES cells using H3K4me3 (initiation-associated chromatin modification) and H3K79me2 (elongation-associated chromatin modification), as determined by ChIP-seq (Marson et al., 2008), as markers of transcriptional state. Active genes with both H3K4me3 and H3K79me2 chromatin modifications, non-productive genes had only

H3K4me3 and inactive genes did not have H3K4me3 or H3K79me2 chromatin modifications. When showing *Rpl3* as the example of the active gene and *Surb7* as the example of the non-productive gene in Figure 1A and Figure 3A, the Pol II (all) dataset used for generating the gene plots was Pol II (all) shControl (from shSpt5 and shNelfA experiment).

Traveling Ratio Calculation

Pol II levels peak in the 5′ region of many genes. To quantify this effect, we have developed a measure called traveling ratio (TR) that compares the ratio between Pol II density in the promoter region and in the gene region.

We first defined the promoter region from −30 to +300 relative to the TSS and the gene body as the remaining length of the gene. We next calculated the average density/nt from rank normalized ChIP-seq density files (described in Bilodeau et al., 2009) for each region and computed the TR as the ratio between the two. Following perturbation from either shRNA knockdown or small molecule inhibition, TR can shift either through changes in the density of promoter proximal Pol II or changes in the gene body Pol II density. For example, Figure S4 shows examples of genes where knockdown of Spt5 or NelfA cause a decrease in promoter proximal Pol II density while Figure 4B shows examples where knockdown of Spt5 causes increases in Pol II density in gene bodies, sometimes accompanied by lower amounts of promoter proximal density.

TR values were calculated for all Pol II bound genes. Pol II occupancy profiles and their corresponding TR value for several example genes are shown in Figure S1. For this figure, the Pol II (all) shControl (from shNelfA and shSpt5 experiment) was used for the gene plots. TR values also show strong agreement between the two control datasets (Pol II shControl and Pol II DMSO), as distributions of TR values are not statistically different (p value < 0.5) as determined by a Welch's T-Test (Figure S1).

TR values were plotted as a function of gene expression in Figure S1D. Expression data used was from (Hailesellasse Sene et al., 2007). No statistically significant correlation was found between TR and gene expression. In contrast, when Pol II Ser2P occupancy at the gene end (the region ± 1 kb from the 3′ end of the gene) is plotted as a function of gene expression (Figure S1E), we see a weak correlation between the two variables.

Mapping DSIF and NELF Peaks Relative to Pol II Peaks and Correlation Calculations

We find that NelfA and Spt5 enrichment spatially overlaps distributions of Pol II near TSS. To determine the spatial relationships between Pol II, NELF, and DSIF, at individual genes, we mapped the location of significant peaks of Spt5 and NelfA relative to locations of downstream Pol II peaks, as determined by a high resolution peak finding algorithm. This algorithm is designed to operate on top of traditional bound region enrichment models—such as the one used in this study—to precisely map binding sites from ChIP-Seq data at higher resolution. The method operates on the assumption that broader ''bound regions'' of statistical enrichment have already been defined, and that within these ''bound regions,'' exists one or more true binding events. Our method is similar to methods used to map ChIP-Seq peaks (Zhang et al., 2008; Marson et al., 2008) in that it primarily utilizes the requirement of a paired forward read and reverse read peak. We searched 500nt upstream and downstream of the Pol II peak for the most significant peak in NelfA and Spt5 ChIP-Seq enrichment that was called enriched above 1e-9 threshold by our gene calling algorithm. If a gene was not enriched for NelfA or Spt5 at this threshold, then it was automatically determined as not co-localized with the Pol II peak.

Eighty-eight percent of downstream Pol II peaks were colocalized by a corresponding peak of NelfA. Similarly, 61% of Pol II peaks were co-localized by a corresponding Spt5 peak. By visual inspection, we found a majority of Pol II peaks where the peak-finding algorithm failed to identify colocalized Spt5 and NelfA peaks to be co-occupied by NelfA or Spt5 peaks, just below cutoff thresholds for calling a bound gene. This suggests that the high threshold of the peak finding algorithm under reported the number of true instances of Pol II and NELF/DSIF colocalization, and indeed NelfA and Spt5 occupancy positively correlate with Pol II occupancy (Figure S3A) further support for general colocalization of NELF and DSIF to Pol II. Almost all identified promoter proximal peaks of Spt5 and NelfA occurred within ± 50 nt of the Pol II peak. In contrast, when we applied the same analysis to nearest peaks of H3K4me3, the majority of peaks were found more than ±50 nt from the Pol II peak.

Heatmap Analysis on ChIP-seq Occupancy

ChIP-seq enrichment for the indicated factor or modification was determined in 50 bp bins (enrichment in the bin as counts per million), centered on each transcriptional start site. Generally, the gene list for each representation was rank ordered based on the amount of Pol II (all) in mES cells, from most to least to correlate the enrichment of the given factor with the amount of Pol II at each gene. Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm) and Java Treeview (http://www.jtreeview.sourceforge.net) were used to visualize the data and generate figures shown in this manuscript.

To determine the transcriptional state of Oct4, Nanog, and c-Myc target genes in mES cells, enriched genes were determined for each factor at p value 1e-8. Many studies have found that the distance between an Oct4 and Nanog binding site and the gene transcriptional start site can vary widely (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; Marson et al., 2008). However, c-Myc binding is generally centered within 1 kb of the transcriptional start site (Chen et al., 2008; Kim et al., 2008). To take this into account Oct4 and Nanog target genes were determined ±5 kb around each transcriptional start site and c-Myc target genes were determined ±1 kb around each transcriptional start site. The transcriptional state was then determined for each bound gene set through determining the occupancy of Pol II Ser5P, H3K4me3 (initiation-associated chromatin modification), H3K79me2 (elongation-associated chromatin modification), and H3K27me3 (repression-associated chromatin modification). The data were displayed at each gene

centered on the TSS from −2.5kb to +3kb, rank ordered based amount of Pol II Ser5P, using Cluster3.0 and Treeview (described above).

Determining Target and Nontarget Gene Sets

To determine c-Myc and Oct4 target and nontarget genes for analysis on changes in Pol II ChIP-seq occupancy, we first ordered all active genes by c-Myc or Oct4 occupancy near the promoter (±1 kb for c-Myc and ±5 kb for Oct4). For c-Myc genes, the top 1000 and lowest 1000 genes by c-Myc occupancy were demarcated as target and nontarget gene,s respectively. For Oct4, we selected as targets the top 100 Oct4 bound genes that also showed at least a −0.2-fold change in expression following Oct4 shutdown (Matoba et al., 2006). The Oct4 nontargets were determined by taking the lowest 100 Oct4-bound genes for which expression data existed.

Global Analysis of Pol II Occupancy Changes following c-Myc/Max Inhibition

To determine the effect and significance of changes in Pol II occupancy at c-Myc target genes following treatment with 10058-F4, we first derived sets of actively elongating c-Myc targets and non-targets. Actively elongating genes were defined as those bound by Pol II within 1kb of the TSS at a p value of 1e-9 and also bound by H3K79me3 in the first 5 kb of the gene, again at a p value of 1e-9.

Actively elongating genes were ranked by c-Myc levels within ± 1 kb of the TSS, as c-Myc generally associates with its target genes close to the TSS (Chen et al., 2008; Kim et al., 2008). We analyzed the effects on transcription following c-Myc/Max inhibition by generating average Pol II occupancy in the promoter or the gene body region for the high-confidence c-Myc targets using rank normalized datasets. For the average Pol II occupancy analysis shown in Figure 6C, we used the high-confidence set of c-Myc targets and non-c-Myc targets containing the top 100 c-Myc target genes and the bottom 100 genes from the rank ordering. This set of target genes and non-target genes were also used for the Pol II TR analysis following 10058-F4 treatment using rank normalized ChIP-seq density data (Figure 6D).

In order to determine the significance of these changes, we calculated the Pol II density for each gene in the promoter, and the gene body plus gene end region, expanding the target genes to the top 1000 and bottom 1000 genes ranked by c-Myc levels were respectively used as a target and nontarget set for further statistical analysis. Regions are delineated in Figure 6. Averages of the Pol II occupancy at c-Myc targets show a loss of Pol II in the gene body and gene end following 10058-F4 treatment. No such loss is seen in nontarget genes. Additionally, levels of Pol II at the promoter appear unchanged following 10058-F4 treatment in both target and nontarget genes. We used a Welch's T-Test to determine whether distributions of gene densities in control versus 10058-F4 treated cells could be generated by the same distribution in each region for target and nontarget genes. We found significant changes in Pol II occupancy only in the gene body of c-Myc target genes ($p = 7.341e-06$) but not the promoter region ($p = 0.4536$) of c-Myc genes. Non-c-Myc target genes did not have a statistically significant change in either gene region ($p < 0.001$). These data suggest that 10058-F4 causes a loss in Pol II occupancy specifically at c-Myc target genes.

Coimmunoprecipitation Analysis

Coimmunoprecipitation studies were done in mES cells using IgG: Upstate/Millipore, anti-c-Myc: Santa Cruz sc-764, anti-Max: Santa Cruz sc-197, anti-Cdk9 Santa Cruz sc-484, anti-CycT1: gift from David Price (Peng et al., 1998). Max was immunoprecipitated from mES cell lysates made in Lysis Buffer (20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, 1% NP-40, 2 mM EDTA, and protease inhibitors) using anti-IgG control, anti-c-Myc, anti-Max, anti-Cdk9, and anti-CycT1 and incubated overnight at 4°C. CycT1 and Cdk9 were immunoprecipitated with mES cell lysates made in Lysis Buffer using anti-IgG control, anti-c-Myc, anti-Max and anti-Cdk9 and incubated for 3 hr at 4°C. Immunoprecipitates were washed three times with lysis buffer and proteins were analyzed by SDS-PAGE gel electrophoresis followed by western blot analysis.

Western Blots

Western blots were done following standard protocols. Antibodies used for western blots were as follows: Spt5: gift of Yuki Yamaguchi and Hiroshi Handa (Wada et al., 1998); NelfA: gift of Yuki Yamaguchi and Hiroshi Handa; Cdk9: Santa Cruz sc-484 and sc-8338, Ser2P Pol II: Abcam (H5 clone) ab24758; Ser5P Pol II: Abcam ab5131; Brg1: Santa Cruz sc-10768; Max: Santa Cruz sc-197; c-Myc: Santa Cruz-764; Cyclin T1: gift of David Price; Oct4: Santa Cruz sc-5279; and TBP: Abcam ab818. Secondary antibodies used were anti-rabbit IgG (HRP-conjugated; GE Healthcare - NA934V), anti-mouse IgG (HRP-conjugated; GE Healthcare – NA931V), anti-Protein A (HRP-conjugated; GE Healthcare – NA9120V), anti-sheep (HRP-conjugated; Santa Cruz – sc2770) and anti-mouse IgG+IgM (HRP-conjugated; Jackson ImmunoResearch – 115-035-044).

Expression Analysis

RNA was extracted from mES cells treated with 10058-F4 or DMSO alone with biological replicates using QIAGEN Qiashredder and RNeasy kits. Residual DNA was degraded using DNA-free kit for DNase treatment (Ambion). cDNA was generated from the DNA-free RNA using Superscript III First Strand reverse transcriptase PCR kits. Expression was determined using quantitative PCR analysis using Taqman assays in triplicate against Gapdh (Mm99999915_g1), Rnf2 (Mm00803321_m1), Brg1 (Mm01151944_m1), Atic (Mm00546566_m1), Bax (Mm00432050_m1), Nol5 (Mm00479705_m1), Brca2 (Mm00464784_m1), and Zfp451 (Mm00659728_m1). Expression was normalized against Gapdh internal control and displayed as a percentage of expression in

the DMSO alone control. c-Myc mRNA levels were quantitated following sh c-Myc knockdown analysis using duplicate Taqman assays for c-Myc (Mm00487804_m1) and normalized against Gapdh.

SUPPLEMENTAL REFERENCES

Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Mol. Cell. Biol. *26*, 250–260.

Arabi, A., Wu, S., Ridderstrale, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlen, S., Hydbring, P., Soderberg, O., Grummt, I., et al. (2005). c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. Nat. Cell Biol. *7*, 303–310.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. *23*, 2484–2489.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–111117.

Fang, Z.H., Dong, C.L., Chen, Z., Zhou, B., Liu, N., Lan, H.F., Liang, L., Liao, W.B., Zhang, L., and Han, Z.C. (2008). Transcriptional regulation of survivin by c-Myc in BCR/ABL-transformed cells: implications in anti-leukemic strategy. J. Cell. Mol. Med. *13*, 2039–2052.

Faumont, N., Durand-Panteix, S., Schlee, M., Gromminger, S., Schuhmacher, M., Holzel, M., Laux, G., Mailhammer, R., Rosenwald, A., Staudt, L.M., et al. (2009). c-Myc and Rel/NF-kappaB are the two master transcriptional systems activated in the latency III program of Epstein-Barr virus-immortalized B cells. J. Virol. *83*, 5014–5027.

Follis, A.V., Hammoudeh, D.I., Wang, H., Prochownik, E.V., and Metallo, S.J. (2008). Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules. Chem. Biol. *15*, 1149–1155.

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev. *22*, 3403–3408.

Hailesellasse Sene, K., Porter, C.J., Palidwor, G., Perez-Iratxeta, C., Muro, E.M., Campbell, P.A., Rudnicki, M.A., and Andrade-Navarro, M.A. (2007). Gene function in early mouse embryonic stem cell differentiation. BMC Genomics *8*, 85.

Hammoudeh, D.I., Follis, A.V., Prochownik, E.V., and Metallo, S.J. (2009). Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J. Am. Chem. Soc. *131*, 7390–7401.

Khanna, A., Bockelman, C., Hemmes, A., Junttila, M.R., Wiksten, J.P., Lundin, M., Junnila, S., Murphy, D.J., Evan, G.I., Haglund, C., et al. (2009). MYC-dependent regulation and prognostic role of CIP2A in gastric cancer. J. Natl. Cancer Inst. *101*, 793–805.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol. Cell *11*, 721–729.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Lee, W.H., Liu, F.H., Lin, J.Y., Huang, S.Y., Lin, H., Liao, W.J., and Huang, H.M. (2009). JAK pathway induction of c-Myc critical to IL-5 stimulation of cellproliferation and inhibition of apoptosis. J. Cell. Biochem. *106*, 929–936.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonicstem cells. Cell *134*, 521–533.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A., and Ko, M.S. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stemcells by expression profiling. PLoS ONE *1*, e26.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of EScells. Nat. Genet. *24*, 372–376.

Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev. *12*, 755–762.

Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol. Cell *9*, 799–809.

Sampson, V.B., Rong, N.H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., Petrelli, N.J., Dunn, S.P., and Krueger, L.J. (2007). MicroRNA let-7a downregulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. *67*, 9762–9770.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol. *7*, 557–567.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Wang, H., Hammoudeh, D.I., Follis, A.V., Reese, B.E., Lazo, J.S., Metallo, S.J., and Prochownik, E.V. (2007). Improved low molecular weight Myc-Max inhibitors. Mol. Cancer Ther. *6*, 2399–2408.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5C-terminal repeats is critical for processive transcription elongation. Mol. Cell *21*, 227–237.

Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev. *19*, 1668–1673.

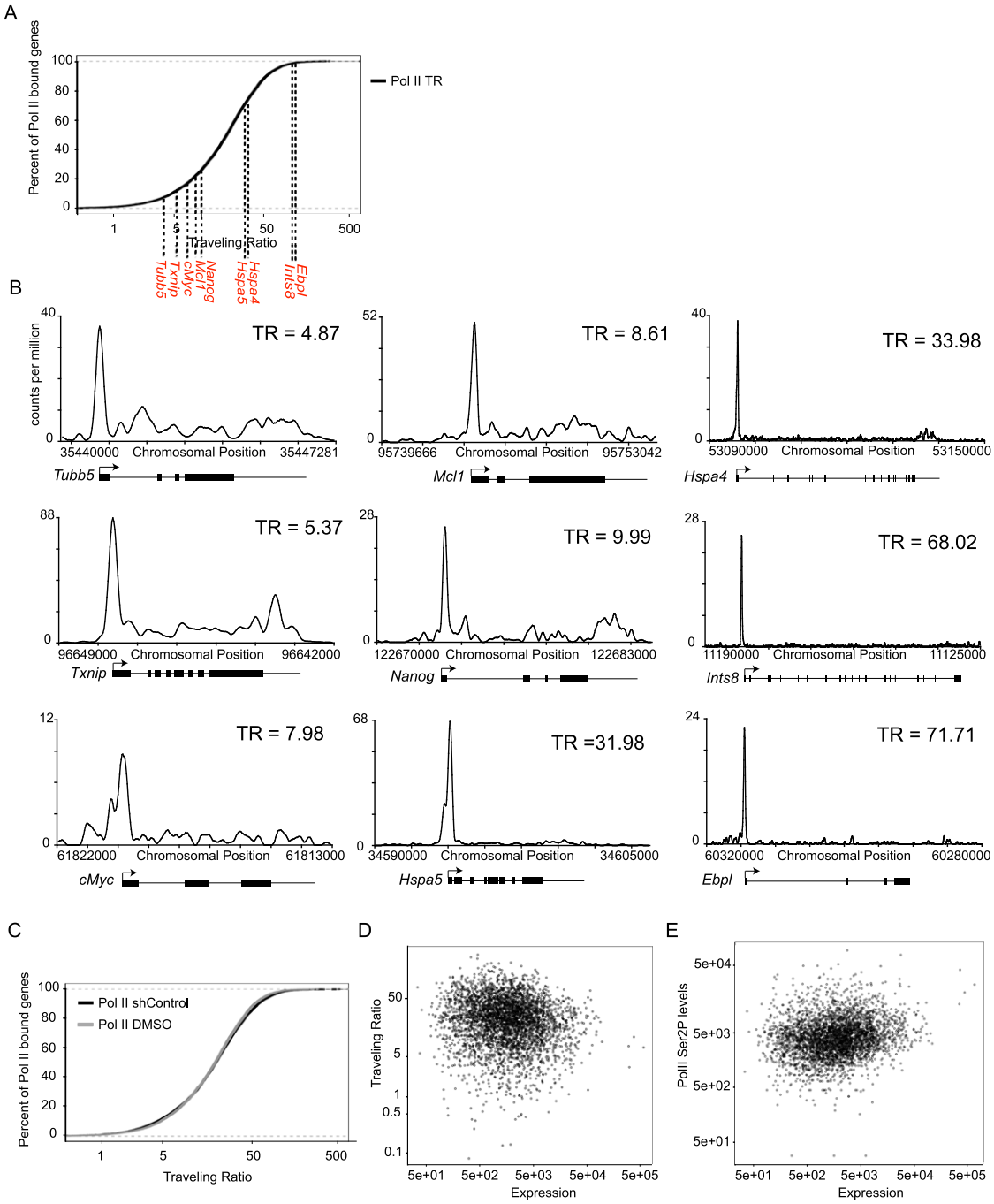

Figure S1. Example Genes with Varying Traveling Ratios, Related to Figure 1

(A) A plot of the percent of Pol II-bound genes with the indicated TR value or less. Each example gene depicted below is shown in red.

(B) Gene plots showing the Pol II occupancy at a given gene and the corresponding TR value to provide the reader with a general idea of how Pol II occupancy at a gene is related to TR. There is a very small correlation (R^2 = 0.17) between gene length and TR, but gene length normalized versions of the TR produce no meaningful changes to our results.

(C) Traveling ratios plots for Pol II occupancy in two mES control datasets (mES cells +shControl (from shNelfA and shSpt5 experiment) and mES cells +DMSO).

(D) Traveling ratio as a function of gene expression, finding no statistically significant correlation between TR and gene expression.

(E) Amount of Pol II Ser2P in the gene end region (±1 kb from the 3′ end of gene) as a function of gene expression, finding a weak correlation between the two.

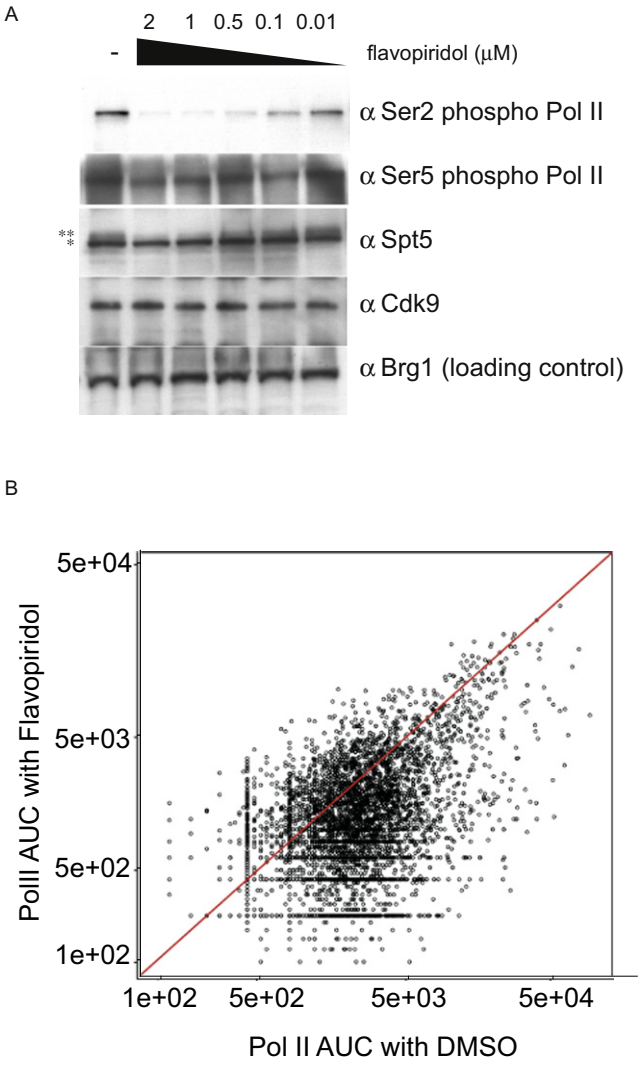

Figure S2. Flavopiridol Treatment in mES Cells Results in Loss of Phosphorylation at P-TEFb Targets, Related to Figure 2

(A) mES cells were treated with the indicated flavopiridol concentration or DMSO alone (–), for 60 min. Nuclear extracts were analyzed by western blot using specific antibodies against Ser2P Pol II, Ser5P Pol II, Spt5, Cdk9, and Brg1 (loading control). **—higher-molecular-weight Spt5 species, most likely the phosphorylated C-terminal repeat form as reported in Yamada et al. (2006), that is flavopiridol sensitive. *—lower-molecular-weight Spt5 species.

(B) Pol II density at the gene end (region defined by ±1 kb from the 3′ end of the gene) is plotted from mES cells treated with DMSO (*x* axis) versus flavopiridol (*y* axis) for all active, non-overlapping genes. AUC is area under the curve, representing the ChIP-seq density for each gene. The vast majority of genes have reduced Pol II occupancy in this region following P-TEFb inhibition.

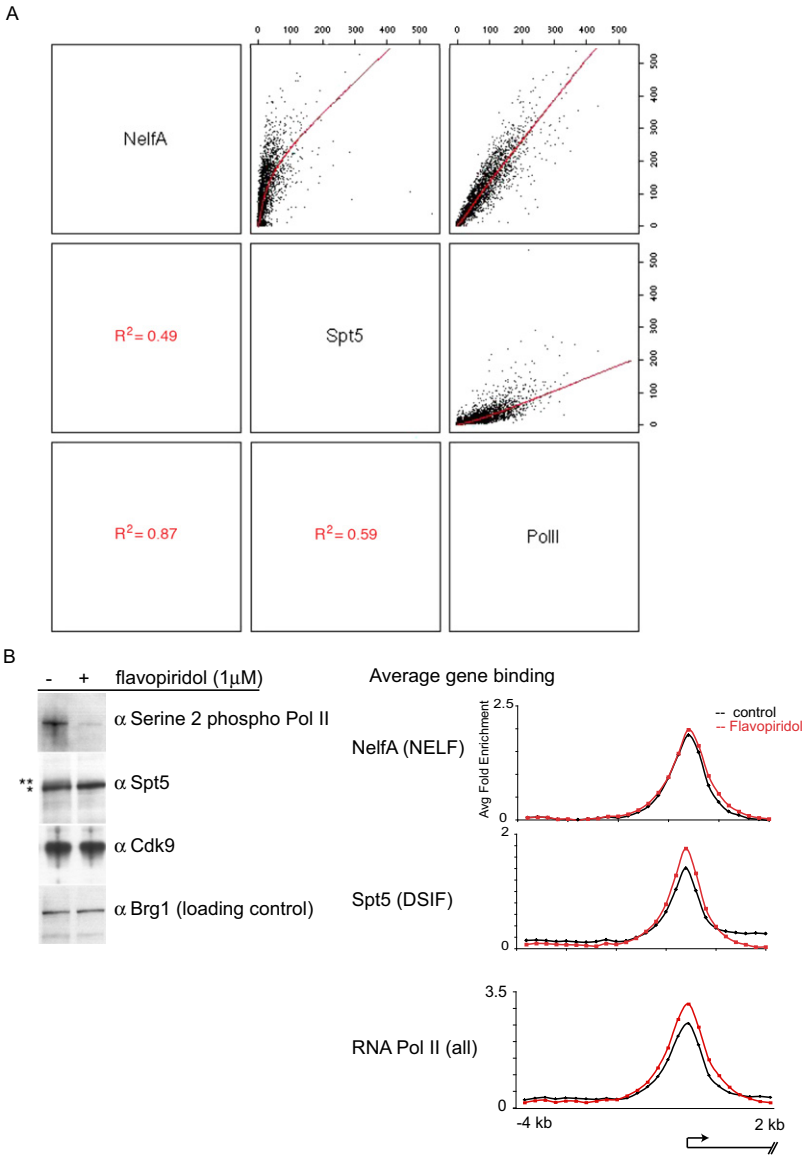

Figure S3. DSIF and NELF Co-occupy Pol II in the Promoter-Proximal Region following P-TEFb Inhibition; Spt5, NelfA, and Pol II ChIP-seq Occupancy Is Highly Correlative; Related to Figure 3

(A) Pairwise correlation analysis of Spt5, NelfA and Pol II (all) ChIP-seq occupancy in mES cells. This demonstrates that Spt5 and NelfA ChIP-seq occupancy positively correlates with Pol II occupancy.

(B) Left—Western blot analysis of mES cells treated with 1 μM flavopiridol or control for 60 min prior. Nuclear extracts were analyzed with specific antibodies against Ser2P Pol II, Spt5, Cdk9, and Brg1 (loading control). **—higher-molecular-weight Spt5 species, most likely the phosphorylated C-terminal repeat form, that is flavopiridol sensitive. *—lower-molecular-weight Spt5 species. Right—Average gene binding for NelfA and Spt5 following 1 μM flavopiridol (red) or control (black) for 60 min in mES cells. The average ChIP-chip enrichment was determined in each bin (250 bp) in each condition and plotted from −4 kb to +2 kb.

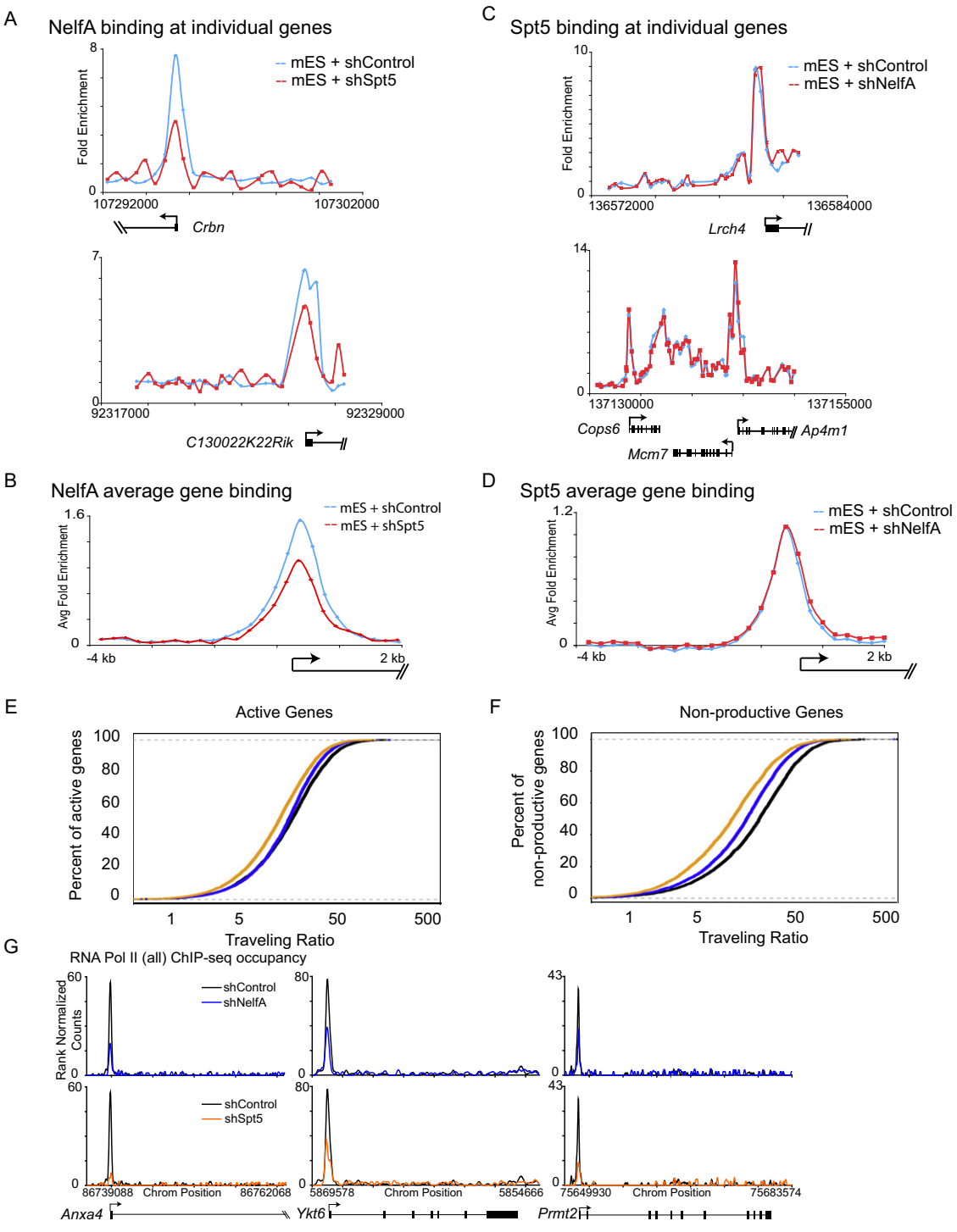

Figure S4. Spt5 Associates with Chromatin following NelfA Knockdown and NelfA Associates with Chromatin following Spt5 Knockdown; Spt5 and NelfA Knockdown Cause a Decrease in Promoter-Proximal Pol II Occupancy at a Subset of Genes; Related to Figure 4

(A) Individual gene examples of NelfA binding in shControl (blue) and shSpt5 (red) mES cells using ChIP-chip.

(B) Average gene binding for NelfA in shControl (blue) and shSpt5 (red) mES cells. The average NelfA enrichment was determined in each bin (250 bp) in each cell type and plotted from −4 kb to +2 kb.

(C) Individual gene examples of Spt5 binding in shControl (blue) and shNelfA (red) mES cells using ChIP-chip. Fold enrichment is plotted over the indicated chromosomal region.

(D) Average gene binding for Spt5 in shControl (blue) and shNelfA (red) mES cells. The average Spt5 enrichment was determined in each bin (250bp) in each cell

type and plotted from −4 kb to +2 kb.
(E) Pol II TR analysis at active genes in mES cells following shControl (black), shSpt5 (orange), and shNelfA (blue) knockdown. Lower TR values indicate a lower degree of pausing and a shift in TR curve to the left indicates a general trend in this gene class to become less paused.
(F) Pol II TR analysis at nonproductive genes in mES cells following shControl (black), shSpt5 (orange), and shNelfA (blue) knockdown. Lower TR values indicate a lower degree of pausing and a shift in TR curve to the left indicates a general trend in this gene class to become less paused. Spt5 knockdown effects Pol II occupancy both active and nonproductive genes and NelfA knockdown mainly effects Pol II occupancy at nonproductive genes.
(G) RNA Pol II (all) ChIP-seq occupancy at three nonproductive genes in mES cells in shControl (black), shSpt5 (orange), or shNelfA (blue).

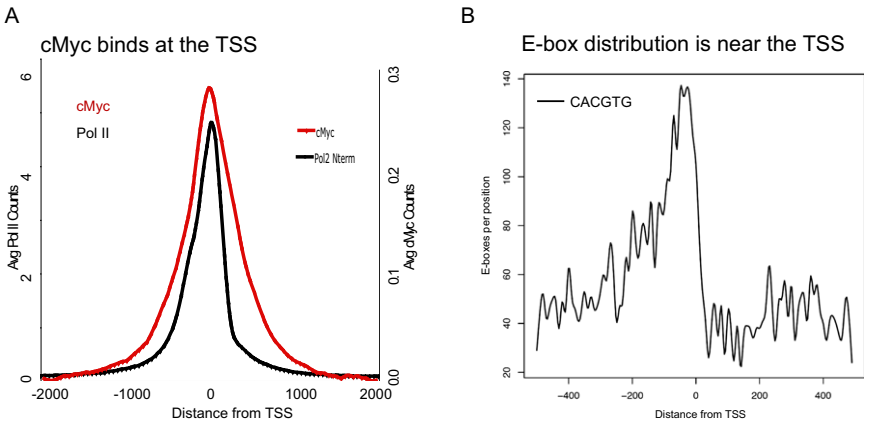

Figure S5. c-Myc Occupies Regions near the Transcriptional Start Site, Related to Figure 5

(A) c-Myc (red) ChIP-seq occupancy is close to the TSS and the average Pol II promoter-proximal peak (black).

(B) Distribution of the canonical E-box core sequence motif (CACGTG), which is recognized by c-Myc, near the TSS.

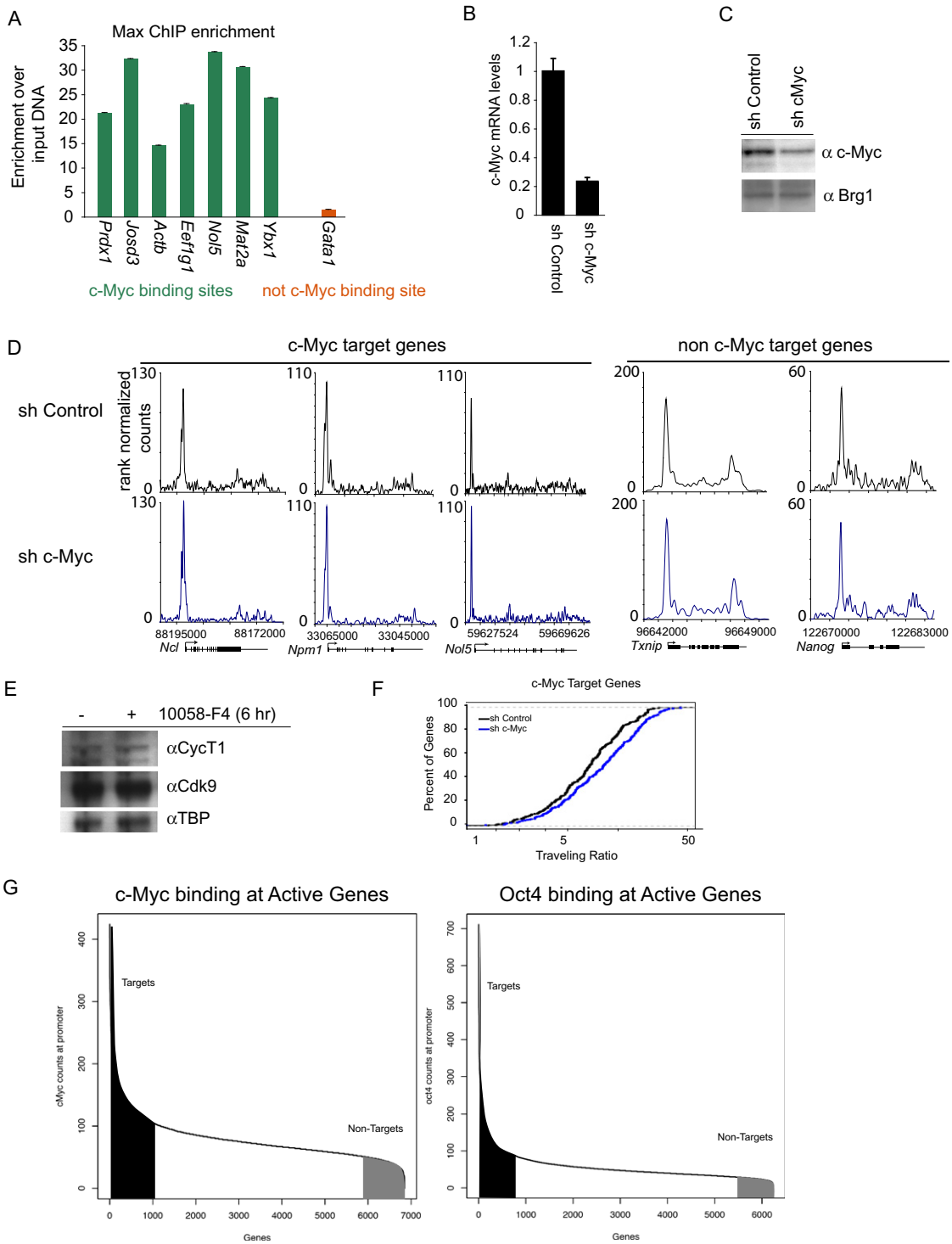

Figure S6. Max Occupies c-Myc-Binding Sites and Pol II Occupancy Is Altered in the Transcribed Region following c-Myc shRNA Knockdown, Related to Figure 6

(A) Max ChIP enrichment in mES cells at seven c-Myc-binding sites (*Prdx1, Josd3, Actb, Eef1g1, Nol5, Mat2a*, and *Ybx1*) and one non c-Myc-binding site (*Gata1*). Max ChIP DNA was analyzed using qPCR at the selected binding sites and enrichment was calculated for replicate PCR reactions over input DNA at each region. This demonstrates that Max binds to c-Myc-binding sites in mES cells, consistent with a model where c-Myc and Max heterodimerize and bind the same sites. Error bars represent SD from duplicate PCR reactions.

(B) Levels of c-Myc mRNA levels in mES cells infected with sh Control or sh c-Myc shRNA constructs. ES cells were infected for 24 hr, then selected for 72 hr prior

to harvesting. c-Myc levels were determined using RT-PCR analysis and normalized against a Gapdh control. Error bars represent SD from duplicate PCR reactions.
(C) Levels of c-Myc protein levels in mES cells infected with sh Control or sh c-Myc shRNA constructs. ES cells were infected for 24 hr, then selected for 72 hr prior to harvesting. c-Myc levels were determined using western blot analysis using an antibody against c-Myc. Brg1 is used as a loading control.
(D) Pol II ChIP-seq binding plots in mES cells with sh Control or sh c-Myc at three c-Myc target genes (*Npm1*, *Ncl,* and *Nol5*) and two non-c-Myc taget genes (*Txnip* and *Nanog*). This panel is demonstrating that at these genes c-Myc knockdown has a similar phenotype to 10058-F4 treatment where Pol II density is reduced in the gene body and the density in the promoter proximal region is unaffected.
(E) Protein levels of P-TEFb components cyclin T1 and Cdk9 with and without 10058-F4 treatment for 6 hr. Protein extracts were analyzed using western blots and probed with antibodies against CycT1 and Cdk9. TBP was used as a loading control. This analysis demonstrates that the phenotype observed following 10058-F4 treatment (similar to P-TEFb inhibition with flavopiridol) is not a result of decreased levels of P-TEFb.
(F) TR analysis on c-Myc target genes in mES cells + sh Control (black) and sh c-Myc (blue), displaying the percent of genes with a given TR. This analysis shows that genes generally become more paused following c-Myc shRNA knockdown, as indicated by the shift in TR to the right.
(G) c-Myc and Oct4 binding at active genes ordered by amount of binding at the promoter (c-Myc ±1 kb, Oct4 ±5 kb). Target (black) and nontarget (gray) gene sets are demarcated, which were used for subsequent analysis on Pol II ChIP-seq occupancy.