Title
Decoupling Epigenetic and Genetic Effects Through Systematic Analysis of Gene Position

Permalink
https://escholarship.org/uc/item/5vz7q8r4

Author
Chen, Menzies

Publication Date
2015-06-29

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Decoupling Epigenetic and Genetic Effects Through Systematic Analysis of Gene Position

A dissertation submitted in partial satisfaction of the Requirements for the degree Doctor of Philosophy in

Bioengineering

by

Menzies Chen

Committee in charge:

Professor Trey Ideker, Chair
Professor Jim Kadonaga
Professor Lorraine Pillus
Professor Bing Ren
Professor Kun Zhang

2013
Copyright

Menzies Chen, 2013

All rights reserved.
The Dissertation of Menzies Chen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013
DEDICATION

To my patient wife and family, past, present, and future, I dedicate this thesis.
Life has many ways of testing a person’s will, either by having nothing happen at all or by having everything happen all at once.

– Paulo Coelho, *The Winner Stands Alone*
# TABLE OF CONTENTS

Signature Page ............................................................................................................. iii

Dedication ................................................................................................................... iv

Table of Contents ...................................................................................................... v

List of Figures ............................................................................................................. viii

List of Tables ............................................................................................................. x

Acknowledgements .................................................................................................... xi

Vita ............................................................................................................................... xii

Abstract of the Dissertation .................................................................................... iii

Chapter 1: Introduction ............................................................................................ 1
   Chapter 1.1: Technologies ....................................................................................... 4
   Chapter 1.1.1: Position-effect assays ...................................................................... 4
   Chapter 1.1.2: The YKO library ............................................................................ 5
   Chapter 1.1.3: Massively parallel barcode sequencing ........................................ 6
   Chapter 1.1.4: Chromatin immunoprecipitation .................................................. 6

Chapter 2: Decoupling Epigenetic and Genetic Effects through Systematic Analysis of Gene Position ............................................................. 13
   Chapter 2.1: Abstract ............................................................................................ 13
   Chapter 2.2: Introduction ....................................................................................... 13
   Chapter 2.3: Results ............................................................................................... 16
      Chapter 2.3.1: A Systematic Position-Effect Screen Using the *Saccharomyces* Gene Deletion Library ................................................................. 16
      Chapter 2.3.2: Insertion of *kanMX* Does Not Significantly Perturb the Chromatin Landscape ................................................................. 17
      Chapter 2.3.3: *kanMX* Expression Is Negatively Correlated with H3K36me3 at the Promoter 18
      Chapter 2.3.4: H3K36me3 Occupancy Is Predictive of *kanMX* Expression .......... 19
      Chapter 2.3.5: H3K36me3 Antagonism of *kanMX* Expression Depends on the Set2-Rpd3 Pathway ................................................................. 20
Chapter 3.5.1: Histone methylations govern gene architecture by defining gene boundaries .......................................................... 63
Chapter 3.5.2: Histone acetylations govern chromatin structure and contribute to gene activation .............................................. 64
Chapter 3.5.3: Histone variant H2A.Z (Htz1) contributes negatively at gene promoters ......................................................................................... 65

Chapter 4: Conclusion .................................................................................. 77
References .................................................................................................. 81
LIST OF FIGURES

Figure 1.1: Position-effect assays. Adapted from Molecular Biology of the Cell, 4th edition. ............................................................... 9

Figure 1.2: Saccharomyces Genome Deletion Project. ........................................ 10

Figure 1.3: Barcode sequencing. ................................................................. 11

Figure 1.4: Chromatin Immunoprecipitation. .................................................. 12

Figure 2.1: kanMX Expression on Chromosome I............................................. 34

Figure 2.2: Insertion of kanMX Does Not Perturb Chromatin but Chromatin Does Perturb kanMX Expression. ................................................................. 35

Figure 2.3: kanMX Is Repressed by H3K36me3 via the Set2-Rpd3 Pathway........... 36

Figure 2.5: H3K36me3 Correlates with Unsuccessful Yeast Knockouts. ............... 38

Supplemental Figure 2.1: RNA-Seq in Four Heterozygous Diploid Strains and Statistics of kanMX Expression on Chromosome I, Related to Figure 2.1. ....................... 39

Supplemental Figure 2.2: Assessing Chromatin Changes due to kanMX and Method for Correlating Expression with Histone Modifications, Related to Figure 2.2........ 41

Supplemental Figure 2.3: Histone Acetylation Affects kanMX Expression at H3K36me3-Regulated Loci, Related to Figure 2.3. ....................................................... 43

Figure 3.1: Schematic illustrating Barcode immunoprecipitation with analysis by high-throughput sequencing. ................................................................. 67

Figure 3.2: BIP-seq measurements of histone acetylation and methylation............. 68

Figure 3.3: Correlation between histone modification and expression before and after insertion of kanMX........................................................................... 69

Figure 3.4: Analysis of transcription factor and chromatin modifier binding and BIP-seq histone modifications. ........................................................................ 70

Figure 3.5: BIP-seq analysis of histone variant H2A.Z (Htz1). .......................... 71
Supplemental Figure 3.1: Replicate plots of the BIP-seq and ChIP-seq measurements in this study. ................................................................. 72

Supplemental Figure 3.2: Validation of ChIP-seq and BIP-seq measurements and analysis of the effect of TSS positioning on H3K4me3 measurements. .............................................. 73

Supplemental Figure 3.3: Analysis of transcription factor and chromatin modifier binding using ChIP-seq histone modifications. ................................................................. 74

Supplemental Figure 3.4: Interactions between transcription factors and chromatin modifiers ordered by mean expression level of each factor’s bound set of genes............ 75

Supplemental Figure 3.5: Analysis of wild-type Htz1 enrichments and overlap of promoters bound by either Nhp6A or Htz1. ................................................................. 76
LIST OF TABLES

Table 2.1: Primers used for quantification of kanMX mRNA. ........................................... 45

Table 2.2: Primers used in ChIP-qPCR to quantify histone modifications at wild-type loci................................................................. 45

Table 2.3: Primers used to quantify enrichment of histone modifications along the kanMX gene. .................................................................................................................. 47

Table 2.4: Primers used as positive and negative controls for all ChIP-qPCR measurements......................................................................................................................... 48
ACKNOWLEDGEMENTS

I would like to thank many who have helped me through the completion of this dissertation. I would first like to acknowledge my thesis advisor, Trey Ideker, who is a true embodiment of a mentor and whose scientific intuition amazes me to this day, for his support and encouragement during my tenure in his lab. His faith and his patience have given me the wherewithal to pursue my ideas, and his guidance has led me in following them through. I would also like to thank my committee members Dr. Jim Kadonaga, Dr. Bing Ren, and Dr. Kun Zhang. In particular, I would like to thank Dr. Lorraine Pillus, without whom much of this work would not have been possible. Her insights into epigenetic transcriptional regulation and her support have enriched my studies and expanded my academic horizons.

I would also like to thank members of the Ideker lab for their support, insights, and friendships, which have given me strength in times of need, laughter in times of grief, and wisdom in times of confusion. In particular I would like to thank Dr. Rohith Srivas, for always sharing the latest papers that have just been published, and his sharp insights that belie his youth on a daily basis. I would like to thank Gordon Bean for his always-unique perspective on life and math. I would like to thank Matan Hofree for his opinions on anything and everything.

Finally, I would like to thank my wife, who is patient, kind, patient, my ever-present reality check, and my rock. I could not have done this without you.

Chapter 2, in full, is a reprint of the material as it appears in: “Chen M, Licon K, Otsuka R, Pillus L, Ideker T. Decoupling epigenetic and genetic effects through
systematic analysis of gene position. Cell Rep. Jan 2 (2013)”. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of a manuscript currently in submission: “Chen M, Ideker T. Systematic gene replacement imparts experimental control for genome-wide analysis of histone-promoter interactions. PLoS ONE. In submission.” The dissertation author was the primary investigator and author of this paper.
VITA

2005 Bachelor of Science Bioengineering University of California, Berkeley
2007 Master of Science Bioengineering University of California, San Diego
2013 Doctor of Philosophy Bioengineering University of California, San Diego

PUBLICATIONS


FIELDS OF STUDY

Major Field: Bioengineering

Studies in Bioinformatics and Transcriptional Regulation
Professor Trey Ideker
ABSTRACT OF THE DISSERTATION

Decoupling Epigenetic and Genetic Effects
Through Systematic Analysis of Gene Position

by

Menzies Chen

Doctor of Philosophy in Bioengineering
University of California, San Diego, 2013

Professor Trey Ideker, Chair

Transcription involves the precise coordination of genetic signals – specific sequences encoded in DNA – with epigenetic signals such as chemical modification of histones. Genome-scale techniques for profiling these signals provide a wealth of information defining how gene sequence, expression and histone modification all vary throughout a genome. However, high-throughput tools that collect measurements for one type of data while explicitly controlling for others are lacking. Classic “position-effect” assays are a low-throughput means for controlling for gene sequence and observing
epigenetic effects on transcription. However, large-scale position-effect screening has remained out of reach due to the seemingly intractable task of generating a large strain library in which a controlled gene is repositioned throughout the genome. Here, we show that systematic gene knockout collections provide an exceptional resource for conducting large-scale position effect studies. Because a single reporter gene replaces each deleted gene, interrogating this reporter provides a sensitive probe into different chromatin environments while controlling for genetic context.

In Chapter 2, I describe the use of the *S. cerevisiae* gene deletion collection to query position effects on histone modification and expression of the *kanMX* gene positioned throughout chromosome I. Using this approach we identify a distinct epigenetic-genetic interaction between H3K36me3 and Rap1 that is not observed when averaging measurements across all genes.

In Chapter 3, I describe the development of a high-throughput technique called BIP-seq to simultaneously query histone modifications at each *kanMX* cassette within a pool of many yeast strains. This technique uses molecular barcodes that were incorporated at the interface between *kanMX* and each wild-type promoter sequence to measure promoter histone modifications while controlling for downstream sequence and expression. Measurements of over 4,800 strains reveals a novel exclusive interaction between the HMGB protein Nhp6A and the histone variant H2A.Z at promoters.
Chapter 1: Introduction

Eukaryotic transcription involves the spatial and temporal coordination of a large number of proteins, including sequence-specific DNA-binding factors, chromatin modifiers, and the general transcriptional machinery and their associated cofactors\(^1\). Eukaryotic DNA is complexed with histone proteins to form chromatin, which serves to package DNA and whose conformation is coordinated with the transcription process\(^2\). As a ubiquitous non-sequence feature of DNA, histones have been extensively studied as an epigenetic feature to understand what role, if any, they might play in the regulation of transcription.

Decades of research have resulted in a model of epigenetic transcriptional regulation in which epigenetic features such as histone occupancy, DNA methylation, and covalent modification of histone tails render genes active or inactive by modulating transcription factor access to underlying DNA\(^3,4\). This model of regulation is primarily built upon correlations between the location and density of epigenetic features and transcriptional outcomes such as mRNA expression level\(^5,6\). For example, hyperacetylation of lysines in N-terminal tails of histones is a feature that strongly correlates with gene activity\(^7\). In its simplest view, this observation has been explained by noting that acetylation contributes a negative charge to the histone tail, decreasing its affinity for negatively charged DNA. The decreased affinity enhances the accessibility of the transcriptional machinery to chromatin.

In recent years, technologies to profile epigenetic features throughout the genome have seen tremendous growth, and numerous genome-wide maps have been generated for multiple organisms across different conditions\(^7-17\). To a large degree, these maps have
validated existing models using high-resolution data that confirm the general correlations that had been previously observed. They have also identified particular combinations of features that appear to associate with distinct levels of gene expression\textsuperscript{7,11,18}. This has led some researchers to hypothesize the existence of a deterministic “histone code” whereby specific combinations of epigenetic features result in discrete transcriptional outcomes\textsuperscript{19}. Other studies illustrate a blended, continuous landscape of epigenetic markers where multiple modifications may share the same if not similar roles in modulating transcription\textsuperscript{7,11}.

To date, a major challenge facing these epigenetic profiling experiments has been lack of a well-controlled environment for isolating epigenetic from genetic regulatory effects. Previous studies have attempted to separate the two by averaging epigenetic states over groups of promoters bound by common TFs\textsuperscript{11,14} or over clusters of genes that are correlated in expression\textsuperscript{7,11,13,14}. However, each cluster is nonetheless comprised of unique gene and promoter sequences that confound the detection of regulatory effects due solely to epigenetics.

To isolate the specific influence of epigenetics, new methods are needed to control for the major genetic inputs to gene transcription, i.e. the DNA sequence of the gene and promoter regions. In this dissertation, I develop an experimental approach that provides this control through the use of position-effect assays in which a controlled gene is repositioned in the genome to vary epigenetic context.

In Chapter 2, I describe the use of the \textit{Saccharomyces} Gene Deletion library (a.k.a. yeast knockout or YKO library) as an existing resource of yeast strains to use as a position-effect screening library. Each gene deletion strain is in fact a gene replacement
wherein a *kanMX* reporter gene replaces a native yeast gene. Because each strain carries *kanMX* in a different genomic position by design, the library forms an excellent resource available for immediate use for querying position effects. To this end, we measured gene expression in 90 strains that position *kanMX* on chromosome I, and related the resulting measurements to the histone modifications found at each position.

In Chapter 3, I develop a genome-wide assay to query protein-DNA interactions at each *kanMX* gene cassette from a pool of thousands of yeast strains. The technique, called barcode immunoprecipitation and analysis by high-throughput sequencing or BIP-seq, resolves measurements from different strains by specifically sequencing unique molecular barcodes that were included directly upstream of each *kanMX* gene cassette in the YKO library. Each barcode uniquely identifies the strain, and was intended to identify the presence of a particular gene deletion strain in a pool of many YKO strains. For the purpose of analyzing position effects, each unique barcode represents a specific position in the genome. We applied BIP-seq to analysis of histone modifications at *kanMX*, and assessed how histone modifications at the interface between a native promoter and *kanMX* are distributed throughout the genome. BIP-seq leverages the finding from Chapter 2 that *kanMX* is more similarly expressed regardless of gene position than wild-type genes to control for gene expression level. Thus, BIP-seq measurements of histone modification do not reflect histone states that arise due to transcription, but rather reflect histone states that result from activity with native promoter regions.
Chapter 1.1: Technologies

Chapter 1.1.1: Position-effect assays

A position-effect assay measures changes in a genetic output (i.e., gene expression) that result from repositioning a gene to new genomic positions. In classic position-effect experiments that demonstrated how epigenetics can have a profound effect on gene expression, reporter genes were translocated to new chromosomal positions near telomeres, where DNA takes on an epigenetically distinct form\textsuperscript{20}. Reporter gene activity was monitored in each translocation experiment and was observed to differ markedly depending on whether the gene was positioned near hetero- or euchromatic regions. For instance, when \textit{ADE2} was translocated from euchromatin to subtelomeric heterochromatin in yeast cells, resulting colonies were sectored red and white in appearance, indicating a variable decrease in \textit{ADE2} expression (Figure 1.1). These results suggested that translocated genes inherit the epigenetic state defined at its new chromosomal locus and undergo epigenetic changes in gene expression\textsuperscript{20}.

This seminal work demonstrates a well-controlled experiment that effectively decouples epigenetic and genetic effects by controlling for gene sequence, and observing how the distinct chromatin context influences gene expression. The experimental design itself was elegant, and involved genetic engineering to create new strains that contain the gene of interest in its new location. However, the range of loci that could be tested in the study was limited to the handful of strains that could feasibly be engineered by the small team of authors, and the methods of quantitation which were state-of-the-art in 1990 are now antiquated by current standards in 2013. In the 23 years since Gottschling, et al.\textsuperscript{20}, much more quantitative and precise methods for quantifying gene expression have been
developed, including quantitative PCR, microarray technologies, and high-throughput DNA sequencing. Though genetic engineering techniques have also advanced\textsuperscript{21}, the generation of new strains for position-effect assays remain a high barrier to large-scale position-effect screening. In this dissertation, I repurpose an existing resource to overcome the genetic engineering barrier to position-effect screening.

Chapter 1.1.2: The YKO library

The YKO library is an openly-available collection of mutant yeast strains \textit{(Saccharomyces cerevisiae)} containing individual start- to stop-codon deletions of all non-essential open reading frames (ORFs) throughout the yeast genome (Open Biosystems, Huntsville, AL). Each ORF is “knocked out” using a gene deletion cassette that replaces the gene via homologous recombination\textsuperscript{22} (Figure 1.2). Positive selection is performed using a reporter gene, \textit{kanMX}, that confers antibiotic resistance to its host but is otherwise physiologically neutral\textsuperscript{23}. As an additional feature, each strain contains a unique molecular barcode that allows its relative abundance to be quantified in pooled cultures using microarrays containing probes complementary to each barcode or high-throughput sequencing\textsuperscript{24,25}.

Although the YKO library was originally created for functional genomics, it has resulted in a fortuitous secondary effect that forms the basis for this dissertation: each yeast strain from the library contains a single reporter gene precisely positioned to a different locus within the genome. Furthermore, the molecular barcodes that uniquely identify each strain also identifies the precise position of the \textit{kanMX} cassette within that strain.
A principal disadvantage for using YKO strains for position-effect screen is clearly the absence of the associated wild-type gene that may be necessary for normal biological function. To address this concern, we utilize heterozygous diploid strains from the library so that each strain possesses at least one full wild-type copy of every gene.

Chapter 1.1.3: Massively parallel barcode sequencing

The inclusion of unique molecular barcodes in the YKO library enables the relative abundance of each strain in a pooled culture to be quantified \(^\text{25-27}\). After genomic DNA is harvested from a pool of yeast, barcodes are prepared by PCR-amplification using common primers (U1, U2, D1, and D2, see Fig. 1.2) that flank each barcode. The relative abundance of each yeast pool is then determined by analyzing each prepared sample by microarray \(^\text{25,28}\) or high-throughput sequencing \(^\text{26,27}\) (Figure 1.3).

In this dissertation, we use the molecular barcodes in each strain to enable pooled analysis of protein-DNA interactions in BIP-seq. Following immunoprecipitation from a pool of many yeast knockout strains, BIP-seq specifically amplifies barcode sequences to measure the relative abundance of a protein of interest at each \textit{kanMX} cassette.

Chapter 1.1.4: Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) remains a principal technology for analyzing protein-DNA interactions. ChIP begins by crosslinking proteins and DNA, and is followed by fragmentation of chromatin into DNA segments ~500 bp in length. Antibodies are then added to purify and precipitate a specific protein of interest and any
associated DNA. This DNA is then analyzed to determine the genomic position where the protein of interest is bound (Figure 1.4; for review, see 29).

Though ChIP has existed since 198430, analysis of the resulting immunoprecipitated DNA was limited to analyses of individual loci until the late 1990s, when microarray technology enabled simultaneous analysis of thousands of gene loci from one immunoprecipitate31-33. This technique combined chromatin immunoprecipitation (“ChIP”) with microarray analysis (“chip”) and became known as ChIP-chip, or ChIP-on-chip.

With the development of second-generation sequencing, ChIP-chip has since been replaced in many studies by ChIP-seq, which uses second-generation high-throughput sequencing (“seq”) to analyze immunoprecipitated DNA in place of microarray analysis.

The earliest applications for ChIP-chip and ChIP-seq have ascertained the genomic locations where sequence-specific transcription factors are bound34-36. However, ChIP has also been used in other applications, including nucleosomal or histone ChIP, in which the protein of interest is not a transcription factor, but a histone protein with or without specific chemical modifications7,9,11,15. These techniques have been widely used to define histone state across the entire epigenome, and remain a principal technique for studying epigenetics.

In this dissertation, ChIP is used to monitor protein-DNA binding in both wild-type and knockout strains to determine whether and how kanMX may influence histone modifications (Chapter 2), and also how native promoter sequence contributes to histone state upstream of a controlled DNA sequence (Chapter 3).
Together, these technologies have increased our understanding of histone modifications and gene expression. I present my thesis work in the following chapters, in which I leverage these technologies to decouple epigenetic and genetic effects on the expression of the *kanMX* gene, and then build on these technologies to develop a new tool for analyzing position effects on a genome-wide scale.
Figure 1.1: Position-effect assays. Adapted from Molecular Biology of the Cell, 4th edition.

(A) Yeast colonies bearing the ADE2 gene in its wild-type position appear white when grown on rich media. When ADE2 is repositioned near telomeres, colonies appear red with white sectors. A red color arises in the absence of adequate ADE2 expression, indicating a variable decrease in ADE2 expression upon gene repositioning. (B) A similar position effect was observed earlier in Drosophila melanogaster. Flies whose genomes had undergone a rare chromosome inversion repositioned the white gene from euchromatin to a position near heterochromatin. These flies bore a mosaic of red and white eye pigmentation, indicating reduced expression of white in a subpopulation of eye cells. Adapted from Alberts, et al., Molecular Biology of the Cell, 4th Edition.
Figure 1.2: Saccharomyces Genome Deletion Project.

The \textit{kanMX} deletion cassette is comprised of a \textit{kan}^R gene fused to a constitutive promoter, \textit{pTEF}, from the related fungus \textit{Ashbya gossypii}. The \textit{kan}^R reporter gene confers antibiotic resistance to G418. \textit{pTEF} contains two binding sites recognized by the Rap1 TF in \textit{S. cerevisiae} and drives expression of \textit{kan}^R. Adapted from the \textit{Saccharomyces} Genome Deletion Project.
Figure 1.3: Barcode sequencing.

Flanking each promoter-gene pair are two oligonucleotide tags, each containing common primer sequences surrounding a 20-bp oligonucleotide “barcode” that uniquely identifies the strain in the library. The tags are labeled UPTAG and DOWNTAG to describe their positions with respect to $kan^R$ \(^{22}\). Each tag is surrounded by universal priming sites, U1/U2 and D1/D2, which allow amplification of the barcode tag features from pooled cultures. The Rap1 binding sites reside approximately 275 base pairs from the UPTAG molecular barcode, and approximately 330 bp downstream of the cassette insertion site \(^{38}\), i.e. the transition from native DNA/chromatin.
Figure 1.4: Chromatin Immunoprecipitation.

After live cells are crosslinked with formaldehyde, chromatin is fragmented, typically using enzymatic or mechanical shearing methods. Antibodies specific to a protein of interest bind their target in solution and are precipitated along with associated DNA. Crosslinks are reversed to release DNA, after which it is purified and labeled (Cy5 channel, red). A portion of the whole-cell extract is saved for measuring background hybridization (Cy3 channel, green). The two labeled samples can then be analyzed on a two-color microarray. Adapted from Ren, et al.\textsuperscript{31}. 
Chapter 2: Decoupling Epigenetic and Genetic Effects through Systematic Analysis of Gene Position

Chapter 2.1: Abstract

Classic “position-effect” experiments repositioned genes near telomeres to demonstrate that the epigenetic landscape can dramatically alter gene expression. Here, we show that systematic gene knockout collections provide an exceptional resource for interrogating position effects, not only near telomeres but at every genetic locus. Because a single reporter gene replaces each deleted gene, interrogating this reporter provides a sensitive probe into different chromatin environments while controlling for genetic context. Using this approach, we find that, whereas systematic replacement of yeast genes with the \textit{kanMX} marker does not perturb the chromatin landscape, chromatin differences associated with gene position account for 35\% of \textit{kanMX} activity. We observe distinct chromatin influences, including a Set2/Rpd3-mediated antagonistic interaction between histone H3 lysine 36 trimethylation and the Rap1 transcriptional activation site in \textit{kanMX}. This interaction explains why some yeast genes have been resistant to deletion and allows successful generation of these deletion strains through the use of a modified transformation procedure. These findings demonstrate that chromatin regulation is not governed by a uniform “histone code” but by specific interactions between chromatin and genetic factors.

Chapter 2.2: Introduction

Transcription requires the precise coordination of genetic signals encoded in DNA with epigenetic signals such as modification of histones\textsuperscript{2,39}. To study which chromatin
modification signals are most informative, powerful genome-scale methods have been applied for correlating profiles of histone modification state with profiles of gene expression measured over all genes\textsuperscript{40}. These studies have identified a number of histone states that associate with transcriptional activity, such as trimethylation of lysine 4 in histone H3 (H3K4), which is found preferentially at the 5′ regions of highly expressed genes\textsuperscript{41}. Both histone states and gene expression state vary along a genome, however, making it difficult to discern which of these states is the cause and which is the effect. Moreover, methods based on genome-wide correlation identify only the most general chromatin effects and miss those that apply preferentially to subsets of genes or promoters, i.e., epigenetic-genetic interactions. For example, the genome-wide-positive association between H3K4me3 and transcription contradicts a previously identified role for H3K4me3 in promoting gene silencing at telomeres, silent mating-type loci, or rDNA regions\textsuperscript{42,43}. Such interactions are increasingly important for understanding human diseases such as cancer, in which both genetic and epigenetic alterations can specifically enable oncogenes and tumor suppressor genes\textsuperscript{44,45}.

Isolating the pure chromatin contribution to gene expression would mean controlling for the genetic sequence as the chromatin context was varied. This is precisely the means by which position-effect variegation was first observed in \textit{Drosophila} (reviewed by\textsuperscript{46}). Gottschling et al.\textsuperscript{20} went on to establish the now classic “position effect” in yeast, in which relocating genes from their wild-type loci to positions near telomeric heterochromatin revealed repressive effects on gene expression due to the distinct chromatin landscape.
Ideally, such position-effect experiments could be performed systematically by measuring the expression of the same gene positioned at each chromatin context, i.e., across all gene positions in the genome. Such a systematic screen has never been performed, perhaps because of the perceived difficulty of such a task. We reasoned that this task might be feasible, however, using the gene knockout library constructed in budding yeast by the *Saccharomyces* Genome Deletion Project\textsuperscript{22}. This project targeted each yeast open reading frame (ORF) for replacement with the *kanMX* cassette, which contains the *TEF* promoter from *Ashbya gossypii* upstream of the *kan*\(^R\) gene conferring resistance to the antibiotic G418\textsuperscript{23}. Deletion strains have been constructed for approximately 6,000 yeast genes representing >90\% of known or suspected ORFs\textsuperscript{28}. Although this deletion library was originally constructed to study gene function, it also possesses the critical feature needed for a systematic position-effect assay: each strain carries the same promoter and gene positioned over the range of chromatin environments presented by a genome.

Here, we show that the *Saccharomyces* Genome Deletion library indeed provides a foundation for systematic gene position experiments. These experiments, which effectively separate epigenetic from genetic effects, permit estimates of the total genome-wide contribution of chromatin to gene transcription while preserving genetic and epigenetic integrity far from the site of gene replacement. Integration of the resulting data with genome-wide maps of histone modifications leads us to propose a specific role for histone H3 lysine 36 trimethylation (H3K36me3) in transcriptional control, via an epigenetic-genetic interaction with the Rap1 transcriptional activation site in the *TEF* promoter.
Chapter 2.3: Results

Chapter 2.3.1: A Systematic Position-Effect Screen Using the *Saccharomyces* Gene Deletion Library

To explore the use of gene deletion libraries for position-effect studies, we selected yeast strains from the heterozygous diploid collection corresponding to all *kanMX*-mediated gene replacements on chromosome I. Heterozygous diploids retain one functional copy of the deleted gene and thus minimize unwanted effects of gene deletion on cell function\(^47\). This assumption was supported by the finding that these strains have wild-type growth phenotypes and mRNA-expression profiles (Supplemental Figures 2.1A–2.1L). Next, we used quantitative RT-PCR (qRT-PCR) to obtain sextuplicate measurements of *kanMX* expression in each of the chromosome I deletion strains (Figure 2.1). As a group, the expression measurements showed significant variation from locus to locus (F test, \(p < 2.04 \times 10^{-11}\)). This variation was due to at least 19 loci that had significantly higher or lower expression (Figures 3.1B and 3.1C), with a 4-fold dynamic range between the highest and lowest expressing loci. A comparison of expression variance at each position with variance across all measurements revealed that gene position accounts for approximately 35% of variation in *kanMX* expression (Experimental Procedures). In addition, decreased expression was observed at telomere-proximal loci (i.e., subtelomeric loci defined by\(^48\)), showing that our assay recapitulated well-established results (Supplemental Figure 2.1M). In contrast, the two pericentric loci were expressed at near-average levels (Supplemental Figure 2.1M), also consistent with previous observations that *S. cerevisiae* centromeric regions remain somewhat
transcriptionally active\textsuperscript{49}. Because each \textit{kanMX} insertion is directly downstream of the native wild-type gene promoter, it is possible that transcriptional machinery recruited by the native promoter directly influences \textit{kanMX} expression. To assess this possibility, we compared our \textit{kanMX} expression measurements to the expression levels of the wild-type genes being replaced and found no relationship between wild-type and \textit{kanMX} expression and no difference in \textit{kanMX} expression between loci that are silenced versus actively expressed in wild-type (Supplemental Figure 2.1N). Nonetheless, it remains conceivable that the native promoter contributes to chromatin state at \textit{kanMX} that then indirectly alters expression—we consider this a part of the position effect being examined.

Chapter 2.3.2: Insertion of \textit{kanMX} Does Not Significantly Perturb the Chromatin Landscape

Although it is assumed in position-effect assays that the inserted construct inherits the chromatin landscape of its new position, we sought to test this assumption directly by comparing the levels of different histone modifications along the \textit{kanMX} cassette with their corresponding levels along the wild-type gene. The technique of chromatin immunoprecipitation followed by qRT-PCR (ChIP-qRT-PCR) was used to quantify levels of five different histone modifications: trimethylation of histone H3 lysines 4, 36, and 79 (H3K4me3, H3K36me3, H3K79me3, respectively) and acetylation of histone H3 lysines 9 and 14 (H3K9ac and H3K14ac, respectively). Measurements were made at sites along the promoter and gene at each of ten different gene knockout positions on chromosome I, two of which had been observed to express \textit{kanMX} at significantly higher or lower levels than average (Supplemental Figure 2.2A; Supplemental Table 2.2). For all
histone modifications tested, the chromatin landscape was found to be very similar between the kanMX cassette and wild-type (Figures 2.2A–2.2J and Supplemental Figure 2.2). Correlations were particularly strong over the kanMX regulatory region and the corresponding 5′ region of the wild-type gene (Pearson r ≥ 0.82, Figures 2.2A–2.2E), suggesting that the insertion of kanMX does not significantly perturb the histone modification landscape in heterozygous diploid gene deletion strains.

Chapter 2.3.3: kanMX Expression Is Negatively Correlated with H3K36me3 at the Promoter

Because the insertion and expression of kanMX do not appear to perturb histone modifications, we next sought to test the converse hypothesis: that histone modifications are predictive of kanMX expression. We used a sliding-window approach to compute the correlation between the kanMX expression levels we had measured along chromosome I and the published occurrence at these loci of seven different histone modifications: H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K79me3, H3K9ac, and H3K14ac7 (Supplemental Figures 2.2D–2.2F; Experimental Procedures). Previous studies have shown that H3K4me3, H3K9ac, and H3K14ac are enriched at the 5′ region of yeast genes at levels that correlate with transcription, whereas H3K4me2 and H3K4me1 are enriched in the gene body and 3′ regions, respectively7,41. H3K36me3 is found over middle and 3′ regions of genes, where it is thought to repress spurious intragenic transcription50. H3K79me3 is enriched within gene bodies, but its deposition is not closely linked with transcription7.
Of the modifications examined, we identified strong anticorrelation between *kanMX* gene expression and H3K36me3 occupancy at the promoter (Figure 2.2K). Such an association was not identified previously in genome-wide histone profiling studies. In contrast, the expression profile of the wild-type genes along chromosome I showed no correlation with H3K36me3 but was positively correlated with other chromatin states such as histone acetylation and H3K4me3 (Figure 2.2L), relationships that have been previously well established\(^6\). Thus, it appears that H3K36me3 has a negative association with expression of the *kanMX* gene, but not with genes in general, which tend to be associated with a variety of other modifications (Figure 2.2M).

Chapter 2.3.4: H3K36me3 Occupancy Is Predictive of *kanMX* Expression

If the negative interaction we have identified between H3K36me3 and *kanMX* expression on chromosome I is general, we reasoned that levels of this histone modification should be predictive of *kanMX* expression on the other 15 yeast chromosomes (II–XVI). Among these chromosomes, ten loci were randomly selected from genomic regions with either reduced or elevated H3K36me3 occupancy. Measurements by qRT-PCR revealed that *kanMX* expression levels were substantially higher when *kanMX* is positioned in regions of reduced H3K36me3 in comparison to regions of elevated H3K36me3 occupancy (Figure 2.3A). Therefore, the dependency of *kanMX* expression on the absence of H3K36me3, a relationship inferred from loci on chromosome I, is indeed predictive of *kanMX* expression throughout the genome.
Chapter 2.3.5: H3K36me3 Antagonism of kanMX Expression Depends on the Set2-Rpd3 Pathway

H3K36me3, a modification catalyzed by the Set2 methyltransferase, is a known element contributing to the repression of spurious transcription initiation via recruitment of the Rpd3 histone deacetylase complex. To test the hypothesis that this mechanism of regulation may also play a role in relation to kanMX, we measured kanMX expression in strains without SET2 or RPD3. We observed increased kanMX expression compared to wild-type at each of ten gene positions in an rpd3Δ background and eight out of nine positions in a set2Δ background (p < 0.003 and p < 0.05, respectively, paired t test; see Figure 2.3B). These results suggest a causal role for H3K36me3 in the regulation of kanMX gene expression and that this regulation is mediated through Set2 and Rpd3.

To further explore the connection to SET2 and RPD3, we looked for differences in chromatin organization that might co-occur with H3K36me3 to explain significantly higher or lower kanMX expression. We found additional support for a connection to the Set2-Rpd3 pathway by comparing histone acetylation levels between loci for which kanMX is expressed at either high or low levels. Five acetylated lysines (H2AK7, H2BK11, H2BK16, H4K12, and H4K16) were significantly associated with kanMX expression (Mann-Whitney U test, p < 0.05; Supplemental Figures 2.3A–2.3E) and significantly anticorrelated with H3K36me3 (Pearson r < −0.7, p < 0.05; Supplemental Figures 2.3F–2.3J, red dots). The anticorrelation between H3K36me3 and the acetylated lysines was particularly striking in comparison to the background correlation among all loci, which was insignificant in all five cases (Supplemental Figures 2.3F–2.3J, black dots). Notably, each of these lysine residues except H4K16 is a
known deacetylation target of Rpd3 and lends further support to a role for the Set2-Rpd3 pathway in kanMX expression.

Chapter 2.3.6: An Antagonistic Interaction between H3K36me3 and Rap1 Binding

The canonical mechanism for chromatin-mediated transcriptional regulation involves modulation of transcription factor (TF) binding upstream of a gene. Because the kanMX cassette is activated by binding of the Rap1 TF to an upstream activation sequence in the TEF promoter, we hypothesized that the repressive interaction between H3K36me3 and kanMX might act through modulation of Rap1 binding. To test this idea, we used antibodies to either Rap1 or H3K36me3 to perform ChIP-qRT-PCR in five knockout strains. We found that Rap1 binding is indeed elevated at loci with low levels of H3K36me3 occupancy and depressed at loci with high levels of H3K36me3 (Pearson $r = -0.97$; Figure 2.4A). Thus, elevated levels of H3K36me3 are predictive not only of reduced kanMX expression but also of reduced binding of Rap1 to the kanMX promoter.

Another test of the apparent antagonism between H3K36me3 and Rap1 is to examine whether this interaction takes place not only at the kanMX locus but also at the numerous other Rap1 binding sites encoded across the genome. For this purpose, we compared the genome-wide binding profiles of Rap1 and H3K36me3, both of which have been published previously by Koerber et al. and Pokholok et al. We found that the genomic regions associated with lower levels of H3K36me3 do indeed tend to be bound more frequently by Rap1 (Figure 2.4B).
Chapter 2.3.7: Gene Expression Downstream of a Rap1 Motif Is Inversely Correlated with H3K36me3

Rap1 recognizes multiple *cis*-regulatory motifs in DNA, and it is known to take on different conformations depending on the sequence to which it is bound. Further investigation showed that the Rap1/H3K36me3 association is strongest at promoters containing an identical Rap1 binding motif to the one carried in the *kanMX* cassette (GCCCATACAT, henceforth called the Rap-kan box). Indeed, among genes downstream of a Rap-kan box, expressed transcripts have lower occupancy of H3K36me3 relative to nonexpressed transcripts (Mann-Whitney U test, $p < 5.4 \times 10^{-4}$, Figure 2.4C, $n = 30$ expressed, 13 nonexpressed), whereas H3K4me3 levels are no different ($p = 0.89$, Figure 2.4D). Interestingly, a slightly different histone distribution is observed when considering genes bearing a more general motif (ACACCCRYACAY, henceforth called the Rap box; In this more general set of genes ($n = 130$ expressed, 134 nonexpressed), the distribution of H3K36me3 among nonexpressed genes appears bimodal, with roughly half of the genes associated with high H3K36me3 occupancy and half of the genes associated with low H3K36me3 occupancy. Even for this expanded Rap box motif, however, it is still the case that H3K36me3 is significantly elevated in nonexpressed genes relative to expressed genes ($p < 3.2 \times 10^{-4}$), whereas H3K4me3 is significantly depressed ($p < 5.9 \times 10^{-3}$, Supplemental Figure 2.4).

Chapter 2.3.8: H3K36me3 Has Likely Interfered with Deletion of Some Yeast Genes

The *Saccharomyces* Genome Deletion Project constructed knockout strains covering many but not all genic positions throughout the genome. Of the 528 yeast ORFs
that have not yet been included in the collection, 321 were attempted but did not yield successful transformants (A. Chu, personal communication). Given the observed H3K36me3-mediated transcriptional repression of \textit{kanMX}, we postulated that this interaction might explain why certain ORFs failed the deletion process. In support of this hypothesis, we observed that H3K36me3 is significantly enriched at ORFs that failed deletion (Figure 2.5A, $p < 2 \times 10^{-16}$). Different ORF deletions were attempted a different number of times, however, potentially introducing a sampling bias. To guard against such bias, we also examined the set of “Dubious ORFs” for which deletion was attempted once and only once. Even in that restricted set, H3K36me3 remained significantly enriched at loci that failed deletion (Figure 2.5A, $p < 5 \times 10^{-6}$).

To accommodate reduced \textit{kanMX} expression during the gene deletion procedure, we developed a modified transformation protocol in which the drug-selection condition is milder than in the original (Experimental Procedures). Four ORFs were selected from loci with elevated H3K36me3 levels that were also unsuccessfully deleted in the \textit{Saccharomyces} Genome Deletion Project. Using the modified protocol, we successfully generated gene deletion strains for three of these four ORFs on our first attempt. To test whether these loci express \textit{kanMX} at particularly low levels due to H3K36me3 occupancy, we measured \textit{kanMX} expression in these strains relative to a strain from chromosome I that expresses \textit{kanMX} at average levels (with low variance across replicates). We found that all three of the newly constructed strains expressed \textit{kanMX} at significantly lower levels than average (Figure 2.5B, $p < 0.01$, one-sample t test). All three \textit{kanMX} transformants also had low expression relative to loci with elevated H3K36me3 that were nonetheless successfully targeted by
the *Saccharomyces* Genome Deletion Project (Figure 2.3A). Thus, H3K36me3-mediated suppression appears to explain, at least in part, why some ORFs failed in the systematic gene deletion process.

Chapter 2.4: Discussion

More than 20 years have passed since the classic position-effect experiment in yeast, in which genes were repositioned to the telomere to show that the epigenetic landscape dramatically alters gene expression\(^1\). Here, we have explored proof of principle that gene knockout libraries can be “repurposed” as a resource to study the effects of gene position not only at the telomere but systematically across an entire eukaryotic chromosome. Using the *Saccharomyces* Gene Deletion collection in this mode, we have identified an antagonistic interaction involving a chromatin mark (i.e., the H3K36me3 histone modification) and a genetic element (i.e., the Rap1 binding site on the *kanMX* gene cassette).

Chapter 2.4.1: H3K36me3 and Rap1-Activated Gene Expression: An Epigenetic-Genetic Interaction

Although the presence of H3K36me3 at the promoter does not generally repress gene expression, our results show that it is repressive in the context of a Rap1 binding site (Figure 2.2). In the same context, other marks such as H3K4me3 that correlate with expression in general are not correlated for a Rap1-driven gene. These differences point to specific chromatin effects on transcription that depend on the context of the gene being transcribed.
The idea that epigenetic-genetic interactions can differ according to the identity of the bound TF has previously been explored by Guccione et al.\textsuperscript{59}, and in one instance, a Rap1 promoter was shown to be particularly sensitive to Rpd3-mediated repression relative to promoters driven by other TFs.\textsuperscript{60} Genome-wide studies have also identified chromatin-TF interactions that operate on only a subset of genes.\textsuperscript{61} In support of distinct interactions within a cohort of genes bound by a common TF, Lickwar et al.\textsuperscript{62}, recently found that subsets of Rap1 bound genes with different binding motifs (one of which is a near-perfect match for the Rap1 motif in \textit{kanMX}) are associated with different functional outcomes such as gene activation and local nucleosome positioning. Here, we show that genes with an identical \textit{kanMX} motif bear a distinct interaction with chromatin modifications not found when considering larger sets of genes (Figures 2.4C, 2.4D, and Supplemental Figure 2.4) and that the Set2-Rpd3 pathway is likely involved in modulating \textit{kanMX} expression. Thus, it appears that H3K36me3- and Set2-Rpd3-mediated gene repression have a greater effect on \textit{kanMX} expression relative to their effect on average genes.

Chapter 2.4.2: Causality in Chromatin-Mediated Transcriptional Regulation

Recently, Henikoff and Shilatifard\textsuperscript{63} have countered the popular histone code hypothesis—whereby histone modifications play a causal role in regulating transcription—with the idea that the data equally support a “reverse model” in which DNA binding regulatory factors modulate the landscape of histone modifications. Under the reverse model, one would expect insertion of a genetic sequence such as \textit{kanMX} to induce concurrent changes in the chromatin landscape. In this study, we did not observe
such changes: levels of histone modification at the kanMX gene cassette remained largely unchanged in comparison to wild-type (Figure 2.2). Conversely, differences in levels of H3K36me3 at different gene positions were found to inversely correlate with kanMX expression. Moreover, disruption of the genes whose protein products catalyze and interact with H3K36me3 (SET2 and RPD3, respectively) results in increased levels of kanMX expression (Figure 2.3B). The clearest interpretation of these results is that H3K36me3 plays a causal role in regulating kanMX gene expression, not vice versa.

It remains to be seen whether such a causal link can be generalized to expression of other genes or whether it is specific to kanMX. Recent findings do suggest different models of chromatin regulation for different genes. A study of the GAL1/10 promoter exemplifies the argument that DNA sequence determines chromatin architecture\textsuperscript{64}. Studies of MYC1 TF binding in human B cells demonstrate how a TF can induce specific histone modifications at its binding target\textsuperscript{59,65}. A study in yeast separates TFs into two groups: those that are histone sensitive, and those that are histone insensitive\textsuperscript{66}. It is possible that the eukaryotic genome may take advantage of multiple modes of regulation, in which case, systematic position-effect screening may provide a suitable method for establishing directionality in the epigenetic-genetic relationship.

Chapter 2.4.3: Gene Knockout Libraries as General Resources for Epigenomics

The demonstration that yeast knockout libraries can be used for position-effect screening opens the door for future studies, which we foresee falling along several lines. First, the positioning of a single gene into many different chromosomal locations is a general feature of gene knockout collections for many eukaryotic organisms,
including *Schizosaccharomyces pombe*\(^{67}\), *Neurospora crassa*\(^{68}\), *Caenorhabditis elegans*\(^{69}\), *Arabidopsis thaliana*\(^{70}\), *Drosophila melanogaster*\(^{71,72}\), and *Mus musculus*\(^{73}\). These other species present modes of epigenetic regulation not present in *Saccharomyces*, such as DNA methylation and RNAi targeting, which the gene knockout collections may help elucidate. Second, the relative ease of genetic manipulation in *S. cerevisiae*, as well as the growing use of zinc finger and TALE nucleases for targeted genome editing in higher eukaryotes\(^{74,75}\), may allow the study of position effects involving other genes beyond that presented by the *kanMX* cassette. In yeast, employing well-established methods for exchanging the *kanMX* marker with an exogenous DNA sequence\(^{76,77}\), one can envision screening for epigenetic interactions with well-conserved candidate human disease genes and promoters\(^{78,79}\) positioned at loci encompassing a wide array of epigenetic states. The resulting strains could be assayed for any desired output, such as candidate gene expression or interaction with regulatory factors. Third, we foresee novel uses of the molecular barcodes included in each knockout strain, which may enable position-effect assays in pooled cultures for parallel analysis. Whereas each bar code represents a missing gene in a functional genomics assay\(^{28}\), each bar code in a position-effect assay represents a distinct position in the genome. For example, analysis of individual bar codes following ChIP in pooled cultures might characterize how a TF-DNA interaction influences or is influenced by epigenetic context. Thus, the wide availability of gene knockout libraries may, in turn, enable researchers to deploy a variety of position-effect analyses and to develop innovative position-effect techniques.
Chapter 2.5: Experimental Procedures

Chapter 2.5.1: Strains and Growth

All strains are from the *S. cerevisiae* heterozygous diploid gene deletion collection in the BY4743 background (Open Biosystems). Constant growth rates were observed for all heterozygous gene deletion strains regardless of chromosomal position (data not shown). Strains were grown to saturation in a 96-deep-well plate (Nunc) in rich medium overnight before transferring to a new 96-well plate on the day of measurement. Cultures were grown to mid-log phase ($1 \times 10^7$ cells/ml), and $\sim 2 \times 10^6$ cells was harvested in each sample. Confirmations of *kanMX* positioning via PCR were performed for a selection of strains, including those exhibiting the most extreme expression levels. To guard against genetic mutations that may have arisen to produce aberrant expression, we performed two tests to validate extreme-expressing strains. First, we sequenced each promoter to look for significant mutations. Second, each strain was remade, and expression in the remade strain was compared to the strain used in this study. One extremely low-expressing strain, *yal040cΔ*, failed both tests and was therefore excluded from our analysis. *rpd3Δ* and *set2Δ* strains were created using standard yeast transformation techniques with the *natMX* gene cassette and selected using rich media supplemented with 50 mg/l of clonNAT. Each deletion was confirmed via primers that flank the targeted *natMX* insertion site, and for *set2Δ*, was further confirmed by protein immunoblot using an antibody to H3K36me3. Each of these 2 strains was then crossed with the 10 strains in Figure 2.3A to produce 19 strains that are each homozygous for deletion of either *SET2* or *RPD3*, and heterozygous for gene deletion by *kanMX* (1 strain could not be recovered in a *set2Δ* background).
Chapter 2.5.2: Heterozygous Diploid Strain Analysis

mRNA-seq was performed in four different strains representing different kanMX positions along chromosome I. Total RNA was isolated as described by Wong et al.\textsuperscript{80}, yielding RNA with an integrity number of at least 7. mRNA was purified and fragmented to an average of 300 bp as described by Yoon and Brem\textsuperscript{81}. First-strand cDNA was then reverse transcribed using Superscript III, followed by RNase H digestion and second-strand synthesis. Libraries were then prepared for analysis on an Illumina HiSeq2000 sequencer. Data were filtered to eliminate clonal reads and aligned using Bowtie\textsuperscript{82} to the S. cerevisiae genome. Coordinates for each ORF were downloaded from Saccharomyces Genome Database, and counts for each ORF were computed as the median number of reads aligned to the last 200 bp of each ORF. Differential expression was determined using the “edgeR” package in R\textsuperscript{83}.

Chapter 2.5.3: RNA Quantification

To minimize batch effects, the mRNA for each replicate of the chromosome I data was isolated and reverse transcribed into cDNA in parallel in a 96-well plate format. Cultures were treated with zymolyase (Seikagaku) for 30 min at 30°C, and total RNA was isolated and reverse transcribed in 96-well format using a Cells-2-Ct kit (Ambion), with the following modifications. Lysis Buffer with DNase I was briefly warmed to 25°C immediately before use, and incubation time in lysis buffer was extended from 5 to 8 min. qRT-PCR was conducted with a Bio-Rad iCycler using Bio-Rad SYBR-Green I Supermix. Primers used to quantify expression are listed in Supplemental Table 2.1.
Primers Used for Quantitative PCR, Related to Experimental Procedures. qRT-PCR measurements were analyzed using the Pfaffl method\textsuperscript{84}, with \textit{kanMX} transcripts quantified relative to \textit{ACT1}. Results for each of six replicates were $\log_2$ transformed and plate and median normalized. Centromeric loci were defined as positions within 1 kb of the centromere. To produce Figure 2.1C, we constructed an empirical null model without position effects by sampling, with replacement, 1,000 data sets of equal size (90 sets of six measurements) from all measurements. Each data set was then ranked. The values displayed in Figure 2.1C summarize the distribution of expression observed at each rank across the null data set. Of the observed measurements, 79\% (71 out of 90) was more extreme than the 99\% confidence intervals defined in our null model without position effects. The contribution of position effects to expression variance was computed as $100 \times (1 - r)$, where $r$ is the ratio of the within-locus variance (six replicates per locus, variance averaged over 90 loci) to the total variance over all 540 measurements. We also assessed whether essentiality of the replaced gene or proximity to a native binding site for Rap1 may have influenced \textit{kanMX} expression. We found that neither correlated significantly with \textit{kanMX} expression (data not shown).

Chapter 2.5.4: Chromatin immunoprecipitation (ChIP)

Each immunoprecipitation (IP) was performed as previously described by Lee et al.\textsuperscript{35}, with the following modifications. For each replicate, 300 ml of yeast was prepared for cell lysis and sonication. Following formaldehyde treatment, crosslinking was quenched with addition of glycine to a final concentration of 400 mM. Cell lysate was collected into a 14 ml tube and sonicated using a Misonix 3000 (power 8, six cycles,
30 s per cycle) to obtain fragments in the range 300–600 bp. Whole-cell extract was collected for multiple IPs using different antibodies on aliquots of the same lysate. The antibodies used were specific to endogenous H3K4me3 (ab8580; Abcam), H3K36me3 (ab9050; Abcam), H3K79me3 (ab2621; Abcam), H3K9ac (39137; Active Motif), H3K14ac (ab52946; Abcam), histone H3 (ab1791; Abcam), or Rap1 (y-300; Santa Cruz Biotechnology).

Chapter 2.5.5: Quantitative ChIP Scoring

At each gene knockout position, qRT-PCR primers were designed to amplify five loci representing one position upstream of the kanMX insertion site, one position with primers flanking the insertion site, and three positions in the kanMX promoter, the 5′ region, and the gene body, respectively (Supplemental Figure 2.2A; Supplemental Table 2.3). To compare these measurements with wild-type, three additional primer pairs were designed for use on wild-type ChIP extracts, to measure the corresponding histone modification enrichments on the native gene sequence for each of ten gene loci. The wild-type primers were positioned at the insertion site, in the 5′ region of the gene, and in the body of the gene. The quality of each IP was assessed by evaluating the enrichment of a DNA sequence known to be bound to each protein (positive control) relative to mitochondrial DNA (negative control) (Supplemental Table 2.4). Immunoprecipitated DNA samples with at least 20-fold enrichment were selected for further analysis. In these samples, enrichments for the kanMX sequence were quantified relative to whole-cell extract using positive control primers as a reference and expressed as a log2 ratio of enrichment. Normalization using a positive control accounts for experimental differences
between replicates (i.e., how well nonspecifically bound DNA was washed away), as well as differences in protein abundance that may arise in different strains. Enrichments for each antibody were normalized separately. Each set of replicate measurements was quantile normalized before subtracting histone H3 enrichments.

Chapter 2.5.6: Correlation with Histone Modifications

To calculate a value representing histone modification levels at the *kanMX* promoter, we averaged previously published histone measurements within a 500 bp window centered at the transcription start site (TSS). A 500 bp window size recapitulated known genome-wide correlations with native gene expression most faithfully. Pokholok et al.\(^7\), examined chromatin sheared randomly by sonication and, thus, introduce the possibility that measurements may include modifications from other regions on random, long DNA fragments. To guard against such noise, we employed an approach that examined histone modifications within a 500 bp sliding window centered at positions from 2 kb upstream of the TSS to 2 kb downstream (Supplemental Figures 2.2D–2.2F). We then searched for peaks of correlation between expression and histone modification that localized over the promoter and TSS.

Chapter 2.5.7: Yeast Transformation

Transformation to produce yeast strains was performed as described by the *Saccharomyces* Genome Deletion Consortium (a derivative of the method developed in \(^21\) with the following modifications. First, we added a short incubation (5–15 min) in 5 mM calcium chloride following heat shock\(^85\). Next, whereas the cited protocol calls for
strong selection with G418 (300 µg/ml) after 3 hr of recovery post heat shock, we plated transformants directly onto rich medium post heat shock and allowed for recovery overnight at 30°C in order to allow transformants to generate sufficient kanMX gene product to promote G418 resistance. Transformants were then exposed to a graded selection procedure, in which cells were first replica plated onto rich media bearing 50 µg/ml of G418 antibiotic, followed by replica plating 2 days later onto rich media bearing 200 µg/ml of G418 and growth at 30°C for 2 more days. Colonies larger than 1 mm in diameter were assayed for correct integration of the kanMX cassette via PCR.

Chapter 2.6: Acknowledgements

We thank L. Clark and F. Zhang for technical assistance, R. Srivas for critical reading of the manuscript, R. Deconde for helpful discussions regarding statistics, A. Shah for laboratory assistance, A. Chu for detailed information about the Saccharomyces Genome Deletion Project, and J. Boeke for helpful comments. This research was supported by grants R21HG005252, R01GM084279, and P50GM085764 from the National Institutes of Health.

Chapter 2, in full, is a reprint of the material as it appears in: “Chen M, Licon K, Otsuka R, Pillus L, Ideker T. Decoupling epigenetic and genetic effects through systematic analysis of gene position. Cell Rep. Jan 2 (2013)”. The dissertation author was the primary investigator and author of this paper.

Chapter 2.7: Accession Numbers

All mRNA-seq data are deposited in GEO under accession number GSE42554.
Figure 2.1: kanMX Expression on Chromosome I.

(A) A systematic position-effect screen in yeast. An antibiotic-resistance gene, \( kan^R \), driven by a Rap1-activated promoter, \( pTEF \), together referred to as \( kanMX \), is inserted at each locus on chromosome I (Chr I) to produce a library of yeast strains (colored cells). Expression of the reporter gene is measured by qRT-PCR. (B) Expression of \( kanMX \) as a function of position on chromosome I. Each point represents the mean of six independent biological replicates; error bars indicate 95% confidence intervals. Positions that are significantly over- or underexpressed (\( p < 0.05 \), one-sample t test) are colored orange. Subtelomeric regions are shaded gray; telomeres are shaded black. The centromere is marked by a dotted vertical line. (C) Position effects account for increased variability in gene expression. To visualize the contribution of position effects to expression variation, we sorted the observed \( kanMX \) expression at each position (black dots) for comparison with an empirical null model without position effects constructed using random samples from all collected data (gray squares; Experimental Procedures). Each gray data point represents the mean expression at that rank across all sampled data sets; error bars represent 95% confidence intervals. See also Supplemental Figure 2.1.
Figure 2.2: Insertion of kanMX Does Not Perturb Chromatin but Chromatin Does Perturb kanMX Expression.

Comparisons of histone modification enrichment between wild-type and knockout strains show high correlation across all ten strains at a given primer position (A–E) and across gene promoter and body positions for any individual knockout strain (F–J). (A–E) Scatterplots comparing ChIP-qRT-PCR enrichment of histone modifications at the kanMX promoter (y axes) and the 5′ region of the corresponding native gene (x axes). Measurements were made using antibodies to H3K4me3 (A), H3K36me3 (B), H3K79me3 (C), H3K9ac (D), and H3K14ac (E). All measurements were normalized by H3 content. (F–J) Traces depicting histone modification enrichment at the yal003wΔ:kanMX cassette (summarized by dotted line; individual data points denoted with “X”) and at the YAL003W native gene (solid line; triangles). Histone modifications are ordered as in (A)–(E). The x axis indicates position relative to start codon. See also Supplemental Figures 2.2A–2.2C. (K and L) Correlation significance of H3K36me3 (K) and H3K4me3 (L) as a function of position around the transcription start site (TSS) (from 2 kb upstream to 2 kb downstream; see also Supplemental Figures 2.2D–2.2F). Black lines represent correlation with wild-type expression; orange lines represent correlation with kanMX expression. Positive correlations rise above the x axis; negative correlations fall below. (M) Summary of correlations at the TSS for seven histone modifications. Colors as in (K) and (L). Asterisks denote significant correlations (p < 0.05).
Figure 2.3: *kanMX* Is Repressed by H3K36me3 via the Set2-Rpd3 Pathway.

(A) H3K36me3 levels predict *kanMX* expression on other chromosomes. Ten loci on yeast chromosomes II–XVI were predicted to have either high (red) or low (blue) *kanMX* expression based on wild-type H3K36me3 enrichment. Expression was measured relative to a strain from chromosome I that expresses *kanMX* at average levels with low variance across replicates. Each bar represents the mean of three independent biological replicates, and error bars represent SEMs. (B) *kanMX* expression increases upon deletion of RPD3 or SET2. To probe the mechanism for the effects observed relative to H3K36me3 in (A), *rpd3Δ* and *set2Δ* strains were constructed using the natMX marker and crossed with strains in (A). The *ybr191wΔ* strain could not be recovered in the *set2Δ* background. mRNA expression of *kanMX* was then assayed by qRT-PCR. All ten assayed strains in an *rpd3Δ* background and eight out of nine strains in *set2Δ* background showed higher mean *kanMX* expression compared to wild type (**p < 0.003 and *p < 0.05, respectively, paired t test). Each point represents the mean of three independent biological replicates. See also Supplemental Figure 2.3.
Figure 2.4: H3K36me3 Is Inversely Proportional to Rap1 Binding at the kanMX Promoter.

(A) ChIP-qRT-PCR measurements of Rap1 occupancy are plotted against measurements of H3K36me3 at the kanMX locus in five knockout strains (two replicate growths per strain). Error bars represent SEs. (B) ChIP-seq counts (log2 transformed) of Rap1 bound nucleosomal sequences56 plotted against H3K36me3 in wild-type cells86. Many loci with low H3K36me3 levels are enriched for Rap1 binding (red circle). (C) Among yeast genes that possess a kanMX binding motif, expressed genes (red, n = 31) are associated with low H3K36me3 enrichment relative to silenced genes (blue, n = 13, p < 3 × 10^{-4}). (D) Among the same set of genes as (C), expressed genes exhibit no significant differences in H3K4me3 compared with silenced genes (p = 0.89). See also Supplemental Figure 2.4.
Figure 2.5: H3K36me3 Correlates with Unsuccessful Yeast Knockouts.

(A) ORFs resistant to deletion in the *Saccharomyces* Genome Deletion Project\(^{28}\) are significantly enriched for H3K36me3 relative to successfully deleted ORFs (Mann-Whitney U test, \(^* p < 2.2 \times 10^{-16}\)). Among dubious ORFs, which were uniformly attempted once and only once, H3K36me3 remains enriched for ORFs that did not produce successful transformants (\(^** p < 5.1 \times 10^{-6}\)). (B) Of the ORFs that were resistant to deletion, three were selected for elevated levels of wild-type H3K36me3 at the ORF promoter and, subsequently, knocked out using a modified selection strategy (Experimental Procedures). Expression from all three loci is significantly depressed (\(p < 0.005\), one-sample t test) relative to the chromosome I strain used as a reference in Figure 2.3A. Each bar represents the mean of ten independent biological replicates, and error bars represent SEMs.
Supplemental Figure 2.1: RNA-Seq in Four Heterozygous Diploid Strains and Statistics of \textit{kanMX} Expression on Chromosome I, Related to Figure 2.1.

To evaluate the extent to which heterozygous gene deletion affects mRNA expression patterns, mRNA-Seq was performed on four heterozygous diploid knockouts strains from chromosome I ($SEO1^{+/−}$, $GIP4^{+/−}$, $FUN26^{+/−}$, and $FUN14^{+/−}$) to quantify their similarity in mRNA gene expression to wild-type cells (Experimental Procedures). Three independent biological replicates were assayed for each knockout strain, and two sets of wild-type yeast were assayed in parallel. (A–E) Scatter plot of whole-genome mRNA counts comparing (A) wild-type and wild-type, (B) wild-type and $SEO1^{+/−}$, (C) wild-type and $GIP4^{+/−}$, (D) wild-type and $FUN26^{+/−}$, and (E) wild-type and $FUN14^{+/−}$. (F–J) Same as (A–E), but showing only data from chromosome I.
Supplemental Figure 2.1: RNA-Seq in Four Heterozygous Diploid Strains and Statistics of kanMX Expression on Chromosome I, Related to Figure 2.1. (continued)

(K) Correlation in expression levels between wild-type and each of the heterozygous diploids was not significantly different than the correlation between two independent sets of wild-type measurements. Shown are Pearson correlation coefficients for the comparisons in (A–J). Error bars denote 95% confidence intervals. (L) Very few genes are differentially expressed genes in the four heterozygous diploid knockouts measured in this study. For comparison, six data sets were downloaded from GEO that comprise all available mRNA-seq haploid gene knockout expression data (accession numbers GSE27934, GSE30703, GSE31300, GSE33497, accessed July 30, 2012). In all cases, the haploid data sets identified greater numbers of differentially-expressed genes than the heterozygous diploid strains tested. These results suggest that the underlying biology within heterozygous diploid strains is highly similar to the biology in wild-type S. cerevisiae. Differential expression was called uniformly using a Bonferroni-corrected p-value < 0.01. (M) kanMX expression on chromosome I grouped by chromosomal region. Subtelomeric expression is significantly lower than expression elsewhere (*, Mann-Whitney p < 0.01). Pericentric loci are not repressed (p = 0.85). (N) Observed kanMX expression (y axis) plotted against RNA-sequencing measurements of expression of wild-type genes at the same position87 (x axis). The box plot (left) summarizes kanMX expression from loci that are not expressed in wild-type (inactive loci). No detectable trend was observed among active loci (Pearson’s r = 0.07, p = 0.57), or between active and inactive loci (Mann-Whitney p = 0.48).
Supplemental Figure 2.2: Assessing Chromatin Changes due to *kanMX* and Method for Correlating Expression with Histone Modifications, Related to Figure 2.2.

ChIP-qPCR was performed in 10 chromosome I knockout strains and wild-type using antibodies to the histone modifications shown in Figure 2.2. (A) In each knockout strain, five measurements were taken using primers positioned at approximately −500, 0, +300, +800, and +1300 bp relative to the *kanMX* insertion site. In wild-type controls, measurements were taken at four positions at −500, 0, +300, and +800 bp relative to the *kanMX* insertion position. (B) Traces of H3K4me3 are shown for each of the ten knockout strains tested. Knockout enrichments are summarized by a dotted line, with individual data points denoted with an “X.” Corresponding wild-type enrichments are summarized with a solid line, with individual data points denoted with a triangle. (C) Same as (B), but for H3K36me3.
Supplemental Figure 2.2: Assessing Chromatin Changes due to \textit{kanMX} and Method for Correlating Expression with Histone Modifications, Related to Figure 2.2. (continued)

(D) Correlation is computed between a vector of gene expression values and a vector of histone modification values for all \( n \) genes on Chromosome I. The histone modification value for each gene is averaged over a 500-bp window around the transcription start site (TSS). (E) The 500-bp window is slid across a 4-kb region surrounding the TSS, resulting in a different histone modification vector at each window position. (F) This process yields the correlation between gene expression and histone modification as a function of window position relative to the TSS.
Supplemental Figure 2.3: Histone Acetylation Affects kanMX Expression at H3K36me3-Regulated Loci, Related to Figure 2.3.

(A–E) Comparisons of histone lysine acetylation at loci identified as expressing kanMX at significantly higher or lower levels (see Figures 3.1 and 3.3) reveal a positive association with expression for specific residues. Measurements of (A) H2AK7ac, (B) H2BK11ac, (C) H2BK16ac, (D) H4K12ac, (E) H4K16ac (from 53) were normalized by nucleosome occupancy54 and represented on the y-axes as log2 ratios of the indicated enrichments. Each of these sites except H4K16 are known to be targeted by Rpd36. Asterisks denote significantly different groups (p < 0.05). (F–J) Among loci with significantly different kanMX expression (orange dots), histone lysine acetylation (y-axes) at the residues from (A–E) is negatively correlated with H3K36me37, whereas among all loci (black dots), there is no such association. Each solid orange line and dotted black line denotes a linear model fitted to H3K36me3-regulated loci and all loci, respectively. Model coefficients and significance are listed below each plot. All measurements are represented as log2 ratios of the indicated enrichments.
Supplemental Figure 2.4: H3K36me3 and H3K4me3 Averaged over Heterogeneous Groups of Genes Reflect Genome-wide Averages, Related to Figure 2.4.

(A and B) Distribution of H3K36me3 (A) and H3K4me3 (B) within yeast genes that possess a consensus Rap1 binding motif (n = 134 and 130 in the blue and red groups respectively). Labels for all panels as in Figures 2.4C and 2.4D. (C and D) Distribution of H3K36me3 (C) and H3K4me3 (D) for all yeast genes.
Table 2.1: Primers used for quantification of kanMX mRNA.

<table>
<thead>
<tr>
<th>Name</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT1-expression</td>
<td>CCTACGTTGGTGATGAAGCT</td>
<td>GTCAGTCAAATCTCTACCGG</td>
</tr>
<tr>
<td>kanMX-expression</td>
<td>GCGAGTGATTTTGATGACGA</td>
<td>ACTGAATCCCGGTGAGAATGG</td>
</tr>
</tbody>
</table>

Table 2.2: Primers used in ChIP-qPCR to quantify histone modifications at wild-type loci.

Histone modification enrichments were measured at ten loci, at four positions along the length of each gene, positioned approximately −500 bp, +0 bp, +400 bp, and +900 bp relative to the TSS. Only three points were sampled for one wild-type gene, YAL046C, which was too short to be sampled at the fourth position (+900 bp).

<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Position</th>
<th>Direction</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPr</td>
<td>YAL067C</td>
<td>-500 bp</td>
<td>Forward</td>
<td>TGCGTCTGCTTTATGTCTTT</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL067C</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>GCGTCCACCCCTTTAAAACAT</td>
</tr>
<tr>
<td>B</td>
<td>YAL067C</td>
<td>+0 bp</td>
<td>Forward</td>
<td>GATGAATGGACATTCGCTACA</td>
</tr>
<tr>
<td>B</td>
<td>YAL067C</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>CCTGCCGCTTTACTGGAATA</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL067C</td>
<td>+500 bp</td>
<td>Forward</td>
<td>TTACGTTTCGGGAATGAAGG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL067C</td>
<td>+500 bp</td>
<td>Reverse</td>
<td>GCAGCAACACCGGTAAAAAG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL067C</td>
<td>+1000 bp</td>
<td>Forward</td>
<td>AATCTTTTCCTCGCCAACCT</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL067C</td>
<td>+1000 bp</td>
<td>Reverse</td>
<td>CTGTTACGCTTCGCAAGG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL046C</td>
<td>-500 bp</td>
<td>Forward</td>
<td>TGGAACAGGTCGAGGGTCTA</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL046C</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>CCCACAGGAAGTATGAATCCAA</td>
</tr>
<tr>
<td>B</td>
<td>YAL046C</td>
<td>+0 bp</td>
<td>Forward</td>
<td>TCCACTTTTCGTCGCTA</td>
</tr>
<tr>
<td>B</td>
<td>YAL046C</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>CCGGACATCGCTTCAGATA</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL046C</td>
<td>+200 bp</td>
<td>Forward</td>
<td>CAATGACTCCGGAGGAAAG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL046C</td>
<td>+200 bp</td>
<td>Reverse</td>
<td>CTGTGTCCTGTTATGAGAC</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL038W</td>
<td>-500 bp</td>
<td>Forward</td>
<td>AAACTGGGACACTTGAGCCAC</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL038W</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>ACGGTCGACAGCCACACAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL038W</td>
<td>+0 bp</td>
<td>Forward</td>
<td>CAGACCTGGAGATCCAAAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL038W</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>TCAATGACACACACACAG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL038W</td>
<td>+300 bp</td>
<td>Forward</td>
<td>TCAAGTGACACACACACAG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL038W</td>
<td>+300 bp</td>
<td>Reverse</td>
<td>CGTACTTTGTCATCGTGAG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL038W</td>
<td>+800 bp</td>
<td>Forward</td>
<td>GGTGTTATGGTGCCAGAGG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL038W</td>
<td>+800 bp</td>
<td>Reverse</td>
<td>CTGGAACCTTGGAGAG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL031C</td>
<td>-500 bp</td>
<td>Forward</td>
<td>ACCACATGGGTTTTCTGACC</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL031C</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>TGATGATCCGGAGAATAAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL031C</td>
<td>+0 bp</td>
<td>Forward</td>
<td>CAATGTTGAACAGACAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL031C</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>TGATGATCCGGAATAGAAGC</td>
</tr>
</tbody>
</table>
Table 2.2: Primers used in ChIP-qPCR to quantify histone modifications at wild-type loci. (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Position</th>
<th>Direction</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N5P</td>
<td>YAL031C</td>
<td>+300 bp</td>
<td>Forward</td>
<td>CCCGTATTGGCTAAGAGGT</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL031C</td>
<td>+300 bp</td>
<td>Reverse</td>
<td>GGTACCTCCTCGGGAAACAT</td>
</tr>
<tr>
<td>NBODY</td>
<td>YAL031C</td>
<td>+800 bp</td>
<td>Forward</td>
<td>CAAATGGCAAGTGATGGAGAA</td>
</tr>
<tr>
<td>NBODY</td>
<td>YAL031C</td>
<td>+800 bp</td>
<td>Reverse</td>
<td>CCCGCAAGTGGGAAATGAT</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL022C</td>
<td>-500 bp</td>
<td>Forward</td>
<td>GTGGGCTATTGGGTGAAGTG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL022C</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>GGTACCTCCTCGGGAAACAT</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL022C</td>
<td>+800 bp</td>
<td>Forward</td>
<td>CAATGGCAAGTGATGGAGAA</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL022C</td>
<td>+800 bp</td>
<td>Reverse</td>
<td>CCCGCAAGTGGGAAATGAT</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL008W</td>
<td>-500 bp</td>
<td>Forward</td>
<td>GTGGGCTATTGGGTGAAGTG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL008W</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>TAAATGCTCTGGGAAATGAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL008W</td>
<td>+0 bp</td>
<td>Forward</td>
<td>CATAACGCTAGAGGGCAAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL008W</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>CCATAAGGGGACGTTGAT</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL008W</td>
<td>+250 bp</td>
<td>Forward</td>
<td>ATGAATCGCCAGCCTTCTAA</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL008W</td>
<td>+250 bp</td>
<td>Reverse</td>
<td>TGTTCCTCCTTCCGGTGA</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL008W</td>
<td>+500 bp</td>
<td>Forward</td>
<td>GGTACCGTACAAGGGGATG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL008W</td>
<td>+500 bp</td>
<td>Reverse</td>
<td>CCCAATAACCCCATCAATAAG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL003W</td>
<td>-500 bp</td>
<td>Forward</td>
<td>CCAATGACGTCTTGAGAAG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAL003W</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>GAAGAGAAGCAGCAGGAAAG</td>
</tr>
<tr>
<td>B</td>
<td>YAL003W</td>
<td>+0 bp</td>
<td>Forward</td>
<td>CAAACCGACGCTTGAAAGAA</td>
</tr>
<tr>
<td>B</td>
<td>YAL003W</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>CTTGGAGAAATCGGTGATG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL003W</td>
<td>+300 bp</td>
<td>Forward</td>
<td>GTTTGACTATCAGTACGAG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL003W</td>
<td>+300 bp</td>
<td>Reverse</td>
<td>GACGTCAGCTTGAGAAACAG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL003W</td>
<td>+600 bp</td>
<td>Forward</td>
<td>TCAAGAGCTTCTCCTGCT</td>
</tr>
<tr>
<td>NBody</td>
<td>YAL003W</td>
<td>+600 bp</td>
<td>Reverse</td>
<td>TCGGAACCCGATAAATCGAC</td>
</tr>
<tr>
<td>NPr</td>
<td>YAR002W</td>
<td>-500 bp</td>
<td>Forward</td>
<td>TTAAGCACTAGGGCGGAG</td>
</tr>
<tr>
<td>NPr</td>
<td>YAR002W</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>GATTTCTGGCCAATTCTTA</td>
</tr>
<tr>
<td>B</td>
<td>YAR002W</td>
<td>+0 bp</td>
<td>Forward</td>
<td>CATCAAATAAGCAACAGGCAAG</td>
</tr>
<tr>
<td>B</td>
<td>YAR002W</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>TTTGGACTCAGTGCTGAA</td>
</tr>
<tr>
<td>N5P</td>
<td>YAR002W</td>
<td>+300 bp</td>
<td>Forward</td>
<td>GTTTTCTGACGTGGGAG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAR002W</td>
<td>+300 bp</td>
<td>Reverse</td>
<td>AGTTTGCAAGCTCTCATTG</td>
</tr>
<tr>
<td>NBody</td>
<td>YAR002W</td>
<td>+800 bp</td>
<td>Forward</td>
<td>CTAACGCTCCTCAGAAT</td>
</tr>
<tr>
<td>NBody</td>
<td>YAR002W</td>
<td>+800 bp</td>
<td>Reverse</td>
<td>CAAATGCGACGTGATATT</td>
</tr>
<tr>
<td>NPr</td>
<td>YAR019C</td>
<td>-500 bp</td>
<td>Forward</td>
<td>GAAAGGGGCTTATCGTGCTT</td>
</tr>
<tr>
<td>NPr</td>
<td>YAR019C</td>
<td>-500 bp</td>
<td>Reverse</td>
<td>CGACCACAGAAAACGGAAGC</td>
</tr>
</tbody>
</table>


Table 2.2: Primers used in ChIP-qPCR to quantify histone modifications at wild-type loci. (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Position</th>
<th>Direction</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>YAR019C</td>
<td>+0 bp</td>
<td>Forward</td>
<td>TCTACTTGGAGGCTCGAGGA</td>
</tr>
<tr>
<td>B</td>
<td>YAR019C</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>CCCAATGACCTGCTTCAAGT</td>
</tr>
<tr>
<td>N5P</td>
<td>YAR019C</td>
<td>+400 bp</td>
<td>Forward</td>
<td>GCGCTAATGGTTCTTTGAGG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAR019C</td>
<td>+400 bp</td>
<td>Reverse</td>
<td>GGATGACTCCTTCACCGTG</td>
</tr>
<tr>
<td>NB</td>
<td>YAR019C</td>
<td>+900 bp</td>
<td>Forward</td>
<td>GAGGACTTTACCGAGCTGA</td>
</tr>
<tr>
<td>NB</td>
<td>YAR019C</td>
<td>+900 bp</td>
<td>Reverse</td>
<td>GCAAGTCCGTCAGGCTTAA</td>
</tr>
<tr>
<td>NP</td>
<td>YAL016W</td>
<td>+500 bp</td>
<td>Forward</td>
<td>AGAGACCCACCATTGACAGC</td>
</tr>
<tr>
<td>NP</td>
<td>YAL016W</td>
<td>+500 bp</td>
<td>Reverse</td>
<td>CGACATAACAAAGCTTGGTC</td>
</tr>
<tr>
<td>B</td>
<td>YAL016W</td>
<td>+0 bp</td>
<td>Forward</td>
<td>GTGGTCAGAAAAGGAGCGAA</td>
</tr>
<tr>
<td>B</td>
<td>YAL016W</td>
<td>+0 bp</td>
<td>Reverse</td>
<td>ATCCCTAAACAGAGCCAAGG</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL016W</td>
<td>+400 bp</td>
<td>Forward</td>
<td>GATGAGGTGTTTGGCCGTTT</td>
</tr>
<tr>
<td>N5P</td>
<td>YAL016W</td>
<td>+400 bp</td>
<td>Reverse</td>
<td>GGGCCACGTTATCCAGGAA</td>
</tr>
<tr>
<td>NB</td>
<td>YAL016W</td>
<td>+900 bp</td>
<td>Forward</td>
<td>GCCTTGGAGGGTACGTTACAT</td>
</tr>
<tr>
<td>NB</td>
<td>YAL016W</td>
<td>+900 bp</td>
<td>Reverse</td>
<td>GAACTTGGCAAACCGAGAAA</td>
</tr>
</tbody>
</table>

Table 2.3: Primers used to quantify enrichment of histone modifications along the kanMX gene.

Alternate “B-ko” primers were used in some cases where multiple qPCR products were obtained when paired with the wild-type “B” primer.

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPr</td>
<td>+300 bp</td>
<td>GACATGGAGGCCCAGAATAC</td>
<td>GAGCCGTAAATTTTTGCTTCG</td>
</tr>
<tr>
<td>K5P</td>
<td>+800 bp</td>
<td>TTATGCCTCTTCCGACCATC</td>
<td>AACAGGAATCGAATGCAACC</td>
</tr>
<tr>
<td>KBody</td>
<td>+1300 bp</td>
<td>TGGCTCCTGGTAGTTTCTC</td>
<td>CGAGCATCAAATGAAACTGC</td>
</tr>
<tr>
<td>B_ko</td>
<td>+0 bp</td>
<td>&lt;strain-specific, see Table 2.2&gt;</td>
<td>GGACGAGGCAACGCTAAACAG</td>
</tr>
<tr>
<td>B_ko_1</td>
<td>+0 bp</td>
<td>&lt;used if double products are formed with other B_ko primers&gt;</td>
<td>CGCCTAATTAACCGGGA</td>
</tr>
<tr>
<td>B_ko_2</td>
<td>+0 bp</td>
<td></td>
<td>AGGGTATTCCTGGGCTCCTCAT</td>
</tr>
</tbody>
</table>
Table 2.4: Primers used as positive and negative controls for all ChIP-qPCR measurements.

Enrichments were computed against positive and negative controls, which are listed in this section. Positive control primers were chosen as follows depending on the immunoprecipitated protein. For Rap1, primers target the RPL2B promoter, a known Rap1 binding site\(^{32}\). For H3K4me3, H3K9ac, and H3K14ac, primers target the 5’ region of the actively-transcribed RPL30 gene. For H3K36me3, primers target the coding region of PMA1, a locus enriched for H3K36me3\(^{51}\). For H3K79me3, primers target the promoter region of FUN12, a region of relatively high H3K79me3\(^{7}\).

<table>
<thead>
<tr>
<th>Control</th>
<th>Antibody</th>
<th>Name</th>
<th>Direction</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>anti-H3</td>
<td>YFR057W</td>
<td>Forward</td>
<td>ATGCCGAGCTAGGATGAGAA</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3</td>
<td>YFR057W</td>
<td>Reverse</td>
<td>CTGCAATTGAGGCAGGATTT</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K4me3</td>
<td>RPL30_5p</td>
<td>Forward</td>
<td>GGATTAGCAAGAAGGCTTTGG</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K4me3</td>
<td>RPL30_5p</td>
<td>Reverse</td>
<td>TAACTGGGCGCTGTTGAAT</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K36me3</td>
<td>PMA1_mid</td>
<td>Forward</td>
<td>TGGCCTCTGCTATCTGCT</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K36me3</td>
<td>PMA1_mid</td>
<td>Reverse</td>
<td>CATTAAACGGGCTTTGGGA</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K79me3</td>
<td>FUN12_5p</td>
<td>Forward</td>
<td>GTGCTAAGAAGCCAGCAAG</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K79me3</td>
<td>FUN12_5p</td>
<td>Reverse</td>
<td>GCAGAATTTTCTTCTCTT</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K9ac</td>
<td>RPL30_5p</td>
<td>Forward</td>
<td>GGATTAGCAAGAAGGCTTTGG</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K9ac</td>
<td>RPL30_5p</td>
<td>Reverse</td>
<td>TAACTGGGCGCTGTTGAAT</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K14ac</td>
<td>RPL30_5p</td>
<td>Forward</td>
<td>GGATTAGCAAGAAGGCTTTGG</td>
</tr>
<tr>
<td>positive</td>
<td>anti-H3K14ac</td>
<td>RPL30_5p</td>
<td>Reverse</td>
<td>TAACTGGGCGCTGTTGAAT</td>
</tr>
<tr>
<td>negative</td>
<td>all</td>
<td>mito</td>
<td>Forward</td>
<td>GCAGCAAGTAAAGACTGTC</td>
</tr>
<tr>
<td>negative</td>
<td>all</td>
<td>mito</td>
<td>Reverse</td>
<td>ACAGGAGGCTCAATGTCAAC</td>
</tr>
</tbody>
</table>
Chapter 3: Systematic gene replacement imparts experimental control for genome-wide analysis of histone-promoter interactions

Chapter 3.1: Abstract

Transcription involves the precise coordination of genetic signals – specific sequences encoded in DNA – with epigenetic signals such as chemical modification of histones. Genome-scale techniques for profiling these signals have provided a wealth of data defining how gene sequence, expression and histone modification all vary throughout a genome. Current techniques, however, collect data for one signal without control over the others, and thus confound the interpretation of specific interactions between various factors. To provide experimental control over gene sequence and expression, we used the *S. cerevisiae* gene deletion collection, in which each strain possesses an identical gene, *kanMX*, in a different genomic position and chromatin environment, to measure interactions between histone modification and wild-type gene promoters. Here we present measurements of histone modification at the interface between a wild-type promoter and each *kanMX* insertion in over 4,800 strains using a new technique we have developed called Barcode immunoprecipitation and analysis by high-throughput sequencing (BIP-seq). Because each *kanMX* sequence is identical and expression is very similar, difference in histone modification due to gene sequence and expression are mitigated, whereas signals due to promoter interactions remain. Overall, histone states at *kanMX* were consistent with wild-type levels, suggesting that activities downstream of transcription start sites play a more limited role in determining histone state than upstream activities in gene promoters. Remarkably, histone acetylation levels measured at *kanMX* were more predictive of wild-type levels of gene expression than
wild-type measurements of histone acetylation. In contrast, histone methylations at
\textit{kanMX} were less predictive of expression, suggesting that gene promoters are primarily
regulated by histone acetylation and not methylations. Finally, we relate histone
modifications to protein-DNA binding maps and find a novel exclusive interaction
between the high mobility group protein Nhp6a and the histone variant H2A.Z.
Subsequent analyses suggest that H2A.Z inhibits binding and gene activation by Nhp6a
and not \textit{vice versa}.

Chapter 3.2: Introduction

Histone modifications play a key role in a variety of processes, including
transcription, replication, and other chromatin transactions\textsuperscript{2}. Correlations between
genome-wide maps of histone modifications and expression have revealed general trends
that describe how various factors co-vary across a genome, and have given rise to
descriptors for some histone modifications as 'activating', 'silencing', or 'poised'\textsuperscript{88}. It
remains difficult, however, to extract specific interactions from these data, because
typical genome-wide data for one factor are collected without control over others. Indeed,
a wealth of data exist to demonstrate the co-occurrence of histone modifications and
expression states\textsuperscript{7,41,53,86,89}, but the resulting genome-wide outputs are necessarily the sum
total of many effects that overlap and thus may obscure interesting biology, such as
regulatory interactions between transcription factor binding and histone modification
states.

We have previously used the \textit{S. cerevisiae} gene deletion library as a resource to
control for gene sequence\textsuperscript{90}, taking advantage of the fact that each gene deletion is a gene
replacement wherein the *kanMX* gene cassette replaces a native gene in a different genomic position and epigenetic context (Figure 3.1A)\(^\text{22}\). In prior work, we observed a novel interaction with a histone mark that influenced gene expression in a modest but significant way\(^\text{90}\). Importantly, we showed that for any given gene replacement, the expression level of *kanMX* was in general more similar to *kanMX* expression at other loci than to the expression level of the gene that had been replaced (Figure 3.1B). We reasoned, then, that the yeast knockout library might be adapted to address a lack of control inherent in analyses involving genome-wide mapping of histone modifications.

The replacement of a native gene by *kanMX* in each strain represents a common, controlled genetic perturbation that homogenizes gene sequence as well as the level of gene expression at each gene locus. Whereas the histone state at each genomic locus may involve interactions with promoter sequence and activity, as well as gene sequence and level of gene expression, measurements of histone state at the *kanMX* cassette in different yeast knockout strains may bring some previously uncontrolled variables under control to reveal specific interactions between histone modifications and gene promoter sequences (Figure 3.1C).

To this end, we have developed barcode immunoprecipitation with analysis by high-throughput sequencing, or BIP-seq. BIP-seq quantifies levels of histone modification using short molecular barcodes that reside at each interface between a native gene promoter and a *kanMX* gene cassette (Figure 3.1D). The presence of *kanMX* controls for gene sequence and expression, in contrast to genome-wide mapping of histone modifications in which these variables are uncontrolled. Each molecular barcode is unique within the strain library, and permits simultaneous analyses of thousands of
strains within one pooled culture and one immunoprecipitation procedure. Measurements for different strains are resolved using the barcode sequence to unambiguously determine the genomic position of each measurement.

In this report, we use BIP-seq to probe histone modifications at 4,822 genomic loci where *kanMX* has been inserted directly downstream of a native promoter. We show that homogenization of gene sequence and expression has limited influence on the global distributions of the chromatin marks examined. We show that BIP-seq measurements of histone acetylation are more predictive of wild-type gene expression than ChIP-seq measurements, whereas BIP-seq measurements of methylation are slightly less predictive than ChIP-seq. Finally, we relate chromatin marks to protein-DNA binding maps to identify interactions between transcription factors and histone modifications, and identify a new exclusive interaction between the structural protein Nhp6a and the histone variant H2A.Z.

Chapter 3.3: Materials and Methods

The BIP-seq assay marries a conventional ChIP-seq procedure\textsuperscript{91} with barcode sequencing analysis\textsuperscript{27} to quantify enrichment of histone modification at the interface between a native gene promoter and a *kanMX* cassette. The key differences between BIP-seq and ChIP-seq are its application to a large pool of many yeast strains, and the resolution of each measurement by high-throughput barcode sequencing\textsuperscript{27}. Pooling is enabled by unique molecular barcodes that reside at each native promoter-*kanMX* cassette interface and whose sequence determines the genomic position of each BIP-seq measurement (Figure 3.1A).
Briefly, we began with a pool of over 6,000 yeast strains, and used antibodies to histone modifications in conventional ChIP to enrich DNA associated with histone modifications of interest. We sheared chromatin randomly using sonication to an average fragment size of 500-bp. The enriched DNA is then split into two fractions: one is processed through a barcode sequencing library preparation\textsuperscript{27} where each molecular barcode is PCR-amplified to attach the appropriate adapters for next-generation sequencing to yield BIP-seq measurements, and the other is processed through a standard next-generation sequencing library preparation to yield conventional ChIP-seq measurements for comparison (Figure 3.1D). A small sample of whole-cell extract was saved for each immunoprecipitation and run through barcode sequencing to estimate the starting population of each yeast strain present in each pool.

BIP-seq generates barcode counts, which are normalized by the distribution of strains in the starting pool, and also by nucleosome occupancy (H3), which is directly measured in parallel with each experiment. Comparable ChIP counts are generated by counting the number of reads that correspond to positions within 500-bp of the start codon of each ORF present in the deletion collection. Detailed experimental and data analysis methods can be found in the Supporting Information.

Chapter 3.4: Results
Chapter 3.4.1: Gene replacement by \textit{kanMX} does not produce wholesale changes in chromatin state

We performed this assay for seven different histone modifications, including trimethylation of histone H3 lysines 4, 36, and 79 (H3K4me3, H3K36me3, and
H3K79me3, respectively), acetylation of histone H3 lysines 9, 27, and 56 (H3K9ac, H3K27ac, and H3K56ac, respectively), and H3 for normalization. We also performed BIP-seq using antibodies to nuclear pore complex proteins as a negative control (Figure 3.2A). 4,822 strains were detectable in our starting pool across all three biological replicates obtained for each histone modification, with Pearson r values between replicates ranging from 0.51 – 0.92 (Figure 3.2A and Supplemental Figure 3.1). As validation of the technique, we compared our results with data collected using ChIP-qPCR as a part of a prior study, and found that both ChIP-seq and BIP-seq measurements corresponded with the ChIP-qPCR data collected in low-throughput using individual strains (Supplemental Figure 3.2). As expected given our preliminary data, correlations between barcode immunoprecipitation (BIP) and chromatin immunoprecipitation (ChIP) are extremely significant (Figure 3.2B), and indicate that kanMX is not causal for any of the assayed histone modifications on a global scale. Of the modifications examined, histone acetylations were more consistently correlated with wild-type measurements than histone methylations.

Chapter 3.4.2: kanMX repositions the transcription start site and perturbs the distribution of some histone methylations

Each insertion of kanMX also perturbs the position of the transcription start site (TSS) from its wild-type position to one ~500 bp downstream. Thus, the comparisons made in Figure 3.2B reflect measurements made at identical genomic positions, but also different positions relative to a functional TSS. To probe for effects related to the shift in TSS position, we recomputed ChIP-seq measurements using a 500-bp sliding window.
We compared each window to BIP-seq measurements to identify the offset in genomic position that produced the maximum correlation between the two datasets. For histone acetylations and H3K36me3, BIP-seq and ChIP-seq measurements were most correlated with no offset in genomic position (Figure 3.2C). In contrast, H3K4me3 and H3K79me3 were most correlated with an offset of ~300 bp upstream of the TSS, and suggest that deposition of these two marks may depend somewhat upon gene architecture and/or transcription (Figure 3.2C).

Chapter 3.4.3: *kanMX* enhances the association between histone acetylation and native promoter strength

Histone lysine acetylation and methylations at lysine 4 and 36 are associated with actively transcribed genes, but it has been unclear whether these associations are due to elements contained within the upstream promoter or due to the process of transcription itself. Because *kanMX* has perturbed expression at each locus to be similar (Figure 3.1B) without perturbing the native promoter sequence, we hypothesized that measurements at *kanMX* would distinguish between these two possibilities. To test this, we correlated RNA-seq expression measurements in wild-type diploid yeast with measurements at *kanMX*, and compared these against correlations with the corresponding measurements on native chromatin. Because of the *kanMX* insertion, we expected to disrupt correlations between histone modifications and expression that were due to the transcription process, whereas correlations due to elements within the promoter would remain. Across the different marks tested, we found that insertion and expression of *kanMX* decreases the association of wild-type gene expression and histone methylations (H3K4me3,
H3K36me3, and H3K79me3; Figure 3.3), as expected. Remarkably, the associations between expression and histone acetylations (H3K9ac, H3K27ac, H3K56ac; Figure 3.3) are each significantly enhanced. Thus, histone acetylation directly upstream of a uniformly expressed *kanMX* gene was more predictive of wild-type levels of expression than measurements of wild-type histone acetylation.

Chapter 3.4.4: BIP-seq identifies interactions between protein binding at promoters and histone state

Together, the data show that the perturbations to expression and gene sequence have modest global effects on histone modifications, and may enhance observation of activity at native gene promoters. This led us to explore promoter sequence features that might explain the maintenance of histone state despite the insertion and expression of *kanMX*. We downloaded a comprehensive survey of genome-wide localizations of over 200 transcription factors and chromatin modifiers that define which protein factors were bound at each promoter in wild-type yeast. To identify potential interactions between protein binding and histone state, we asked whether protein-binding status significantly partitioned histone modification data collected by either BIP-seq or ChIP-seq (Figure 3.4A). For each protein-binding factor, we computed differences in the levels of histone modification at bound genes and non-bound genes, assessing significance using a Mann-Whitney U-statistic. Across different significance thresholds, BIP-seq data reveal more significant interactions than ChIP-seq data for all histone acetylations and H3K36me3, whereas for H3K4me3 and H3K79me3, BIP-seq generated approximately equal or fewer significant interactions than ChIP-seq, respectively (Figure 3.4B). These results suggest
that histone acetylations and H3K36me3 may play important roles in regulation of gene promoters, whereas H3K4me3 and H3K79me3 are active elsewhere.

Of the transcription factors and chromatin modifiers examined by Venters, et al.\textsuperscript{92}, we found numerous factors where protein binding was highly correlated with levels of histone modification. The most significant interactions were found to be among members of the preinitiation complex (PIC)\textsuperscript{93} that assembles the enzymatic machinery for transcription by RNA polymerase II (all p < 1e-20; Figure 3.4C, left). The promoters targeted by these factors contained much higher levels of histone acetylation, and even showed some level of coherence among the subcomplexes within the PIC, for example within members of the RNA Polymerase II holoenzyme. In contrast, the interactions observed for the PAF complex involved in transcriptional elongation\textsuperscript{94} were much less significant (Figure 3.4C, right). In addition to significant interactions for members of the transcriptional machinery, BIP-seq data also identified important subunits within known chromatin modifying complexes, such as SAGA and NuA4\textsuperscript{95}. Within SAGA, BIP-seq identified significant interactions with histone acetylations for SPT3, SPT7, ADA2, and GCN5, which together form the core subunits of SAGA (Figure 3.4D, left). Of the subunits tested from the NuA4 acetyltransferase, BIP-seq identified the catalyzing acetyltransferase component ESA1 with significant interactions with histone acetylation (Figure 3.4D, right). The same analyses using ChIP-seq data reveal far fewer and less significant interactions with histone acetylations, but equally strong interactions with H3K4me3 (Supplemental Figure 3.3).

Since our BIP-seq measurements were more predictive of wild-type gene expression than wild-type ChIP-seq measurements, we were curious to see if this result
would also manifest in our analysis of transcription factor binding. We organized all of the interactions for all binding factors from Venters, et al.\textsuperscript{92} according to the mean expression level of the genes targeted by each TF. Organizing a heatmap in this way, it is apparent that the BIP-seq dataset (Supplemental Figure 3.4A) more clearly reflects the genome-wide positive correlation between acetylation and expression than the ChIP-seq dataset (Supplemental Figure 3.4B).

Chapter 3.4.5: BIP-seq analysis of H2A.Z

In contrast to histone acetylations and methylations, whose roles in transcriptional activation are well established, the histone variant H2A.Z (called Htz1 in yeast) appreciates a more colorful history in the scientific literature. Genome-wide studies of Htz1 have come to varying conclusions about its role in transcriptional regulation. Mapping studies find that Htz1 is most abundant at inactive genes, suggesting a repressive role\textsuperscript{96,97}, but deletion of Htz1 hinders proper activation of some genes and causes spread of heterochromatic silencing near telomeres, suggesting an activating role\textsuperscript{98}. Current models have proposed that Htz1 “poises” inactive genes for rapid activation upon induction, and does so through a mechanism of histone loss\textsuperscript{99}. Other models propose that Htz1 serves as transcriptional memory to mark instances of RNA polymerase initiation\textsuperscript{100}.

Such models are difficult to test, because the proposed mechanisms may yield different effects for different subsets of genes, and averaging measurements from a genome-wide map across many genes combines effects on chromatin that are uncontrolled and challenging to interpret. BIP-seq, however, controls for gene sequence
and expression, so we performed BIP-seq using antibodies to Htz1 to measure H2A.Z abundance at wild-type promoters upstream of a uniformly expressed \textit{kanMX} gene cassette.

In a model where Htz1 serves as transcriptional memory, BIP-seq measurements upstream of a uniformly expressed \textit{kanMX} cassette should reveal uniformly distributed levels of Htz1 across all positions. However, our direct observations of Htz1 at wild-type and \textit{kanMX} loci show a high degree of correlation (Figure 3.5A), comparable to the correlations observed for histone acetylations and methylations (Figure 3.2B).

Within certain promoters, Htz1-containing nucleosomes flank a nucleosome-depleted region at the TSS\textsuperscript{101}, so we were curious whether the perturbation to the position of the TSS in \textit{kanMX} strains influenced Htz1 deposition. We conducted a sliding-window analysis similar to Figure 3.2C to assess the whether Htz1 measured at \textit{kanMX} sites and at wild-type loci is most correlated at identical genomic positions or at offset positions. We found that Htz1 is most correlated at identical genomic positions, providing evidence that Htz1 deposition is not dependent on the relative position of the TSS (Figure 3.5B).

Htz1 occupies the promoters of inactive genes, yet deletion studies have demonstrated its importance in the activation of some genes. In one model, the role of Htz1 in activation occurs through accelerated loss of Htz1-containing nucleosomes\textsuperscript{99}, and provides a potential explanation for why Htz1 may appear more abundant at inactive genes, as histone loss is proportional to the rate of transcription. Under this model, one may expect that controlling for gene sequence and expression would thus reveal an activating role for Htz1. However, in our measurements at \textit{kanMX} sites, we found that a modest anticorrelation observed between Htz1 at wild-type promoters and gene
expression is significantly more anticorrelated when comparing Htz1 at *kanMX* with gene expression (Figure 3.5C). Similar to our observations of histone acetylation, measurements of Htz1 at *kanMX* are thus more predictive of the wild-type expression of the downstream gene than wild-type measurements, though as a gene expression inhibitor. These results suggest that Htz1 deposition is indeed a feature encoded by promoter sequence and is not dependent upon the transcriptional machinery or a function of transcriptional memory.

Htz1 tends to be localized to genes lacking a TATA box \(^1\), and indeed, we observe a very significant difference in Htz1 enrichment between TATA and TATA-less genes defined by Basehoar, et al.\(^2\) (p < 4.6e-31, Figure 3.5D). A more recent model characterizes two types of promoter architectures, in which an “open” architecture typically carries two H2A.Z-containing nucleosomes flanking a ~150 bp nucleosome-depleted region of DNA, and another “covered” architecture is more occupied by canonical histones and require activator binding and chromatin remodeling for gene activation\(^3\). Consistent with this model, H2A.Z is significantly more abundant at “open” promoters relative to “covered” promoters (p < 7.6e-55, Figure 3.5D). These results are more significant than similar comparisons using wild-type measurements (Supplemental Figure 3.5A), and are consistent with our claim that BIP-seq measurements enhance the observation of chromatin characteristics that are specific to promoters.
Chapter 3.4.6: Nhp6a is negatively associated with H2A.Z

We then related Htz1 abundance to promoter binding events, and found many expected interactions in which higher levels of Htz1 are associated with binding by members of the SWR1 complex (Figure 3.5E), which is the complex responsible for exchanging canonical H2A with Htz1 subunits$^{104,105}$. We also observed negative interactions, in which protein binding is associated with decreased levels of Htz1. In some instances, such as members of the RNA polymerase II holoenzyme (Rpb3 and Rpo21) or factors related to ribosomal protein gene expression (Ifh1 and Rap1), the negative result might be a simple extension of our finding that Htz1 occupies inactive genes. For the two most significant interacting factors, Bur6 and Nhp6A, the connections were not as clear. Bur6 encodes one part of the two-member NC2 complex that is primarily associated with transcriptional repression$^{106}$, though instances of activation have also been reported$^{107}$. Nhp6A encodes a high mobility group B (HMGB) protein that binds and remolds nucleosomes to modulate chromatin structure$^{108}$, and is also a member of the FACT complex, which has roles in transcriptional elongation and initiation. Below, we follow up specifically on the negative interaction between Nhp6A and Htz1 where Nhp6A binding is associated with the absence of Htz1.

Nhp6A is described in the literature as a structural protein that binds 22% of promoters in yeast$^{108}$, and has been shown to stabilize chromatin and assist in the assembly of the pre-initiation complex for transcriptional initiation. As both Nhp6a and Htz1 are found preferentially at promoters, we analyzed the promoters targeted by each protein using whole-genome ChIP-chip binding maps published in earlier studies for both
Nhp6a\textsuperscript{109} and Htz1\textsuperscript{96}. Using the genes defined in each study to be bound by each protein (962 Nhp6a-bound and 1455 Htz1-containing promoters), we determined the likelihood of the observed overlap if the binding of each were independent events. We observed an overlap of only 59 promoters, which is significantly lower than was expected by random chance (p < 6.5e-48, hypergeometric test; Figure 3.5F). Nhp6a is a part of the FACT complex, and FACT has been shown to promote binding of TATA-binding protein (TBP) to a TATA box\textsuperscript{110}. As Htz1 is found preferentially at TATA-less genes, a possible explanation for this negative interaction is that Nhp6A and Htz1 operate exclusively at the two different types of promoters. However, even when accounting for the presence of a TATA box, binding of Nhp6a and presence of Htz1 are significantly exclusive within the set of TATA-containing genes (p < 5.2e-07, hypergeometric test, Supplemental Figure 3.5B) and TATA-less genes (p < 2.6e-31, Supplemental Figure 3.5C, TATA-containing genes defined by Basehoar, et al.\textsuperscript{102}). Thus, the interaction cannot be attributable solely to action at different types of promoters.

Curiously, genome-wide transcriptional profiling in \textit{htz1}\textsuperscript{Δ} and \textit{nhp6a}\textsuperscript{Δ} strains\textsuperscript{109} reveals a significant overenrichment in the overlap of differentially expressed genes between the two mutants (p < 1.6e-05, hypergeometric test; Figure 3.5G), despite the two proteins’ exclusive patterns of promoter binding. To investigate this further, we examined whether promoters bound by Nhp6a or that contain Htz1 were particularly sensitive to deletion of \textit{htz1}\textsuperscript{Δ} and \textit{nhp6a}\textsuperscript{Δ}, respectively. 6,202 \textit{S. cerevisiae} genes were sorted by their expression in either \textit{htz1}\textsuperscript{Δ} or \textit{nhp6a}\textsuperscript{Δ}. Sliding 500-gene windows across each data set were examined. For each window, the percentage of gene promoters either bound by Nhp6A or that contain Htz1 was plotted as a function of average gene
expression change in the \(htz1\Delta\) or \(nhp6a\Delta\) mutant. We found that as genes are more differentially overexpressed in an \(htz1\Delta\) mutant, gene promoters are more likely to be bound by Nhp6a (Figure 3.5H). In contrast, as genes are more differentially expressed in an \(nhp6a\Delta\) mutant, they are less likely to possess promoters that contain Htz1 (Figure 3.5I). Taken together, the evidence suggests that Htz1 incorporation into a promoter inhibits binding and gene activation by Nhp6a, though not \textit{vice versa}.

Chapter 3.5: Discussion

Here we have developed a genome-wide assay to quantify the impact of a common genetic perturbation on histone-DNA interactions. The common insertion of the \textit{kanMX} gene has standardized the levels of expression at each locus (Supplemental Figure 3.1), and the BIP-seq technique we present enables the simultaneous measurement of chromatin interactions with the same gene sequence in thousands of yeast strains. Our results show only modest differences in histone state due to perturbation of gene sequence and expression, and show distinct patterns of association with expression for histone acetylations and methylations. We also find a novel epigenetic-genetic interaction between the HMGB protein Nhp6A and histone variant Htz1 which encodes the histone variant H2A.Z in yeast.

Chapter 3.5.1: Histone methylations govern gene architecture by defining gene boundaries

The histone methylations H3K4me3 and H3K36me3 are each catalyzed exclusively by Set1p and Set2p, respectively\textsuperscript{43,51,112}. Both Set1p and Set2p are known to
associate directly with RNA Polymerase II during the process of transcription\textsuperscript{50,113,114}, so it is not surprising that the insertion of \textit{kanMX} results in a significant decrease in each correlation with expression (Figure 3.3). H3K4me3 was especially perturbed in the presence of \textit{kanMX}, and our results suggest that this is partially due to the repositioning of a functional TSS (Figure 3.2C). For a vast majority of gene loci, measurements of H3K4me3 at the TSS appear to be compressed within a narrow range relative to measurements at \textit{kanMX} (Figure 3.2B). Though a similar result is obtained when comparing wild-type H3K4me3 at the TSS with measurements at 500 bp upstream of the TSS (Supplemental Figure 3.2C), comparison of the upstream measurements with \textit{kanMX} show little correspondence (Supplemental Figure 3.2D). This suggests that H3K4me3 may serve a general feature of gene architecture as opposed to a regulatory feature of the upstream promoter.

Chapter 3.5.2: Histone acetylations govern chromatin structure and contribute to gene activation

In contrast, histone acetylations are catalyzed by a variety of chromatin-remodeling complexes that interact directly with sequence-specific transcription factors that generally bind upstream of each \textit{kanMX} perturbation\textsuperscript{6,53}. Furthermore, global acetylation levels are thought to be independent of transcriptional status\textsuperscript{6}. Accordingly, we observe a stronger association between histone acetylation and wild-type gene expression in the presence of \textit{kanMX} than in wild type, presumably because wild-type transcription is non-uniform and imposes additional chromatin transactions that introduce noise in ChIP measurements over thousands of unique genetic elements.
Chapter 3.5.3: Histone variant H2A.Z (Htz1) contributes negatively at gene promoters

Htz1 has been reported to localize to distinct sets of promoters that are either active or inactive\textsuperscript{96,115}. Subsequent follow-up studies have proposed that Htz1 facilitates histone turnover during expression at active loci\textsuperscript{116,117}, and that Htz1 at inactive loci serves to ‘poise’ downstream genes for rapid expression\textsuperscript{117}. In our data, whereas measurements on native chromatin showed modest negative correlation with promoter strength, we observed a more significant inverse association between measurements of Htz1 at \textit{kanMX} and promoter strength (Figure 3.5A). In the context of what is known about Htz1, our results suggest that the genome-wide insertion and expression of \textit{kanMX} may have perturbed the association of Htz1 with active loci.

Here we report a novel assay called barcode immunoprecipitation and analysis by high-throughput sequencing (BIP-seq) to quantify protein-DNA interactions at a defined genetic perturbation in a pooled format containing many different yeast strains. Our system has quantified histone and histone modification interactions at a controlled gene sequence, \textit{kanMX}, which has been precisely positioned at thousands of loci throughout the yeast genome as part of the \textit{S. cerevisiae} genome deletion project. Because the \textit{kanMX} gene is expressed at similar levels at each gene locus, we were able to assess how normalizing for gene sequence and expression altered the levels of histone state, and found only modest differences in histone state due to \textit{kanMX}. As the only sequence differences at each \textit{kanMX} gene are due to the upstream wild-type promoter, we related our measurements to features found within the upstream promoter to find significant interactions between components of the transcriptional machinery and histone
acetylation, as well as a novel negative interaction between Nhp6A and Htz1. BIP-seq provides a unique experimental system in which promoter activity and transcriptional activity may be decoupled to dissect how histone modifications influence gene expression.

Chapter 3, in full, is a reprint of a manuscript currently in submission: “Chen M, Ideker T. Systematic gene replacement imparts experimental control for genome-wide analysis of histone-promoter interactions. PLoS ONE. In submission.” The dissertation author was the primary investigator and author of this paper.
Figure 3.1: Schematic illustrating Barcode immunoprecipitation with analysis by high-throughput sequencing.

(A) The yeast knockout library is a collection of yeast strains, each with a defined native gene that has been replaced by kanMX. Each kanMX gene is flanked by unique 20-bp molecular barcodes that uniquely identifies the strain and the position of kanMX integration. (B) Comparison of wild-type expression levels with expression levels for kanMX. The dynamic range of expression observed for wild-type genes on chromosome I is much wider than the range of expression levels observed for kanMX at 90 positions throughout chromosome I. The range of expression for all genes is also shown (data from 90). (C) A multitude of factors can influence histone state, including promoter sequence and activity, as well as gene sequence, length and activity. Measurements of histone state across different kanMX strains controls for differences in gene sequence, length, and activity, and should thus reveal promoter interactions with histone state more clearly. (D) BIP-seq follows a conventional ChIP protocol, but is performed on a pool of many yeast strains. Immunoprecipitated DNA is analyzed by conventional library preparation methods, as well as a Bar-seq library preparation to specifically amplify the kanMX-flanking molecular barcodes. A sample of whole cell extract is also run through Bar-seq to estimate the relative abundance of each strain in the yeast pool.
Figure 3.2: BIP-seq measurements of histone acetylation and methylation.

(A) BIP-seq measurements for a pair of biological replicates are shown as a positive control. All biological replicates are shown in Supplemental Figure 3.1. A negative control using antibodies to nuclear pore complex proteins is also shown. BIP-seq data are plotted on the x-axis, and wild-type ChIP-seq data are plotted on the y-axis. (B) Each plot shows the level of the indicated histone modification recovered with Barcode IP compared to the levels recovered using conventional ChIP-seq. Each dataset has been normalized by histone occupancy using BIP and ChIP measurements collected in parallel. All comparisons are significantly correlated (p < 2.2e-16), though histone acetylations are more consistently correlated than histone methylations. (C) To assess whether different histone modifications were perturbed by the change in the relative position of the functional TSS, we compared BIP-seq measurements to a sliding window of ChIP-seq measurements to determine the base-pair offset that produces the maximum correlation. Histone acetylations and H3K36me3 are most correlated with no offset. H3K4me3 and H3K79me3 are most correlated with a ~300 bp upstream offset.
Figure 3.3: Correlation between histone modification and expression before and after insertion of $\textit{kanMX}$.

Shown are Spearman $\rho$ values (analogous to Pearson’s $r$) representing correlations between wild-type gene expression and histone modification levels as sampled by chromatin and barcode immunoprecipitation (dark and light bars, respectively).
Figure 3.4: Analysis of transcription factor and chromatin modifier binding and BIP-seq histone modifications.

(A) Interactions between binding factors and histone modifications are identified by assessing whether gene targets bound by a factor have significantly more or less histone modification enrichment than non-bound targets. Significance is assessed using a non-parametric rank-sum test (Mann-Whitney U). (B) BIP-seq data recover more significant interactions than ChIP-seq for histone acetylations and H3K36me3, and fewer interactions for H3K79me3. Interactions were assessed using binding maps from Venters, et al. and histone modification measurements at kanMX loci (BIP-seq, black) or wild-type (ChIP-seq, red). Each plot shows the number of significant interactions found (y-axis) at a given significance threshold (x-axis). (C) Significant interactions clustered into complexes associated with the transcriptional machinery. Components of the pre-initiation complex (left) are grouped by subcomplex membership, indicated with colored bars. Similar grouping of members of the PAF complex (right), involved in transcriptional elongation, were less significant. (D) BIP-seq measurements recover interactions that define the acetyltransferase activities of the SAGA and NuA4 complexes.
Figure 3.5: BIP-seq analysis of histone variant H2A.Z (Htz1).

(A) BIP vs ChIP. (B) maximum correlation with no offset. (C) Htz1 at kanMX is much more anticorrelated with expression than wild-type Htz1. (D) Htz1 is more enriched at TATA-less promoters than TATA-containing promoters\(^{102}\) (* p < 4.6e-31, Mann-Whitney), and is also more abundant at ‘open’ promoters than ‘covered’ promoters\(^{103,118}\) (** p < 7.6e-55). Note that analogous differences using ChIP-seq measurements are less significant (Supplemental Figure 3.5A). (E) Htz1 interactions with DNA-binding factors show positive interactions (pink) with members of the SWR1 complex, and negative interactions (blue) with components of the RNA Polymerase II holoenzyme (Rpb3 and Rpo21) and activators of ribosomal protein gene expression (Ifh1 and Rap1). We also identify previously unreported negative interactions between Htz1 and Bur6 and Nhp6A. (F) ChIP-chip data measuring wild-type binding of Nhp6A\(^{109}\) and incorporation of Htz1\(^{96}\) show a very significant underenrichment in the overlap of targeted promoters. (G) Transcriptional profiling of htz1D and nhp6aD show a surprising overenrichment in the identities of differentially expressed genes. (H) Genes with higher levels of differential expression in a htz1D background are associated with a higher likelihood of being a promoter targeted by Nhp6A. (I) Differentially expressed genes in an nhp6aD background are associated with a lower likelihood of incorporating Htz1 in the promoter.
Supplemental Figure 3.1: Replicate plots of the BIP-seq and ChIP-seq measurements in this study.

Three biological replicates were performed in total; replicates of the first two are shown. BIP-seq data has been normalized by background pool population, and normalized by H3 content at each locus. ChIP-seq data have been normalized by H3 content. See Methods and Supplemental Information.
Supplemental Figure 3.2: Validation of ChIP-seq and BIP-seq measurements and analysis of the effect of TSS positioning on H3K4me3 measurements.

ChIP-qPCR was performed in both wild-type diploids (BY4743) and four different knockout strains as a part of \(^9\). Shown are ChIP-seq counts at the loci sampled in \(^9\) plotted against wild-type ChIP-qPCR data (A), and BIP-seq counts of the strains sampled in \(^9\) plotted against \textit{kanMX} ChIP-qPCR data (B). (C) The correlation between BIP-seq and ChIP-seq measurements of H3K4me3 is very similar to the correlation between ChIP-seq measurements at the TSS with measurements at a position 500 bp upstream of the TSS. (D) BIP-seq measurements are somewhat correlated to H3K4me3 measurements at positions 500 bp upstream of the TSS.
Supplemental Figure 3.3: Analysis of transcription factor and chromatin modifier binding using ChIP-seq histone modifications.

Significant interactions were detected using the analysis shown in Figure 3.4A, but using measurements from ChIP-seq instead of BIP-seq. Promoter-binding factors are listed in rows, and histone modifications comprise the columns. The color of each cell represents the difference in mean histone modification between bound and non-bound gene targets. Note the stronger differences in H3K4me3 relative to the other modifications. (A) Components of the pre-initiation complex are grouped by subcomplex membership, indicated with colored bars. (B) Similar grouping of members of the PAF complex, involved in transcriptional elongation, were less significant. (C) ChIP-seq measurements show strong interactions between H3K4me3 and histone acetyltransferase complexes (C) SAGA and (D) NuA4.
Supplemental Figure 3.4: Interactions between transcription factors and chromatin modifiers ordered by mean expression level of each factor’s bound set of genes.

Interactions between all promoter-binding factors and histone modifications as measured by BIP-seq (left) and ChIP-seq (right) are visualized as a heatmap. Promoter-binding factors are listed in rows and ordered by the average expression level of the genes bound by each factor. Each cell is colored by the mean difference in histone modification enrichment between each promoter-binding factors bound targets and its unbound targets.
Supplemental Figure 3.5: Analysis of wild-type Htz1 enrichments and overlap of promoters bound by either Nhp6A or Htz1.

(A) Wild-type levels of Htz1 at TATA-less genes are significantly higher than at TATA-containing genes (* p < 2.1e-22, Mann-Whitney), as previously reported[99]. Htz1 is also enriched in “open” promoters compared to “covered” promoters (*** p < 1.7e-35), as defined by Tirosh and Barkai[118]. Note that analogous differences using BIP-seq measurements are more significant (Figure 3.5D). (B, C) To assess whether the negative interaction between Nhp6A and Htz1 is due to differences in the types of promoters that are bound, we performed the same analysis as in Figure 3.5F within the sets of TATA-containing genes (B) or TATA-less genes (C). We found that the negative interaction was still significant even after accounting for the type of promoters that are bound.
Chapter 4: Conclusion

Epigenetic features have been implicated in critical cellular functions and pathologies, ranging from gene silencing and DNA repair in yeast\textsuperscript{5} to tissue-specific gene expression, cell differentiation, carcinogenesis, and aging in humans and mammals\textsuperscript{119}. Throughout development, differentiating cells accumulate epigenetic instructions that ultimately determine fully differentiated patterns of expression. Many disease phenotypes, including cancer, stem from fundamental epigenetic changes that inactivate critical genes or activate disruptive genes\textsuperscript{78,120,121}.

Identifying disease-causing epigenetic changes\textsuperscript{122} and finding ways to mitigate, alter, or reverse deleterious ones will be the subject of biomedical research for the foreseeable future. The work presented in this dissertation assists these efforts by providing a platform for assessing epigenetic effects within a controlled genetic environment. In Chapter 2, I demonstrated how the YKO library might be used as a readily available position-effect screening library, and discovered a distinct epigenetic-genetic interaction between H3K36me3 and the Rap1-driven \textit{kanMX} gene. This association is notably distinct from genome-wide patterns that have previously been reported, and suggest distinct epigenetic-genetic interactions may exist that differ significantly from the general associations produced from genome-wide maps. The principle limitation with this work is that the specific knowledge gained from any position-effect screen is limited to the gene of interest (i.e., \textit{kanMX}, in this case). A natural next step beyond studying \textit{kanMX} is the study of additional genes of interest. However, this requires the replacement of \textit{kanMX} by the gene to be studied at each position of interest, and though the common \textit{kanMX} sequence may simplify some
aspects, at this time such an undertaking would still require substantial effort and time. More advanced techniques may be developed in the future, and indeed there may be promise in adapting mass transformation technologies such as those employed in diploid-based synthetic lethality analysis (“dSLAM”)\(^5\), but at present such technologies carry considerable procedural variability that would hamper subsequent analyses (data not shown).

In Chapter 3, I developed a high-throughput technology called BIP-seq that took advantage of second-generation DNA sequencing technologies and molecular barcoding that had been built into the YKO library. We took advantage of the placement of each molecular barcode, which resides at the interface between a native wild-type promoter and a common \textit{kanMX} gene cassette, to control for downstream effects and analyze how differences in wild-type promoters upstream of each \textit{kanMX} influenced histone state. We observed a striking dichotomy in how histone acetylations and methylations related to promoter activity. On the whole, our results suggested that histone methylations in general play a role in defining gene architecture, whereas histone acetylations are more associated with gene regulation. Furthermore, we were able to apply BIP-seq to study Htz1 in yeast, and found a novel negative interaction between the HMG protein Nhp6A and deposition of Htz1 at gene promoters. The development of BIP-seq represents one of the first uses of barcoding technology for an application other than population monitoring, and a wide array of other immunoprecipitation targets are now within reach. Other targets are limited by the availability of suitable antibodies for precipitating a target of interest. For the study of other histone modifications, BIP-seq will benefit from the rapidly growing catalog of ChIP-grade antibodies targeting other histone modifications.
For transcription factors, however, the availability of suitable antibodies is much more restricted. Typical studies of transcription factor binding in yeast use epitope-tagged forms of each transcription factor, in order to standardize which antibodies and immunoprecipitation procedures are used. The YKO library carries only wild-type transcription factors, and introducing an epitope-tagged version for BIP-seq analysis would require transformations in every strain within the YKO pool. Analysis of transcription factor binding by BIP-seq, then, would necessitate antibodies specific to native versions of a yeast transcription factor, which are not readily available from commercial sources.

My work has generated a number of benefits that have advanced the field. First, position-effect measurements effectively control for gene sequence, and are thus more able to address questions of cause and effect that often arise when genome-wide mapping efforts report a correlation between expression and an epigenetic feature. Whereas genome-wide correlations may show correlation between a histone modification and expression, it is unclear whether histone modifications are the cause or the effect. Because the YKO library has introduced a known perturbation to gene sequence, examinations of histone modifications and expression using the YKO library are better able to distinguish directionality in the epigenetic-genetic relationship.

Second, though we specifically address histone modifications, the technologies we present are general techniques that can be adapted to study other epigenetic phenomena, such as chromatin accessibility. The key advantage of using the YKO library for these assays is the experimental control of gene sequence and precise knowledge of genomic positions for each strain. For example, BIP-seq technology could be adapted for
studying chromatin accessibility by integrating DNase digestion and techniques from DNase-seq\textsuperscript{123}. Within a pool of yeast strains, DNase digestion within the molecular barcode or priming regions should prevent amplification. Thus, detection of molecular barcodes following DNase digestion would suggest a relatively protected region of the genome. Such an extension of BIP-seq may address whether the previously observed association between chromatin accessibility and transcription is a regulatory feature, or whether it is a downstream consequence of transcription.

Third, our approach repurposes a gene knockout library in \textit{S. cerevisiae}, but many of these advances are also applicable to strain libraries in other model organisms. To date, knockout collections exist in \textit{S. pombe}\textsuperscript{67}, \textit{N. crassa}\textsuperscript{68}, \textit{C. elegans}\textsuperscript{69}, \textit{A. thaliana}\textsuperscript{70}, \textit{D. melanogaster}\textsuperscript{71,72}, and \textit{M. musculus}\textsuperscript{73}. These collections employed similar genetic engineering techniques in their construction, and are thus amenable to analyses similar to those presented here. These other model organisms also enable investigation of epigenetic mechanisms not present in budding yeast. For example, \textit{S. pombe} would enable the study of RNA interference\textsuperscript{124}, and \textit{N. crassa} would enable the study of DNA methylation\textsuperscript{125}. With strain collections in multicellular organisms such as \textit{D. melanogaster} and \textit{M. musculus}, along with new technologies to edit genomes using TALE nucleases\textsuperscript{74} or Cas9\textsuperscript{126}, position-effect screening may serve as an additional tool for epigeneticists to more completely unravel the mechanisms that govern gene expression.
References


Cheng, C., Shou, C., Yip, K. Y. & Gerstein, M. B. Genome-wide analysis of chromatin features identifies histone modification sensitive and insensitive yeast


