

BOSTON UNIVERSITY
COLLEGE OF ENGINEERING

Dissertation

THE TITLE IS WASDA

by

JAMES CHUANG

B.S., Johns Hopkins University, 2013
M.S., Boston University, 2018

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2019

© 2019 by
JAMES CHUANG
All rights reserved

Approved by

First Reader

Fred Winston, PhD
Professor of Genetics
Harvard Medical School

Second Reader

Ahmad Khalil, PhD
Assistant Professor of Biomedical Engineering

Third Reader

L. Stirling Churchman, PhD
Assistant Professor of Genetics
Harvard Medical School

Fourth Reader

John T. Ngo, PhD
Assistant Professor of Biomedical Engineering

Fifth Reader

Wilson Wong, PhD
Assistant Professor of Biomedical Engineering

Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

James Chuang

THE TITLE IS WASDA

JAMES CHUANG

Boston University, College of Engineering, 2019

Major Professors: Fred Winston, PhD
Professor of Genetics
Harvard Medical School

Ahmad Khalil, PhD
Assistant Professor of Biomedical Engineering

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

List of Tables	ix
List of Figures	xiii
1 Introduction	1
1.1 A brief introduction to transcription	1
1.2 Transcription elongation factors Spt6 and Spt5	2
1.3 Reproducible data analysis for genomics	2
2 Genomics of transcription elongation factor Spt6	4
2.1 Collaborators	4
2.2 Introduction to Spt6 and intragenic transcription	4
2.3 Data analysis pipelines for TSS-seq and ChIP-nexus	7
2.3.1 TSS-seq peak calling	7
2.3.2 A note on ChIP-nexus peak calling	7
2.4 TSS-seq and TFIIB ChIP-nexus results for <i>spt6-1004</i>	7
2.5 MNase-seq results from <i>spt6-1004</i>	12
2.5.1 Clustering of MNase-seq profiles at <i>spt6-1004</i> -induced intra- genic TSSs	16
2.6 Other features of <i>spt6-1004</i> intragenic promoters	19
2.6.1 Information content and sequence preference of intragenic TSSs	19

2.6.2	Sequence motifs enriched at intragenic TSSs	19
2.7	Summary	20
2.8	Methods	20
2.8.1	Yeast strain construction and grown conditions	20
2.8.2	Sequencing library preparation (TSS-seq, ChIP-nexus, MNase-seq, NET-seq)	21
2.8.3	Genome builds	21
2.8.4	TSS-seq data analysis	21
2.9	Bibliography	23
3	Genomics of transcription elongation factor Spt5	26
3.1	Collaborators	26
3.2	Introduction to Spt5 and prior work	26
3.3	An aside on spike-in normalization for ChIP-seq	30
3.4	TSS-seq results from Spt5 depletion	30
3.5	MNase-seq results from Spt5 depletion	31
3.5.1	MNase-seq profile at Spt5-depletion-induced antisense TSSs .	32
3.6	Sequence motifs enriched at antisense TSSs	32
3.7	Summary	32
3.8	Bibliography	33
4	Stress-responsive intragenic transcription	34
4.1	Collaborators	34
4.2	Possible functions for intragenic transcription in wild-type cells	34
4.3	Discovery of stress-induced intragenic promoters by TFIIB ChIP-nexus and TSS-seq	34

4.4	Chromatin landscape of oxidative-stress-induced promoters	34
4.5	Polysome enrichment of oxidative-stress-induced intragenic transcripts	34
4.6	TSS-seq analysis of oxidative stress in <i>Saccharomyces sensu stricto</i> species	34
4.7	Functions of intragenic DSK2 expression in oxidative stress	34
4.8	Summary	34
4.9	Bibliography	36
	Bibliography	40

List of Tables

List of Figures

2.1	Western blot showing Spt6 protein levels in wild-type and <i>spt6-1004</i> cells, at 30 °C and after 80 minutes at 37 °C.	4
2.2	Diagram of transcript classes.	4
2.3	RNA-seq, TSS-seq, and TFIIB ChIP-nexus signal at the <i>AAT2</i> gene, in <i>spt6-1004</i> after 80 minutes at 37 °C.	5
2.4	Heatmaps of sense and antisense TSS-seq signal from wild-type and <i>spt6-1004</i> cells, over non-overlapping coding genes.	8
2.6	Bar plot of the number of TSS-seq peaks of various genomic classes differentially expressed in <i>spt6-1004</i> versus wild-type.	9
2.7	Set diagram of the number of genes with <i>spt6-1004</i> -induced intragenic transcripts reported in Cheung et al. (2008), Uwimana et al. (2017), and our TSS-seq data.	9
2.5	Heatmaps of TFIIB ChIP-nexus protection from wild-type and <i>spt6-1004</i> cells, over non-overlapping coding genes	10
2.8	Violin plots of expression level distributions for genomic classes of TSS-seq peaks in wild-type and <i>spt6-1004</i> cells.	10
2.9	TFIIB ChIP-nexus protection over the 20 kb flanking the gene <i>SSA4</i> , in wild-type and <i>spt6-1004</i> cells.	11

2.10 Scatterplots of fold-change in <i>spt6-1004</i> over wild-type, comparing TSS-seq and TFIIB ChIP-nexus.	12
2.11 Average MNase-seq dyad signal in wild-type and <i>spt6-1004</i> , over non-overlapping genes aligned by wild-type +1 nucleosome dyad.	13
2.12 Contour plot of nucleosome occupancy and fuzziness in wild-type and <i>spt6-1004</i>	13
2.13 Heatmaps of sense NET-seq signal, MNase-seq dyad signal, nucleosome occupancy changes, and nucleosome fuzziness changes over non-overlapping coding genes, aligned by genic TSS and arranged by sense NET-seq signal.	15
2.14 Average MNase-seq dyad signal around all <i>spt6-1004</i> -induced intragenic TSSs, grouped by a self-organizing map of the MNase-seq signal.	17
2.15 Average wild-type and <i>spt6-1004</i> MNase-seq dyad signal and GC content for three clusters of <i>spt6-1004</i> -induced intragenic TSSs, as well as wild-type genic TSSs.	18
2.16 Sequence logos of TSS-seq reads overlapping genic and intragenic TSS-seq peaks in <i>spt6-1004</i>	19
2.17 Kernel density estimate of matches to a consensus TATA-box motif upstream of genic and <i>spt6-1004</i> -induced intragenic TSSs.	20
2.18 Sequence logos of motifs discovered by MEME upstream of <i>spt6-1004</i> -induced intragenic and antisense TSSs.	21
3.1 Diagram of the dual-shutoff system used to deplete Spt5 from <i>S. pombe</i>	26

3.2	Average Spt5 ChIP-seq, RNAPII ChIP-seq, and sense NET-seq signal over non-overlapping coding genes, from Spt5 depleted and non-depleted cells.	29
3.3	Enrichment of RNAPII phospho-serine 5 and phospho-serine 2 over non-overlapping coding genes, in Spt5 depleted and non-depleted cells.	29
3.4	Heatmaps of antisense RNA-seq signal from Spt5 depleted and non-depleted cells, over non-overlapping coding genes.	30
3.5	Bar plot of the number of TSS-seq peaks of various genomic classes differentially expressed in Spt5 depleted versus non-depleted cells.	30
3.6	Heatmaps of antisense TSS-seq, RNA-seq, and NET-seq signal from Spt5 depleted and non-depleted cells, over genes with Spt5-depletion-induced antisense TSSs.	31
3.7	Average MNase-seq dyad signal from Spt5 depleted and non-depleted cells, over non-overlapping coding genes.	31
3.8	A figure showing MNase-seq signal around Spt5-depletion-induced antisense TSSs.	32
3.9	A figure showing motifs enriched upstream of Spt5-depletion-induced antisense TSSs.	32
4.1	TFIIB ChIP-nexus protection over all genes with stress-induced intragenic TFIIB peaks.	35
4.2	TFIIB ChIP-nexus protection over four genes with stress-induced intragenic TFIIB peaks.	36
4.3	Bar plot of the number of promoters from various genomic classes differentially expressed in oxidative stress.	37

4.4	TSS-seq expression levels in oxidative stress of oxidative-stress-induced genic and intragenic promoters.	38
4.5	A figure showing TSS-seq, TFIIB ChIP-nexus, and MNase-ChIP-seq for the oxidative-stress-induced promoters.	38
4.6	Polysome enrichment in oxidative stress, for oxidative-stress-induced genic and intragenic promoters.	39
4.7	A figure showing TSS-seq coverage over oxidative-stress-induced TSSs in the three species.	39
4.8	A figure showing TSS-seq coverage over DSK2 in the three species, possibly with the corresponding northern blot.	39
4.9	A figure showing TSS-seq, TFIIB ChIP-nexus, and MNase-ChIP-seq at DSK2.	39
4.10	A figure showing DSK2 fitness competition results.	39

Chapter 1

Introduction

1.1 A brief introduction to transcription

In eukaryotic cells, transcription of protein-coding genes is carried out by the protein complex RNA polymerase II

 Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.2 Transcription elongation factors Spt6 and Spt5

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.3 Reproducible data analysis for genomics

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet

orci dignissim rutrum.

Chapter 2

Genomics of transcription elongation factor Spt6

2.1 Collaborators

Steve Doris optimized TSS-seq and ChIP-nexus protocols
generated TSS-seq and ChIP-nexus libraries

Olga Viktorovskaya generated MNase-seq libraries

Magdalena Murawska generated NET-seq libraries

Dan Spatt various experiments for publication

2.2 Introduction to Spt6 and intragenic transcription

Studies in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* have previously examined the requirement for Spt6 in normal transcription (Cheung et al., 2008; DeGennaro et al., 2013; Kaplan et al., 2003; Pathak et al., 2018; Uwimana et al., 2017; van Bakel et al., 2013). As Spt6 is essential for viability in *S. cerevisiae*,

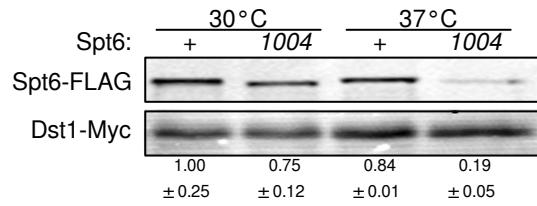
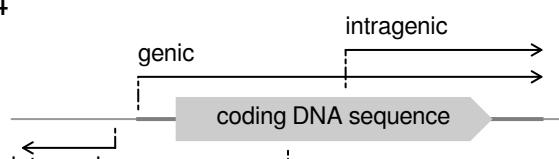


Figure 2.1: Western blot showing Spt6 protein levels in wild-type and *spt6-1004* cells, at 30°C and after 80 minutes at 37 °C. Immunoblotting was performed using α -FLAG antibody to detect Spt6 and α -Myc antibody to detect Dst1 from a spike-in strain. The quantification shown is the mean \pm standard deviation of three blots.



many of these studies use the same temperature-sensitive *spt6* mutant used in this project, ***spt6-1004***, which encodes an in-frame deletion of a helix-hairpin-helix domain within Spt6 (Kaplan et al., 2003). When *spt6-1004* cells are shifted from 30°C to 37°C for 80 minutes, bulk Spt6 protein levels are depleted to about 20% of wild-type levels (Figure 2.1). The most notable phenotype of the *spt6-1004* mutant is the appearance of **intragenic transcripts**, transcripts which appear to arise from within protein-coding sequences, in both sense and antisense orientations relative to the coding gene (Figure 2.2) (Cheung et al., 2008; DeGennaro et al., 2013; Kaplan et al., 2003; Uwimana et al., 2017).

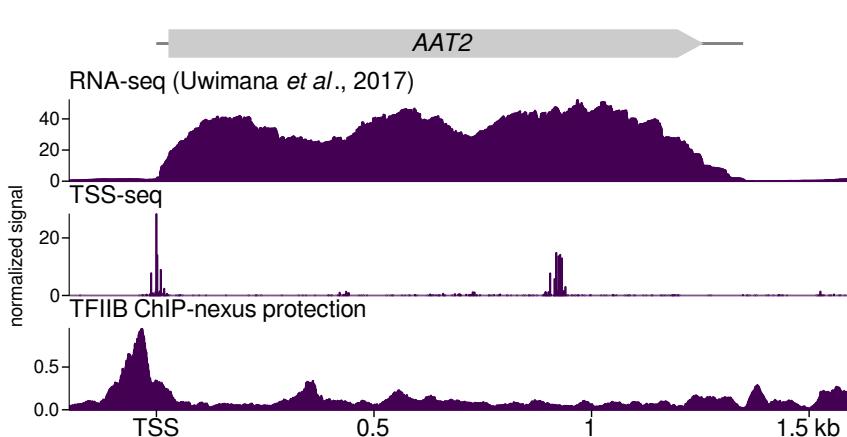


Figure 2.3: Sense strand RNA-seq signal, sense strand TSS-seq signal, and TFIIB ChIP-nexus protection at the *AAT2* gene, in *spt6-1004* after 80 minutes at 37°C.

Previous genome-wide measurements of transcript levels in *spt6-1004* relied on tiled microarrays (Cheung et al., 2008) and RNA sequencing (Uwimana et al., 2017). Studying intragenic transcription is difficult with these methods, since the signal for an

intragenic transcript in the same orientation as the gene it overlaps is convoluted with the signal from the full-length ‘genic’ transcript (Figure 2.3) (Cheung et al., 2008; Lickwar et al., 2009). Therefore, these methods can only discover intragenic transcripts which are highly expressed relative to the corresponding genic transcript, and are likely to find many false positives. Additionally, these methods are assays of steady-state RNA levels, which makes them unable to distinguish whether the intragenic transcripts observed in *spt6-1004* result from: A) new intragenic transcription initiation in the mutant, B) reduced decay of intragenic transcripts which are rapidly degraded in wild-type, or C) processing of full-length protein-coding RNAs.

To address these challenges to studying intragenic transcription, we applied two genomic assays to *spt6-1004*: transcription start-site sequencing (**TSS-seq**), and **ChIP-nexus of TFIIB**, a component of the RNA polymerase II pre-initiation complex (PIC). TSS-seq sequences the 5' end of capped and polyadenylated RNAs (Arribere and Gilbert, 2013; Malabat et al., 2015), allowing separation of intragenic from genic RNA signals and identification of intragenic transcript starts with single-nucleotide resolution (Figure 2.3). ChIP-nexus is a high-resolution chromatin immunoprecipitation technique, in which the immunoprecipitated DNA is exonuclease digested up to the bases crosslinked with the protein of interest before sequencing (He et al., 2015). When applied to the PIC component TFIIB, ChIP-nexus reports where transcription initiation is occurring, thus allowing us to determine if intragenic transcripts in *spt6-1004* result from new transcription initiation.

2.3 Data analysis pipelines for TSS-seq and ChIP-nexus

2.3.1 TSS-seq peak calling

2.3.2 A note on ChIP-nexus peak calling

2.4 TSS-seq and TFIIB ChIP-nexus results for *spt6-1004*

To study the relationship between Spt6 and transcription, TSS-seq and TFIIB ChIP-nexus libraries were prepared from wild-type and *spt6-1004* cells, both shifted from 30°C to 37°C for 80 minutes. In wild-type cells, TSS-seq and TFIIB ChIP-nexus recapitulate their expected distributions over the genome: Most TSS signal is restricted to annotated genic TSSs, while most TFIIB signal is localized just upstream of the TSS (Figures 2.4, 2.5). In *spt6-1004*, the signal for both assays infiltrates gene bodies, already suggesting that new transcription initiation does contribute to the intragenic transcription phenotype. Notably, sense strand TSS-seq signal in *spt6-1004* tends to occur towards the 3' end of genes, while antisense strand TSS-seq signal tends to occur towards the 5' end of genes.

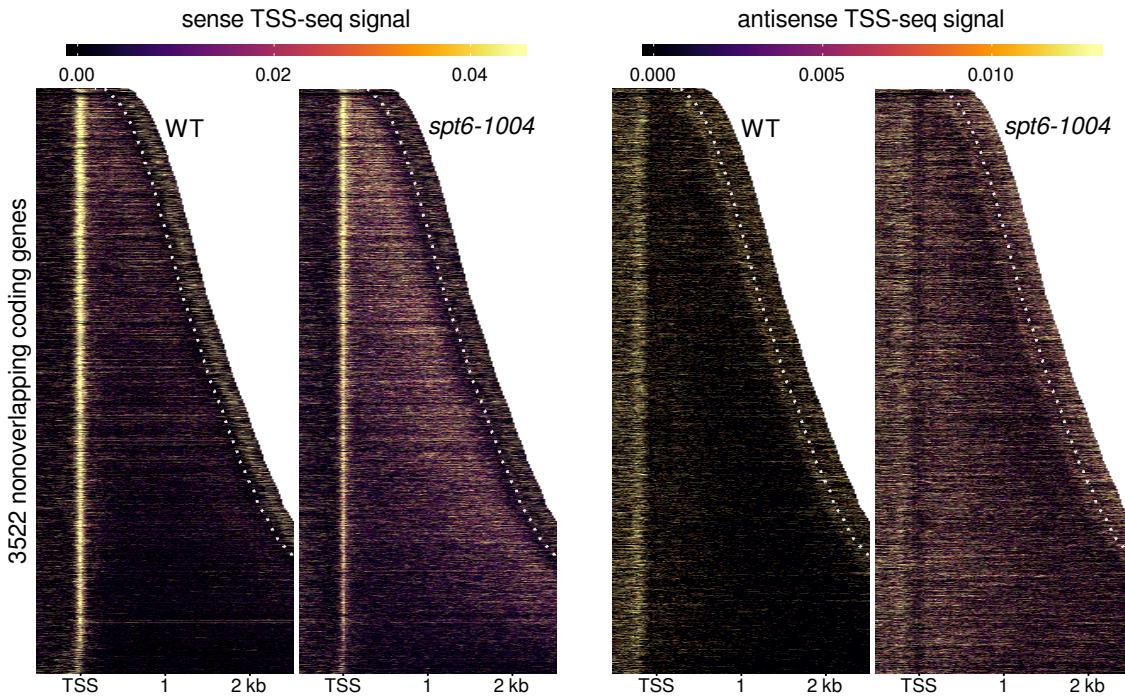


Figure 2.4: Heatmaps of sense and antisense TSS-seq signal from wild-type and *spt6-1004* cells, over 3522 non-overlapping genes aligned by wild-type genic TSS and sorted by annotated transcript length. Data are shown for each gene up to 300 nucleotides 3' of the cleavage and polyadenylation site (CPS), indicated by the white dotted line. Values are the mean of spike-in normalized coverage in non-overlapping 20 nucleotide bins, averaged over two replicates. Values above the 92nd percentile are set to the 92nd percentile for visualization.

The TSS-seq data were quantified by peak calling and differential expression analysis, and classified into genomic categories based on their position relative to coding genes. As suggested by the heatmap visualization (Figure 2.4), we detect significant induction of over 4000 intragenic and antisense TSSs in *spt6-1004* (Figure 2.6). Compared to previous studies identifying *spt6-1004* intragenic transcription by tiled microarray and RNA-seq, we identify intragenic transcription at over 1000 additional genes (Figure 2.7) and have the exact start sites of all identified TSSs.

The TSS-seq data also revealed an unexpected downregulation of most genic TSSs: In this experiment, we detected a significant downregulation to levels below 67% of wild-type levels at 75% (3579/4792) of genic TSSs (Figure 2.6). As a result of intra-genic/antisense induction and genic repression, expression levels in *spt6-1004* of all classes of transcripts become similar to one another (Figure 2.8).

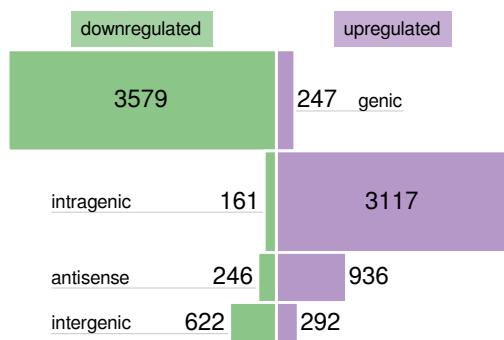


Figure 2.6: Bar plots of the number of TSS-seq peaks differentially expressed in *spt6-1004* after 80 minutes at 37°C versus wild-type after 80 minutes at 37°C. The height of each bar is proportional to the total number of peaks in the category, including those not found to be significantly differentially expressed.

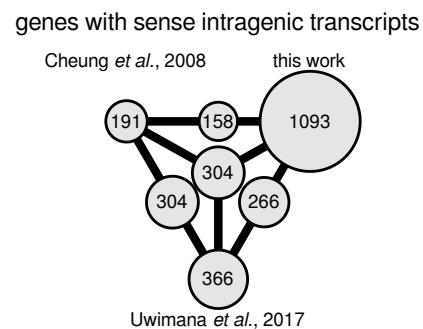


Figure 2.7: Set diagram of the number of genes reported to have *spt6-1004*-induced intragenic transcripts using tiled arrays (Cheung et al., 2008), RNA-seq (Uwimana et al., 2017), and TSS-seq (this work).

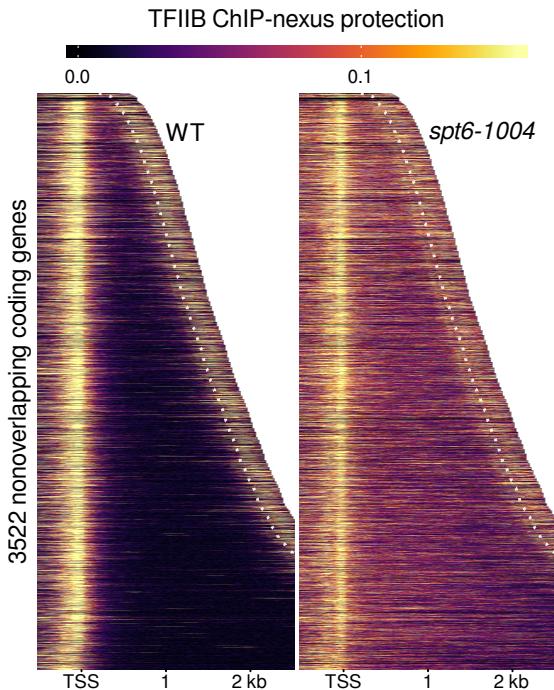


Figure 2.5: Heatmaps of TFIIB binding measured by ChIP-nexus, over the same regions shown in Figure 2.4. Values are the mean of library-size normalized coverage in non-overlapping 20 bp bins, averaged over two replicates. Values above the 85th percentile are set to the 85th percentile for visualization.

The changes in transcript levels in *spt6-1004* observed by TSS-seq correspond with substantial differences in the pattern of TFIIB binding on the genome. In contrast to the discrete peaks in promoter regions seen in wild-type, TFIIB in *spt6-1004* binds much more promiscu-

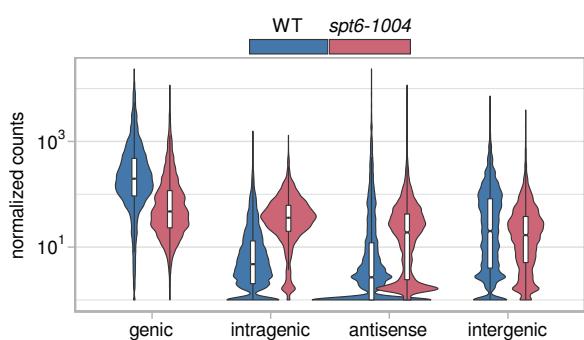


Figure 2.8: Violin plots of expression level distributions for genomic classes of TSS-seq peaks in wild-type and *spt6-1004*, both after 80 minutes at 37°C. Normalized counts are the mean of spike-in size factor

ously, with many loci having TFIIB signal spread over broad regions of the genome (Figure 2.9). This difference in binding pattern makes peak calling ineffective for quantifying TFIIB signal in this case: ChIP-seq peak callers generally use different algorithms for calling ‘narrow’ peaks (e.g. for sequence-specific transcription factors) and ‘broad’ peaks (e.g. for histone modifications), meaning that a single algorithm is unable to call peaks that are meaningful for differential binding analyses between wild-type and *spt6-1004*. Therefore, to see if changes in transcript levels in *spt6-1004* correspond to changes in transcription initiation, we compared the change in TSS-seq signal at TSS-seq peaks in *spt6-1004* to the change in TFIIB ChIP-nexus signal in the window extending 200 bp upstream of the TSS-seq peak. Changes in TSS-seq signal in *spt6-1004* are associated with a change in TFIIB signal of the same sign at over 82% of TSSs of any genomic class, indicating that the increase in intragenic transcript levels and decrease in genic transcript levels observed in *spt6-1004* are in large part explained by changes in transcription initiation.

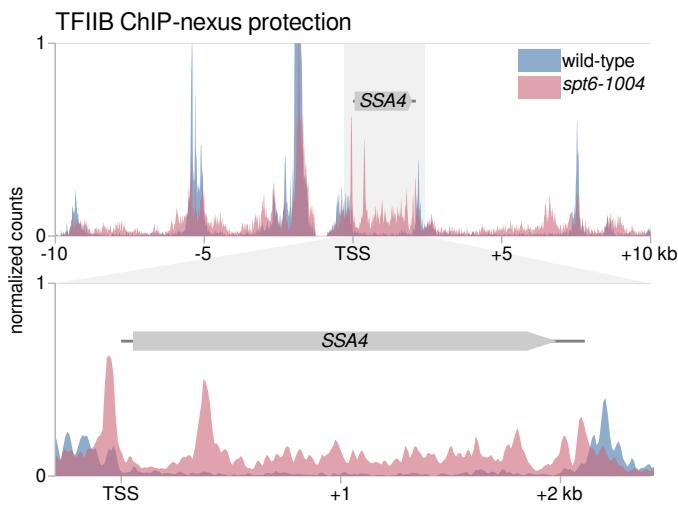


Figure 2.9:

- top) TFIIB ChIP-nexus protection in wild-type and *spt6-1004*, over 20 kb of chromosome II flanking the *SSA4* gene.
- bottom) Expanded view of TFIIB protection over the *SSA4* gene.

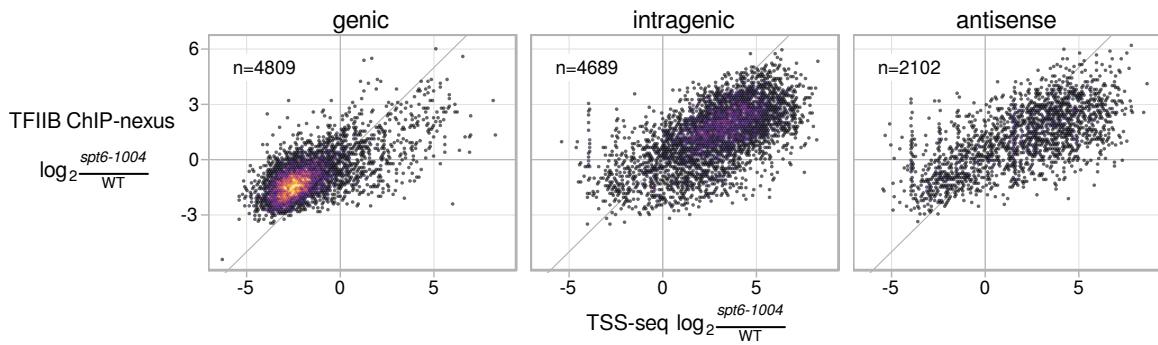


Figure 2.10: Scatterplots of fold-change in *spt6-1004* over wild-type, comparing TSS-seq and TFIIB ChIP-nexus. Each dot represents a TSS-seq peak paired with the window extending 200 bp upstream of the TSS-seq peak summit for quantification of TFIIB ChIP-nexus signal. Fold-changes are regularized fold-change estimates from DESeq2, with size factors determined from the *S. pombe* spike-in (TSS-seq), or *S. cerevisiae* counts (ChIP-nexus).

2.5 MNase-seq results from *spt6-1004*

Because a primary function of Spt6 is to act as histone chaperone that reassembles nucleosomes in the wake of transcription (Duina, 2011), it is reasonable to expect that the transcriptional changes seen in *spt6-1004* would be associated with changes in chromatin structure. The requirement for Spt6 in maintaining normal chromatin structure has been demonstrated in previous studies (Bortvin and Winston, 1996; Ivanovska et al., 2011; Jeronimo et al., 2015; Kaplan et al., 2003; Perales et al., 2013; van Bakel et al., 2013). To re-examine this requirement in higher resolution, we assayed nucleosome protection genome-wide using micrococcal nuclease digestion of chromatin followed by sequencing (MNase-seq).

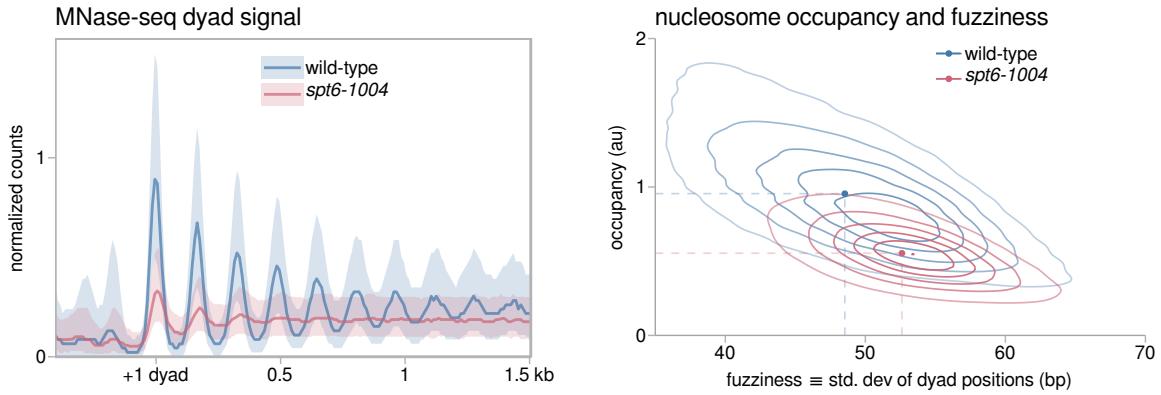


Figure 2.11: Average MNase-seq dyad signal in wild-type and *spt6-1004*, over 3522 non-overlapping genes aligned by wild-type +1 nucleosome dyad. Values are the mean of spike-in normalized coverage in non-overlapping 20 bp bins, averaged over two replicates (*spt6-1004*) or one experiment (wild-type). The solid line and shading are the median and the inter-quartile range.

Figure 2.12: Contour plot of the global distributions of nucleosome occupancy and fuzziness in wild-type and *spt6-1004*. Dashed lines indicate median values.

In wild-type, the MNase-seq data recapitulate the expected signature over genes, with a nucleosome-depleted region upstream of a strongly positioned ‘+1’ nucleosome, and a regularly phased array of nucleosomes over the gene body (Figure 2.11). In *spt6-1004*, nucleosome signal is severely reduced at canonical nucleosome positions and spreads into inter-nucleosome regions. Changes in aggregate nucleosome signal such as those observed in Figure 2.11 are the combination of changes to nucleosome occupancy (the number of reads assigned to a nucleosome), fuzziness (the standard deviation of read positions for a nucleosome), and position (the coordinate with the maximum reads for a nucleosome) (Chen et al., 2013). Using DANPOS2

(Chen et al., 2013), we called nucleosome positions and quantified these metrics for wild-type and *spt6-1004*. Wild-type nucleosomes span a relatively wide range of occupancy and fuzziness space, with highly occupied nucleosomes tending to be less fuzzy (i.e., more well-positioned) (Figure 2.12). In *spt6-1004*, the population of nucleosomes is much more homogeneous: nucleosome occupancy is decreased globally, and nucleosome fuzziness is restricted to the high end of the wild-type distribution.

Previous studies observed two trends: 1) In wild-type cells, nucleosome positioning is weaker over highly transcribed genes than over moderately transcribed genes (Shivaswamy et al., 2008), and 2) In *spt6-1004* cells, the decrease in nucleosome occupancy is greater for highly transcribed genes (Ivanovska et al., 2011). To re-examine these trends, we looked at the MNase-seq data in the context of NET-seq data, which reports the position of actively transcribing RNAPII and reflects a gene's level of transcription (Figure 2.13) (Churchman and Weissman, 2012). The data support the first trend: in wild-type, genes with the strongest NET-seq signal have decreased MNase-seq signal. However, there is no obvious relationship between transcription level and the nucleosome changes observed in *spt6-1004* (Figure 2.13). The apparent discrepancy might be explained by the improved resolution and breadth of MNase-seq versus the MNase and microarray of chromosome III used in the previous study (Ivanovska et al., 2011).

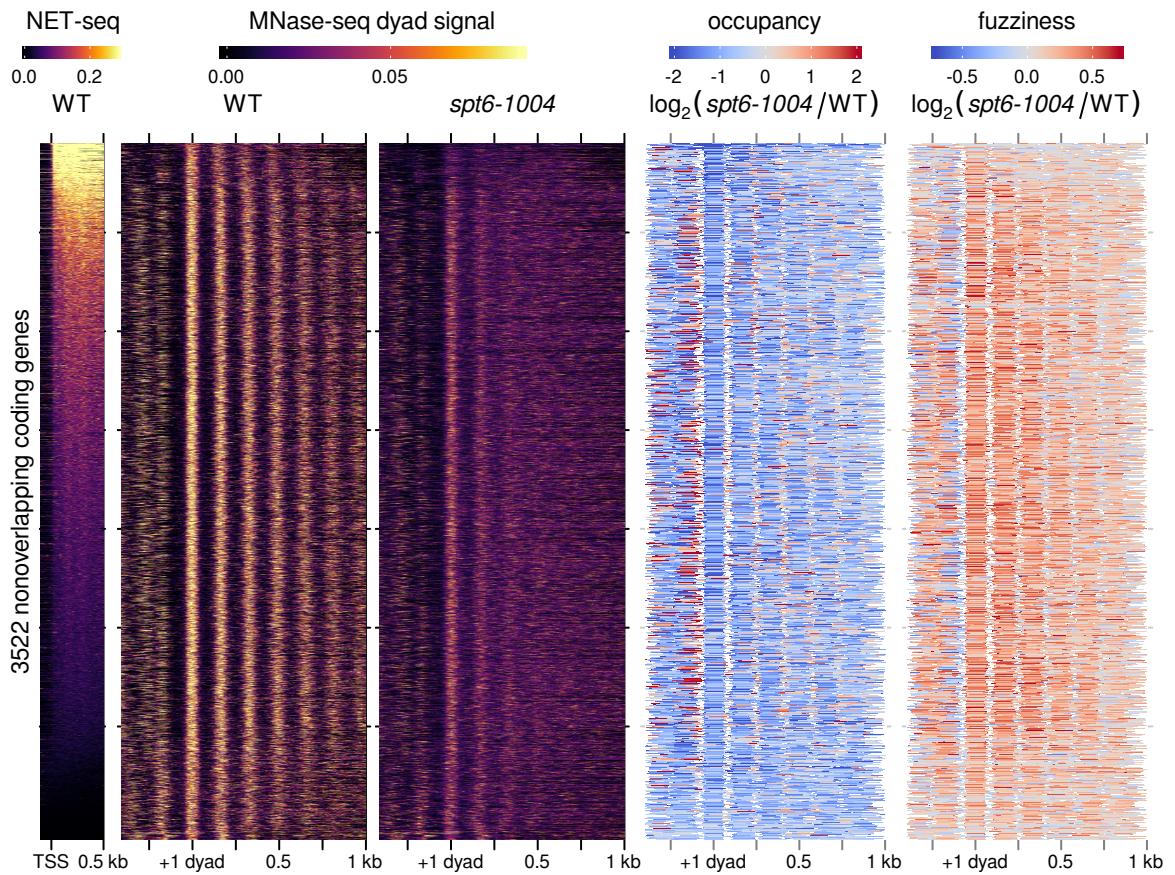


Figure 2.13:

- left) Heatmap of sense strand NET-seq signal for 3522 non-overlapping genes, aligned by genic TSS and sorted by total sense strand NET-seq signal in the window extending 500 nt downstream from the genic TSS. Values are the mean of library-size normalized coverage in non-overlapping 20 nt bins, averaged over two replicates.
- middle) Heatmaps of MNase-seq dyad signal in wild-type and *spt6-1004* for the same genes, aligned by wild-type +1 nucleosome dyad and arranged by sense NET-seq signal as in the leftmost panel. Values are the mean of spike-in normalized coverage in non-overlapping 20 bp bins, averaged over two replicates (*spt6-1004*) or one experiment (wild-type).
- right) Heatmaps of fold-change in nucleosome occupancy and fuzziness for the same genes, aligned by wild-type +1 nucleosome dyad and arranged by sense NET-seq signal as in the leftmost panel.

2.5.1 Clustering of MNase-seq profiles at *spt6-1004*-induced intragenic TSSs

The aggregate MNase-seq dyad signal around all *spt6-1004* intragenic TSSs is aperiodic (Figure 2.15, top left panel), which occurs as a result of destructive interference from offset nucleosome phasing patterns. To discover these phasing patterns, we used the wild-type and *spt6-1004* MNase-seq data flanking intragenic TSSs to train a self-organizing map to assign TSSs with similar MNase-seq patterns to nearby nodes in a rectangular grid (Figure 2.14). This allowed us to see that, although there is considerable diversity in the nucleosome pattern surrounding intragenic TSSs, most intragenic TSSs occur in areas between the positions of nucleosome dyads. By hierarchically clustering the ‘code vectors’ representing each node of the self-organizing map, we further grouped intragenic TSSs into three major clusters differing primarily by the phasing of the nucleosome array relative to the TSS, as shown in Figure 2.15.

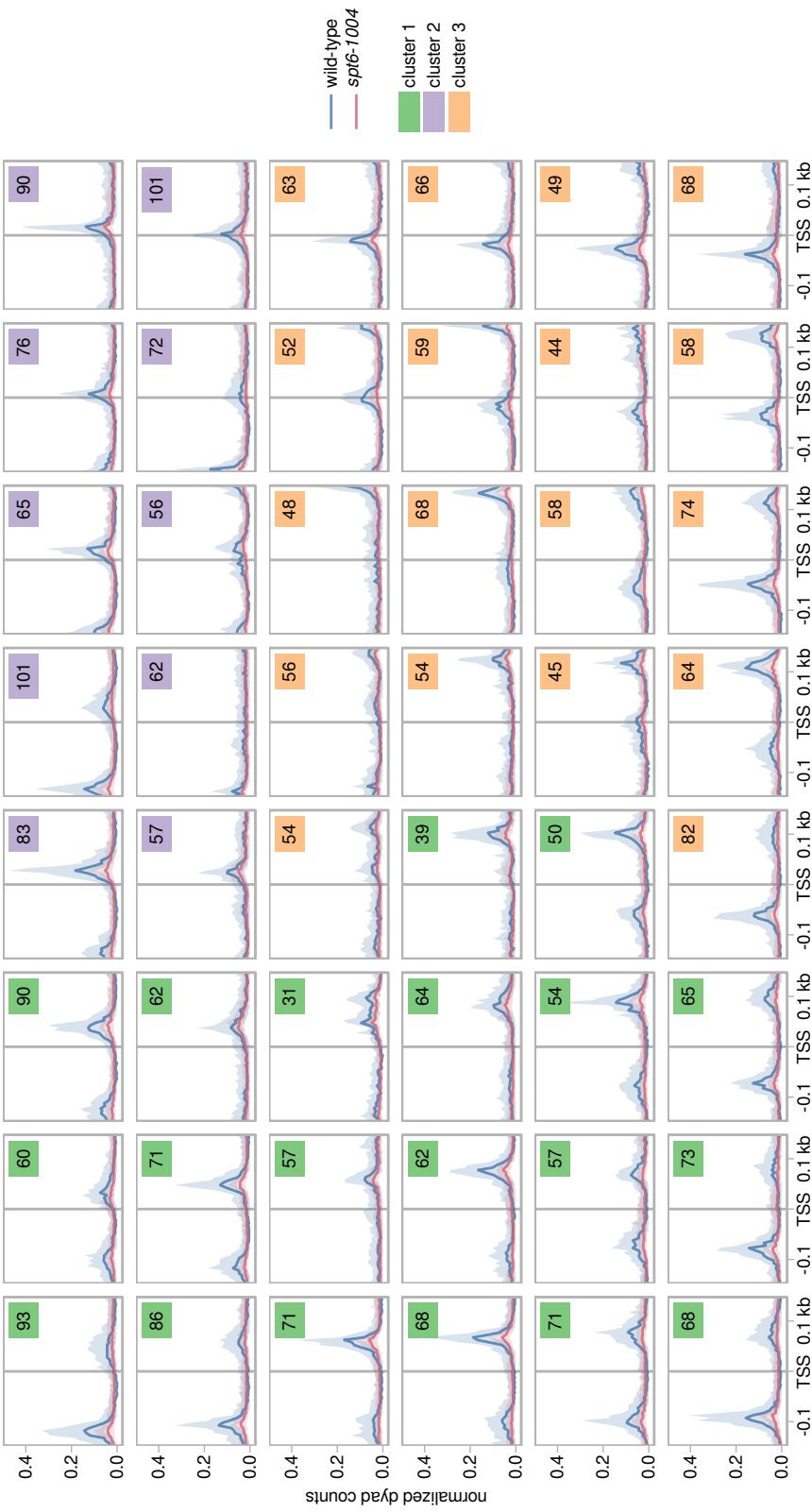


Figure 2.14: Average MNase-seq dyad signal around all *spt6-1004*-induced intragenic TSSs, grouped by assignment to nodes of a 6x8 super-organizing map (SOM). The number of TSSs assigned to each node is shown in the upper right of each panel, and is shaded by the node's assignment to a cluster determined by agglomerative hierarchical clustering of the nodes. The solid line and shading are the median and inter-quartile range of the mean spike-in normalized coverage over two replicates (*spt6-1004*) or one experiment (wild-type), in non-overlapping 5 bp bins.

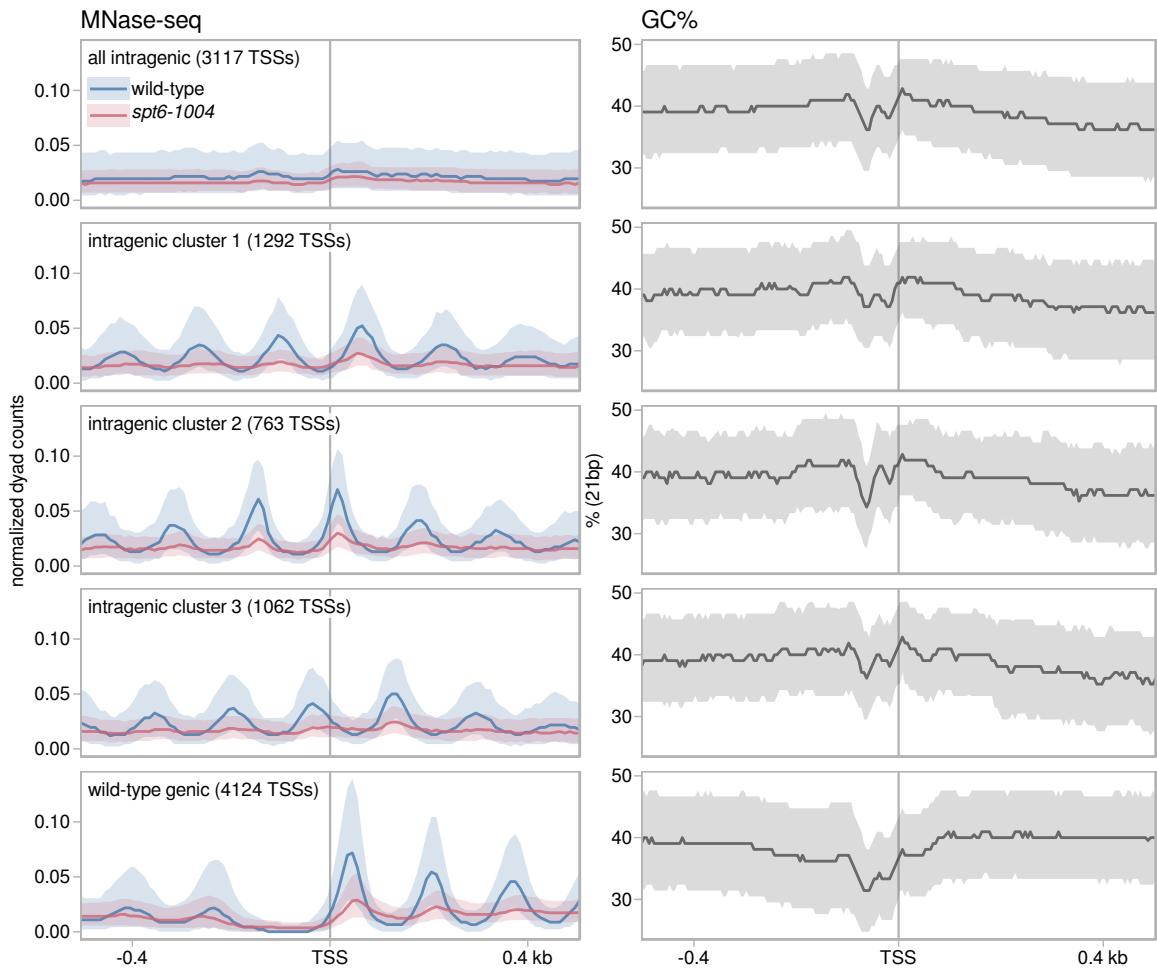


Figure 2.15:

- left column) Average MNase-seq dyad signal for *spt6-1004* intragenic TSSs, both aggregated and grouped into three clusters by the wild-type and *spt6-1004* MNase-seq dyad signal flanking the TSS, as well as all genic TSSs detected in wild-type. Values are the mean of spike-in normalized dyad coverage in non-overlapping 10 bp bins, averaged over two replicates (*spt6-1004*) or one experiment (wild-type). The solid line and shading are the median and inter-quartile range.
- right column) Average GC content of the DNA sequence in a 21 bp window, as above.

2.6 Other features of *spt6-1004* intragenic promoters

The resolution with which we were able to identify intragenic TSSs allowed us to closely examine their sequence features and compare them to genic TSSs.

2.6.1 Information content and sequence preference of intragenic TSSs

To examine the DNA sequence preference of intragenic and genic TSSs in *spt6-1004*, we aligned the sequences of all TSS-seq reads overlapping TSS-seq peaks of each class, and calculated the information content and sequence distribution for each class (Figure 2.16). Intragenic TSSs have a sequence preference almost identical to previously observed sequence preference of genic TSSs (Malabat et al., 2015), suggesting that RNA polymerase initiates transcription similarly at genic and intragenic TSSs, and that the lack of intragenic initiation in wild-type is due to inaccessibility of the initiation sequence, possibly due to histones.

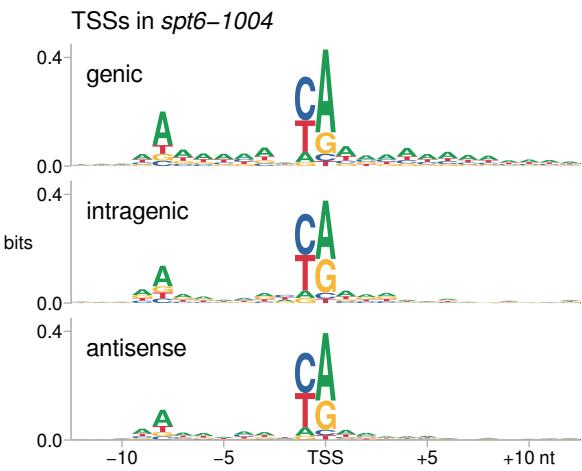


Figure 2.16: Sequence logos depicting information content and sequence preference of TSS-seq reads overlapping genic and intragenic TSS-seq peaks in *spt6-1004*.

2.6.2 Sequence motifs enriched at intragenic TSSs

To examine whether sequence-specific transcription factors contribute to the expression of intragenic transcripts in *spt6-1004*, we looked for enrichment or depletion of

the DNA sequence motifs associated with these factors upstream of intragenic TSSs. Exact matches to the TATA element consensus sequence TATAWAWR are enriched upstream between 100 and 150 nt upstream of intragenic TSSs, in the same position but to a lesser degree than the TATA enrichment observed upstream of genic TSSs (Figure 2.17). This further supports the model that *spt6-1004* intragenic promoters are sequences similar to canonical genic promoters, which become accessible for transcription initiation when the normal chromatin state is disturbed.

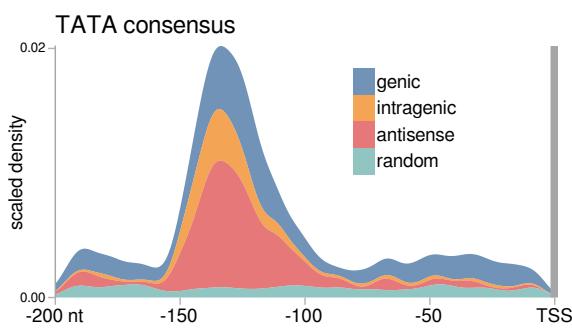


Figure 2.17: Scaled density of occurrences of exact matches to the motif TATAWAWR upstream of TSSs. For each category, a Gaussian kernel density estimate of the positions of motif occurrences is scaled by the number of motif occurrences per region.

2.7 Summary

2.8 Methods

2.8.1 Yeast strain construction and growth conditions

All yeast strains were constructed by standard yeast transformation or crosses. The *spt6-1004* and wild-type strains were grown as previously described (Cheung et al., 2008): Cells were grown in YPD at 30 °C to a concentration of approximately 1×10^7 cells/ml ($\text{OD}_{600} = 0.6$), at which point an equal volume of YPD medium pre-warmed to 44 °C was added, and the cultures were shifted to 37 °C for 80 minutes.

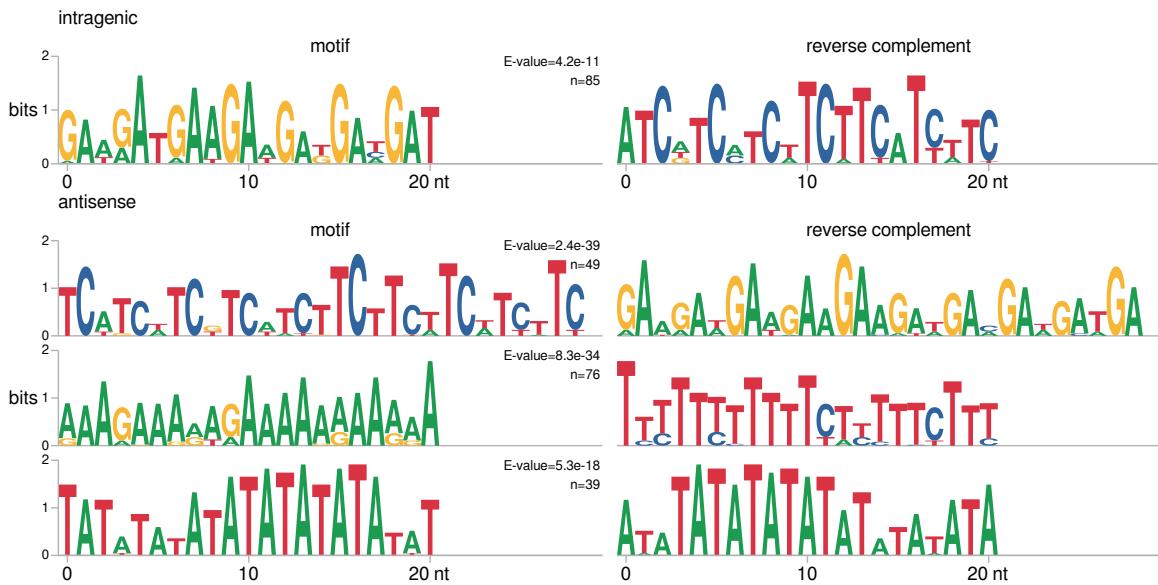


Figure 2.18: Sequence logos of motifs discovered by MEME (Bailey et al., 2015) in the window -100 to +30 bp relative to *spt6-1004* intragenic and antisense TSSs. The number of motif occurrences and the E-value, indicating the expected number of motif occurrences if the input sequences were scrambled, are shown for each motif.

2.8.2 Sequencing library preparation (TSS-seq, ChIP-nexus, MNase-seq, NET-seq)

All library preparation methods are detailed in Doris et al. (2018).

2.8.3 Genome builds

The genome build used for *S. cerevisiae* was R64-2-1. The genome build used for *S. pombe* was ASM294v2.

2.8.4 TSS-seq data analysis

An up-to-date version of the Snakemake (Köster and Rahmann, 2012) workflow used to process TSS-seq libraries is maintained at github.com/winston-lab/tss-seq. At the

time of writing, removal of adapter sequences and random hexamer sequences from the 3' end of the read and 3' quality trimming were performed using cutadapt (Martin, 2011). The random hexamer molecular barcode on the 5' end of the read was then removed and processed using a custom Python script (adapted from ?). Reads were aligned to the combined *S. cerevisiae* and *S. pombe* reference genomes using Tophat2 (Kim et al., 2013) without a reference transcriptome, and uniquely mapping reads were selected using SAMtools (?). Reads mapping to the same location as another read with the same molecular barcode were identified as PCR duplicates and removed using a custom Python script (adapted from ?). Coverage of the 5'-most base, corresponding to the TSS, was extracted using bedtools genomecov (Quinlan and Hall, 2010) and normalized to the total number of reads uniquely mapping to the *S. pombe* genome. Quality statistics of raw, cleaned, non-aligning, and uniquely aligning non-duplicate reads were assessed using FastQC (?).

The pipeline additionally performs TSS-seq peak calling, differential expression, classification of peaks into genome categories, gene ontology analysis, motif enrichment analysis, *de novo* motif discovery, sequence logo visualization, and data visualization with the option to separate data into clusters of similar signal.

2.9 Bibliography

- Arribere, J. A. and Gilbert, W. V. (2013). Roles for transcript leaders in translation and mrna decay revealed by transcript leader sequencing. *Genome Research*, 23(6):977–987. 2.2
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The meme suite. *Nucleic Acids Research*, 43(W1):W39–W49. 2.18
- Bortvin, A. and Winston, F. (1996). Evidence that spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267):1473–1476. 2.5
- Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). Danpos: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Research*, 23(2):341–351. 2.5
- Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *saccharomyces cerevisiae* genome. *PLOS Biology*, 6(11):1–13. (document), 2.2, 2.2, 2.7, 2.8.1
- Churchman, L. S. and Weissman, J. S. (2012). Native elongating transcript sequencing (net-seq). *Current Protocols in Molecular Biology*, 98(1):14.4.1–14.4.17. 2.5
- DeGennaro, C. M., Alver, B. H., Marguerat, S., Stepanova, E., Davis, C. P., Bähler, J., Park, P. J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Molecular and Cellular Biology*, 33(24):4779–4792. 2.2
- Doris, S. M., Chuang, J., Viktorovskaya, O., Murawska, M., Spatt, D., Churchman, L. S., and Winston, F. (2018). Spt6 is required for the fidelity of promoter selection. *bioRxiv*. 2.8.2
- Duina, A. A. (2011). Histone chaperones spt6 and fact: Similarities and differences in modes of action at transcribed genes. *Genet Res Int*, 2011:625210. 22567361[pmid]. 2.5
- He, Q., Johnston, J., and Zeitlinger, J. (2015). Chip-nexus enables improved detection of in vivo transcription factor binding footprints. *Nature Biotechnology*, 33:395 EP –. 2.2
- Ivanovska, I., Jacques, P.-♦., Rando, O. J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: Different consequences in coding and regulatory regions. *Molecular and Cellular Biology*, 31(3):531–541. 2.5, 2.5

- Jeronimo, C., Watanabe, S., Kaplan, C., Peterson, C., and Robert, F. (2015). The histone chaperones fact and spt6 restrict h2a.z from intragenic locations. *Molecular Cell*, 58(6):1113 – 1123. 2.5
- Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–1099. 2.2, 2.5
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36. 2.8.4
- Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522. 2.8.4
- Lickwar, C. R., Rao, B., Shabalin, A. A., Nobel, A. B., Strahl, B. D., and Lieb, J. D. (2009). The set2/rpd3s pathway suppresses cryptic transcription without regard to gene length or transcription frequency. *PLOS ONE*, 4(3):1–7. 2.2
- Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mrna decay. *eLife*, 4:e06722. 2.2, 2.6.1
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12. 2.8.4
- Pathak, R., Singh, P., Ananthakrishnan, S., Adamczyk, S., Schimmel, O., and Govind, C. K. (2018). Acetylation-dependent recruitment of the fact complex and its role in regulating pol ii occupancy genome-wide in saccharomyces cerevisiae. *Genetics*, 209(3):743–756. 2.2
- Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. *The EMBO Journal*, 32(19):2645–2656. 2.5
- Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842. 2.8.4
- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V. R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLOS Biology*, 6(3):1–13. 2.5
- Uwimana, N., Collin, P., Jeronimo, C., Haibe-Kains, B., and Robert, F. (2017). Bidirectional terminators in saccharomyces cerevisiae prevent cryptic transcription

from invading neighboring genes. *Nucleic Acids Research*, 45(11):6417–6426. (document), 2.2, 2.2, 2.7

van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T. R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. *PLOS Genetics*, 9(5):1–18. 2.2, 2.5

Chapter 3

Genomics of transcription elongation factor Spt5

3.1 Collaborators

Ameet Shetty generated TSS-seq, MNase-seq, NET-seq, RNA-seq, and ChIP-seq libraries

3.2 Introduction to Spt5 and prior work

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer

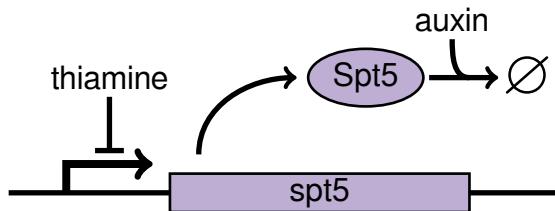


Figure 3.1: Diagram of the dual-shutoff system used to deplete Spt5 from *S. pombe*. Spt5 is expressed from a thiamine-repressible promoter, and tagged with an auxin-inducible degron tag for specific degradation upon addition of auxin.

sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed grida sollic-

itudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis,

fermentum vel, eleifend faucibus, vehicula eu, lacus.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus

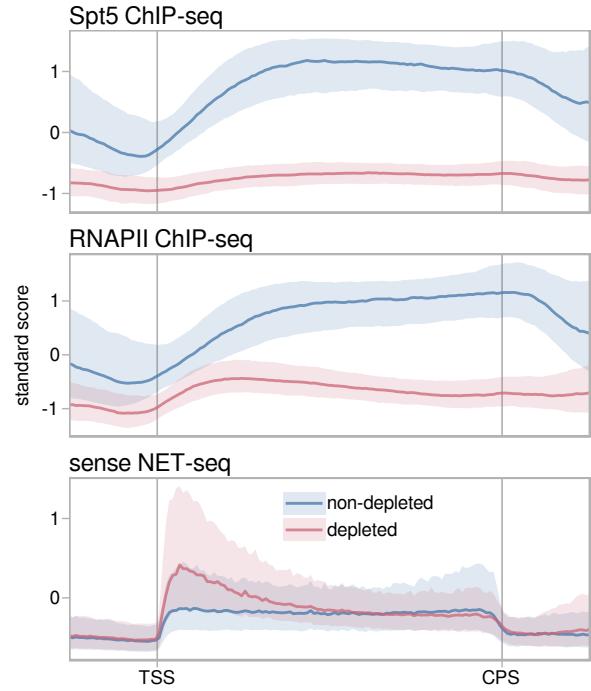
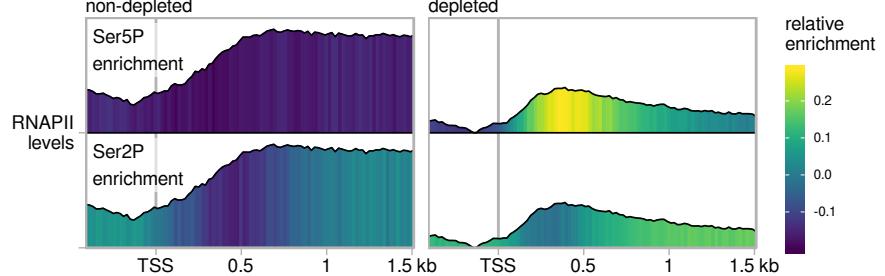


Figure 3.2: Average Spt5 ChIP-seq, RNAPII ChIP-seq, and sense NET-seq signal in Spt5 non-depleted and depleted cells, over 1989 non-overlapping coding genes scaled from TSS to CPS. The solid



libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla.

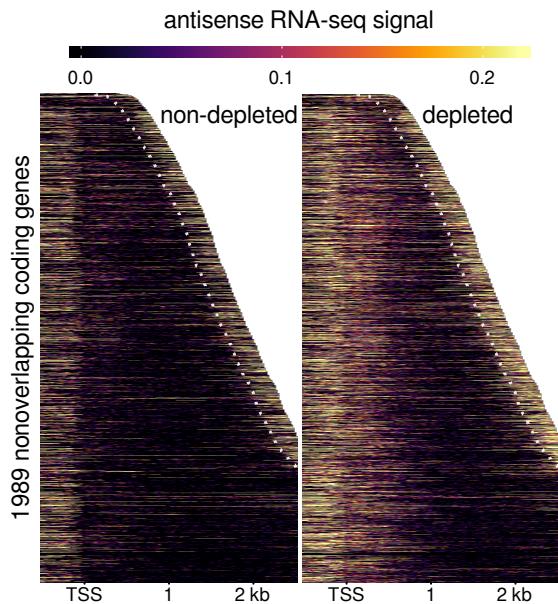


Figure 3.4: Caption wsdasdr zzzz.

Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

3.3 An aside on spike-in normalization for ChIP-seq

3.4 TSS-seq results from Spt5 depletion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec

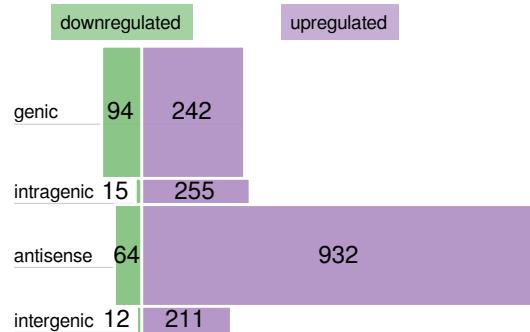


Figure 3.5: Caption wsdasdr zzzz.

Figure 3.6: Caption wsdasdr zzzz.

vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.5 MNase-seq results from Spt5 depletion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi do-

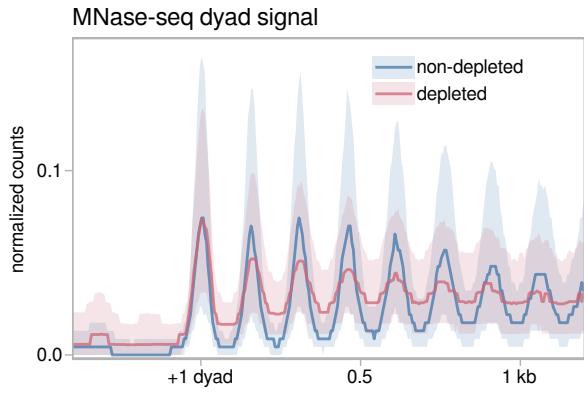


Figure 3.7: Caption wsdasdr zzzz.

Figure 3.8: Caption wsdasdr zzzz.

Figure 3.9: Caption wsdasdr zzzz.

lor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.5.1 MNase-seq profile at Spt5-depletion-induced antisense TSSs

3.6 Sequence motifs enriched at antisense TSSs

3.7 Summary

3.8 Bibliography

Chapter 4

Stress-responsive intragenic transcription

4.1 Collaborators

Steve Doris generated TSS-seq and ChIP-nexus libraries

Dan Spatt polyribosome fractionation, fitness competitions,
and other experiments

James Warner fitness competitions and other experiments

4.2 Possible functions for intragenic transcription in wild-type cells

4.3 Discovery of stress-induced intragenic promoters by TFIIB ChIP-nexus and TSS-seq

4.4 Chromatin landscape of oxidative-stress-induced promoters.

4.5 Polysome enrichment of oxidative-stress-induced intragenic transcripts

4.6 TSS-seq analysis of oxidative stress in *Saccharomyces sensu stricto* species

4.7 Functions of intragenic DSK2 expression in oxidative stress

4.8 Summary

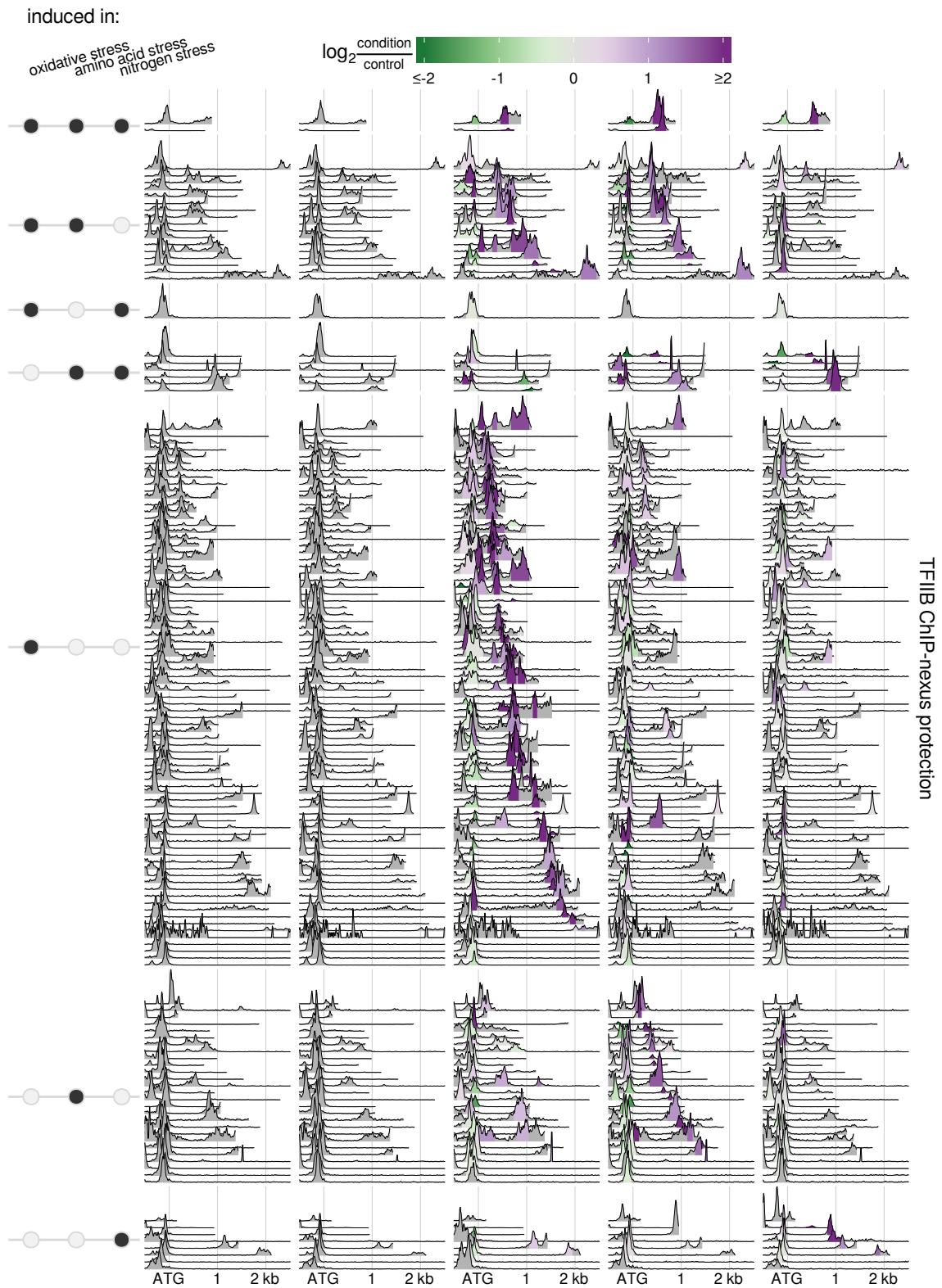


Figure 4.1: Wasldfkjlk asldkfj.

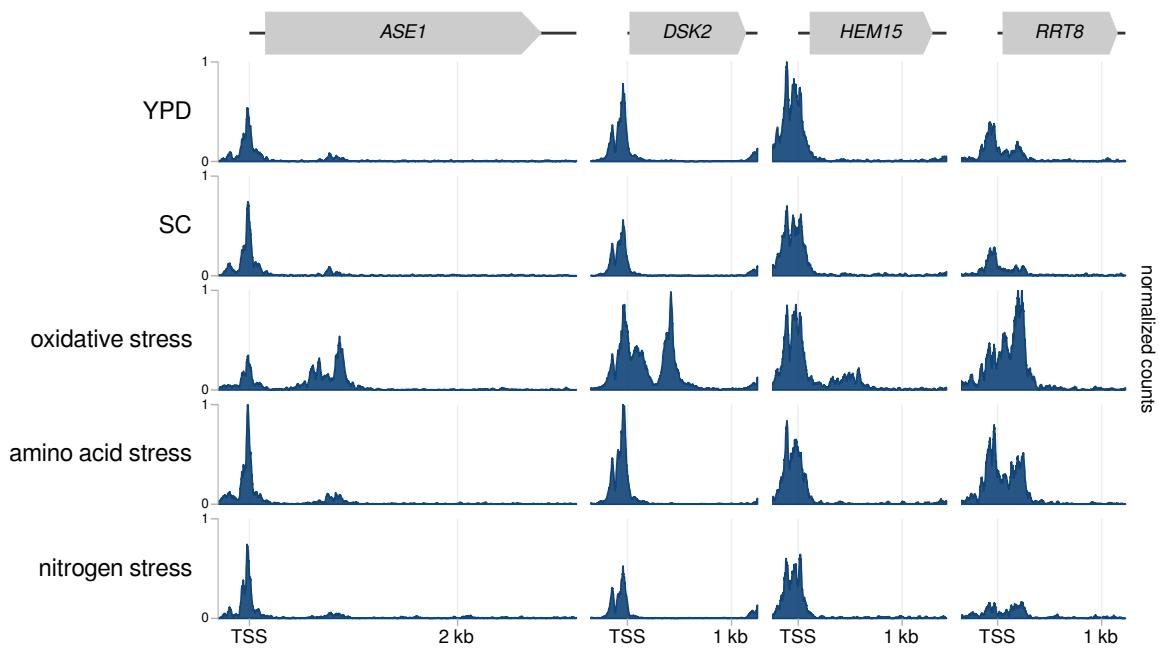


Figure 4.2: Caption asdflkj asldkfjlkj.

4.9 Bibliography

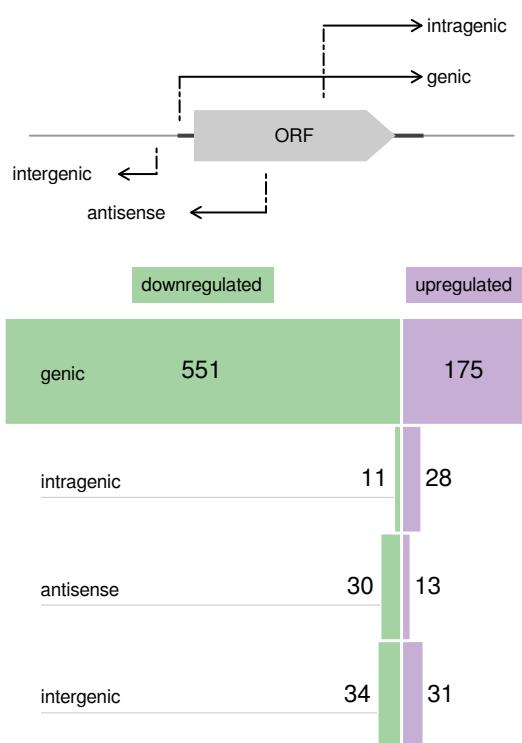


Figure 4.3: Caption dsafklj asldkfjlkj.

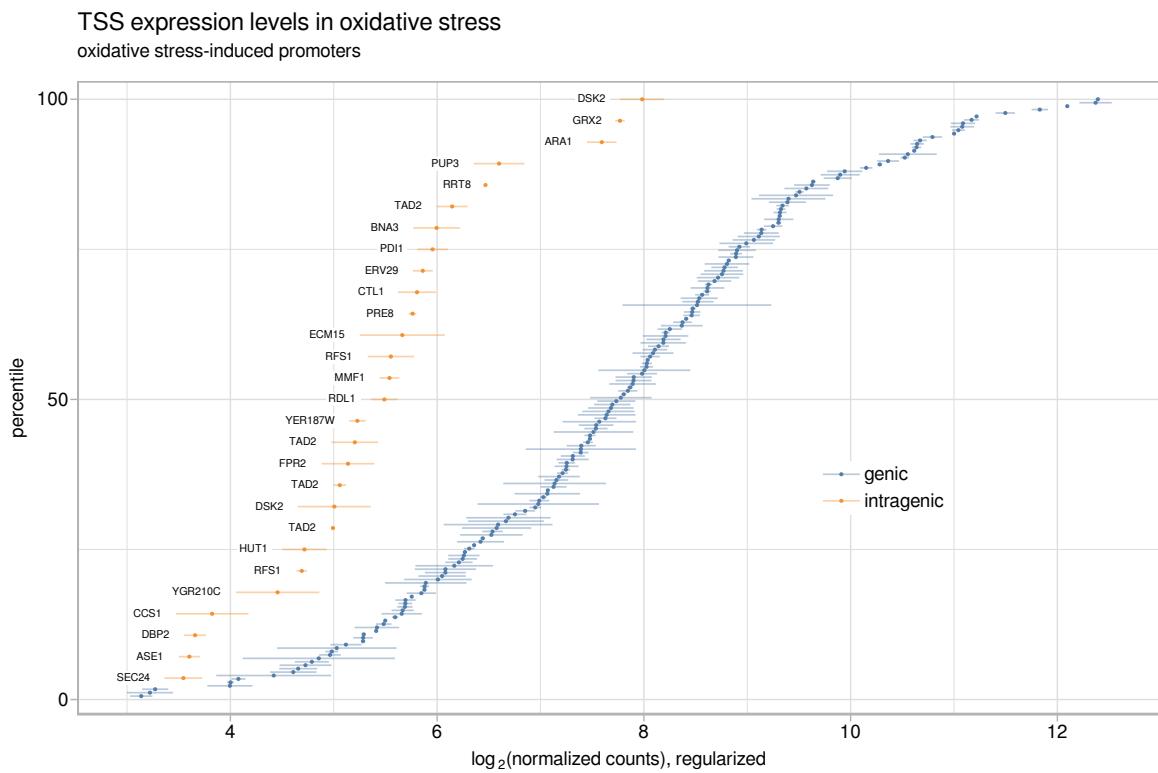


Figure 4.4: Caption dsafklj zzzz.

Figure 4.5: Caption dsafklj .

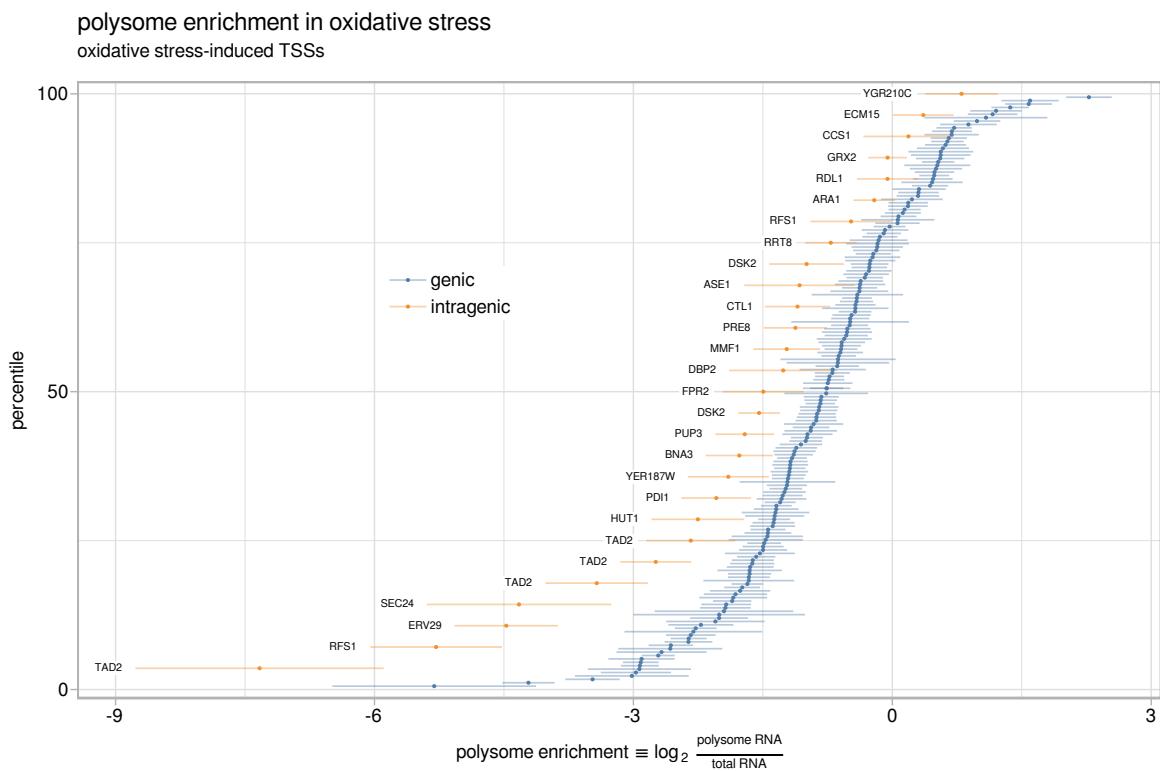


Figure 4.6: Caption wsadasdr zzzz.

Figure 4.7: Caption dsafklj .

Figure 4.8: Caption dsafklj .

Figure 4.9: Caption dsafklj .

Figure 4.10: Caption dsafklj .

Bibliography

- Arribere, J. A. and Gilbert, W. V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Research*, 23(6):977–987.
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The meme suite. *Nucleic Acids Research*, 43(W1):W39–W49.
- Bortvin, A. and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267):1473–1476.
- Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). Danpos: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Research*, 23(2):341–351.
- Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLOS Biology*, 6(11):1–13.
- Churchman, L. S. and Weissman, J. S. (2012). Native elongating transcript sequencing (net-seq). *Current Protocols in Molecular Biology*, 98(1):14.4.1–14.4.17.
- DeGennaro, C. M., Alver, B. H., Marguerat, S., Stepanova, E., Davis, C. P., Bähler, J., Park, P. J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Molecular and Cellular Biology*, 33(24):4779–4792.
- Doris, S. M., Chuang, J., Viktorovskaya, O., Murawska, M., Spatt, D., Churchman, L. S., and Winston, F. (2018). Spt6 is required for the fidelity of promoter selection. *bioRxiv*.
- Duina, A. A. (2011). Histone chaperones Spt6 and FACT: Similarities and differences in modes of action at transcribed genes. *Genet Res Int*, 2011:625210. 22567361[pmid].

- He, Q., Johnston, J., and Zeitlinger, J. (2015). Chip-nexus enables improved detection of *in vivo* transcription factor binding footprints. *Nature Biotechnology*, 33:395 EP –.
- Ivanovska, I., Jacques, P.-♦., Rando, O. J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: Different consequences in coding and regulatory regions. *Molecular and Cellular Biology*, 31(3):531–541.
- Jeronimo, C., Watanabe, S., Kaplan, C., Peterson, C., and Robert, F. (2015). The histone chaperones fact and spt6 restrict h2a.z from intragenic locations. *Molecular Cell*, 58(6):1113 – 1123.
- Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–1099.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36.
- Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522.
- Lickwar, C. R., Rao, B., Shabalin, A. A., Nobel, A. B., Strahl, B. D., and Lieb, J. D. (2009). The set2/rpd3s pathway suppresses cryptic transcription without regard to gene length or transcription frequency. *PLOS ONE*, 4(3):1–7.
- Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mrna decay. *eLife*, 4:e06722.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12.
- Pathak, R., Singh, P., Ananthakrishnan, S., Adamczyk, S., Schimmel, O., and Govind, C. K. (2018). Acetylation-dependent recruitment of the fact complex and its role in regulating pol ii occupancy genome-wide in *saccharomyces cerevisiae*. *Genetics*, 209(3):743–756.
- Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. *The EMBO Journal*, 32(19):2645–2656.
- Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V. R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLOS Biology*, 6(3):1–13.
- Uwimana, N., Collin, P., Jeronimo, C., Haibe-Kains, B., and Robert, F. (2017). Bidirectional terminators in *Saccharomyces cerevisiae* prevent cryptic transcription from invading neighboring genes. *Nucleic Acids Research*, 45(11):6417–6426.
- van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T. R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. *PLOS Genetics*, 9(5):1–18.

Vita

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tin-

cidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellen-
tesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam.
Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia.
Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula
feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pel-
lentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu
purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi.
Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tin-
cidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac
habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc
elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollic-
itudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor.
Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus
semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Ali-
quam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hen-
drerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum
porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in
dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris
tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla.
Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus
vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet
vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie

non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.