

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
Introduction	1
A brief introduction to transcription	1
Transcription elongation factors Spt6 and Spt5	1
Reproducible data analysis for genomics	1
1 Genomics of transcription elongation factor Spt6	2
1.1 Collaborators	2
1.2 Introduction to Spt6 and intragenic transcription	2
1.3 Data analysis pipelines for TSS-seq and ChIP-nexus	5
1.3.1 TSS-seq peak calling	5
1.3.2 A note on ChIP-nexus peak calling	5
1.4 TSS-seq and TFIIB ChIP-nexus results for <i>spt6-1004</i>	5
1.5 MNase-seq results from <i>spt6-1004</i>	11
1.5.1 Clustering of MNase-seq profiles at <i>spt6-1004</i> -induced intra- genic TSSs	15
1.6 Other features of <i>spt6-1004</i> intragenic promoters	17
1.6.1 Information content of intragenic TSSs	17

1.6.2	Sequence motifs enriched at intragenic TSSs	18
1.7	Summary	19
1.8	Bibliography	20
2	Genomics of transcription elongation factor Spt5	21
2.1	Collaborators	21
2.2	Introduction to Spt5 and prior work	21
2.3	An aside on spike-in normalization for ChIP-seq	21
2.4	TSS-seq results from Spt5 depletion	21
2.5	MNase-seq results from Spt5 depletion	21
2.5.1	MNase-seq profile at Spt5-depletion-induced antisense TSSs .	21
2.6	Sequence motifs enriched at antisense TSSs	21
2.7	Summary	21
2.8	Bibliography	23
3	Stress-responsive intragenic transcription	24
3.1	Collaborators	24
3.2	Possible functions for intragenic transcription in wild-type cells	24
3.3	Discovery of stress-induced intragenic promoters by TFIIB ChIP-nexus and TSS-seq	24
3.4	Chromatin landscape of oxidative-stress-induced promoters.	24
3.5	Polysome enrichment of oxidative-stress-induced intragenic transcripts	24
3.6	TSS-seq analysis of oxidative stress in <i>Saccharomyces sensu stricto</i> species	24
3.7	Functions of intragenic DSK2 expression in oxidative stress	24
3.8	Summary	24

3.9 Bibliography	26
Bibliography	30

List of Tables

List of Figures

1.1	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.	2
1.2	Diagram of transcript classes.	2
1.3	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.	3
1.4	Heatmaps of sense and antisense TSS-seq signal from wild-type and <i>spt6-1004</i> cells, over non-overlapping coding genes.	6
1.6	Bar plot of the number of TSS-seq peaks of various genomic classes differentially expressed in <i>spt6-1004</i> versus wild-type.	7
1.7	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.	7
1.5	Heatmaps of TFIIB ChIP-nexus protection from wild-type and <i>spt6-1004</i> cells, over non-overlapping coding genes	8
1.8	Violin plots of expression level distributions for genomic classes of TSS-seq peaks in wild-type and <i>spt6-1004</i> cells.	8
1.9	TFIIB ChIP-nexus protection over the 20 kb flanking the gene <i>SSA4</i> , in wild-type and <i>spt6-1004</i> cells.	9

1.10 Scatterplots of fold-change in <i>spt6-1004</i> over wild-type, comparing TSS-seq and TFIIB ChIP-nexus.	10
1.11 Average MNase-seq dyad signal in wild-type and <i>spt6-1004</i> , over non-overlapping genes aligned by wild-type +1 nucleosome dyad.	11
1.12 Contour plot of nucleosome occupancy and fuzziness in wild-type and <i>spt6-1004</i>	11
1.13 Average wild-type and <i>spt6-1004</i> MNase-seq dyad signal around all <i>spt6-1004</i> -induced intragenic TSSs, grouped by a self-organizing map of the MNase-seq signal.	13
1.14 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
1.15 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 genic TSSs.	16
1.16 Sequence logos of TSS-seq reads overlapping genic and intragenic TSS-seq peaks in <i>spt6-1004</i>	17
1.17 Kernel density estimate of matches to a consensus TATA-box motif upstream of genic and <i>spt6-1004</i> -induced intragenic TSSs.	18
1.18 Sequence logos of motifs enriched upstream of <i>spt6-1004</i> -induced intragenic and antisense TSSs.	19
2.1 Diagram of the dual-shutoff system used to deplete Spt5 from <i>S. pombe</i>	21

2.2	Average Spt5 ChIP-seq, RNAPII ChIP-seq, and sense NET-seq signal over non-overlapping coding genes, from Spt5 depleted and non-depleted cells.	22
2.3	Enrichment of RNAPII phospho-serine 5 and phospho-serine 2 over non-overlapping coding genes, in Spt5 depleted and non-depleted cells.	22
2.4	Heatmaps of antisense RNA-seq signal from Spt5 depleted and non-depleted cells, over non-overlapping coding genes.	22
2.5	Bar plot of the number of TSS-seq peaks of various genomic classes differentially expressed in Spt5 depleted versus non-depleted cells.	22
2.6	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.	22
2.7	Average MNase-seq dyad signal from Spt5 depleted and non-depleted cells, over non-overlapping coding genes.	22
2.8	A figure showing MNase-seq signal around Spt5-depletion-induced antisense TSSs.	22
2.9	A figure showing motifs enriched upstream of Spt5-depletion-induced antisense TSSs.	22
3.1	TFIIB ChIP-nexus protection over all genes with stress-induced intragenic TFIIB peaks.	25
3.2	TFIIB ChIP-nexus protection over four genes with stress-induced intragenic TFIIB peaks.	26
3.3	Bar plot of the number of promoters from various genomic classes differentially expressed in oxidative stress.	27

3.4	TSS-seq expression levels in oxidative stress of oxidative-stress-induced genic and intragenic promoters.	28
3.5	A figure showing TSS-seq, TFIIB ChIP-nexus, and MNase-ChIP-seq for the oxidative-stress-induced promoters.	28
3.6	Polysome enrichment in oxidative stress, for oxidative-stress-induced genic and intragenic promoters.	29
3.7	A figure showing TSS-seq coverage over oxidative-stress-induced TSSs in the three species.	29
3.8	A figure showing TSS-seq coverage over DSK2 in the three species, possibly with the corresponding northern blot.	29
3.9	A figure showing TSS-seq, TFIIB ChIP-nexus, and MNase-ChIP-seq at DSK2.	29
3.10	A figure showing DSK2 fitness competition results.	29

Introduction

A brief introduction to transcription

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

Transcription elongation factors Spt6 and Spt5

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 1

Genomics of transcription elongation factor Spt6

1.1 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

1.2 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*,

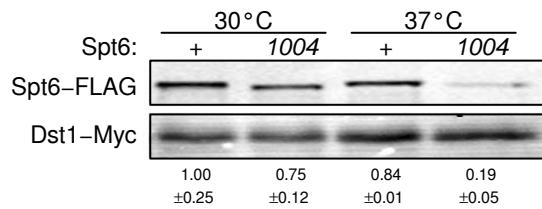
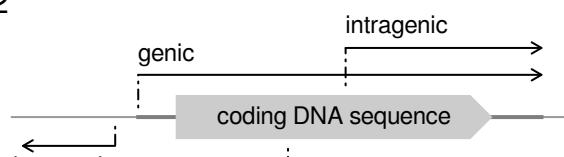


Figure 1.1: Western blot showing Spt6 protein levels in wild-type and *spt6-1004* cells, at 30°C and after 80 minutes at 37 °C. Immunoblotting was performed using α -FLAG antibody to detect Spt6 and α -Myc antibody to detect Dst1 from a spike-in strain. The quantification shown is the mean \pm standard deviation of three blots.

2



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 1.1). 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 1.2) (Cheung et al., 2008, DeGennaro et al., 2013, Kaplan et al., 2003, Uwimana et al., 2017).

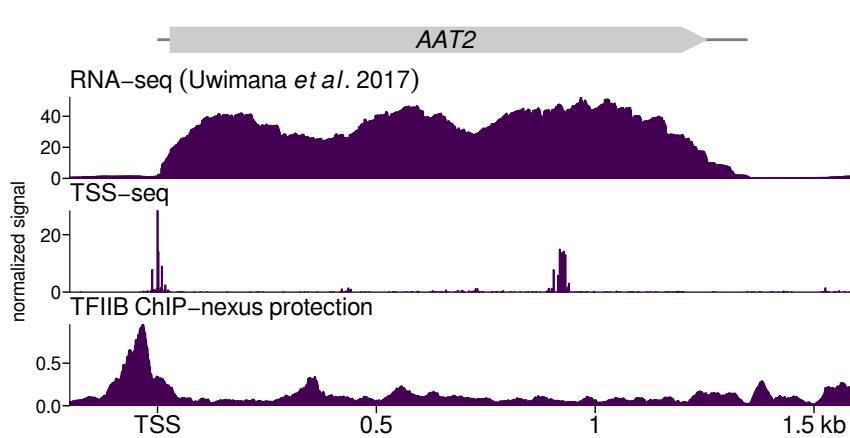


Figure 1.3: Sense strand RNA-seq signal, sense strand TSS-seq signal, and TFIIB ChIP-nexus protection at the AAT2 gene, in *spt6-1004* after 80 minutes at 37°C.

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 1.3) (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 1.3). 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.

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

1.3.1 TSS-seq peak calling

1.3.2 A note on ChIP-nexus peak calling

1.4 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 1.4, 1.5). 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.

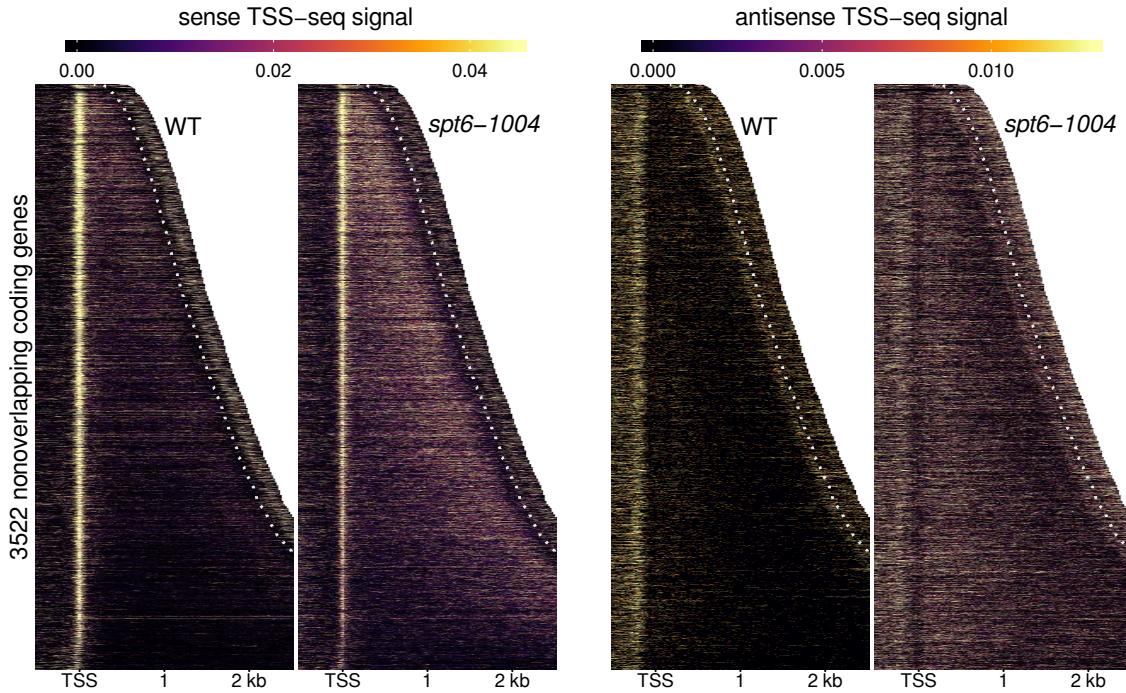


Figure 1.4: Heatmaps of sense and antisense TSS-seq signal from wild-type and *spt6-1004* cells, over 3522 non-overlapping genes aligned by wild-type genic TSS and sorted by annotated transcript length. Data are shown for each gene up to 300 nucleotides 3' of the cleavage and polyadenylation site (CPS), 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 1.4), we detect significant induction of over 4000 intragenic and antisense TSSs in *spt6-1004* (Figure 1.6). 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 1.7) 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 1.6). As a result of intragenic/antisense induction and genic repression, expression levels of all classes of transcripts become similar in *spt6-1004* (Figure 1.8).

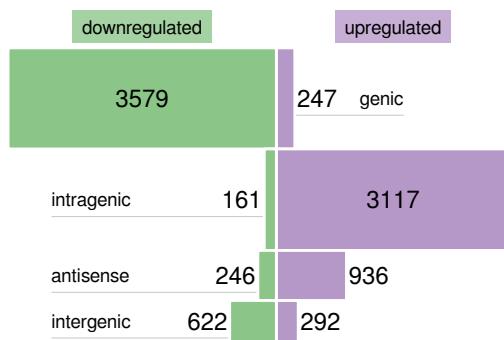


Figure 1.6: Bar plots of the number of TSS-seq peaks differentially expressed in *spt6-1004* after 80 minutes at 37°C versus wild-type after 80 minutes at 37°C. The height of each bar is proportional to the total number of peaks in the category, including those not found to be significantly differentially expressed.

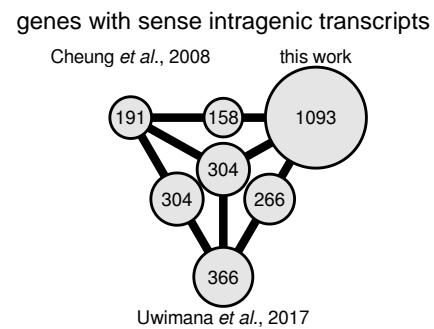


Figure 1.7: Set diagram of the number of genes reported to have *spt6-1004*-induced intragenic transcripts using tiled arrays Cheung et al. (2008), RNA-seq Uwimana et al. (2017), and TSS-seq (this work).

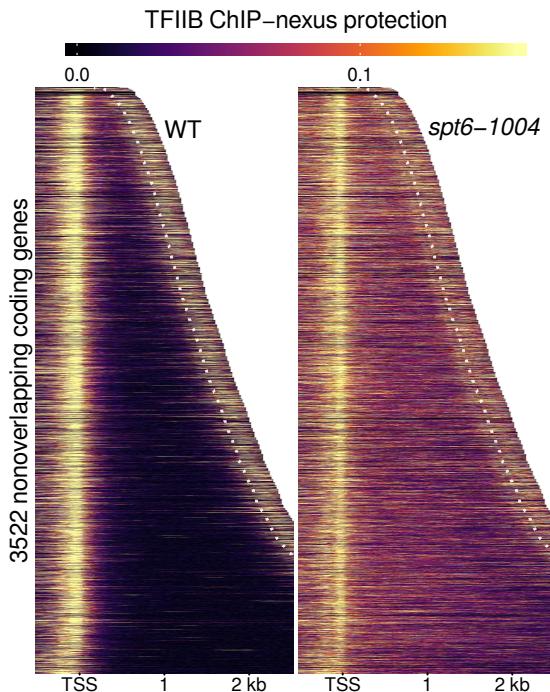


Figure 1.5: Heatmaps of TFIIB binding measured by ChIP-nexus, over the same regions shown in Figure 1.4. Values are the mean of library-size normalized coverage in non-overlapping 20 bp bins, averaged over two replicates. Values above the 85th percentile are set to the 85th percentile for visualization.

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 vehicula augue eu neque. Pellent

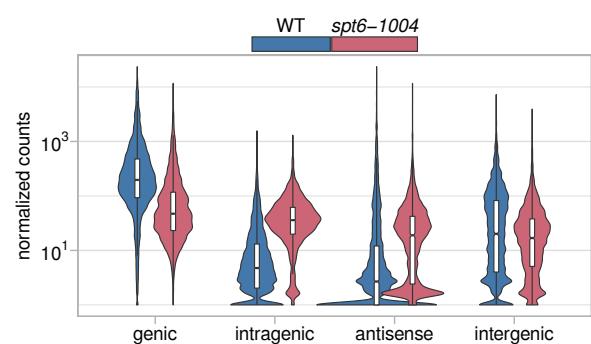


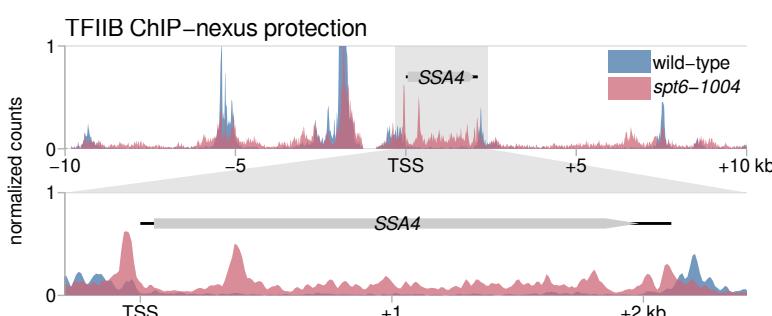
Figure 1.8: Violin plots of expression level distributions for genomic classes of TSS-₈ seq peaks in wild-type and *spt6-1004*, both after 80 minutes at 37°C. Normalized counts are the mean of spike-in size factor

tesque habitant morbi tristique senectus
 et netus et malesuada fames ac turpis
 egestas. Mauris ut leo. Cras viverra me-
 tus 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-
 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.

Figure 1.9:

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

bottom) Expanded view
of TFIIB protection over the
SSA4 gene.



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

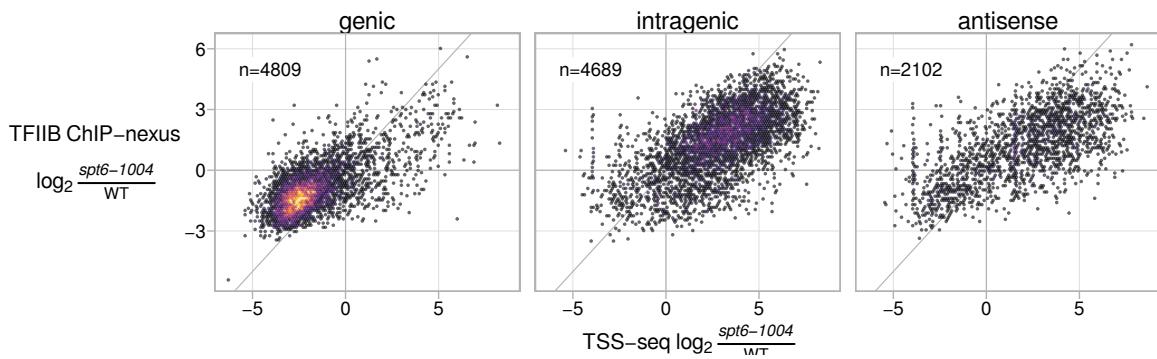


Figure 1.10: Scatterplots of fold-change in *spt6-1004* over wild-type, comparing TSS-seq and TFIIB ChIP-nexus. Each dot represents a TSS-seq peak paired with the window extending 200 bp upstream of the TSS-seq peak summit for quantification of TFIIB ChIP-nexus signal. Fold-changes are regularized fold-change estimates from DESeq2, with size factors determined from the *S. pombe* spike-in (TSS-seq), or *S. cerevisiae* counts (ChIP-nexus).

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.

1.5 MNase-seq results from *spt6-1004*

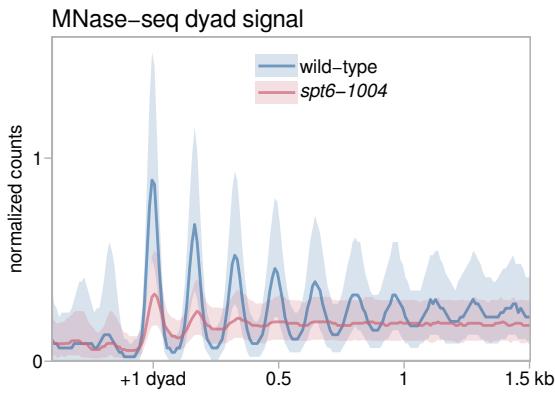


Figure 1.11: Average MNase-seq dyad signal in wild-type and *spt6-1004*, over 3522 non-overlapping genes aligned by wild-type +1 nucleosome dyad. Values are the mean of spike-in normalized coverage in non-overlapping 20 bp bins, averaged over two replicates (*spt6-1004*) or one experiment (wild-type). The solid line and shading are the median and the inter-quartile range.

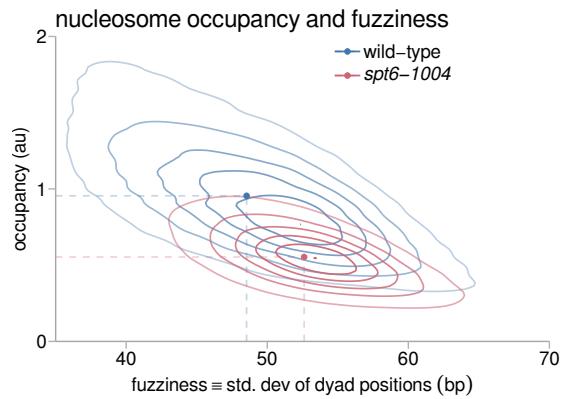


Figure 1.12: Contour plot of the global distributions of nucleosome occupancy and fuzziness in wild-type and *spt6-1004*. Dashed lines indicate median values.

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.

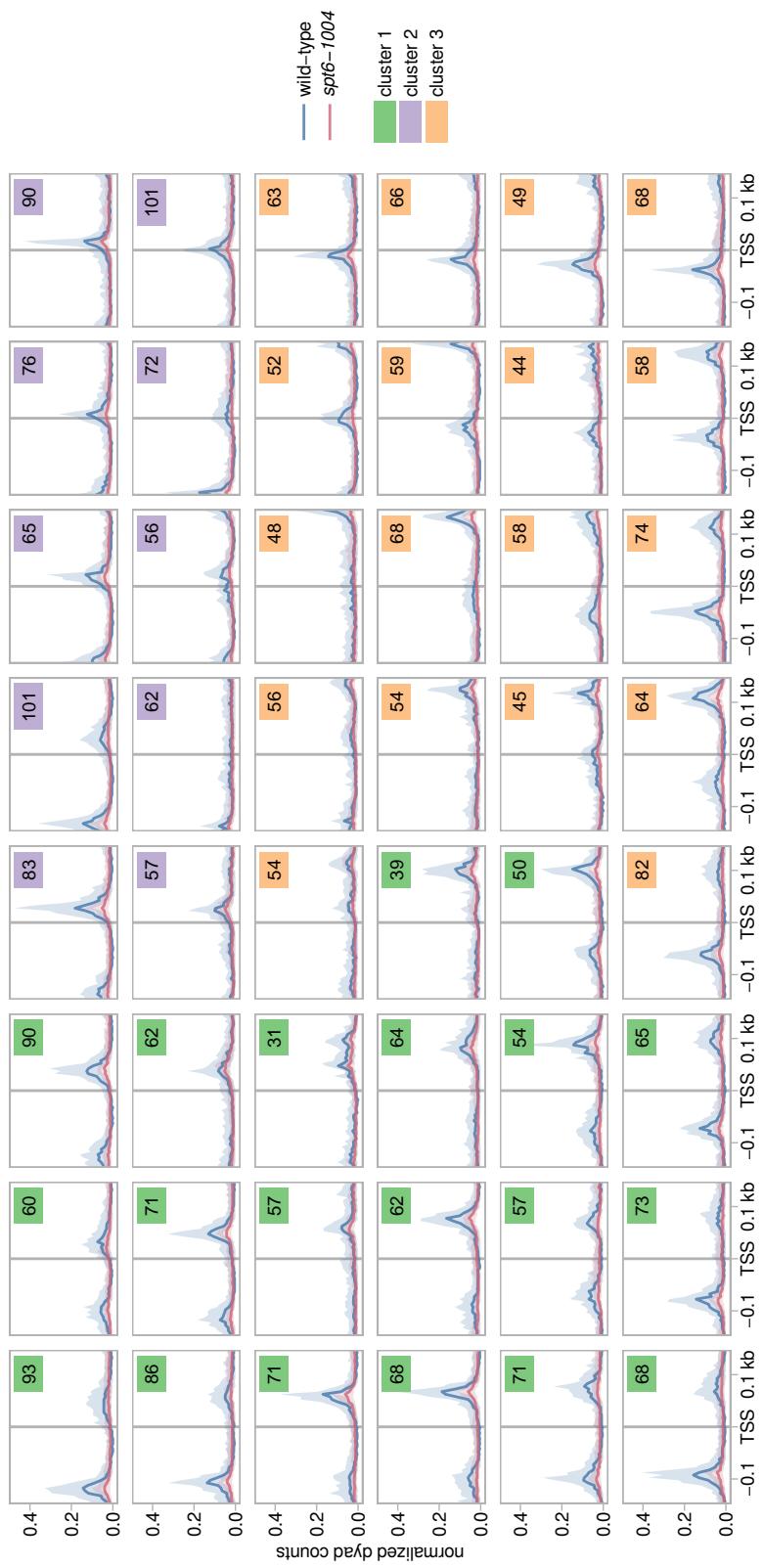


Figure 1.13: bar

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

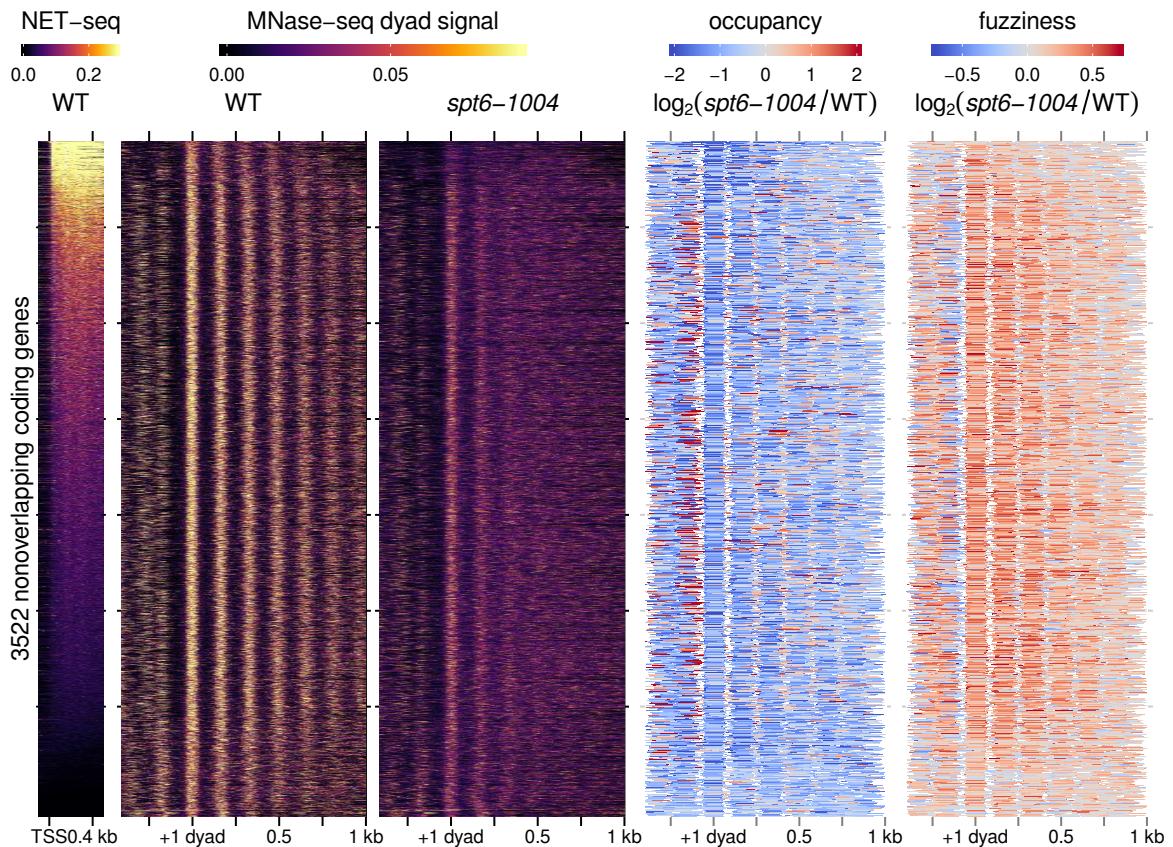


Figure 1.14:

- 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 nucleotides 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.

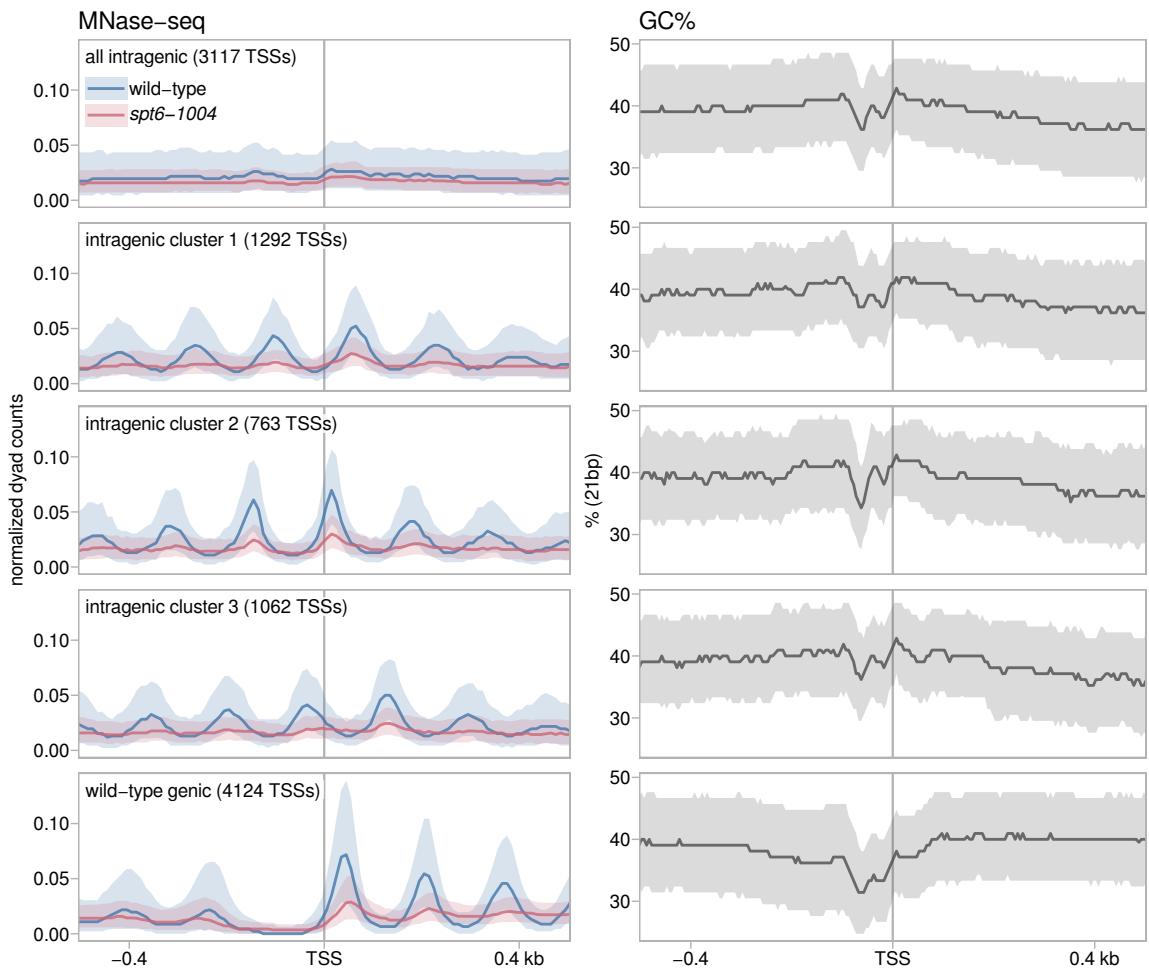


Figure 1.15:

- top row) Average MNase-seq dyad signal for two clusters of *spt6-1004* intragenic TSSs (clustered by wild-type and *spt6-1004* MNase-seq dyad signal flanking the TSS), as well as all genic TSSs detected in wild-type and *spt6-1004*. 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.
- bottom row) Average GC content of the DNA sequence, as above.

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.

1.6 Other features of *spt6-1004* intragenic promoters

1.6.1 Information content of intragenic TSSs

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.

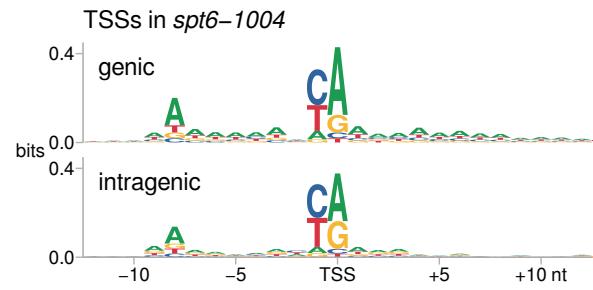


Figure 1.16: Sequence logos of the information content of TSS-seq reads overlapping genic and intragenic TSS-seq peaks in *spt6-1004*.

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.

1.6.2 Sequence motifs enriched at intragenic TSSs

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.

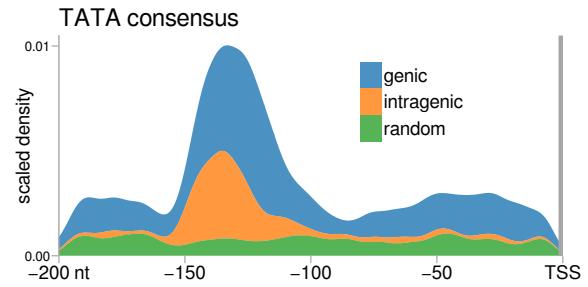


Figure 1.17: Scaled density of occurrences of exact matches to the motif TATAAWR upstream of TSSs. For each category, a Gaussian kernel density estimate of the positions of motif occurrences is multiplied by the number of motif occurrences in the genomic category and divided by the number of regions in the category.

Figure 1.18

1.7 Summary

1.8 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. 1.2
- 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), 1.2, 1.2, 1.7
- 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. 1.2
- 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 –. 1.2
- Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–1099. 1.2
- 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. 1.2
- 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. 1.2
- 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. 1.2
- 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), 1.2, 1.2, 1.7
- 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. 1.2

Chapter 2

Genomics of transcription elongation factor Spt5

2.1 Collaborators

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

2.2 Introduction to Spt5 and prior work

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

2.4 TSS-seq results from Spt5 depletion

2.5 MNase-seq results from Spt5 depletion

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

2.6 Sequence motifs enriched at antisense TSSs

2.7 Summary

Figure 2.1: Caption wsdasdr zzzz.

Figure 2.2: Caption wsdasdr zzzz.

Figure 2.3: Caption wsdasdr zzzz.

Figure 2.4: Caption wsdasdr zzzz.

Figure 2.5: Caption wsdasdr zzzz.

Figure 2.6: Caption wsdasdr zzzz.

Figure 2.7: Caption wsdasdr zzzz.

Figure 2.8: Caption wsdasdr zzzz.

Figure 2.9: Caption wsdasdr zzzz.

2.8 Bibliography

Chapter 3

Stress-responsive intragenic transcription

3.1 Collaborators

Steve Doris generated TSS-seq and ChIP-nexus libraries

Dan Spatt polyribosome fractionation

3.2 Possible functions for intragenic transcription in wild-type cells

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

3.4 Chromatin landscape of oxidative-stress-induced promoters.

3.5 Polysome enrichment of oxidative-stress-induced intragenic transcripts

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

3.7 Functions of intragenic DSK2 expression in oxidative stress

3.8 Summary

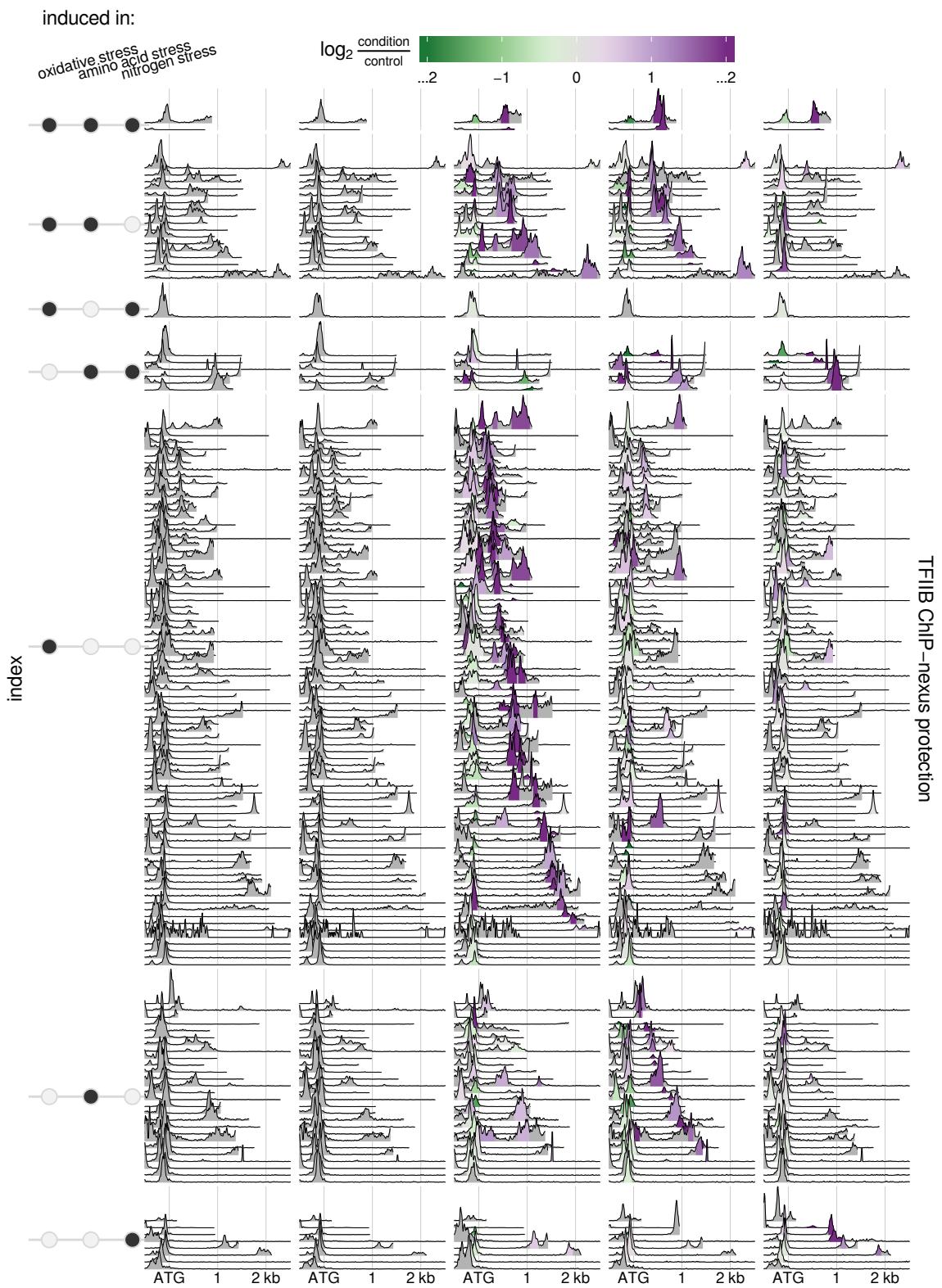


Figure 3.1: Wasldfkjlk asldkfj.

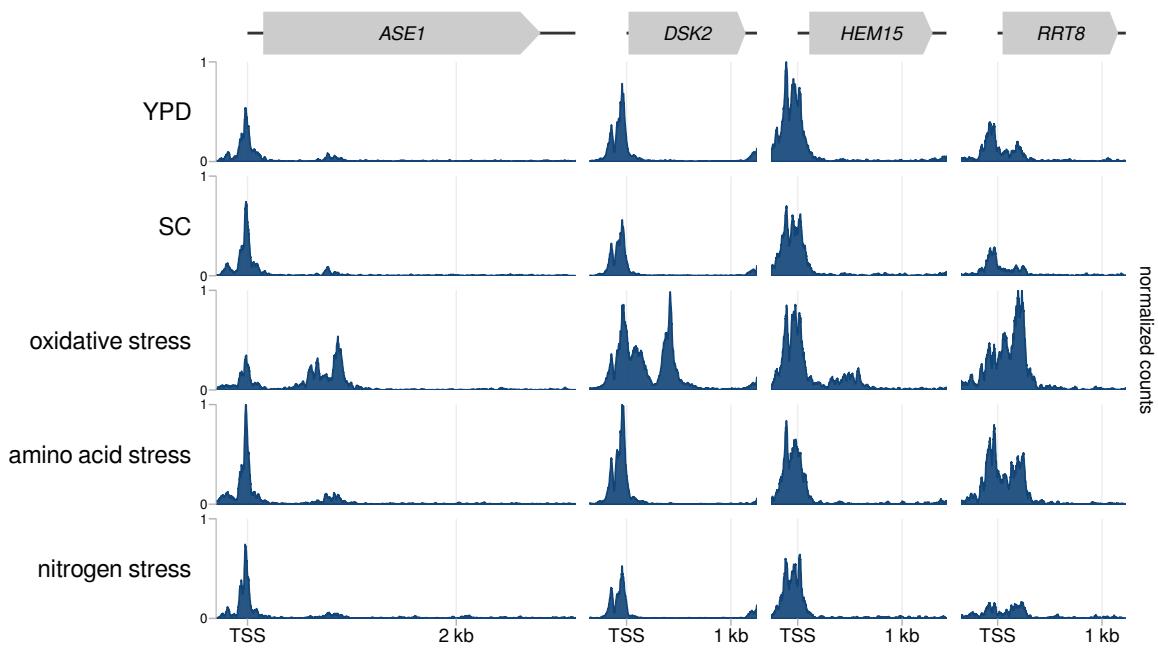


Figure 3.2: Caption asdflkj asldkfjlkj.

3.9 Bibliography

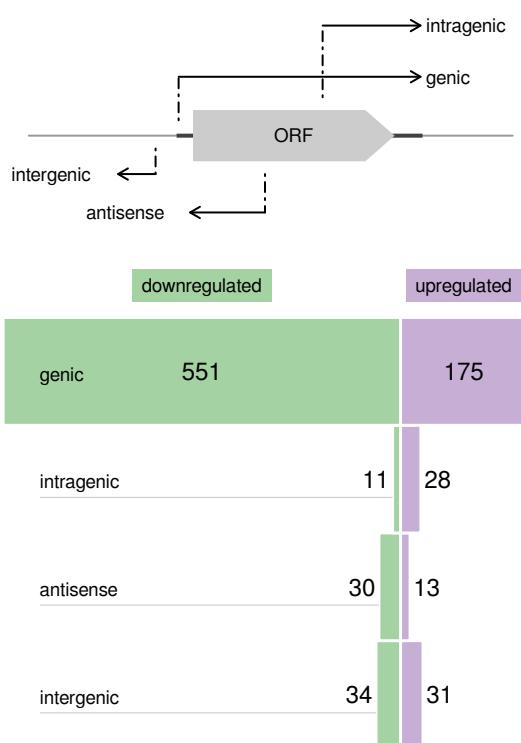


Figure 3.3: Caption dsafklj asldkfjlkj.

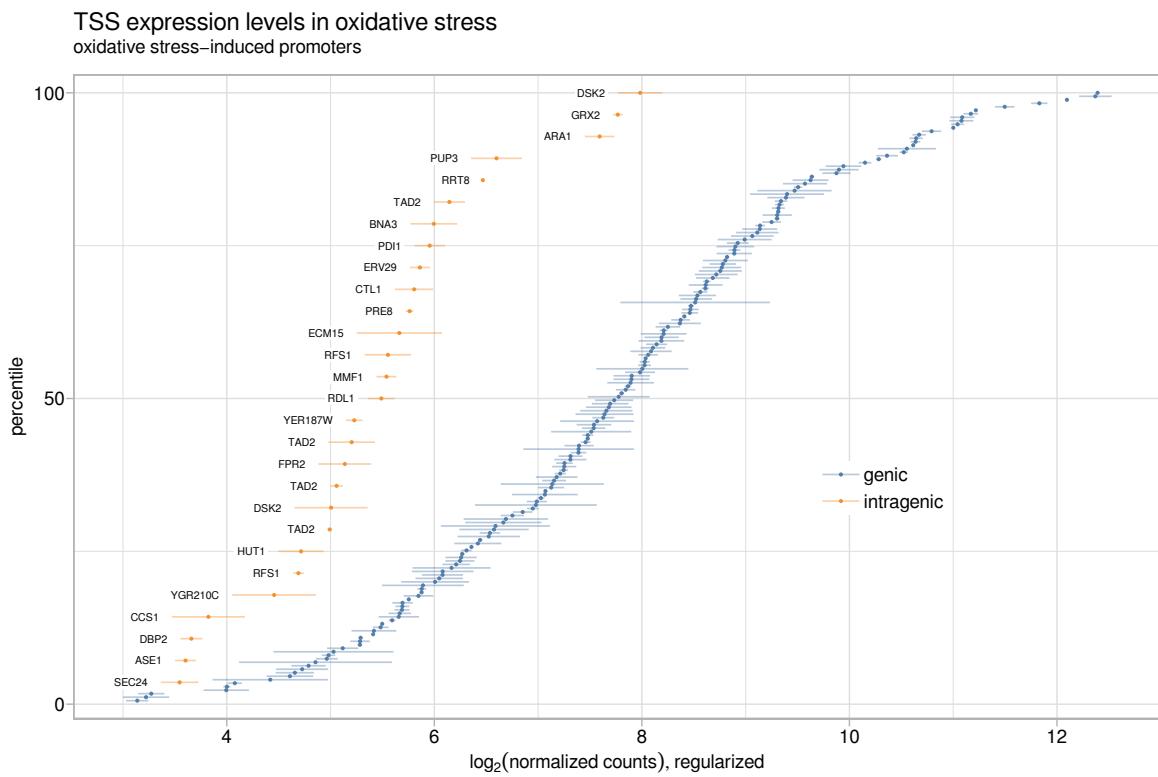


Figure 3.4: Caption dsafklj zzzz.

Figure 3.5: Caption dsafklj .

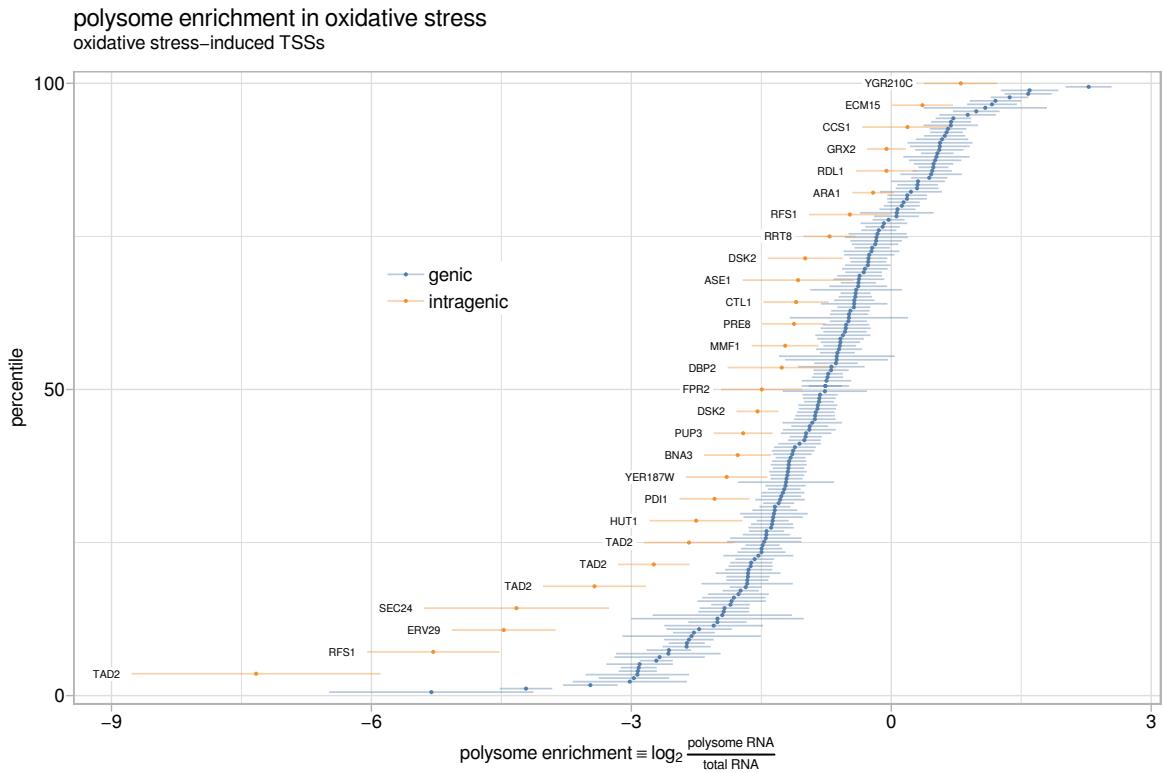


Figure 3.6: Caption wsadasdr zzzz.

Figure 3.7: Caption dsafklj .

Figure 3.8: Caption dsafklj .

Figure 3.9: Caption dsafklj .

Figure 3.10: 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.
- 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.
- 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.
- 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 –.
- Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–1099.
- 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.
- 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.

- 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.