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

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

List of Figures

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 (Pol II), and broadly occurs in three sequential stages of transcription initiation, elongation, and termination (Shandilya and Roberts, 2012). During each of these stages, the Pol II complex is associated with distinct sets of factors which modulate the activity of Pol II and carry out co-transcriptional processes such as RNA capping, RNA splicing, histone modification, RNA cleavage, and RNA polyadenylation. Given how fundamental transcription is to gene expression, it is unsurprising that every stage of transcription is highly regulated.

To get a rough idea of just how tightly transcription is regulated, it is useful to consider a back-of-the-envelope calculation of the specificity of transcription initiation in the human genome. That is, what proportion of the human genome at which transcription could initiate does transcription initiation actually occur?

The number of positions at which transcription could theoretically initiate is simply the size of the genome: The human genome is approximately three billion base pairs in length (BNID 111378, Weber et al. (2009)), and since each base pair can be transcribed from each of its two strands, there are 6×10^9 available positions.

The number of positions at which transcription *does* initiate can be estimated from the number of genes transcribed by Pol II and the number of positions that Pol II initiates from for each gene. At last count, the human genome contains about twenty thousand protein-coding genes (Consortium et al., 2012). To be conservative in our estimate with regards to specificity, we will assume that all twenty thousand genes are expressed. We also know that protein-coding genes are only a subset of the genes transcribed by Pol II: Pol II also transcribes multiple classes of non-coding genes, including enhancers and long non-coding RNAs (Kaikkonen and Adelman, 2018). Compared to protein-coding genes, the number of non-coding genes is less certain. If we assume that there are five non-coding genes for each coding gene, this brings our estimate of the number of genes transcribed by Pol II to 1.2×10^5 genes.

As you will see from yeast transcription start site data in later chapters, transcription initiation for a single gene generally occurs at multiple nucleotides, generating multiple major transcript isoforms per gene. Assuming that there are five major transcription start sites (TSSs) per gene, the proportion of the human genome at which transcription initiation occurs is

$$\frac{(1.2 \times 10^5 \text{ genes}) \left(5 \frac{\text{TSSs}}{\text{gene}}\right)}{(6 \times 10^9 \text{ possible TSSs})} = 1 \times 10^{-4}.$$

Our rough estimate says that, when presented with ten thousand positions to choose from, RNA polymerase starts transcription from only one![1]

Many factors are known to contribute to this remarkable specificity. Most notably, transcription initiation requires the presence of specific DNA sequence motifs, which

[1] A similar conclusion is reached by examining ENCODE CAGE-seq data: At the time of writing, ENCODE reports roughly 150,000 TSS peaks across 30 cell types/cell lines. Assuming the signal is concentrated at 5 nucleotides per peak, then $\frac{\left(1.5 \times 10^5 \text{peaks}\right)\left(5 \frac{\text{nt}}{\text{peaks}}\right)}{6 \times 10^9 \text{nt}} = \frac{1}{8000}$.

increase the probability of Pol II binding to DNA together with necessary initiation factors (Haberle and Stark, 2018). That factors known to associate with Pol II during transcription initiation control transcription initiation is hardly surprising. A less obvious fact is that some transcription *elongation* factors, including histone chaperones and histone modification enzymes, also play a role in restricting where transcription initiation is allowed to occur (Cheung et al., 2008; Hennig and Fischer, 2013; Kaplan et al., 2003). Evidence suggests that these elongation factors are likely required to maintain normal chromatin structure over transcribed regions, and that the disruption of normal chromatin structure allows Pol II to initiate transcription in regions which are normally inacessible (). Chapters 2 and 3 of this dissertation describe our studies of **Spt6** and **Spt5**, two of the transcription elongation factors involved in this process. One phenotype observed when these factors are disrupted is **intragenic transcription**, transcription appearing to arise from within protein-coding sequences. In chapter 4, I describe our efforts to understand how intragenic transcription might play a role in the cellular response to various stress conditions. The remainder of this introduction provides a brief overview of the considerations taken into account in order to make the data analyses behind this dissertation as transparent and reproducible as possible.

1.2 Reproducible data analysis for genomics

My role in the projects in this dissertation is a mix of **data scientist** and **data engineer**: I build pipelines for processing (usually genomic) datasets, taking raw data through processing, statistical analysis, and data visualization. This mostly entails surveying available tools, selecting the tools most suitable for the task, and coding solutions to problems when existing tools are insufficient.

The analysis of complex datasets like those generated by genomic assays presents challenges to achieving transparency and reproducibility when reporting methods and results. In building the data analysis pipelines behind the results of this dissertation, I have tried to meet these challenges by following best practices that would be standards for publication in an ideal world. All of my data analyses are open source (github.com/winston-lab), and are designed to be reproducible by others: For all publications, an self-contained archive is uploaded which includes everything needed to go from raw data to the figures and results of the publication (e.g. `https://doi.org/10.5281/zenodo.1409826`). This level of accessibility is greatly facilitated by building data analyses using Snakemake (Köster and Rahmann, 2012), one of several available frameworks for workflow management (Di Tommaso et al., 2017; Voss et al., 2017). Snakemake's scalable execution and its ability to specify dependencies in virtual environments allow workflows to truly be reproducible: data analyses can be re-run on personal computers, computing clusters, or cloud environments, and the exact versions of the software used when initially running the data analysis will automagically be deployed.

Open sharing of data and code like this is essential to the scientific process. When analysis pipelines routinely consist of tens of steps with tens of parameters each, seeing the data and code is the only way for those interested to know exactly how the data were handled. Altogether, this allows for more informed evaluation of results from the literature, as well as the possibility of finding and correcting errors in analysis.

1.3 Bibliography

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. 1.1

Consortium, T. E. P., Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kellis, M., Kheradpour, P., Lassmann, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C. J., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. A. L., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Good, P. J., Feingold, E. A., Crawford, G. E., Dekker, J., Elnitski, L., Farnham, P. J., Giddings, M. C., Gingeras, T. R., Guigó, R., Hubbard, T. J., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Stamatoyannopoulos, J. A., Tenenbaum, S. A., Weng, Z., White, K. P., Wold, B., Yu, Y., Wrobel, J., Risk, B. A., Gunawardena, H. P., Kuiper, H. C., Maier, C. W., Xie, L., Chen, X., Mikkelsen, T. S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M. J., Durham, T., Ku, M., Truong, T., Eaton, M. L., Dobin, A., Tanzer, A., Lagarde, J., Lin, W., Xue, C., Williams, B. A., Zaleski, C., Röder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O. J., Park, E., Preall, J. B., Presaud, K., Ribeca, P., Robyr, D., Ruan, X., Sammeth, M., Sandhu, K. S., Schaeffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Hayashizaki, Y., Reymond, A., Antonarakis, S. E., Hannon, G. J., Ruan, Y., Carninci, P., Sloan, C. A., Learned, K., Malladi, V. S., Wong, M. C., Barber, G. P., Cline, M. S., Dreszer, T. R., Heitner, S. G., Karolchik, D., Kirkup, V. M., Meyer, L. R., Long, J. C., Maddren, M., Raney, B. J., Grasfeder, L. L., Giresi, P. G., Battenhouse, A., Sheffield, N. C., Showers, K. A., London, D., Bhinge, A. A., Shestak, C., Schaner, M. R., Ki Kim, S., Zhang, Z. Z., Mieczkowski, P. A., Mieczkowska, J. O., Liu, Z., McDaniell, R. M., Ni, Y., Rashid, N. U., Kim, M. J., Adar, S., Zhang, Z., Wang, T., Winter, D., Keefe,

D., Iyer, V. R., Zheng, M., Wang, P., Gertz, J., Vielmetter, J., Partridge, E., Varley, K. E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K. M., Anaya, M., Cross, M. K., Muratet, M. A., Newberry, K. M., McCue, K., Nesmith, A. S., Fisher-Aylor, K. I., Pusey, B., DeSalvo, G., Parker, S. L., Balasubramanian, S., Davis, N. S., Meadows, S. K., Eggleston, T., Newberry, J. S., Levy, S. E., Absher, D. M., Wong, W. H., Blow, M. J., Visel, A., Pennachio, L. A., Petrykowska, H. M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Davidson, C., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J. M., Griffiths, E., Harte, R., Hendrix, D. A., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Leng, J., Lin, M. F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J. M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tapanari, E., Tress, M. L., van Baren, M. J., Washietl, S., Wilming, L., Zadissa, A., Zhang, Z., Brent, M., Haussler, D., Valencia, A., Addleman, N., Alexander, R. P., Auerbach, R. K., Balasubramanian, S., Bettinger, K., Bhardwaj, N., Boyle, A. P., Cao, A. R., Cayting, P., Charos, A., Cheng, Y., Eastman, C., Euskirchen, G., Fleming, J. D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyengar, S., Jin, V. X., Karczewski, K. J., Kasowski, M., Lacroute, P., Lam, H., Lamarre-Vincent, N., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X. J., O'Geen, H., Ouyang, Z., Patacsil, D., Raha, D., Ramirez, L., Reed, B., Shi, M., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K.-K., Yang, X., Struhl, K., Weissman, S. M., Penalva, L. O., Karmakar, S., Bhanvadia, R. R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Victorsen, A., Auer, T., Centanin, L., Eichenlaub, M., Gruhl, F., Heermann, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Partridge, E. C., Jain, G., Balasundaram, G., Bates, D. L., Byron, R., Canfield, T. K., Diegel, M. J., Dunn, D., Ebersol, A. K., Frum, T., Garg, K., Gist, E., Hansen, R. S., Boatman, L., Haugen, E., Humbert, R., Johnson, A. K., Johnson, E. M., Kutyavin, T. V., Lee, K., Lotakis, D., Maurano, M. T., Neph, S. J., Neri, F. V., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Rynes, E., Sanchez, M. E., Sandstrom, R. S., Shafer, A. O., Stergachis, A. B., Thomas, S., Vernot, B., Vierstra, J., Vong, S., Wang, H., Weaver, M. A., Yan, Y., Zhang, M., Akey, J. M., Bender, M., Dorschner, M. O., Groudine, M., MacCoss, M. J., Navas, P., Stamatoyannopoulos, G., Beal, K., Brazma, A., Flicek, P., Johnson, N., Lukk, M., Luscombe, N. M., Sobral, D., Vaquerizas, J. M., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M. W., Schaub, M. A., Miller, W., Bickel, P. J., Banfai, B., Boley, N. P., Huang, H., Li, J. J., Noble, W. S., Bilmes, J. A., Buske, O. J., Sahu, A. D., Kharchenko, P. V., Park, P. J., Baker, D., Taylor, J., and Lochovsky, L. (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489:57 EP –. Article. 1.1

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319. 1.2

Haberle, V. and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, 19(10):621–637. 1.1

Hennig, B. P. and Fischer, T. (2013). The great repression: chromatin and cryptic transcription. *Transcription*, 4(3):97—101. 1.1

Kaikkonen, M. U. and Adelman, K. (2018). Emerging roles of non-coding rna transcription. *Trends in Biochemical Sciences*, 43(9):654–667. 1.1

Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–1099. 1.1

Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522. 1.2

Shandilya, J. and Roberts, S. G. (2012). The transcription cycle in eukaryotes: From productive initiation to rna polymerase ii recycling. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(5):391 – 400. 1.1

Voss, K., Gentry, J., and Van der Auwera, G. (2017). Full-stack genomics pipelining with gatk4 + wdl + cromwell. In *18th Annual Bioinformatics Open Source Conference (BOSC 2017)*. 1.2

Weber, G., Springer, M., Jorgensen, P., Milo, R., and Moran, U. (2009). Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(suppl_1):D750–D753. 1.1

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

The conserved transcription elongation factor Spt6 interacts directly with RNA polymerase II (Close et al., 2011; Diebold et al., 2010b; Liu et al., 2011; Sdano et al., 2017; Sun et al., 2010; Yoh et al., 2007), histones (Bortvin and Winston, 1996; McCullough et al., 2015), and another elongation factor called Spn1/Iws1 (Diebold et al., 2010a; Li et al., 2018; McDonald et al., 2010). The classification of Spt6 as a transcription elongation factor is based on its association with elongating Pol II (Andrulis et al., 2000; Ivanovska et al., 2011; Kaplan et al., 2000; Mayer et al., 2010), and its ability to enhance elongation both *in vitro* (Endoh et al., 2004) and *in vivo* (Ardehali et al., 2009), though Spt6 has also been shown to regulate initiation in a small number

of cases (Adkins and Tyler, 2006; Ivanovska et al., 2011). Evidence suggests that as Spt6 travels with elongating Pol II, it acts as a histone chaperone, reassembling nucleosomes in the wake of transcription (Duina, 2011). Consistent with its histone chaperone function, Spt6 influences chromatin structure (Bortvin and Winston, 1996; DeGennaro et al., 2013; Ivanovska et al., 2011; Jeronimo et al., 2015; Kaplan et al., 2003; Perales et al., 2013; van Bakel et al., 2013); Spt6 is also required for some histone modifications, including H3K36 methylation (Carrozza et al., 2005; Chu et al., 2006; Yoh et al., 2008; Youdell et al., 2008), and, in some organisms, H3K4 and H3K27 methylation (Begum et al., 2012; Chen et al., 2012; DeGennaro et al., 2013; Wang et al., 2017, 2013).

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*, 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). A notable phenotype of the *spt6-1004* mutant is the appearance of **intragenic transcripts**, transcripts which

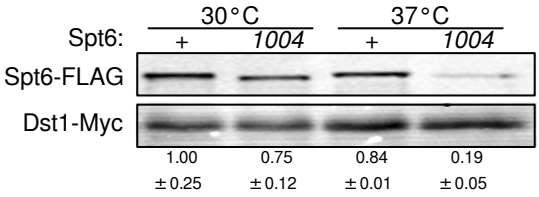


Figure 2.1: Western blot for Spt6 in wild-type and *spt6-1004* cells, at 30°C and after 80 minutes at 37°C. Spt6 and Dst1 from a spike-in were detected using α-FLAG and α-Myc antibodies, respectively. The mean $\pm$ standard deviation of three blots are shown below each lane.

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).

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 (Figures 2.2, 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.

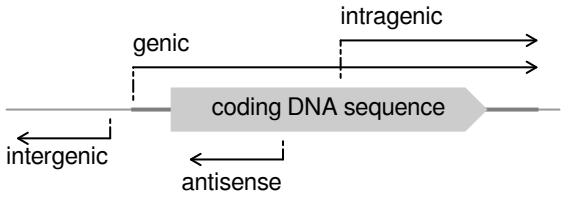


Figure 2.2: Diagram of transcript orientation with respect to coding DNA sequences, for the categories of transcripts referred to in this document.

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

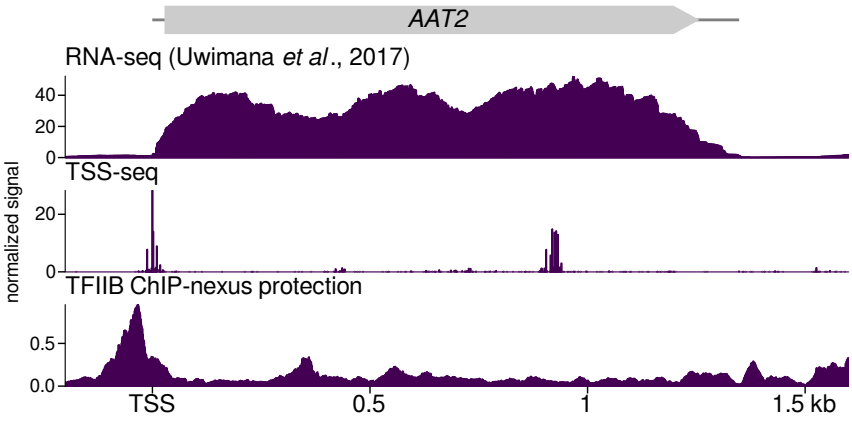


Figure 2.3: Sense strand RNA-seq signal, sense strand TSS-seq signal, and TFIIB ChIP-nexus protection at the *AAT2* gene, in *spt6-1004* after 80 minutes at 37°C.

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 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, reflecting widespread intragenic expression of capped and polyadenylated transcripts, and suggesting that new transcription initiation contributes to the intra-

genic 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.

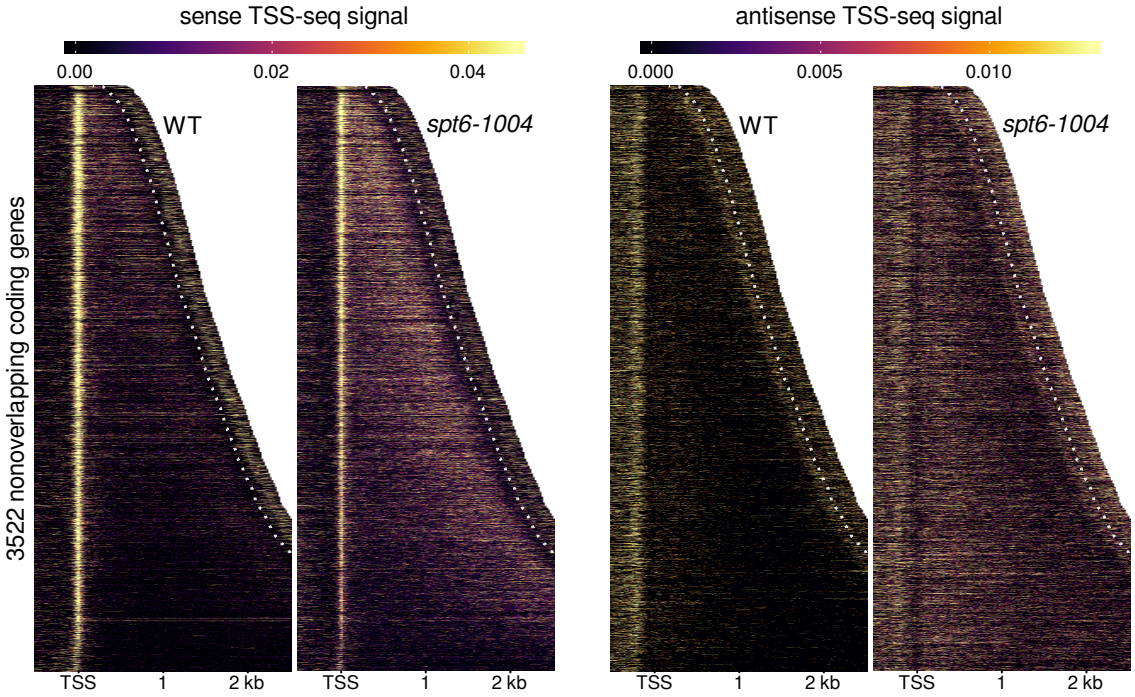


Figure 2.4: Heatmaps of sense and antisense TSS-seq signal from wild-type and *spt6-1004* cells, over 3522 non-overlapping genes aligned by wild-type genic TSS and sorted by annotated transcript length. Data are shown for each gene up to 300 nucleotides 3′ of the cleavage and polyadenylation site (CPS), 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* (Fig-

ure 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.

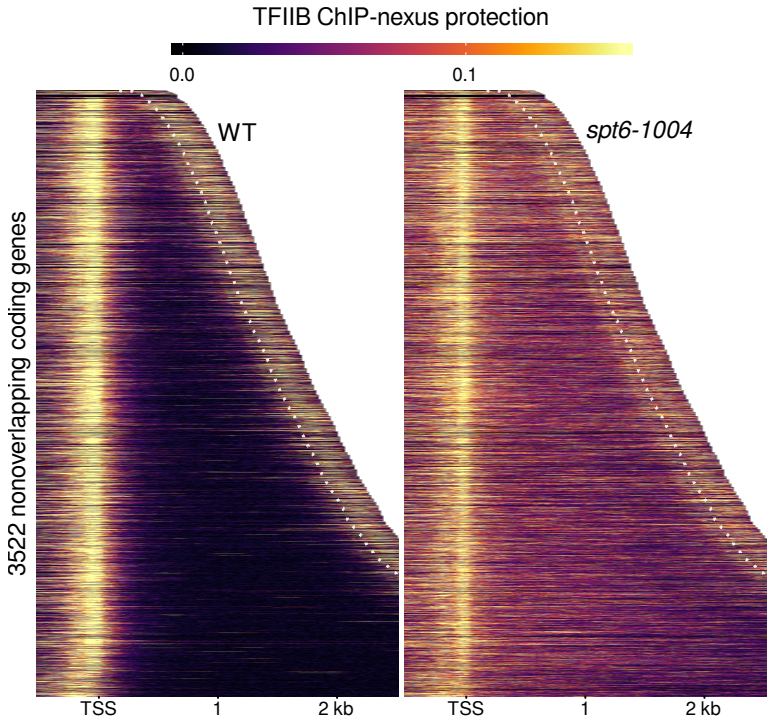


Figure 2.5: Heatmaps of TFIIB binding measured by ChIP-nexus, over the same regions shown in Figure 2.4. Values are the mean of library-size normalized coverage in non-overlapping 20 bp bins, averaged over two replicates. Values above the 85th percentile are set to the 85th percentile for visualization.

The 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 intragenic/antisense induction and genic repression, expression levels in *spt6-1004* of all classes of transcripts become similar to one another (Figure 2.8).

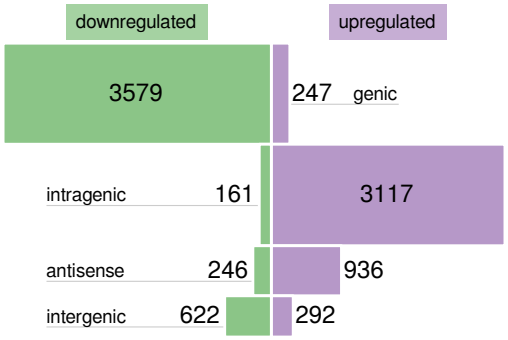


Figure 2.6: Bar plots of the number of TSS-seq peaks differentially expressed in *spt6-1004* after 80 minutes at 37°C versus wild-type after 80 minutes at 37°C. The height of each bar is proportional to the total number of peaks in the category, including those not found to be significantly differentially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

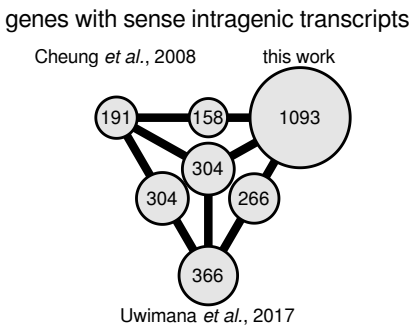


Figure 2.7: Set diagram of the number of genes reported to have *spt6-1004*-induced intragenic transcripts using tiled arrays (Cheung et al., 2008), RNA-seq (Uwimana et al., 2017), and TSS-seq (this work).

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 promiscuously, 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

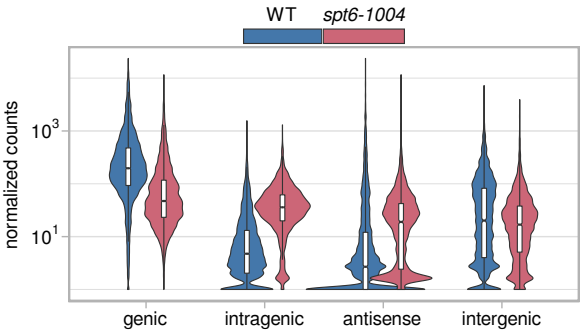


Figure 2.8: Violin plots of expression level distributions for genomic classes of TSS-seq peaks in wild-type and *spt6-1004*, both after 80 minutes at 37°C. Normalized counts are the mean of spike-in size factor normalized counts from two replicates.

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.

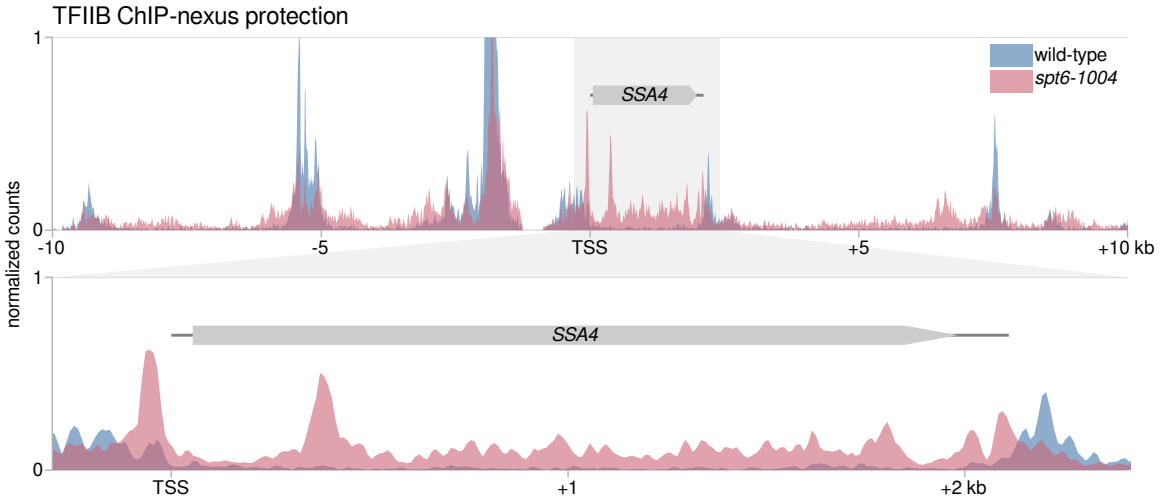


Figure 2.9:

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.

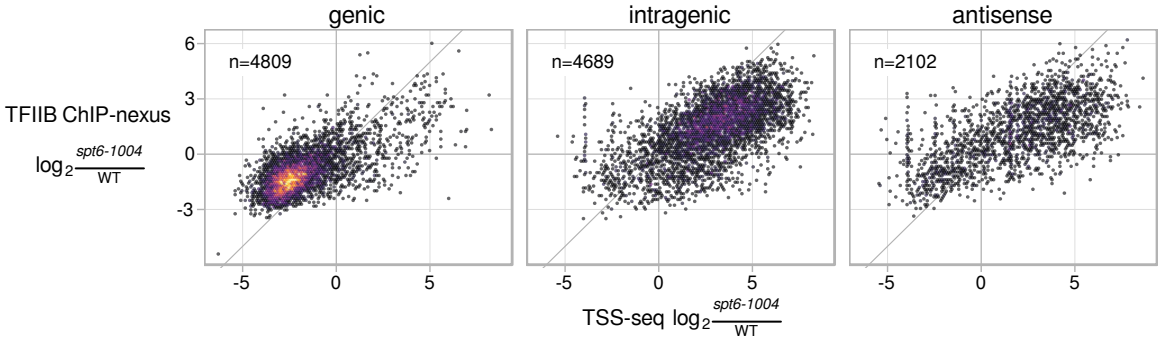


Figure 2.10: Scatterplots of fold-change in *spt6-1004* over wild-type, comparing TSS-seq and TFIIB ChIP-nexus. Each dot represents a TSS-seq peak paired with the window extending 200 bp upstream of the TSS-seq peak summit for quantification of TFIIB ChIP-nexus signal. Fold-changes are regularized fold-change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

2.4 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).

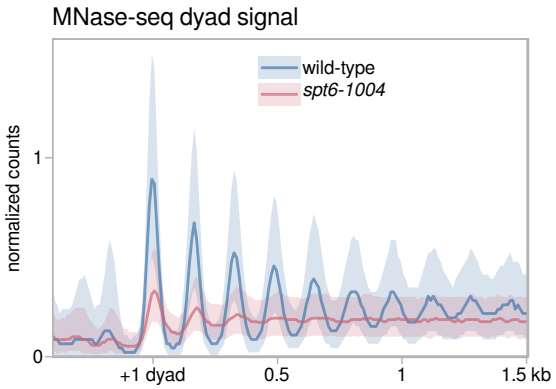


Figure 2.11: Average MNase-seq dyad signal in wild-type and *spt6-1004*, over 3522 non-overlapping genes aligned by wild-type +1 nucleosome dyad. Values are the mean of spike-in normalized coverage in non-overlapping 20 bp bins, averaged over two replicates (*spt6-1004*) or one experiment (wild-type).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

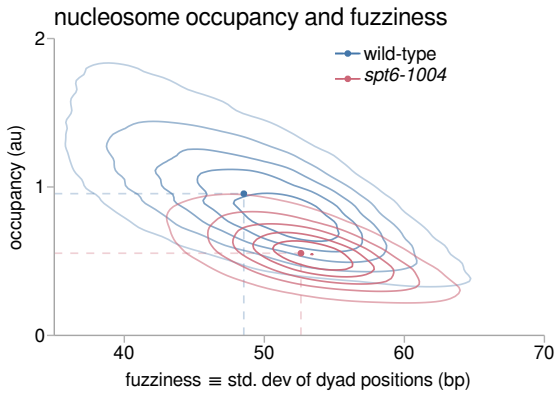


Figure 2.12: Contour plot of the global distributions of nucleosome occupancy and fuzziness in wild-type and *spt6-1004*. Dashed lines indicate median values.

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).

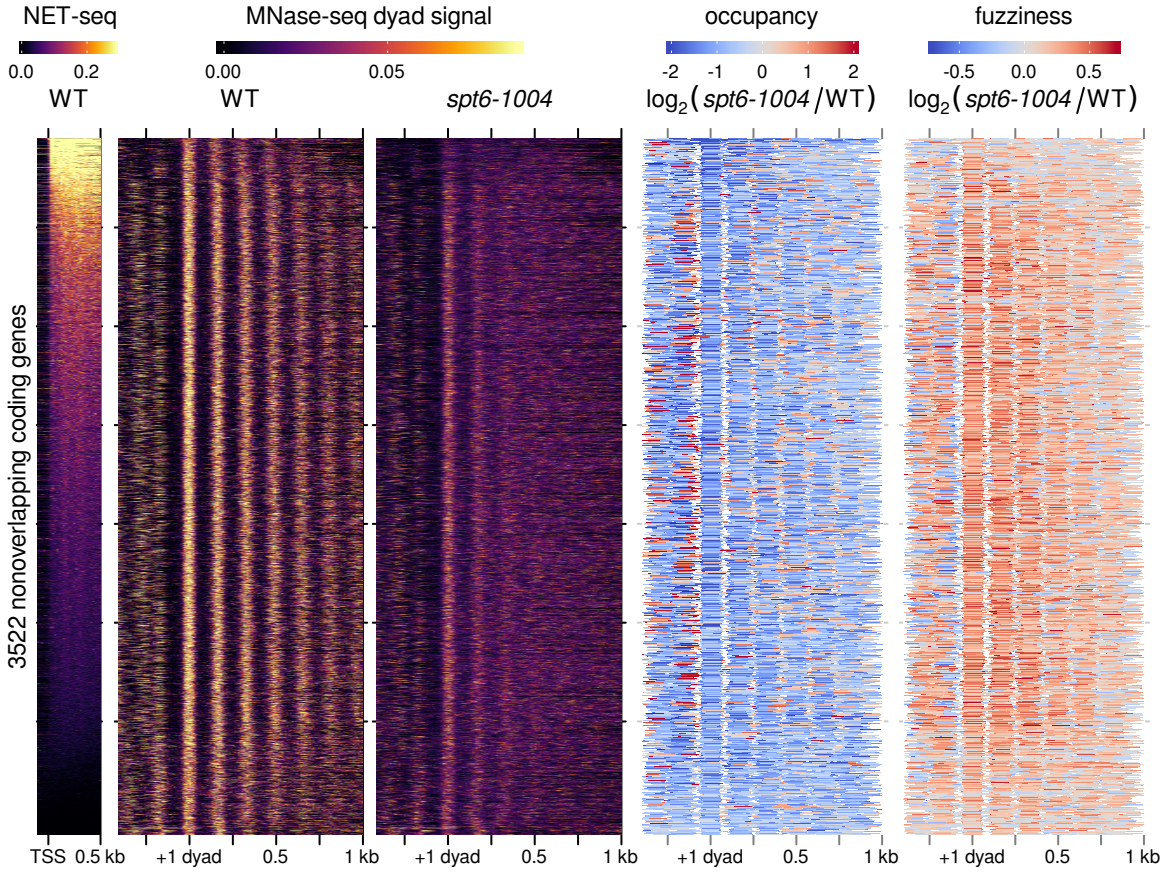


Figure 2.13:

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.

2.4.1 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 nodes 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.

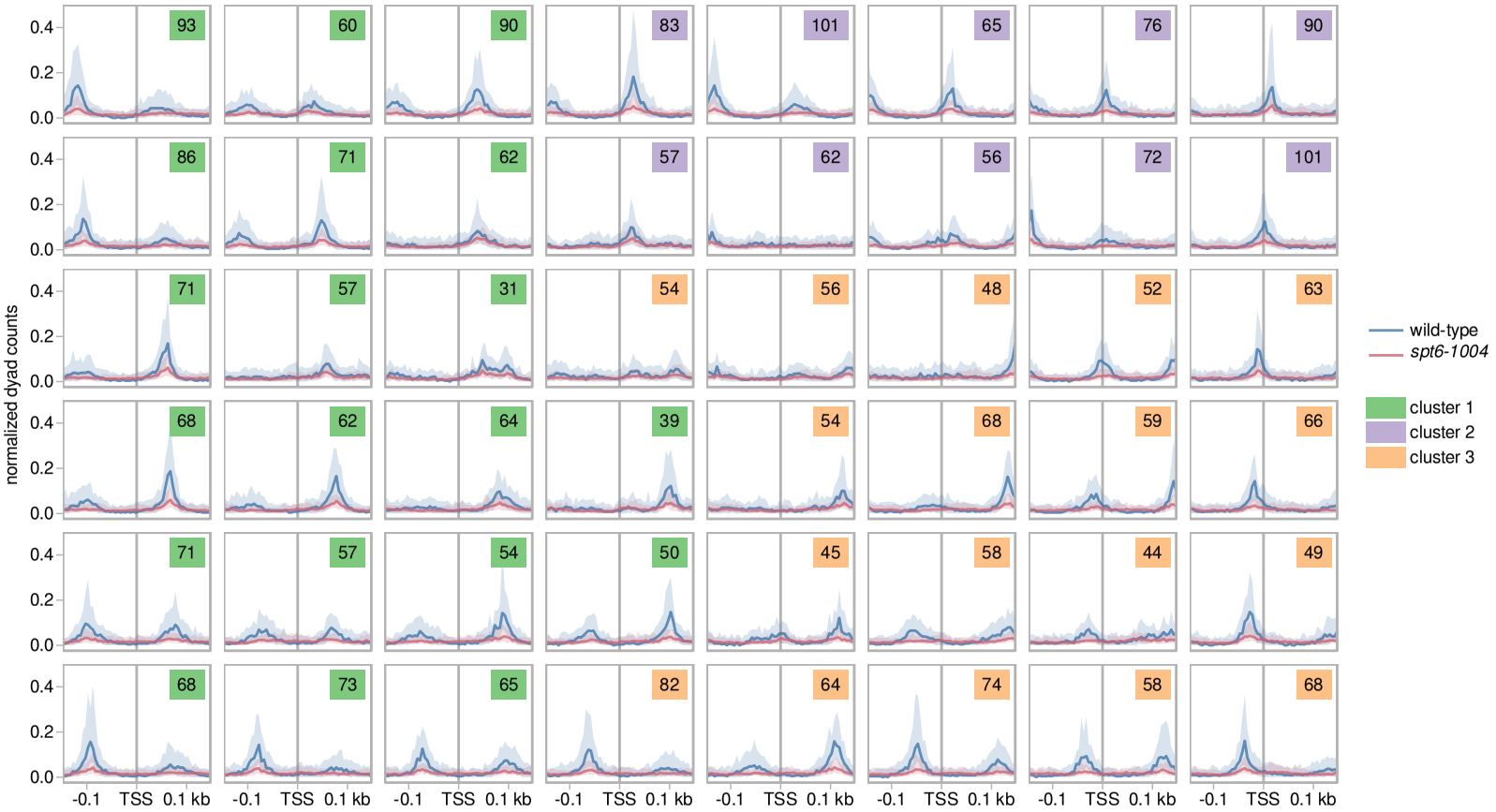


Figure 2.14: Average MNase-seq dyad signal around all *spt6-1004*-induced intragenic TSSs, grouped by assignment to nodes of a 6x8 super-organizing map (SOM). The number of TSSs assigned to each node is shown in the upper right of each panel, and is shaded by the node's assignment to a cluster determined 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.

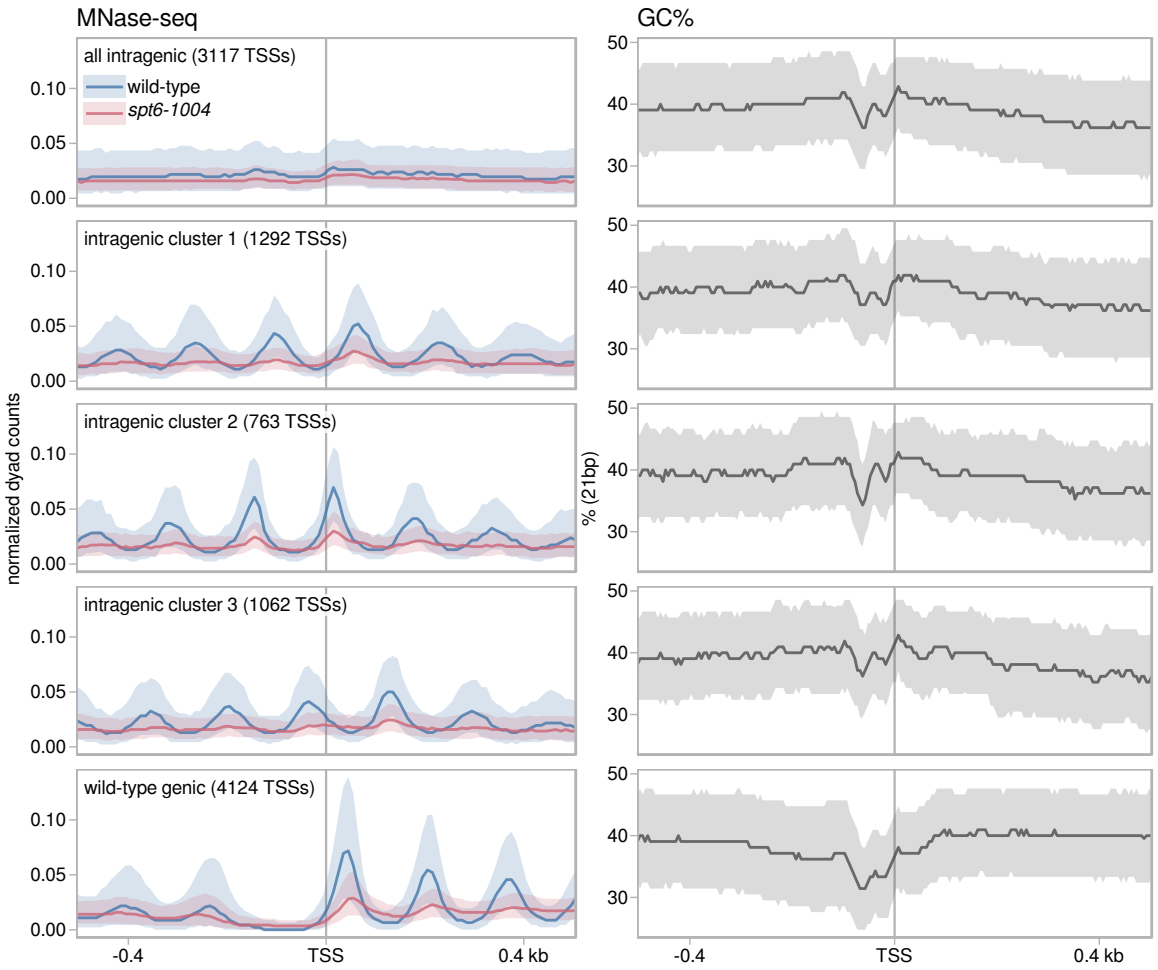


Figure 2.15:

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.

2.5 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.

2.5.1 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.

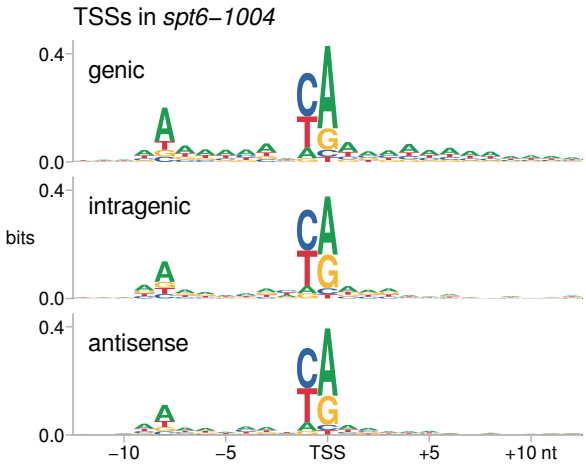


Figure 2.16: Sequence logos depicting information content and sequence preference of TSS-seq reads overlapping genic and intragenic TSS-seq peaks in *spt6-1004*.

2.5.2 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.

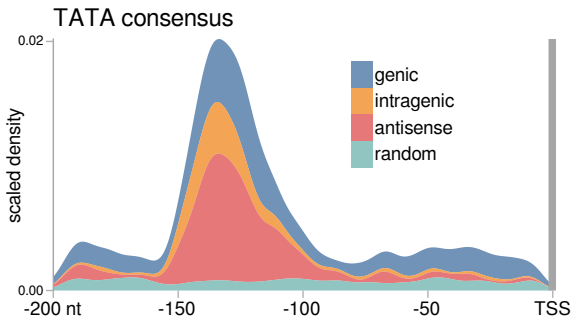


Figure 2.17: Scaled density of occurrences of exact matches to the motif TATAWAWR upstream of TSSs. For each category, a Gaussian kernel density estimate of the positions of motif occurrences is scaled by the number of motif occurrences per region.

2.6 Discussion

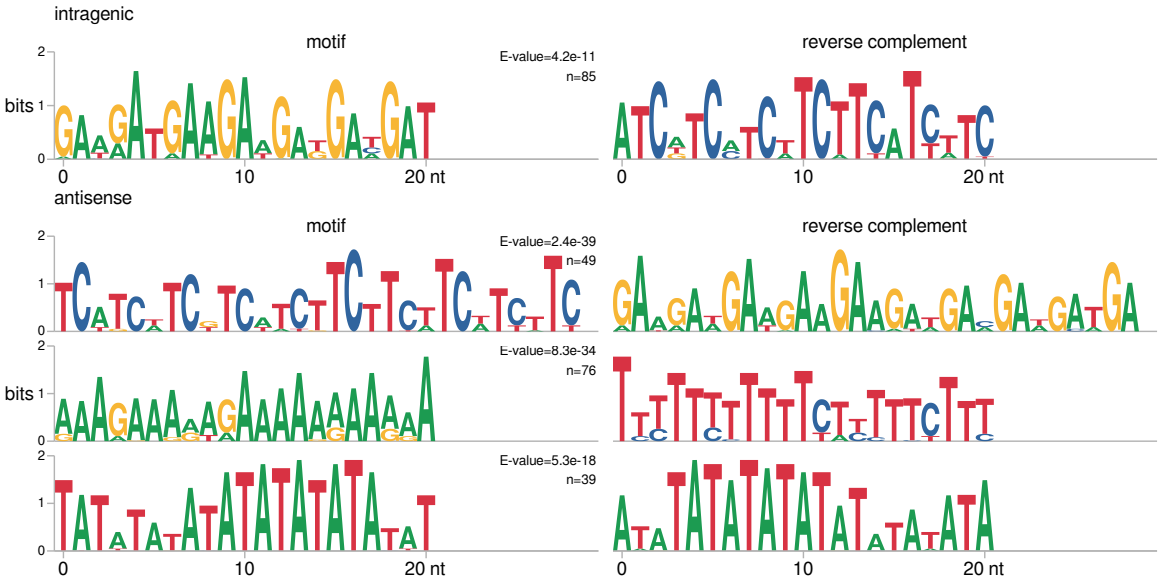


Figure 2.18: Sequence logos of motifs discovered by MEME (Bailey et al., 2015) in the window -100 to +30 bp relative to *spt6-1004* intragenic and antisense TSSs. The number of motif occurrences and the E-value, indicating the expected number of motif occurrences if the input sequences were scrambled,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

2.7 Methods

2.7.1 Yeast strain construction and grown 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 ($OD_{600} = 0.6$), 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.

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

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

2.7.3 Genome builds

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

2.7.4 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 uniquely mapping, non-duplicate *S. pombe* reads. 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.

2.7.4.1 Reannotation of *S. cerevisiae* TSSs using TSS-seq data

TSS-seq coverage from two replicates of a wild-type *S. cerevisiae* strain grown at 30 °C in YPD was averaged and used to adjust the 5′ ends of an annotation of major transcript isoforms based on TIF-seq data (**?**). The 5′ end of the original annotation was changed to the position of maximum TSS-seq signal in a window $\pm$ 250 nt of the original 5′ end if the maximum TSS-seq signal was greater than the 95th percentile of all non-zero TSS-seq signal.

2.7.4.2 TSS-seq peak calling

TSS-seq data representing transcription from a single promoter tends to occur as a cluster of signal distributed over a range of positions, rather than a single nucleotide (Arribere and Gilbert, 2013; Malabat et al., 2015). It is reasonable to consider such a cluster of TSS-seq signal as a single entity, because the signals within the cluster

are usually highly correlated to one another across different conditions. Therefore, to identify TSSs from TSS-seq data and quantify them for downstream analyses such as differential expression, it is necessary to annotate these groups of TSS-seq signal by using the data to perform peak-calling.

At the time of writing, TSS-seq peak calling for a given experimental group was performed by 1-D watershed segmentation of the data for each sample in the group, followed by filtering for reproducibility within the group by the Irreproducible Discovery Rate (IDR) method (Li et al., 2011). First, a smoothed version of the TSS-seq coverage is generated for each sample using an adaptive two-stage kernel density estimation with a discretized Gaussian kernel (**?**). For a given nucleotide, the adaptive kernel bandwidth, σ_{adaptive}, is given by

$$\sigma_{\text{adaptive}} = \sigma_{\text{pilot}} \left(\frac{\rho_{\text{pilot}}}{g} \right)^{-\alpha},$$

where σ_{pilot} is the standard, fixed bandwidth of a Gaussian kernel used to calculate the pilot signal density ρ_{pilot} at that nucleotide, g is the geometric mean of ρ_{pilot} over the whole genome, and α is a parameter in $[0, 1]$ that determines the degree to which the pilot density ρ_{pilot} affects σ_{adaptive}. The adaptive kernel adjusts the kernel bandwidth to be smaller in regions of high signal density and larger in regions of lower signal density, allowing the smoother to better accommodate both 'sharp' TSSs where the signal is distributed over a relatively small window, as well as 'broad' TSSs where the signal is more dispersed. For all analyses in this document, adaptive smoothing was performed with $\sigma_{\text{pilot}} = 10$ and $\alpha = 0.2$.

Following smoothing, an initial set of peaks is formed by assigning all nonzero signal in the original, unsmoothed coverage to the nearest local maximum of the smoothed coverage, and taking the minimum and maximum genomic coordinates of

the original coverage as the peak boundaries for each local maximum of the smoothed coverage. Peaks are then trimmed to the smallest genomic interval that includes 95% of the original coverage, and the probability of the peak begin generated by noise is estimated by a Poisson model where λ, the expected coverage, is the maximum of the expected coverage over the chromosome and the expected coverage in the 2 kb window upstream of the peak (à la the ChIP-seq peak caller MACS2 (Zhang et al., 2008)). Finally, peaks are ranked by their significance under the Poisson model, and a final list of peaks for the group is generated using the IDR method (IDR $= 0.1$) (Li et al., 2011). In brief, IDR compares ranked lists of regions in order to set a cutoff, beyond which the regions are no longer consistent between replicates.

The python script used for 1-D watershed segmentation of TSS-seq data is available as part of the TSS-seq pipeline, and the IDR implementation used in the pipeline is also available on GitHub.

2.7.4.3 TSS differential expression analysis

For TSS-seq differential expression analysis, TSS-seq peak-calling was performed as described above for both *S. cerevisiae* and the *S. pombe* spike-in. The read counts for each peak in each condition were used as the input to differential expression analysis by DESeq2 (Love et al., 2014), with the alternative hypothesis $|\log_2(\text{fold-change})| > 1.5$ and a false discovery rate of 0.1. To normalize by spike-in, the size factors of the *S. pombe* spike-in counts were used as the size factors for *S. cerevisiae*, although we note that due to the median of ratios normalization used in DESeq2, the major TSS-seq results of this work are still observed when *S. cerevisiae* size factors are used.

2.7.4.4 Classification of TSS-seq peaks into genomic categories

TSS-seq peaks were assigned to genomic categories based on their position relative to the transcript annotation described above and an annotation of all verified open reading frames (ORFs) and blocked reading frames in *S. cerevisiae* (**??**). First, 'genic' regions were defined as follows: If a gene was present in both the transcript and ORF annotations, the genic region was defined as the interval (annotated TSS-30 nt, start codon). If gene was present in the transcript annotation but not the ORF annotation, the genic region was defined as the interval (annotated TSS - 30 nt, annotated TSS + 30 nt). If a gene was present only in the ORF annotation, the genic region was defined as the interval (start codon - 30 nt, start codon). For the purposes of peak classification, regions were considered overlapping if they had at least one base of overlap. TSS-seq peaks were classified as genic if they overlapped a genic region on the same strand. Peaks were classified as intragenic if they were not classified as as genic peak, and their summit position overlapped an open or closed reading frame on the same strand. Peaks were classified as antisense if their summit position overlapped a transcript on the opposite strand. Finally, peaks were classified as intergenic if they did not overlap a transcript, reading frame, or genic region on either strand.

2.7.4.5 TSS information content and sequence composition

TSS-seq alignments were pooled for all replicates in a condition, and the DNA sequence flanking the position of every read overlapping TSS-seq peaks of a particular genomic category was extracted using SAMtools (**?**) and bedtools (Quinlan and Hall, 2010). The information content and sequence composition was quantified using WebLogo (**?**), with the zeroth-order Markov model of the *S. cerevisiae* genomic se-

quence as the background composition. Sequence logos were plotting using helper functions from ggseqlogo (**?**).

2.7.5 ChIP-nexus data analysis

An up-to-date version of the Snakemake (Köster and Rahmann, 2012) workflow used to process ChIP-nexus libraries is maintained at github.com/winston-lab/chip-nexus. At the time of writing, filtering for reads containing the constant region of the adapter on the 5′ end of the read, 3′ adapter removal, and 3′ quality trimming were performed using cutadapt (Martin, 2011). The random pentamer molecular barcode on the 5′ end of the read was then removed and processed using a custom Python script modified from **?**. Reads were aligned to the combined *S. cerevisiae* and *S. pombe* genomes using Bowtie2 (Langmead and Salzberg, 2012), 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 modified from **?**. Coverage of the 5′-most base, corresponding to the point of crosslinking, was extracted using bedtools genomecov (Quinlan and Hall, 2010). The median fragment size estimated by MACS2 (Zhang et al., 2008) over all samples was used to generate coverage of factor protection and fragment midpoints, by extending reads to the fragment size, or by shifting reads by half the fragment size, respectively. Coverage was normalized to the total number of reads uniquely mapping to *S. cerevisiae*. Quality statistics of raw, cleaned, non-aligning, and uniquely aligning non-duplicate reads were assessed using FastQC **?**.

2.7.5.1 A note on ChIP-nexus peak calling

2.7.5.2 TFIIB ChIP-nexus differential binding analysis

For TFIIB ChIP-nexus differential binding analysis, TFIIB peaks were called by MACS2 and IDR filtering as described above. A non-redundant list of peaks called in the condition and control groups being compared was generated using bedtools multiinter (Quinlan and Hall, 2010), and the counts of fragment midpoints for each peak in each sample were used as the input to differential binding analysis by DESeq2 (Love et al., 2014), with the alternative hypothesis $|\log_2(\text{fold-change})| > 1.5$ and a false discovery rate of 0.1. For estimation of change in TFIIB binding upstream of TSS-seq peaks, TFIIB fragment midpoint counts in the window extending 200 bp upstream of the TSS-seq peak summit were used as the input to DESeq2. *S. cerevisiae* counts were used for size factor calculation.

2.7.5.3 Classification of TFIIB ChIP-nexus peaks into genomic categories

As for TSS-seq peaks, TFIIB ChIP-nexus peaks were assigned to genomic categories based on their position relative to the transcript annotation described above, an annotation of all verified open reading frames (ORFs) and blocked reading frames (**??**), and an annotation of 'genic' regions derived from the transcript and ORF annotations. TFIIB ChIP-nexus peaks were classified as genic if they overlapped a genic region. Peaks were classified as intragenic if they were not classified as a genic peak, and the entire peak overlapped an open or closed reading frame. Finally, peaks were classified as intergenic if they did not overlap a transcript, reading frame, or genic region.

2.7.6 MNase-seq data analysis

2.7.6.1 Nucleosome quantification

2.7.6.2 Clustering of MNase-seq signal at *spt6-1004* intragenic TSSs

2.7.7 Motif enrichment

2.7.8 *De novo* motif discover motif discovery

2.8 Bibliography

Adkins, M. W. and Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of spt6-mediated chromatin reassembly of promoter regions. *Molecular Cell*, 21(3):405 – 416. 2.2

Andrulis, E. D., Guzmán, E., Döring, P., Werner, J., and Lis, J. T. (2000). High-resolution localization of drosophila spt5 and spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes & Development*, 14(20):2635–2649. 2.2

Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., and Lis, J. T. (2009). Spt6 enhances the elongation rate of rna polymerase ii in vivo. *The EMBO Journal*, 28(8):1067–1077. 2.2

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, 2.7.4.2

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

Begum, N. A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone spt6 is required for activation-induced cytidine deaminase target determination through h3k4me3 regulation. *Journal of Biological Chemistry*, 287(39):32415–32429. 2.2

Bortvin, A. and Winston, F. (1996). Evidence that spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267):1473–1476. 2.2, 2.4

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005). Histone h3 methylation by set2 directs deacetylation of coding regions by rpd3s to suppress spurious intragenic transcription. *Cell*, 123(4):581 – 592. 2.2

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. 2.4

Chen, S., Ma, J., Wu, F., Xiong, L.-j., Ma, H., Xu, W., Lv, R., Li, X., Villen, J., Gygi, S. P., Liu, X. S., and Shi, Y. (2012). The histone h3 lys 27 demethylase jmjd3 regulates gene expression by impacting transcriptional elongation. *Genes & Development*, 26(12):1364–1375. 2.2

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), 2.2, 2.2, 2.7, 2.7.1

Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. (2006). The bur1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by set2. *Molecular and Cellular Biology*, 26(8):3029–3038. 2.2

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. 2.4

Close, D., Johnson, S. J., Sdano, M. A., McDonald, S. M., Robinson, H., Formosa, T., and Hill, C. P. (2011). Crystal structures of the s. cerevisiae spt6 core and c-terminal tandem sh2 domain. *Journal of Molecular Biology*, 408(4):697 – 713. 2.2

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, 2.2

Diebold, M.-L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an iws1/spt6 complex reveals an interaction domain conserved in tfiis, elongin a and med26. *The EMBO Journal*, 29(23):3979–3991. 2.2

Diebold, M.-L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem sh2 enables interaction of elongation factor spt6 with rna polymerase ii. *Journal of Biological Chemistry*, 285(49):38389–38398. 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*. 2.7.2

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]. 2.2, 2.4

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., Sato, H., Yamaguchi, Y., Mandal, S. S., Reinberg, D., Wada, T., and Handa, H. (2004). Human spt6 stimulates transcription elongation by rna polymerase ii in vitro. *Molecular and Cellular Biology*, 24(8):3324–3336. 2.2

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 –. 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. 2.2, 2.4, 2.4

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. 2.2, 2.4

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, 2.2, 2.4

Kaplan, C. D., Morris, J. R., Wu, C.-t., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in d. melanogaster. *Genes & Development*, 14(20):2623–2634. 2.2

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. 2.7.4

Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522. 2.7.4, 2.7.5

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9:357 EP –. 2.7.5

Li, Q., Brown, J. B., Huang, H., and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.*, 5(3):1752–1779. 2.7.4.2

Li, S., Almeida, A. R., Radebaugh, C. A., Zhang, L., Chen, X., Huang, L., Thurston, A. K., Kalashnikova, A. A., Hansen, J. C., Luger, K., and Stargell, L. A. (2018). The elongation factor spn1 is a multi-functional chromatin binding protein. *Nucleic Acids Research*, 46(5):2321–2334. 2.2

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. 2.2

Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011). Solution structure of tandem sh2 domains from spt6 protein and their

binding to the phosphorylated rna polymerase ii c-terminal domain. *Journal of Biological Chemistry*, 286(33):29218–29226. 2.2

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15(12):550. 2.7.4.3, 2.7.5.2

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, 2.5.1, 2.7.4.2

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12. 2.7.4, 2.7.5

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general rna polymerase ii transcription complex. *Nature Structural & Molecular Biology*, 17:1272–1278. 2.2

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones spt6 and fact collaborate to assemble, inspect, and maintain chromatin structure in saccharomyces cerevisiae. *Genetics*, 201(3):1031–1045. 2.2

McDonald, S. M., Close, D., Xin, H., Formosa, T., and Hill, C. P. (2010). Structure and biological importance of the spn1-spt6 interaction, and its regulatory role in nucleosome binding. *Molecular Cell*, 40(5):725 – 735. 2.2

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. 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. 2.2, 2.4

Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842. 2.7.4, 2.7.4.5, 2.7.5, 2.7.5.2

Sdano, M. A., Fulcher, J. M., Palani, S., Chandrasekharan, M. B., Parnell, T. J., Whitby, F. G., Formosa, T., and Hill, C. P. (2017). A novel sh2 recognition mechanism recruits spt6 to the doubly phosphorylated rna polymerase ii linker at sites of transcription. *eLife*, 6:e28723. 2.2

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. 2.4

Sun, M., Larivière, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem sh2 domain in transcription elongation factor spt6 binds the phosphorylated rna polymerase ii c-terminal repeat domain (ctd). *Journal of Biological Chemistry*, 285(53):41597–41603. 2.2

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), 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, 2.2, 2.4

Wang, A. H., Juan, A. H., Ko, K. D., Tsai, P.-F., Zare, H., Dell'Orso, S., and Sartorelli, V. (2017). The elongation factor spt6 maintains esc pluripotency by controlling super-enhancers and counteracting polycomb proteins. *Molecular Cell*, 68(2):398 – 413.e6. 2.2

Wang, A. H., Zare, H., Mousavi, K., Wang, C., Moravec, C. E., Sirotkin, H. I., Ge, K., Gutierrez-Cruz, G., and Sartorelli, V. (2013). The histone chaperone spt6 coordinates histone h3k27 demethylation and myogenesis. *The EMBO Journal*, 32(8):1075–1086. 2.2

Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., and Jones, K. A. (2007). The spt6 sh2 domain binds ser2-p rnapii to direct iws1-dependent mrna splicing and export. *Genes & Development*, 21(2):160–174. 2.2

Yoh, S. M., Lucas, J. S., and Jones, K. A. (2008). The iws1:spt6:ctd complex controls cotranscriptional mrna biosynthesis and hypb/setd2-mediated histone h3k36 methylation. *Genes & Development*, 22(24):3422–3434. 2.2

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., and Mellor, J. (2008). Roles for ctk1 and spt6 in regulating the different methylation states of histone h3 lysine 36. *Molecular and Cellular Biology*, 28(16):4915–4926. 2.2

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of chip-seq (macs). *Genome Biology*, 9(9):R137. 2.7.4.2, 2.7.5

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

Relevant information about Spt5 is summarized as follows (Shetty et al., 2017):

- Spt5 is the only transcription factor known to be conserved in all three domains of life (Hartzog and Fu, 2013; Werner, 2012).
- Spt5 co-localizes with elongating RNA Pol II (Mayer et al., 2010; Rahl et al., 2010).
- Spt5 binds over the Pol II clamp domain, likely stabilizing the elongation complex (Hirtreiter et al., 2010; Klein et al., 2011; Martinez-Rucobo et al., 2011).
- Spt5 physically recruits factors to the elongating transcription complex, in a manner dependent on the modification status of its C-terminal region (CTR) (Hartzog and Fu, 2013):
  - in its unphosphorylated state, the CTR aids in recruiting the mRNA capping enzyme (Doamekpor et al., 2014, 2015; Schneider et al., 2010; Wen

and Shatkin, 1999)

- in its phosphorylated state, the CTR recruits the Paf1 complex, which is important for Pol II elongation (Liu et al., 2009; Mbogning et al., 2013; Wier et al., 2013; Zhou et al., 2009)
- Spt5 helps to recruit mRNA 3' end processing factors (Mayer et al., 2012; Stadelmayer et al., 2014; Yamamoto et al., 2014).
- Spt5 helps to recruit the Rpd3S histone deacetylase complex (Drouin et al., 2010).

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.

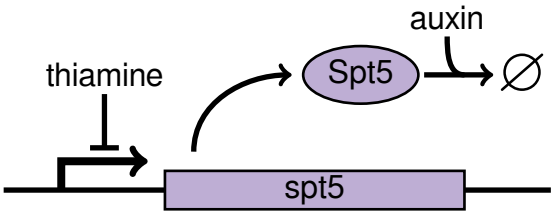


Figure 3.1: Diagram of the dual-shutoff system used to deplete Spt5 from *S. pombe*. Spt5 is expressed from a thiamine-repressible promoter, and tagged with an auxin-inducible degron tag for specific degradation upon addition of auxin.

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 gravida sollicitudin, 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.

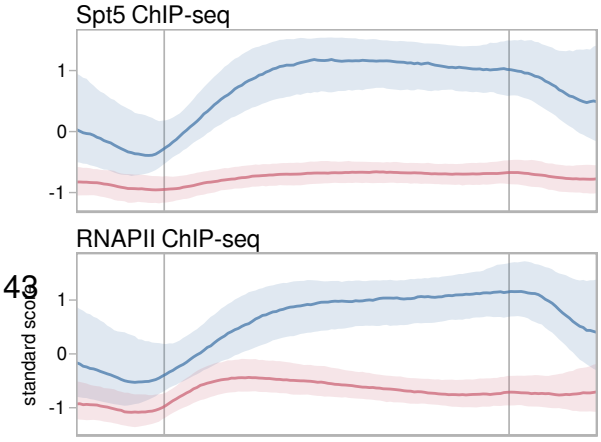
Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hen-

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 consectetuer.

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, tel-



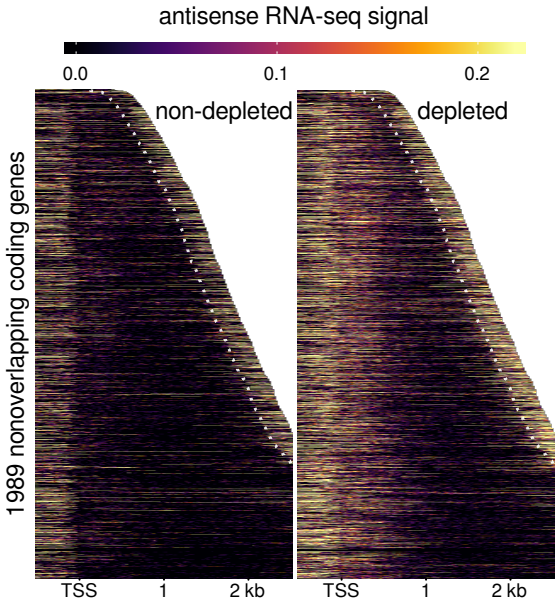


Figure 3.4: Caption wsdasdr zzzz.

lus. 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

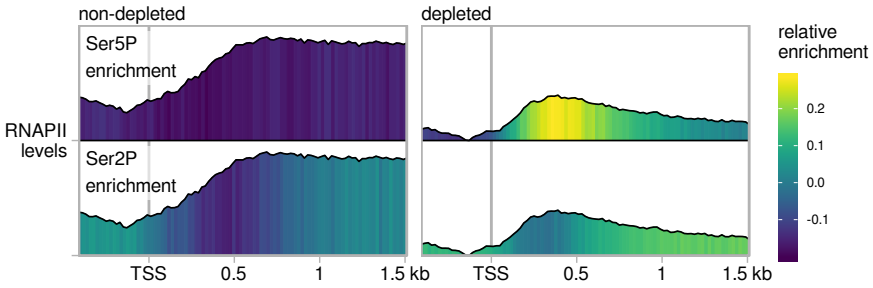


Figure 3.3: Caption wsdasdr zzzz.

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.

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

3.4 TSS-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

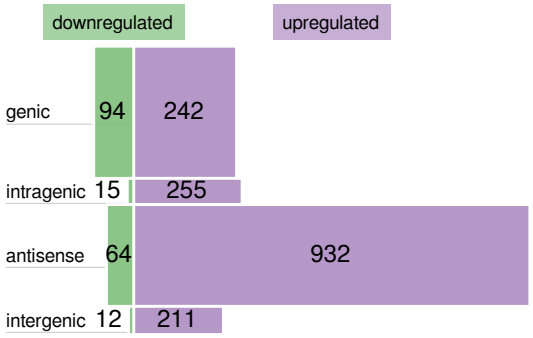


Figure 3.5: Caption wsdasdr zzzz.

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

Figure 3.6: Caption wsdasdr zzzz.

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 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.

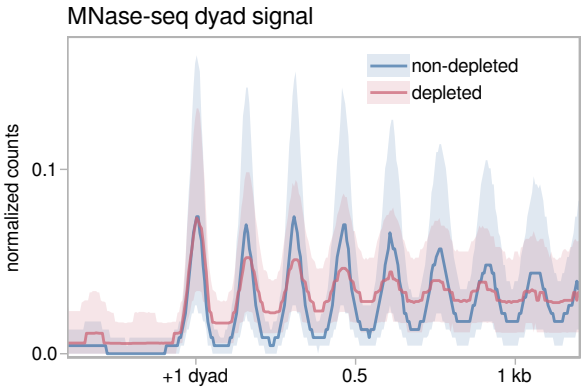


Figure 3.7: Caption wsdasdr zzzz.

Figure 3.8: Caption wsdasdr zzzz.

Figure 3.9: Caption wsdasdr zzzz.

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

3.6 Sequence motifs enriched at antisense TSSs

3.7 Discussion

3.8 Methods

3.8.1 A note on spike-in normalization for ChIP-seq experiments with input samples

In the course of determining how to do spike-in normalization for ChIP-seq libraries, I discovered the following error in a published spike-in normalization method. Throughout the following explanation, I use 'experimental' and 'spike-in' to refer to the two genomes present in the experiment, e.g., experimental signal and spike-in signal.

The goal when including spike-ins in a ChIP-seq experiment is to be able to normalize the experimental signal, such that the normalized signal is proportional to the absolute abundance of the factor being immunoprecipitated. A straightforward method to accomplish this normalization is to linearly scale the experimental signal of a library by a normalization factor, which we will call α. To calculate α for each library, we can use the fact that a normalized 'spike-in signal' should be the same for all libraries, since the biological state of the spike-in cells is the same in all libraries. The key to correctly determining α is defining exactly what this spike-in signal is.

The measurement we begin with for determination of the spike-in signal of a library is the number of reads in the library which map uniquely to the spike-in genome

(R_{spike}). This value will vary based on two factors: the sequencing depth of the library, and the proportion of cells which were spike-in cells (ϕ):

$R_{\text{spike}} \equiv$ the number of reads in the library mapping uniquely to the spike-in genome;

$\phi \equiv$ the proportion of spike-in cells in the sample.

However, the derivation of α is more easily understood in terms of absolute cell numbers rather than ϕ:

$C_{\text{exp}} \equiv$ the number of experimental cells used to prepare a library;

$C_{\text{spike}} \equiv$ the number of spike-in cells used to prepare a library.

We can express the **number of spike-in reads per spike-in cell** by simply taking the fraction $\frac{R_{\text{spike}}}{C_{\text{spike}}}$. We know that the biological state of a spike-in cell is the same regardless of which sample it belongs to, so we *could* set $\frac{R_{\text{spike}}}{C_{\text{spike}}}$ equal to all samples in order to calculate α. However, this would not account for differences in ϕ between samples: Two libraries representing the same condition and sequenced to the same depth should have equivalent values of $\frac{R_{\text{spike}}}{C_{\text{spike}}}$, which does not hold true if they differed in the proportion of spike-in added.

The metric for 'spike-in signal' that leads to the correct expression for α is the **number of spike-in reads per spike-in cell *per experimental cell***:

$$\frac{\frac{R_{\text{spike}}}{C_{\text{spike}}}}{C_{\text{exp}}} = \frac{R_{\text{spike}} C_{\text{exp}}}{C_{\text{spike}}}.$$

From here, it's simple to calculate α by setting this value to be equal for all samples. Since the actual value of the spike-in signal doesn't matter as long as it is equal for all libraries, we can arbitrarily set it to 1 for convenience:

$$\alpha \frac{R_{\text{spike}} C_{\text{exp}}}{C_{\text{spike}}} = 1$$

$$\alpha = \frac{C_{\text{spike}}}{R_{\text{spike}} C_{\text{exp}}}.$$

Notice that only the ratio of spike-in to experimental cells is needed to calculate α, and not the absolute number of spike-in and experimental cells. We can rewrite this expression in terms of ϕ, the proportion of the sample that was spike-in cells:

$$\phi = \frac{C_{\text{spike}}}{C_{\text{spike}} + C_{\text{exp}}}$$

$$C_{\text{spike}} = \phi \left(C_{\text{spike}} + C_{\text{exp}} \right)$$

$$C_{\text{spike}} \left(1 - \phi \right) = \phi C_{\text{exp}}$$

$$\frac{C_{\text{spike}}}{C_{\text{exp}}} = \frac{\phi}{1 - \phi}$$

$$\alpha = \frac{C_{\text{spike}}}{R_{\text{spike}} C_{\text{exp}}}$$

$$\alpha = \frac{\phi}{R_{\text{spike}} \left(1 - \phi \right)}.$$

This form for α differs from the one presented in **?** with no derivation:

$$\alpha = \frac{\phi}{R_{\text{spike}} \left(1 - \phi \right)} \qquad \alpha_{\text{orlando}} = \frac{\phi}{R_{\text{spike}}}.$$

Working through a few examples with both versions of α will reveal that α_{orlando} leads to incorrect normalization when ϕ is not equivalent for all samples.

In the first example, we will vary sequencing depth between two libraries, keeping everything else constant. Consider a single ChIP library prep in which 20% of the cells were spike-in cells (i.e., $\phi = 0.2$). The library is then unevenly split into two aliquots

and sequenced. One library has four times the reads of the other library.

$$R_{\text{spike}_1} = 1 \qquad R_{\text{spike}_2} = 4$$

$$R_{\text{exp}_1} = 4 \qquad R_{\text{exp}_2} = 16$$

$$\alpha_1 = \frac{\phi}{R_{\text{spike}_1}(1-\phi)} \quad \alpha_2 = \frac{\phi}{R_{\text{spike}_2}(1-\phi)} \quad \alpha_{\text{orlando}_1} = \frac{\phi}{R_{\text{spike}_1}} \quad \alpha_{\text{orlando}_2} = \frac{\phi}{R_{\text{spike}_2}}$$

$$\alpha_1 = \frac{0.2}{1\,(0.8)} \quad \alpha_2 = \frac{0.2}{4\,(0.8)} \quad \alpha_{\text{orlando}_1} = \frac{0.2}{1} \quad \alpha_{\text{orlando}_2} = \frac{0.2}{4}$$

$$\alpha_1 = \frac{4}{16} \quad \alpha_2 = \frac{1}{16} \quad \alpha_{\text{orlando}_1} = \frac{4}{20} \quad \alpha_{\text{orlando}_2} = \frac{1}{20}.$$

The total levels of spike-in normalized experimental signal can be found for each library by multiplying α by R_{exp}, for our version of α,

$$\text{signal}_1 = \alpha_1 R_{\text{exp}_1} \qquad \text{signal}_2 = \alpha_2 R_{\text{exp}_2}$$

$$\text{signal}_1 = \frac{4}{16}(4) \qquad \text{signal}_2 = \frac{1}{16}(16)$$

$$\text{signal}_1 = 1 \qquad \text{signal}_2 = 1$$

and for α_{orlando}:

$$\text{signal}_{\text{orlando}_1} = \alpha_{\text{orlando}_1} R_{\text{exp}_1} \qquad \text{signal}_{\text{orlando}_2} = \alpha_{\text{orlando}_2} R_{\text{exp}_2}$$

$$\text{signal}_{\text{orlando}_1} = \frac{4}{20}(4) \qquad \text{signal}_{\text{orlando}_2} = \frac{1}{20}(16)$$

$$\text{signal}_{\text{orlando}_1} = 0.8 \qquad \text{signal}_{\text{orlando}_2} = 0.8$$

Only the relative abundances within normalization methods matter, so in this case both calculations correctly normalized for library size and say that the normalized signal in the two libraries are the same.

Now let's consider two libraries from two different conditions with $\phi = 0.1$. In condition 2, there is a known global decrease in experimental signal expected. This time, we will skip the algebra:

$$R_{\text{spike}_1} = 1 \qquad R_{\text{spike}_2} = 4$$

$$R_{\text{exp}_1} = 9 \qquad R_{\text{exp}_2} = 6$$

$$\alpha_1 = \frac{4}{36} \qquad \alpha_2 = \frac{1}{36} \qquad \alpha_{\text{orlando}_1} = \frac{4}{40} \qquad \alpha_{\text{orlando}_2} = \frac{1}{40}$$

$$\text{signal}_1 = 1 \qquad \text{signal}_2 = 1/6 \qquad \text{signal}_{\text{orlando}_1} = 0.9 \qquad \text{signal}_{\text{orlando}_2} = 0.15$$

Both methods correctly detect that experimental signal levels in library 2 are 1/6th that of library 1.

Finally, let's consider two libraries from the same condition which were spiked in with different amounts of spike-in cells. Both libraries are sequenced to the same depth. Since the libraries are from the same condition, we expect their total experimental signal to be the same after normalization, even though they had different

amounts of spike-in added.

$$\phi_1 = 0.2 \qquad \phi_2 = 0.4$$

$$R_{\text{spike}_1} = 2 \qquad R_{\text{spike}_2} = 4$$

$$R_{\text{exp}_1} = 8 \qquad R_{\text{exp}_2} = 6$$

$$\alpha_1 = \frac{\phi_1}{R_{\text{spike}_1}(1-\phi_1)} \quad \alpha_2 = \frac{\phi_2}{R_{\text{spike}_2}(1-\phi_2)} \quad \alpha_{\text{orlando}_1} = \frac{\phi_1}{R_{\text{spike}_1}} \quad \alpha_{\text{orlando}_2} = \frac{\phi_2}{R_{\text{spike}_2}}$$

$$\alpha_1 = \frac{0.2}{2(0.8)} \quad \alpha_2 = \frac{0.4}{4(0.6)} \quad \alpha_{\text{orlando}_1} = \frac{0.2}{2} \quad \alpha_{\text{orlando}_2} = \frac{0.4}{4}$$

$$\alpha_1 = \frac{3}{24} \quad \alpha_2 = \frac{4}{24} \quad \alpha_{\text{orlando}_1} = \frac{1}{10} \quad \alpha_{\text{orlando}_2} = \frac{1}{10}$$

$$\text{signal}_1 = \alpha_1 R_{\text{exp}_1} \qquad \text{signal}_2 = \alpha_2 R_{\text{exp}_2}$$

$$\text{signal}_1 = \frac{3}{24}(8) \qquad \text{signal}_2 = \frac{4}{24}(6)$$

$$\text{signal}_1 = 1 \qquad \text{signal}_2 = 1$$

$$\text{signal}_{\text{orlando}_1} = \alpha_{\text{orlando}_1} R_{\text{exp}_1} \qquad \text{signal}_{\text{orlando}_2} = \alpha_{\text{orlando}_2} R_{\text{exp}_2}$$

$$\text{signal}_{\text{orlando}_1} = \frac{1}{10}(8) \qquad \text{signal}_{\text{orlando}_2} = \frac{1}{10}(6)$$

$$\text{signal}_{\text{orlando}_1} = 0.8 \qquad \text{signal}_{\text{orlando}_2} = 0.6$$

Here, our method correctly normalizes the two samples to the same total experimental signal while using the Orlando α results in an apparent decrease in signal in library

2. This is because the Orlando α fails to account for the fact that when you add more spike-in to a sample, you necessarily decrease the proportion of the sample that is experimental. In most experiments with spike-ins, this isn't really a problem because we assume that ϕ is the same for all samples. However, with ChIP-seq experiments that include input samples, if we assume that the experimental and spike-in input sample read counts are proportional to the amounts of experimental and spike-in cells mixed, we can plug these values in for values of ϕ to get a more reliable estimation of experimental signal levels. In this case, it becomes important to use the correct equation for α.

So, putting everything together, here's how I use the spike-in to normalize an IP ChIP-seq library paired with an input ChIP-seq library.

As stated above, we assume that the experimental and spike-in read counts in the input sample are proportional to the numbers of experimental and spike-in cells used to prepare the library:

$$R_{\text{input}_{\text{exp}}} \propto C_{\text{exp}},$$

$$R_{\text{input}_{\text{spike}}} \propto C_{\text{spike}}$$

Therefore, we can plug these values in for C for both the input and IP libraries (using the form of α without ϕ):

$$\alpha_{\text{input}} = \frac{C_{\text{input}_{\text{spike}}}}{R_{\text{input}_{\text{spike}}} C_{\text{input}_{\text{exp}}}} \qquad \alpha_{\text{IP}} = \frac{C_{\text{input}_{\text{spike}}}}{R_{\text{IP}_{\text{spike}}} C_{\text{input}_{\text{exp}}}}$$

$$\alpha_{\text{input}} \propto \frac{R_{\text{input}_{\text{spike}}}}{R_{\text{input}_{\text{spike}}} R_{\text{input}_{\text{exp}}}} \qquad \alpha_{\text{IP}} \propto \frac{R_{\text{input}_{\text{spike}}}}{R_{\text{IP}_{\text{spike}}} R_{\text{input}_{\text{exp}}}}$$

$$\alpha_{\text{input}} \propto \frac{1}{R_{\text{input}_{\text{exp}}}}$$

Notice how α_{input} reduces down to normalizing by the experimental library size, with no dependence at all on the spike-in. This makes sense because the input always represents the same state, regardless of how much spike-in is added to it. The function of the spike-in in the input is only to allow us to estimate abundances in the corresponding IP library. Rewriting α_{IP} in the form

$$\alpha_{\text{IP}} \propto \frac{1}{R_{\text{IP}_{\text{spike}}} \frac{R_{\text{input}_{\text{exp}}}}{R_{\text{input}_{\text{spike}}}}}$$

shows that α_{IP} will basically scale the experimental IP signal to the same scale as the experimental input signal, using the spike-in as a link between the two samples. This makes it natural to subtract the normalized input signal from the normalized IP signal: since they are on the same scale, the resulting coverage can be interpreted as reporting how much more IP signal was detected than was expected based on the input.

3.9 Bibliography

Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S., and Lima, C. D. (2014). How an mrna capping enzyme reads distinct rna polymerase ii and spt5 ctd phosphorylation codes. *Genes & Development*, 28(12):1323–1336. 3.2

Doamekpor, S. K., Schwer, B., Sanchez, A. M., Shuman, S., and Lima, C. D. (2015). Fission yeast rna triphosphatase reads an spt5 ctd code. *RNA*, 21(1):113–123. 3.2

Drouin, S., Laramée, L., Jacques, P.-�., Forest, A., Bergeron, M., and Robert, F. (2010). Dsif and rna polymerase ii ctd phosphorylation coordinate the recruitment of rpd3s to actively transcribed genes. *PLOS Genetics*, 6(10):1–12. 3.2

Hartzog, G. A. and Fu, J. (2013). The spt4–spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(1):105 – 115. RNA polymerase II Transcript Elongation. 3.2

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the rna polymerase clamp coiled-coil motif. *Nucleic Acids Research*, 38(12):4040–4051. 3.2

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., and Murakami, K. S. (2011). Rna polymerase and transcription elongation factor spt4/5 complex structure. *Proceedings of the National Academy of Sciences*, 108(2):546–550. 3.2

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor spt5 by yeast bur1 kinase stimulates recruitment of the paf complex. *Molecular and Cellular Biology*, 29(17):4852–4863. 3.2

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., and Cramer, P. (2011). Architecture of the rna polymerase–spt4/5 complex and basis of universal transcription processivity. *The EMBO Journal*, 30(7):1302–1310. 3.2

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general rna polymerase ii transcription complex. *Nature Structural & Molecular Biology*, 17:1272–1278. 3.2

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D. E., and Cramer, P. (2012). The spt5 c-terminal region recruits yeast 3' rna cleavage factor i. *Molecular and Cellular Biology*, 32(7):1321–1331. 3.2

Mbogning, J., Nagy, S., Pagé, V., Schwer, B., Shuman, S., Fisher, R. P., and Tanny, J. C. (2013). The paf complex and prf1/rtf1 delineate distinct cdk9-dependent pathways regulating transcription elongation in fission yeast. *PLOS Genetics*, 9(12):1–14. 3.2

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., and Young, R. A. (2010). c-myc regulates transcriptional pause release. *Cell*, 141(3):432 – 445. 3.2

Schneider, S., Pei, Y., Shuman, S., and Schwer, B. (2010). Separable functions of the fission yeast spt5 carboxyl-terminal domain (ctd) in capping enzyme binding and transcription elongation overlap with those of the rna polymerase ii ctd. *Molecular and Cellular Biology*, 30(10):2353–2364. 3.2

Shetty, A., Kallgren, S. P., Demel, C., Maier, K. C., Spatt, D., Alver, B. H., Cramer, P., Park, P. J., and Winston, F. (2017). Spt5 plays vital roles in the control of sense and antisense transcription elongation. *Molecular Cell*, 66(1):77 – 88.e5. 3.2

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., Parrinello, H., Cuvier, O., and Benkirane, M. (2014). Integrator complex regulates nelf-mediated rna polymerase ii pause/release and processivity at coding genes. *Nature Communications*, 5:5531 EP –. Article. 3.2

Wen, Y. and Shatkin, A. J. (1999). Transcription elongation factor hspt5 stimulates mrna capping. *Genes & Development*, 13(14):1774–1779. 3.2

Werner, F. (2012). A nexus for gene expression—molecular mechanisms of spt5 and nusg in the three domains of life. *Journal of Molecular Biology*, 417(1):13 – 27. 3.2

Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M., and VanDemark, A. P. (2013). Structural basis for spt5-mediated recruitment of the paf1 complex to chromatin. 3.2

Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., and Yamaguchi, Y. (2014). Dsif and nelf interact with integrator to specify the correct post-transcriptional fate of snrna genes. *Nature Communications*, 5:4263 EP –. Article. 3.2

Zhou, K., Kuo, W. H. W., Fillingham, J., and Greenblatt, J. F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast bur kinase substrate spt5. *Proceedings of the National Academy of Sciences*, 106(17):6956–6961. 3.2

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 Discussion

4.9 Methods

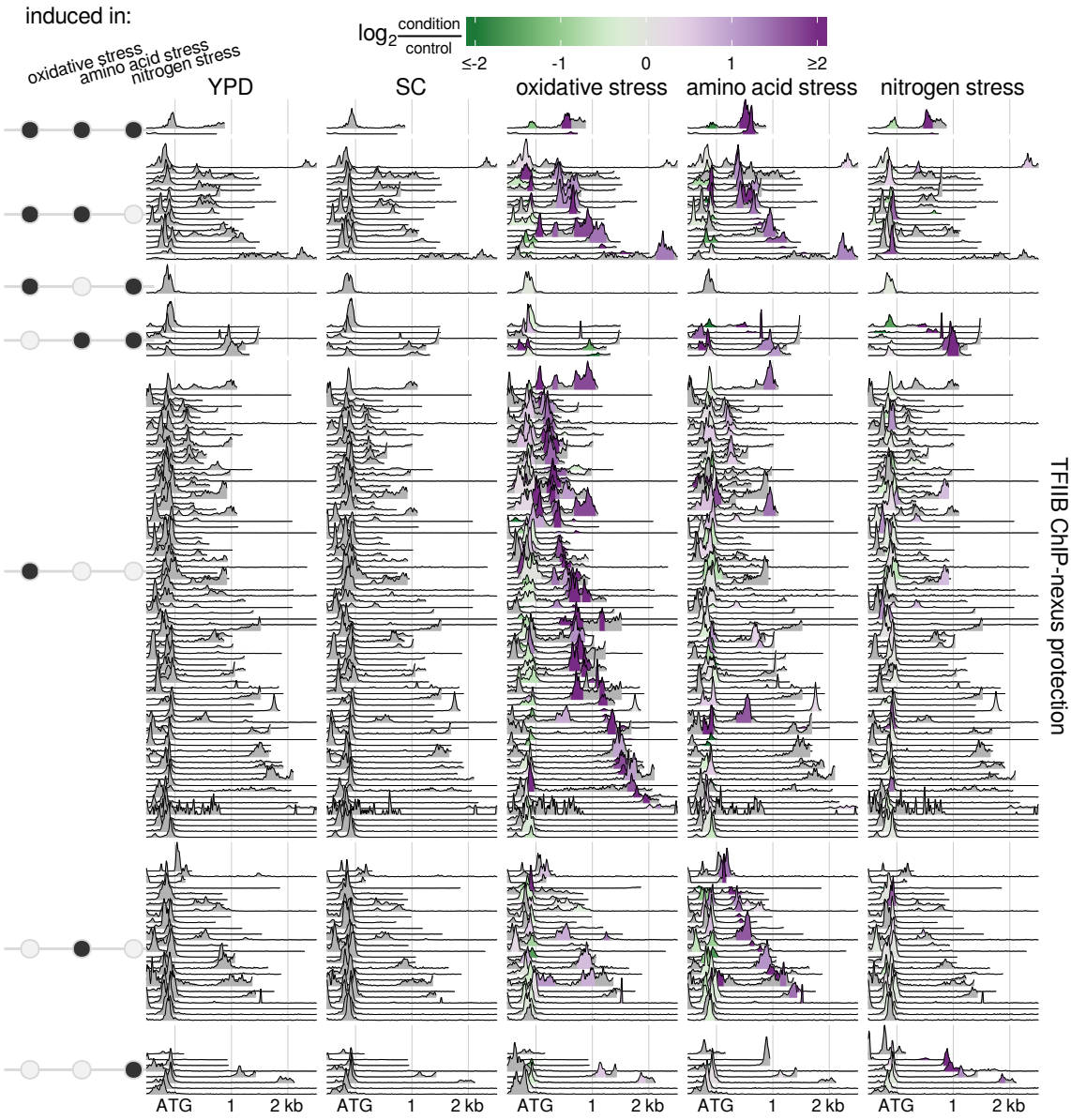


Figure 4.1: Relative TFIIB ChIP-nexus protection over all genes with an intragenic TFIIB peak significantly induced in one or more of the stress conditions tested, as depicted in the left panel. Genes are aligned by start codon, and are sorted within each group by the distance from the start codon to the summit of the induced intragenic TFIIB peak. Data are shown for each gene up to the stop codon of the gene. Regions where TFIIB peaks are called are shaded in the stress conditions according to the fold-change of the peak relative to the corresponding control condition.

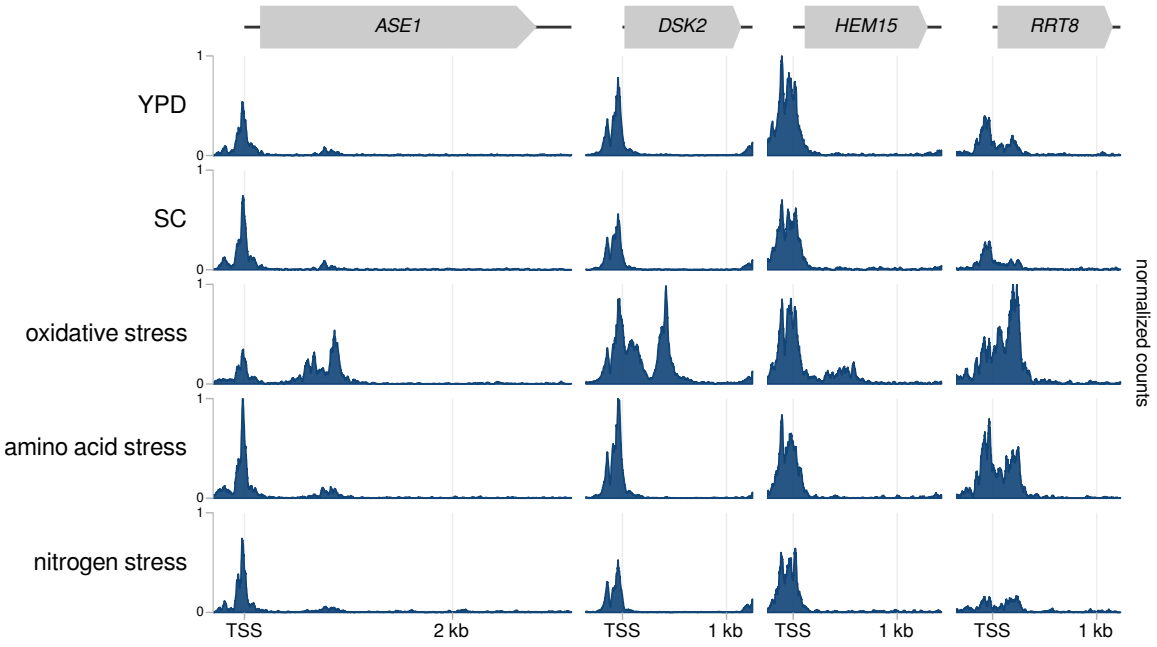


Figure 4.2: Caption asdflkj asldkfjlkj.

4.10 Bibliography

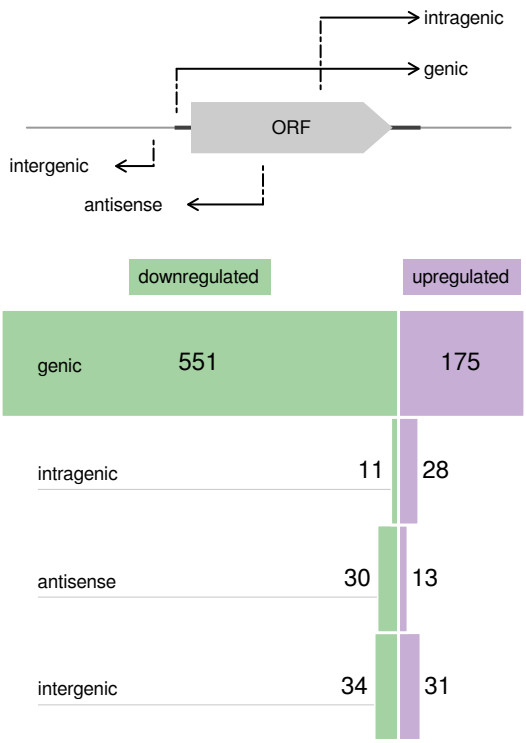


Figure 4.3: Caption dsafklj asldkfjlkj.

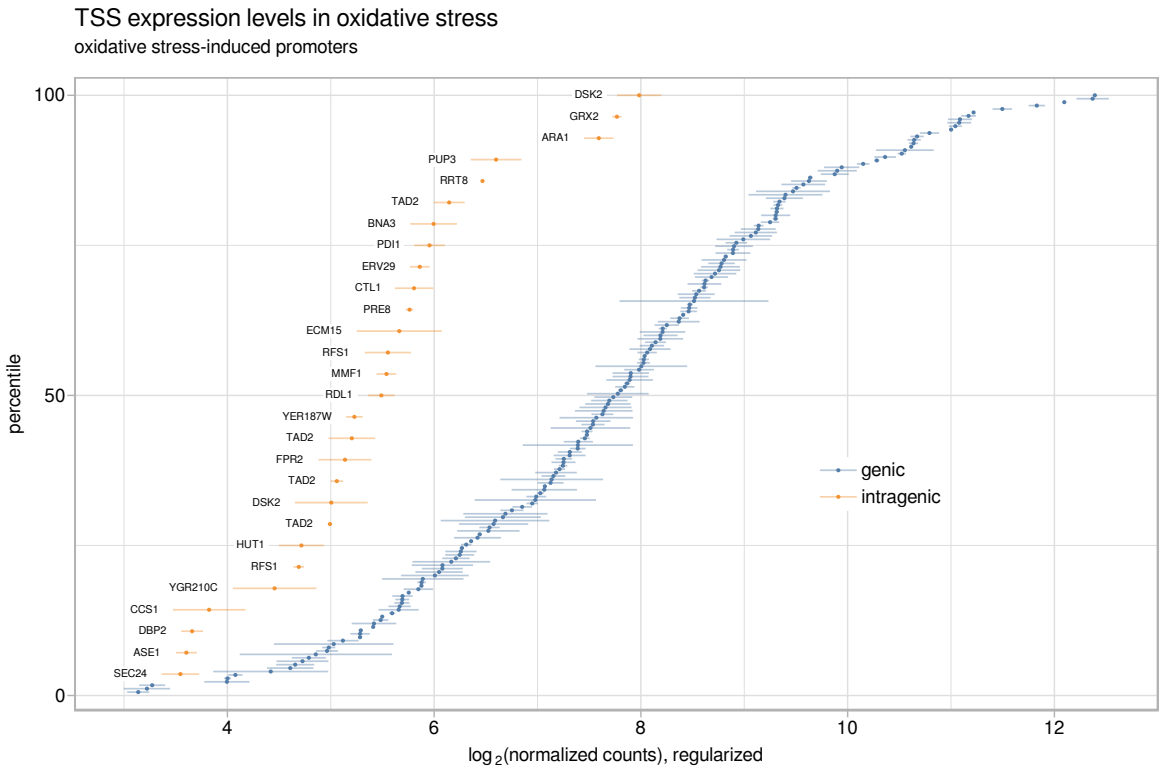


Figure 4.4: Caption dsafklj zzzz.

Figure 4.5: Caption dsafklj .

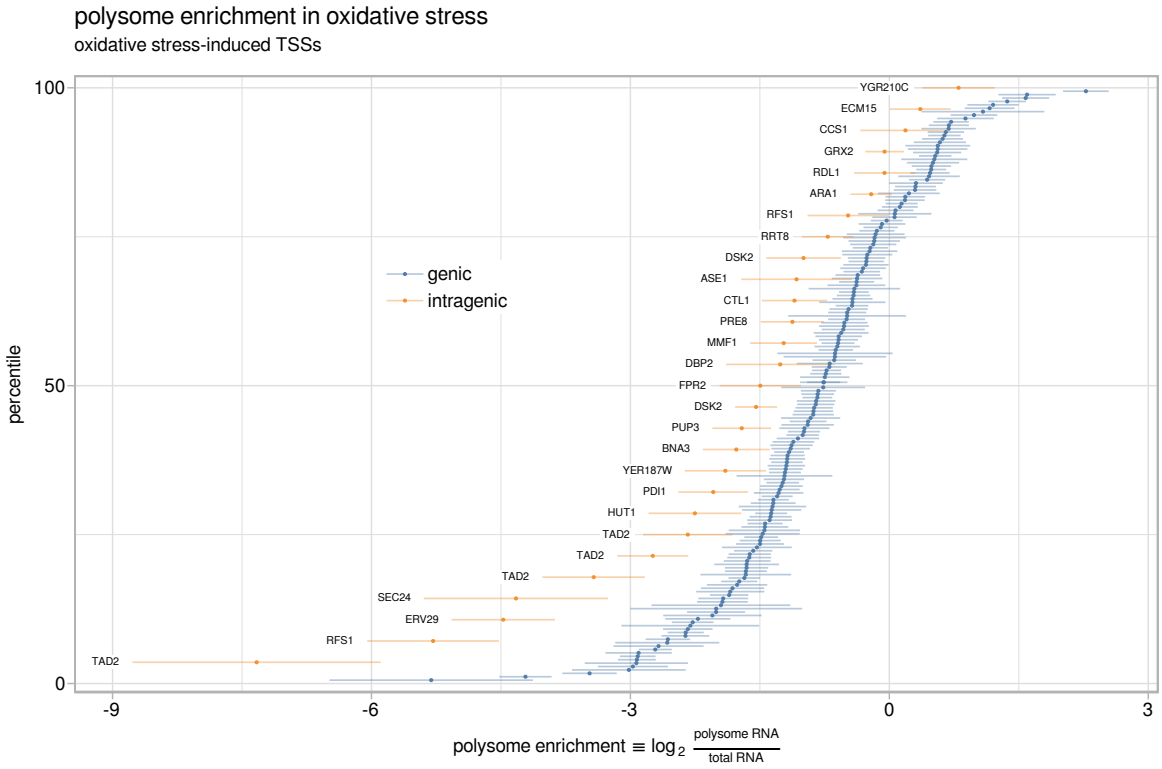


Figure 4.6: Caption wsdasdr zzzz.

Figure 4.7: Caption dsafklj .

Figure 4.8: Caption dsafklj .

Figure 4.9: Caption dsafklj .

Figure 4.10: Caption dsafklj .

Bibliography

Adkins, M. W. and Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of spt6-mediated chromatin reassembly of promoter regions. *Molecular Cell*, 21(3):405 – 416.

Andrulis, E. D., Guzmán, E., Döring, P., Werner, J., and Lis, J. T. (2000). High-resolution localization of drosophila spt5 and spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes & Development*, 14(20):2635–2649.

Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., and Lis, J. T. (2009). Spt6 enhances the elongation rate of rna polymerase ii in vivo. *The EMBO Journal*, 28(8):1067–1077.

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.

Begum, N. A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone spt6 is required for activation-induced cytidine deaminase target determination through h3k4me3 regulation. *Journal of Biological Chemistry*, 287(39):32415–32429.

Bortvin, A. and Winston, F. (1996). Evidence that spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267):1473–1476.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005). Histone h3 methylation by set2 directs deacetylation of coding regions by rpd3s to suppress spurious intragenic transcription. *Cell*, 123(4):581 – 592.

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.

Chen, S., Ma, J., Wu, F., Xiong, L.-j., Ma, H., Xu, W., Lv, R., Li, X., Villen, J., Gygi, S. P., Liu, X. S., and Shi, Y. (2012). The histone h3 lys 27 demethylase jmjd3 regulates gene expression by impacting transcriptional elongation. *Genes & Development*, 26(12):1364–1375.

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.

Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. (2006). The bur1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by set2. *Molecular and Cellular Biology*, 26(8):3029–3038.

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.

Close, D., Johnson, S. J., Sdano, M. A., McDonald, S. M., Robinson, H., Formosa, T., and Hill, C. P. (2011). Crystal structures of the s. cerevisiae spt6 core and c-terminal tandem sh2 domain. *Journal of Molecular Biology*, 408(4):697 – 713.

Consortium, T. E. P., Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kellis, M., Kheradpour, P., Lassmann, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C. J., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. A. L., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Good, P. J., Feingold, E. A., Crawford, G. E., Dekker, J., Elnitski, L., Farnham, P. J., Giddings, M. C., Gingeras, T. R., Guigó, R., Hubbard, T. J., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Stamatoyannopoulos, J. A., Tenenbaum, S. A., Weng, Z., White, K. P., Wold, B., Yu, Y., Wrobel, J., Risk, B. A., Gunawardena, H. P., Kuiper, H. C., Maier, C. W., Xie, L., Chen, X., Mikkelsen, T. S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M. J., Durham, T., Ku, M., Truong, T., Eaton, M. L., Dobin, A., Tanzer, A., Lagarde, J., Lin, W., Xue, C., Williams, B. A., Zaleski, C., Röder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira,

P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O. J., Park, E., Preall, J. B., Presaud, K., Ribeca, P., Robyr, D., Ruan, X., Sammeth, M., Sandhu, K. S., Schaeffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Hayashizaki, Y., Reymond, A., Antonarakis, S. E., Hannon, G. J., Ruan, Y., Carninci, P., Sloan, C. A., Learned, K., Malladi, V. S., Wong, M. C., Barber, G. P., Cline, M. S., Dreszer, T. R., Heitner, S. G., Karolchik, D., Kirkup, V. M., Meyer, L. R., Long, J. C., Maddren, M., Raney, B. J., Grasfeder, L. L., Giresi, P. G., Battenhouse, A., Sheffield, N. C., Showers, K. A., London, D., Bhinge, A. A., Shestak, C., Schaner, M. R., Ki Kim, S., Zhang, Z. Z., Mieczkowski, P. A., Mieczkowska, J. O., Liu, Z., McDaniell, R. M., Ni, Y., Rashid, N. U., Kim, M. J., Adar, S., Zhang, Z., Wang, T., Winter, D., Keefe, D., Iyer, V. R., Zheng, M., Wang, P., Gertz, J., Vielmetter, J., Partridge, E., Varley, K. E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K. M., Anaya, M., Cross, M. K., Muratet, M. A., Newberry, K. M., McCue, K., Nesmith, A. S., Fisher-Aylor, K. I., Pusey, B., DeSalvo, G., Parker, S. L., Balasubramanian, S., Davis, N. S., Meadows, S. K., Eggleston, T., Newberry, J. S., Levy, S. E., Absher, D. M., Wong, W. H., Blow, M. J., Visel, A., Pennachio, L. A., Petrykowska, H. M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Davidson, C., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J. M., Griffiths, E., Harte, R., Hendrix, D. A., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Leng, J., Lin, M. F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J. M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tapanari, E., Tress, M. L., van Baren, M. J., Washietl, S., Wilming, L., Zadissa, A., Zhang, Z., Brent, M., Haussler, D., Valencia, A., Addleman, N., Alexander, R. P., Auerbach, R. K., Balasubramanian, S., Bettinger, K., Bhardwaj, N., Boyle, A. P., Cao, A. R., Cayting, P., Charos, A., Cheng, Y., Eastman, C., Euskirchen, G., Fleming, J. D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyengar, S., Jin, V. X., Karczewski, K. J., Kasowski, M., Lacroute, P., Lam, H., Lamarre-Vincent, N., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X. J., O'Geen, H., Ouyang, Z., Patacsil, D., Raha, D., Ramirez, L., Reed, B., Shi, M., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K.-K., Yang, X., Struhl, K., Weissman, S. M., Penalva, L. O., Karmakar, S., Bhanvadia, R. R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Victorsen, A., Auer, T., Centanin, L., Eichenlaub, M., Gruhl, F., Heermann, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Partridge, E. C., Jain, G., Balasundaram, G., Bates, D. L., Byron, R., Canfield, T. K., Diegel, M. J., Dunn, D., Ebersol, A. K., Frum, T., Garg, K., Gist, E., Hansen, R. S., Boatman, L., Haugen, E., Humbert, R., Johnson, A. K., Johnson, E. M., Kutyavin, T. V., Lee, K., Lotakis, D., Maurano, M. T., Neph,

S. J., Neri, F. V., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Rynes, E., Sanchez, M. E., Sandstrom, R. S., Shafer, A. O., Stergachis, A. B., Thomas, S., Vernot, B., Vierstra, J., Vong, S., Wang, H., Weaver, M. A., Yan, Y., Zhang, M., Akey, J. M., Bender, M., Dorschner, M. O., Groudine, M., MacCoss, M. J., Navas, P., Stamatoyannopoulos, G., Beal, K., Brazma, A., Flicek, P., Johnson, N., Lukk, M., Luscombe, N. M., Sobral, D., Vaquerizas, J. M., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M. W., Schaub, M. A., Miller, W., Bickel, P. J., Banfai, B., Boley, N. P., Huang, H., Li, J. J., Noble, W. S., Bilmes, J. A., Buske, O. J., Sahu, A. D., Kharchenko, P. V., Park, P. J., Baker, D., Taylor, J., and Lochovsky, L. (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489:57 EP –. Article.

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.

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319.

Diebold, M.-L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an iws1/spt6 complex reveals an interaction domain conserved in tfiis, elongin a and med26. *The EMBO Journal*, 29(23):3979–3991.

Diebold, M.-L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem sh2 enables interaction of elongation factor spt6 with rna polymerase ii. *Journal of Biological Chemistry*, 285(49):38389–38398.

Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S., and Lima, C. D. (2014). How an mrna capping enzyme reads distinct rna polymerase ii and spt5 ctd phosphorylation codes. *Genes & Development*, 28(12):1323–1336.

Doamekpor, S. K., Schwer, B., Sanchez, A. M., Shuman, S., and Lima, C. D. (2015). Fission yeast rna triphosphatase reads an spt5 ctd code. *RNA*, 21(1):113–123.

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*.

Drouin, S., Laramée, L., Jacques, P.-�., Forest, A., Bergeron, M., and Robert, F. (2010). Dsif and rna polymerase ii ctd phosphorylation coordinate the recruitment of rpd3s to actively transcribed genes. *PLOS Genetics*, 6(10):1–12.

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].

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., Sato, H., Yamaguchi, Y., Mandal, S. S., Reinberg, D., Wada, T., and Handa, H. (2004). Human spt6 stimulates transcription elongation by rna polymerase ii in vitro. *Molecular and Cellular Biology*, 24(8):3324–3336.

Haberle, V. and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, 19(10):621–637.

Hartzog, G. A. and Fu, J. (2013). The spt4–spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(1):105 – 115. RNA polymerase II Transcript Elongation.

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 –.

Hennig, B. P. and Fischer, T. (2013). The great repression: chromatin and cryptic transcription. *Transcription*, 4(3):97—101.

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the rna polymerase clamp coiled-coil motif. *Nucleic Acids Research*, 38(12):4040–4051.

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.

Kaikkonen, M. U. and Adelman, K. (2018). Emerging roles of non-coding rna transcription. *Trends in Biochemical Sciences*, 43(9):654–667.

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

Kaplan, C. D., Morris, J. R., Wu, C.-t., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in d. melanogaster. *Genes & Development*, 14(20):2623–2634.

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.

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., and Murakami, K. S. (2011). Rna polymerase and transcription elongation factor spt4/5 complex structure. *Proceedings of the National Academy of Sciences*, 108(2):546–550.

Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9:357 EP –.

Li, Q., Brown, J. B., Huang, H., and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.*, 5(3):1752–1779.

Li, S., Almeida, A. R., Radebaugh, C. A., Zhang, L., Chen, X., Huang, L., Thurston, A. K., Kalashnikova, A. A., Hansen, J. C., Luger, K., and Stargell, L. A. (2018). The elongation factor spn1 is a multi-functional chromatin binding protein. *Nucleic Acids Research*, 46(5):2321–2334.

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.

Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011). Solution structure of tandem sh2 domains from spt6 protein and their binding to the phosphorylated rna polymerase ii c-terminal domain. *Journal of Biological Chemistry*, 286(33):29218–29226.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor spt5 by yeast bur1 kinase stimulates recruitment of the paf complex. *Molecular and Cellular Biology*, 29(17):4852–4863.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15(12):550.

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.

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., and Cramer, P. (2011). Architecture of the rna polymerase–spt4/5 complex and basis of universal transcription processivity. *The EMBO Journal*, 30(7):1302–1310.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general rna polymerase ii transcription complex. *Nature Structural & Molecular Biology*, 17:1272–1278.

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D. E., and Cramer, P. (2012). The spt5 c-terminal region recruits yeast 3' rna cleavage factor i. *Molecular and Cellular Biology*, 32(7):1321–1331.

Mbogning, J., Nagy, S., Pagé, V., Schwer, B., Shuman, S., Fisher, R. P., and Tanny, J. C. (2013). The paf complex and prf1/rtf1 delineate distinct cdk9-dependent pathways regulating transcription elongation in fission yeast. *PLOS Genetics*, 9(12):1–14.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones spt6 and fact collaborate to assemble, inspect, and maintain chromatin structure in saccharomyces cerevisiae. *Genetics*, 201(3):1031–1045.

McDonald, S. M., Close, D., Xin, H., Formosa, T., and Hill, C. P. (2010). Structure and biological importance of the spn1-spt6 interaction, and its regulatory role in nucleosome binding. *Molecular Cell*, 40(5):725 – 735.

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.

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., and Young, R. A. (2010). c-myc regulates transcriptional pause release. *Cell*, 141(3):432 – 445.

Schneider, S., Pei, Y., Shuman, S., and Schwer, B. (2010). Separable functions of the fission yeast spt5 carboxyl-terminal domain (ctd) in capping enzyme binding and transcription elongation overlap with those of the rna polymerase ii ctd. *Molecular and Cellular Biology*, 30(10):2353–2364.

Sdano, M. A., Fulcher, J. M., Palani, S., Chandrasekharan, M. B., Parnell, T. J., Whitby, F. G., Formosa, T., and Hill, C. P. (2017). A novel sh2 recognition mechanism recruits spt6 to the doubly phosphorylated rna polymerase ii linker at sites of transcription. *eLife*, 6:e28723.

Shandilya, J. and Roberts, S. G. (2012). The transcription cycle in eukaryotes: From productive initiation to rna polymerase ii recycling. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(5):391 – 400.

Shetty, A., Kallgren, S. P., Demel, C., Maier, K. C., Spatt, D., Alver, B. H., Cramer, P., Park, P. J., and Winston, F. (2017). Spt5 plays vital roles in the control of sense and antisense transcription elongation. *Molecular Cell*, 66(1):77 – 88.e5.

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.

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., Parrinello, H., Cuvier, O., and Benkirane, M. (2014). Integrator complex regulates nelf-mediated rna polymerase ii pause/release and processivity at coding genes. *Nature Communications*, 5:5531 EP –. Article.

Sun, M., Larivière, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem sh2 domain in transcription elongation factor spt6 binds the phosphorylated rna polymerase ii c-terminal repeat domain (ctd). *Journal of Biological Chemistry*, 285(53):41597–41603.

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.

Voss, K., Gentry, J., and Van der Auwera, G. (2017). Full-stack genomics pipelining with gatk4 + wdl + cromwell. In *18th Annual Bioinformatics Open Source Conference (BOSC 2017)*.

Wang, A. H., Juan, A. H., Ko, K. D., Tsai, P.-F., Zare, H., Dell'Orso, S., and Sartorelli, V. (2017). The elongation factor spt6 maintains esc pluripotency by controlling super-enhancers and counteracting polycomb proteins. *Molecular Cell*, 68(2):398 – 413.e6.

Wang, A. H., Zare, H., Mousavi, K., Wang, C., Moravec, C. E., Sirotkin, H. I., Ge, K., Gutierrez-Cruz, G., and Sartorelli, V. (2013). The histone chaperone spt6 coordinates histone h3k27 demethylation and myogenesis. *The EMBO Journal*, 32(8):1075–1086.

Weber, G., Springer, M., Jorgensen, P., Milo, R., and Moran, U. (2009). Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(suppl_1):D750–D753.

Wen, Y. and Shatkin, A. J. (1999). Transcription elongation factor hspt5 stimulates mrna capping. *Genes & Development*, 13(14):1774–1779.

Werner, F. (2012). A nexus for gene expression—molecular mechanisms of spt5 and nusg in the three domains of life. *Journal of Molecular Biology*, 417(1):13 – 27.

Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M., and VanDemark, A. P. (2013). Structural basis for spt5-mediated recruitment of the paf1 complex to chromatin.

Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., and Yamaguchi, Y. (2014). Dsif and nelf interact with integrator to specify the correct post-transcriptional fate of snrna genes. *Nature Communications*, 5:4263 EP –. Article.

Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., and Jones, K. A. (2007). The spt6 sh2 domain binds ser2-p rnapii to direct iws1-dependent mrna splicing and export. *Genes & Development*, 21(2):160–174.

Yoh, S. M., Lucas, J. S., and Jones, K. A. (2008). The iws1:spt6:ctd complex controls cotranscriptional mrna biosynthesis and hypb/setd2-mediated histone h3k36 methylation. *Genes & Development*, 22(24):3422–3434.

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., and Mellor, J. (2008). Roles for ctk1 and spt6 in regulating the different methylation states of histone h3 lysine 36. *Molecular and Cellular Biology*, 28(16):4915–4926.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of chip-seq (macs). *Genome Biology*, 9(9):R137.

Zhou, K., Kuo, W. H. W., Fillingham, J., and Greenblatt, J. F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast bur kinase substrate spt5. *Proceedings of the National Academy of Sciences*, 106(17):6956–6961.

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 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 gravida sollicitudin, 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 consectetuer.

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.