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

<div align="center">Approved by</div>

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
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

Introduction

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

Chapter 1

Genomics of transcription elongation factor Spt6

1.1 Collaborators

Steve Doris	optimized TSS-seq and ChIP-nexus protocols
	generated TSS-seq and ChIP-nexus libraries
Olga Viktorovskaya	generated MNase-seq libraries
Magdalena Murawska	generated NET-seq libraries
Dan Spatt	wetlab experiments for publication

1.2 Introduction to Spt6 and intragenic transcription

The aim of this work is related to understanding how a eukaryotic cell specifies which sites in its genome are permitted to become sites of transcription initiation. From genetic studies in yeast, it is known that some factors controlling the specificity of transcription initiation are actually transcription *elongation* factors, including histone chaperones and histone modification enzymes (Cheung et al., 2008, Hennig and Fischer, 2013, Kaplan et al., 2003). My collaborators on this project are interested in the role of the transcription elongation factor **Spt6** in this process. The following is a quick introduction to Spt6 (Doris et al., 2018):

- Spt6 interacts directly with:

- RNA polymerase II (RNAPII) (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)
  - the essential factor Spn1 (IWS1) (Diebold et al., 2010a, Li et al., 2018, McDonald et al., 2010)
- Spt6 is believed to function primarily as an elongation factor based on:
  - association with elongating RNAPII (Andrulis et al., 2000, Ivanovska et al., 2011, Kaplan et al., 2000, Mayer et al., 2010)
  - ability to enhance elongation in vitro (Endoh et al., 2004) and in vivo (Ardehali et al., 2009)
- Spt6 has been shown to regulate initiation in some cases (Adkins and Tyler, 2006, Ivanovska et al., 2011)
- Spt6 regulates:
  - 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)
  - histone modifications, including:
    * H3K36 methylation (Carrozza et al., 2005, Chu et al., 2006, Yoh et al., 2008, Youdell et al., 2008)
    * in some organisms, H3K4 and H3K27 methylation (Begum et al., 2012, Chen et al., 2012, DeGennaro et al., 2013, Wang et al., 2017, 2013)
- Spt6 is likely a histone chaperone required to reassemble nucleosomes in the wake of transcription (Duina, 2011).

Previous studies in the yeasts *S. cerevisiae* and *S. pombe* have examined the requirement for Spt6 in normal transcription (Cheung et al., 2008, DeGennaro et al.,

3

2013, Kaplan et al., 2003, Pathak et al., 2018, Uwimana et al., 2017, van Bakel et al., 2013). Many of these studies make use of the same temperature-sensitive *S. cerevisiae spt6* mutant used in this project, ***spt6-1004***, in which Spt6 protein is depleted at the non-permissive temperature of 37°C (Kaplan et al., 2003). The most notable phenotype of the *spt6-1004* mutant is the expression of **intragenic transcripts**, transcripts which appear to start within protein-coding genes, both in the same orientation and in the antisense orientation relative to the coding gene (Figure **??**) (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). These methods assay steady-state RNA levels, making them unable to determine 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 turned over in wild-type, or C) processing of full-length protein-coding RNAs. Additionally, these methods are suboptimal for identifying where intragenic transcription occurs, since the signal for an intragenic transcript in the same orientation as the gene it overlaps is convoluted with the signal from the full-length 'genic' transcript (Figure **??**) (Cheung et al., 2008, Lickwar et al., 2009).

To overcome these issues, one of my collaborators applied two assays to study transcription in *spt6-1004*: transcription start-site sequencing (**TSS-seq**), and **ChIP-nexus of TFIIB**, a component of the RNA polymerase II pre-initiation complex (PIC). The TSS-seq technique sequences the 5′ end of capped and polyadenylated RNAs (Arribere and Gilbert, 2013, Malabat et al., 2015), allowing separation of intragenic from genic RNA signals and identification of intragenic transcript starts with single-nucleotide resolution. The ChIP-nexus technique used is a high-resolution chromatin

immunoprecipitation technique in which the ChIPed DNA is exonuclease digested up to the bases crosslinked with the factor of interest before sequencing (He et al., 2015). When applied to the PIC component TFIIB, ChIP-nexus provides a way to determine whether intragenic transcripts result from new intragenic transcription initiation.

1.3   pipeline development for TSS-seq and ChIP-nexus

In order to use TSS-seq and ChIP-nexus to answer questions about Spt6 and intragenic transcription, I developed analysis pipelines for TSS-seq and ChIP-nexus data. The pipelines are written using the Python-based Snakemake workflow specification language (Köster and Rahmann, 2012), and perform steps including read cleaning (Martin, 2011), various quality controls (Andrews, 2012), read alignment (Kim et al., 2013, Langmead and Salzberg, 2012), data normalization, coverage track generation (Quinlan and Hall, 2010), peak calling (Zhang et al., 2008), differential expression/binding analyses (Love et al., 2014), data visualization with clustering, motif enrichment analyses (Bailey et al., 2015), and gene ontology analyses (Young et al., 2010). The Snakemake framework allows these data analyses to be reproducible and scalable from workstations up to computing clusters. Updated versions of these pipelines with more details on their capabilities are available at github.com/winston-lab. In the following subsections I will describe the thought behind only a few of the more novel pipeline steps before moving on to results relating to Spt6 and intragenic transcription.

1.3.1   TSS-seq peak calling

TSS-seq data from a single region of transcription initiation tends to occur as a cluster of signal distributed over a range of positions, rather than a single nucleotide (Figure

??) (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 signal by using the data to perform peak-calling.

In its current state, the TSS-seq pipeline calls peaks using 1-D watershed segmentation, followed by filtering for reproducibility 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 a discretized Gaussian kernel. Next, an initial set of peaks is generated by: 1) assigning all nonzero signal in the original, unsmoothed coverage to the nearest local maximum of the smoothed coverage in the direction of positive derivative, and 2) taking the minimum and maximum genomic coordinates of the original coverage assigned to each local maximum as the peak boundaries. The peaks are then trimmed to the smallest genomic window that includes 95% of the original coverage, and the probability of the peak being 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 a window upstream of the peak (as for the ChIP-seq peak caller MACS2 (Zhang et al., 2008)). The influence of local read density on λ is intended to reduce false positive peaks within gene bodies, especially for highly expressed genes: Since there are more fragments of RNA present for highly expressed genes, more fragments within the gene body will make it into the final library, even if they are not true 5' ends. To generate the final set of peaks, the peaks are ranked by significance under the Poisson model, and filtered by IDR. In brief, IDR attempts to separate true peaks from experimental noise based on

the intuition that, when peaks in each replicate are independently ranked by a metric such as significance, true peaks will have more similar ranks between replicates than peaks representing noise (Li et al., 2011).

The IDR algorithm currently only works for two replicates. Future improvements could include expanding the IDR implementation to handle more replicates and improve the accuracy of peak calling with more data.

1.3.2 ChIP-nexus peak calling

A number of tools have been created specifically for peak-calling using data from high-resolution ChIP techniques such as ChIP-nexus and ChIP-exo (Hansen et al., 2016, Wang et al., 2014). When applied to our TFIIB ChIP-nexus data, these tools tended to split what appeared to be a single TFIIB binding event into multiple peaks. This may be because TFIIB has been observed to crosslink to DNA at multiple points (Figure **??**) (Rhee and Pugh, 2012), which suggests that while these tools may work well for factors that bind symmetrically with a single crosslinking point on either side, there is still room for improvement when it comes to factors with more complex binding patterns. For the purposes of this project, the standard ChIP-seq peak caller MACS2 was used (Zhang et al., 2008).

ChIP-seq peaks lack strand information, as DNA binding factors usually do not bind DNA in a strand-specific manner. Because of this, we could not separate intragenic TFIIB peaks into peaks associated with sense or antisense transcription. The distinctive shape of the aggregate TFIIB ChIP-nexus signal (Figure **??**) suggests that information about the strand of transcription may be present in the ChIP-nexus binding profile. Future work could include learning the direction of transcription from labeled ChIP-nexus training data.

1.4 TSS-seq and TFIIB ChIP-nexus results from *spt6-1004*

To assay transcription start sites and transcription initiation in *spt6-1004*, one of my collaborators performed TSS-seq and ChIP-nexus of TFIIB. In wild-type cells, TSS-seq and TFIIB ChIP-nexus signal has the expected distribution over the genome, with most TSS-seq signal at annotated genic TSSs and most TFIIB signal just upstream (Figures **??**, **??**). In *spt6-1004*, the signal for both assays infiltrates gene bodies, consistent with a role for intragenic initiation in the intragenic transcription phenotype. Notably, sense strand TSS-seq signal in *spt6-1004* tends to occur towards the 3' end of genes, while antisense strand TSS-seq signal tends to occur towards the 5' ends of genes.

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 (Figure **??**). The results from this analysis support the pattern observed in the heatmap visualization (Figure **??**), with most genic TSSs downregulated and almost 8000 TSSs upregulated intragenic or antisense to genes. The overall effect of this on expression levels is to equalize expression levels between the different classes of transcripts (Figure **??**).

1.5 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.

Andrews, S. (2012). FastQC A Quality Control tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

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, 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.

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.

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.

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.

Doris, S. M., Chuang, J., Viktorovskaya, O., Murawska, M., Spatt, D., Churchman, L. S., and Winston, F. (2018). Spt6 is required for the fidelity of promoter selection. *bioRxiv*.

Duina, A. A. (2011). Histone chaperones spt6 and fact: Similarities and differences in modes of action at transcribed genes. *Genet Res Int*, 2011:625210. 22567361[pmid].

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.

Hansen, P., Hecht, J., Ibn-Salem, J., Menkuec, B. S., Roskosch, S., Truss, M., and Robinson, P. N. (2016). Q-nexus: a comprehensive and efficient analysis pipeline designed for chip-nexus. *BMC Genomics*, 17(1):873.

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.

Ivanovska, I., Jacques, P.-�., Rando, O. J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: Different consequences in coding and regulatory regions. *Molecular and Cellular Biology*, 31(3):531–541.

Jeronimo, C., Watanabe, S., Kaplan, C., Peterson, C., and Robert, F. (2015). The histone chaperones fact and spt6 restrict h2a.z from intragenic locations. *Molecular Cell*, 58(6):1113 – 1123.

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

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.

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.

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.

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.

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.

Rhee, H. S. and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483:295 EP –. Article.

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.

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.

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.

Wang, L., Chen, J., Wang, C., Uusküla-Reimand, L., Chen, K., Medina-Rivera, A., Young, E. J., Zimmermann, M. T., Yan, H., Sun, Z., Zhang, Y., Wu, S. T., Huang, H., Wilson, M. D., Kocher, J.-P. A., and Li, W. (2014). Mace: model based analysis of chip-exo. *Nucleic Acids Research*, 42(20):e156.

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.

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for rna-seq: accounting for selection bias. *Genome Biology*, 11(2):R14.

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.

Chapter 2

Genomics of transcription elongation factor Spt5

2.1   Collaborators

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

2.2 Bibliography