The Pennsylvania State University

The Graduate School

Eberly College of Science

CORE PROMOTER FUNCTION IN TRANSCRIPTIONAL REGULATION OF CONSTITUTIVELY EXPRESSED GENES IN *DROSOPHILA*

A Dissertation in

Biochemistry and Molecular Biology

by

Douglas G Baumann

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

May 2018

The dissertation of Douglas G Baumann was reviewed and approved* by the following:

David S. Gilmour
Professor of Biochemistry and Molecular Biology
Dissertation Adviser
Chair of Committee

Lu Bai
Assistant Professor of Biochemistry and Molecular Biology

Joseph C. Reese
Professor of Biochemistry and Molecular Biology

Shaun Mahony
Assistant Professor of Biochemistry and Molecular Biology

Tae-Hee Lee
Associate Professor of Chemistry

Wendy Hanna-Rose
Associate Professor of Biochemistry and Molecular Biology
Interim Department Head, Biochemistry and Molecular Biology

* Signatures are on file in the Graduate School.

Abstract

Regulation of gene expression is a key driver of organismal development and cellular differentiation. Misregulation of gene expression disrupts development and drives many disease states. While cells modulate gene expression through a variety of processes, transcription is thought to be the most critical and highly regulated process in gene regulation. Ultimately, all transcription regulatory signals converge at the core promoter. Here, I investigate the function of the conserved *Drosophila* core promoter element Motif 1. Most of our knowledge of transcription initiation comes from studies on promoters with a TATA box and Initiator. While these studies have yielded invaluable information, most promoters lack this combination of elements. Thus, our knowledge of the mechanisms that drive initiation from promoters with different core promoter elements is limited.

In order to better understand alternative mechanisms of initiation, I investigated ribosomal protein (RP) gene transcription in *Drosophila*. I focused on RP genes for several reasons. First, Motif 1 is enriched at RP gene promoters in *Drosophila.* Second, TBP is not required for transcription of RP genes. Third, in all higher eukaryotes examined, the TCT motif replaces the Initiator sequence at RP gene transcription start sites. Finally, RP genes represent a well-defined gene-regulatory network and are thought to be coordinately expressed. For these reasons, the RP genes offer a compelling model system to study alternative initiation mechanisms and the coordinate transcription of gene networks. Here, I demonstrate that Motif 1 and M1BP are required for RP gene transcription. Additionally, I found that M1BP works at RP genes by recruiting the TBP-related factor 2 (TRF2) to RP gene promoters. My data shows for the first time that TRF2 is recruited to promoters by sequence-specific binding factors. This

resolves a long-standing question in the field since TRF2 does not bind DNA in a sequence-specific manner. Furthermore, I show that the largest subunit of the general transcription factor TFIID associates with all RP gene promoters *in vivo*. Thus, while previous reports called into question TFIID's role in initiation at TCT motif-containing RP genes, this finding further attests to TFIID's key role in transcription.

I also identify an M1BP interacting protein called GFZF and investigated its function. I show that GFZF is a glutathione *S*-transferase (GST) and that it is essential for transcription activation. These results provide the first example of a transcription factor with GST activity. GFZF has appeared in a number of genetic screens that implicate it in processes as disparate as hybrid inviability and positive regulation of RAS/MAPK signaling. Because little was known about GFZF's molecular function, the authors of these reports explained and interpreted GFZF's appearance in these screens in complex and unclear terms. My findings suggest GFZF's involvement in these disparate cellular processes results from its association with over 1800 housekeeping gene promoters. Many of these housekeeping genes are directly involved in the processes that were being investigated. The knowledge of GFZF's molecular and cellular function should guide future interpretations when GFZF appears in large-scale screens. Finally, while it remains unclear why a transcription factor possesses a functional GST domain, I discuss ways that a transcription factor with GST activity might function in gene regulation. This might include roles in the transcriptional response to stress and the maintenance of genome integrity.

Table of Contents

List of Figures ix

List of Tables xii

Abbreviations xiii

Acknowledgments xiv

Chapter 1: Introduction 1

Genome Accessibility 1

Transcription Activators and Mediator 2

The General Transcription Machinery and Preinitiation Complex Formation 5

Core Promoter Elements 13

The TBP-related Factors 18

Additional Core Promoter Sequence Elements 19

Motif 1 and M1BP 20

GFZF and Glutathione S-transferases (GSTs) 21

Questions to Be Addressed 23

Chapter 2: M1BP recruits TRF2 to coordinately transcribe ribosomal protein genes 26

Introduction 27

Results 30

RP genes in *Drosophila* are coordinately transcribed at exceptionally high levels. 30

M1BP activates transcription of RP genes in cells 31

M1BP recruits TRF2 to the RP gene promoter 38

ChIP-exo provides evidence that M1BP recruits TRF2 to the majority of RP genes *in vivo* 42

ChIP-exo analysis detects TAF1 at RP gene promoters. 47

Colocalization of M1BP and TRF2 is largely restricted to the RP genes 51

RP gene promoters act as enhancers of other RP genes 51

Discussion 55

Materials and Methods 60

RNAi Knockdown in S2R+ Cells Followed by Chromatin Immunoprecipitation 60

Western Blots 61

Nuclear Extracts and Immunodepletion of M1BP 61

In vitro Transcription Reactions and Primer Extension Assay 62

Expression and Purification of M1BP 62

Synthesis of radiolabeled TRF2 63

Immobilized template pulldown experiments 64

Maltose-binding protein expression and pulldown experiments 64

ChIP-Exo 65

Bioinformatics 66

Peak Calling and ChIP-seq Analysis 66

Antibodies 67

Chapter 3: GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, is a transcriptional co-activator 68

Abstract 69

Introduction 70

Results and Discussion 72

M1BP associates with GFZF 72

GFZF associates with chromosomes. 75

GFZF colocalizes with M1BP at many promoters. 78

GFZF is a transcriptional co-activator 82

GFZF has glutathione S-transferase activity. 87

Identification of GFZF as a transcription factor provides insights into its roles in a broad spectrum of biological processes. 91

The GST activity of GFZF 96

Materials and Methods 101

Nuclear Extracts 101

Immobilized Template Pulldowns 101

Purification of Maltose-binding protein (Mal) fusions 102

Maltose-binding protein fusion pulldowns with His-GFZF 103

Polytene Chromosome Squashes and Immunofluorescence 104

Purification of full-length His-GFZF 104

ChIP-exo 105

Peak Calling 106

Luciferase Reporter Assays 106

RNAi-depletion of GFZF followed by chromatin Immunoprecipitation (ChIP)... 107

Expression and Purification of the GFZF GST domain..... 107

GST Activity Assay 108

Chapter 4: Additional Considerations and Future Directions..... 110

Nucleosome Clearance from Promoters by M1BP and NURF 110

M1BP and CG3995 as Chromatin Boundary Factors and Genome Organizers.... 112

M1BP and Spectrin 119

M1BP and dTOR 121

Speculation on Biomolecular Condensate Formation for Coordinate Transcription 122

Additional Considerations on M1BP Function..... 126

Materials and Methods..... 127

Peak Calling 127

NURF Analysis..... 127

Preparation of Protein A Magnetic Sepharose Beads 128

Co-Immunoprecipitation Experiments..... 128

References..... 130

List of Figures

Figure 1-1. Representation of the role pioneer factors play in DNA opening 2

Figure 1-2. Model for transcription activation by Mediator 4

Figure 1-3. Structural model of TFIID 9

Figure 1-4. Order-of-assembly model of preinitiation complex formation 12

Figure 1-5. Illustration of GFZF 23

Figure 2-1. Ribosomal Protein genes are highly transcribed at approximately equivalent levels 31

Figure 2-2. Schematic of the RP gene luciferase reporter assay 32

Figure 2-3. Motif 1 is required for RP gene transcription in cells 33

Figure 2-4. Motif 1 contributes to transcription of RpL30 and RpLP1 *in vitro* 34

Figure 2-5. Promoter strength correlates with higher signal-to-noise ratios 36

Figure 2-6. M1BP is required for RP gene transcription *in vitro* 37

Figure 2-7. M1BP recruits TRF2 to RP gene promoters 39

Figure 2-8. M1BP interacts with TRF2 in solution 40

Figure 2-9. TRF2 is dispensable for M1BP's association with RP gene promoters 41

Figure 2-10. TRF2 levels at RP gene promoters decreases with M1BP depletion irrespective of the presence of M1BP 42

Figure 2-11. M1BP and TRF2 co-occupy the majority of RP gene promoters 44

Figure 2-12. M1BP and TRF2 show overlapping ChIP-exo patterns on RP genes 45

Figure 2-13. TAF1 occupies RP gene promoters in regions that overlap with M1BP and TRF2 48

Figure 2-14. TAF1 is associated with actively transcribed promoters. 49
Figure 2-15. Many active TRF2-associated promoters are also bound by TAF1.....50
Figure 2-16. TRF2 and M1BP show greater overlap at RP gene promoters than other active promoters. 51
Figure 2-18. RP gene promoters act as enhancers of the RpS12 promoter 54
Figure 2-19. Model depicting M1BP's recruitment of TRF2 at RP gene promoters 59
Figure 3-1. Putzig, GFZF, CG3995, and M1BP associate with wt Motif 1 immobilized templates 73
Figure 3-2. GFZF is a nuclear protein recruited to promoter DNA by M1BP 74
Figure 3-3. M1BP and GFZF interact in solution.....75
Figure 3-4. GFZF is associated with polytene chromosomes.....76
Figure 3-5. GFZF and M1BP have similar staining patterns on polytene chromosomes.....77
Figure 3-6. Detection of ectopically expressed FLAG-tagged M1BP on polytene chromosomes 77
Figure 3-7. M1BP and GFZF show co-localized staining on polytene chromosome spreads.....78
Figure 3-8. Single- and triple-crosslinked GFZF ChIP-exo datasets correlate well.....79
Figure 3-9. M1BP and GFZF co-localize at promoters with Motif 1 80
Figure 3-10. GFZF and M1BP share similar exonuclease footprints 81
Figure 3-11. GFZF colocalizes with M1BP at promoter 81
Figure 3-12. A luciferase reporter assay reveals that GFZF is a transcriptional co-activator 84
Figure 3-13. GFZF knockdown results in Pol II and M1BP loss at GFZF-bound promoters 85
Figure 3-14. M1BP signals are stronger at GFZF-bound promoters.....86
Figure 3-15. Total cellular levels of M1BP are unaffected following GFZF depletion 87

Figure 3-16. GFZF has GST activity with high affinity for glutathione 88
Figure 3-17. GFZF's GST domain is not required for transcription activation of transfected DNA 90
Figure 3-18. GFZF binds the promoters of genes integral to genetic screens in which GFZF was a "hit" 92
Figure 3-19. GFZF binds the promoters of genes involved in the G2 to M DNA damage checkpoint 94
Figure 3-20. GFZF binds the promoter region of Male Specific Lethal (MSL) complex genes ...95
Figure 3-21. Orthologs of the entirety of GFZF are restricted to Schizophora 98
Figure 3-22. Orthologs of the entirety of GFZF are restricted to Schizophora 99
Figure 4-1. NURF is more highly enriched at M1BP-bound promoters than GAF-bound promoters 112
Figure 4-2. BED finger domain sequence alignment.....114
Figure 4-3. M1BP interacts with CP190.....117
Figure 4-4. M1BP IP identifies GFZF, Spectrin, and *Drosophila* target of rapamycin (dTOR).119
Figure 4-5. M1BP interacts with α-chain Spectrin 121
Figure 4-6. The TRF2 N-terminal domain is predicted to be highly disordered 124

List of Tables

Table 1-1. Conserved eukaryotic core promoter elements 20

Table 3-1. GFZF-associated genes tend to serve housekeeping functions 82

Abbreviations

Pol II	RNA Polymerase II
M1BP	Motif 1-binding protein
TBP	TATA-binding protein
TRF	TBP-related factor
TRF2	TBP-related factor 2
GFZF	glutathione S-transferase FLYWCH zinc finger protein
TAF	TBP-associated factor
DREF	DNA replication-related element factor
TFIIA	Transcription factor IIA
TFIIB	Transcription factor IIB
TFIID	Transcription factor IID
TFIIE	Transcription factor IIE
TFIIF	Transcription factor IIF
TFIIH	Transcription factor IIH
PIC	Pre-initiation complex
GST	Glutathione S-Transferase
GSH	Glutathione
RP	Ribosomal Protein
Inr	Initiator element
DPE	Downstream promoter element
DCE	Downstream core element

Acknowledgments

First I would like to thank Dave. He has been a tremendous positive influence throughout my time in his lab. He has joined in my successes and offered great encouragement through failures and setbacks. He has guided my research, challenged me to think in new ways, and has made me a better scientist. He has pushed me to achieve success in my professional endeavors while also enabling me to maintain health and balance in my personal relationships and pursuits. Like the many others he has mentored over the years, Dave has provided the support and stability to enable me to test my own limits without fear of retribution or criticism.

I also thank the members of my committee. This includes Lu Bai, Tae-Hee Lee, Shaun Mahony, Joseph Reese, and Yanming Wang. Their constructive feedback has catalyzed much of this work. Their insights have improved my science. Through their suggestions and encouragement, they have guided my professional development. The generosity they've displayed in sharing their time and ideas have made my exams and committee meetings highly positive experiences.

I appreciate the feedback and contributions from my contemporaries in the Gilmour lab. This includes Dr. Bede Portz, Dr. Jian Li, Dr. Yijun Qiu, Dr. Michael Fisher, Dr. Bhavana Achary, Laura Stefanik, Feiyou Liu, and Roberta Dollinger. Jian's M1BP work provided a wonderful foundation upon which to build my own body of research.

Most of all, I thank my wife and son, Rheanna and McKay, for their unwavering support. Through their unconditional love, sacrifice, and patience they have motivated me to be my best self.

This work was supported by the National Institutes of Health (NIH) (grant R01GM047477 to D.S.G.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Chapter 1: Introduction

The correct spatiotemporal regulation of gene expression is a nuanced process that underpins vast swaths of biology and its misregulation can contribute to disease states and problems in development. Regulation of Pol II-mediated transcription is one of the primary ways in which gene expression is modulated. Despite there being a number of steps at which transcription is known or thought to be regulated, initiation is perhaps the most rate-limiting step (Cisse *et al*, 2013).

Genome Accessibility

A number of steps must take place prior to transcription initiation. To start, since the eukaryotic genome is packaged in nucleosomes, distal regulatory (enhancers) and proximal regulatory (promoters) regions must be made accessible to transcription factors and to the general transcription machinery. While it has been postulated, based on studies in yeast, that the underlying AT-rich sequence of promoter regions dictates nucleosome positioning and occlusion by disfavoring nucleosome association (Segal *et al*, 2006; Ioshikhes *et al*, 2006; Yuan & Liu, 2008; Peckham *et al*, 2007), it has been demonstrated that such sequence preference can only partially explain nucleosome depletion in these regions (Zhang *et al*, 2011; Krietenstein *et al*, 2016; Valouev, 2011). Specialized transcription factors, termed pioneer factors, facilitate the removal of nucleosomes from enhancers and promoters. Pioneer factors bind their recognition sequence while the DNA is still wrapped around the nucleosome and ultimately destabilize or alter histone-DNA contacts (Cirillo *et al*, 2002; Cuesta *et al*, 2007; Hatta & Cirillo, 2007). This alteration enables other transcription factors or histone remodeling complexes to bind chromatin

and causes a cascade of events including further nucleosome destabilization, promoter opening, and transcription (Zaret & Carroll, 2011) (Figure 1-1).

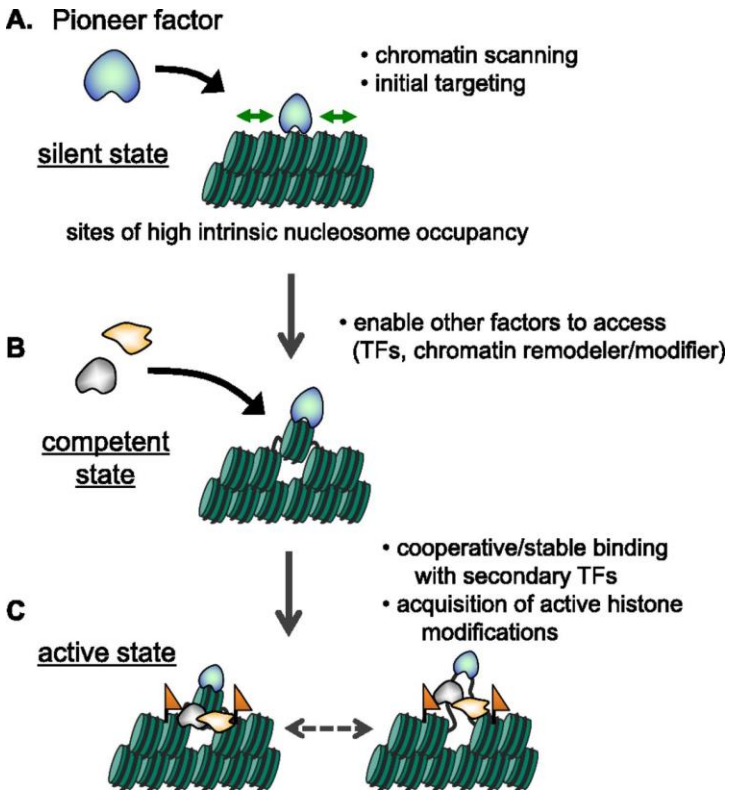

Figure 1-1. Representation of the role pioneer factors play in DNA opening.
(A) The pioneer factor (blue) binds DNA in a condensed chromatin environment and (B) enables the binding of other transcription factors and/or histone remodeling and modifying complexes (shown in orange and gray). (C) Regions where pioneer factors bind are "primed" for activation by the recruitment of additional cofactors and the acquisition of histone modifications associated with active transcription. Reproduced from (Iwafuchi-Doi & Zaret, 2014).

Transcription Activators and Mediator

During or following the process of nucleosome depletion from promoters and enhancers, additional sequence-specific transcription factors (activators) bind both distal and proximal cis-regulatory elements. Co-activators associate indirectly with chromatin through interactions with sequence-specific transcription factors. Activators and coactivators are thought to drive

expression as they are brought into close spatial proximity with the promoter region through a process termed looping (Szutorisz *et al*, 2005). The Mediator complex is perhaps the most important or well-studied facilitator of the process. Mediator was first identified as a complex required to overcome squelching effects when transcription activators were over-expressed in yeast (Kelleher *et al*, 1990). It was demonstrated that this overexpression caused these activators to compete with each other for binding sites within the Mediator complex that are necessary for activated transcription. This competition reduces Pol II-mediated transcription (Flanagan *et al*, 1991). Mediator makes contacts with components of the general transcription machinery (discussed below) as well as transcription factors bound to distal regulatory elements thereby facilitating the process of activated transcription (Petrenko *et al*, 2016). One recent review concludes that Mediator plays a role in most rate-limiting steps in transcription (Jeronimo & Robert, 2017). Figure 1-2 summarizes the current understanding of some of Mediator's roles in transcription initiation and looping.

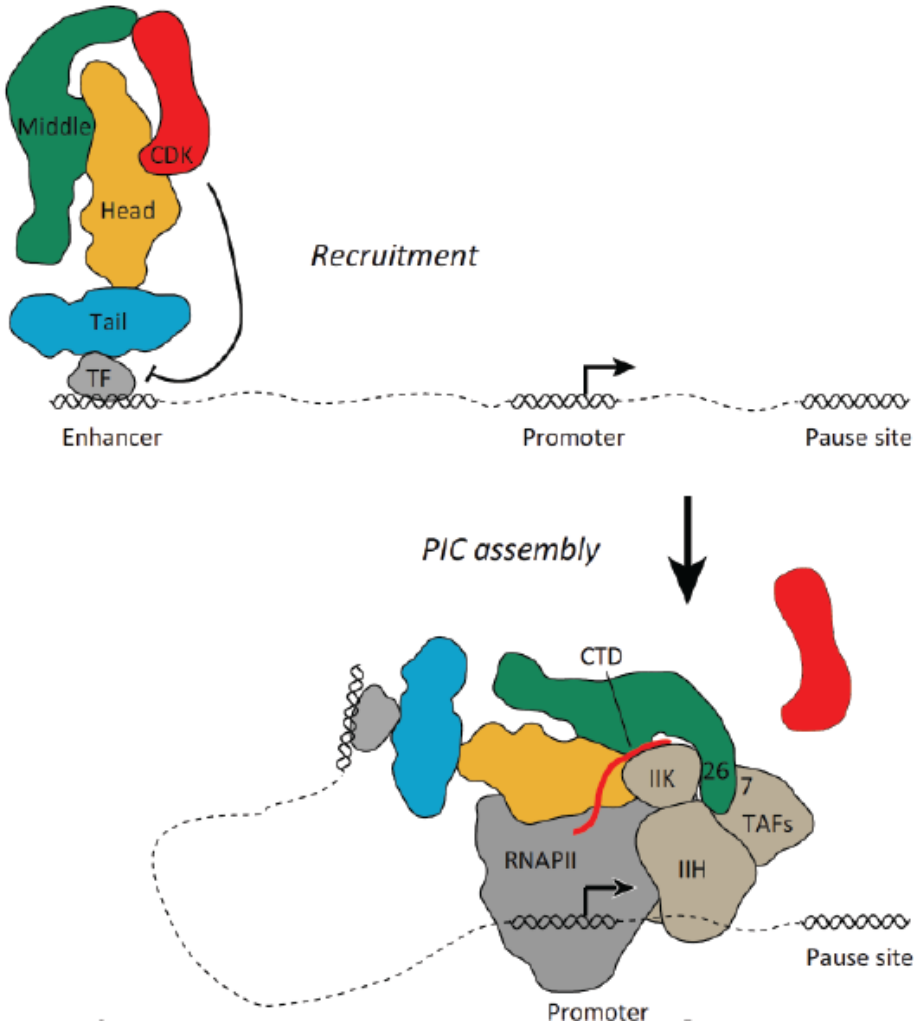

Figure 1-2. Model for transcription activation by Mediator.

Enhancers and promoters are bound by transcription activators and the general transcription factors (GTFs), respectively (shown in gray). Mediator (four modules represented in color) is shown bridging the interactions between the distal- and proximal-bound factors, thereby facilitating activated transcription by Pol II. In this model, the GTF's association with the core promoter is facilitated or enhanced by Mediator. Notably, the kinase module (shown in red) and Pol II interact in a mutually exclusive manner. Additionally, the kinase module can interfere with the tail module's recruitment by transcription factors. Thus, the kinase module has an inhibitory effect on transcription. Reproduced from (Jeronimo & Robert, 2017).

The General Transcription Machinery and Preinitiation Complex Formation

The core promoter ultimately integrates all upstream regulatory signals, thus making this region a transcriptional gateway (Heintzman & Ren, 2007). The core promoter is defined as the 100 bp region centered on the transcription start site (TSS). It consists of conserved DNA elements that direct the association of proteins and complexes necessary for transcription. Early studies on the highly active adenovirus major late (AdML) promoter revealed that purified Pol II could not accurately initiate transcription *in vitro* unless it was supplemented with subcellular fractions (Weil *et al*, 1979). Further fractionation of these subcellular fractions revealed a total of five biochemically distinct, evolutionarily conserved proteins. These proteins are required for the accurate initiation of transcription from the AdML promoter (Matsui *et al*, 1980; Sawadogo & Roeder, 1985; Flores *et al*, 1989, 1992). These proteins are now commonly referred to as general transcription factors (GTFs). They include TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. Altogether, they form and constitute the pre-initiation complex (PIC).

By interrogating the function of the core promoter and GTFs, subsequent studies provided a greater mechanistic understanding of transcription initiation. Using electrophoretic mobility shift assays (EMSAs) and DNase I footprinting, one particularly notable study determined both the binding order and nuclease protected regions for each GTF (Buratowski *et al*, 1989). The events identified in this study, in order of occurrence, include: binding of TFIID to the promoter followed by the recruitment of TFIIA and TFIIB, then TFIIF and Pol II, and finally TFIIE and TFIIH. This order-of-assembly model of PIC formation has held up well to additional scrutiny from studies using biochemistry, genomics, and structural analysis (Sainsbury *et al*, 2015). It should be noted that there are reports of a Pol II holoenzyme consisting of Pol II, TFIIB, TFIIE, TFIIF, TFIIH, a histone acetyltransferase, and a histone remodeler (Ossipow *et al*,

1995; Maldonado *et al*, 1996; Wu & Chiang, 1998; Wu *et al*, 1999). The composition of this complex varies depending on the purification scheme and source material used. The Pol II holoenzyme complex has been proposed to function in a manner that is analogous to prokaryotic RNA polymerase wherein TFIID (σ factors would be the functional counterpart in bacteria) recruits the nearly fully assembled Pol II complex. Once recruited, the Pol II holoenzyme could then initiate transcription without the sequential addition of other factors (Koleske & Young, 1994, 1995; Kim *et al*, 1994). Despite these early reports, far less is known about this complex and any putative functions in cells. Evidence from a recent single-molecule study supports both the order-of-assembly and holoenzyme models of initiation (Horn *et al*, 2016). This report makes the future study of holoenzyme complexes a worthwhile endeavor as such mechanisms may confer additional layers of regulatory complexity. Nevertheless, the evidence from most biochemical and structural studies (detailed below) is consistent with the order-of-assembly model. The data from such studies reveals the molecular functions carried out by each factor that ultimately enable Pol II to catalyze the formation of the first phosphodiester bond. Thus, in this dissertation, I will focus primarily on the order-of-assembly model, particularly emphasizing some early steps in initiation. Since the entire processes of PIC formation is critical to our understanding of the regulation of transcription initiation, a more detailed review of the relevant literature is fitting.

TFIID is the first complex to bind the promoter during transcription initiation (Nakajima *et al*, 1988; Buratowski *et al*, 1989). The TATA box was one of the first core promoter elements discovered and both the TATA element and TBP are conserved from archaea to humans (Patikoglou *et al*, 1999). This conservation hints at the importance of the interaction between the factor and element. Structural studies of TBP binding during initiation revealed that TBP induces

a nearly 90 degree bend in the DNA at the TATA sequence (Kim & Burley, 1994; Kim *et al*, 1993). Additional studies suggested this bending is critical for subsequent steps in PIC formation and initiation (Juo *et al*, 1996; Nikolov *et al*, 1995; Tan *et al*, 1996). TBP was once believed to be the sole factor responsible for conferring the biochemical function seen with early TFIID fractions (Cavallini *et al*, 1989). However, further studies demonstrated that TBP exists in a biochemically defined complex with 13 TBP-associated factors (TAFs) (Chiang *et al*, 1993; Poon & Weil, 1993; Reese *et al*, 1994). Early *in vitro* transcription studies demonstrated that signals from transcription activators were only transduced when the entire TFIID complex was present in the reactions, but not when TBP was purified away from the rest of the complex (Pugh & Tjian, 1990; Smale *et al*, 1990; Dynlacht *et al*, 1991; Tanese *et al*, 1991). Thus, these studies demonstrated that the entire TFIID complex was required for activated, but not basal transcription. Additional reports revealed that this effect is a result of interactions between transcription activators or co-activators and some TAFs (Hoey *et al*, 1993; Goodrich *et al*, 1993). Thus, it was proposed that TFIID acts as a coactivator either by being recruited to promoters through its interaction with transcription activators or through conformational changes that occur in the TFIID complex after contact with the activator (Hoey *et al*, 1993; Goodrich *et al*, 1993). Interestingly, later studies demonstrated that activated transcription can occur in a TAF-independent manner at some gene promoters (Walker *et al*, 1996; Oelgeschläger *et al*, 1998). Thus, these findings suggest that TAF and TFIID function may be context dependent.

In addition to the contact TBP makes with the TATA box, a series of elegant DNase I footprinting and crosslinking studies revealed that TFIID, particularly the TAF1 and TAF2 subunits, makes extensive contacts with the TSS and downstream promoter regions (Nakatani *et al*, 1990; Purnell & Gilmour, 1993; Sypes & Gilmour, 1994; Purnell *et al*, 1994; Knutson *et al*,

2000). Subsequent studies showed that TAF6 and TAF9 also make contact with downstream elements (Burke & Kadonaga, 1996; Theisen *et al*, 2010). Evidence from a recent high-resolution cryogenic electron microscopy (Cryo-EM) structure of TFIID bound to a synthetic core promoter sequence (termed SCP1) supports the results from the early crosslinking studies. These Cryo-EM structures revealed that TAF1, TAF7, and TAF2 are able to make extensive contacts with downstream DNA elements (Louder *et al*, 2016). TAF6 does not contact the promoter directly, but rather bridges the interactions between the B and C lobes of TFIID that contact the upstream and downstream sequences. Also, TAF9 was unable to be assigned to any density in this structure. These findings once again raise the possibility that TAF functions are promoter-context dependent. The overall structure of TFIID bound to the SCP1 core promoter highlights the extent of the contacts TFIID makes throughout the core promoter region (Figure 1-3). Notably, all of these studies were performed on promoters with the TATA box and Initiator (Inr) elements (discussed later in greater detail). This is an important point since the composition and combination of core promoter elements and the factors that bind to them will be a topic of major consideration throughout this dissertation.

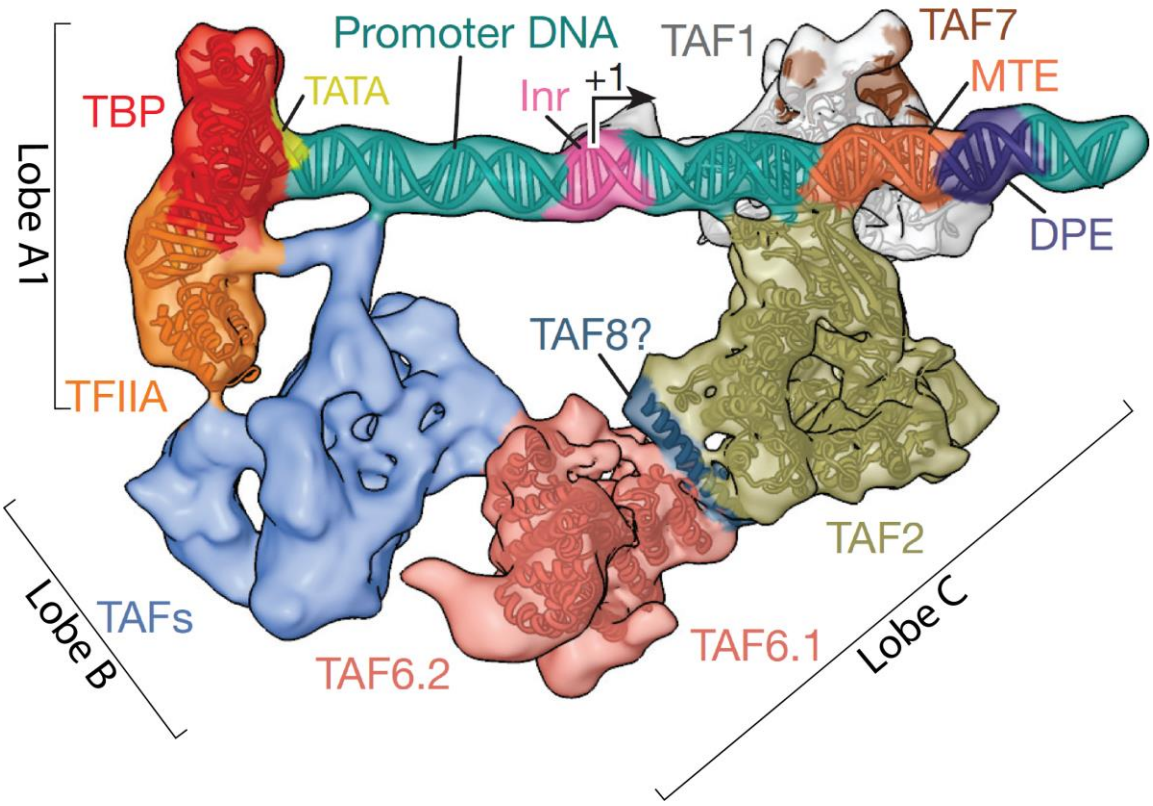

Figure 1-3. Structural model of TFIID.

Displayed are Cryo-EM densities of TFIID bound to SCP1 DNA. TFIID consists of three lobes: A, B, and C. Lobe A is further divided into two lobes, with the A2 lobe being a dynamic section of the A lobe. The lobes of TFIID that were visible in the structure are labeled and delimited accordingly. The densities and docked crystal structures of Lobe A1 consisting of TBP and TFIIA are colored red and orange, respectively. Lobe B TAF densities are represented in blue. Lobe C consists of TAFs 1 (white), 2 (green), 7 (brown), 6 (maroon), and 8 (dark blue) and is shown with crystal structures docked into the Cryo-EM densities. The various conserved promoter elements of the synthetic super core promoter are labeled along the DNA double helix. Lobe A2 was not seen in the Cryo-EM images likely as a result of its highly dynamic nature. Adapted from (Louder et al, 2016a).

After TFIID associates with the promoter, TFIIA and TFIIB associate with TBP and subsequently bind promoter DNA or the early initiation complex. Although early experiments with TFIIA indicated that it played an essential role in initiation (Reinberg *et al*, 1987), it was later shown, in reactions of higher purity, to be dispensable for initiation *in vitro* (Wu *et al*, 1998;

Van Dyke *et al*, 1988). Notably though, other reports argue that TFIIA stabilizes the TBP-DNA complex, particularly in conditions that are suboptimal for TBP-TATA binding (Imbalzano *et al*, 1994). Despite not being required for *in vitro* initiation, TFIIA is essential in yeast (Ranish *et al*, 1992). This finding suggests TFIIA might play a critical role in PIC formation in a cellular context where additional activation barriers exist. In agreement with this possibility, TFIIA derepresses negative regulation of TBP. For example, TFIIA impedes the action of factors that inhibit TBP function including NC2 (Xie *et al*, 2000), the N-terminal region of TAF1 (Kokubo *et al*, 1998), and HMGB1 (Ge & Roeder, 1994a). TFIIA also promotes the dissociation of TBP dimers to their active, monomeric form (Coleman *et al*, 1999). Finally, TFIIA bridges the interaction between the general transcription machinery and transcription factors including NTF-11, VP16, SP1, AP-11, PC4, and CTF (Ge & Roeder, 1994b; Yokomori *et al*, 1994; Ozer *et al*, 1996). Thus, TFIIA is considered a coactivator (Thomas & Chiang, 2008).

Unlike TFIIA, TFIIB is required for initiation *in vitro* and is conserved in all eukaryotes and in archaea (Ouzounis & Sander, 1992). TFIIB interacts with TBP and the DNA sequences flanking the TATA box. TFIIB stabilizes the ternary complex consisting of TBP, DNA, and TFIIA (Bagby *et al*, 1995; Nikolov *et al*, 1995). Through its N-terminal B-ribbon and a core cyclin domain, TFIIB interacts with and recruits Pol II (Bushnell *et al*, 2004). TFIIB stabilizes Pol II's association with promoter DNA through contacts with both the template strand and Pol II. It also contributes to transcription by facilitating Pol II clamp closure and DNA-RNA hybrid strand separation as it guides RNA to the exit channel of Pol II (Sainsbury *et al*, 2013). Finally, TFIIB also facilitates DNA bending by TBP (Zhao & Herr, 2002). These studies highlight TFIIB's fundamental role in early initiation.

In the next step of PIC formation, TFIIB recruits Pol II and TFIIF to promoters (Ha *et al*, 1993). TFIIF prevents Pol II from binding DNA non-specifically, likely through its own interaction with DNA (Sopta *et al*, 1989; McCracken & Greenblatt, 1991; Conaway *et al*, 1991). This property of TFIIF is consistent with its distant evolutionary conservation with bacterial σ factors (Sopta *et al*, 1989). Additionally, TFIIF stabilizes the early PIC through contacts with DNA, both upstream and downstream of the TATA box, and with Pol II (He *et al*, 2013). TFIIF induces changes in DNA topology, which contributes to its function in initiation (Robert *et al*, 1998). Furthermore, it enhances phosphodiester bond formation, participates in TSS selection, and enhances early elongation (Ren *et al*, 1999; Ghazy *et al*, 2004; Price *et al*, 1989; Funk *et al*, 2002).

In the last step of PIC formation, TFIIF recruits TFIIE and TFIIH. Both TFIIE and TFIIH are required for DNA opening. TFIIE bridges the interaction between TFIIH and Pol II (Flores *et al*, 1989; Maxon *et al*, 1994). TFIIE also has ssDNA binding activity (Yokomori *et al*, 1998; Okamoto *et al*, 1998). It has been proposed that this ssDNA binding activity accounts for TFIIE's ability to open DNA and facilitate the transition from initiation to elongation (Holstege *et al*, 1995, 1996; Watanabe *et al*, 2003; Forget *et al*, 2004). TFIIH contains three modules, all of which possess enzymatic activity. The XPB module is a 3'-5' helicase that functions in promoter opening (Holstege *et al*, 1996; Tirode *et al*, 1999; Guzmán & Lis, 1999). Mutations in the XPB module revealed that the ATPase activity is required for promoter opening while the helicase activity is required for promoter escape (Lin *et al*, 2005). The TFIIH XPD module contains 5'-3' helicase activity and is required for TFIIH function in the nucleotide excision DNA repair pathway (Coin *et al*, 2007). CDK7, the catalytic subunit of the kinase module, phosphorylates the Carboxy-terminal domain of Pol II (Feaver *et al*, 1991; Serizawa *et al*, 1995; Adamczewski

et al, 1996), a domain which is itself an important regulatory module of Pol II (Harlen & Churchman, 2017). Figure 1-4 represents the current order-of-assembly model of initiation based on decades of biochemical and structural data.

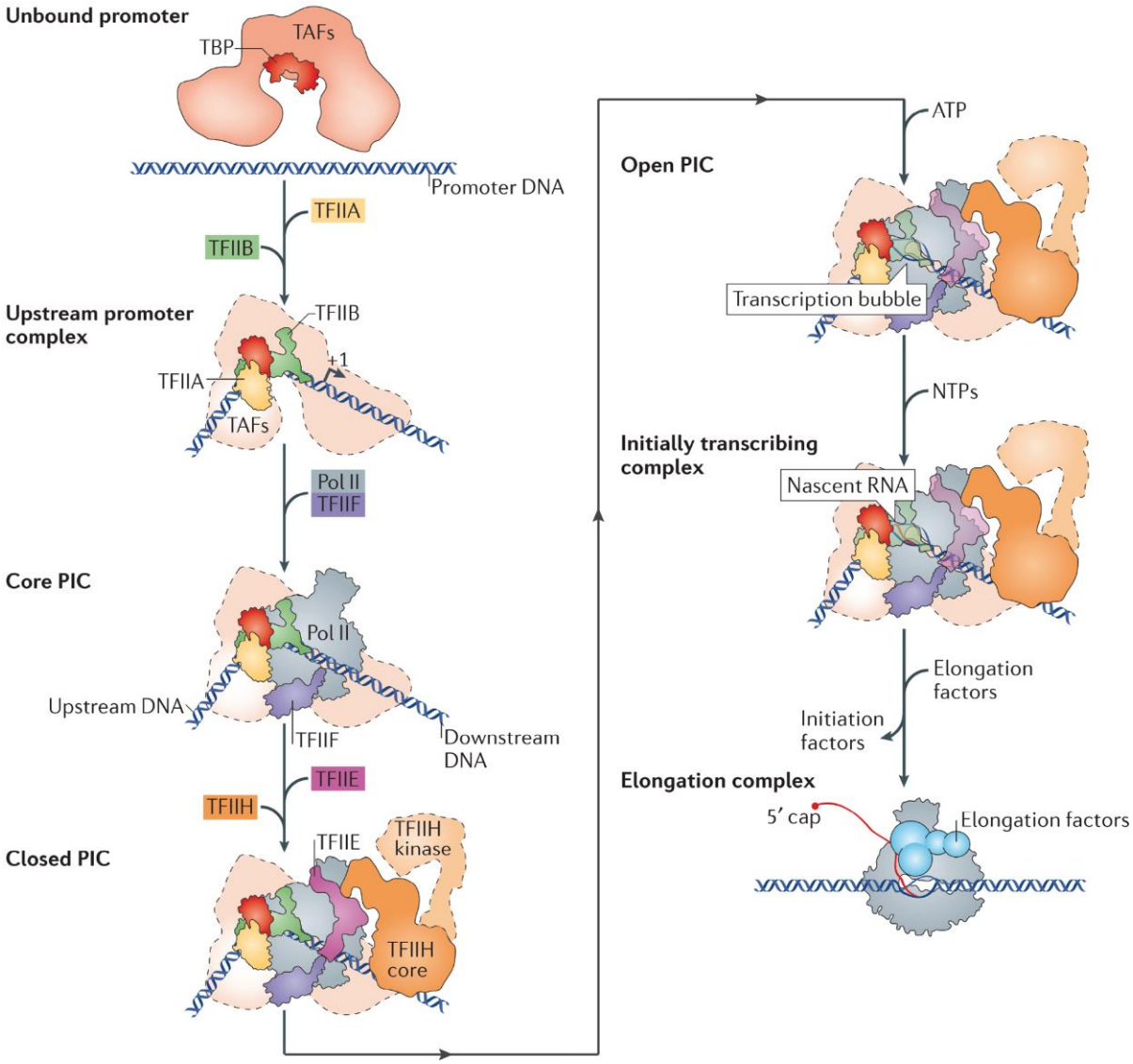

Figure 1-4. Order-of-assembly model of preinitiation complex formation.
TFIID binds promoter DNA and facilitates recruitment of TFIIA and TFIIB. TFIIF and Pol II follow. Promoter DNA is melted and initial transcription occurs through the combined effects of TFIIE and TFIIH. Elongation factors such as Spt4/5 associate following promoter clearance. Adapted from (Sainsbury *et al*, 2015).

Core Promoter Elements

The TATA box was the first core promoter element described in the literature when the flanking sequences upstream of *Drosophila*, mammalian, and viral genes were compared and the TATA box was observed in all the genes examined (Goldberg, 1979; Breathnach & Chambon, 1981). The TATA box has been a major focal point in the study of core promoters and transcription initiation ever since. The pyrimidine-rich TATA sequence confers biochemical properties that enable TBP to associate with the minor groove and induce a kink (Kim & Burley, 1994; Kim *et al*, 1993)). In yeast, the distance between the TATA box and the TSS is determined by TFIIB and Pol II (Li *et al*, 1994). However, additional reports contradict this finding and rather suggest the position of the TATA box is not a determinant of TSS selection (Chen & Struhl, 1985; Hahn *et al*, 1985). Some studies have called into question whether the stably-induced kink of TATA DNA by TBP is necessary for TBP function in initiation (Hoopes *et al*, 1992; Zhao & Herr, 2002). A more recent study concludes that sequence-specific TBP-DNA interactions are not required for transcription of TATA-less genes in yeast (Kamenova *et al*, 2014). Thus, it seems our understanding TATA and TBP function in transcription may still be incomplete.

The Inr was one of the first core promoter elements described and is important for *in vitro* transcription (Corden *et al*, 1980; Talkington & Leder, 1982; Dierks *et al*, 1983; Concino *et al*, 1984; Smale & Baltimore, 1989). Mutating the Inr sequence causes a change in transcription initiation sites (Chen & Struhl, 1985; Hahn *et al*, 1985; McNeil & Smith, 1985). Therefore, the Inr is thought to be important for TSS selection (Grosschedl & Birnstiel, 1980; Kuehner & Brow, 2006). Additional studies identified sequences downstream of the TSS that are important for *in vitro* transcription (Lewis & Manley, 1985; Nakatani *et al*, 1990). Conserved elements in this

region include the downstream promoter element (DPE) and motif ten element (MTE). The spacing between the +1 position in the Inr and the DPE and MTE is exquisitely important as a shift of one or two base pairs alters the transcriptional output of promoters possessing these elements (Burke & Kadonaga, 1996; Lim *et al*, 2004). When the DPE and MTE exist together at the same promoter, which they most often do, transcriptional output is further enhanced (Theisen *et al*, 2010) . Another downstream element, named the downstream core element (DCE) was first discovered in the promoter region of the β-globin gene when mutations in the element reduced expression of β-globin leading to β-thalassemia. β-thalassemia is a disease caused by mismatched expression of α- and β-globin (Öner *et al*, 1991; Cai *et al*, 1992; Ho *et al*, 1996). It was later shown that mutations in the DCE reduced TFIID's affinity for the promoter and decreased *in vitro* transcription activity (Lewis *et al*, 2000). The presence of these conserved and functional motifs suggests that TFIID's extensive interactions with the core promoter are important. The results also suggest that the function of any given promoter element is context dependent.

Interestingly, despite the fact that the general transcription machinery is highly conserved, there are no global promoter elements (Vo Ngoc *et al*, 2017b). For example, the Inr, the most abundant core promoter element, is present on roughly 40% of human focused promoters (Vo Ngoc *et al*, 2017a) and 26% of *Drosophila* promoters (Ohler *et al*, 2002b). The TATA box is present even less frequently (~3%) at such promoters, while TATA-like sequences only appear about 28% of the time. TATA-like sequences exist at roughly 20% of focused promoters with an Inr sequence, thus, Inr and TATA-like sequences show an inverse correlation (Vo Ngoc *et al*, 2017a). This study confirmed the findings of previous analyses of metazoan core promoters which demonstrated that promoters with a well-defined TATA box and Inr sequence

only accounted for a fraction of promoters throughout the genome (Ohler *et al*, 2002b; FitzGerald *et al*, 2006). The downstream MTE and DPE are found less frequently than the TATA box and Initiator (Ohler *et al*, 2002b). The TFIIB-recognition elements exist at only a fraction of promoters and are poorly conserved sequences that require TFIIB binding in order to be considered functional elements (Vo Ngoc *et al*, 2017b).

Despite the diversity in core promoter composition it is clear that TFIID makes extensive contacts with the core promoter and these contacts are important in transcription initiation. The extent and importance of the contacts between TFIID and promoter DNA lead to the proposal that TFIID is primarily responsible for conferring promoter specificity in PIC formation (Sainsbury *et al*, 2015). Indeed, it was proposed early on that the Initiator can fulfill a compensatory role at promoters that lack a TATA box (Smale *et al*, 1990). Consistent with this notion, work performed two decades ago demonstrated that sequences at and downstream of the TSS dictate TFIID's affinity for the hsp70 promoter (Purnell *et al*, 1994; Emanuel & Gilmour, 1993). Further experimentation is needed to determine the extent to which many of the previously mentioned downstream elements contribute to TFIID's affinity for promoters.

Of course, biology and evolution are complex and new studies are beginning to reveal just how diverse transcription initiation mechanisms might be. For instance, one study reported that a polypyrimidine-rich motif (termed the TCT motif) was present, in lieu of an Inr, at almost every RP gene promoter in both *Drosophila* and humans (Parry *et al*, 2010). The authors determined that the TCT motif is distinct from the Inr. They demonstrated that the TFIID complex containing TBP shows markedly reduced DNase I protection on TCT motif promoters compared to promoters where the sequence is mutated to TCA. Notably, the TCA mutation more closely matches the consensus Inr sequence. The reduced DNase I protection on TCT motif

promoters is presumed to be the result of weaker affinities (Parry *et al*, 2010). Thus, this study raises the question whether a non-canonical TFIID complex acts in lieu of TFIID to initiate transcription at these promoters. Reports of TAF-containing non-TFIID complexes hint at the possibility of such mechanisms. For example, TAF homologs such as *Drosophila* cannonball (TAF5 homolog), no hitter (TAF4 homolog), mia (TAF6 homolog), and rye (TAF12 homolog) all play essential roles in spermatocyte differentiation (Hiller *et al*, 2001, 2004). These factors associate with the promoter region of genes required for spermatocyte development and likely function by reducing occupancy of polycomb repressive complex 1 at these sites (Chen *et al*, 2005). These studies suggest that tissue-specific non-canonical TFIID complexes may confer special regulatory properties. When considered with the fact that core promoters show considerable diversity in composition, such studies have led some to question how general the GTFs truly are. Consequently, it has been proposed that the name be changed from general transcription factors to basal transcription factors (Danino *et al*, 2015). TFIID's association with promoters lacking canonical core elements will be more fully addressed in Chapter 2.

The majority of studies on PIC formation use strong promoters including the AdML, CMV, or SCP1 promoters as well as the yeast CYC1, ADH1, CYC7, HIS3, and HIS4 promoters (Lewis & Manley, 1985; He *et al*, 2013; McNeil & Smith, 1985; Buratowski *et al*, 1988; Chen & Struhl, 1985). These promoters contain canonical elements including a consensus TATA box, an Inr sequence, and some combination of downstream promoter elements including the motif ten element (MTE) and DPE. While these studies have yielded troves of information, as noted above, the fact remains that most promoters lack many or most of the elements highlighted in these studies. As mentioned above, the Inr encompasses the TSS and is the most common core promoter sequence element, yet, it is found less than half of active genes with a focused

promoter (Vo Ngoc *et al*, 2017a). Furthermore, in addition to lacking a TATA element, several studies have demonstrated that not all promoters are bound by TBP (Isogai *et al*, 2007; Mokry *et al*, 2010; Wang *et al*, 2014; Pugh & Venters, 2016). All of this ultimately raises the question, if TBP and TFIID are critical for PIC formation, how is initiation achieved at genes lacking a TATA box or TBP and Inr? All the evidence suggests that additional initiation mechanisms exist in eukaryotes, yet such mechanisms on genes lacking the canonical elements have been poorly studied. Are different combinations of basal transcription factors required to direct initiation at genes with different promoter architecture and composition?

Expounding a bit further, TBP's high degree of conservation and seemingly essential structural role in PIC formation might lead one to reasonably assume that every active gene promoter would possess a TATA box and be bound by TBP. Indeed, it has even been proposed that all promoters use TBP during initiation (Sainsbury *et al*, 2015). Yet, the lack of a TATA box at most core promoters and the fact that TBP appears to crosslink at only about 60% of Pol II-associated genes in human embryonic stem cells (Maston *et al*, 2012) suggests this may not be the case. Indeed, the first evidence of TBP-independent initiation of a Pol II-transcribed gene was provided over two decades ago when it was discovered that the zinc-finger protein YY1 drove initiation *in vitro* with only TFIIB and Pol II (Usheva & Shenk, 1994). YY1 is able to bend DNA and, like TBP, this ability is likely critical for its function in initiation (Kim & Shapiro, 1996). Notably however, this mode of initiation has only been observed when supercoiled templates were used. So, it's not clear whether the process occurs in cells. However, the finding that accurate initiation is adversely impacted when YY1's binding site is mutated suggests that YY1 may play a role in PIC formation in cells (Athanikar *et al*, 2004).

The TBP-related Factors

Like YY1, TBP-like or TBP-related factors (TRFs) offer an alternative model for initiation that may explain, in some cases, the lack of a conserved TATA box and the lack of TBP binding at thousands of active gene promoters. Four TRFs have been described in eukaryotes. Given the high sequence homology with TBP, TRFs are thought to adopt the same saddle-shaped structure as TBP (Rabenstein *et al*, 1999). TRF1, the first TRF identified (Crowley *et al*, 1993), is found only in insects and is identical to TBP at 63% of amino acid residues in the C-terminal DNA-binding domain (Rabenstein *et al*, 1999). Like all other TRFs, TRF1 is capable of interacting with TFIIA and TFIIB (Rabenstein *et al*, 1999; Moore *et al*, 1999; Teichmann *et al*, 1999). The interactions with TFIIA and TFIIB are thought to be critical for TRF1's ability to substitute for TBP from select promoters in *in vitro* transcription experiments (Hansen *et al*, 1997). While TRF1 guides transcription from alternate promoters of some Pol II-transcribed genes (Holmes & Tjian, 2000), its main function may relate to its indispensable role in Pol III-dependent transcription of tRNAs, 5S RNA, and U6 RNA (Takada *et al*, 2000). TRF3, a vertebrate-specific TRF widely expressed in all tissues, is 93% identical to TBP (Persengiev *et al*, 2003). TRF3 exists in complex with TAF3 and drives expression of Pol II-transcribed genes during myogenesis (Deato & Tjian, 2007). Notably, TRF3 has retained at least some capacity to interact with the TATA box or TATA-like sequences and is thought to alter gene specificities during development (Bártfai *et al*, 2004; Jallow *et al*, 2004; Deato & Tjian, 2007)

TRF2 is present in all metazoans and it has been proposed that TRF2 played a key role in the evolution of bilateral symmetry in animals (Duttke *et al*, 2014). Unlike TRF1 and TRF3, TRF2 is only 41% identical to TBP. The inherent differences in the DNA-binding domains TBP and TRF2 would likely make the initiation mechanisms of TRF2 incongruent with the

mechanism by which TBP initiates transcription. Specifically, like the other TRFs, TRF2 is thought to adopt the same saddle-shaped structure as TBP (Rabenstein *et al*, 1999). However, TRF2 diverges from TBP in the residues that are important for TBP's association with the TATA box. Specifically, many of the aromatic residues in TBP that are thought to be responsible for inducing the 90 degree bend in the TATA box are not present in TRF2 (Rabenstein *et al*, 1999). Thus, while TRF2 is required for transcription of select TATA-less promoters *in vitro*, how TRF2 associates with promoters remains a mystery as no binding element has been determined (Hochheimer *et al*, 2002; Ohbayashi *et al*, 2003; Chong *et al*, 2005; Isogai *et al*, 2007). Importantly though, TRF2 has retained the ability to interact with TFIIA and TFIIB (Rabenstein *et al*, 1999). One study demonstrated that TRF2, but not TBP, is involved in transcription of the histone H1 gene but not the core histone genes while other promoters showed ChIP signal for both TRF2 and TBP (Isogai *et al*, 2007). The same study concluded that TRF2 was required for the transcription of *Drosophila* Ribosomal protein (RP) genes *in vivo* (Isogai *et al*, 2007). More recent work confirmed that TRF2, but not TBP, was required for the *in vivo* and *in vitro* transcription of RP genes (Wang *et al*, 2014). Thus, TRF2 fulfills a critical, TBP-independent role at many promoters. Therefore, the study of TRF2-mediated transcription may reveal novel initiation mechanisms.

Additional Core Promoter Sequence Elements

A recent review article outlines the various elements present in metazoan core promoters. Table 1-1 lists those promoter elements, identifies their position relative to the TSS, and provides their consensus sequence. Though I haven't discussed all the elements present in

the table, the information conveys that core promoters are heterogeneous assemblies of numerous functional sequence elements.

Table 1. *Consensus sequences of some core promoter elements*

Motif	Location	Consensus
TATA box	Upstream T at −32 to −28	TATAWR
BRE^u	Upstream of TATA box	SSRCGCC
BRE^d	−23 to −17	RTDKKKK
Inr	−2 to +4	$TCA_{+1}GTY$ (*Drosophila*)
	−3 to +3	$BBCA_{+1}BW$ (human)
TCT	−2 to +6	$YYC_{+1}TTTYY$ (*Drosophila*)
	−1 to +6	$YC_{+1}TYTYY$ (human)
XCPE1	−8 to +2	$DSGYGGRAS_{+1}M$
XCPE2	−9 to +2	$VCYCRTTRCM_{+1}Y$
MTE	+18 to +22	CGANC
	+27 to +29	CGG
DPE	+28 to +32	RGWYV
DCE	Box I: +6 to +11	CTTC
	Box II: +16 to +21	CTGT
	Box III: +30 to +34	AGC
DTIE	+23 to +31	GSGRDNHGG

(W) A or T; (R) A or G; (S) G or C; (D) A, G, or T (not C); (K) G or T; (Y) C or T; (B) C, G, or T (not A); (M) A or C; (V) A, C, or G (not T); (N) A, C, G, or T (any base); (H) A, C, or T (not G). (BRE) TFIIB recognition element upstream (u) or downstream (d); (Inr) initiator; (XCPE1) X core promoter element 1; (MTE) motif ten element; (DPE) downstream core promoter element; (DCE) downstream core element; (DTIE) downstream transcription initiation element.

Table 1-1. Conserved eukaryotic core promoter elements. Reproduced from (Vo Ngoc *et al*, 2017b).

Motif 1 and M1BP

The advent of whole-genome sequencing contributed new information about the composition of core promoters. Two studies detailed the prevalence of core promoter elements throughout the genome. They also identified several other unknown or understudied motifs present in *Drosophila* promoters (Ohler *et al*, 2002b; FitzGerald *et al*, 2006). Among the elements identified was Motif 1. Motif 1 was intriguing because genes with Motif 1 tend to lack a TATA box and canonical Initiator (Ohler *et al*, 2002b). Notably, this study estimated that Motif 1 was present at 25% of the gene promoters they examined. For comparison sake, the Inr was found at 26% of core promoters investigated in this study. The prevalence of Motif 1-containing genes and their distinct composition suggests that they could serve as a valuable model for non-

canonical transcription initiation. Additionally, Motif 1-containing genes are essential to study in order to gain a complete, global understanding of transcription regulation in this important model organism. Interestingly, Motif 1 showed up in a scan of Ribosomal Protein (RP) gene core promoters (Isogai *et al*, 2007), a point that will be explored more extensively in Chapter 2. Another study reported that Motif 1 is present at many genes that are upregulated in response to amino-acid starvation. This study also demonstrated, through mutational analysis, that Motif 1 is necessary and sufficient to drive the transcriptional response to amino acid starvation (Li *et al*, 2010). Intrigued and spurred on by the scarcity of information about Motif 1-containing promoters, our lab identified the factor that binds Motif 1, named it M1BP, and began to characterize it (Li & Gilmour, 2013). This work demonstrated that M1BP binds the core promoter of nearly 2000 *Drosophila* genes and is required for Pol II-mediated transcription. Our lab also reported that Motif 1 and M1BP tend to be present at and occupy the promoters of housekeeping genes. Notably, M1BP-bound genes tend to be expressed at relatively steady levels through all stages of development and in all tissues. Furthermore, nearly all the Motif 1 sites within core promoters are bound by M1BP. This study also demonstrated M1BP genes have a strong nucleosome-depleted promoter region with a highly ordered array of nucleosomes in the body of the genes (Li & Gilmour, 2013). These findings provided the onus to pursue a more thorough characterization of M1BP in order to better understand the regulation of this prevalent and important gene class.

GFZF and Glutathione S-transferases (GSTs)

My early exploration of M1BP and Motif 1 lead to the discovery of several interacting factors. Experiments exploring the implications of two of these factors will be reported in Chapters 2 and 3. One of the factors, GFZF, has an array of zinc fingers fused to a glutathione S-

transferase (GST) homology domain (Illustrated in Figure 1-5). GSTs catalyze the nucleophilic addition of the activated thiol group of glutathione to an electrophilic substrate (Hayes *et al*, 2005). Glutathione (GSH) is a tripeptide and is the most abundant non-protein thiol in cells (Salinas & Wong, 1999). GSTs have primarily been studied for their roles in response to chemical insult and drug resistance (Hayes & Pulford, 1995). Living organisms are constantly exposed to mutagenic compounds present in the foods they consume and in the environment (Ames *et al*, 1990). Additionally, reactive oxygen species (ROS) are natural byproducts of aerobic respiration, ionizing radiation, and cellular inflammation (Hayes & Pulford, 1995). Naturally, cellular mechanisms have evolved to cope with these persistent chemical insults. GSTs play a key role in the cellular processes responsible for neutralizing these chemical threats (Mannervik & Danielson, 1988; Lavoie *et al*, 1992; Berhane *et al*, 1994). GSTs exhibit broad substrate specificity (Mannervik & Danielson, 1988) and have even been shown to sequester non-substrate compounds (Listowsky *et al*, 1988). Furthermore, it's been shown that GST expression levels correlate with cellular sensitivity to a wide spectrum of toxic compounds (Hayes & Pulford, 1995). Thus, GSTs play a clear and important role in the cellular response to toxic stressors. More recent studies have demonstrated that a π-class GST in mammals catalyzes the glutathionylation of cysteine residues of other proteins (Townsend *et al*, 2009). This is believed to be a protective mechanism that ensures these redox-sensitive cysteine residues aren't irreversibly oxidized and damaged during oxidative stress (Townsend *et al*, 2009). Additionally, studies have shown that α- and μ-class GSTs are present in the nucleus and are bound in layers to the nuclear periphery. It is believed this localization enables these GSTs to neutralize chemical threats before or upon reaching the nucleus and inflicting DNA damage (Stella *et al*, 2007). These more recent studies suggest a previously unappreciated variety in the mechanisms

employed by GSTs to protect cells during cellular stress. In Chapter 3, I provide the first evidence of a GST acting as a transcription factor (Baumann *et al*, 2017). This finding expands our knowledge about the multifaceted roles of GSTs in the cell and has interesting implications for gene regulation fluctuations following chemical insult. Furthermore, these results suggest that GSTs may have additional, undiscovered cellular functions.

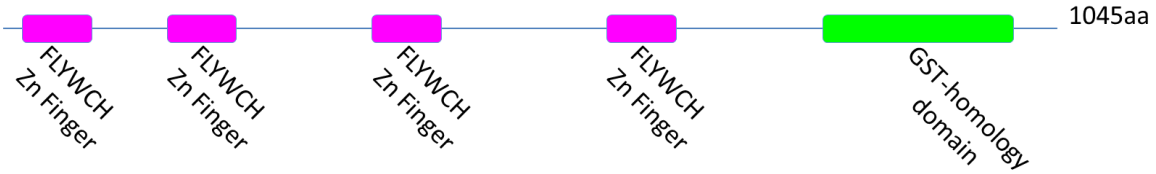

Figure 1-5. Illustration of GFZF.
GFZF is 1045 amino acid residues in length and has four N-terminal FLYWCH Zinc finger domains upstream of a GST-homology domain. The FLYWCH domain is a subtype of the abundant C2H2 zinc finger family, but little is known about FLYWCH zinc finger function.

Questions to Be Addressed

Evidence from previous studies suggested that Motif 1 and, by extension, M1BP might play a role in activation at the housekeeping gene promoters to which it binds. Motif 1 was returned from a MEME search using sequences obtained from a TRF2 ChIP-chip experiment (Isogai *et al*, 2007). Additionally, our own scanning assessment of M1BP ChIP-seq signals on the genome browser revealed that M1BP was present at many of the RP gene promoters. Thus, I formulated a hypothesis that M1BP was required for transcription of RP genes. Furthermore, since M1BP resides in the core promoter, I posited that it might drive expression of RP genes by recruiting TRF2 to RP gene promoters. I demonstrate in Chapter 2 that this is the case. The importance of this finding might extend beyond *Drosophila* since all metazoans possess TRF2;

thus, it seems plausible that analogous mechanisms for TRF2 recruitment to gene promoters exist in higher eukaryotes.

In the same study on RP genes, I also sought to determine whether TAF1 was involved in transcription from RP gene promoters. It was unclear whether it was binding RP gene promoters since the purified TFIID complex showed little to no association with TCT motif promoters as measured by DNase I footprinting. Also, the existence of TAF variants and non-canonical TAF complexes mentioned above provided a potential model for TFIID- or TAF1-independent initiation. My results show that TAF1 associates readily with RP genes and correlates with Pol II levels on virtually all active genes. This is an important clarification in the literature as it seems TAF1, with its extensive promoter contacts, functions at all Pol II-transcribed gene promoters.

In addition to understanding the potential interplay between M1BP and TRF2, I sought to identify other factors that associate with M1BP in the context of core promoter sequences. I devised an immobilized template pulldown strategy and identified several factors that are pulled down in a Motif 1-dependent fashion. One of the factors identified, Putzig, had previously been shown to interact with TRF2 in a multi-subunit complex, thus providing a potential biochemical connection between M1BP and TRF2. Another factor I identified was the GST-containing FLYWCH zinc-finger protein (GFZF). GFZF was of particular interest because it appeared as a "hit" in a number of genetic screens, yet little to nothing was known about its molecular and cellular functions. This gap in our understanding of GFZF function seemed important to address as the various screens and reports implicated GFZF in medically-relevant cellular processes including the G2 to M DNA damage checkpoint and regulation of the RAS/MAPK signaling pathway.

Additionally, GFZF was very interesting from a molecular perspective since it had an

unusual architecture consisting of four zinc fingers connected to a C-terminal glutathione S-transferase (GST) domain. Since zinc-fingers are typically thought to be involved in nucleic acid binding, this would implicate GFZF as a potential transcription regulator. After performing a number of experiments, I was able to conclude that GFZF is a transcription coactivator. This is the first report of a factor with GST activity directly affecting transcriptional output. This finding could open the door to studying unexplored mechanisms by which cells modulate their transcriptional response to oxidative stress or changes in metabolite (glutathione) concentrations in the cell. While there has been some literature describing redox regulation of proteins in cells, none have described the potential modulation of a Pol II-transcription coactivator. GFZF could thus serve as a great model system for studying the role of oxidative stress in the regulation of Pol II-mediated transcription.

Chapter 2: M1BP recruits TRF2 to coordinately transcribe ribosomal protein genes

Most of the content in this chapter has been published in Nucleic Acids Research (NAR), which allows publication of this material in my dissertation.

Introduction

The tight coordinate regulation of ribosomal protein genes is common to all organisms (Perry, 2007). These regulatory mechanisms must ensure that each of the ribosomal proteins are synthesized in the appropriate amounts to assemble ribosomes. Additionally, they need to be synthesized at levels needed to meet the translational demands of each cell (Perry, 2007). In bacteria, coordinate regulation is achieved by organizing the RP genes into operons. In the yeast *S. cerevisiae*, transcription of the ribosomal genes is coordinated by a collection of well-characterized DNA binding proteins (Reja *et al*, 2015) and references therein). Some of these proteins like Rap1, which is bound to essentially all of the RP promoters in *S. cerevisiae*, function at many genes in addition to the ribosomal protein genes. This indicates that the regulatory network controlling the RP genes involves combinatorial control rather than a single master regulator. The evolution of these mechanisms is quite fluid because the Rap1 binding sites are absent from the promoters of RP genes in several strains of yeast that are evolutionarily distant from *S. cerevisiae* (Hu & Li, 2007; Tanay *et al*, 2005). Sequence comparisons of RP genes from other eukaryotes indicate there could be considerable diversity in the mechanisms that regulate RP genes transcription (Li *et al*, 2005); (Ma *et al*, 2009), a result that is somewhat surprising given the functional conservation of the ribosomal proteins.

Much less is known about the proteins that regulate the RP genes in higher eukaryotes. Conserved sequences shared among subsets of RP genes allude to several candidates but few of these have been tested directly (Li *et al*, 2005; Hu & Li, 2007; Perry, 2005; Roepcke *et al*, 2006). One conserved element whose function has been explored is the DRE (Perry, 2005). Available evidence indicates that the DNA replication-related element (DRE)-binding factor called DREF activates RP genes in human cells (Yamashita *et al*, 2007). Other candidate proteins implicated

by the presence of conserved sequences include Sp1, NRF-2, Myc, and YY1 (Li *et al*, 2005; Hu & Li, 2007; Perry, 2007; Roepcke *et al*, 2006; Brown *et al*, 2008).

Most of the sequences that have been implicated in regulating the RP genes in animals are only present in a subset of RP genes. One exception is a pyrimidine-rich sequence called the TCT motif that encompasses the Initiator of virtually all RP genes found in *Drosophila* and humans (Parry *et al*, 2010). This sequence might restrict RP genes from using the general transcription factor, TFIID, since TFIID binds poorly to RP gene promoters from *Drosophila* and swapping the TCT motif for the Initiator sequence of the Adenovirus major late promoter greatly reduces the affinity of TFIID for this mutated promoter even though a TATA box is still present (Parry *et al*, 2010; Wang *et al*, 2014).

Recently, the TBP-related factor called TRF2 was shown to be directly involved in transcription of RP genes (Wang *et al*, 2014). In addition to regulating RP genes, TRF2 is involved in regulating the histone H1 gene (but not the genes encoding core histones) and several genes involved in development (Isogai *et al*, 2007; Kedmi *et al*, 2014). Like TBP, TRF2 associates with the general transcription factors, TFIIB and TFIIA; therefore it is likely to provide a foundation much like TBP for assembling a preinitiation complex (Rabenstein *et al*, 1999). However, there is no evidence that TRF2 binds DNA and this lack of DNA binding activity can be attributed to amino acid substitutions on the face of TRF2 that are predicted to correspond to the DNA binding face of TBP (Isogai *et al*, 2007; Moore *et al*, 1999; Rabenstein *et al*, 1999; Wang *et al*, 2014). Thus, the mechanisms by which TRF2 associates with promoters are poorly understood. TRF2 has been detected in a complex that contains DREF so the DRE found in a subset of RP genes could recruit TRF2 via DREF (Hochheimer *et al*, 2002). An uncharacterized TRF2 complex has been shown to exhibit selective binding for the canonical

Initiator sequence and downstream promoter element (DPE) found in many *Drosophila* promoters (Kedmi *et al*, 2014). Since these elements are absent from most of the RP gene promoters, the recruitment mechanism for TRF2 to those RP promoters lacking the DRE is not known.

In this chapter I present my investigation of the mechanism of transcriptional control of RP genes in *Drosophila* to gain insight into their coordinate regulation and how TRF2 associates with these promoters. This investigation was prompted by the observation that over half of the RP genes have a conserved core promoter motif known as Motif 1 (Ma *et al*, 2009). Our lab recently identified a transcription factor, dubbed M1BP, which associates with Motif 1 (Li & Gilmour, 2013). M1BP is member of the ZAD-Znf family of zinc-finger proteins that has undergone a lineage specific expansion in arthropods and could be the counterpart of KRAB-Znf or SCAN-Znf family of proteins that are prevalent in vertebrates (Tadepally *et al*, 2008; Stubbs *et al*, 2011). M1BP associates with over 1500 promoters and most of these promoters drive constitutive expression of housekeeping genes (Li & Gilmour, 2013). Here, I show that M1BP activates transcription of RP genes and that it could do so by directly interacting with TRF2 and recruiting TRF2 to the RP gene promoters. I also discover that although recent evidence indicates TBP and TFIID are not involved in RP gene transcription (Wang *et al*, 2014), TAF1, the largest subunit of TFIID, associates with all of the RP gene promoters in cells. The presence of TAF1 at RP gene promoters suggests the involvement of a TAF complex lacking TBP in transcribing the RP gene network. The specificity of this regulatory network appears to be defined in part by the combination of M1BP and TRF2 since this combination of factors is largely restricted to the RP genes. This work provides a mechanism for TRF2 recruitment to RP gene promoters and implicates a novel combination of both well conserved transcription factors

(TAF1, TRF2, and DREF) and a lineage specific transcription factor (M1BP), converging at core promoters to coordinately regulate this network of essential genes.

Results

RP genes in *Drosophila* are coordinately transcribed at exceptionally high levels

The conclusion that RP genes are highly expressed in coordinate fashion in metazoans is based largely on extrapolation of measurements of steady state mRNA levels in yeast (Warner, 1999). To more accurately assess RP gene transcription levels, I calculated PRO-seq read densities for each gene in *Drosophila* Schneider 2 (S2R+) cells (Kwak *et al*, 2013). The region from the transcription start site (TSS) to +100 was excluded to prevent bias arising from genes that are highly paused, but lowly transcribed (Adelman & Lis, 2012). I calculated the PRO-seq reads per kilobase for all genes in the active gene list provided in (Kwak *et al*, 2013) and assessed RP gene transcription activity relative to other actively transcribed genes. 62 RP genes appear in the active gene list. The other 17 have been filtered from the list either because they were not active or because their proximity to other genes could confound the bioinformatic analysis of PRO-seq signals (Kwak *et al*, 2013). Of the 62 genes on the list, 59 are transcribed in the top 10% of all active genes with the other 3 genes falling in the next decile (Figure 2-1). Thus, most of the RP genes are transcribed at roughly equivalent high levels. This correlation suggests, but cannot prove, that regulation at the level of transcription is important for coordinate RP expression**.** However, the factors involved in achieving this high level of coordinate transcription are largely unknown.

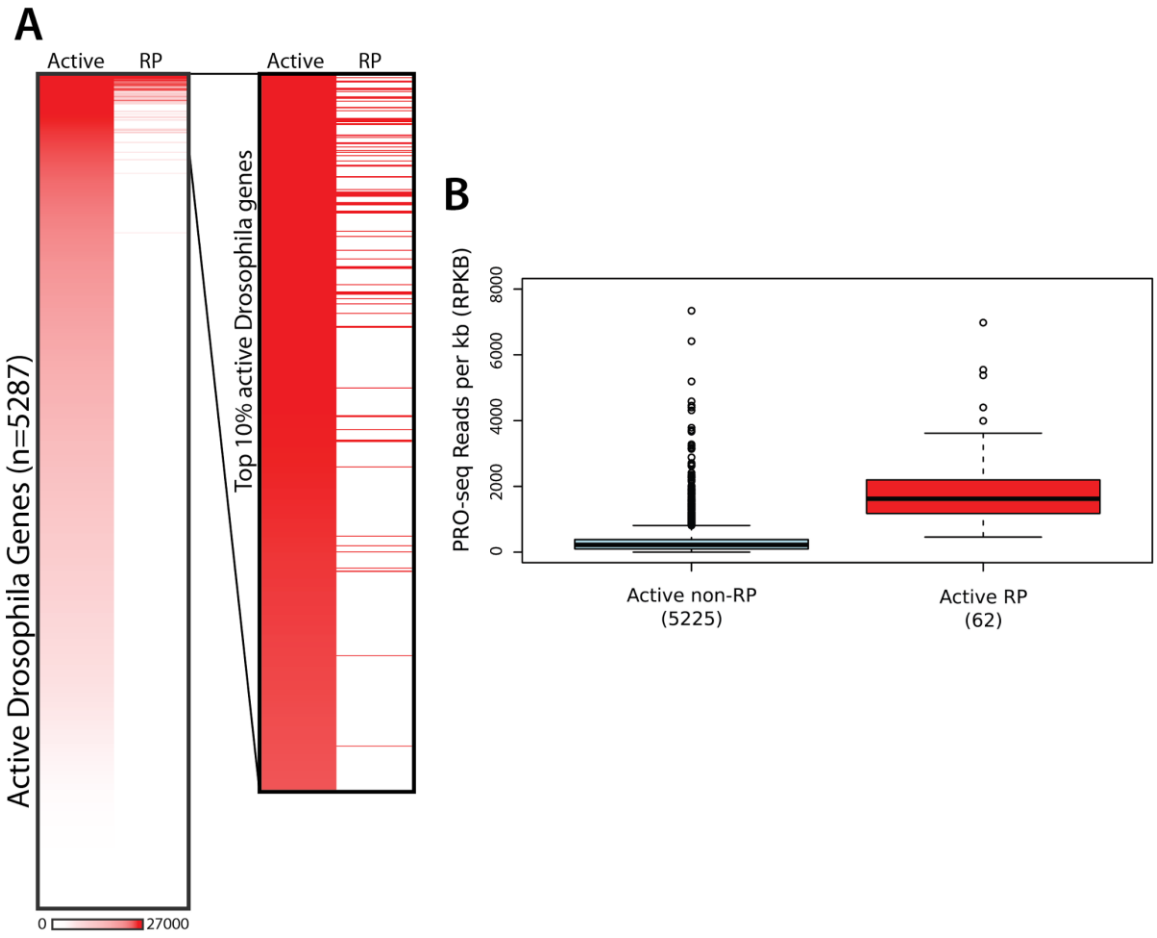

Figure 2-1. Ribosomal Protein genes are highly transcribed at approximately equivalent levels. A list of 5287 active *Drosophila* genes derived from (Kwak *et al*, 2013) were ranked by PRO-seq read densities in gene bodies. (A) RP genes are highlighted on the right-hand side in each panel. The second panel is an expanded section of the heatmap corresponding to the top 10% of active *Drosophila* genes (n=528). The bar below the left panel shows the color corresponding to the PRO-seq read densities (reads per kilobase) for each gene. (B) Box plot for the active genes with RP genes removed or RP genes only. The interquartile range is contained in the boxed region. The plots were generated with the boxplot function in R. The data point for one non-RP gene is not shown because it falls above the y-axis cutoff.

M1BP activates transcription of RP genes in cells

Previous analyses identified a conserved sequence called Motif 1 among the RP gene promoters (Ma *et al*, 2009) and our laboratory recently discovered a protein, named M1BP, that associates with Motif 1 (Li & Gilmour, 2013). To determine if Motif 1, and by extension M1BP,

is involved in transcribing RP genes, I used a luciferase reporter assay (Illustrated in Figure 2-2) with the promoter sequences (-500+50) from either *RpLP1* or *RpL30*.

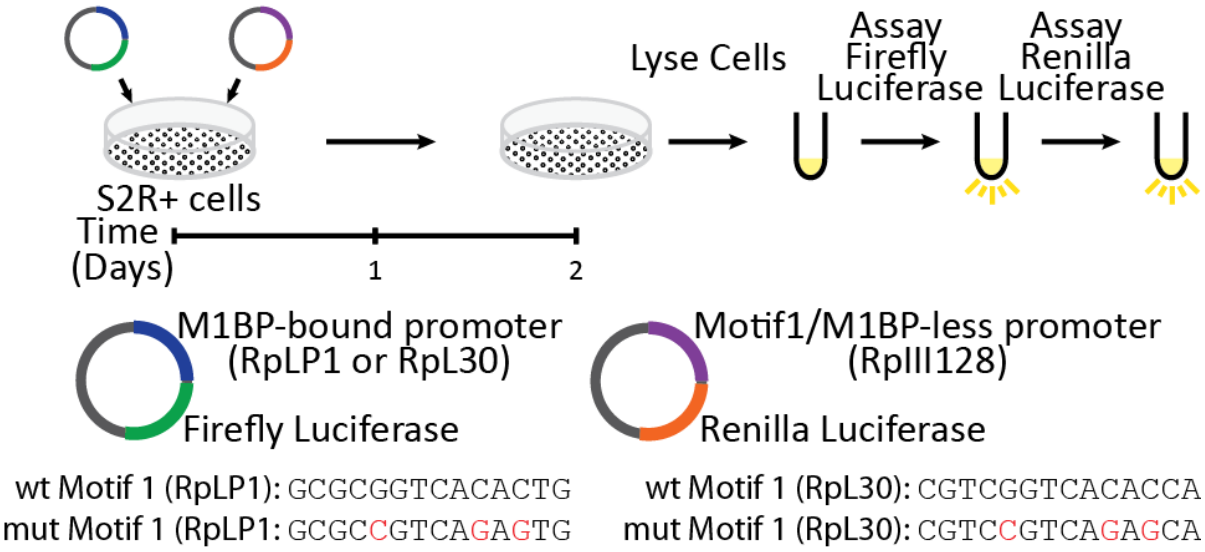

Figure 2-2. Schematic of the RP gene luciferase reporter assay.
Mutant *RpLP1* and *RpL30* have 3 highly conserved nucleotides in Motif 1 mutated. The wt and mutant Motif 1 sequences for both promoters are shown below the illustration. The RpIII128 promoter lacks Motif 1 and serves an internal control.

The *RpLP1* and *RpL30* gene promoters had previously served as models for RP gene transcription (Wang *et al*, 2014; Parry *et al*, 2010). I also prepared mutant counterparts with 3 point mutations in Motif 1 that are known to abolish M1BP binding (Li & Gilmour, 2013). Following transfection into *Drosophila* S2R+ cells, I saw greater than 5-fold decline in luciferase levels when Motif 1 was mutated (Figure 2-3). The results demonstrate that Motif 1 contributes to transcription at RP gene promoters and implicates M1BP in transcriptional activation of RP genes.

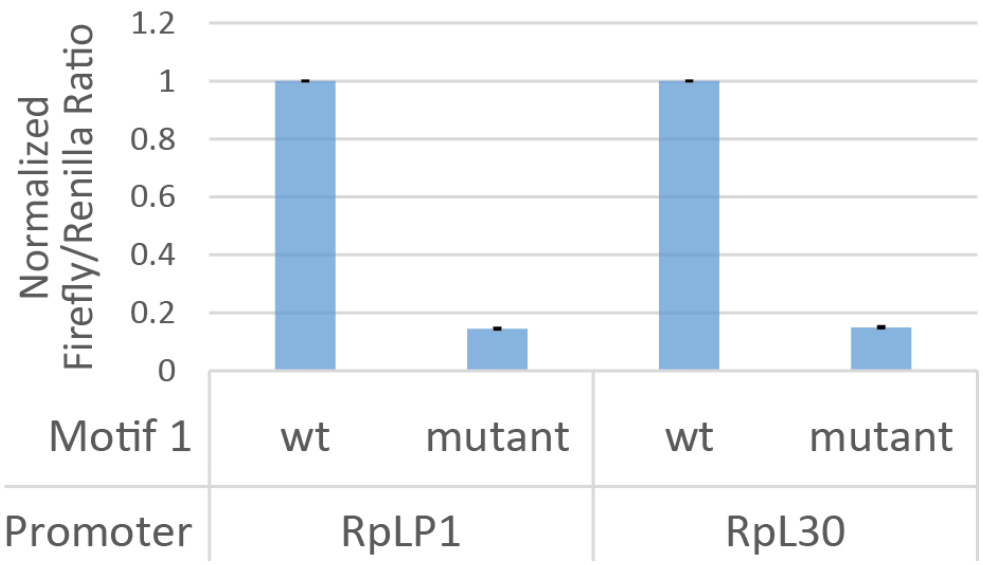

Figure 2-3. Motif 1 is required for RP gene transcription in cells.

Firefly/Renilla luciferase ratio of relative light unit measurements. Ratios are normalized to the wt Motif 1 sample for each promoter. Error bars represent standard deviation (n=3 biological replicates).

To directly test the role of M1BP in RP gene activation, I performed *in vitro* transcription in nuclear extracts which allowed me to determine the effects on RP gene transcription of both mutating Motif 1 and depleting M1BP. Mutating Motif 1 in the *RpLP1* and *RpL30* promoter caused about a 2 to 4-fold decrease in transcription which approximates the effect seen in cells (Figure 2-4 A-B, cf. lanes 2-4 with 5-7 and Figure 2-4 C). α-amanitin inhibited transcription of the *RpL30* promoter indicating that the transcription was mediated by Pol II (Figure 2-4 A-B, cf. lane 1 with lanes 2-4).

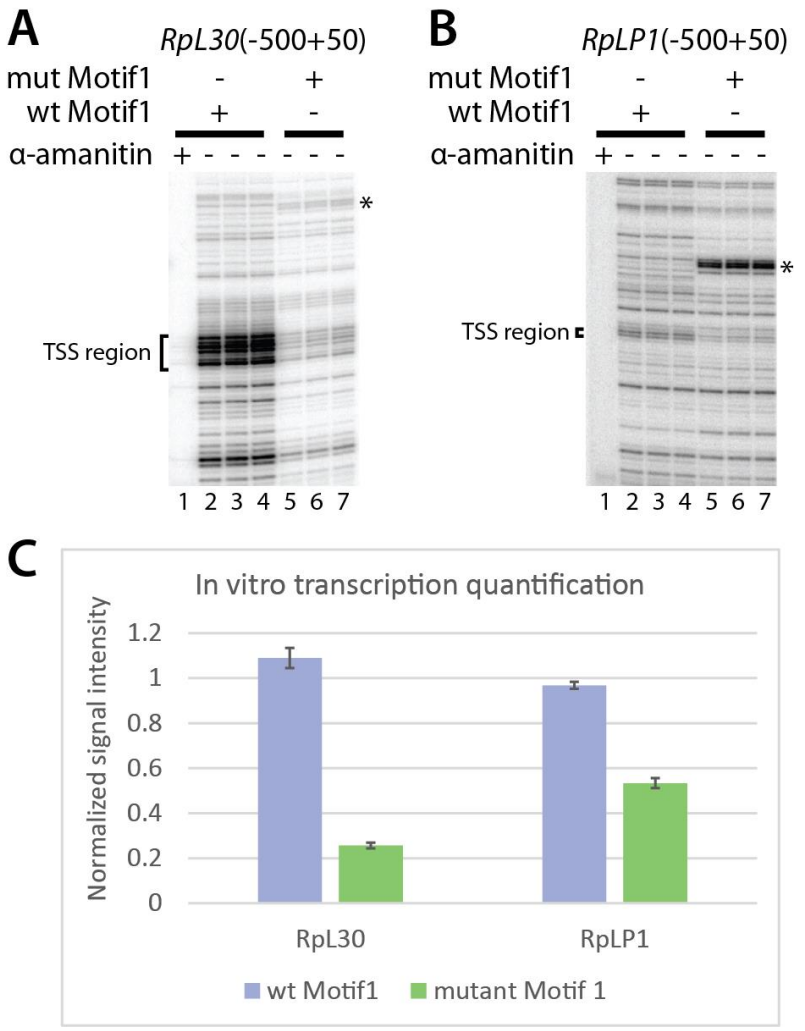

Figure 2-4. Motif 1 contributes to transcription of *RpL30* and *RpLP1 in vitro*.

(A-B) Primer extension analysis of transcripts produced from the (A) *RpL30* and (B) *RpLP1* promoters (-500 to ~+50) during transcription in *Drosophila* embryo nuclear extracts. Transcription reactions lacking α-amanitin were performed in triplicate. The bracketed region encompasses the M1BP-dependent TSS region and a portion of the TCT motif (Parry *et al*, 2010). The M1BP-dependent transcription start sites observed *in vitro* correspond to the transcription start sites detected *in vivo* using PRO-cap (30). The bands outside the bracketed region are produced by Pol II and likely represent nonspecific transcription, which occurs at levels comparable to transcription of the RP genes. (C) Quantification of bracketed TSS region transcripts from (A-B). Error bars represent standard deviation (n=3 biological replicates). Samples have been normalized to the first wt replicate for each promoter. * denotes an artifact band arising in the Motif 1 region following mutation.

The spurious pattern of bands existing outside the TSS region could be attributed to either weak promoter activity or to nuclear extracts of poor quality. To assess this, I performed primer extension analysis of *in vitro* transcription reactions using the *Hsp70Bc* promoter as a template and compared these results with those from the *RpLP1* and *RpL30* promoter templates. The *Hsp70Bc* promoter is highly active *in vitro*, thus, it should have a high ratio of bands from the TSS region compared to regions outside the TSS. Figure 2-5 demonstrates that the extracts are of good quality as the ratio of bands in the TSS region compared to non-TSS regions is a function of promoter strength. Thus, the *Hsp70Bc* and *RpL30* promoters have considerably higher intensity bands in the TSS region than *RpLP1*. Nevertheless, *RpLP1* does behave in a manner consistent with what is observed with *RpL30* both *in vivo* (Figure 2-3) and *in vitro* (Wang *et al*, 2014; Parry *et al*, 2010), thus, while higher activity from *RpLP1* would be desirable, it still serves as an additional, valuable model of initiation from RP gene promoters.

To determine if Motif 1 was functioning through M1BP, I immunodepleted M1BP from the nuclear extract (Figure 2-6 A). Immunodepletion of M1BP caused a decrease in RP gene transcription (Figure 2-6 B-C, lanes 1-2 vs 5-6; Quantifications in D-E). To establish that the immunodepletion of M1BP itself, rather than some associated protein was responsible for inhibiting transcription of the RP genes *in vitro*, I expressed and purified M1BP from *E. coli* (Figure 2-6 F). Addition of recombinant M1BP to the M1BP-depleted nuclear extract restored RP gene transcription to its normal level (Figure 2-6 B-C, cf. lanes 5 and 6 to lanes 7 and 8) indicating that M1BP activates transcription of the RP gene *in vitro*. Addition of recombinant M1BP to the mock-depleted extract had no effect on RP gene transcription suggesting the M1BP is not limiting in the mock-depleted extract (Figure 2-6 B-C, cf. lanes 1 and 2 to lanes 3 and 4).

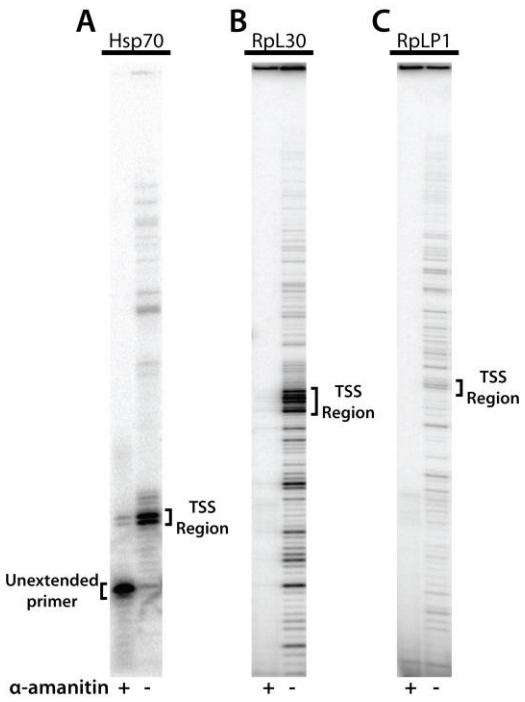

Figure 2-5. Promoter strength correlates with higher signal-to-noise ratios.

(A-C) Primer extension analysis of transcripts produced from the (A) *Hsp70Bc*, (B) *RpL30*, and (C) *RpLP1* promoters in *Drosophila* embryo nuclear extracts. The bracketed regions to the right of each phosphorimage delimits the TSS regions for each promoter. The bracket to the left of the phosphorimage in panel A delimits the unextended primer band for the *Hsp70Bc* reactions. The unextended primer bands are not shown for the RP gene promoters.

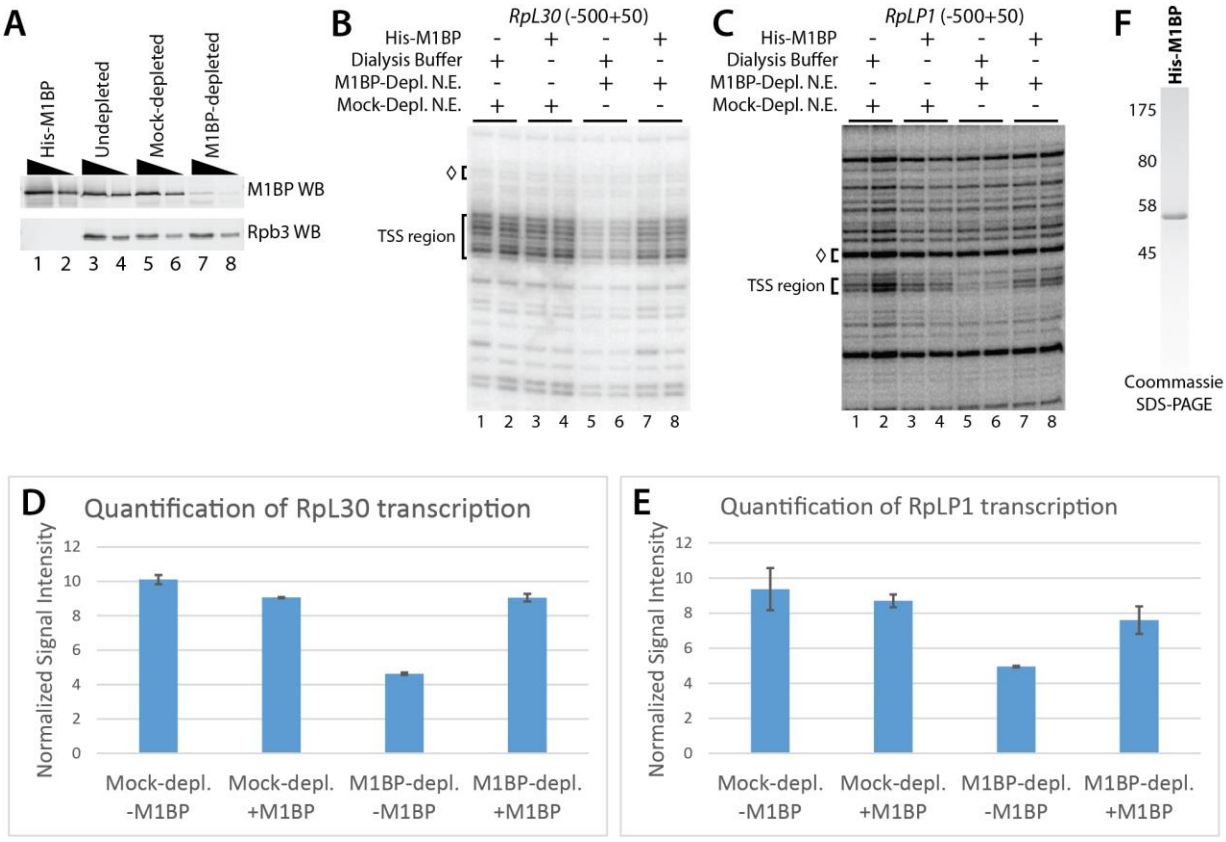

Figure 2-6. M1BP is required for RP gene transcription *in vitro*.

(A) M1BP-probed (top) and Rpb3-probed (bottom) western blot of purified His-M1BP and undepleted, mock-depleted, or M1BP-depleted nuclear extracts from 0-12 hour embryos. 10 ng or 30 ng purified His-M1BP and 20 µg or 60 µg of each extract type was loaded for SDS-PAGE western blot analysis. (B-C) Primer extension analysis of transcripts produced from the (B) *RpL30* or (C) *RpLP1* promoter in embryo nuclear extracts. The bracketed region denotes the same TSS region described in Figure 2-4. Each transcription reaction was performed in duplicate. Lanes 1 and 2: Mock-depleted extract supplemented with dialysis buffer. Lanes 3 and 4: Mock depleted extract supplemented with recombinant M1BP. Lanes 5 and 6: M1BP-depleted extract supplemented with dialysis buffer. Lanes 7 and 8: M1BP-depleted extract supplemented with enough recombinant M1BP to replace the amount that was immunodepleted. (D-E) Normalized quantifications of the bracketed TSS regions for the gels shown in B-C. Plotted are the average signals of the two replicates with the range represented by the error bars. The bracketed region denoted by ◊ in B-C corresponds to the band(s) used for sample normalizations between lanes. (F) Coomassie-stained SDS-PAGE analysis of purified, N-terminally His-tagged M1BP expressed in and purified from *E. coli*.

M1BP recruits TRF2 to the RP gene promoter

Recently the TBP-related factor, TRF2, was found to be involved in transcription of RP genes (Wang *et al*, 2014). How TRF2 associates with promoters is enigmatic since, unlike its counterpart TBP, TRF2 has not been observed to bind directly to DNA. Motif 1, the binding site for M1BP, is typically located within 50 nucleotides of the transcription start site of RP genes, making it a core promoter element (Ohler *et al*, 2002b). In addition, I had observed that Putzig, a protein previously shown to associate with TRF2, co-immunoprecipitated with M1BP (Figure 3-1). Hence, I investigated the possibility that M1BP might be recruiting TRF2 to RP gene promoters by performing immobilized template pulldown experiments with DNA template sequences corresponding to the core promoter regions of two RP genes. *RpLP1* and *RpL30* promoter sequences harboring normal or mutated Motif 1 elements were immobilized on beads. Following immobilization, purified recombinant M1BP, and *in vitro* translated ^{35}S-labeled TRF2 were added either together or separately to the immobilized DNA and proteins that associated with the DNA templates were assessed by SDS-PAGE (Figure 2-7).

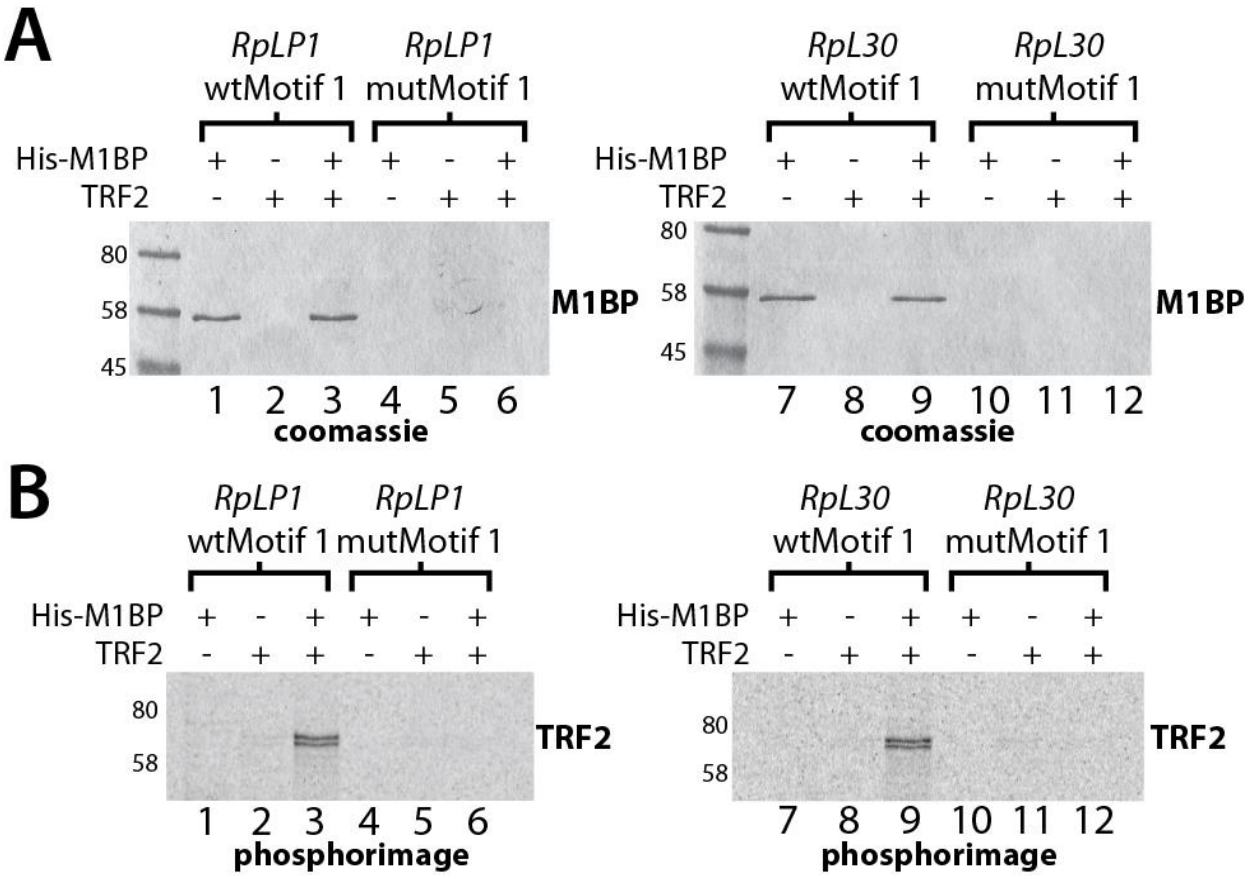

Figure 2-7. M1BP recruits TRF2 to RP gene promoters.

Immobilized template pull-down assay. ^{35}S-labeled TRF2 was synthesized *in vitro* using Promega's TnT T7 Quick Coupled Transcription/Translation System. His-M1BP was expressed in and purified from *E. coli*. His-M1BP, TRF2, or TRF2 and His-M1BP were added to either *RpLP1* or *RpL30* template-bound streptavidin Dynabeads containing either a wild-type or mutant Motif 1 sequence (wt Motif 1 or mut Motif 1, respectively). Panel A shows coomassie-stained images from SDS-PAGE analysis of bound protein recovered from *RpLP1* and *RpL30* immobilized templates. Panel B shows phosphorimager scans of the same gels in panel A. M1BP binds only to the wt Motif 1 template regardless of whether TRF2 is present in the reaction. TRF2 only binds to the wt Motif 1 promoter template when M1BP is present.

M1BP binds in a Motif 1-dependent manner to both promoter sequences in the absence of TRF2 (Figure 2-7 A, cf. lanes 1, 4, 7, and 10), while TRF2 alone does not bind to any of the promoter fragments (Figure 2-7 B, lanes 2, 5, 8, and 11). In contrast, when M1BP is added

together with TRF2, TRF2 associates with the Motif 1-containing M1BP-bound promoters (Figure 2-7 A-B, lanes 3 and 9). Additionally, I expressed Maltose-binding protein (Mal) fused to M1BP and was able to pull down TRF2 indicating that TRF2 and M1BP interact in solution as well as on DNA templates (Figure 2-8, A-B). Altogether, these results show that M1BP can recruit TRF2 to the promoter but TRF2 is not required for M1BP to bind the promoter.

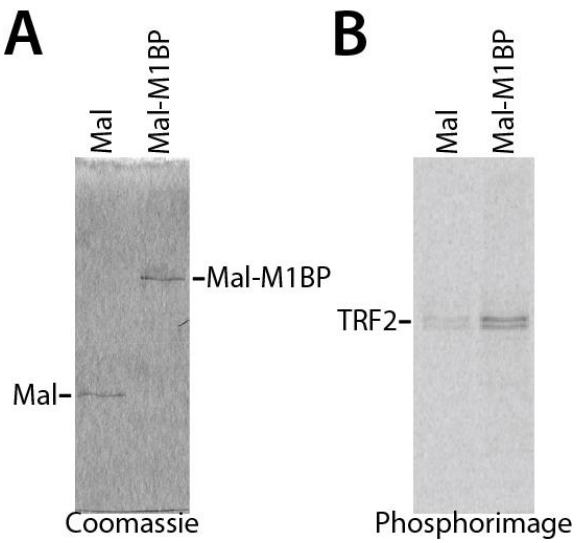

Figure 2-8. M1BP interacts with TRF2 in solution.

Maltose-binding protein (Mal) and Mal-M1BP fusion pulldown assay. ^{35}S-labeled TRF2 was synthesized as described in Figure 2-7 and added to either purified, amylose resin-bound Mal or Mal-M1BP. Panel A shows coomassie-stained images from SDS-PAGE analysis of bound protein recovered after binding and washing. Panel B shows the phosphorimage scans of the same gels in panel A. Recovery of TRF2 is increased with the Mal-M1BP fusion compared to the Mal alone.

In order to determine if the immobilized template pulldowns reflect the binding properties in cells, I depleted TRF2 or M1BP using RNAi and monitored the association of each protein with representative RP genes using chromatin immunoprecipitation. I found that TRF2 depletion caused significant loss of TRF2 from RP gene promoters while M1BP levels remained unchanged (Figure 2-9 A-B). A western blot confirmed that the RNAi worked as expected since

I observed total cellular TRF2 levels decrease, while the levels of M1BP and NELF-E remained unaffected (Figure 2-9 C).

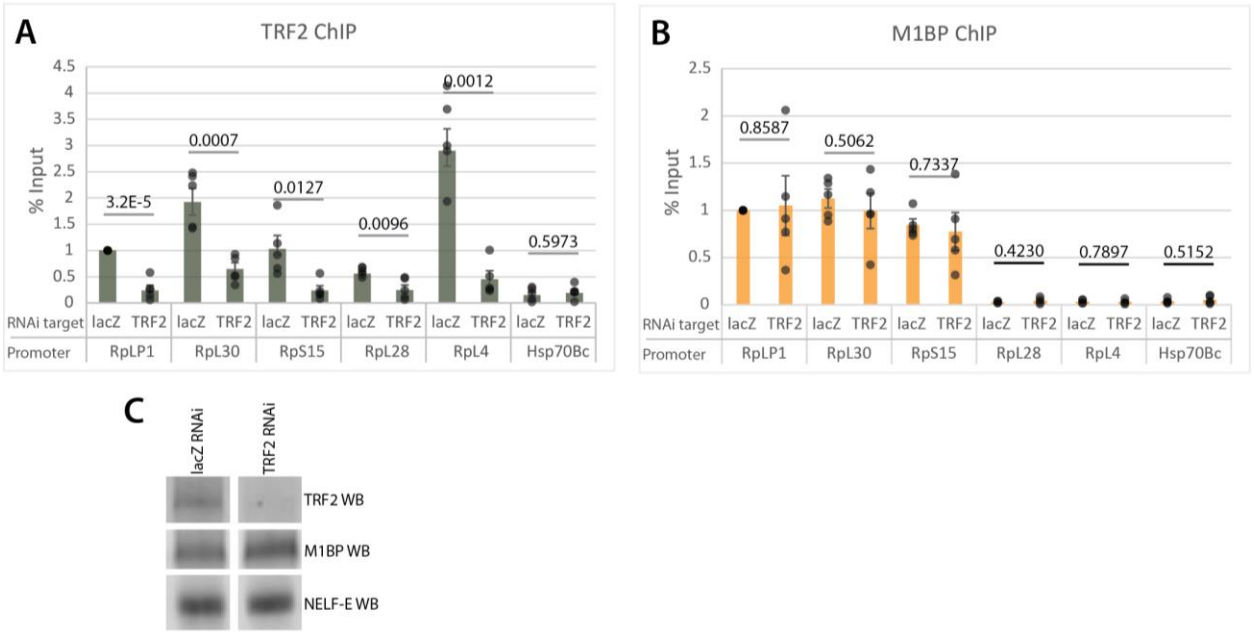

Figure 2-9. TRF2 is dispensable for M1BP's association with RP gene promoters.
(A-B) RNAi-mediated depletion of TRF2. Following 3-day RNAi knockdown using dsRNA targeting either *lacZ* (negative control) or *Trf2*, cells were lysed and ChIP experiments were performed for TRF2 or M1BP. qPCR quantifications were normalized to *lacZ* RNAi signal at *RpLP1*. *Hsp70Bc* lacks both factors and thus serves as a negative control. *RpL28* and *RpL4* lack M1BP. Individual data points are displayed as gray dots. Each experiment was performed at least 4 times. Error bars indicate standard deviation. p-values from two-tailed T-tests are provided for each promoter. (C) Western blots from S2R+ chromatin lysates used for ChIP following 3-day RNAi. The RNAi targets are indicated above the blot images. *lacZ* RNAi served as a negative control.

Depletion of M1BP using RNAi caused a decrease of both M1BP and TRF2 from RP genes (Figure 2-10 A-C). However, I was unable to conclude that TRF2's association was directly linked to M1BP promoter binding in cells because depletion of M1BP also resulted in loss of TRF2 from *RpL28* and *RpL4*. These two RP genes were not bound by M1BP. Hence, the loss of TRF2 that occurred when M1BP was depleted could be a direct effect of the loss of

M1BP or an indirect effect of the coordinate repression of RP genes that likely occurs as depletion of M1BP diminishes the rate of cell proliferation (Li & Gilmour, 2013).

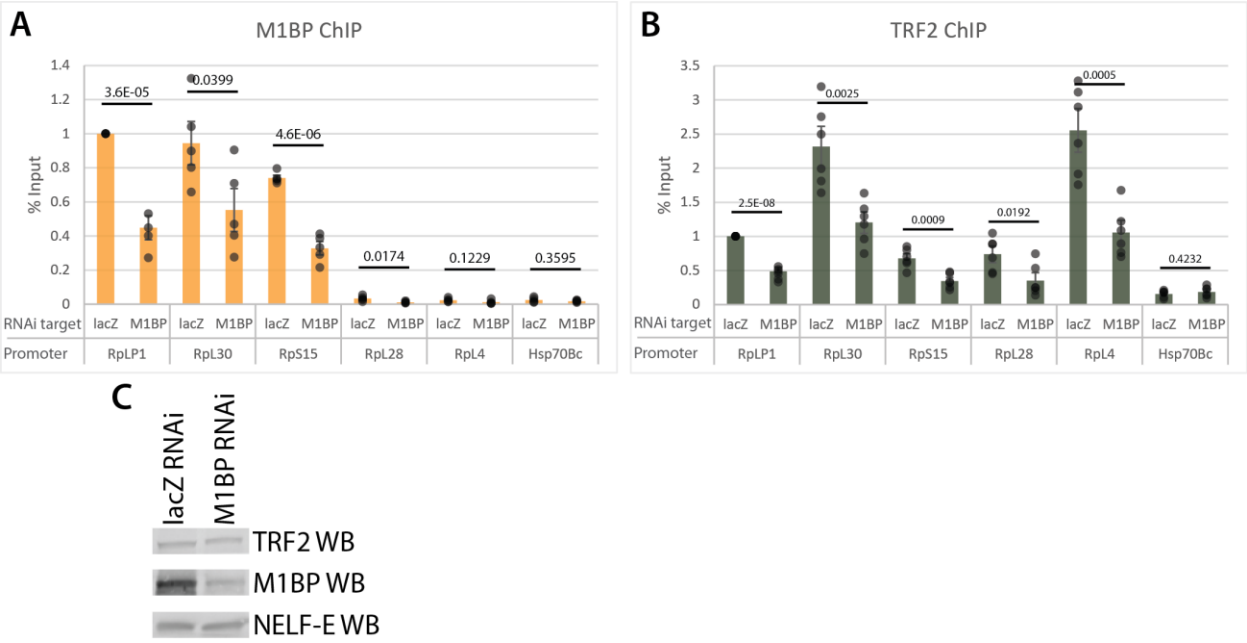

Figure 2-10. TRF2 levels at RP gene promoters decreases with M1BP depletion irrespective of the presence of M1BP.

(A) M1BP and (B) TRF2 ChIP levels at the indicated promoters from S2R+cells following 5-day treatment with dsRNA targeting either *lacZ* (control) or *M1BP*. Individual data points are displayed as gray dots. Each experiment was performed 6 times. Error bars indicate the standard deviation. p-values from two-tailed T-tests are provided for each promoter. (C) Western blot of TRF2, M1BP, and NELF-E following the depletion of M1BP by *M1BP* RNAi. Despite having M1BP at its promoter, TRF2 levels are unaffected following M1BP depletion. NELF-E expression is not regulated by M1BP and thus serves as a control.

ChIP-exo provides evidence that M1BP recruits TRF2 to the majority of RP genes *in vivo*

Because of the pleiotropic effects that might accompany depleting M1BP from cells, I turned to ChIP-exo analysis to investigate the relationship between M1BP and TRF2 in cells. Recently, ChIP-exo analyses of factors associated with the RP genes in yeast provided insight

into the protein-protein interactions that are involved in regulating these genes (Reja *et al*, 2015). ChIP-exo analysis maps at near single nucleotide resolution the sites where a protein crosslinks to DNA by treating immunoprecipitated protein-DNA adducts with lambda exonuclease and subjecting the digested DNA to high-throughput sequencing (Rhee & Pugh, 2012). Since lambda exonuclease digests the DNA in a 5' to 3' direction and is blocked by protein-DNA crosslinks, protein binding sites are demarcated by sequencing tags on opposite strands that manifest as peak pairs.

M1BP binding sites on the RP gene promoters in cells were readily detected by ChIP-exo and are strikingly different from a mock ChIP-exo pattern (Figure 2-11).

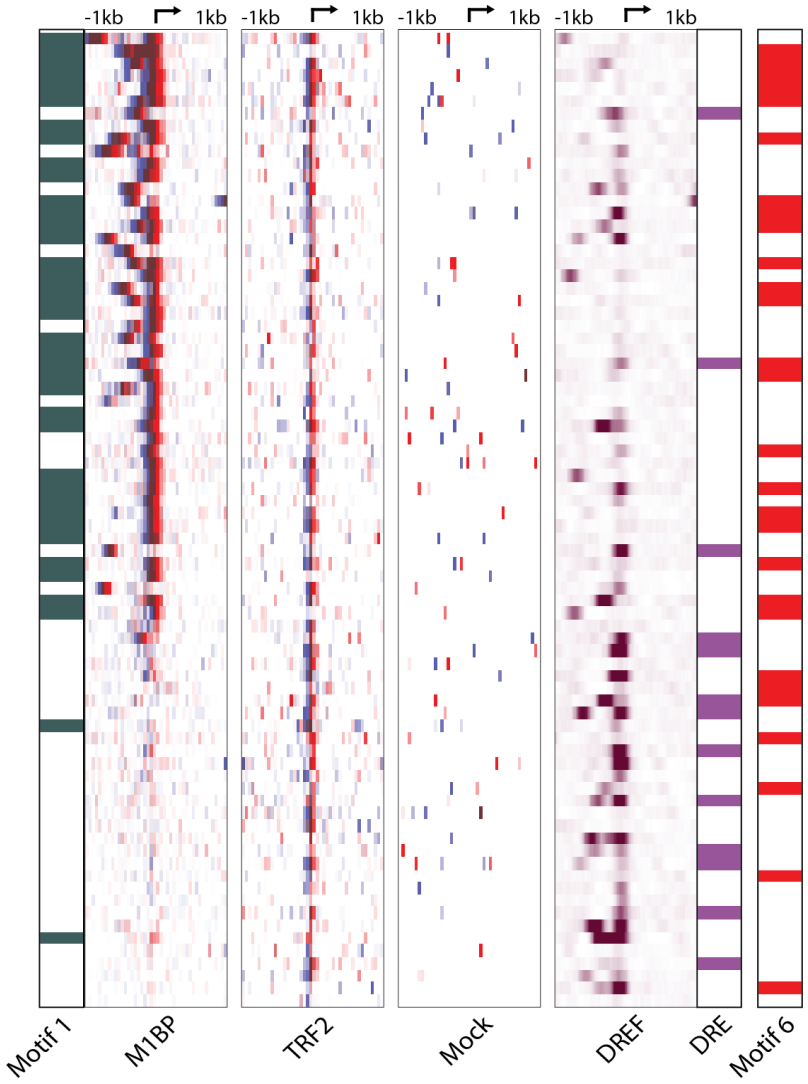

Figure 2-11. M1BP and TRF2 co-occupy the majority of RP gene promoters.

M1BP, TRF2, and Mock ChIP-exo reads and DREF ChIP-seq reads mapped relative to the TSS of 78 RP genes and sorted by M1BP ChIP-exo reads summed in a 2kb window centered on the TSS. RP genes having Motif 1, DRE, or Motif 6 within 200 bp of the TSS are indicated in green in the far left panel or purple or red in the two far right panels, respectively. The arrow at the top of each heatmap marks the transcription start site. 8 paralogs lacking a TCT motif or TRF2 peak have been removed. This criteria was used since TRF2 was shown to be required for RP gene transcription, and the TCT motif is present in lieu of an Inr sequence in at least one paralog at 77 out of 79 RP genes. Finally, *RpL15* is present on Chr3LHet, a region that was excluded in my mapping and was thus removed from this RP gene list. Therefore, the final number of RP genes used for this and subsequent heatmap and composite plot analysis is 78.

I called peaks and the data confirmed that M1BP is highly enriched at RP gene promoters when compared to all active genes ($p<0.0001$, Fisher's exact test, two-tailed). Composite plots using the TSS as a reference point reveal a complex pattern of crosslinking that extends from -150 on the top strand to +50 on the bottom strand (Figure 2-12 A). The M1BP pattern is unlikely to be due solely to M1BP crosslinking directly to DNA since M1BP has 5 zinc fingers which are predicted to be just enough to bind the 10-15 nucleotide long Motif 1 (Figure 2-12 B).

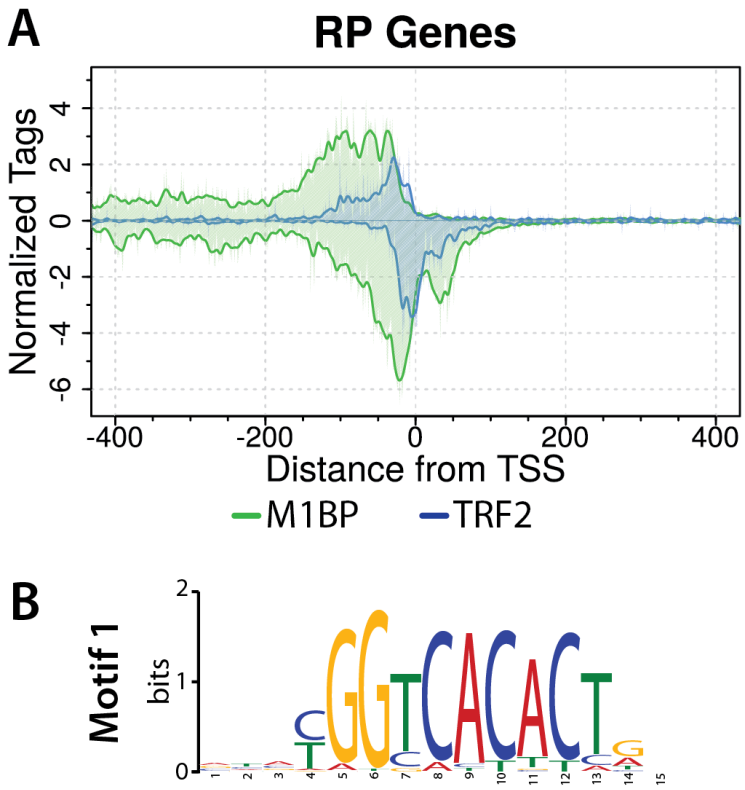

Figure 2-12. M1BP and TRF2 show overlapping ChIP-exo patterns on RP genes.
(A) ChIP-exo analysis of M1BP (green trace) and TRF2 (blue trace) for RP genes. Composite plots in single nucleotide bins were generated from the same RP gene list used for the heatmap in Figure 2-11. (B) Logo representation of the Motif 1 position weight matrices used to identify genomic motif locations.

This ChIP-exo pattern resembles the broad pattern recently described for the yeast RP genes and was attributed to multi-protein complexes crosslinking to DNA (Reja *et al,* 2015). ChIP-exo analysis of TRF2 revealed a more compact pattern of crosslinks than M1BP. The majority of TRF2 crosslinks occurred immediately upstream from the TSS (Figure 2-12 A). Comparison of the composite plots of TRF2 and M1BP revealed striking overlap of the TRF2 ChIP-exo pattern with the M1BP peak pair most proximal to the TSS. This type of overlap in ChIP-exo patterns has been interpreted to reflect the overlap between two factors binding in concert with one another (Reja *et al*, 2015), thus the data provides evidence consistent with our binding assays which indicates that TRF2 interacts with M1BP.

TRF2 is present at almost all of the RP gene promoters yet M1BP is detected at approximately two-thirds of them (See Figure 2-11). TRF2 was previously found to be in a complex with DREF (Hochheimer *et al*, 2002) so I wondered if the DRE and, by extension, DREF might function in recruiting TRF2 to those RP genes that lack M1BP. To explore this possibility, I used previously published DREF ChIP-seq data (Li *et al*, 2015) and determined that DREF is enriched at RP genes when compared to all active genes ($p=0.0394$, Fisher's exact test, two tailed). Notably, DREF was further enriched among those RP gene promoters that lack M1BP (Figure 2-12, $p=0.0009$, Fisher's exact test, two-tailed). Thus, two mechanisms appear to function to recruit TRF2 to RP gene promoters.

(Ohler *et al*, 2002b) reported that Motif 1 is enriched at genes that possess Motif 6, which is another conserved promoter element in *Drosophila*. The factor that recognizes Motif 6 is unknown. Given the connection between Motif 6 and Motif 1, I wondered if Motif 6 might also contribute to RP gene transcription. To that end, I identified genes that have Motif 6 within 200 bp of the TSS and found that Motif 6 is statistically enriched at RP gene promoters compared to

all active genes (p=0.001, Fisher's exact test, two tailed). Additionally, Motif 6 is often found at genes that lack M1BP and/or DREF (Figure 2-11, far right panel). Thus, Motif 6, or the factor that binds it, might also contribute to TRF2 recruitment and RP gene transcription.

ChIP-exo analysis detects TAF1 at RP gene promoters.

My ChIP-exo pattern for M1BP revealed a peak of crosslinks on the bottom strand, 30 to 50 nucleotides downstream from the transcription start site. Since this region of the promoter was not required to bind M1BP to immobilized DNA, I suspected that these crosslinks might involve another protein that interacts with M1BP. Crosslinking and Cryo-EM studies have shown that TAF1 contacts this region of the promoter (Sypes & Gilmour, 1994; Wu *et al*, 2001; Louder *et al*, 2016). My ChIP-exo analysis reveals that TAF1 is present at virtually all of the RP gene promoters (Figure 2-13 A). On RP genes, the downstream peak of TAF1 coincides well with the downstream peak of M1BP (Figure 2-13 B, orange arrowhead) raising the possibility that this M1BP peak is the result of M1BP crosslinking to TAF1 which is in turn crosslinked to this downstream region.

I was interested in determining if there was any relationship between the presence of TAF1 and the level of paused Pol II. TAF1 appears to be present at virtually all genes with Pol II and most TRF2-associated promoters (Figure 2-14 and 2-15). I used PRO-seq reads as a readout of Pol II levels for several reasons. PRO-seq is quantitative, exhibits high dynamic range, and provides precise strand-specific position information (Kwak *et al*, 2013). Also, high-quality PRO-seq data was available from S2R+ cells grown under the same conditions I used for my ChIP-exo analysis. It should be noted that PRO-seq cannot detect initiating Pol II since 15-20 nucleotides of RNA are necessary for genome mapping.

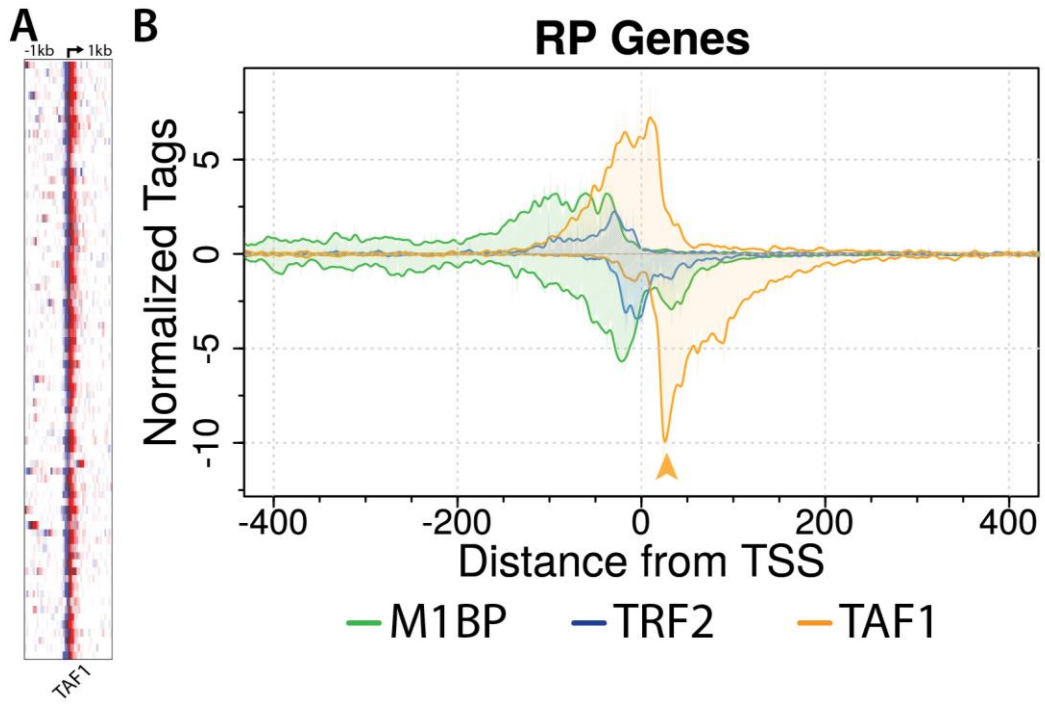

Figure 2-13. TAF1 occupies RP gene promoters in regions that overlap with M1BP and TRF2. (A) Heatmaps display TAF1 ChIP-exo reads from S2R+ cells piled from -1 kb to +1 kb relative to RP gene transcription start sites (TSS). Rows represent individual genes and are sorted by M1BP reads summed in a 2kb window as in Figure 2-11. The transcription start site position is indicated by the arrow at the top of the panel. Duplicate genes were refined to a single isoform by removing 8 paralogs lacking TRF2 or a TCT motif in the promoter. (B) Composite plots for M1BP, TRF2, and TAF1 were generated from the same RP gene list used for the heatmaps. The orange arrow highlights a TAF1 peak that aligns with an M1BP peak.

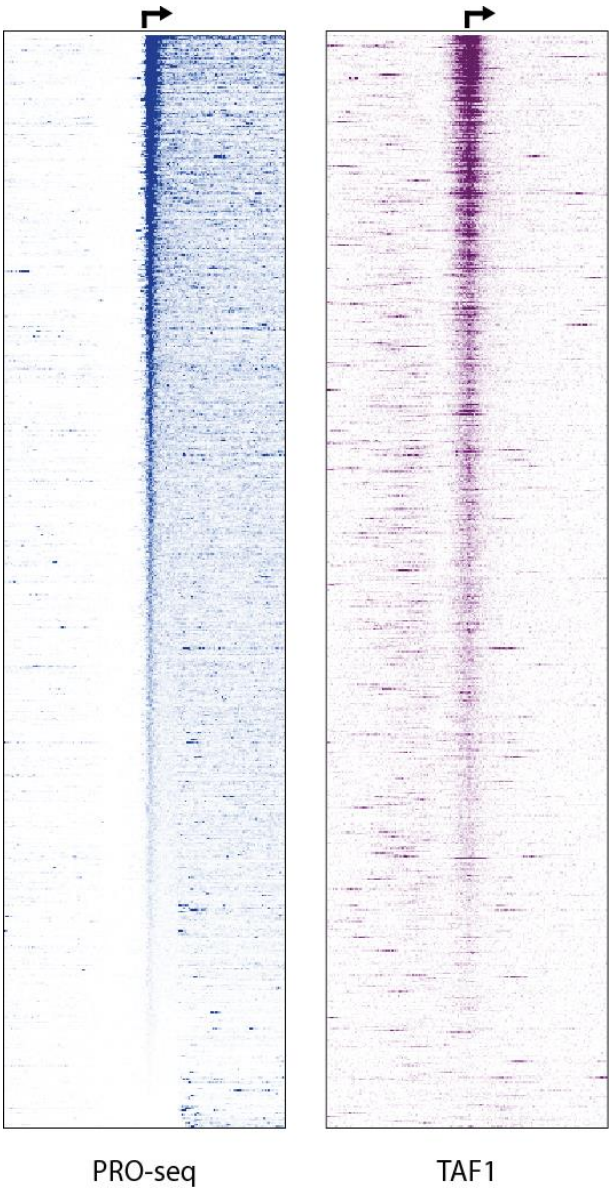

Figure 2-14. TAF1 is associated with actively transcribed promoters.

PRO-seq (left) and TAF1 ChIP-exo (right) reads mapped in a 2 kb window relative to the TSS of active non-RP genes (n=5225). Arrows above the plots represent the TSS. Genes are sorted by PRO-seq reads summed in a -50 to +250 window relative to the TSS. Top and bottom reads from the ChIP-exo analysis are displayed in the same color.

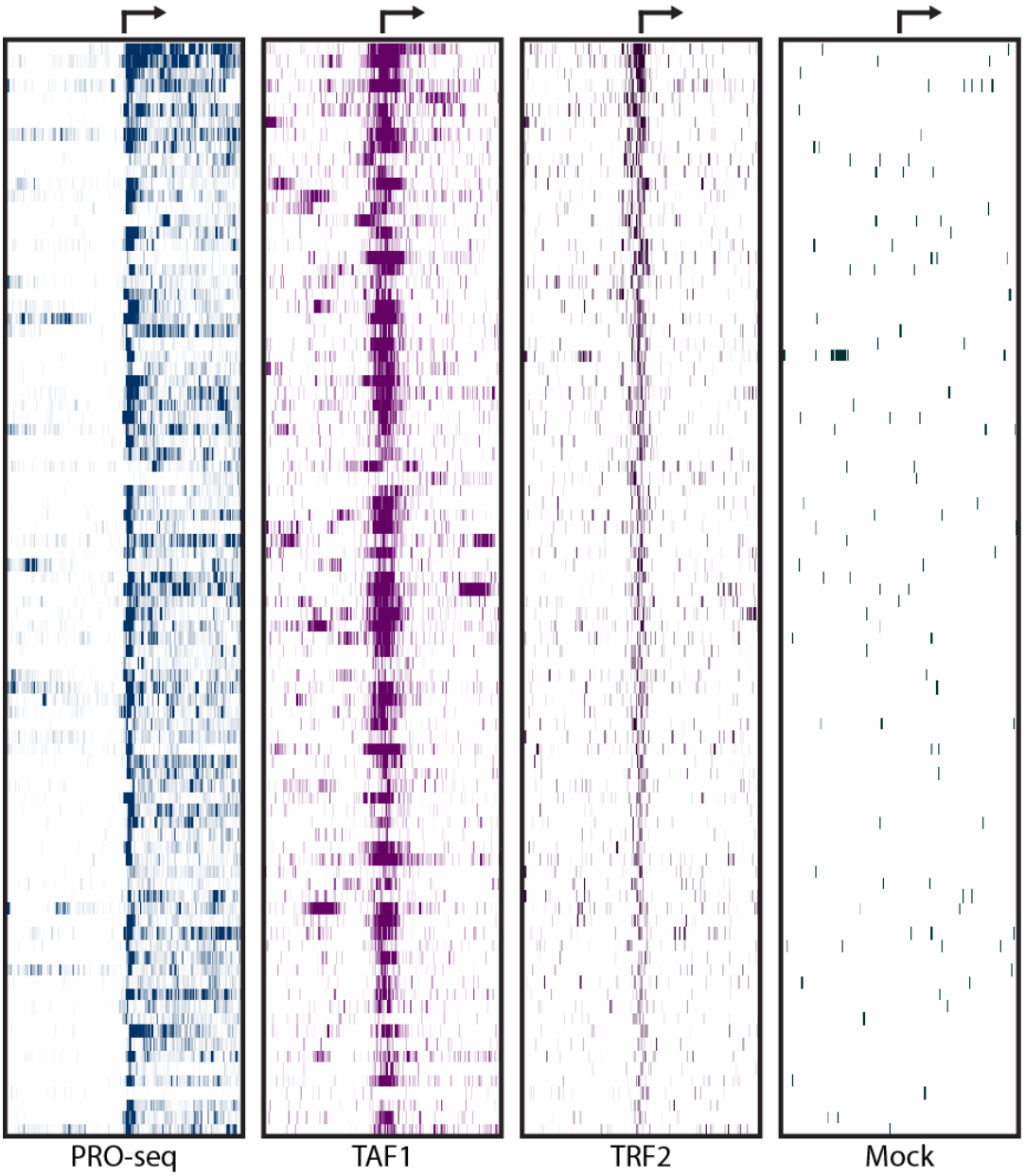

Figure 2-15. Many active TRF2-associated promoters are also bound by TAF1.

PRO-seq, TAF1, TRF2, and mock ChIP-exo reads mapped in a 2kb window relative to the TSS of active non-RP genes with a TRF2 peak within 200 bp of a TSS (n=89). Genes are sorted by TRF2 ChIP-exo read densities in a 400 bp window centered on the TSS. Top and bottom reads from the ChIP-exo analysis are displayed in the same color.

Colocalization of M1BP and TRF2 is largely restricted to the RP genes

Genomic analysis indicates that M1BP and TRF2 associate with many genes (Li & Gilmour, 2013; Wang *et al*, 2014). To determine if M1BP might function at other promoters by recruiting TRF2, I compared the distributions of M1BP and TRF2. M1BP and TRF2 show far less co-occupancy among non-RP genes (Figure 2-16). Thus, these two factors appear to have converged on the RP genes to help drive their robust and coordinated expression. However, the association of TRF2 with the ribosomal protein genes cannot be solely dependent on M1BP; otherwise, TRF2 would be present at other M1BP genes. A possible contributor to the specificity of TRF2 for M1BP-occupied RP genes is the TCT motif that is largely restricted to the RP gene promoters (Parry *et al*, 2010).

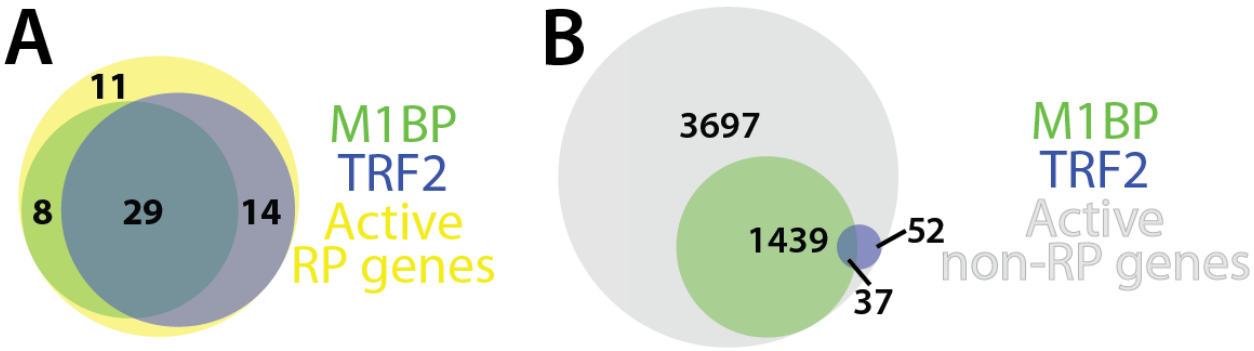

Figure 2-16. TRF2 and M1BP show greater overlap at RP gene promoters than other active promoters.

(A, B) Venn diagrams showing the overlap between M1BP and TRF2 peaks present at (A) all active RP gene promoters (n=62) or (B) all other active gene promoters (n=5225).

RP gene promoters act as enhancers of other RP genes

M1BP and TRF2, together with the TCT motif could play an essential role in coordinating expression of RP genes. Recently developed enhancer screens have offered new

information about how enhancers contribute to expression of different promoters or promoter classes. One screen, termed self-transcribing active regulatory region sequencing (STARR-seq) (Overview in Figure 2-17), was developed in order to identify novel enhancer sequences and determine core promoter-enhancer specificities (Arnold *et al*, 2013).

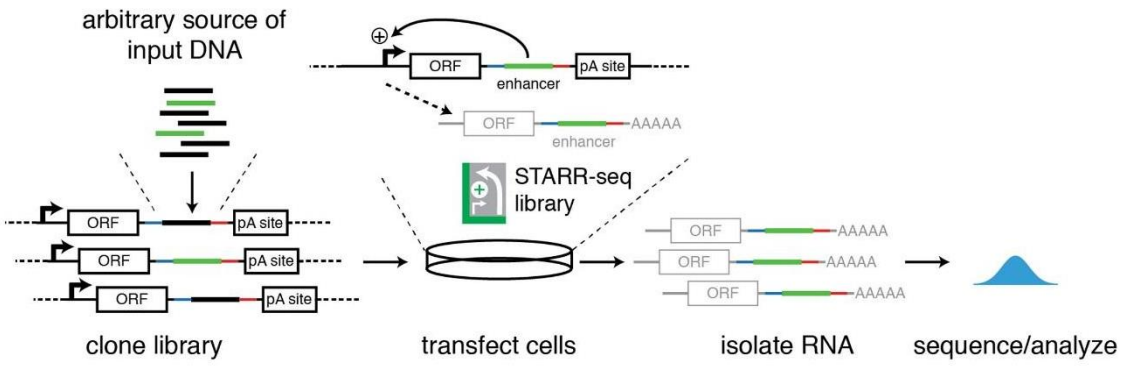

Figure 2-17. Overview of the STARR-seq method.
In STARR-seq, clonal libraries are generated by inserting arbitrary DNA sequences downstream of defined promoter and ORF sequences in the 3'UTR. The original report used sonicated genomic DNA from *Drosophila* as the insert. Following transfection into cultured cells, the reporter transcripts are isolated from total RNA and RNA-seq is performed. Because enhancers can act on upstream promoters, any inserted DNA that enhances transcription will be overrepresented in the RNA-seq output. The level of insert enrichment in the RNA-seq output correlates with the sequence's enhancer strength for the tested promoter. Reproduced from (Muerdter *et al*, 2015).

In this assay, plasmids are constructed that contain random genomic regions inserted downstream of known promoter sequences. After transfection, the genomic regions inserted downstream of the promoter that activate transcription are self-transcribed. The sequence and relative abundance of the transcripts containing the random genomic sequence are identified and measured using RNA-seq. Thus, in addition to identifying new enhancer sequences that activate a promoter sequence of interest, the assay also provides a measure of fold activation. While it is an artificial system, measurements of fold activation provide a meaningful measurement of enhancer strength in the assay. A recent analysis of published STARR-seq (Zabidi *et al*, 2015)

data concluded that housekeeping promoters themselves act as enhancers (Cubeñas-Potts *et al*, 2016). I initially analyzed the STARR-seq data to see if I could detect enhancers that act upon the RP genes but found that the RP gene promoters themselves function as enhancers (Figure 2-18). Thus, my analysis suggests that RP genes promoters could serve as enhancers of other RP genes. This potentially explains how RP genes are coordinately transcribed.

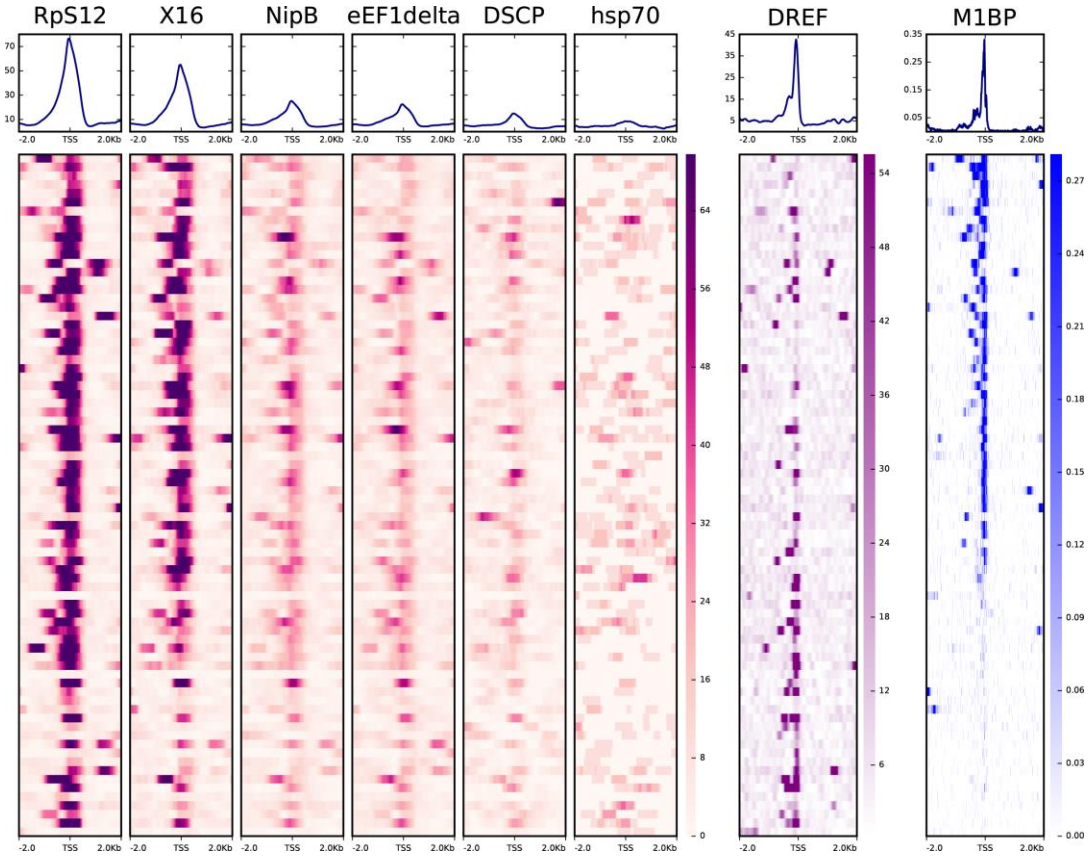

Figure 2-18. RP gene promoters act as enhancers of the RpS12 promoter.

STARR-seq (Zabidi *et al*, 2015), DREF ChIP-seq (Li *et al*, 2015), and M1BP ChIP-exo reads were mapped relative to RP gene TSSs in a 4 kb window centered on the TSS. The genes are ranked according to the M1BP ChIP-exo read total within the 4 kb window. Higher signal intensities reflect the ability of the inserted genomic region to enhance transcription from the promoter listed at the top of each panel. Composite plots for each data set are shown above the heatmaps. STARR-seq data measures the ability of segments of DNA to activate candidate core promoters. *RpS12*, *X16*, *NipB*, and *eEF1delta* are candidate housekeeping gene core promoters. Specifically, *X16* and *NipB* possess Motif 1 and Motif 6. *RpS12* and *eEF1delta* have both the DRE and a TCT motif. The *DSCP* is a developmentally regulated core promoter and *hsp70* is a candidate stress inducible core promoter, both of which have a canonical Initiator and TATA box. The scale bars on the right-hand side represent the normalized mean read values for each bin in the 4 kb genomic region around the TSS.

Discussion

Here, I show that M1BP activates transcription of RP genes in *Drosophila* and that it can do so by recruiting TRF2 to RP gene promoters in cells. These conclusions are based on my demonstration that M1BP is detected in the core promoter region of the majority of RP genes in cells and that mutation of Motif 1 diminished the level of expression from RP reporter genes. Additionally, I have demonstrated that M1BP activates transcription of RP gene promoters in nuclear extracts. Also, I show that M1BP recruits TRF2 to promoter DNA *in vitro* and that M1BP and TRF2 colocalize on the RP gene promoters in cells. M1BP, therefore, is the first sequence-specific DNA binding protein that has been directly shown to activate RP gene transcription in metazoans. DREF is possibly the only other protein, but it remains to be determined if it activates RP genes *in vitro*. Since these transcription factors associate with a broad spectrum of genes, loss of function assays in cells must be viewed with caution as it is difficult to distinguish between direct and indirect effects regardless of whether the protein can be detected at a particular gene. Consequently, the demonstration that depletion of M1BP inhibits RP gene transcription *in vitro* and that this inhibition can be reversed by adding back M1BP is of utmost importance in establishing that M1BP directly activates RP gene transcription.

Mechanisms by which TRF2 associates with promoters are not well understood. DREF was purified in a complex with TRF2 but no direct measurement of TRF2 recruitment to DNA by this complex was provided (Hochheimer *et al*, 2002). An uncharacterized TRF2 complex associates with promoters bearing the downstream promoter element (DPE) and canonical Initiator (Kedmi *et al*, 2014), but RP genes lack both of these DNA elements. Here I provide a direct mechanism that involves M1BP associating with its cognate binding site and interacting

directly with TRF2. Since there is little overlap between M1BP and TRF2 outside of RP gene promoters, it follows that additional cis-elements are required for TRF2's association with M1BP. I suspect that the TCT motif, along with M1BP and DREF, may be additional key contributors to TRF2's association with gene promoters.

The total number of promoter-associated TRF2 peaks that I observe (n=132) is considerably lower than observed previously (Wang *et al*, 2014). A couple of reasons could account for this discrepancy. First, the previous study used 2-4 hour embryos, whereas I used S2R+ cells. It is possible that TRF2 functions at a broader spectrum of developmentally regulated genes in the early embryo than in S2R+ cells. Additionally, the difference could be due to the increased signal to noise ratio afforded by ChIP-exo which results in more reliable peak detection.

Yeast RP gene transcription requires TFIID (Shen & Green, 1997; Kuras *et al*, 2000; Mencía *et al*, 2002). However, TFIID's role in transcription of RP genes in higher eukaryotes was unclear and called into question when the Kadonaga lab demonstrated that a TFIID complex containing TBP fails to protect promoters bearing a TCT motif from DNase I digestion (Parry *et al*, 2010). Since almost all RP genes have a TCT motif in *Drosophila*, it followed that TFIID may not associate with RP gene promoters. Thus, detection of TAF1, and by extension TFIID, on *Drosophila* RP gene promoters was unexpected, but significant. Moreover, previous analysis of the PCNA promoter showed that immunodepletion of TFIID with TAF1 antibody from a *Drosophila* transcription reaction did not inhibit transcription of a TRF2-dependent promoter (Hochheimer *et al*, 2002). Nevertheless, our ChIP-exo data provides evidence for M1BP being in close proximity to, and potentially interacting with, TRF2 and TAF1 on RP gene promoters. The ChIP-exo data showed M1BP contacts downstream of the TSS yet Motif 1 typically resides

upstream from the TSS. Since the ChIP-exo data for M1BP and TAF1 display overlapping peaks in the +30 to +50 region, I propose that M1BP is in contact with TAF1 and that the ChIP-exo signal for M1BP in this region is a consequence of M1BP crosslinking to TAF1 and TAF1 in turn crosslinking to the +30 to +50 region. In contrast, the ChIP exo signals for M1BP and TRF2 are shifted relative to each other by approximately 10 nucleotides suggesting that M1BP might position TRF2 on the DNA adjacent to M1BP. Though this data may seem incongruous with the finding that DNase I footprinting of purified TFIID on TCT promoters showed little to no protection, it should be noted that the TFIID purification scheme in this study involved the use of a TBP antibody (Parry *et al*, 2010). Thus, the population of TFIID used for the DNase I footprinting analysis contained stably-associated TBP. It would be interesting to test if a more heterogeneous mixture of TFIID would show stronger protection or greater reactivity at hypersensitive sites in a DNase I assay than the monotypic population used in the study. In future studies, it would be beneficial to interrogate whether RP genes and other genes with high downstream TAF1 signals in higher eukaryotes also correlate with higher rates of reinitiation. Higher rates of reinitiation could, at least in part, explain how high and steady rates of transcription are achieved at RP genes and other housekeeping genes.

Subsequent to publishing our Nucleic Acids Research article, the Buratowski lab published a study that agrees with our TAF1 results as they observe extensive TAF1 contacts downstream of the TSS in yeast. Interpreting data from both biochemical and genomic experiments, they conclude that reinitiation is one major function of the TAFs (Joo *et al*, 2017). The authors show that in yeast nuclear extracts, TAF contacts downstream of the TSS enable activator-independent reinitiation. Notably, they provide evidence that the relative abundance of TAF contacts downstream of the TSS varies between genes. Using TAF1 mutant yeast strains,

they observed that genes whose expression was most affected by the TAF1 mutant had a higher proportion of downstream TAF1 contacts relative to upstream contacts. They argue that these differences in downstream contacts confer specialized properties that ultimately affect the way different gene promoters respond to input from other transcription factors (Joo *et al*, 2017). In light of these findings, it would be insightful to compare the variability in TAF1 downstream contacts between different gene classes in *Drosophila*. If RP genes show particularly high levels of downstream contacts, this may provide a rationale to explain how RP genes maintain high and consistent levels of transcription. Understanding whether M1BP or TRF2 play a role in enabling or altering TAF's downstream contacts might reveal additional principles that govern the expression of RP genes or other gene regulatory networks.

A unique feature of the RP gene promoters in *Drosophila* and humans is the presence of the TCT motif located at the transcription start site (Parry *et al*, 2010). What recognizes this motif is currently not known. Since TAF1 is known to recognize the canonical Initiator element (Verrijzer *et al*, 1995; Wu *et al*, 2001; Parry *et al*, 2010), its presence at RP gene promoters raises the possibility that TAF1 also recognizes the TCT motif. DNAse I footprinting analysis of TFIID binding to RP gene promoters indicated that binding was extremely weak. However, close inspection of the DNase I cutting patterns in the absence and presence of TFIID reveals the appearance of weak hypersensitive cut sites near the TCT motif (Parry *et al*, 2010). One possibility is that M1BP together with TRF2 enhance the affinity of TAF1 for the RP gene TCT motif. If reinitiation were taking place through TAF1's downstream interactions as discussed above (Joo *et al*, 2017), it could be that the combination of M1BP and TRF2 functions to maintain TFIID's association with the promoter for additional rounds of transcription.

I propose that M1BP functions as a hub to recruit TRF2 (Figure 2-19). Additionally, since the only known TAF1-containing complex in metazoans is TFIID, I propose that TFIID still binds to RP gene promoters along with TRF2. One possibility is that TRF2 displaces TBP at RP gene promoters. A recent model of TFIID bound to promoter DNA indicates that TFIIA is involved in connecting TBP to TAF1 (Louder *et al*, 2016). Since TRF2 associates with TFIIA (Rabenstein *et al*, 1999), displacement of TBP from TAF1 by TRF2 is tenable.

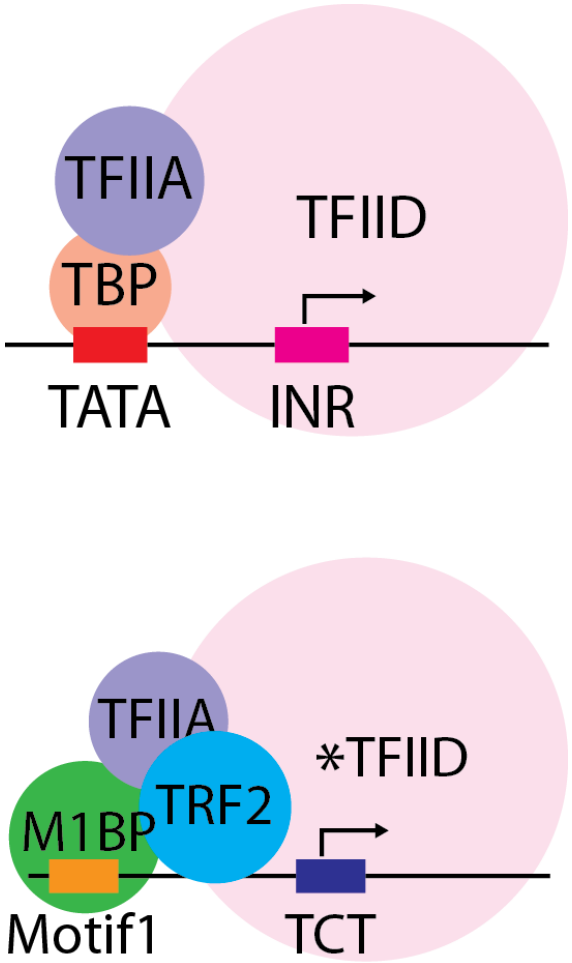

Figure 2-19. Model depicting M1BP's recruitment of TRF2 at RP gene promoters.
At TATA-containing promoters, TBP-bound TFIID engages with promoter sequences both up and downstream of the TSS. At the majority of RP gene promoters, which lack both a TATA box and Initiator sequence, M1BP and TRF2 bind the core promoter upstream of the TSS. The asterisk (*) denotes a non-canonical TFIID complex proposed to have TRF2 substituting for TBP.

Our analysis of STARR-seq data indicates that RP gene promoters can act as enhancers and that they are selective in activating housekeeping gene core promoters and not core promoters of developmental and stress-responsive genes. The RP gene promoters more strongly activated the candidate RP gene promoter over all the other tested candidates. This selectivity could establish a network in which active RP genes and other housekeeping genes act reciprocally to activate each other. In addition, the selectivity of the enhancer activity of these RP promoters would prevent them from inadvertently activating nearby developmentally regulated genes.

Materials and Methods

RNAi Knockdown in S2R+ Cells Followed by Chromatin Immunoprecipitation

dsRNA was generated by *in vitro* transcription with T7 polymerase on templates flanked by T7 promoter sequences. After the *in vitro* transcription reactions, the samples were centrifuged at 16000 x *g* for 2 minutes to pellet the magnesium pyrophosphate. The supernatant was transferred to a new tube and incubated at 37℃ for 30 minutes with 1 unit RQ1 DNase per 20 µl of reaction mix. Finally, after adding 50 µl of 7.5 M LiCl and 50 mM EDTA per 20 µl reaction, the samples were incubated on dry ice for 10 minutes then spun at 4℃ for 20 minutes at 16000 x *g* to pellet the RNA. The precipitation with LiCl specifically pellets RNA while leaving free nucleotide and DNA in the supernatant thus enabling the accurate determination of RNA concentration by spectrophotometry. For the RNAi, 10^6 cells were plated in a 10 cm culture dish in 6 ml of serum-free M3+BPYE media (*Drosophila* Genome Resource Center) and were treated with 180 µg of the indicated dsRNA for 1 hour after which the total media volume was brought up to 12 ml with a final fetal bovine serum concentration of 10%. Following this treatment, cells

were allowed to incubate for the time specified in the figure legends. Following completion of the knockdown, cells were crosslinked with formaldehyde, chromatin was prepared and immunoprecipitations were performed as previously described (Li & Gilmour, 2013). Percent recovery at designated genomic locations was determined by qPCR. Primers for dsRNA generation and qPCR are listed in (Baumann & Gilmour, 2017). qPCR reactions were assembled using Bioline SensiMix SYBR Hi-ROX (QT605-20) master mix with reaction conditions matching the manufacturer's recommendations. qPCR and analysis was performed using ABI StepOnePlus system. Reactions were heated to 95℃ for 10 minutes, then underwent 40 cycles of 95℃ for 15 seconds and 60℃ for 60 seconds with data acquisition taking place during the 60℃ step.

Western Blots

Formaldehyde crosslinked chromatin lysates from RNAi-treated cells were heated to 75°C for 10 min in SDS-PAGE sample buffer (50 mM Tris-Cl, pH 6.8, 2% SDS, 10% glycerol, and 100 mM dithiothreitol). Crosslinks were then reversed overnight at 65°C. Samples were electrophoresed on a 10% polyacrylamide gel, blotted to nitrocellulose, and probed with antibodies against the indicated factors.

Nuclear Extracts and Immunodepletion of M1BP

Nuclear extracts were prepared from 0-12 hour Oregon R embryos as previously described (Biggin & Tjian, 1988). Immunodepletion of nuclear extracts was performed as previously described (Wang *et al*, 2014).

In vitro Transcription Reactions and Primer Extension Assay

In vitro transcription reactions and the primer extension assay were performed essentially as previously described (Wampler & Kadonaga, 1992). Twenty-five microliter reactions containing 125 µg nuclear extract, 32.5 mM HEPES (pH 7.6, K^+), 20 mM KCl, 6.25 mM $MgCl_2$, 0.05 mM EDTA, 5% glycerol, 1 mM DTT, 1% PEG (Sigma product number P2263, MW:15-20kD), 10 µg/ml α-amanitin (where indicated), 2 units Promega Recombinant RNasin, 20 ng/µl plasmid template, and 4.8 ng/µl recombinant M1BP (where indicated) were incubated at 24°C for 30 minutes. Twenty-five microliter *Hsp70Bc* reactions shown in Figure 2-5 contained 125 µg nuclear extract, 32.5 mM HEPES (pH 7.6, K^+), 50 mM KCl, 6.25 mM $MgCl_2$, 0.05 mM EDTA, 5% glycerol, 1 mM DTT, 10 µg/ml α-amanitin (where indicated), 2 units Promega Recombinant RNasin, and 10 ng/µl plasmid template were incubated at 24°C for 20 minutes. After PIC formation, NTPs were added to a concentration of 0.5 mM and transcription occurred for 20 minutes (or 10 minutes for Hsp70Bc reactions) at 24°C. Reactions were stopped by addition of 0.8% SDS, 16 mM EDTA, 160 mM NaCl, 0.2 mg/mL Torula yeast RNA, and 0.08 mg/mL Proteinase K and incubated for at least 5 minutes at room temperature. Samples were extracted with 25:24:1 phenol:chloroform:isoamyl alcohol followed by chloroform alone. Primer extension assays were then performed as previously described (Carey *et al*, 2013) and analyzed on a 10% sequencing gel containing 8M urea.

Expression and Purification of M1BP

Rosetta (DE3) pLysS competent cells (EMD Millipore) were transformed with a previously reported M1BP expression vector (Li & Gilmour, 2013). 0.5 liters of transformed cells were grown in LB media at 37°C to an OD_{600} of 0.4. IPTG was added to a final

concentration of 0.25 mM IPTG and incubated overnight at 11°C. Cells were collected, resuspended, and flash-frozen with liquid nitrogen in 25 mL Lysis buffer (25 mM HEPES (pH 7.6, K+), 300 mM NaCl, 10 mM imidazole, 10 mM beta-mercaptoethanol, 0.1% Triton X-100, 10 uM $ZnCl_2$ with Protease inhibitors and 1 mg/mL lysozyme). Cells were thawed, incubated on ice for 15 minutes, sonicated and finally cleared by centrifugation at 20,000 x g. One-half milliliter of TALON (Clontech) resin previously equilibrated with lysis buffer was added to the cleared lysate and incubated with end over end mixing at 4°C for 1 hour. Resin was collected and washed with lysis buffer for 15 minutes at 4°C, then poured into a column and washed with an additional 50 mL lysis buffer. Samples were eluted from the column in Elution buffer (25 mM HEPES (pH 7.6, K^+), 150 mM NaCl, 250 mM imidazole, 10 mM beta-mercaptoethanol, 0.1% Triton X-100, 10% glycerol, and 10 uM $ZnCl_2$ plus protease inhibitors). The samples were then dialyzed at 4°C overnight in Elution buffer lacking imidazole. Samples were centrifuged at 16000 x g for 10' at 4°C and the supernatant was analyzed by SDS-PAGE and used in the indicated experiments.

Synthesis of radiolabeled TRF2

A DNA fragment encoding TRF2 was amplified from S2R+ cDNA (See (Baumann & Gilmour, 2017) for primer sequences). 100 ng of the T7-flanked TRF2 coding region PCR product was added to the TnT Quick Coupled Transcription/Translation System (Promega) rabbit reticulocyte lysates and the reactions were carried out per the manufacturer's protocol. A parallel reaction was done without adding the PCR template to produce a negative control for the immobilized template pulldown experiments.

Immobilized template pulldown experiments

Oligonucleotides corresponding to -52 to +8 of RpL30 or -36 to +14 of RpLP1 were annealed in 1 mM EDTA, 10 mM Tris-Cl pH 8.0, 100 mM NaCl. The annealed oligonucleotides were purified from a polyacrylamide gel to ensure only hybridized oligonucleotides were used in the pulldowns. One oligonucleotide was biotinylated so that the template could be bound to Streptavidin Dynabeads as per the manufacturer's protocol. Template bound beads were equilibrated in binding buffer consisting of 25 mM HEPES (pH 7.6, K^+), 150 mM NaCl, 10 mM beta-mercaptoethanol, 0.2 mM PMSF, 0.1% Triton X-100, 10% glycerol, and 10 µM $ZnCl_2$. 3.5 µg of recombinant His-M1BP or control buffer and 10 µl ^{35}S-TRF2 or control (No PCR template) TnT reactions were added as indicated. All components were incubated for 15 minutes at room temperature, then washed extensively with binding buffer. Beads were boiled in Laemmli sample buffer and the samples were analyzed by 10% SDS-PAGE. ^{35}S-labelled TRF2 was detected with a phosphorimager.

Maltose-binding protein expression and pulldown experiments

BL21DE3 *E. coli* cells expressing Maltose binding protein (Mal) alone or Mal fused to M1BP were grown to an OD600 of 0.4-0.5, IPTG was added and proteins were expressed in LB media overnight at 11°C. Cells were lysed then cleared by spinning at 100,000xg for 30 minutes. Mal was bound to amylose resin and washed with a buffer consisting of 50 mM HEPES, pH 7.6, 500 mM NaCl, 10% glycerol, 1% NP-40 and 1 mM DTT. Mal-M1BP was bound to amylose resin and washed with a buffer consisting of 25 mM HEPES, pH 7.6, 1 M NaCl, 200 mM KCl, 10% glycerol, 0.1% NP-40, 10 µM $ZnCl_2$, and 1 mM DTT. 10 µl of each type of protein-bound resin was equilibrated with several washes of pulldown buffer (25 mM HEPES, pH 7.6, 0.1 mM

EDTA, 12.5 mM $MgCl_2$, 10% Glycerol, 100 mM NaCl, 0.1% CHAPS, and 0.1% NP-40). Following equilibration, 100 µl of pulldown buffer, 2 µl of ^{35}S-TRF2, and 20 µg BSA was added to each sample. Mixtures were incubated for 15 minutes at room temperature while rotating. Samples were washed for 5 minutes with end-over-end rotation in 100 µl of pulldown buffer. A total of 5 washes were performed for each sample. Beads were boiled in Laemmli sample buffer and the samples were analyzed by 10% SDS-PAGE. ^{35}S-labelled TRF2 was detected with a phosphorimager.

ChIP-Exo

ChIP-Exo was performed essentially as described in (Rhee & Pugh, 2012) with minor modifications. Libraries were quantified by qPCR and sequenced on an Illumina NextSeq 500. Basecalls were performed using Bcl2FastQ version 2.16.0. Sequenced reads were masked for low-quality sequence, then mapped to the *Drosophila melanogaster* dm3 whole genome using BWA mem (versions 0.7.9a, 0.7.12) with the default parameters. Heatmaps were generated with HOMER bioinformatics software (Heinz *et al*, 2010) and java Treeview (Saldanha, 2004). Tables for composite plots were generated with HOMER and plots were visualized using R (Venables & Smith, 2010). Position weight matrices (PWM) for Motif 1 and DRE were obtained by performing a MEME-ChIP search of 200 bp regions centered around M1BP ChIP-exo or DREF ChIP-seq peak centers as determined by GEM using default settings (Guo *et al*, 2012). The Motif 6 PWM was taken from the MEME-ChIP search of 200 bp regions centered on M1BP ChIP-exo peak centers. Motif 6 was the second most enriched motif in that search (after Motif 1).The PWMs were fed into the FIMO tool (Grant *et al*, 2011) to identify motif locations genome-wide with a p-value cutoff <1E-04.

Bioinformatics

My list of active genes was derived from the active gene list provided in (Kwak *et al*, 2013). RP genes were selected and isolated from the list using their flybase annotation symbol. There are a total of 87 RP genes. Eight RP genes have duplicate isoforms and I eliminated one isoform of each duplicate if it lacked a TCT motif (Parry et al, 2010) or a TRF2 ChIP-seq peak bringing the total number of genes analyzed to 79. *RpL15* resides on Chr3LHet and, since our ChIP-exo data was not mapped to those regions, it was removed from the ChIP-exo heatmaps and composite plots thus bringing the final RP gene number to 78. PRO-seq bedgraph files were obtained from (Duarte *et al*, 2016). Read pileups for heatmaps were performed with the HOMER bioinformatics tool using the annotatePeaks.pl script (Heinz *et al*, 2010). Composite plots were generated in R. Venn Diagrams were generated with Biovenn (Hulsen *et al*, 2008). Heatmaps were generated with Java Treeview (Saldanha, 2004). STARR-seq heatmaps were generated using deepTools (Ramírez *et al*, 2016).

Peak Calling and ChIP-seq Analysis

DREF ChIP-seq (GSM1535985) (Li *et al*, 2015), and input control experiment reads (Li *et al*, 2015) were downloaded from http://www.ebi.ac.uk/ena in fastq format. DREF and Input reads were mapped to the dm3 genome in Galaxy with the BWA read aligner using default parameters. The GEM peak caller (Guo *et al*, 2012) was used to call peaks from the experiment and control bed files. Genes having transcription start sites within 200 bp of peak centers were designated M1BP-, TRF2- or DREF-associated.

Antibodies

The M1BP antibody was initially described and characterized in (Li & Gilmour, 2013). The preimmune sera comes from the same rabbit used to produce the M1BP antibody prior to injection with purified M1BP. The TRF2 antibody was described in (Wang *et al*, 2014). The TAF1 antibody was described in (Maile *et al*, 2004).

Chapter 3: GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, is a transcriptional co-activator

Most of the content in this chapter has been published in Molecular and Cellular Biology (MCB), which allows publication of this material in my dissertation.

Abstract

The core promoter of protein-encoding genes plays a central role in regulating transcription. M1BP is a transcriptional activator that associates with a core promoter element known as Motif 1 that resides at thousands of genes in *Drosophila*. To gain insight into how M1BP functions, I employed immobilized template pulldowns and identified an interacting protein called GFZF. GFZF, which is notable for its unusual combination of FLYWCH zinc finger domains fused to a glutathione S-transferase (GST) homology domain, had been previously identified in genetic screens for factors involved in maintenance of hybrid inviability, the G2-M DNA damage checkpoint, and RAS/MAPK signaling but its contribution to these processes was unknown. Here, using a combination of biochemistry, immunofluorescence microscopy, and high-resolution genome-wide approaches, I show that GFZF resides in the nucleus and binds the promoter region of over one thousand actively transcribed genes. Reporter assays and chromatin immunoprecipitation (ChIP) experiments following RNAi-mediated depletion of GFZF reveal that GFZF functions as a transcriptional co-activator. In addition, I demonstrate that GFZF is a glutathione S-transferase with a physiologically-relevant affinity for glutathione. Altogether, I conclude that GFZF is the first reported transcriptional co-activator with intrinsic GST activity, and its identification as a transcriptional co-activator provides an explanation for its role in numerous biological processes. I also speculate about the exciting potential implications for a GST protein acting at this fundamental step in gene regulation.

Introduction

Regulation of RNA Pol II-transcribed genes is one of the primary mechanisms by which cells coordinate the processes required for survival, proliferation, and development. The core promoter, defined as the 80-100 base pair region centered on the transcription start site, is the hub of transcription regulation (Danino *et al*, 2015). Transcription initiates when general transcription factors (GTFs) bind elements within the core promoter region forming a complex consisting of Pol II and other highly conserved Pol II-associated transcription factors (Thomas & Chiang, 2008). In recent years, our understanding has advanced from a model where the core promoter and the GTFs act as static integrators of signals from sequence specific transcription factors that bind enhancer regions and modulate transcription levels, to one where the core promoter and its machinery is a more dynamic assembly with different enhancer specificities (Zabidi & Stark, 2016; Ohler & Wassarman, 2010) and intrinsic regulatory properties (Danino *et al*, 2015).

One particular core promoter element has emerged that provides a clear contrast to the models arising from canonical promoters. The element named Motif 1 (FitzGerald *et al*, 2006; Ohler *et al*, 2002a) is present in the promoter region of thousands of genes in *Drosophila*. We identified and characterized a factor that binds this conserved element and named it M1BP (Li & Gilmour, 2013). M1BP is enriched at housekeeping gene promoters and M1BP-bound genes tend to have moderate-to-high levels of paused Pol II, are constitutively expressed, and show little spatiotemporal fluctuation in transcription levels (Li & Gilmour, 2013). Additionally, Motif 1, and by extension, M1BP-bound promoters tend to lack many of the elements once thought to be essential for initiation such as the TATA box and Initiator (Ohler *et al*, 2002a), so how initiation

occurs at these promoters remains a mystery. Thus, the study of M1BP promoters might provide insights into previously unknown mechanisms of transcription initiation and activation.

Here I characterize a factor called GFZF that M1BP recruits to promoters. GFZF turns out to be a novel transcriptional coactivator that has glutathione S-transferase (GST) activity. GFZF has been identified in many genetic screens since its initial characterization (Dai *et al*, 2004). These screens have implicated GFZF in a wide variety of processes including regulation of the cell cycle (Ambrus *et al*, 2009), DNA damage checkpoints during the transition from G2 to M phase (Kondo & Perrimon, 2011), transcriptional and splicing control of RAS/MAPK signaling (Ashton-Beaucage *et al*, 2014), response to oxidative stress (Li *et al*, 2008), three-dimensional organization of polycomb complexes (Gonzalez *et al*, 2014), and speciation (Phadnis *et al*, 2015) among other processes (Provost *et al*, 2006; Barth *et al*, 2014; Ranson *et al*, 2001). Despite its involvement in these critical cellular processes, little is known about the mechanism by which it carries out these seemingly disparate functions. Early work reported that GFZF resides in the cytoplasm (Dai *et al*, 2004). Here, I present data supporting a parsimonious conclusion that GFZF is a transcription factor required for expression of the many factors that carry out the functions described in the above screens.

Historically, GSTs have been studied for their role in cellular detoxification (Salinas & Wong, 1999). However, there are notable examples of GSTs performing additional cellular functions which include the regulation of signal transduction (Adler *et al*, 1999), inhibition of apoptosis (Kamada *et al*, 2004), and the response to oxidative stress (Hayes *et al*, 2005). Thus, it seems that GSTs play a critical, and perhaps underappreciated, role in cellular function and homeostasis. Our unprecedented finding of a transcription factor with GST activity raises the

possibility of additional layers of complexity in the already complex process of metazoan transcriptional regulation.

Results and Discussion

M1BP associates with GFZF

In order to understand the function of the core promoter, it is essential to know what factors associate with it. To identify factors that associate with Motif 1-containing core promoters, I immobilized a DNA sequence that corresponds to the promoter of a mitochondrial ribosomal protein gene subunit (*mRpS30*). This promoter was chosen because it contains a strong consensus Motif 1 and our lab previously observed that *mRpS30* transcript levels decreased significantly upon M1BP depletion (Li & Gilmour, 2013). Additionally, another study used this promoter to drive GFP expression in flies and observed a loss of GFP expression when Motif 1 was removed (Li *et al*, 2010). To perform my immobilized template pulldown experiment, I incubated the *mRpS30* promoter template with *Drosophila* embryo nuclear extracts. As a negative control, I also incubated extracts with a promoter sequence containing a mutant version of Motif 1 that no longer binds M1BP. Bound proteins were then detected by SDS-PAGE and identified by mass spectrometry. Comparison of the factors bound to these two promoters identified several factors including Putzig, GFZF, and CG3995 (Figure 3-1). The identification of Putzig is consistent with previous findings that Putzig exists in a complex with TRF2 (Hochheimer *et al*, 2002) and that TRF2 interacts with M1BP (Baumann & Gilmour, 2017). The function of CG3995 is unknown. However, CG3995 is predicted to bind nucleic acids because it possesses both a BED-type zinc finger and Myb/Sant-like binding domain (Finn *et al*, 2016). I discuss the potential implications for M1BP's interaction with CG3995 and Putzig in Chapter 4.

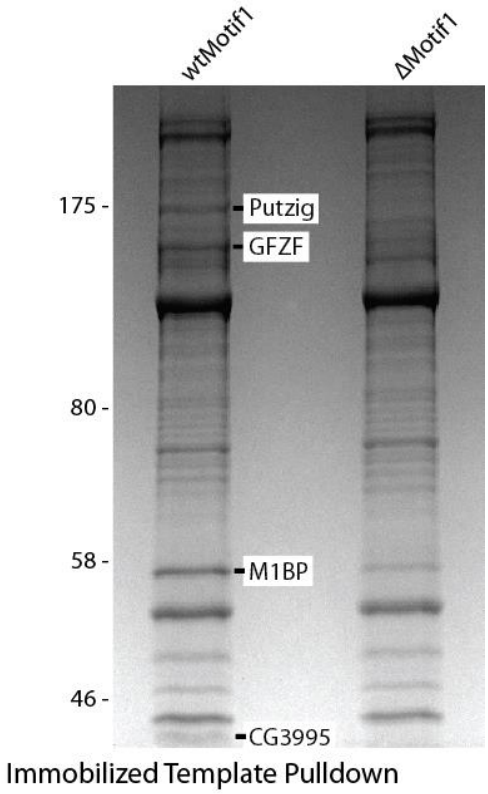

Figure 3-1. Putzig, GFZF, CG3995, and M1BP associate with wt Motif 1 immobilized templates. Silver-stained SDS-PAGE analysis of eluates from immobilized template pulldowns using the *mRpS30* promoter DNA with either the wild-type Motif 1 sequence (wt Motif 1) or 3 mutated nucleotides (Δ Motif 1). wt and Δ Motif 1 sequences for the *mRpS30* promoter are shown below the gel. Nuclear extracts were incubated with the immobilized templates and bound proteins were recovered and analyzed by 8% SDS-PAGE then identified by mass spec.

To determine if GFZF interacts directly with M1BP, I performed the immobilized template pull-down analysis with recombinant proteins expressed in and purified from *E. coli*. The data are consistent with my results from extracts and show, as expected, neither of the two factors associated with the mutated Motif 1 template (Figure 3-2, see lanes 3, 5, 7). In contrast, M1BP was able to associate with the wild-type Motif 1 template regardless of whether GFZF

was present in the reaction (Figure 3-2, cf. lanes 4 and 8). Conversely, GFZF was only able to associate with the wild-type Motif 1 template in the presence of M1BP (Figure 3-2, cf. lanes 6 and 8). Thus, direct interaction between M1BP and GFZF is likely to be involved in recruiting GFZF to promoter DNA.

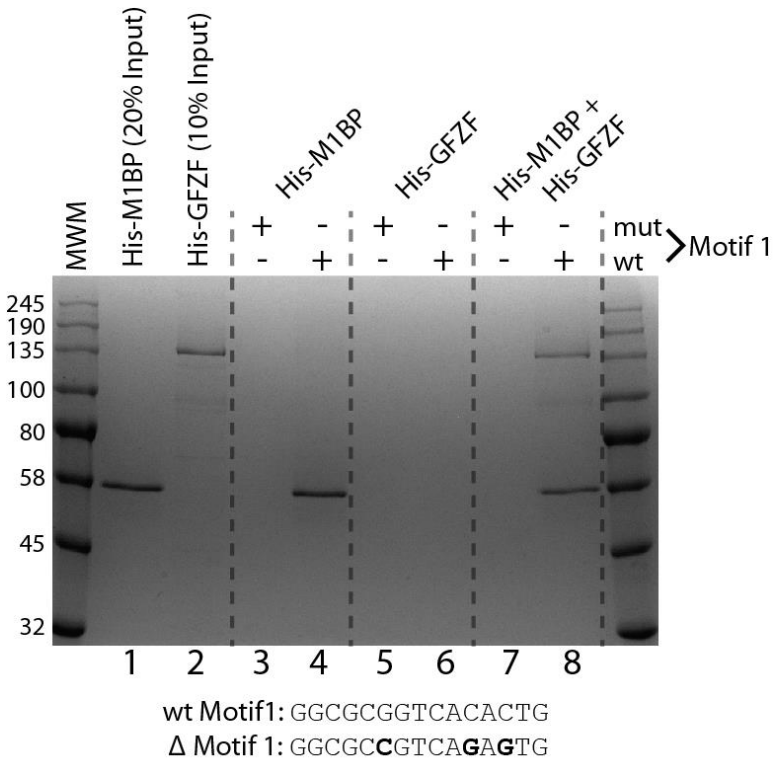

Figure 3-2. GFZF is a nuclear protein recruited to promoter DNA by M1BP.
Coomassie-stained SDS-PAGE analysis of immobilized template pulldown experiments using the *RpLP1* promoter sequence performed with recombinant, purified His-M1BP and His-GFZF. wt and Δ Motif 1 sequences for the *RpLP1* promoter are shown below the gel.

To determine if M1BP and GFZF interact in the absence of a DNA template, I performed pulldowns with purified maltose binding protein (Mal) fusions. Using either the alpha fragment of lacZ as a control or full-length M1BP fused to Mal, I determined that GFZF interacts specifically with M1BP (Figure 3-3, cf. lanes 3 and 5). Notably, the immobilized template pulldown (Figure 3-2) showed a roughly stoichiometric recovery of both GFZF and M1BP;

whereas, in the case of the Mal fusion pulldowns, GFZF is recovered sub-stoichiometrically. This suggests that GFZF may have a greater propensity to bind M1BP in a DNA-templated context.

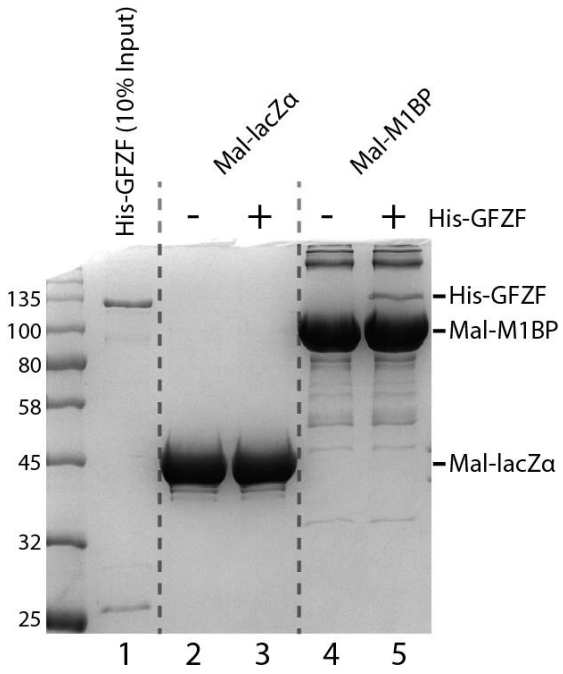

Figure 3-3. M1BP and GFZF interact in solution.
Coomassie-stained, SDS-PAGE analysis of Mal fusion pulldown experiments performed with recombinant, purified Mal-lacZα or Mal-M1BP fusion proteins immobilized on amylose beads and His-GFZF.

GFZF associates with chromosomes.

While GFZF was originally reported to be primarily a cytoplasmic protein (Dai *et al*, 2004), the results of my immobilized template pulldown experiments indicated that GFZF might associate with chromosomes. To test this, we used immunofluorescence microscopy with GFZF antibody to detect GFZF on polytene chromosomes. Antibody against GFZF localized it to distinct bands broadly distributed across each chromosome (Figure 3-4).

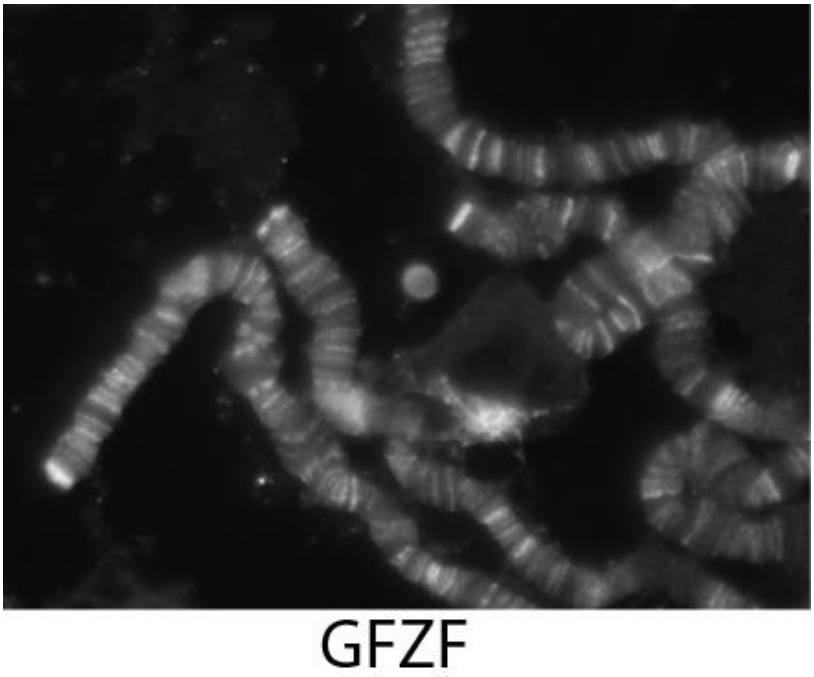

Figure 3-4. GFZF is associated with polytene chromosomes.
Polytene chromosomes from third instar larvae were spread and stained with antibody against GFZF. Polytene chromosome spreads were prepared and stained by David Gilmour.

Since my pull-down analysis indicated that M1BP and GFZF associate with each other, I compared their distributions on chromosomes. A comparison of M1BP and GFZF staining patterns on different polytene chromosome spreads revealed very similar staining patterns (Figure 3-5). Since both M1BP and GFZF antisera were prepared in rabbits, I could not detect both proteins at the same time on the same specimens. To circumvent this problem, I constructed a transgenic fly line that expresses FLAG-tagged M1BP (Figure 3-6) and localized the two proteins with a mouse monoclonal antibody targeting the FLAG epitope on M1BP and rabbit antibody targeting GFZF. This revealed significant overlap in staining for the two factors (Figure 3-7), suggesting that M1BP and GFZF bind the same genomic regions.

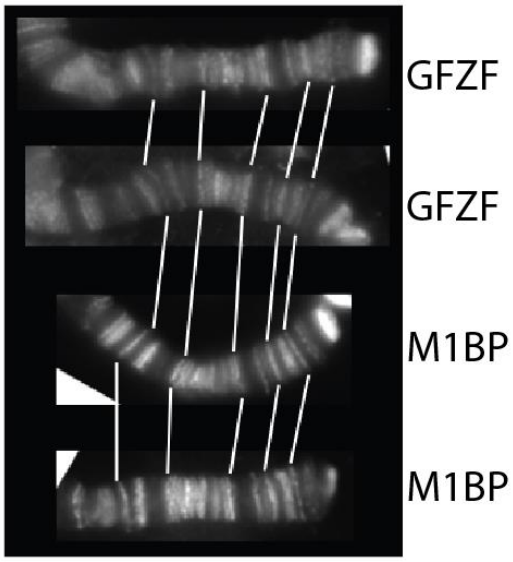

Figure 3-5. GFZF and M1BP have similar staining patterns on polytene chromosomes.
Comparison of anti-M1BP and anti-GFZF staining patterns at the end of chromosome 2R on separately stained samples. White lines guide an alignment of the patterns. Polytene chromosome spreads were prepared and stained by David Gilmour.

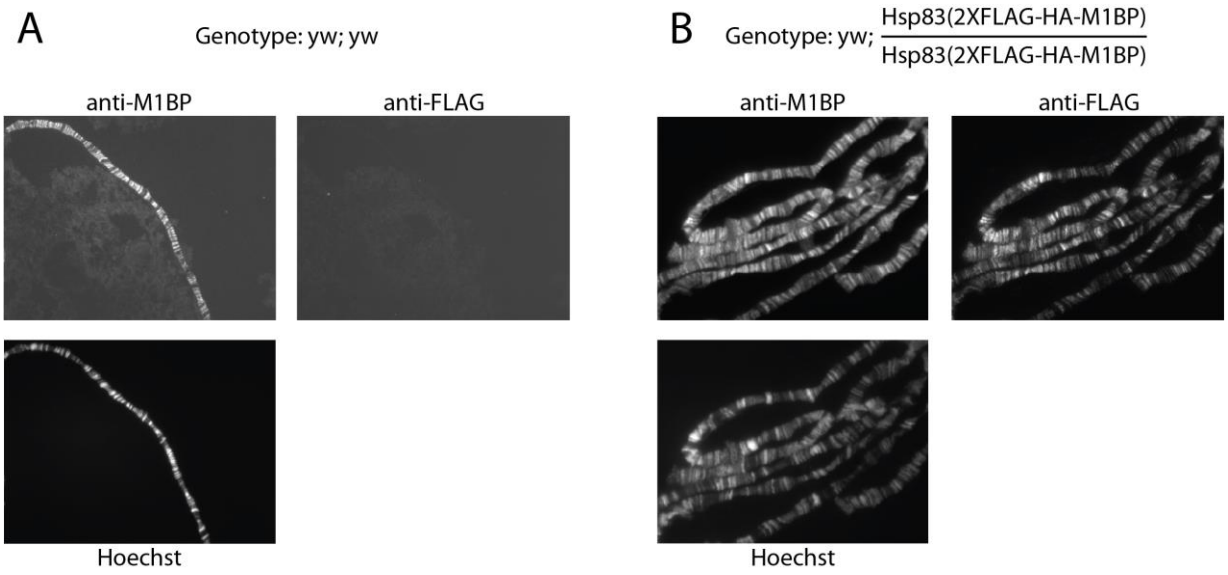

Figure 3-6. Detection of ectopically expressed FLAG-tagged M1BP on polytene chromosomes.
Polytene chromosome squashes from (A) yellow-white third instar larvae or (B) larvae expressing M1BP with N-terminal double FLAG and single HA tags under control of the *Hsp83* Promoter were stained with antibodies against M1BP, the FLAG epitope, and with Hoechst dye. Polytene chromosome spreads were prepared by David Gilmour.

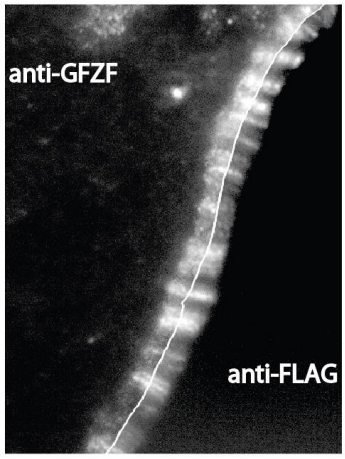

Figure 3-7. M1BP and GFZF show co-localized staining on polytene chromosome spreads. Polytene chromosomes from flies expressing 2XFLAG-HA-M1BP were spread and stained with mouse antibody against the FLAG epitope and rabbit antibody against GFZF. For ease of comparison, the images were split and aligned so that GFZF is to the left of the thin white line and FLAG-M1BP is to the right. Polytene chromosome spreads were prepared by David Gilmour.

GFZF colocalizes with M1BP at many promoters.

To gain further insights into GFZF's role, I mapped the distribution of GFZF on the genome using ChIP-exo. ChIP-exo reads from GFZF using only formaldehyde resulted in enrichment in promoter regions, but a low signal-to-noise ratio. I reasoned that the low signals were a result of either poor antibody performance or GFZF indirectly associating with chromatin through M1BP's interaction with promoters. Thus, in an attempt to address the second issue and potentially increase the signal-to-noise for the GFZF samples, I prepared the chromatin using two additional crosslinkers (described in the materials and methods section). This, in theory, should increase the likelihood of capturing M1BP-GFZF interactions and, if GFZF associates with chromatin exclusively through M1BP, should ultimately amplify the GFZF ChIP-exo signals. Using DeepTools2 (Ramírez *et al*, 2016), I compared the ChIP-exo data from this additionally cross-linked material with our formaldehyde-only material and found a good

correlation between the datasets (Figure 3-8) (Pearson R^2=0.88, summed reads surrounding gene promoters). I also noticed increased signal intensities using the chromatin with extra cross-linkers. Thus, I used the data from this chromatin preparation for our GFZF analysis and found that GFZF was present on over 1000 promoters in proliferating *Drosophila* S2R+ cells (Figure 3-9).

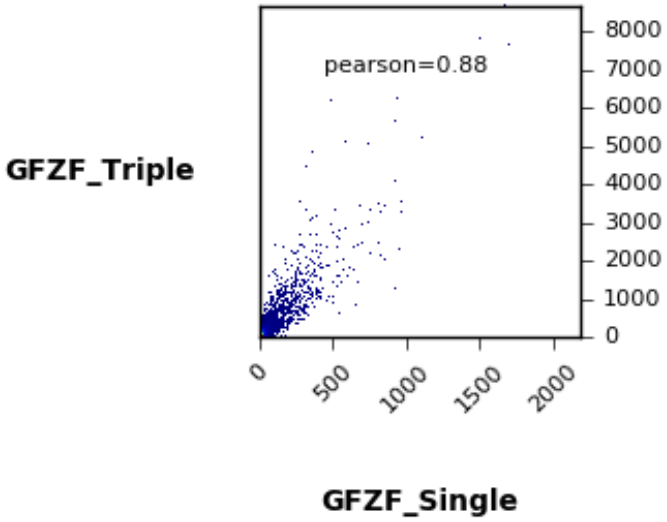

Figure 3-8. Single- and triple-crosslinked GFZF ChIP-exo datasets correlate well.
Pairwise comparison of GFZF ChIP-exo reads from single- and triple-crosslinked chromatin preparations summed in a 500 bp windows centered on 5339 active gene TSSs. These 5339 genes lack a neighboring gene TSS within 500 bp. The strength of the normalized signal intensities between the datasets differ, but the Pearson correlation is strong (R^2=0.88).

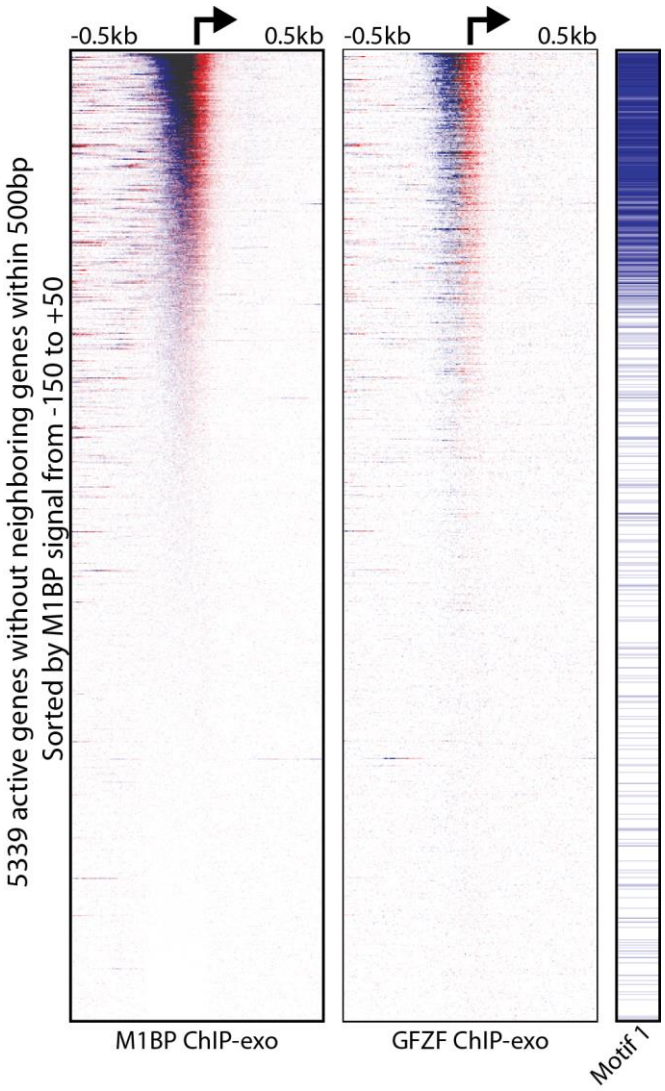

Figure 3-9. M1BP and GFZF co-localize at promoters with Motif 1.

Heatmaps showing M1BP (left panel) or GFZF (center panel) ChIP-exo reads mapped in a 1000 bp window centered on the transcription start site (TSS) of 5339 active genes lacking a neighboring gene TSS within 500 bp. ChIP-exo reads for the forward and reverse strands have been separated and displayed in blue and red, respectively. The presence of Motif 1 within 100 bp of the TSS is indicated by blue lines in the right panel.

A composite plot shows that the ChIP-exo signal for GFZF largely overlaps with M1BP and is concentrated in a 100 bp region just upstream from the transcription start site (Figure 3-10). Additionally after calling peaks for both factors, I identified genes that have a M1BP or

GFZF peak within 100 bp of the TSS. 3013 genes are bound by M1BP, while 1885 are bound by GFZF. Furthermore, both factors are almost exclusively bound to the promoter region of active genes (Figure 3-11).

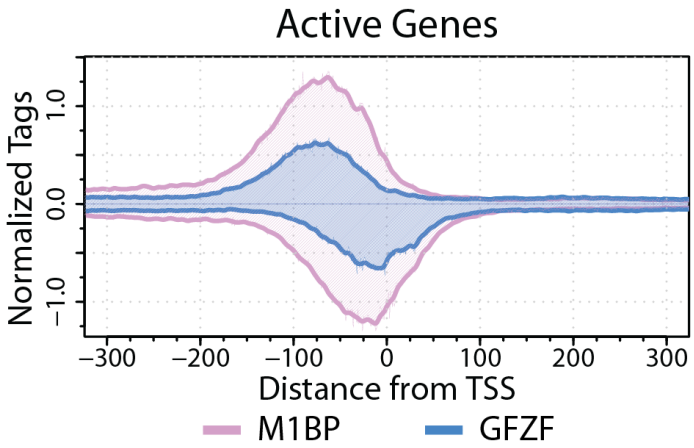

Figure 3-10. GFZF and M1BP share similar exonuclease footprints.
Composite plot of the heatmaps displayed in a 600 bp window centered on the TSS. ChIP-exo reads for the forward and reverse strands are displayed above and below the x-axis, respectively.

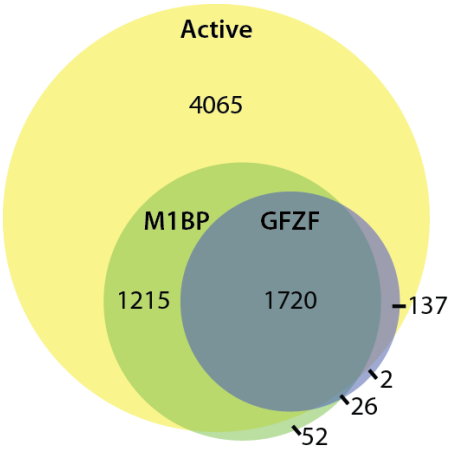

Figure 3-11. GFZF colocalizes with M1BP at promoter.
Venn diagram showing the overlap between active genes with M1BP and GFZF peak calls within 100 bp of a TSS. Genes have not been filtered for neighboring gene TSSs.

Gene ontology analysis of genes with a GFZF peak within 100 bp of the TSS revealed that, like M1BP (Li & Gilmour, 2013), GFZF is highly enriched at the promoters of genes that perform housekeeping functions (i.e. metabolism, organization, and cellular physiology) (Table 3-1). Thus, I conclude that GFZF and M1BP show remarkable overlap throughout the genome.

Term Category (GOTERM_BP_1)	P-value (EASE)
cellular process	2.4E-61
cellular component organization or biogenesis	1.1E-54
biological regulation	3.8E-41
regulation of biological process	5.1E-41
localization	1.2E-21
metabolic process	9.3E-21
developmental process	1.3E-20
growth	4.7E-15
signaling	5.5E-12
response to stimulus	1E-11

Table 3-1. GFZF-associated genes tend to serve housekeeping functions.
The table shows enriched gene ontology terms for GFZF-bound promoters.

GFZF is a transcriptional co-activator

The extensive co-localization of GFZF with M1BP, a known transcription factor (Li & Gilmour, 2013), raises the possibility that GFZF is a transcription factor. To test if GFZF activates transcription, I performed a dual-luciferase reporter assay following GFZF depletion in S2R+ cells (Figure 3-12 A).

I used the GFZF-associated promoters for the Ribosomal Protein gene, *RpLP1*, Sex-lethal (*Sxl*) gene, *roX2* gene, or abnormal wing discs gene (*awd*) to drive transcription of a Firefly luciferase reporter. These promoters were chosen because previous studies had linked GFZF to processes and pathways in which their genes or gene products are involved (Phadnis *et al*, 2015; Provost *et al*, 2006). As an internal control for the transfection efficiency, I used the *RpIII128* promoter, which lacks M1BP and GFZF, to drive expression of a sequence coding Renilla

luciferase. Both Firefly- and Renilla-coding plasmids were transfected with either an empty expression vector, or one that expressed a FLAG-tagged version of GFZF. Cells were treated for 1 day with dsRNA targeting either *lacZ* as a control, exon 2 of *GFZF*, or the 5' untranslated region (5'UTR) of *GFZF* and subsequently transfected with reporter plasmids. Two days later, cells were lysed and assayed for Firefly and Renilla luciferase activity. Ectopically expressed FLAG-GFZF activated each of the promoters in the presence of the lacZ control RNAi (Figure 3-12 B-E, cf. bars 1 and 2). This suggests that GFZF levels in the cell are limiting. RNAi targeting exon 2 of both the endogenous and ectopic GFZF inhibited GFZF dependent activation (Figure 3-12 B-E, cf. bars 2 and 4). In contrast, RNAi targeting the 5' UTR of endogenous GFZF, which is different from the 5' UTR of ectopic GFZF, did not inhibit activation by FLAG-GFZF (Figure 3-12 B-E, cf. bars 2 and 6). Instead, the level of expression mediated by endogenous GFZF was diminished (Figure 3-12 B-E, cf. bars 1 and 5).

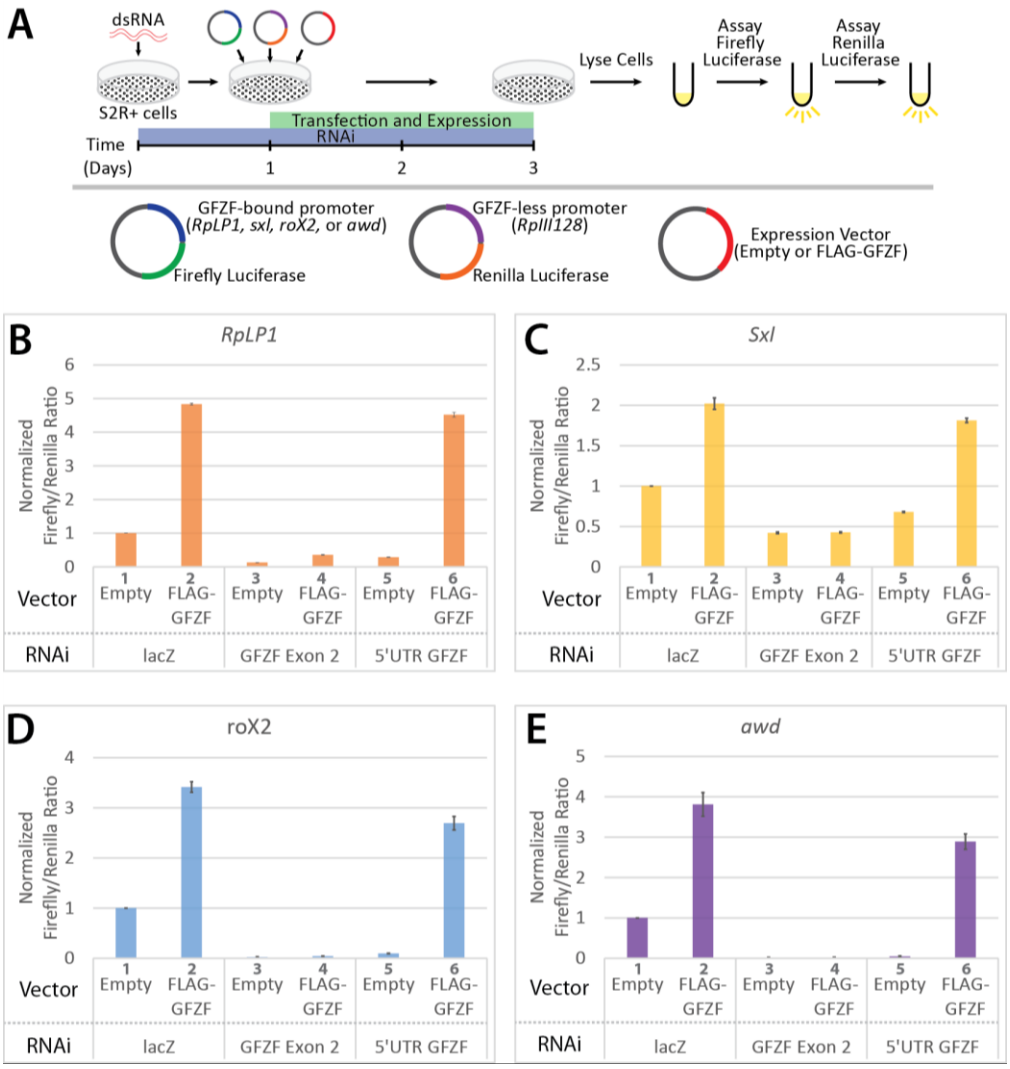

Figure 3-12. A luciferase reporter assay reveals that GFZF is a transcriptional co-activator.
(A) S2R+ cells were treated with dsRNA targeting either *lacZ* (RNAi control), Exon 2 of *GFZF*, or the 5' UTR of endogenous *GFZF*. The Exon 2 RNAi targets both the endogenous and transfected version of *GFZF* and the 5' UTR RNAi targets only the endogenous *GFZF*. After 24 hour incubation with dsRNA, luciferase reporter plasmids and a vector that expresses either empty or wild-type GFZF were transfected into the cells. After an additional 48 hour incubation, the cells were lysed then assayed in tandem for Firefly and Renilla luciferase activity. Firefly luciferase expression was driven by a GFZF-bound promoter while Renilla luciferase expression was driven by the GFZF- and M1BP-less RpIII128 promoter. (B-E) Firefly/Renilla ratios are displayed for *RpLP1*, *Sxl*, *roX2*, and *awd*. All values are normalized to the *lacZ* dsRNA, empty vector sample at each promoter.

To determine if GFZF is involved in activation of endogenous genes, I knocked down the level of GFZF and used ChIP to monitor the association of GFZF, M1BP, and Pol II with the same promoters that were tested in our transient expression assay. RNAi targeting *GFZF* caused significant decreases in the level of GFZF associated with the *RpLP1*, *Sxl*, *roX2*, and *awd* promoters (Fig. 3-13A).

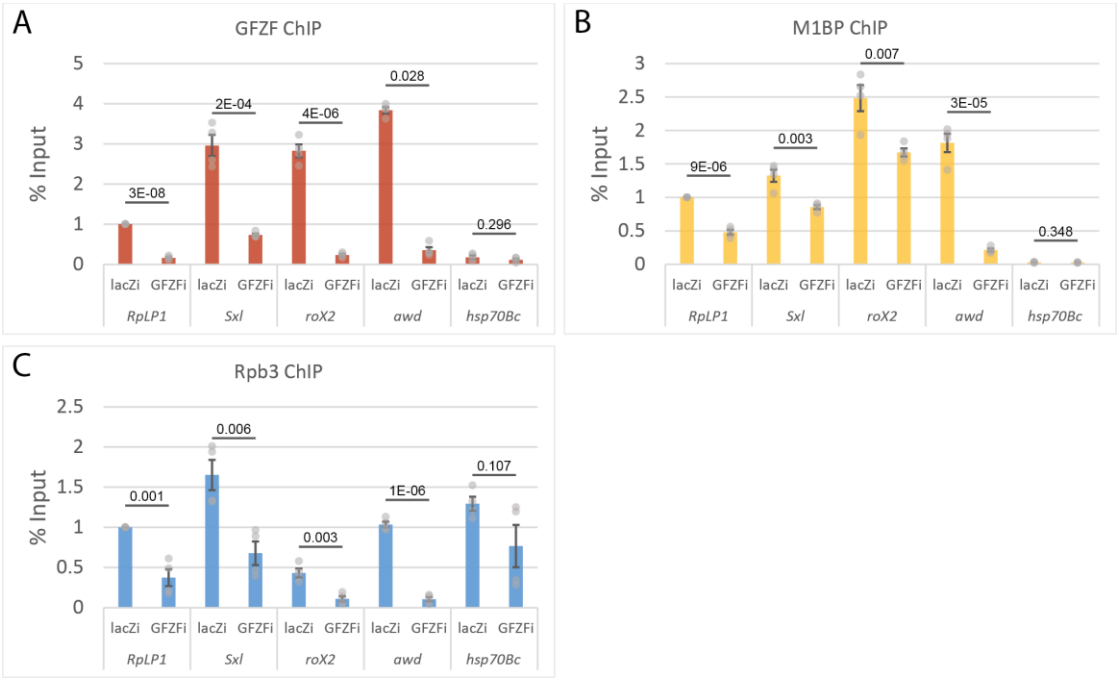

Figure 3-13. GFZF knockdown results in Pol II and M1BP loss at GFZF-bound promoters. qPCR quantification of (A) GFZF, (B) M1BP, and (C) Rpb3 ChIP samples following 3-day RNAi treatment with dsRNA targeting either *lacZ* (lacZi) as a control or *GFZF* (GFZFi). Error bars show standard deviation (n=4 biological replicates). Two-tailed t-tests assuming equal variance were used to generate p-values (reported for each loci tested).

These results confirm that my ChIP-exo analysis with the GFZF antibody indeed monitors GFZF. Knock-down of GFZF also caused decreases in the level of M1BP associating with the promoters (Fig. 3-13 B). This was unexpected since our biochemical analysis showed that M1BP bound a promoter fragment independently of GFZF (Fig. 3-2). Notably, an examination of M1BP ChIP-exo reads on M1BP-bound genes with or without GFZF reveals that

averaged M1BP signal intensities are higher on genes that have a GFZF peak (Figure 3-14). Western blot analysis shows that the knockdown of GFZF does not affect the level of M1BP (Fig. 3-15). Thus, the contribution of GFZF to M1BP promoter-occupancy must reflect some role for GFZF contributing to M1BP binding in a cellular context. GFZF might be stabilizing the binding of M1BP to Motif 1. Because ChIP experiments ultimately rely on crosslinking, another explanation could be that GFZF induces some conformational change in M1BP that affects the crosslinking efficiency. Alternatively, GFZF might only associate with stably-bound M1BP. In accordance with transient expression data, the knock down of GFZF caused a marked decrease in the level of Rpb3, a Pol II subunit, detected at GFZF-associated promoters. As expected, GFZF depletion had an insignificant impact on Rpb3 associated with the *hsp70* promoter (Fig. 3-13C). Taken together, the transient expression data and the ChIP analysis establish that GFZF is a transcriptional co-activator.

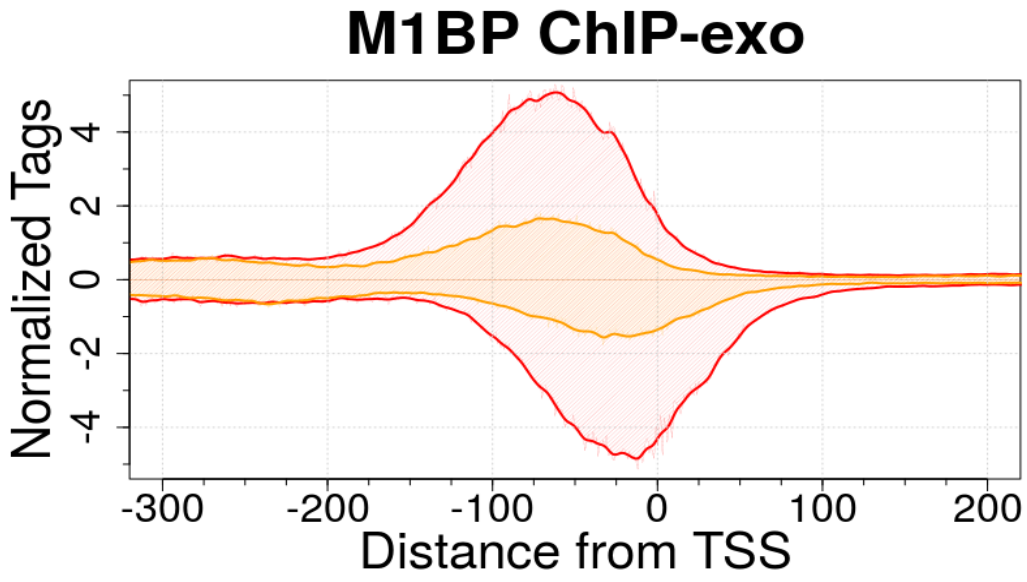

Figure 3-14. M1BP signals are stronger at GFZF-bound promoters.
Normalized M1BP ChIP-exo are mapped relative to the TSS of genes with (red trace, n=1766) or without (orange trace, n=1304) a GFZF peak within 100 bp of the TSS.

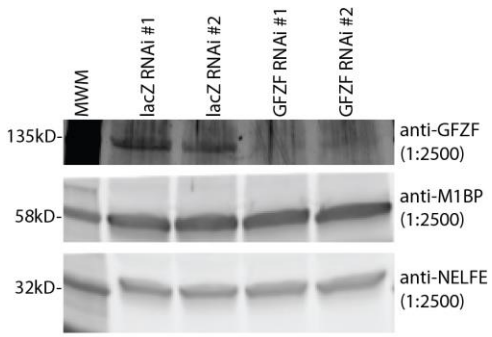

Figure 3-15. Total cellular levels of M1BP are unaffected following GFZF depletion.
Anti-GFZF, anti-M1BP, and anti-NELF E probed western blots of S2R+ cell lysates following 3-day RNAi. Western blots are from RNAi-treated cells performed as biological replicates and are annotated as #1 and #2. MWM: molecular weight marker. The antibody dilution ratio in 2% milk TBS-T is indicated in parentheses on the right-hand side of each membrane section.

I attempted to establish an *in vitro* transcription system that was dependent upon GFZF, but depletion of GFZF had no effect on transcriptional output from the gene promoters tested *in vitro*. Perhaps activation by GFZF requires a chromatin context that is not provided by my *in vitro* transcription system.

GFZF has glutathione S-transferase activity.

An intriguing feature of GFZF is its GST homology region which is unprecedented for a transcription factor. A previous study demonstrated that GFZF binds a glutathione (GSH) column and can be eluted with GSH in a dose-dependent fashion (Dai *et al*, 2004). To test if GFZF functions as a glutathione S-transferase and to measure its affinity for GSH, I expressed the GST domain of GFZF with a His-tag in *E. coli* and purified it using metal-affinity and ion-exchange chromatography (Figure 3-16 A). I also designed and expressed a catalytic mutant (S876A) of GFZF using the structure of a related GST in silkworm for reference (Kakuta *et al*, 2011). I next assayed GST activity by monitoring the increase in absorbance at 340 nm that

results when glutathione (GSH) is conjugated to 1-Chloro-2,4-dinitrobenzene (CDNB). Based on initial reaction velocities, the K_m for glutathione for the wild-type (wt) and S876A mutant was determined to be 0.07 mM and 3.28 mM, respectively (Figure 3-16 B). The K_m of wt GFZF falls well below the physiological range of GSH concentrations which has been reported to be between 1-10 mM (Montero *et al*, 2013), though it has been reported that GSH concentrations are lower in the nucleus (Söderdahl *et al*, 2003). Thus, GFZF's high affinity for GSH suggests that it is probably almost always bound in a cellular context.

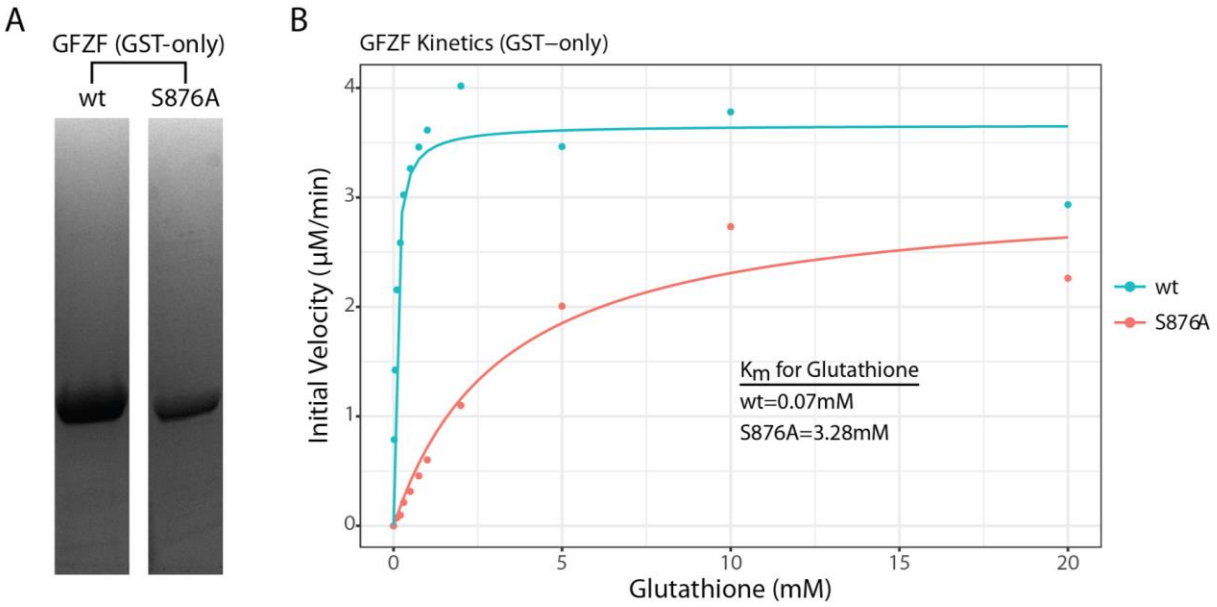

Figure 3-16. GFZF has GST activity with high affinity for glutathione.
(A) Coomassie-stained SDS-PAGE analysis of the His-tagged GST homology domain of GFZF (wt and S876A mutant). (B) Kinetic analysis of GFZF's GST domain. The K_m was obtained by altering GSH concentration while keeping CDNB constant.

To determine if the GST activity was involved in transcriptional activation, I measured activation of the luciferase reporter genes in the presence of a wild-type, mutant (S876A), or truncated GFZF which has the GST domain deleted (Figure 3-17 A). The wt and S876A mutant activate transcription to a similar extent while the truncated GST-less mutant had approximately

half as much activity (Figure 3-17 B). While wt GFZF activates transcription more robustly than the GST-less mutant in the luciferase assay, it was critical to assess whether differences in protein expression could account for the differences between those samples. To that end, I performed western blots against the FLAG epitope to quantify ectopic GFZF expression in cells (representative western blots shown in Figure 3-17 C). Upon comparing the fold increase in luciferase activity with the fold increase in ectopic protein expression, I conclude that the GST portion of GFZF does not contribute to its ability to activate transcription in this assay (Figure 3-17 D). (p-value=0.78, two tailed t-test assuming equal variance comparing the difference in fold increase of wt/ΔGST for the luciferase activity and ectopic GFZF expression, n=12).

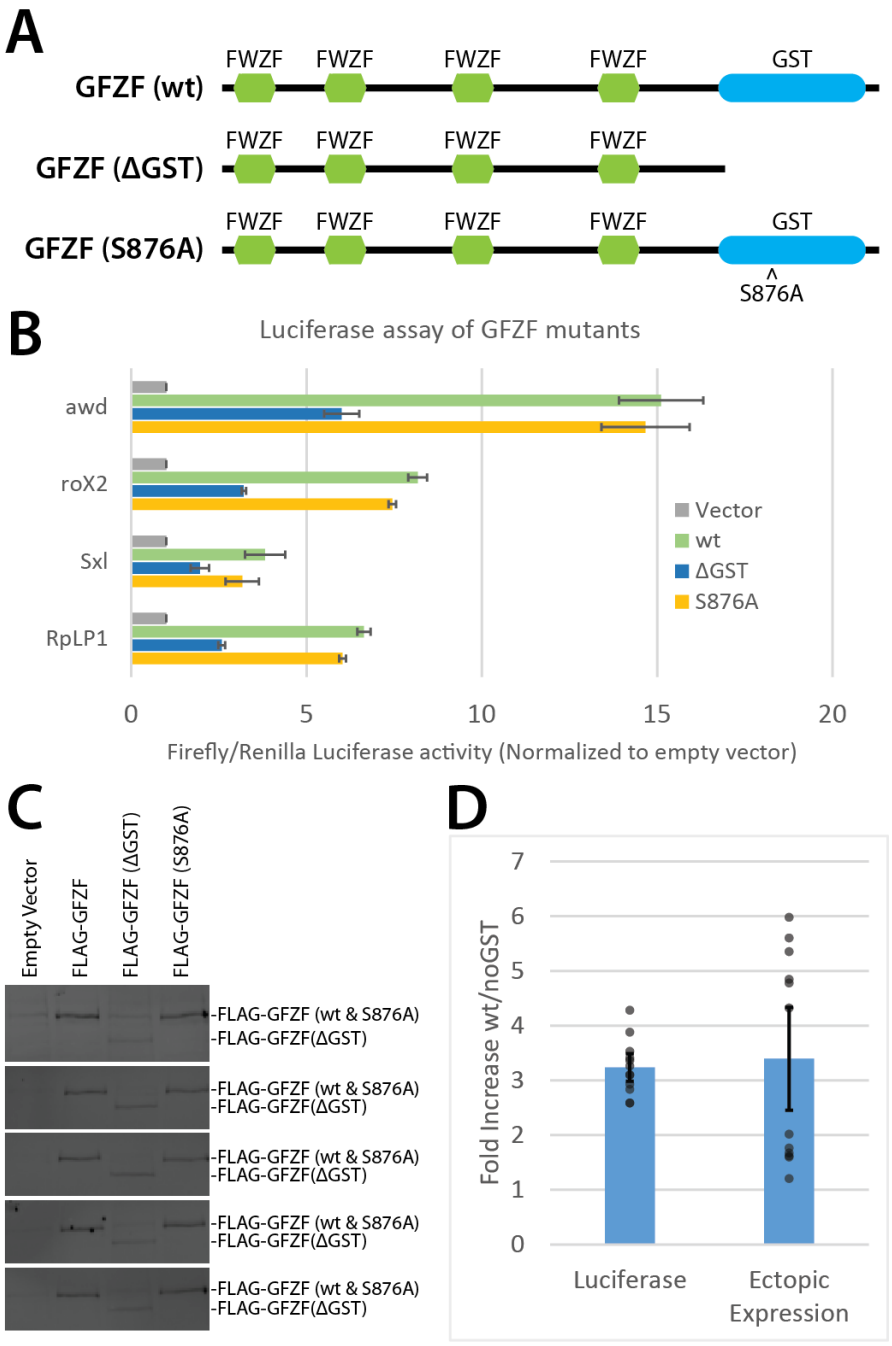

Figure 3-17. GFZF's GST domain is not required for transcription activation of transfected DNA. (A) Schematic of GFZF and the mutants used to test the function of GFZF's GST domain in transcription activation. FWZF stands for FLYWCH zinc finger domain. (B) Normalized Firefly/Renilla luciferase activities are displayed for four different reporter promoters with wt, ΔGST, and S876A mutant versions of GFZF ectopically expressed (n=3 biological replicates). (C) Representative western blots of luciferase lysates following ectopic expression of wt, ΔGST, or S876A GFZF. Western blots from 5 of 12 independent experiments are shown. The upper and lower band in each blot corresponds to full-length (wt

or S876A) and truncated (ΔGST) GFZF, respectively. Chart comparing the wt over ΔGST fold increase of the luciferase activity and ectopic expression levels. To calculate the fold difference in luciferase activity, I subtracted the endogenous (Empty vector) luciferase values from the ectopically expressed GFZF luciferase values then plotted the wt/ΔGST ratio. For the fold difference in ectopic expression levels, I calculated the wt/ΔGST ectopic expression ratios after quantifying the wt and ΔGST anti-FLAG western blot signal intensities in ImageJ. There is no significant difference between the fold increase in luciferase activity and the fold increase in ectopic expression for the wt vs ΔGST sample ($p=0.78$). Error bars represent standard deviation. n=12 (3 biological replicates from 4 promoters).

Identification of GFZF as a transcription factor provides insights into its roles in a broad spectrum of biological processes.

Since its initial discovery, GFZF has appeared as a "hit" in numerous screens (Kondo & Perrimon, 2011; Ashton-Beaucage *et al*, 2014; Gonzalez *et al*, 2014; Phadnis *et al*, 2015; Provost *et al*, 2006). While possible explanations for GFZF's appearance in these screens have been put forth, they have lacked a unifying cellular function that could explain GFZF's seemingly disparate roles. Here I show that GFZF binds approximately 1800 genes and functions as a transcriptional co-activator. This new information can explain the broad functionality of GFZF. *GFZF* was first identified in *Drosophila* as a suppressor of a gene called *killer of prune* (also known as *awd*). Mutations in *awd* alone cause no phenotype but are lethal in flies that are homozygous for nonlethal mutations in another gene called *prune*. It was proposed that mutations in *GFZF* suppressed the lethality caused by the combination of mutations in *awd* and *prune* because wild-type GFZF was generating something toxic by conjugating glutathione to a metabolic product derived from the activities of mutant *prune* (a cyclic AMP phosphodiesterase) and mutant *awd* (a nucleoside diphosphate kinase) (Provost *et al*, 2006). However, my results provide a simpler explanation: GFZF associates with the *awd* promoter and activates transcription (Figure 3-18 A, Figure 3-12 E, and Figure 3-13 A-C).

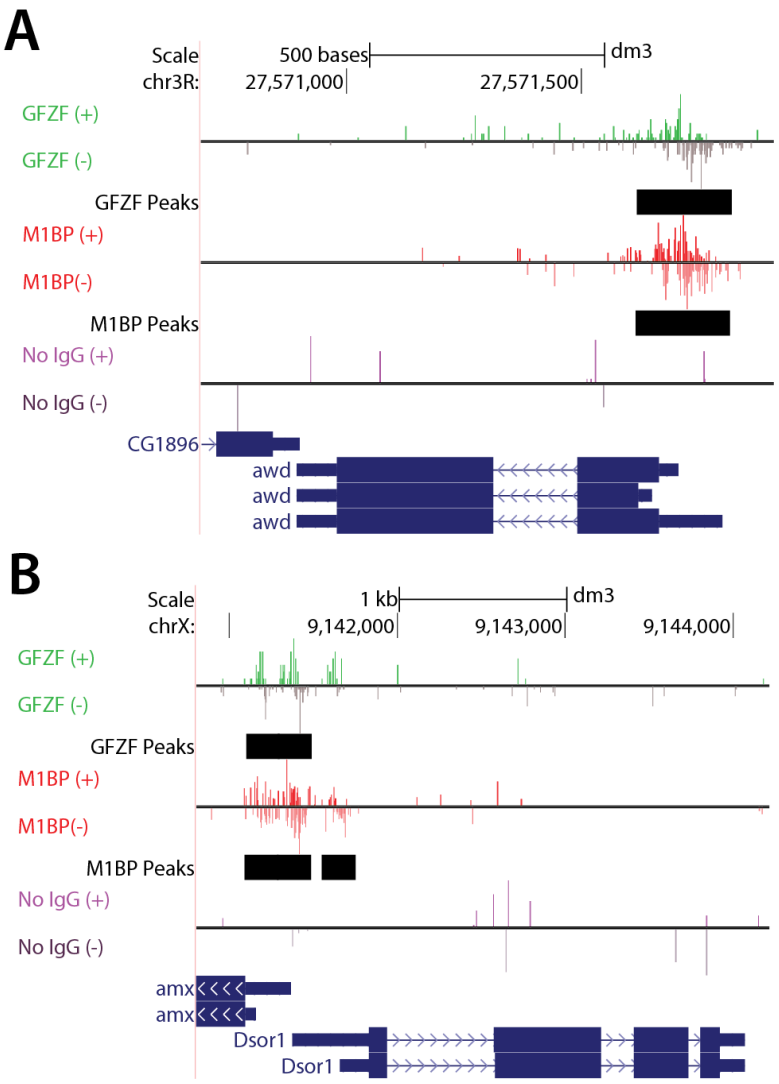

Figure 3-18. GFZF binds the promoters of genes integral to genetic screens in which GFZF was a "hit".

(A) UCSC genome browser shot showing the association of GFZF and M1BP with the promoter of the gene, *awd*. *awd* is also known as *killer of prune*. Mutations in *awd* alone cause no phenotype but are lethal in flies that are homozygous for nonlethal mutations in another gene called prune (Provost *et al*, 2006). The reads above and below the horizontal line correspond to ChIP-exo reads from the forward (+) and reverse (-) strands, respectively. (B) UCSC genome browser shot showing GFZF's association with the promoter of Dsor1 which is critical for the RAS/MAPK signaling pathway (Ashton-Beaucage *et al*, 2014). Dsor1 encodes the MEK kinase. Shown are forward (+) and reverse (-) strand ChIP-exo reads for M1BP, GFZF, or a No IgG ChIP control. Bars in black show the genomic location of MultiGPS-generated peaks for either GFZF or M1BP.

In this scenario, GFZF mutants defective in transcription would reduce the level of expression of mutant *awd*. The drop in *awd* transcription means there would no longer be sufficient mutant Awd protein to cause lethality with mutant Prune protein. In another case, GFZF's appearance in a screen for RAS-mediated MAPK activation can be explained by it binding to the core promoter region of *mek* (Figure 3-18 B, *Dsor1*). In accordance with GFZF's function as a transcriptional co-activator, the authors demonstrate that knockdown of GFZF results in reduced levels of *mek* transcripts (Ashton-Beaucage *et al*, 2014). Likewise, GFZF's appearance in the G2-M DNA damage checkpoint screen could be simply explained by GFZF being required for the transcription of other factors involved in this DNA damage checkpoint. Our ChIP-exo analysis indicates that GFZF associates with 22 of the 64 genes that were identified in this screen, including the promoters of factors known to have roles in this DNA damage checkpoint including *myt1*, *14-3-3ε*, and *tefu* (Kondo & Perrimon, 2011) (Figure 3-19).

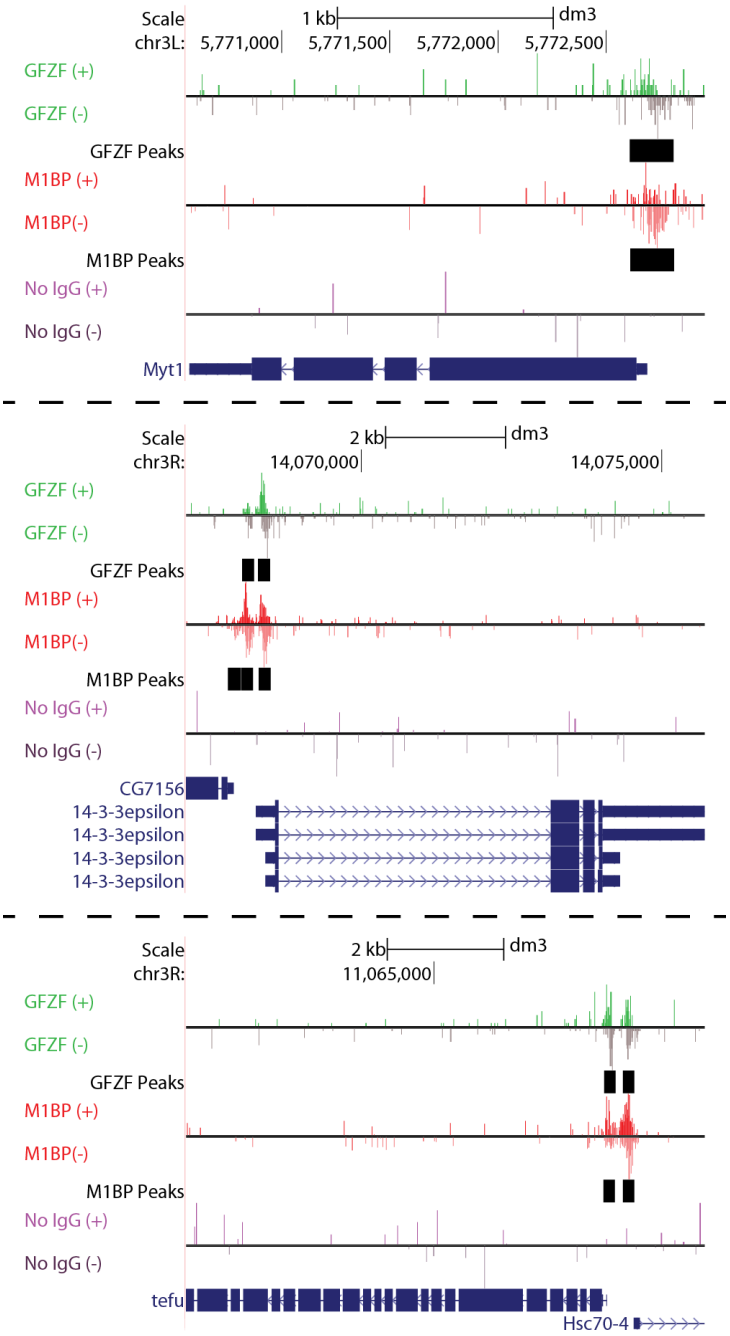

Figure 3-19. GFZF binds the promoters of genes involved in the G2 to M DNA damage checkpoint. Based on an RNAi screen, *Myt1*, *14-4-4ε, and tefu* were among the top genes to affect the G2 to M DNA damage checkpoint (Kondo & Perrimon, 2011). UCSC Genome browser shots of ChIP-exo data show that GFZF associates with the promoters of all three genes. Shown are forward (+) and reverse (-) strand ChIP-exo reads for M1BP, GFZF, or a No IgG ChIP control. Bars in black show the genomic location of MultiGPS-generated peaks for either GFZF or M1BP.

GFZF was also identified in a screen for mutations that affect hybrid inviability (Phadnis *et al*, 2015). When female *D. melanogaster* are mated to male *D. simulans*, no male progeny are produced. Mutations in GFZF in male *D. simulans* allowed production of male progeny in this interspecies mating. GFZF binds to the promoter of three (*msl-1*, *msl-2*, and *mle*) out of five subunits that comprise the male-specific lethal (MSL) complex in flies. Additionally, it binds to the promoter region of *roX2*, one of the ncRNAs that is part of the MSL complex (Figure 3-20).

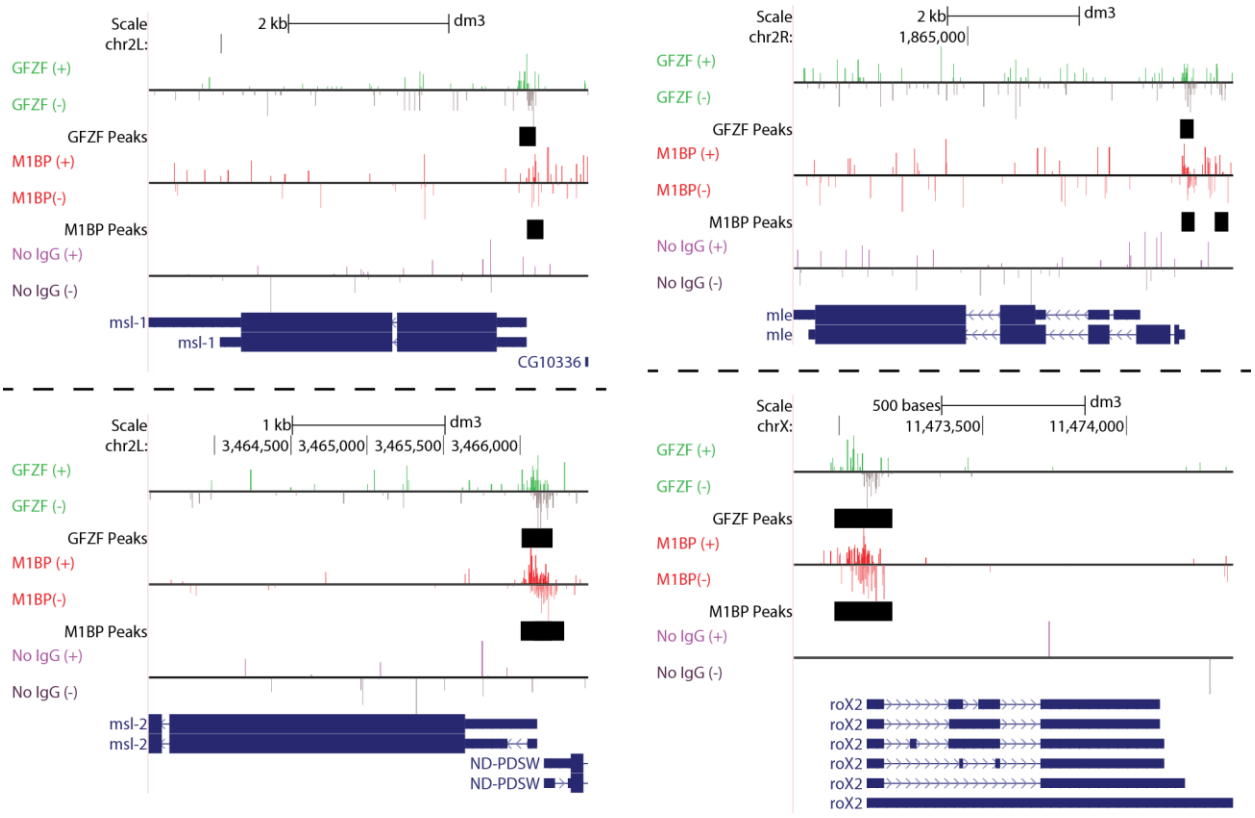

Figure 3-20. GFZF binds the promoter region of Male Specific Lethal (MSL) complex genes. UCSC Genome browser shots of genes encoding components of the MSL complex. GFZF's appearance in a screen for hybrid inviability, which leads to male lethality, could pertain to its role in expression of components of the MSL complex. Shown are forward (+) and reverse (-) strand ChIP-exo reads for M1BP, GFZF, or a No IgG ChIP control. Bars in black show the genomic location of MultiGPS-generated peaks for either GFZF or M1BP.

The MSL complex functions in dosage compensation in male flies by doubling the amount of transcription arising from genes on the X chromosome; disrupting the function of the MSL complex causes male lethality. Since GFZF is a transcriptional co-activator and binds the promoters of several genes encoding the MSL complex, I speculate that hybrid-specific GFZF-mediated misregulation of MSL components might be contributing to male lethality. This would be consistent with others who have provided evidence that defects in dosage compensation contribute to hybrid inviability (Rodriguez *et al*, 2007; Chatterjee *et al*, 2007; Bachtrog, 2008). However, a follow-up study which tested the hypothesis that defects in the MSL complex contribute to hybrid inviability concluded that defects in MSL function cannot fully explain hybrid inviability (Barbash, 2010). It could be that GFZF's role in hybrid inviability is more nuanced than misregulation of MSL complex components and might involve mis-expression of other factors involved in maintaining incompatibility. Whatever the case, it is reasonable to speculate that GFZF's role will involve mis-regulation of genes required for maintenance of hybrid inviability.

The GST activity of GFZF

GFZF is unusual because of its unique combination of zinc fingers and a functional GST domain. Our search for homologous genes in other organisms indicates that genes sharing homology to the entirety of GFZF are limited to Schizophora, the section of true flies which includes the common house fly. Since other neopterans, including mosquitoes, have GST proteins that share homology with GFZF's GST domain but lack GFZF's zinc fingers, it is likely that GFZF evolved recently as a result of a gene fusion (Clustal alignments in Figure 3-21 and Cladogram in Supplemental Figure 3-22). In accordance with this hypothesis, mRNA expression

data shows that there is a second promoter in the intron of *GFZF* that immediately precedes the GST domain of the full-length *GFZF* gene and the resulting transcript is predicted to encode a functional GST. This transcript is detected from 14 hour old embryos to adults whereas the full-length GFZF is detected throughout development beginning with 0-2 hour old embryos (Gramates *et al*, 2017).

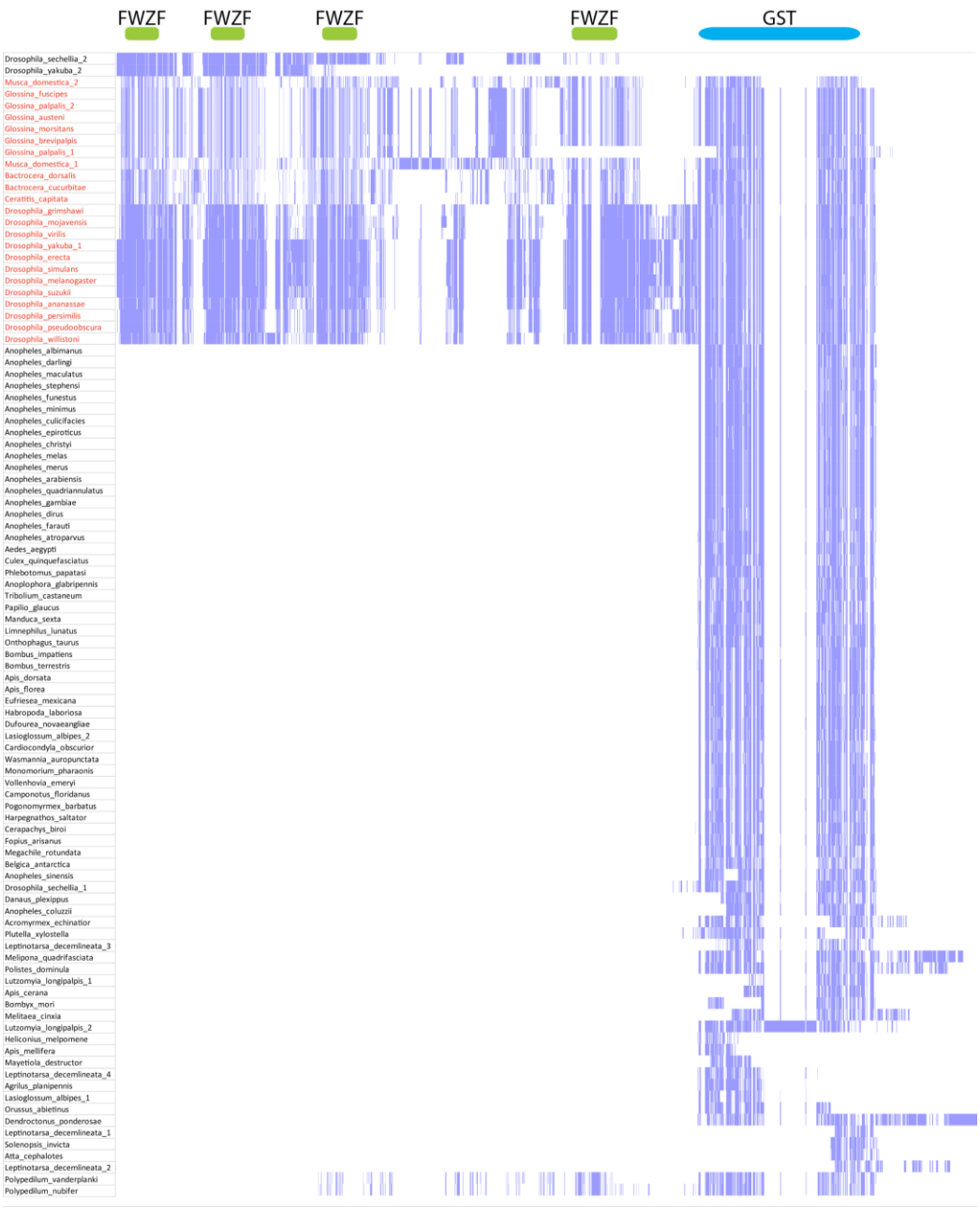

Figure 3-21. Orthologs of the entirety of GFZF are restricted to Schizophora.

GFZF orthologs and sequences were obtained from OrthoDB by providing the amino acid sequence of GFZF and alignments were generated using Clustal Omega (Sievers *et al*, 2011). The image was generated using Jalview with the "BIosum62" coloring mode (Waterhouse *et al*, 2009). The zinc finger (FWZF) and GST domains are delineated above the alignments. Species with full-length GFZF are highlighted in red. Only Schizophorans have full-length GFZF.

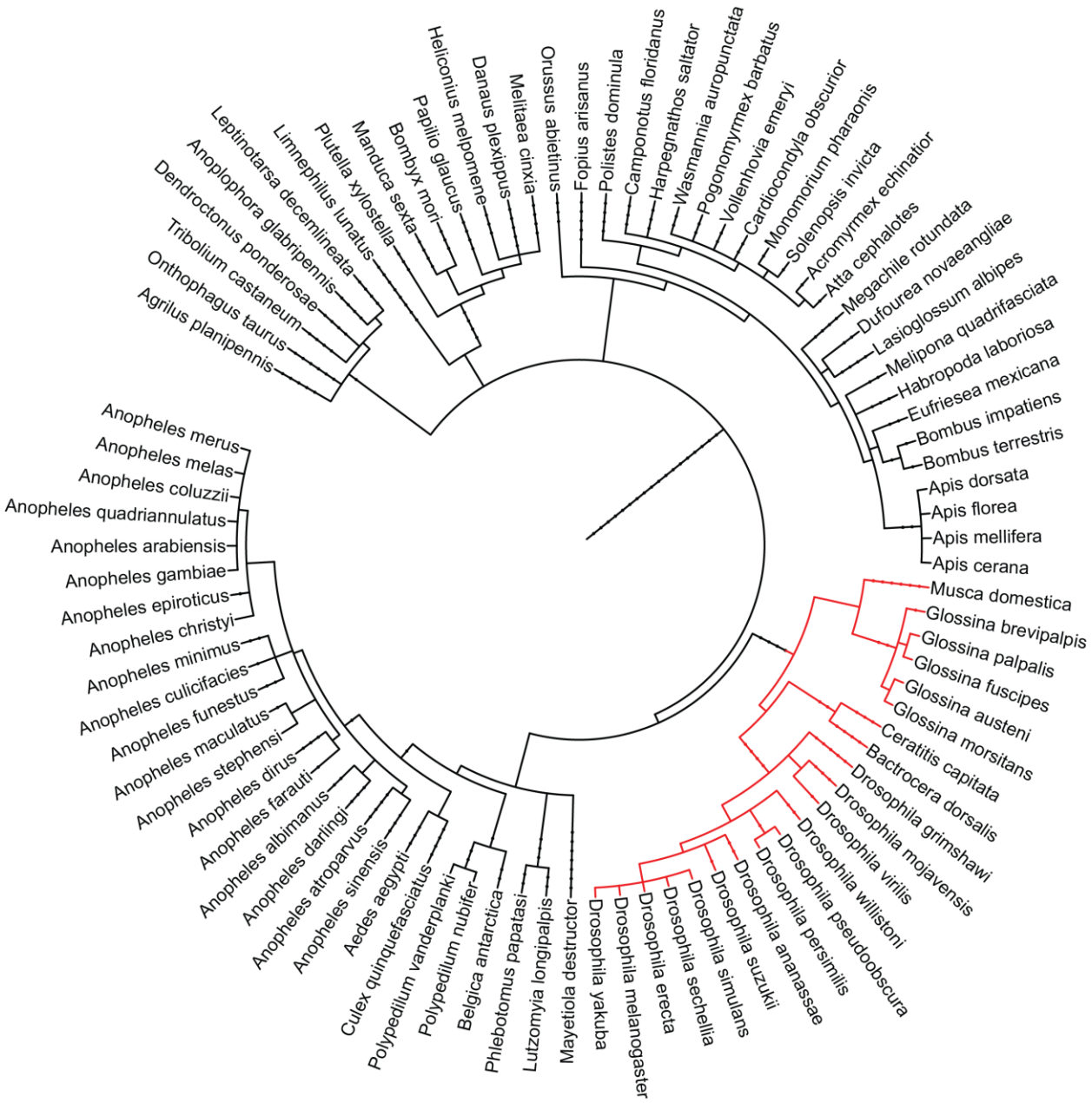

Figure 3-22. Orthologs of the entirety of GFZF are restricted to Schizophora.

Cladogram showing Schizophora cladogram was generated with Interactive Tree of Life (iTOL) server (Letunic & Bork, 2016) by feeding the phyloT server with the list of organisms that possess GFZF orthologs according to OrthoDB as shown in Figure 3-19. Species with orthologs possessing N-terminal FLYWCH zinc fingers, like GFZF, are highlighted in red.

At this point, the function of the GST domain is unclear. I observed that deletion of this domain reduced the level of expression of the remainder of the protein and that the remaining part still activated transcription. Since I only assayed for function on transiently transfected DNA, it remains possible that the GST activity is important in a natural chromatin context, which is not formed on transiently transfected DNA. Mutations in the GST domain of GFZF have been identified that cause larval lethality so the domain appears to be essential (Provost *et al*, 2006).

It is possible that the gene fusion resulting in GFZF is fortuitous and that the GST domain's function is not linked to gene regulation. On the other hand, this fusion raises the intriguing possibility that GST activity is important for gene expression and that other organisms bring GST activity to a gene's promoter through protein-protein interactions. GST proteins are best known for their roles in protecting cells from toxic endogenous and xenobiotic compounds so GST might function at promoters to inhibit DNA damage (Hayes *et al*, 2005). Another possibility is that GFZF could serve as a sensor of the redox potential of the cell. Having a GST transcription factor act as a nuclear sensor of the redox state of the cell could ensure that cells can quickly alter their transcriptional output in response to stress and chemical insult. There is precedent for redox regulation of transcription factors, both directly (Brigelius-Flohé & Flohé, 2011), and through signal transduction (Adler *et al*, 1999). Brf2, a Pol III core transcription factor, has a single oxidation-prone cysteine residue that when oxidized, inhibits Brf2's ability to form a complex with TBP at some Pol III-dependent promoters. In cells, oxidative stress caused a sharp decline in Brf2-dependent gene transcripts (Gouge *et al*, 2015). In an example of redox regulation through signal transduction, a GST protein acts to inhibit c-Jun N-terminal kinase (JNK) activity under normal physiological conditions. However, when cells are treated with

hydrogen peroxide or UV irradiation, the GST dimerizes and no longer inhibits JNK, thus allowing the signaling cascade to commence (Adler *et al*, 1999). As further evidence of redox-driven transcriptional regulation, sublethal levels of hydrogen peroxide globally reduce the turnover rate of Pol II paused in the promoter proximal region of genes (Nilson *et al*, 2017). Finally, PrfA, a protein in the intracellular pathogenic bacteria *Listeria monocytogenes* appears to be allosterically regulated by glutathione (Reniere *et al*, 2015). If, as in the above examples, such a molecular switch regulates GFZF function in response to redox perturbations, it would represent an elegant means of quickly altering gene expression of a multitude of genes in response to stress.

Materials and Methods

Nuclear Extracts

Nuclear extracts were prepared from 0-12 hour Oregon R embryos as previously described (Biggin & Tjian, 1988; Li & Gilmour, 2013).

Immobilized Template Pulldowns

Immobilized template pulldowns were performed by annealing oligonucleotides (sequences listed in (Baumann *et al*, 2017)) corresponding to the core promoter sequence of *mRpS30* (-32 to +18) or *RpLP1* (-37 to +13) and gel purifying the annealed templates from a polyacrylamide gel. One strand of the annealed template was biotinylated at the 5’ end. Oligonucleotides used for pulldowns from nuclear extracts have an inverted 3’ dT to inhibit degradation by 3’ exonuclease present in the nuclear extracts. For immobilized template pulldowns from nuclear extracts, 3.5 µg of wt or mutant motif 1 *mRpS30* template was immobilized on 100 µl streptavidin Dynabeads (ThermoFisher - 11205D) according to the manufacturer’s instructions. Template-bound beads

were equilibrated in 0.18 M HEMGN (180 mM KCl, 25 mM HEPES pH 7.6, 12.5 mM $MgCl_2$, 0.1 mM EDTA pH 7.9, 10% glycerol, 0.1% NP-40). The beads were then incubated at room temperature for 45 minutes with 250 µl of nuclear extract and 40 µg of HaeIII-digested *E. coli* DNA. Beads were washed 5 times at room temperature for 10 minutes each wash with 0.18 M HEMGN. Finally, beads were transferred to a new tube, boiled in gel loading buffer, and the resulting proteins were analyzed on an 8% SDS-polyacrylamide gel.

For immobilized template pulldowns using purified factors, 800 ng wt or mutant Motif 1 *RpLP1* template was immobilized on 20 µl streptavidin Dynabeads. 150 µl binding reactions had the following composition: 25 mM HEPES (pH 7.6), 200 mM NaCl, 40 mM KCl, 10% glycerol, 10 µM $ZnCl_2$, 0.1% NP-40, 1 mM DTT, 3.5 µg His-GFZF (or dialysis buffer), 3.5 µg His-M1BP (purification previously described in (Baumann & Gilmour, 2017), and 25 µg sonicated Salmon sperm DNA. Samples were incubated at room temperature for 1 hour and washed 3 times with wash buffer consisting of 25 mM HEPES (pH 7.6), 200 mM NaCl, 40 mM KCl, 10% glycerol, 10 µM $ZnCl_2$, 0.1% NP-40, and 1 mM DTT then transferred to a new tube and eluted with 15 µl gel loading buffer for 10 minutes at 75°C. Samples were loaded on an 8% SDS-PAGE and stained with coomassie brilliant blue. 10% (15 µl) of the unbound fractions were also analyzed by SDS-PAGE as above.

Purification of Maltose-binding protein (Mal) fusions

One liter of BL21 (DE3) *E. coli* culture transformed with plasmids encoding N-terminal Mal fused with a rigid linker to the lacZα fragment or M1BP were grown at 37°C to an OD600 of 0.8. Expression of the Mal fusion protein was induced by addition of IPTG to a final

concentration of 300 µM. Cultures were incubated at 37°C for an additional 2 hours, then placed on ice and harvested by centrifugation for 10 minutes at 7500 x *g*. The pellet was frozen in liquid nitrogen, then resuspended in 125 ml Mal lysis buffer (25 mM HEPES, pH 7.6, 500 mM NaCl, 10% glycerol, 0.1% NP40, 1 mM DTT, 0.4 mM PMSF, and a protease inhibitor cocktail). All subsequent steps were performed at 4°C. Cells were lysed by passing the cell suspension 3 times through a microfluidizer. Lysates were cleared by ultracentrifugation for 1 hour in a Beckman Type 70 Ti rotor at 35,000 rpm. 50 ml of supernatant was passed through 0.5 ml of amylose resin (NEB E8021) packed in a Poly-Prep gravity column (Bio-Rad 7311550). The resin was washed with 12 column volumes of Mal lysis buffer and finally with 4 column volumes of Mal lysis buffer with 180 mM NaCl. Protein-bound resin was stored at 4°C.

Maltose-binding protein fusion pulldowns with His-GFZF

500 µl binding reactions consisted of 25 mM HEPES, pH 7.6, 180 mM NaCl, 10% glycerol, 0.1% NP40, 1 mM DTT, 0.4 mM PMSF, 2 µl Mal-bound amylose resin (~9 µg Mal fusion protein), and 3.5 µg His-GFZF. Reactions were incubated at 4°C for 2 hours with end-over-end rotation, then washed with 25 mM HEPES, pH 7.6, 180 mM NaCl, 10% glycerol, 0.1% NP40, 1 mM DTT, 0.4 mM PMSF at 4°C for 2 hours with end-over-end rotation. Resin was transferred to a new tube and material was eluted in 20 µl gel loading buffer by heating to 75°C for 10 minutes. Material was analyzed by 10% SDS-PAGE and stained with coomassie brilliant blue.

Polytene Chromosome Squashes and Immunofluorescence

Salivary glands were dissected from third instar larvae and stained with antibodies against GFZF (Dai *et al*, 2004), M1BP (Li & Gilmour, 2013), or anti-FLAG M2 antibody (Sigma - F1804) as previously described (Ghosh *et al*, 2011). The FLAG+HA-tagged M1BP transgenic flies were generated by amplifying the M1BP coding sequence from the plasmid described in (Li & Gilmour, 2013) and inserting it downstream of the hsp83 promoter in the pCaSpeR-hs83 (Missra & Gilmour, 2010). The resulting plasmid encoded M1BP with two consecutive FLAG tags and an HA tag at the N-terminus. *Drosophila* transformation was performed by Rainbow Transgenic Flies, Inc.

Purification of full-length His-GFZF

The coding sequence of GFZF was cloned from S2R+ cDNA into the NheI and EcoRI restriction sites of the pET28 expression vector. Rosetta (DE3) pLysS cells were grown in 1 liter of LB media at 37°C to an OD_{600} of 0.4. IPTG was added to a final concentration of 1 mM and cells were incubated at 18°C for 24 hours. Cells were collected, lysed in lysis buffer (50 mM HEPES (pH7.6), 500 mM NaCl, 10 mM imidazole, 10% glycerol, 0.1% Triton X-100, 2 mM PMSF), sonicated, and centrifuged at 20,000 x *g* for 20 minutes. Lysates were applied to 300 µl Ni-Nta resin (Qiagen - 30210) and bound in batch for 1 hour at 4°C. Resin was collected and washed in batch at 4°C for 15 minutes with 30 ml lysis buffer. Resin was packed in a column, washed with an additional 5 ml lysis buffer, and eluted with 50 mM HEPES (pH 7.6), 150 mM NaCl, 10% glycerol, 200 mM imidazole, 0.1% Triton X-100, and 2 mM PMSF. Eluates were further purified through Mono Q using a buffer consisting of 50 mM HEPES (pH 7.6), 10% glycerol, 0.1 mM $ZnCl_2$, 0.1% Triton X-100 with a NaCl gradient from 150 mM to 550 mM.

ChIP-exo

ChIP-exo was performed with antibodies against GFZF (Dai *et al*, 2004) and M1BP (Li & Gilmour, 2013) essentially as described in (Rhee & Pugh, 2012) with minor modifications. Libraries were quantified by qPCR and sequenced on an Illumina NextSeq 500. Basecalls were performed using Bcl2FastQ version 2.16.0. Sequenced reads were masked for low-quality sequence, then mapped to the *D. melanogaster* dm3 whole genome using BWA mem (versions 0.7.9a, 0.7.12) with the default parameters. Heatmaps were generated with HOMER bioinformatics software (Heinz *et al*, 2010) and Java Treeview (Saldanha, 2004). Tables for composite plots were generated with HOMER and plots were visualized using R (R Core Team). Genome browser images were generated with the UCSC genome browser (Kent *et al*, 2002).

Based on the knowledge that M1BP mediates GFZF's association with chromatin, I reasoned I might be able to boost GFZF's ChIP-exo signal intensity by using additional cross-linking reagents which would increase the likelihood that GFZF and M1BP form protein-protein cross links. To test this, I rinsed the cells twice with PBS then added 5 mM dimethyl adipimidate (DMA) and 1.5 mM ethylene glycol bis(succinimidyl succinate) (EGS). After incubating the cells with DMA and EGS for 10 minutes at room temperature, 1% formaldehyde was added and cells were incubated for an additional 10 minutes. I chose to use DMA and EGS because a previous report demonstrated that these crosslinkers enhanced the ChIP signals of proteins indirectly associated with chromatin (Zeng *et al*, 2006). I performed ChIP-exo with this triple crosslinked material and noticed that signal intensities improved and accuracy was retained. Thus, I used the data from this chromatin preparation for our GFZF analysis.

Peak Calling

The 5' ends of reads were obtained and MultiGPS (Mahony *et al*, 2014) was used to call peaks using the default settings with Pre-immune ChIP-exo libraries serving as control samples. A list of genes with a GFZF peak within 100 bp of its TSS was used for gene ontology (GO) analysis. GO analysis was performed using DAVID with the GOTERM_BP_1 list (Huang *et al*, 2009b, 2009a). The Venn diagram was generated with BioVenn (Hulsen *et al*, 2008). The active gene list derived from (Nechaev *et al*, 2010) as described in (Li & Gilmour, 2013).

Luciferase Reporter Assays

S2R+ cells were grown in M3+BPYE containing 10% FBS and at 25°C. For luciferase assays without dsRNA treatment, 1.7 million S2R+ cells in 600 µl of media were seeded per well in 24-well plates (Corning 353047) and incubated overnight. The following day, cells in a well were transfected with a mixture of: 25 ng pGL3-(*RpLP1* (-500 to +50), *Sxl* (-500 to +53), *roX2* (-258 to +60), or *awd* (-500 to +25)), 25 ng pRL-polIII-*Renilla* (Gilchrist *et al*, 2008), 50 ng pAc5.1 (empty or 2XFLAG-GFZF wt, ΔGST, or S876A mutant), 20 µl serum-free media, and 0.2 µl fuGENE HD (Promega - E2311). The plasmids and media were premixed prior to the addition of fuGENE HD as indicated in the manufacturer's protocol. Two days after transfection, cells were lysed and assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega - E1910) according to the manufacturer's protocol. To monitor relative levels of transiently expressed derivatives of GFZF, a portion of cell lysate from the Dual-Luciferase assay containing 20 µg of total protein (as determined by Bio-Rad Bradford Assay #5000006) was subjected to western blot analysis with M2 anti-FLAG antibody (Sigma - F1804). For luciferase assays with dsRNA treatment, dsRNA was generated by *in vitro* transcription with T7

polymerase on PCR-generated templates flanked by T7 promoter sequences (primer sequences used to generate each PCR template are included in (Baumann *et al*, 2017)). 1.7 million cells in 300 µl serum-free media were seeded per well in 24-well plates. Cells were then treated with 3 µg of dsRNA for 1 hour, after which 300 µl of media supplemented with 20% FBS and 2X antibiotic + antifungal (Corning 30-004-CI) was added. dsRNA-treated cells were then incubated overnight and transfected with DNA the following day as described above.

RNAi-depletion of GFZF followed by chromatin Immunoprecipitation (ChIP)

S2R+ cells were maintained in M3+BPYE containing 10% FBS and grown at 25°C. Ten milliliters of cells at 2.8 million cells per ml were plated in a 10 cm dish and incubated overnight. The following day, the media was removed, the cells were rinsed with PBS, and 6 ml of serum-free media was added. 60 µg dsRNA targeting either lacZ or Exon 2 of GFZF was added to the culture and incubated for 1 hour. 6 ml of media containing 20% FBS and 2X antibiotic + antifungal (Corning 30-004-CI) was added and cells were incubated for an additional 3 days. ChIP experiments were performed as described previously (Li & Gilmour, 2013) using rabbit polyclonal antisera against GFZF, M1BP, or Rpb3. Primers used to quantify percent recovery are listed in (Baumann *et al*, 2017). For western blots, cells were lysed in LDS sample buffer and material equivalent to 2 million lysed cells was subjected to SDS-PAGE, transferred to nitrocellulose, and probed with antibodies against GFZF, M1BP, or NELF-E.

Expression and Purification of the GFZF GST domain

For the purification of the His-tagged GST domain of GFZF, the DNA sequence encoding residues 800-1045 of GFZF was cloned into pET28 using the NheI and EcoRI

restriction sites. The S876A mutation was introduced using the In-Fusion cloning kit (Clontech 638910) with primers bearing the desired mutations. Rosetta (DE3) pLysS were transformed and cells were grown at 37°C in 4 liters of LB media to an OD600 of 0.8-0.9 and induced with 0.5 mM IPTG. Following induction, cells were incubated at 15°C for 24 hours. Cell were harvested and resuspended in 150 ml of lysis buffer (25 mM HEPES (pH7.6), 500 mM NaCl, 5 mM imidazole, 10% glycerol, 0.1% Nonidet P-40, 1 µM $ZnCl_2$, 0.1 mM PMSF, 20 mM 2-mercaptoethanol, and a protease inhibitor cocktail). The cells were lysed by passing the cell suspension through a microfluidizer three times, then ultracentrifuged at 125000 x *g* for 30 minutes. The cleared lysate was bound in batch with 1.5 ml of Talon resin at 4°C for 1 hour and washed in batch with 45 ml of lysis buffer for 30 minutes at 4°C. The Talon resin was packed into a column, washed with an additional 5 column volumes and eluted with buffer consisting of 50 mM HEPES (pH 7.6), 100 mM NaCl, 200 mM imidazole, 10% glycerol, 0.1% Nonidet P-40, 1 µM $ZnCl_2$, 0.1 mM PMSF, and 20 mM 2-mercaptoethanol. The eluates were dialyzed overnight at 4°C in a buffer consisting of 25 mM HEPES (pH7.6), 100 mM NaCl, 10% glycerol, 0.1% Nonidet P-40, 0.1 mM PMSF, and 2 mM DTT. Purified proteins were flash frozen in liquid nitrogen and stored at -80°C.

GST Activity Assay

1 mL reactions consisting of 100 mM Potassium Phosphate (pH6.5), 1 mM 1-Chloro-2,4-dinitrobenzene (CDNB), 130 nM (8 µg) purified His-GFZF truncations, and varying amounts of glutathione were assayed for absorbance at 340 nm every 20 seconds for 10 minutes. Automated readings were taken on a Pharmacia Biotech Ultrospec 3000. The slope for the linear part of the

curve (typically 0-240 seconds) was taken as the initial velocity. R was used to generate the graph and determine the V_{max} and K_m using non-linear least squares regression.

Chapter 4: Additional Considerations and Future Directions

The value of studying M1BP has been manifested in the previous chapters. Future studies will undoubtedly further elucidate the mechanisms of M1BP's function in RP gene transcription and activation with GFZF. Below, I discuss other potential cellular roles for M1BP, describe preliminary experimental results, and speculate about the significance of M1BP's additional putative functions.

Nucleosome Clearance from Promoters by M1BP and NURF

The immobilized template pulldown experiment I performed and described in Chapter 3 (Figure 3-1) identified Putzig, GFZF, and CG3995 as factors that interact with promoter sequences in a Motif 1-dependent manner. Previous reports demonstrated that Putzig exists in a complex with TRF2 (Hochheimer *et al*, 2002). The work I presented in Chapter 2 demonstrates that TRF2 interacts with M1BP (Baumann & Gilmour, 2017). Thus, a biochemical connection between Putzig and M1BP could exist via both factors' interactions with TRF2. My results with recombinant factors indicate that TRF2 is able to associate with promoters in an M1BP-dependent manner without Putzig. However, it could be that in cells, M1BP interacts with both TRF2 and Putzig as part of a larger complex. If multiple M1BP interaction sites do exist within a TRF2-Putzig complex, this might enhance the putative recruitment of the complex by M1BP.

Putzig also exists in a complex with the histone remodeler NURF (Kugler *et al*, 2011). NURF was first identified through its interaction with GAGA factor (GAF) and is required for activated transcription *in vitro* (Tsukiyama *et al*, 1994; Tsukiyama & Wu, 1995; Xiao *et al*, 2001). In an attempt to determine whether a connection exists between M1BP and NURF, I

compared the abundance of NURF at M1BP- and GAF-bound gene promoters. To do this, I first called peaks using GAF and M1BP ChIP-exo data and determined which genes had GAF or M1BP peaks within 100 bp of the TSS. Using previously published genome-wide NURF ChIP-chip data (GEO accession: GSE20829), I mapped the levels of NURF enrichment relative to GAF- and M1BP-bound gene TSSs. Interestingly, I observed that NURF is more highly enriched at M1BP-bound gene promoters than GAF-bound gene promoters (Figure 4-1). This is somewhat surprising given the well-studied connection between GAF and NURF. This result suggests that the Putzig-M1BP interaction may direct NURF to M1BP-bound locations to deplete nucleosomes around the TSS. The significant enrichment of NURF at M1BP-bound promoters might account for the high degree of nucleosome depletion around M1BP-bound promoters as observed in (Li & Gilmour, 2013). It should be noted, the Putzig-M1BP interaction appears to be dependent on the presence of DNA, since my immunoprecipitations of M1BP from the same *Drosophila* embryo nuclear extracts failed to detect Putzig (Figure 4-4; immunoprecipitation results further discussed later in this chapter).

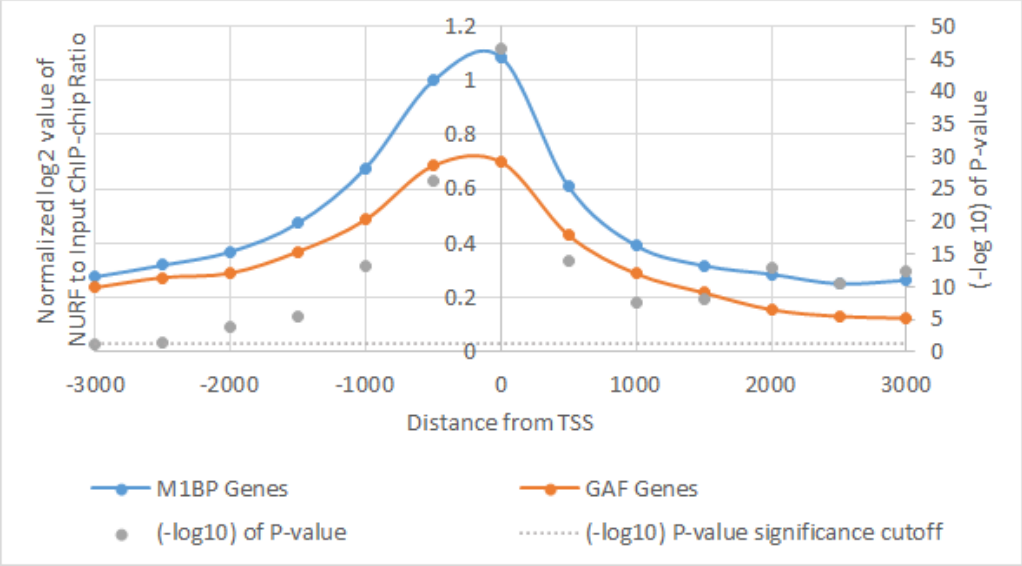

Figure 4-1. NURF is more highly enriched at M1BP-bound promoters than GAF-bound promoters. log2 values of the NURF/Input ChIP-chip ratio were mapped relative to the TSSs of M1BP-bound (n=1741) or GAF-bound (n=921) gene promoters that lack neighboring TSSs within 500 bp. Data were binned in 500 bp regions in accordance with the low-resolution nature of ChIP-chip data. p-values for each binned region (plotted as gray dots) were calculated using a two-tailed T-test assuming equal variance. The negative $\log_{10}$ transformation of each p-value is reported with the dashed line representing the $-\log_{10}$ transformation of the 0.05 statistical significance cutoff.

M1BP and CG3995 as Chromatin Boundary Factors and Genome Organizers

In addition to Putzig and GFZF, the mass spectrometry results from the immobilized template pulldown experiments from nuclear extracts also revealed CG3995 as a putative M1BP interaction partner. Though largely uncharacterized, it has been inferred from sequence analysis that CG3995 possesses a BEAF- and DREF-related (BED) finger DNA binding domain (Aravind, 2000). Like the more extensively studied C2H2 zinc fingers, the BED finger is predicted to chelate divalent cations and recognize specific DNA sequences. According to the protein family (Pfam) database (Finn *et al*, 2016), only 9 *Drosophila* proteins are predicted to

possess BED finger domains. Thus, it is not a particularly common domain. This scarcity might suggest the domain carries out rather specialized functions.

The fact that CG3995 has a BED finger domain is interesting because DREF and the *Drosophila* insulator protein BEAF-32 each possess a BED finger (Alignment shown in Figure 4-2). Notably, DREF and BEAF-32 are the only factors with a BED finger that have been characterized. Based on this connection with DREF and BEAF-32, I hypothesize that CG3995 might function in RP gene transcription and chromatin boundary formation. In support of the first point, I demonstrated in Chapter 2 that DREF is enriched at many RP gene promoters. Furthermore, DREF regulates expression of RP genes in humans (Yamashita et al. 2007). Thus, since CG3995 and DREF represent 2 of the 9 proteins with a BED finger, it seems reasonable to speculate that CG3995 might also contribute to coordinated RP gene expression. One possibility is that CG3995 recognizes Motif 6. Motif 6 is enriched at RP gene promoters (see Figure 2-11), but the factor that recognizes this element remains unknown.

The argument that CG3995 might act as a boundary factor is supported by the fact that BEAF-32 and DREF are known boundary factors (Hart *et al*, 1999; Gilbert *et al*, 2006). The gene regulation field is very interested in boundary elements and factors since they are thought to divide genomes into functionally-related regions. It is believed that partitioning the genome into functionally-related domains drives the precise transcriptional programs necessary for development and differentiation (Pope *et al*, 2014). The available literature suggests the BED finger domain is specialized for binding chromatin boundary elements (Aravind, 2000). If CG3995 does bind a boundary element (putatively Motif 6), then characterizing CG3995 would likely offer novel insights into genome organization and could further elucidate the mechanisms by which transcription is impacted by genome organization.

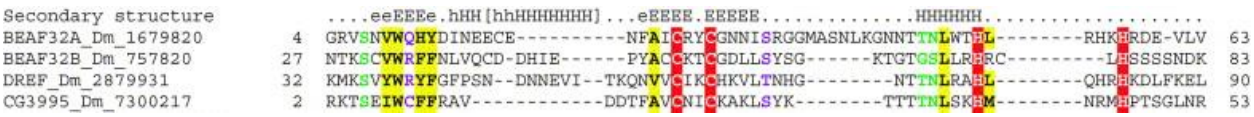

Figure 4-2. BED finger domain sequence alignment.

Sequence alignments of the BED finger domain of BEAF-32A, BEAF-32B, DREF, and CG3995. BEAF-32A and -32B are differentially spliced versions of the same gene that differ in their 80 N-terminal amino acid residues (Hart *et al*, 1997). Both BEAF-32A and -32B are thought to contribute to genome organization (Gilbert *et al*, 2006). I reproduced this figure from (Aravind, 2000). They constructed the alignment by parsing the highest-scoring pairs generated in PSI-BLAST searches, then adjusted the alignment based on secondary-structure prediction for related subsets. Secondary structure was predicted using the PHD program (Rost & Sander, 1993; Rost *et al*, 1994). H/h and E/e above the alignments represent predicted α-helices and β-strands, respectively. Upper case letters denote strong (90%) prediction and lowercase letters denote moderate (72%) prediction. The shading indicates the following: aromatic residues (YFWLIVMA; yellow); small residues (SAGTVPNHD; green); polar residues (STQNEDRKH; purple). The metal-chelating cysteine and histidine residues are shown in red

Determining whether CG3995 binds Motif 6 should be relatively straightforward. Jian Li, a former graduate student in the lab, identified M1BP using DNA affinity chromatography (Li & Gilmour, 2013). Additionally, I've demonstrated that immobilized template pulldowns on a smaller scale can be used to identify factors that interact with DNA templates in a sequence-specific manner (Figure 3-1). Either approach should be suitable for determining the Motif 6 recognition factor. Using my approach, a biotinylated template containing Motif 6, but lacking Motif 1, could be immobilized on streptavidin Dynabeads. Factors that associate specifically with a consensus Motif 6 and not a mutant Motif 6 template could then be pulled out of *Drosophila* embryo nuclear extracts and assessed by SDS-PAGE then identified by mass spectrometry. If CG3995 does bind Motif 6, I would expect CG3995 to appear in the case where wt Motif 6 was used, but not in the case where Motif 6 is mutated. To reproduce the results in a

reconstituted system, CG3995 could be expressed and purified from *E. coli*. Whether purified CG3995 binds a Motif 6-containing template could then be assessed by an immobilized template pulldown or gel-shift assay. It would be interesting and important to test whether the combination of M1BP and CG3995 increases affinity when the binding factors and recognition elements are present either alone or in combination.

M1BP's role in Genome Organization

In addition to M1BP's association with CG3995, several lines of evidence indicate that M1BP contributes to genome organization. First, two recent reports demonstrated that Motif 1 and M1BP are enriched at topologically-associated domains (TAD) boundaries (Hug *et al*, 2017; Ramírez *et al*, 2018). TADs are regions of the genome that often reside in close proximity in the nucleus. This conclusion has been inferred from ligation-based genome-wide sequencing data that shows sequences within TADs interact more frequently than sequences delimited by TAD boundaries (Pope *et al*, 2014). The genes present in regions of the genome encompassed in a single TAD often possess similarities in their transcriptional activity and their associated histone modifications.

As the name suggests, TAD boundaries delineate neighboring TADs. Most TAD boundaries occur at sites of active transcription (Ramírez *et al*, 2018), and are enriched in *Drosophila* for core promoter elements including: Motif 1, the DRE, and Motif 6. Notably, there is a significant enrichment ($p<0.0001$, fisher's exact test) of Motif 1 at promoters containing Motif 6, suggesting that M1BP might facilitate the recruitment of the factor that binds Motif 6. Perhaps the combination of Motif 1 and Motif 6 at promoters drives TAD boundary formation.

This could be interrogated using single-locus or genome-wide assays to evaluate chromatin conformation following deletion or insertion of Motif 1 and Motif 6 at select promoters.

One additional piece of evidence I generated further implicates M1BP playing a role in genome organization. I determined that M1BP interacts with the insulator-bridging factor CP190 (Figure 4-3). Recent studies have shown that the ZAD-Znf proteins ZIPIC and Pita also interact with CP190 (Maksimenko *et al*, 2015). Through additional experiments, the study demonstrates that ZIPIC and Pita are required for proper genome organization and could thus be classified as chromatin boundary factors. As mentioned in Chapter 2, M1BP is also a member of the ZAD-Znf family of transcription factors. This family of transcription factors has undergone a lineage-specific expansion in *Drosophila*, analogous to the expansion of KRAB-Znf or SCAN-Znf factors in humans (Chung *et al*, 2007; Nowick & Stubbs, 2010; Stubbs *et al*, 2011). Given the similarities between these factors and M1BP, I suspect Pita and ZIPIC may bind distal regulatory elements and associate indirectly with M1BP via their common interaction with CP190. In support of this notion, it was demonstrated that different portions of CP190 were responsible for interacting with ZIPIC and Pita (Maksimenko *et al*, 2015). Thus, CP190 may be able to accommodate interactions from multiple ZAD-Znf transcription factors and other boundary factors at the same time. The indirect association of sequence-specific binding factors through CP190 could be a key mechanism that directs genome organization and TAD establishment. These CP190-mediated interactions could also reveal the mechanisms that direct promoter-enhancer interactions. A deeper understanding of the promoter-enhancer interactions that drive housekeeping gene expression would be of value to the field since the majority of enhancer studies have focused on developmental or hormone-regulated genes (MacArthur *et al*, 2009;

Zaret & Carroll, 2011). Thus, it remains unclear to what extent housekeeping genes rely on input from distal regulatory sites.

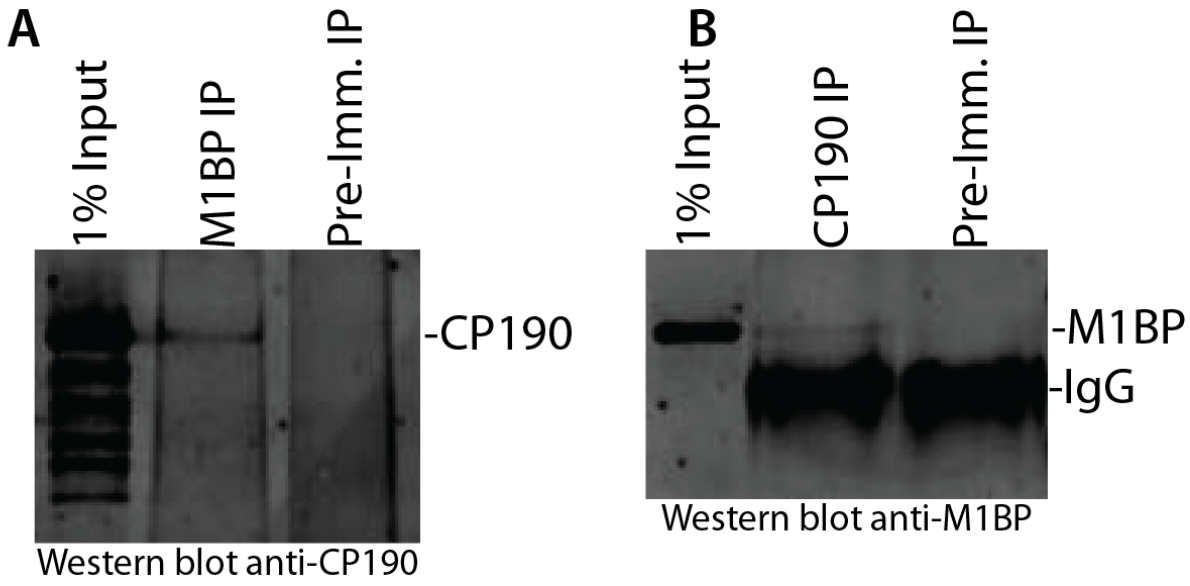

Figure 4-3. M1BP interacts with CP190.

(A) M1BP or Preimmune immunoprecipitates from nuclear extracts were separated by SDS-PAGE and analyzed by probing western blots with antibody against CP190. (B) CP190 or Preimmune immunoprecipitates from nuclear extracts were separated by SDS-PAGE and analyzed by probing western blots with antibody against M1BP.

An open question in the field of genome organization and gene regulation is whether active transcription drives TAD formation or TAD formation directs active transcription. M1BP may offer a unique opportunity to directly investigate this question. M1BP acts as both a transcription activator and, putatively, as a delimiter of TADs. So, if we could eliminate M1BP's ability to associate with CP190 without affecting its ability to activate transcription, we could potentially address this issue. Obviously, there are clear technical obstacles to test such a hypothesis. For one, it may not even be possible to separate the two functions. However, following the lead of (Maksimenko *et al*, 2015), one could quickly determine the region(s) of

M1BP that are responsible for its interaction with CP190. To facilitate such a project, I have cloned, expressed, and purified full-length and truncated versions of M1BP fused to maltose-binding protein. A mutant version of M1BP that is incapable of interacting with CP190 could then be tested for transcription activity *in vitro*. If M1BP's transcription activation and CP190-interaction functions could be separated, then it might be possible to determine whether TAD boundaries are established as a result of active transcription or active transcription ensues after CP190-mediated TAD boundary formation. If such a system could be established, then I could test the chromatin organization of the system by chromatin conformation capture (3C) (Dekker *et al*, 2002) and assess transcriptional activity by qPCR. If TADs arise as a result of active transcription, then TAD boundaries should remain if the wild-type copy of M1BP is swapped with a transcriptionally competent version of M1BP that no longer interacts with CP190 in cells. If these results were not observed, it would suggest that TAD formation is driven by mechanisms independent of transcription.

As a side note, there is one ZAD-Znf present in humans (ZFP276) of unknown function (Wong *et al*, 2003, 2000). Thus, a more complete understanding of ZAD-Znf function could provide useful insight into the function of ZFP276. Interestingly, (Stampfel *et al*, 2015) tested the activities of a number of *Drosophila* transcription factors in mammalian cells. They found that, among the factors tested, the *Drosophila* transcription factors exerted the same functions in mammalian cell lines as they did in *Drosophila* cells. Therefore, understanding the properties of transcription factors that drive regulation in *Drosophila* could offer direct insights about regulatory principles in humans.

M1BP and Spectrin

In addition to immobilized template pulldown experiments, I performed immunoprecipitation (IP) experiments as an orthogonal approach to identify factors that interact with M1BP. Briefly, *Drosophila* embryo nuclear extracts were incubated with antisera raised against M1BP or Preimmune sera, the bound material was washed, eluted, and separated by SDS-PAGE. Finally, the bands that were enriched in the M1BP IP lanes were identified using liquid chromatography-tandem mass spec (LC-MS/MS). These experiments once again identified GFZF as an M1BP interactor. Additionally, I identified α- and β-Spectrin and *Drosophila* Target of Rapamycin (dTOR) as potential M1BP interactors (Figure 4-4).

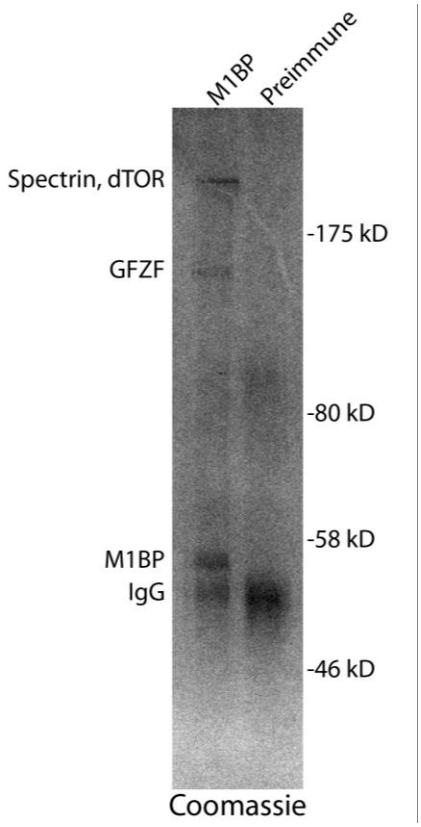

Figure 4-4. M1BP IP identifies GFZF, Spectrin, and *Drosophila* target of rapamycin (dTOR). M1BP and Preimmune sera were used for immunoprecipitations from *Drosophila* embryo nuclear extracts. The samples were analyzed by 8% SDS-PAGE, stained with coomassie, and the identity of the labeled bands excised from the M1BP lane (left) was revealed through LC-MS/MS.

Since Spectrin is traditionally studied for its role in cell structure, my detection of Spectrin interacting with M1BP was unexpected. Nevertheless, I confirmed the results of the mass spec identification by probing M1BP IP western blots with anti-Spectrin antibody. This experiment reveals clear enrichment of Spectrin in the M1BP IP sample compared to the Preimmune IP sample (Figure 4-5, cf. lanes 3 and 4). Spectrin proteins were originally discovered fifty years ago as factors critical for the maintenance of red blood cell structure (Marchesi & Steers, 1968). Since then, studies have identified and described roles for both α- and β-Spectrin in the nucleus (Sridharan *et al*, 2006). For instance, α-Spectrin plays an essential role in DNA interstrand crosslink repair (Brois *et al*, 1999; McMahon *et al*, 1999). In patients with Fanconi's Anemia, the loss of α-Spectrin results in chromosome instability and telomere dysfunction (Sridharan *et al*, 2003; Zhang *et al*, 2010, 2013). Interestingly, M1BP resides at the telomeres of polytene chromosomes (Li & Gilmour, 2013). Thus, the observation that Spectrin also binds telomeres is of particular interest (Zhang *et al*, 2013). It should be noted that Spectrin binding telomeres would need to be confirmed in *Drosophila* as this has not been reported in flies. If Spectrin does interact with telomeres in *Drosophila*, it could be that M1BP recruits Spectrin to telomeres. This recruitment would enable Spectrin to carry out its putative function of recruiting repair proteins to sites of telomere damage (Zhang *et al*, 2013). If such a mechanism did exist, it would demonstrate an additional, non-transcriptional function for M1BP. Further interrogation of the functional significance of this association would be interesting given M1BP's function in genome organization and Spectrin's well-established role in structuring the cell.

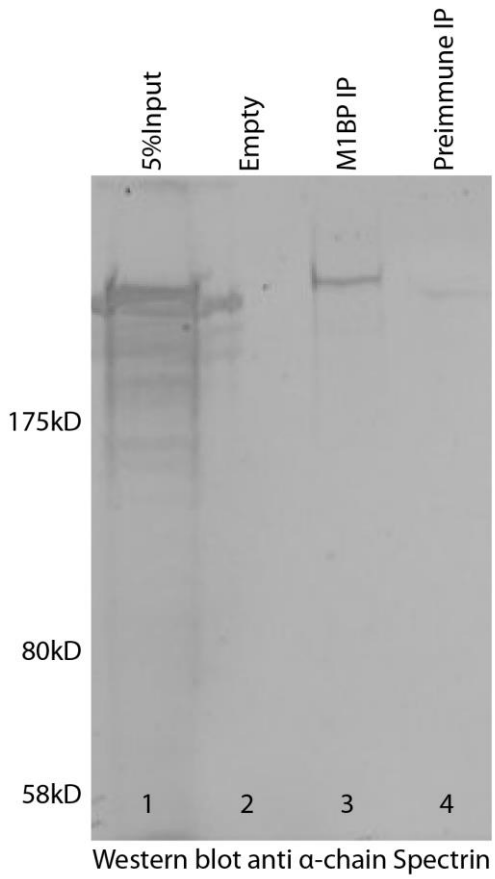

Figure 4-5. M1BP interacts with α-chain Spectrin.

M1BP and Preimmune immunoprecipitates (lanes 3 and 4) from nuclear extracts were separated by SDS-PAGE and analyzed by probing the western blot with an antibody against α-chain Spectrin.

M1BP and dTOR

The conserved serine/threonine kinase TOR and the pathways dependent upon it have been extensively studied. TOR-dependent signaling pathways are involved in regulating processes such as transcription, ribosome biogenesis, and nutrient shuttling (Hall, 2008). Since M1BP drives the expression of housekeeping genes that carry out such functions, it would make a great deal of sense for TOR to target M1BP. I attempted to confirm the mass spectrometry results by immunoprecipitating M1BP then probing for TOR by western blot, but I was unable to detect TOR. Thus, confirming the biochemical interaction between M1BP and TOR should

precede a more careful interrogation of any putative functional connection between the two factors. Nevertheless, a potential biological connection between the factors is logical and begs for further exploration. One intriguing possibility is that TOR phosphorylates M1BP to control M1BP function in cells. As an example, phosphorylation of M1BP by TOR might target M1BP to the nucleus or otherwise affect its subcellular localization. Alternatively, phosphorylation of M1BP might alter M1BP's affinity for other factors (i.e. GFZF or TRF2) that contribute to transcription of M1BP-bound genes. It is also possible that phosphorylation alters M1BP's affinity for Motif 1. However, the fact recombinant M1BP purified from *E. coli* is able to bind Motif 1 (Li & Gilmour, 2013; Baumann & Gilmour, 2017) argues against this last possibility. If the interaction between M1BP and TOR were confirmed, determining the biological implications of the interaction would be interesting since it links a master regulator of housekeeping gene transcription (M1BP) with a signaling kinase best known for its role in cell growth, cell survival, and nutrient signaling (TOR) (Kennedy & Lamming, 2016).

Speculation on Biomolecular Condensate Formation for Coordinate Transcription

Recently, biomolecular condensates (also known as liquid-liquid phase separations) have vaulted into the scientific communities' collective awareness. Pioneers in the field have concluded that biomolecular condensates drive the formation of membrane-less organelles (Banani *et al*, 2017). They and others also postulate that biomolecular condensates can account for poorly understood phenomena such as transcriptional bursting and nuclear organization (Banani *et al*, 2017; Hnisz *et al*, 2017). Condensate formation is driven by the physical properties of RNA and proteins possessing high valency. Proteins with low-complexity, intrinsically-disordered regions (IDRs) exhibit this high valency. The C-terminal domain (CTD) of Pol II is a

notable example of an IDR. The fact that the CTD is an IDR has led some to speculate that transcriptional bursting is mediated through the multivalent interactions of the CTD, which might be regulated by specific phospho-isoforms of the CTD (Hnisz *et al*, 2017). This is an intriguing notion and is likely being tested by other labs. However, this might not explain how gene classes are differentially regulated. How, for example, are RP genes so highly and coordinately expressed? Can biomolecular condensates provide a rational hypothesis? I think one potentially valid explanation for their coordinate regulation is that the factors bound to RP genes drive the formation of highly specific biomolecular condensates.

There is a key characteristic of TRF2 that may drive biomolecular condensate formation and ultimately lead to coordinate transcription of RP genes. Specifically, *Drosophila* TRF2 has a long N-terminal region that is predicted to be highly disordered. Notably, it is predicted to be even more highly disordered than the CTD of Pol II (Figure 4-6). Long IDRs are thought to predispose factors to form or enter liquid-liquid phase-separated states (Banani *et al*, 2017). Thus, it seems possible that this long IDR of TRF2 could promote the formation of biomolecular condensates. In such a scenario, the unique combination of M1BP with TRF2 might confer molecular specificity that allows other RP gene promoters to coalesce and drive robust transcription in a coordinated fashion. As discussed in Chapter 2, STARR-seq data shows that RP gene promoters serve as enhancers of the RpS12 gene which suggests that RP gene promoters interact to modulate the transcription of other RP genes. This spatial proximity further strengthens the argument that RP genes are coalescing in nuclear space. It is interesting to think that this coalescence might be driven by biomolecular condensates.

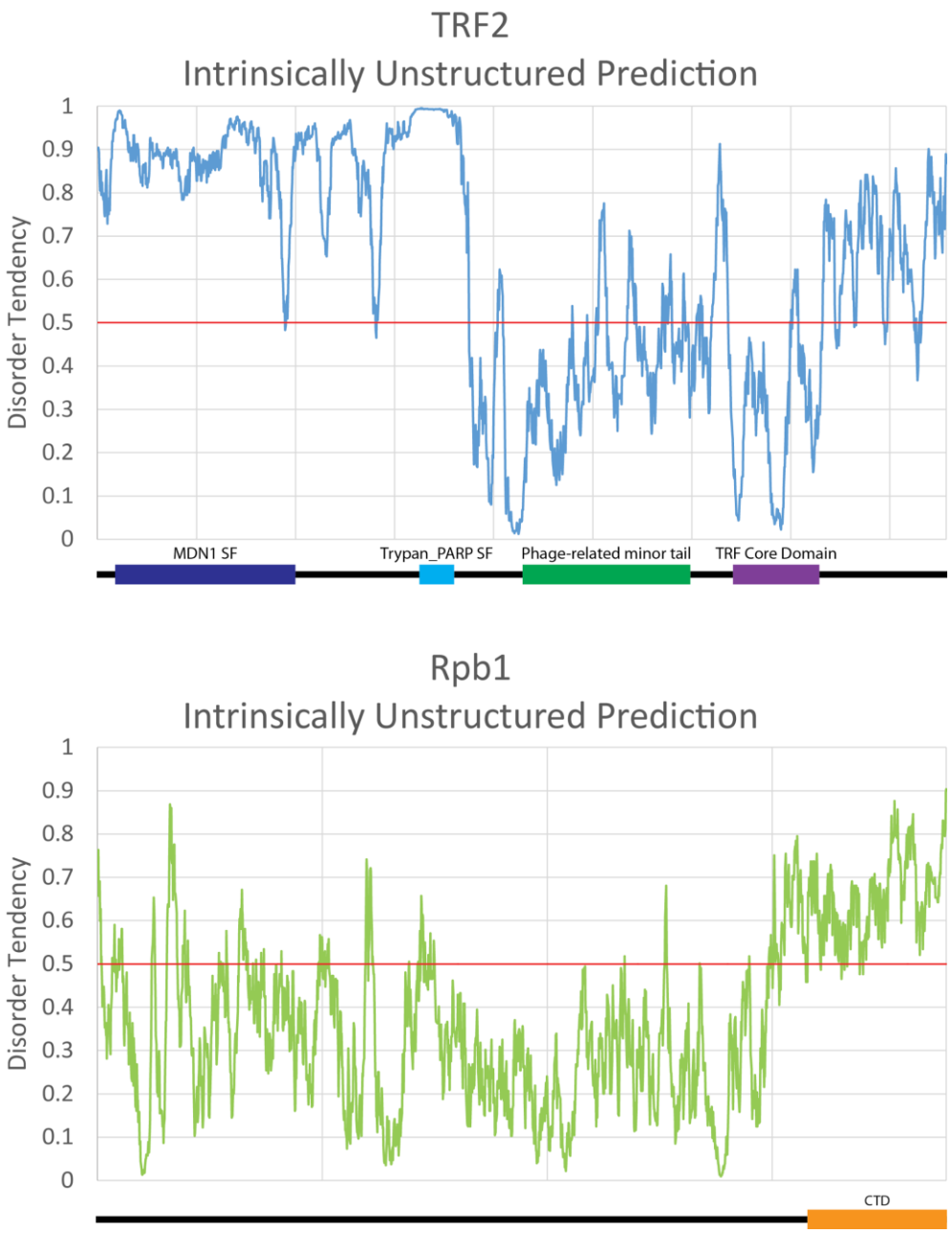

Figure 4-6. The TRF2 N-terminal domain is predicted to be highly disordered.

Sequences corresponding to TRF2 and Rpb1 were fed into the IUPred server (Dosztányi *et al*, 2005b, 2005a) and the outputs were plotted in Excel. For reference, a graphical representation of each protein with predicted or known domains is included below each chart. Only the CTD of Pol II is delimited in the Rpb1 graphic. According to IUPred, scores above 0.5 indicate disorder. A red line at 0.5 has thus been included to highlight this cutoff.

Biochemical experiments could be performed with purified TRF2 and M1BP to see whether they are able to form droplets or condensates *in vitro*. Notably, in my experiments and in the experiments that the Kadonaga lab performed, the short isoform that lacks this N-terminal extension was used. This isoform is able to rescue *in vitro* transcription activity following TRF2 depletion (Wang *et al*, 2014) and can interact with M1BP (Baumann & Gilmour, 2017). It would be interesting to test whether RP gene transcription is affected in cells when the N-terminal IDR is removed or replaced with a well-folded domain. These early explorations could initially rely upon genome-wide ChIP and RNA-seq methods, but ultimately fluorescence microscopy methods will be required to observe the dynamic behavior of these putative droplets in single cells. Using microscopy techniques, one could also test the hypothesis that the unique combination of M1BP or DREF and TRF2 at RP gene promoters enables the coalescence and coordinate transcription of RP genes through phase-separated states. Super-resolution microscopy measurements of fluorescently-labeled TRF2 or M1BP could be used in conjunction with RNA-FISH and oligopaint FISH or related sequence-specific genome localization techniques (Beliveau *et al*, 2015). RNA-FISH enables the quantification and localization of specific mRNAs. Thus, RNA-FISH could be used to quantify the abundance and location of RP gene transcripts in the nucleus. Oligopaint FISH enables the high-resolution localization of genomic regions using sequence-specific DNA probes. Thus, oligopaint FISH could be used to determine whether RP gene promoters and TRF2 coalesce in a manner predicted by phase separation. So far, oligopaint FISH has only successfully been used to identify the location of repetitive DNA sequence elements (Beliveau *et al*, 2012). However, recent advances using catalytically inactive Cas9/CRISPR have increased the detection sensitivity of techniques aimed at visualizing targeted genomic loci (Ma *et al*, 2015; Shechner *et al*, 2015; Shao *et al*, 2016; Qin

et al, 2017; Neguembor *et al*, 2017). These advances and continued future improvements should enable the detection and super-resolution localization of non-repetitive genomic sequences.

Additional Considerations on M1BP Function

Another interesting question to be addressed with regards to M1BP-bound genes is whether or to what extent these genes are driven by enhancers. While it would make sense to have housekeeping gene transcription regulation independent of, or less-dependent on, signals from enhancers, there is evidence that suggests housekeeping gene promoters do receive signals from regions scattered throughout the genome (Zabidi *et al*, 2015). As demonstrated in Chapter 2, RP genes are co-regulated in a manner that might be explained by RP gene promoters acting as enhancers of other RP genes. This is evidenced by the fact that most RP genes promoters serve to activate transcription arising from the *RpS12* gene promoter (Baumann & Gilmour, 2017). Additionally, more recent studies suggest that there are classes of transcription factors that act at enhancers to activate or repress transcription from the *RpS12* promoter (Stampfel *et al*, 2015). This study identified 15 different classes of transcription factors with unique regulatory properties. Of the fifteen classes, four classes were found to enhance transcription from distal regulatory elements while one class was found to be repressive. Notably, one class was found to enhance transcription from the *RpS12* promoter, but was repressive for all other developmental promoters tested. The factors in this particular class belong to the ZAD-Znf family. As noted previously, M1BP is part of the ZAD-Znf class of transcription factors (Chung *et al*, 2007). This question of enhancer-mediated activation of M1BP-bound genes is likely intimately related to M1BP's function in genome organization. The fact that many ZAD-Znf factors interact with

CP190 suggests they might form a unique assembly required to drive constitutive expression of housekeeping genes.

Since its initial report in 2013 (Li & Gilmour, 2013), there has been considerable interest in understanding M1BP's function in transcription initiation (Baumann & Gilmour, 2017), pausing (Zouaz *et al*, 2017), and genome organization (Ramírez *et al*, 2018). M1BP's contributions to these and other fundamental cellular processes are still likely underappreciated or unknown. Thus, the further study of M1BP is likely to shed additional light on these processes. M1BP and Motif 1 are critical to our understanding of many nuclear processes in *Drosophila* and could serve as valuable model to study genome organization, nucleosome depletion, and biomolecular condensate-driven transcription of gene networks.

Materials and Methods

Peak Calling

The GEM (Guo *et al*, 2012) or MultiGPS (Mahony *et al*, 2014) peak callers were used to call GAF or M1BP peaks (respectively) using the default settings with Pre-immune and No-IgG ChIP-exo libraries serving as control samples. Genes having a GAF or M1BP peak within 100 bp of the TSS were identified and used for the NURF mapping.

NURF Analysis

NURF ChIP-chip data was downloaded from the GEO accession: GSE20829. The data I obtained was in bedgraph format with the $\log_2$ value of NURF ChIP/Input serving as the score in the fourth column. I used HOMER's annotatePeaks function to map the $\log_2$ NURF values relative to the M1BP- and GAF-bound gene TSSs.

Preparation of Protein A Magnetic Sepharose Beads

50 µl of protein A magnetic sepharose resin was collected and washed three times with 500 µl 180 mM HEMG (180 mM KCl, 25 mM HEPES pH 7.6, 12.5 mM $MgCl_2$, 0.1 mM EDTA pH 7.9, 10% glycerol). Each wash was performed at room temperature for 5 minutes with constant rotation. The resin was suspended in 250 µl 180 mM HEMG. 20 µl of M1BP, CP190, or pre-immune antisera was then added and the mixture was incubated for 2 hours at room temperature with constant rotation. The resin was rinsed with 180 mM HEMG three times as before and transferred to a new tube. The antibody-bound resin was rinsed three times with 500 µl 200 mM sodium borate (pH 9.0). The antibody was then crosslinked to the resin by adding 100 µl of 20 mM dimethyl pimelimidate in 200 mM sodium borate (pH 9.0) and incubating the mixture for 30 minutes at room temperature with constant rotation. The crosslinking reaction was stopped by decanting the solution and rinsing the resin with 100 µl 200 mM ethanolamine. Another 100 µl of 200 mM ethanolamine was added and the resin was incubated for 2 hours at room temperature to quench all crosslinking. Finally, the antibody-conjugated resin was washed three times with 180 mM HEMG as before. 0.02% sodium azide was added if the resin was to be stored for an extended period.

Co-Immunoprecipitation Experiments

Nuclear extracts from 0-12 hour *Drosophila* embryos were pre-cleared by incubating 1 ml nuclear extract with 250 µl sepharose CL-6B resin for 1 hour at 4°C with constant rotation. 250 µl of the pre-cleared nuclear extract was added to 10 µl (dry volume) of antibody-conjugated resin (preparation described above). Before the IP, the antibody-conjugated resins were pre-eluted two times by incubation with 500 µl 100 mM Glycine (pH 2.5) for 15 minutes at room temperature with constant rotation. The pre-cleared extract and pre-eluted resins were incubated

overnight at 4°C with constant rotation. The following day the resin was washed at room temperature five times for five minutes per wash using 1 ml 180 mM HEMG. The resin was washed one additional time with 180 mM HEMG with a HEPES concentration of 5 mM instead of 25 mM. This was done in order to reduce the buffering capacity of the solution prior to eluting with glycine at low pH. Before elution, the resin was transferred to a new tube to reduce non-specific background arising from proteins that might be stuck to the walls of the tube. To elute, the resin was incubated with 30 µl 100 mM Glycine (pH 2.5) for five minutes. The eluted material (and, in some cases, the unbound material) was subjected to 8% SDS-PAGE. For mass spectrometry analysis, the bands present in the M1BP IP lane, but not the Mock (pre-immune) IP lane were excised and submitted to the Penn State core facility for identification by LC-MS/MS. For western blot analysis, the material separated by SDS-PAGE was transferred to nitrocellulose membranes and probed with antibody against either α-chain Spectrin (1:50,000 dilution) or CP190 (1:10,000 dilution). α-chain Spectrin antibody was a gift from Claire Thomas. CP190 antibody was a gift from Victor Corces.

References

Adamczewski JP, Rossignol M, Tassan JP, Nigg EA, Moncollin V & Egly JM (1996) MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH. *EMBO J.* **15:** 1877–1884

Adelman K & Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13:** 720–731

Adler V, Yin Z, Fuchs SY, Benezra M, Rosario L, Tew KD, Pincus MR, Sardana M, Henderson CJ, Wolf CR, Davis RJ & Ronai Z (1999) Regulation of JNK signaling by GSTp. *EMBO J.* **18:** 1321–1334

Ambrus AM, Rasheva VI, Nicolay BN & Frolov MV (2009) Mosaic genetic screen for suppressors of the de2f1 mutant phenotype in Drosophila. *Genetics* **183:** 79–92

Ames BN, Profet M & Gold LS (1990) Dietary pesticides (99.99% all natural). *Proc. Natl. Acad. Sci. U. S. A.* **87:** 7777–7781

Aravind L (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25:** 421–423

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M & Stark A (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339:** 1074–1077

Ashton-Beaucage D, Udell CM, Gendron P, Sahmi M, Lefrançois M, Baril C, Guenier A-S, Duchaine J, Lamarre D, Lemieux S & Therrien M (2014) A functional screen reveals an extensive layer of transcriptional and splicing control underlying RAS/MAPK signaling in Drosophila. *PLoS Biol.* **12:** e1001809

Athanikar JN, Badge RM & Moran JV (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32:** 3846–3855

Bachtrog D (2008) Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in Drosophila. *Genetics* **180:** 1123–1129

Bagby S, Kim S, Maldonado E, Tong KI, Reinberg D & Ikura M (1995) Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. *Cell* **82:** 857–867

Banani SF, Lee HO, Hyman AA & Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18:** 285–298

Barbash DA (2010) Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. *Genetics* **184:** 313–316

Bártfai R, Balduf C, Hilton T, Rathmann Y, Hadzhiev Y, Tora L, Orbán L & Müller F (2004) TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol.* **14:** 593–598

Barth TK, Schade GOM, Schmidt A, Vetter I, Wirth M, Heun P, Thomae AW & Imhof A (2014) Identification of novel Drosophila centromere-associated proteins. *Proteomics* **14:** 2167–2178

Baumann DG, Dai M-S, Lu H & Gilmour DS (2017) GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, is a transcriptional co-activator. *Mol. Cell. Biol.* Available at: http://dx.doi.org/10.1128/MCB.00476-17

Baumann DG & Gilmour DS (2017) A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. *Nucleic Acids Res.* Available at: https://academic.oup.com/nar/article/doi/10.1093/nar/gkx676/4080187/A-sequence-specific-core-promoter-binding [Accessed August 16, 2017]

Beliveau BJ, Boettiger AN, Avendaño MS, Jungmann R, McCole RB, Joyce EF, Kim-Kiselak C, Bantignies F, Fonseka CY, Erceg J, Hannan MA, Hoang HG, Colognori D, Lee JT, Shih WM, Yin P, Zhuang X & Wu C-T (2015) Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6:** 7147

Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard J-M & Wu C-T (2012) Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 21301–21306

Berhane K, Widersten M, Engström A, Kozarich JW & Mannervik B (1994) Detoxication of base propenals and other alpha, beta-unsaturated aldehyde products of radical reactions and lipid peroxidation by human glutathione transferases. *Proc. Natl. Acad. Sci. U. S. A.* **91:** 1480–1484

Biggin MD & Tjian R (1988) Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* **53:** 699–711

Breathnach R & Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–383

Brigelius-Flohé R & Flohé L (2011) Basic principles and emerging concepts in the redox control of transcription factors. *Antioxid. Redox Signal.* **15:** 2335–2381

Brois DW, McMahon LW, Ramos NI, Anglin LM, Walsh CE & Lambert MW (1999) A deficiency in a 230 kDa DNA repair protein in fanconi anemia complementation group A cells is corrected by the FANCA cDNA. *Carcinogenesis* **20:** 1845–1853

Brown SJ, Cole MD & Erives AJ (2008) Evolution of the holozoan ribosome biogenesis regulon. *BMC Genomics* **9:** 442

Buratowski S, Hahn S, Guarente L & Sharp PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56:** 549–561

Buratowski S, Hahn S, Sharp PA & Guarente L (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334:** 37–42

Burke TW & Kadonaga JT (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10:** 711–724

Bushnell DA, Westover KD, Davis RE & Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303:** 983–988

Cai SP, Eng B, Francombe WH, Olivieri NF, Kendall AG, Waye JS & Chui DH (1992) Two novel beta-thalassemia mutations in the 5’ and 3' noncoding regions of the beta-globin gene. *Blood* **79:** 1342–

1346

Carey MF, Peterson CL & Smale ST (2013) The primer extension assay. *Cold Spring Harb. Protoc.* **2013:** 164–173

Cavallini B, Faus I, Matthes H, Chipoulet JM, Winsor B, Egly JM & Chambon P (1989) Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. *Proc. Natl. Acad. Sci. U. S. A.* **86:** 9803–9807

Chatterjee RN, Chatterjee P, Pal A & Pal-Bhadra M (2007) Drosophila simulans Lethal hybrid rescue mutation (Lhr) rescues inviable hybrids by restoring X chromosomal dosage compensation and causes fluctuating asymmetry of development. *J. Genet.* **86:** 203–215

Chen W & Struhl K (1985) Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. *EMBO J.* **4:** 3273–3280

Chen X, Hiller M, Sancak Y & Fuller MT (2005) Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. *Science* **310:** 869–872

Chiang CM, Ge H, Wang Z, Hoffmann A & Roeder RG (1993) Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. *EMBO J.* **12:** 2749–2762

Chong JA, Moran MM, Teichmann M, Kaczmarek JS, Roeder R & Clapham DE (2005) TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Mol. Cell. Biol.* **25:** 2632–2643

Chung H-R, Löhr U & Jäckle H (2007) Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol. Biol. Evol.* **24:** 1934–1943

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M & Zaret KS (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9:** 279–289

Cisse II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, Dugast-Darzacq C, Hajj B, Dahan M & Darzacq X (2013) Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341:** 664–667

Coin F, Oksenych V & Egly J-M (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* **26:** 245–256

Coleman RA, Taggart AK, Burma S, Chicca JJ 2nd & Pugh BF (1999) TFIIA regulates TBP and TFIID dimers. *Mol. Cell* **4:** 451–457

Conaway RC, Garrett KP, Hanley JP & Conaway JW (1991) Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A.* **88:** 6205–6209

Concino MF, Lee RF, Merryweather JP & Weinmann R (1984) The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. *Nucleic Acids Res.* **12:** 7423–7433

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C & Chambon P (1980) Promoter sequences of eukaryotic protein-coding genes. *Science* **209:** 1406–1414

Crowley TE, Hoey T, Liu JK, Jan YN, Jan LY & Tjian R (1993) A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. *Nature* **361:** 557–561

Cubeñas-Potts C, Rowley MJ, Lyu X, Li G, Lei EP & Corces VG (2016) Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45:** 1714–1730

Cuesta I, Zaret KS & Santisteban P (2007) The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol. Cell. Biol.* **27:** 7302–7314

Dai M-S, Sun X-X, Qin J, Smolik SM & Lu H (2004) Identification and characterization of a novel Drosophila melanogaster glutathione S-transferase-containing FLYWCH zinc finger protein. *Gene* **342:** 49–56

Danino YM, Even D, Ideses D & Juven-Gershon T (2015) The core promoter: At the heart of gene expression. *Biochim. Biophys. Acta* **1849:** 1116–1131

Deato MDE & Tjian R (2007) Switching of the core transcription machinery during myogenesis. *Genes Dev.* **21:** 2137–2149

Dekker J, Rippe K, Dekker M & Kleckner N (2002) Capturing chromosome conformation. *Science* **295:** 1306–1311

Dierks P, van Ooyen A, Cochran MD, Dobkin C, Reiser J & Weissmann C (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* **32:** 695–706

Dosztányi Z, Csizmók V, Tompa P & Simon I (2005a) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347:** 827–839

Dosztányi Z, Csizmok V, Tompa P & Simon I (2005b) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21:** 3433–3434

Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ & Lis JT (2016) Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev.* **30:** 1731–1746

Duttke SHC, Doolittle RF, Wang Y-L & Kadonaga JT (2014) TRF2 and the evolution of the bilateria. *Genes Dev.* **28:** 2071–2076

Dynlacht BD, Hoey T & Tjian R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66:** 563–576

Emanuel PA & Gilmour DS (1993) Transcription factor TFIID recognizes DNA sequences downstream of the TATA element in the Hsp70 heat shock gene. *Proc. Natl. Acad. Sci. U. S. A.* **90:** 8449–8453

Feaver WJ, Gileadi O, Li Y & Kornberg RD (1991) CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67:** 1223–1230

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M,

Sangrador-Vegas A, Salazar GA, Tate J & Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44:** D279–85

FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B & Vinson C (2006) Comparative genomics of Drosophila and human core promoters. *Genome Biol.* **7:** R53

Flanagan PM, Kelleher RJ 3rd, Sayre MH, Tschochner H & Kornberg RD (1991) A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350:** 436–438

Flores O, Lu H & Reinberg D (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.* **267:** 2786–2793

Flores O, Maldonado E & Reinberg D (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.* **264:** 8913–8921

Forget D, Langelier M-F, Thérien C, Trinh V & Coulombe B (2004) Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell. Biol.* **24:** 1122–1131

Funk JD, Nedialkov YA, Xu D & Burton ZF (2002) A Key Role for the α1 Helix of Human RAP74 in the Initiation and Elongation of RNA Chains. *J. Biol. Chem.* **277:** 46998–47003

Ge H & Roeder RG (1994a) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J. Biol. Chem.* **269:** 17136–17140

Ge H & Roeder RG (1994b) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78:** 513–523

Ghazy MA, Brodie SA, Ammerman ML, Ziegler LM & Ponticelli AS (2004) Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol. Cell. Biol.* **24:** 10975–10985

Ghosh SKB, Missra A & Gilmour DS (2011) Negative elongation factor accelerates the rate at which heat shock genes are shut off by facilitating dissociation of heat shock factor. *Mol. Cell. Biol.* **31:** 4232–4243

Gilbert MK, Tan YY & Hart CM (2006) The Drosophila boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics* **173:** 1365–1375

Gilchrist D a., Nechaev S, Lee C, Ghosh SKB, Collins JB, Li L, Gilmour DS & Adelman K (2008) NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* **22:** 1921–1933

Goldberg ML (1979) Sequence analysis of Drosophila histone genes.

Gonzalez I, Mateos-Langerak J, Thomas A, Cheutin T & Cavalli G (2014) Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. *Mol. Cell* **54:** 485–499

Goodrich JA, Hoey T, Thut CJ, Admon A & Tjian R (1993) Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75:** 519–530

Gouge J, Satia K, Guthertz N, Widya M, Thompson AJ, Cousin P, Dergai O, Hernandez N & Vannini A (2015) Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell* **163:** 1375–1387

Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, Falls K, Goodman JL, Hu Y, Ponting L, Schroeder AJ, Strelets VB, Thurmond J, Zhou P & the FlyBase Consortium (2017) FlyBase at 25: looking to the future. *Nucleic Acids Res.* **45:** D663–D671

Grant CE, Bailey TL & Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27:** 1017–1018

Grosschedl R & Birnstiel ML (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **77:** 1432–1436

Guo Y, Mahony S & Gifford DK (2012) High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* **8:** e1002638

Guzmán E & Lis JT (1999) Transcription factor TFIIH is required for promoter melting in vivo. *Mol. Cell. Biol.* **19:** 5652–5658

Hahn S, Hoar ET & Guarente L (1985) Each of three 'TATA elements' specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **82:** 8562–8566

Ha I, Roberts S, Maldonado E, Sun X, Kim LU, Green M & Reinberg D (1993) Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* **7:** 1021–1032

Hall MN (2008) mTOR-what does it do? *Transplant. Proc.* **40:** S5–8

Hansen SK, Takada S, Jacobson RH, Lis JT & Tjian R (1997) Transcription Properties of a Cell Type–Specific TATA-Binding Protein, TRF. *Cell* **91:** 71–83

Harlen KM & Churchman LS (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18:** 263–273

Hart CM, Cuvier O & Laemmli UK (1999) Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma* **108:** 375–383

Hart CM, Zhao K & Laemmli UK (1997) The scs' boundary element: characterization of boundary element-associated factors. *Mol. Cell. Biol.* **17:** 999–1009

Hatta M & Cirillo LA (2007) Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. *J. Biol. Chem.* **282:** 35583–35593

Hayes JD, Flanagan JU & Jowsey IR (2005) Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **45:** 51–88

Hayes JD & Pulford DJ (1995) The Glut athione S-Transferase Supergene Family: Regulation of GST and the Contribution of the lsoenzymes to Cancer Chemoprotection and Drug Resistance Part II.

Crit. Rev. Biochem. Mol. Biol. **30:** 521–600

Heintzman ND & Ren B (2007) The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* **64:** 386–400

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38:** 576–589

He Y, Fang J, Taatjes DJ & Nogales E (2013) Structural visualization of key steps in human transcription initiation. *Nature* **495:** 481–486

Hiller MA, Lin TY, Wood C & Fuller MT (2001) Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.* **15:** 1021–1030

Hiller M, Chen X, Pringle MJ, Suchorolski M, Sancak Y, Viswanathan S, Bolival B, Lin T-Y, Marino S & Fuller MT (2004) Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* **131:** 5297–5308

Hnisz D, Shrinivas K, Young RA, Chakraborty AK & Sharp PA (2017) A Phase Separation Model for Transcriptional Control. *Cell* **169:** 13–23

Hochheimer A, Zhou S, Zheng S, Holmes MC & Tjian R (2002) TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature* **420:** 439

Hoey T, Weinzierl RO, Gill G, Chen JL, Dynlacht BD & Tjian R (1993) Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell* **72:** 247–260

Holmes MC & Tjian R (2000) Promoter-selective properties of the TBP-related factor TRF1. *Science* **288:** 867–870

Holstege FC, Tantin D, Carey M, van der Vliet PC & Timmers HT (1995) The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* **14:** 810–819

Holstege FC, van der Vliet PC & Timmers HT (1996) Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15:** 1666–1677

Hoopes BC, LeBlanc JF & Hawley DK (1992) Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway. *J. Biol. Chem.* **267:** 11539–11547

Ho PJ, Rochette J, Fisher CA, Wonke B, Jarvis MK, Yardumian A & Thein SL (1996) Moderate reduction of beta-globin gene transcript by a novel mutation in the 5' untranslated region: a study of its interaction with other genotypes in two families. *Blood* **87:** 1170–1178

Horn AE, Kugel JF & Goodrich JA (2016) Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity. *Nucleic Acids Res.* **44:** 7132–7143

Huang DW, Sherman BT & Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4:** 44–57

Huang DW, Sherman BT & Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37:** 1–13

Hug CB, Grimaldi AG, Kruse K & Vaquerizas JM (2017) Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169:** 216–228.e19

Hu H & Li X (2007) Transcriptional regulation in eukaryotic ribosomal protein genes. *Genomics* **90:** 421–423

Hulsen T, de Vlieg J & Alkema W (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9:** 488

Imbalzano AN, Zaret KS & Kingston RE (1994) Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol. Chem.* **269:** 8280–8286

Ioshikhes IP, Albert I, Zanton SJ & Pugh BF (2006) Nucleosome positions predicted through comparative genomics. *Nat. Genet.* **38:** 1210–1215

Isogai Y, Keles S, Prestel M, Hochheimer A & Tjian R (2007) Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21:** 2936–2949

Iwafuchi-Doi M & Zaret KS (2014) Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28:** 2679–2692

Jallow Z, Jacobi UG, Weeks DL, Dawid IB & Veenstra GJC (2004) Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 13525–13530

Jeronimo C & Robert F (2017) The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. *Trends Cell Biol.* **27:** 765–783

Joo YJ, Ficarro SB, Soares LM, Chun Y, Marto JA & Buratowski S (2017) Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes Dev.* **31:** 2162–2174

Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ & Dickerson RE (1996) How proteins recognize the TATA box. *J. Mol. Biol.* **261:** 239–254

Kakuta Y, Usuda K, Nakashima T, Kimura M, Aso Y & Yamamoto K (2011) Crystallographic survey of active sites of an unclassified glutathione transferase from Bombyx mori. *Biochim. Biophys. Acta* **1810:** 1355–1360

Kamada K, Goto S, Okunaga T, Ihara Y, Tsuji K, Kawai Y, Uchida K, Osawa T, Matsuo T, Nagata I & Kondo T (2004) Nuclear glutathione S-transferase π prevents apoptosis by reducing the oxidative stress-induced formation of exocyclic dna products. *Free Radical Biology and Medicine* **37:** 1875–1884

Kamenova I, Warfield L & Hahn S (2014) Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. *Mol. Cell. Biol.* **34:** 2929–2943

Kedmi A, Zehavi Y, Glick Y, Orenstein Y, Ideses D, Wachtel C, Doniger T, Waldman Ben-Asher H, Muster N, Thompson J, Anderson S, Avrahami D, Yates JR, Shamir R, Gerber D & Juven-Gershon T (2014) Drosophila TRF2 is a preferential core promoter regulator. *Genes Dev.* **28:** 2163–2174

Kelleher RJ 3rd, Flanagan PM & Kornberg RD (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61:** 1209–1215

Kennedy BK & Lamming DW (2016) The Mechanistic Target of Rapamycin: The Grand ConducTOR of Metabolism and Aging. *Cell Metab.* **23:** 990–1003

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002) The human genome browser at UCSC. *Genome Res.* **12:** 996–1006

Kim JL & Burley SK (1994) 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1:** 638–653

Kim J & Shapiro DJ (1996) In simple synthetic promoters YY1-induced DNA bending is important in transcription activation and repression. *Nucleic Acids Res.* **24:** 4341–4348

Kim Y, Geiger JH, Hahn S & Sigler PB (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365:** 512–520

Kim YJ, Björklund S, Li Y, Sayre MH & Kornberg RD (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77:** 599–608

Knutson A, Castaño E, Oelgeschläger T, Roeder RG & Westin G (2000) Downstream promoter sequences facilitate the formation of a specific transcription factor IID-promoter complex topology required for efficient transcription from the megalin/low density lipoprotein receptor-related protein 2 promoter. *J. Biol. Chem.* **275:** 14190–14197

Kokubo T, Swanson MJ, Nishikawa JI, Hinnebusch AG & Nakatani Y (1998) The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* **18:** 1003–1012

Koleske AJ & Young RA (1994) An RNA polymerase II holoenzyme responsive to activators. *Nature* **368:** 466–469

Koleske AJ & Young RA (1995) The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20:** 113–116

Kondo S & Perrimon N (2011) A genome-wide RNAi screen identifies core components of the G_2 -M DNA damage checkpoint. *Sci. Signal.* **4:** rs1

Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF & Korber P (2016) Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167:** 709–721.e12

Kuehner JN & Brow DA (2006) Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* **281:** 14119–14128

Kugler SJ, Gehring E-M, Wallkamm V, Krüger V & Nagel AC (2011) The Putzig-NURF nucleosome remodeling complex is required for ecdysone receptor signaling and innate immunity in Drosophila melanogaster. *Genetics* **188:** 127–139

Kuras L, Kosa P, Mencia M & Struhl K (2000) TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. *Science* **288:** 1244–1248

Kwak H, Fuda NJ, Core LJ & Lis JT (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339:** 950–953

Lavoie L, Tremblay A & Mirault ME (1992) Distinct oxidoresistance phenotype of human T47D cells transfected by rat glutathione S-transferase Yc expression vectors. *J. Biol. Chem.* **267:** 3632–3636

Letunic I & Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44:** W242–5

Lewis BA, Kim T-K & Orkin SH (2000) A downstream element in the human β-globin promoter: Evidence of extended sequence-specific transcription factor IID contacts. *Proceedings of the National Academy of Sciences* **97:** 7172–7177

Lewis ED & Manley JL (1985) Control of adenovirus late promoter expression in two human cell lines. *Mol. Cell. Biol.* **5:** 2433–2442

Li H-M, Buczkowski G, Mittapalli O, Xie J, Wu J, Westerman R, Schemerhorn BJ, Murdock LL & Pittendrigh BR (2008) Transcriptomic profiles of Drosophila melanogaster third instar larval midgut and responses to oxidative stress. *Insect Mol. Biol.* **17:** 325–339

Li J & Gilmour DS (2013) Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J.* **32:** 1829–1841

Li L, Edgar B a. & Grewal SS (2010) Nutritional control of gene expression in Drosophila larvae via TOR, Myc and a novel cis-regulatory element. *BMC Cell Biol.* **11:** 7

Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong C-T, Cubeñas-Potts C, Hu M, Lei EP, Bosco G, Qin ZS & Corces VG (2015) Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol. Cell* **58:** 216–231

Lim CY, Santoso B, Boulay T, Dong E, Ohler U & Kadonaga JT (2004) The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* **18:** 1606–1617

Lin YC, Choi WS & Gralla JD (2005) TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat. Struct. Mol. Biol.* **12:** 603–607

Listowsky I, Abramovitz M, Homma H & Niitsu Y (1988) Intracellular binding and transport of hormones and xenobiotics by glutathione-S-transferases. *Drug Metab. Rev.* **19:** 305–318

Li X, Zhong S & Wong WH (2005) Reliable prediction of transcription factor binding sites by phylogenetic verification. *Proc. Natl. Acad. Sci. U. S. A.* **102:** 16945–16950

Li Y, Flanagan PM, Tschochner H & Kornberg RD (1994) RNA polymerase II initiation factor interactions and transcription start site selection. *Science* **263:** 805–807

Louder RK, He Y, López-Blanco JR, Fang J, Chacón P & Nogales E (2016) Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531:** 604–609

MacArthur S, Li X-Y, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SVE, Knowles DW, Stapleton M, Bickel P, Biggin MD & Eisen MB (2009) Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10:** R80

Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S & Pederson T (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **112:** 3002–3007

Mahony S, Edwards MD, Mazzoni EO, Sherwood RI, Kakumanu A, Morrison CA, Wichterle H & Gifford DK (2014) An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLoS Comput. Biol.* **10:** e1003501

Maile T, Kwoczynski S, Katzenberger RJ, Wassarman DA & Sauer F (2004) TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science* **304:** 1010–1014

Maksimenko O, Bartkuhn M, Stakhov V, Herold M, Zolotarev N, Jox T, Buxa MK, Kirsch R, Bonchuk A, Fedotova A, Kyrchanova O, Renkawitz R & Georgiev P (2015) Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res.* **25:** 89–99

Maldonado E, Shiekhattar R, Sheldon M, Cho H, Drapkin R, Rickert P, Lees E, Anderson CW, Linn S & Reinberg D (1996) A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* **381:** 86–89

Mannervik B & Danielson UH (1988) Glutathione transferases--structure and catalytic activity. *CRC Crit. Rev. Biochem.* **23:** 283–337

Marchesi VT & Steers E Jr (1968) Selective solubilization of a protein component of the red cell membrane. *Science* **159:** 203–204

Maston GA, Zhu LJ, Chamberlain L, Lin L, Fang M & Green MR (2012) Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *Elife* **1:** e00068

Matsui T, Segall J, Weil PA & Roeder RG (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* **255:** 11992–11996

Maxon ME, Goodrich JA & Tjian R (1994) Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev.* **8:** 515–524

Ma X, Zhang K & Li X (2009) Evolution of Drosophila ribosomal protein gene core promoters. *Gene* **432:** 54–59

McCracken S & Greenblatt J (1991) Related RNA polymerase-binding regions in human RAP30/74 and Escherichia coli sigma 70. *Science* **253:** 900–902

McMahon LW, Walsh CE & Lambert MW (1999) Human α Spectrin II and the Fanconi Anemia Proteins FANCA and FANCC Interact to Form a Nuclear Complex. *J. Biol. Chem.* **274:** 32904–32908

McNeil JB & Smith M (1985) Saccharomyces cerevisiae CYC1 mRNA 5'-end positioning: analysis by in vitro mutagenesis, using synthetic duplexes with random mismatch base pairs. *Mol. Cell. Biol.* **5:** 3545–3551

Mencía M, Moqtaderi Z, Geisberg JV, Kuras L & Struhl K (2002) Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol. Cell* **9:** 823–833

Missra A & Gilmour DS (2010) Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc. Natl. Acad. Sci. U. S. A.* **107:** 11301–11306

Mokry M, Hatzis P, de Bruijn E, Koster J, Versteeg R, Schuijers J, van de Wetering M, Guryev V, Clevers H & Cuppen E (2010) Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. *PLoS One* **5:** e15092

Montero D, Tachibana C, Rahr Winther J & Appenzeller-Herzog C (2013) Intracellular glutathione pools are heterogeneously concentrated. *Redox Biol* **1:** 508–513

Moore PA, Ozer J, Salunek M, Jan G, Zerby D, Campbell S & Lieberman PM (1999) A Human TATA Binding Protein-Related Protein with Altered DNA Binding Specificity Inhibits Transcription from Multiple Promoters and Activators. *Mol. Cell. Biol.* **19:** 7610–7620

Muerdter F, Boryń ŁM & Arnold CD (2015) STARR-seq - principles and applications. *Genomics* **106:** 145–150

Nakajima N, Horikoshi M & Roeder RG (1988) Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* **8:** 4028–4040

Nakatani Y, Horikoshi M, Brenner M, Yamamoto T, Besnard F, Roeder RG & Freese E (1990) A downstream initiation element required for efficient TATA box binding and in vitro function of TFIID. *Nature* **348:** 86–88

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y & Adelman K (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327:** 335–338

Neguembor MV, Sebastian-Perez R, Aulicino F, Gomez-Garcia PA, Cosma MP & Lakadamyali M (2017) (Po)STAC (Polycistronic SunTAg modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. *Nucleic Acids Res.* Available at: http://dx.doi.org/10.1093/nar/gkx1271

Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG & Burley SK (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377:** 119–128

Nilson KA, Lawson CK, Mullen NJ, Ball CB, Spector BM, Meier JL & Price DH (2017) Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Res.* Available at: https://academic.oup.com/nar/article/doi/10.1093/nar/gkx724/4084663/Oxidative-stress-rapidly-stabilizes-promoter [Accessed October 3, 2017]

Nowick K & Stubbs L (2010) Lineage-specific transcription factors and the evolution of gene regulatory networks. *Brief. Funct. Genomics* **9:** 65–78

Oelgeschläger T, Tao Y, Kang YK & Roeder RG (1998) Transcription activation via enhanced preinitiation complex assembly in a human cell-free system lacking TAFIIs. *Mol. Cell* **1:** 925–931

Ohbayashi T, Shimada M, Nakadai T, Wada T, Handa H & Tamura T (2003) Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic Acids Res.* **31:** 2127–2133

Ohler U, Liao G-C, Niemann H & Rubin GM (2002a) Computational analysis of core promoters in the Drosophila genome. *Genome Biol.* **3:** RESEARCH0087

Ohler U, Liao G-C, Niemann H, Rubin GM, Latchman DS, Lemon B, Tjian R, Smale ST, Butler J, Kadonaga JT, Kutach AK, Kadonaga JT, Fickett JW, Hatzigeorgiou AG, Davuluri RV, Grosse I, Zhang MQ, Down TA, Hubbard TJ, Hannenhalli S, et al (2002b) Computational analysis of core promoters in the Drosophila genome. *Genome Biol.* **3:** research0087.1

Ohler U & Wassarman DA (2010) Promoting developmental transcription. *Development* **137:** 15–26

Okamoto T, Yamamoto S, Watanabe Y, Ohta T, Hanaoka F, Roeder RG & Ohkuma Y (1998) Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. *J. Biol. Chem.* **273:** 19866–19876

Öner R, Agarwal S, Dimovski AJ, Efremov GD, Petkov GH, Altay C, Gurgey A & Huisman THJ (1991) The G→A Mutation at Position +22 31 to the Cap Site of the β-Globin Gene as a Possible Cause for a β-Thalassemia. *Hemoglobin* **15:** 67–76

Ossipow V, Tassan JP, Nigg EA & Schibler U (1995) A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* **83:** 137–146

Ouzounis C & Sander C (1992) TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell* **71:** 189–190

Ozer J, Bolden AH & Lieberman PM (1996) Transcription factor IIA mutations show activator-specific defects and reveal a IIA function distinct from stimulation of TBP-DNA binding. *J. Biol. Chem.* **271:** 11182–11190

Parry TJ, Theisen JWM, Hsu J-Y, Wang Y-L, Corcoran DL, Eustice M, Ohler U & Kadonaga JT (2010) The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev.* **24:** 2013–2018

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T & Burley SK (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13:** 3217–3230

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K & Weng Z (2007) Nucleosome positioning signals in genomic DNA. *Genome Res.* **17:** 1170–1177

Perry RP (2005) The architecture of mammalian ribosomal protein promoters. *BMC Evol. Biol.* **5:** 15

Perry RP (2007) Balanced production of ribosomal proteins. *Gene* **401:** 1–3

Persengiev SP, Zhu X, Dixit BL, Maston GA, Kittler ELW & Green MR (2003) TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc. Natl. Acad. Sci. U. S. A.* **100:** 14887–14891

Petrenko N, Jin Y, Wong KH & Struhl K (2016) Mediator Undergoes a Compositional Change during Transcriptional Activation. *Mol. Cell* **64:** 443–454

Phadnis N, Baker EP, Cooper JC, Frizzell KA, Hsieh E, de la Cruz AFA, Shendure J, Kitzman JO & Malik HS (2015) An essential cell cycle regulation gene causes hybrid inviability in Drosophila. *Science* **350:** 1552–1555

Poon D & Weil PA (1993) Immunopurification of yeast TATA-binding protein and associated factors. Presence of transcription factor IIIB transcriptional activity. *J. Biol. Chem.* **268:** 15325–15328

Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, Thurman RE, Cheng Y, Gülsoy G, Dennis JH, Snyder MP, Stamatoyannopoulos JA, Taylor J, Hardison RC, Kahveci T, Ren B, et al (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* **515:** 402–405

Price DH, Sluder AE & Greenleaf AL (1989) Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Mol. Cell. Biol.* **9:** 1465–1475

Provost E, Hersperger G, Timmons L, Ho WQ, Hersperger E, Alcazar R & Shearn A (2006) Loss-of-function mutations in a glutathione S-transferase suppress the prune-Killer of prune lethal interaction. *Genetics* **172:** 207–219

Pugh BF & Tjian R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61:** 1187–1197

Pugh BF & Venters BJ (2016) Genomic Organization of Human Transcription Initiation Complexes. *PLoS One* **11:** e0149339

Purnell BA, Emanuel PA & Gilmour DS (1994) TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes Dev.* **8:** 830–842

Purnell BA & Gilmour DS (1993) Contribution of sequences downstream of the TATA element to a protein-DNA complex containing the TATA-binding protein. *Mol. Cell. Biol.* **13:** 2593–2603

Qin P, Parlak M, Kuscu C, Bandaria J, Mir M, Szlachta K, Singh R, Darzacq X, Yildiz A & Adli M (2017) Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.* **8:** 14725

Rabenstein MD, Zhou S, Lis JT & Tjian R (1999) TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proceedings of the National Academy of Sciences* **96:** 4791–4796

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A & Manke T (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9:** 189

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F & Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44:** W160–5

Ranish JA, Lane WS & Hahn S (1992) Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science* **255:** 1127–1129

Ranson H, Rossiter L, Ortelli F, Jensen B, Wang X, Roth CW, Collins FH & Hemingway J (2001) Identification of a novel class of insect glutathione S-transferases involved in resistance to DDT in the malaria vector Anopheles gambiae. *Biochem. J* **359:** 295–304

R Core Team R: A Language and Environment for Statistical Computing. Available at: https://www.R-project.org

Reese JC, Apone L, Walker SS, Griffin LA & Green MR (1994) Yeast TAFIIS in a multisubunit complex required for activated transcription. *Nature* **371:** 523–527

Reinberg D, Horikoshi M & Roeder RG (1987) Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J. Biol. Chem.* **262:** 3322–3330

Reja R, Vinayachandran V, Ghosh S & Pugh BF (2015) Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev.* **29:** 1942–1954

Ren D, Lei L & Burton ZF (1999) A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly. *Mol. Cell. Biol.* **19:** 7377–7387

Reniere ML, Whiteley AT, Hamilton KL, John SM, Lauer P, Brennan RG & Portnoy DA (2015) Glutathione activates virulence gene expression of an intracellular pathogen. *Nature* **517:** 170–173

Rhee HS & Pugh BF (2012) ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr. Protoc. Mol. Biol.* **Chapter 21:** Unit 21.24

Robert F, Douziech M, Forget D, Egly JM, Greenblatt J, Burton ZF & Coulombe B (1998) Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell* **2:** 341–351

Rodriguez MA, Vermaak D, Bayes JJ & Malik HS (2007) Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 15412–15417

Roepcke S, Zhi D, Vingron M & Arndt PF (2006) Identification of highly specific localized sequence motifs in human ribosomal protein gene promoters. *Gene* **365:** 48–56

Rost B & Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232:** 584–599

Rost B, Sander C & Schneider R (1994) PHD--an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10:** 53–60

Sainsbury S, Bernecky C & Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16:** 129–143

Sainsbury S, Niesser J & Cramer P (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493:** 437–440

Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20:** 3246–3248

Salinas AE & Wong MG (1999) Glutathione S-transferases--a review. *Curr. Med. Chem.* **6:** 279–309

Sawadogo M & Roeder RG (1985) Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci. U. S. A.* **82:** 4394–4398

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang J-PZ & Widom J (2006) A genomic code for nucleosome positioning. *Nature* **442:** 772–778

Serizawa H, Mäkelä TP, Conaway JW, Conaway RC, Weinberg RA & Young RA (1995) Association of Cdk-activating kinase subunits with transcription factor TFIIH. *Nature* **374:** 280–282

Shao S, Zhang W, Hu H, Xue B, Qin J, Sun C, Sun Y, Wei W & Sun Y (2016) Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res.* **44:** e86

Shechner DM, Hacisuleyman E, Younger ST & Rinn JL (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12:** 664–670

Shen W-C & Green MR (1997) Yeast TAFII145 Functions as a Core Promoter Selectivity Factor, Not a General Coactivator. *Cell* **90:** 615–624

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7:** 539

Smale ST & Baltimore D (1989) The 'initiator' as a transcription control element. *Cell* **57:** 103–113

Smale ST, Schmidt MC, Berk AJ & Baltimore D (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. U. S. A.* **87:** 4509–4513

Söderdahl T, Enoksson M, Lundberg M, Holmgren A, Ottersen OP, Orrenius S, Bolcsfoldi G & Cotgreave IA (2003) Visualization of the compartmentalization of glutathione and protein-glutathione mixed disulfides in cultured cells. *FASEB J.* **17:** 124–126

Sopta M, Burton ZF & Greenblatt J (1989) Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* **341:** 410–414

Sridharan D, Brown M, Lambert WC, McMahon LW & Lambert MW (2003) Nonerythroid alphaII spectrin is required for recruitment of FANCA and XPF to nuclear foci induced by DNA interstrand cross-links. *J. Cell Sci.* **116:** 823–835

Sridharan DM, McMahon LW & Lambert MW (2006) alphaII-Spectrin interacts with five groups of functionally important proteins in the nucleus. *Cell Biol. Int.* **30:** 866–878

Stampfel G, Kazmar T, Frank O, Wienerroither S, Reiter F & Stark A (2015) Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528:** 147–151

Stella L, Pallottini V, Moreno S, Leoni S, De Maria F, Turella P, Federici G, Fabrini R, Dawood KF, Bello ML, Pedersen JZ & Ricci G (2007) Electrostatic association of glutathione transferase to the nuclear membrane. Evidence of an enzyme defense barrier at the nuclear envelope. *J. Biol. Chem.* **282:** 6372–6379

Stubbs L, Sun Y & Caetano-Anolles D (2011) Function and Evolution of C2H2 Zinc Finger Arrays. *Subcell. Biochem.* **52:** 75–94

Sypes MA & Gilmour DS (1994) Protein/DNA crosslinking of a TFIID complex reveals novel interactions downstream of the transcription start. *Nucleic Acids Res.* **22:** 807–814

Szutorisz H, Dillon N & Tora L (2005) The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem. Sci.* **30:** 593–599

Tadepally HD, Burger G & Aubry M (2008) Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol.*

Biol. **8:** 176

Takada S, Lis JT, Zhou S & Tjian R (2000) A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. *Cell* **101:** 459–469

Talkington CA & Leder P (1982) Rescuing the in vitro function of a globin pseudogene promoter. *Nature* **298:** 192–195

Tanay A, Regev A & Shamir R (2005) Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast. *Proceedings of the National Academy of Sciences* **102:** 7203–7208

Tanese N, Pugh BF & Tjian R (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev.* **5:** 2212–2224

Tan S, Hunziker Y, Sargent DF & Richmond TJ (1996) Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381:** 127–151

Teichmann M, Wang Z, Martinez E, Tjernberg A, Zhang D, Vollmer F, Chait BT & Roeder RG (1999) Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. Natl. Acad. Sci. U. S. A.* **96:** 13720–13725

Theisen JWM, Lim CY & Kadonaga JT (2010) Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol. Cell. Biol.* **30:** 3471–3479

Thomas MC & Chiang C-M (2008) The General Transcription Machinery and General Cofactors. *Crit. Rev. Biochem. Mol. Biol.* Available at: http://www.tandfonline.com/doi/full/10.1080/10409230600648736#.Vq9y77IrKVM

Tirode F, Busso D, Coin F & Egly JM (1999) Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* **3:** 87–95

Townsend DM, Manevich Y, He L, Hutchens S, Pazoles CJ & Tew KD (2009) Novel role for glutathione S-transferase pi. Regulator of protein S-Glutathionylation following oxidative and nitrosative stress. *J. Biol. Chem.* **284:** 436–445

Tsukiyama T, Becker PB & Wu C (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367:** 525–532

Tsukiyama T & Wu C (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83:** 1011–1020

Usheva A & Shenk T (1994) TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76:** 1115–1121

Valouev A (2011) Determinants of nucleosome organization in primary human cells. *Nature* **474:** 516–520

Van Dyke MW, Roeder RG & Sawadogo M (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241:** 1335–1338

Venables WN & Smith DM (2010) the R Development Core Team.(2008) An Introduction to R. *The R Development Core Team* **2:** 1–90

Verrijzer CP, Chen J-L, Yokomori K & Tjian R (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81:** 1115–1125

Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SHC & Kadonaga JT (2017a) The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev.* **31:** 6–11

Vo Ngoc L, Wang Y-L, Kassavetis GA & Kadonaga JT (2017b) The punctilious RNA polymerase II core promoter. *Genes Dev.* **31:** 1289–1301

Walker SS, Reese JC, Apone LM & Green MR (1996) Transcription activation in cells lacking TAFIIS. *Nature* **383:** 185–188

Wampler SL & Kadonaga JT (1992) Functional analysis of Drosophila transcription factor IIB. *Genes Dev.* **6:** 1542–1552

Wang Y-L, Duttke SHC, Chen K, Johnston J, Kassavetis GA, Zeitlinger J & Kadonaga JT (2014) TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev.* **28:** 1550–1555

Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24:** 437–440

Watanabe T, Hayashi K, Tanaka A, Furumoto T, Hanaoka F & Ohkuma Y (2003) The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol. Cell. Biol.* **23:** 2914–2926

Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25:** 1189–1191

Weil PA, Luse DS, Segall J & Roeder RG (1979) Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* **18:** 469–484

Wong JC, Alon N, Norga K, Kruyt FA, Youssoufian H & Buchwald M (2000) Cloning and analysis of the mouse Fanconi anemia group A cDNA and an overlapping penta zinc finger cDNA. *Genomics* **67:** 273–283

Wong JCY, Gokgoz N, Alon N, Andrulis IL & Buchwald M (2003) Cloning and mutation analysis of ZFP276 as a candidate tumor suppressor in breast cancer. *J. Hum. Genet.* **48:** 668–671

Wu CH, Madabusi L, Nishioka H, Emanuel P, Sypes M, Arkhipova I & Gilmour DS (2001) Analysis of core promoter sequences located downstream from the TATA element in the hsp70 promoter from Drosophila melanogaster. *Mol. Cell. Biol.* **21:** 1593–1602

Wu SY & Chiang CM (1998) Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. *J. Biol. Chem.* **273:** 12492–12498

Wu SY, Kershnar E & Chiang CM (1998) TAFII-independent activation mediated by human TBP in the presence of the positive cofactor PC4. *EMBO J.* **17:** 4478–4490

Wu SY, Thomas MC, Hou SY, Likhite V & Chiang CM (1999) Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription. *J. Biol. Chem.* **274:** 23480–23490

Xiao H, Sandaltzopoulos R, Wang HM, Hamiche A, Ranallo R, Lee KM, Fu D & Wu C (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8:** 531–543

Xie J, Collart M, Lemaire M, Stelzer G & Meisterernst M (2000) A single point mutation in TFIIA suppresses NC2 requirement in vivo. *EMBO J.* **19:** 672–682

Yamashita D, Sano Y, Adachi Y, Okamoto Y, Osada H, Takahashi T, Yamaguchi T, Osumi T & Hirose F (2007) hDREF regulates cell proliferation and expression of ribosomal protein genes. *Mol. Cell. Biol.* **27:** 2003–2013

Yokomori K, Verrijzer CP & Tjian R (1998) An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. *Proc. Natl. Acad. Sci. U. S. A.* **95:** 6722–6727

Yokomori K, Zeidler MP, Chen JL, Verrijzer CP, Mlodzik M & Tjian R (1994) Drosophila TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev.* **8:** 2313–2323

Yuan G-C & Liu JS (2008) Genomic sequence is highly predictive of local nucleosome depletion. *PLoS Comput. Biol.* **4:** e13

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O & Stark A (2015) Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518:** 556–559

Zabidi MA & Stark A (2016) Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet.* **32:** 801–814

Zaret KS & Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25:** 2227–2241

Zeng P-Y, Vakoc CR, Chen Z-C, Blobel GA & Berger SL (2006) In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* **41:** 694, 696, 698

Zhang P, Herbig U, Coffman F & Lambert MW (2013) Non-erythroid alpha spectrin prevents telomere dysfunction after DNA interstrand cross-link damage. *Nucleic Acids Res.* **41:** 5321–5340

Zhang P, Sridharan D & Lambert MW (2010) Knockdown of mu-calpain in Fanconi anemia, FA-A, cells by siRNA restores alphaII spectrin levels and corrects chromosomal instability and defective DNA interstrand cross-link repair. *Biochemistry* **49:** 5570–5581

Zhang Z, Wippo CJ, Wal M, Ward E, Korber P & Pugh BF (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332:** 977–980

Zhao X & Herr W (2002) A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. *Cell* **108:** 615–627

Zouaz A, Auradkar A, Delfini MC, Macchi M, Barthez M, Ela Akoa S, Bastianelli L, Xie G, Deng W-M, Levine SS, Graba Y & Saurin AJ (2017) The Hox proteins Ubx and AbdA collaborate with the transcription pausing factor M1BP to regulate gene transcription. *EMBO J.* **36:** 2887–2906

Curriculum Vitae

DOUG BAUMANN
462 N Frear, University Park, PA, 16802 | 814.865.3795 | dgb152@psu.edu

RESEARCH EXPERIENCE

Dr. David Gilmour, Penn State University, University Park, PA
Ph.D. Student **Dec 2011 – Current**

Myriad Genetics, Salt Lake City, UT
Laboratory Technician, Level II **Dec 2010 – Aug 2011**
Laboratory Technician, Level I **May 2010 – Dec 2010**

- Processed clinical samples to detect mutations in the BRCA1 and BRCA2 genes, mutations in these genes are known to predispose individuals to a type of malignant breast cancer
- Worked in a specialty clinical diagnostic laboratory to troubleshoot problems and challenges related to problematic patient samples or processes

Dr. Brian Poole, Brigham Young University, Provo, UT
Mentored Research Student **Aug 2008 – Apr 2010**

- Received ORCA grant from BYU to perform mentored research studying Systemic Lupus Erythematosus
- Analyzed inflammation levels following drug treatments of lymphoblastoid cell lines derived by Epstein - Barr virus infection of B-cells isolated from patient blood samples

PUBLICATIONS

GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, is a transcriptional co-activator. Baumann DG, Dai MS, Lu H, and Gilmour DS. Molecular and Cellular Biology 2017. doi: 10.1128/MCB.00476-17

A sequence-specific core promoter binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. Baumann DG, Gilmour DS. Nucleic Acids Research 2017. doi.org/10.1093/nar/gkx676

Four Promoters of IRF5 Respond Distinctly to Stimuli and are Affected by Autoimmune-Risk Polymorphisms. Clark D, Read R, Mayhew V, Petersen S, Argueta L, Stutz L, Till R, Bergsten S, Robinson B, Baumann D, Heap J, Poole B. Frontiers in immunology 2013 vol: 4 pp: 360

AWARDS

Richard L. and Norma L. McCarl Graduate Endowment in Biochemistry, Microbiology, & Molecular Biology **Oct 2017**

- Annual scholarship awarded to an outstanding graduate student

Braddock-Roberts Graduate Fellowship, Penn State University **Aug 2011 – Aug 2012**

- Incentive fellowship awarded to outstanding program applicants

Office of Research and Creative Activities (ORCA) Grant recipient, Brigham Young University **Aug 2009 – May 2010**
Full Tuition Scholarship, Utah Valley University **Aug 2002 – Apr 2003**

MEETINGS and CONFERENCES

Evolution And Core Processes In Gene Expression, ASBMB Special Symposia. Stowers Institute, Kansas City, MO. **July 2017**

- Baumann DG, Dai MS, Lu H, and Gilmour DS. GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, acts as a transcription factor at hundreds of genes. Poster

Mechanisms of Eukaryotic Transcription, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. **August 2015**

- Baumann DG, and Gilmour DS. The *Drosophila* general regulatory factor M1BP drives ribosomal protein gene expression by recruiting TRF2 to their promoters. Poster

Chromatin and Epigenetic Regulation of Transcription. Penn State University, University Park, PA. **July 2015**

- Baumann DG, and Gilmour DS**.** The *Drosophila* general regulatory factor M1BP drives ribosomal protein gene expression by recruiting TRF2 to their promoters. Presentation & Poster