TRANSCRIPTIONAL REGULATION IN DROSOPHILA -- FROM GENOME TO GENE

By

Yiliang Wei

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology – Doctor of Philosophy

2016

ABSTRACT

TRANSCRIPTIONAL REGULATION IN DROSOPHILA
-- FROM GENOME TO GENE

By

Yiliang Wei

Transcriptional regulation plays a major role in gene expression, and is critical for development and diseases. To understand the molecular mechanisms of transcriptional control, I took advantage of the classic gene regulation model, *Drosophila melanogaster*, and developed approaches from both a genome wide angle and zooming in to the specific gene. On the genome level, I carried out genome-wide studies to characterize binding profiles for Drosophila retinoblastoma family proteins. I identified novel, yet conserved roles for retinoblastoma proteins in regulation of signaling pathways and ribosomal biosynthesis. I discovered that the retinoblastoma cofactors might be an important regulator of cellular growth through control of ribosomal gene expression, an unrecognized feature of this tumor suppressor protein that is apparently evolutionarily conserved in mammals. On the specific gene level, I characterized the *cis*-regulatory landscape of the Drosophila insulin receptor gene, a gene that is frequently involved in many diseases such as diabetes, Alzheimer's, and cancer. I identified many dynamically and redundantly regulated enhancers embedded in the large insulin receptor gene locus, and demonstrated that a "housekeeping" gene can be subject to control by extraordinarily complex regulatory circuitry.

This thesis is dedicated to my parents.
Thank you for always supporting me.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Prof. David Arnosti for his continuous support through my PhD study, and for his patience, kindness, motivation, and immense knowledge. His guidance not only helped me with research and writing of this thesis, but also shaped me into an independent scientific researcher with critical and creative thinking. The training and lessons I received from Dr. Arnosti is truly a life-long benefit.

I would also like to thank Dr. Henry, for being a passionate mentor and a good friend. I would like to thank my committee members, Dr. Burton, Dr. Xiao, Dr. Floer, and Dr. Knott, for their helpful guidance. I would like to thank all former and present members of the Arnosti and Henry laboratories, for their contribution and inspiration to my study.

I would deeply thank my parents, Gang Wei and Chunyan Yu, for their constant support and love.

Lastly, I would like to thank Michigan State University. As written in its motto - 'Advancing Knowledge, Transforming Lives', MSU has truly “transformed” my life, and I will always be proud to be a Spartan.

TABLE OF CONTENTS

LIST OF FIGURES

KEY TO ABBREVIATIONS

20E	20-Hydroxyecdysone
3’UTR	Three prime untranslated region
5’UTR	Five prime untranslated region
AP	Anterior-posterior
AP-1	Activator protein 1
ATF	Activating transcription factor
BAC	Bacterial artificial chromosome
BEAF-32	Boundary element associated factor of 32kD
bp	base pair
CCG	Cell cycle gene
CDK	Cyclin-dependent kinase
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation with DNA microarray
ChIP-exo	Chromatin immunoprecipitation with the use of exonucleases
ChIP-seq	Chromatin immunoprecipitation sequencing

CP190	Centrosomal protein 190kD
CRE	*Cis*-regulatory element
CRISPR	Cluster regularly interspaced short palindromic repeats
CRP	Cytoplasmic ribosomal protein
CTCF	CCCTC-binding factor
dFOX	Drosophila FOXO
DHS-seq	DNase I hypersensitivie site sequencing
dILP	Drosophila insulin-like peptide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DP	E2F dimerization partner
DPE	Downstream promoter element
DRE	DNA replication-related element
dREAM	dimerization partner, retinoblastoma-like, E2F and MuvB complex
Dref	DNA replication related element factor
DV	Dorsal-ventrol
E	Embryo

EcR	Ecdyone receptor
Eip74EF	Ecdysone-induced protein 74EF
Eip75EF	Ecdysone-induced protein 75EF
EtOH	Ethanol
eve	*even skipped*
Ex	The largest exon
FAIRE-seq	Formaldehyde assisted isolation of regulatory element sequencing
FBS	Fetal bovine serum
FOXJ2	Forkhead box J2
FOXO	Forkhead box, sub-group O
FPKM	Fragments per kilobase of transcript per million mapped reads
ftz	*fushi tarazu*
GFP	Green fluorescent protein
GO	Gene ontology
GRN	Gene regulatory network
H3K27Ac	Histone H3 lysine residue 27 acetylation
H3K4Me3	Histone H3 lysine residue 4 tri-methylation

H3K4Me1	Histone H3 lysine residue 4 mono-methylation
HMGI	High mobility group I
HOMER	Hypergeometric optimization of motif enrichment
hr	Hour
IE	Instability element
IGF1	Insulin-like growth factor 1
IGF1R	IGF1 receptor
IGV	Integrative genomics viewer
Inr	Initiator
InR	Drosophila insulin receptor
INT	Intergenic region
IR/INSR	Insulin receptor
JDP	Jun dimerization protein
Kbp	Kilo base pair
L	Larval
Loc.	Locus
MACS	Model-based analysis for ChIP-seq

MAST	Motif alignment & search tool
Mbp	Mega base pair
MCM5	Minichromosome maintenance complex component 5
MEME	Mutiple Em for motif elicitation
min	Minute
Mip	Myb-interacting protein
mL	Milliliter
mRNA	Messenger RNA
MRP	Mitochondria ribosomal protein
mRpL	Mitochondria ribosomal protein large subunit
mRpS	Mitochondria ribosomal protein small subunit
ncRNA	non-coding RBA
ng	nanogram
OSC	Ovarian stem cell
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDK	Pyruvate dehydrogenase kinase

pdm	*POU domain protein*
PI3K	Phosphatidylinositol 3-kinase
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PTEN	Phosphatase and tensin homolog
PWM	Position weight matrix
qPCR	Quantitative PCR
qRT-PCR	Quantitative real-time PCR
RB	Retinoblastoma
Rfb1	Retinoblastoma family protein 1
Rbf2	Retinoblastoma family protein 2
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RP	Ribosomal protein
RpL	Ribosomal protein large subunit
RpS	Ribosomal protein small subunit
S	Stage

S6K	Ribosomal protein S6 kinase
s.d.	Standard deviation
sna	*snail*
SNP	Single nucleotide polymorphism
SPG	Signaling pathway gene
STAP	Sequence to affinity prediction
STARR-seq	Self-transcribing active regulatory region sequencing
T2D/TIID	Type 2 diabetes
TAF	TBP associated factor
TBP	TATA-binding protein
TFIID	Transcription factor II D
tko	*technical knockout*
TOR	Target of rapamycin
TSC1	Tuberous sclerosis compex 1
TSC2	Tuberous sclerosis complex 2
TSS	Transcription start site
TTS	Transcription termination site

TU	Transcription unit
twi	*twist*
UAS	Upstream activation sequence
usp	*ultraspiracle*
WPP	White pre-pupae
yw	*yellow-while*
µg	Microgram
µL	Microlite

CHAPTER 1

Introduction

Mechanism of transcription control

Transcriptional regulation and transcription factors

Transcription, the first step in expression of a gene, plays important roles in development and disease. Stemming from pioneering work by Jacob and Monod over half a century ago, the principles of transcriptional control have been established by studies in bacterial and eukaryotic organisms (Jacob and Monod, 1961). Across all domains of life, transcriptional regulation involves DNA binding transcription factors (trans-acting factors) and their co-factors interacting with specific DNA sequences (cis-regulatory elements) to regulate the basal transcription machinery, which includes the multisubunit RNA polymerase enzyme and factors that interact with polymerase at many promoters. The process of transcriptional regulation in eukaryotes involves complex mechanisms, including direct binding between transcriptional activators and surfaces of the basal machinery; DNA looping between distal regulatory sites and the transcriptional start sites of regulated genes; chromatin remodeling, nucleosome positioning, and the still poorly understood action of noncoding RNA (ncRNA) (Lee and Young, 2013). Most eukaryotic transcription activators regulate transcription initiation by recruiting coactivators, such as the mediator complex, p300, and other widely-active factors (Lee and Young, 2013). Transcription factors also impact

transcription elongation by helping RNA polymerase II release from the pause sites (Rahl et al., 2010). At the chromatin level, ATP-dependent chromatin remodeling complexes are recruited for gene activation to remove nucleosomes and create access for transcription machinery and regulatory factors (Clapier and Cairns, 2009). Transcription factors also recruit histone-modifying enzymes to add or remove specific chemical modifications at histone tails. These modifications further create surface for other transcription factors (Portela and Esteller, 2010).

Cis-regulatory elements – Promoters and insulators

Another important aspect of transcriptional regulation is mediated through DNA elements, which include basal promoters, enhancers, and boundary elements or insulators (Arnosti, 2003). Basal promoters are typically ~100 bp in size and are located directly at the transcription initiation sites. Basal promoter regions may contain Initiator (Inr) sites overlapping +1 site of transcriptional initiation, 5' TATA-box sequences positioned at -30bp and 3' promoter element such as the DPE (Ohler et al., 2003). The TATA-box interacts with TATA-binding protein (TBP), an important part of basal transcription machinery that helps the binding of RNA polymerase and other basal transcription factors. TBP is part of the multisubunit general transcription factor TFIID, which contains TBP-associated proteins (TAFs). The Inr is bound by RNA polymerase and TAFs, while the DPE serves to provide additional contacts with TAFs to help anchor TFIID (Arnosti, 2003). However, not all basal promoters contain the same elements. The so-called "housekeeping" genes lack these elements in their promoters, but instead are enriched with DRE and other motifs, while in mammals CpG-rich promoters feature dispersed initiation patterns that are not guided by these canonical

elements (Maunakea et al., 2010; Zabidi et al., 2014). Many basal promoters appear to act in an interchangeable, promiscuous manner, but differences in the sequences of basal promoters can contribute to their specific interactions with enhancers that regulate the gene expression, facilitating the proper interactions between transcriptional regulators and target genes (Marinic et al., 2013; Zabidi et al., 2014). Another level of transcriptional specificity is generated by boundary elements or insulators, which are sequences bound by regulatory proteins that create barriers for regulatory elements (Arnosti, 2003; Wood et al., 2011). Insulators provide the "traffic control" function necessary for proper action of the complex HOX gene *cis* regulatory areas, where multiple enhancers are linked to the correct transcription units for correct tissue- and temporal regulation (Hagstrom et al., 1996; Zhou et al., 1996). A recent study in Drosophila showed that insulators, together with an assembly of DNA binding factors and associated cofactors termed the "dREAM complex" act at divergently transcribed genes to concentrate transcriptional activity to one side, and block transcriptional activation from the less active neighbors (Korenjak et al., 2014). Other studies suggested that insulators not only function as barriers, but also bring distal enhancers to their target genes through formation of DNA loops (Yang and Corces, 2012).

Cis-regulatory element – Enhancers

In eukaryotes, enhancers are distally-acting regulatory sequences that associate with sequence-specific transcription factors that recognize motifs within the enhancer to control gene expression (Spitz and Furlong, 2012). Multiple transcription factors and their cofactors typically bind cooperatively to individual enhancers to regulate nearby or distant gene expression through enhancer-promoter interactions. Enhancers are typically

a few hundred base pairs (bp) in length (Arnosti, 2003). They can be located close to their target genes, or up to 1Mbp distant in higher eukaryotes. Most commonly, an enhancer contains multiple binding sites for different transcription factors. Two distinct models have been proposed to explain the functional roles of enhancers, the "enhanceosome" model and the "billboard" model. The former one suggests that specific, highly spatially-sensitive interactions between transcription factors within an enhancer are critical for the overall output; while the latter suggests that an enhancer acts more like an information display, where the positioning of individual factors is flexible, and overall output is driven by multiple, specific interactions with basal transcription machinery (Arnosti and Kulkarni, 2005).

Identification of enhancers

Enhancers and their associated transcription factors play a leading role in regulating transcription, and identification of enhancers has been a major focus for studying transcriptional regulation. A classic way to identify enhancers is to use reporters to test whether specific DNA sequences are transcriptionally active. This method allows for detailed analysis of the enhancers, and can involve tracking their expression pattern in vivo. However, applying this method at the whole genome level requires a tremendous amount of work (Kvon et al., 2014). Taking advantage of high-throughput sequencing technologies, genome-wide approaches to identify putative enhancers have been widely employed, and include finding the in vivo profiles of transcription factors by ChIP-seq. In addition, chromatin features associated with regulatory regions have been studied, focusing on specific enhancer-associated histone modifications (H3K27-acetylation; H3K4-methylation) and open chromatin defined by DNAse I hypersensitivity and other

methods including FAIRE-seq (McKay and Lieb, 2013). More recently, the Stark laboratory devised a method (STARR-seq) to test the activity of fragments covering the entire genome, using high throughput sequencing to identify those fragments that are intrinsically active in specific cell types, or are responsive to hormonal treatment (Arnold et al., 2013; Shlyueva et al., 2014). Although these genome-wide approaches offer a global picture of putative enhancer distribution, they do not describe molecular details of transcriptional regulation of a given gene. For instance, in many cases, changes in histone modifications that would mark an enhancer are not associated with measured effects on gene expression, suggesting that there may be many false positives associated with such data sets (Kok et al., 2015). Additionally, functional assays such as STARR-seq must identify enhancer action in the context of a specific basal promoter, overlooking important enhancer-basal promoter specificity. Thus, to characterize *cis*-regulatory elements at the level of the individual gene, targeted reporter assays are still valuable.

Transcriptional regulation in development and normal physiology

Transcriptional regulation in development

Unlike broadly-expressed "housekeeping" genes, "developmental" genes display specific spatial and temporal patterns of expression, which are reflective of precise transcriptional controls. The specific temporal and spatial patterning is often regulated by the binding of cell-type specific transcription factors to developmental enhancers, or the activation of transcription factors in cell type specific manners by signaling pathways (Spitz and Furlong, 2012). Drosophila embryogenesis is a well-studied

example. There, both anterior-posterior (AP) and dorsal-ventral (DV) pattern formations require that broadly expression activators and spatially restricted repressors interact on specific enhancers to produce refined gene expression patterns, which will direct the developmental fate of each segment. In both cases, complex gene regulatory networks (GRNs) control the precise output of gene expression. On the AP axis, maternal gradients of transcription factor mRNA, such as *bicoid* and *caudal* regulate gap genes, which regulate pair-rule genes that control the initial expression of segment-polarity genes. Most of the segmentation genes also show cross-regulation to provide robustness to the system (Rivera-Pomar and Jäckle, 1996; Jaeger, 2011). Similarly, on the DV axis, transcription factors Dorsal, Twi, and Sna regulate a number of genes to establish boundaries of gene expression (Levine and Davidson, 2005).

During development, transcription factors can often bind to diverse sets of *cis*-regulatory elements (CREs) to generate temporal gene expression patterns. As mentioned above, Twi is one of the key factors that regulate DV pattern formation in Drosophila. Genome-wide studies of Twi targets at different developmental stages reveal that Twi binds to different sets of target genes in a temporally regulated manner, though Twi is expressed through those developmental time periods, indicating that other factors may aid or inhibit its bindings at specific developmental stages. Further bioinformatics analysis of Twi bound sequences suggested that Dorsal motifs are enriched in early-bound Twi regions, while Sna motifs associate with Twi binding through all assayed stages (Sandmann et al., 2007). These indicate that, depending on the cellular context, a particular type of transcription factor may bind to distinct enhancers or interact with different cofactors to generate diverse outputs.

Key transcription factors such as Bicoid and Dorsal lie at the apex of complex GRNs that contain numerous downstream transcriptional modules that propagate and create temporal and spatial patterns by combinatorial interactions. To integrate numerous regulatory inputs, developmental genes often contain multiple modular enhancers, which are "hardwired" in their flanking genomic regions and/or within intragenic sequences. A well-studied example is the crucial pair-rule gene *even-skipped* (*eve*). In early blastoderm embryos, this gene is expressed in a seven-stripe pattern, which is established by five modular enhancers, whereby each controls the expression of a subset of the stripes (Macdonald et al., 1986; Frasch et al., 1987; Goto et al., 1989; Harding et al., 1989; Fujioka et al., 1999). Sometimes the regulatory landscape for an individual gene is complex; a recent study of Drosophila *pdm* gene, which encodes a POU homeodomain transcription factor involved in neurogenesis, identified 77 unique enhancer modules positioned through the gene region of *pdm* locus. These modular enhancers direct tissue and developmental stage specific gene expression (Ross et al., 2015).

Transcriptional regulation in physiology

In addition to providing developmental specificity, transcriptional enhancers and promoters endow the cell with the ability to mount appropriate physiological responses to changing conditions. Some well-studied examples in eukaryotes include responses to heat shock, cytokine signals and other immune challenges, hypoxia, and xenobiotics (Pirkkala et al., 2001; Hoffmann, 2003; Cummins and Taylor, 2005; Hennighausen and Robinson, 2008; Pavek and Dvorak, 2008). One universally important stimulus is that of changing cellular nutritional status. In mammals, changes in blood glucose levels impact

the insulin signaling pathway; upon binding of insulin to the insulin receptor, intercellular kinases phosphorylate a variety of target proteins including the FOXO transcription factor to regulate its activity. Interestingly, FOXO is involved in feedback regulation that affects expression of the insulin receptor itself, a type of homeostatic control that is often found in regulatory circuits (Puig et al., 2003). Genome-wide studies by RNA-seq and ChIP-seq reveal that FOXO also controls the transcription of many other transcription factors and translation factors (Alic et al., 2011). Thus starting from the binding of insulin to its receptor, the signaling cascade regulates FOXO activity, triggering a complex multiple responses that impact cellular growth and physiology, as well as sensitivity of the signaling pathway itself.

Transcriptional control and evolution

Considering the central role that transcriptional regulation plays in development and physiology, it is not surprising that *cis*- and *trans*-acting components show substantial modification through evolutionary time as biological systems have evolved in complexity from single cell bacteria to multicellular organisms. A well-studied case in metazoans concerns the *HOX* genes, a group of highly conserved paralogous genes encoding transcription factors that control body plan formation. *HOX* genes are typically arranged in genomic clusters, and their differential expression patterns in the embryo from anterior to posterior are colinear with their 5' to 3' arrangement on the chromosome. In invertebrates such as Drosophila, there is only one set of *HOX* genes, while in mice and humans, there are four sets, reflecting duplication of the locus and elaborate regulation that corresponds to the more complex body plans and embryonic development of mammals (Nolte et al., 2012). The expansion and duplication of *HOX*

gene clusters suggest they have played a major force in shaping animal diversity during metazoan evolution. Diversification of individual *HOX* gene sequences and regulatory sequences have been linked to morphological evolution, accounting for striking diversity in appendage development in vertebrates and invertebrates, for instance (Merabet et al., 2009; Nolte et al., 2012). Such evolutionary changes are observed for many classes of transcription factors, with individual gene families undergoing substantial diversification and amplification in different lineages, both in bacteria and eukaryotes (Babu and Teichmann, 2003; Mendoza et al., 2013).

Changes in protein coding sequences of transcription factors can underlie evolution of phenotypic diversity, since such alterations tend to be more pleiotropic. However, due to their pleiotropic nature, mutations in the protein coding sequences are also more deleterious (Wittkopp and Kalay, 2011). Numerous examples of evolution of *cis*-regulatory elements have been documented, as they are in general less likely to have pleiotropic deleterious effects, such mutations are thought to be the most prevalent cause of phenotypic divergence (Wittkopp and Kalay, 2011). One recent example comes from limb evolution in bats; changes in a limb-expressed enhancer that controls the *Prx1* transcription factor gene contributes to the elongated forelimbs found in bats compared to the mouse (Cretekos et al., 2008). Similarly, changes in *cis*-regulatory elements of the *optix* gene encoding a homeodomain-containing transcription factor lead to different wing color patterns among *Heliconis* butterfly species (Reed et al., 2011). Changes in *cis*-regulatory elements are not only seen between different species, but also commonly exist at population level. For instance, a bioinformatics analysis of sea urchin populations revealed that variation within *cis*-regulatory elements is very common in

this species, even at essential transcription factor binding sites (Garfield et al., 2012). Genome-wide surveys involving crosses between different inbred lines and even different species of yeast, flies, and higher eukaryotes permit the estimation of evolutionary divergence in gene expression due to *cis*-acting and *trans*-acting mutations (Yvert et al., 2003; Wittkopp et al., 2004; Ronald et al., 2005). However, it has been difficult to ascertain at a molecular level how individual examples of sequence variation in population and species impacts gene expression and phenotypes, because mapping DNA sequence variation to phenotype requires deep knowledge of gene expression processes that is often lacking.

Transcription and disease

Trans-factors and disease

Somatic and germline mutations in both *trans*-acting factors and *cis*-regulatory elements have been widely implicated in human diseases, such as developmental syndromes, cancer and diabetes (Lee and Young, 2013). In fact, many transcription factors and co-factors were first characterized as tumor suppressors and oncogenes. For example, AP-1, a heterodimer that consists of members of c-Fos, c-Jun, ATF, and JDP families, was among the first eukaryotic transcription factors characterized biochemically. The proteins components turned out to be encoded by proto-oncogenes that had been genetically identified in separate studies (Ozanne et al., 2007). AP-1 is widely involved in regulation of cell growth and proliferation, cellular differentiation, and apoptosis, and is rapidly and transiently induced by serum or growth factors via oncogene mediated signaling transduction pathways (Shaulian and Karin, 2001; Ozanne et al., 2007).

Overexpression of AP-1 members is strongly oncogenic, inducing cellular transformation, and mutations affecting AP-1 genes are associated with many cancers, such as osteosarcoma, skin, and liver tumors (Eferl and Wagner, 2003). A distinct type of genetic action is observed with tumor suppressor proteins, many of which are also function as transcription factors or cofactors. Loss of function, rather than activation, of these genes is associated with cancer. The Retinoblastoma protein (RB) was the first transcription co-factor found to act as tumor suppressor, as I discuss below.

Cis-elements and disease

In addition to the lesions effecting trans-acting factors, many genome-wide studies have identified associations between mutations in *cis*-regulatory elements and human diseases, such as cancer, diabetes, β-thalassemia, hemophilia, atherosclerosis, and Alzheimer's (Villard, 2004; Epstein, 2009; Lee and Young, 2013). Single nucleotide polymorphisms (SNPs) associated with disease may be located at the promoter of the target genes, or distal sites, but they share the features of significantly reducing target gene expression by altering the binding sites for transcription factors and RNA polymerase II (Epstein, 2009). To identify the significance of particular mutations, functional tests are necessary to discern whether such SNPs are causal, or are merely linked to nearby functional alleles. Overall, many genome-wide studies indicate that mutations associated with regions bearing characteristics of enhancers are enriched in SNPs linked to human disease, lending support to the idea that such changes affect transcriptional regulation (Mathelier et al., 2015).

Retinoblastoma family proteins

One of the most intensively studied transcription cofactor associated with human disease is the retinoblastoma tumor suppressor protein. Retinoblastoma tumor suppressor proteins are transcriptional co-repressors have been described as cofactors that are preferentially bound by E2F transcription factors to regulate the cell cycle, an ancient function that appears to be widely conserved across eukaryote lineages (van den Heuvel and Dyson, 2008; Cao et al., 2010). Genetic disruptions to RB and its regulators occur in a wide variety of human tumors (Nevins, 2001).

In addition to cell cycle control, the biological roles of RB proteins appear to extend to cellular differentiation, senescence, and apoptosis, but the molecular targets and mechanisms for these processes are less well understood. The Arnosti and Henry laboratories have conducted biochemical and genetic studies to further characterize the regulation and targeting of RB in Drosophila (Acharya et al., 2012; Wei et al., 2015). Similar to their human counterparts, the Drosophila Retinoblastoma family members (Rbfs) control cell cycle and developmentally regulated gene expression (Du and Pogoriler, 2006). RB proteins function by interacting with E2F and its heterodimeric partner protein DP; these transcription factors regulate cell-cycle specific gene expression, and are repressed by RB prior to entry into S phase (Du and Pogoriler, 2006). RB proteins are also involved in a repressor complex, the dREAM complex that is involved in repression of developmental genes (Korenjak et al., 2004; Lewis et al., 2004). RB association with E2F/DP is regulated through phosphorylation by cell cycle controlled kinases, Cyclin/CDK complexes, during G1/S phase. Hyper-phosphorylated RB proteins exhibit a decrease ability to interact with their binding partners, allowing

the target genes to be de-repressed (Du and Pogoriler, 2006). There are multiple phosphorylation sites within the RB proteins, which may act as a phosphorylation code to control specific activities of RB proteins (Rubin, 2013).

Besides phosphorylation, our laboratories identified an autonomous degron located in Drosophila Rbf1 C-terminus that contributes to the ubiquitination and stability of the protein. Paradoxically, the stability of Rbf1 is inversely related to the activity of the protein; mutations in this domain that stabilize Rbf1 reduce its activity as a co-repressor (Acharya et al., 2010; Raj et al., 2012). This regulatory feature is conserved in mammalian RB family proteins (Sengupta et al., 2015). Phosphorylation sites within this C-terminus degron, as well as in N-terminus also play a role in controlling the stability of the protein, besides their well-studied functions in controlling RB protein activities (Zhang et al., 2014). Thus there is a tight link between the phosphorylation, stability, and activity of the RB proteins.

RB proteins are well known for their functions in cell cycle regulation. However, many studies suggest that RB also functions in other aspects of cellular physiology, such as protein synthesis, apoptosis, and metabolism (Harbour and Dean, 2000; Genovese et al., 2006; Gjidoda and Henry, 2013; Nicolay et al., 2013; Reynolds et al., 2013). Using genome-wide approaches, we and others have shown that the Drosophila RB proteins associate with many ribosomal protein (RP) and signaling pathway gene promoters (Acharya et al., 2012; Korenjak et al., 2012; Wei at al., 2015). These associations appear to be conserved for mammalian RB proteins as well, although this aspect of mammalian RB biology was not recognized until my studies in Drosophila (Chicas et al., 2010). The link of RB proteins to RP gene regulation suggests that these promoters may be directly

targeted by repressive signals, a feature that has not been previously observed in higher eukaryotes. Corepressors acting on RP genes in yeast are not conserved in higher eukaryotes (Hu and Li, 2007). Thus the RB regulation of RP gene promoters may provide the first, or few, evidence of negative regulation mechanism of these "housekeeping" genes. The studies contribute to a new understanding of RB function as tumor suppressor, combining regulation of both cell cycle and cell growth.

Insulin signaling pathway and regulation of insulin receptor gene

Insulin signaling pathway and insulin receptor

An unexpected discovery from our genome-wide studies of Drosophila RB proteins is that nearly half of the genes directly involved in the insulin signaling pathway are bound by this tumor suppressor, with especially robust Rbf1 occupancy observed at the promoter of the insulin receptor (*InR*) gene (Acharya et al., 2012; Wei et al., 2015). The insulin-signaling pathway is a conserved cascade that plays a major role in regulating metabolism and growth in diverse metazoan species. The insulin receptor (IR) functions to directly bind to insulin at specific target tissues, and to initiate the response to the hormone. In mammals, the *IR* and the IGF1 receptor (*IGF1R*) are evolved from a common ancestor gene, and they play fundamental regulatory roles in glucose metabolism and growth (Belfiore and Malaguarnera, 2011). Deregulation of *IR* has a role in both type 2 diabetes (T2D) and cancers. In one study, reduced *IR* gene expression was observed in pancreatic islets isolated from T2D patients, indicating that levels of *IR* expression are associated with this disease (Gunton et al., 2005). A causative role is suggested from genetic studies in the mouse; mice with a tissue-specific *IR* knock-out in

pancreatic cells developed a T2D-like phenotype (Kitamura et al., 2003). Misregulation of insulin signaling also links diabetes and cancer; T2D is associated with hyperinsulinemia, and because cancer cells can exhibit increased levels of IR, elevated insulin in T2D patients may affect cancer cell growth (Belfiore and Malaguarnera, 2011). In fact, overexpressing *IR* alone can induce a transformed phenotype in mice fibroblasts and a human breast cell line, whereas IR inhibition was sufficient to block mammary tumor progression (Hofmann et al., 1989; Belfiore and Malaguarnera, 2011). In addition, many epidemiological studies have demonstrated that T2D is an important risk factor for a variety of cancers. Emerging studies are targeting IR, an important link between these two diseases, to develop anticancer therapies (Belfiore and Malaguarnera, 2011).

The Drosophila insulin receptor

Drosophila has one insulin-like receptor (InR) that is 35% identical to human IR. The single archetypal insulin pathway in Drosophila is functionally analogous to the dual mammalian insulin/IGF system, regulating cell metabolism, growth, survival and proliferation (Figure 1-1) (Petruzzelli et al., 1986; Fernandez et al., 1995; Brogiolo et al., 2001; Teleman, 2010). The Drosophila *InR* is essential for development and is required for the formation of the epidermis and nervous system during embryogenesis (Fernandez et al., 1995). Heterozygous flies show severe developmental delays and reduced body size, similar to the phenotype observed in mice with ablation of *IR* and *IGFR* (Brogiolo et al., 2001; Kitamura et al., 2003). The importance of *InR* is seen not only in studies that have experimentally tracked the metabolic and growth regulation in individual flies. Natural variation in the *InR* gene found in different populations of Drosophila revealed that this gene associated with regional size variation. The *InR* locus shows evidence of

positive selection, with clines in allele frequency across latitude in both North America and Australia (Paaby et al., 2010). Moreover, *InR* sequence variation greatly affects fecundity and stress tolerance, suggesting that the polymorphism at *InR* is functionally significant (Paaby et al., 2010). The effects of natural population variations in the *IR* and *IGF1R* genes are also seen in humans, where the DNA variants in *IR* and *IGF1R* are linked to longevity (Kenyon, 2010). The insulin receptor association with longevity was also experimentally identified in worms, flies, and mice, indicating that the insulin pathway's effect on lifespan has been evolutionarily conserved (Tatar et al., 2001; Giannakou and Partridge, 2007; Kenyon, 2010).

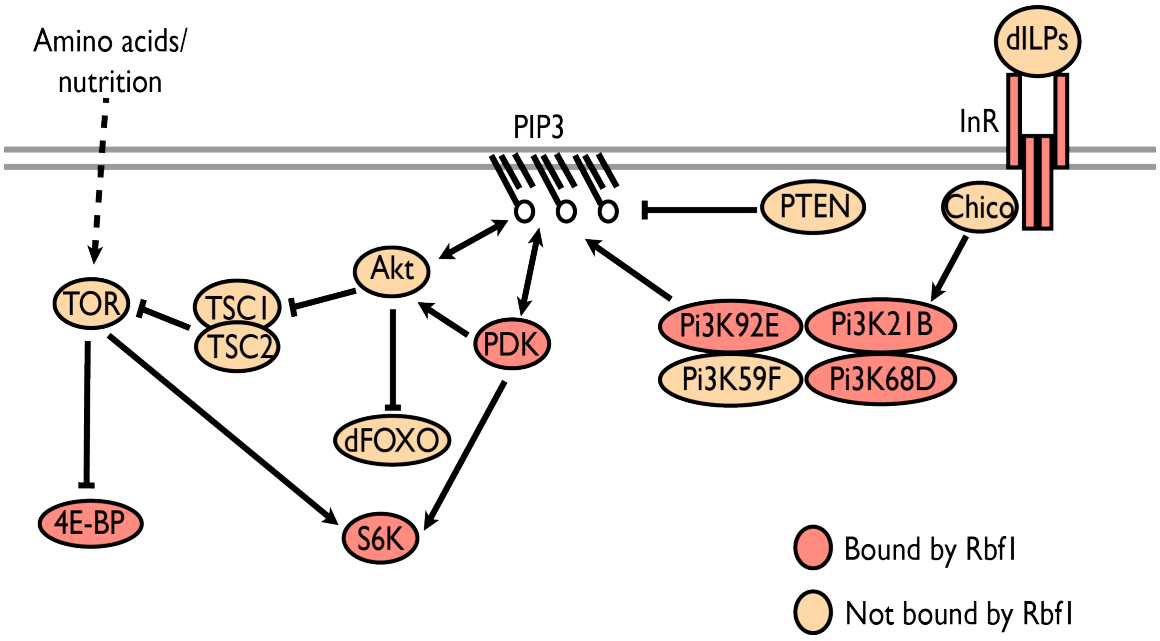


Figure 1-1: Overview of insulin signaling pathway in Drosophila. Model of insulin and TOR signaling in *Drosophila*. Abbreviations: dILPs, *Drosophila* insulin-like peptides 1-7; InR, insulin receptor; PI3K, phosphoinositide 3-kinase; PDK, phosphoinositide-dependent kinase; PIP3, phosphatidylinositol (3,4,5)-trisphosphate; Akt/PKB, protein kinase B; S6K, S6 kinase; TOR, target of rampamycin; TSC, tuberous sclerosis complex; 4E-BP, 4E-binding protein; dFOXO, *Drosophila* forkhead box, sub-group O. The genes promoters that have been identified associated with Rbf1 from ChIP-seq are indicated in red (Archarya et al., 2012).

Transcriptional regulation of InR

Although the frequency of T2D and its link to cancer has renewed interest in studying the role of IR in cancer progression, the mechanisms that regulate IR and IGFR levels still remain unclear in mammalian system. However, the Drosophila system has provided key insights into regulatory mechanisms affecting expression of the insulin receptor gene. This model system has provided strong clues that transcriptional regulation plays a key role in regulating insulin receptor gene expression. Previous studies have shown that the Drosophila forkhead protein FOXO (dFOXO) and ecdysone receptor (EcR) directly target the *InR* gene, and effectively regulate its gene transcription in response to nutrient and steroid hormone (Puig et al., 2003; Gauhar et al., 2009). Our research has focused on identifying the genome-wide targets of a cancer suppressor, Rbf1 in developing fly embryos, and for the first time Rbf1 was found strongly associated with the *InR* promoter, and to significantly repress its activity in cells (Acharya et al., 2012; Raj et al., 2012). A shared feature for the insulin receptor gene in Drosophila and human is that it contains large introns. In Drosophila, these introns are nearly 40 kbp, while they span nearly 200 kbp in humans (Casas-Tinto et al., 2007). These large introns contain many putative enhancers, as indicated by STARR-seq and other methods, suggesting the gene is subjected to complex regulatory control (Kaplan et al., 2011; Li et al., 2011; Nègre et al., 2011; Thomas et al., 2011; Arnold et al., 2013; McKay and Lieb, 2013). In both the fly and humans, only small (~2 kb) regions of the insulin receptor gene have been previously characterized for *cis*-regulatory activity (Leal et al., 1992; Lee and Tsai., 1994; García-Arencibia et al., 2005; Casas-Tinto et al., 2007;

Calle et al., 2008). Considering the overall size of the gene, the *cis*-regulatory control of the mammalian insulin receptor gene remains largely uncharacterized.

Thesis preview

In this thesis, I explore molecular characterization of transcriptional controls that apply to understanding the genome-wide activities of a conserved transcriptional corepressor, the RB protein, as well as detailed and specific *cis*- regulatory controls of one of the targets of RB regulation, the gene encoding the insulin receptor. I identified novel functions of Drosophila RB family proteins (Rf1 and Rbf2), transcription cofactors and tumor suppressors, whose human counterparts are widely involved in many type of cancers. Besides their well-known roles in cell cycle regulation, I found that their target genes include many involved in ribosomal synthesis and signaling pathways. These novel roles of Rbf1 and Rbf2 appear to be conserved in mammals. Further analysis of the ribosomal targets revealed a noncanonical role for Rbf regulation. The RB regulation of ribosomal proteins and signaling pathways would potentially give a means to integrate growth control and cell cycle control, which may have significant implications to understanding a more general role of RB in cancer and other diseases.

One especially intriguing target of Rbf1 in signaling pathways is the insulin receptor gene. The *InR* gene is highly conserved; it is critical for metabolism and growth, and essential for development. As a "housekeeping" gene, *InR* contains unexpected long intragenic regions, embedded with many putative *cis*-elements. This feature appears to be conserved in mammalian insulin receptor gene, indicating this "housekeeping" gene is, in fact, subjected to complex transcriptional control. A major part of my thesis is

dedicated to identify and characterize the *cis*- regulatory elements associated with the *InR* gene using reported-based assays, and the mapping their responses to dFOXO and ecdysone, two key transcriptional components regulating gene expression in response to nutrient status and growth. My detailed mutagenic studies of the active enhancers identified specific elements and motifs required for enhancer activity. The dynamic regulation of multiple enhancers within this gene by dFOXO and ecdysone indicates these enhancers may play a role in temporal, spatial, and fine-tuning control of *InR* gene expression. My study indicates that this gene is subject to a complex transcriptional circuit extending far beyond the previously described simple model of the FOXO-feedback loop mechanism. This study will be a fundamental guideline for designing genetic assays to understand the transcriptional regulation of the insulin receptor gene, which will give a better understanding of the role of insulin receptor in metabolism, growth control and cancer.

REFERENCES

REFERENCES

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

Acharya, P., Raj, N., Buckley, M. S., Zhang, L., Duperon, S., Williams, G., Henry, R. W. and Arnosti, D. N. (2010). Paradoxical instability-activity relationship defines a novel regulatory pathway for retinoblastoma proteins. *Mol. Biol. Cell* **21**, 3890–3901.

Alic, N., Andrews, T. D., Giannakou, M. E., Papatheodorou, I., Slack, C., Hoddinott, M. P., Cocheme, H. M., Schuster, E. F., Thornton, J. M. and Partridge, L. (2011). Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. *Mol. Syst. Biol.* **7**, 502.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077.

Arnosti, D. N. (2003). Analysis and function of transcriptional regulatory elements: insights from Drosophila. *Annu. Rev. Entomol.* **48**, 579–602.

Arnosti, D. N. and Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* **94**, 890–898.

Belfiore, A. and Malaguarnera, R. (2011). Insulin receptor and cancer. *Endocr. Relat. Cancer* **18**, R125–47.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R. and Hafen, E. (2001). An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–21.

Calle, C., Maestro, B. and Garcia-Arencibia, M. (2008). Genomic actions of 1,25-dihydroxyvitamin D3 on insulin receptor gene expression, insulin receptor number

and insulin activity in the kidney, liver and adipose tissue of streptozotocin-induced diabetic rats. *BMC Mol. Biol.* **9**, 65.

Cao, L., Peng, B., Yao, L., Zhang, X., Sun, K., Yang, X. and Yu, L. (2010). The ancient function of RB-E2F pathway: insights from its evolutionary history. *Biol. Direct* **5**, 55.

Casas-Tinto, S., Marr, M. T., Andreu, P. and Puig, O. (2007). Characterization of the Drosophila insulin receptor promoter. *Biochim. Biophys. Acta* **1769**, 236–43.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R. a, Narita, M., Zhang, M. and Lowe, S. W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–87.

Clapier, C. R. and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304.

Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J. 4th and Behringer, R. R. (2008). Regulatory divergence modifies limb length between mammals. *Genes Dev.* **22**, 141–151.

Cummins, E. P. and Taylor, C. T. (2005). Hypoxia-responsive transcription factors. *Pflugers Arch.* **450**, 363–371.

de Mendoza, A., Sebe-Pedros, A., Sestak, M. S., Matejcic, M., Torruella, G., Domazet-Loso, T. and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4858–66.

Du, W. and Pogoriler, J. (2006). Retinoblastoma family genes. *Oncogene* **25**, 5190–200.

Eferl, R. and Wagner, E. F. (2003). AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3**, 859–868.

Epstein, D. J. (2009). Cis-regulatory mutations in human disease. *Brief. Funct. Genomic. Proteomic.* **8**, 310–316.

Fernandez, R., Tabarini, D., Azpiazu, N., Frasch, M. and Schlessinger, J. (1995). The Drosophila insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. *EMBO J.* **14**, 3373–84.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (1987). Characterization and localization of the even-skipped protein of Drosophila. *EMBO J.* **6**, 749–759.

Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A. A., Weir, M. and Goto, T. (1996). Drosophila Paired regulates late even-skipped expression through a composite binding site for the paired domain and the homeodomain. *Development* **122**, 2697–2707.

Garcia-Arencibia, M., Davila, N., Campion, J., Carmen Carranza, M. and Calle, C. (2005). Identification of two functional estrogen response elements complexed with AP-1-like sites in the human insulin receptor gene promoter. *J. Steroid Biochem. Mol. Biol.* **94**, 1–14.

Garfield, D., Haygood, R., Nielsen, W. J. and Wray, G. A. (2012). Population genetics of cis-regulatory sequences that operate during embryonic development in the sea urchin Strongylocentrotus purpuratus. *Evol. Dev.* **14**, 152–167.

Gauhar, Z., Sun, L. V, Hua, S., Mason, C. E., Fuchs, F., Li, T.-R., Boutros, M. and White, K. P. (2009). Genomic mapping of binding regions for the Ecdysone receptor protein complex. *Genome Res.* **19**, 1006–13.

Genovese, C., Trani, D., Caputi, M. and Claudio, P. P. (2006). Cell cycle control and beyond: emerging roles for the retinoblastoma gene family. *Oncogene* **25**, 5201–5209.

Giannakou, M. E. and Partridge, L. (2007). Role of insulin-like signalling in Drosophila lifespan. *Trends Biochem. Sci.* **32**, 180–188.

Gjidoda, A. and Henry, R. W. (2013). RNA polymerase III repression by the retinoblastoma tumor suppressor protein. *Biochim. Biophys. Acta* **1829**, 385–392.

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413–422.

Gunton, J. E., Kulkarni, R. N., Yim, S., Okada, T., Hawthorne, W. J., Tseng, Y.-H., Roberson, R. S., Ricordi, C., O'Connell, P. J., Gonzalez, F. J., et al. (2005). Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122**, 337–349.

Harbour, J. W. and Dean, D. C. (2000). The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev.* **14**, 2393–2409.

Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *EMBO J.* **8**, 1205–1212.

Hennighausen, L. and Robinson, G. W. (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes Dev.* **22**, 711–721.

Hoffmann, J. A. (2003). The immune response of Drosophila. *Nature* **426**, 33–38.

Hofmann, C., Goldfine, I. D. and Whittaker, J. (1989). The metabolic and mitogenic effects of both insulin and insulin-like growth factor are enhanced by transfection of insulin receptors into NIH3T3 fibroblasts. *J. Biol. Chem.* **264**, 8606–8611.

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

JACOB, F. and MONOD, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

Jaeger, J. (2011). The gap gene network. *Cell. Mol. Life Sci.* **68**, 243–274.

Kaplan, T., Li, X.-Y., Sabo, P. J., Thomas, S., Stamatoyannopoulos, J. A., Biggin, M. D. and Eisen, M. B. (2011). Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet.* **7**, e1001290.

Kenyon, C. J. (2010). The genetics of ageing. *Nature* **464**, 504–512.

Kitamura, T., Kahn, C. R. and Accili, D. (2003). Insulin receptor knockout mice. *Annu. Rev. Physiol.* **65**, 313–332.

Kok, K., Ay, A., Li, L. M. and Arnosti, D. N. (2015). Genome-wide errant targeting by Hairy. *Elife* **4**,.

Korenjak, M., Kwon, E., Morris, R. T., Anderssen, E., Amzallag, A., Ramaswamy, S. and Dyson, N. J. (2014). dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. *Nucleic Acids Res.* **42**, 8939–8953.

Korenjak, M., Taylor-Harding, B. and Binné, U. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell* **119**, 181–193.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95.

Leal, M. A., Cabanas, C., Rius, C., Aller, P. and Calle, C. (1992). Modulation by dexamethasone of insulin binding and insulin receptor mRNA levels in U-937 human promonocytic cells. *Biochimie* **74**, 545–549.

Lee, J. K. and Tsai, S. Y. (1994). Multiple hormone response elements can confer glucocorticoid regulation on the human insulin receptor gene. *Mol. Endocrinol.* **8**, 625–634.

Lee, T. I. and Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251.

Levine, M. and Davidson, E. H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4936–4942.

Lewis, P. W., Beall, E. L., Fleischer, T. C., Georlette, D., Link, A. J. and Botchan, M. R. (2004). Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev.* **18**, 2929–2940.

Li, X.-Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. A. and Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol.* **12**, R34.

Macdonald, P. M., Ingham, P. and Struhl, G. (1986). Isolation, structure, and expression of even-skipped: a second pair-rule gene of Drosophila containing a homeo box. *Cell* **47**, 721–734.

Madan Babu, M. and Teichmann, S. A. (2003). Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res.* **31**, 1234–1244.

Marinic, M., Aktas, T., Ruf, S. and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542.

Mathelier, A., Shi, W. and Wasserman, W. W. (2015). Identification of altered cis-regulatory elements in human disease. *Trends Genet.* **31**, 67–76.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257.

McKay, D. J. and Lieb, J. D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. *Dev. Cell* **27**, 306–318.

Merabet, S., Hudry, B., Saadaoui, M. and Graba, Y. (2009). Classification of sequence signatures: a guide to Hox protein function. *Bioessays* **31**, 500–511.

Negre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527–531.

Nevins, J. R. (2001). The Rb/E2F pathway and cancer. *Hum. Mol. Genet.* **10**, 699–703.

Nicolay, B. N., Gameiro, P. a, Tschöp, K., Korenjak, M., Heilmann, A. M., Asara, J. M., Stephanopoulos, G., Iliopoulos, O. and Dyson, N. J. (2013). Loss of RBF1 changes glutamine catabolism. *Genes Dev.* **27**, 182–96.

Nolte, C., Ahn, Y. and Krumlauf, R. (2001). Evolutionary Developmental Biology:Hox Gene Evolution. In *eLS*, Chichester, UK: John Wiley & Sons, Ltd.

Ohler, U., Liao, G., Niemann, H. and Rubin, G. M. (2002). Computational analysis of core promoters in the Drosophila genome. *Genome Biol.* **3**, RESEARCH0087.

Ozanne, B. W., Spence, H. J., McGarry, L. C. and Hennigan, R. F. (2007). Transcription factors control invasion: AP-1 the first among equals. *Oncogene* **26**, 1–10.

Paaby, A. B., Blacket, M. J., Hoffmann, A. A. and Schmidt, P. S. (2010). Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents. *Mol. Ecol.* **19**, 760–774.

Pavek, P. and Dvorak, Z. (2008). Xenobiotic-induced transcriptional regulation of xenobiotic metabolizing enzymes of the cytochrome P450 superfamily in human extrahepatic tissues. *Curr. Drug Metab.* **9**, 129–143.

Petruzzelli, L., Herrera, R., Arenas-Garcia, R., Fernandez, R., Birnbaum, M. J. and Rosen, O. M. (1986). Isolation of a Drosophila genomic sequence homologous to the kinase domain of the human insulin receptor and detection of the phosphorylated Drosophila receptor with an anti-peptide antibody. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 4710–4714.

Pirkkala, L., Nykanen, P. and Sistonen, L. (2001). Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **15**, 1118–1131.

Portela, A. and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068.

Puig, O., Marr, M. T., Ruhf, M. L. and Tjian, R. (2003). Control of cell number by Drosophila FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* **17**, 2006–2020.

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

Raj, N., Zhang, L., Wei, Y., Arnosti, D. N. and Henry, R. W. (2012). Ubiquitination of retinoblastoma family protein 1 potentiates gene-specific repression function. *J. Biol. Chem.* **287**, 41835–41843.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., Jiggins, C. D., Chamberlain, N. L., Kronforst, M. R., Chen, R., et al. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141.

Reynolds, M. R., Lane, a N., Robertson, B., Kemp, S., Liu, Y., Hill, B. G., Dean, D. C. and Clem, B. F. (2014). Control of glutamine metabolism by the tumor suppressor Rb. *Oncogene* **33**, 556–66.

Rivera-Pomar, R. and Jackle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. *Trends Genet.* **12**, 478–483.

Ronald, J., Brem, R. B., Whittle, J. and Kruglyak, L. (2005). Local regulatory variation in Saccharomyces cerevisiae. *PLoS Genet.* **1**, e25.

Ross, J., Kuzin, A., Brody, T. and Odenwald, W. F. (2015). cis-regulatory analysis of the Drosophila pdm locus reveals a diversity of neural enhancers. *BMC Genomics* **16**, 700.

Rubin, S. M. (2013). Deciphering the retinoblastoma protein phosphorylation code. *Trends Biochem. Sci.* **38**, 12–19.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. M. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev.* **21**, 436–449.

Sengupta, S., Lingnurkar, R., Carey, T. S., Pomaville, M., Kar, P., Feig, M., Wilson, C. A., Knott, J. G., Arnosti, D. N. and Henry, R. W. (2015). The Evolutionarily Conserved C-terminal Domains in the Mammalian Retinoblastoma Tumor Suppressor Family Serve as Dual Regulators of Protein Stability and Transcriptional Potency. *J. Biol. Chem.* **290**, 14462–14475.

Shaulian, E. and Karin, M. (2001). AP-1 in cell proliferation and survival. *Oncogene* **20**, 2390–2400.

Shlyueva, D., Stelzer, C., Gerlach, D., Yanez-Cuna, J. O., Rath, M., Boryn, L. M., Arnold, C. D. and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol. Cell* **54**, 180–192.

Spitz, F. and Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626.

Tatar, M., Kopelman, A., Epstein, D., Tu, M. P., Yin, C. M. and Garofalo, R. S. (2001). A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107–110.

Teleman, A. A. (2010). Molecular mechanisms of metabolic regulation by insulin in Drosophila. *Biochem. J.* **425**, 13–26.

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., Giste, E., Fisher, W., Hammonds, A., Celniker, S. E., et al. (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biol.* **12**, R43.

van den Heuvel, S. and Dyson, N. J. (2008). Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell Biol.* **9**, 713–24.

Villard, J. (2004). Transcription regulation and human diseases. *Swiss Med. Wkly.* **134**, 571–579.

Wei, Y., Mondal, S. S., Mouawad, R., Wilczynski, B., Henry, R. W. and Arnosti, D. N. (2015). Genome-Wide Analysis of Drosophila Rbf2 Protein Highlights Diversity of RB Family Targets and Possible Role in Regulation of Ribosome Biosynthesis. *G3 (Bethesda)*.

Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88.

Wittkopp, P. J. and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69.

Wood, A. M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B. C., Jones, K. C. and Corces, V. G. (2011). Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Mol. Cell* **44**, 29–38.

Yang, J. and Corces, V. G. (2012). Insulators, long-range interactions, and genome function. *Curr. Opin. Genet. Dev.* **22**, 86–92.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R. and Kruglyak, L. (2003). Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat. Genet.* **35**, 57–64.

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

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W. and Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J. Biol. Chem.* **289**, 24863–24873.

CHAPTER 2

Genome-wide analysis of Drosophila Rbf2 protein highlights diversity of RB family targets and possible role in regulation of ribosome biosynthesis

Abstract

Rbf2 is a recently evolved Retinoblastoma family member in Drosophila, differing from Rbf1 especially in the C-terminus. To investigate whether the unique features of Rbf2 contribute to diverse roles in gene regulation, we performed ChIP-Seq for both Rbf2 and Rbf1 in embryos. A previous model for Rb-E2F interactions suggested that Rbf1 binds dE2F1 or dE2F2, while Rbf2 is restricted to binding to dE2F2, however, we found that Rbf2 targets approximately twice as many genes as Rbf1. Highly enriched among the Rbf2 targets were ribosomal protein genes. We tested the functional significance of this finding by assessing Rbf activity on ribosomal protein promoters and the endogenous genes. Rbf1 and Rbf2 significantly repressed expression of some ribosomal protein genes, although not all bound genes showed transcriptional effects. Interestingly, many ribosomal protein genes are similarly targeted in human cells, indicating that these interactions may be relevant for control of ribosome biosynthesis and growth. We carried out bioinformatic analysis to investigate the basis for differential targeting by these two proteins, and found that Rbf2-specific promoters have distinct sequence motifs, suggesting unique targeting mechanisms. Association of Rbf2 with these promoters appears to be independent of dE2F2/dDP, although promoters bound by both Rbf1 and Rbf2 require dE2F2/dDP. The presence of unique Rbf2 targets suggest that evolutionary

appearance of this corepressor represents the acquisition of potentially novel roles in gene regulation for the RB family.

This work was published as the following manuscript:

Wei, Y., Mondal, S. S., Mouawad, R., Wilczynski, B., Henry, R. W. and Arnosti, D. N. (2015). Genome-Wide Analysis of Drosophila Rbf2 Protein Highlights Diversity of RB Family Targets and Possible Role in Regulation of Ribosome Biosynthesis. *G3 (Bethesda)*.

My contribution to this study was execution of all ChIP experiments, functional assays of reporters and qPCR measurement of gene expression, and mapping and analysis of ChIP-seq and RNA-seq results.

Introduction

Retinoblastoma (Rb) tumor suppressor proteins, including vertebrate RB, p130, and p107, are important regulators of the cell cycle, apoptosis, differentiation, genomic stability, and metabolism (Weinberg, 1995; Dyson, 1998; Norton et al., 1998; Fan and Steer, 1999; Zheng and Lee, 2002; Hernando *et al*., 2004; Giacinti and Giordano, 2006; Nicolay *et al*., 2013; Reynolds et al., 2014; and references therein). These proteins function as transcriptional co-repressors that bind to E2F and DP proteins, and control transcription of a diverse set of target genes, in many cases in a cell cycle dependent manner (reviewed in Weinberg, 1995; Classon and Harlow, 2002; Du and Pogoriler, 2006; van den Heuvel and Dyson, 2008; and references therein). The Drosophila Retinoblastoma family members Rbf1 and Rbf2 are structurally similar to the vertebrate proteins and possess functionally conserved activities in control of cell cycle and developmental genes (reviewed in Du and Pogoriler, 2006). The RB-E2F pathway is conserved in most eukaryotic lineages, especially in multicellular organisms (Cao et al., 2010). Most arthropod genomes encode a single *RB* gene, easily distinguishable by conserved sequences encoding the core "pocket domain" essential for E2F interaction. Interestingly, the genus *Drosophila* contains an additional retinoblastoma family member, *rbf2* (Stevaux et al., 2002). The Rbf2 protein possesses a conserved pocket domain, similar to that of Rbf1. It also contains a distinct C-terminus that lacks the conserved instability element, which has been shown to control both stability and activity of Rbf1 (Acharya et al., 2010; Raj et al., 2012). Both Rbf1 and Rbf2 mediate transcriptional repression, however, these proteins have different inherent ability to interact with E2F proteins; Rbf1 has been found to functionally interact with both the

activator dE2F1 as well as the repressor dE2F2, while Rbf2 is found to interact with dE2F2, but not dE2F1 (Frolov et al., 2001; Stevaux et al., 2002). Cell-based *in vitro* assays suggested Rbf1 acts as a strong repressor of dE2F1 targets. By contrast, the action of Rbf2 appears to be weaker, and requires co-expression of dE2F2 for maximal repression (Stevaux et al., 2002).

Rbf1 and Rbf2 are co-expressed at many points in development, but there are important differences. In contrast to the relatively stable expression of Rbf1 during embryonic development, the Rbf2 protein levels vary considerably, with a peak at early stages (Stevaux et al., 2002; Keller et al., 2005). In contrast to broadly overlapping patterns early in embryogenesis, the proteins show tissue-specific expression in the developing central nervous system. The Rbf1 and Rbf2 proteins are co-expressed in larval imaginal discs, but Rbf1 is the main family member expressed in adults with the exception of the ovary, where Rbf2 is also expressed at high levels (Stevaux et al., 2002; Keller et al., 2005). Consistent with its expression profile, Rbf2 was found to repress differentiation markers in embryos and ovaries. Although unlike *rbf1* mutants, *rbf2* null flies are viable, *rbf2* mutant females laid eggs at a four-fold higher rate than wild-type individuals (Stevaux et al., 2005). Interestingly, this phenotype was not seen in *de2f2* mutant flies (Stevaux et al., 2005), although dE2F2 has been suggested to be the mediator of Rbf2 interactions with DNA (Stevaux et al., 2002).

The genome binding profile of Rbf1 has been characterized in both Drosophila embryos and larvae, and both studies revealed that Rbf1 interacts with numerous genes related to cellular signaling pathways, in addition to previously characterized cell cycle genes (Acharya et al., 2012; Korenjak et al., 2012). Although the genome-wide binding of

Rbf2 has not previously been reported, ChIP-qPCR of individual target genes has revealed that Rbf2 is present at Rbf1-bound loci, suggesting that these proteins may regularly co-occupy promoter regions (Korenjak et al., 2012). Considering the evolutionary conservation within the genus *Drosophila* of *rbf2* and the pervasive co-occupancy of Rbf1 and Rbf2, the modest phenotype of *rbf2* mutants presents a conundrum regarding the selection pressure for this gene over large evolutionary periods within *Drosophila*.

Previous studies suggested that Rbf1 and Rbf2 targeting is mediated via dE2F/dDP (Stevaux et al., 2002). Biochemical as well as genetic information supports this view; the larval lethality phenotype of *rbf1* inactivation can be rescued by a mutation in dE2F1 that disrupts this protein's activation domain (Du, 2000). In the larva, a *dDP* null mutation abolishes the genome-wide association of Rbf1, as well as the association of Rbf2 to several tested target genes (Korenjak et al., 2012). In contrast, the mammalian RB protein does not interact exclusively with E2F family proteins, but also physically and functionally interacts with diverse transcription factors and regulatory proteins (as reviewed in Classon and Dyson, 2001; Morris and Dyson, 2001; Chinnam and Goodrich, 2011), as well as components of the RNA polymerase I and III basal transcription machinery (Cavanaugh et al., 1995; Larminie et al., 1997; White, 1997; Hirsch et al., 2000 and 2004; Gjidoda and Henry, 2013). RB proteins in flies, worms, and vertebrates are frequently complexed with additional promoter-associated regulatory factors, including components of the evolutionarily-conserved dREAM complex, which has been shown to regulate developmental gene expression (Korenjak et al., 2004; Lewis et al., 2004). In Drosophila, a majority of the Rbf1-bound regions are also occupied by one

or more proteins of this multi-protein complex (Acharya et al., 2012). Genetically, the dREAM complex functions not only as a repressor, but also appears to recruit insulator proteins to block enhancer activity on divergently transcribed genes (Bohla et al., 2014; Korenjak et al., 2014).

In mammals, individual RB family proteins have distinct molecular targets. This targeting is influenced by structural differences in the RB proteins, particularly in the C-terminus, which allow them to bind preferentially to distinct E2F factors (Rubin et al., 2005; Julian et al., 2008; Cecchini and Dick, 2011; Dick and Rubin, 2013). In Drosophila, the C-terminus of Rbf2 is structurally divergent from that of Rbf1, which affects the regulation of this protein, and potentially influences promoter targeting (N. Raj and R.W. Henry, unpublished). To determine how this structurally divergent protein interacts with genomic targets, we carried out parallel ChIP-seq analysis of Rbf1 and Rbf2 in developing embryos, followed by bioinformatic and functional analysis of target genes. Here, we discuss how distinct genome-wide interactions of Rbf2 point to possible diversification in functions for these Rbf proteins. Ribosomal protein genes are one class not previously considered as RB targets, pointing to a potentially important role in growth control as well as cell cycle regulation. Analysis of newly identified Rbf targets suggest that the canonical RB-E2F model may not describe the full spectrum of interactions found for the derived Rbf2 protein.

Results

Genome-wide Rbf1 and Rbf2 association

To identify the genomic targets of RB family proteins in Drosophila, we used ChIP-exo analysis to measure binding profiles for both Rbf1 and Rbf2 in 12- to 18-hr embryos (Rhee and Pugh, 2011; Figure 2-1A). The canonical Rbf-E2F interaction model holds that Rbf1 binds to both dE2F1 and dE2F2 proteins while Rbf2 binds only dE2F2 (Frolov et al., 2001; Stevaux et al., 2002). Therefore, it was surprising that there were substantially more peaks identified for Rbf2 (4708) than for Rbf1 (2356); this corresponds to 3945 and 1955 genes, respectively. As noted previously for Rbf1, Rbf2 binding was also localized primarily to promoter-specific regions (Acharya et al., 2012; Figure 2-2, 2-3).

To measure the overlap between the Rbf and E2F genomic binding profiles, we compared the Rbf genomic targets to those associated with dE2F1 and dE2F2 previously identified in larvae (Korenjak et al., 2012). In the larvae, dE2F2 was found to have nearly 4000 binding sites, compared to dE2F1, which only has less than 300 binding sites (Korenjak et al., 2012). We mapped the dE2F1 and dE2F2 peaks to the nearest genes, and compared to genes bound by Rbf1 and Rbf2. Over half of Rbf1 target genes were bound by dE2F1 or dE2F2, while less than half of Rbf2 target genes were bound by any E2F factor (Figure 2-1A). The discrepancy between Rbf protein binding and E2F factor binding may reflect the two different developmental stages used for measuring binding, although many individual genes are similarly bound in both stages (Acharya et al., 2012; Korenjak et al., 2012). To directly compare Rbf and E2F targets at the same

developmental stage, we conducted ChIP-qPCR analysis using Rbf1, Rbf2, dE2F1, dE2F2 and dDP antibodies in 12- to 18-hr embryos. We checked selected targets that were previously found bound or not bound by Rbf1, Rbf2 and dE2F2 (Korenjak et al., 2012, and this study) (Figure 2-4). We noticed weak dE2F2 and dDP bindings on some ribosomal protein gene promoters that were previously shown to be bound by Rbf2, but not dE2F2 (Korenjak *et al*. 2012). However, these ChIP signals were also close to signals from non-specific promoters that were unlikely to be targeted by Rbf or E2F (Figure 2-4). Thus whether these Rbf2-only targets are bound by dE2F2 or dDP needs to be determined by the global background of the dE2F2 and dDP antibodies. But it is possible that some Rbf2 binding is directed by E2F-independent mechanisms, which we explore below.

Rbf1 and Rbf2 were found to co-occupy many genes, either through simultaneous binding to multiple transcription factors on a given promoter, or perhaps in a competitive manner (Figure 2-1B). A small number of promoters were bound only by Rbf1 (Figure 2-1C), while others featured significant Rbf2 binding and no trace of Rbf1, suggesting that these promoters may recruit Rbf factors in a different fashion from the genes bound by both Rbf1 and Rbf2 (Figure 2-1C). Indeed, motif searches of Rbf1/Rbf2 peak areas compared to Rbf2-only peaks showed that E2F-like sequences were enriched in those areas co-bound by Rbf1/Rbf2. Motifs enriched under Rbf2-alone peaks did not contain E2F-like sequences, but instead contained distinct sequences (Figure 2-5).

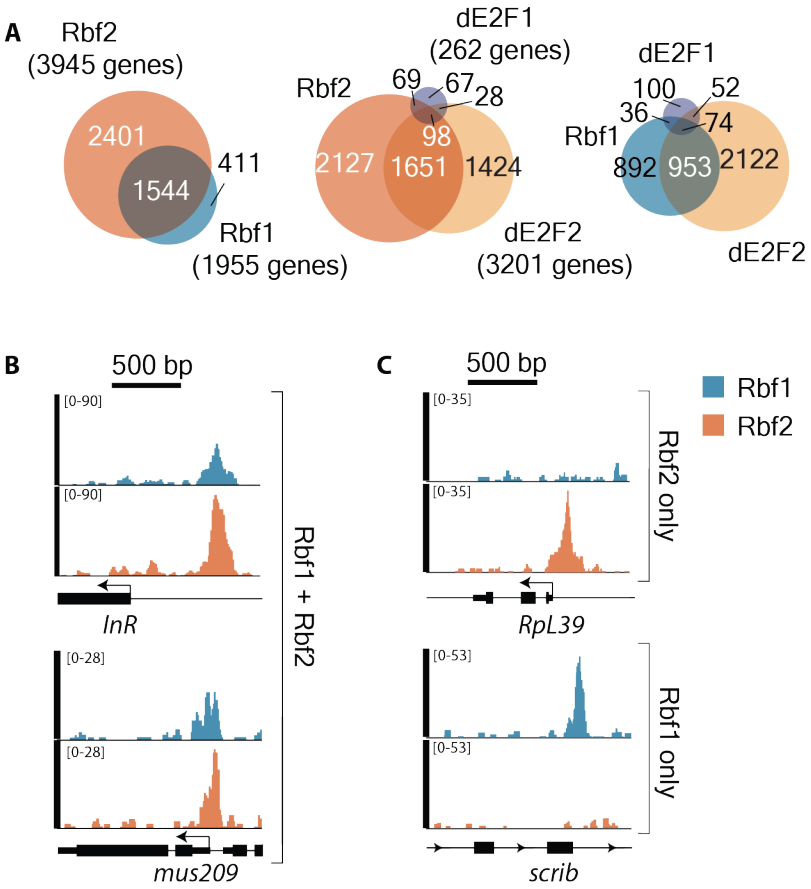


Figure 2-1: Rbf2 binds to a large number of unique targets in the *Drosophila* genome. (A) Visualization of number of genes bound by Rbf1 or Rbf2 peaks, and overlap of these genes with those targetd by dE2F1 and dE2F2 (Korenjak et al., 2012). (B) Examples of promoter regions co-occupied by both Rbf1 and Rbf2. (C) Examples of genes bound uniquely by Rbf1 or Rbf2.

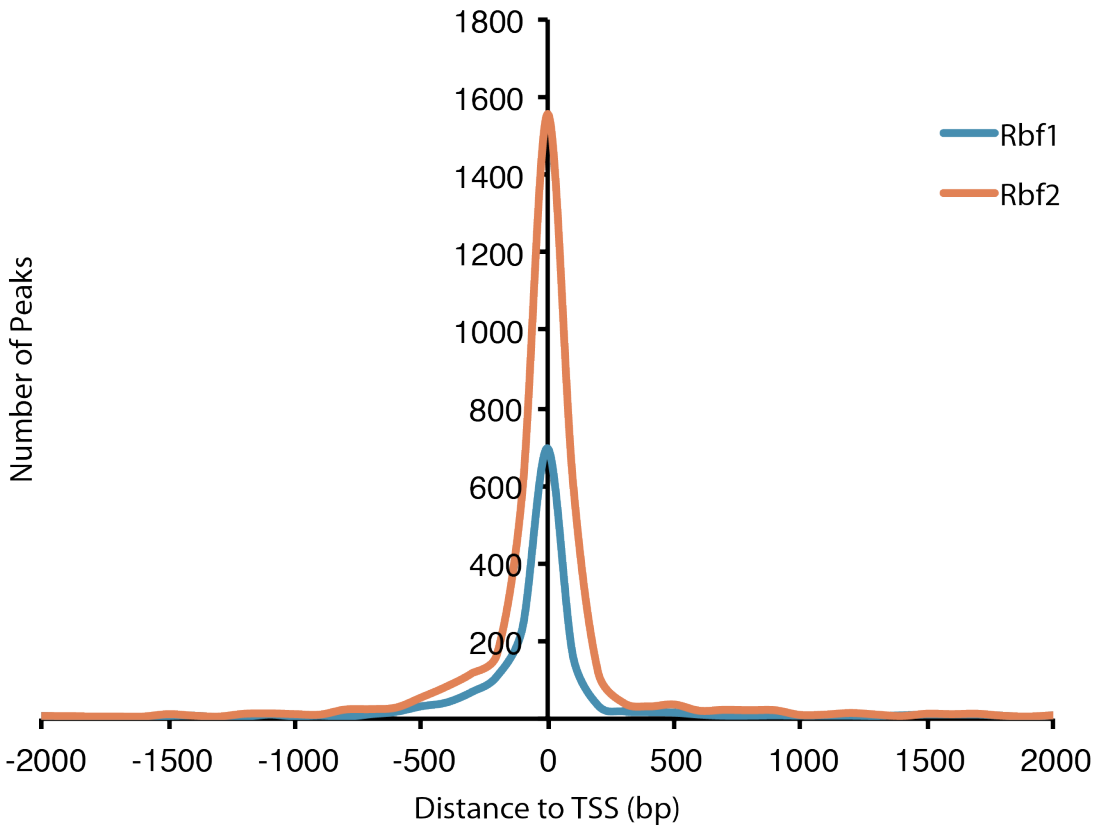


Figure 2-2: Distribution of distances of Rbf1/2 peaks to the nearest Transcription Start Sites (TSS). Majority of the Rbf1/2 peaks were located within 500 bp of TSS.

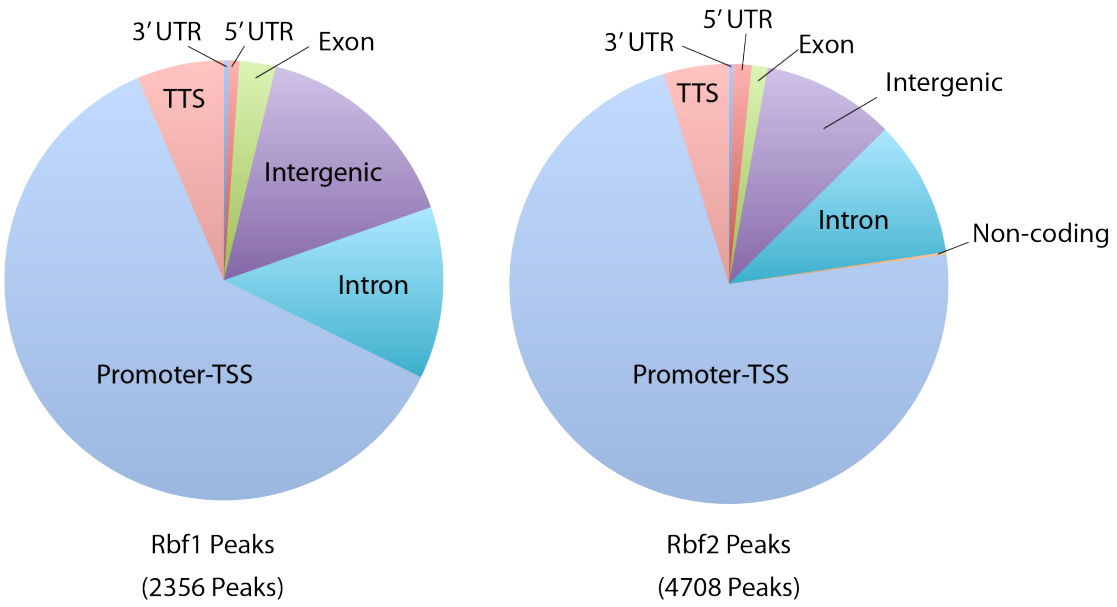


Figure 2-3: Annotations of genomic positions of Rbf1/2 peaks. A heavy bias to promoter-proximal (-1 kp to +100 bp of TSS) regions is noted. (TSS, transcription start site; TTS, transcription termination site)

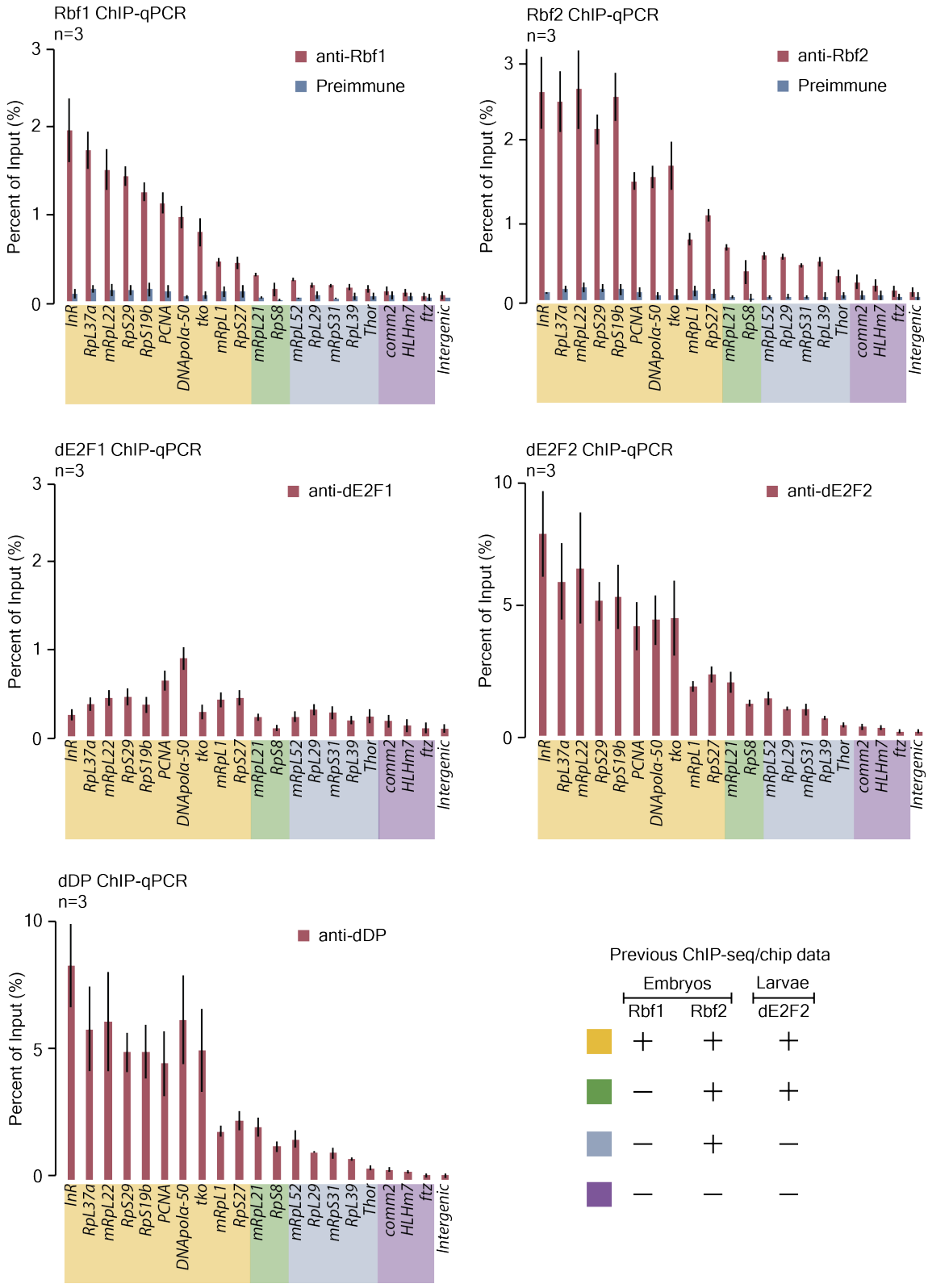


Figure 2-4: ChIP-qPCR for Rbf1, Rbf2, dE2F1, dE2F2, and dDP on chromatin from 12-18 hr embryos.

Figure 2-4 (cont'd)

The selected targets are grouped in different colors based on whether they are bound by Rbf1/Rbf2 (from this study) or dE2F2 (Korenjak et al., 2012).

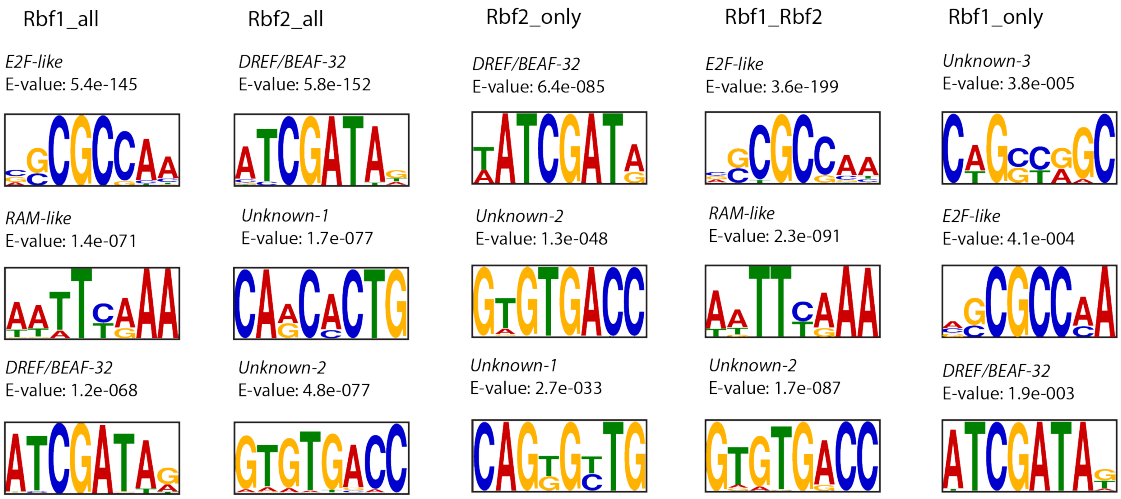


Figure 2-5: Motifs enriched in sequences associated with Rbf1- and Rbf2-bound regions of the genome that were located near TSS. MEME-ChIP was used to identify *de novo* overrepresented motifs. Top three motifs (ranked by E-value) were shown. E2F-like sequences were not enriched in Rbf2-associated regions, except where Rbf1 was also present.

Rbf2-alone targets include most ribosomal protein genes

We analyzed the nature of genes bound by Rbf1, Rbf2, or both Rbf1/Rbf2 using the DAVID gene ontology annotation database (Huang et al., 2009a and 2009b). Consistent with the known importance of RB proteins for cell cycle regulation, genes bound by both Rbf1/Rbf2 were significantly enriched for this category. In contrast, cell cycle related genes were not enriched in the set of genes bound solely by Rbf2; instead, the most significantly enriched category was that of ribosomal proteins (Figure 2-6A). The Rbf1-only group showed no significant enrichment of any gene class in this analysis (data not shown). To further characterize this enriched feature, we manually inspected Rbf1 and Rbf2 peaks on each of the 94 cytoplasmic ribosomal protein gene promoters (CRP) and 75 mitochondrial ribosomal protein gene promoters (MRP), observing that Rbf2 associated with a majority of the ribosomal protein gene promoters (Figure 2-6B). We also compared our results with the previous dREAM complex ChIP-chip study (Georlette et al., 2007), and found that some, but not all dREAM complex components co-occupy with Rbf1/2 on ribosomal protein gene promoters (Figure 2-7). In our earlier study (Acharya et al., 2012), Rbf1 was found to bind multiple genes encoding components of conserved signaling pathways. In the current study, we found that Rbf2 also associates with a significant number of signaling pathway gene promoters (Figure 2-6B). ChIP-qPCR assays were performed on selected cell cycle, signaling pathway, and ribosomal protein targets, confirming the enrichment found in the ChIP-exo experiments (Figure 2-6C). Thus, Rbf2 appears to occupy a greater fraction of noncanonical targets such as signaling pathway and ribosomal protein genes, compared to Rbf1, which is present together with Rbf2 at many canonical cell cycle related genes.

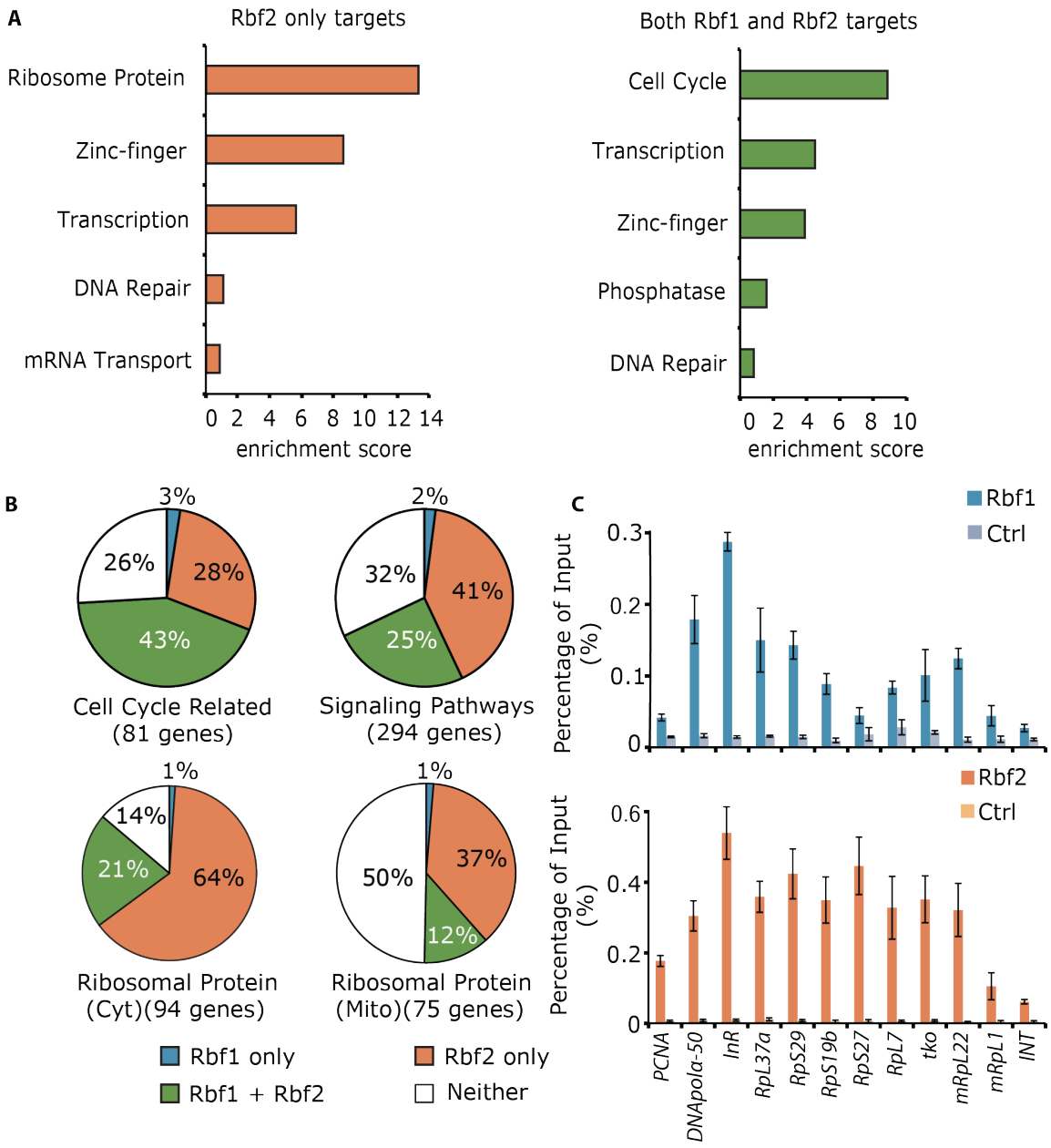


Figure 2-6: Enriched targeting by Rbf1 and Rbf2. (A) Genes bound by Rbf2 alone or by both Rbf1 and Rbf2 were functionally annotated using the DAVID database (Huang et al., 2009); values indicate enrichment scores. (B) Promoters of cell cycle related genes (annotated by flybase.org, The Interactive Fly), signaling pathway genes (Acharya et al., 2012), cytoplasmic ribosomal protein genes, and mitochondria ribosomal protein genes

Figure 2-6 (con't)

(Marygold et al., 2007) were manually inspected for Rbf1 and Rbf2 binding sites. (C) To validate ChIP-Seq results, manual ChIP of Rbf1 and Rbf2 on chromatin from 12-18 hr embryos was carried out on selected cell cycle (*PCNA*, *DNApolα-50*), signaling pathway (*InR*), cytoplasmic ribosomal protein (*RpL37a*, *RpS29*, *RpS19b*, *RpS27*, *RpL7*), and mitochondrial ribosomal protein (*mRpS12/tko*, *mRpL22*, *mRpL1*) targets using anti-Rbf1, anti-Rbf2, and pre-immune serum. An intergenic region (INT) was used as negative control.

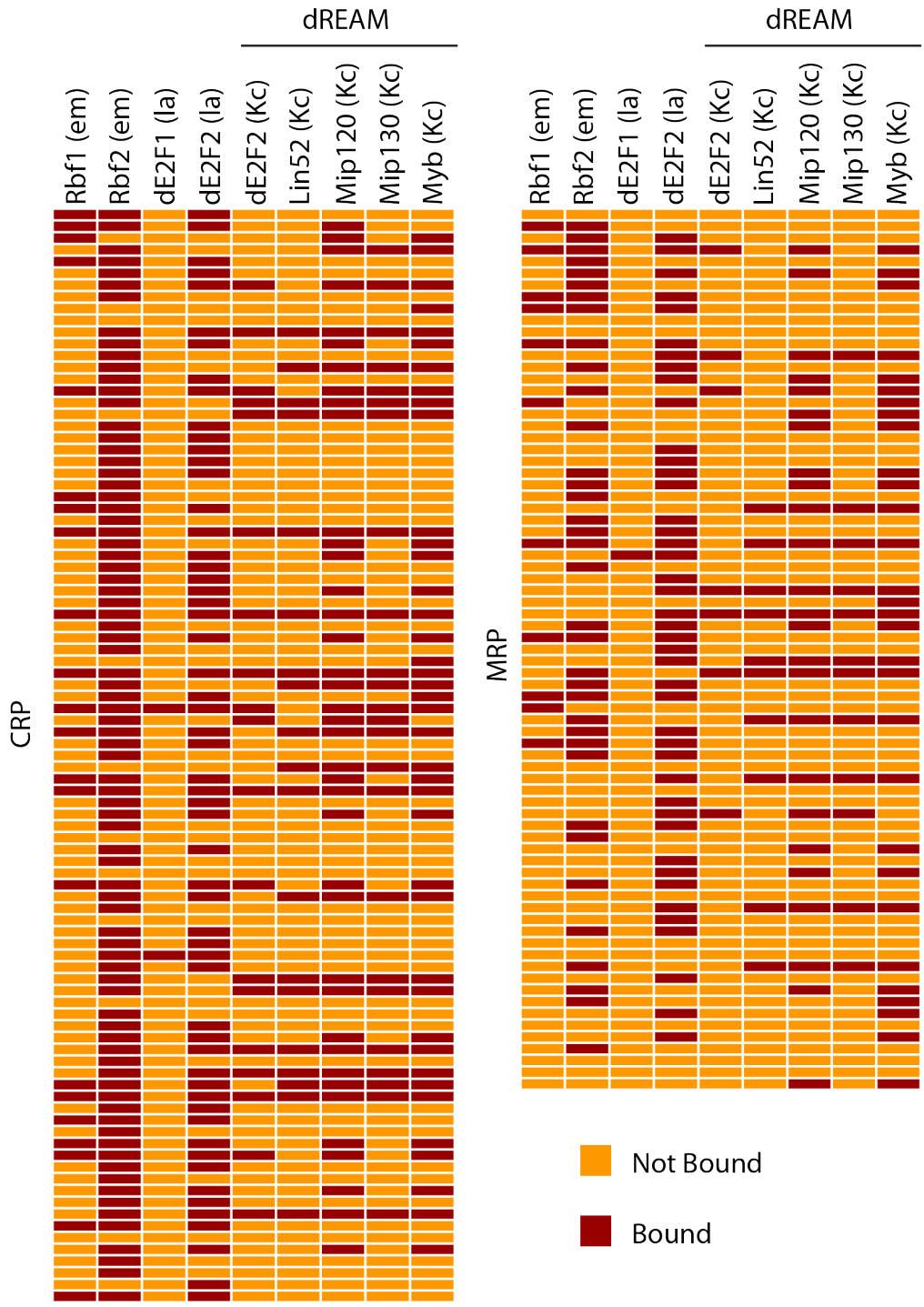


Figure 2-7: Association of Rbf proteins, E2F proteins, and dREAM complex proteins with ribosomal protein gene promoters. Rbf1 and Rbf2 data were from this study, dE2F1, dE2F2, and dREAM data were from previous studies (Georlette et al.,

Figure 2-7 (con't)

2007; Korenjak et al., 2012). Bound promoters were indicated in dark red color, unbound promoters were indicated in yellow color. ChIP data is from embryo (em), larva (la), and Kc cells (Kc).

Rbf2 shows differential repression activity on ribosomal protein gene promoters

To determine the regulatory significance of Rbf1 and Rbf2 binding at ribosomal protein promoters, we selected several genes for further functional characterization. Six promoter-proximal regions from cytoplasmic and mitochondrial ribosomal protein genes were cloned into a luciferase reporter, and the effects of Rbf1, Rbf2, dE2F2, or a combination of Rbf2 plus dE2F2 were tested in Drosophila S2 cells. As expected, transcription from the *PCNA-luc* reporter was repressed by Rbf1, dE2F2, and Rbf2/dE2F2 (Figure 2-8). In contrast, none of the ribosomal protein gene promoters were repressed by Rbf1, even though these particular promoters have robust Rbf1 signals in the embryo. Notably, overexpression of Rbf2 alone repressed the *mRpS12/tko* promoter ~25%, with repression increasing to ~50% with co-expression of dE2F2. Overexpression of dE2F2 alone decreased *RpL37a* promoter activity by about one-third, with a modest but reproducible ~15-20% repression observed on *RpS29* and *mRpL22* promoters. These latter promoters were not sensitive to Rbf1 or Rbf2 overexpression alone. The *mRpL1* or *RpS14b* promoters were not repressed to any extent by any of the overexpressed proteins, and in fact transcription of these reporters was mildly stimulated. Thus, unlike the classical RB cell cycle target *PCNA*, whose expression dynamically varies during cell growth, regulation of these non-canonical ribosomal protein gene promoters is more restrained. This behavior is consistent with the similarly modest but reproducible regulation of these genes under growth-limiting or stress conditions (Gasch et al., 2000; Causton et al., 2001; Gershman et al., 2007; Miller et al., 2011). As central mediators of global protein expression, small changes in ribosomal protein expression

are predicted to have significant and pleiotropic effects (Steffen et al., 2012; Xue and Barna, 2012; Woolford and Baserga, 2013; Hasygar and Hietakangas, 2014).

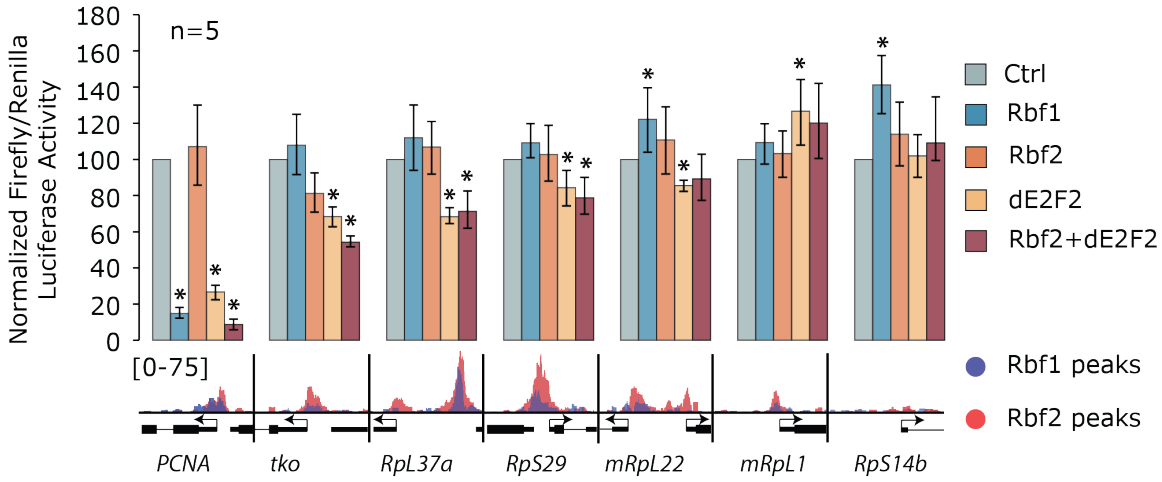


Figure 2-8: Transcriptional responses of Rbf targeted genes in reporter gene assays. Rbf1, Rbf2, dE2F2, or Rbf2/dE2F2 were overexpressed in cells containing reporters with promoter regions of indicated genes. ChIP occupancy by Rbf1 and Rbf2 is shown along with gene structure. Rbf1 showed repression activity only on *PCNA*. Rbf2 and/or E2F2 significantly repressed *PCNA*, *mRpS12*/*tko*, *RpL37a*, *mRpL22*, and had modest repression on *RpS29*. Activity of *mRpL1* and *RpS14b* promoters was not significantly repressed by any treatment. The increase in expression may be due to indirect effects, particularly for *RpS14b*, which is not bound by these proteins in ChIP assays. (*p-value < 0.05)

Rbf1 represses ribosomal protein gene expression in vivo

To further examine the significance of Rbf1 association with ribosomal protein gene promoters, we performed RNA-seq of larval wing discs that were engineered to overexpress Rbf1 (Elenbaas et al., 2014). Globally, a majority of the ribosomal protein genes showed modest reductions in expression, with only a few showing an increase (Figure 2-9). Six ribosomal protein genes were significantly repressed by Rbf1 in this developmental context, showing decreases of 20-35% (Figure 2-10A), similar to the repression observed on cell cycle genes, including *PCNA*, *DNApolα-50*, and *Mcm5* (Figure 2-10B). Consistent with the Rbf1 overexpression data, knocking-down *rbf1* alone, or *rbf1* with *rbf2* in cell culture significantly increased cell cycle genes expression, and widely induced ribosomal protein genes expression (Figure 2-10C). Knocking-down *rbf2* alone or together with *de2f2* did not have much impact on the ribosomal protein genes, although some of these gene promoters were significantly repressed by Rbf2/dE2F2 *in vitro* (Figure 2-8). Interestingly, among this set of ribosomal protein genes, only *RpL13* was bound by Rbf1 and Rbf2 in embryos and larvae (Acharya et al., 2012; Korenjak et al., 2012; and this study). We speculate that some of these genes not found to bind the corepressor in the embryo may bind Rbf1 specifically in the rapidly proliferating cells of the wing disc, or alternatively, these genes may harbor lower levels of Rbf1 that were not called as peaks in our analysis. Indeed, a number of these promoters contain DNA motifs such as DREF and RAM that were also enriched under Rbf1 peaks, and which may be diagnostic of Rbf1 function (Acharya et al., 2012).

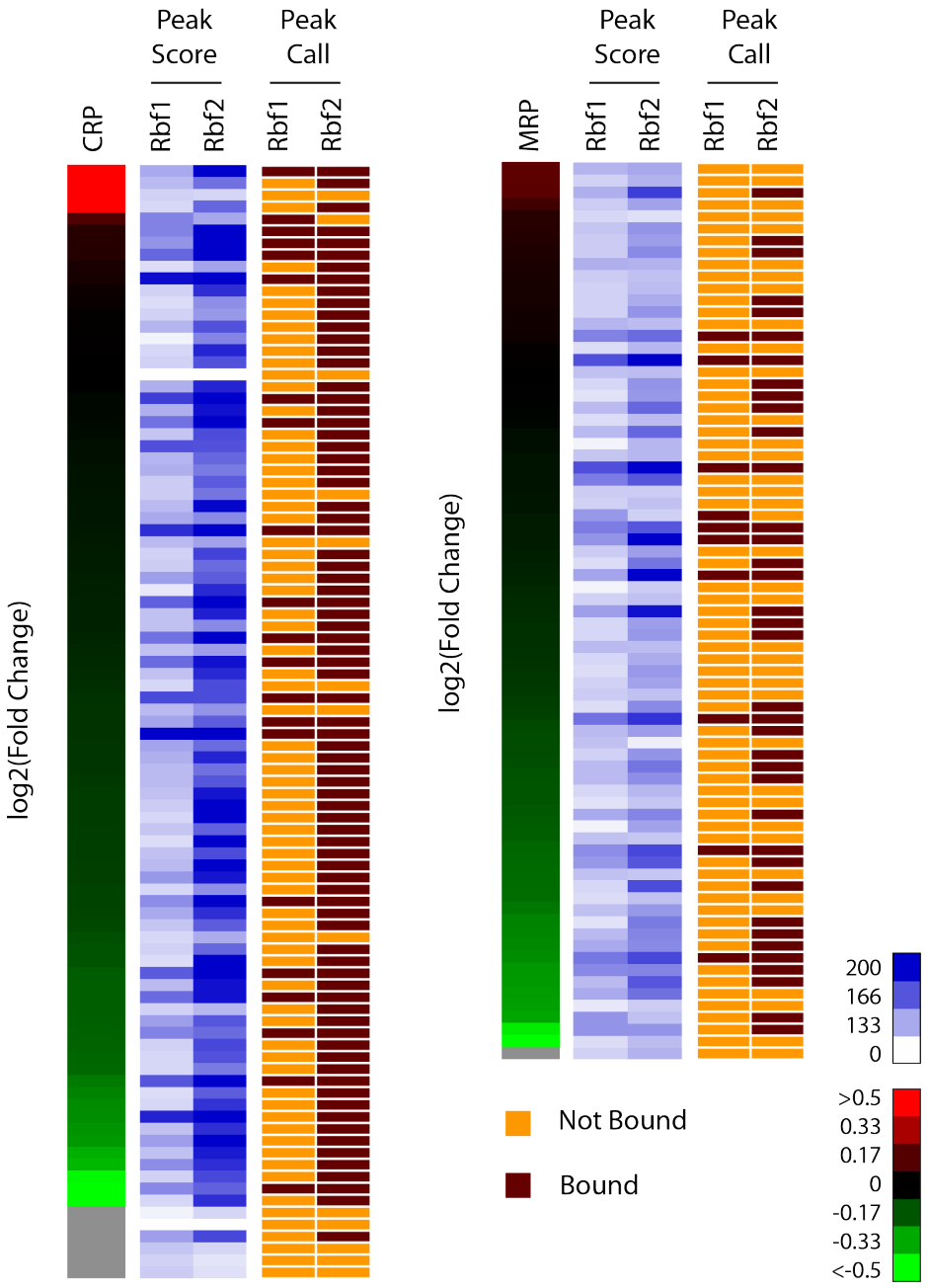


Figure 2-9: Global analysis of ribosomal protein gene expression in larval wing imaginal discs with overexpression of Rbf1 (red, increased expression; green, decreased expression). For each gene, information about Rbf1 and Rbf2 association shown at right, including peak scores and peak call, determined by HOMER software as described in Materials and Methods.

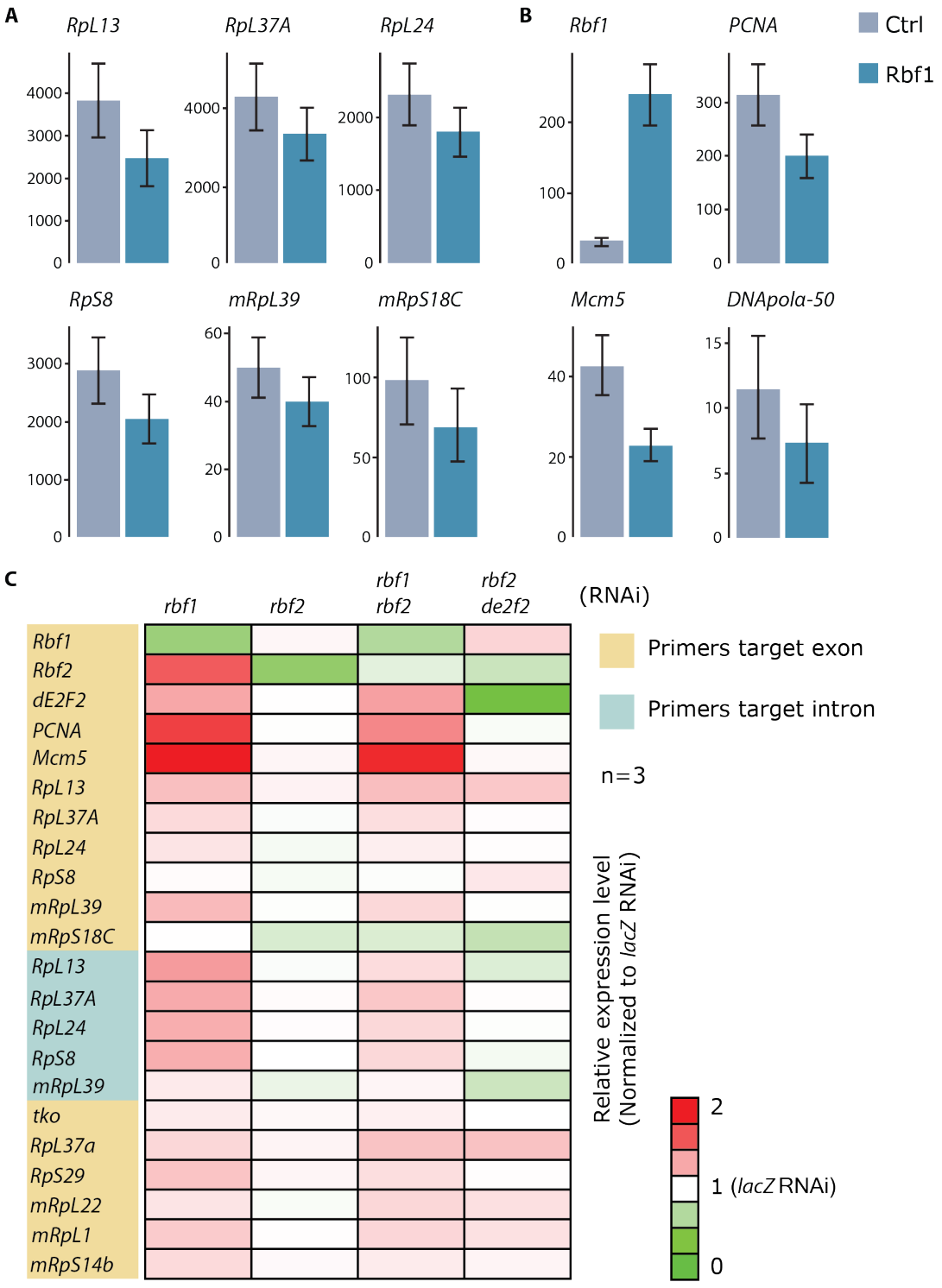


Figure 2-10: Rbf1-mediated reduction of ribosomal protein gene expression in wing imaginal disc tissue.

Figure 2-10 (con't)

(A) Six ribosomal protein genes were significantly repressed in response to overexpression of Rbf1 (p-value < 0.007, q-value < 0.05) (B) Cell cycle genes were repressed by Rbf1 overexpression. The y-axis indicates FPKM value (fragments per kilobase of transcript per million), error bars indicate cross-replicate variability and measurement uncertainty (Trapnell et al., 2012). Rbf1 was expressed in wing discs of third instar larvae under control of the *Pen*>Gal4 driver; three biological replicates were conducted and analyzed by RNA-seq, as described in Materials and Methods.

Enrichment of BEAF-32 motifs in Rbf bound promoters

Our analyses of the Rbf and E2F genomic binding profiles revealed many Rbf2 target genes that were not bound by E2F factors, and therefore we tested whether there was evidence for other transcription factors associated with Rbf2 bound regions on target promoters. To identify relevant motifs, we used the STAP program, which correlates ChIP signal intensity with presence of overrepresented motifs for known transcription factors (He et al., 2009). Globally, a few motifs showed strong correlation with Rbf1 and Rbf2 peaks, including the E2F and DREF motifs that we previously demonstrated to be enriched at Rbf1 binding sites (Acharya et al., 2012) (Figure 2-11). Viewed as separate classes, those genes annotated as "cell cycle related", "signaling" and "cytoplasmic ribosomal protein" also showed a strong enrichment for the E2F motif (Figure 2-12~14). Promoters from cell cycle and ribosomal protein genes were also enriched in a variety of other motifs, presumably related to their unique regulation (Figure 2-12, 2-14~15). However, genes representing conserved signaling pathways were not strongly enriched for additional motifs, likely because the divergent promoter sequences have very diverse regulatory properties (Figure 2-13). Unexpectedly, we found motifs for BEAF-32, an insulator binding protein, significantly correlated with both Rbf1 and especially with Rbf2 peaks (Figure 2-11). BEAF-32 binding sites measured in Drosophila 0-8-hr old embryos significantly overlap with Rbf1 and Rbf2 peaks (Figure 2-16A, B), with co-occupancy found for one-third of the Rbf2 and just over one-quarter of Rbf1 sites (Figure 2-16B) (Yang et al., 2012). A similar overlapping was also observed for BEAF-32 binding sites in S2 cells (Figure 2-16B) (Schwartz et al., 2012). Other insulator proteins, such as CP190 also co-occupy Rbf2 binding sites similar to BEAF-32, while the overlapping between Rbf2 and CTCF was less significant (Figure 2-16B) (Schwartz

et al., 2012). Focusing specifically on ribosomal protein gene promoters, BEAF-32 binding sites were significantly enriched, especially on Rbf2-bound genes (Figure 2-16C).

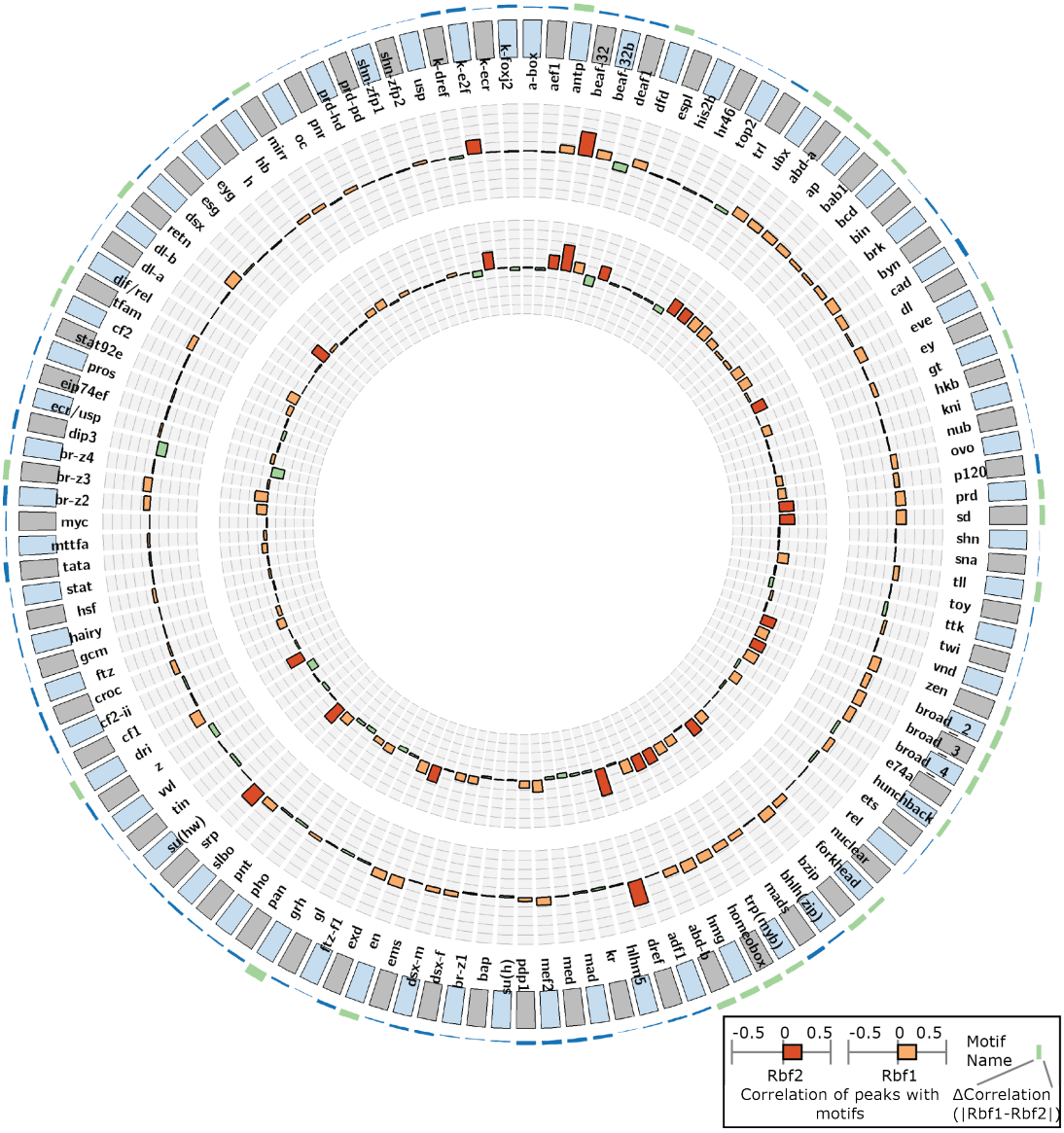


Figure 2-11: Enrichment of specific DNA binding protein motifs under peaks for Rbf1 and Rbf2 genome-wide. STAP results from 127 motifs were plotted in circular diagrams. The first histogram below the circumference shows Pearson correlation scores of individual motifs with Rbf1 ChIP-exo peak intensity, and the inner histogram in the circle shows the Rbf2 data. Strong enrichment for E2F, DREF, and BEAF-32 motifs is noted for both Rbf1 and Rbf2. The scale is from -0.5 to 0.5 with baseline of 0 in the

Figure 2-11 (con't)

middle, high scores (>0.19) are indicated in dark orange color (implying strong correlation), medium scores are in orange, and negative correlations are in green. The histogram outside the circumference shows the score differences between Rbf1 and Rbf2. The histogram is of light green color by default. Correlation-difference values lower than 0.04 are in blue color, implying those motifs are correlated with both Rbf1 and Rbf2 at similar level; correlation-difference values higher than 0.14 are in dark red color as seen in Figure 2-12~15, implying those motifs correlate with Rbf1-binding, but not Rbf2-binding, and vice versa. The Pearson correlation scores were calculated on the whole fly genome. For data in the four functional classes (cell cycle, signaling, and ribosomal protein genes - cytoplasmic and mitochondrial), see Figure 2-12~15.

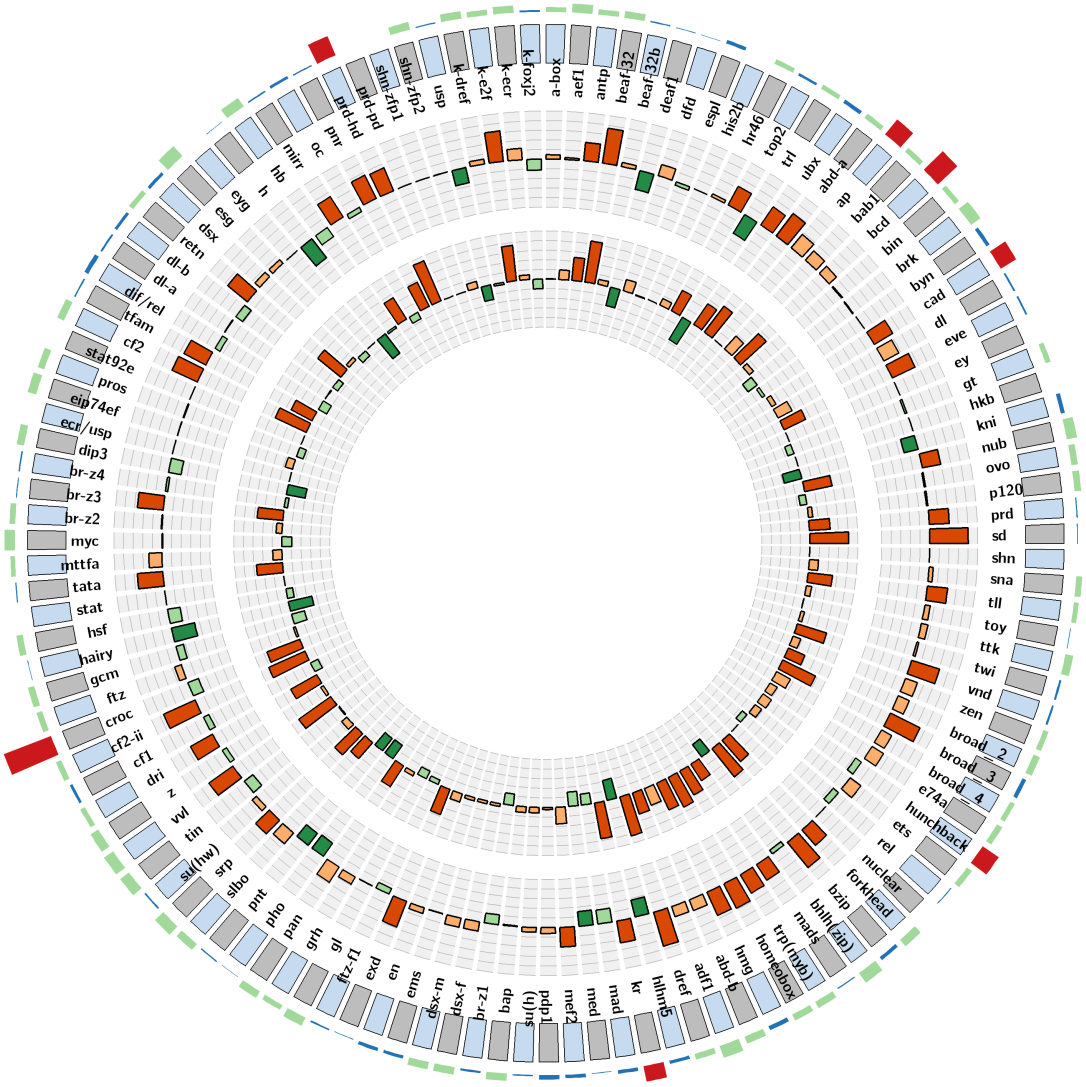


Figure 2-12: Enrichment of 127 motifs on selected cell cycle gene promoters. Inner wheel, Rbf2; middle wheel, Rbf1, outer wheel, differences between Rbf1 and Rbf2.

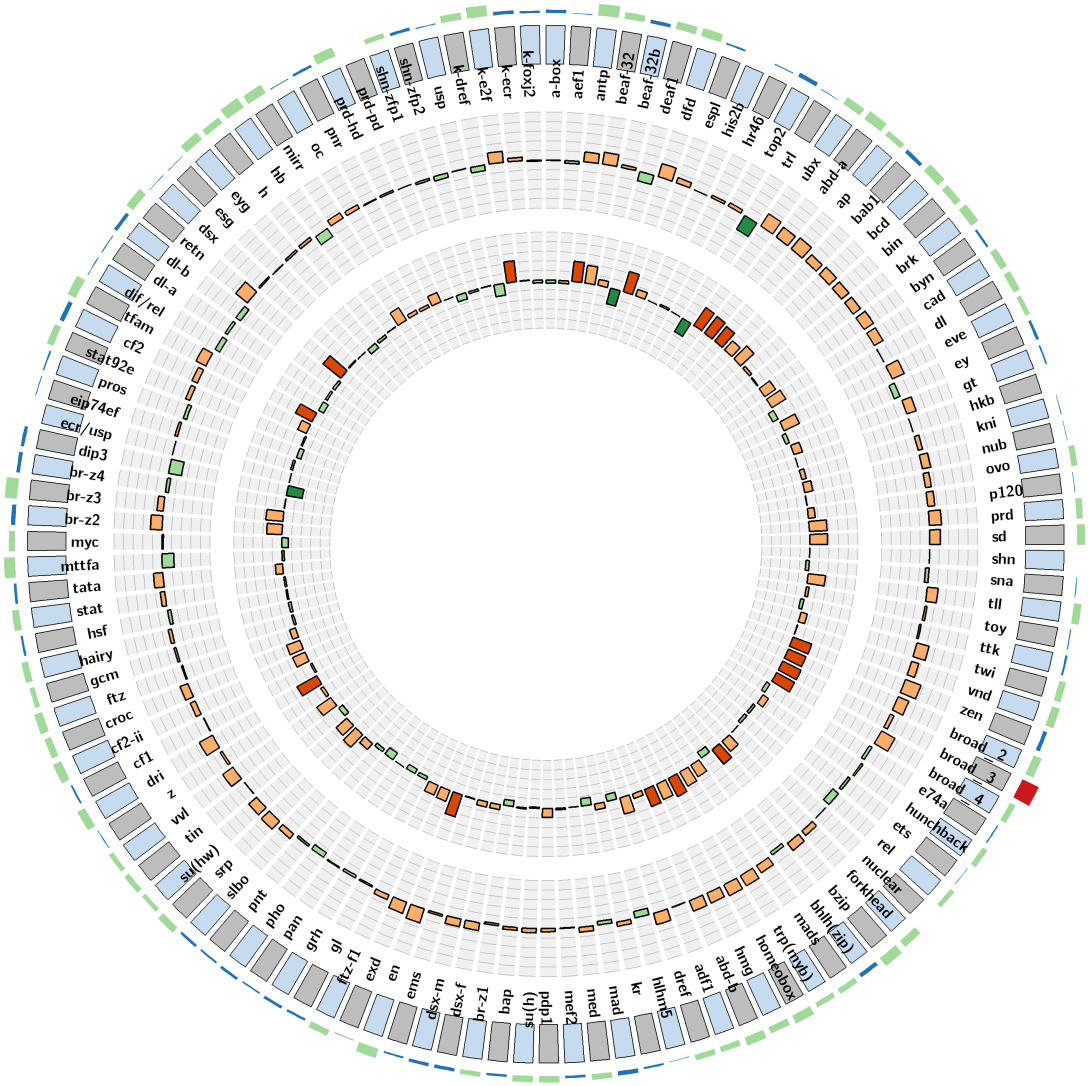


Figure 2-13: Enrichment of 127 motifs on selected signaling pathway gene promoters. Inner wheel, Rbf2; middle wheel, Rbf1, outer wheel, differences between Rbf1 and Rbf2.

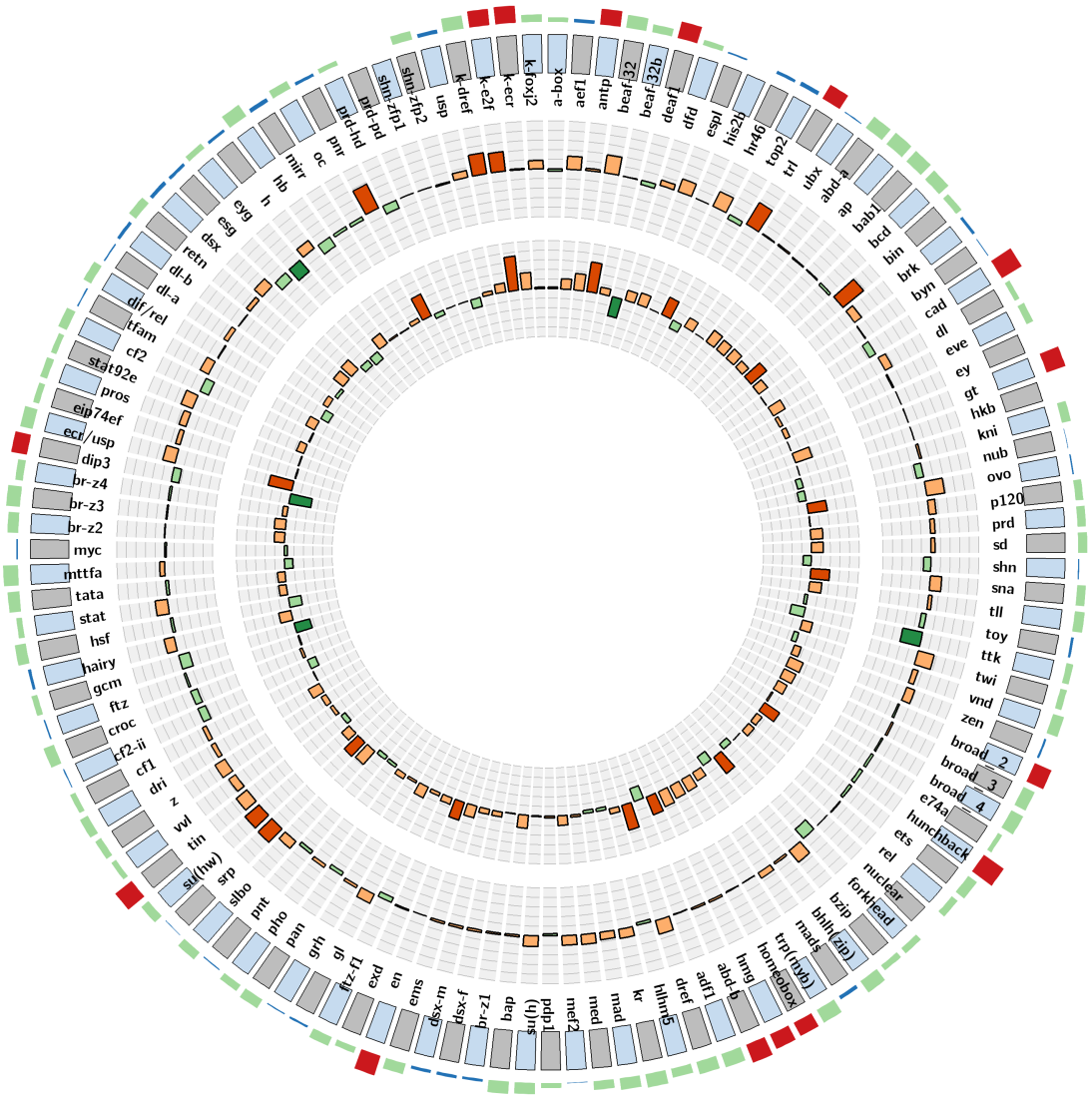


Figure 2-14: Enrichment of 127 motifs on selected cytoplasmic ribosomal gene promoters. Inner wheel, Rbf2; middle wheel, Rbf1, outer wheel, differences between Rbf1 and Rbf2.

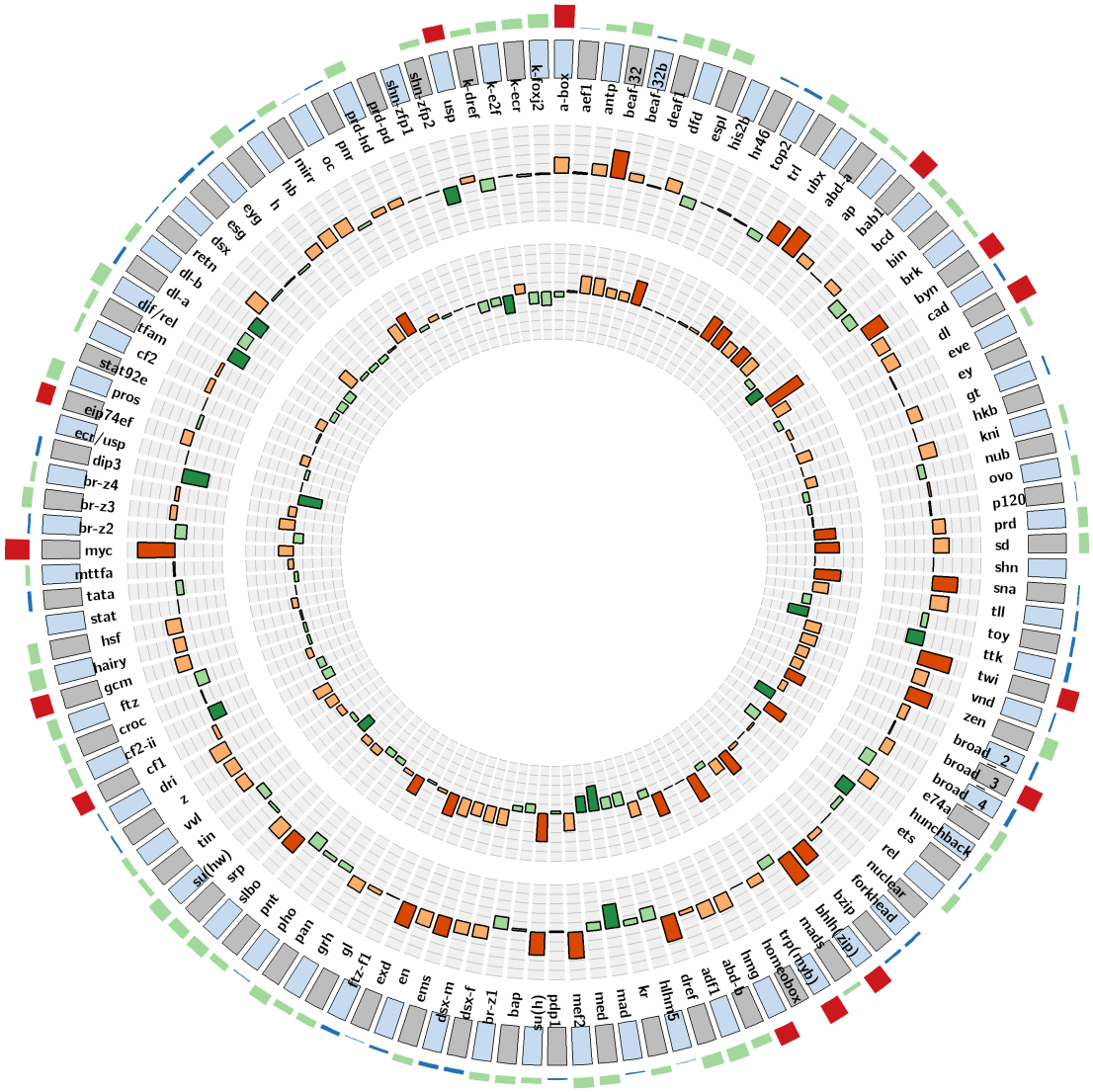


Figure 2-15: Enrichment of 127 motifs on selected mitochondrial ribosomal gene promoters. Inner wheel, Rbf2; middle wheel, Rbf1, outer wheel, differences between Rbf1 and Rbf2.

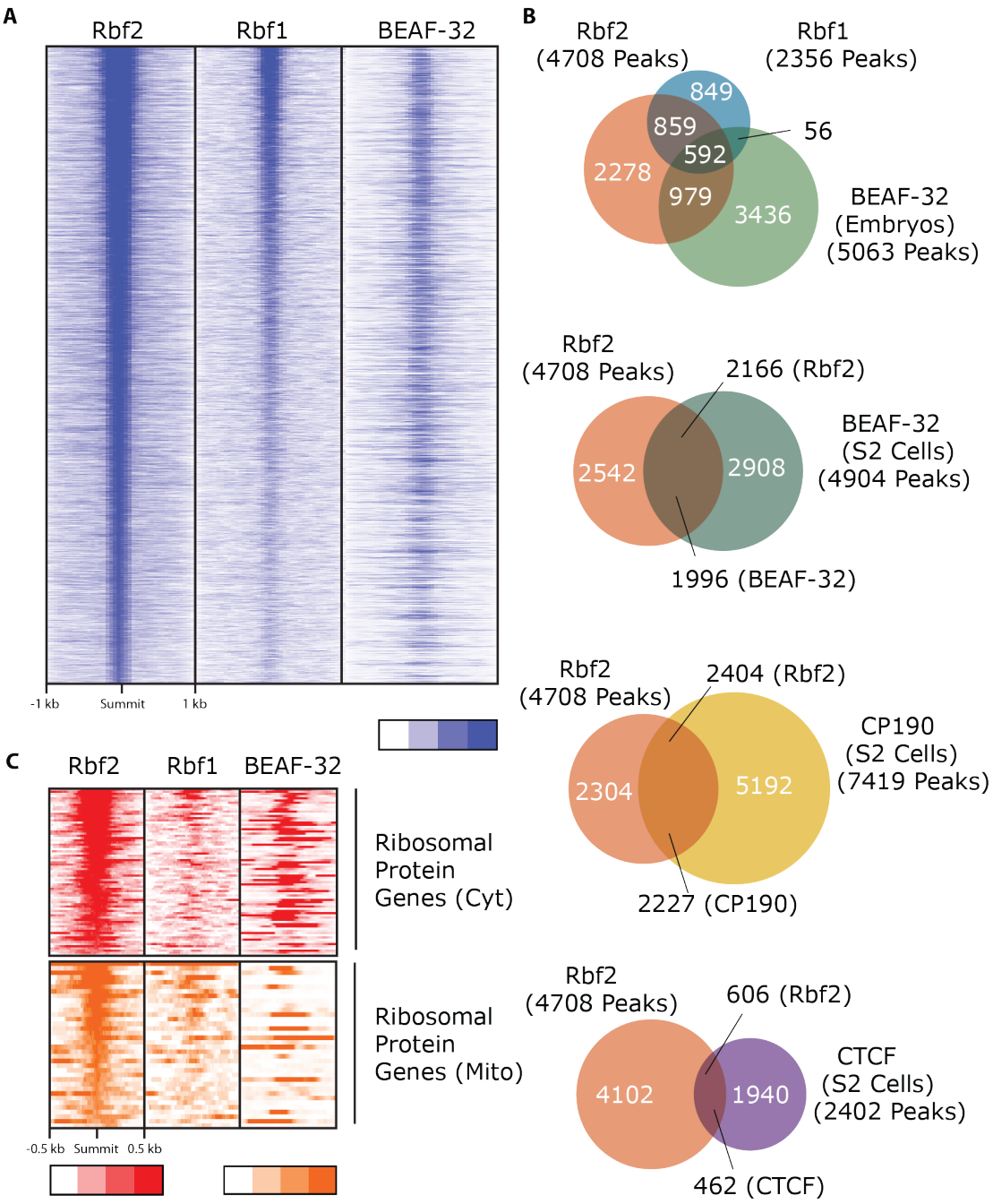


Figure 2-16: Correlation between Rbf2, Rbf1, and BEAF-32 ChIP signals (Yang et al., 12). (A) Heat map centered on the Rbf2 peak summits, and sorted by Rbf2 peak scores on all genomic regions bound by Rbf2. (B) Comparison of Rbf2, Rbf1, and BEAF-32 (Yang et al., 12) binding sites. The overlap between Rbf2 and BEAF-32 was

Figure 2-16 (con't)

statistically significant (log(p) = -5763). (C) Correlations between positions of binding of Rbf2, Rbf1, and BEAF-32 are shown in ribosomal protein promoter regions, using heat maps centered on the Rbf2 peak summits, and sorted by Rbf2 peak scores.

Some ribosomal protein gene promoters exhibit non-canonical Rbf recruitment

Our discovery of genes uniquely bound by Rbf2 but not Rbf1, and the lack of E2F-like motifs within these promoter regions, prompted us to test whether Rbf2 might be recruited to promoters by alternative factors. We tested whether Rbf2 recruitment would therefore be dependent on dE2F/dDP proteins or BEAF-32 in cultured cells. We depleted *de2f1*, *de2f2*, *dDP* or *BEAF-32* in Drosophila Kc cells with double-stranded RNA, followed by ChIP for Rbf2. (Figure 2-17A, B). The knockdown was sufficient to substantially deplete endogenous gene expression levels, leading to loss of expression of cell cycle genes *PCNA* and *Mcm5* in the cases of *de2f1* and *dDP* knockdown (Figure 2-17C). We examined promoters from cell cycle genes (*DNApolα-50*, *PCNA*), signaling pathway genes (*InR*, *Thor*), and ribosomal protein targets either bound by both Rbf1 and Rbf2 or Rbf2 alone. Knockdown of *BEAF-32* had no effect on Rbf2 recruitment on any promoter, even those with the highest BEAF-32 binding signals (Figure 2-17A, B). Thus, Rbf2 and BEAF-32 may bind to these promoters independently. By contrast, knockdown of *de2f2* or *dDP* substantially reduced the Rbf2 signal on the *RpS19b*, *RpS29*, *mRpL22*, *mRpS12/tko InR*, *PCNA*, *DNApolα-50*, and *RpL37a* promoters (Figure 2-17A), consistent with the previously described Rbf2-dE2F2-dDP recruitment mechanism (Stevaux et al., 2002). Significantly, for the *Thor* gene and eight other ribosomal gene promoters tested, the *de2f2/dDP* knockdown showed little to no effect on Rbf2 interaction (Figure 2-17B). It is interesting that most of those promoters were not bound by Rbf1, and a previous study also suggested they were not bound by dDP (Ambrus et al., 2013). Interestingly, on a number of promoters, we observed a modest increase of Rbf2 signal upon *de2f1* knockdown, possibly because of competition

between dE2F2/Rbf2 and dE2F1 on some Rbf targets. We repeated this ChIP experiment in Drosophila S2 cells, and found that Rbf2 binding on these Rbf2-alone ribosomal protein gene promoters was also not affected by *de2f2/dDP* knockdown (data not shown). These results suggest that Rbf2 interacts with some promoters via an E2F/DP-independent mechanism.

To determine whether ribosomal protein gene promoters bound preferentially by Rbf2 may have different transcription factor binding sites, we analyzed the occurrences and affinities of E2F-, DREF-, and FOXJ2-like motifs that previously had been shown to be enriched on Rbf1 bound regions (Acharya et al., 2012). We found that promoters bound uniquely by Rbf2 have lower binding scores for E2F, DREF, and FOXJ2 (Figure 2-18A-C). Surveying the entire set of sites uniquely bound by Rbf2 genome-wide, we found a similar lack of strong E2F sites (Figure 2-18D).

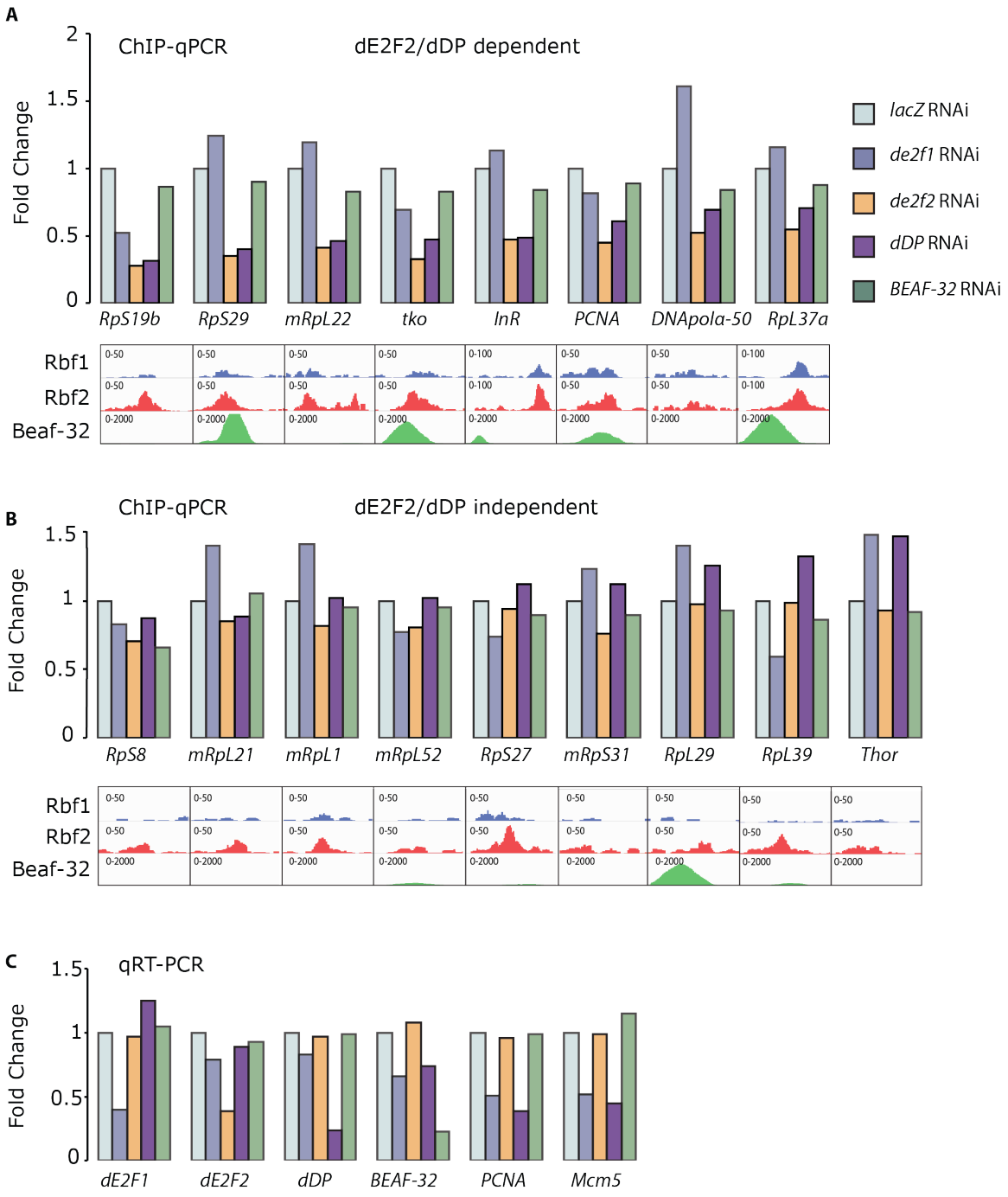


Figure 2-17: RNAi depletion reveals E2F/DP-dependent and –independent Rbf2 binding in cultured cells. (A) ChIP results for genes on which Rbf2 binding to promoters was affected by *de2f2* or *dDP* knockdown. (B) ChIP results for genes on

Figure 2-17 (con't)

which Rbf2 showed little or no loss of binding by similar depletions. These promoter had weak or nonexistent Rbf1 binding. ChIP recovery for factor depletion was normalized to levels obtained for *lacZ* control knockdown. (C) Knockdown efficiency of targeted mRNAs was ~60-70%, as revealed by RT-PCR. Consistent with this depletion, the *de2f1* or *dDP* knockdown strongly affects the expression of *PCNA* and *Mcm5* cell cycle genes.

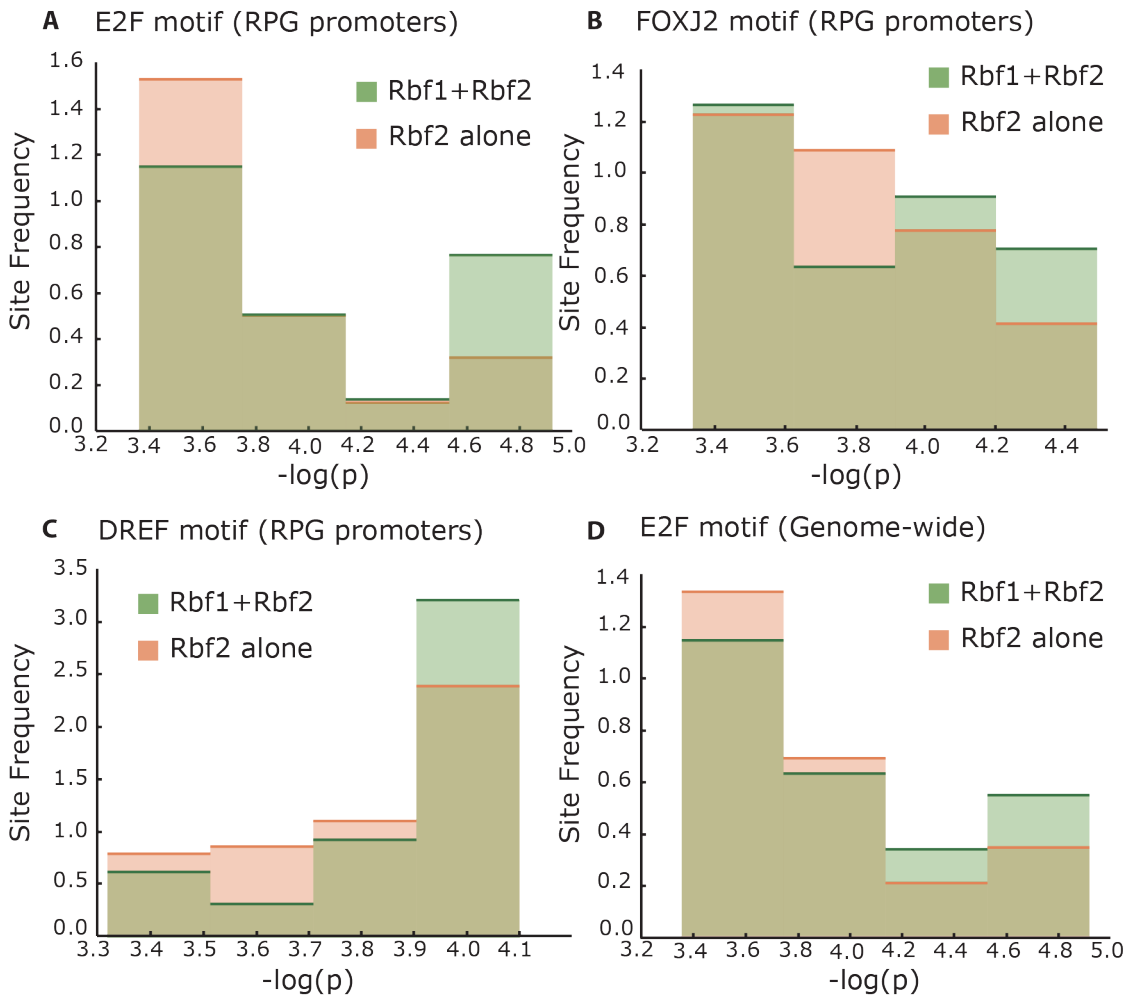


Figure 2-18: Distinct qualities of motifs associated with Rbf1+Rbf2 bound promoters, vs. those bound solely by Rbf2. (A) The E2F motif quality was highest on ribosomal promoters bound by both Rbf1 and Rbf2; -log p values indicated on horizontal axis, and frequency of occurrence on vertical axis. (B, C) Previously identified Rbf1-associated motifs DREF and FOXJ2 also show a tendency towards stronger sites in co-bound sequences. (D) The site strength of E2F motif was also found significantly shifted towards stronger sites in the Rbf1+Rbf2 promoters, compared to the Rbf2-alone promoters, when assessed genome-wide (p=1.48 e-09). A total of 120 motifs were tested for differential representation in the two classes of Rbf2 alone vs. Rbf1+2.

Association with ribosomal protein gene promoters is a conserved character for the RB family

To determine whether the widespread Rbf association with ribosomal protein gene promoters represents conserved regulatory interactions, we surveyed human RB and p130 protein ChIP-seq data in fibroblasts (Chicas et al., 2010), and *C.elegans* RB homolog protein Lin-35 ChIP-seq in larvae (Latorre et al., 2015). We inspected all human and *C.elegans* orthologs of *Drosophila* ribosomal protein genes, observing that a majority of the ribosomal protein gene promoters were bound by RB, p130, or Lin-35 (Figure 2-19). The high proportion of ribosomal protein genes targeted by these corepressors suggests that there may be a conserved role for these RB family proteins in regulating protein synthesis and growth.

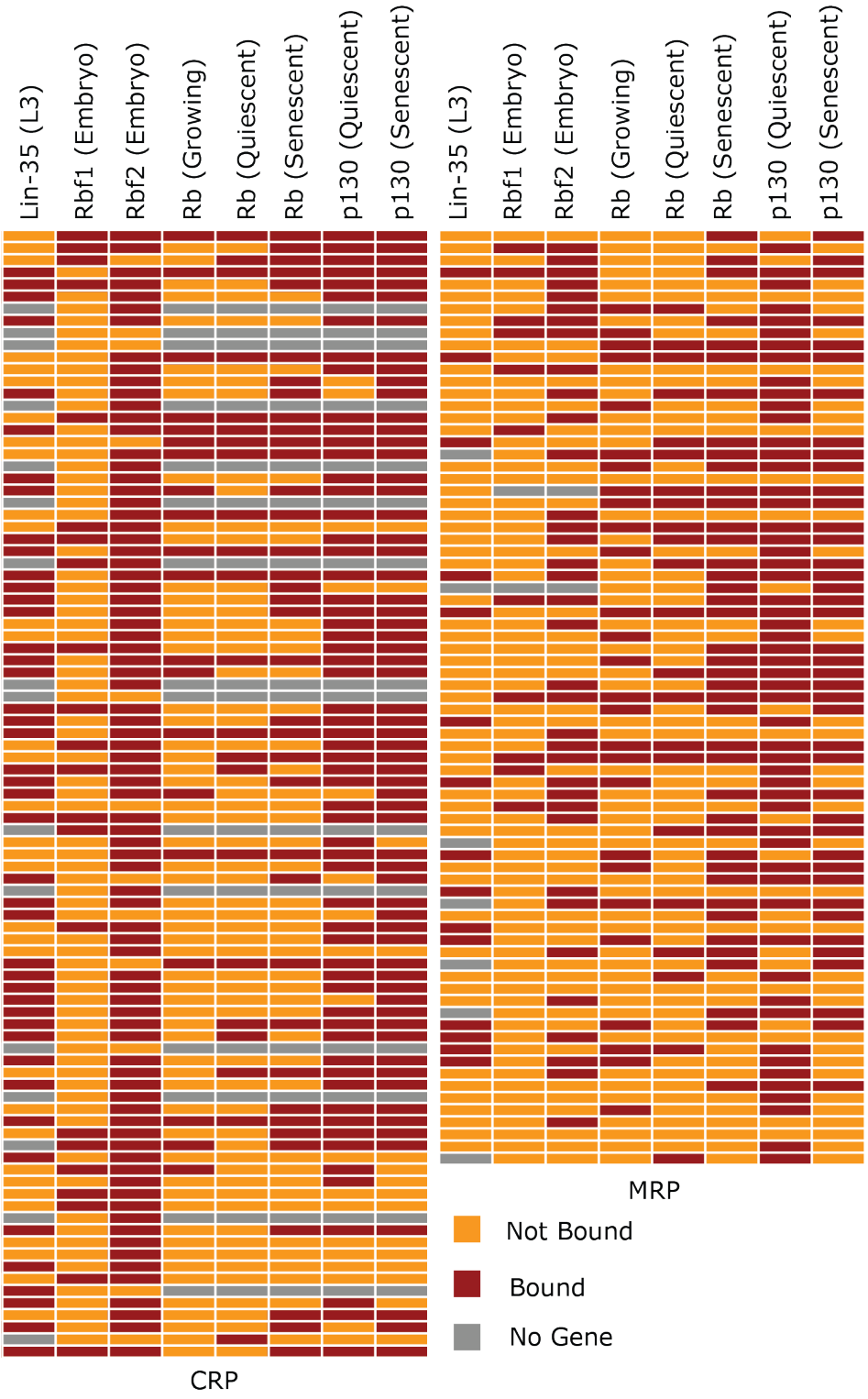


Figure 2-19: Retinoblastoma corepressor association with ribosomal protein genes is a conserved feature of RB proteins. Orthologous ribosomal protein genes were

Figure 2-19 (con't)

identified in *C.elegans*, *Drosophila* and human, and association with retinoblastoma family proteins in promoter regions noted by colored lines. Association of retinoblastoma proteins is noted for a large fraction of cytoplasmic and mitochondrial promoters. Binding to each promoter from the *C.elegans* and human data set was analyzed by uploading peak calling files from Latorre et al., 2015, and Chicas et al., 2010, and manually annotating peaks within 500 bp of the TSS.

Discussion

Retinoblastoma gene families have undergone diversification in multiple lineages (Cao et al., 2010; Gutzat et al., 2012). In metazoans, the RB family proteins of *Drosophila* and vertebrates have independently diversified; in the case of flies, Rbf2 has substantial differences in the C-terminus, which is thought to be a key domain for regulation and binding specificity. In vertebrates, RB similarly exhibits substantial differences in the C-terminus compared to the more ancestral p107 and p130 paralogs (Classon and Dyson, 2001). Thus, Rbf2 and RB represent evolutionary innovations, which may direct the regulation of unique sets of genes or respond to different environmental and developmental signals.

The ChIP-Seq comparison of Rbf1 and Rbf2 binding profiles revealed several unexpected features, given previous findings that Rbf2 co-occupies a number of promoters with Rbf1. First, there were approximately 2000 genes targeted uniquely by Rbf2. This pattern either represents the neofunctionalization of Rbf2 with acquisition of novel gene targets, or alternatively, many of these genes may be bound by the Rbf1 homolog in sister species, with a Rbf2 acquiring some of these interactions through subfunctionalization of Rbf1. Comparative functional studies will help to clarify this point. The unique binding of Rbf2 to some promoters runs contrary to an earlier model that suggested that Rbf2 would only interact with a subset of the genes bound by Rbf1, because Rbf2 was thought to bind preferentially to dE2F2, while Rbf1 was more promiscuous. Our bioinformatic analysis indicates that there are indeed distinct patterns of motifs present on Rbf2-only regions, including a depletion of strong E2F-like sites, suggesting that other transcription factors may direct Rbf2 recruitment. In mammals, RB

and p107 specificity is driven partially by differential contacts mediated by the C-terminal regulatory domains (Rubin et al., 2005; Julian et al., 2008; Cecchini and Dick, 2011; Dick and Rubin, 2013). Likewise, the unique C terminal domain of Rbf2 may allow interactions with different types of regulators. Previous genetic experiments showed a genome-wide depletion of Rbf1 binding in *dDP* mutant larvae, as well as loss of Rbf2 from select genes. Our results are consistent with these findings, in that those specific genes tested for Rbf2 association (such as *InR*) are E2F-dependent genes that are also bound by Rbf1. Just as mammalian RB has diversified its interactions with the genome through association with non-E2F factors, Drosophila Rbf2 may have alternative binding partners whose identities remain to be determined.

Despite the widespread binding of Rbf2 in the *Drosophila* genome, genetic analysis of *rbf2* has shown that flies lacking this gene are viable, unlike the lethal phenotype of *rbf1* mutants. Why is the *rbf2* gene evolutionarily retained throughout the *Drosophila* lineage, despite the modest phenotype? The genes exhibit similar, although not identical expression patterns, suggesting that both proteins are likely to be present in many tissues. One clue comes from the adult pattern of *rbf2* expression, which is concentrated in the ovary (Stevaux et al., 2002; Keller et al., 2005). Although *rbf2* nulls were healthy and viable, these mutants lay eggs at a considerably higher rate than wild-type controls (Stevaux et al., 2005). Reproductive output is doubtlessly under strong selection, and Drosophila egg laying is in fact tightly coupled to nutritional signals. Excessive resource allocation represented by high rates of egg laying under laboratory conditions may be reproductively disadvantageous over the life span of the individual. Thus, the presence of Rbf2 may modulate egg laying through fine-tuned transcriptional control of cellular

signaling genes, as well as control of core biosynthetic components, such as the ribosomal protein gene family.

Our study suggests that Rbf corepressors may be directly repressing transcription of ribosomal protein genes; interestingly, there is no precedent for direct negative regulation of this class of genes by transcriptional repressors. Previous studies have focused on the engagement of transcriptional activators at ribosomal protein gene promoters. In light of the central role that ribosome biogenesis plays in controlling global gene expression, it is rather surprising that this regulon would be controlled solely by positive inputs. Almost every regulated gene, from phages to bacteria to eukaryotic cells, features the combined action of both activators and repressors to achieve fine-tuned gene expression. The ribosomal protein genes represent a unique class that typically exhibits less variation in expression levels than developmentally-regulated genes, which may be completely silenced in many settings. Thus, typical transcriptional regulation of ribosomal protein genes may be rather subtle, but such modulation would nevertheless have pleiotropic consequences if not correctly executed. Global gene analyses typically focus on more dramatic fold changes than we observe here, thus this response may have been previously below the threshold considered to be significant (Dimova et al., 2003). The selective regulation of ribosomal protein genes noted in our study, whereby only a subset of promoters was bound or regulated, is consistent with previous findings that the regulation of mRNA levels of some ribosomal protein genes is more dynamic than others, likely because other layers of regulation ensure stoichiometric production of ribosome components (Miller et al., 2011). The heterogeneous composition of activators at ribosomal promoters may contribute to this

differential regulation; in mammals, the DRE motif for the DREF factor is found at many but not all ribosomal protein promoters, suggesting that common but not identical levels of regulation are probably at work (Yamashita et al., 2007). It is interesting that mammalian RB has been reported to directly regulate the activity of RNA polymerase I and III, providing a link for this cell-cycle regulatory protein to control the biosynthetic capacity of cells (Cavanaugh et al., 1995; Larminie et al., 1997; White, 1997; Hirsch et al., 2000 and 2004; Felton-Edkins et al., 2003; Gijdoda and Henry, 2013). A regulatory connection with ribosomal protein genes would ensure that all facets of ribosome production would be influenced by RB signaling. Just as misregulation of c-Myc, which plays a positive role in ribosome synthesis, is linked to cancer, this model provides a new perspective to the impact of retinoblastoma proteins in cancer, where both disturbances to cell cycle control as well as accumulation of biomass through control of ribosome genes would play critical roles in tumorigenesis (White, 2004 and 2005).

Materials and Methods

ChIP-exo

ChIP-exo was conducted using 12-18 hr old *yw* *Drosophila melanogaster* embryos (strain yw^{67}) collected and aged at room temperature. Fixing and chromatin preparation was carried out as described before (Acharya et al., 2012). Immunoprecipitations and sequencing were carried out by Peconic LLC (State College, PA), using highly specific polyclonal rabbit anti-Rbf1 or anti-Rbf2 serum as described (Keller et al., 2005).

Read mapping, peaking finding, visualization, and annotation

We obtained 13,453,984 reads for Rbf1 ChIP-exo, and 12,596,328 reads for Rbf2 ChIP-exo. Read mapping was conducted by Peconic LLC. using *Drosophila melanogaster* genome version R5/dm3. To identify Rbf1 and Rbf2 bound regions, assign these peaks to nearest genes, and classify these peaks to specific genomic regions, HOMER v3.12 software was used, with default settings for peak calling and annotation (Heinz et al., 2010). The peaks were visualized using IGV browser v2.2.5. We identified 2356 peaks for Rbf1 ChIP-exo, which were mapped to 1955 genes, and 4708 peaks for Rbf2 ChIP-exo, which were mapped to 3945 genes. To compare with dE2F1 and dE2F2 targets, peak information from dE2F1 and dE2F2 ChIP-chip in *Drosophila melanogaster* larvae (Korenjak et al., 2012) was annotated using HOMER. To compare Rbf1 and Rbf2 peaks with BEAF-32 peaks, raw bed file data for binding of BEAF-32 protein in 0-8 hr old *Drosophila melanogaster* embryos was obtained from (Yang et al., 2012) and peaks were calculated by HOMER using default settings. To compare Rbf1

and Rbf2 targets with human RB targets, human RB and p130 ChIP-seq peaks (Chicas et al., 2010) were annotated using HOMER with the hg18 genome, and their association with human ribosomal protein gene promoters were manually inspected by browsing the peak-calling files in IGV browser with hg18 genome. To analyze *C.elegans* RB homolog Lin-35 association with ribosomal protein gene promoters, peak-calling file for Lin-35 (Latorre et al., 2015) was visualized in IGV browser with WS220 genome, and the ribosomal protein gene promoters were manually inspected for Lin-35 binding. To compare overlapping peaks between different data sets, HOMER was used with overlapping threshold set at 100 bp.

De novo motif searching

To identify motifs associated with Rbf1 and Rbf2 targets indicated in Figure 2-5, the sequences of Rbf1 and Rbf2 binding regions that associate with TSS/promoter (by HOMER default, -1 kb to +100 bp) were extracted from *Drosophila* genome R5/dm3 on the UCSC Genome Browser and subjected to *de novo* motif searching using MEME-ChIP with default settings (Machanick and Bailey, 2011).

Validation of ChIP-exo peaks

To validate the enrichment of Rbf1 and Rbf2 on their canonical and non-canonical target genes, several genes were selected and association with Rbf proteins tested by ChIP-qPCR. 12-18 hr *yw Drosophila melanogaster* embryos were used to prepare chromatin for the immunoprecipitation, and three biological replicates were conducted as previously described (Acharya et al., 2012). Pre-immune sera Rbf1-226.0 and Rbf2-4.0 were used for negative controls (Keller et al., 2005). To directly compare Rbf1, Rbf2

targets with dE2F1, dE2F2 and dDP at the same developmental stage, ChIP-qPCR analysis was performed using the Rbf1, Rbf2 antibodies and pre-immune sera as described above, along with dE2F1, dE2F2, and dDP antibodies (gifts from Dr. Nicholas Dyson lab), in 12-18 hr embryos.

Gene ontology (GO) analysis

Rbf1 and Rbf2 associated genes identified using HOMER were subjected to GO analysis. The enrichment of GO terms was performed using the online tool DAVID (Huang et al., 2009a, 2009b) with KEGG_PATHWAY and SP_PIR_KEYWORDS. Eleven annotation clusters were identified for Rbf1-and-Rbf2 targets, three were identified for Rbf1-only targets, and seventeen were identified for Rbf2-only targets. The enrichment scores of the top five annotation clusters for Rbf1-and-Rbf2 targets, and Rbf2-only targets were plotted as shown in Figure 2-6A. The automated gene assignments by HOMER can arbitrarily assign peaks to one of two divergently transcribed genes, although the distance of Rbf peak to the more distal TSS may be close enough to be functionally important. Therefore, to identify all genes that may be likely transcriptional targets of the Rbf proteins, and to calculate the percentage of genes bound by Rbf1 or Rbf2 in different functional groups, Rbf1 and Rbf2 binding regions were manually inspected in the promoter regions of 81 selected Cell Cycle Genes (CCG), 294 Signaling Pathway Genes (SPG) (Acharya et al., 2012), 94 Cytoplasmic Ribosomal Protein (CRP) genes, and 75 Mitochondrial Ribosomal Protein (MRP) genes (Marygold et al., 2007). A few additional genes were therefore added to the dataset of Rbf1 or Rbf2 potential targets from HOMER.

Reporter constructs and luciferase assay

To analyze Rbf1 and Rbf2 activity on ribosome protein gene promoters, promoter regions of *RpL37a* -788 to +132, *RpS29* -369 to +60, *mRpS12/tko* -1074 to +155, *mRpL22* -478 to +79, and *mRpL1* -420 to +47 containing Rbf1 or Rbf2 binding regions with transcription initiation sites were cloned into *Asc*I and *Sal*I sites in pAC2T-luciferase vector (Acharya et al., 2010). The *PCNA*-luciferase reporter was used as a positive control (Acharya et al., 2010), and promoter region of *RpS14b* -348 to +33 that is bound by neither Rbf1 nor Rbf2 was used as a negative control. 100 ng of the reporters were co-transfected with 250 ng of pRL-CMV Renilla luciferase reporter and 250 ng pAX-*rbf1* (Acharya et al., 2010), or 250 ng pAX-*rbf2*, with or without 200 ng pIE4-*myc-de2f2*. For the control group, equal amounts of pAX were used instead of pAX-*rbf1*, pAX-*rbf2*, or pIE4-*myc-de2f2*. Luciferase assays were conducted as described before with three biological replicates (Acharya et al., 2010), a *t*-test was used to analyze the statistical significance.

RNAi and ChIP

Double-stranded RNA (dsRNA) for *lacZ*, *de2f1*, *de2f2*, *dDP*, and *BEAF-32* were prepared as described before (Ullah et al., 2007). 40 million Drosophila Kc cells were treated with dsRNA at concentration of 10 μg/ml for four days. ChIP from Kc cells was performed as described (Hirsch et al., 2004). For qPCR analysis shown in Figure 2-10C, 1 million Drosophila S2 cells were treated with dsRNA for *lacZ*, *rbf1*, *rbf2*, *rbf1+rbf2,* and *rbf2+de2f2* at concentration of 10 μg/ml for four days. Total RNA was isolated

using TRIzol (Invitrogen), cDNA was prepared using ABI High Capacity cDNA RT Kit (Life Technologies) following the manual with 2 μg of total RNA.

RNA-seq

The UAS-*rbf1* fly line was constructed as previously described (Zhang et al., 2014). *Pendulin*-Gal4 driver line (Stock Number: 113920) and UAS-*GFP* line (Stock Numbers: 35786) were obtained from Bloomington Stock Center. 100-150 wing imaginal discs were dissected from third-instar larvae of *Pen*Gal4>UAS *rbf1* and *PenGal4*>UAS *GFP* flies. Total RNA was isolated using TRIzol (Invitrogen) followed by cleanup steps using RNeasy Mini kit (Qiagen). 1-4 μg total RNA from three biological replicates was collected. Library preparation and sequencing was conducted by the Research Technology Support Facility (Michigan State University) using an Illumina HiSeq2500. All standard libraries were created using Illumina TruSeq kits and reagents following the manufacturer's protocols. In brief, polyA mRNA was isolated from total RNA, chemically fragmented, and then reverse transcribed to form double stranded cDNA. The cDNA was then end repaired, A-tailed, adapter ligated and amplified to create the final library. A bead-based size selection was performed to target final library molecules with a mean size of 500 base pairs. All libraries were then quantified on a Qubit Fluorometer (Life Technologies) and run on an Agilent BioAnalyzer to determine final size and purity of the library. Final concentration was then determined by qPCR using the KAPA Illumina Library Quantification Kit (KAPABiosystems). Libraries were appropriately diluted and loaded onto the flow cell for sequencing on the Illumina HiSeq2500 following the manufacturer's protocols. RNA-seq reads were mapped using TopHat v2.0.13 and analyzed using Cufflinks v2.2.1 (Trapnell et al., 2012). Analyzed

results were visualized using R v2.15.3 with CummeRbund package as described (Trapnell et al., 2012).

Data set preparation for STAP analysis

For all (15,829) *D. melanogaster* genes, their locations and DNA sequences from 500bp upstream to Transcription Start Site (TSS) were retrieved from Flybase and UCSC database (dmel-5.48 Flybase release). For the four functional groups: Cytoplasmic Ribosomal Protein (CRP) genes, Mitochondrial Ribosomal Protein (MRP) genes, Cell Cycle Genes (CCG) and Signaling Pathway Genes (SPG) the same data was extracted and processed separately. The quantitative ChIP enrichments were calculated from the .wiggle files computed by MACS v1.4.2 (Zhang et al., 2008) by taking a maximum average signal over a sliding window within the 500bp upstream of the TSS both for Rbf1 and Rbf2 ChIP experiments. Position Weight Matrices (PWMs) of 127 motifs of Transcription Factors Binding Sites (TFBS) compiled from literature were used.

Testing for motif association with ChIP enrichment

The STAP program was used to test which TFBS affinity scores correlate with ChIP enrichment for the DNA sequences upstream of the TSS (He et al., 2009). For individual motif analysis, STAP was run with default parameters (sequence file, data file and motif file) with the option of co-operative binding set to 0 for each of the 127 motifs. The Pearson correlation between predicted binding and observed binding (in the cases of both Rbf1 binding and Rbf2 binding) for each of the 127 motifs was plotted using Circos (Krzywinski et al., 2009).

Motif strength assessment

Using a pipeline programmed in Python, MAST (Bailey and Gribskov, 1998) was run for each of the 127 motifs on the database of 15829 sequences to obtain each motif's occurrences, with maximal p-value=0.0005 and E-value=10000. All motif occurrences for each TFBS were extracted from the mast output file. Then we divided the sequences into two groups: co-bound by Rbf1+Rbf2 and bound by Rbf2 only. This procedure was repeated for the genome wide set of sequences as well as the ribosome associated sequences only. We compared the distribution of the strength of non-overlapping binding sites reported by MAST (as p-values). Negative logarithms (-log10) of those p-values (the lower p-value the stronger value, hence the reverse logarithm) were plotted as histograms for both "Rbf1+Rbf2" and "Rbf2-only". Mann-Whitney rank test was performed on the observed two groups with the threshold of one-sided p-value<0.05.

REFERENCES

REFERENCES

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

Acharya, P., Raj, N., Buckley, M. S., Zhang, L., Duperon, S., Williams, G., Henry, R. W. and Arnosti, D. N. (2010). Paradoxical instability-activity relationship defines a novel regulatory pathway for retinoblastoma proteins. *Mol. Biol. Cell* **21**, 3890–3901.

Ambrus, A. M., Islam, A. B. M. M. K., Holmes, K. B., Moon, N. S., Lopez-Bigas, N., Benevolenskaya, E. V and Frolov, M. V (2013). Loss of dE2F compromises mitochondrial function. *Dev. Cell* **27**, 438–451.

Bailey, T. L. and Gribskov, M. (1998). Methods and statistics for combining motif match scores. *J. Comput. Biol.* **5**, 211–221.

Bohla, D., Herold, M., Panzer, I., Buxa, M. K., Ali, T., Demmers, J., Krüger, M., Scharfe, M., Jarek, M., Bartkuhn, M., et al. (2014). A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. *PLoS One* **9**, e107765.

Cao, L., Peng, B., Yao, L., Zhang, X., Sun, K., Yang, X. and Yu, L. (2010). The ancient function of RB-E2F pathway: insights from its evolutionary history. *Biol. Direct* **5**, 55.

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S. and Young, R. A. (2001). Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337.

Cavanaugh, a H., Hempel, W. M., Taylor, L. J., Rogalsky, V., Todorov, G. and Rothblum, L. I. (1995). Activity of RNA polymerase I transcription factor UBF blocked by Rb gene product. *Nature* **374**, 177–180.

Cecchini, M. J. and Dick, F. a (2011). The biochemical basis of CDK phosphorylation-independent regulation of E2F1 by the retinoblastoma protein. *Biochem. J.* **434**, 297–308.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R. a, Narita, M., Zhang, M. and Lowe, S. W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–87.

Chinnam, M. and Goodrich, D. W. (2011). RB1, development, and cancer. *Curr. Top. Dev. Biol.* **94**, 129–169.

Classon, M. and Dyson, N. (2001). P107 and P130: Versatile Proteins With Interesting Pockets. *Exp. Cell Res.* **264**, 135–47.

Classon, M. and Harlow, E. (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* **2**, 910–917.

Dick, F. A. and Rubin, S. M. (2013). Molecular mechanisms underlying RB protein function. *Nat. Rev. Mol. Cell Biol.* **14**, 297–306.

Dimova, D. K., Stevaux, O., Frolov, M. V and Dyson, N. J. (2003). Cell cycle-dependent and cell cycle-independent control of transcription by the Drosophila E2F/RB pathway. *Genes Dev.* **17**, 2308–2320.

Du, W. and Pogoriler, J. (2006). Retinoblastoma family genes. *Oncogene* **25**, 5190–200.

Du, W. (2000). Suppression of the rbf null mutants by a de2f1 allele that lacks transactivation domain. **379**, 367–379.

Dyson, N. (1998). The regulation of E2F by pRB-family?proteins. *Genes Dev.* **12**, 2245–2262.

Elenbaas, J. S., Mouawad, R., Henry, W. R., Arnosti, D. N. and Payankaulam, S. (2014). Role of Drosophila Retinoblastoma Protein Instability Element in Cell Growth and Proliferation. *Cell Cycle* 37–41.

Fan, G. and Steer, C. J. (1999). The role of retinoblastoma protein in apoptosis. *Apoptosis* **4**, 21–29.

Felton-Edkins, Z. A., Kenneth, N. S., Brown, T. R. P., Daly, N. L., Gomez-Roman, N., Grandori, C., Eisenman, R. N. and White, R. J. (2003). Direct regulation of RNA polymerase III transcription by RB, p53 and c-Myc. *Cell Cycle* **2**, 181–184.

Frolov, M. V, Huen, D. S., Stevaux, O., Dimova, D., Balczarek-Strang, K., Elsdon, M. and Dyson, N. J. (2001). Functional antagonism between E2F family members. *Genes Dev.* **15**, 2146–2160.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. and Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257.

Georlette, D., Ahn, S., MacAlpine, D. M., Cheung, E., Lewis, P. W., Beall, E. L., Bell, S. P., Speed, T., Manak, J. R. and Botchan, M. R. (2007). Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. *Genes Dev.* **21**, 2880–96.

Gershman, B., Puig, O., Hang, L., Peitzsch, R. M., Tatar, M. and Garofalo, R. S. (2007). High-resolution dynamics of the transcriptional response to nutrition in Drosophila: a key role for dFOXO. *Physiol. Genomics* **29**, 24–34.

Giacinti, C. and Giordano, a (2006). RB and cell cycle progression. *Oncogene* **25**, 5220–5227.

Gjidoda, A. and Henry, R. W. (2013). RNA polymerase III repression by the retinoblastoma tumor suppressor protein. *Biochim. Biophys. Acta* **1829**, 385–392.

Gutzat, R., Borghi, L. and Gruissem, W. (2012). Emerging roles of RETINOBLASTOMA-RELATED proteins in evolution and plant development. *Trends Plant Sci.* **17**, 139–148.

Hasygar, K. and Hietakangas, V. (2014). p53- and ERK7-Dependent Ribosome Surveillance Response Regulates Drosophila Insulin-Like Peptide Secretion. *PLoS Genet.* **10**, e1004764.

He, X., Chen, C.-C., Hong, F., Fang, F., Sinha, S., Ng, H.-H. and Zhong, S. (2009). A biophysical model for analysis of transcription factor interaction and binding site arrangement from genome-wide binding data. *PLoS One* **4**, e8155.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–89.

Hernando, E., Nahle, Z., Juan, G., Diaz-Rodriguez, E., Alaminos, M., Hemann, M., Michel, L., Mittal, V., Gerald, W., Benezra, R., et al. (2004). Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature* **430**, 797–802.

Hirsch, H. a, Gu, L. and Henry, R. W. (2000). The retinoblastoma tumor suppressor protein targets distinct general transcription factors to regulate RNA polymerase III gene expression. *Mol. Cell. Biol.* **20**, 9182–9191.

Hirsch, H. A., Jawdekar, G. W., Lee, K.-A., Gu, L. and Henry, R. W. (2004). Distinct mechanisms for repression of RNA polymerase III transcription by the retinoblastoma tumor suppressor protein. *Mol. Cell. Biol.* **24**, 5989–5999.

Huang, D. W., Sherman, B. T. and Lempicki, R. a (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.

Julian, L. M., Palander, O., Seifried, L. A., Foster, J. E. G. and Dick, F. A. (2008). Characterization of an E2F1-specific binding domain in pRB and its implications for apoptotic regulation. *Oncogene* **27**, 1572–1579.

Keller, S. A., Ullah, Z., Buckley, M. S., Henry, R. W. and Arnosti, D. N. (2005). Distinct developmental expression of Drosophila retinoblastoma factors. *Gene Expr. Patterns* **5**, 411–421.

Korenjak, M., Anderssen, E., Ramaswamy, S., Whetstine, J. R. and Dyson, N. J. (2012). RBF binding to both canonical E2F targets and noncanonical targets depends on functional dE2F/dDP complexes. *Mol. Cell. Biol.* **32**, 4375–87.

Korenjak, M., Kwon, E., Morris, R. T., Anderssen, E., Amzallag, A., Ramaswamy, S. and Dyson, N. J. (2014). dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. *Nucleic Acids Res.* **42**, 8939–8953.

Korenjak, M., Taylor-Harding, B. and Binné, U. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell* **119**, 181–193.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

Larminie, C. G., Cairns, C. A., Mital, R., Martin, K., Kouzarides, T., Jackson, S. P. and White, R. J. (1997). Mechanistic analysis of RNA polymerase III regulation by the retinoblastoma protein. *EMBO J.* **16**, 2061–2071.

Latorre, I., Chesney, M. A., Garrigues, J. M., Stempor, P., Appert, A., Francesconi, M., Strome, S. and Ahringer, J. (2015). The DREAM complex promotes gene body H2A.Z for target repression. *Genes Dev.* **29**, 495–500.

Latorre, I., Chesney, M. a, Garrigues, J. M., Stempor, P., Appert, A., Francesconi, M., Strome, S. and Ahringer, J. (2013). The DREAM complex promotes gene body H2A . Z for target repression. 495–500.

Lewis, P. W., Beall, E. L., Fleischer, T. C., Georlette, D., Link, A. J. and Botchan, M. R. (2004). Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev.* **18**, 2929–2940.

Machanick, P. and Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697.

Marygold, S. J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G. H., Harrison, P. M., Yu, Z., Kenmochi, N., Kaufman, T. C., et al. (2007). The ribosomal protein genes and Minute loci of Drosophila melanogaster. *Genome Biol.* **8**, R216.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **7**, 458.

Morris, E. J. and Dyson, N. J. (2001). Retinoblastoma protein partners. *Adv. Cancer Res.* **82**, 1–54.

Nicolay, B. N., Gameiro, P. a, Tschöp, K., Korenjak, M., Heilmann, A. M., Asara, J. M., Stephanopoulos, G., Iliopoulos, O. and Dyson, N. J. (2013). Loss of RBF1 changes glutamine catabolism. *Genes Dev.* **27**, 182–96.

Norton, J. D., Deed, R. W., Craggs, G. and Sablitzky, F. (1998). Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol.* **8**, 58–65.

Raj, N., Zhang, L., Wei, Y., Arnosti, D. N. and Henry, R. W. (2012). Ubiquitination of retinoblastoma family protein 1 potentiates gene-specific repression function. *J. Biol. Chem.* **287**, 41835–41843.

Reynolds, M. R., Lane, a N., Robertson, B., Kemp, S., Liu, Y., Hill, B. G., Dean, D. C. and Clem, B. F. (2014). Control of glutamine metabolism by the tumor suppressor Rb. *Oncogene* **33**, 556–66.

Rhee, H. S. and Pugh, B. F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419.

Rubin, S. M., Gall, A. L., Zheng, N. and Pavletich, N. P. (2005). Structure of the Rb C-terminal domain bound to E2F1-DP1: A mechanism for phosphorylation-induced E2F release. *Cell* **123**, 1093–1106.

Schwartz, Y. B., Linder-Basso, D., Kharchenko, P. V, Tolstorukov, M. Y., Kim, M., Li, H.-B., Gorchakov, A. A., Minoda, A., Shanower, G., Alekseyenko, A. A., et al. (2012). Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res.* **22**, 2188–2198.

Steffen, K. K., McCormick, M. A., Pham, K. M., MacKay, V. L., Delaney, J. R., Murakami, C. J., Kaeberlein, M. and Kennedy, B. K. (2012). Ribosome deficiency protects against ER stress in Saccharomyces cerevisiae. *Genetics* **191**, 107–118.

Stevaux, O., Dimova, D. K. and Dyson, N. J. (2005). Retinoblastoma Family 2 is Required In Vivo for the Tissue-Specific Repression of dE2F2 Target Genes ND ES. **4**, 1272–1280.

Stevaux, O., Dimova, D., Frolov, M. V, Taylor-harding, B., Morris, E. and Dyson, N. (2002). Distinct mechanisms of E2F regulation by Drosophila RBF1 and RBF2. *EMBO J.* **21**, 4927–4937.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578.

Ullah, Z. and Buckley, M. (2007). Retinoblastoma protein regulation by the COP9 signalosome. *Mol. Biol. ...* **18**, 1179–1186.

van den Heuvel, S. and Dyson, N. J. (2008). Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell Biol.* **9**, 713–24.

Wei, Y., Mondal, S. S., Mouawad, R., Wilczynski, B., Henry, R. W. and Arnosti, D. N. (2015). Genome-Wide Analysis of Drosophila Rbf2 Protein Highlights Diversity of RB Family Targets and Possible Role in Regulation of Ribosome Biosynthesis. *G3 (Bethesda).*

Weinberg, R. a (1995). The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330.

White, R. J. (1997). Regulation of RNA polymerases I and III by the retinoblastoma protein: a mechanism for growth control? *Trends Biochem. Sci.* **22**, 77–80.

White, R. J. (2004). RNA polymerase III transcription and cancer. *Oncogene* **23**, 3208–3216.

White, R. J. (2005). RNA polymerases I and III, growth control and cancer. *Nat. Rev. Mol. Cell Biol.* **6**, 69–78.

Woolford, J. L. and Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics* **195**, 643–81.

Xue, S. and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369.

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

Yang, J., Ramos, E. and Corces, V. G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. *Genome Res.* **22**, 2199–2207.

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W. and Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J. Biol. Chem.* **289**, 24863–24873.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

Zheng, L. and Lee, W.-H. (2002). Retinoblastoma tumor suppressor and genome stability. *Adv. Cancer Res.* **85**, 13–50.

CHAPTER 3

Complex cis-regulatory landscape of the insulin receptor gene reflects dynamic regulation of a "housekeeping" gene

Abstract

Insulin signaling plays key roles in development, growth and metabolism, through dynamic control of glucose uptake, global protein translation, and transcriptional regulation. Altered levels of insulin signaling are known to play critical roles in development and disease, yet the molecular basis of such differential signaling is obscure. Expression of the insulin receptor (*InR*) gene itself appears to play an important role, but the nature of the molecular wiring controlling *InR* transcription remains to be elucidated. We characterized the complex regulatory elements contributing to *InR* expression in Drosophila, and found that dynamic regulation of this gene reflects direct and indirect contributions of the dFOXO, EcR, Rbf, and additional transcription factors through redundant elements dispersed throughout ~40 kbp of noncoding regions. The dynamic regulation of this broadly expressed gene's transcription in response to nutritional and tissue-specific inputs represents an integration of multiple *cis* regulatory elements, whose structure and function appear to be sculpted by evolutionary selection to provide a highly tailored set of signaling responses on developmental and tissue-specific levels.

This work is currently in review as the following manuscript:

Wei, Y., Gokhale, R., Sonnenschein, A., Montgomery, K., Ingersoll, A. and Arnosti, D. N. (2016). Complex cis-regulatory landscape of the insulin receptor gene underlies the broad expression of a central signaling regulator. *Development.*

My contribution to this study was the generation of the reporter library, execution of all reporter assays in different cell types and of dFOXO ChIP, and analysis of reporter data.

Introduction

The insulin signaling pathway plays a major role in growth and metabolism of metazoans. In mammals, signaling at the cellular level involves the insulin receptor (INSR), a member of the ligand-activated receptor kinase superfamily, which also includes the homologous insulin-like growth factor 1 receptor (IGF-1R) (Ebina et al., 1985; Ullrich et al., 1985, 1986; Ullrich and Schlessinger, 1990). These receptors bind to insulin and IGF-1/2 peptides (Rechler and Nissley, 1985; Yamaguchi et al., 1993; Frasca et al., 1999). Upon ligand activation, the receptor undergoes a conformational change, leading to autophosphorylation, followed by phosphorylation of adaptor proteins, including insulin receptor substrate (IRS) and Shc. Subsequent activation of the PI3K-Akt and Ras-Raf-MAPK pathways then ensues, influencing activity of the TOR pathway to regulate protein synthesis (Oldham and Hafen, 2003). Akt propagates the metabolic effects of the signaling cascade by targeting downstream substrates, including the glucose transporter GLUT4 (Bertrand et al., 2008; Gonzalez et al., 2011). A critical transcription factor that is regulated by Akt is FOXO, which is phosphorylated and excluded from the nucleus as a result of insulin signaling (Puig and Tjian, 2005). This aspect of insulin signaling appears to be widely conserved in metazoans, with FOXO homologs serving to mediate the effect of insulin signaling on growth, aging and metabolism in *C. elegans* and *Drosophila* (Taguchi and White, 2008). Interestingly, the FOXO transcription factor has been shown to directly activate *INSR* gene expression, representing a transcriptional feedback loop of this pathway (Puig and Tjian, 2005).

Although circulating insulin levels dictate overall pathway activity, the different levels of the receptor protein itself may also influence signaling in a tissue-specific manner.

Such tissue- and stage-specific differences in *INSR* expression may be developmentally "hardwired", and subject to evolutionary modification, changing the impact of insulin signaling in control of body size and morphology (Goldstein et al., 1987; Belfiore et al., 2009). In addition to developmentally programmed expression of *INSR*, a variety of physiological stimuli influence expression of this gene, although the consequence of this regulation has been little explored. In the context of normal homeostasis, diet, hormone levels and other signals impact transcription of the mammalian *INSR*. In disease states, expression of the receptor is influenced by viral infection and diabetes, and elevated levels of *INSR* are observed in numerous cancers, leading to an insulin-dependent growth phenotype (Kriauciunas et al., 1993; Gunton et al., 2005; Belfiore and Malaguarnera, 2011). Deregulation of *INSR* and *IGF-1R* plays an important role in cancer progression, as cancer cells often overexpress these receptors (Belfiore and Malaguarnera, 2011). Patients with type II diabetes or obesity, conditions associated with hyperinsulinemia, have elevated risk of cancer (Taubes, 2012). On the other hand, low levels of *INSR*/*IGF-1R* expression as well as INSR/IGF-1R signaling in the brain are found to be associated with Alzheimer's disease (Frölich et al., 1998; Moloney et al., 2010). However, despite the epidemiological and experimental evidence for misregulation of these receptors in cancer and other diseases, we have limited knowledge about their transcriptional regulation.

Although the *INSR* gene is very broadly expressed - unlike highly tissue-specific developmental genes - its regulation is complex and may involve multiple *cis* and *trans* components (Lee et al., 1992). In humans and mice, *INSR* is expressed at different levels in a tissue- and temporal-specific manner (Goldstein et al., 1987; Belfiore et al., 2009).

A limited two kbp promoter-proximal region of the human *INSR* gene was shown to mediate hormonal response, with induction of reporter genes noted after treatment with dexamethasone, glucocorticoids, vitamin D, and estrogen (Leal et al., 1992; Lee and Tsai, 1994; García-Arencibia et al., 2005; Calle et al., 2008). Sp1, HMGI, p53, and Rb were found to associate with and regulate this region of the *INSR* gene (Cameron et al., 1992; Shen et al., 1995; Webster et al., 1996; Brunetti et al., 2001). Although the promoter proximal region is regulated, the transcriptional significance of other regions of the gene has been largely unexplored. A conserved feature of this gene is the presence of large introns that may be of regulatory importance; while the coding sequence encompasses less then 10 kbp, the human transcription unit spans nearly 200 kbp. Invertebrate *InR* genes are smaller, but still sizable, featuring large intronic regions (Casas-Tinto et al., 2007). Genomic surveys of the mammalian gene reveal a plethora of functionally uncharacterized chromatin marks and structures consistent with putative enhancers (Pasquali et al., 2014; Roadmap Epigenomics Consortium et al., 2015). A recent study of mouse T-cells identified multiple enhancers in the large introns of the insulin receptor gene, although the properties of these elements was unexplored (Vanhille et al., 2015). Similar to *INSR*, the *IGF-1R* gene also possesses large introns, yet only the promoter region has been studied for transcriptional regulation (Sarfstein et al., 2006; Schayek et al., 2010; Oberbauer, 2013).

Drosophila has a single insulin receptor gene, *InR*, and this receptor is activated by a family of insulin-like peptides (dILPs) to control growth and homeostasis (Oldham and Hafen, 2003). The *InR* gene is critical to embryonic development, function of the nervous system, and regulation of growth by controlling final body/organ size

(Petruzzelli et al., 1986; Garofalo and Rosen, 1988; Fernandez et al., 1995; Brogiolo et al., 2001; Song et al., 2003; Wong et al., 2014). Loss of function of *InR* mutations result in pleiotropic recessive phenotypes, leading to embryonic lethality (Fernandez et al., 1995). An intriguing aspect of *InR* gene expression is the role played by insulin-signaling itself (Jünger et al., 2003; Kramer et al., 2003; Puig et al., 2003; Casas-Tinto et al., 2007). The *Drosophila* forkhead protein FOXO (dFOXO) is a key component of the insulin-signaling pathway; this transcription factor not only regulates growth-control genes such as *Thor/4EBP* but also feedback regulates *InR* (Puig et al., 2003). An internal *InR* promoter was found to contain dFOXO response elements that were activated by dFOXO (Casas-Tinto et al., 2007). By controlling levels of *InR* expression, dFOXO is thus thought to play a role in setting the sensitivity of this signaling pathway. The feedback mechanism appears to be conserved in mammals (Puig and Tjian, 2005). The functional relevance of this feedback mechanism in the context of the whole organism is not known, however. Studies in *Drosophila* have also provided mechanistic information about control of the insulin receptor gene transcription by nutrition and the steroid hormone ecdysone. In *Drosophila*, the ecdysone receptor is a transcription factor that controls molting and development, and with its co-factor ultraspiracle (USP) binds to the steroid hormone 20-hydroxyecdysone (20E) to mediate its transcriptional effects (Koelle et al., 1991; Riddiford et al., 2000; Hu et al., 2003). This hormone stimulates expression of *InR* in *Drosophila* Kc cells as well as in the fat body of the larval silk moth *Bombyx* (Gauhar et al., 2009, Liu et al., 2010). ChIP-Seq studies have provided evidence that this regulation is direct; the ecdysone receptor, EcR and USP binds to the *InR* locus in *Drosophila* (Gauhar et al., 2009). More recently, enhancer studies using STARR-seq

technology in *Drosophila* cell lines identified a number of 20E-responsive elements in *InR* gene locus (Shlyueva et al., 2014). Interestingly, ecdysone signaling and insulin signaling also regulate each other. Insulin signaling affects ecdysone synthesis, while ecdysone impedes PI3K activity to stimulate dFOXO nuclear localization (Tu et al., 2002; Colombani et al., 2005; Koyama et al., 2014; Herboso et al., 2015).

We have recently reported that the retinoblastoma transcriptional co-repressor proteins also associate with the *Drosophila InR* promoter, although its significance for regulation is not well understood (Acharya et al., 2012; Korenjak et al., 2012; Wei et al., 2015). Retinoblastoma (RB) tumor suppressor proteins are key regulators of the cell cycle, as well as serving roles in control of cellular signaling and development, and loss of RB function is implicated in a wide variety of human tumors (Giacinti and Giordano, 2006). Similar to their human counterparts, the *Drosophila* retinoblastoma family members (Rbfs) control cell cycle and developmentally regulated gene expression (Du and Pogoriler, 2006). Although it was not commented on when the human ChIP-seq studies were published, we note that RB family proteins are similarly found to be associated with the human *INSR* gene (Chicas et al., 2010). These data indicate that the RB protein may directly target insulin pathway genes to transcriptionally regulate their gene expression, changing the inherent "set point" of insulin responsiveness in different tissues. The RB regulation of insulin signaling pathway would potentially link growth and cell cycle control, of potential importance in explaining links between type II diabetes and cancer.

Similar to its mammalian counterpart, the *Drosophila InR* gene is a large locus, with nearly 40 kbp of introns (Casas-Tinto et al., 2007). Characteristic histone modifications,

DNase I hypersensitivity, FAIRE-seq, and STARR-seq suggest that multiple putative enhancers are located in these large introns, however we lack an integrated understanding of functional regulation of this central player in cell metabolism and development (Kaplan et al., 2011; Li et al., 2011; Nègre et al., 2011; Thomas et al., 2011; Arnold et al., 2013; McKay and Lieb, 2013; Shlyueva et al., 2014). To elucidate the molecular underpinnings of transcriptional control of this key regulatory gene, here we describe a comprehensive identification and characterization of *cis*-regulatory elements associated with the *InR* gene, mapping their responses to dFOXO, ecdysone and RB. Our detailed mutagenic studies of the active enhancers identify specific elements and motifs required for enhancer activity, and demonstrate that these elements work in a non-additive fashion. The dynamic regulation of these enhancers by transcriptional inputs indicates these enhancers play a role in temporal, spatial, and critical fine-tuning control of the *InR* gene expression. Our study indicates that this gene is subject to a complex transcriptional circuit extending far beyond the previously described simple model of the FOXO-feedback loop mechanism. This gene circuit analysis transforms our understanding of the insulin receptor gene, in that even such a broadly expressed gene requires exquisite controls, whose functions have critical relevance to the roles of this signaling pathway in metabolism, growth control and cancer.

Results

Genomic rescue construct identifies regulatory regions of InR

The *Drosophila InR* gene occupies a 50 kbp region on the 3R chromosome including ~ 40 kbp of introns (Figure 3-1A). To identify the genomic region that is responsible for *InR* expression, we used an 80 kbp BAC (InR-BAC) that spans the *InR* locus, generated a rescue transgene, and tested its ability to rescue the lethality of an *InR* mutant. The BAC construct includes the entire *InR* transcription unit as well as the 3' *CG15498* gene, and portions of the 5'*E2F1* and 3' *slou* genes (Figure 3-1A). We crossed this construct into a background containing the temperature-dependent conditional lethal trans-heterozygous alleles (InR^{GC25}/InR^{E19}) (Shingleton et al., 2005). InR^{GC25}/InR^{E19} flies are not viable when raised at 27 °C. The BAC was able to rescue this lethality and *InR-BAC/+ ;* InR^{GC25}/InR^{E19} survivors were obtained (Figure 3-1B, Figure 3-2). The presence of two copies of the BAC increases *InR* gene expression in these flies about two to three fold (Figure 3-1C). The relevant *cis* regulatory sequences for *InR* expression are thus located within this region; we therefore investigated the short 5' intergenic sequence and the sizeable introns of *InR* to uncover relevant *cis* regulatory elements.

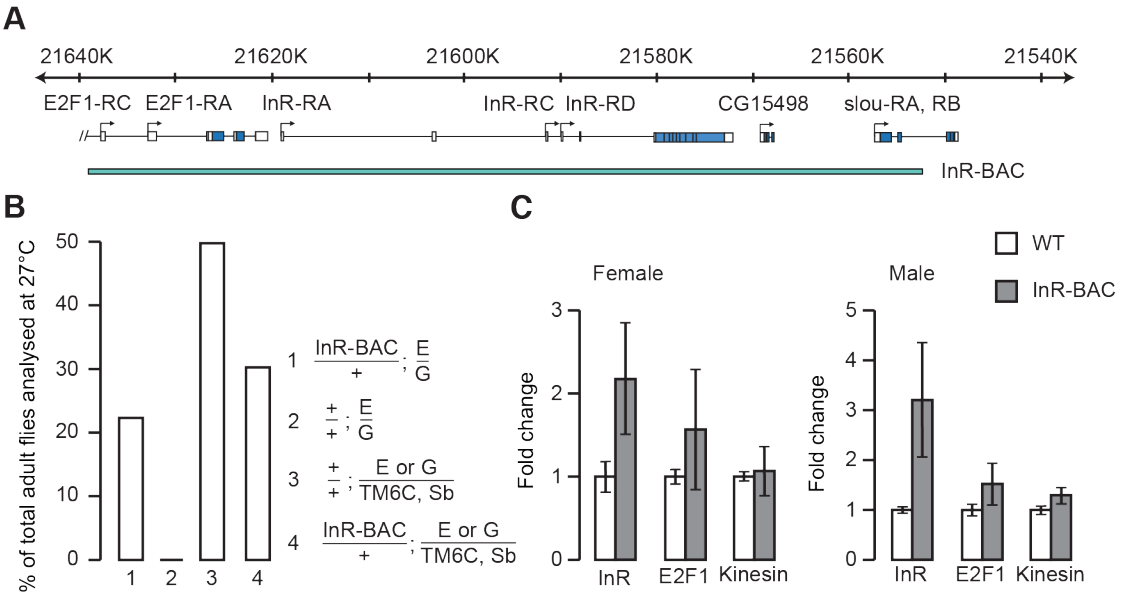


Figure 3-1: An 80 kbp region contains *cis* regulatory information necessary for genomic rescue of *InR* mutants. (A) The *Drosophila InR* gene spans ~50 kbp and contains multiple large introns. The entire gene along with its 5' and 3' regions are contained in an *InR*-BAC transgene (indicated by green line) inserted into chromosome 2. (B) The *InR*-BAC transgene rescues the lethality of the transheterozygous InR^{E19}/InR^{GC25} allele combination (represented here showing second and third chromosome genotypes as *InR*-BAC/+; *E/G*, where E represents InR^{E19}, and G represents InR^{GC25}) when raised at the nonpermissive temperature of 27 °C. No *E/G* survivors lacking the rescue BAC transgene were found; a total of 89 flies were analyzed. The rescue was also performed at 18 °C as shown in Figure 3-2. (C) Transcript levels of *InR* and *E2F1* measured from 3-day adult homozygous *InR*-BAC females and males. Both females and males showed 2~3 fold increase in transcripts of *InR* and ~1.5 fold for *E2F1*. No significant change was observed for *kinesin* (Kinesin heavy chain, also *khc*) as a negative control. All transcripts were normalized to *28S* transcript levels.

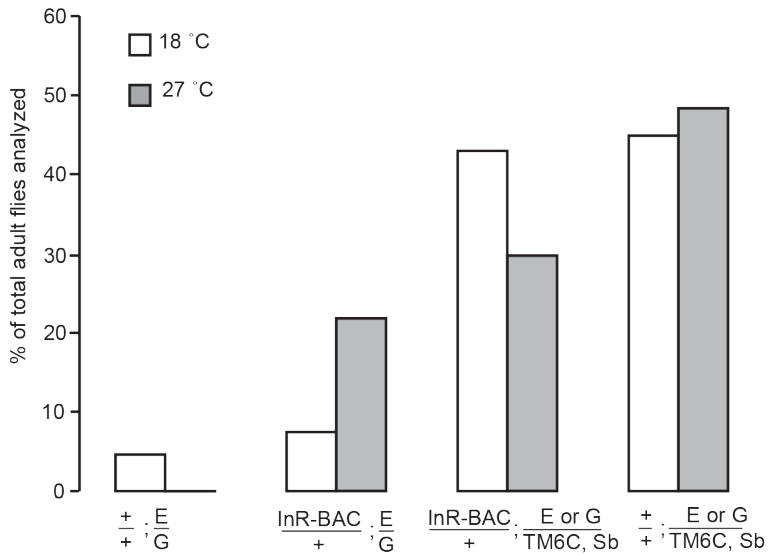


Figure 3-2: Number and genotype of survivors from *InR* BAC rescue experiment. *InR* BAC rescue experiment showing fractions of progeny with and without *InR* transgene at permissive (18°C) and nonpermissive (27°C) temperatures. No survivors lacking the rescue transgene were found at the higher temperature. The transheterozygous InR^{E19}/InR^{GC25} allele combination is lethal (represented here showing second and third chromosome genotypes as *InR*-BAC/+; *E/G*, where E represents InR^{E19}, and G represents InR^{GC25}) when raised at the nonpermissive temperature of 27 °C. 97 individuals were analyzed for crosses at 27°C, and 107 individuals for crosses at 18°C.

Evidence of tissue-specific enhancers in InR introns

The *InR* gene is flanked by only short intergenic regions, thus important regulatory sequences may be located within its sizable introns, although *cis* regulatory elements may reside anywhere within the 80 kbp region defined by the BAC. Data from cell-type specific enhancer analysis using STARR-seq technology, as well as DNase hypersensitivity data and measurement of open chromatin using FAIRE-seq support the notion that *InR* introns are likely to harbor relevant *cis*-regulatory elements (Figure 3-3) (Kaplan et al., 2011; Li et al., 2011; Thomas et al., 2011; Arnold et al., 2013; McKay and Lieb, 2013). To evaluate the regulatory potential of intronic regions in the whole fly, we tested ten GAL4 lines bearing genomic fragments derived from the *InR* gene (Figure 3-4A, Figure 3-5) (Pfeiffer et al., 2008). Previous measurements in the embryo indicated that some of these elements drive GFP expression in dynamic and cell-type specific patterns (Jenett et al., 2012; Jory et al., 2012; Manning et al., 2012; Li et al., 2014). We found that three of the fragments also express GFP in larvae and adults, in either ubiquitous or tissue-specific patterns (Figure 3-4A, Figure 3-5). To obtain more insight on possible *cis* regulatory elements, we surveyed extant datasets for information about chromatin accessibility and ChIP-seq information that may reveal active regulatory regions in this locus. To identify possible correlated features, we plotted the results of genome-wide enhancer surveys (from S2 and ovarian stem cells (OSC)), chromatin accessibility in different developmental stages and tissues as measured by FAIRE-seq, and enhancer-associated histone modifications H3K27Ac, H3K4Me1, and the p300 coactivator (Kaplan et al., 2011; Li et al., 2011; Nègre et al., 2011; Thomas et al., 2011; Arnold et al., 2013; McKay and Lieb, 2013; Shlyueva et al., 2014), The resultant

patterns do not provide a consistent, easily interpretable set of correlations across different developmental times. Enhancers found using STARR-seq do point to apparently redundantly-acting enhancers in *InR* introns with either shared or cell-type specific patterns (Figure 3-4B) (Arnold et al., 2013). These enhancers overlap some of the fragments tested as GAL4 drivers, but there was not a complete agreement between these different methods. The two types of assays relied on distinct basal promoters, which may have biased detection because of enhancer-promoter specificity (Marinić et al., 2013; Zabidi et al., 2015).

STARR-seq analysis is limited to activity measurements in two cell types, but more general regulation of the locus may be revealed by chromatin marks tested in the whole animal in different developmental stages. Chromatin accessibility measured by DNaseI hypersensitivity in S2 and OSC cells correlates partially with STARR-seq enhancers measured in these two cell types, and some of these regions overlap with hypersensitive regions identified by DNaseI hypersensitivity in the developing embryo (Figure 3-4B) (Kaplan et al., 2011; Li et al., 2011; Thomas et al., 2011; Arnold et al., 2013). Similar patterns are observed between DNaseI and FAIRE, a complementary method for measuring open chromatin. FAIRE data indicate that certain features are present from embryonic through pupal development, but these regions do not align with enhancers found to drive ubiquitous expression at many time points. Many intronic regions also exhibit transient FAIRE signals, consistent with findings that some tissue-specific enhancers are detected by changes in this signal (Figure 3-4B). In sum, although some regulatory regions overlap with DNaseI and FAIRE signals, these features alone do not appear to provide a simple or comprehensive indication of the *cis* regulatory elements.

Important independent marks of eukaryotic enhancers are H3K27Ac and H3K4Me1 chromatin modifications (Rajagopal et al., 2014). The developmental patterns of these marks, which are present at varying levels over most of the intronic regions of *InR*, were much more dynamic than those for chromatin accessibility, possibly reflecting the dynamics of *InR* gene expression in development (modENCODE). Parts of the chromatin marks overlapped with STARR-seq enhancers and fragments tested in the whole fly as GAL4 drivers, however, these marks are also evident on regions that were apparently devoid of activity (Figure 3-4B). Consistent with the dynamic pattern of histone modifications, the histone acetyltransferase coactivator p300 also showed dynamic pattern, with some overlap with H3K27Ac patterns (Figure 3-4B) (modENCODE). Together, these data provide evidence that there are multiple, dynamically active enhancers embedded in *InR* introns, however, their regulation or structure is not revealed by simple inspection of chromatin features of the locus.

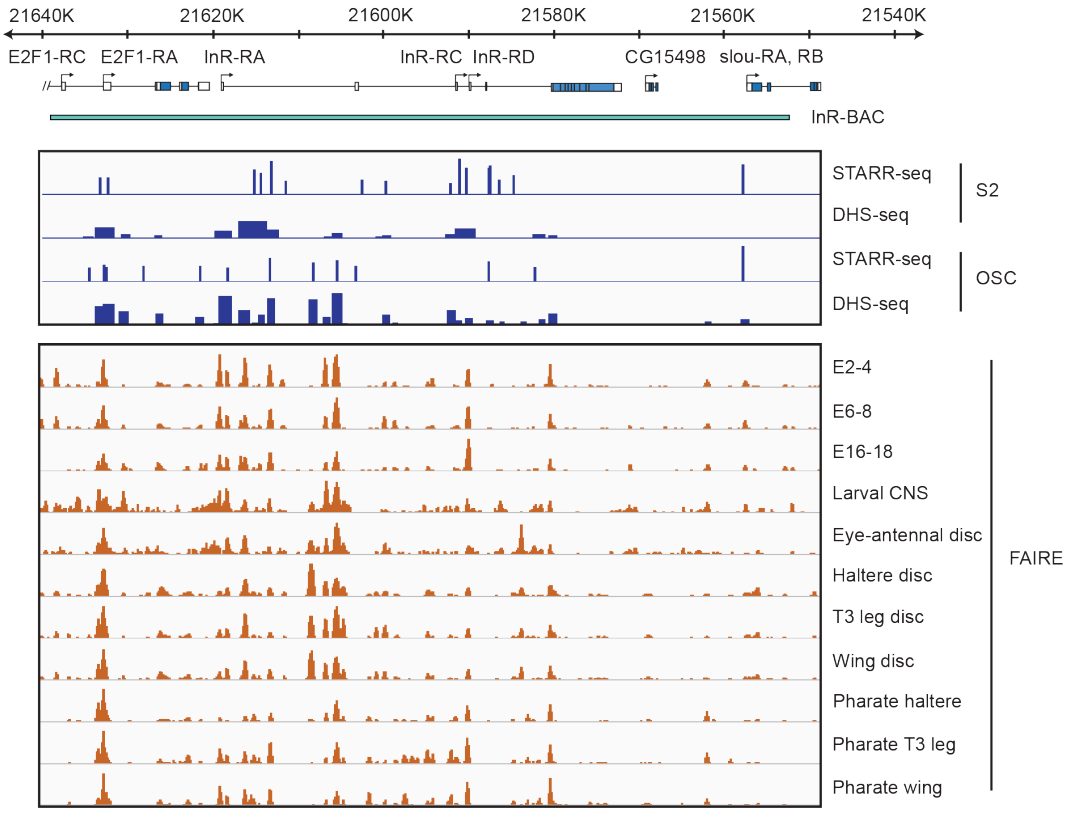


Figure 3-3: Enhancers identified by STARR-seq, DHS-seq and FAIRE at *InR* gene locus. STARR-seq, DHS-seq and FAIRE signals were aligned with *InR* region spanned by the *InR*-BAC construct. Signals were enriched in the *InR* intron regions, suggesting that *cis*-regulatory elements may be present.

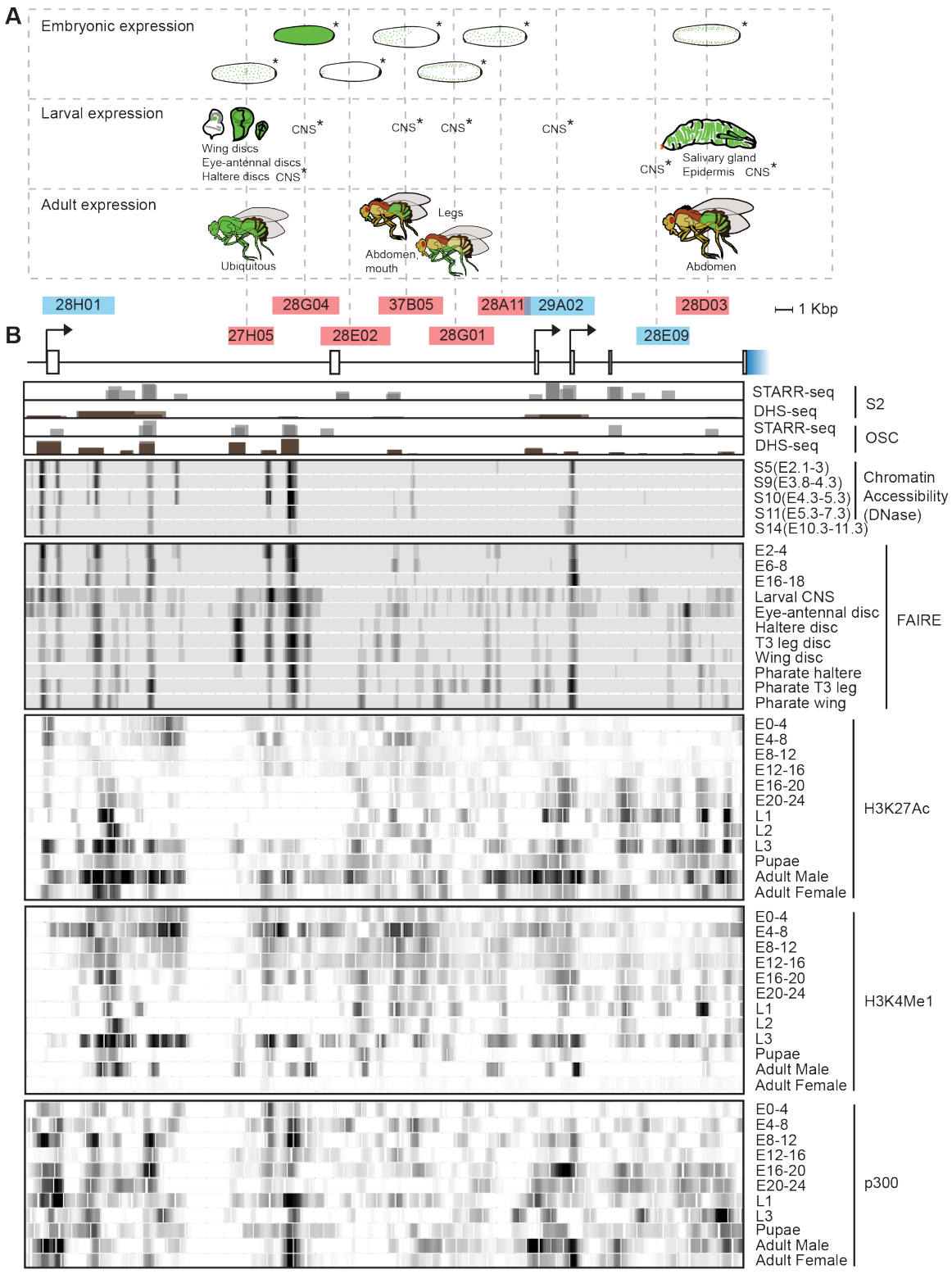


Figure 3-4: Regulatory landscape of the *InR* locus. (A) Transcriptional output of genomic fragments associated with the *InR* locus. To assess larval and adult activity of

Figure 3-4 (con't)

Janelia GAL4 lines that contain genomic fragments in the *InR* locus, we crossed these transcriptional drivers to a UAS-GFP line. Extant information collected from FlyLight database for embryonic and larval activity is also shown (*Data from FlyLight, www.janelia.org/project-team/flylight). Fragments labeled in red showed GFP signal in embryos, larval, or adult flies. Fragments labeled in blue showed limited or no expression. Representative images in larvae and adult flies from this study are shown in Figure 3-5. (B) Alignment of the *InR* gene locus with previously identified enhancers in S2 and OSC cells (Arnold et al., 2013), and with various chromatin features, including chromatin accessibility identified by DNase-seq in cell lines (Arnold et al., 2013), and in developing embryos (Berkeley Drosophila Transcription Network Project Chromatin Accessibility, Kaplan et al., 2011; Li et al., 2011; Thomas et al., 2011), FAIRE in different developmental stages and tissues (McKay and Lieb, 2013), enhancer-associated modifications H3K27Ac, H3K4Me1 and p300 binding in different developmental stages (modENCODE). For STARR-seq and DHS-seq, data from two replicates are shown, and the darker areas indicate overlapping results from the replicate experiments. The heights of the bars indicate enhancer activities or reads for the peaks. For STARR-seq data, the width of the signal is set as 600 bp, the average length of STARR-seq fragments (Arnold et al., 2013). Chromatin accessibility data from Berkeley Drosophila Transcription Network Project are presented according to different developmental stages, the corresponding developing times are indicated on the side (S is abbreviation for stage, and E is abbreviation for embryo). The data for FAIRE, H3K27Ac, H3K4Me1, and p300 are

Figure 3-4 (con't)

presented as peak intensities, with darker shades indicating higher peaks. All data scales are normalized to local maximum. Genome version is *Drosophila* genome dm3/R5.

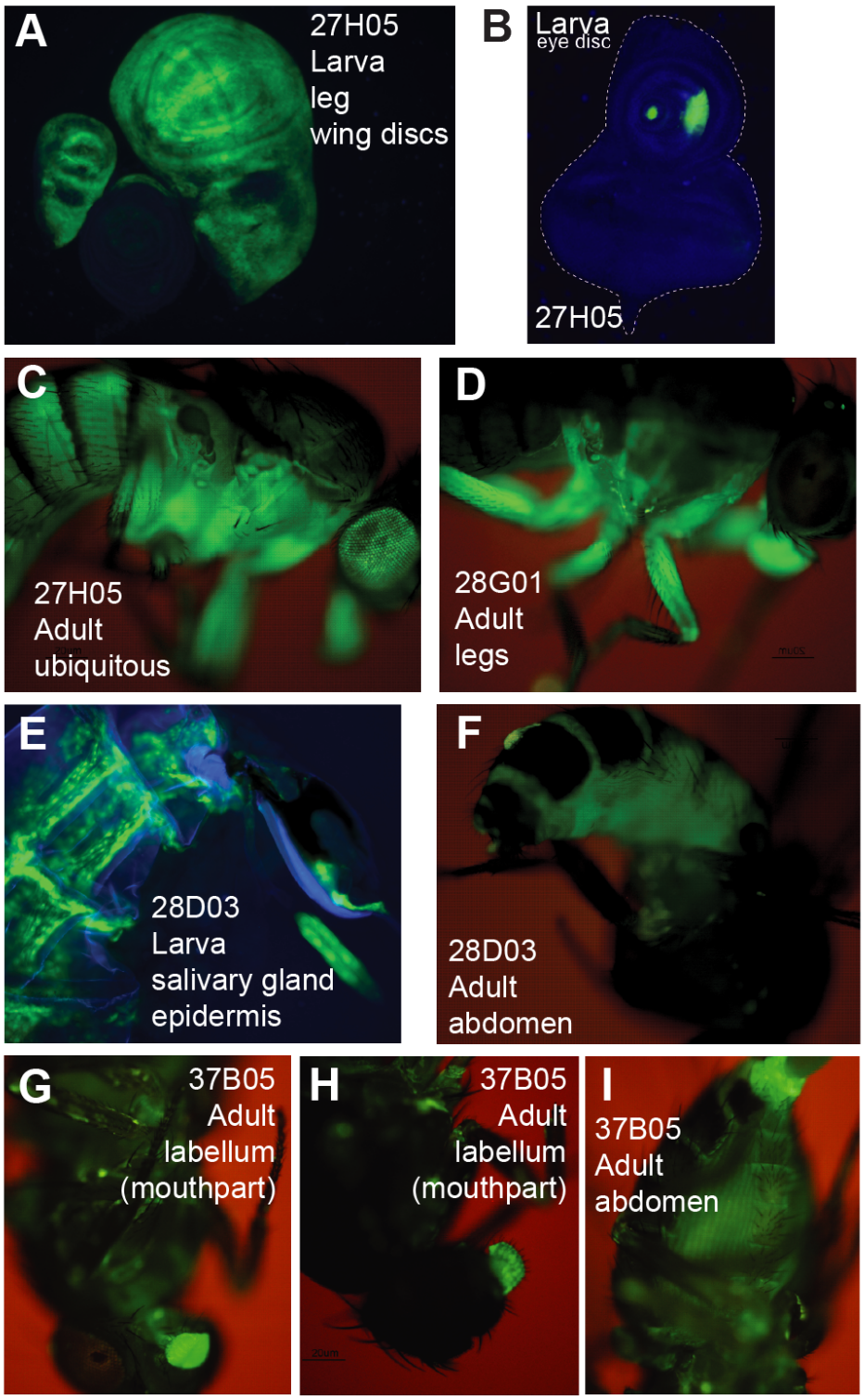


Figure 3-5: Images of larvae and adult flies expressing UAS-GFP under the control of indicated Janelia Gal4 driver lines. (A) GFP expression in larval leg and wing discs

Figure 3-5 (con't)

with Gal4 driver 27H05. (B) GFP expression in larval eye disc with Gal4 driver 27H05. (C) Ubiquitous GFP expression in adult with Gal4 driver 27H05. (D) GFP expression in adult legs with Gal4 driver 28G01. (E) Larval salivary gland and epidermis expression with Gal4 driver 28D03. (F) Adult abdominal expression with Gal4 driver 28D03. (G) GFP expression in adult mouthparts with Gal4 driver 37B05. (H)GFP expression in adult mouthparts with Gal4 driver 37B05. (I) GFP expression in adult abdomen with Gal4 driver 37B05.

Identification of active enhancers located within InR introns

We wished to delineate the exact structure of *InR* regulatory regions and identify possible regulation by the dFOXO transcription factor and the steroid hormone ecdysone, two important transcriptional inputs for *InR* gene regulation (Puig et al., 2003; Puig and Tjian, 2005; Gauhar et al., 2009). To identify enhancers in the *InR* introns and study their activities at molecular level, we dissected the *InR* introns into 25 fragments, each about 1.5 kbp in size (Figure 3-6A). As noted above, genome-wide assays for *cis* regulatory elements used synthetic basal promoters, which may lack functional compatibility with the endogenous enhancers. Therefore we tested these elements in combination with the endogenous basal promoter regions. The *InR* gene has three annotated transcription start sites, T1, T2 and T3 (Casas-Tinto et al., 2007). Genome-wide RNA polymerase II occupancy and the H3K4Me3 histone modification, linked with transcriptional start sites, showed strong association with T1 promoter, throughout different developmental stages (Figure 3-7) (modENCODE). Previous studies also showed that T1 promoter, as well as the mRNA isoform transcribed from T1 promoter, was the dominant one (Casas-Tinto et al., 2007). To compare promoter activities, we cloned the T1, T2 or T3 promoter into a luciferase reporter construct. As a negative control, a similar size intronic fragment (PT) was also tested (Figure 3-6A). We assayed all three promoters and the negative control fragment in both S2 and Kc cells (Figure 3-6B, C), and found T1 promoter activity was much higher than T2 and T3. In fact, T2 and T3 had similar activity as the negative control (PT) in the two cell types. Because this basal promoter appears to be the location for the majority of *in vivo* initiation, the T1 promoter was then used to assay the 25 intron fragments in reporter constructs.

Fragments driving higher levels of expression than the T1 promoter itself were considered potential active enhancers. To explore possible cell-type specificity, we tested the reporters in both S2 and Kc cells. Intron fragments 2, 3, 20, and 22 were found active in both two cell types, and the levels of activity varied. Fragments 4, 12, 15 were active in one of the two cell types, indicating cell-type specific enhancers (Figure 3-6B, C).

We find active elements in S2 cells in regions 2, 3, and 12 similar to findings from STARR-seq in S2 cells. However, in some cases there are disparities; we find no activity in region 6 or 23-25, where possible enhancers were detected in some STARR-seq assays, while regions 20 and 22 were robust activators, but not consistently identified in STARR-seq (Figure 3-6A, B, C). These differences may reflect differences between these assays; we tested longer fragments (1.5 kbp vs. 600 bp), thus were less likely to divide and inactivate an enhancer element, and our assays relied on enhancers communicating with the endogenous basal promoter region, which may provide compatibility lacking in the genome-wide approach. Our assay was also able to detect possible repressor activity, because the T1 basal promoter region provides a strong enough signal to detect such interference. For example, fragments 6 and 7 showed evidence for putative repressors; fusions with these elements generated signals lower than that of the T1 promoter alone (Figure 3-6B, C). Such activity would not be identified by STARR-seq.

We asked whether the alternative T2 or T3 promoters, which exhibited intrinsically low activity in reporter assays, might be stimulated by regulatory elements identified above. Fusion constructs containing regions of 2 or 3 robustly activated transcription from T1,

but candidate promoter regions T2 and T3 did not result in significant gene expression, suggesting that these basal elements are not likely to generate much of the overall transcriptional output, a conclusion supported by RNA-seq analysis (Figure 3-8) (Graveley et al., 2011; Brown et al., 15AD; Attrill et al., 2016). Previous studies have focused largely on the regulatory potential of T2 (Puig et al., 2003; Casas-Tinto et al., 2007); our analysis indicates that much of the regulatory activity of this locus is likely channeled through the distal T1 promoter.

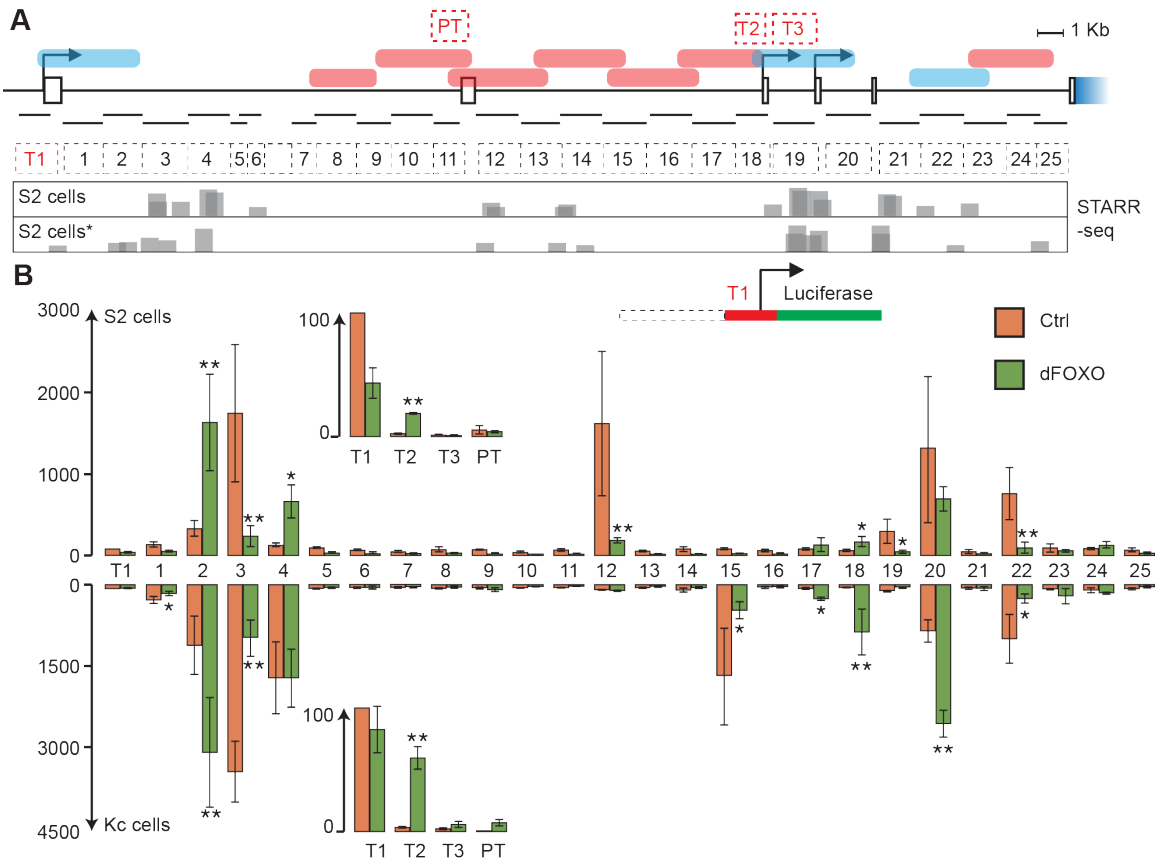
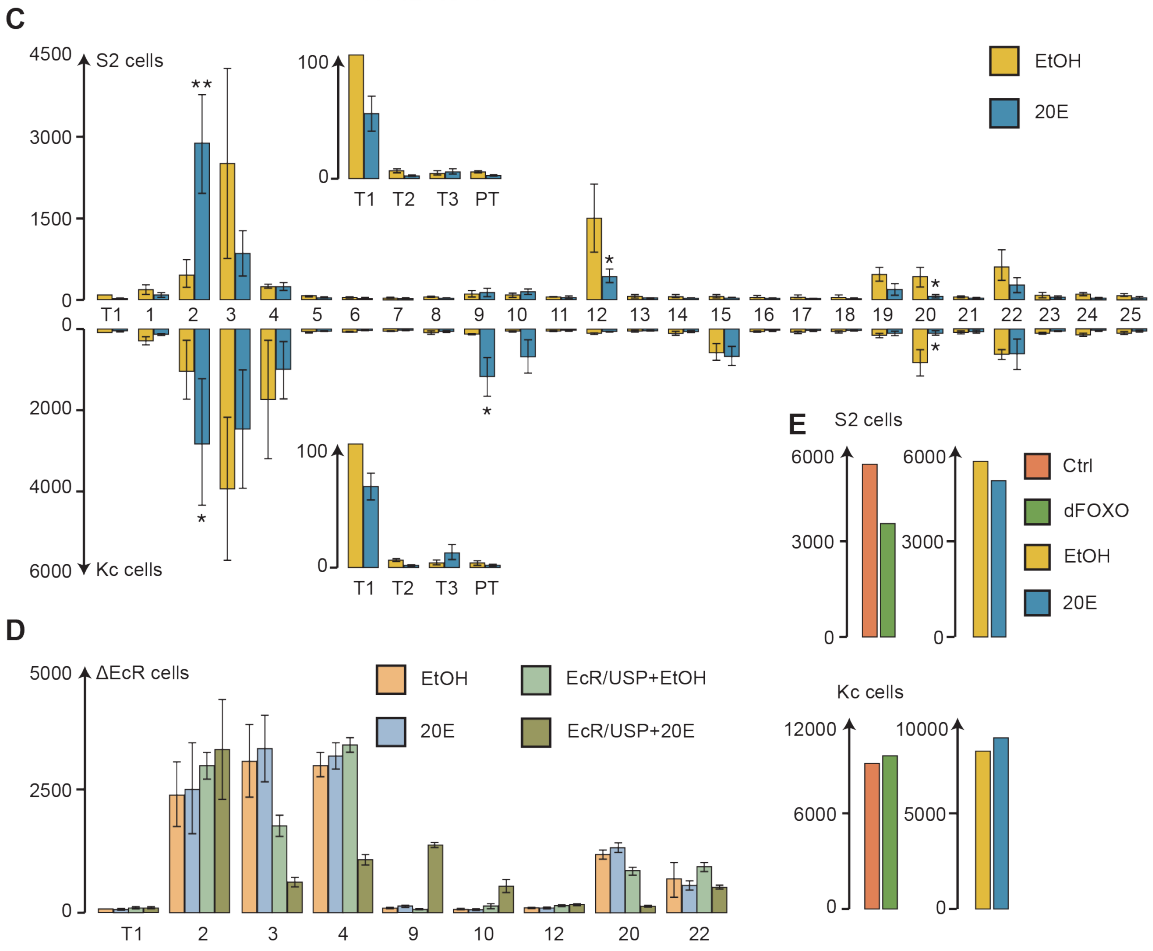


Figure 3-6: *InR* introns contain *cis*-regulatory elements that respond to dFOXO and/or ecdysone signaling.

Figure 3-6 (con't)



(A) Overall structure of the *InR* gene and scheme for reporter library. The intronic regions of the *InR* gene were divided into 25 fragments. Each fragment (~1.5 kb) was fused to 5' of promoter proximal region of the first transcription start site (T1) to drive the luciferase reporter gene expression. The promoter proximal regions of second (T2) and third (T3) transcription start sites were also used to drive luciferase gene expression. As a negative control for promoter activity, an internal region that overlaps with fragment 11, and contains a portion of second exon was tested (PT). The regulatory output of putative enhancer fragments from the Janelia Gal4 collection are indicated

Figure 3-6 (con't)

above the gene structure (Figure 3-4A) with red for active enhancers, and blue for less active or inactive elements. STARR-seq data from two independent studies are presented below the gene structure (Figure 3-4B), with two replicates from each study, and darker areas indicating overlapping signals from the replicates (Arnold et al., 2013; *Shlyueva et al., 2014). A small repeat-rich region between fragments 6 and 7 was unclonable. (B) Reporter activities in S2 or Kc cells, and their responses to dFOXO expression. The reporters were co-transfected in S2 or Kc cells with dFOXO expression construct. As a control, an empty expression vector (pAX vector) was transfected. Individual elements showed either increased or decreased activity in response to dFOXO expression. The luciferase activity of T1 reporter was set as 100, and all other reporters were normalized to T1, since they are all fused to the 5' of T1 (except for T2, T3, and PT). In this and following figures, at least three biological replicates with three technical replicates were performed for each reporter. Error bars indicate s.d. from biological replicates. (* $p<0.05$, ** $p<0.01$) (C) Reporter activities in S2 or Kc cells, and their responses to 20 hydroxyecdysone (20E). The cells were treated with 20E (10^{-5} M) in ethanol 24 hours after transfecting the reporters. Control cells were treated with ethanol. (D) Transcriptional responses to 20E treatment require the presence of EcR. Reporters that respond to 20E treatment in S2/Kc cells fail to respond in an EcR-deficient cell line (Kc-derived ΔEcR cells) (also Figure 3-9). However, these responses were restored by transfecting ΔEcR cells with EcR/USP expression vectors. (E) Test for additivity of enhancer function; overall activity and response to dFOXO or 20E was calculated from output of individual reporters from (B) and (C) (above activity of T1 alone). The

Figure 3-6 (con't)

summed values of the 25 fragments would indicate that dFOXO treatment would have a net repressive effect, and 20E no net change, if enhancers work additively.

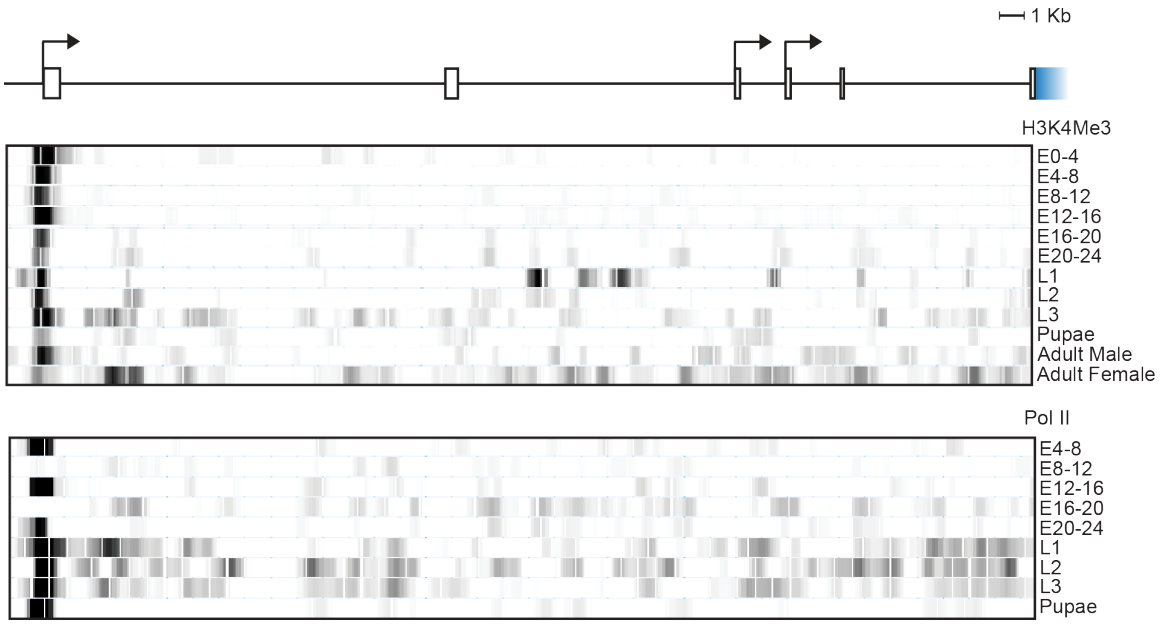


Figure 3-7: RNA polymerase II and H3K4Me3 signals at *InR* locus. An enrichment of RNA polymerase II and histone modification H3K4Me3 at T1 suggests that this site is the major locus for initiation of transcription. Data are presented as peak intensity, and the scale is normalized to local maximum (Berkeley Drosophila Transcription Network Project Chromatin Accessibility, Kaplan et al., 2011; Li et al., 2011; Thomas et al., 2011).

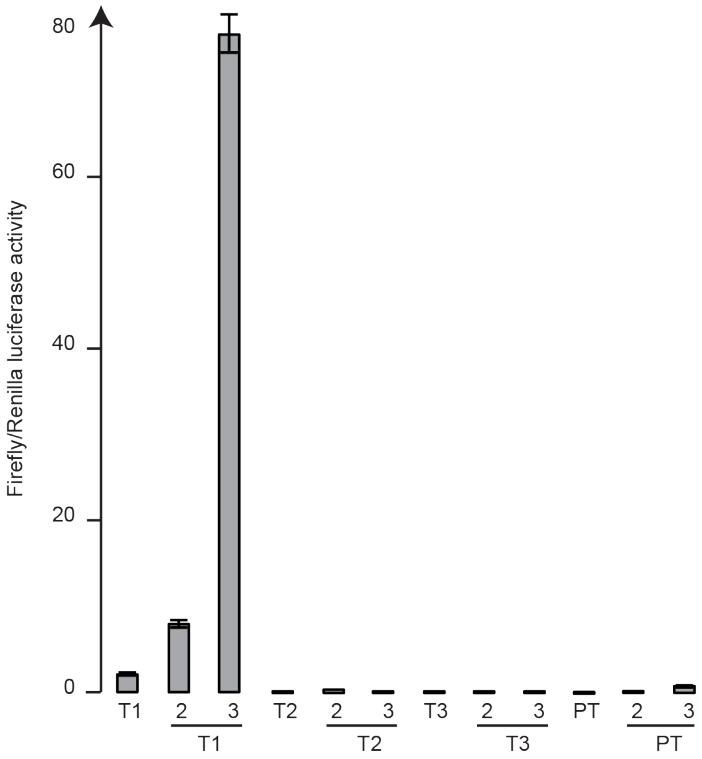


Figure 3-8: Preferential activation of T1 promoter element by *InR* intronic enhancers 2 and 3. Intron fragments 2 or 3 were fused to T1, T2, T3, or PT. The enhancer fragments were active when fused to T1, but were largely unable to stimulate reporter gene transcription when fused to T2, T3 or the negative control PT. Error bars indicate s.d. from three technical replicates.

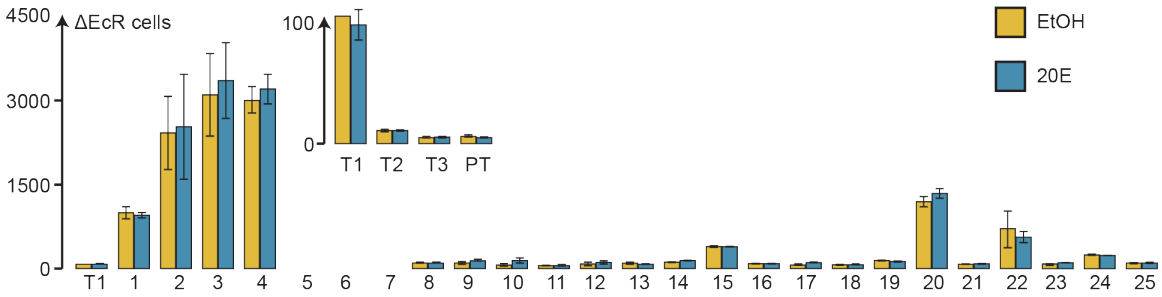


Figure 3-9: Requirement for EcR protein for 20E response in Kc cells. Fragments found active in wild-type cells do not respond to 20E treatment in the EcR-deficient cell line. Error bars indicate s.d. from three technical replicates.

Dynamic regulation of InR enhancers by dFOXO

In addition to its role in mediating the downstream activities of insulin signaling, previous studies have highlighted the potential for direct transcriptional activity of the dFOXO transcription factor on *InR* expression (Jünger et al., 2003; Puig et al., 2003; Casas-Tinto et al., 2007). Binding sites for dFOXO are present at T2, and reporter genes containing these sequences are activated by dFOXO (Jünger et al., 2003; Puig et al., 2003; Casas-Tinto et al., 2007). In light of the overall stimulatory effect on *InR* expression by dFOXO, it has been suggested that this T2 activation is path by which dFOXO acts on the gene (Puig et al., 2003; Casas-Tinto et al., 2007). We assayed each element surveyed above for activation by dFOXO by co-transfection of the reporters with a dFOXO expression vector. Consistent with previous reports, the weak T2 promoter was activated by dFOXO in both S2 and Kc cells, although overall activity was modest (Figure 3-6B). By contrast, robust activation by dFOXO was observed with region 2, and region 4 in S2 cells (the T1 promoter was itself slightly repressed by dFOXO expression in S2 cells, but this effect was apparently not dominant on all enhancers). Strikingly, expression of dFOXO had a strong and significant negative effect on other elements, including regions 3 and 22, which were repressed in both cell types. Fragment 20 was repressed in S2 cells, whereas it was activated by dFOXO expression in Kc cells (Figure 3-6B).

Most dFOXO response fragments may be indirectly regulated by dFOXO

To determine if the transcriptional effects mediated by dFOXO were a consequence of direct interaction of the protein with these regulatory elements, we performed ChIP

analysis using anti-dFOXO serum. A previously characterized direct target of dFOXO, the *Thor* (*4EBP*) promoter was used as a positive control, which showed strong endogenous dFOXO binding (Figure 3-10A, B) (Teleman et al., 2008; Alic et al., 2011; Bai et al., 2013). The previously characterized dFOXO-bound T2 (region 18) also showed lower but significant dFOXO enrichment (Figure 3-10A, B) (Puig et al., 2003; Casas-Tinto et al., 2007). Surprisingly, none of the other elements that were transcriptionally regulated by dFOXO expression exhibited strong binding by the factor (Figure 3-10A, B). A prominent peak was observed on fragment 10, an element that however was not activated or repressed by dFOXO (Figure 3-10A, B). To further ascertain whether the signals we observed represented dFOXO binding, we treated cells with insulin to activate the signaling pathway, which should result in phosphorylation and exclusion of endogenous dFOXO from the nucleus, or subjected cells to serum starvation, which should reduce signaling and increase dFOXO activity (Puig et al., 2003). Indeed, insulin treatment resulted in reduced dFOXO ChIP signals on *Thor*, as well as regions 10 and 18, while starvation increased the ChIP signal as expected (Figure 3-10C). These observations are consistent with the ChIP signals representing dFOXO interaction.

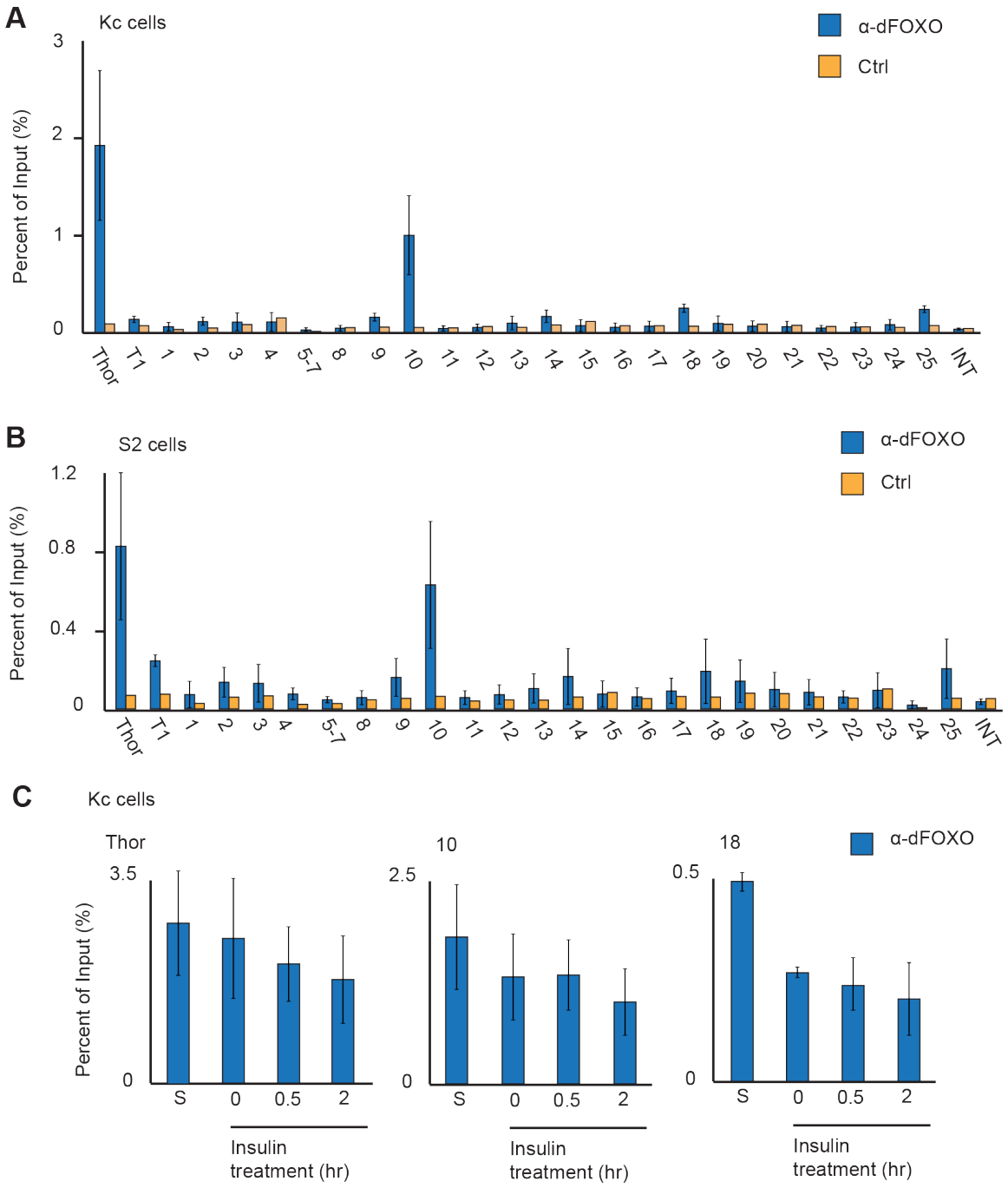


Figure 3-10: In vivo occupancy of dFOXO provides evidence for direct and indirect regulation of *InR*. (A) Binding of dFOXO to *InR* genomic regions measured by chromatin immunoprecipitation using chromatin from Kc cells. The *Thor/4EBP*

Figure 3-10 (con't)

promoter served as a positive control. Regions 10, 18, and 25 showed the highest signals. Error bars show s.d. for five biological replicates. (B) Similar association of dFOXO to the *InR* locus measured by chromatin immunoprecipitation in S2 cells. Error bars show s.d. for four biological replicates. (C) Starvation (S) enhances dFOXO binding on *Thor* promoter, and on *InR* intron fragments 10 and 18 (which overlaps with T2), whereas insulin treatment (0.5 and 2 hours) reduces the measured occupancy.

Ecdysone induces similar responses to dFOXO

Ecdysone treatment increases expression of *InR*, however the molecular mechanism of this regulation has not been elucidated (Gauhar et al., 2009). To determine how this important regulatory hormone might affect the transcriptional elements of *InR*, we treated cells with 20-hydroxyecdysone (20E) 24 hours after transfecting the cells with *InR* reporters, and measured activities after another 24 hours. The T1 promoter was slightly repressed by 20E treatment in both cell types, while T2 and T3 showed no significant changes (Figure 3-6C). Fragment 2 was robustly activated, whereas 3 and 20 were significantly repressed by 20E treatment in both cell types. In Kc cells, we observed cell-type specific activation of elements 9 and 10, which alone had not shown significant transcriptional potential. A greater number of elements showed reduction in activity after 20E treatment, though some of these effects were modest (Figure 3-6C). To determine whether these 20E responses required the ecdysone receptor (EcR), we assayed reporters in an *EcR* deficient cell line (Δ*EcR*) that was derived from Kc cells (Swevers et al., 1996). None of the fragments showed any response to 20E treatment in Δ*EcR* cells, including all three promoters (Figure 3-6D, Figure 3-9). We transfected the Δ*EcR* cells with EcR and USP (the heterodimeric partner of EcR) expression constructs, and assessed 20E activation or repression. In this setting, regulation similar to that observed in 20E treated wild type Kc cells was now observed, confirming the role of EcR in this regulation (Figure 3-6D).

Interestingly, many of the elements tested showed similar responses to 20E and dFOXO, including 2, 3, 12 and 20, suggesting the involvement of linked pathways (Figure 3-6B, C). Significantly, 20E signaling has been shown to affect dFOXO localization by

regulating PI3K activity, suggesting that some of the 20E effects may be mediated by dFOXO activity (Colombani et al., 2005). In addition, dFOXO has been reported to bind directly to the USP co-factor of EcR (Koyama et al., 2014). Thus, 20E may regulate some enhancers in the *InR* gene via dFOXO activity. To test if dFOXO regulation is dependent on EcR, we measured the effect of dFOXO overexpression on reporters in Δ*EcR* cells, and found that the transcription factor was able to regulate these elements as in wild-type Kc cells (Figure 3-11). Thus, 20E and dFOXO may share downstream pathways to regulate several enhancers in *InR* gene, but EcR is not required for dFOXO activity itself.

To gain further insight into the regulation by 20E, we compared our data in S2 cells to STARR-seq analysis in the same cell line treated with 20E (Figure 3-12) (Shlyueva et al., 2014). Essential features are confirmed in both studies; region 2 is activated by 20E, which corresponds to a direct binding site of EcR (Figure 3-12). Upon hormone addition, this binding is lost, indicating that EcR acts here as a repressor. Interestingly, regions 9 and 10 were not found to act as transcriptional control elements in S2 cells, but they do exhibit a loss of EcR binding upon 20E treatment. We find that in Kc cells, these regions in fact function in a cell-type specific manner, possibly because Kc cells possess an additional activator for these elements. Regarding repression of elements by 20E, both our data and STARR-seq are in agreement that regions 3, 12, 19, and 22 have reduced activity after 20E treatment. Less reproducible signals from STARR-seq (regions 13 and 14) were not supported by our assays, suggesting that these STARR-seq signals may be spurious. None of the 20E repressed areas correlated with directly bound EcR peaks, thus these elements may be subject to indirect regulation. One gene induced

by EcR is *Eip74EF*, which functions as a repressor (Shlyueva et al., 2014). ChIP-seq analysis of this protein in embryos indicates that this protein may interact with repressed regions 19 and 20, but other repressed elements may be repressed by a different factor (modENCODE). Eip74EF is also found to bind to region 2, which was activated by 20E. We speculate that this binding may represent a progressive gene switch, in which initial derepression after loss of EcR binding is later followed by repression, as EcR-driven Eip74EF repressor levels increase.

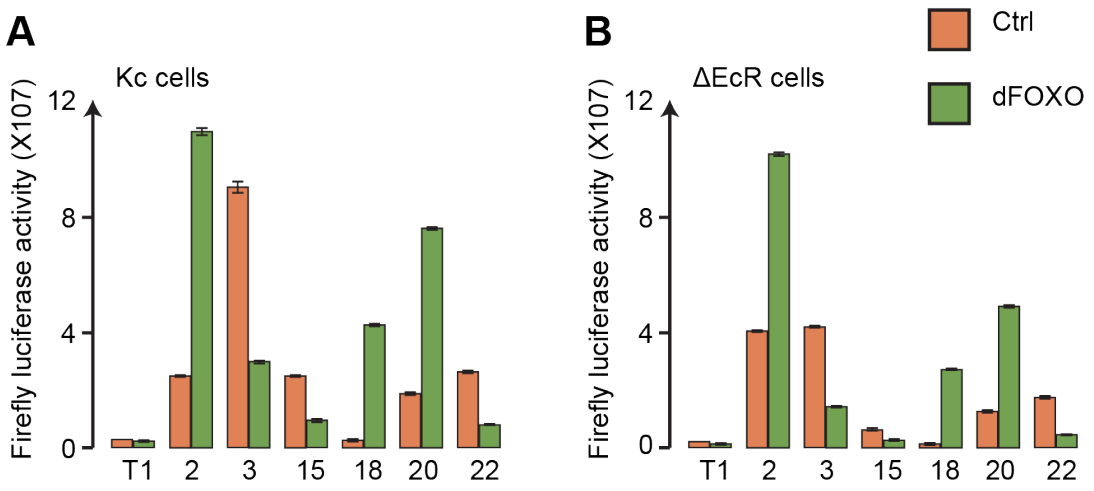


Figure 3-11: EcR is not required for dFOXO activity on *InR* enhancers. (A) Activities of selected reporters in response to dFOXO overexpression in Kc cells. (B) Activities of selected reporters in response to dFOXO overexpression in ΔEcR cells. Error bars indicate s.d. from three technical replicates.

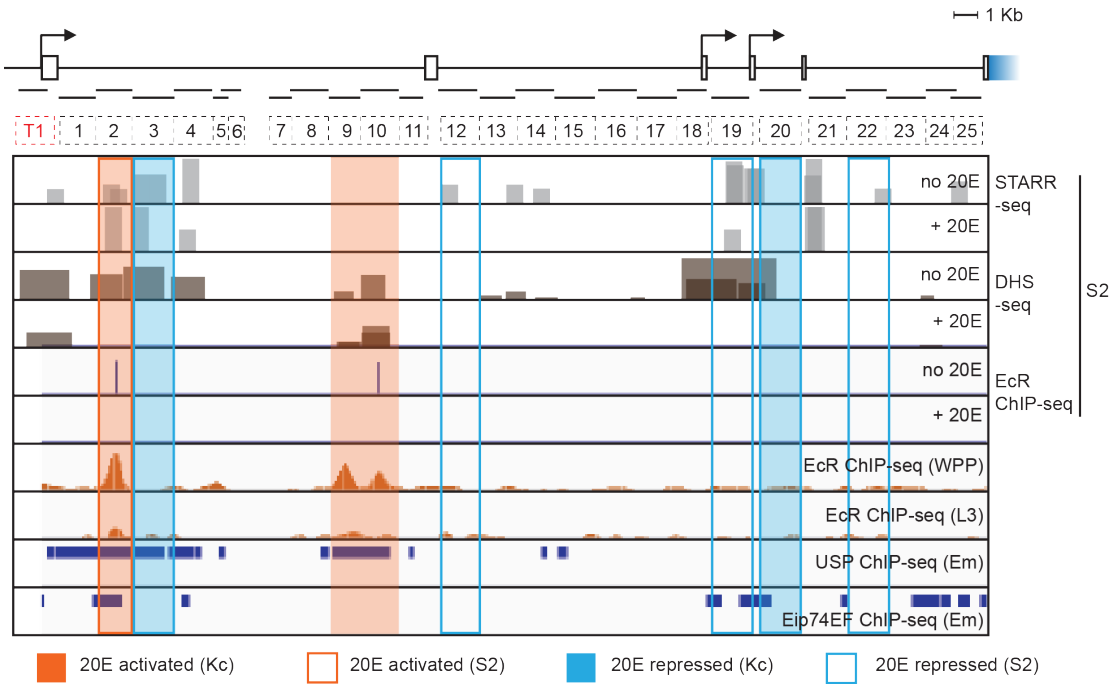


Figure 3-12: Ecdysone signaling involves direct activation and indirect repression of *InR* enhancers. 20E responsive enhancers identified in this study are plotted in orange (activated) and blue (repressed). 20E responsive enhancers identified from STARR-seq are aligned below structure of the gene, with the corresponding DNase I-seq data (Shlyueva et al., 2014). Gray bars indicate identified enhancer or DNase I site. For each of the datasets, results from the two replicates are shown, with darker areas indicating reproducible signals. Below, thin blue bars show two ChIP-seq peaks for EcR in S2 cells without or with 20E treatment; their loss identifies 20E-sensitive binding (Shlyueva et al., 2014). These peaks align with EcR ChIP-seq data in pupae (WPP) and larvae (L3) (modENCODE). The regions bound by EcR partner protein USP and EcR-regulated repressor Eip47EF (modENCODE) are shown as blue bars. The 20E activated fragments colored in orange overlap with EcR/USP binding sites, while 20E repressed fragments colored in blue lack EcR binding; about half of them overlap with Eip74EF binding.

Impact of RB binding site on InR promoter and enhancers

The direct actions of dFOXO and EcR in *InR* expression had previously been supported by genetic and biochemical evidence. More recently, we and the Dyson laboratory noted that the T1 proximal promoter region of the *InR* gene is occupied *in vivo* by the Rbf1 retinoblastoma tumor suppressor protein, the homolog of mammalian RB (Acharya et al., 2012; Korenjak et al., 2012; Wei et al., 2015). This promoter region is also bound by the Rbf1 paralog, Rbf2, and a set of proteins that frequently associate with E2F/Rb proteins, the dREAM complex (Georlette et al., 2007; Wei et al., 2015). The binding of Rbf1 appears to be of functional significance, for a reporter gene driven by the T1 promoter region is repressed by Rbf1 expression in S2 cells (Raj et al., 2012). To further explore the significance of Rbf protein interaction with the *InR* T1 promoter, we removed a 100 bp fragment centered underneath the Rbf1 binding peak (ΔRbf1) (Figure 3-13A). This mutant ΔRbf1 promoter showed modest but reproducibly higher activity than the wild-type T1 promoter, indicating that this Rbf1-binding portion of the promoter acts to downregulate expression (Figure 3-13B). Because the T1 promoter proximal region has itself a relatively modest transcriptional output, we explored the significance of Rbf1 in the context of more active reporters, containing the active elements 2, 3 or 12. Particularly for the fusion containing region 3, the transcriptional impact of the small T1 deletion was much larger in absolute terms than that observed for just the basal promoter itself, suggesting that Rbf1 may reduce the functionality not just of local activators within T1, but compromise the utility of the basal promoter for element 3 (Figure 3-13C). Similar "booster" roles for basal elements have been noted in developmentally active genes (Yuh and Davidson, 1996). The removal of the Rbf1

binding region did not change the effects of dFOXO expression, which activated element 2 and repressed 3 and 12 (Figure 3-13C).

As a co-repressor, Rbf1 binds to dE2F1 to block its activation function (Du and Pogoriler, 2006). Removal of the Rbf1-binding element, which includes E2F motifs, does not abrogate function of T1, suggesting that other regulatory sites contribute to this promoter's activity. We tested whether dE2F1 activates the T1 promoter, and whether this required the region involved in Rbf1 recruitment. Cotransfection of dE2F1 significantly upregulated reporters containing the wild-type as well as the ΔRbf1 T1 promoter; the fold stimulation was similar to that observed for the control *PCNA* promoter (Figure 3-13D). *PCNA* was maximally stimulated at lower concentrations of transfected dE2F1, suggesting that it may have different affinities for dE2F1 binding (Figure 3-14). The ΔRbf1 T1 promoter was activated by dE2F1 to a higher level than the wild-type T1 promoter, consistent with the removal of the repressive function of Rbf1, and indicating that there might be additional dE2F1 binding sites that are not suitable for Rbf1 recruiting, or that the activation occurs through an indirect effect via other transcription factors. These data suggest that Rbf1 has a repressive function on the T1 promoter, and more than merely interfering with local activators, the action of this corepressor may generally influence the ability of linked regulatory regions to fully engage and stimulate transcription from the T1 start site. This mode of regulation contrasts to the all-or-nothing effect observed for Rbf1 and Rb family proteins in general on cell-cycle promoters (Raj et al., 2012).

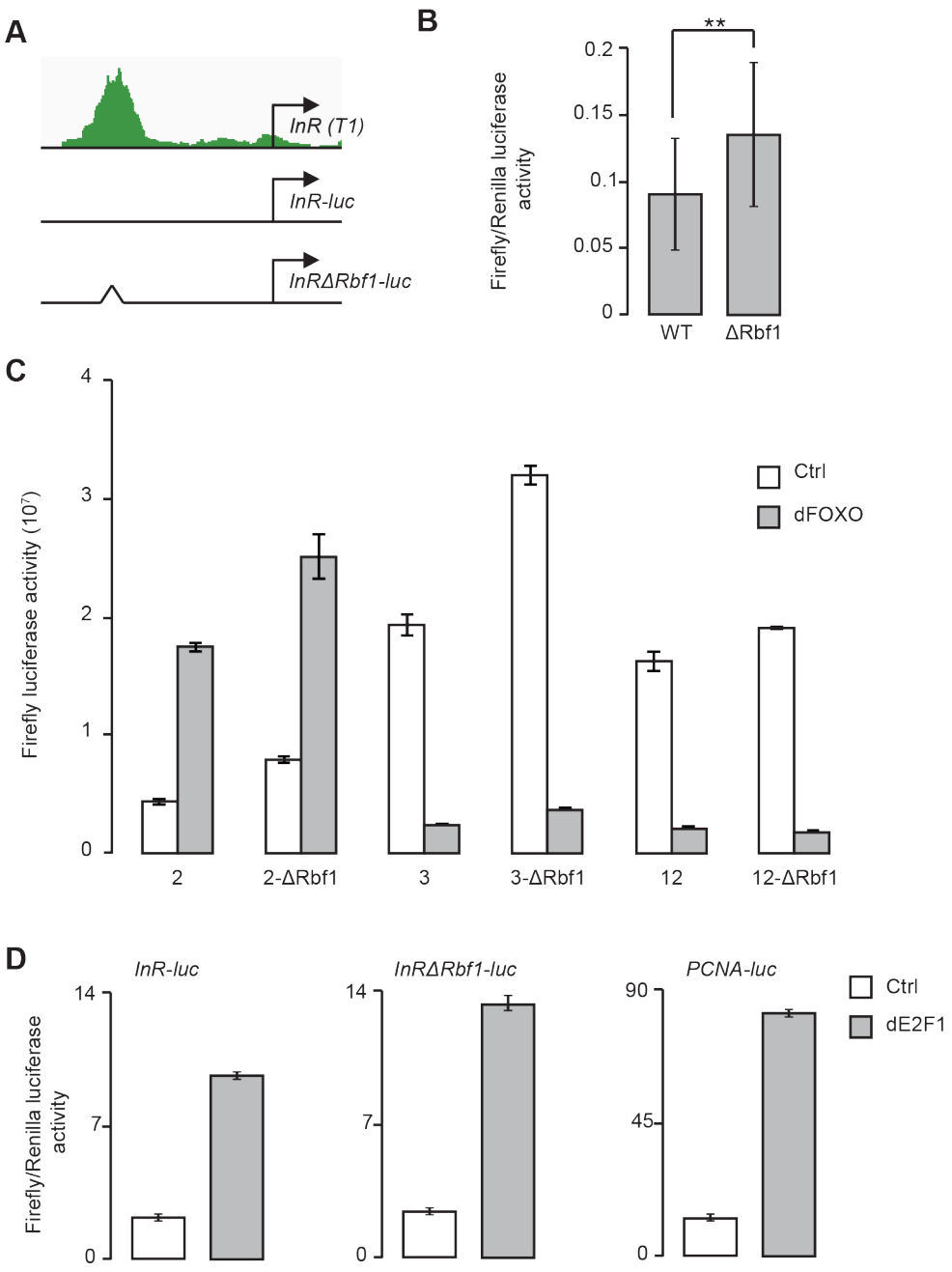


Figure 3-13: Promoter-proximal Rbf1 binding site role in transcriptional regulation. (A) Position of interaction of Rbf1 with *InR* T1 promoter proximal region (Acharya et al., 2012). A mutant T1 promoter-luciferase reporter was generated by deleting a 100 bp region associated with the Rbf1 binding peak (*InRΔRbf1-luc*). (B)

Figure 3-13 (con't)

Deletion of Rbf1-associated region modestly increases T1 basal promoter activity (**p<0.01). Error bars indicate s.d. from four biological replicates. (C) Transcriptional activation by three *InR* intronic enhancers are potentiated by removal of Rbf1 binding region. The deletion of the Rbf1 binding site had a greater impact on proximal enhancers 2 and 3 than distal enhancer 12, but all showed higher levels of activities on the *InRΔRbf1-luc* construct (white bars). dFOXO-mediated activation or repression was unaffected by loss of Rbf1 binding region (gray bars). (D) dE2F1 activation of the T1 *InR* promoter. The *InRΔRbf1-luc* was activated by dE2F1 to higher levels than the wild-type promoter, indicating there may be dE2F1 binding sites outside the Rbf1 peak, or indirect dE2F1 regulation. A canonical Rbf1/dE2F1 target, the *PCNA* promoter, was used as a positive control. For C and D, error bars indicate s.d. for three technical replicates.

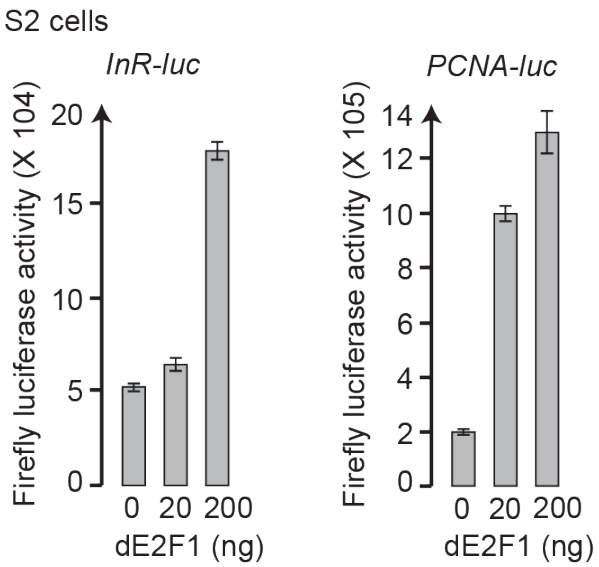


Figure 3-14: Differential sensitivity of *InR* T1 promoter and *PCNA* promoter to dE2F1. To achieve maximal induction of the *InR* T1 promoter, transfection of 200 ng of dE2F1-expression vector is required, while 20 ng of the dE2F1 vector is sufficient to reach near maximal activation of the more active *PCNA* promoter. Error bars indicate s.d. from three technical replicates.

Transcriptional circuitry of the InR gene revealed by precise-mapping of CREs

To achieve a higher resolution and better understanding of the transcription circuitry regulating *InR* gene, we further analyzed each of the active enhancers and dFOXO/20E response enhancers by making serial deletions (~300 bp each, M1~M5) in each of the selected intron fragments (Figure 3-15A). Each fragment was tested in S2 and Kc cells and assessed for response to dFOXO expression or 20E treatment. The deletion series revealed portions of each enhancer necessary for baseline activation, as well as regions that possess inherent repressive potential. Some deletions attenuated or abrogated response to dFOXO or 20E. With enhancer 2, a dFOXO- and 20E-activated element, removal of region M1 reduced the basal activity to the same level as T1, suggesting that the region contains an essential activator binding site(s) (Figure 3-15B). Removal of M3 greatly induces the basal activity, indicating the presence of a repressor binding site(s). dFOXO induction was somewhat attenuated by removal of either M3 or M5, indicating potential dFOXO-dependent activator binding sites. We summarized the effects of these mutations as symbols for constitutive activators or repressors, or dFOXO- or 20E-dependent activator or repressor effects (Figure 3-15C). For enhancer 2, removal of M3 produced a complex effect with 20E treatment; the baseline expression increases, but the ability for 20E to activate is lost, rather, this treatment causes repression (Figure 3-15B). Repression on this element is almost certainly due to the direct binding of EcR, as this protein has been found to bind within this region (Figure 3-12) and the removal of segment M3 has no derepressive effect in cells lacking EcR (Figure 3-15B). We propose that the M3 mutant is repressed rather than activated by 20E treatment because the region contains activator sites, in addition to EcR binding sites, and these activator sites

are important for overall enhancer activity. 20E-treatment removes EcR and simultaneously triggers expression of repressors (such as Eip74EF, which may act on 20E-repressed enhancers 3, 12 etc.). The weaker complement of activators left on this version of enhancer 2 would in this view be dominantly suppressed by action of these 20E-induced repressors, while a wild-type enhancer would not. We analyzed all of the regulatory fragments using the same deletion analysis in both cell types, as well as *ΔEcR* cells, and tested for responses to dFOXO and 20E (Figure 3-16). The results are summarized for all elements, using symbols to indicate the presence of activator or repressor activities in sub-regions M1-M5 (Figure 3-15 for results in Kc cells; Figure 3-17 for results in S2 cells; Figure 3-16 for all data).

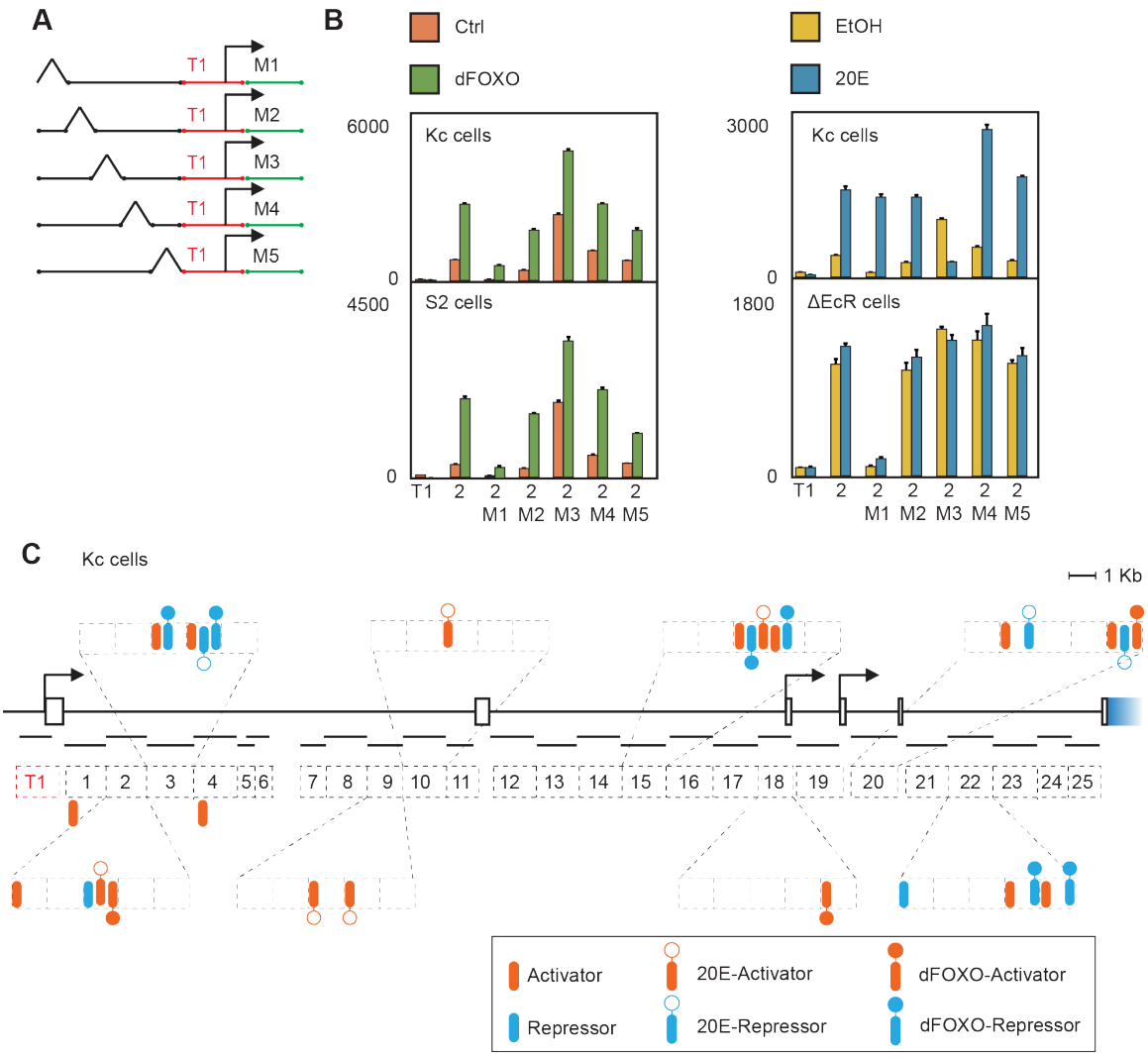


Figure 3-15: High-resolution analysis of the *InR* gene intronic enhancers unveils complex molecular logic circuits for *cis*-regulation. (A) Deletional analysis of active intronic enhancers. Constitutive active or inducible elements were subjected to serial deletions, removing blocks of ~300 bp to yield derivatives M1-M5. The truncated reporters were then tested in S2 and Kc cells as shown in Figure 3-6. (B) Activities of truncated reporters derived from intronic fragment 2. M1 removes a constitutive activator activity, while M3 removes a constitutive repressor, as well as 20E and

Figure 3-15 (con't)

dFOXO activation potential. Entire dataset shown in Figure 3-16. Error bars indicate s.d. for three technical replicates. (C) A comprehensive map of enhancers and imputed regulatory factors based on analysis of all mutant reporters in Kc cells. Results for S2 cells are shown in Figure 3-17.

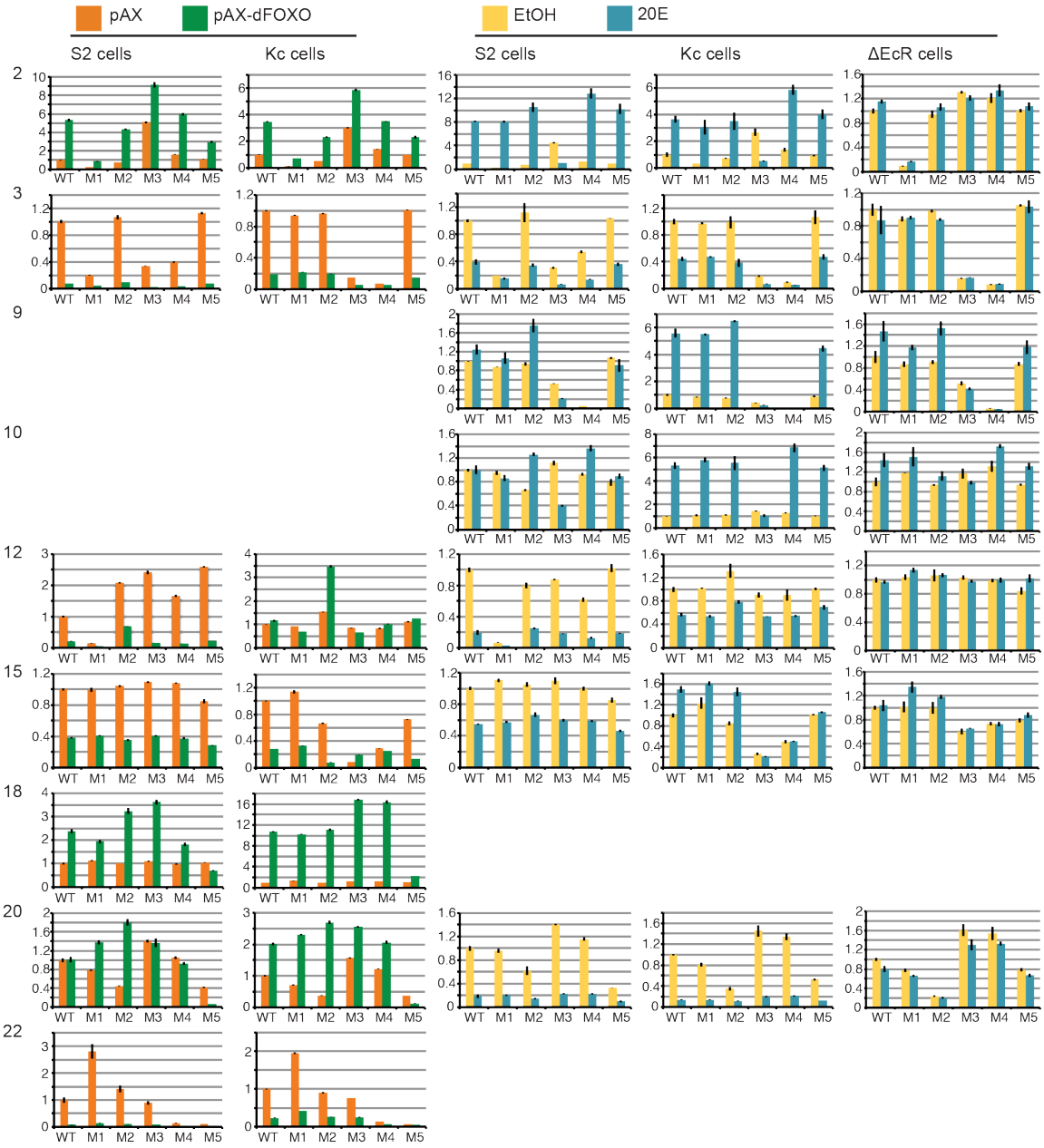


Figure 3-16: Complete data for deletional analysis outlined in Figure 3-15. Enhancer numbers indicated at left. Enhancers 9 and 10 were not dFOXO responsive, thus were not tested for dFOXO effect. Similarly, 18 and 22 were not 20E regulated, thus were not tested for 20E effect. Error bars indicate s.d. from three technical replicates.

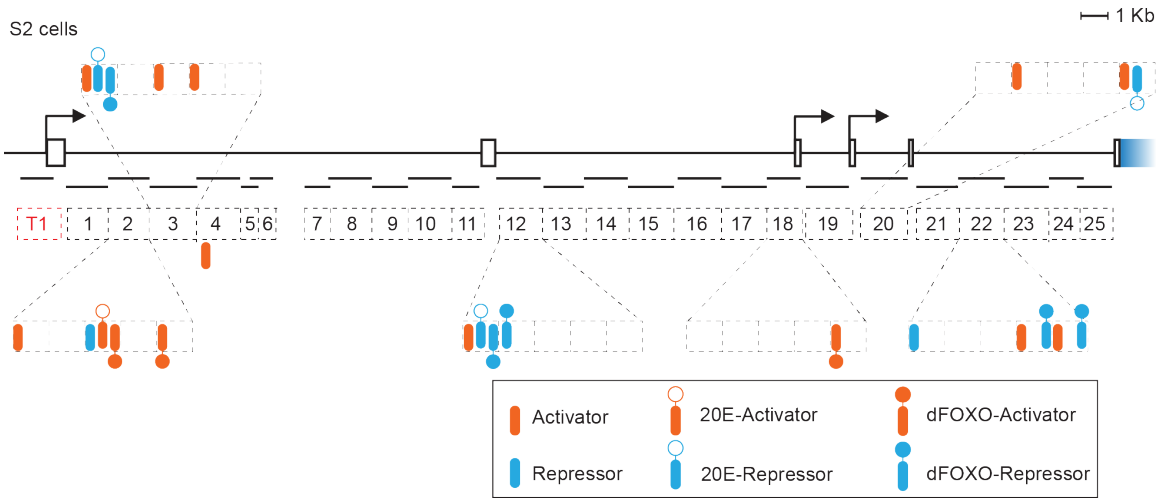


Figure 3-17: Summary of high-resolution analysis of the *InR* gene intronic enhancers and imputed regulatory factors in S2 cells.

Combinatorial interactions of InR regulatory elements

Our detailed analysis of the *cis*-regulatory landscape of the *InR* gene indicates that multiple, parallel-acting elements contribute to overall regulation of expression. Early studies emphasized the modularity of multiple enhancers acting on developmental genes, but a number of studies have shown how some discrete *cis* regulatory elements function in combinatorial manners (Small et al., 1993; Marinić et al., 2013; Bothma et al., 2015). Are the regulatory units identified in the *InR* gene independently-acting units that function in an additive manner, or might there be higher-order interactions? To test this possibility, we compared the sum of transcriptional output of the mapped *cis* regulatory elements with the responses of the endogenous gene as a function of dFOXO or ecdysone regulation. Two lines of evidence indicate that overall, dFOXO signaling activates the endogenous *InR* gene; first, nutritional limitation activates dFOXO and is correlated with increased levels of *InR* mRNA, and second, overexpression of dFOXO increases levels of endogenous *InR* mRNA (Kramer et al., 2003; Puig et al., 2003; Casas-Tinto et al., 2007; Gershman et al., 2007). We summed the transcriptional outputs of all elements before and after expression of dFOXO in S2 or Kc cells (Figure 3-6E). The overall summed output indicated that if regulatory elements work in a simple, additive manner, dFOXO should lead to a net loss of *InR* expression in S2 cells and no change in Kc cells, which contradicts the previous observations from the endogenous gene. Three possibilities would explain why the simple summation model is incorrect: (1) The relative levels of activity of the enhancers may be position dependent; if element 2 plays a disproportion role, then the overall output would be activation. (2) The direct induction of *InR* represents a short-term response measured within hours, while the

activities measured by reporter assay may reflect longer-term adaptations, including secondary gene circuit activity involving indirect regulation. (3) When positioned together, enhancers may exhibit different activities through synergistic activities. Similar explanations may account for the net effects of 20E treatment; simply summing enhancer activities, 20E would have little effect on *InR* expression, but endogenous *InR* mRNA levels increase within a few hours of 20E treatment (Figure 3-18). The temporal control model (2) suggested above is supported by the time series expression of *InR*; after an initial peak at 3-6 hours, levels of the mRNA decrease again to near initial levels, presumably under the influence of repressors such as Eip74EF that are EcR targets. To test for potential combinatorial interactions, we fused *cis* regulatory regions together and compared their activities to the individual parts (Figure 3-19). For regions 2 and 3, the enhancers showed sub-additive behavior, meaning that the sum was somewhat less than the individual activities. This effect may be simply a function of distance-dependent activation, a well-known property of *cis* regulatory elements (despite the generalization that enhancers should work in a distance-independent manner) (Banerji et al., 1981). While our reductionist analysis of the *cis* regulatory elements of this gene serve to identify key properties of each of these molecular switches, a quantitative combinatorial understanding will come from re-integrating this information in the intact locus.

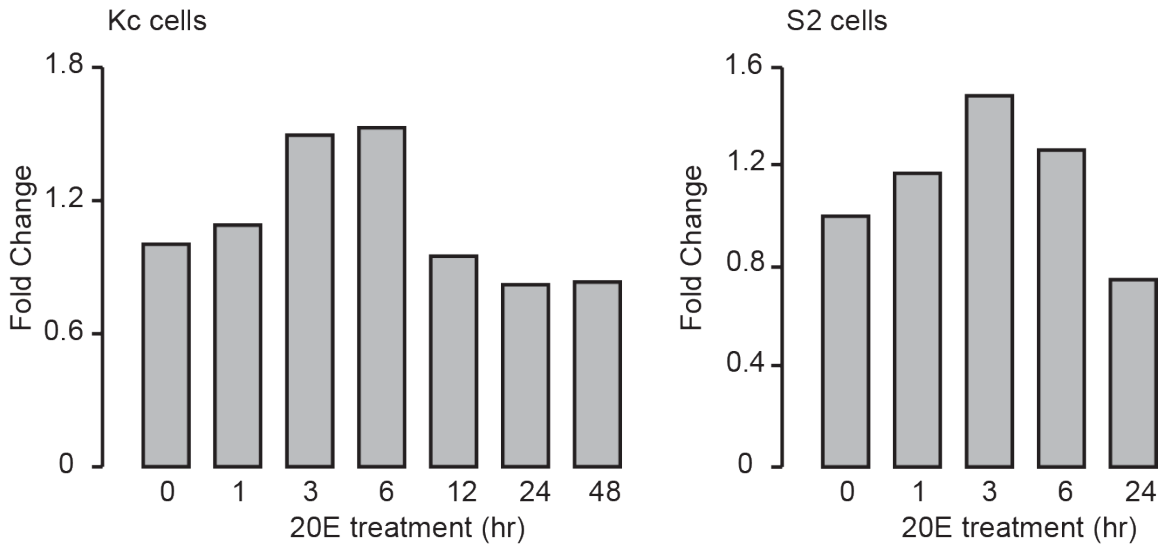


Figure 3-18: Kinetics of induction of endogenous *InR* transcript levels upon 20E treatment. Kc cells were treated 20E for 1, 3, 6, 12, 24, and 48 hours. Endogenous *InR* transcript levels reached a maximum (~1.5 fold induction) between 3 and 6 hours after treatment, and then decreased approximately to levels seen in untreated cells after 24 hours. The primers used in the qPCR analysis were designed to measure *InR* transcript A, the major isoform transcribed from the T1 promoter. The data was normalized to transcript levels of *RpS13*. We also observe this transient induction in S2 cells (right panel). An induction of the native *InR* gene in Kc cells three hours after 20E treatment was also noted in a previous study (Gauhar et al., 2009).

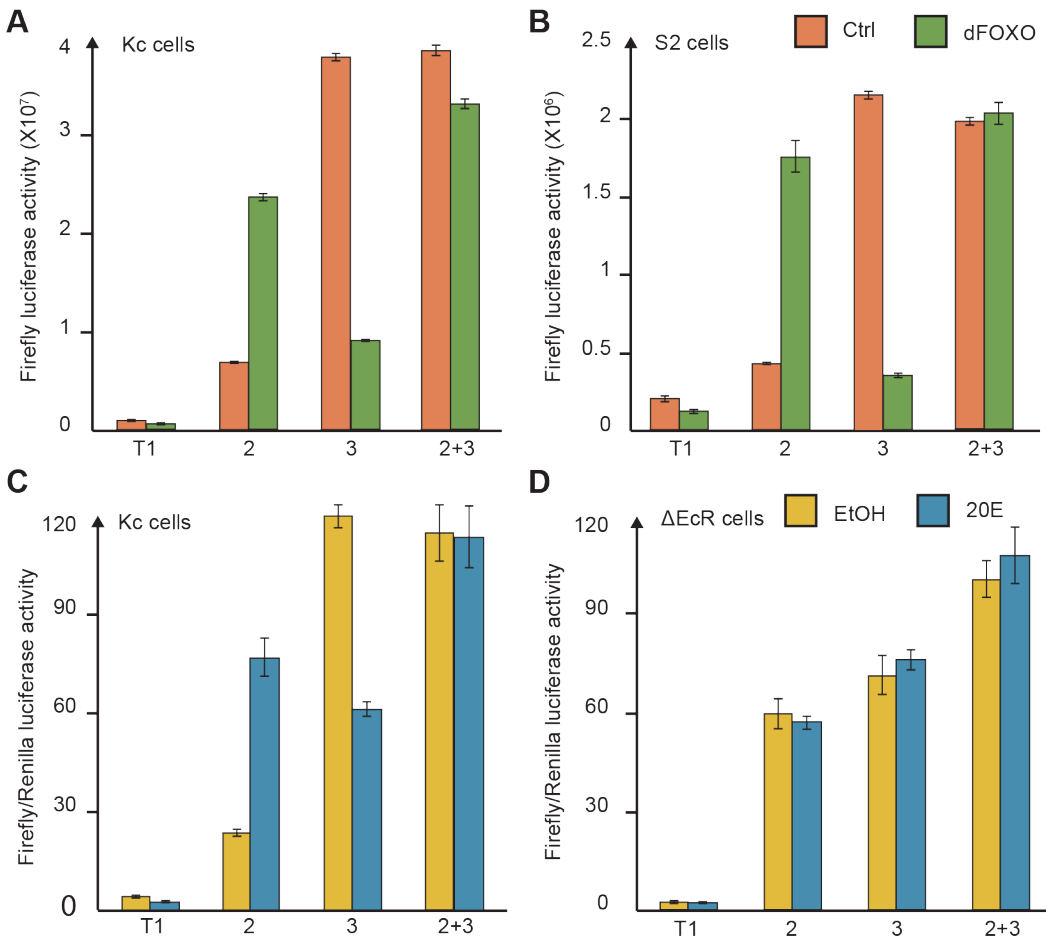


Figure 3-19: Non-additive behavior of *InR* enhancers. (A) Activity of 2, 3, or 2+3 enhancers and response to dFOXO in Kc cells. (B) Activity of 2, 3, or 2+3 enhancers and response to dFOXO in S2 cells. (C) Activity of 2, 3, or 2+3 enhancers and response to 20E in Kc cells. (D) Activity of 2, 3, or 2+3 enhancers and response to 20E in ΔEcR cells. Consistently, enhancer 3 was stronger than 2. In most cases, when fused together as 2+3, the overall activity was less than the sum of 2 and 3. Upon dFOXO overexpression or 20E treatment, enhancer 2 is activated, while enhancer 3 is repressed; the activity of 2+3 was more similar to 2, indicating that possible repressors acting on enhancer 3 are short-range, and not dominant. Error bars indicate s.d. from three technical replicates.

Analysis of population- and species-level variation

The density of regulatory sequences within the *InR* gene prompted us to look for evidence of conservation, or perhaps generation of novel positively-selected variants. A previous study found evidence for strong purifying selection on protein-coding exons of *InR* in *D. melanogaster* and *D. simulans* (Guirao-Rico and Aguadé, 2009). We surveyed a recently completed analysis of wild-type populations of *D. melanogaster* representing ~340 individuals from 24 populations in east, west and southern Africa, as well as samples from France that were analyzed by whole genome sequencing (Lack et al., 2015). In considering the frequency of single nucleotide polymorphisms (SNP) across a 150 kbp region incorporating the *InR* locus, we note that the Tajima's D score was approximately -2, theoretically an indication for purifying selection. However, this value was not substantially different from that for the entire 1 Mbp region including *InR*, therefore this value of -2 may indicate population bottlenecks rather than strong selection against mutations in enhancers (Figure 3-20A, Figure 3-21) (Nielsen, 2005). Considering insertion/deletion patterns, the protein-coding portion of *InR* is strongly depleted for indels, as would be expected for an essential gene. There are many indels distributed throughout the introns of the gene, including within regulatory elements that we describe above (Figure 3-20A). Indel frequency within the *InR* introns is similar to that of the entire 150 kbp locus (Figure 3-21). Some of these indels found within *InR* regulatory regions may have functional consequences. For example, in enhancer 2, we found a 6 bp deletion in 13% of the surveyed genomes; this deletion overlaps a TG repeat that is of functional importance when tested in reporter assays (Figure 3-15B, Figure 3-22). This specific deletion appears to be distributed throughout a range of

sampled populations. Other indels within *InR* are considerably larger, up to 45 bp, which are even more likely to impact functional binding sites (Figure 3-20A). The frequency of larger indels (>10 bp) is in some cases >50%; the distribution of some of these variations is skewed toward specific geographic regions, possibly a result of selection for specific *InR* regulatory properties.

Direct observation of diverse metazoan genomes confirms that the general structures of intron-rich *InR* genes are conserved, however conservation of putative regulatory regions is often difficult to discern, as illustrated in Figure 3-20A. To test for conservation within the functional elements we identified within the *InR* introns, we measured conservation in the locus using pairwise alignments from the UCSC genome browser between *D. melanogaster* and other species in the Drosophilidae family (Rosenbloom et al., 2015). *D. simulans* is ~1.4 million years diverged, while *D. grimshawi* is 40 million years diverged (Obbard et al., 2012). We measured the average BLASTZ score per 100 bp for regions of interest (Figure 3-20B). The degree to which regulatory sequences are conserved is quite variable between regions and between species, especially at greater evolutionary distances. Neither genome-wide survey of chromatin marks as represented in Figure 3-3 nor the overall levels of nucleotide conservation reveal the functional properties of this locus, emphasizing the necessity for fine-scale functional measurements as carried out here.

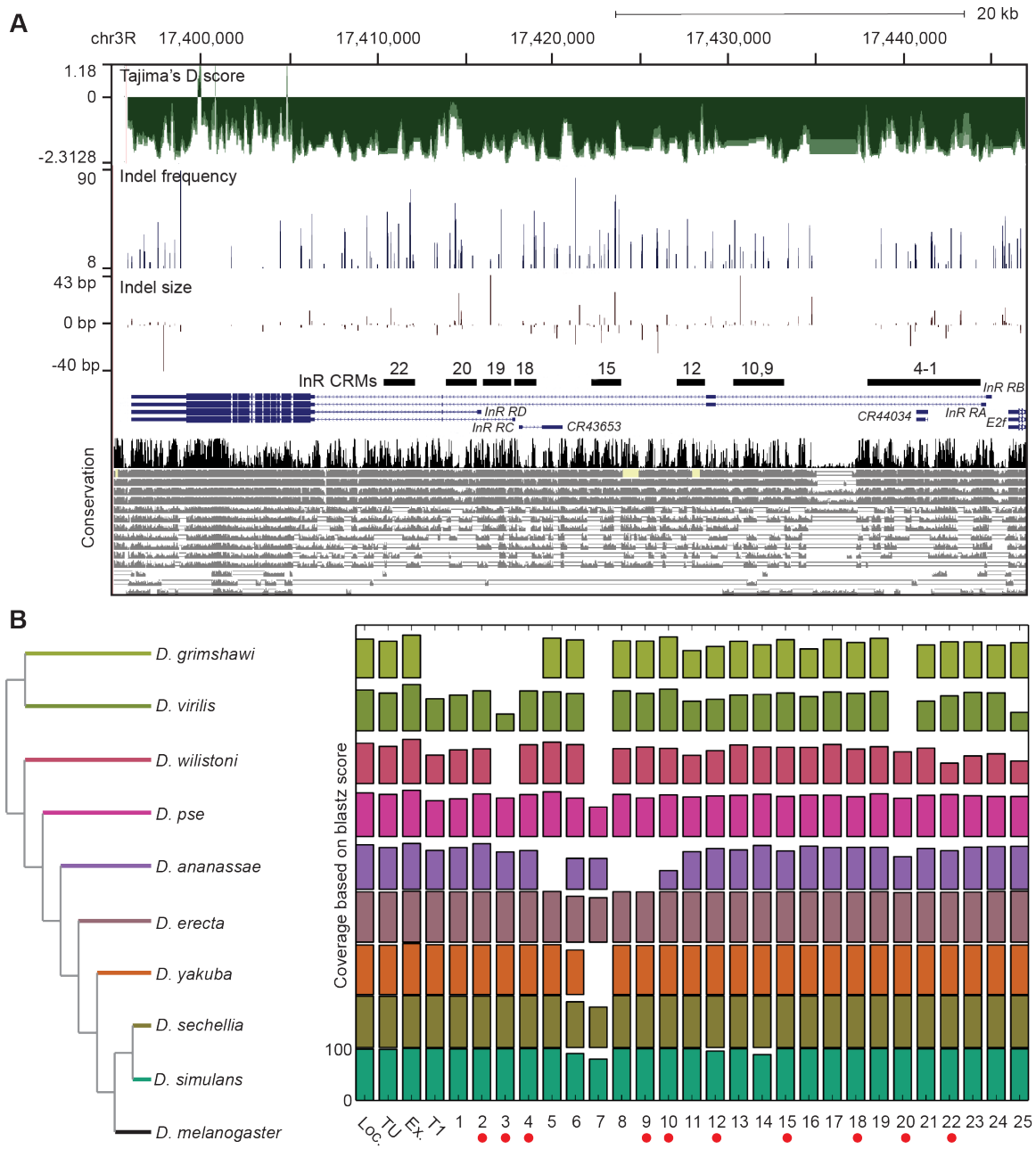


Figure 3-20: Genetic variation at *InR* locus across different species and among different populations. (A) Indel/SNP frequencies at the *InR* locus across different *D. melanogaster* populations. 340 genomes from 24 African *D. melanogaster* populations were analyzed. Tajima's D score indicates proportion of high and low frequency SNPs. Low frequency indels, present in less than 30 genomes (8% frequency), are not listed.

Figure 3-20 (con't)

The *InR* intron fragments that are either constitutively active or respond to dFOXO overexpression or 20E treatment are indicated with bars above the *InR* gene structure. Conservation at the *InR* locus across different *Drosophila* and other insect species are shown below the gene structure. The species include *D.simulans*, *D.sechellia*, *D.yakuba*, *D.erecta*, *D.ananassae*, *D.pseudoobscura*, *D.persimilis*, *D.willistoni*, *D.virilis*, *D.mojavensis*, *D.grimshawi*, *A.gambiae*, *A.mellifera*, and *T.castaneum* (from top to bottom). Species conservation is obtained from USCS genome browser. (B) Sequence conservation indicated by BLASTZ scores at individual intron fragment across different *Drosophila* species, plotted on a $\log_{10}$ scale. Fragments that are either constitutively active or respond to dFOXO overexpression or 20E treatment are indicated with red dot. (Loc., the *InR* locus; TU, the *InR* transcription unit; Ex., the largest *InR* exon, containing protein coding region).

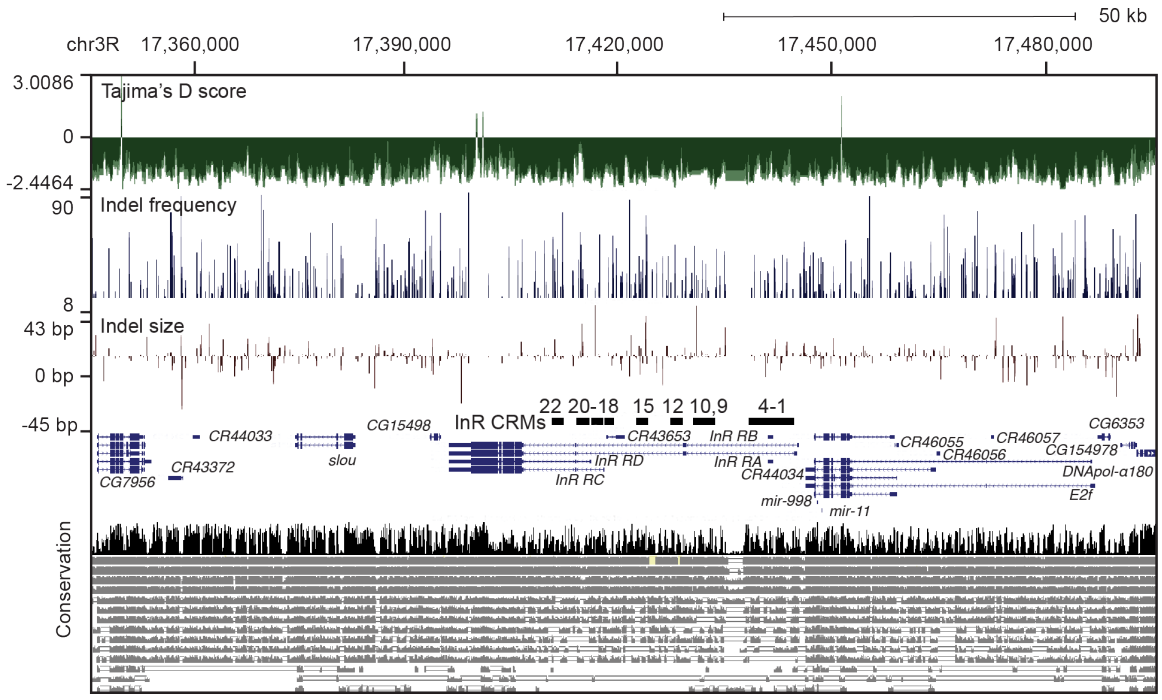


Figure 3-21: Population variation and species conservation at the *InR* gene locus as shown in Figure 3-20 with larger genome area. Note indel depletion at the long introns of *InR*, site of protein-coding sequences. The small gap in aligned genomes at bottom between enhancer regions 4 and 9 correlates to repeat sequences that are not uniquely mappable.

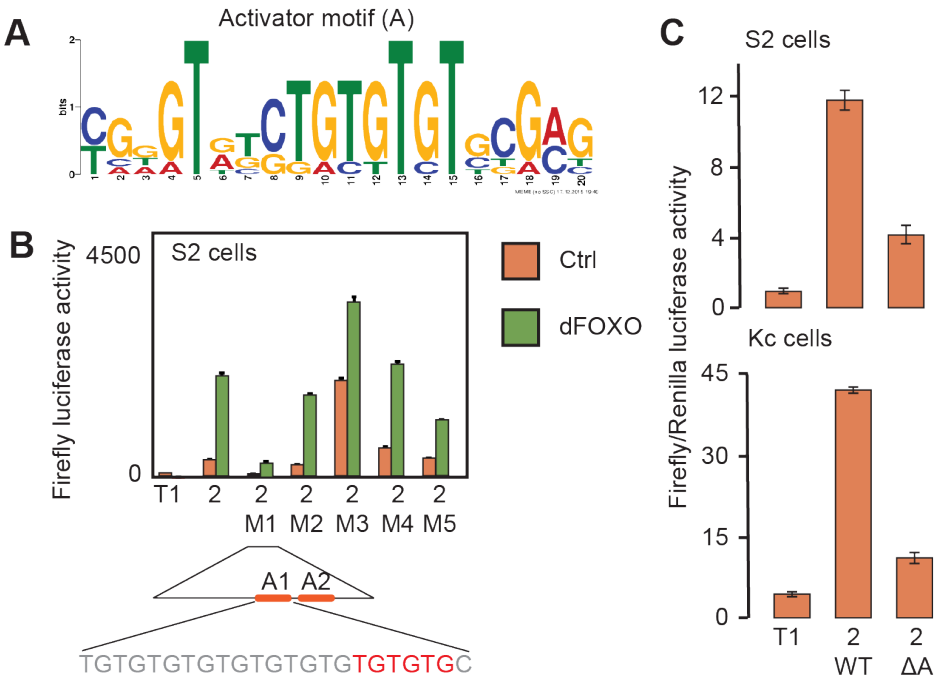


Figure 3-22: Sequence variation found within the *InR* gene in natural *Drosophila* populations lies within GT element important for *InR* enhancer activity. (A) A putative activator motif (A) containing GT repeats was identified by searching constitutively active *InR* enhancer regions with the *de novo* motif finding program MEME (Bailey et al., 2009). (B) Two activator motifs (A1 and A2) were found within the M1 region of fragment 2. Remove of M1 disrupts the enhancer activity as shown in Fig. 7. One deletion identified from African *D. melanogaster* populations (frequency 13%) is found in the A1 motif (indicated by red letters). (C) Removal of A1 and A2 motifs from fragment 2 disrupts enhancer activity in both S2 and Kc cells. Error bars indicate s.d. from three technical replicates.

Discussion

Feedback regulation by FOXO transcription factors has been demonstrated to play a key role in controlling expression of *InR* (Kramer et al., 2003; Jünger et al., 2003; Puig et al., 2003; Puig and Tjian, 2005; Casas-Tinto et al., 2007; Gershman et al., 2007). This aspect of transcriptional regulation of *InR* has been studied in molecular detail; based on genetic perturbation studies and transcriptional reporter assays with short segments of the *InR* gene, a previous model suggested that dFOXO directly binds and activates one of several insulin receptor gene basal promoters (Puig et al., 2003; Puig and Tjian, 2005; Casas-Tinto et al., 2007). Our study indicates that dFOXO regulation appears to be far more complex than suggested by this earlier model. By extensively surveying the ~40 kbp insulin receptor gene introns, we confirmed the direct, if modest, activation role for dFOXO on the internal promoter, but we also found that dFOXO activates or represses at least a half dozen additional enhancers located within introns of the *InR* gene (Figure 3-6B, Figure 3-23A). The majority of this regulation appears to rely on transcriptional intermediates and multiple regulatory layers, forming a complex regulatory circuit (Figure 3-23A). How common is such concerted direct and indirect regulation by dFOXO? Hundreds of genes are suggested to be direct targets of dFOXO regulation in *Drosophila*, however most of them have not been investigated further for transcriptional regulation (Alic et al., 2011; Bai et al., 2013). In some cases, genes such as *Thor/4EBP* with small promoter regions, direct activation by dFOXO may represent the bulk of the regulation. However, other genes appear to be subject to so-called incoherent feed-forward regulation, in which a factor confers both positive and negative effects. The *RpL24-like* ribosomal protein gene promoter is directly repressed by dFOXO, and is

activated by the transcription factor dMyc, which is in turn activated by dFOXO, establishing a two-layer regulation of this ribosomal protein gene (Teleman et al., 2008; Alic et al., 2011; Herter et al., 2015). The compact promoter of *RpL24-like* probably does not approach the complexity of regulation seen with *InR*, which may reflect the importance of fine control of the receptor gene at the apex of this signaling cascade. Thus, it remains to be established how often dFOXO target genes are regulated via multiple enhancers through complex direct and indirect paths, but simple direct activation may represent only one class of important FOXO effects.

Why does the *InR* gene contain so many *cis* regulatory elements, when the mRNA is characterized by widespread expression and relatively less variation in transcript levels than that observed for many developmentally expressed genes? Indeed, "housekeeping" genes such as ribosomal protein genes have relatively compact structure, yet are capable of exhibiting significant regulation in response to environmental signals (Teleman et al., 2008). A simple promoter such as those driving this class of gene may provide the high levels of activity needed for very abundant transcripts, but may lack the ability to achieve precise control from distinct regulatory inputs and to buffer perturbations. The regulatory elements of *InR* appear to be tuned to maintain moderate responses to signals. For example, both the *InR* gene and *E74/E75* genes are regulated by 20E and the ecdysone receptor (Gauhar et al., 2009; Bernardo et al., 2014). The *InR* gene contains elements that either activated or repressed by 20E, while the *E74*/*75* genes contains multiple copies of 20E activator elements. Upon exposure to 20E, *E74*/*75* levels increase dramatically, while *InR* levels increase much more modestly (Figure 3-18) (Bernardo et al., 2014; Mirth et al., 2014). The incoherent feed-forward properties of the *InR* gene

may ensure more precise changes in gene expression, preventing pleiotropic impacts on the downstream signaling pathway. Similarly, the co-existence of dFOXO activated and repressed enhancers may allow dFOXO to achieve precise temporal and spatial control of *InR*. In one model, the incoherent signaling may enable a temporally complex expression pattern, whereby the direct action of dFOXO may first transiently upregulate *InR* gene expression, followed by a delayed turn-down due to the indirectly repressed enhancers (Figure 3-23B). In addition, the multiple layers of regulation by dFOXO may also provide tissue-specific regulation, whereby certain enhancers have a dominant role in different cellular contexts (Figure 3-23C). The physical occupancy of promoters of many genes in the insulin signaling pathway by dFOXO, EcR and Rbf1 hints at further complexity in fine-tuning transcript levels of this signaling pathway (Gauhar et al., 2009; Alic et al., 2011; Acharya et al., 2012). Although we have used a reductionist approach to identify key elements in the control of *InR* expression, it is very likely that the action of these enhancers combined is nonlinear, so that placement and combinatorial interaction influences the overall output. Such complex enhancer dynamics have been observed for regulation of transcription factor and signaling molecules in *Drosophila* and in mammals (Marinić et al., 2013; Bothma et al., 2015). The complexity of *InR* regulation is suggested by the large size of this intron-rich gene, which spans nearly 50 kbp in *Drosophila* and ~200 kbp in mammals. In addition to direct identification of numerous regulatory elements in our fine-structure/function analysis, the dense population of suggestive chromatin modifications within introns of the mammalian insulin receptor genes supports the idea that this locus is a "superenhancer", defined as a broad region of regulatory DNA on which histone

modifications consistent with enhancers are widely deposited (Andersson et al., 2014; Pasquali et al., 2014; Arner et al., 2015; Roadmap Epigenomics Consortium et al., 2015; Vanhille et al., 2015; Wei et al., 2016).

The deep analysis of insulin receptor gene *cis* regulation is critical for molecular interpretation of disease states, for misregulation of this gene has been reported in cancer, type II diabetes, and Alzheimer's disease (Gunton et al., 2005; Freude et al., 2009; Belfiore and Malaguarnera, 2011). This change in expression is likely to have functional consequences, because the increase in cancer risk for individuals with type II diabetes may involve the expression of the insulin receptor in cancer cells. In human pancreatic islet cells, candidate *INSR* enhancers were identified based on chromatin marks and transcription factor binding sites, some of which overlap with sequence variants associated with type II diabetes and Alzheimer's disease (Pasquali et al., 2014). In a *Drosophila* cancer model system, a high sugar diet facilitates tumorigenesis, whereby the *InR* gene was upregulated via the Wnt signaling pathway in the tumors (Hirabayashi et al., 2013). With regards to another well-studied cancer pathway, the importance of retinoblastoma protein in directly regulating *InR* in such settings is still poorly understood, but it may represent a pathway by which cell cycle and signaling are coordinately misregulated. At the same time, it appears that Rbf1 regulation of the *InR* promoter is independent of the C-terminal degron that has recently been shown to impact repression on cell cycle promoters, suggesting that there may be differential regulation of *InR* and cell cycle genes (Raj et al., 2012). Taken together with the recent finding that retinoblastoma protein may also play a role in regulating ribosomal protein genes, linked to growth regulation, it appears that retinoblastoma proteins may play key

roles in cell cycle, signaling and growth control through direct transcriptional regulation of such genes (Wei et al., 2015).

We note that there is evidence of extensive population-level variation throughout the *Drosophila InR* locus, with only the protein coding portion showing a depletion in indels. The identified intronic regulatory elements are not similarly depleted for indels or SNP. This variation may reflect the great degree of sequence plasticity observed in many enhancers, even as function is conserved, consistent with a "billboard" model of enhancer structure (Arnosti and Kulkarni, 2005). In addition, the apparent redundancy in regulatory regions found in this locus may enable robustness of output, allowing more variation in the functional output of individual elements. However, some of the variation observed at a population and species level may be indicative of meaningful, functional changes in *InR* expression that may lead to significant phenotypic differences. Insulin signaling itself is subject to extensive evolutionary sculpting, influencing the nutritionally-driven body allometry in *Drosophila* and the baroque regulation of the male weapon of the rhinoceros beetle (Shingleton et al., 2005; Emlen et al., 2012). A naturally-occurring indel affecting InR protein structure was previously described to correlate with body size and stress tolerance associated with population clines in *Drosophila* (Paaby et al., 2014). Our fine-structure map of *InR* regulatory regions will for the first time allow for interpretation of population variation within the large noncoding portions of the gene.

Some of the best-characterized complex metazoan regulatory systems involve transcriptional enhancers controlling the expression of cell-type specific transcription factors and secreted ligands (Fujioka et al., 1999; Marinić et al., 2013). Widely

expressed genes have in some studies been considered to have simpler transcriptional regulatory potential, being classified as “housekeeping” promoters. However, it is clear that most genes, whether universally expressed or limited to very specific spatial and temporal patterns, show dynamic levels of expression, and we have argued that a simple “housekeeping/developmental” dichotomy fails to capture the complexity, and importance of dynamic regulation, including for genes not previously considered to have important transcriptional regulation (Wei and Arnosti, 2015; Payankaulam et al., 2016). Across the spectrum of transcriptional control, *InR* may represent one of the more elaborately regulated wide-expressed genes, because of its pleiotropic nature at the apex of a signaling cascade. Having extensively characterized the *cis*-regulatory landscape of this gene, it is clear that a major further objective will be a complete identification of transcription factors and *cis* elements contributing to regulation. Our fine-mapping the *cis* regulatory elements, will permit construction of computational models to demonstrate the importance of *cis* regulatory variation within this locus for insulin receptor gene regulation in the context of development, disease and evolution. Many genome-wide studies utilize chromatin marks as proxies for active enhancers, however, our analysis of the *InR* locus suggests that chromatin marks alone are insufficient for characterization of *cis*-regulatory elements. In fact, even dynamic chromatin features can sometimes reflect off-target effects of transcription factors, rather than functional interactions (Kok et al., 2015). Thus for specific disease-relevant genes such as *InR*, a detailed functional analysis is still critical to obtain the level of understanding necessary to understand the impact of regulatory sequence variation and signaling events associated with disease and evolution.

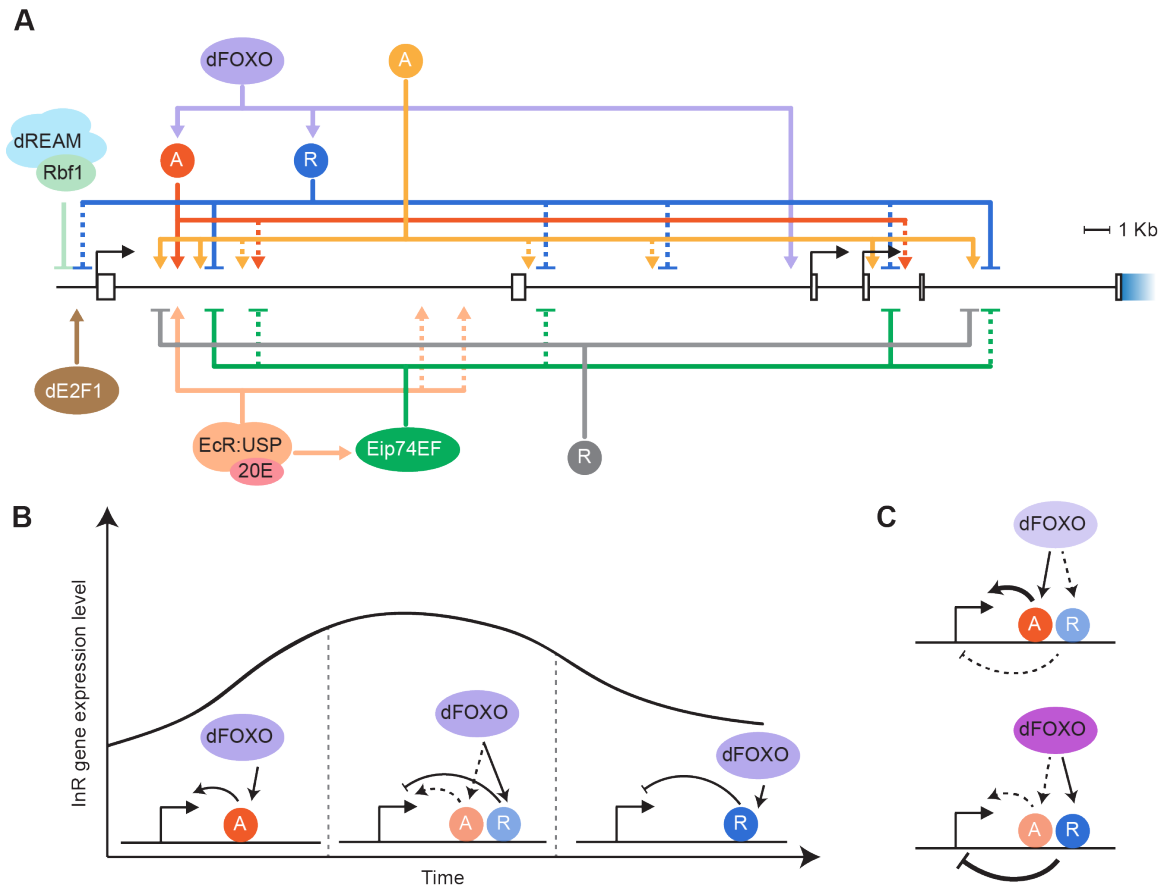


Figure 3-23: Transcription circuitry indicates complex controls of the *InR* gene for temporal or spatial regulation. (A) Transcriptional circuitry of the *InR* gene. Regulatory inputs relating to dFOXO, ecdysone, dE2F1, Rbf1 and additional activators and repressors are indicated. Arrows indicate activation, lines repression. Broken lines indicate elements that are active in a cell-type specific manner. The dREAM complex proteins are associated with Rbf1 at the promoter, but their functional relevance for *InR* expression is unknown (Acharya et al., 2012; Wei et al., 2015). Uncharacterized activator(s) indicated by A, and repressor(s) by R. Eip74EF-mediated repression role is speculative. (B) Model for temporal control of the *InR* gene expression by dFOXO, explaining activating and repressing role of this transcription factor in regulation. In this

Figure 3-23 (con't)

model, dFOXO first activates the *InR* gene through direct binding of the gene and by inducing the expression of transcription activators that bind *InR*. Through indirect means, in a later phase, dFOXO induces expression of transcription repressors to reduce gene expression. By controlling both activator and repressor, dFOXO precisely regulates the *InR* gene enhancers to achieve a fine-tuned expression. (C) Model for spatial control of the *InR* gene expression by dFOXO. In this model, dFOXO activates the *InR* gene in one tissue by primarily stimulating expression of *InR*-binding activators, while in another cell type, the predominant expression of dFOXO-regulated repressors allows for reduction in *InR* expression. Both processes may be partially active to achieve an intermediate level of expression, allowing different equilibrium expression levels to be reached.

Materials and Methods

Fly strains and reporter analysis

The following fly strains were obtained from the Bloomington Stock Center (stock numbers in parentheses): pBac{attp-3B}VK00001 (9722), InR^{GC25}(9554), InR^{E19}(9646), Putative *InR* enhancers: GMR27H05-Gal4(47519), GMR28A11-Gal4(45164), GMR28D03-Gal4(47521), GMR28E02-Gal4(49458), GMR38E09-Gal4(48080), GMR28G01-Gal4 (45547), GMR28G04-Gal4(45548), GMR28H01-Gal4(45947), GMR29A02-Gal4(45175), GMR37B05-Gal4(47564), UAS-GFP (1521). Each putative enhancer line was crossed to the UAS-GFP line. Larval tissues and adult flies were imaged on an Olympus BX-41 microscope.

Generation of transgenic flies

BAC construct CH321-24D17 containing the entire *InR* locus was obtained from the BacPac Resources Center (Oakland, CA). BACs were grown overnight for 16-20h and high copy number was induced using Epicentre BAC autoinduction solution (Illumina). DNA was prepared using the HiPure Midiprep kit following the manufacturer's instructions (Invitrogen). DNA was diluted to a final concentration of ~1µg μL^{-1} and 400 embryos were injected by Rainbow Transgenics Inc (Camarillo, CA). Landing site line VK00001, containing an *attp* site at location 59D3 on chromosome 2, was used for injection and integration of the BAC.

qPCR analysis of mRNA from transgenic flies

Three-day old adult males and virgin females were collected and flash frozen at -80^{0}C. Total RNA was extracted using Trizol Reagent (Invitrogen) and subjected to DNaseI treatment (Ambion DNaseI Kit) at room temperature for 15 minutes to eliminate genomic DNA contamination. Reverse transcription for first strand synthesis was carried out using random primers and Multiscribe Reverse Transcriptase (ABI Biosystems). Real-time PCR was performed using POWER SYBR Green Master Mix (ABI Biosystems) and analyzed on Eppendorf Mastercycler Realplex. Gene expression was assayed in 3-5 biological replicates of 8-10 flies each and normalized against expression of 28S rRNA. Standard curves were generated using six serial dilutions of total RNA extracted from two individuals of Samarkand wild-type first, second and third instar larvae, pupae (male) and adult flies (male). Gene expression fold changes were calculated by normalizing to WT.

Genome data visualization

STARR-seq and DHS-seq data were obtained from NCBI GEO database (GSE40739, GSE47691) (Arnold et al., 2013; Shlyueva et al., 2014). To visualize the peak files, bedgraph files were generated by expanding the peaks 300 bp downstream and 300 bp upstream. The heights of the peaks indicate peak scores as shown in the original peak files. For Figure 3-4, the height was set as local maximum. For Figure 3-12, the heights of peaks were fixed at maximum of no-20E samples. Regions identified in the two replicate studies are displayed on Figure 3-4, 3-6, and 3-12; some regions were found in only one of the two experiments. Reproducibly identified peaks are indicated by the

darker color, indicating overlap between two replicates. FAIRE data were obtained from NCBI GEO database (GSE38727) (McKay and Lieb, 2013). The WIG files for peaks were visualized on USCS genome browser with peak setting as intensity, and height setting as local maximum. ChIP-seq for H3K27Ac, H3K4Me1, H2KMe3, p300, and RNA polymerase II were obtained from modENCODE. Data were visualized with built-in genome browser on modENCODE with peak setting as density, and peak height was set as local maximum. The genome version used in this study is dm3/R5.

Luciferase reporter library

The *InR* T1, T2, and T3 basal promoter elements, as well as a non-promoter control region (PT) in the first intron of *InR* were cloned into *Asc*I and *Sal*I sites of the p2T-Luc vector (Ryu and Arnosti, 2003). To generate the luciferase reporter library, the *InR* intron regions were divided into 25 ~1.5 kbp fragments. Each fragment was cloned into the upstream of T1 promoter in the p2T-Luc vector, using the *Kpn*I and *Asc*I sites. To generate the serial deletions on selected reporters, primers creating specific deletions were designed. The mutagenesis cloning was done as previously described (Zhang et al., 2014). To generate the expression vectors for dFOXO, EcR, and USP, the cDNA of these genes were cloned from cDNA vectors (obtained from Drosophila Genomic Resource Center, LD19191/dFOXO, LD09973/USP, RE33854/EcR). The cDNA was then cloned into *Kpn*I and *Xba*I sites of pAX-Flag vector as described before (Zhang et al., 2014).

Cell culture and transfection

Drosophila S2 cells, Kc cells (Kc167) and ΔEcR cells (derived from Kc cells, obtained from Drosophila Genomic Resource Center, ID: L57-3-11) were cultured in Schneider

medium (Gibco) supplied with 10% FBS (Gibco) and penicillin-streptomycin (100 Unit ml^{-1} penicillin, 100 ug ml^{-1} streptomycin, Gibco). For dFOXO overexpression, 100 ng luciferase reporter was co-transfected with 200 ng pAX-dFOXO-Flag vector and 250 ng CMV-Renilla luciferase reporter (Promega). As a control, the reporter was co-transfected with 200 ng empty pAX vector. The vectors were transfected into 1.5 million S2 or Kc cells in 6-well plates using Transfectene reagent (QIAGEN). For luciferase assays, the cells were harvested three days after transfection, cells were pelleted and resuspended in 300 μl PBS solution (Sigma). 75 μl of re-suspended cells was used for each luciferase assay, and three technical replicates were performed. The luciferase assay was performed using Promega Dual-Glo luciferase reagent with luminometer (Veritas) following the manufacturer's instructions. For 20E treatment, 100 ng of the luciferase reporter was co-transfected with 200 ng pBluescript (pBS) vector (an empty vector to make up the total transfected DNA amount to 550 ng) and 250 ng CMV-Renilla luciferase reporter in S2, Kc or ΔEcR cells as described above. Twenty-four hours after transfection, the cells were treated with 20E (Sigma) at the final concentration of 10^{-5} M in ethanol or ethanol alone as control. Luciferase assays were conducted 24 hours after 20E treatment as described above. For EcR/USP overexpression, 100 ng luciferase reporter was co-transfected with 200 ng pAX-EcR and 200 ng pAX-USP, and 250 ng CMV-Renilla luciferase reporter in ΔEcR cells. As a control, the luciferase reporter was co-transfected with 400 ng empty pAX vector. The 20E treatment was performed 24 hours after transfection, and luciferase assay was conducted 24 hours after 20E treatment as described above. Because dFOXO overexpression affected CVM-Renilla luciferase activity, data for dFOXO activity were

not normalized to CMV-Renilla luciferase, and were presented as the firefly luciferase activity. For the 20E treatment, CMV-Renilla luciferase activity was not affected, and data were normalized to CMV-Renilla luciferase readings. To compare the WT T1 promoter activity with mutant T1 promoter lacking the Rbf1 binding site (as shown in Figure 3-13B), 10 ng of the WT and mutant reporters with 250 ng of CMV-Renilla luciferase reporter were transfected into S2 cells. The luciferase readings were collected three days after transfection.

FOXO ChIP-qPCR

ChIP-qPCR assays in cell culture were conducted as described before (Carey et al., 2009) using 100 million cells. dFOXO antibody (polyclonal rabbit antiserum 524.4 and 524.5) was a gift from Dr. Carla Margulies (Ludwig Maximilian University of Munich, Germany). 5 μl of anti-dFOXO serum and non-immune serum was used for each immunoprecipitation. ChIP DNA was purified using PCR purification Kit (QIAGEN). qPCR was conducted using SYBR-Green reagent (Quanta) on a ABI 7500 machine. For starvation treatment, Kc cells were grown in Schneider's medium without FBS for 48 hours. For insulin treatment, Kc cells were grown in Schneider's medium supplied with 10% FBS, and then treated with human insulin solution (Sigma) at the final concentration of 0.013 mg ml^{-1} for 30 minutes or 2 hours.

SNP analysis of population variation

Genome assemblies for the DPGP2 and DPGP3 were downloaded in .seq format from the Drosophila Genome Nexus at http://johnpool.net/genomes.html (Lack et al., 2015). Fasta sequences for specific regions were extracted using a custom python script using

Python 2.7, available at https://github.com/arnosti-lab/InR/tree/master/Population_analysis. Tagima's D scores were calculated using DnaSP version 5 (Librado and Rozas, 2009), and converted into .bedgraph format. Intermediate files are available on the github page cited above. Bedgraph files were visualized using the UCSC Genome Browser (Kent et al., 2002).

Indel analysis of population variation

Indel files were downloaded as .vcf files from the Drosophila Genome Nexus. The 'round 1' batch of indels were used for further analysis. The VCF files were converted into a simplified format using a unix command available on the github page described above. Specific alleles that occurred in at least 30 of the 349 DPGP genomes were extracted and converted to bedgraph format using a custom python script, also available on the github page described above. Bedgraph files were visualized using the UCSC Genome Browser.

Analysis of sequence conservation between species

Pairwise alignments were downloaded in axt file format from the University of California Santa-Cruz Genome Browser. Alignments used were *D. melanogaster* version dm3 to *D. grimshawi* droGri2, *D. willistoni* droWil1, *D. ananasse* droAna3, *D. pseudoobscura* dp4, *D. erecta* droEre2, *D. yakuba* droYak2, *D. sechellia* droSec1, and *D. simulans* droSim1. These were obtained from http://hgdownload.soe.ucsc.edu/downloads.html#fruitfly. Summary lines from each chromosome axt file were used to create a single summary file for each genome. A custom python script was used to determine the average BLASTZ score per 100 base

pairs for a region, which was plotted by species on a $\log_{10}$ scale Perfect conservation over 100 base pairs would yield a score near 10,000. Regions used included the 100kb loci around the gene *InR*, the transcriptional unit of the gene, and the largest exons for the gene, and each of the tested fragments within *InR*. This python script, summary command and summary files are available at https://github.com/arnosti-lab/InR/tree/master/Interspecies_analysis.

REFERENCES

REFERENCES

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

Alic, N., Andrews, T. D., Giannakou, M. E., Papatheodorou, I., Slack, C., Hoddinott, M. P., Cocheme, H. M., Schuster, E. F., Thornton, J. M. and Partridge, L. (2011). Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. *Mol. Syst. Biol.* **7**, 502.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077.

Arnosti, D. N. and Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* **94**, 890–898.

Attrill, H., Falls, K., Goodman, J. L., Millburn, G. H., Antonazzo, G., Rey, A. J. and Marygold, S. J. (2016). FlyBase: establishing a Gene Group resource for Drosophila melanogaster. *Nucleic Acids Res.* **44**, D786–92.

Bai, H., Kang, P., Hernandez, A. M. and Tatar, M. (2013). Activin signaling targeted by insulin/dFOXO regulates aging and muscle proteostasis in Drosophila. *PLoS Genet.* **9**, e1003941.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308.

Belfiore, A. and Malaguarnera, R. (2011). Insulin receptor and cancer. *Endocr. Relat. Cancer* **18**, R125–47.

Belfiore, A., Frasca, F., Pandini, G., Sciacca, L. and Vigneri, R. (2009). Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocr. Rev.* **30**, 586–623.

Bernardo, T. J., Dubrovskaya, V. A., Xie, X. and Dubrovsky, E. B. (2014). A view through a chromatin loop: insights into the ecdysone activation of early genes in Drosophila. *Nucleic Acids Res.* **42**, 10409–10424.

Bertrand, L., Horman, S., Beauloye, C. and Vanoverschelde, J.-L. (2008). Insulin signalling in the heart. *Cardiovasc. Res.* **79**, 238–248.

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T. and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* **4**,.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R. and Hafen, E. (2001). An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–221.

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399.

Brunetti, A., Manfioletti, G., Chiefari, E., Goldfine, I. D. and Foti, D. (2001). Transcriptional regulation of human insulin receptor gene by the high-mobility group protein HMGI(Y). *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **15**, 492–500.

Calle, C., Maestro, B. and Garcia-Arencibia, M. (2008). Genomic actions of 1,25-dihydroxyvitamin D3 on insulin receptor gene expression, insulin receptor number and insulin activity in the kidney, liver and adipose tissue of streptozotocin-induced diabetic rats. *BMC Mol. Biol.* **9**, 65.

Cameron, K. E., Resnik, J. and Webster, N. J. (1992). Transcriptional regulation of the human insulin receptor promoter. *J. Biol. Chem.* **267**, 17375–17383.

Carey, M. F., Peterson, C. L. and Smale, S. T. (2009). Chromatin immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.* **2009**, pdb.prot5279.

Casas-Tinto, S., Marr, M. T. 2nd, Andreu, P. and Puig, O. (2007). Characterization of the Drosophila insulin receptor promoter. *Biochim. Biophys. Acta* **1769**, 236–243.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R. a, Narita, M., Zhang, M. and Lowe, S. W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–87.

Colombani, J., Bianchini, L., Layalle, S., Pondeville, E., Dauphin-Villemant, C., Antoniewski, C., Carre, C., Noselli, S. and Leopold, P. (2005). Antagonistic actions of ecdysone and insulins determine final size in Drosophila. *Science* **310**, 667–670.

Du, W. and Pogoriler, J. (2006). Retinoblastoma family genes. *Oncogene* **25**, 5190–200.

Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J. H., Masiarz, F., Kan, Y. W. and Goldfine, I. D. (1985). The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling. *Cell* **40**, 747–758.

Emlen, D. J., Warren, I. A., Johns, A., Dworkin, I. and Lavine, L. C. (2012). A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. *Science* **337**, 860–864.

Fernandez, R., Tabarini, D., Azpiazu, N., Frasch, M. and Schlessinger, J. (1995). The Drosophila insulin receptor homolog: a gene essential for embryonic

development encodes two receptor isoforms with different signaling potential. *EMBO J.* **14**, 3373–3384.

Frasca, F., Pandini, G., Scalia, P., Sciacca, L., Mineo, R., Costantino, A., Goldfine, I. D., Belfiore, A. and Vigneri, R. (1999). Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Mol. Cell. Biol.* **19**, 3278–3288.

Freude, S., Schilbach, K. and Schubert, M. (2009). The role of IGF-1 receptor and insulin receptor signaling for the pathogenesis of Alzheimer's disease: from model organisms to human disease. *Curr. Alzheimer Res.* **6**, 213–223.

Frolich, L., Blum-Degen, D., Bernstein, H. G., Engelsberger, S., Humrich, J., Laufer, S., Muschner, D., Thalheimer, A., Turk, A., Hoyer, S., et al. (1998). Brain insulin and insulin receptors in aging and sporadic Alzheimer's disease. *J. Neural Transm.* **105**, 423–438.

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. and Jaynes, J. B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* **126**, 2527–2538.

Garcia-Arencibia, M., Davila, N., Campion, J., Carmen Carranza, M. and Calle, C. (2005). Identification of two functional estrogen response elements complexed with AP-1-like sites in the human insulin receptor gene promoter. *J. Steroid Biochem. Mol. Biol.* **94**, 1–14.

Garofalo, R. S. and Rosen, O. M. (1988). Tissue localization of Drosophila melanogaster insulin receptor transcripts during development. *Mol. Cell. Biol.* **8**, 1638–1647.

Gauhar, Z., Sun, L. V, Hua, S., Mason, C. E., Fuchs, F., Li, T.-R., Boutros, M. and White, K. P. (2009). Genomic mapping of binding regions for the Ecdysone receptor protein complex. *Genome Res.* **19**, 1006–1013.

Georlette, D., Ahn, S., MacAlpine, D. M., Cheung, E., Lewis, P. W., Beall, E. L., Bell, S. P., Speed, T., Manak, J. R. and Botchan, M. R. (2007). Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. *Genes Dev.* **21**, 2880–96.

Gershman, B., Puig, O., Hang, L., Peitzsch, R. M., Tatar, M. and Garofalo, R. S. (2007). High-resolution dynamics of the transcriptional response to nutrition in Drosophila: a key role for dFOXO. *Physiol. Genomics* **29**, 24–34.

Giacinti, C. and Giordano, a (2006). RB and cell cycle progression. *Oncogene* **25**, 5220–5227.

Goldstein, B. J., Muller-Wieland, D. and Kahn, C. R. (1987). Variation in insulin receptor messenger ribonucleic acid expression in human and rodent tissues. *Mol. Endocrinol.* **1**, 759–766.

Gonzalez, E., Flier, E., Molle, D., Accili, D. and McGraw, T. E. (2011). Hyperinsulinemia leads to uncoupled insulin regulation of the GLUT4 glucose transporter and the FoxO1 transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10162–10167.

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., van Baren, M. J., Boley, N., Booth, B. W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473–479.

Guirao-Rico, S. and Aguade, M. (2009). Positive selection has driven the evolution of the Drosophila insulin-like receptor (InR) at different timescales. *Mol. Biol. Evol.* **26**, 1723–1732.

Gunton, J. E., Kulkarni, R. N., Yim, S., Okada, T., Hawthorne, W. J., Tseng, Y.-H., Roberson, R. S., Ricordi, C., O'Connell, P. J., Gonzalez, F. J., et al. (2005). Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122**, 337–349.

Herboso, L., Oliveira, M. M., Talamillo, A., Perez, C., Gonzalez, M., Martin, D., Sutherland, J. D., Shingleton, A. W., Mirth, C. K. and Barrio, R. (2015). Ecdysone promotes growth of imaginal discs through the regulation of Thor in D. melanogaster. *Sci. Rep.* **5**, 12383.

Herter, E. K., Stauch, M., Gallant, M., Wolf, E., Raabe, T. and Gallant, P. (2015). snoRNAs are a novel class of biologically relevant Myc targets. *BMC Biol.* **13**, 25.

Hirabayashi, S., Baranski, T. J. and Cagan, R. L. (2013). Transformed Drosophila cells evade diet-mediated insulin resistance through wingless signaling. *Cell* **154**, 664–675.

Hu, X., Cherbas, L. and Cherbas, P. (2003). Transcription activation by the ecdysone receptor (EcR/USP): identification of activation functions. *Mol. Endocrinol.* **17**, 716–731.

Jenett, A., Rubin, G. M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep.* **2**, 991–1001.

Jory, A., Estella, C., Giorgianni, M. W., Slattery, M., Laverty, T. R., Rubin, G. M. and Mann, R. S. (2012). A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of Drosophila melanogaster. *Cell Rep.* **2**, 1014–1024.

Junger, M. A., Rintelen, F., Stocker, H., Wasserman, J. D., Vegh, M., Radimerski, T., Greenberg, M. E. and Hafen, E. (2003). The Drosophila forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signaling. *J. Biol.* **2**, 20.

Kaplan, T., Li, X.-Y., Sabo, P. J., Thomas, S., Stamatoyannopoulos, J. A., Biggin, M. D. and Eisen, M. B. (2011). Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet.* **7**, e1001290.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.

Koelle, M. R., Talbot, W. S., Segraves, W. A., Bender, M. T., Cherbas, P. and Hogness, D. S. (1991). The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67**, 59–77.

Kok, K., Ay, A., Li, L. M. and Arnosti, D. N. (2015). Genome-wide errant targeting by Hairy. *Elife* **4**,.

Korenjak, M., Anderssen, E., Ramaswamy, S., Whetstine, J. R. and Dyson, N. J. (2012). RBF binding to both canonical E2F targets and noncanonical targets depends on functional dE2F/dDP complexes. *Mol. Cell. Biol.* **32**, 4375–87.

Koyama, T., Rodrigues, M. A., Athanasiadis, A., Shingleton, A. W. and Mirth, C. K. (2014). Nutritional control of body size through FoxO-Ultraspiracle mediated ecdysone biosynthesis. *Elife* **3**,.

Kramer, J. M., Davidge, J. T., Lockyer, J. M. and Staveley, B. E. (2003). Expression of Drosophila FOXO regulates growth and can phenocopy starvation. *BMC Dev. Biol.* **3**, 5.

Kriauciunas, K. M., Goldstein, B. J., Lipes, M. A. and Kahn, C. R. (1993). Modulation of expression of insulin and IGF-I receptor by Epstein-Barr virus and its gene products LMP and EBNA-2 in lymphocyte cell lines. *J. Cell. Physiol.* **154**, 486–495.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., Langley, C. H. and Pool, J. E. (2015). The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241.

Leal, M. A., Cabanas, C., Rius, C., Aller, P. and Calle, C. (1992). Modulation by dexamethasone of insulin binding and insulin receptor mRNA levels in U-937 human promonocytic cells. *Biochimie* **74**, 545–549.

Lee, J. K. and Tsai, S. Y. (1994). Multiple hormone response elements can confer glucocorticoid regulation on the human insulin receptor gene. *Mol. Endocrinol.* **8**, 625–634.

Lee, J. K., Tam, J. W., Tsai, M. J. and Tsai, S. Y. (1992). Identification of cis- and trans-acting factors regulating the expression of the human insulin receptor gene. *J. Biol. Chem.* **267**, 4638–4645.

Li, X.-Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. A. and Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol.* **12**, R34.

Li, H.-H., Kroll, J. R., Lennox, S. M., Ogundeyi, O., Jeter, J., Depasquale, G. and Truman, J. W. (2014). A GAL4 driver resource for developmental and behavioral studies on the larval CNS of Drosophila. *Cell Rep.* **8**, 897–908.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Liu, Y., Zhou, S., Ma, L., Tian, L., Wang, S., Sheng, Z., Jiang, R.-J., Bendena, W. G. and Li, S. (2010). Transcriptional regulation of the insulin signaling pathway genes by starvation and 20-hydroxyecdysone in the Bombyx fat body. *J. Insect Physiol.* **56**, 1436–1444.

Manning, L., Heckscher, E. S., Purice, M. D., Roberts, J., Bennett, A. L., Kroll, J. R., Pollard, J. L., Strader, M. E., Lupton, J. R., Dyukareva, A. V, et al. (2012). A resource for manipulating gene expression and analyzing cis-regulatory modules in the Drosophila CNS. *Cell Rep.* **2**, 1002–1013.

Marinic, M., Aktas, T., Ruf, S. and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542.

McKay, D. J. and Lieb, J. D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. *Dev. Cell* **27**, 306–318.

Mirth, C. K., Tang, H. Y., Makohon-Moore, S. C., Salhadar, S., Gokhale, R. H., Warner, R. D., Koyama, T., Riddiford, L. M. and Shingleton, A. W. (2014). Juvenile hormone regulates body size and perturbs insulin signaling in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7018–7023.

Moloney, A. M., Griffin, R. J., Timmons, S., O'Connor, R., Ravid, R. and O'Neill, C. (2010). Defects in IGF-1 receptor, insulin receptor and IRS-1/2 in Alzheimer's disease indicate possible resistance to IGF-1 and insulin signalling. *Neurobiol. Aging* **31**, 224–243.

Negre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527–531.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218.

Obbard, D. J., Maclennan, J., Kim, K.-W., Rambaut, A., O'Grady, P. M. and Jiggins, F. M. (2012). Estimating divergence dates and substitution rates in the Drosophila phylogeny. *Mol. Biol. Evol.* **29**, 3459–3473.

Oberbauer, A. M. (2013). The Regulation of IGF-1 Gene Transcription and Splicing during Development and Aging. *Front. Endocrinol. (Lausanne).* **4**, 39.

Oldham, S. and Hafen, E. (2003). Insulin/IGF and target of rapamycin signaling: a TOR de force in growth control. *Trends Cell Biol.* **13**, 79–85.

Paaby, A. B., Bergland, A. O., Behrman, E. L. and Schmidt, P. S. (2014). A highly pleiotropic amino acid polymorphism in the Drosophila insulin receptor contributes to life-history adaptation. *Evolution* **68**, 3395–3409.

Pasquali, L., Gaulton, K. J., Rodriguez-Segui, S. A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J. J., Moran, I., Gomez-Marin, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143.

Payankaulam, S., Yeung, K., McNeill, H., Henry, R. W. and Arnosti, D. N. (2016). Regulation of cell polarity determinants by the Retinoblastoma tumor suppressor protein. *Sci. Rep.* **6**, 22879.

Petruzzelli, L., Herrera, R., Arenas-Garcia, R., Fernandez, R., Birnbaum, M. J. and Rosen, O. M. (1986). Isolation of a Drosophila genomic sequence homologous to the kinase domain of the human insulin receptor and detection of the phosphorylated Drosophila receptor with an anti-peptide antibody. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 4710–4714.

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., Scully, A., Carlson, J. W., Wan, K. H., Laverty, T. R., et al. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9715–9720.

Puig, O. and Tjian, R. (2005). Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. *Genes Dev.* **19**, 2435–2446.

Puig, O., Marr, M. T., Ruhf, M. L. and Tjian, R. (2003). Control of cell number by Drosophila FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* **17**, 2006–2020.

Raj, N., Zhang, L., Wei, Y., Arnosti, D. N. and Henry, R. W. (2012). Ubiquitination of retinoblastoma family protein 1 potentiates gene-specific repression function. *J. Biol. Chem.* **287**, 41835–41843.

Rajagopal, N., Ernst, J., Ray, P., Wu, J., Zhang, M., Kellis, M. and Ren, B. (2014). Distinct and predictive histone lysine acetylation patterns at promoters, enhancers, and gene bodies. *G3 (Bethesda).* **4**, 2051–2063.

Rechler, M. M. and Nissley, S. P. (1985). The nature and regulation of the receptors for insulin-like growth factors. *Annu. Rev. Physiol.* **47**, 425–442.

Riddiford, L. M., Cherbas, P. and Truman, J. W. (2000). Ecdysone receptors and their biological actions. *Vitam. Horm.* **60**, 1–73.

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–81.

Ryu, J.-R. and Arnosti, D. N. (2003). Functional similarity of Knirps CtBP-dependent and CtBP-independent transcriptional repressor activities. *Nucleic Acids Res.* **31**, 4654–4662.

Sarfstein, R., Maor, S., Reizner, N., Abramovitch, S. and Werner, H. (2006). Transcriptional regulation of the insulin-like growth factor-I receptor gene in breast cancer. *Mol. Cell. Endocrinol.* **252**, 241–246.

Schayek, H., Bentov, I., Rotem, I., Pasmanik-Chor, M., Ginsberg, D., Plymate, S. R. and Werner, H. (2010). Transcription factor E2F1 is a potent transactivator of the insulin-like growth factor-I receptor (IGF-IR) gene. *Growth Horm. IGF Res.* **20**, 68–72.

Shen, W. J., Kim, H. S. and Tsai, S. Y. (1995). Stimulation of human insulin receptor gene expression by retinoblastoma gene product. *J. Biol. Chem.* **270**, 20525–20529.

Shingleton, A. W., Das, J., Vinicius, L. and Stern, D. L. (2005). The temporal requirements for insulin signaling during development in Drosophila. *PLoS Biol.* **3**, e289.

Shlyueva, D., Stelzer, C., Gerlach, D., Yanez-Cuna, J. O., Rath, M., Boryn, L. M., Arnold, C. D. and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol. Cell* **54**, 180–192.

Small, S., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* **119**, 762–772.

Song, J., Wu, L., Chen, Z., Kohanski, R. A. and Pick, L. (2003). Axons guided by insulin receptor in Drosophila visual system. *Science* **300**, 502–505.

Swevers, L., Cherbas, L., Cherbas, P. and Iatrou, K. (1996). Bombyx EcR (BmEcR) and Bombyx USP (BmCF1) combine to form a functional ecdysone receptor. *Insect Biochem. Mol. Biol.* **26**, 217–221.

Taguchi, A. and White, M. F. (2008). Insulin-like signaling, nutrient homeostasis, and life span. *Annu. Rev. Physiol.* **70**, 191–212.

Taubes, G. (2012). Cancer research. Unraveling the obesity-cancer connection. *Science* **335**, 28,30–32.

Teleman, A. A., Hietakangas, V., Sayadian, A. C. and Cohen, S. M. (2008). Nutritional control of protein biosynthetic capacity by insulin via Myc in Drosophila. *Cell Metab.* **7**, 21–32.

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., Giste, E., Fisher, W., Hammonds, A., Celniker, S. E., et al. (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biol.* **12**, R43.

Tu, M.-P., Yin, C.-M. and Tatar, M. (2002). Impaired ovarian ecdysone synthesis of Drosophila melanogaster insulin receptor mutants. *Aging Cell* **1**, 158–160.

Ullrich, A. and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203–212.

Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y. C. and Tsubokawa, M. (1985). Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature* **313**, 756–761.

Ullrich, A., Gray, A., Tam, A. W., Yang-Feng, T., Tsubokawa, M., Collins, C., Henzel, W., Le Bon, T., Kathuria, S. and Chen, E. (1986). Insulin-like growth

factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity. *EMBO J.* **5**, 2503–2512.

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T. M., Fernandez, N., Ballester, B., Andrau, J. C. and Spicuglia, S. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.* **6**, 6905.

Webster, N. J., Resnik, J. L., Reichart, D. B., Strauss, B., Haas, M. and Seely, B. L. (1996). Repression of the insulin receptor promoter by the tumor suppressor gene product p53: a possible mechanism for receptor overexpression in breast cancer. *Cancer Res.* **56**, 2781–2788.

Wei, Y. and Arnosti, D. N. (2015). Enhancer trafficking: free throws and three-pointers. *Dev. Cell* **32**, 135–137.

Wei, Y., Mondal, S. S., Mouawad, R., Wilczynski, B., Henry, R. W. and Arnosti, D. N. (2015). Genome-Wide Analysis of Drosophila Rbf2 Protein Highlights Diversity of RB Family Targets and Possible Role in Regulation of Ribosome Biosynthesis. *G3 (Bethesda)*.

Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., Wang, F., Su, J., Wu, Q., Liu, H., et al. (2016). SEA: a super-enhancer archive. *Nucleic Acids Res.* **44**, D172–9.

Wong, D. M., Shen, Z., Owyang, K. E. and Martinez-Agosto, J. A. (2014). Insulin- and warts-dependent regulation of tracheal plasticity modulates systemic larval growth during hypoxia in Drosophila melanogaster. *PLoS One* **9**, e115297.

Yamaguchi, Y., Flier, J. S., Benecke, H., Ransil, B. J. and Moller, D. E. (1993). Ligand-binding properties of the two isoforms of the human insulin receptor. *Endocrinology* **132**, 1132–1138.

Yuh, C. H. and Davidson, E. H. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069–1082.

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

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W. and Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J. Biol. Chem.* **289**, 24863–24873.

CHAPTER 4

Future Directions

Biochemical characterization of Rb family proteins connected these corepressors with E2F family transcription factors, which were known to control transcription of cell cycle genes. Indeed, a variety of experiments have demonstrated that genetic disruption of RB function can promote activation of S-phase genes (reviewed in Du and Pogoriler, 2006). However, deletion of *RB* gene in mouse models reveal multiple tissue-specific defects in neuronal cells, erythropoiesis, stem cells, and progenitors, suggesting that RB proteins have additional functions beyond that of cell cycle control (Jacks et al., 1992). At the molecular level, in addition to DNA replication control, RB proteins are have been implicated in regulation of apoptosis, polarity, cell differentiation, mitochondria function, and metabolism (Du and Pogoriler, 2006; Ambrus et al., 2013; Reynolds et al., 2013; Nicolay et al., 2013; Payankaulam et al., 2016). However, we still lack specific knowledge of how RB might be directly contributing to these processes. Genome-wide studies have provided clues to direct transcriptional targets whose regulation may underlie these processes. In both Drosophila and human, RB proteins have been found to associate with many genes involved in many other cellular functions besides cell cycle control (Chicas et al., 2010; Acharya et al., 2012; Korenjak et al., 2012; Wei et al., 2015). Two intriguing categories of genes that are directly bound by RB proteins are ribosomal protein genes and signaling pathway genes. These classes of genes display strong association with RB proteins, yet have not been described as or investigated previously as RB regulated genes.

Although genome-wide studies provide candidate targets, physical binding is not proof that regulation occurs. Thus in following up on these studies, it is important to understand the functional significance of RB interactions with those target genes. In my thesis research, I have shown that Rbf proteins have modest but significant repression activities on specific ribosomal protein gene promoters, as well as an important signaling gene, *InR*. In addition, with misexpression of Rbf1 in larval wing disc tissue, our lab has carried out RNA-seq analysis and found that a full spectrum of ribosomal protein genes were deregulated, although the changes noted were less than two-fold, and clearly of lesser degree than that of cell cycle promoters (Wei et al., 2015). These data suggest that the impact of RB on these genes may be less dramatic than that for canonical cell cycle roles. However, is this "modest" activity biologically significant? Ribosomal genes and many of the signaling pathway genes are central "housekeeping" genes that play critical roles in cellular homeostasis. Thus, modest changes in expression of these genes may lead to pleotropic consequences. Studies have shown that the ribosomal gene expression is highly synchronized, and disturbance of few genes' expression can cause severe defects in ribosomal assembly (Steffen et al., 2012; Xue and Barna, 2012; Woolford and Baserga, 2013; Hasygar and Hietakangas, 2014). Thus, a fine-tuning function of RB proteins on ribosomal protein genes may contribute to an important part maintaining cellular homeostasis. Previous transcriptomic studies have generally neglected such events by setting a threshold that overlooked changes of less than 1.5-2 fold, thus high precision studies to study the effects of RB overexpression or knock-down in diverse conditions may reveal more of these fine-tuning functions of Rb in ribosomal biosynthesis and other signaling pathways (Dimova et al., 2003).

My research suggests that RB regulation may involve more on/off switch like function for cell cycle promoters, and a more incremental "governor" function for many other types of genes. One complication for discerning these diverse types of regulation is that complete ablation of an RB gene, or significant overexpression may trigger complex transcriptional responses directly or indirectly associated with RB, making it difficult to isolate the direct RB function on a given group of genes. Therefore to characterize the RB function on specific genes such as ribosomal genes or components of the insulin signaling pathway, an approach emphasizing the *cis*-elements, rather than the *trans*-factor, would be more beneficial. The recently-developed CRISPR genome editing technique offers such a tool (Hsu et al., 2014). By deleting RB *cis*-elements on the target gene promoters in the context of the whole organism, we will be able to specify and characterize the impact of RB on these genes' expression, and give rise to reveal RB function in protein synthesis and cellular signaling.

Another path to explore RB function on these noncanonical targets is to characterize how RB proteins change the local chromatin conformation. It is possible that RB on such promoters does not have a role in masking an E2F activation domain, but rather induces chromatin marks that are inhibitory. Many studies have described RB function in the regulation of chromatin structure (as reviewed in Talluri and Dick, 2012). Also, human RB has been shown to interact with hundreds of proteins, many of which are involved in histone modifications and nucleosome positioning (Morris and Dyson, 2001). Thus Rb binding to these gene promoters may give an onset of chromatin modifications that creating binding surface for other cofactors. In addition, my study shows that ribosomal and signaling pathway gene promoters lack of strong E2F site, indiating that

RB function on these promoters may involve a different sets of co-factors. Thus further sophisticated bioinformatics analysis would be helpful to determine poten RB co-factors on these promoters.

As another member of Drosophila RB family protein, the Rbf2 protein has been neglected, due to its "insignificant" knockout phenotype (no loss of viability, but an increase in egg laying) (Stevaux et al., 2005). Rbf2 is specifically highly expressed in the ovaries, therefore this phenotype might indicate that Rbf2 regulates protein production in this very active biosynthetic context via control of ribosomal protein gene promoters. Nutritional signals are directly sensed by the female to control egg production, which is important for the animal's adaptation to the environment. Thus another long-term goal would be to characterize Rbf2 function, especially in female reproduction. The previous genetic knock-out of *rbf2* gene was not a satisfying approach, as the process also introduced disturbance to a nearby gene (Stevaux et al., 2005). As mentioned above, CRISPR would be an ideal tool to generate a clean genetic knock-out of *rbf2* to understand its function in fly development and physiology.

An intriguing aspect of the genome-wide targets of Rbf1 is its strong association with many signaling pathway genes. Many developmental studies have identified how spatial and temporal specificity is a reflection of the correct engagement of signaling systems in tissues and distinct cell types. However, less focus has been devoted to differential expression of the components of signaling systems, a process that may change the long-term sensitivity of such pathways in physiological adaptation. My study aimed at describing the complete transcriptional circuit for one of the most important signaling receptors, which also associates with strong RB binding, the Drosophila insulin receptor

(InR). I showed that this widely expressed "housekeeping" gene has the elaborate controls that are normally associated with developmental genes, indicating that the precise modulation of this gene's expression is crucial. This finding may provide a new paradigm for the analysis of cis regulation, for it suggests that there may be a large set of genes sitting at critical points in pleiotropic pathways whose regulation is extremely finely-tuned, yet rarely dramatically up or down regulated in the manner of some of the classic developmental genes. We would expect that *cis*- regulatory variation mapped to these loci may be equally important in understanding population variation, and human disease. A broader question is whether this regulatory paradigm applies to other "housekeeping" genes. One clue might be that some genes, such as *dFOXO* (with ~20 kbp introns), and *EcR* (with ~50 kbp introns), have rather sizeable intergenic or intragenic regions which may be replete with regulatory elements. STARR-seq and other genome-wide studies have offered convenient resources to start searching these features at genome level (Arnold et al., 2013). This would also change our way to view "housekeeping" and "developmental" genes, as this simple binary classification may obscure important transcriptional regulation that would be shaped by evolution.

Following my study of the *cis*-regulatory elements of the *InR* gene, an immediate question is what are the *trans*-acting factors in this circuitry? We are using bioinformatics tools to predict potential transcription factors that associate the *cis*-elements, and will further characterize their molecular function in regulating *InR* gene expression. Another aspect remains to be understood is the physiological relevance of these CREs. One intriguing study would be to genetically modify these CREs with CRISPR, to characterize their function *in vivo*, and understand how they impact the

feedback regulation and the plasticity of the *InR* gene expression. In addition, we could also introduce sequences representing natural population variation into the gene, to understand whether these SNPs and indels have measureable, functional impact on the *InR* gene expression, and physiology in the whole organism. On a broader view and from the aspect of human genetics, comparative genomic studies would be useful to characterize the association between the population variation at the insulin receptor locus and diabetes, which has shown strong propensity in certain populations (Cornelis and Hu, 2012).

REFERENCES

REFERENCES

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

Ambrus, A. M., Islam, A. B. M. M. K., Holmes, K. B., Moon, N. S., Lopez-Bigas, N., Benevolenskaya, E. V and Frolov, M. V (2013). Loss of dE2F compromises mitochondrial function. *Dev. Cell* **27**, 438–451.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077.

Carneiro, T., Carvalho, C., Braga, J., Rino, J., Milligan, L., Tollervey, D. and Carmo-Fonseca, M. (2008). Inactivation of cleavage factor I components Rna14p and Rna15p induces sequestration of small nucleolar ribonucleoproteins at discrete sites in the nucleus. *Mol. Biol. Cell* **19**, 1499–1508.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R. a, Narita, M., Zhang, M. and Lowe, S. W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–87.

Cornelis, M. C. and Hu, F. B. (2012). Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. *Annu. Rev. Nutr.* **32**, 245–259.

Dimova, D. K., Stevaux, O., Frolov, M. V and Dyson, N. J. (2003). Cell cycle-dependent and cell cycle-independent control of transcription by the Drosophila E2F/RB pathway. *Genes Dev.* **17**, 2308–2320.

Du, W. and Pogoriler, J. (2006). Retinoblastoma family genes. *Oncogene* **25**, 5190–5200.

Hasygar, K. and Hietakangas, V. (2014). p53- and ERK7-Dependent Ribosome Surveillance Response Regulates Drosophila Insulin-Like Peptide Secretion. *PLoS Genet.* **10**, e1004764.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278.

Korenjak, M., Anderssen, E., Ramaswamy, S., Whetstine, J. R. and Dyson, N. J. (2012). RBF binding to both canonical E2F targets and noncanonical targets depends on functional dE2F/dDP complexes. *Mol. Cell. Biol.* **32**, 4375–87.

Morris, E. J. and Dyson, N. J. (2001). Retinoblastoma protein partners. *Adv. Cancer Res.* **82**, 1–54.

Nicolay, B. N., Gameiro, P. a, Tschöp, K., Korenjak, M., Heilmann, A. M., Asara, J. M., Stephanopoulos, G., Iliopoulos, O. and Dyson, N. J. (2013). Loss of RBF1 changes glutamine catabolism. *Genes Dev.* **27**, 182–96.

Payankaulam, S., Yeung, K., McNeill, H., Henry, R. W. and Arnosti, D. N. (2016). Regulation of cell polarity determinants by the Retinoblastoma tumor suppressor protein. *Sci. Rep.* **6**, 22879.

Reynolds, M. R., Lane, a N., Robertson, B., Kemp, S., Liu, Y., Hill, B. G., Dean, D. C. and Clem, B. F. (2014). Control of glutamine metabolism by the tumor suppressor Rb. *Oncogene* **33**, 556–66.

Steffen, K. K., McCormick, M. A., Pham, K. M., MacKay, V. L., Delaney, J. R., Murakami, C. J., Kaeberlein, M. and Kennedy, B. K. (2012). Ribosome deficiency protects against ER stress in Saccharomyces cerevisiae. *Genetics* **191**, 107–118.

Stevaux, O., Dimova, D. K. and Dyson, N. J. (2005). Retinoblastoma Family 2 is Required In Vivo for the Tissue-Specific Repression of dE2F2 Target Genes ND ES. **4**, 1272–1280.

Talluri, S. and Dick, F. A. (2012). Regulation of transcription and chromatin structure by pRB: here, there and everywhere. *Cell Cycle* **11**, 3189–3198.

Wei, Y., Mondal, S. S., Mouawad, R., Wilczynski, B., Henry, R. W. and Arnosti, D. N. (2015). Genome-Wide Analysis of Drosophila Rbf2 Protein Highlights Diversity of RB Family Targets and Possible Role in Regulation of Ribosome Biosynthesis. *G3 (Bethesda)*.

Xue, S. and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369.

APPENDICES

APPENDIX A

Role of Rbf1/E2F1 in regulating signaling pathway gene promoters

In Chapter 2, I showed that Rbf1 binds to many signaling pathway gene promoters. Rbf1 is a cofactor for transcription activator dE2F1. To test if dE2F1 functions as an activator on some of these gene promoters, I performed luciferase assays using S2 cells cotransfected with luciferase reporters driven by selected signaling pathway gene promoters, and dE2F1 expression vector (as described in Material and Methods in Chapter 2, except that 20 ng dE2F1 expression vector was cotransfected with 600 ng luciferase reporters). Results are shown in Figure A-1. Only *PCNA* promoter, as a positive control, was activated by dE2F1. All other tested promoters were not activated, or even repressed by dE2F1.

This result was published in the following manuscript:

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

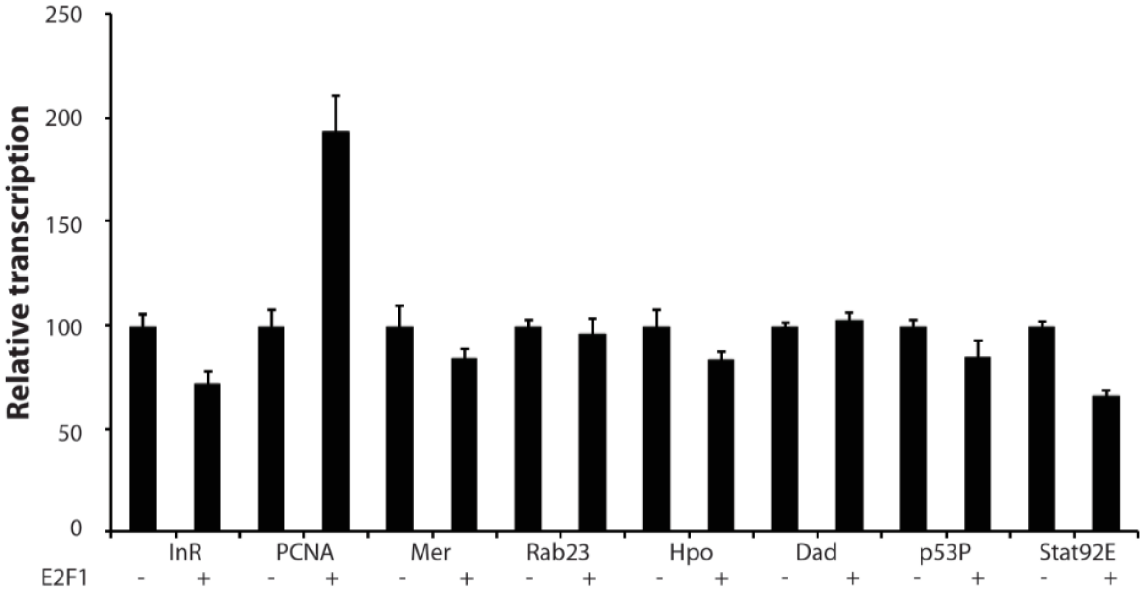


Figure A-1: E2F responsiveness of promoters of selected genes in signaling pathways. Drosophila S2 cells were cotransfected with *InR, PCNA, Merlin, Rab23, Hippo, Dad, p53-proximal,* or *Stat92E* luciferase reporters, with (+) or without (-) a plasmid overexpressing E2F1. Only *PCNA* luciferase expression was elevated by E2F1.

To compare Rbf1 activity on signaling pathway gene promoters and cell cycle gene promoters, I performed luciferase assays on three selected signaling pathway gene promoters (*InR*, *wts*, and *Pi3K68D*), and three selected cell cycle gene promoters (*PCNA*, *Polα*, and *Mcm7*). The results are shown in Figure A-2. Rbf1 shows strong repression on all three cell cycle gene promoters, while modest repression on signaling pathway gene promoters. However, the Rbf1 "IE" domain, as described in Chapter 1, was required for Rbf1 activity on the cell cycle gene promoters, but not on signaling pathway gene promoters.

This result was published in the following manuscript:

Raj, N., Zhang, L., Wei, Y., Arnosti, D. N. and Henry, R. W. (2012). Ubiquitination of retinoblastoma family protein 1 potentiates gene-specific repression function. *J. Biol. Chem.* **287**, 41835–41843.

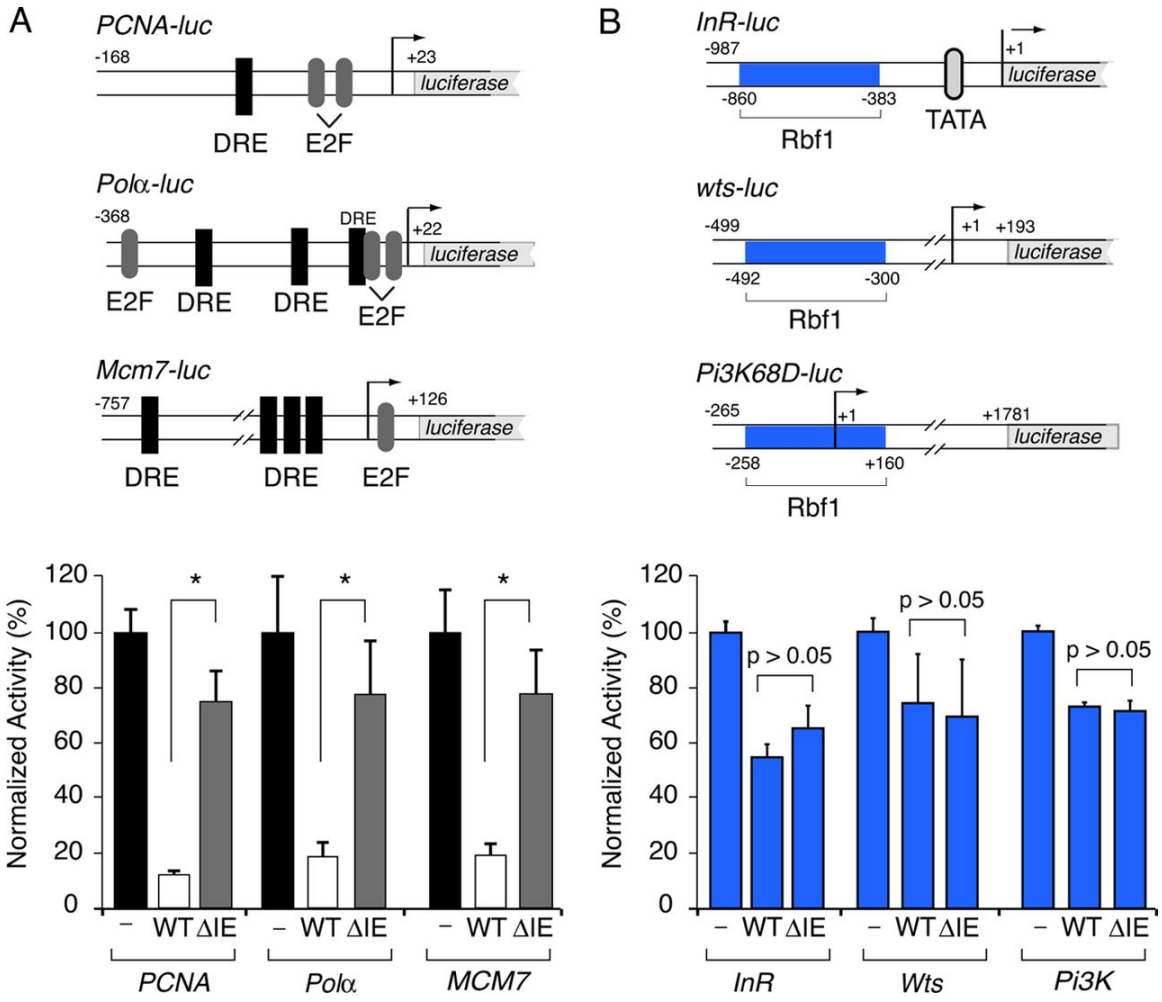


Figure A-2: Context dependence of the Rbf1-IE for transcriptional repression. Rbf1 WT and Rbf1 ΔIE showed dissimilar repression activities on the E2F1-dependent reporters as compared with the E2F1-independent promoters. Data are from at least three biological replicates. *, $p < 0.05$, *error bars* indicate S.D.

REFERENCES

REFERENCES

Acharya, P., Negre, N., Johnston, J., Wei, Y., White, K. P., Henry, R. W. and Arnosti, D. N. (2012). Evidence for autoregulation and cell signaling pathway regulation from genome-wide binding of the Drosophila retinoblastoma protein. *G3 (Bethesda)*. **2**, 1459–1472.

Raj, N., Zhang, L., Wei, Y., Arnosti, D. N. and Henry, R. W. (2012). Ubiquitination of retinoblastoma family protein 1 potentiates gene-specific repression function. *J. Biol. Chem.* **287**, 41835–41843.

APPENDIX B

Rbf1 phosphoryation and stability

As described in Chapter 1 that Rbf1 protein activity is regulated by phosphorylation via Cdk/Cyc complexes. To test the association between Rbf1 phosphorylation and protein stability, I cotransfected S2 cells with Rbf1 and Cdk/Cyc expression vectors. As shown in Figure B-1, expression of Cdk/Cyc complexes stalized Rbf1 protein by increasing its protein half-life. Using a phospho-gel assay, I demonstrated that Cdk/Cyc complexes were able to phosphorlate and stabilize Rbf1 protein (Figure B-2).

These results were published in the following manuscript:

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W. and Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J. Biol. Chem.* **289**, 24863–24873.

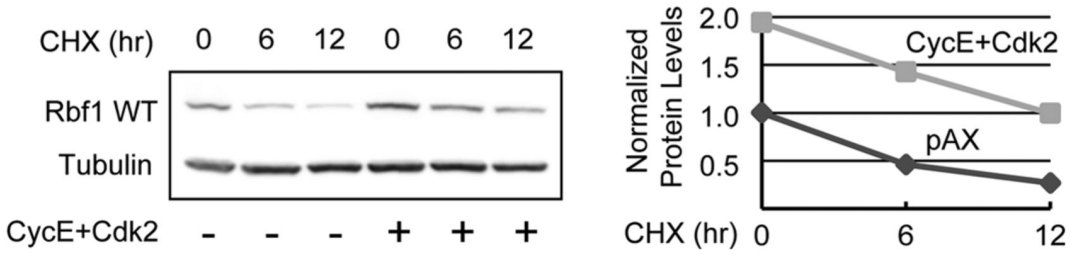


Figure B-1: Drosophila Rbf1 was subjected to Cyc-Cdk-mediated stabilization. Rbf1 protein stability is increased with concomitant expression of CyclinE and Cdk2. In this representative experiment, Rbf1 protein exhibited a half-life of 6 h, *versus* 12 h in the presence of Cyc-Cdk. Similar 2-fold or greater differences were noted in five independent experiments. Protein levels were quantitated by photon capture analysis with a Fuji LAS-3000 Imager and normalized to tubulin levels. *CHX*, cycloheximide.

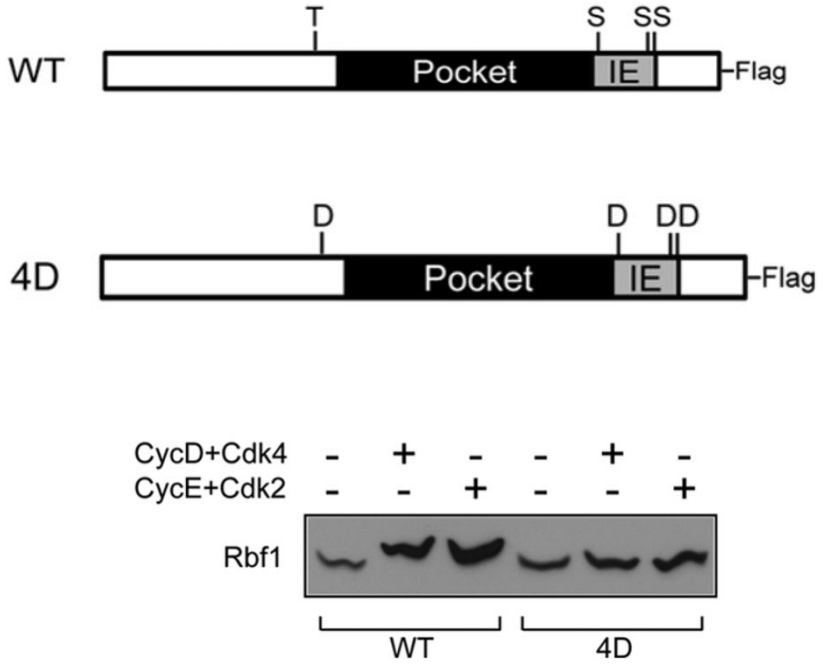


Figure B-2: Rbf1 protein is phosphorylated and stabilized by Cyc-Cdk complexes. Stabilization of Rbf1 under conditions of Cyc-Cdk overexpression is associated with direct modification of the protein and is dependent on conserved threonine/serine residues. Wild-type Rbf1 protein exhibits a mobility shift when run on the PhosTag™ gel system, indicative of phospho-protein, associated with its increased abundance. The Rbf1 4D mutant exhibits no shift and no significant increase in protein level.

REFERENCES

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W. and Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J. Biol. Chem.* **289**, 24863–24873.

APPENDIX C

Role of Rbf1 and RBf2 C-terminus in protein stability and activity

Rbf1 and Rbf2 proteins differ in their C-terminus domains. For Rbf1, the C-terminus contains IE, which is critical for both Rbf1 stability and activity. However, Rbf2 lacks of this feature. Biochemical analysis showed that Rbf1 binds to both dE2F1 and dE2F2, while Rbf2 is restricted to dE2F2; Rbf2 binds to many more promoters than Rbf1, however (as described in Chapter 2). Rbf1 acts as a dominant repressor, while the full repression function of Rbf2 relies on the presence of dE2F2. Thus the unique C-terminus of these Rb proteins might be critical separating Rbf1 and Rbf2. From an evolution perspective, the evolvement of Rbf C-terminus separated these two proteins, and may also allow the emergence of a new Rb protein that was not present in other insect species. This may contribute to unique lineage characters of Drosophila.

To test the unique features of C-terminus of Rb proteins, I swapped the C-terminus of Rbf1 with Rbf2, and created Rbf1 protein with Rbf2 C-terminus (Rbf1-Rbf2C), and Rbf2 protein with Rbf1 C-terminus (Rbf2-Rbf1C) (Figure C-1A). Since Rbf2 C-terminus does not contain IE domain, I expected that Rbf1-Rbf2C would be like Rbf1ΔC, which is more stable than the Rbf1 WT (Figure C-1B). However, Rbf1-Rbf2C showed the same protein level as Rbf1 WT, indicating the C-terminus of Rbf2 was able to act as "IE" in the context with the rest part of Rbf1 (Figure C-1B). Previous study showed that Rbf1 IE is an autonomous degron, and was able to decrease the protein stability when fused to GFP (Acharya et al., 2010). Thus I expected that Rbf2-Rbf1C

might destabilize Rbf2. However, the results indicated the opposite, that Rbf2-Rbf1C showed much higher protein level than Rbf2, suggesting that Rbf1C, in the context of Rbf2, stabilized Rbf2 protein (Figure C-1B). Why Rbf1 IE destabilized GFP, but not Rbf2? Maybe GFP lacks of pocket domains that can interact with Rbf C-terminus. In addition, Rbf2ΔC is less stable than Rbf2 WT, indicating the Rbf2 C-terminus contains a "stability element" (SE) (Figure C-1B).

The test how C-terminus affect Rbf protein activities, I compared the repression activities of Rbf1-Rbf2C and Rbf2-Rbf1C with the wild-type proteins. Rbf1-Rbf2C is able to repress *PCNA* promoter, but at much weaker level compared to Rbf1 WT. Rbf2-Rbf1C is a little stronger repressor compared to Rbf2 WT, which does not repress *PCNA* promoter by itself. Rbf2-Rbf1 showed modest repression activity on PCNA. In the presence of dE2F2, Rbf1-Rbf2C repressed *PCNA* promoter as good as RBf1 WT, a reminiscent that Rbf2 requires dE2F2 for full repression activity (Stevaux et al., 2005); Rbf2-Rbf1C repressed *PCNA* promoter as the WT Rbf2, while the Rbf2ΔC was slightly weaker repressor, suggesting Rbf1C may rescue Rbf2ΔC since Rbf1 is also able to interact with dE2F2. In the presence of dE2F1, Rbf1-Rbf2C repressed *PCNA* promoter a little better than Rbf1ΔC, suggesting the Rbf1-Rbf2C may not interact with dE2F1 as Rbf1WT, a reminiscent that Rbf2 does not biochemically interact with dE2F1 (Stevaux et al., 2005); Rbf2-Rbf1C failed to repress *PCNA* promoter, indicating that the presence of Rbf1C may not allow Rbf2-Rbf1C to interact with dE2F1 (Figure C-2).

The current results created a paradox: Rbf1C destabilize Rbf1, but stabilize Rbf2; Rbf2C stabilize Rbf2, but destabilize Rbf1. Thus the "IE" of Rbf1, or the "SE" of Rbf2 may not function autonomously, but rather has close interaction with the rest part of the protein,

possibly the pocket domains or N-terminus. In the context of Rbf1, the Rbf1C interacts with N or pocket domain to direct Rbf1 degradation; when replaced by Rbf2C, Rbf2C may still interact with Rbf1 pocket or N (since the pocket domains are highly conserved between Rbf1 and Rbf2), and this interaction mimics the overall structural conformation of Rbf1WT. In the context of Rbf2, the Rbf2C interacts with N or pocket domain to protect Rbf2 from degradation; when replaced by Rbf1C, it can still interact with the other parts of Rbf2 to keep the overall conformation to stabilize Rbf2-Rbf1C (Figure C-3). Although in the context of protein stability, Rbf2C was able to destabilize Rbf1 as Rbf1C was, and Rbf1C was able to stabilize Rbf2 as Rbf2C was, Rbf1-Rbf2C and Rbf2-Rbf1C did not maintain their repression activities as the wild type proteins. This may be due to the possibility that the C-terminus of Rbf proteins contribute to their interaction with specific E2F proteins. The Rbf1-Rbf2C acts more like Rbf2 WT, that its full repression activity requires co-expression with dE2F2. However, Rbf2-Rbf1C still acts Rbf2WT, indicating the interaction between Rbf1C with Rbf2 may still mimics the overall conformation of Rbf2 WT (as indicated by protein stability), which does not allow it to interact with dE2F1 (Figure C-3).

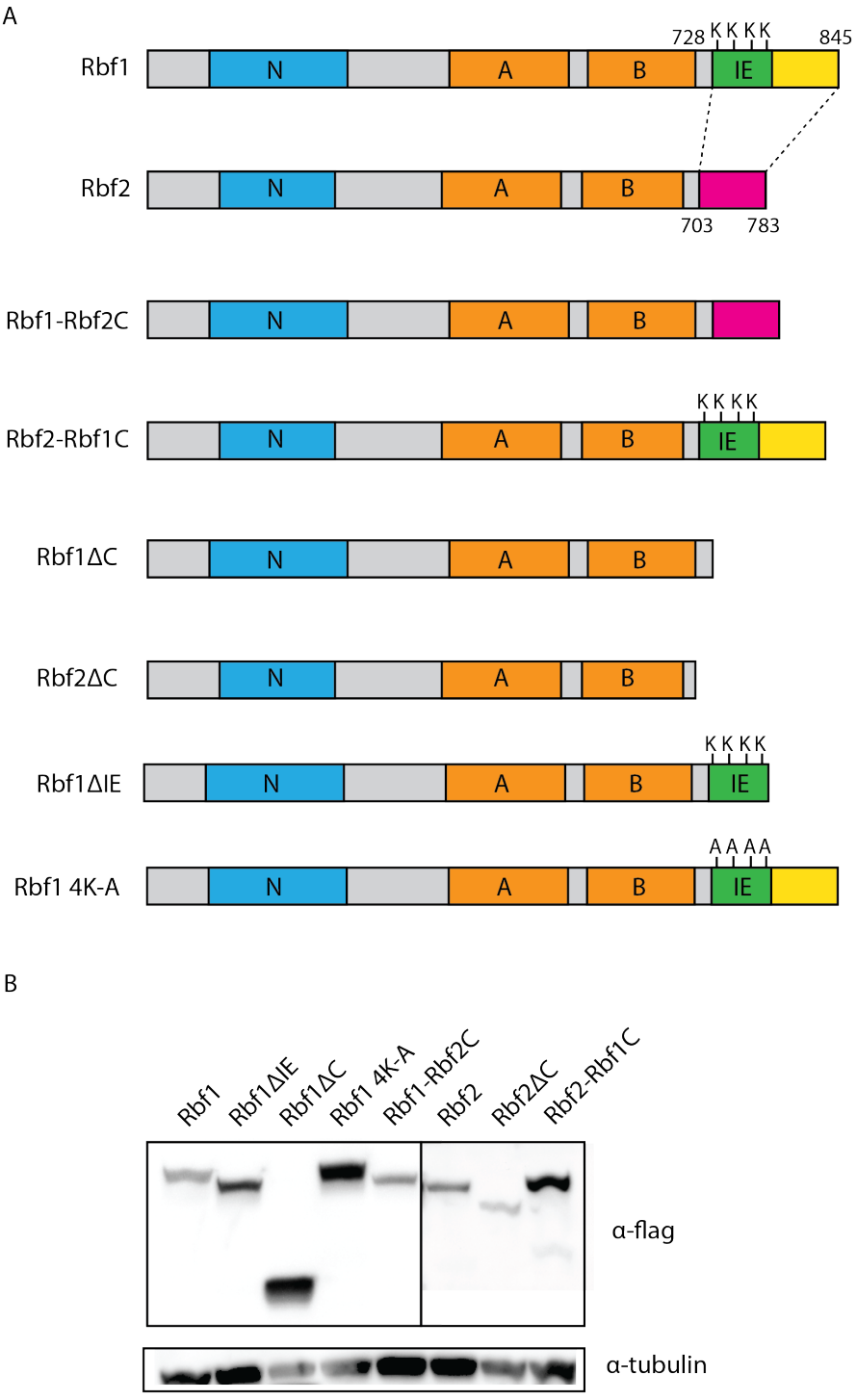


Figure C-1: C-terminus of Rbf1 and Rbf2 proteins contribute to the protein stabilities. (A) Overview of different Rbf1 and Rbf2 constructs. (B) Protein levels of different Rbf1 and Rbf2 constructs when expressed in S2 cells.

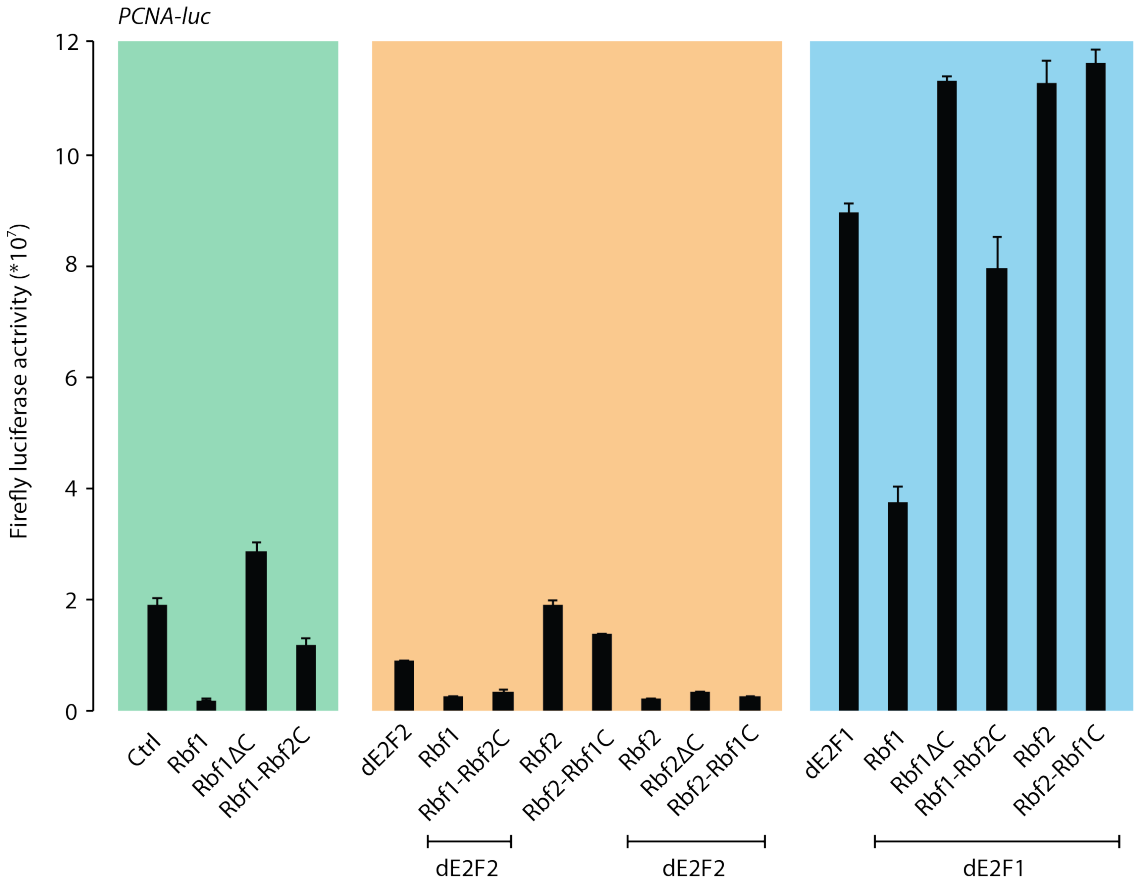


Figure C-2: C-terminus of Rbf1 and Rbf2 proteins partially contribute to their activities.

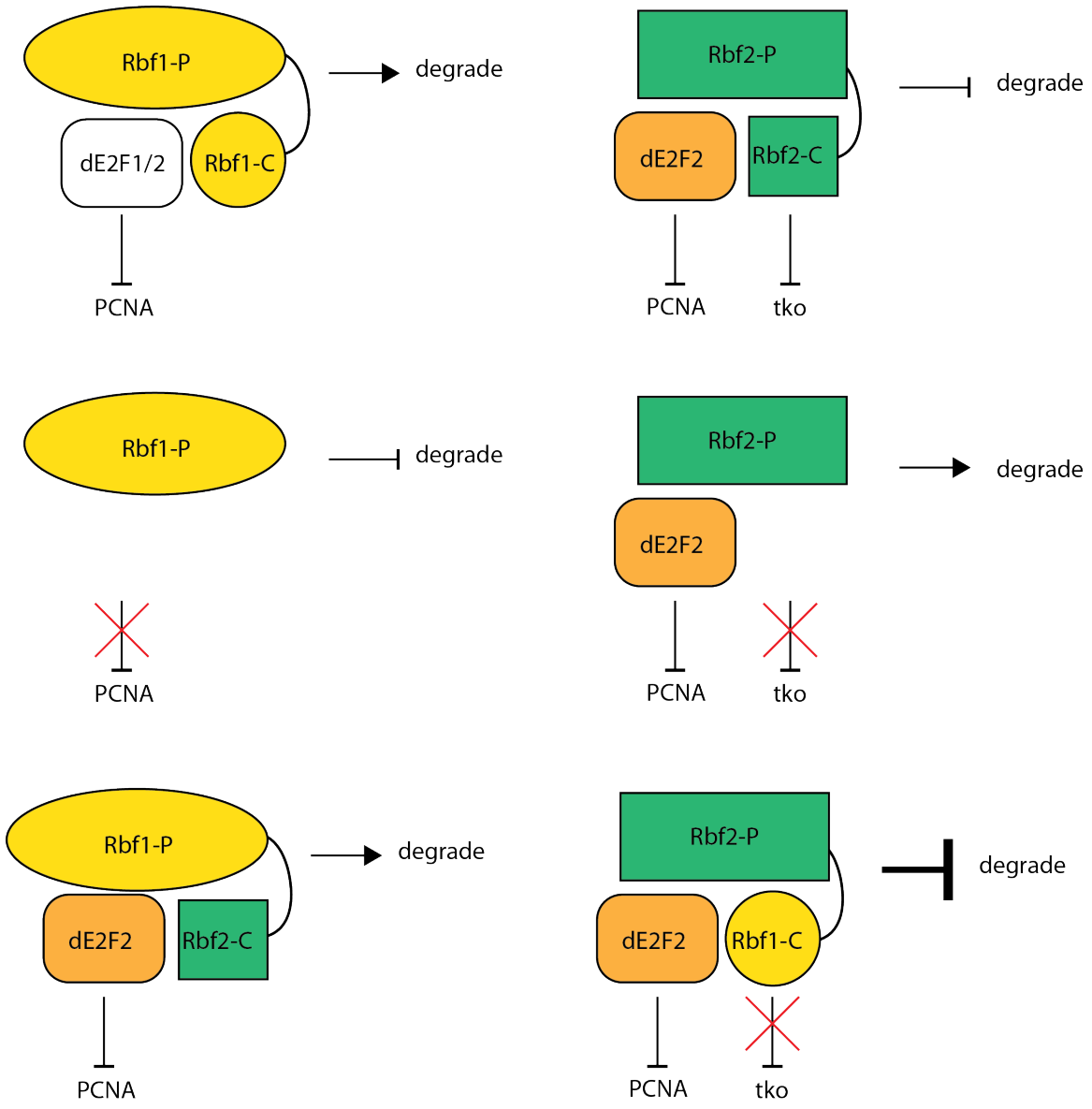


Figure C-3: A model showing the specific interation between Rb C-terminus and the rest of the protein contributes to the protein stability and activity.

REFERENCES

REFERENCES

Acharya, P., Raj, N., Buckley, M. S., Zhang, L., Duperon, S., Williams, G., Henry, R. W. and Arnosti, D. N. (2010). Paradoxical instability-activity relationship defines a novel regulatory pathway for retinoblastoma proteins. *Mol. Biol. Cell* **21**, 3890–3901.

Stevaux, O., Dimova, D. K. and Dyson, N. J. (2005). Retinoblastoma Family 2 is Required In Vivo for the Tissue-Specific Repression of dE2F2 Target Genes ND ES. **4**, 1272–1280.