

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

The Central Region of the Drosophila Co-repressor Groucho as a Regulatory Hub

**Permalink**

<https://escholarship.org/uc/item/9q43k50r>

**Author**

Kwong, Pak Ning

**Publication Date**

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
Los Angeles

The Central Region of the Drosophila Co-repressor Groucho as a Regulatory Hub

A dissertation submitted in partial satisfaction  
of the requirements for the degree Doctor of Philosophy  
in Biochemistry and Molecular Biology

by

Pak Ning Kwong

2016

© Copyright by

Pak Ning Kwong

2016

## **ABSTRACT OF THE DISSERTATION**

The Central Region of the Drosophila Co-repressor Groucho as a Regulatory Hub

by

Pak Ning Kwong

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, CA 2009

Professor Albert J. Courey, Chair

Groucho (Gro) is a transcriptional corepressor that plays a critical role in Drosophila embryonic development. It contains ordered C- and N-terminal domains required, respectively, for repressor binding and tetramerization in addition to a disordered, but essential, central region. While many mechanisms of Gro mediated repression have been proposed, none fully account for the functional importance of the

disordered central domains. In this thesis, I attempt to elucidate the function of this region.

The unstructured Gro central domains may mediate a wide range of protein-protein interactions. To identify Gro interacting proteins, I performed affinity purification followed by mass spectrometry analysis. I recovered over 160 potential interacting proteins that function in chromatin organization, developmental processes, and mRNA splicing. In addition, a co-IP experiment was performed to confirm that one of the interacting complexes, the U1 snRNP complex, physically interacts with Gro. I also employed a high throughput RNAi screen involving Gal4-Gro-responsive luciferase reporter to systematically test the functional relevance of the 160 potential interacting proteins. Two of the U1 specific proteins were shown to be positive regulators of Gro function. As the affinity purification and reporter assay results imply regulation of Gro activity by the U1 snRNP complex, I analyzed the expression profile of cells subjected to knockdown of Gro or snRNP U1C. The two data sets revealed a significant overlap between genes regulated by these two factors. Furthermore, comparison of our RNA-seq data with Gro and RNA polymerase II ChIP data led to number of insights, including the finding that Gro-repressed genes are enriched for promoter-proximal RNA polymerase II. In conclusion, these findings suggest roles for spliceosomal components and paused Pol II in Gro mediated repression.

The dissertation of Pak Ning Kwong is approved.

Jorge Torres

Frank A. Laski

Albert J. Courey, Committee Chair

University of California, Los Angeles

2016

## TABLE OF CONTENTS

<b>Abstract of the Dissertation .....</b>	<b>ii</b>
<b>Acknowledgements .....</b>	<b>vii</b>
<b>Vita.....</b>	<b>ix</b>
<b>Chapter 1: Introduction: Groucho as transcriptional co-repressor .....</b>	<b>1</b>
Figures.....	16
References.....	28
<b>Chapter 2: The Central Region of the Drosophila Co-Repressor Groucho as a Regulatory Hub .....</b>	<b>37</b>
Introduction .....	39
Materials and Methods.....	40
Results.....	42
Discussion.....	47
References.....	49
<b>Chapter 3: Interaction between Groucho and histone deacetylase .....</b>	<b>51</b>
Abstract.....	52
Introduction .....	53
Materials and Methods.....	55
Results.....	60
Discussion.....	62
Figures.....	65
References.....	73
<b>Chapter 4: A revised model of Gro mediated repression .....</b>	<b>74</b>
Figures.....	80
References.....	84

## LIST OF FIGURES

Figure 1-1 .....	16
Figure 1-2 .....	18
Figure 1-3 .....	20
Figure 1-4 .....	22
Figure 1-5 .....	24
Figure 1-6 .....	26
Figure 3-1 .....	65
Figure 3-2 .....	67
Figure 3-3 .....	69
Figure 3-4 .....	71
Figure 4-1 .....	80
Figure 4-2 .....	82



## ACKNOWLEDGEMENTS

I would like to thank my parents for bringing me to America. They made a lot of sacrifices to provide a better education opportunity for me and my brother. They were extremely supportive on every decision I made, including the decision to come to UCLA for graduate studies.

I would like to thank my mentor Albert Courey for providing the opportunity to work in his lab. Through his guidance, I have the chance to formulate hypothesis and analysis data independently, which led to the discovery of Gro interacting with spliceosome that no one in the lab has expected. The training I received allows me to become a well-rounded scientist, with a lot of valuable experience that I would not have received otherwise.

I would like to thank Wiam Turki-Judeh for her assistance on various experiments presented in chapter 2 and chapter 3, and her effort to create a framework of my project. I would like to thank Michael Chambers for providing his expertise of bioinformatics to generate all the results presented in chapter 2. I would like to thank Ajay Vashisht and James Wohlschlegel for the collaboration on the mass spectrometry analysis presented in chapter 2. I would like to thank Tak Yu Yau for his assistance on the co-IP experiment presented in chapter 2. I would like to thank all former Courey lab members for sharing the laughter and tears throughout my graduate career.

I would like to thank my committee members, Guillaume Chanfreau, Jorge Torres, Frank Laski, and Siavash Kurdistani for their advice. My publication would not have happened without their insightful guidance.

I would like to thank my undergraduate research mentor Wolf Heyer for allowing me to work in his lab when I knew nothing about research. I would specifically thank my post graduate mentor Simon Chan for taking me as a lab technician. It was his influence that I became engaged in science. In addition, I would not have been able to get into UCLA without his help.

I would like to acknowledge the American Society for Biochemistry and Molecular Biology for allowing me to reprint “The central region of the Drosophila co-repressor Groucho as a regulatory hub” Journal of biological chemistry 2015;290(50):30119-30130.

## **Vita**

- 2007  
B.S., Biochemistry and Molecular Biology  
University of California, Davis  
Davis, CA
- 2007-2009  
Laboratory Assistant  
Dr. Simon Chan Laboratory  
University of California, Davis  
Davis, CA
- 2009-2016  
Graduate Student Researcher  
Dr. Albert Courey Laboratory  
University of California, Los Angeles  
Los Angeles, CA
- 2009-2013  
Teaching Assistant  
Protein Purification and Characterization  
Cold Spring Harbor Laboratory  
Cold Spring Harbor, NY
- 2009-2016  
Teaching Assistant  
University of California, Los Angeles  
Los Angeles, CA

## Publications

Kwong P N, Chambers M, Vashisht A, Turki Judeh W, Yau T Y, Wohlschlegel J A, Courey A J. The Central Region of the Drosophila Co-repressor Groucho as a Regulatory Hub. *Journal of biological chemistry* 2015;290(50):30119-30130.

Ravi M, Kwong P N, Menorca R M, Valencia J T, Ramahi J S, Stewart J L, Tran R K, Sundaresan V, Comai L, Chan S W. The rapidly evolving centromere-specific histone has stringent functional requirements in *Arabidopsis thaliana*. *Genetics* 2010;186(2):461-471.

## **Chapter 1**

### **Introduction:**

**Groucho as transcriptional co-repressor**

## **Introduction**

Proper development relies on correct temporal and spatial control of gene expression, which is tightly regulated by both activators and repressors. Co-repressors are regulatory factors that are essential for repression, but that lack DNA binding domains. Rather than binding to DNA directly, they are recruited to genes via protein-protein interactions with DNA-bound repressors. Groucho (Gro) is a co-repressor conserved in metazoans. Although mounting evidence illustrates the global roles of Gro in many different developmental contexts, we still do not fully understand the mechanisms underlying Gro repressive activity. Characterization of proteins that directly or indirectly associate with Gro is crucial for insight into the mechanisms of Gro-mediated repression. The central focus of this thesis will be on discovering proteins that regulate the activity of Gro.

## **Gro and its homologs**

Gro is encoded by a *Drosophila* gene first discovered through a viable hypomorphic allele that results in extra bristles above the eyes, a phenotype resembling the bushy eyebrows of the American movie star Groucho Marx [1]. The *gro* gene is located in the Enhancer of split (*E(spl)*) gene complex, which includes genes encoding seven basic helix-loop-helix (bHLH) transcription factors of the Hairy-Enhancer of Split (HES) subfamily, although Gro itself does not contain a bHLH domain [2]. Instead, it contains a WD-repeat domain, which is also present in the  $\beta$  subunit of trimeric G proteins including transducin

[3]. For these reasons, the human homologs of Gro were named transducin-like Enhancer of Split (TLE) proteins [4]. While there is only one Gro family protein in *Drosophila*, four non-redundant Gro homologs are encoded in both the human genome (TLE1-4) [5] and the mouse genome [Groucho-like genes (Grg) 1-4] [6]. Gro homologs are also found in *C. elegans* (UNC-37) [7] and *Xenopus* [Enhancer of Split Groucho 1 (ESG1)] [8]. In addition, functionally analogous proteins containing WD-repeat domains are found in yeast (Tup1) [9] and in plants (TPL/TPR/WSIP and LUG/LUH) [10], although it is not clear if the yeast and plant proteins should be considered true Gro orthologs.

### **Structure and function of Gro**

Gro has a five domains structure based on sequence homology: the Q, GP, CcN, SP, and WD-repeat domains (Figure 1-1) [11]. Crystal structures of the Q and WD-repeat domains are available, and their functions are well characterized [12, 13]. In contrast to the Q and WD-repeat domains, the central GP, CcN, and SP domains are predicted to be disordered by their hydrophobicity and charge density to databases of ordered and disordered proteins [14], and they were thought to be functionally dispensable as none of the known *gro* point mutations was mapped to these regions [15]. However, recent study has demonstrated their functional importance in Gro-mediated repression [14]. This section will provide an overview of each domain.

#### *1. Q domain*

The Q domain includes the first 133 amino acids of Gro and is so-named for its high glutamine content. It forms a 70 amino acids long helix followed by two short helices. The Q domain can self-associate, forming an extensive parallel coiled-coil dimer. The dimer can further self-associate, resulting in an antiparallel tetramer [12].

Tetramerization of the Q domain is essential to Gro function. Point mutations in the Q domain that disrupt the tetramerization structure abolish Gro repression in tissue culture cell luciferase reporter assay [16], while deletion of the first 38 residues of the long helix results in the up-regulation of a vgQ-LacZ reporter that is normally repressed by Gro acting through the repressor Brinker in the wing imaginal disk [15]. Furthermore, reporter assays in HEK293T cell have shown that only the TLE tetramer, but not the dimer exhibits repression activity [12].

Besides promoting tetramerization, the Q domain can also bind to Tcf/Lef and negatively regulate Wnt-signaling [17, 18]. Binding of Tcf is down regulated by ubiquitylation of the Q domain and is independent of tetramerization, since ubiquitylation did not disrupt tetramerization [19], and both the TLE tetramer and dimer have the same affinity for Tcf [12].

## *2. GP domain*

The GP domain includes amino acid residues 134-194 of Gro and is so-named for its glycine/proline-rich composition. It is necessary for binding of Gro to histone



deacetylase Rpd3/HDAC1 [20], which may allow Gro to repress gene expression by promoting the deacetylation of histone tails and increasing nucleosome density [21]. A recent study has also showed an unexpected role of the GP domain in nuclear localization, in which deletion of the GP domain resulted in relocalization of Gro to the cytoplasm [14].

### *3. CcN domain*

The CcN domain includes amino acid residues 195-257 of Gro and is so-named for its putative nuclear localization signal and several putative phosphorylation sites for casein kinase 2 (CK2) and cdc2. S239 and S253 of TLE1 were shown to be phosphorylated by CK2, and mutation of S239A resulted in a reduction of TLE1 activity in a reporter assay (S239 is conserved in *Drosophila* Gro but not S253) [22]. Although the CcN domain was long believed to be responsible for nuclear localization, the majority of Gro remains in the nucleus when this domain is deleted [14].

### *4. SP domain*

The SP domain includes amino acid residues 258-390 of Gro and is so-named for its serine/proline-rich composition. Phosphorylation of this domain by MAPK resulted in reduction of Gro activity, suggesting negative regulation of Gro activity by the receptor tyrosine kinase (RTK)/Ras signaling pathway [23, 24]. In addition, the SP domain seems to be important for target gene specificity, as deletion of this domain resulted in repression of non-Gro target genes [14]. However, RNA-seq of overexpressed SP deleted Gro did not

lead to a major shift in the transcriptome compared to that resulting from overexpression of wild type Gro (unpublished data).

### *5. WD-repeat domain*

The WD-repeat domain includes amino acid residues 391-719 of Gro and forms a seven bladed  $\beta$ -propeller [13]. It binds to the WRPW/Y or eh1 (FxlxxIL) motif, allowing Gro to be recruited to DNA by repressors [25]. Examples of WRPW/Y motif-containing repressors that recruit Gro include Hes1/Hairy, Ripply1, Bowline, Hucklebein, Runt, and Brinker [26-31], while examples of eh1 motif-containing repressors that recruit Gro include Engrailed, Goosecoid, Dorsal, Knirps, and PRH/Hex [32-37]. As an artificially recruited Gro lacking the WD-repeat domain can repress reporter expression to the same degree as wild type Gro, and it thus appears that this domain is dispensable for repression activity [38].

### **The intrinsically disordered domains of Gro in protein-protein interactions**

The central domains of Gro were shown to be disordered and yet essential for repression function. To aid in an understanding of these domains, I will therefore provide a brief introduction to intrinsically disordered protein domains.

Protein-protein interactions are essential to the function of proteins. Some interactions are needed for protein activity, while others regulate protein activity in response to environmental clues. Protein interactions are generally believed to follow the

“lock and key” model, in which proteins assume a unique 3D shape before coming together [39]. Disordered regions were thought to serve as flexible linkers between structured domains and to lack their own specific functions. However, increasing evidence suggests some disordered regions, termed intrinsically disordered regions (IDR), exert critical roles in biological processes [40, 41].

IDR lack rigid 3D structures under native conditions, but undergo a disordered to ordered transition when they bind to interacting partners (Figure 1-2) [42]. These interactions can be highly specific but reversible, allowing proteins to serve as regulatory hubs and link multiple processes together. They can serve as hubs of large regulatory networks by mediating a wide array of highly specific protein interactions [40].

IDR can be identified by computational prediction and extensive work in crystallization or NMR spectrometry. Disordered regions failed to scatter X-rays coherently, resulting in missing electron density in crystal structures [42]. NMR spectrometry is useful for uncovering the existence of transient tertiary structure by measuring the shift in conformational states [43]. Disordered regions are usually deficient in bulky hydrophobic amino acids, which are frequently associated with the hydrophobic cores of globular domains [41, 44]. Early attempts to discover IDR looked for consecutive charged residues in the primary structure, and predicted that about 25-41% of eukaryotic proteins have disordered regions of over 50 amino acids [45]. Recently, more advanced software has been developed to predict disordered regions using a combination of

primary and tertiary structure alignment [42]. It is based on the assumption that the position of an ordered residue should be conserved within multiple tertiary structures, while disordered residues will demonstrate a variation in position when multiple tertiary structures are aligned [46].

IDR are commonly found in eukaryotic transcription factors [47, 48]. They have been shown to play important roles in transcriptional regulation in several aspects. First, some DNA binding domains like the AT hook and basic motif are intrinsically disordered and go through a disorder to order transition when they interact with DNA [47, 49]. Second, they can modulate the affinity of DNA binding domain for different DNA sequences, which allow the same transcription factor to bind a variety of sequences with context dependent specificity [50]. Third, they can recruit co-regulators and co-factors that regulate the activity of transcription factors, such as the recruitment of the co-activator CBP to HIF- $\alpha$  or the recruitment of the initiation complex to p53 to initiate transcription [51, 52]. Since Gro itself does not bind DNA, it is likely that the central domains regulate Gro activity by recruiting co-regulators and co-factors.

### **Mechanisms of Gro-mediated repression**

One way to classify a repressor is by whether it mediates long-range or short-range repression (Figure 1-3). By definition, long-range repression refers to a repressor that makes a promoter resistant to the influence of all enhancers, even if the repressor binding sites are thousands of base pairs away from the enhancers. On the other hand, short-

range repression refers to a repressor that blocks the function of nearby enhancers without interfering more distant enhancers, thus maintaining enhancer autonomy [53, 54]. Gro was initially classified as long-range repressor as it can repress gene expression indiscriminately over long distances [9, 55]. However, recent studies suggested that it can also mediate short-range repression [32, 36].

Three non-exclusive models have been proposed for Gro-mediated repression (Figure 1-4) [56]. The first model is suggested by the ability of the Q domain to mediate Gro oligomerization, the interaction of HDAC1 with the GP domain, and the affinity of Gro for deacetylated histone tails. In this model, repressors recruit Gro through its WD-repeat domain, and Gro then recruits HDAC1, which induces local histone deacetylation. Oligomerization via the Q domain and binding of Gro to deacetylated histone tails recruit additional Gro to the region, and this results in a higher order chromatin structure that is inaccessible to transcription machinery. The second model is suggested by the interaction of Tup1 (a yeast analog of Gro) with several components of the Mediator complex [57-59]. Tup1 may prevent the Mediator complex from interacting with other transcription factors to assemble the pre-initiation complex. Additional evidence supporting interaction between Gro and Mediator came from a genetic study in *C. elegans*, in which combined mutation of UNC-37 (a homolog of Gro) and Mediator components resulted in a synergistic effect on the development of adult male tail [60]. The third model is suggested by interactions of Grg4 with transcription factor Pax2 [61]. Binding of Grg4 is thought to

mask the Pax2 activation domain, preventing the phosphorylation of the activation domain that is necessary for transcription activation.

Given the various results presented in literature, Gro probably silences by multiple mechanisms, including both histone deacetylase dependent and independent mechanisms. In particular, it appears that the mechanism of repression may vary depending upon the repressor and the target gene. In addition to utilizing different mechanism of repression, Gro activity can be regulated in several ways, which are discussed in the next section.

### **Functional roles of Gro**

#### *Cell signaling*

Gro participates in a number of signaling pathways (Figure 1-5) [62]. The repression activity of Gro in these pathways are regulated through various means. In this section, I will briefly discuss Gro involvement in these signaling pathways with an emphasis on its regulation.

During neurogenesis, Gro interacts with Hairless and Suppressor of Hairless to repress the expression of E(Spl) repressors in the absence of the Notch signal [63, 64]. Upon activation, the intracellular domain of the Notch receptor displaces Gro and allows expression of E(Spl) repressors, which then recruit Gro to repress expression of proneural genes [65]. Thus, the Notch signaling pathway provides an example of Gro recruitment to

different target genes through the availability of repressors. In addition, phosphorylation of the SP domain by MAPK attenuates Gro repression in the Notch pathway [23]. This provides another means of regulation through post-translational modification. Moreover, it demonstrates an example of crosstalk between the EGFR and Notch pathways through the regulation of Gro activity.

Gro also negatively regulates the Wnt signaling pathway. In the absence of the Wnt signal, Gro binds to Tcf/Lef to repress target gene expression. Activation of signaling results in  $\beta$ -catenin displacing Gro and activating target gene expression [18]. It should be noted that while  $\beta$ -catenin and Gro are mutually exclusive in ChIP [66],  $\beta$ -catenin does not reduce the affinity of Gro to Tcf/Lef [12]. These data argue against a simple model of direct competition between Gro and  $\beta$ -catenin, and affinity of Gro for its repressors is likely to play a role in deciding whether target genes will be expressed.

Finally, Gro is recruited by Brinker to repress wing patterning genes when Dpp signaling is inactive [27]. Brinker also has a separate motif to recruit another co-repressor CtBP. While some target genes require both co-repressors to be present for full repression, others can be repressed by Gro alone [27]. Thus, this demonstrates a way to regulate repression activity by cooperation between different co-repressors.

Given the involvement of Gro in various pathways, it is not surprising that Gro plays important roles in both normal development and in disease. In *Drosophila*, Gro is essential in sex determination and neurogenesis [56, 67]. The vertebrate homolog also acts in

myogenesis and haematopoiesis. [11]. In addition, increased Gro function correlates with the appearance of certain forms of cancer including lung cancer [5, 68].

### *Embryonic patterning*

During embryonic patterning, cells rely on morphogen gradients, which provide positional information. The concentration of a morphogen determines recruitment of transcription factors, which alters gene expression pattern and cell fate (Figure 1-6). For example, in anterior-posterior patterning of the *Drosophila* blastoderm, the morphogen Bicoid is produced at the anterior end and diffuses posteriorly, forming a gradient with the highest concentration at the anterior end. Bicoid activates the transcription of a number of target genes, including the repressors sloppy-paired 1, giant, and hunchback, which will eventually determine the development of cephalic, thoracic, and abdominal segments [69, 70].

Gro has been known to play important roles in patterning of the embryo and imaginal discs. In dorsoventral patterning, Gro interacts with the morphogen Dorsal to restrict expression of *dpp* and *zen* to the dorsal region [71]. In patterning of the unsegmented termini, Gro interacts with Capicua to restrict the expression of *hkb* and *tll* to embryonic termini [72, 73]. In anteroposterior patterning of *Drosophila* wing, Gro represses expression of hedgehog and engrailed in the anterior compartment, hence restricting expression of the morphogen Dpp to the posterior compartment [74].



## Splicing machinery in transcription

An updated version of the central dogma of molecular biology states that genes are transcribed into pre-mRNA, which is then spliced into mRNA and translated into proteins. While each step in the process has been studied separately in great details, it has been known that they are not independent events [75-77]. Splicing often occurs co-transcriptionally [78], and transcription factors frequently associate with splicing factors [79-82]. In this context, the finding to be presented in chapter 2 that Gro interacts with certain splicing factors is not completely surprising. In this section, I will therefore provide a brief introduction to the current evidence for interactions between transcription and splicing factors.

It is widely accepted that splicing events, such as recognition of splice sites by spliceosome components, occurs co-transcriptionally when RNA Polymerase II (Pol II) is synthesizing nascent pre-mRNA [83, 84]. The central focus of this process is the COOH-terminal repeat domain (CTD) of the Pol II large subunit, which is a disordered region located at the RNA exit channel [85, 86]. CTD has been shown to interact with various factors that are important for capping, splicing, and polyadenylation of mRNA [87-89]. By positioning splicing factors near the nascent RNA, the spliceosome can assemble more efficiently to recognize splice sites [89]. In addition, splicing decision can be affected by Pol II elongation rate, allowing alternative splicing for weak splice sites to occur when Pol II slows down [90].

Nucleosome density, DNA methylation, and histone modification have been known to play important roles in transcriptional regulation. Recent studies suggest that they may also affect splicing. Gcn5, a well-known histone acetyltransferase that activates gene expression, is shown to affect the recruitment of specific splicing factors to pre-mRNA [91]. Moreover, exons are enriched in nucleosomes and are methylated more frequently than introns [92, 93]. In addition, different promoters can yield transcripts that are subject to differential alternative splicing [94, 95].

While most research has focused on the effect of transcription factors in splicing, there is also increasing evidence that promoter proximal splicing elements can influence transcription. It is well established that promoter proximal introns stimulate transcription [96, 97]. The length of first exon is inversely proportional to the expression level, and removal of introns leads to reduction of H3K4me3, which is usually associated with gene activation [98]. In addition, splicing factors have been shown to associate with transcription factors and stimulate transcriptional initiation [99].

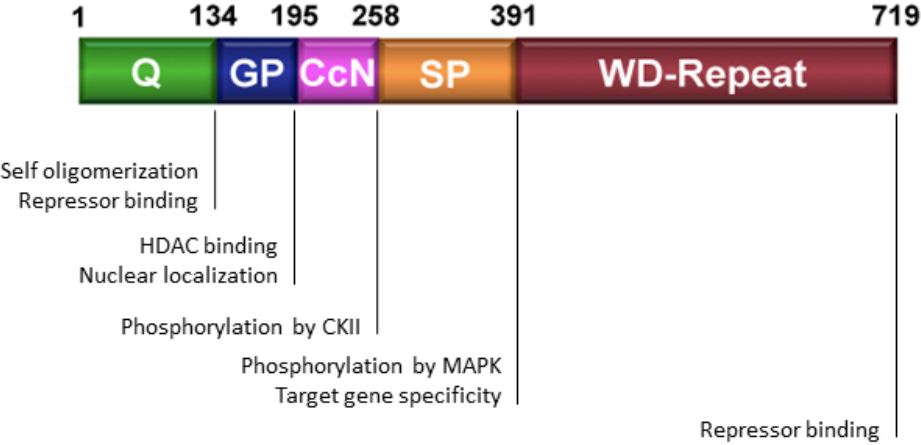
### **Concluding remarks:**

Gro is a conserved corepressor involved in multiple developmental processes. While there are several proposed mechanisms of Gro-mediated repression, they do not fully account for the functional importance of the disordered central domains. Thus, I have attempted to elucidate the mechanism by identifying potential Gro interacting proteins and validating the interactions (Chapter 2). Our results indicate that the central domains

of Gro mediate multiple interactions required for repression and reveal a possible mechanism of Gro-mediated repression through an interaction with the spliceosome complex. In addition, a chimeric Gro was constructed to study the role of histone deacetylase interactions in Gro activity (Chapter 3). Our results corroborate and extend previous findings that Gro interacts with histone deacetylase through its GP domain.

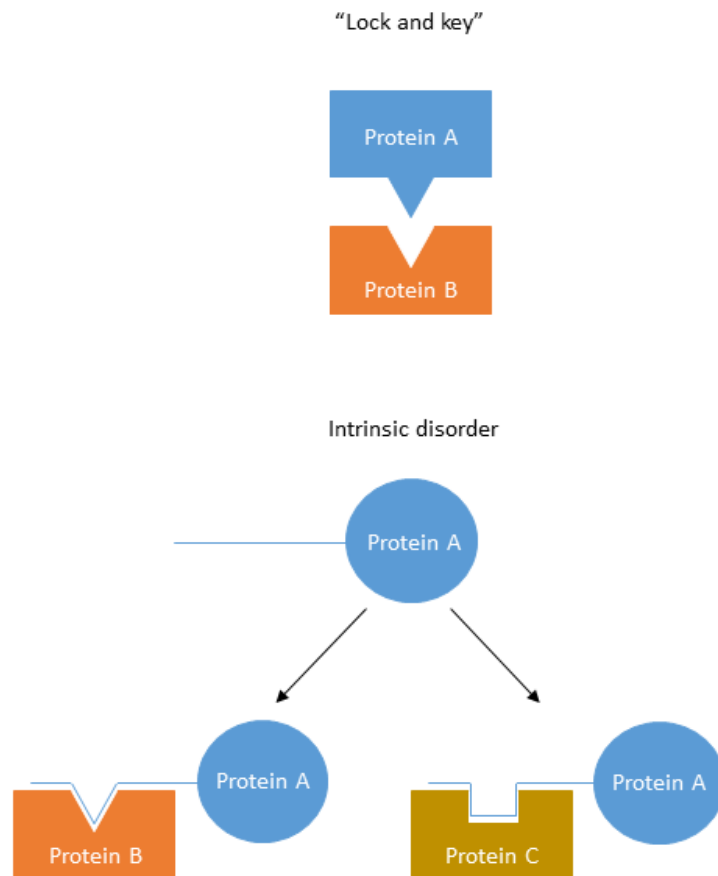
**Figure 1-1. Schematic representation of Gro.** Gro has a five domains structure based on sequence homolog. Refer to main text for detail explanation of each domain.

Figure 1-1



**Figure 1-2. Models of protein-protein interactions.** Protein-protein interactions can be mediated by the matching of rigid tertiary structures (“lock and key” model) or conformational change of the disorder region (intrinsic disorder model).

Figure 1-2



**Figure 1-3. Classification of repression.** Repressor can be classified as long range or short range. In long range repression, such as chromatin modification, the recruitment of repressor will result in repression of all enhancers. In short range repression, the recruitment of repressor will only repress nearby enhancers without affecting more distant enhancers.



**Figure 1-3**

Long range repression



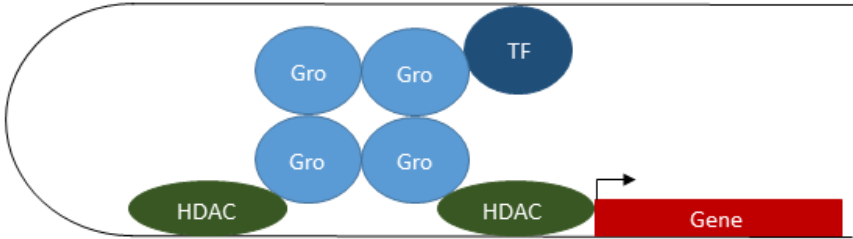
Short range repression



**Figure 1-4. Three proposed models of Gro mediated repression.** (a) Gro self-oligomerization and recruitment of histone deacetylase induced higher order chromatin formation that is inaccessible to transcription machinery. (b) Gro interacts with Mediator to prevent formation of pre-initiation complex. (c) Gro blocks the phosphorylation of Pax2 by JNK that activates expression.

**Figure 1-4**

A. Histone deacetylation and silent state



B. Interaction with Mediator complex



C. Masking of transcriptional activator



**Figure 1-5. Involvement of Gro in various developmental pathways.** In Notch signaling, Gro represses the expression of E(spl) repressors to promote expression of proneural genes. Upon activation, the intracellular domain of Notch receptor (NICD) displaces Gro and Hairless, allowing expression of E(spl) repressors, which then recruit Gro to repress expression of proneural genes. In Wnt signaling, Tcf/Lef interact with Gro through the Q domain to repress target gene expression. Upon activation,  $\beta$ -catenin enters into nucleus and displaces Gro to activate target gene expression. In dpp signaling, brinker recruits Gro and/or the short range repressor CtBP to repress target gene expression. Upon activation, phosphorylated Mad and Medea form a complex to repress expression of brinker [100].

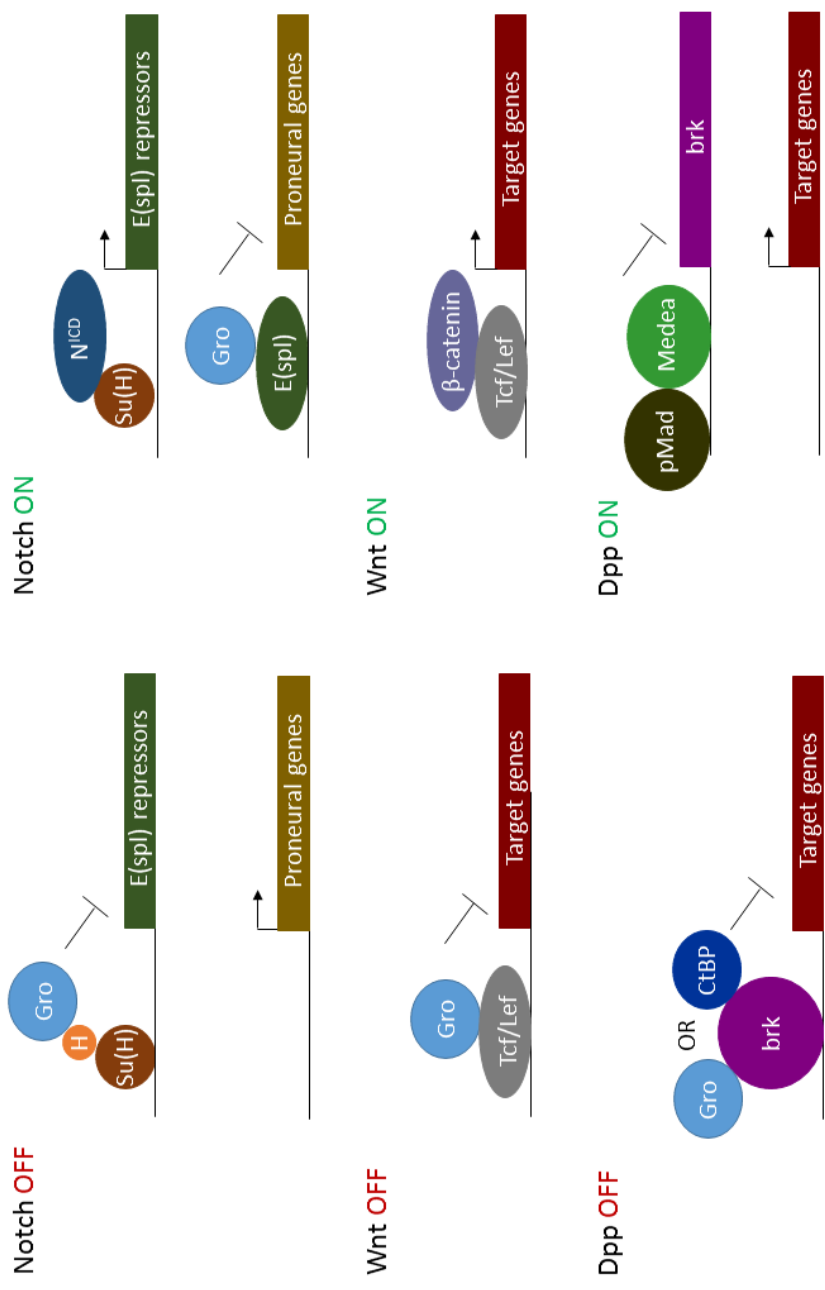
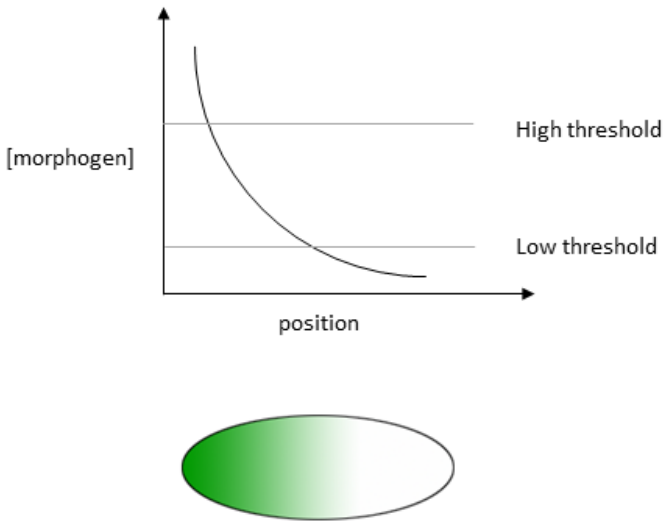


Figure 1-5

**Figure 1-6. Morphogen gradient determines gene expression.** Genes that require high concentration of morphogen (high threshold) will only be expressed near the origin of morphogen, while genes that require low concentration of morphogen (low threshold) can be expressed in more distant position.

Figure 1-6



## References

1. Lindsley, D.L., E.H. Grell, and C.B. Bridges, Genetic variations of *Drosophila melanogaster*. Carnegie Institution of Washington Publication. 1967, Washington? 471 p.
2. Wurmbach, E., I. Wech, and A. Preiss, The Enhancer of split complex of *Drosophila melanogaster* harbors three classes of Notch responsive genes. *Mech Dev*, 1999. 80(2): p. 171-80.
3. Hartley, D.A., A. Preiss, and S. Artavanis-Tsakonas, A deduced gene product from the *Drosophila* neurogenic locus, enhancer of split, shows homology to mammalian G-protein beta subunit. *Cell*, 1988. 55(5): p. 785-95.
4. Stifani, S., et al., Human homologs of a *Drosophila* Enhancer of split gene product define a novel family of nuclear proteins. *Nat Genet*, 1992. 2(2): p. 119-27.
5. Gasperowicz, M. and F. Otto, Mammalian Groucho homologs: redundancy or specificity? *J Cell Biochem*, 2005. 95(4): p. 670-87.
6. Mallo, M., F. Franco del Amo, and T. Gridley, Cloning and developmental expression of *Grg*, a mouse gene related to the groucho transcript of the *Drosophila* Enhancer of split complex. *Mech Dev*, 1993. 42(1-2): p. 67-76.
7. Pflugrad, A., et al., The Groucho-like transcription factor UNC-37 functions with the neural specificity gene *unc-4* to govern motor neuron identity in *C. elegans*. *Development*, 1997. 124(9): p. 1699-709.
8. Choudhury, B.K., et al., Cloning and developmental expression of *Xenopus* cDNAs encoding the Enhancer of split groucho and related proteins. *Gene*, 1997. 195(1): p. 41-8.
9. Smith, R.L. and A.D. Johnson, Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem Sci*, 2000. 25(7): p. 325-30.
10. Liu, Z. and V. Karmarkar, Groucho/Tup1 family co-repressors in plant development. *Trends Plant Sci*, 2008. 13(3): p. 137-44.
11. Buscarlet, M. and S. Stifani, The 'Marx' of Groucho on development and disease. *Trends Cell Biol*, 2007. 17(7): p. 353-61.



12. Chodaparambil, J.V., et al., Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J*, 2014. 33(7): p. 719-31.
13. Pickles, L.M., et al., Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. *Structure*, 2002. 10(6): p. 751-61.
14. Turki-Judeh, W. and A.J. Courey, The unconserved groucho central region is essential for viability and modulates target gene specificity. *PLoS One*, 2012. 7(2): p. e30610.
15. Jennings, B.H., S.M. Wainwright, and D. Ish-Horowicz, Differential in vivo requirements for oligomerization during Groucho-mediated repression. *EMBO Rep*, 2008. 9(1): p. 76-83.
16. Chen, G., P.H. Nguyen, and A.J. Courey, A role for Groucho tetramerization in transcriptional repression. *Mol Cell Biol*, 1998. 18(12): p. 7259-68.
17. Arce, L., K.T. Pate, and M.L. Waterman, Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. *BMC Cancer*, 2009. 9: p. 159.
18. Daniels, D.L. and W.I. Weis, Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol*, 2005. 12(4): p. 364-71.
19. Hanson, A.J., et al., XIAP monoubiquitylates Groucho/TLE to promote canonical Wnt signaling. *Mol Cell*, 2012. 45(5): p. 619-28.
20. Chen, G., et al., A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in *Drosophila* development. *Genes Dev*, 1999. 13(17): p. 2218-30.
21. Winkler, C.J., A. Ponce, and A.J. Courey, Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density. *PLoS One*, 2010. 5(4): p. e10166.
22. Nuthall, H.N., K. Joachim, and S. Stifani, Phosphorylation of serine 239 of Groucho/TLE1 by protein kinase CK2 is important for inhibition of neuronal differentiation. *Mol Cell Biol*, 2004. 24(19): p. 8395-407.
23. Hasson, P., et al., EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. *Nat Genet*, 2005. 37(1): p. 101-5.

24. Helman, A., et al., Phosphorylation of Groucho mediates RTK feedback inhibition and prolonged pathway target gene expression. *Curr Biol*, 2011. 21(13): p. 1102-10.
25. Jennings, B.H., et al., Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell*, 2006. 22(5): p. 645-55.
26. Goldstein, R.E., et al., Hucklebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. *Development*, 1999. 126(17): p. 3747-55.
27. Hasson, P., et al., Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. *EMBO J*, 2001. 20(20): p. 5725-36.
28. Kawamura, A., et al., Groucho-associated transcriptional repressor ripply1 is required for proper transition from the presomitic mesoderm to somites. *Dev Cell*, 2005. 9(6): p. 735-44.
29. Kondow, A., et al., Bowline, a novel protein localized to the presomitic mesoderm, interacts with Groucho/TLE in *Xenopus*. *Int J Dev Biol*, 2006. 50(5): p. 473-9.
30. Nuthall, H.N., et al., Role for Hes1-induced phosphorylation in Groucho-mediated transcriptional repression. *Mol Cell Biol*, 2002. 22(2): p. 389-99.
31. Aronson, B.D., et al., Groucho-dependent and -independent repression activities of Runt domain proteins. *Mol Cell Biol*, 1997. 17(9): p. 5581-7.
32. Andrioli, L.P., et al., Groucho-dependent repression by sloppy-paired 1 differentially positions anterior pair-rule stripes in the *Drosophila* embryo. *Dev Biol*, 2004. 276(2): p. 541-51.
33. Flores-Saaib, R.D., S. Jia, and A.J. Courey, Activation and repression by the C-terminal domain of Dorsal. *Development*, 2001. 128(10): p. 1869-79.
34. Jimenez, G., Z. Paroush, and D. Ish-Horowicz, Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev*, 1997. 11(22): p. 3072-82.
35. Jimenez, G., C.P. Verrijzer, and D. Ish-Horowicz, A conserved motif in gooseoid mediates groucho-dependent repression in *Drosophila* embryos. *Mol Cell Biol*, 1999. 19(3): p. 2080-7.
36. Payankulam, S. and D.N. Arnosti, Groucho corepressor functions as a cofactor for the Knirps short-range transcriptional repressor. *Proc Natl Acad Sci U S A*, 2009. 106(41): p. 17314-9.

37. Swingler, T.E., et al., The proline-rich homeodomain protein recruits members of the Groucho/Transducin-like enhancer of split protein family to co-repress transcription in hematopoietic cells. *J Biol Chem*, 2004. 279(33): p. 34938-47.
38. Fisher, A.L., S. Ohsako, and M. Caudy, The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Mol Cell Biol*, 1996. 16(6): p. 2670-7.
39. Caceres, R.A., et al., Molecular recognition models: a challenge to overcome. *Curr Drug Targets*, 2008. 9(12): p. 1077-83.
40. Dunker, A.K., et al., Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J*, 2005. 272(20): p. 5129-48.
41. Uversky, V.N. and A.K. Dunker, Understanding protein non-folding. *Biochim Biophys Acta*, 2010. 1804(6): p. 1231-64.
42. Atkins, J.D., et al., Disorder Prediction Methods, Their Applicability to Different Protein Targets and Their Usefulness for Guiding Experimental Studies. *Int J Mol Sci*, 2015. 16(8): p. 19040-54.
43. Chen, J., Towards the physical basis of how intrinsic disorder mediates protein function. *Arch Biochem Biophys*, 2012. 524(2): p. 123-31.
44. Vihinen, M., E. Torkkila, and P. Riihonen, Accuracy of protein flexibility predictions. *Proteins*, 1994. 19(2): p. 141-9.
45. Dunker, A.K., et al., Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform*, 2000. 11: p. 161-71.
46. Deng, X., J. Eickholt, and J. Cheng, A comprehensive overview of computational protein disorder prediction methods. *Mol Biosyst*, 2012. 8(1): p. 114-21.
47. Liu, J., et al., Intrinsic disorder in transcription factors. *Biochemistry*, 2006. 45(22): p. 6873-88.
48. Minezaki, Y., et al., Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol*, 2006. 359(4): p. 1137-49.
49. Weiss, M.A., et al., Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, 1990. 347(6293): p. 575-8.

50. Bondos, S.E. and H.C. Hsiao, Roles for intrinsic disorder and fuzziness in generating context-specific function in Ultrabithorax, a Hox transcription factor. *Adv Exp Med Biol*, 2012. 725: p. 86-105.
51. Ema, M., et al., Molecular mechanisms of transcription activation by HLF and HIF1alpha in response to hypoxia: their stabilization and redox signal-induced interaction with CBP/p300. *EMBO J*, 1999. 18(7): p. 1905-14.
52. Uesugi, M. and G.L. Verdine, The alpha-helical FXXPhiPhi motif in p53: TAF interaction and discrimination by MDM2. *Proc Natl Acad Sci U S A*, 1999. 96(26): p. 14801-6.
53. Courey, A.J. and S. Jia, Transcriptional repression: the long and the short of it. *Genes Dev*, 2001. 15(21): p. 2786-96.
54. Gray, S. and M. Levine, Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in *Drosophila*. *Genes Dev*, 1996. 10(6): p. 700-10.
55. Barolo, S. and M. Levine, hairy mediates dominant repression in the *Drosophila* embryo. *EMBO J*, 1997. 16(10): p. 2883-91.
56. Turki-Judeh, W. and A.J. Courey, Groucho: a corepressor with instructive roles in development. *Curr Top Dev Biol*, 2012. 98: p. 65-96.
57. Kuchin, S. and M. Carlson, Functional relationships of Srb10-Srb11 kinase, carboxy-terminal domain kinase CTDK-I, and transcriptional corepressor Ssn6-Tup1. *Mol Cell Biol*, 1998. 18(3): p. 1163-71.
58. Papamichos-Chronakis, M., et al., Hrs1/Med3 is a Cyc8-Tup1 corepressor target in the RNA polymerase II holoenzyme. *J Biol Chem*, 2000. 275(12): p. 8397-403.
59. Gromoller, A. and N. Lehming, Srb7p is a physical and physiological target of Tup1p. *EMBO J*, 2000. 19(24): p. 6845-52.
60. Zhang, H. and S.W. Emmons, *Caenorhabditis elegans* unc-37/groucho interacts genetically with components of the transcriptional mediator complex. *Genetics*, 2002. 160(2): p. 799-803.
61. Cai, Y., et al., Groucho suppresses Pax2 transactivation by inhibition of JNK-mediated phosphorylation. *EMBO J*, 2003. 22(20): p. 5522-9.

62. Cinnamon, E. and Z. Paroush, Context-dependent regulation of Groucho/TLE-mediated repression. *Curr Opin Genet Dev*, 2008. 18(5): p. 435-40.
63. Paroush, Z., et al., Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell*, 1994. 79(5): p. 805-15.
64. Grbavec, D. and S. Stifani, Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. *Biochem Biophys Res Commun*, 1996. 223(3): p. 701-5.
65. Delidakis, C., et al., Two genetically and molecularly distinct functions involved in early neurogenesis reside within the Enhancer of split locus of Drosophila melanogaster. *Genetics*, 1991. 129(3): p. 803-23.
66. Sierra, J., et al., The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes. *Genes Dev*, 2006. 20(5): p. 586-600.
67. Chen, G. and A.J. Courey, Groucho/TLE family proteins and transcriptional repression. *Gene*, 2000. 249(1-2): p. 1-16.
68. Allen, T., et al., Grg1 acts as a lung-specific oncogene in a transgenic mouse model. *Cancer Res*, 2006. 66(3): p. 1294-301.
69. Porcher, A. and N. Dostatni, The bicoid morphogen system. *Curr Biol*, 2010. 20(5): p. R249-54.
70. Briscoe, J. and S. Small, Morphogen rules: design principles of gradient-mediated embryo patterning. *Development*, 2015. 142(23): p. 3996-4009.
71. Dubnicoff, T., et al., Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. *Genes Dev*, 1997. 11(22): p. 2952-7.
72. Jimenez, G., et al., Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. *Genes Dev*, 2000. 14(2): p. 224-31.
73. Paroush, Z., S.M. Wainwright, and D. Ish-Horowicz, Torso signalling regulates terminal patterning in Drosophila by antagonising Groucho-mediated repression. *Development*, 1997. 124(19): p. 3827-34.

74. de Celis, J.F. and M. Ruiz-Gomez, groucho and hedgehog regulate engrailed expression in the anterior compartment of the *Drosophila* wing. *Development*, 1995. 121(10): p. 3467-76.
75. Maniatis, T. and R. Reed, An extensive network of coupling among gene expression machines. *Nature*, 2002. 416(6880): p. 499-506.
76. Merkhofer, E.C., P. Hu, and T.L. Johnson, Introduction to cotranscriptional RNA splicing. *Methods Mol Biol*, 2014. 1126: p. 83-96.
77. de Almeida, S.F. and M. Carmo-Fonseca, Reciprocal regulatory links between cotranscriptional splicing and chromatin. *Semin Cell Dev Biol*, 2014. 32: p. 2-10.
78. Khodor, Y.L., et al., Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev*, 2011. 25(23): p. 2502-12.
79. Monsalve, M., et al., Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell*, 2000. 6(2): p. 307-16.
80. Ge, H., Y. Si, and A.P. Wolffe, A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol Cell*, 1998. 2(6): p. 751-9.
81. Hallier, M., et al., The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *J Biol Chem*, 1998. 273(9): p. 4838-42.
82. Davies, R.C., et al., WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev*, 1998. 12(20): p. 3217-25.
83. Goldstrohm, A.C., A.L. Greenleaf, and M.A. Garcia-Blanco, Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene*, 2001. 277(1-2): p. 31-47.
84. Napolitano, G., L. Lania, and B. Majello, RNA polymerase II CTD modifications: how many tales from a single tail. *J Cell Physiol*, 2014. 229(5): p. 538-44.
85. Chen, C.Y., et al., Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci U S A*, 2009. 106(1): p. 127-32.
86. Cramer, P., D.A. Bushnell, and R.D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 2001. 292(5523): p. 1863-76.

87. Schwer, B. and S. Shuman, Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell*, 2011. 43(2): p. 311-8.
88. Mischo, H.E. and N.J. Proudfoot, Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta*, 2013. 1829(1): p. 174-85.
89. David, C.J., et al., The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev*, 2011. 25(9): p. 972-83.
90. Munoz, M.J., et al., DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell*, 2009. 137(4): p. 708-20.
91. Gunderson, F.Q. and T.L. Johnson, Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet*, 2009. 5(10): p. e1000682.
92. Amit, M., et al., Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep*, 2012. 1(5): p. 543-56.
93. Chodavarapu, R.K., et al., Relationship between nucleosome positioning and DNA methylation. *Nature*, 2010. 466(7304): p. 388-92.
94. Cramer, P., et al., Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*, 1999. 4(2): p. 251-8.
95. Cramer, P., et al., Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A*, 1997. 94(21): p. 11456-60.
96. Damgaard, C.K., et al., A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*, 2008. 29(2): p. 271-8.
97. Furger, A., et al., Promoter proximal splice sites enhance transcription. *Genes Dev*, 2002. 16(21): p. 2792-9.
98. Bieberstein, N.I., et al., First exon length controls active chromatin signatures and transcription. *Cell Rep*, 2012. 2(1): p. 62-8.
99. Kwek, K.Y., et al., U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol*, 2002. 9(11): p. 800-5.

100. Affolter, M. and K. Basler, The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. *Nat Rev Genet*, 2007. 8(9): p. 663-74.



## **Chapter 2**

### **The Central Region of the *Drosophila* Co-Repressor Groucho as a Regulatory Hub**

In this chapter I present data from a publication that identified Gro interacting proteins. The functional relevance of these proteins on Gro-mediated repression was investigated through a reporter assay and transcriptome analyses of RNAi treated cells. I carried out the purification of Gro interacting proteins presented in figure 1, the co-IP experiment presented in figure 2, the development and execution of the reporter assay presented in figure 3, and the sample preparation of RNAi treated cells presented in figure 4. This research was originally published in the Journal of biological chemistry 2015;290(50):30119-30130.

## The Central Region of the *Drosophila* Co-repressor Groucho as a Regulatory Hub<sup>\*[5]</sup>

Received for publication, July 28, 2015, and in revised form, October 12, 2015. Published, JBC Papers in Press, October 19, 2015, DOI 10.1074/jbc.M115.681171

Pak N. Kwong<sup>†</sup>, Michael Chambers<sup>‡</sup>, Ajay A. Vashisht<sup>§</sup>, Wiam Turki-Judeh<sup>†¶</sup>, Tak Yu Yau<sup>‡</sup>, James A. Wohlschlegel<sup>§¶</sup>, and Albert J. Courey<sup>†¶1</sup>

From the Departments of <sup>†</sup>Chemistry and Biochemistry and <sup>§</sup>Biological Chemistry and <sup>¶</sup>Molecular Biology Institute, UCLA, Los Angeles, California 90095

**Background:** The co-repressor Groucho has an essential, but disordered, central region.

**Results:** We identified over 160 central region-binding proteins, many of which, including components of the spliceosome, modulate Groucho-mediated repression.

**Conclusion:** Groucho regulates transcription by multiple mechanisms and may link the transcriptional and splicing machineries.

**Significance:** Its central region may serve as the hub of a regulatory network.

Groucho (Gro) is a *Drosophila* co-repressor that regulates the expression of a large number of genes, many of which are involved in developmental control. Previous studies have shown that its central region is essential for function even though its three domains are poorly conserved and intrinsically disordered. Using these disordered domains as affinity reagents, we have now identified multiple embryonic Gro-interacting proteins. The interactors include protein complexes involved in chromosome organization, mRNA processing, and signaling. Further investigation of the interacting proteins using a reporter assay showed that many of them modulate Gro-mediated repression either positively or negatively. The positive regulators include components of the spliceosomal subcomplex U1 small nuclear ribonucleoprotein (U1 snRNP). A co-immunoprecipitation experiment confirms this finding and suggests that a sizable fraction of nuclear U1 snRNP is associated with Gro. The use of RNA-seq to analyze the gene expression profile of cells subjected to knockdown of Gro or snRNP-U1-C (a component of U1 snRNP) showed a significant overlap between genes regulated by these two factors. Furthermore, comparison of our RNA-seq data with Gro and RNA polymerase II ChIP data led to a number of insights, including the finding that Gro-repressed genes are enriched for promoter-proximal RNA polymerase II. We conclude that the Gro central domains mediate multiple interactions required for repression, thus functioning as a regulatory hub. Furthermore, interactions with the spliceosome may contribute to repression by Gro.

Groucho (Gro)<sup>2</sup> is a conserved metazoan co-repressor that may be particularly critical for long range repression whereby repressors are able to establish large transcriptionally silent domains that can spread over many thousands of base pairs (1–3). Gro is essential in many developmental processes, including sex determination, neurogenesis, and pattern formation in *Drosophila* as well as myogenesis and hematopoiesis in vertebrates (2, 4, 5). Gro also has roles in multiple signal transduction pathways, including the Ras and Notch pathways (6–8). Furthermore, increased Gro activity correlates with the appearance of certain forms of cancer such as lung cancer (9, 10). Thus, understanding the mechanism of Gro-mediated repression should contribute to our understanding of long range repression and its role in development, signaling, and disease.

Sequence comparison of Gro family proteins reveals five domains (2, 10). The C-terminal WD repeat domain forms a  $\beta$ -propeller that interacts with the WRPW and eh1 motifs found in many Gro-dependent DNA-binding repressors (11). The N-terminal Q domain folds into a coiled coil structure that forms tetramers and perhaps higher order oligomers, and this self-association is required for robust repression (12–15). The central GP, CcN, and SP domains are believed to have essential functions even though their primary sequences are not well conserved. The GP domain interacts with the histone deacetylase Rpd3/HDAC1 (16, 17). Histone deacetylation is broadly associated with gene silencing, and treatment of flies with histone deacetylase inhibitors attenuates Gro-mediated repression (18). In addition, the GP domain is essential for nuclear localization because deletion of this domain prevents Gro nuclear uptake (19). The SP domain regulates Gro function negatively as its deletion leads to promiscuous repression and

\* This work was supported by National Institutes of Health Grants GM44522 (to A. J. C.) and GM089778 (to J. A. W.). The authors declare that they have no conflict of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[5] This article contains supplemental Tables S1 and S2.

<sup>1</sup> To whom correspondence should be addressed: Dept. of Chemistry and Biochemistry, UCLA, 607 Charles E. Young Dr. East, Los Angeles, CA 90095-1569. Tel.: 310-825-2530; E-mail: courey@chem.ucla.edu.

<sup>2</sup> The abbreviations used are: Gro, Groucho; MudPIT, multidimensional protein identification technology; snRNP, small nuclear ribonucleoprotein; seq, sequencing; Pol II, RNA polymerase II; qPCR, quantitative PCR; dsRNA, double-stranded RNA; CKII, casein kinase II; ACF, ATP-dependent chromatin remodeling and assembly factor; Su(H), Suppressor of Hairless; Brk, Brinker; UAS, upstream activating sequence; EWS/FLI, Ewing sarcoma breakpoint region 1/Friend leukemia integration 1 transcription factor.

## Groucho/Spliceosome Interactions

**TABLE 1**  
PCR primers used in construction of plasmids encoding GST fusion proteins

Domain	Sequence
Q	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAATATCCCTCACCGGTGCGCCACCCC ATTATTACTCGAGTCACTGCTGGGCGTGGATCTGTGCCCA
GP	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGTGCCAGGTGGACCACCTCAGCCGA ATTATTACTCGAGTCAAGAAATTGAGCAATCGCTCCGCG
CcN	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGTTTCGCCGCGCATCGTGAAGAAT ATTATTACTCGAGTCAATAGACACGTCGCTCGCCGTTGGGA
SP	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGAGGTGCGCGATCGGGAAAGCTTGA ATTATTACTCGAGTCAACCCGTTAGGCGCGAGGGATGTGGA

developmental defects (19). Phosphorylation of the SP domain by Ras/MAPK signaling was shown to attenuate repression, providing a mechanism for regulating repression in response to environmental cues (20). Finally, the CcN domain is also targeted for phosphorylation by protein kinases and is required for repression by Gro (19, 21).

Sequence analysis of the Gro central domains strongly suggests that they are intrinsically disordered (19). Intrinsically disordered regions in proteins lack rigid three-dimensional structures under native conditions and can serve as hubs of large regulatory networks by mediating a wide array of highly specific protein interactions (22, 23). Increasing evidence suggests that intrinsically disordered domains have critical functions in transcriptional regulation (24, 25).

In this study, we set out to illuminate the mechanisms of Gro-mediated repression by identifying proteins that interact with the N-terminal Q domain and the three central domains. A proteomic screen revealed over 160 interacting proteins, many of which are components of protein complexes in a variety of functional categories such as chromatin remodeling and RNA processing. Perhaps most notably, the interactors included multiple components of the spliceosome, and a co-immunoprecipitation experiment suggests that a sizable fraction of U1 snRNP (a subcomplex of the spliceosome) is associated with Gro in embryonic nuclei.

As a means of systematically validating the functional significance of these interactions, we carried out a novel reporter assay using three different luciferase reporters that could be monitored simultaneously. These assays showed that many of the interacting proteins, including the protein components of U1 snRNP, are required for optimal Gro-mediated repression. Lastly, we compared the effects on the gene expression profile of Gro and U1 snRNP knockdown, finding a significant overlap in the regulated genes. Our results indicate that the central domains of Gro mediate multiple interactions required for repression and reveal a possible mechanism of Gro-mediated repression through an interaction with the spliceosome complex or subcomplexes. This reinforces previous studies suggesting that the spliceosome has roles in transcriptional regulation in addition to its roles in RNA processing (26–30).

### Experimental Procedures

**Plasmids**—To generate plasmids for expression of glutathione *S*-transferase (GST) fusion proteins, sequences encoding the Gro domains were amplified by PCR and inserted between the BamHI and XhoI sites of pGEX4T (GE Healthcare). The Q domain included Gro amino acids 1–133, the GP domain included amino acids 134–194, the CcN domain included

amino acids 195–257, and the SP domain included amino acids 258–390. Sequences of PCR primers are provided in Table 1.

Plasmids used in the reporter assay were generated as follows. The red luciferase plasmid, G5DE5-pCBR, was generated by inserting the G5 DE5 enhancer region (14) into pCBR-Basic vector (Promega catalog number E1411) between the KpnI and XhoI sites. The green luciferase plasmid, DE5G5-pCBG68, was generated by inserting the luciferase gene using NcoI and Sall from pCBG68-basic vector (Promega catalog number E1431) into the DE5 G5 vector, which has UAS elements downstream of the reporter.<sup>3</sup> Actin promoter-driven Dorsal (pPac DI), Twist (pPac Twi), and Gal4-Gro (pAct Gal4-Gro) plasmids have been described previously (14). The RplIII128 promoter-driven *Renilla* luciferase plasmid, RplIII128-Rluc, was obtained from Addgene (ID number 37380) (31).

**Affinity Purification and Identification of Gro-interacting Proteins**—Plasmids encoding the recombinant domains fused to GST or GST alone were transformed into BL21 cells. 250 ml of midlog cells were induced with 0.25 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside for an hour. Cells were pelleted at 4,000  $\times$  g, resuspended in 25 ml of salty TE (0.15 M NaCl, 10 mM Tris, pH 8, 1 mM EDTA) with protease inhibitor (Life Technologies catalog number 88266), and incubated on ice for 30 min. Samples were incubated at 4 °C for 15 min after DTT and Triton X-100 were added to final concentrations of 5 mM and 1%, respectively. Cells were then disrupted through a microfluidizer (Microfluidics M110L) using standard conditions. The lysate was collected and centrifuged at 14,000  $\times$  g for 10 min at 4 °C. Supernatant was collected, and 1 ml of glutathione-agarose resin (50% slurry) was added. After overnight incubation, the resin was washed with ice-cold PBS three times and stored at 4 °C.

**Drosophila** embryo nuclear extracts were prepared as described previously (32). To isolate Gro-interacting proteins, 20  $\mu$ g of glutathione bead-immobilized recombinant domains was mixed with nuclear extract containing 30 mg of protein (20 mg/ml) in 8 ml of HEMNK buffer (40 mM HEPES, pH 7.5, 5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 1 mM DTT, 0.5% Nonidet P-40, 0.1 M KCl) at 4 °C overnight. Samples were washed six times for 15 min with 5 ml of HEMNK buffer. Proteins were first eluted with 5 ml of 2 M NaCl in HEMNK buffer and then with 2.5 ml of 2 M NaCl in HEMNK buffer for 20 min each. Eluted proteins were subjected to TCA precipitation prior to multidimensional protein identification technology (MudPIT) analysis. MudPIT analysis was performed as described previously (33). Peptide

<sup>3</sup> A. J. Courey, unpublished data.

**TABLE 2**  
RT-qPCR primers

Gene	Sequence
Gro	TTTATTACAACATGTTTCGAAATCATGC TTCGCTTTTGTATGCGTTGCTAC
snRNP-U1-C	CTCAGGAACGGCATCAACGTT TATAATTAATTGTTTTCGCTATCGGG
Rpl32	CCCAAGGGTATCGACAACAGA CGATTCGCCGCGAGTAAAC
U1 snRNA	ATACTTACCTGGCGTAGAGGTTAACC AACGCCATTCGCCGCTA

identifications were filtered using a false discovery rate cutoff of 0.05 as determined by the decoy database approach. Protein level false positive rates were less than 0.03 for all individual runs.

Supplemental Table S1C includes all the mass spectroscopy data for the two independent replicate screens carried out with each GST fusion protein and GST alone, and supplemental Tables S1, A and B includes selective data for 159 proteins that were detected in both replicates as well as three proteins (Histone H3, Caf1, and Bic) that were only detected in one replicate but for which other data confirm the significance of the interaction (supplemental Table S1B, notes 2 and 3). Ribosomal proteins were excluded from the lists in supplemental Table S1, A and B.

**Gro Immunoprecipitation and Reverse Transcription-qPCR (RT-qPCR) Analysis of U1 snRNA**—500  $\mu$ g of nuclear extract was incubated with 1.875  $\mu$ g of affinity-purified rabbit antibody against the Gro GP domain or rabbit IgG in a final volume of 250  $\mu$ l of HEMNK buffer overnight at 4 °C. 225  $\mu$ g of Protein A Dynabeads (Invitrogen catalog number 10001D) were incubated with the samples at 4 °C for 1 h. Samples were then washed with HEMNK buffer three times for 10 min each. For RT-qPCR, RNA was eluted in 10  $\mu$ l of water by heating to 80 °C for 2 min. Samples were treated with DNase I according to the manufacturer's protocol (Promega catalog number M6101). Reverse transcription was performed with 300 ng of random primer (Invitrogen catalog number 48190-011), and qPCR was performed using primers amplifying U1 snRNA (Table 2). Threshold cycle values were converted to percent input values by comparison with a standard curve generated from multiple serial dilutions of RNA isolated by TRIzol (Life Technologies catalog number 10296010) extraction from the input nuclear extract. Primer specificity was validated by melting curve analysis of the amplification products (data not shown).

For immunoblotting, samples were eluted in SDS-PAGE loading buffer. Proteins were detected with a mixture of mouse anti-Gro (Developmental Studies Hybridoma Bank; 1:650 dilution) and affinity-purified rabbit anti-GP domain (1:100 dilution) antibodies. Immunoblots were subsequently probed with goat anti-mouse 680 and goat anti-rabbit 800 IRDye-coupled secondary antibodies (LI-COR) and imaged with a LI-COR Odyssey imager.

**Three-reporter Luciferase Assay**—To guard against off-target effects, each candidate gene was knocked down with three non-overlapping dsRNAs when possible (the complete list of dsRNAs used is available upon request). Each dsRNA was tested in triplicate. dsRNA was synthesized by the *Drosophila* RNAi Screening Center and realiquoted into white flat bottom

96-well plates (USA Scientific catalog number CC7682-7968) at 150 ng/well in 10  $\mu$ l of water using a Beckman Coulter BioMek FX work station.

Transfections were carried out with Effectene reagent (Qiagen catalog number 301425). 6  $\mu$ g each of G5DE5-pCBR and DE5G5-pCBG68, 0.6  $\mu$ g of RpIII128-Rluc, 1  $\mu$ g of pPac D1, 0.3  $\mu$ g of pPac Twi, and 1.2  $\mu$ g of pAct Gal4-Gro were suspended in 600  $\mu$ l of buffer EC. 33  $\mu$ l of this mixture was added to 25  $\mu$ l of enhancer. After 2–3 min, 7.5  $\mu$ l of Effectene was added and mixed by pipetting up and down. 6  $\mu$ l of this mixture was immediately added into each well of a 96-well plate containing 150 ng of dsRNA. 4–8 min later, 100  $\mu$ l of S2 cells (diluted to  $1 \times 10^6$  cell/ml) was added to each well. Cells were incubated at 24 °C for 2 days before assaying.

The luminescence signal was measured with a Molecular Devices LJJ Analyst HT microplate reader using emission filters ET510/80m and E610LP (Chroma catalog numbers S-022658 and 138951). 50  $\mu$ l of D-luciferin (Chroma-Glo system, Promega catalog number E2980) was added to each well. Five minutes later, the reaction was stopped by the addition of 50  $\mu$ l of stop buffer containing coelenterazine (Dual-Luciferase system, Promega catalog number E1960). The luminescence signal was measured immediately without applying a filter.

To address the issue of signal overlap, raw signals were subjected to filter correction. The corrected red luminescence signal,  $R'$ , and green luminescence signal,  $G'$ , were calculated according to the following equations.

$$R' = \frac{\text{Lrf} - \text{Lgf} \times \left(\frac{\text{Grf}}{\text{Ggf}}\right)}{\left(\frac{\text{Rrf}}{R}\right) - \left(\frac{\text{Rgf}}{R}\right) \times \left(\frac{\text{Grf}}{\text{Ggf}}\right)} \quad (\text{Eq. 1})$$

$$G' = \frac{\text{Lgf} - R' \times \left(\frac{\text{Rgf}}{R}\right)}{\left(\frac{\text{Ggf}}{G}\right)} \quad (\text{Eq. 2})$$

Parameters were determined by expressing the individual luciferases and recording the luminescence signals with red and green filters and with no filter (data not shown). The ratio of green signal passed through the red filter,  $\text{Grf}/\text{Ggf}$ , was determined to be 0.0975; the ratio of red signal passed through the red filter,  $\text{Rrf}/R$ , was determined to be 0.42; the ratio of red signal passed through the green filter,  $\text{Rgf}/R$ , was determined to be 0; and the ratio of green signal passed through the green filter,  $\text{Ggf}/G$ , was determined to be 0.47. Lrf and Lgf are luminescence signals in which cells are co-transfected with both red and green luciferases. Lrf is the signal recorded with the red filter, and Lgf is the signal recorded with the green filter.

The signal from untransfected cells was then subtracted from the corrected data to eliminate background. Processed data were then normalized to the internal control *Renilla* luciferase. Finally, data were compared with the signal from cells in the same plate that were treated with control GFP dsRNA. A change in long or short range repression was considered significant if the  $p$  value was  $<0.1$ . If multiple dsRNAs were tested for



## Groucho/Spliceosome Interactions

a given gene (as was true in most cases; supplemental Table S2), then a change is only listed if the  $p$  value was  $<0.1$  for at least two separate dsRNAs.

**RNA-seq Library Preparation**—Gro dsRNA was generated by PCR amplification of the first 800 nucleotides of the coding sequence using primers containing T7 promoters followed by *in vitro* transcription with T7 RNA polymerase. snRNP-U1-C dsRNA was generated by PCR and *in vitro* transcription of the snRNP-U1-C coding sequencing with primers 5'-taatacactcactatagggtactCAAAGTACTATTGCGACTACTGC and 5'-taatacactcactatagggtactCTTGGGTCGGTTCATGATTCC (lowercase letters represent the T7 promoter sequences). Transfection was carried out as described previously (34). RT-qPCR was used to determine the knockdown efficiency prior to RNA-seq library preparation. RT-qPCR primers targeted the 3'-UTRs of Gro and snRNP-U1-C. Rpl32 was used as a reference gene. The specificity of all primers was validated by melting curve analysis of the amplification products (data not shown). Sequences of the qPCR primers are listed in Table 2.

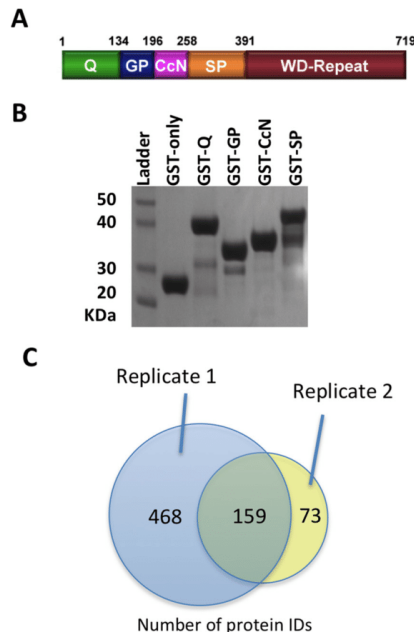
Total RNA was extracted with TRIzol according to the manufacturer's protocol. RNA integrity was determined with an Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent catalog number 5067-1511). Isolation of mRNA was carried out as follows. Streptavidin magnetic beads (Promega catalog number Z5481) were prepared in aliquots of 120 and 60  $\mu$ l in  $0.5\times$  SSC with 10 mM EDTA. 15  $\mu$ g of total RNA was mixed with 1.5  $\mu$ M biotinylated 15-mer poly(T) oligonucleotide in  $0.5\times$  SSC with 10 mM EDTA. Samples were first incubated at 75  $^{\circ}$ C for 5 min followed by 15  $^{\circ}$ C for 10 min and 10  $^{\circ}$ C for 10 min. Samples were then incubated with 120  $\mu$ l of magnetic beads at 4  $^{\circ}$ C for 2 h followed by 60  $\mu$ l of magnetic beads at 4  $^{\circ}$ C for 30 min. The two aliquots of beads were combined and washed four times with 300  $\mu$ l of ice-cold  $0.1\times$  SSC containing 10 mM EDTA. mRNA was first eluted with 100  $\mu$ l of water followed by 150  $\mu$ l of water at 37  $^{\circ}$ C for 10 min each. Samples were precipitated with ethanol and stored at  $-80^{\circ}$ C. Pulldown efficiency of mRNA and depletion efficiency of 18S rRNA were determined by RT-qPCR (data not shown).

The RNA-seq library was prepared according to the manufacturer's protocol (Epicenter, catalog numbers SSV21124 and RSBC10948). The concentration of the library was determined with Pico Green (Life Technologies catalog number Q32851) according to the manufacturer's directions. Fluorescence signal was measured using a TECAN M1000 fluorescence plate reader.

**Bioinformatics**—Alignment of paired end reads to the *Drosophila melanogaster* genome (assembly BDGP 5/dm3) was performed with Tophat2 (v2.0.9) (35) using default parameters. DESeq2 (v1.6.3) (36) was used for gene expression level normalization and differential expression significance testing. Histone modification and motif enrichment analysis were carried with i-cisTarget (37) using default parameters. Enriched gene ontology analysis was done with FlyMine (v31.0) (38) using default parameters.

## Results

**Identification of Gro-interacting Proteins**—A previous study showed that deletion of the GP or CcN domain in the Gro central region led to a loss of Gro-mediated repression and to



**FIGURE 1. Purification of Gro-interacting proteins.** *A*, schematic representation of Gro. The Q, GP, CcN, and SP domains were tagged with GST. *B*, the GST-tagged domains were expressed in *E. coli* and purified with glutathione-agarose beads. They were then resolved by 10% SDS-PAGE and visualized by Coomassie Blue staining. These proteins were then used as affinity reagents in the purification of Gro-interacting proteins from embryonic nuclear extracts that were subsequently identified by MudPIT (see Table 4 and supplemental Table S1). *C*, Venn diagram showing overlap between the non-ribosomal proteins identified in two replicate sets of affinity purification experiments. Fisher's exact test indicates that the overlap between the two sets is highly significant ( $p < 2.2 \times 10^{-16}$ ).

lethality, whereas deletion of the SP domain led to reduced specificity of Gro-mediated repression and to reduced viability (19). To identify possible regulatory partners of these domains, we used them as affinity reagents to purify interacting proteins, which were then identified by mass spectrometry. The three central domains of Gro were expressed as GST-tagged proteins and purified from *Escherichia coli* lysates (Fig. 1, *A* and *B*). We also constructed a similarly tagged form of the N-terminal Q domain because previous studies suggested that, in addition to mediating Gro oligomerization, the Q domain engages in interactions with regulatory targets (39, 40).

The glutathione bead-immobilized GST-fused domains (or, as a negative control, immobilized unfused GST) were incubated with a *Drosophila* embryo nuclear extract. After extensive washing, interacting proteins were eluted with 2 M salt and analyzed by MudPIT (33) (supplemental Table S1C). Duplicate extract preparations and affinity purifications were carried out and analyzed on separate dates, and there was a high degree of overlap between the sets of proteins identified in these duplicate experiments (Fig. 1C). With three exceptions (see "Experimental Procedures"), only proteins that appeared in both rep-

**TABLE 3**  
Enriched gene ontology groups of Gro-interacting proteins  
 $p < 0.05$ .

Enriched gene ontology	No. of genes
Gene expression	83
Chromosome organization	21
Chromatin modification	11
mRNA processing	53
Cell cycle	30
Cell differentiation	66
Developmental process	76
Neurogenesis	57
Anatomical structure development	74

licates were included in our list of Gro-interacting proteins (Fig. 1C and supplemental Table S1, A and B). Gene ontology analysis of this list of 162 proteins revealed a variety of functions, including regulation of gene expression, RNA processing, and developmental processes (Table 3).

89 of the 162 Gro-interacting proteins associated uniquely with one domain (in all but one case, the SP domain), whereas 32 interacted with two domains. In the case of 23 of the 32 proteins that interacted with two domains, one of these domains was the Q domain (supplemental Table S1A). This is consistent with the known role of the Q domain in homo-oligomerization (12–15). In accord with this role, chromatography using GST-Q as the affinity reagent resulted in the purification of some full-length endogenous Gro (supplemental Table S1C and data not shown). This could lead to the co-purification of Gro-interacting proteins that bind to regions outside the Q domain. Thus, 112 (89 plus 23) of the 162 detected interacting proteins can, in principal, be accounted for by the binding of Gro to a single central domain. However, at least 50 proteins (162 minus 112) are able to bind independently to two or three central domains. The ability to interact with multiple Gro domains could allow tighter binding or more versatile control of binding.

The list of interacting proteins (Table 4 and supplemental Table S1, A and B) contains multiple components of known multisubunit protein complexes. For example, we identified the  $\alpha$  and  $\beta$  subunits of casein kinase II (CKII), a previously identified regulator of Gro activity (21). We also detected protein complexes involved in chromosome organization, including both components of the ATP-dependent chromatin remodeling and assembly factor (ACF), Acl1 and Iswi (41). Our proteomic screens also identified all the core protein components of the nucleosome (the core histones) as well as histone variant H2Av, consistent with previous studies demonstrating functional interactions between Gro and nucleosomes (42–44).

Perhaps most surprisingly, we discovered a number of components of the spliceosome among the group of Gro-interacting proteins, including all three proteins unique to U1 snRNP, components of U4/U6 snRNP, U2 snRNP, and the Sm complex (45, 46). To validate the interaction between Gro and U1 snRNP, *Drosophila* embryo nuclear extracts were subjected to immunoprecipitation using an affinity-purified antibody against the Gro GP domain or, as a negative control, rabbit IgG. An anti-Gro immunoblot of the immunoprecipitated material demonstrates the efficiency of the immunoprecipitation (Fig. 2A). RNA was extracted from the immunoprecipitates and ana-

**TABLE 4**  
Representative Gro-interacting proteins

For the complete list, full gene/protein names, and UniProt identifiers, see supplemental Table S1.

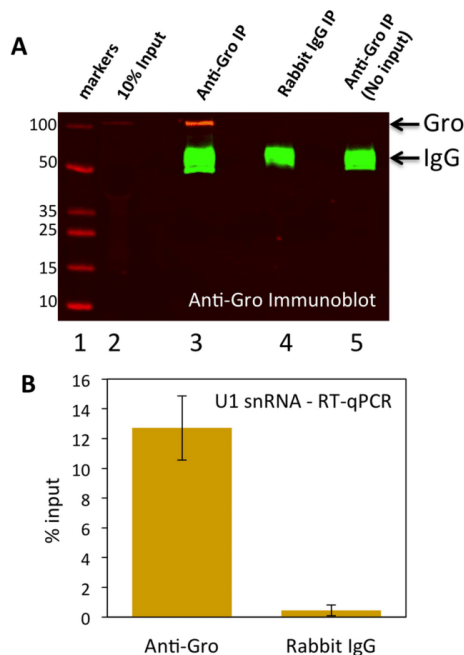
Protein	Description
<b>Chromosome organization</b>	
Acl1	ACF chromatin remodeling complex
Iswi	ACF chromatin remodeling complex
Caf1	dNuRD chromatin remodeling complex
Nap1	Histone chaperone
Ball	H2A Thr-119 kinase
JIL-1	H3 Ser-10 kinase
Top1	Topoisomerase
Top2	Topoisomerase
<b>Developmental process</b>	
CKII $\alpha$	CKII complex
CKII $\beta$	CKII complex
Nopp140	Negative regulator of CKII
Nito	Positive regulator of Wnt signaling pathway
Rm62	DEAD box helicase
Fmr1	Fragile X protein; interacting partner of Rm62
Vir	Involvement in sex determination
Snama	Involvement in eye morphogenesis
NonA	Involvement in visual perception
<b>mRNA processing</b>	
snRNP-U1-A	U1 snRNP complex
snRNP-U1-C	U1 snRNP complex
snRNP-U1-70K	U1 snRNP complex
U4-U6-60K	U4/U6 snRNP complex
CG7028	U4/U6 snRNP complex
Prp31	U4/U6 snRNP complex
Prp8	U5 snRNP complex
U2af38	U2 snRNP complex
U2af50	U2 snRNP complex
SF2	U2 snRNP complex
Smd2	Sm complex
Smd3	Sm complex
Nop60B	H/ACA ribonucleoprotein complex subunit
NHP2	H/ACA ribonucleoprotein complex subunit

lyzed by RT-qPCR with primers specific for U1 snRNA (a component of U1 snRNP). The results show that ~13% of the U1 snRNA in the nuclei of 0–12-h embryos is associated with Gro (Fig. 2B).

**Functional Analysis of Gro-interacting Proteins**—We next carried out functional assays to determine whether the interacting proteins are required for regulation of a Gro-responsive reporter gene. Previous studies established a reliable reporter assay for Gro function using a luciferase reporter containing Gal4 binding sites (UAS elements) as well as an artificial enhancer containing binding sites for the Dorsal and Twist activators (14, 16, 18, 47). Dorsal/Twist-activated transcription of this reporter is strongly repressed upon introduction of a Gal4-Gro fusion protein. By altering the position of UAS elements relative to the artificial enhancer, we were able to examine both short range and long range Gro-mediated repression simultaneously (Fig. 3, A and B). The reporter system relied on two variants of click beetle luciferase that use D-luciferin as a substrate and emit either red or green light (48). In addition, a plasmid encoding *Renilla* luciferase, which uses coelenterazine as a substrate, was used as an internal control for transfection efficiency, cell viability, and general effects on transcription and translation. We validated the three-reporter system using dsRNA against Dorsal, Gro, and Rpd3 (which is partially required for Gro-mediated repression (18)) (Fig. 3C). As predicted, Dorsal knockdown resulted in a complete loss of activation, Gro knockdown resulted in a complete loss of repression, and Rpd3 knockdown resulted in a partial loss of repression.

Each of the candidates from the screen for Gro-interacting proteins was knocked down by RNAi using up to three dsRNAs

## Groucho/Spliceosome Interactions



**FIGURE 2. Validation of the interaction between Gro and U1 snRNP.** *A*, 0–12-h *Drosophila* embryo nuclear extracts were subjected to immunoprecipitation using an affinity-purified polyclonal antibody directed against the Gro GP domain or, as a control, rabbit IgG. To assess immunoprecipitation efficiency and specificity, immunoprecipitates were subjected to SDS-PAGE and immunoblotting. The blot was probed with a mixture of the rabbit anti-Gro domain antibody and a mouse monoclonal anti-Gro antibody and with IRDye-labeled secondary antibodies. The signal from the rabbit antibody was detected in the green channel of the IR imager, whereas the signal from the mouse antibody was detected in the red channel. Rabbit IgG heavy chain (IgG) and Gro bands are indicated with arrows on the right. The orange-yellow color of the Gro band is indicative of the overlap between the red and green signals. Lane 1, markers labeled in kDa; lane 2, 10% input; lane 3, anti-Gro immunoprecipitate (IP); lane 4, rabbit IgG immunoprecipitate; lane 5, mock anti-Gro immunoprecipitate from which input nuclear extract was omitted. *B*, RNA was extracted from immunoprecipitates prepared as described in *A*. The RNA from the immunoprecipitates as well as the RNA extracted from the input nuclear extracts was analyzed by RT-qPCR as described under “Experimental Procedures” to determine U1 snRNA levels. Error bars based on two independent biological replicates indicate S.D. A two-tailed *t* test gives  $p = 0.016$ .

per gene to guard against off-target effects. We excluded the histones from this analysis under the assumption that knockdown of these essential chromatin components would have pleiotropic deleterious effects on cell metabolism and because each histone is encoded by multiple genes, making efficient knockdown problematic. We therefore tested 157 genes in this S2 cell luciferase assay in most cases with multiple dsRNAs per gene (three if available), and each dsRNA was tested in triplicate. In total, we carried out ~1,300 assays (including controls) in a 96-well plate format using a partially automated approach (see “Experimental Procedures”).

A candidate was scored as a regulator of Gro-mediated repression if knockdown reproducibly resulted in either an

increase or a decrease in the level of repression (see “Experimental Procedures” for explanation of the statistical test of significance). 44 candidates met these criteria of which 28 interfered with optimal repression (*i.e.* repression increased upon knockdown; these were termed “negative regulators of Gro”), and 16 were required for optimal repression (*i.e.* repression decreased upon knockdown; these were termed “positive regulators of Gro”). We provide representative data for one negative regulator (Vir), one positive regulator (snRNP-U1-C), and one protein that is neither a positive nor a negative regulator (SR protein kinase, SRPK) (Fig. 3D); a list of all the positive and negative regulators (Table 5); and a separate list showing the quantitative effect of RNAi knockdown of each of the 44 regulators on repression by Gal4-Gro (supplemental Table S2). Of particular interest, four spliceosomal proteins, including two components of U1 snRNP, act as positive regulators of Gro, confirming the functional significance of the interaction between Gro and U1 snRNP. A few other noteworthy examples among the Gro regulators (Table 5 and supplemental Table S2) include both components of the CKII complex (CKII $\alpha$  and CKII $\beta$ ), which act as negative regulators, and the chromatin remodeling factor Acf1, which acts as a positive regulator (see “Discussion”).

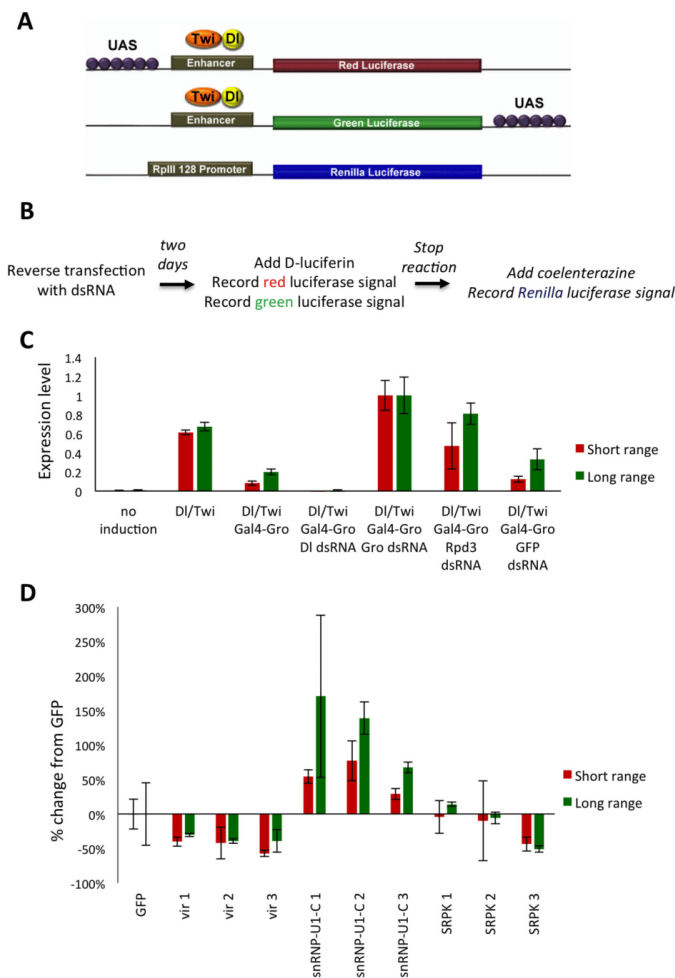
**Expression Profiling of Gro and snRNP-U1-C Knockdown Cells**—snRNP-U1-C is one of the components of the U1 snRNP complex, which is responsible for 5' splice site recognition (46). In addition to its role in RNA processing, it has been shown to repress transcription of EWS/Fli1-transactivated genes (30). Because our data indicated that snRNP-U1-C may also modulate Gro function, we examined the genome-wide role of snRNP-U1-C in Gro-mediated repression. Using RNA-seq, we compared the effects of snRNP-U1-C knockdown with those of Gro knockdown on the gene expression profile in S2 cells. Cells were treated with Gro or snRNP-U1-C dsRNA for 4 days, leading to 4-fold or greater knockdown of the Gro and snRNP-U1-C mRNA (Fig. 4A). The transcriptomes in wild-type and Gro knockdown S2 cells were quantitatively similar to those published previously (49, 50) (Fig. 4, B and C). We note that the genes differentially expressed in the snRNP-U1-C knockdown are enriched for genes containing introns as would be expected given the role of U1 snRNP in splicing. However, this set of genes also contains a number of intronless genes, consistent with the idea that snRNP-U1-C has roles in gene regulation apart from its role in splicing (Fig. 4D). We note that changes in the expression of an intronless gene could also reflect a requirement for the product of an intron-containing gene in the expression of the intronless gene.

98 genes were differentially expressed in both Gro and snRNP-U1-C knockdown cells (Fig. 4E) of which 36 were up-regulated in either case. These coordinately up-regulated targets included genes in various signaling pathways such as the Wnt, Notch, and Toll pathways (Table 6). Comparison with publicly available ChIP-seq data on histone modification and transcription factor binding revealed that these coordinately regulated genes were most enriched for histone H3K36 methylation and the H3K36 methyltransferase ASH1 (Fig. 4F).

To determine whether the regulatory effects of knocking down Gro are likely to be direct, we compared our RNA-seq



## Groucho/Spliceosome Interactions



**FIGURE 3. The three-reporter high throughput luciferase assay.** *A*, schematic representation of the three reporters. Constructs are not drawn to scale. In the red luciferase reporter, the Gal4 binding sites (UAS elements) are immediately upstream of the enhancer, whereas in the green luciferase reporter, the UAS elements are about 2 kb downstream of the transcriptional start site. Expression is induced by the Dorsal (*Dl*) and Twist (*Twi*) activators and repressed by Gal4-Gro. The *Renilla* luciferase reporter under control of the class II promoter from the gene encoding RNA polymerase III subunit (RpIII128) was used as an internal control for transfection efficiency, cell viability, and general effects on transcription and translation. *B*, flow chart of the reporter assay. *C*, validation of the reporter assay. Co-transfection with Dorsal and Twist (*Dl/Twi*)-encoding plasmids activated both the red and green reporters, whereas addition of a plasmid encoding the Gal4-Gro fusion resulted in repression of the reporters. Dorsal, Gro (including Gal4-Gro), and the histone deacetylase Rpd3, which is partially required for Gro-mediated repression (18), were knocked down by RNAi. Data are normalized to the red and green signals from the Gro dsRNA sample. Error bars based on triplicate transfection assays represent S.D. *D*, representative results of the reporter assay. The luciferase reporter assay was carried out using three non-overlapping dsRNAs from the genes encoding Vir, snRNP-U1-C, and SR protein kinase (*SRPK*). The result of transfection with each dsRNA was compared with that of transfection with GFP dsRNA. Error bars based on triplicate transfection assays represent S.D.

data from Gro knockdown S2 cells with available S2 cell Gro ChIP data (49). Gro appears to bind many genes that it does not repress (Fig. 5A). This is consistent with observations made with numerous regulatory factors (51, 52) and suggests that binding, although required, is not sufficient for regulation. We observed an enrichment of Suppressor of Hairless (Su(H)) and Brinker (Brk) binding motifs within Gro ChIP-

seq peaks in the differentially expressed genes but not in the non-differentially expressed genes (Fig. 5B). Comparison of our RNA-seq data from Gro knockdown cells with available Pol II ChIP-chip data (53) also reveals an enrichment in Pol II pausing near the transcriptional start site in genes that are up-regulated upon Gro knockdown (*i.e.* genes that are repressed by Gro; Fig. 6).

## Groucho/Spliceosome Interactions

**TABLE 5**

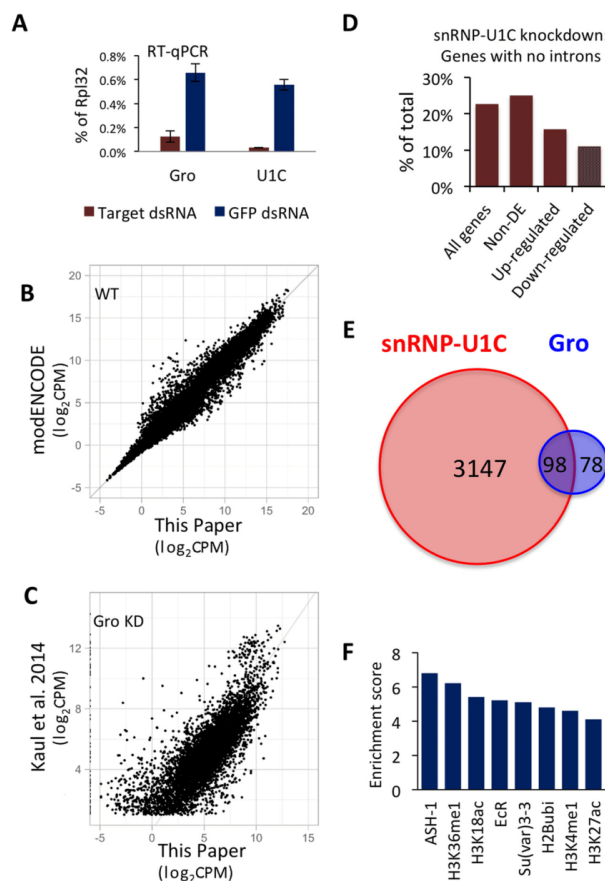
**Positive and negative Gro regulators**

See supplemental Table S2 for quantitative information on positive and negative regulation by these factors. See "Experimental Procedures" for an explanation of the test of statistical significance that genes had to pass to be included in this list.

Potential negative regulators of Gro <sup>a</sup>	Potential positive regulators of Gro <sup>b</sup>
CKII $\alpha$ , CKII $\beta$ , Nopp140, fl(2)d, l(2)35Df, l(3)72Ab, Vir, NonA, Nito, X16, Nap1, JIL-1, Nop5, NHP2, FK506-bp1, CG3605, Prp31, Fmr1, Cdk12, CG6418, CG7372, CG7946, Srp68, Srp72, Ssrp, Pitslre, Pep, Nab2	snRNP-U1-C, snRNP-U1-70K, U2af50, U4-U6-60K, Rm62, Orc2, Smid, Acn, Acl1, Snama, CG1622, ZCHC8, CG4709, CG4806, Lat, Srp19

<sup>a</sup> Negative regulators are defined as the products of those genes the knockdown of which led to increased repression by Gal4-Gro in the reporter assay.

<sup>b</sup> Positive regulators are defined as the products of those genes the knockdown of which led to decreased repression by Gal4-Gro in the reporter assay.



**FIGURE 4. Genome-wide expression profiling reveals co-regulation of genes by Gro and snRNP-U1-C.** *A*, expression of Gro and snRNP-U1-C mRNA after dsRNA treatment. RT-qPCR was performed after extraction of total RNA. Data were normalized to reference gene Rpl32. *Error bars* based on duplicate experiments represent S.D. *B*, comparison of transcriptomes from our wild-type S2 cell RNA-seq data and the modENCODE S2 cell RNA-seq data. *C*, comparison of transcriptomes from our Gro knockdown RNA-seq data and previously published Gro knockdown RNA-seq data (49). The transcripts that were detected at significant levels in only the previously published Gro knockdown study (represented by the *points* in contact with the *vertical axis*) correspond primarily to non-polyadenylated transcripts. In *B* and *C*, the scale on both axes is log<sub>2</sub>(CPM) where CPM is counts per million sequence reads. *D*, based on RNA-seq analysis of wild-type and snRNP-U1-C knockdown cells, genes were categorized as non-differentially expressed upon knockdown (*Non-DE*; 12,028 genes), up-regulated upon knockdown (1,431 genes), and down-regulated upon knockdown (1,691 genes). The percentage of genes in each category with no introns is shown. Some *Drosophila* genes lack annotated transcripts, and thus it was not possible to determine their intron count. This results in a small numerical discrepancy between the number of differentially expressed genes included in this analysis and the number of snRNP-U1-C differentially expressed genes shown in *E*. *E*, Venn diagram showing numbers of differentially expressed genes in Gro and snRNP-U1-C knockdown cells and the overlap between these sets. Fisher's exact test indicates that the overlap is highly significant ( $p < 2.2 \times 10^{-16}$ ). *F*, enrichment of Gro/snRNP co-regulated genes for various features. Normalized enrichment scores were calculated using cumulative recovery curves (37). Scores above 2.5 are considered significant.

**TABLE 6**  
Genes up-regulated upon knockdown of either Gro or snRNP-U1-C

Name	Function
Secreted Wg-interacting molecule	Wnt signaling pathway
Wnt oncogene analog 5	Wnt signaling pathway
E(spl)m2-BFM	Notch signaling
Spatzle	Toll signaling pathway
SH2 ankyrin repeat kinase	JNK cascade
Dawdle	SMAD protein signal transduction
CG33275	Rho protein signal transduction
Epac	Rap protein signal transduction
Boundary element-associated factor of 32kD	H3K9 methylation
Syncrinp	Dorsal/ventral axis specification
Fasciclin 1	Neuron recognition
Axotactin	Transmission of nerve impulse
Muscle-specific protein, 300 kDa	Skeletal muscle tissue development
Cheerio	Lamellocyte differentiation

## Discussion

Previous studies showed that the disordered Gro central domains are essential for properly regulated transcriptional repression (2, 19). To shed light on the mechanism by which these domains function, we used them as affinity reagents to purify interacting proteins in *Drosophila* embryo nuclear extracts that were then identified by MudPIT. We identified over 160 interacting polypeptides, many of which associate with one another in a variety of multiprotein complexes. Several of these interacting proteins (*e.g.* the core histones and CKII) were previously characterized as Gro interactors, thus partially validating the screen. In addition, we validated the interaction between Gro and U1 snRNP by demonstrating the presence of U1 snRNA in an anti-Gro immunoprecipitate of embryonic nuclear extracts.

As a means of systematically validating interactions, we used a functional assay in *Drosophila* cells in which 157 of the interactors were each knocked down by RNAi to determine their requirement for Gal4-Gro-mediated repression of a luciferase reporter. In this way, we obtained evidence that 44 of the interactors have functional roles in Gro-mediated repression. 28 of these are required for repression, whereas 16 of them antagonize repression. The number 44 is probably an underestimate of the true number of functional interactors due to the artificiality of the reporter assay. For example, because we artificially recruit Gro to the reporter by tethering it to the Gal4 DNA binding domain, any interactions that work to help recruit Gro to the template will not be required. In addition, the reporters are introduced by transient transfection, and certain chromatin structures or modifications that contribute to Gro-mediated repression may not be reproduced in this context.

*Gro Interactors Include Chromatin Remodelers, Protein Kinases, and Protein Complexes Involved in RNA Processing*—Gro-mediated repression may be associated with changes in chromatin structure, including histone deacetylation and possibly increased nucleosome density (3, 18, 54). Consistent with this possibility, our proteomic screen identified a number of histone modifiers and ATP-dependent chromatin remodelers, including subunits of the ACF chromatin remodeling complexes (Acf1 and Iswi), the histone chaperone NAP1, and the histone kinases JIL-1 and Ball. Consistent with the idea that chromatin remodelers may be required for Gro-mediated repression by catalyzing changes in nucleosome density or

higher order chromatin structure, our reporter assay showed that Acf1 is required for optimal repression by Gro.

CKII is a heterotetrameric complex consisting of two copies of a catalytic subunit (CKII $\alpha$ ) and two copies of a regulatory subunit (CKII $\beta$ ) (55, 56). A previous study showed that CKII phosphorylates Gro at multiple sites, including serines 239 and 253, to promote repression (21). We identified both the  $\alpha$  and  $\beta$  subunits of CKII and the CKII negative regulator Nopp140 in our proteomic screen, but our findings are inconsistent with the view that CKII is a positive regulator of Gro and that Nopp140 acts by inhibiting CKII. This is because our reporter assays show that CKII $\alpha$ , CKII $\beta$ , and Nopp140 are all negative regulators of Gro. However, our results are consistent with other findings showing that Gro phosphorylation can block repression (2). Furthermore, the effect we observed due to Nopp140 knockdown could reflect the role of this factor in processes other than CKII regulation (57).

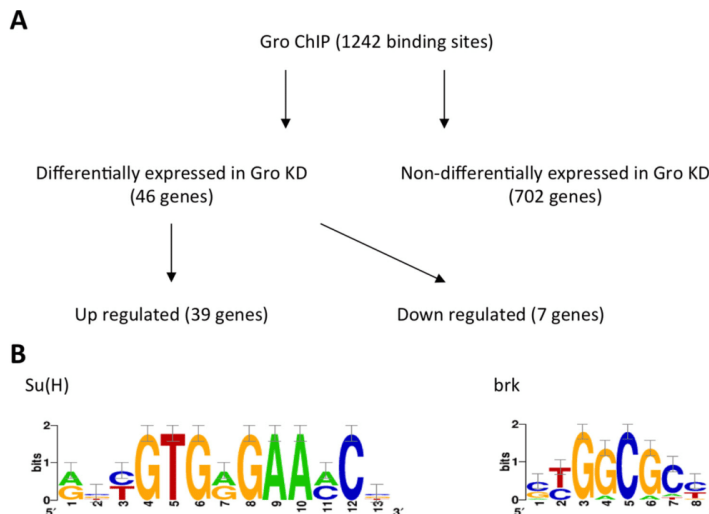
In addition to several expected protein complexes, we also isolated many novel Gro-interacting proteins, one of which is the RNA helicase Rm62 (also known as p68). Rm62 is a DEAD box RNA helicase that has multiple functions, including roles in RNA processing, RNAi, and transcriptional regulation (58). Previous studies have shown a dual role for Rm62 in transcriptional regulation; its interaction with coactivator CBP/p300 may lead to gene activation (59), whereas its interaction with HDAC1 may lead to repression (60, 61). Our reporter assay confirms its function as a positive regulator of Gro-mediated repression as knocking down Rm62 resulted in attenuated Gro activity. Interestingly, Rm62 was also shown to be an essential splicing component through its action on the U1 snRNP (62, 63). The possible significance of the spliceosome in Gro-mediated repression is discussed below.

*An Unanticipated Role for the Spliceosome in Gro-mediated Repression*—One of the most surprising findings from our proteomic screen was the purification of a significant portion of the spliceosome complex, which suggests a potential role for the spliceosome in transcriptional regulation.

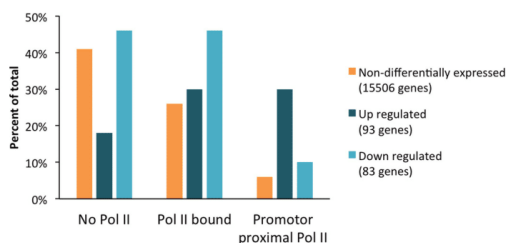
Pre-mRNA processing frequently occurs co-transcriptionally (64–66). Splicing factors are often recruited to nascent transcripts by the C-terminal domain of the Pol II large subunit and elongation factors (67, 68). In addition, there is evidence that co-activators are able to interact with splicing factors (27). The interaction between the transcriptional and splicing machinery may be functionally relevant because different promoters can yield transcripts that are subject to differential alternative splicing (69, 70). Although many studies have focused on the effect of transcription factors in splicing, there is also increasing evidence that promoter-proximal splicing elements can influence transcription (26, 28, 71).

U1 snRNP, a part of the spliceosome, consists of U1 snRNA, three U1 snRNP-specific proteins, and the seven-subunit Sm complex (46). Our list of 162 Gro-interacting proteins (supplemental Table S1, B and C) includes all three U1 snRNP-specific proteins (snRNP-U1C, snRNP-U1-70K, and snRNP-U1-A) as well as two subunits of the Sm complex (Sm-D2 and Sm-D3). We note that we also detected at least four other Sm complex subunits in one of the two replicate screens (Sm-B, Sm-F, Sm-D1, and Sm-G) (supplemental Table S1C). Additionally, we showed by co-immu-

## Groucho/Spliceosome Interactions



**FIGURE 5. Gro binding regions in differentially expressed genes.** A, S2 cell ChIP-seq data (49) identified 1,242 Gro binding sites, which map to 748 genes, 46 of which were differentially expressed when we knocked down Gro. Of the 46 differentially expressed genes, 39 were up-regulated and seven were down-regulated in response to Gro knockdown (KD). B, Gro binding regions in the 46 differentially expressed genes are significantly enriched for Su(H) and Brk binding sites.



**FIGURE 6. Gro-repressed genes are enriched for promoter-proximal Pol II.** The percentage of non-differentially expressed genes and genes that are either up-regulated or down-regulated in Gro knockdown cells containing no Pol II bound or Pol II bound or enriched for promoter-proximal Pol II as ascertained by Pol II ChIP-chip analysis (53).

nonprecipitation that ~13% of U1 snRNA, the RNA component of the U1 snRNP, is associated with Gro in embryonic nuclei. Thus, we have detected essentially the entire U1 snRNP in our proteomic screens for Gro-interacting proteins.

Data from our reporter assay suggest that the U1 snRNP complex is required for optimal Gro-mediated repression as snRNP-U1-C and snRNP-U1-70K knockdown attenuated repression. Consistent with our finding, it has been shown that snRNP-U1-C overexpression can decrease EWS/FLI-activated transcription (30). It is worth noting that the U1 snRNA is known to associate with transcription factor IIH and promote transcriptional initiation *in vitro* (29). Thus, the effect of the U1 snRNP complex in transcription regulation may be context-dependent.

**Gro Recruitment Is Insufficient for Repression**—The available S2 cell Gro ChIP-seq data (49) reveal 1,242 Gro binding sites in the S2 cell genome associated with 748 genes, whereas our RNA-seq analysis revealed that only 46 of these 748 genes are

differentially expressed in Gro knockdown S2 cells, implying that Gro binds to many genes that it does not regulate. The apparent contradiction could be explained by the absence of a required transcriptional activator in S2 cells to activate these genes upon Gro depletion. Regardless of the reason for the finding that Gro binds to many more genes than it regulates, this is a phenomenon that is common to many (perhaps most) eukaryotic gene-specific transcriptional regulators (51, 52). Gro ChIP-seq peaks associated with genes differentially expressed upon Gro knockdown are enriched for Su(H) and Brk binding motifs. This is in agreement with the known roles of Su(H) and Brk in the recruitment of Gro to target genes in the Notch and Dpp signaling pathways, respectively (72–74).

Genes that are up-regulated in Gro knockdown cells (and that are therefore candidate Gro repression targets) exhibit enrichment in Pol II pausing near the transcriptional start site. This finding is in agreement with the hypothesis that Pol II pausing is one mechanism to repress gene expression (75, 76). We note that our proteomic screen revealed the Pol II C-terminal domain kinase Cdk12 as a Gro-interacting protein (supplemental Table S1). By phosphorylating the C-terminal domain on Ser-2, Cdk12 may function to allow release of paused Pol II (77). Consistent with this idea, our reporter assay shows that Cdk12 functions to alleviate Gro-mediated repression (Table 5 and supplemental Table S2).

Genes that are differentially expressed in Gro and snRNP-U1-C knockdown cells are enriched for H3K36me1 as well as the H3K36 methyltransferase ASH1. Although H3K36me is involved in multiple functions, including transcriptional regulation, splicing, and DNA repair (78, 79), these findings suggest a previously unknown role for this histone mark in Gro-mediated repression.



**The Gro Central Region as a Regulatory Hub of Repression Activity**—In conclusion, our findings reinforce the idea that the Gro central domains, which are intrinsically disordered, are indispensable for repression (19). Previous studies from our laboratory and other laboratories show that the GP domain interacts with the histone deacetylase Rpd3/HDAC1, which may promote local histone deacetylation and alter nucleosome density (16, 18). The identification of the ACF chromatin remodeling complexes as a central region-interacting protein complex and our demonstration that knockdown of this protein attenuates Gro-mediated repression provide further support for the idea that regulation of chromatin structure is a critical aspect of Gro-mediated repression. Conversely, modulation of chromatin structure is likely not the only mechanism of Gro-mediated repression as histone deacetylase inhibitors and Rpd3 knockdown reduce, but do not abolish, Gro-mediated repression (16, 18) (Fig. 3C). Through a combination of proteomic screening, reporter assays, and genome-wide expression profiling, our results suggest a possible new mechanism of Gro-mediated repression involving the action of the spliceosome. Future experiments will focus on elucidating the underlying mechanisms by which these interacting partners act in Gro-mediated repression.

**Author Contributions**—A. J. C., P. N. K., and W. T.-J. conceived and planned the study, which was coordinated by A. J. C. P. N. K. and A. J. C. wrote the manuscript. P. N. K. and W. T.-J. conducted most of the experiments. T. Y. Y. assisted with the co-immunoprecipitation study. M. C. carried out the bioinformatics analysis. A. A. V. and J. A. W. carried out the MudPIT analysis. All authors reviewed and approved the manuscript.

**Acknowledgments**—We thank Robert Damoiseaux, Scientific Director of the UCLA Molecular Screening Shared Resource Center, for assistance in developing the mechanized multiple luciferase reporter screen. We are grateful to all the members of the Courey laboratory (past and present) for insight and advice.

## References

- Courey, A. J., and Jia, S. (2001) Transcriptional repression: the long and the short of it. *Genes Dev.* **15**, 2786–2796
- Turki-Judeh, W., and Courey, A. J. (2012) Groucho: a corepressor with instructive roles in development. *Curr. Top. Dev. Biol.* **98**, 65–96
- Martinez, C. A., and Arnosti, D. N. (2008) Spreading of a corepressor linked to action of long-range repressor hairy. *Mol. Cell. Biol.* **28**, 2792–2802
- Chen, G., and Courey, A. J. (2000) Groucho/TLE family proteins and transcriptional repression. *Gene* **249**, 1–16
- Gasperowicz, M., and Otto, F. (2005) Mammalian Groucho homologs: redundancy or specificity? *J. Cell. Biochem.* **95**, 670–687
- Nagel, A. C., Krejci, A., Tenin, G., Bravo-Patiño, A., Bray, S., Maier, D., and Preiss, A. (2005) Hairless-mediated repression of notch target genes requires the combined activity of Groucho and CtBP corepressors. *Mol. Cell. Biol.* **25**, 10433–10441
- Blair, S. S. (2007) Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. *Annu. Rev. Cell Dev. Biol.* **23**, 293–319
- Hasson, P., and Paroush, Z. (2006) Crosstalk between the EGFR and other signalling pathways at the level of the global transcriptional corepressor Groucho/TLE. *Br. J. Cancer* **94**, 771–775
- Allen, T., van Tuyll, M., Iyengar, P., Jothy, S., Post, M., Tsao, M. S., and Lobe, C. G. (2006) Grg1 acts as a lung-specific oncogene in a transgenic mouse model. *Cancer Res.* **66**, 1294–1301
- Buscarlet, M., and Stifani, S. (2007) The 'Marx' of Groucho on development and disease. *Trends Cell Biol.* **17**, 353–361
- Jennings, B. H., Pickles, L. M., Wainwright, S. M., Roe, S. M., Pearl, L. H., and Ish-Horowitz, D. (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol. Cell* **22**, 645–655
- Song, H., Hasson, P., Paroush, Z., and Courey, A. J. (2004) Groucho oligomerization is required for repression *in vivo*. *Mol. Cell. Biol.* **24**, 4341–4350
- Chodaparambil, J. V., Pate, K. T., Hepler, M. R., Tsai, B. P., Muthurajan, U. M., Luger, K., Waterman, M. L., and Weis, W. I. (2014) Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J.* **33**, 719–731
- Chen, G., Nguyen, P. H., and Courey, A. J. (1998) A role for Groucho tetramerization in transcriptional repression. *Mol. Cell. Biol.* **18**, 7259–7268
- Pinto, M., and Lobe, C. G. (1996) Products of the grg (Groucho-related gene) family can dimerize through the amino-terminal Q domain. *J. Biol. Chem.* **271**, 33026–33031
- Chen, G., Fernandez, J., Mische, S., and Courey, A. J. (1999) A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in *Drosophila* development. *Genes Dev.* **13**, 2218–2230
- Choi, C. Y., Kim, Y. H., Kwon, H. J., and Kim, Y. (1999) The homeodomain protein NK-3 recruits Groucho and a histone deacetylase complex to repress transcription. *J. Biol. Chem.* **274**, 33194–33197
- Winkler, C. J., Ponce, A., and Courey, A. J. (2010) Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density. *PLoS One* **5**, e10166
- Turki-Judeh, W., and Courey, A. J. (2012) The unconserved groucho central region is essential for viability and modulates target gene specificity. *PLoS One* **7**, e30610
- Hasson, P., Egoz, N., Winkler, C., Volohonsky, G., Jia, S., Dinur, T., Volk, T., Courey, A. J., and Paroush, Z. (2005) EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. *Nat. Genet.* **37**, 101–105
- Nuthall, H. N., Joachim, K., and Stifani, S. (2004) Phosphorylation of serine 239 of Groucho/TLE1 by protein kinase CK2 is important for inhibition of neuronal differentiation. *Mol. Cell. Biol.* **24**, 8395–8407
- Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148
- Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding. *Biochim. Biophys. Acta* **1804**, 1231–1264
- Tantos, A., Han, K. H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.* **348**, 457–465
- Bondos, S. E., and Hsiao, H. C. (2012) Roles for intrinsic disorder and fuzziness in generating context-specific function in Ultrathorax, a Hox transcription factor. *Adv. Exp. Med. Biol.* **725**, 86–105
- Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., and Kjems, J. (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors *in vivo*. *Mol. Cell* **29**, 271–278
- Monsalve, M., Wu, Z., Adelmant, G., Puigserver, P., Fan, M., and Spiegelman, B. M. (2000) Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol. Cell* **6**, 307–316
- Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A., and Proudfoot, N. J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.* **16**, 2792–2799
- Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N. J., and Akoulitchev, A. (2002) U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat. Struct. Biol.* **9**, 800–805
- Knoop, L. L., and Baker, S. J. (2000) The splicing factor U1C represses EWS/FLI-mediated transactivation. *J. Biol. Chem.* **275**, 24865–24871
- Armknicht, S., Boutros, M., Kiger, A., Nybakken, K., Mathey-Prevot, B., and Perrimon, N. (2005) High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392**, 55–73
- Soeller, W. C., Poole, S. J., and Kornberg, T. (1988) *In vitro* transcription of the *Drosophila* engrailed gene. *Genes Dev.* **2**, 68–81
- Wohlschlegel, J. A. (2009) Identification of SUMO-conjugated proteins and their SUMO attachment sites using proteomic mass spectrometry.

## Groucho/Spliceosome Interactions

- Methods Mol. Biol.* **497**, 33–49
34. Zhou, R., Mohr, S., Hannon, G. J., and Perrimon, N. (2013) Inducing RNAi in *Drosophila* cells by transfection with dsRNA. *Cold Spring Harb. Protoc.* **2013**, 461–463
  35. Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36
  36. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550
  37. Herrmann, C., Van de Sande, B., Potier, D., and Aerts, S. (2012) i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* **40**, e114
  38. Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., McLaren, P., North, P., Rana, D., Riley, T., Sullivan, J., Watkins, X., Woodbridge, M., Lilley, K., Russell, S., Ashburner, M., Mizuguchi, K., and Micklem, G. (2007) FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biol.* **8**, R129
  39. Brantjes, H., Roose, J., van De Wetering, M., and Clevers, H. (2001) All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res.* **29**, 1410–1419
  40. Daniels, D. L., and Weis, W. I. (2005)  $\beta$ -Catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12**, 364–371
  41. Bouazoune, K., and Brehm, A. (2006) ATP-dependent chromatin remodeling complexes in *Drosophila*. *Chromosome Res.* **14**, 433–449
  42. Flores-Saab, R. D., and Courey, A. J. (2000) Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho- and Tup1-mediated repression. *Nucleic Acids Res.* **28**, 4189–4196
  43. Sekiya, T., and Zaret, K. S. (2007) Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin *in vitro* and impairs activator binding *in vivo*. *Mol. Cell* **28**, 291–303
  44. Edmondson, D. G., Smith, M. M., and Roth, S. Y. (1996) Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. *Genes Dev.* **10**, 1247–1259
  45. Patel, S. B., and Bellini, M. (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* **36**, 6482–6493
  46. Will, C. L., and Lüthmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707
  47. Fisher, A. L., Ohsako, S., and Caudy, M. (1996) The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino acid transcription repression and protein-protein interaction domain. *Mol. Cell Biol.* **16**, 2670–2677
  48. Thorne, N., Inglese, J., and Auld, D. S. (2010) Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem. Biol.* **17**, 646–657
  49. Kaul, A., Schuster, E., and Jennings, B. H. (2014) The Groucho co-repressor is primarily recruited to local target sites in active chromatin to attenuate transcription. *PLoS Genet.* **10**, e1004595
  50. Celniker, S. E., Dillon, L. A., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K. P., Waterston, R. H., and modENCODE Consortium (2009) Unlocking the secrets of the genome. *Nature* **459**, 927–930
  51. Fisher, W. W., Li, J. J., Hammonds, A. S., Brown, J. B., Pfeiffer, B. D., Weizmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J. A., Eisen, M. B., Bickel, P. J., Biggin, M. D., and Celniker, S. E. (2012) DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21330–21335
  52. Walhout, A. J. (2011) What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol.* **12**, 109
  53. Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., and Adelman, K. (2007) RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511
  54. Li, L. M., and Arnosti, D. N. (2011) Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. *Curr. Biol.* **21**, 406–412
  55. Niefind, K., Guerra, B., Ermakowa, I., and Issinger, O. G. (2001) Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. *EMBO J.* **20**, 5320–5331
  56. Niefind, K., Raaf, J., and Issinger, O. G. (2009) Protein kinase CK2 in health and disease: protein kinase CK2: from structures to insights. *Cell. Mol. Life Sci.* **66**, 1800–1816
  57. Meier, U. T., and Blobel, G. (1992) Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* **70**, 127–138
  58. Fuller-Pace, F. V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.* **34**, 4206–4215
  59. Rossow, K. L., and Janknecht, R. (2003) Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene* **22**, 151–156
  60. Buszczak, M., and Spradling, A. C. (2006) The *Drosophila* P68 RNA helicase regulates transcriptional deactivation by promoting RNA release from chromatin. *Genes Dev.* **20**, 977–989
  61. Wilson, B. J., Bates, G. J., Nicol, S. M., Gregory, D. J., Perkins, N. D., and Fuller-Pace, F. V. (2004) The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol. Biol.* **5**, 11
  62. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**, 46–50
  63. Liu, Z. R. (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol. Cell Biol.* **22**, 5443–5450
  64. Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* **15**, 326–331
  65. Han, J., Xiong, J., Wang, D., and Fu, X. D. (2011) Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343
  66. Bentley, D. L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17**, 251–256
  67. Maniatis, T., and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506
  68. Bentley, D. (2002) The mRNA assembly line: transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14**, 336–342
  69. Cramer, P., Cáceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., and Kornblihtt, A. R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* **4**, 251–258
  70. Cramer, P., Pesce, C. G., Baralle, F. E., and Kornblihtt, A. R. (1997) Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11456–11460
  71. Merkhofer, E. C., Hu, P., and Johnson, T. L. (2014) Introduction to cotranscriptional RNA splicing. *Methods Mol. Biol.* **1126**, 83–96
  72. Hasson, P., Müller, B., Basler, K., and Paroush, Z. (2001) Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. *EMBO J.* **20**, 5725–5736
  73. Paroush, Z., Finley, R. L., Jr., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R., and Ish-Horowitz, D. (1994) Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**, 805–815
  74. Grbavec, D., and Stifani, S. (1996) Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. *Biochem. Biophys. Res. Commun.* **223**, 701–705
  75. Adelman, K., and Lis, J. T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731
  76. Gaertner, B., and Zeitlinger, J. (2014) RNA polymerase II pausing during development. *Development* **141**, 1179–1183
  77. Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., and Greenleaf, A. L. (2010) CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316
  78. Lee, J. S., and Shilatifard, A. (2007) A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat. Res.* **618**, 130–134
  79. Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126

Downloaded from <http://www.jbc.org/> at UCLA-Louisie Darling Biomed Lib on March 7, 2016

## **Chapter 3**

### **Interaction between Groucho and histone deacetylase**

## **Abstract**

The *Drosophila* co-repressor Gro interacts with class I histone deacetylase Rpd3 through its central disordered domain. Previous studies demonstrated the functional significance of this interaction through the use of histone deacetylase inhibitors, but the conclusions were not definitive as they did not rule out the possibility that the role of Rpd3 was indirect. Using a Gro mutant that lacks the Rpd3 interacting domain, I attempted to demonstrate further the connection between Rpd3 and Gro-mediated repression. Transgenic flies expressing this mutant was used to measure repression activity by phenotypic analysis of wing developmental defect and by a vgQ-LacZ reporter assay. We were unable to draw a definitive conclusion about the role of Rpd3 in Gro-mediated repression due to technical difficulties. Nevertheless, our results corroborate and extend previous findings that Gro interacts with histone deacetylase through its central disordered domain.



## Introduction

Histone deacetylase has been associated with various biological processes, from development to tumorigenesis and aging [1, 2]. There are 4 classes of histone deacetylases conserved from yeast to human, classified by sequence homology to their yeast orthologues: Class I is homologous to yeast Rpd3; Class II to yeast Hda1; and Class III to yeast Sirtuin; Class IV is structurally related to Class I and II, but phylogenetic analysis suggests that it belongs to a separate class [3, 4].

Gro is known to interact with the *Drosophila* class I histone deacetylase Rpd3 through its GP domain [5]. A previous study demonstrated the functional importance of Rpd3 in Gro mediated repression using a histone deacetylase inhibitor, as well as overexpression experiments in tissue culture cells [6]. While these approaches are informative, they are not definitive, as the inhibitor may act indirectly and overexpression could lead to artifacts due to superphysiological concentrations of the histone deacetylase. Thus, to demonstrate further the connection between Rpd3 and Gro-mediated repression, we created a Gro variant that does not interact with Rpd3 and measured its repression activity. Loss in repression activity of this mutant would demonstrate a direct role of Rpd3 to Gro mediated repression. Using transgenic flies expressing this Gro mutant, we attempted to determine its activity by in development with various phenotypic assays. As a result of unexpected technical difficulties, we were unable to draw a definitive conclusion about Rpd3 in Gro-mediated repression.

Nevertheless, our results corroborate previous findings that the GP domain is necessary for both interaction with Rpd3 and nuclear localization.

## Materials and methods

### *Plasmid:*

The GP deletions were generated by PCR: which GP2 lacks amino acids 134-148, GP3 lacks amino acids 149-163, GP4 lacks amino acids 164-179, GP5 lacks amino acids 180-194, GP6 lacks amino acids 134-163, and GP7 lacks amino acids 164-194. PCR reactions was setup to amplify the whole plasmid except the indicated regions using wild type Gro in pET17b as template. All primers contained an Asc I restriction site, so that the amplified products could be circularized after restriction digestion. These plasmids are referred to as pET17b-GP variants.

Plasmids used for subcellular localization immunofluorescence in S2 cells were generated by PCR of the Gro gene with pET17b-GP variants as the template. PCR products were cloned into S2 cell expression vector pMK33-BD. The primers also encode for a N-terminal flag tag (DYKDDDDK) for immunostaining. These plasmids will be referred to as pMK33-GP variants.

To generate chimeric Gro, PCR reactions was setup to amplify the whole plasmid except the GP domain using wild type Gro in pET17b. The primers also contain a coding sequence for the T-antigen NLS sequence (PKKKRKVEDP) [7]. The PCR product was digested with Asc I and re-ligated as previously mentioned. This plasmid will be referred to as pET17b-NLS Gro. The coding region of pET17b-NLS Gro was amplified and cloned into

the pUASP expression vector using NotI and BamHI, and the resulting plasmid is named pUASP-NLS Gro. pUASP- NLS Gro was used to transform the Drosophila germ line by standard procedures (Rainbow Transgenic Flies, Inc).

The plasmid used for the pulldown assay was generated by PCR of the gene encoding Rpd3 pGEM3Zf(+)-Rpd3 as template [5]. The PCR product was inserted into pET16b using NdeI and XhoI. This plasmid will be referred to as pET16b-Rpd3. pET16b-Rpd3 was generated by Wiam Turki-Judeh.

*Fly strains:*

The pUASP expression vector contains the white gene (*w*) that generates red eyes in transgenic *w*- flies. To transgenes inserted into the 2nd chromosome, transgenic flies were crossed with CyO/tft. The F1 progeny containing CyO were crossed with CyO/tft again. If the transgene is on the 2nd chromosome, all red eye F2 progeny will either have curly wings (CyO) or ectopic bristles in the dorsal mesothorax (tft) but not both. A similar approach was used to map transgene to the 3rd chromosome using a TM3/TM6B tester strain.

To generate flies homozygous transgenes on both the 2nd and 3rd chromosomes, 2nd or 3rd chromosome were introduced into a background with chromosome balancers. These flies were then crossed with each other, and F2 progeny were screened for the desired doubly homozygous flies.

*Cell culture:*

To establish the stable S2 cell lines used for the subcellular localization immunofluorescence experiments, the pMK33-GP variants were transfected into S2 cells using QIAGEN Effectene according to the manufacturer's protocol. Transfected cells were then selected with hygromycin B until stable cell lines were established. After the selection process, GP variant expression was induced by addition of 0.5 mM CuSO<sub>4</sub> to the cell culture medium.

*Subcellular localization immunofluorescence:*

Induced S2 cells were stained with 1:250 diluted mouse anti-Flag antibodies (Sigma). Secondary antibodies were goat-anti-mouse conjugated with Alexa Fluor 568 (Molecular Probes). DNA was stained with 1 ug/ml DAPI. Confocal images of S2 cells were obtained on a TCS SPE confocal laser scanning microscope (Leica Microsystem, Heidelberg) using a 100X objective. These experiments were carried out with assistance from Wiam Turki-Judeh.

*Pulldown assays with His-tagged Rpd3:*

GP variants and His-Rpd3 were translated in vitro in the presence of [35S]-methionine with the TNT T7 coupled reticulocyte lysate system (Promega) using the pET17b-GP variants and pET16b-Rpd3. 2% of the translated product was reserved for analysis as input, while the remaining 98% was diluted into 200 ul of binding buffer (25

mM HEPES pH 7.6, 450 mM NaCl, 10 mM imidazole, 0.1% Tween 20, 1 mM dithiothreitol) and incubated with Ni-NTA beads (QIAGEN) overnight at 4°C. Beads were then washed six times for 15 mins with binding buffer. Proteins bound to the beads were eluted with Laemmli sample buffer (60 mM Tris-Cl pH 6.8, 2% SDS, 10% glycerol, 5%  $\beta$ -mercaptoethanol, 0.01% bromophenol blue). Samples and reserved input were analyzed by 10% SDS-PAGE and autoradiography. This experiment was carried out with assistance from Wiam Turki-Judeh.

#### *Overexpression of chimeric Gro*

Transgenic male flies were crossed with virgin Serrate-Gal4 (ser-Gal4) females. 40 third instar wing imaginal discs from the progeny were placed into 30  $\mu$ l of Laemmli sample buffer and boiled for 5 mins. Samples were run for 2 hours in 8% SDS-PAGE and transferred to Immobilon-FL PVDF Membranes (Millipore). Proteins were detected with a mixture of mouse anti-Gro (Developmental Studies Hybridoma Bank, 1:200 dilution) and rabbit anti-EPRS (1:10000 dilution) antibodies. Immunoblots were subsequently probed with goat anti-mouse 680 and goat anti-rabbit 800 IR-dye coupled secondary antibodies (Li-Cor) and imaged with a Li-Cor Odyssey imager.

#### *Generation of Flp-FRT clones of flies overexpressing chimeric Gro*

Clonal overexpression of chimeric Gro was achieved as described previously [8, 9]. Briefly, hs-flippase; vgQ-LacZ; Actin>CD2>Gal4 virgin females were crossed with

homozygous pUASP-NLS Gro. After 2-3 days of mating, progeny were heat-shocked at 35°C for 40 mins as second instar larvae. Third instar wing imaginal discs were probed with mouse anti-CD2 (Serotec, 1:500 dilution) and rabbit anti- $\beta$ -galactosidase (ICN/Cappel, 1:10000 dilution) antibodies. Samples were subsequently probed with goat anti-mouse Alexa Fluor 488 and goat anti-rabbit Alexa Fluor 568 secondary antibodies (Molecular Probes). DNA was stained with 1  $\mu$ g/ml DAPI. Confocal images of wing imaginal discs were obtained on a TCS SPE confocal laser scanning microscope (Leica Microsystem, Heidelberg) using a 20X objective.

## Results

Since the GP domain is necessary both for the interaction with Rpd3 and nuclear localization, deletion of the entire GP domain results in cytoplasmic Gro [10]. Therefore, we first mapped the regions within GP domain that interacts with Rpd3 and that are required for nuclear localization. A series of deletion constructs were created that covered the entire GP domain (Figure 3-1A). The constructs were transfected into S2 cell to determine their cellular localization. We found that the last 14 residues of the GP domain is necessary for correct nuclear localization (Figure 3-1B). We then performed a pulldown assay and found that the last 30 residues of GP domain is necessary for the Rpd3 interaction (Figure 3-1C).

Since it is not possible to create a nuclear Gro that does not interact with Rpd3 by simply deleting the GP domain, we created a chimeric Gro with its GP domain replaced by a T-antigen NLS sequence. In addition, the construct contains a UAS binding site upstream of the promoter, which allows expression of protein in the presence of Gal4. Transgenic flies were then generated with the construct. We identified 5 lines containing 2nd chromosome insertions and 5 line containing 3rd chromosome insertions.

A previous study showed that overexpression of Gro led to abnormal wing development in a dosage dependent manner, with two-fold overexpression yielding mild to moderate wing veination defects and four-fold overexpression showing yielding severe wing deformation (unpublished data). Thus, we attempted to determine the adult wing



phenotype resulting from overexpression of the chimeric Gro with the wing specific ser-Gal4. No obvious wing developmental defect was observed in these overexpression, likely explained by low levels of transgene expression (Figure 3-2). To increase the expression level, flies were generated that were homozygous for both 2nd and 3rd chromosome transgenes. While the positive controls expressed the transgene at previously characterized level, the chimeric Gro yielded was not overexpressed, but was only expressed at levels comparable to that of wild-type Gro (Figure 3-3), precluding any conclusion from the phenotypic analysis.

Since the low level of overexpression precluded any conclusion about the repression activity of the chimeric Gro, we attempted to investigate the function of chimeric Gro with the vestigial quadrant LacZ reporter system (vgQ-LacZ). In wing development, vg is repressed by Gro. By putting the vg enhancer upstream of the LacZ reporter, the Gro repression activity can be determined by examining LacZ expression in these flies. We used a heat induced Flp-FRT system to drive the expression of Gal4, which in turn activates the expression of chimeric Gro. When CD2 is flipped out, Gro is expressed, which represses the expression of LacZ (Figure 3-4). Thus, the expression of CD2 and LacZ are mutually exclusive if Gro is functional. However, our result was inconclusive because the immunostaining for CD2 gave unexpected uniform staining of the entire wing imaginal disc, while we expected unstained patches in regions where CD2 was flipped out (data not shown).

## Discussion

We attempted to demonstrate a direct relationship of Rpd3 in Gro mediated repression through a Gro deletion that does not interact with Rpd3. We successfully generated a Gro chimera in which the GP domain was replaced with the T-antigen NLS sequence. However, we decided not to proceed further with the project due to technical difficulties.

We tried to measure the repression activity of the chimeric Gro in two different ways unsuccessfully. In the first method, we attempted to overexpress the chimeric protein and observe the phenotypic change in wing development. This experiment is highly dependent on the level of overexpression, as it has been shown that a 4 fold overexpression of wild type Gro was required to observe a significant phenotype (unpublished data). Unfortunately, we only observed levels of expression equal to that of endogenous Gro even with flies containing four copies of the transgene. The location of transgene is an important factor that affects the level of overexpression, as the transgene may land in a region with strong enhancer or vice versa. Unfortunately, the method we used to generate our transgenic flies would result in random insertion of transgene. To improve the experiment, I propose we should generate our transgenic flies with the CRISPR-Cas9 system, in which transgene can be specifically inserted into a highly expressed region [11].

In the second method, we attempted to employ the vgQ-LacZ reporter system to measure the activity of the chimeric Gro. We did not come to a conclusion because the CD2 immunostaining resulted in an unexpected uniform staining of the whole wing imaginal disc. It should be noted that the LacZ immunostaining gave an expected “lip” shape in the wing imaginal disc that represents the expression pattern of the vg enhancer (data not shown). Thus, it is unlikely that the uniform CD2 immunostaining was due to insufficient washing of antibodies. Possible improvements of the immunostaining include optimization of the CD2 antibodies concentration and search for an alternative CD2 antibodies. Another possible source of error would be the flippase failed to flip out CD2, which should be easily confirmed by PCR using primers flanking the FRT sites.

Based on our results, there is also a possibility that overexpression of chimeric Gro results in dominant negative lethality by disrupting histone deacetylation. First, any repressors recruiting the chimeric Gro will not be able to promote histone deacetylation. Second, the chimeric Gro can oligomerize with endogenous Gro that is already associated with the chromatin, reducing the level of histone deacetylation. Third, the chimeric Gro can also oligomerize with endogenous Gro that is not associated with the chromatin, preventing the functional from being recruited by repressors. Overexpression of the chimeric Gro will thus compete away functional Gro and drastically change the histone modification state, which may result in lethality. This could explain why we were not able to obtain the desired expression level in the phenotypic analysis or obtain clones in the

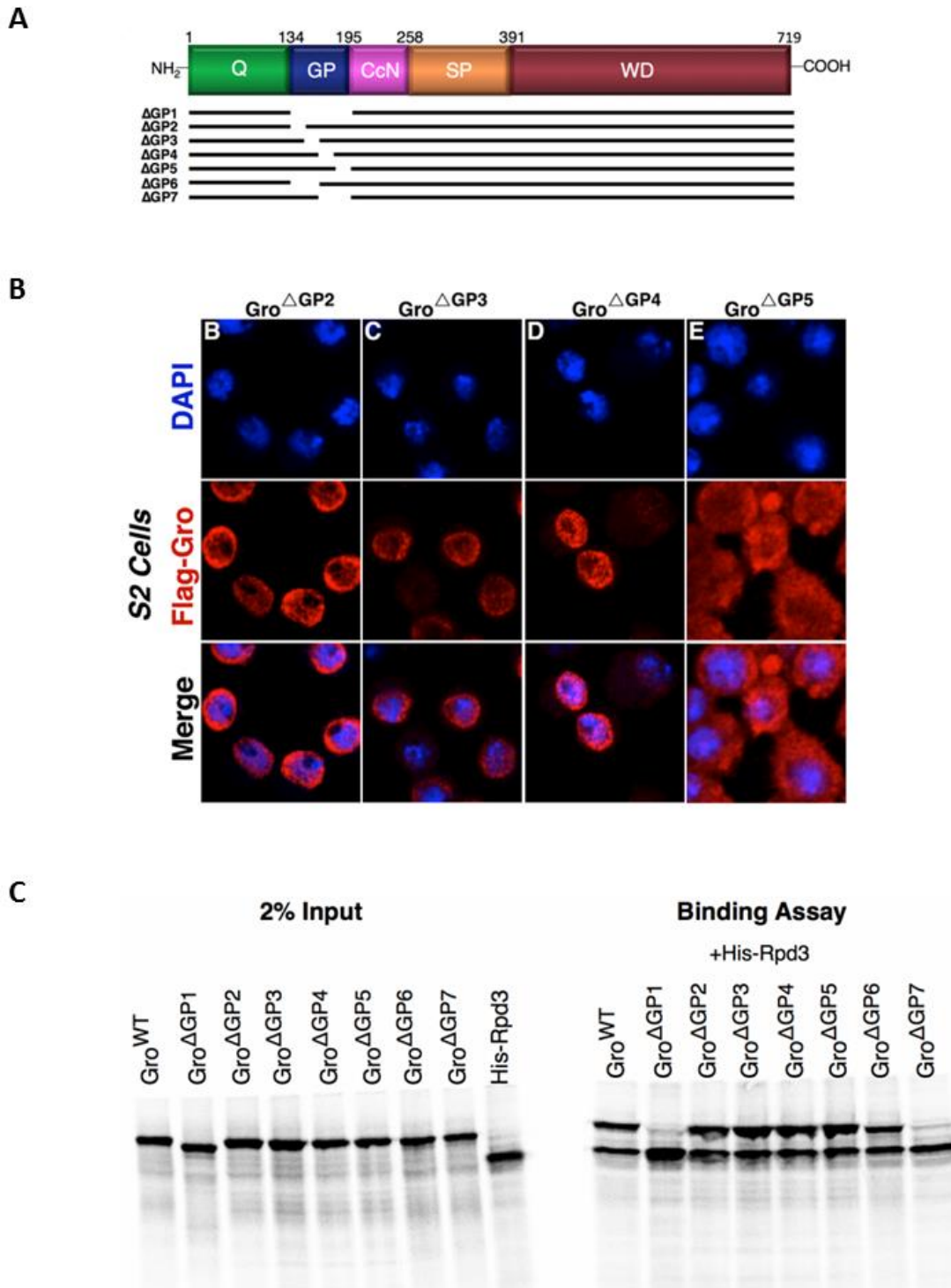
vgQ-LacZ reporter assay. It should be noted that neither cytoplasmic Gro nor nuclear Gro lacking other central domains caused lethality [10]. Therefore, if the negative dominant hypothesis is true, it requires Gro to be both nuclear and lack the histone deacetylase interacting domain.

Although the primary objective of the experiment was not fulfilled, our results are consistent with previous findings that the GP domain is necessary both for interaction with Rpd3 and nuclear localization.

**Figure 3-1. GP domain is necessary for nuclear localization and Rpd3 binding. (A)**

Schematic representation of GP deletion mutations. A series of GP deletion mutation was created to map the regions within GP domain that interacts with Rpd3 and nuclear localization.  $\Delta$ GP1 has the whole GP domain removed,  $\Delta$ GP2-5 each has a quarter of the GP domain removed, and  $\Delta$ GP6-7 have the first and second half of the GP domain removed, respectively. **(B)** S2 cells overexpressing the deletion constructs.  $\Delta$ GP5 results in mislocalization of Gro. **(C)** Pulldown assays with *in vitro* transcribed Rpd3 and GP deletion mutations.  $\Delta$ GP1 and  $\Delta$ GP7 are unable to bind Rpd3.

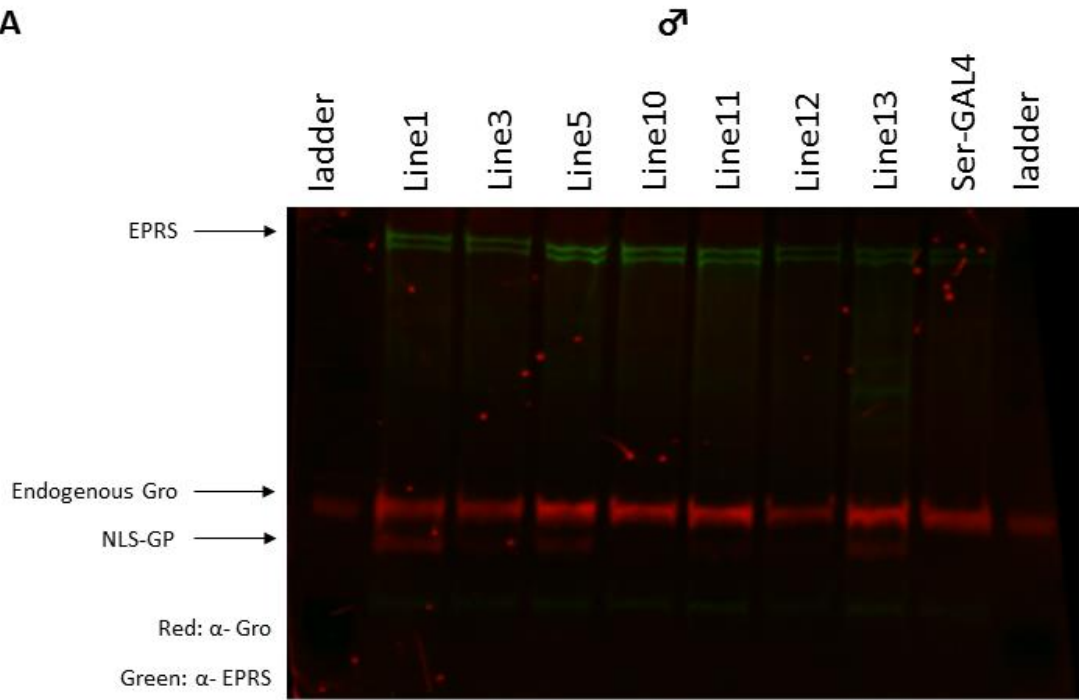
Figure 3-1



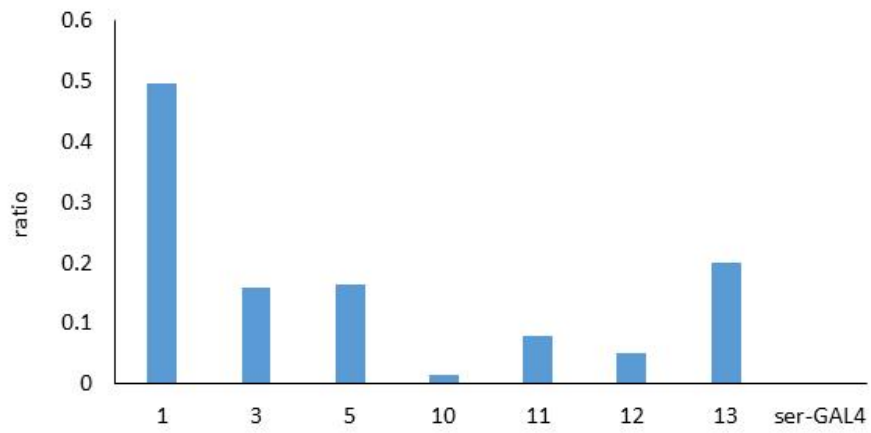
**Figure 3-2. Overexpression of chimeric Gro. (A)** Male transgenic flies containing chimeric Gro (or ser-GAL4 as negative control) were crossed with virgin ser-GAL4 females. Each line represents an independent insertion event. EPRS serves as loading control. **(B)** Quantitation of of the immunoblot. Ratio represents transgene expression over endogenous Gro expression.

Figure 3-2

A



B

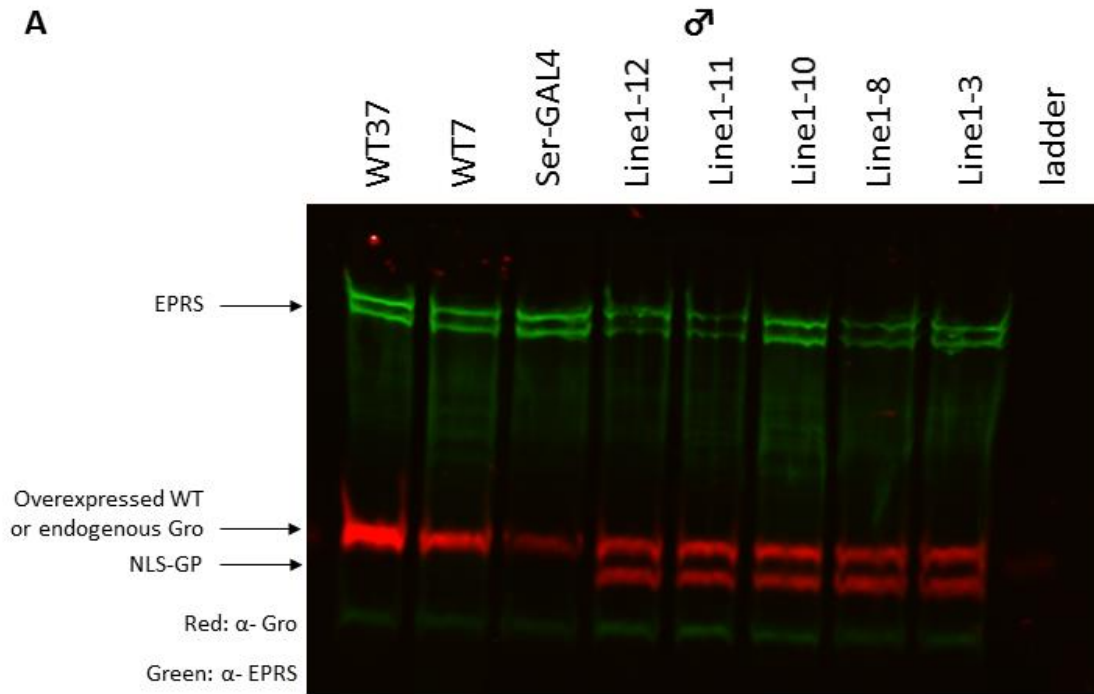




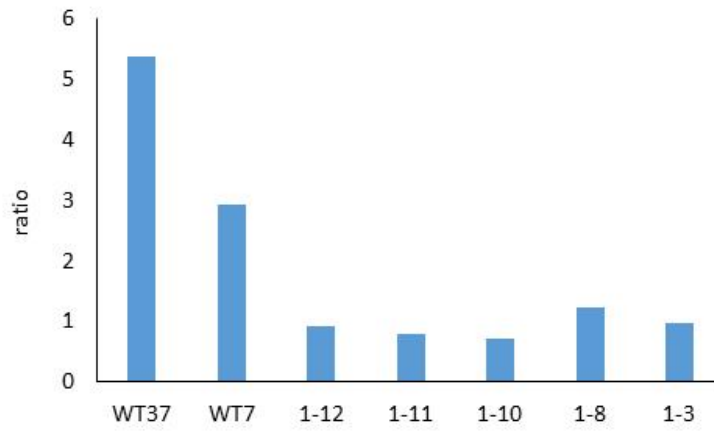
**Figure 3-3. Overexpression of chimeric Gro in flies containing two copies of the transgene. (A)** Male transgenic flies containing homozygous transgene in both 2nd and 3rd chromosome were crossed with virgin ser-GAL4. Each line represents a different cross of transgenic flies (eg. Line 1-10 is generated by crossing line 1 with line 10). WT37 and WT7 are two overexpression lines that were previous characterized to overexpress full length Gro by 4-fold and 2-fold, respectively. **(B)** Quantitation of overexpression from immunoblot. Ratio represents transgene expression over endogenous Gro expression. Ratio of WT37 and WT7 were determined by dividing their signals over that of ser-GAL4 (negative control).

Figure 3-3

A

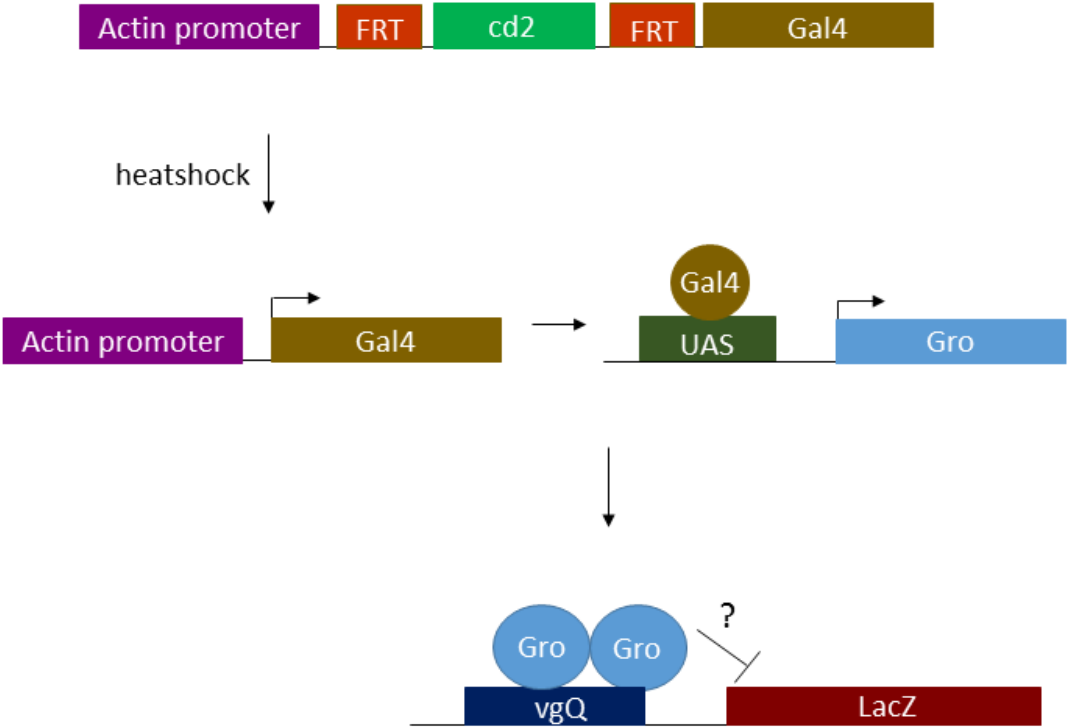


B



**Figure 3-4. Assaying Gro repression activity through the vgQ-LacZ reporter.** Schematic representation of the vgQ-LacZ reporter system. When CD2 is flipped out, Gro is expressed, which represses expression of LacZ. Thus, the expression of CD2 and LacZ are mutually exclusive if Gro is functional.

Figure 3-4



## References

1. Jurkin, J., et al., *Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis*. Cell Cycle, 2011. **10**(3): p. 406-12.
2. Poulouse, N. and R. Raju, *Sirtuin regulation in aging and injury*. Biochim Biophys Acta, 2015. **1852**(11): p. 2442-55.
3. Witt, O., et al., *HDAC family: What are the cancer relevant targets?* Cancer Lett, 2009. **277**(1): p. 8-21.
4. Yang, X.J. and E. Seto, *The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men*. Nat Rev Mol Cell Biol, 2008. **9**(3): p. 206-18.
5. Chen, G., et al., *A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development*. Genes Dev, 1999. **13**(17): p. 2218-30.
6. Winkler, C.J., A. Ponce, and A.J. Courey, *Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density*. PLoS One, 2010. **5**(4): p. e10166.
7. Wolff, B., et al., *Antibodies against the SV40 large T antigen nuclear localization sequence*. Arch Biochem Biophys, 1991. **288**(1): p. 131-40.
8. Hasson, P., et al., *Brinker requires two corepressors for maximal and versatile repression in Dpp signalling*. EMBO J, 2001. **20**(20): p. 5725-36.
9. Song, H., et al., *Groucho oligomerization is required for repression in vivo*. Mol Cell Biol, 2004. **24**(10): p. 4341-50.
10. Turki-Judeh, W. and A.J. Courey, *The unconserved groucho central region is essential for viability and modulates target gene specificity*. PLoS One, 2012. **7**(2): p. e30610.
11. Gratz, S.J., et al., *CRISPR-Cas9 Genome Editing in Drosophila*. Curr Protoc Mol Biol, 2015. **111**: p. 31.2.1-31.2.20.

## **Chapter 4**

### **A revised model of Gro mediated repression**

## **Introduction**

As mentioned in previous chapters, Gro is a corepressor that silences gene expression by multiple mechanisms. Numerous models have been proposed for the mechanism of Gro-mediated repression, including: changes higher order chromatin structure formation through histone deacetylation and Gro oligomerization, interaction with Mediator complex to prevent assembly of the pre-initiation complex, and masking of activation domain of other transcription factors [1]. However, these models do not fully explain the functional importance of the disordered central domains. Deletion of the central domains resulted in changes the ability of Gro to direct repression, suggesting that these domains are essential for properly regulated transcriptional repression [2]. Computational analysis of their structures led us to hypothesize that they are intrinsically disordered, allowing them to mediate multiple protein-protein interaction [2]. In chapter 2, I have identified over 160 Gro interacting proteins through mass spectrometry and validated the interactions through both reporter assays and RNA-seq. Many of these proteins associate with one another in a variety of multi-protein complexes. Several of them (e.g., the core histones, CKII) were previously characterized as Gro interactors thus partially validating the screen. In addition, we reveal a possible mechanism of Gro mediated repression through an interaction with the spliceosome complex or subcomplexes.

### **Gro interacts with chromatin remodeling factors**

One of the proposed mechanisms of Gro-mediated repression suggest that recruitment and oligomerization of Gro will promote the formation of a higher order chromatin structure that is inaccessible to transcription machinery. Previous studies from our lab and other labs show that the GP domain interacts with the histone deacetylase Rpd3/HDAC1, which may promote local histone deacetylation and alter nucleosome density [3, 4]. Although we failed to identify Rpd3 in our GST-GP sample, the identification of the ACF chromatin remodeling complexes as a central region interacting protein complex provides further support for the idea that regulation of chromatin structure is a critical aspect of Gro mediated repression. However, modulation of chromatin structure is likely not the only mechanism of Gro mediated repression. Rpd3 knockdown did not fully restore reporter expression in our RNAi screen, consistent with published data that histone deacetylase inhibitors and Rpd3 knockdown reduce but do not abolish Gro-mediated repression [3, 4].

### **Regulation of Gro activity by the CKII complex**

A previous study showed that the CKII complex phosphorylates Gro at multiple sites including serines 239 and 253, and that mutagenesis of Ser 239 to alanine attenuates repression presumably by preventing phosphorylation of this residue [5]. We have identified both the catalytic subunit (CKII $\alpha$ ) and regulatory subunit (CKII $\beta$ ) of the CKII complex as Gro interacting proteins. In addition, we have also identified Nopp140, a well-known negative regulator of CKII [6]. However, our reporter assay data is inconsistent with



the view that CKII is a positive regulator of Gro and that Nopp140 acts by inhibiting CKII. This is because our reporter assays show that CKII $\alpha$ , CKII $\beta$ , and Nopp140 are all negative regulators of Gro. Thus, consistent with other findings showing that Gro phosphorylation can block repression, CKII may primarily be a negative regulator of Gro [1]. Furthermore, the effect we observe due to Nopp140 knockdown could be due to the role of this factor in processes other than CKII regulation [7].

### **A novel interaction between Gro and U1 snRNP complex**

The U1 snRNP complex consists of U1 snRNA, seven Sm proteins (B/B', D1, D2, D3, E, F, G) and three U1 specific proteins (U1-C, U1-70K, and U1-A) [8]. We were able to identify almost all of these components as Gro interacting proteins. In addition, we independently confirmed the U1 snRNP interaction through Gro co-immunoprecipitation. Data from our reporter assay further suggests that the U1 snRNP complex is required for optimal Gro mediated repression, as U1-C and U1-70K knockdown attenuated repression. The N-terminal region of U1-70K is necessary and sufficient to recruit U1-C [9], which may explain our finding that loss in either of the proteins reduced Gro activity. Consistent with our finding, it has been shown that U1-C overexpression can repress expression activated by EWS/FLI [10]. It is worth noting that the U1 snRNA is known to associate with TFIIH and promote transcriptional initiation in vitro [11]. Thus, the effect of the U1 snRNP complex in transcription regulation may be context dependent.

### **Interaction of Gro with CTD kinase may promote paused Pol II**

Pol II pausing is one mechanism to repress gene expression [12, 13]. Genes that are up-regulated in Gro knockdown cells exhibit enrichment in Pol II pausing near the transcriptional start site. In addition, our proteomic screen revealed the CTD kinase Cdk12 as a Gro-interacting protein. Cdk12 promotes paused Pol II to restart elongation by phosphorylating Ser2 of CTD [14]. Thus, it is possible that inhibition of Cdk12 by Gro leads to Pol II pausing.

### **A revised model of Gro repression**

How might the spliceosome contribute to Gro mediated repression? A recent study has shown that nascent RNA transcripts are able to recruit transcription factors to their binding sites [15]. Thus, I propose that the interaction between Gro and U1 snRNP complex will recruit Gro to its binding sites. Nascent RNA containing promoter proximal introns recruits the U1 complex, which brings Gro in close proximity to its repressors. This is particularly important if the interaction between Gro and the transcription factor is weak. In addition, by bringing Gro near the CTD through interaction with the U1 complex, Gro may inhibit Cdk12, thus preventing CTD Ser2 phosphorylation and resulting in Pol II pausing (Figure 4-1). To test this hypothesis, I propose to knockdown the U1 snRNP complex and perform ChIP on Gro and Pol II. We would expect a decrease in promoter proximal Gro occupancy and Pol II pausing if our hypothesis is true.

Phosphorylation of the SP domain by MAPK resulted in reduction of Gro activity, suggesting a negative regulation of Gro activity by the EGFR signaling pathway [16]. Based

on the finding that U1 snRNP complex interacts with Gro primarily through the SP domain, I proposed that phosphorylation of the SP domain, by increasing the negative charge of the domain, could impair the interaction with the negatively charged snRNA. Since interaction of Gro and the snRNP complex positively regulates Gro activity, phosphorylation of the SP domain might disrupt this interaction interfering with Gro-mediated repression (Figure 4-2). To test this hypothesis, I propose to repeat the Gro co-IP experiment with mutations in the phosphorylation sites. Phosphomimetic substitutions (e.g., glutamate substitutions) should result in reduced snRNA pulldown, while phosphorylation-defective substitutions (e.g., alanine substitutions) should result in increased snRNA pulldown.

### **Concluding remarks**

Our experiments corroborate and extend published findings showing that Gro promotes higher order chromatin structure formation through interactions with chromatin remodeling factors. In addition, we confirmed a previous proposed interaction of Gro and the CKII complex. We discovered a novel interaction between Gro and the spliceosome complex, and we propose a revised model that incorporates our finding with the current models of Gro repression mechanism.

**Figure 4-1. Revised model of Gro mediated repression.** Interaction of Gro with the snRNP U1 complex may recruit Gro to its binding site, which then promotes the formation of higher order chromatin structure. In addition, the U1 complex may allow Gro to be in close contact with CTD, which then inhibits phosphorylation of Ser2 by Cdk12 and results in paused Pol II.

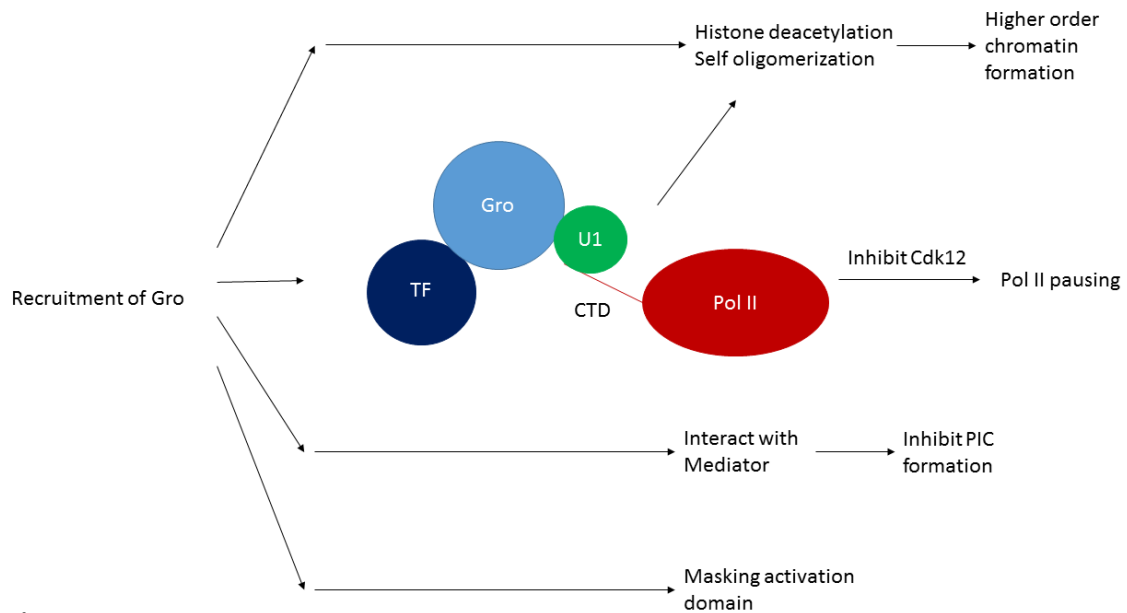


Figure 4-1

**Figure 4-2. Proposed model of phosphorylation in Gro mediated repression.**

Phosphorylation of Gro by EGFR pathway may disrupt the interaction between Gro and the snRNP-U1 complex, leading to attenuation of repression.

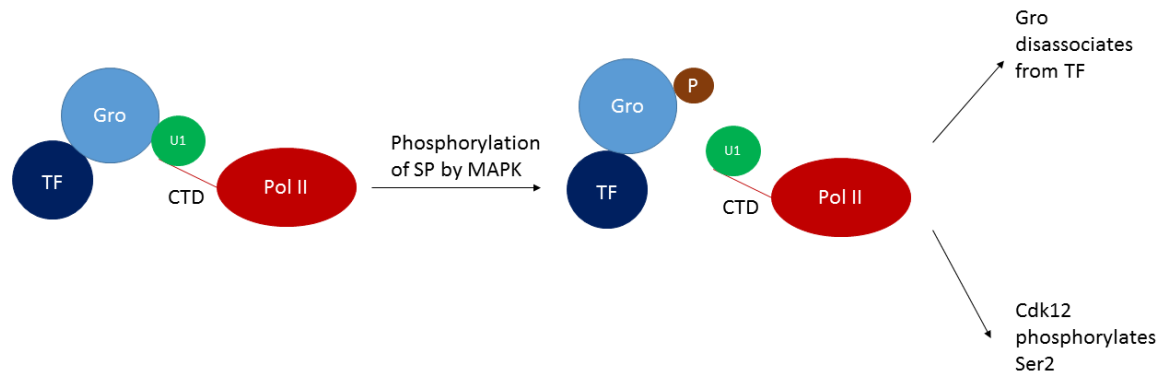


Figure 4-2

## References

1. Turki-Judeh, W. and A.J. Courey, Groucho: a corepressor with instructive roles in development. *Curr Top Dev Biol*, 2012. 98: p. 65-96.
2. Turki-Judeh, W. and A.J. Courey, The unconserved groucho central region is essential for viability and modulates target gene specificity. *PLoS One*, 2012. 7(2): p. e30610.
3. Chen, G., et al., A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in *Drosophila* development. *Genes Dev*, 1999. 13(17): p. 2218-30.
4. Winkler, C.J., A. Ponce, and A.J. Courey, Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density. *PLoS One*, 2010. 5(4): p. e10166.
5. Nuthall, H.N., K. Joachim, and S. Stifani, Phosphorylation of serine 239 of Groucho/TLE1 by protein kinase CK2 is important for inhibition of neuronal differentiation. *Mol Cell Biol*, 2004. 24(19): p. 8395-407.
6. Lee, W.K., et al., Structural and functional insights into the regulation mechanism of CK2 by IP6 and the intrinsically disordered protein Nopp140. *Proc Natl Acad Sci U S A*, 2013. 110(48): p. 19360-5.
7. Meier, U.T. and G. Blobel, Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell*, 1992. 70(1): p. 127-38.
8. Pomeranz Krummel, D.A., et al., Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, 2009. 458(7237): p. 475-80.
9. Nelissen, R.L., et al., The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J*, 1994. 13(17): p. 4113-25.
10. Knoop, L.L. and S.J. Baker, The splicing factor U1C represses EWS/FLI-mediated transactivation. *J Biol Chem*, 2000. 275(32): p. 24865-71.
11. Kwek, K.Y., et al., U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol*, 2002. 9(11): p. 800-5.



12. Adelman, K. and J.T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet*, 2012. 13(10): p. 720-31.
13. Gaertner, B. and J. Zeitlinger, RNA polymerase II pausing during development. *Development*, 2014. 141(6): p. 1179-83.
14. Malumbres, M., Cyclin-dependent kinases. *Genome Biol*, 2014. 15(6): p. 122.
15. Sigova, A.A., et al., Transcription factor trapping by RNA in gene regulatory elements. *Science*, 2015. 350(6263): p. 978-81.
16. Hasson, P., et al., EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. *Nat Genet*, 2005. 37(1): p. 101-5.