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

Pak N. Kwong[1], Michael Chambers[1], Ajay A. Vashisht[2], Wiam Turki-Judeh[1,3], Tak Yu Yau,[1] James A. Wohlschlegel[2,3], and Albert J. Courey[1,3]

[1]Department of Chemistry and Biochemistry
[2]Department of Biological Chemistry
[3]Molecular Biology Institute
University of California, Los Angeles 90095

Running title: *Groucho/Spliceosome Interactions*

To whom correspondence should be addressed: Albert J. Courey, Department of Chemistry and Biochemistry, 607 Charles E. Young Drive East, Los Angeles, CA 90095-1569; Telephone: (310) 825-2530; Email: courey@chem.ucla.edu

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.

ABSTRACT

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 to Gro and Pol II ChIP data led to a number of insights including the finding that Gro-repressed genes are enriched for promoter proximal Pol 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) 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 basepairs (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 β-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, since deletion of this domain prevents Gro nuclear uptake (19). The SP domain regulates Gro function negatively, as its deletion leads to promiscuous repression and 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 employing 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 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 GST fusion proteins, sequences encoding the Gro domains were amplified by PCR and inserted between the BamHI and XhoI sites of pGEX4T (GE Healthcare Life Sciences). 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 Cat.# E1411) between the KpnI and XhoI sites. The green luciferase plasmid, DE5G5-pCBG68, was generated by inserting the luciferase gene using NcoI and SalI from pCBG68-basic vector (Promega Cat.# E1431) into the DE5 G5 vector, which has UAS elements downstream of the reporter (unpublished data). Actin promoter driven Dorsal (pPac Dl), Twist (pPac Twi) and Gal4-Gro (pAct Gal4-Gro) plasmids have been previously described (14). The RpIII128 promoter driven Renilla luciferase plasmid, RpIII128-Rluc, was obtained from Addgene (ID #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 mid-log cells were induced with 0.25 mM IPTG for an hour. Cells were pelleted at 4000 x g, resuspended in 25 mL of Salty TE (0.15 M NaCl, 10 mM Tris pH8, 1 mM EDTA) with protease inhibitor (Life Technologies, Cat.# 88266), and incubated on ice for 30 min. Samples were incubated at 4° 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 14000 x g for 10 min at 4°. Supernatant was collected, and 1 ml of glutathione agarose resin (50% slurry) was added. After overnight incubation, the resin was washed with cold PBS three times and stored at 4°.

Drosophila embryo nuclear extracts were prepared as previously described (32). To isolate Gro-interacting proteins, 20 μg of glutathione bead-immobilized recombinant domains were 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 MgCl2, 0.2 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.1 M KCl) at 4° overnight. Samples were washed six times for 15 minutes 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 MudPIT analysis. MudPIT analysis was performed as previously described (33). Peptide identifications were filtered using a false discovery rate (FDR) 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.

Table S1C includes all the mass spectroscopy data for the two independent replicate screens carried out with each GST fusion protein and GST alone, while Tables S1A and B include 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 (see notes 2 and 3 to Table S1B). Ribosomal proteins were excluded from the lists in Tables S1A and B.

Gro immunoprecipitation and reverse transcriptase qPCR (RT-qPCR) analysis of U1 snRNA–500 μg of nuclear extract was incubated with 1.875 μg of affinity purified rabbit antibody against the Gro GP domain or rabbit IgG in a final

volume of 250 μl HEMNK buffer overnight at 4˚. 225 μg of Protein A Dynabeads (Invitrogen Cat.# 10001D) were incubated with the samples at 4˚for 1 hour. Samples were then washed with HEMNK buffer three times for 10 minutes each. For RT-qPCR, RNA was eluted in 10 μl water by heating to 80˚ for 2 minutes. Samples were treated with DNase I according to the manufacturer's protocol (Promega Cat.# M6101). Reverse transcription was performed with 300 ng of random primer (Invitrogen Cat.# 48190-011), and qPCR was performed using primers amplifying U1 snRNA (Table 2). Threshold cycle values were converted to percent input values by comparison to a standard curve generated from multiple serial dilutions of RNA isolated by Trizol extraction (Life Technologies, Cat.# 10296010) from the input nuclear extract. Primer specificity was validated by melting curve analysis of the amplification products (data now 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 IR-dye 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 re-aliquoted into white flat bottom 96 well plates (USA scientific Cat.# CC7682-7968) with 150 ng/well in 10 μl of water using a Beckman Coulter BioMek FX Workstation.

Transfections were carried out with Effectene reagent (Qiagen Cat.# 301425). 6 μg each of G5DE5-pCBR and DE5G5-pCBG68, 0.6 μg of RpIII128-Rluc, 1 μg of pPac Dl, 0.3 μg of pPac Twi, and 1.2 μg of the pAct Gal4-Gro were

suspended in 600 μl buffer EC. 33 μl of this mixture was added to 25 μl of enhancer. After 2-3 minutes, 7.5 μl Effectene was added and mixed by pipetting up and down. 6 μl of this mixture was immediately added into each well of a 96-well plate containing 150 ng of dsRNA. 4-8 minutes later, 100 μl of S2 cells (diluted to 1 x 10^6 cell/ml) was added to each well. Cells were incubated at 24˚ for 2 days before assaying.

The luminescence signal was measured with a Molecular Devices LJL Analyst HT microplate reader using emission filters ET510/80m and E610LP (Chroma Cat.#S-022658 and #138951). 50 μl of D-luciferin (Chroma-Glo system, Promega Cat.# E2980) was added to each well. Five minutes later the reaction was stopped by the addition of 50 μl of stop buffer containing coelenterazine (Dual-Luciferase system, Promega Cat.# 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:

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

$$G' = \frac{Lgf - R' \times \left(\frac{Rgf}{R}\right)}{\left(\frac{Ggf}{G}\right)}$$

Parameters were determined by expressing the individual luciferases and recording the luminescence signals with red and green filters, and without filter (data not shown). The ratio of green signal passed through the red filter, Grf/Ggf, was determined to be 0.0975; the ratio of red signal passed through the red filter, Rrf/R, was determined to be 0.42; the ratio of red signal passed through the green filter, Rgf/R, was determined to be 0; the ratio of green signal passed through the green filter, Ggf/G, was determined to be 0.47. Lrf and Lgf are luminescence signals in

4

which cells are co-transfected with both red and green luciferases. Lrf is the signal recorded with red filter, and Lgf is the signal recorded with 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 to 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 p < 0.1. If multiple dsRNAs were tested for a given gene (as was true in most cases, Table S2), then a change is only listed if p<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'-taatacgactcactatagggtactCAAAGTACTATTGCGACTACTGC and 5'-taatacgactcactatagggtactCTTGGGTCCGTTCATGATTCC. Transfection was carried out as previously described (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 (Life Technologies, Cat.# 10296010). RNA integrity was determined with an Agilent 2100 Bioanalyzer using the RNA 6000 Nano Kit (Agilent Cat.# 5067-1511). Isolation of mRNA was carried out as follows. Streptavidin magnetic beads (Promega

Cat.# Z5481) were prepared in aliquots of 120 μl and 60 μl in 0.5X SSC with 10 mM EDTA. 15 μg of total RNA was mixed with 1.5 μM of biotinylated 15-mer poly(T) oligonucleotide in 0.5X SSC with 10 mM EDTA. Samples were first incubated at 75° for 5 minutes, followed by 15° for 10 minutes and 10° for 10 minutes. Samples were then incubated with 120 μl of magnetic beads at 4° for 2 hours, followed by 60 μl of magnetic beads at 4° for 30 min. The two aliquots of beads were combined and washed four times with 300 μl of ice cold 0.1X SSC containing 10 mM EDTA. mRNA was first eluted with 100 μl of water followed by 150 μl of water at 37° for 10 min each. Samples were precipitated with ethanol and stored at -80°. 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 (Epicentre, Cat.# SSV21124 and Cat.# RSBC10948). The concentration of the library was determined with Pico Green (Life Technologies, Cat.# Q32851) according to the manufacturer's directions. Fluorescence signal was measured using a TECAN M1000 fluorescent plate reader.

Bioinformatics–Alignment of paired-end reads to the D. 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 was 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 domains in the Gro central region led to a loss of Gro-mediated repression and to lethality, while 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 glutathione-S-transferase (GST)-tagged proteins and purified from E. coli lysates (Figure 1A, B). We also constructed a similarly tagged form of the N-terminal Q domain since previous studies suggested that, in addition to mediating Gro oligomerization, the Q domain engages in interactions with regulatory targets (39,40).

The glutathione bead-immoblized 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 multi-dimensional protein identification technology (MudPIT) (33) (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 (Figure 1C). With three exceptions (see Experimental Procedures), only proteins that appeared in both replicates were included in our list of Gro interacting proteins (Figure 1C, Table S1A, 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 the 162 Gro-interacting proteins associated uniquely with one domain (in all but one case, the SP domain), while 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 (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 (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, Table S1A, B) contains multiple components of known multisubunit protein complexes. For example, we identified the α and β 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), Acf1 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 immunopreciptated material demonstrates the efficiency of the immunoprecipitation (Figure 2A). RNA was extracted from the immunoprecipitates and analyzed 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 hour embryos is associated with Gro (Figure 2B).

Functional analysis of Gro interacting proteins–We next carried out functional assays to determine if the interacting proteins are required for regulation of a Gro-responsive reporter gene. Previous studies established a reliable reporter assay for Gro function employing 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 (Figure 3A, 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)) (Figure 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 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 approximately 1300 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). Forty-four 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 (Figure 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 (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 (Tables 5 and S2) include both components of the CKII complex (CKIIα, CKIIβ), 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/FLI-transactivated genes (30). Since 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 to that of Gro knockdown on the gene expression profile in S2 cells. Cells were treated with Gro or snRNP-U1-C dsRNA for four days, leading to four-fold or greater knockdown of the Gro and snRNP-U1-C mRNA (Figure 4A). The transcriptomes in wild-type and Gro knockdown S2 cells were quantitatively similar to those published previously (49,50) (Figure 4B, 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 intron-less genes consistent with the idea that snRNP-U1-C has roles in gene regulation apart from its role in splicing (Figure 4D). We note that changes in the expression of an intron-less gene could also reflect a requirement for the product of an intron-containing gene in the expression of the intron-less gene.

98 genes were differentially expressed in both Gro and snRNP-U1-C knockdown cells (Figure 4E), of which 36 were upregulated in either case. These coordinately upregulated targets included genes in various signaling pathways, such as the Wnt, Notch, and Toll pathways (Table 6). Comparison with publically 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 (Figure 4F).

To determine if the regulatory effects of knocking down Gro are likely to be direct, we compared our RNA-seq data from Gro knockdown S2 cells to available S2 cell Gro ChIP data (49). Gro appears to bind many genes that it does not repress (Figure 5A). This is consistent with

observations made with numerous regulatory factors (51,52) and suggests that binding, while required, is not sufficient for regulation. We observe 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 (Figure 5B). Comparison of our RNA-seq data from Gro knockdown cells to 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; Figure 6).

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, which were then identified by MuDPIT. We identified over 160 interacting polypeptides, many of which associate with one another in a variety of multi-protein complexes. Several of these interacting proteins (e.g., the core histones, 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 employed 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 while 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 (CKIIα) and two copies of a regulatory subunit (CKIIβ) (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 α and β 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 CKIIα, CKIIβ, 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 observe 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 have 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), while 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 (CTD) of the RNA 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 since different promoters can yield transcripts that are subject to differential alternative splicing (69,70). While 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 (Tables S1B 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) (Table S1C). Additionally, we showed by co-immunoprecipitation that approximately 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 suggests 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 TFIIH 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) reveals 1242 Gro binding sites in the S2 cell genome associated with 748 genes, while 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 recruit 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 which 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 (CTD) kinase Cdk12 as a Gro-interacting protein (Table S1). By phosphorylating the CTD 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 (Tables 5 and 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. While 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 of that the Gro central domains, which are intrinsically disordered, are indispensable for repression (19). 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 (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, provides further support for the idea that regulation of chromatin structure is a critical aspect of Gro mediated repression. On the other hand, 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) (Figure 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.

Conflict of interest: The authors declare that they have no conflict of interest with the contents of this article.

Author contributions: AJC, PNK, and WT-J conceived and planned the study, which was coordinated by AJC. PNK and AJC wrote the manuscript. PNK and WT-J conducted most of the experiments. TYY assisted with the co-immunoprecipitation study. MC carried out the bioinformatics analysis. AAV and JAW carried out the MuDPIT analysis. All authors reviewed and approved the manuscript.

REFERENCES

1. Courey, A. J., and Jia, S. (2001) Transcriptional repression: the long and the short of it. *Genes Dev* **15**, 2786-2796
2. Turki-Judeh, W., and Courey, A. J. (2012) Groucho: a corepressor with instructive roles in development. *Current topics in developmental biology* **98**, 65-96
3. 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
4. Chen, G., and Courey, A. J. (2000) Groucho/TLE family proteins and transcriptional repression. *Gene* **249**, 1-16
5. Gasperowicz, M., and Otto, F. (2005) Mammalian Groucho homologs: redundancy or specificity? *J Cell Biochem* **95**, 670-687
6. Nagel, A. C., Krejci, A., Tenin, G., Bravo-Patino, 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
7. Blair, S. S. (2007) Wing vein patterning in Drosophila and the analysis of intercellular signaling. *Annual review of cell and developmental biology* **23**, 293-319
8. Hasson, P., and Paroush, Z. (2006) Crosstalk between the EGFR and other signalling pathways at the level of the global transcriptional corepressor Groucho/TLE. *British journal of cancer* **94**, 771-775
9. Allen, T., van Tuyl, 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

10. Buscarlet, M., and Stifani, S. (2007) The 'Marx' of Groucho on development and disease. *Trends Cell Biol* **17**, 353-361

11. Jennings, B. H., Pickles, L. M., Wainwright, S. M., Roe, S. M., Pearl, L. H., and Ish-Horowicz, D. (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell* **22**, 645-655

12. 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

13. 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

14. Chen, G., Nguyen, P. H., and Courey, A. J. (1998) A role for Groucho tetramerization in transcriptional repression. *Mol Cell Biol* **18**, 7259-7268

15. 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

16. 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

17. 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

18. 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

19. 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

20. 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

21. 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

22. 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

23. Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding. *Biochim Biophys Acta* **1804**, 1231-1264

24. Tantos, A., Han, K. H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription. *Mol Cell Endocrinol* **348**, 457-465

25. Bondos, S. E., and Hsiao, H. C. (2012) Roles for intrinsic disorder and fuzziness in generating context-specific function in Ultrabithorax, a Hox transcription factor. *Adv Exp Med Biol* **725**, 86-105

26. 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

27. 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

28. 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

29. 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 TFIIH and regulates transcriptional initiation. *Nat Struct Biol* **9**, 800-805

30.   Knoop, L. L., and Baker, S. J. (2000) The splicing factor U1C represses EWS/FLI-mediated transactivation. *J Biol Chem* **275**, 24865-24871

31.   Armknecht, 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 in enzymology* **392**, 55-73

32.   Soeller, W. C., Poole, S. J., and Kornberg, T. (1988) In vitro transcription of the Drosophila engrailed gene. *Genes Dev* **2**, 68-81

33.   Wohlschlegel, J. A. (2009) Identification of SUMO-conjugated proteins and their SUMO attachment sites using proteomic mass spectrometry. *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 Harbor protocols* **2013**, 461-463

35.   Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, 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 biology* **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 biology* **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. *Nature structural & molecular biology* **12**, 364-371

41.   Bouazoune, K., and Brehm, A. (2006) ATP-dependent chromatin remodeling complexes in Drosophila. *Chromosome Res* **14**, 433-449

42.   Flores-Saaib, 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 Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**

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. *Chemistry & biology* **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 genetics* **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 mod, E. C. (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., Weiszmann, 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. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21330-21335

52.   Walhout, A. J. (2011) What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome biology* **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. *Current biology : CB* **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 molecular biology* **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., Caceres, 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. *Proceedings of the National Academy of Sciences of the United States of America* **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., Muller, 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-Horowicz, 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. *Mutation research* **618**, 130-134

79.   Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nature reviews. Molecular cell biology* **13**, 115-126

TABLES

Table 1: PCR primers used in construction of plasmids encoding GST fusion proteins

Domain	Sequence
Q	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAATATCCCTCACCGGTGCGCCACCCC
	ATTATTACTCGAGTCACTGCTGGGCGTGGATCTGTTGCCCA
GP	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGTGCCAGGTGGACCACCTCAGCCGA
	ATTATTACTCGAGTCACGAATTGAGCAATCGCTCCTCGGC
CcN	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGTTTCGCCGGCCGATCGTGAGAAGT
	ATTATTACTCGAGTCACATAGACACGTGCTCGCCGTTGGGA
SP	ATTATTAGGATCCATGGATTACAAGGACGATGACGATAAGAGGTGCGCGATCGGGAAAGCTTGA
	ATTATTACTCGAGTCAACCCGTTAGGGCCGAGGGATGTGGA

Table 2: RT-qPCR primers

Gene	Sequence
Gro	TTTATTACAACATGTTCGAAATCATGC
	TTCGCTTTTTGATGCGTTGCTAC
snRNP-U1-C	CTCAGGAACGGCATCAACGTT
	TATAATTAATTGTTTTCGCTATCGGG
Rpl32	CCCAAGGGTATCGACAACAGA
	CGATCTCGCCGCAGTAAAC
U1 snRNA	ATACTTACCTGGCGTAGAGGTTAACC
	AACGCCATTCCCGGCTA

Table 3. Enriched gene ontology groups of Gro-interacting proteins[1]

Enriched gene ontology	# 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

[1]$p<0.05$

Table 4. Representative Gro-interacting proteins[1]

Chromosome organization	
Acf1	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
Developmental process	
CKIIα	CKII complex
CKIIβ	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
mRNA processing	
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

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

17

Table 5. Positive and negative Gro-regulators[1]

Potential negative regulators of Gro[2]	Potential positive regulators of Gro[3]
CKIIα, CKIIβ, 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, Acf1, snama, CG1622, ZCHC8, CG4709, CG4806, lat, Srp19

[1]See 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.

[2]Negative regulators are defined as the products of those genes the knock down of which led to increased repression by Gal4-Gro in the reporter assay.

[3]Positive regulators are defined as the products of those genes the knock down of which led to decreased repression by Gal4-Gro in the reporter assay.

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
Syncrip	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

Table S1: Gro interacting proteins

See Excel spreadsheet

Table S2: Gro-interacting proteins that have significant effects on repression by Gal4-Gro

See Excel spreadsheet

FIGURE LEGENDS

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, which were subsequently identified by MuDPIT (see Tables 4 and 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 ($p < 2.2 \times 10^{-16}$).

Figure 2. Validation of the interaction between Gro and U1 snRNP. (A) 0-12 hour 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-GP domain antibody and a mouse monoclonal anti-Gro antibody, and IR-dye labeled secondary antibodies. The signal from the rabbit antibody was detected in the green channel of the IR imager, while 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 kD; Lane 2) 10% input; Lane 3) Anti-Gro immunoprecipitate; 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 in Experimental Procedures to determine U1 snRNA levels. Error bars based on two independent biological replicates indicate standard deviation. A two-tailed t-test gives $p = 0.016$.

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, while in the green luciferase reporter, the UAS elements are about 2 kb downstream of the transcriptional start site. Expression is induced by the Dorsal and Twist activators and repressed by Gal4-Gro. The Renilla luciferase reporter was used as an internal control for transfection efficiency and cell viability. (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, while addition of a plasmid encoding the Gal4-Gro fusion resulted in repression of the reporters. Dorsal, Gro (including Gal4-Gro), and the histone deactylase 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 standard deviation. (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 SRPK. The result of transfection with each dsRNA was compared to that of transfection with GFP dsRNA. Error bars based on triplicate transfection assays represent standard deviation.

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 was normalized to reference gene Rpl32. (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 $\log_2(CPM)$ 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 (1,691genes). Percent 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-U1C differentially expressed genes shown in part E of this figure. (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 are calculated using cumulative recovery curves (37). Scores above 2.5 are considered significant.

Figure 5. Gro binding regions in differentially expressed genes. (A) S2 cell ChIP-seq data (49) identified 1242 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 7 were down regulated in response to Gro knock-down. (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. Percent of non-differentially expressed genes, and genes that are either up-regulated or down regulated in Gro knockdown cells containing no Pol II bound, Pol II bound, or enriched for promoter proximal Pol II as ascertained by Pol II ChIP-chip analysis (53).
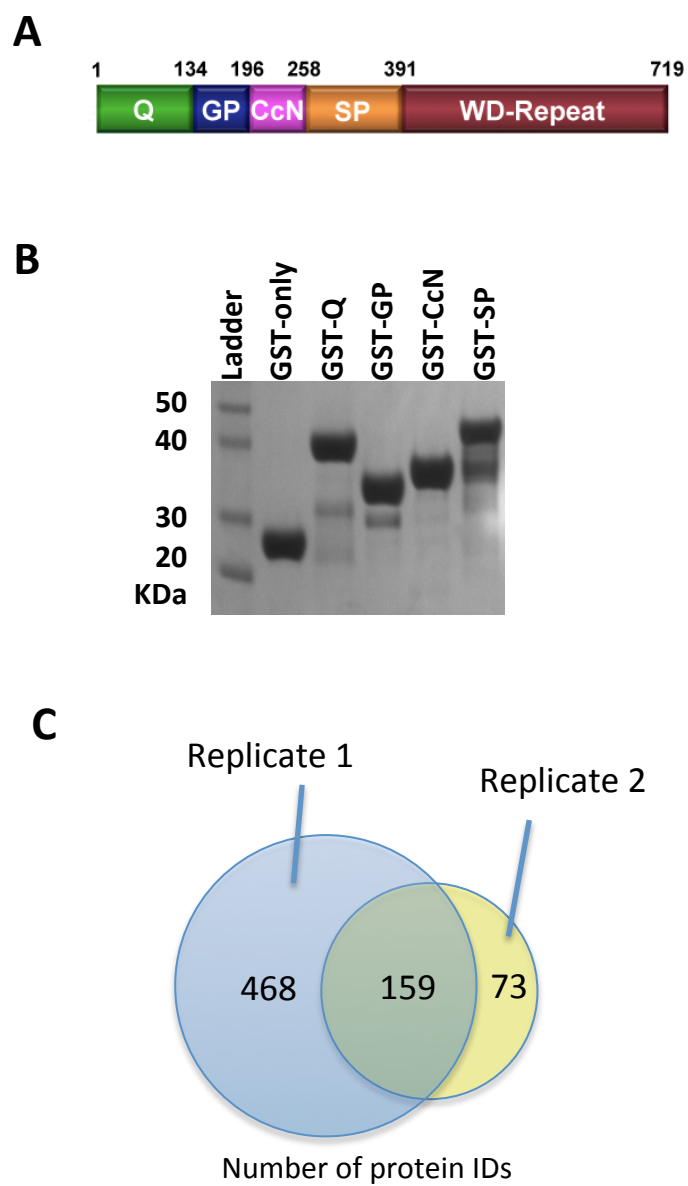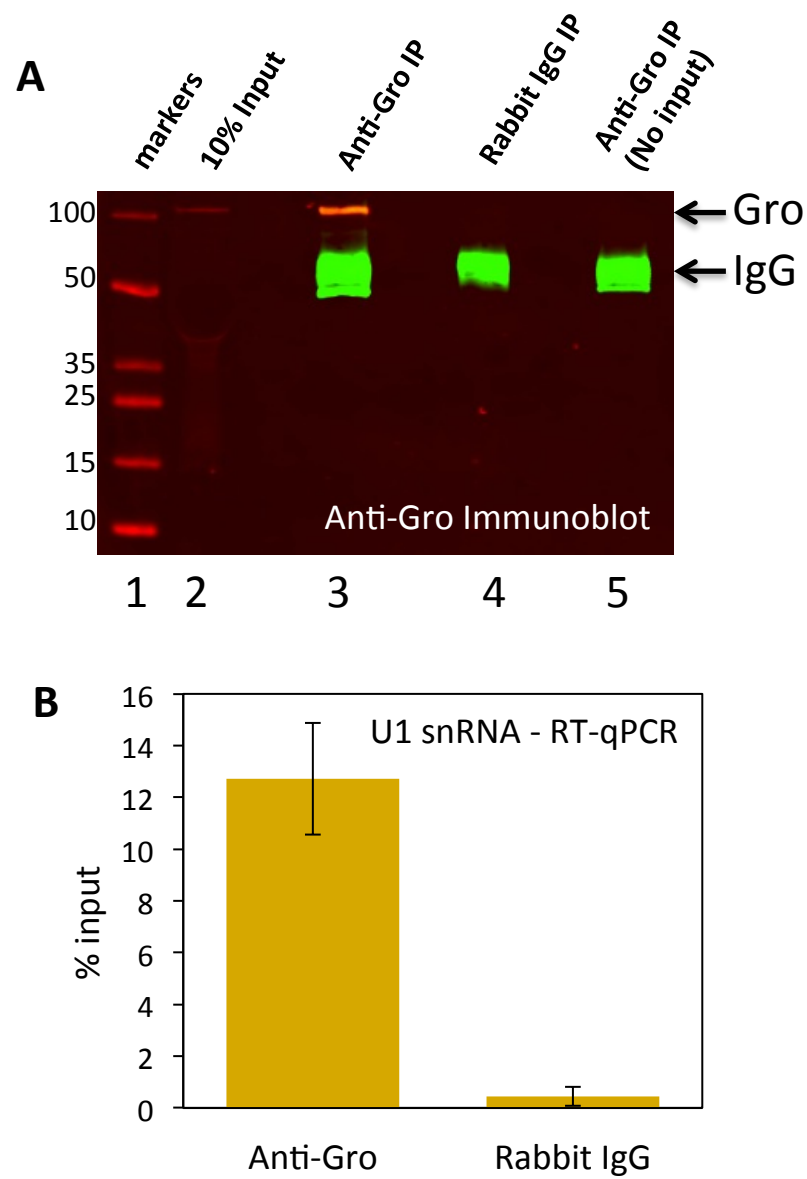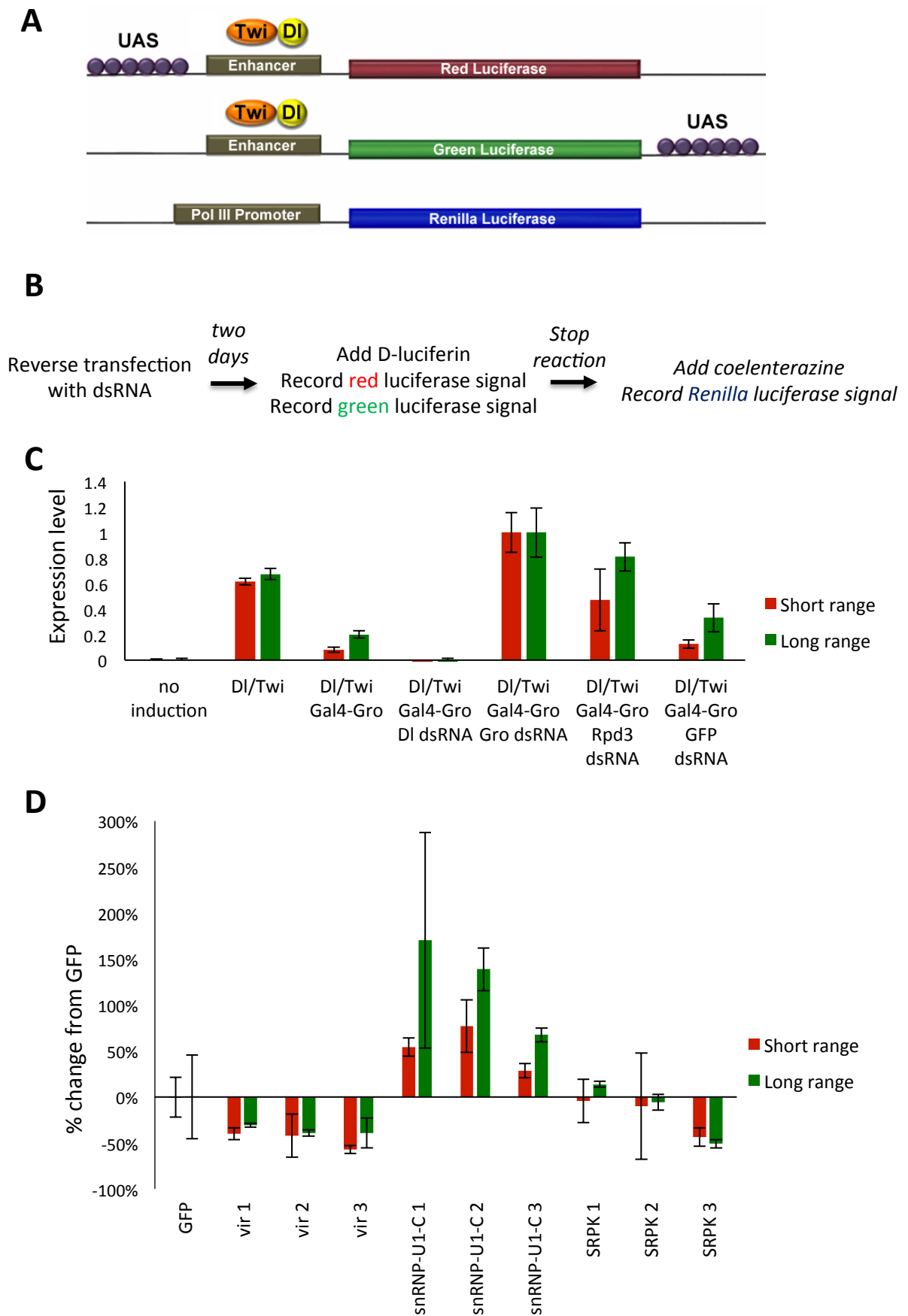
A

B

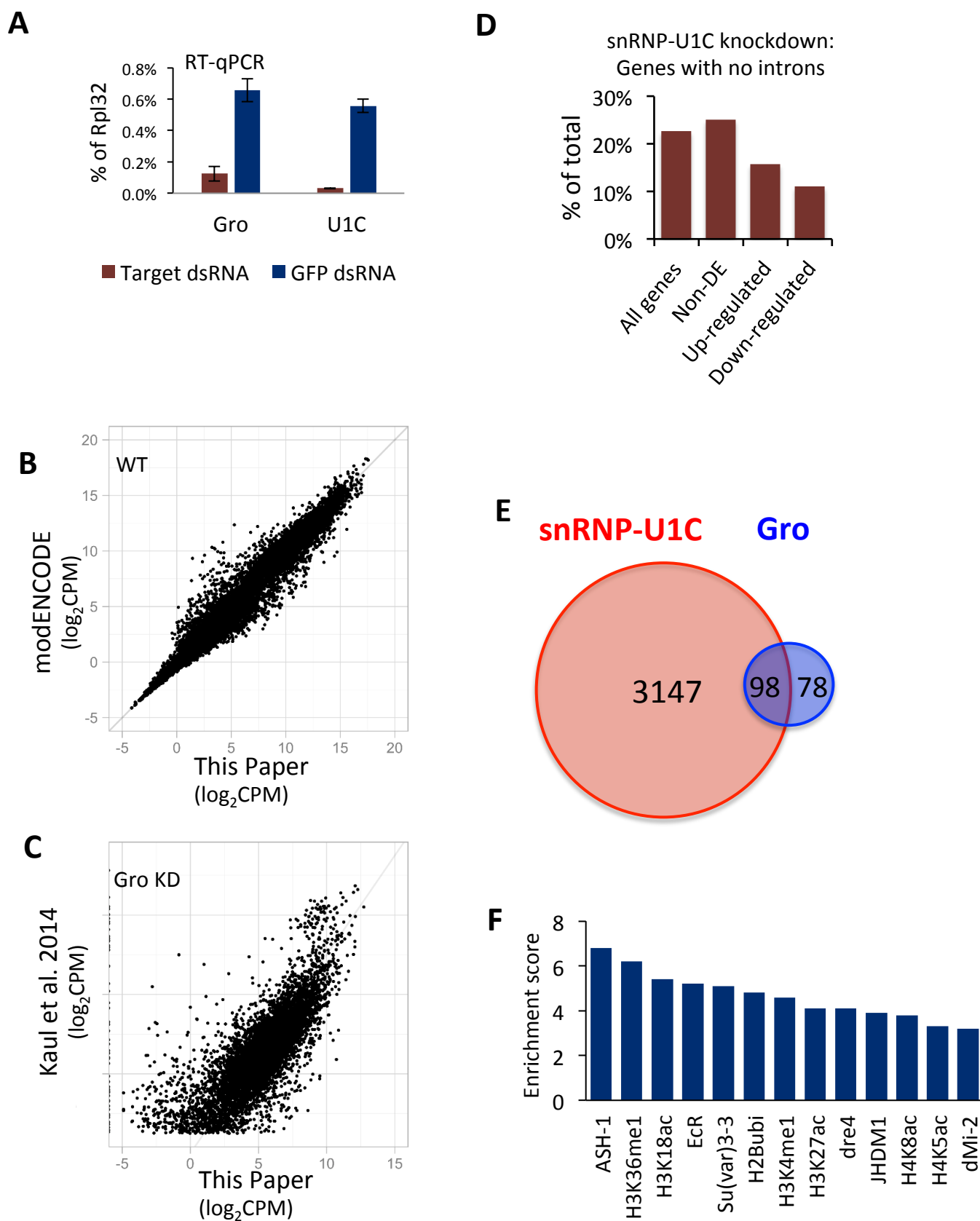
C

Number of protein IDs

Fig 1

A

markers | 10% Input | Anti-Gro IP | Rabbit IgG IP | Anti-Gro IP (No input)

100 ← Gro

50 ← IgG

35
25

15

10

Anti-Gro Immunoblot

1   2   3   4   5

B

U1 snRNA - RT-qPCR

% input

16
14
12
10
8
6
4
2
0

Anti-Gro   Rabbit IgG

Fig 2

Fig 3

A RT-qPCR

% of Rpl32

■ Target dsRNA ■ GFP dsRNA

B WT
modENCODE (log$_2$CPM)
This Paper (log$_2$CPM)

C Gro KD
Kaul et al. 2014 (log$_2$CPM)
This Paper (log$_2$CPM)

D snRNP-U1C knockdown: Genes with no introns
% of total
All genes Non-DE Up-regulated Down-regulated

E snRNP-U1C   Gro
3147   98   78

F
Enrichment score
ASH-1, H3K36me1, H3K18ac, EcR, Su(var)3-3, H2Bubi, H3K4me1, H3K27ac, dre4, JHDM1, H4K8ac, H4K5ac, dMi-2

Fig 4

A

Gro ChIP (1242 binding sites)

Differentially expressed in Gro KD
(46 genes)

Non-differentially expressed in Gro KD
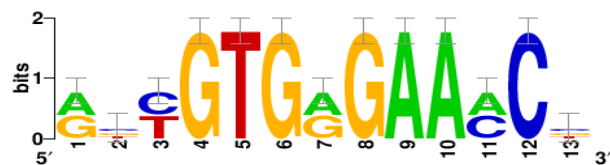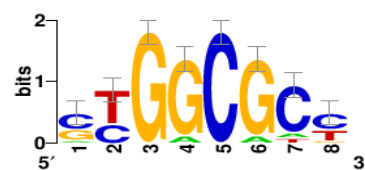(702 genes)

Up regulated (39 genes)

Down regulated (7 genes)

B

Su(H)

brk

Fig 5

Fig 6

Gene Regulation:

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

GENE REGULATION

GENOMICS AND PROTEOMICS

Pak N. Kwong, Michael Chambers, Ajay A. Vashisht, Wiam Turki-Judeh, Tak Yu Yau, James A. Wohlschlegel and Albert J. Courey

Access the most updated version of this article at doi: 10.1074/jbc.M115.681171

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2015/10/19/M115.681171.DC1.html

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2015/10/19/jbc.M115.681171.full.html#ref-list-1