AN INVESTIGATION OF THE MECHANISMS OF GROUCHO MEDIATED TRANSCRIPTIONAL REPRESSION

María Lorena Martínez Quiles

OXFORD BROOKES UNIVERSITY

Department of Biological and Medical Sciences

This thesis is submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

September 2021

Abstract

Groucho (Gro) is the *Drosophila* member of an evolutionarily conserved family of proteins that act as co-repressors for a wide range of transcription factors and includes four orthologs in humans, known as Transducin-like enhancer of split proteins (TLE1-4). This family of co-repressors play crucial roles in many developmental pathways (including Notch, Wingless and BMP/Dpp), and have been implicated in the pathogenesis of some human cancers.

Despite the important roles that Gro proteins play during development and in adult life, there is not an established model that describes the mechanism underpinning Gro-mediated transcriptional repression. Recent studies that used Gro ChIP-seq data from *Drosophila* cells and embryos indicated that Gro could mediate transcriptional repression by promoting Pol II promoter-proximal pausing.

In this work, I took three different approaches to test this model, and gain understanding of Gro-mediated repression. For the first approach, genetic assays exploiting the GAL4/UAS system were used to drive the tissue-specific RNAi downregulation of *gro* and genes involved in Pol II pausing in the developing wing and eye. These assays revealed that *gro* interacts genetically with the pausing factors *Trl*, *Nelf*, *bin3*, and *Larp7,* indicating that they work together during *Drosophila* development, and provide the first *in vivo* evidence that Gro promotes Pol II pausing. Additionally, genetic assays in which the levels of Gro were decreased or increased in a Notch-dependent *Drosophila* tumour model showed the essential role that Gro has during normal growth and tumorigenesis during eye development.

Secondly, I performed a bioinformatic analysis of several published ChIP-seq datasets in *Drosophila* Kc167 to study the genome-wide overlap between Gro peaks and the binding sites of factors involved in Pol II pausing (including GAF, Nelf-E, Cdk9, and Myc), as well as Pol II peaks. It was also explored the co-localization between Gro and Psq, which together with GAF is known to bind to GAGA motifs, as well as with factors that associate with the two different Psq isoforms, containing and lacking the BTB domain. This analysis revealed a large overlap of Gro-bound regions with the binding sites of GAF and Nelf-E, which are two factors key in the establishment of the poised Pol II. A large overlap was also observed between Gro peaks and the binding sites of Myc, and Cdk9, which are involved in the release of the paused Pol II. This indicates that Gro may block the positive action of these factors on transcription and

provides new hints on the possible mechanism by which Gro promotes Pol II pausing. Gro also showed different degrees of co-localization with the two Psq ChIP-seq datasets analysed, as well as with their associated factors. A very high ChIP-seq signal was observed at Gro-bound regions for Psq total (antibody against all Psq isoforms), ISWI, and CBP. A weaker signal was observed for BTB-containing Psq, and Pc, a very weak signal for the architectural proteins Su(Hw), CP190 and Mod(mdg4). Thus, these results suggest a preferent co-localization of Gro with non-BTB Psq and their associated proteins (ISWI, CBP, and Pc).

For the third approach, I performed two unbiased high-throughput screens to find novel proteins that interact with Gro. Specifically, I used the full-length *Drosophila* Gro protein as bait in the yeast 2-hybrid screen, and in a TurboID-based proximity labelling assay in *Drosophila* Kc167 cells. However, these approaches resulted in unfruitful finding novel interacting partners.

Taking together all the evidence from this work I proposed a model of Gro-mediated repression, in which Gro "competes" with activators to regulate transcription.

Acknowledgements

I would like to give a big thanks to my supervisor Dr Barbara Jennings for her great guidance and support during this journey. Thanks for giving me the opportunity to do this research project, in which I have grown professionally and personally.

Many thanks to all my lab and office mates on the first floor of the Sinclair Building for all the invaluable help and support over these years, and most importantly for the great friendly atmosphere in the lab.

Thanks also to the fantastic people in the lab of Dr Maria Domínguez in Alicante. Thanks for including me in the team during the research visit, in which I have learnt so much.

I am also very grateful to all my friends in Oxford, and in Spain, who have made this time better than I could have hoped for. Thanks for being there during these years, and for sharing trips, beers, meals, and “covid” walks.

Por último, y más importante, me gustaría agradecerle a mi familia por su apoyo incondicional y comprensión, sin el cual no podría haber realizado esta tesis.

Table of contents

Table of figures

Table of tables

Abbreviations and key terms

7sk snRNP	7SK small nuclear ribonucleoprotein complex
aa	Amino acids
ACV	Anterior cross vein
AD	Activation domain
BDB	DNA-binding domain
Bin3	Bicoid-interacting protein 3
BioID	Proximity labelling assay of protein interactors
BirA	*E. coli* biotin ligase
BSA	Bovine serum albumin
BTB	Broad-Complex, Tramtrack and Bric a brac domain
CBP	CREB-binding protein
Cdk9	Cyclin-dependent kinase 9
cDNA	Complementary Deoxyribonucleic acid
Chip-chip	Chromatin Immunoprecipitation followed by DNA microarray
Chip-seq	Chromatin Immunoprecipitation followed by Sequencing
CNS	Central Nervous System
co-IP	Co-Immunoprecipitation
CTD	C-terminal domain of Polymerase II
CycT	Cyclin-T
DAPI	6-diamidino-2-phenylindole
DHSs	DNase I hypersensitive sites
Dl	Delta
Dpp	Decapentaplegic pathway
DSIF	DRB Sensitivity Inducing Factor
E(spl) bHLH	Enhancer of Split family of basic helix-loop-helix (bHLH) proteins
ECL	Enhanced chemiluminescent
GFP	Green fluorescent protein
eGFP	Enhanced green fluorescent protein
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N’,N’-tetraacetic acid

GAF	GAGA factor
GO	Gene ontology
Gro	Gro
H	Hairless
HDAC	Histone deacetylase
HEXIM	HEXamethylene-bis-acetamide-Inducible protein
LARP7	La related protein 7
LC-MS/MS	Liquid chromatography coupled to tandem-mass spectrometry
LEF1	Lymphoid enhancer-binding factor-1
LFQ	Label free quantification
MAPK	Mitogen-activated protein kinase
miniTurbo	Promiscuous biotin ligase developed by (Branon et al., 2018)
mRNA	messenger RNA
NCoR	Nuclear receptor corepressor
SMRT	Silencing mediator of retinoic acid and thyroid hormone receptor
NELF	Negative elongation factor
N^{ICD}	Notch intracellular domain
NuRD	Nucleosome remodeling and deacetylase complex
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCV	Posterior cross vein
Pol II	RNA polymerase II
Psq	Pipsqueak
P-TEFb	Positive transcription elongation factor
RNAi	Interference RNA
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Su(H)	Suppressor of Hairless
Su(Hw)	Suppressor of Hairy-wing
TCF	T cell factor/lymphoid enhancer factor
TF	Transcription factor
TLE	Transducin-Like Enhancer of Split

Tris	Trisaminomethane
Trl	Trithorax-like
TSC	Total spectral counts
TSS	Transcription start site
TurboID	Promiscuous biotin ligase developed by (Branon et al., 2018)
Wnt	Wingless
X-a-gal	Chromogenic substrate for yeast galactosidase
Y2H	Yeast 2-hybrid

1 Introduction

1.1 Regulation of gene expression during *Drosophila* development

All the cells within a mature animal arise from a single cell, the fertilized egg, which during development divides, and ultimately creates every type of cell in the body. The great variety of cell types present in multicellular organisms are the consequence of the expression of specific sets of genes at various times. This complexity is achieved by the sophisticated regulatory mechanisms that tightly regulate gene expression at many stages during development and in adult life.

Gene expression can be regulated at many different levels during the process of transcription and translation, which include transcript stability and processing, localization of transcripts to specific sites for translation, translation efficiency, and post-translation modifications. However, all genes are regulated at the level of transcription via complex mechanisms that allow the cell to control gene expression in a precise temporally and spatially manner. Hence, transcriptional regulation plays essential roles during development, growth, and response to environmental changes.

To understand these complicated processes, *Drosophila melanogaster*, commonly known as the fruit fly, has been extensively used as a model organism over more than a century [reviewed in (Hales et al., 2015; Jennings, 2011)]. Although simpler than mammals, *Drosophila* is a complex organism, and it is well-established that most of the developmental processes and pathways are highly conserved between them. The use of *Drosophila* as a model organism has many technical advantages as they are easy to culture in the laboratory, have a short life cycle, and each female can produce dozens of offspring. In addition, they have a compact genome of four chromosomes, which can be genetically modified by the many available genetic tools.

All these features make the fruit fly a powerful model organism that has enabled important advances in developmental biology, neurobiology, transcriptional regulation, and population genetics, among other fields. The study of the fruit fly has raised key concepts in transcriptional regulation, such as the cis-regulatory elements that act over long distances, the hierarchical cascades of transcription factors (TFs), and the position effect variegation.

The mechanisms underlying transcriptional regulation in flies are generally well conserved with mammals, although there is an exception with the epigenetic modification of DNA methylation. In mammals, DNA methylation is commonly associated with gene silencing, and it is best characterized at CpG dinucleotides. Despite the important role that this epigenetic mark has in gene regulation in mammals, it has not been conserved in several animals, including the invertebrate model organisms *Drosophila melanogaster* and *Caenorhabditis elegans* [reviewed in (Greenberg and Bourc'his, 2019)].

1.1.1 *Drosophila* life cycle

As mentioned in the previous section, *Drosophila* has a short life cycle, which allows the rapid generation of large numbers of offspring that can be used in genetic crosses.

In the fruit fly, development from a fertilized egg to an adult requires an average of 9-10 days at 25^{o}C (Figure 1 A), however, the speed of this process is highly sensitive to temperature and it extends to 19 days when flies are cultured at 18^{o}C. The first developmental stage after fertilization is embryogenesis, which is completed in ~24 hr and is followed by three larval stages, which take five days in total. Once the larval development is completed, the animal undergoes metamorphosis within a pupal case for 4-5 days to achieve the final adult form. During the pupal stage, many larval tissues degenerate and the imaginal discs present in the larvae are transformed into most of the adult structures, including eyes, legs, and wings (Figure 1 B) [reviewed in (Jennings, 2011)].

Imaginal discs are defined as epithelial tissues, which arise from a cluster of cells in the embryo and develop during larval stages in insects that undergo metamorphosis to generate the majority of adult tissues. They have been broadly used as experimental systems to investigate multiple aspects of development, and their study has led to many important discoveries in developmental and cell biology. Thus, the use of imaginal disc by developmental biologists is another reason that makes the fly fruit an excellent model organism [reviewed in (Beira and Paro, 2016)].

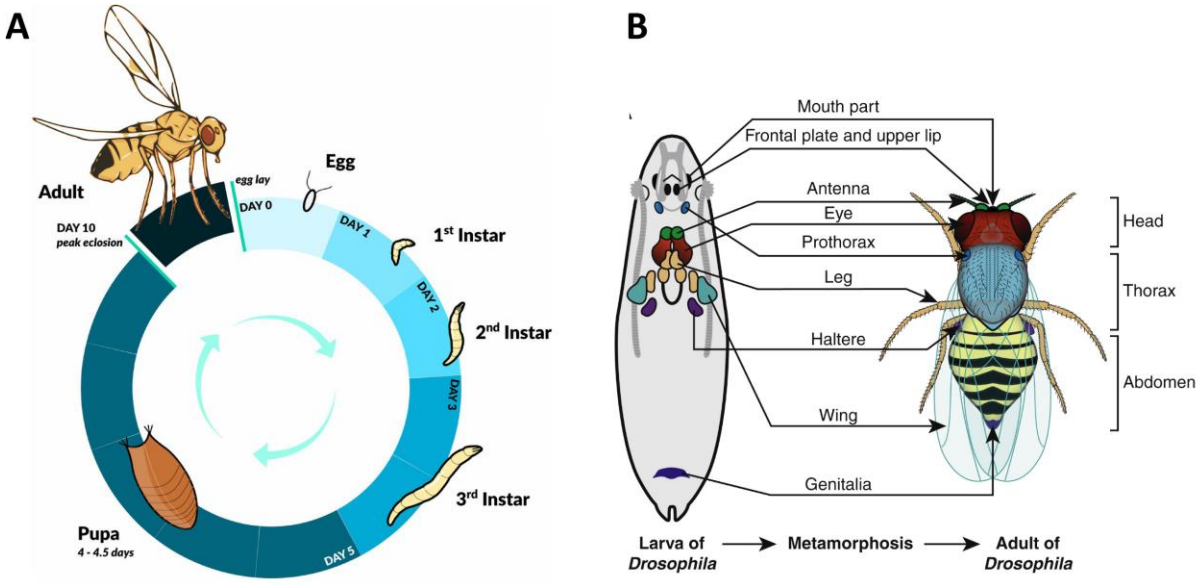


Figure 1: Schematic representation of the *Drosophila* life cycle, and the larval imaginal discs.
(A) Scheme of the *Drosophila* melanogaster life cycle at 25°C, which takes approximately 10 days. It shows the developmental cycle, from the fertilized egg (day 0) to the adult fly (day 10), and proceeds through the three larval instar stages (1st Instar, 2nd Instar, and 3rd Instar), and the pupal stage. Image adapted from (Flatt, 2020). **(B)** Scheme showing the localization of the imaginal discs inside the third instar larva (on the left) and the adult structures they transform into (on the right). Image adapted from (Aldaz and Escudero, 2010).

1.1.2 Pioneer and non-pioneer transcription factors

Transcriptional activity during animal development is primary regulated by the cooperative action among DNA-binding TFs. TFs bind to specific sequence motifs at proximal and distal regulatory elements. According to their function in development, TFs can be classified into pioneer and non-pioneer (or canonical).

Pioneer factors are master TFs that recognize and bind DNA wrapped up in chromatin, triggering the opening of the local chromatin, and enabling the binding of canonical TFs (Zaret and Carroll, 2011). Pioneer TFs play a unique role during development, promoting a transcriptionally competent chromatin landscape, and implementing cellular programs. In the early *Drosophila* embryo, pioneer TFs are essential during the maternal-to-zygotic transition (MTZ); the process in which the control of development progressively transitions from the maternal genome to the zygotic genome. This transition is essential in embryogenesis and involves cellular reprograming and the initiation of transcription at thousands of genes. In this

context, the pioneer factor Zelda is key in promoting chromatin accessibility and transcription at the initiation of zygotic genome activation (ZGA) (McDaniel et al., 2019).

Although Zelda was the first global genomic activator identified in *Drosophila* embryogenesis (Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011), recent studies have implicated other factors with a pioneering activity that are involved in this process. In this context, GAGA factor (GAF) (Figure 2), encoded by the *Tritorax-like gene* (*Trl*), is expressed and required to open chromatin during MTZ (Gaskill et al., 2021). GAF binds to GA-dinucleotides repeats and has diverse additional roles in transcriptional regulation, such as transcriptional activation, silencing, and higher-order chromatin structure [reviewed in (Chetverina et al., 2021)]. In addition to GAF, two other proteins are expressed in the early embryo and are known to bind to GA repeats: chromatin-linked adaptor for MSL proteins (CLAMP), and Pipsqueak (Psq). CLAMP has some overlapping functions with GAF; evidence shows that both are pioneer factors that function independently and are essential during embryonic development (Gaskill et al., 2021). Meanwhile, evidence points to Psq acting as a transcriptional repressor and also having a role in chromatin looping (Gutierrez-Perez et al., 2019; Huang et al., 2002).

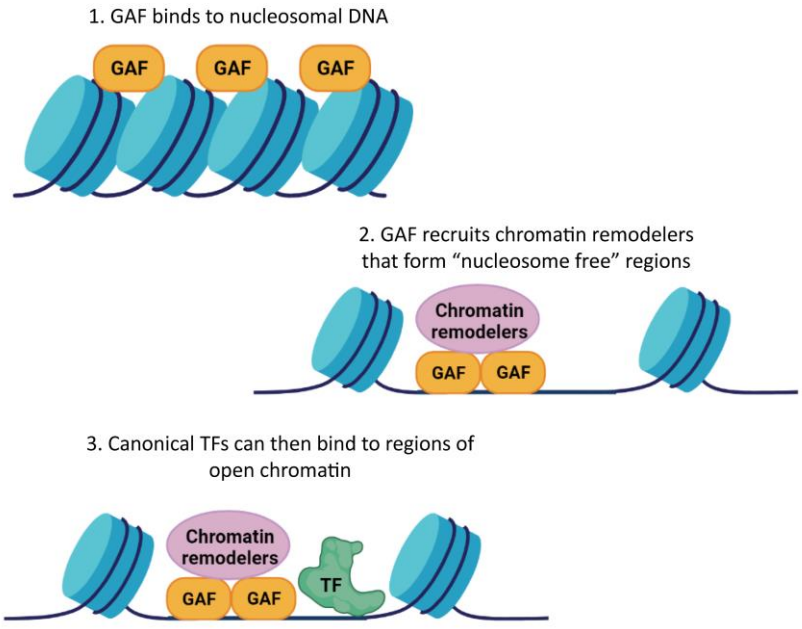


Figure 2: Schematic representation of chromatin opening by the pioneer factor GAF.
Created with BioRender.com.

Another pioneer factor essential during *Drosophila* development is Grainy head (Grh), which was first identified as a pioneer factor in the larval eye disc (Jacobs et al., 2018). In contrast to Zelda and GAF, Grh is not required in the early embryo but later in the embryonic and larval development (Nevil et al., 2020; Uv et al., 1997).

After pioneer factors expose the regulatory chromatin domains, non-pioneer TFs can access and bind to those sequences, contributing to the maintenance of the open chromatin landscape. In addition, non-pioneer factors activate transcription via the recruitment of the transcriptional apparatus, and they can be subclassified into general and context-specific TFs.

In eukaryotes, general transcription factors (GTFs) are present in all cells and are necessary for transcription to occur. GTFs together with the RNA Polymerase II (Pol II) constitute the preinitiation complex (PIC) at promoters, which enables transcription initiation. The minimal PIC includes at least six key GTFs: TFIID, TFIIB, TFIIH, TFIIE, TFIIF, and TFIIA (Sainsbury et al., 2015). Importantly, TFIID is a multiunit complex composed of a TATA-box binding protein (TBP), and up to 15 TBP-associated factors (TAFs). This complex functions via the binding of TBP to the "TATA box" sequence upstream of the transcription start site (TSS) and plays an important role in the formation of PIC.

In addition to GTFs, other factors, known as context-specific TFs, are required only in certain conditions to regulate the transcription of a specific set of genes. Those TFs bind to cis-regulatory elements at certain places and spaces and frequently initiate patterns of gene expression. The large number of specific TFs in eukaryotes can be classified into five groups based on certain structural similarities (motifs): zing finger, leucine zipper, helix-turn-helix, helix-loop-helix, and homeodomain TFs [(Wingender et al., 2015)].

1.1.3 *Drosophila* embryogenesis

During embryonic development, *Drosophila* progress through several complex stages and processes that begin with the fertilized egg. Following fertilization, the zygotic nuclei undergo fourteen divisions that follow a precise pattern of migration and movement. The first thirteen cycles are syncytial divisions regulated by maternal genes that load their gene products into the egg during oogenesis (maternal-effect genes). There are at least 38 maternal-effect genes identified in flies that are important for the dorsal-ventral and anterior-posterior polarity in

the embryo and include *bicoid*, *toll*, and *nanos* (Lehmann and Nüsslein-Volhard, 1991; Morisalo and Anderson, 1995; St Johnston et al., 1989).

During the fourteenth nuclei division, cell membranes form between nuclei, generating the cellularized blastoderm. At this developmental stage occurs the major transcriptional wave of zygotic genome activation, and the embryo starts to develop a pattern of body segments that are controlled by the zygotic-segmentation genes. This group of genes encode TFs that can be subdivided into gap, pair-rule, and segment-polarity. Gap proteins control the expression of pair-rule genes, which divide the embryo into segments and regulate the expression of segment polarity genes. The latter group of genes finally establish the anterior-posterior axis of each segment (Figure 3 A) [reviewed in (Hoy, 2003; Tadros and Lipshitz, 2009)].

In this context, transcriptional repression is known to play an important role in the regulation of these regulatory networks. Thus, many gap and pair-rule genes encode repressive TFs that mutually repress each other to establish sharp boundaries between adjacent segments. Among the gap genes that encode repressors are *tailless*, *giant*, *kruppel*, *knirps*, and *huckebein* (Figure 3 B). A combination of these repressors is known to regulate the repressive pair-rule genes *hairy* and *even-skipped* (*eve*), which in turn, regulate other pair-rule genes including *fushi tarazu* (*ftz*), *odd-skipped* (*odd*), and *runt* (*run*) (Clark, 2017). Mutations in any of these genes cause serious effects on segment patterning that lead to embryo death (Bucher and Klingler, 2004; Janssens et al., 2013; Nüsslein-Volhard and Wieschaus, 1980).

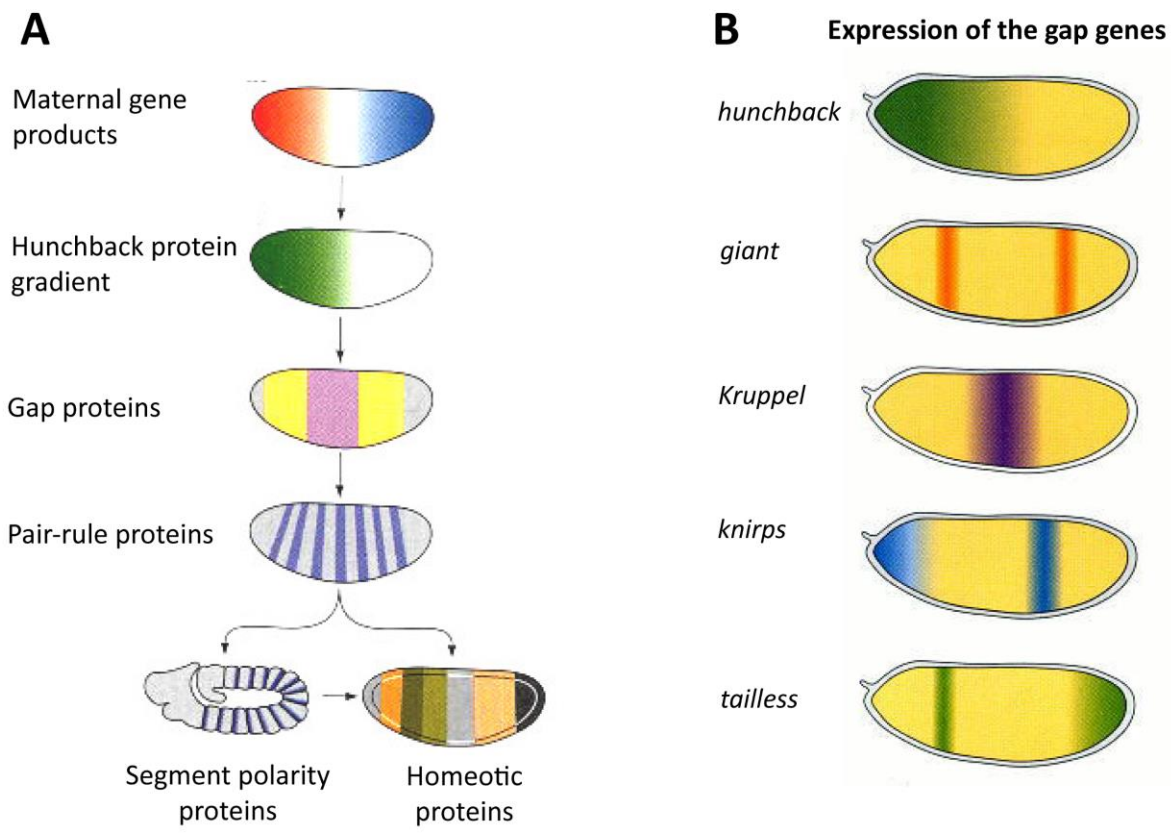


Figure 3: Schematic representation of the hierarchy of genes that establish the *Drosophila* anterior-posterior pattern, and expression of zygotic genes.
(A) Model of anterior-posterior pattern formation. In brief, the maternal effect genes create gradients of proteins, that form a gradient of the Hunchback protein, which differentially activates the gap genes. The gap genes divide the embryo in broad territories, activating the expression of the pair-rule genes. Then, the segment polarity genes divide the embryo into segments and define the spatial domains of the homeotic genes. **(B)** Expression pattern of gap genes. Image adapted from (Developmental Biology.6th edition).

Importantly, sequence-specific DNA-binding transcriptional repressors, such as the previously mentioned gap and pair-rule repressive factors do not have any intrinsic activity that inhibits transcription but function via recruitment of co-factors. In this framework, one co-factor that is recruited by many TFs during *Drosophila* embryogenesis is Groucho.

1.2 The Groucho family of co-repressor proteins

Drosophila Groucho (Gro) is the founding member of the evolutionarily conserved family of corepressor proteins [reviewed in (Buscarlet and Stifani, 2007; Jennings and Ish-Horowicz, 2008)]. The *gro* gene was named after the phenotype of the first mutation identified in

Drosophila, *groucho*^{1} (*gro*^{1}). When homozygous, the hypomorphic allele *gro*^{1} displays clumps of extra bristles over the eyes, resembling the distinctive bushy eyebrows of the film star Groucho Marx (Lindsley and Grell 1968). The *gro g*ene was mapped via mutations to a transcript initially considered to lie within the *Enhancer of Split complex* [*E(spl)-C*], *m9/10*. This transcript was sequenced and the protein it encoded was found to have homology (through what is now known as the WD-domain, see below) to the G-protein beta subunit Transducin (Hartley et al., 1988). Subsequently, cDNAs encoding four *gro* orthologues were identified in humans and designated as *Transducin-like Enhancer of split* (*TLE*) genes (*TLE1-4*) (Stifani et al., 1992). Gro family proteins are found across eukaryotic evolution and include Tup1 in *Saccharomyces cerevisiae*, TOPLESS and LEUNIG in *Arabidopsis,* and GRG1-4 in mice [reviewed in (Jennings and Ish-Horowicz, 2008)].

This family of proteins do not bind DNA directly as they lack any DNA-binding domain but act as a repressive co-factor for a wide variety of DNA-binding proteins. Among the first DNA-binding proteins that demonstrated to be Gro-dependent repressors were members of the Hairy and Enhancer of Split (HES) family of basic helix-loop-helix (bHLH) proteins (Paroush, 1994). The Enhancer of split [E(spl)] bHLH proteins are encoded within the *Enhancer of split complex* [*E(spl)-C*], a cluster containing 13 genes that are well-characterized effectors of the Notch pathway (Delidakis et al., 2014). Seven of these genes, named *E(spl)mδ, mγ, mβ, m3, m5, m7*, and *m8*, encode basic-Helix-Loop-Helix (bHLH) proteins that physically interact with Groucho (Fisher et al., 1996; Paroush, 1994). Subsequent studies showed that Gro interacts with other repressors, including Engrailed, Dorsal, Runt, Huckebein, Capicua (Cic) and Brinker (Brk) (Aronson et al., 1997, 1997; Dubnicoff et al., 1997; Goldstein et al., 1999; Hasson, 2001; Jimenez et al., 1997).

1.2.1 Biological roles of Gro family proteins

There is a large body of evidence supporting the crucial role of Gro proteins during animal development. This family of proteins is known to interact with a variety of developmental TFs and pathways, including Wnt/Wingless (Wg), Hedgehog, Decapentaplegic/Bone Morphogenetic Protein (Dpp/BMP), Notch, and Epidermal Growth Factor (EGF).

The developmental roles of this family of corepressors have been extensively studied in the fruit fly, in which many Gro interactors have been identified to date (Table 1).

In flies, Gro associates with several key downstream effectors of developmental pathways to repress target genes. In this frame, the E(spl) bHLH proteins, Brk, and T-cell factor (TCF) proteins are Gro-dependent repressors with important roles in the Notch, Dpp and Wnt pathways, respectively. Gro also interacts with EGF signalling, however, it does it differently from the previously mentioned pathways. Gro is a target of the EGF pathway, but its function is downregulated upon pathway activation [reviewed in (Buscarlet and Stifani 2007, Hasson and Paroush 2006)].

Proteins are known to interact with Groucho in *Drosophila melanogaster*			
Alhambra (Alh)	Hairless (H)	Optix	Intermediate neuroblast defective (Ind)
BarH1 (B-H1)	Ecdysone-induced protein 78C (Eip78C)	Protein on ecdysone puffs (Pep)	Knirps (Kni)
Chip (Chi)	Histone deacetylase 1 (HDAC1)	Bother of odd with entrails limited (Bowl)	Midline (Mid)
C-terminal binding protein (CtBP)	HGTX	Brinker (Brk)	Odd skipped (Odd)
Dichaete (D)	HES-related (Hers)	Capicua (Cic)	Pangolin (Pan)
Dorsal interacting protein 2 (Dlip2)	Holes in muscle (him)	Dorsal	Relative of woc (Row)
Dorsocross1 (Doc1)	Homeodomain interacting protein kinase (Hipk)	Deadpan (Dpn)	Runt (Run)
Enhancer of split m3, helix-loop-helix (E(spl)m3-HLH)	Hormone receptor-like in 96 (Hr96)	Elbow B (ElB)	Spalt major (Salm)
Enhancer of split m4, Bearded family member (E(spl)m4-BFM)	Similar to Deadpan (Sidpn)	Empty spiracles (Ems)	Scribbler (Sbb)
Enhancer of split m5, helix-loop-helix (E(spl)m5-HLH)	Suppressor of variegation 2-10 (Su(var)2-10)	Engrailed (En)	Scute (Sc)
Enhancer of split m7, helix-loop-helix (E(spl)m7-HLH)	Ultrabithorax (Ubx)	Even skipped (Eve)	Schunurri (Shn)
Enhancer of split m8, helix-loop-helix (E(spl)m8-HLH)	Anti-silencing factor 1 (Asf1)	Eyeless (Eye)	Sloppy paired 1 (Slp 1)
Enhancer of split mβ, helix-loop-helix (E(spl)mβ-HLH)	Bagpipe (Bag)	Eyes absent (Eya)	Sine oculis (So)
Enhancer of split mβ, helix-loop-helix (E(spl)mβ-HLH)	Inverted repeat binding protein 18 kDa (Irb18)	Ftz transcription factor 1 (Ftz-f1)	Tiptop (Tio)
Enhancer of split mδ, helix-loop-helix (E(spl)mδ-HLH)	Kruppel homolog 1 (Kr-h1)	Hairy	Ventral nervous system defective (Vnd)
Transcriptional adaptor 2b (Ada2b)	Ets at 65 A (Ets65A)	Myc	Huckebein (Hkb)

Table 1: Table summarizing the known proteins that interact with Gro from the literature.
List of *Drosophila* Gro interactors from Flybase (Larkin et al., 2021) in alphabetical order.

In the Notch pathway, *Drosophila* Gro has both positive and negative effects. In summary, Notch signalling is activated when the Notch (N) receptor binds to its ligands [Delta (Dl) or Serrate (Ser) in *Drosophila*; Delta or Jagged orthologues in mammals], which are present at the surface of contacting cells. This leads to the proteolytic cleavage of the N receptor, and the release of its intracellular domain (N^{ICD}). Then, the N^{ICD} translocates to the nucleus, where it binds to the DNA-binding TF Suppressor of Hairless [Su(H); CBF1 in mammals]. Subsequently, several co-activators [including mastermind (Mam)] associate with the complex, allowing the activation of Notch target genes, including the *E(spl) bHLH* genes (Figure 4 A). In the absence of Notch signal, Gro interacts with the adaptor protein Hairless (H), which is recruited by Su(H). The complex formed by Gro, H, and Su(H) - which also includes the co-repressor CtBP (C-terminal binding protein) - leads to inactivation of the Notch targets, and hence antagonizes Notch signalling (Figure 4 A). Conversely, Gro plays a key role during Notch activation by associating with the E(spl) bHLH repressors, which are the key downstream effectors of the Notch pathway [reviewed in (Bray, 2016; Lai, 2004)].

Notch signalling is central to the determination of many cell fates during development as it is frequently used to establish binary cell fates between adjacent cells. This pathway is best known for the process named 'lateral inhibition' that is classically illustrated in the context of neurogenesis. Neurogenesis is initiated in the *Drosophila* ectoderm by the expression of proneural genes in a small group of cells called proneural clusters (PNCs). These genes include *achaete*, *scute*, *lethal of scute*, and *atonal,* which encode basic bHLH transcriptional activators that confer PNCs with the potential to adopt a neural cell fate. The cell in the PNC that expresses the highest level of proneural proteins adopts the neural fate, and laterally inhibits adjacent cells via the activation of Notch, and hence the expression of the *E(spl) bHLH* genes (Figure 4 B). In this setting, Gro plays an essential role acting as a co-repressor for the E(spl) bHLH repressors, which leads to the suppression of proneural gene expression in these cells [reviewed in (Lai, 2004)].

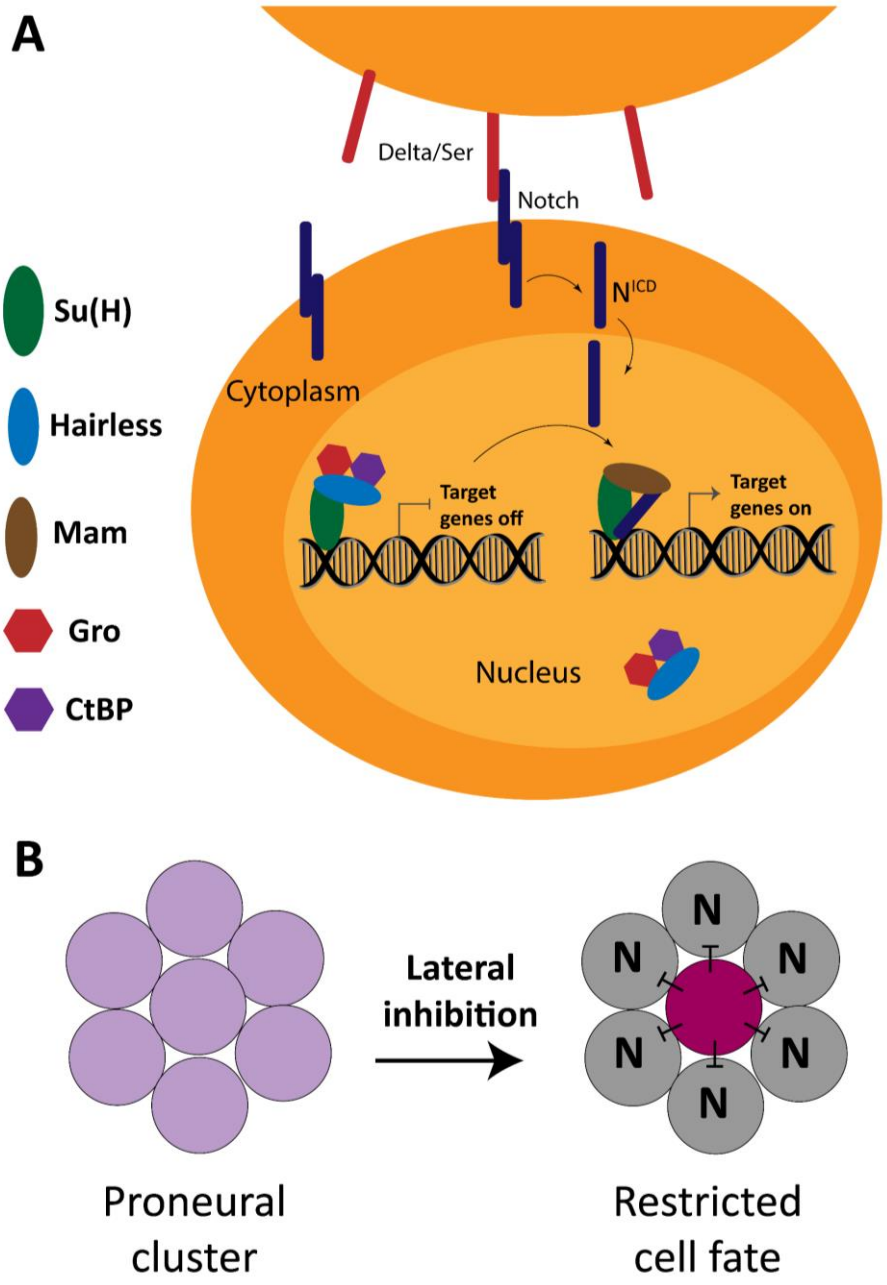


Figure 4: Schematic representations of Notch signalling in the presence and absence of pathway activation, and the process of lateral inhibition.

(A) Scheme illustrating the main components of *Drosophila* Notch signalling when the pathway is OFF (left) or ON (right). **(B)** Scheme illustrating lateral inhibition in *Drosophila* neurogenesis. The cell within the proneural cluster (cells in purple) that has higher levels of Delta adopts a neural fate (magenta cell) and inhibits the neighbouring cells from acquiring a neuronal fate using Notch signalling (grey cells).

Gro plays a similar negative role in Wnt signalling, keeping the Wnt target genes repressed in the absence of pathway activation. When the Wnt signal is off, the key downstream effector of the pathway, Armadillo (β-catenin in mammals) is phosphorylated and degraded in the

cytoplasm. In this setting, the transcription factor TCF (also named Pangolin in *Drosophila*) associates with Gro in the nucleus, converting TCF to a repressor, and keeping Wnt target genes repressed. Upon Wnt activation this process is inhibited, allowing unphosphorylated β-catenin to accumulate, enter the nucleus, and bind to the TCF, converting it to a transcriptional activator that least to the transcription of Wnt target genes.

In *Drosophila*, the role of Gro is also crucial in Wnt signalling, which has been demonstrated in the analysis of gro^{MB5} (encoding a Gro protein with a deletion in the Q domain), and gro^{MB36} (null allele) (Jennings et al., 2008) mutant somatic clones in wing imaginal discs. The analysis of these clones showed a derepression of *senseless (sens)* and *Distal-less (Dll),* two well-characterized Wnt target genes in the wing imaginal disc. Furthermore, this study revealed that Pygopus (Pygo) is not essential for Wingless pathway output when Gro is defective. Pygo is another nuclear component of Wnt signalling Pygo proteins associate with Armadillo/β-catenin by the adapter protein Legless (also named BCL9) (Mieszczanek et al., 2008).

In the context of Dpp signalling, Gro interacts with Brk to antagonize the transduction of this pathway. Brk is a key transcriptional repressor that antagonizes Dpp signalling in many developmental processes and recruits several co-repressors, including Gro and CtBP (Hasson, 2001). Brk acts by blocking the expression of two key TF that mediate cellular response to Dpp activation: Mothers against dpp (Mad) and Medea (Med) (Mad and Med). In brief, the activation of Dpp signalling starts with the assembling of Dpp dimers that form receptor complexes at membranes, including the Thickveins (Tkv) receptor. Tkv is then phosphorylated and activated by the Punt receptor upon ligand binding. Activated Tkv, in turn, phosphorylates Mad, which associates with Med and additional TFs, to form protein complexes that regulate the expression of numerous genes [reviewed in (Hamaratoglu et al., 2014)].

1.2.1.1 Role of Gro in *Drosophila* wing development

A well-studied example of how Gro interacts with several signalling pathways in development is in the formation of the wing veins.

The *Drosophila* adult wing displays a pattern of five longitudinal veins (L1 to L5) and two cross-veins [anterior-cross vein (ACV) and posterior-cross vein (PCV)] that are established during larval and early pupal development (Figure 5 A). This developmental process is coordinated

by the action of the EGF, Notch, Dpp, Hedgehog, and Wnt cascades. The wing imaginal disc is a monolayered epithelium formed by ~ 50.000 cells, in which the central region, known as the pouch, develops into the wing blade. At early stages, the wing imaginal disc is divided by two compartmental boundaries that intersect perpendicularly in the pouch: one anterior-posterior (A-P), and the other dorso-ventral (D-V). Those compartments are formed by different cell populations that do not intermingle and the signals between them result in the positioning and elaboration of early pro-veins [reviewed in (Shimmi et al., 2014)].

The A and P compartments are defined by the expression of the key TFs Engrailed (En) and Invictus in the P cells, and Cubitus interruptus (Ci) in the A cells. The secretion of En in the P compartment induces the expression of the short-range morphogen Hedgehog (Hh), which can cross the A-P boundary to regulate the expression of other key genes in neighbouring cells. In the A-side of the A-P boundary Hh signalling - which is mediated by the TF Ci – creates gradients that induce the expression of several proteins, including Decapentaplegic (Dpp), and Wingless (Wg) (Figure 5 B). Dpp is a long-range morphogen that is induced by moderate doses of Hh and diffuses along both the A and P compartments. In addition, to regulate *dpp*, Hh also induces the expression of Wingless (Wg) at the D-V boundary, which is also key in wing development (Figure 5 B) [reviewed in (Beira and Paro, 2016; Crozatier et al., 2004)].

The EGF and Notch pathways play also essential roles during the development of the wing imaginal disc, a process in which both pathways are known to antagonise each other. Notch functions as an anti-vein, whereas EGF acts overcoming the influence of Notch and enabling vein formation. Several studies detailed the role of Gro in the crosstalk between these pathways in the wing imaginal disc. In brief, activation of EGF by its ligands leads to the phosphorylation of the Mitogen-Activated Protein Kinase (MAPK), which can then enter the nucleus and phosphorylate its target proteins, including Gro (Figure 5 C). The phosphorylation of Gro downregulates its repressive function and leads to relieving Gro-dependent transcriptional repression. In wing development, several known repressors act via Gro recruitment; E(spl) bHLH-mβ, Brk, and Cic (Figure 5 C). This group of repressors act to inhibit vein formation. Thus, the EGF pathway accomplishes the formation, of pro-vein tissue via their inactivation, which is achieved through Gro phosphorylation [reviewed in (Hasson and Paroush 2006)].

This regulatory action of the MAPK pathway on Gro has also been reported in human cells. In those cells, unphosphorylated human TLE1 acts repressing MAPK target genes. Upon activation of MAPK, Gro is phosphorylated and excluded from the nucleus, which leads to derepression of MAPK target genes that are Gro-dependent. Therefore, this study shows that the role of the MAPK pathway on Gro proteins is conserved across evolution (Zahavi et al., 2017).

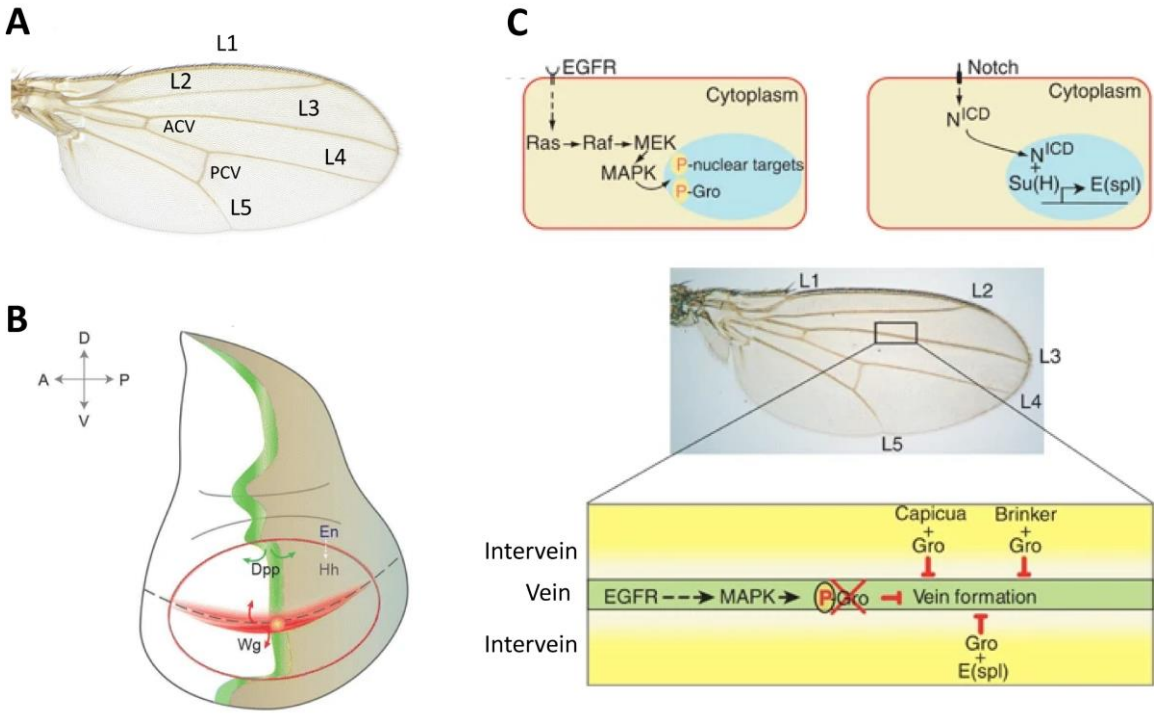


Figure 5: Developmental pathways involved in the formation of *Drosophila* wing, and role of Gro in the interplay between them.
(A) Image of a wild type *Drosophila* adult wing, showing the key structures: longitudinal veins (numbered L1 to L5), anterior cross-vein (ACV), and posterior cross-vein (PCV). **(B)** Scheme showing the expression domains of four key signals in the wing imaginal disc. Image adapted from (Beira and Paro, 2016). **(C)** Schematic representations of the EGF and Notch signalling pathways (top panel), as well as the role of Gro in the interplay between these pathways during wing development. Adapted from (Hasson and Paroush 2006).

1.2.1.2 Roles of Gro proteins in mammalian development

As illustrated in the case of the MAPK pathway, the roles of Gro proteins are well-conserved through evolution from invertebrates to mammals. However, less is known about the function of Gro proteins in mammalian development due to the presence of four *gro* orthologues, and the more restricted array of genetic tools available to study their function. Nevertheless,

several studies shed light on the role of mammalian Gro proteins in the regulation of body patterning, and cancer pathogenesis [reviewed in (Agarwal et al., 2015; Buscarlet and Stifani, 2007; Yuan et al., 2017)].

One of the processes in which Gro proteins have been described to have essential roles across vertebrates, as well as in invertebrates, is neurogenesis. A study in mice revealed that the constitutive expression of TLE1 in the post mitotic neurons of the brain inhibits neuronal development in the embryonic forebrain. In these transgenic mice, it was observed an increase in apoptosis and neural loss in the telencephalon (Yao et al., 2000), showing the central function of Gro proteins in neural differentiation in the mammalian brain. Gro proteins are also known to associate with FoxG1, a key transcriptional repressor involved in cell proliferation and differentiation during mammalian brain development. Moreover, this family of co-repressors acts as an adapter protein between FoxG1 and the Notch effector Hes1, facilitating their interaction and potentiating Hes1-mediated repression in mammalian forebrain development (Roth et al., 2010; Yao et al., 2001). Another class of proteins that interact with Gro are the homeodomain (HD) TFs, which are essential in neural fate specification in the ventral tube. Among the mouse HD proteins that recruit Gro proteins to repress downstream genes are: Nkx2.2, Nkx2.9, Nkx6.1, Dbx2, and Pax7 (Muhr et al., 2001).

Another developmental process where Gro proteins have been well explored in mammals is haematopoiesis. In this context, TLE1 interacts physically with the Runt-related transcription factor (RUNX), and with TCF. RUNX is a transcriptional regulator of the T-cell receptor α and β (TCR α and β), as well as TCF, which is also involved in the Wnt pathway. Both TFs normally function as transcriptional activators in the regulation of TCR expression and Wnt signalling, however, they can also function as repressors via recruitment of Gro (Levanon et al., 1998). Hes1 is another transcriptional repressor that depends on TLE proteins to regulate blood cell development. In this specific developmental process, Hes1 recruits TLE proteins to regulate the expression of the regulator of myeloid development named CCAAT/enhancer binding protein α (C/EBPα) (De Obaldia et al., 2013).

The functions of Gro proteins in neurogenesis and haematopoiesis mentioned above are only a few examples that illustrate the crucial function of this family of proteins in mammalian development. However, there is much more literature on the function of Gro in mammals and their interacting partners, which is summarized in Table 2.

The diverse developmental events and signalling pathways in which Gro co-repressors have a key role highlight the importance of the precise spatiotemporal regulation of their protein levels. In this context, misregulated Gro protein levels have been found in several diseases and types of tumours, such as cervical and colonic carcinoma, lung cancer, and haematological malignancies, glioblastoma, among others [reviewed in (Yuan et al., 2017)].

Developmental process	Gro/TLE family member(s) involved	Interacting partner	Interacting domain	Organism/cell type	References
Forebrain development	TLE1	BF1/FoxG1	YWPPPFLSL	Mouse	(Yao et al., 2001)
Neuronal differentiation	TLE1	Casein Kinase 2 (CK2)		Mouse	(Nuthall et al., 2004)
Neuronal differentiation	TLE1	Pin1 and HIPK2		Mouse	(Ciarapica et al., 2014)
Hematopoiesis	TLE1	AML	VWRPW, C-terminus	Human Jurkat T cells, HeLa cells, and mice	(Imai et al., 1998; Levanon et al., 1998; Wang et al., 1996)
Hematopoiesis	TLE1	LEF1		Human Jurkat T cells	(Levanon et al., 1998)
Hematopoiesis	TLE1	PRH/Hex	Eh1	Human K562 cells	(Swingler et al., 2004)
T-cell development	TLE	Hes1	WRPW	Mouse	(De Obaldia et al., 2013)
B-cell lymphopoiesis	TLE4	Pax5	C-terminal transactivation domain	Mouse and human cell lines	(Eberhard et al., 2000)
B-cell lymphopoiesis	TLE4	PU.1		Mouse and human cell lines	(Linderson et al., 2004)
Osteogenesis	TLE1, TLE2	AML	VWRPY, C-terminus	HeLa, rat osteosarcoma (ROS17/2.8)	(Javed et al., 2000)
Adipogenesis	TLE3	PPARΥ		10T1/2 and 3T3L1 Adipocytes	(Villanueva et al., 2011)
Adipogenesis	TLE3	Tcf4		10T1/2 and 3T3L1 Adipocytes	(Villanueva et al., 2011)

Pituitary development	TLE1	Hesx1	Eh1	293T cells	(Dasen, 2001)
Pituitary development	TLE1	Prop1		293T cells	(Carvalho et al., 2010)
Pancreas endocrine development	TLE3	Nkx2.2	TN domain	Pancreatic PANC1 cells	(Doyle et al., 2007; Papizan et al., 2011)
Kidney development	TLE4	Pax2		HEK 293 cells, mouse	(Cai, 2003)
Kidney development	TLE4	Cux1		Mouse	(Sharma et al., 2009)
Kidney development	TLE2, TLE3	Osr1		HEK293T cells	(Xu et al., 2014)
Heart development	TLE1, TEL3	Tbx20	Eh1	HEK293T cells, mouse	(Kaltenbrun et al., 2013)

Table 2: Table summarising the developmental processes where Gro family members are involved in mammals, and their associated protein interactors.
The table was adapted from (Agarwal et al., 2015).

1.2.2 Structural features of Gro family proteins

Groucho proteins are composed of five domains, two of them highly conserved across orthologues: The N-terminal glutamine-rich domain (also known as Q-domain), and the C-terminal WD-repeat domain. The central Gro domains, termed GP (Glycine/Proline-rich), CcN, and SP (Serine/Proline-rich), are less conserved (Figure 6).

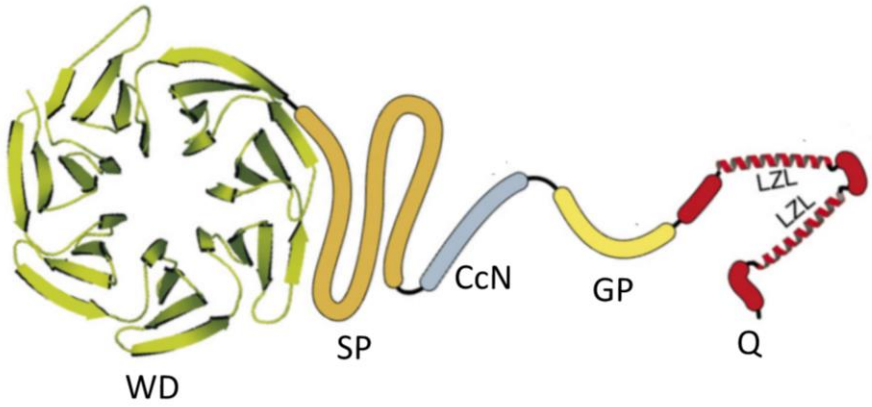


Figure 6: The structure of Gro proteins showing the five evolutionary conserved domains.
From left to right: The C-terminal WD domain, the central domains GP (Glycine/Proline-rich), CcN, and SP (Serine/Proline-rich), and the N-terminal glutamine-rich Q domain. The C-terminal WD domain (green) is the most highly conserved domain of Gro proteins, and it is composed of four-stranded β sheets. The N-terminal Q domain (red) forms two tandem leucine-zipper-like (LZL) motifs that form amphipathic α-helices. Image adapted from (Buscarlet and Stifani, 2007).

The Q-domain forms a two coiled-coil motif that facilitates oligomerization of Gro proteins into tetramers (Chen et al 1998, Pinto and Lobe 1996, Song et al 2004). This domain also mediates Groucho-binding to several repressors, including TCF proteins (Brantjes, 2001; Cavallo et al., 1998; Chodaparambil et al., 2014), and Myc (Orian et al., 2007).

TCF proteins are key downstream effectors of Wnt signalling that can act either as activators or repressors when bound to β-catenin or Gro proteins, respectively (Cavallo et al., 1998). In mammals, there are four TCF proteins; TCF-1, LEF-1, TCF-3, and TCF-4. Although all TCF proteins bind to β-catenin and TLE proteins (Brantjes, 2001), they differ from each other in their affinity for TLE-1, which directly correlates with their repressive capacity. More in detail, it is known that TCF-3 and TCF-4 bind to TLE-1 more rapidly than TCF-1 and LEF-1, and thus

TCF-3 and TCF-4 act most frequently as repressors, whereas TCF-1 and LEF-1 act most frequently as activators. TCF proteins bind to TLE-1 through the Q domain, which also forms tetramers that bind to hypoacetylated and H4K20 tri-methylated histones to promote changes in chromatin structure (Chodaparambil et al 2014). This study, together with others (Sekiya and Zaret, 2007; Song et al., 2004) indicated that oligomerization via the Q domain is important for Gro proteins mediated repression.

However, the Q domain is not always required for Gro activity *in vivo* (Jennings et al 2008) suggesting that there are multiple mechanisms of Gro-mediated repression. Specifically, they showed that *gro* alleles containing mutations in the Q domain that reduce or prevent oligomerization (gro^{MB5}, and gro^{MB12}, respectively) resulted in relatively normal segmentation and terminal patterning phenotypes in *Drosophila* embryos. Furthermore, mutations in the Q domain that abolish oligomerization did not change Gro recruitment to the genome in Kc167 *Drosophila* cells (Kaul et al 2014).

A high-resolution crystal structure of the WD-domain was solved for the human TLE1, revealing that it forms a β-propeller with seven blades around a central pore (Pickles et al., 2002) (Figure 7 A). The WD domain of TLE1 is highly conserved with *Drosophila* and its crystal structure has been used to map several missense *gro* mutations in flies (Jennings et al., 2008, 2006). This study demonstrated that the WD domain, and specifically its central pore, is critical for Gro recruitment by many repressors and TFs which contain short sequence motifs, including the WRPW and eh1 motifs (Jennings et al., 2006) (Figure 7 B). Among the Gro-dependent repressors that contain the Trp-Arg-Pro-Trp (WRPW) motif are Hairy, the seven E(spl) bHLH proteins, and Deadpan (Dpn) (Fisher et al., 1996; Jimenez et al., 1997; Paroush, 1994). The WRPW domain is commonly located at the extreme C-terminal of proteins although there are exceptions, as in the case of Hes1, which can still bind to the WD domain (Jennings et al., 2006). Other TFs that bind to Gro through C-terminal WRPW-like motifs are Runt, and Huckebein (Aronson et al., 1997; Goldstein et al., 1999). On the other hand, the Gro-binding proteins that contain an Engrailed homology-1 (eh1; consensus sequence = FxlxxlL) are Engrailed, Dorsal and Oddskipped among others (Copley, 2005; Dubnicoff et al., 1997; Jimenez et al., 1997; Tolkunova et al., 1998).

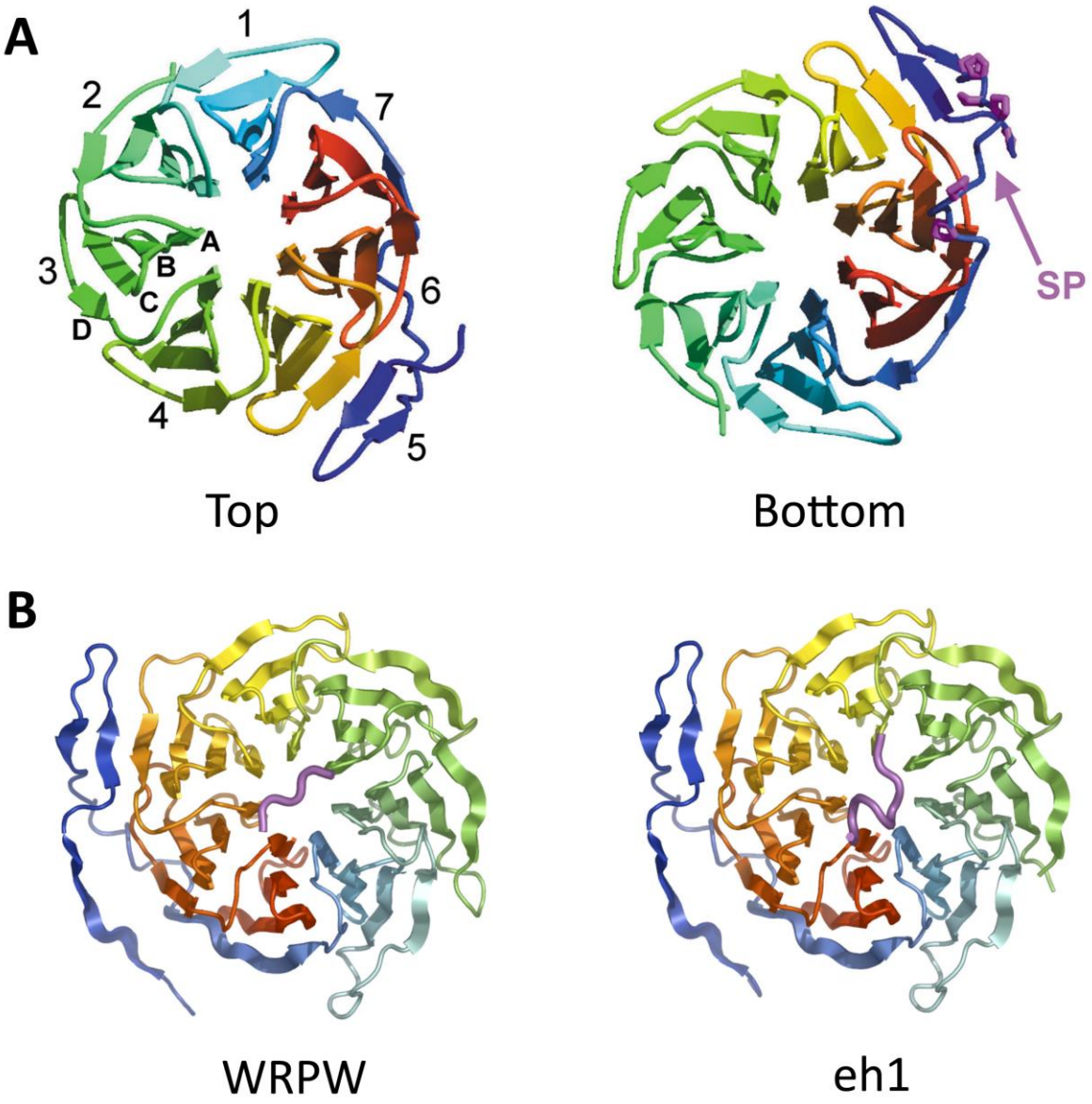


Figure 7: Secondary structure of the human TLE1 WD domain, and binding to the WRPW and eh1 peptide motifs.

(A) Ribbon diagrams of the C-terminal domain of the human TLE1, showing the seven-bladed β propeller structure. Each blade contains four β sheets (A-D in the image), excluding blade 5 which has two additional strands belonging to the C-terminal of the SP domain. The left panel shows the "top" surface, which is defined by D-A loops linking sequential blades, and the left panel shows the opposite view. Arrow points the SP domain, showing how it packs against the edges of blades 5 and 6. Proline residues are highlighted in magenta. Image adapted from (Pickles et al., 2002). **(B)** Same ribbon diagrams than in A but additionally illustrating the WRPW peptide from human Hes1 (backbone in purple on the left panel), and the eh1 peptide from human Goosecoid (backbone in purple on the right panel) on the pore of the WD domain. Image adapted from (Jennings et al., 2006).

The central Groucho domains, termed GP (Glycine/Proline-rich), CcN, and SP (Serine/Proline-rich), are less conserved, however, they have multiple roles in Groucho repression and its regulation. Among these roles are the recruitment of histone deacetylases, in which the GP

domain is involved. The subcellular localization of Groucho is regulated by the nuclear localization signal contained in the CcN domain. And the phosphorylation of Groucho, regulated by the phosphorylation sites contained in the CcN and SP domains, that are the target for a variety of Ser/Thr kinases [reviewed in (Jennings and Ish-Horowicz, 2008) and (Buscarlet and Stifani, 2007)].

1.2.3 Mechanisms of Gro mediated repression

Despite the critical role of Gro mediated repression in many signalling pathways and developmental processes, the precise molecular mechanisms underlying Gro repression remain enigmatic. Evidence from different studies suggests that Gro proteins achieve repression by different mechanisms that are possibly coordinated and include the inhibition of the Pol II complex, and changes in chromatin structure.

The first studies on *Drosophila* Gro after it was established to act as a co-repressor *in vivo* in 1994 (Paroush, 1994) showed that it acts on regulatory regions located over long distances to mediate repression (Barolo, 1997; Dubnicoff et al., 1997). Therefore, Gro was initially classified as a "long-range" repressor that blocks transcription at distances greater than 1 Kb from its genomic location. Simultaneous studies that analysed the function of Gro *in vivo* also revealed that Gro assembles into large oligomeric structures, as well as binds to the Histone Deacetylase 1 (often referred to as Rpd3 in *Drosophila*) and histones. Altogether these observations lead Chen and Courey to propose a model of Gro-mediated repression known as the "spreading model". In this model, after Gro is recruited to regulatory regions by DNA-binding TFs, it oligomerizes and interacts with histones to spread along with the DNA template. Subsequently, the formation of these large Gro domains facilitates the recruitment of chromatin-modifying factors, including Rpd3, which leads to histone deacetylation and chromatin silencing (Figure 8 A) [reviewed in (Chen and Courey 2000)]. However, a subsequent study revealed that Gro can also act locally as a "short-range" co-repressor (Nibu, 2001). Although studies suggested that Gro can function both locally and over long distances, successive models of Gro function assumed that Gro oligomerizes on DNA to mediate silencing at distant regulatory regions.

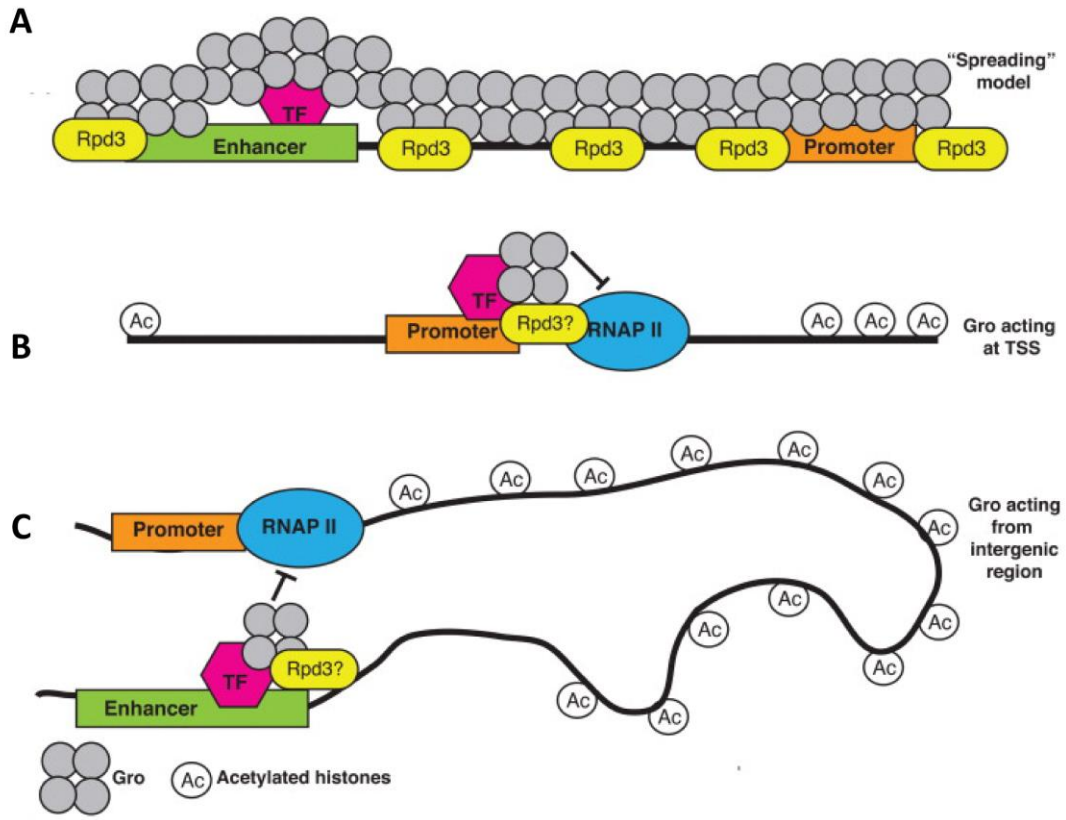


Figure 8: Models of Gro-mediated repression.

(A) Original 'spreading model' proposed by (Chen and Courey 2000). In this model, Gro oligomerises along long regions of DNA, and recruits Rpd3, which leads to histone deacetylation, and hence to closed chromatin. **(B-C)** Models proposed by (Kaul et al 2015), in which Gro can act both (B) locally (C) and from a distance to repress transcription. In both circumstances, DNA-binding proteins (TF in the image) recruit Gro to chromatin, which somehow interacts (directly or indirectly) with the Pol II complex to repress transcription at TSSs. Imaged adapted from (Kaul et al 2015).

Recent genome-wide binding data of Gro in *Drosophila* cell lines (Kaul et al 2014) and embryos (Chambers et al., 2017) have provided new evidence on the mechanisms underlying Gro-mediated repression that contradict the "spreading model". Kaul et al. analysed Gro ChIP-seq data in *Drosophila* cells showing that Gro typically binds to the genome at discrete peaks with a mean width of ~ 830 bp. This pattern of Gro binding was also observed in the human TLE3, indicating that Gro could be recruited similarly in flies than in humans (Kaul et al 2014). Moreover, this study revealed that blocking Gro oligomerization does not change the average width of the peaks. Altogether this data contradicted the "spreading model", in which Gro oligomerizes via the Q domain along several kilobases of DNA to mediate repression at distal regions.

Kaul and colleagues study also demonstrated that Gro frequently binds to TSSs of annotated genes that are not completely silenced, but rather recruit Pol II and express mRNA at some level. These results led the authors to propose a possible model in which Gro acts by promoting Pol II promoter-proximal pausing (Figure 8 B). This model is also supported by the evidence that Gro is recruited to many TSSs that have a very high pause ratio. Moreover, depletion of Gro by RNAi in cells showed a decrease in Pol II pausing at the *E(spl)mβ-HLH* locus, a high confidence Gro target, although this effect was not observed genome-wide (Kaul et al., 2014).

In addition, to find Gro peaks at TSSs, that study also revealed many other Gro peaks at intergenic regions, further from promoters. This, together with the failure to detect large and continuous Gro-bound regions made the authors favour a model where Gro bound to distant regulatory elements is brought to promoters via a looping mechanism (Figure 8 C) (Kaul et al 2015). Interestingly, a study that analysed enhancer three-dimensional (3D) contacts during *Drosophila* embryogenesis revealed that promoters of many differentially expressed genes are found in chromatin loops with enhancers (Ghavi-Helm et al., 2014). Furthermore, they found that Pol II is recruited to these promoters but paused in most cases. It is the recruitment of TFs or additional enhancer at 3D hubs that triggers the activation of transcription by the release of paused Pol II, allowing the rapid activation of those genes (Ghavi-Helm et al., 2014).

Findings in *Drosophila* embryos (Chambers et al., 2017) agree with the Gro ChIP-seq data in culture cell lines. Chambers and colleagues study demonstrated that in embryos Gro associates with chromatin in discrete, and usually transient peaks, often clustered upstream of or within genes. They also identified a set of high confidence Gro target genes that are highly enriched for paused Pol II. Thus, these results, together with the study in *Drosophila* cells, support a model in which Gro plays a role in promoter-proximal pausing.

Another mechanism by which Gro corepressors might mediate repression is the recruitment of factors that modify chromatin structure. There is a body of evidence showing that Gro interacts physically and genetically with Rpd3 in *Drosophila*. Studies in *Drosophila* cells and embryos show the functional interaction between these two proteins (Chen et al 1999, Chen et al 1998, Mannewik et al 1999, Winkler et al 2010), supporting the model in which Gro recruits Rpd3 to deacetylate histones, generating repressive chromatin. One example in which both Gro and Rpd3 are required is to maintain Runt-mediated repression of *engrailed*

in the early *Drosophila* embryo (Wheeler et al., 2002). It is also known that the repression of the Gro target gene *fushi turazu* (*ftz*) is mediated by the ectopic expression of Hairy, via histone deacetylation for several kilobases around *ftz* (Martinez and Arnosti, 2008).

Nevertheless, it is acknowledged that Gro repression could, at most, be only partially dependent on Rpd3 *in vivo*. The Kaul et al. study in *Drosophila* cell lines did not detect any significant change in histone acetylation under *gro* depletion, however, it did observe an association of Gro with regions of low histone acetylation (Kaul et al 2014). Additionally, they showed that 59% of Gro binding sites overlap Rpd3 peaks in Kc167 cells, which is consistent with the model of Gro repression involving Rpd3 in some, but not all contexts. These observations indicate that Gro does not simply repress transcription by recruiting Rpd3 to deacetylate histones driving chromatin compaction.

Importantly, HDACs are frequently part of larger multi-protein complexes that regulate transcription, such as NuRD, Sin3, CoREST, and NCoR/SMRT. NuRD (nucleosome remodelling and deacetylation) is a highly conserved complex that combines chromatin remodelling and histone deacetylation, and it has been implicated in transcriptional repression [reviewed in (Dege and Hagman, 2014)]. In this frame, TLE proteins have been shown to associate with NuRD and the cardiac TF Tbx20 to repress transcription in HEK293 cells (derived from human embryonic kidney cells) (Kaltenbrun et al., 2013). However, it has not yet been established whether the interaction between Gro and NuRD is widespread.

1.2.4 Chromatin environment at Gro-bound genomic regions

One of the mechanisms that the cell uses to control transcriptional activity is the regulation of the local chromatin composition. In this context, chromatin-based silencing mechanisms form chromatin structures that regulate nuclear organisation, repetitive elements, and protein-coding genes, among others. Traditionally, chromatin can be divided into two types that differ on histone tail modifications, transcriptional activity, and protein composition: euchromatin and heterochromatin. Euchromatin is lightly packed and bears genes that are under active (or potentially active) transcription. It is characterized by RNA polymerase, acetylated histones, and trimethylation on histone H3 at lysines 4 and 36 (H3K4me3 and

H3K36me3). In contrast, heterochromatin is densely condensed, creating an environment not permissive to transcription. Heterochromatin is enriched in histone hypoacetylation, heterochromatin protein 1 (HP1), and histone methylations such as H3K9me3 and the Polycomb modification H3K27me3 [reviewed in (Ahringer and Gasser, 2018; Beisel and Paro, 2011)].

Gro binding sites in Kc167 cells are mostly enriched in chromatin associated with active and developmentally regulated genes, defined as "the Red" chromatin (Figure 8 B) by (Filion et al., 2010). Moreover, Gro peaks are excluded from the repressed "Black" and "Green" (HP1) types of chromatin (Figure 8 B). These chromatin types are characterized by the presence of specific combinations of proteins, and differ from each other in their transcriptional activity, histone modifications, and sequence properties, among other features. The "Green", and "Blue" chromatin types are consistent with the constitutive and facultative heterochromatin (explained below in this section), respectively. The "Black" is another type of repressed chromatin. In contrast, the "Yellow" and "Red" are two types of euchromatin characterized by active transcription of housekeeping genes and developmentally regulated genes, respectively (Figure 8 A) (Filion et al., 2010). Importantly, Gro was also found to bind regions associated with DNase I hypersensitivity, indicating that they lie in open chromatin (Kaul et al 2014).

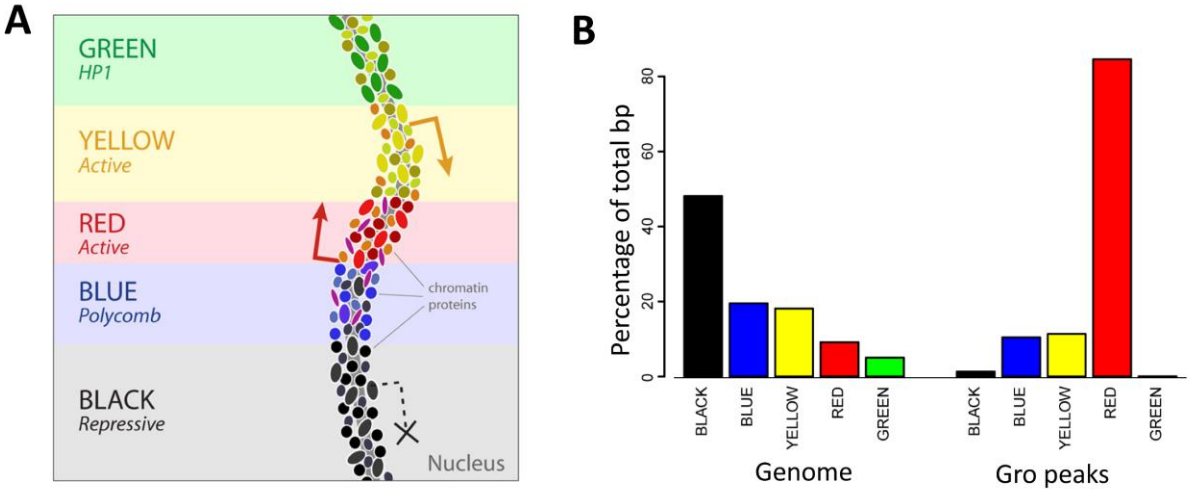


Figure 9: Types of chromatin defined by Filion and colleagues, and Gro binding sites enrichment in each type of chromatin.

(A) Image illustrating the five types of chromatin named by (Filion et al., 2010); the "Green", "Blue", and "Black" are repressive types of chromatin, whereas the "Yellow" and "Red" are active types of chromatin. Image adapted from (Filion et al., 2010). **(B)** Percentage of Gro binding sites within each of the five types of chromatin in Kc167 cells (on the right), as well as the percent of the *Drosophila* genome that map to each chromatin class in Kc167 cells. Image adapted from (Kaul et al 2014).

Taken together the data shown in this section indicates that Gro is likely to function differently from other well-characterized repressors like Polycomb Repressive Complexes (PRC), and heterochromatin.

It is noteworthy to mention that there are two different subtypes of heterochromatin, constitutive and facultative, which differ in their composition. Constitutive heterochromatin contains genome regions that are poor in genes, and rich in repetitive DNA and transposable elements. It is found at the centromeres and telomers of chromosomes, keeping them silent throughout the cell cycle. It is enriched in the histone modifications H3K9me2/3 and its ligand, HP1. Facultative chromatin is also condensed, and transcriptionally repressed, however, it is formed in genomic regions that can decondense, adapting open states within spatial-temporal contexts, such as specific cell-cycle or developmental stages. It is characterized by the histone modification H3K27me and the Polycomb group (PcG) proteins that ensure long-term transcriptional repression [reviewed in (Ahringer and Gasser, 2018; Beisel and Paro, 2011)].

In *Drosophila*, there are a total of five PcG protein complexes that include the two major complexes, Polycomb repressive complex 1 (PRC1) and PRC2. A third important PcG complex, named Pho-repressive complex (PhoRC) contains the zing finger protein Pleiohometic (Pho) that plays a role in the recruitment of PRC1 and PRC2. PcG complexes function by binding to cis-regulatory elements termed PcG response elements (PREs), and also to TSSs (Enderle et al., 2011). PREs are DNA sequences that can function as enhancers or silencers depending on the transcriptional context of their associated gene. They contain different motifs recognized by several sequence-specific DNA-binding proteins, such as GAGA factor (GAF), Zeste, Dorsal switch protein 1 (DSP1), Pipsqueak, and Grainyhead. Importantly, all these proteins also participate in many other processes other than PcG silencing, including general gene repression and activation [reviewed in (Beisel and Paro, 2011)].

1.3 Pol II promoter-proximal pausing

1.3.1 Overview of Pol II promoter-proximal pausing

Transcription occurs generally in three stages: initiation, elongation, and termination. Initiation starts with the binding of an activator TF to a promoter or an enhancer element. This binding leads to the recruitment of co-activators, histone-modifying enzymes, RNA polymerase, and general TFs, that form the preinitiation complex (PIC). The initiation stage is followed by elongation, in which the RNA polymerase moves along the DNA template, synthesising an RNA transcript. During the last stage, termination, the RNA transcript is released, and the transcription complex is dissociated from the DNA template. Each stage of transcription is tightly regulated; thus, transcription is not only regulated by the formation of the PIC but there are additional mechanisms that regulate RNA polymerase activity.

It has been long acknowledged that the transition from transcriptional initiation to elongation can be a rate-limiting step in the expression of some genes in *Drosophila* and mammals. At this checkpoint, Pol II might be at higher levels at the TSSs than within the gene body, a phenomenon known as Pol II promoter-proximal pausing [reviewed in (Core and Adelman, 2019)].

One of the first lines of evidence of genes undergoing promoter-proximal pausing came from studying the genes encoding the *Drosophila* heat-shock proteins (*hsp*). In the late 1980s studies using nuclear run-on assays demonstrated that in the absence of heat shock, transcriptionally engaged Pol II accumulates at the promoter regions of *hsp70,* in association with short nascent RNAs of approximately 25 nucleotides of length (Rougvie and Lis 1988). After the induction of heat shock, activating factors trigger the release of Pol II from promoter-proximal pausing, and there is a rapid increase of full-length transcripts (Rougvie and Lis 1988). Other early examples of genes showing promoter-proximal pausing were reported during the early elongation of the human immunodeficiency virus type 1 (HIV-1), c-*myc,* and *c-fos* genes (Kao et al., 1987; Plet et al., 1995; Strobl and Eick, 1992). Together, these and other studies in the late 1980s and early 1990s showed that there is a critical checkpoint

in gene expression at the level of early elongation and recognized the importance of understanding this checkpoint.

The transcription of the human HIV-1 was one of the classic and important models to study Pol II promoter-proximal pausing and release. HIV-1 is a retrovirus that causes Acquired Immune Deficiency Syndrome (AIDS) in humans. Upon the entry of the HIV-1 virus to the host cell, the viral genomic RNA is transcribed into cDNA and integrated into the host genome. Then, Pol II is recruited to the promoter of the HIV-1 viral gene to enable its transcription by the host transcriptional machinery; however, this process is limited at the early elongation stage. The region of the HIV-1 gene that is initially transcribed forms an RNA stem-loop structure named "transactivation response" or TAR element. The RNA TAR element is thought to interact with the Negative Elongation Factor (NELF), establishing pausing at the HIV promoter. Importantly, the viral transcriptional activator, Tat, is required to overcome this block in transcription, and allow the synthesis of the full-length HIV RNA. The tat protein achieves that by interacting with the TAR RNA, and the human Cyclin T1 (CycT1), which is part of kinase Positive Transcription Elongation Factor b (P-TEFb). This complex enables the recruitment of P-TEFb to the HIV promoter, which is required for the release of poised Pol II (see below for detailed process). The full-length HIV-1 transcripts are produced in these settings, which are required for productive viral replication [reviewed in (Ott et al., 2011)].

In the last 20 years, advances in sequencing techniques have demonstrated that this rate-limiting checkpoint at early elongation is a widespread mechanism in *Drosophila* and mammals, rather than a process restricted to a handful of specific genes (Guenther et al., 2007; Muse et al., 2007; Zeitlinger et al., 2007). Furthermore, Pol II pausing is key during animal development [reviewed in (Gaertner and Zeitlinger, 2014)].

Briefly, in the current model for Pol II promoter-proximal pausing, first, the Pol II and the general transcription factors are recruited to the promoter region of a gene. This forms the Pre-initiation complex that enables transcription initiation and the synthesis of a short, nascent RNA (Figure 10 A). Pol II then pauses close to the promoter, awaiting further signals before entering the gene body and allowing productive elongation. The fate of the pause early elongation complex is decisive for gene output [reviewed in (Chen et al 2018, Core and Adelman 2019)].

Two main protein complexes act together to stabilize Pol II in a paused conformation: the DRB sensitive-inducing factor (DSIF), and the negative elongation factor (NELF) (Core et al 2012, Gilchrist et al 2008). DSIF is a heterodimer formed by the Spt5 and Spt4 proteins that recognise the emerging nascent RNA, associating with the Pol II elongation complex. The NELF complex is made up of four subunits (NELF-A, NELF-B, NELF-C/D, and NELF-E) that recognises and binds to the Pol II-DSIF interface, stabilizing the paused Pol II (Figure 10 B). The release of paused Pol II into productive elongation requires the kinase Positive Transcription Elongation Factor b (P-TEFb), which consists of the Cyclin-dependent kinase 9 (Cdk9) and its regulator, Cyclin T (CycT) (Marshall and Price, 1995; Peng et al., 1998). Cdk9 phosphorylates DSIF, NELF, and the Pol II C-terminal domain (CTD) (Figure 10 C). The phosphorylation of NELF results in its dissociation from the elongation complex. In contrast, the phosphorylation of DSIF converts this complex into a positive elongation factor that travels with Pol II through the gene, allowing productive elongation (Figure 10 D) [reviewed by (Jennings, 2013)].

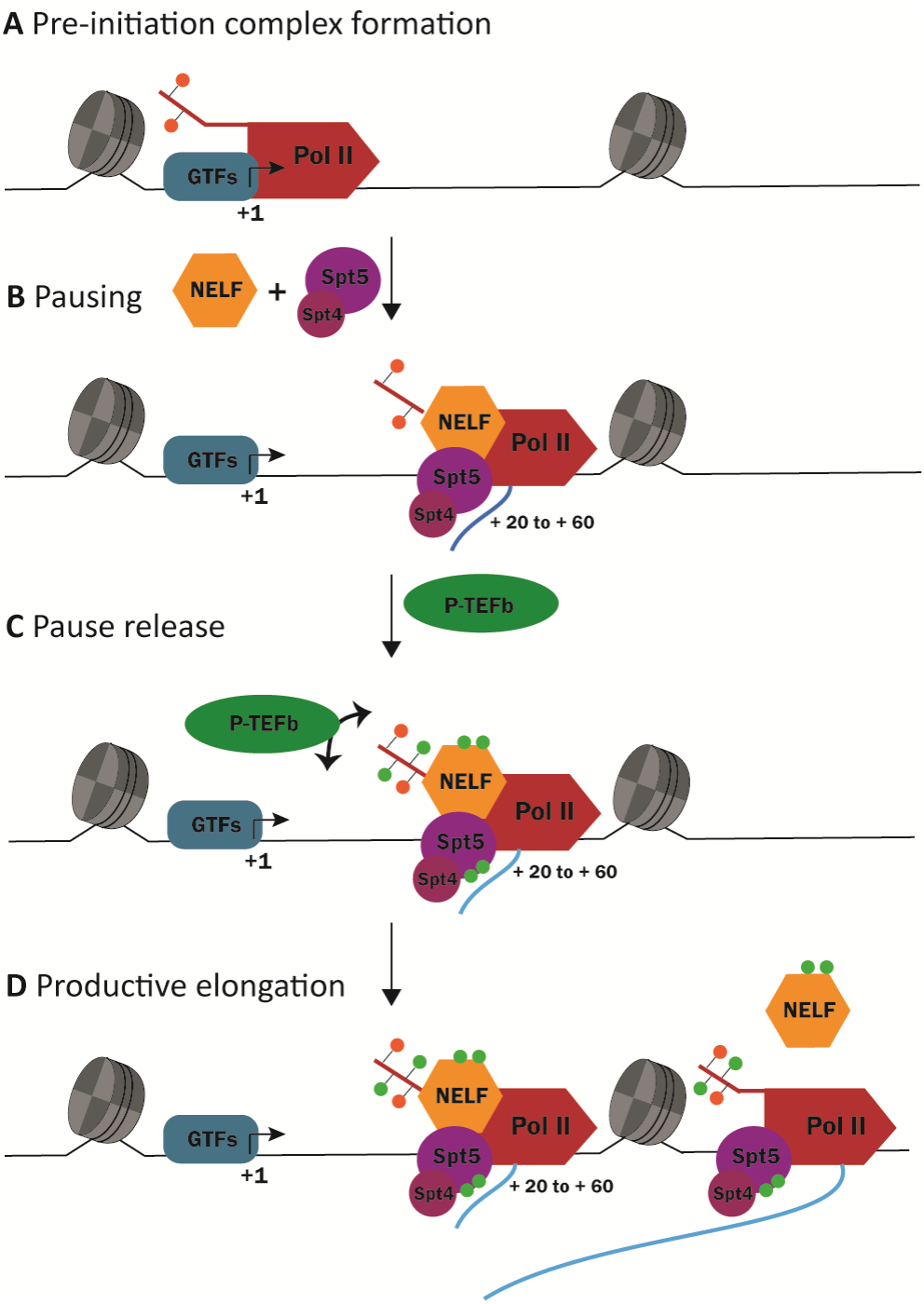


Figure 10: Establishment, maintenance, and release of Pol II pausing.

(A) General transcription factors (GTFs) recruit the Pol II to the promoter of a gene to form the preinitiation complex (TSS labelled with an arrow). **(B)** Pol II pausing occurs shortly after transcribing between 20 and 60 bp (small nascent RNA is shown in blue). The paused Pol II stabilization involves the association of the pausing factors NELF, and DSIF (complex formed by Spt5 and Spt4). **(C)** Release from pausing is driven by the recruitment and activation of P-TEFb, which phosphorylates NELF, Spt5, and Ser2 residues of the Pol II CTD (green dots). **(D)** Phosphorylated NELF dissociates from the elongation complex, whereas phosphorylated DSIF is transformed into a positive elongation factor that associates with Pol II throughout the gene.

1.3.2 Additional factors involved in establishment and release of paused Pol II

In addition to the NELF and DSIF complexes, which stabilize the paused Pol II in most or all genes, there is an increasing number of sequence-specific TFs that have been implicated in Pol II pausing. The *Drosophila* GAF is one of the important factors known to recruit NELF, and promote Pol II pausing at many genes, including *Hsp70* (Hendrix et al., 2008; Lee et al., 2008; Li and Gilmour, 2013). GAF has also been involved in the formation of permissive chromatin via the interaction with remodelling complexes, which include PBAP, NURF, and FACT [reviewed in (Pal 2017)]. A recent study proposed a two-step mechanism of GAF-mediated chromatin opening at promoters (Judd et al., 2021). In this process first PBAP forms nucleosome-free regions at promoters, in which Pol II is recruited to initiate transcription. Then, Pol II initiates transcription and progresses to the promoter-proximal pause region where it pauses. Subsequently, NURF is necessary to ensure chromatin opening at the gene body and facilitate the transcription from paused Pol II to productive elongation. In flies, the Motif Binding Protein 1 (M1BP), is also known to regulate Pol II pausing in sets of promoters distinct to GAF (Li and Gilmour, 2013).

One other factor that has been involved in Pol II pausing in mammals and *Drosophila* is the general transcription factor TFIID. However, the role of TFIID in Pol II pausing remains unclear as this factor has been shown to have roles in both the establishment of pausing (Yadav et al., 2019) and pausing release (Fant et al., 2020). The dual role of this factor in pausing could be explained by the enrichment of specific core promoter elements at some genes, since this factor can bind to different promoter sequences, including Inr, DPE, and TATA box (Shao et al., 2019) (will be discussed later in this section).

Additional complexes have been recently shown to promote pausing in *Drosophila*, such as the pre-exon junction complex (pre-EJC), which regulates splicing of long introns (Akhtar et al., 2019). And the histone chaperon complex Facilitates Chromatin Transcription (FACT), which is thought to maintain Pol II pausing by stabilizing the +1 nucleosome (Tettey et al., 2019).

As explained above, the P-TEFb complex is required for the release of paused Pol II, and it is found in two major forms that are in a reversible equilibrium. An active form, in which P-TEFb is alone or associated with proteins that recruit P-TEFb to its target genes (such as BRD4 or MYC). Alternatively, P-TEFb is an inactive form, in which P-TEFb is bound to the 7SK small nuclear ribonucleoprotein complex (7SK snRNP). This equilibrium regulates the level of active P-TEFb that stimulates transcription elongation, which is necessary to control gene expression [reviewed in (Jennings, 2013; McNamara et al., 2016; Paparidis et al., 2017)].

In mammals, the 7SK snRNP complex is composed of the 7SK small nuclear RNA (7SK snRNA), the kinase inhibitors hexamethylene bisacetamide-inducible 1/2 (HEXIM1 or HEXIM2), and 2 RNA-binding proteins; the La-related protein 7 (Larp 7) and the methylphosphate-capping enzyme (MePCE) [reviewed by (McNamara et al., 2016)]. In *Drosophila,* the 7SK snRNA complex similarly regulates P-TEFb, however, only one member of the HEXIM family is present in *Drosophila* (dHEXIM), which is responsible for blocking Cdk9 activity. There is also a *Drosophila* homolog for LARP7, dLARP7, and for MePCE, named Bicoid interacting protein 3 (Bin3) (Nguyen et al., 2012).

As mentioned previously, when P-TEFb is not bound to the 7SK snRNP complex, it is frequently associated with factors that mediate its activation and recruitment to target genes. The main factors that form this activating complex with P-TEFb are the bromodomain-containing protein 4 (BRD4), and the super elongation complex (SEC). However, the recruitment of those cofactors ultimately depends on TFs that bind to specific promoters or enhancers [reviewed in (Chen et al., 2018)].

A well-known sequence-specific TF that interacts with P-TEFb to promote pausing release at target genes in mammals is MYC (de Pretis et al 2017, Rahl et al 2010). MYC regulates gene expression globally, and it is known to interact physically with a variety of factors. Moreover, MYC function is required for proper cell growth and proliferation and can act as an oncogene in the onset of cancer [reviewed in (Baluapuri et al., 2020)]. In *Drosophila*, Myc has been associated with promoters of paused genes that are highly expressed, suggesting that poised Pol II at genes regulated by Myc is quickly released into productive elongation (Yang et al., 2013). Interestingly, Myc has been shown to interact physically and genetically with Gro to regulate a subset of Myc genes, antagonising Myc function *in vivo* (Orian et al., 2007).

1.3.3 Involvement of promoter sequences in pausing

Despite extensive published research on how Pol II pausing is released by activating factors, the exact mechanisms of pausing establishment at genes where this step is rate-limiting remain enigmatic. The properties of promoters are known to play a key role in the differences in the establishment and stability of paused Pol II among genes. In the early *Drosophila* embryo, Pol II is frequently paused at developmental genes that are enriched with combinations of core promoter elements, including the Downstream Promoter Element (DPE), GAGA motif, and the Initiator (Inr) element (Hendrix et al., 2008; Lee et al., 2008). Thus, combinations of these motifs are thought to work cooperatively to recruit and retain stalled Pol II. It has also been shown that variations in core promoter elements can influence the stability of the paused Pol II (Shao et al., 2019). Shao and colleges revealed that the presence of an Inr variant that contains a G nucleotide at position +2 increases the stability of the paused Pol II and it is found at promoters that lack TATA boxes. In addition, a variant of the DPE element that is highly enriched in CG dinucleotides, named the Pause Button (PB), has been found at approximately 25% of paused genes in the *Drosophila* embryo (Hendrix et al., 2008). The authors also revealed that CG-rich motifs were over-represented in sequences downstream of the TSSs in promoters with paused Pol II compared to constitutive promoters (Hendrix et al., 2008). Even evidence shows that Pol II pausing depends on the promoter elements, it is still unknown the exact mechanisms by which these sequences regulate Pol II pausing.

Furthermore, promoter-swapping experiments demonstrated that minimal promoter sequences are sufficient to establish paused Pol II and mediate synchronous patterns of gene activation *in Drosophila* (Lagha et al., 2013).

1.3.4 Pausing during development

In *Drosophila*, there is a large body of evidence consistent with the enrichment of paused Pol II at developmentally regulated genes (Boettiger and Levine, 2009; Wang et al., 2007; Zeitlinger et al., 2007). There are several possible functions of Pol II pausing during development [reviewed by (Gaertner and Zeitlinger, 2014)]. In this context, the presence of loaded but paused Pol II at developmental genes in the early *Drosophila* embryo could enable

the rapid and synchronous activation of gene expression upon developmental cues (Lagha et al., 2013). In addition, the presence of paused Pol II at genes that are permanently silent across tissues suggests that pausing could be a permissive step that prepares genes for future activation by tissue-specific TFs (Gaertner et al., 2012).

In the absence of activation signals, paused genes are repressed by the action of transcriptional repressors, as the *Drosophila* Snail repressor, which has been found to correlate with paused Pol II (Bothma et al., 2011; Zeitlinger et al., 2007). The mechanisms underlying repression at paused genes might involve the inhibition of activators, or the prevention of chromatin looping formation with core promoters (Bothma et al., 2011). Another mechanism that might keep paused genes from being expressed is repression by Polycomb group proteins (PcG). In flies, genes that are enriched in both paused Pol II and PcG marked genes with highly dynamic expression patterns during development (Gaertner et al., 2012).

A recent study in *Drosophila* cells that investigated gene regulation at the *hsp70* locus proposed a model in which dynamic protein-protein interactions between the PcG proteins and Spt5 could enable rapid switches in gene expression at Pho target genes. The authors revealed that Pho (a DNA-binding PcG member) can dynamically switch between interacting partners in response to the TF Heat Shock factor (HSF). More in detail, they unveiled that Pho interacts with dSfmbt, part of the PhoRC complex, under silencing conditions. Upon heat shock, HSF is expressed and recruited to chromatin, where it induces the dynamic interaction between Pho and the elongation factor Spt5, which leads to gene activation (Pereira and Paro, 2017).

It has been demonstrated that pausing also has essential functions during mammalian development, although its specific roles are still unclear. A study using global run-on sequencing (GRO-seq) in mouse embryonic stem cells (ESCs) and embryonic fibroblasts (MEFs) showed that Pol II pausing plays important role in cell differentiation in both cell types (Min et al., 2011). Interestingly, another study using mouse ESCs cells found that paused Pol II is more frequently found at genes that regulate cell cycle and signal transduction, and not developmental genes, as it would be expected (Williams et al., 2015). In that study, the genetic ablation of NELF in mouse ESCs cells caused proliferation defects, embryonic lethality, and misregulation of signalling pathways (Williams et al., 2015). Additionally, another study

has shown that hematopoietic differentiation is regulated by Pol II pausing in mammalian cells (X. Liu et al., 2017).

Although the role of paused Pol II during development in *Drosophila* has been investigated over many years, there is still little known about the phenotypic effects of mutations in factors that directly regulate this process [reviewed by (Jennings, 2013)]. A study that investigated Hexim (part of the 7SK snRNA complex) function during *Drosophila* development revealed that the knockdown of this factor disrupts organ formation. Moreover, the authors showed that the knockdown of Hexim in the developing wing disrupted the formation of this organ due to the deregulation of the Hedgehog pathway, which provokes failure in cell differentiation, and leads to apoptosis (Nguyen et al., 2016). Another study in flies that investigated Nelf, 7SK snRNP, and P-TEFb *in vivo* provided evidence that Pol II pausing also plays a role in growth control by regulating the output of the Hippo pathway. The authors showed that the RNAi downregulation of *Nelf*, *bin3,* or *Hexim* enhances growth driven by the transcriptional co-activator Yorkie (Yki), leading to neoplastic transformation of *Drosophil*a wing imaginal discs (Nagarkar et al., 2020).

2 Aims

Despite the established observations that Gro has many important roles during development, the mechanism of how it functions remains enigmatic. However, the latest studies using genome-wide techniques provided new insights into Gro-mediated repression that indicated that Gro may promote Pol II pausing, which in turn has opened new areas to research in this field.

The main aim of my project was to determine if Gro promotes promoter-proximal pausing *in vivo* to gain insight into the mechanisms of repression by Gro. To address this aim, the project was subdivided into the following three objectives:

1. Investigate and validate the possible biological interaction between Groucho and candidate factors involved with Pol II pausing derived from the literature by using genetic and biochemical assays.

2. Exploring the overlap between Gro and different key pausing factors at the TSSs and in other parts of the gene bodies by analysing ChIP-seq data deposited in databases.

3. Screen for novel interactions between Groucho and other proteins by carrying out protein-protein interaction assays.

3 Results

3.1 Biological interaction between *gro* and genes involved in Pol II pausing

As described in section 1.2.3 of the Introduction, several models describing the mechanism of Gro repression have been proposed and there is evidence that the mechanisms may differ dependent on context (Jennings et al., 2008). Among the proposed models are the direct inhibition of the transcriptional machinery, and the recruitment of large repressor complexes that contain HDACs [reviewed by (Buscarlet and Stifani, 2007)]. An additional model, supported by more recent data proposes that Gro acts promoting Pol II pausing (Chambers et al., 2017; Kaul et al., 2014). The latter model is based on molecular evidence from *Drosophila* cells but no studies *in vivo* have yet been done to test this model.

To test the model that Gro can mediate repression by promoting Pol II pausing *in vivo* I tested for genetic interactions between *gro* and genes encoding candidate factors involved in Pol II pausing. To this aim, we designed two genetic assays using the UAS/GAL4 binary system for targeted tissue-specific RNAi knockdown (Brand and Perrimon, 1993). We selected the tissue-specific knockdown since the complete depletion of *gro* by null-alleles is known to be lethal (Jennings et al., 2008). Initially, we designed a genetic assay to knock down the genes of interest in the *Drosophila* developing wing by using *nubbin-GAL4* as a driver. We decided to downregulate genes during wing development since the role of Gro in the molecular pathways that give rise to the adult wing is well-established [reviewed in (Hasson and Paroush, 2006)]. To test whether the results we observed in the developing wing could be extended to other tissues, we designed a very similar genetic assay in which we used *eyeless-GAL4* to knock down the same genes in the developing eye.

The first candidate gene that I tested is *Trl*, the gene encoding the *Drosophila* GAGA Factor (GAF); a pioneer factor that has been previously linked to Pol II pausing, among other functions (Gaskill et al., 2021; Hendrix et al., 2008; Lee et al., 2008; Li and Gilmour, 2013; Pal, 2017). We decided to test this gene first since there was already molecular data from genome-wide binding assays that indicated that *gro* and *Trl* may work together to regulate gene expression (Kaul et al., 2014). More in detail, Kaul et al. showed that 82% of Gro ChIP-seq peaks located at TSSs overlap peaks of GAF binding in *Drosophila* cells. Additionally, this study

showed that Gro binding sites were enriched by GAF motifs, among other binding motifs for known Gro partners, which is discussed in section 3.2.1.1 of the results.

In addition to *Trl*, I tested the following key genes involved in Pol II pausing: *Nelf*, *Myc*, as well as components of the P-TEFb and the 7SK complexes.

3.1.1 An RNAi-based genetic assay in *Drosophila* wings to test for interactions with *gro*

To test for genetic interactions between *gro* and genes encoding factors known to regulate Pol II pausing I used *nubbin-GAL4* (*nub-GAL4*) to drive the expression of different *UAS-RNAi* transgenes against *gro* and the candidate genes in the developing wing. *nub-GAL4* drives the expression of GAL4 under the *nubbin* promoter, which is expressed from the early second instar larval stage in the group of cells of the wing imaginal disc (known as wing pouch territory) that give rise to the adult wing blade (Cifuentes and Garcia-Bellido, 1997).

Under our experimental design *gro* and the genes of interest were simultaneously or individually downregulated in the developing wing by *nub-GAL4*, driving the expression of *UAS-RNAi*, and giving rise to double and single knockdowns, respectively. I then analysed and quantified the effects on the adult wing vein pattern in these knockdowns. The comparison of the severity of changes in the wing vein phenotype between the double and single knockdowns enabled us to establish whether *gro* and the candidate genes act together in the developmental pathways that give rise to the adult wing.

The first step to establishing the assay was to validate and characterize an *UAS-RNAi* line against *gro*. To do so I tested three *Drosophila* transgenic RNAi lines against *gro* from the Vienna Drosophila Resource Center (VDRC) (fly stocks 6316 and 6315 from the GD RNAi collection, and 110546 from the KK RNAi collection), and one from the Bloomington Drosophila Stock Center (fly stock 91407). We excluded the fly line from the KK RNAi collection (110546) as it expresses ectopic tiptop (Vissers et al., 2016), as well as the one from Bloomington as it gave no phenotype when driven by *nub-GAL4.* The other tested stocks from the VDRC GD RNAi collection (6316 and 6315) gave a very similar phenotype to each other when driven by *nub-GAL4*; phenotypes that were expected from a knockdown of *gro* (described later in this section). For the initial experiments I used both lines, however, since

the 6315 line was more difficult to maintain as a stock, I focussed on the fly line 6316 (will be referred to as *UAS-gro*^{RNAi} in this thesis). It is noteworthy to mention that in addition to *nub-GAL4*, other drivers were tested, but they gave either a very weak or not phenotype with *UAS-gro*^{RNAi} (B Jennings, personal communication).

I next generated a fly stock that expresses the *UAS-gro*^{RNAi} transgene under the *nub-GAL4* driver [genotype: *nub-GAL4/Cyo; UAS-gro*^{RNAi}*/UAS-gro*^{RNAi} (referred to *nub-GAL4 > UAS-gro*^{RNAi})], which was used for the rest of the experiments performed in this assay. In parallel, I generated other stock containing a mock *UAS* transgene [genotype: *nub-GAL4/nub-GAL4; UAS-GFP/UAS-GFP* (referred to *nub-GAL4 > UAS-GFP*)], which was used for control experiments.

3.1.1.1 Validation and characterization of *gro* RNAi knockdown in the developing wing

To validate the knockdown of *gro* by *UAS-gro*^{RNAi} I tested whether that line was able to rescue the wing phenotype caused by the overexpression of *gro* under the *nub-GAL4* driver. The overexpression of *gro* (via *UAS-gro*) driven in the developing wing by the *MS1096-Gal4* driver was previously reported to cause a phenotype characterized by small and blistered wings (Hasson et al., 2005). As expected, the expression of *UAS-gro* by *nub-GAL*4 resulted in a small and blistered wing phenotype (Figure 11 D). This phenotype was partially rescued by the co-expression with the *UAS-gro*^{RNAi} transgene (Figure 11 E), indicating that the *UAS-gro*^{RNAi} efficiently downregulated *gro*.

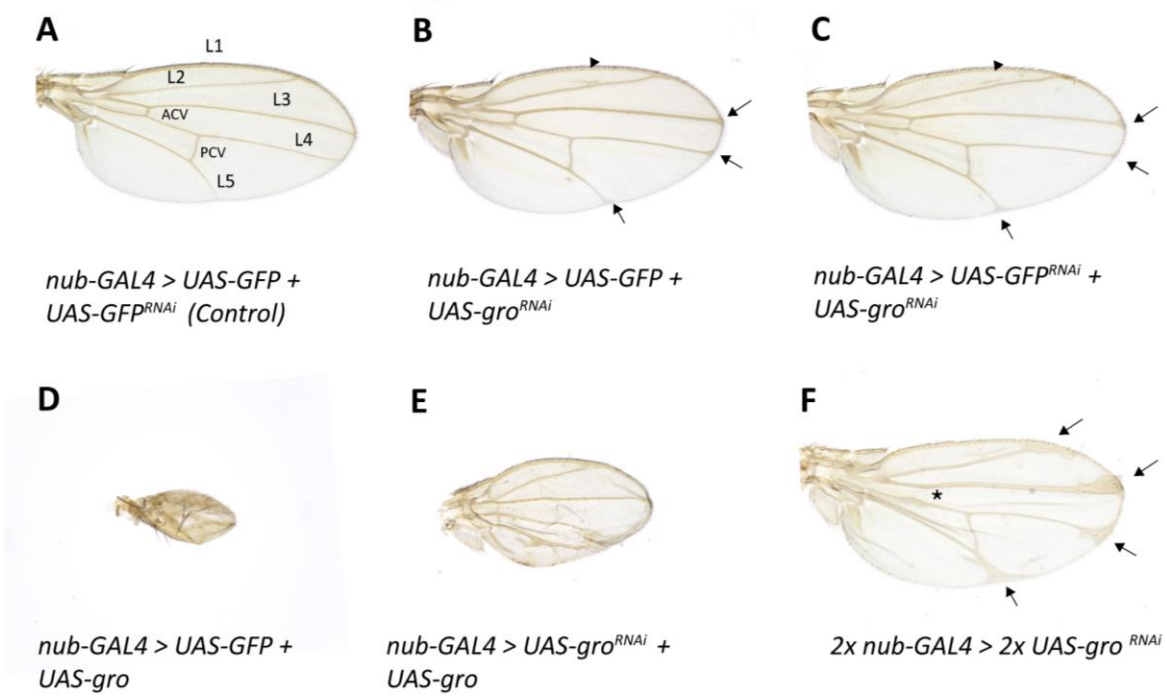


Figure 11: The RNAi downregulation of *gro* in the developing wing results in weak defects in the wing vein pattern, and it partially rescues the strong wing phenotype caused by *gro* overexpression. Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing (genotype: *nub-GAL4/UAS-GFP ^{RNAi}; UAS-GFP/+)* showing key structures: anterior cross vein (ACV), posterior cross vein (PCV), and longitudinal veins L1 to L5. **(B)** Heterozygous *gro* RNAi knockdown (genotype: *nub-GAL4/+; UAS-GFP/UAS-gro^{RNAi}*) showing ectopic venation above L2 (arrowhead), and weak longitudinal vein spreading (arrows). **(C)** Heterozygous *gro* RNAi (genotype: *nub-GAL4/UAS-GFP ^{RNAi}; +/UAS-gro^{RNAi}*) showing ectopic venation above L2 (arrowhead), and weak longitudinal vein spreading (arrows). **(D)** *gro* overexpression (genotype: *nub-GAL4/UAS-gro; UAS-GFP/+*) showing a small and blistered wing. **(E)** Simultaneous *gro* knockdown and overexpression (genotype: *nub-GAL4/UAS-gro; UAS-gro^{RNAi} /+*) showed a partial rescue of the wing size and wing vein pattern. **(F)** Homozygous *gro* RNAi knockdown (genotype: *nub-GAL4/nub-GAL4; UAS-gro^{RNAi} /UAS-gro^{RNAi}*) showing increase of ectopic venation, severely thickened veins, loss of the ACV (asterisk), and severe longitudinal vein spreading (arrows).

I next characterized the adult wing phenotypes generated by the RNAi downregulation of *gro* during wing development in the heterozygous and homozygous states. For the heterozygous, I analysed the effects on the wing vein pattern when a single copy of the *UAS-gro^{RNAi}* transgene is expressed in the developing wing. I analysed heterozygous *gro* knockdown wings with two slightly different genetic backgrounds that were generated by crossing the two main stocks I used in this assay; *nub-GAL4 > UAS-gro^{RNAi}* and *nub-GAL4 > UAS-GFP;* over a mock RNAi line (*UAS-GFP^{RNAi}*) and the *UAS-gro^{RNAi}* line, respectively (crossing scheme shown in Figure 13 C and D). The resulting offspring genotypes contained two identical transgenes; *nub-GAL4* and *UAS-gro^{RNAi}*; and differ in the location of the mock transgene. Whereas one

genotype contains it on the second chromosome (*UAS-GFP* ^{RNAi}), the other contains it on the third (*UAS-GFP*). The analysis of *gro* knockdown in both genetic backgrounds is important since they were both used to generate single and double knockdowns in the genetic assay.

I analysed the adult wing phenotypes of both heterozygous *gro* knockdowns (genotypes: *nub-GAL4/UAS-GFP*^{RNAi}*; UAS-gro*^{RNAi} */+*, and nub-*GAL4 /+; UAS-GFP/ UAS*-gro^{RNAi}). The analysis of adult wings with those genotypes showed weak effects on the wing vein patterning on both genetic backgrounds. These phenotypes were characterized by ectopic venation above the L2 vein, loss of anterior cross veins (ACVs), and weak longitudinal vein spreading (Figure 11 B and C). The wing phenotypes were highly consistent within wings of the same genotype and between genotypes, although small differences were observed in the quantification of the effects between the two genotypes. Specifically, there were small differences between the two genotypes in the percentage of wings that presented loss of ACV, and ectopic venation above L2 (Table 3). Additionally, the wing longitudinal vein spreading was quantified using a system that I designed to classify wings according to the severity of the phenotype. Wings were classified into four different classes (Class 0 to 4) that represent increasing severities of the wing vein spreading (Figure 12 A). The analysis of the longitudinal vein spreading using this system revealed highly consistent phenotypes for both heterozygous *gro* knockdowns (Figure 12 B).

The homozygous *gro* knockdown resulted in stronger defects in the wing vein pattern (Figure 11 F) than the heterozygous, as expected, causing additional loss of the ACV, strong thickened, blistered veins and, increase ectopic venation (Table 3), as well as severe longitudinal vein spreading (Figure 12 B).

The *gro* RNAi knockdown phenotypes shown in this section characterized by thickened veins and ectopic venation suggest loss of Notch signalling (Mummery-Widmer et al., 2009), and gain of EGF (Diaz-Benjumea and Hafen, 1994). Similar wing phenotypes were observed in previous studies that generated cells homozygous for endogenous *gro* mutations using clones in the wing (de Celis and Ruiz-Gómez, 1995).

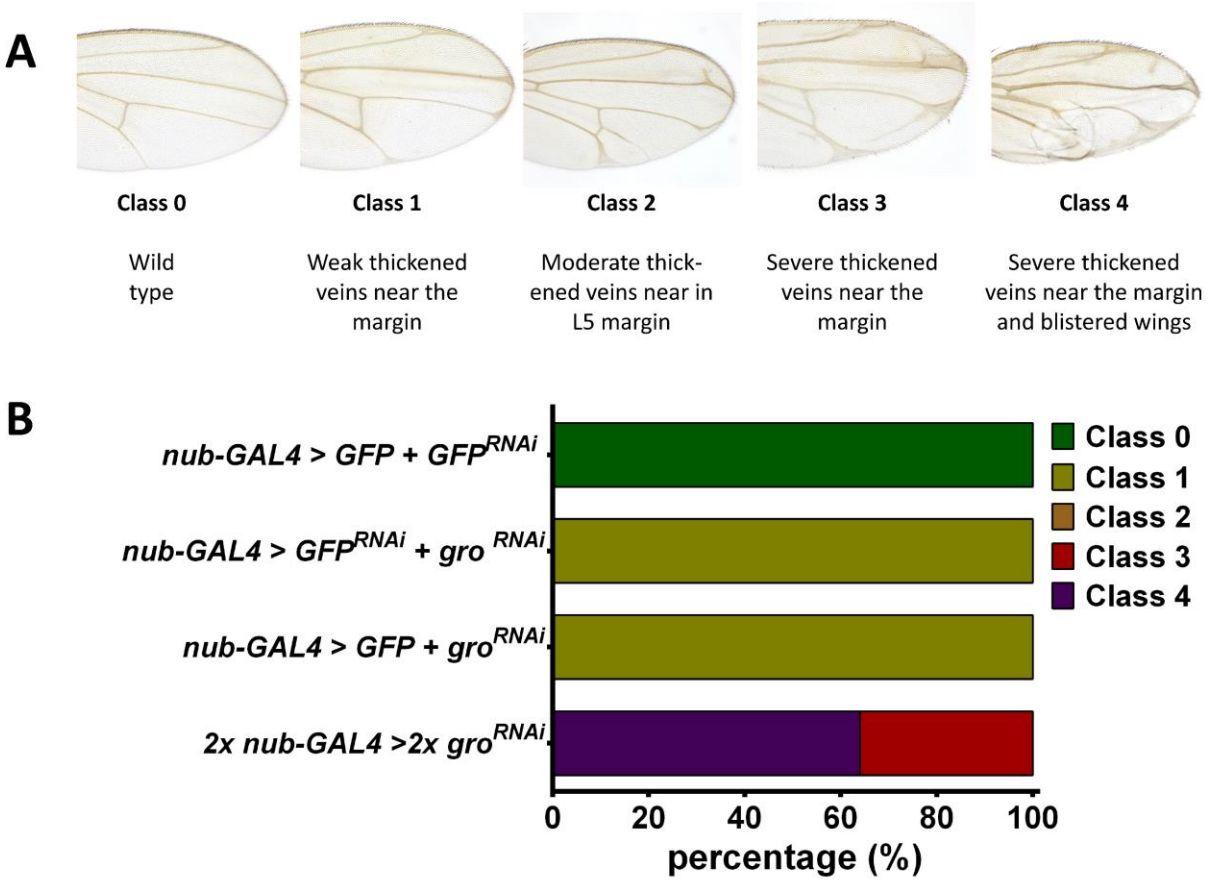


Figure 12: Classification system used to analyse the wing longitudinal vein spreading, and quantification of heterozygous and homozygous *gro* knockdown wing phenotypes.

(A) *Drosophila* adult wing images that are representative for each of the classes established in the classification system in based on the severity of the longitudinal vein spreading, and a brief description of the phenotypes. There are four classes, which represent increasing degrees of severity (Class 0, 1, 2, 3, and 4). **(B)** Quantification of the severity of the wing vein spreading phenotype for each genotype. The graph shows the frequency of each class observed in control wings, as well as in heterozygous and homozygous *gro* knockdown wings. The Y-axis indicates the *UAS* and *UAS-RNAi* transgenes expressed under the *nubbin-GAL4* driver. Genotypes from the top to the bottom of the graph are *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+*, *nub-GAL4/+; UAS-gro^{RNAi} /UAS-GFP, nub-GAL4/nub-GAL4; UAS-gro^{RNAi} /UAS-gro^{RNAi}*. Class 1 predominates in the single *gro* knockdowns, whereas Class 3 and 4 in the homozygous. Flies were raised at 25^{o}C. At least 70 wings of each genotype were scored for quantification in the panel, except for *nub-GAL4/nub-GAL4; UAS-gro^{RNAi} /UAS-gro^{RNAi}* genotype for which only 25 wings were scored.

After this primary analysis of *gro* knockdowns, the *nub-GAL4* assay was extended and used to test for possible cooperation between *gro* and genes encoding key factors involved in Pol II pausing.

Under our experimental design, the *nub-GAL4 > UAS-gro^{RNAi}* and *nub-GAL4 > UAS-GFP* stocks were crossed over several *UAS-RNAi* fly lines against the candidate genes to yield a progeny of double and single knockdowns of *gro* and the candidate genes, respectively (Figure 13 A

and B). It is noteworthy that the *nub-GAL4 > UAS-GFP* line was used to generate single knockdowns containing the same number of UAS sites as the double knockdowns.

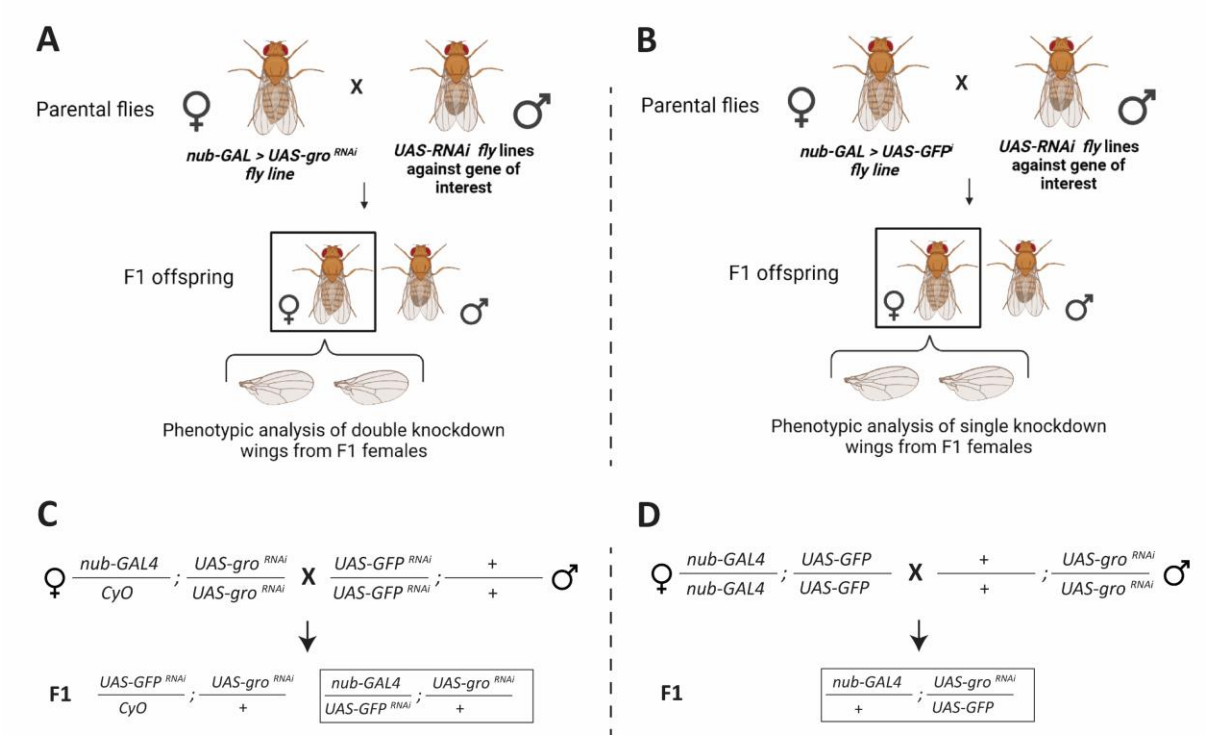


Figure 13: Schematic representation of the genetic assay designed to generate single and double knockdown *Drosophila* wings, and cross scheme used to generate the *gro* knockdown wings in two different genetic backgrounds.

(A) Females from the *nub-GAL4 > UAS-gro^{RNAi}* line were crossed over several *UAS-RNAi* lines against the genes of interest to generate double knockdown flies in the developing wing. Subsequently, the F1 offspring flies were collected and wings from females were mounted for posterior phenotypic analysis. **(B)** Females from the *nub-GAL4 > UAS-GFP* line were crossed over the same *UAS-RNAi* lines to generate single knockdown flies in the developing wing. Subsequently, the F1 offspring were collected and wings from females were mounted for posterior phenotypic analysis. **(C)** Females from the *nub-GAL4 > UAS-gro^{RNAi}* line were crossed over males of a mock RNAi line (genotype: *UAS-GFP ^{RNAi} /UAS-GFP ^{RNAi}; +/+*). F1 flies with the desired genotype (*nub-GAL4/UAS-GFP^{RNAi}; UAS-gro^{RNAi} /+*) were collected and wings from females were mounted for phenotypic analysis of *gro* knockdown wings. **(D)** Females from the *nub-GAL4 > UAS-GFP* line were crossed over males of the *UAS-gro^{RNAi}* line. F1 flies (full genotype: *nub-GAL4 /+; UAS-GFP/ UAS-gro^{RNAi}*) were collected and wings from females were mounted for phenotypic analysis of *gro* knockdown wings. Illustration created with images from BioRender.com.

3.1.1.2 Genetic interaction between *gro* and *Trl* in the developing wing

The first candidate gene I tested for interaction with *gro* was *Trl* as Gro binding data from ChIP-seq assays revealed that Gro is recruited to many sites across the genome that also recruit GAF (Kaul et al., 2014) (see also section 3.2.3.1 of the results).

The *nub-GAL4 > UAS-gro^{RNAi}* line was crossed with flies containing an *UAS-RNAi* transgene against *Trl* (41095 stock line from VDRC; will be referred to as *UAS-Trl ^{RNAi}*) to generate an offspring in which both *gro* and *Trl* are downregulated in the developing wing. In parallel, flies from the *nub-GAL4 > UAS-GFP* stock were crossed over the *UAS-Trl^{RNAi}* line to generate an offspring in which only *Trl* is downregulated in the developing wing.

The *UAS-Trl^{RNAi}* line used in this experiment had been previously validated and characterized (Blanch et al., 2015). Downregulation of *Trl* in the developing wing by this *UAS-RNAi* line resulted in a reduction of the wing blade size but no change in the wing vein patterning. The authors concluded that the reduced wing size was due to a defect in cell proliferation since no significant difference in trichome density was detected (Blanch et al., 2015). Consistent with Blanch et al observations, the only significant effect of the *Trl* knockdown I observed was a reduction of the wing blade area (Figure 14 B).

However, flies in which both *gro* and *Trl* were downregulated in the developing wing showed severe changes in the wing vein patterning (Figure 14 D). Quantification revealed an increase in loss of ACV, wing blistering, vein thickening, and ectopic venation (Table 3). In addition, I observed an increase in severe longitudinal vein spreading (Figure 14 E). These phenotypes were either much weaker or not observed in the single knockdowns of *Trl* and *gro* (Figure 14 B and C, Table 3 and Figure 14 E). Thus, the double knockdown wings showed stronger effects on the wing vein pattern than the single knockdown wings of those genes. Furthermore, the effects observed on the double knockdown wings were very similar to the effects observed in the homozygous *gro* knockdown wings (Table 3). Taken together this data supports a model in which *gro* and *Trl* act together during patterning of the wing.

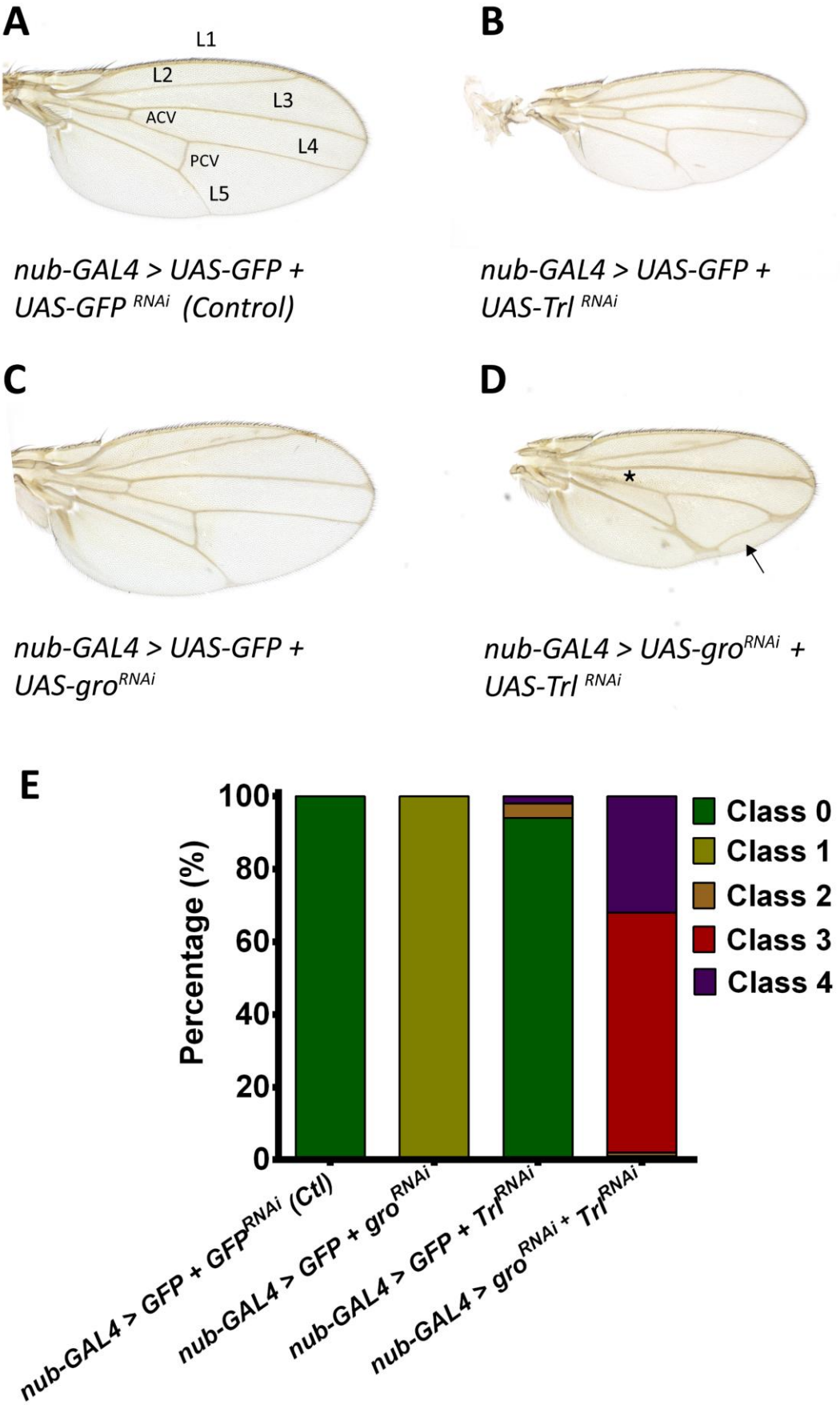

A
L1
L2
L3
ACV
L4
PCV
L5
B
nub-GAL4 > UAS-GFP + UAS-GFP RNAi (Control)
nub-GAL4 > UAS-GFP + UAS-Trl RNAi
C
D
*
nub-GAL4 > UAS-GFP + UAS-groRNAi
nub-GAL4 > UAS-groRNAi + UAS-Trl RNAi
E
Percentage (%)
100
80
60
40
20
0
Class 0
Class 1
Class 2
Class 3
Class 4
nub-GAL4 > GFP + GFPRNAi (Ctl)
nub-GAL4 > GFP + groRNAi
nub-GAL4 > GFP + TrlRNAi
nub-GAL4 > groRNAi + TrlRNAi

Figure 14: The simultaneous RNAi downregulation of *gro*, and *Trl* in the developing wing results in severe changes in the wing structure compared to the single knockdown of these genes.
Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *Trl* RNAi knockdown showing significant reduction of the wing blade size and no obvious changes in the wing vein pattern. **(C)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(D)** *gro* and *Trl* double knockdown showing ectopic venation above L2, additional ectopic venation (arrow), loss of ACV, thickened veins, and severe longitudinal vein spreading. **(E)** Quantification of the severity of the wing vein spreading phenotypes following the classification system explained in Figure 12 A. The graph indicates on the Y-axis the *UAS* and *UAS-RNA*i transgenes expressed under the *nubbin-GAL4* driver. Genotypes from the left to the right: *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/+; UAS-gro^{RNAi}/UAS-GFP, nub-GAL4/+; UAS-GFP/UAS-Trl^{RNAi}*, and *nub-GAL4/+; UAS-gro^{RNAi}/UAS-Trl^{RNAi}*. A predominance of Class 0 and Class 1 was observed in the single knockdowns of *Trl* and *gro,* respectively, whereas a predominance of Class 3 and 4 was observed in the double knockdown and *gro* and *Trl*. At least 80 wings of each genotype were scored for quantification in panel E.

3.1.1.3 Genetic interaction between *gro* and *Nelf* in the developing wing

Since this genetic assay provided biological evidence that Gro and GAF work together to regulate gene expression *in vivo*, we decided to test whether *gro* interacts with other key pausing factors. I next explored the possible genetic interaction between *gro* and *Nelf*. As described in section 1.3 of the Introduction, the recruitment of the Nelf complex (Nelf-A, -B, -D, and -E) to regulatory regions of genes is key in the establishment of paused Pol II.

The *nub-GAL4 > UAS-gro^{RNAi}* line was crossed over lines containing *UAS-RNAi* transgenes against different *Nelf* subunits to generate double knockdown flies in the developing wing. In parallel, flies from the *nub-GAL4 > UAS-GFP* line were crossed to the same *UAS-RNAi* lines against *Nelf* to yield single knockdowns. I used four RNAi lines against *Nelf* from Bloomington (32897, 34847, 38934, 32835 stock lines; will be referred to as *UAS-Nelf-A^{RNAi}*, *UAS-Nelf-B^{RNAi}*, *UAS-Nelf-D^{RNAi}*, and *UAS-Nelf-E^{RNAi}*) which have been previously validated in several studies (K. Liu et al., 2017; Nagarkar et al., 2020; Ngian et al., 2021).

The single *Nelf-B* knockdown wings showed a weak phenotype characterized by bifurcation of the posterior crossvein (PCV) and moderate longitudinal vein spreading in L5 (Figure 15 C, quantified in Figure 15 G). In contrast, the double knockdown phenotype showed a significant increase in ectopic venation, wing blistering, and thickened veins (Table 3), as well as an increase in wing vein spreading (Figure 15 G). Thus, the combined RNAi knockdown of *gro* and

Nelf-B during wing development resulted in adult wings with strong changes in the vein patterning that were not observed or weak in the single knockdown of those genes. A very similar pattern was observed in the single downregulation of *Nelf-D* (Figure 15 E, Figure 15 G), as well as in the double downregulation of *gro* and *Nelf-D* (Figure 15 F, Table 3 and Figure 15 G).

The increase in the severity of the effects on the wing observed in the double knockdowns *gro* and *Nelf-B*, and *gro* and *Nelf-D* with respect to the single knockdowns of those genes indicate that Gro cooperates with NELF to regulate gene expression *in vivo.* Furthermore, the wing phenotype observed in those double knockdowns was highly similar to that observed in the double knockdown of *gro* and *Trl* (Figure 14 D), as well as in the homozygous *gro* knockdown (Figure 11 F). Altogether this data supports that NELF, GAF, and Gro cooperate in the signalling pathways that regulate the patterning of the veins.

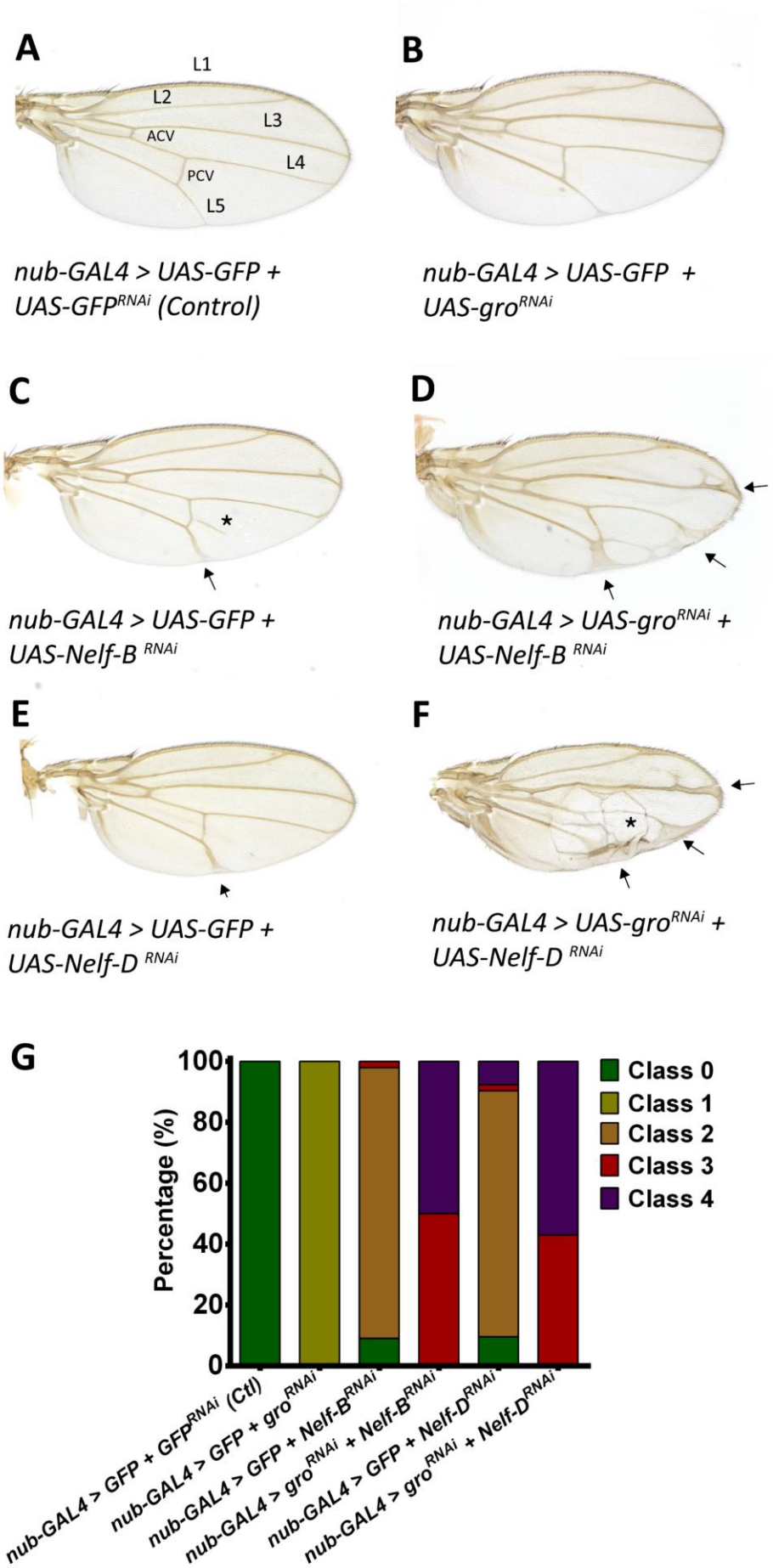

A
L1
L2
L3
ACV
L4
PCV
L5
nub-GAL4 > UAS-GFP + UAS-GFP^RNAi (Control)
B
nub-GAL4 > UAS-GFP + UAS-gro^RNAi
C
nub-GAL4 > UAS-GFP + UAS-Nelf-B RNAi
D
nub-GAL4 > UAS-gro^RNAi + UAS-Nelf-B RNAi
E
nub-GAL4 > UAS-GFP + UAS-Nelf-D RNAi
F
nub-GAL4 > UAS-gro^RNAi + UAS-Nelf-D RNAi
G
Percentage (%)
100
80
60
40
20
0
Class 0
Class 1
Class 2
Class 3
Class 4
nub-GAL4 > GFP + GFP^RNAi (Ctl)
nub-GAL4 > GFP + gro^RNAi
nub-GAL4 > GFP + Nelf-B^RNAi
nub-GAL4 > gro^RNAi + Nelf-B^RNAi
nub-GAL4 > GFP + Nelf-D^RNAi
nub-GAL4 > gro^RNAi + Nelf-D^RNAi

Figure 15: The simultaneous RNAi downregulation of *gro* and the *Nelf* subunits; *Nelf-B*, and *Nelf-D*, in the developing wing results in severe changes in the wing structure compared to the single knockdown of these genes.

Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(C)** *Nelf-B* RNAi knockdown showing bifurcation of the PCV (asterisk), and weak longitudinal vein spreading in the L5 vein (arrow). **(D)** *gro* and *Nelf-B* double knockdown showing ectopic venation, thickened veins, and severe longitudinal vein spreading (arrows). **(E)** *Nelf-D* RNAi knockdown showing weak longitudinal vein spreading in the L5 vein (arrow). **(F)** *gro* and *Nelf-D* double knockdown showing ectopic venation, thickened veins, severe longitudinal vein spreading (arrows), and blistering (asterisk). **(G)** Quantification of the severity of the wing vein spreading phenotypes following the classification system explained in Figure 12 A. Graph shows the frequency of each class observed in the single and double knockdowns wings. The graph indicates on the Y-axis the *UAS* and *UAS-RNA*i transgenes expressed under the *nubbin-GAL4* driver. Genotypes from the left to the right: *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/+; UAS-gro^{RNAi}/UAS-GFP, nub-GAL4/+; UAS-Nelf-B^{RNAi} /+*, and *nub-GAL4/+; UAS-gro^{RNAi}/UAS-Nelf-B^{RNAi},* and *nub-GAL4/UAS-Nelf-D^{RNAi}; UAS-GFP/+*, and *nub-GAL4/UAS-Nelf-D^{RNAi}; UAS-gro^{RNAi}/+.* A predominance of Class 2 and Class 1 was observed in the single knockdowns of *Nelf-B, Nelf-D* and *gro,* respectively, whereas a predominance of Class 3 and 4 was observed in both double knockdowns and *gro* and *Nelf-B,* and gro and *Nelf-D*. At least 80 wings of each genotype were scored for quantification in panel G, except for *nub-GAL4/UAS-Nelf-D^{RNAi}; UAS-GFP/+*, and *nub-GAL4/UAS-Nelf-D^{RNAi}; UAS-gro^{RNAi}/+* for which I scored 48, and 50 wings, respectively.

The combined downregulation of *gro* and *Nelf-A* during wing development resulted in a mild phenotype. Specifically, the *gro/Nelf-A* double-knockdown wings (Figure 16 D) showed an enhancement of thickened veins and wing folding (Table 3) whereas no effect was observed in the single downregulation of *Nelf-A* (Figure 16 C, Table 3 and Figure 16 G). In comparison to the double knockdowns of *gro/Nelf-B* and *gro/Nelf-D*, the double knockdown of *gro* and *Nelf-A* resulted in much milder effects on the wing vein pattern. In this case, we cannot rule out the possibility of *Nelf-A* expression was not effectively reduced during wing development at 25°C since no effect was observed in the single *Nelf-A* knockdown wings at this temperature (Figure 16 C, Table 3).

Finally, the wing phenotype caused by the simultaneous knockdown of *gro* and *Nelf-E* (Figure 16 F) during wing development was very similar to the individual knockdown phenotypes of those genes (Figure 16 B and E). Thus, no significant difference was observed between the single and the double knockdown wings in the phenotypic analysis (Table 3 and Figure 16 G). Interestingly, the single knockdown of *Nelf-E* resulted in notching in approximately 8% of the wings scored and no other obvious effects (Table 3). This effect on the wing vein pattern contrasts with the observed in the single knockdowns of *Nelf-B* and *Nelf-D*, which showed moderate longitudinal vein spreading in L5 instead (Figure 15 C and E). This data suggests that

the role of *Nelf-E* during wing development may differ from the role of Nelf-*B* and *Nelf-D* subunits. In this context, several studies showed that the disruption of the distinct NELF subunits has different impacts on gene expression depending on the cell type and biological process (Gilchrist et al., 2008; Ngian et al., 2021; Wang et al., 2010). Furthermore, the Nelf-E subunit was reported not required for Pol II pausing *in vitro* (Missra and Gilmour, 2010). Altogether this data suggest that *Nelf-E* might not be as necessary as the *Nelf-B* and *Nelf-D* subunits for Pol II pausing, and either for Gro to mediate repression.

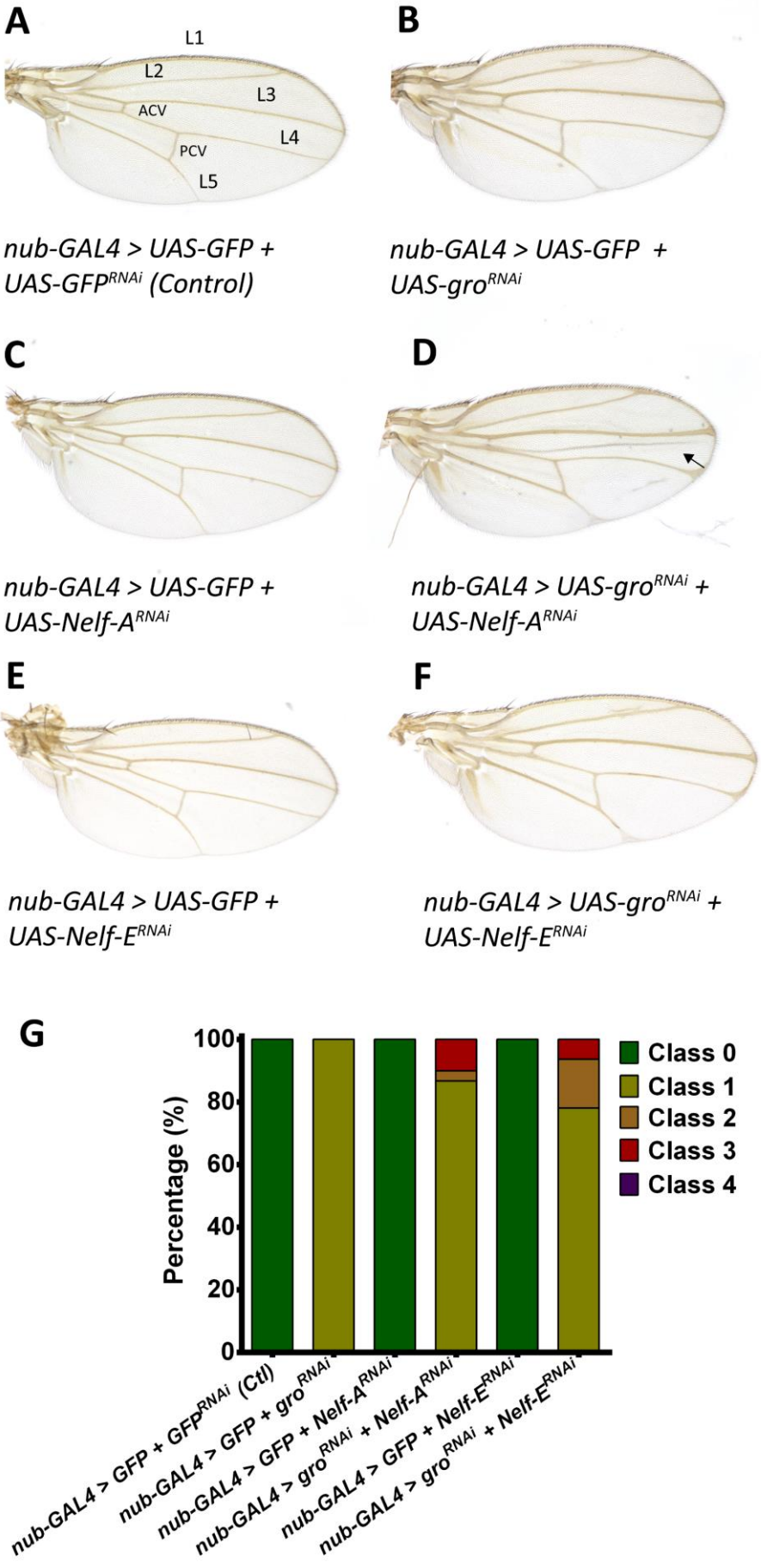

A
L1
L2
L3
ACV
L4
PCV
L5
nub-GAL4 > UAS-GFP + UAS-GFP^RNAi (Control)
B
nub-GAL4 > UAS-GFP + UAS-gro^RNAi
C
nub-GAL4 > UAS-GFP + UAS-Nelf-A^RNAi
D
nub-GAL4 > UAS-gro^RNAi + UAS-Nelf-A^RNAi
E
nub-GAL4 > UAS-GFP + UAS-Nelf-E^RNAi
F
nub-GAL4 > UAS-gro^RNAi + UAS-Nelf-E^RNAi
G
Percentage (%)
100
80
60
40
20
0
Class 0
Class 1
Class 2
Class 3
Class 4
nub-GAL4 > GFP + GFP^RNAi (Ctl)
nub-GAL4 > GFP + gro^RNAi
nub-GAL4 > GFP + Nelf-A^RNAi
nub-GAL4 > gro^RNAi + Nelf-A^RNAi
nub-GAL4 > GFP + Nelf-E^RNAi
nub-GAL4 > gro^RNAi + Nelf-E^RNAi

Figure 16: The simultaneous RNAi downregulation of *gro*, and the Nelf subunits; *Nelf-A*, and *Nelf-E*, in the developing wing does not lead to changes in wing structure compared to the single knockdown of these genes.

Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(C)** *Nelf-A* RNAi knockdown showing no obvious change in the wing vein pattern. **(D)** *gro* and *Nelf-A* double knockdown showing thickened veins and wing blistering (arrow). **(E)** *Nelf-E* RNAi knockdown no obvious change in the wing vein patterning **(F)** *gro* and *Nelf-E* double knockdown showing ectopic venation above L2, and weak longitudinal vein spreading (arrows). **(G)** Quantification of the severity of the wing vein spreading phenotypes following the classification system explained in Figure 12 A. Graph shows the frequency of each class observed in the single and double knockdowns wings. The graph indicates on the Y-axis the *UAS* and *UAS-RNA*i transgenes expressed under the *nubbin-GAL4* driver. Genotypes from the left to the right: *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/+; UAS-gro^{RNAi}/UAS-GFP, nub-GAL4/+; UAS-Nelf-A^{RNAi}/+*, and *nub-GAL4/+; UAS-gro^{RNAi}/UAS-Nelf-A^{RNAi},* and *nub-GAL4/+; UAS-Nelf-E^{RNAi}/+*, and *nub-GAL4/+; UAS-gro^{RNAi}/ UAS-Nelf-E^{RNAi}*. A predominance of Class 1 was observed in the double knockdowns of *gro* and *Nelf-A, and gro and Nelf-E,* as well as in single knockdowns and of those genes. At least 60 wings of each genotype were scored for quantification in panel G, except for *nub-GAL4/+; UAS-gro^{RNAi}/ UAS-Nelf-E^{RNAi}* for which I scored 32 wings.

3.1.1.4 Genetic interaction tested between *gro* and genes encoding components of the P-TEFb and 7SK snRNP complexes in the developing wing

To further explore the role of Gro in Pol II pausing I tested whether Gro cooperates with the P-TEFb complex (Cdk9 and CycT in *Drosophila*), as well as with components of the 7SK snRNP complex. P-TEFb plays a major role in the release of paused Pol II and the 7SK snRNP complex regulates the activity of P-TEFb, in both flies and mammals (see section 1.3 of the Introduction). Gro could potentially promote Pol II pausing by either inhibiting P-TEFb activity directly, or by stabilizing the association of P-TEFb with the inhibitory 7SK snRNP complex.

The RNAi downregulation of *Cdk9* driven by *nub-GAL4* was lethal as no flies were recovered from the cross. This was also observed in the double knockdown of *gro* and *Cdk9*, hence we could not proceed with this experiment (Table 3). In contrast, the downregulation of *CycT* alone or in combination with *gro* did not result in any obvious change in the wing vein pattern, (Table 3). The *UAS-RNAi* fly line against *CycT* I have used in this experiment (35168 stock line from Bloomington; will be referred to *UAS-CycT ^{RNAi}*) was previously validated in another study (Zeng et al., 2015). In that case, the authors observed cell death when *UAS-CycT ^{RNAi}* was expressed in the midgut in flies raised at 29°C. Considering I did not observe any change in

the wing phenotype in the single knockdown of *CycT* at 25ºC despite the critical role of this gene in the regulation of gene expression, it is possible that *CycT* was not properly downregulated at 25ºC. Taken all together, we cannot conclude from this experiment whether *gro* genetically interacts with *CycT* and/or *Cdk9* during wing development.

In addition, I explored the possible cooperation between Gro and three components of the 7SK snRNP complex: Hexim, Larp7, and Bin3.

Two RNAi fly lines containing different *UAS-RNAi* transgenes against *Hexim* (one from Bloomington:34632, and one from VDRC: 100500 KK), which have been reported previously in the literature (Nagarkar et al., 2020; Nguyen et al., 2012) were tested. The single knockdown of *Hexim* by *nub-GAL4* by both RNAi lines was mostly lethal, and the few flies that hatched had very small and blistered wings when raised at 25ºC or 18ºC. As a result of the high lethality, I observed in the RNAi lines against *Hexim* I could not go further with these experiments, and hence I could not test whether *Hexim* genetically interacts with *gro* in this system.

Finally, the single downregulation of the other two components of the 7SK snRNP complex, *Larp7* and *Bin3*, resulted in no obvious change in the wing vein pattern, although I did observe a reduction of the wing blade size (Figure 17 C and 17 E). However, the double knockdowns of *gro* and these genes (Figure 17 D and 17 F) resulted in severe changes in the wing vein pattern. Specifically, the phenotypic analysis of the double knockdown wings *gro* and *Larp7* showed a significant increase in the loss of ACV, vein thickening, and notching (Table 3), as well as an increase in the vein wing spreading (Figure 17 G). A very similar phenotype was observed in the double knockdown of *gro* and *Bin3* (Figure 17 F) showing a significant increase in the loss of ACV and PCV, vein thickening (Table 3), and longitudinal vein spreading (Figure 17 G).

Taking all this data together, we can conclude that Gro acts along with the 7SK snRNP complex to promote Pol II pausing during wing development

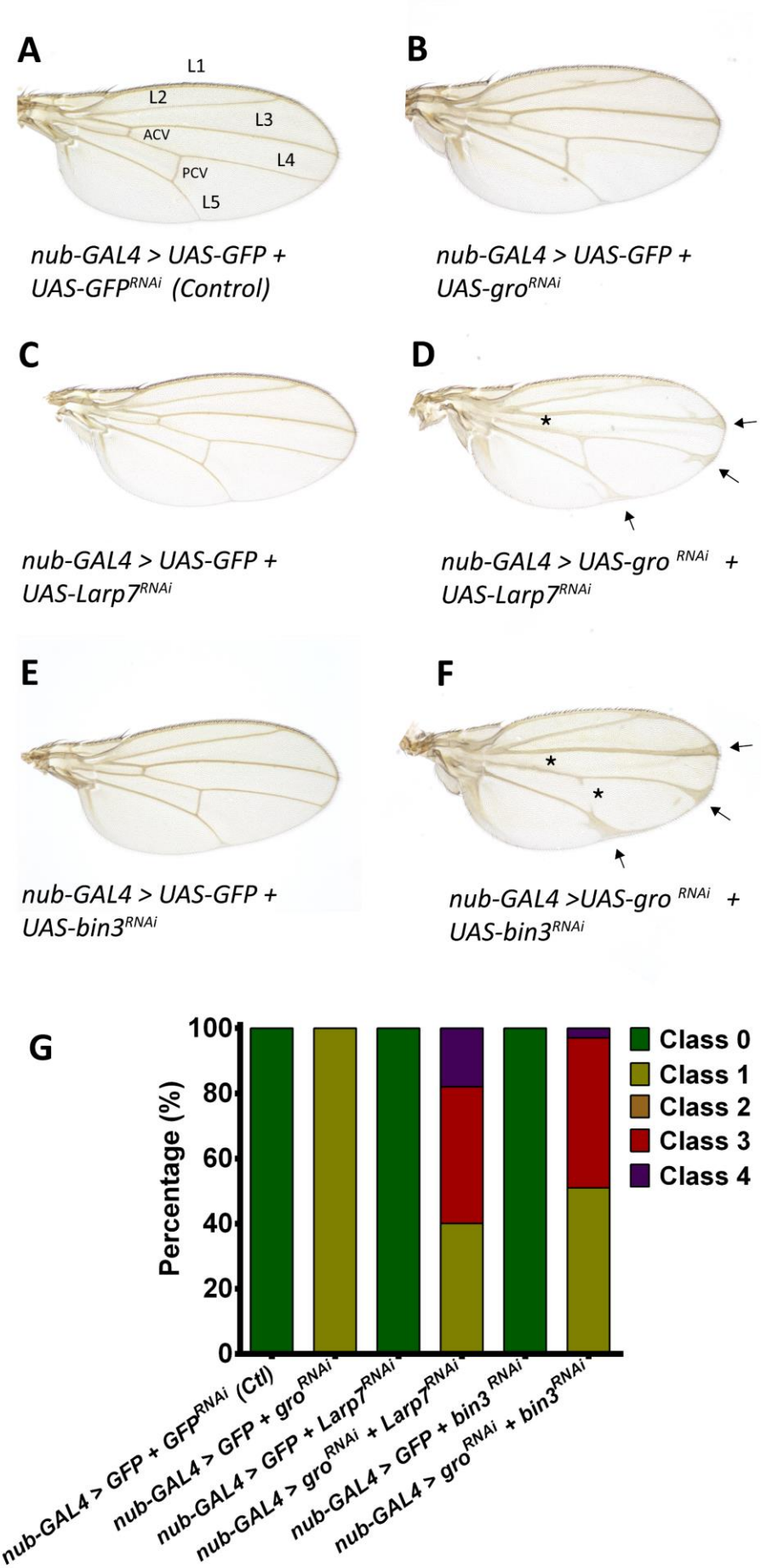

A
L1
L2
L3
ACV
L4
PCV
L5
nub-GAL4 > UAS-GFP + UAS-GFP^RNAi (Control)
B
nub-GAL4 > UAS-GFP + UAS-gro^RNAi
C
nub-GAL4 > UAS-GFP + UAS-Larp7^RNAi
D
nub-GAL4 > UAS-gro ^RNAi + UAS-Larp7^RNAi
E
nub-GAL4 > UAS-GFP + UAS-bin3^RNAi
F
nub-GAL4 >UAS-gro ^RNAi + UAS-bin3^RNAi
G
Percentage (%)
100
80
60
40
20
0
Class 0
Class 1
Class 2
Class 3
Class 4
nub-GAL4 > GFP + GFP^RNAi (Ctl)
nub-GAL4 > GFP + gro^RNAi
nub-GAL4 > GFP + Larp7^RNAi
nub-GAL4 > gro^RNAi + Larp7^RNAi
nub-GAL4 > GFP + bin3 ^RNAi
nub-GAL4 > gro^RNAi + bin3^RNAi

Figure 17: The simultaneous downregulation of *gro* and the components of the 7SK sRNP complex; *Larp7*, and *Bin3* in the developing wing results in severe changes in the wing structure compared to the single knockdown of these genes.

Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(C)** *Larp7* RNAi knockdown showing no obvious change in the wing vein pattern. **(D)** *gro* and *Larp7* double knockdown showing ectopic venation above L2, thickened veins, loss of ACV (asterisk), and severe longitudinal vein spreading **(E)** *bin3* RNAi knockdown showing no obvious change in the wing vein patterning **(F)** *gro* and *bin3* double knockdown showing loss of ACV and PCV (asterisks), thickened veins, and severe longitudinal vein spreading (arrows). **(G)** Quantification of the severity of the wing vein spreading phenotypes following the classification system explained in Figure 12 A. The graph indicates on the Y-axis the *UAS* and *UAS-RNA*i transgenes expressed under the *nubbin-GAL4* driver. Full genotypes from the left to the right: *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/+; UAS-gro^{RNAi}/UAS-GFP, nub-GAL4/UAS-Larp7^{RNAi}; UAS-GFP/+*, *nub-GAL4/ UAS-Larp7^{RNAi}; UAS-gro^{RNAi}/+,* and *nub-GAL4/+; UAS-bin3^{RNAi}/+*, and *nub-GAL4/+; UAS-gro^{RNAi}/UAS-bin3^{RNAi}.* A predominance of Class 0 was observed in the single knockdowns of *Larp7* and *Bin3*; Class1 in the single knockdown of *gro*; Class 3 and 4 in the double knockdown of *gro* and *Larp7;* and Class 1 and 3 in the double knockdown *gro* and *bin3.* At least 80 wings of each genotype were scored for quantification in panel G, except for *nub-GAL4/ UAS-Larp7^{RNAi}; UAS-gro^{RNAi}/+,* for which I scored 50 wings.

3.1.1.5 Genetic interaction between *gro* and *Myc* in the developing wing

The last factor involved in Pol II pausing I tested in detail by this genetic assay was *Myc*. This TF has been found to play a key role in the release of the paused Pol II in mammals via the recruitment of the P-TEFb complex (de Pretis et al., 2017; Rahl et al., 2010) (see section 1.3.2 of Introduction). Myc has also previously been shown to interact with *gro* both physically (by co-immunoprecipitation) and genetically in the developing embryonic nervous system (Orian et al., 2007). In this study, Gro was found to antagonise Myc function, which would be consistent with a role for Gro in inhibiting Myc's ability to promote Pol II pause release.

To test whether *Myc* interacts with *gro* during wing development I used a *UAS-RNAi* fly line against *Myc* from Bloomington (35784 fly stock; will be referred to as *UAS-Myc ^{RNAi}*). The expression of this transgene in the developing wing by *nub-GAL4* did not result in any change in the wing vein pattern, however, the wing size was reduced, as expected (Figure 18 C). In contrast, the simultaneous downregulation of *Myc* and *gro* resulted in strong changes in the wing vein pattern compared to the single knockdown of *gro* and *Myc* (Figure 18 D). *gro* and *Myc* double-knockdown wings showed an increase in loss of ACV and PCV (Table 3), vein

thickening, as well as an increase in the severity of the longitudinal vein spreading (Figure 18 E). This phenotype suggests that there is an interaction between *gro* and *Myc* in the developing wing. The observation that reducing Myc function enhances the *gro* loss of function phenotype is perhaps unexpected as they have previously been shown to have an antagonistic relationship.

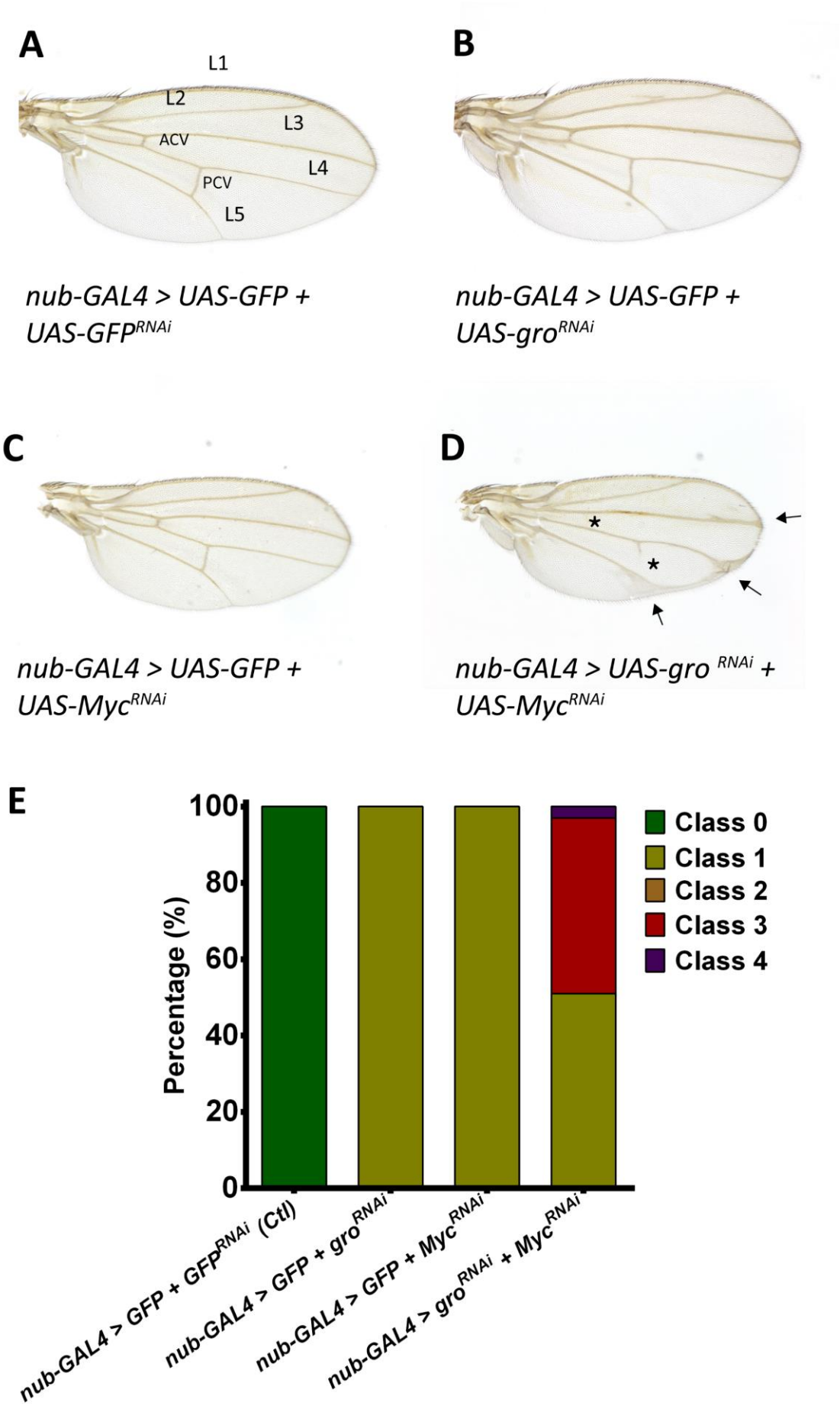

A
L1
L2
L3
ACV
L4
PCV
L5
B
nub-GAL4 > UAS-GFP + UAS-GFP^RNAi
nub-GAL4 > UAS-GFP + UAS-gro^RNAi
C
D
nub-GAL4 > UAS-GFP + UAS-Myc^RNAi
nub-GAL4 > UAS-gro ^RNAi + UAS-Myc^RNAi
E
Percentage (%)
100
80
60
40
20
0
Class 0
Class 1
Class 2
Class 3
Class 4
nub-GAL4 > GFP + GFP^RNAi (Ctl)
nub-GAL4 > GFP + gro^RNAi
nub-GAL4 > GFP + Myc^RNAi
nub-GAL4 > gro^RNAi + Myc^RNAi

Figure 18: The double RNAi knockdown of *gro*, and *Myc* in the developing wing results in severe changes in the wing structure compared to the single knockdown of these genes.
Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *Myc* RNAi knockdown showing a significant reduction of the wing blade size and no obvious changes in the wing vein pattern. **(C)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(D)** *gro* and *Myc* double knockdown showing loss of ACV and PCV (asterisks), thickened veins, and severe longitudinal vein spreading. **(E)** Quantification of the severity of the wing vein spreading phenotypes following the classification system explained in Figure 12 B. Graph shows the frequency of each class observed in the single and double knockdowns wings. The graph indicates on the Y-axis the *UAS* and *UAS-RNA*i transgenes expressed under the *nubbin-GAL4* driver. Genotypes from the left to the right: *nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+*, *nub-GAL4/+; UAS-gro^{RNAi}/UAS-GFP, nub-GAL4/+; UAS-GFP/UAS-Myc^{RNAi}*, and *nub-GAL4/+; UAS-gro^{RNAi}/UAS-Myc^{RNAi}.* A predominance of Class 0 and Class 1 was observed in the single knockdowns of *Myc* and *gro,* respectively, whereas a predominance of Class 1, 3 and 4 was observed in the double knockdown and *gro* and *Myc.* At least 70 wings of each genotype were scored for quantification in panel G, except for *nub-GAL4/+; UAS-gro^{RNAi}/UAS-Myc^{RNAi},* for which I scored 56 wings.

3.1.1.6 Additional factors involved with transcriptional regulation tested with RNAi lines, and endogenous mutations

The wing genetic assay was used to test the possible genetic interaction between *gro* and the *vitamin D3 up-regulated protein 1* (*Vdup1*). To do so, we used a *Vdup1^{RNAi}* line that was already validated in a previous study (Oberacker et al., 2018). The Vdup1 protein was detected as a possible Gro interactor in the Y2H assay, however, this protein resulted in a false positive when the full-length Vdup1 protein was retested (see chapter 3.3.1). Vdup1 is a tumour suppressor protein that acts blocking the cell cycle progression through the repression of AP-1-mediated transcription (Nishinaka et al., 2004). As expected, no genetic interaction was detected between *gro* and *Vdup1* (Figure 57 in the Appendix). Since Vdup1 is known to regulate transcription but has not been involved in Pol II pausing, this negative result makes stronger the evidence shown in this chapter. Thus, it is a good negative control that shows that the genetic interactions observed in the wing are specific to the involvement of *gro* in Pol II pausing and are not observed unspecifically with any transcriptional regulators.

A number of other additional factors implicated in the regulation of Pol II pausing were tested in this assay by using *UAS-RNAi* lines (including *Spt5* and *Pho*); the results are summarized in Table 3, along with results for the factors already mentioned.

In addition, many of the factors tested using *UAS-RNAi* lines were also tested using endogenous mutations in heterozygosity with *gro* knockdown. One of these factors was *Hexim*, which could not be tested using *UAS-RNAi* lines as it resulted lethal. However, none of these gave significant enhancement of the *gro* knockdown phenotype (Table 9 in the Appendix).

	Total or Partial loss of ACV (%)	Total or Partial loss of PCV (%)	Ectopic venation above L2 (%)	Notching (%)	Wing blistering (%)	Strong vein thickening (%)	Other effects observed (%)
nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+	0	0	7.2	0	0	0	
nub-GAL4/+; UAS-GFP/ UAS-gro^{RNAi}	9.1	0	100***	0	0	0	
nub-GAL4/ UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+	26.7***	0	81.3***	0	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-gro^{RNAi}	95.7***	15	100	0	64***	100***	Increased ectopic venation
nub-GAL4/+; UAS-GFP/ UAS-Trl^{RNAi} (41095)	30.6***	0	19***	0	1.0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Trl^{RNAi} (41095)	100***	0	98.3	0	40***	97.8***	Increased ectopic venation
nub-GAL4/+; UAS-GFP/ UAS-Nelf-B^{RNAi} (34847)	0	0	32.7***	0	0		PCV bifurcation, thicker L5 from the PCV
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Nelf-B^{RNAi} (34847)	0	0.9	72.6	0	56.2***	100***	Increased ectopic venation
nub-GAL4/UAS-Nelf-D^{RNAi} (38934)*; UAS-GFP/+*	0	6.1	55.1***	0	6.3	1.9	Thicker L5 from the PCV
nub-GAL4/UAS-Nelf-D^{RNAi} (38934)*; UAS-gro^{RNAi}/+*	14.6	6.25	95.8		59.3***	100***	Increased ectopic venation

nub-GAL4/+; UAS-GFP/ UAS-Myc^{RNAi} (35784)	0	0	5.5	0	0	0	
nub-GAL4/+; UAS-gro^{RNAi} / UAS-Myc^{RNAi} (35784)	76***	22.2***	7.5	0	0	96.8***	Increased ectopic venation
nub-GAL4/UAS-Larp7^{RNAi} (31635)*; UAS-GFP/+*	0	0	0.6	5	0	0	
nub-GAL4/UAS-Larp7^{RNAi} (31635)*; UAS-gro^{RNAi}/+*	92.7***	12.2	89.4	34.7***	0	60***	Pigmentation defects
nub-GAL4/+; UAS-GFP/ UAS-bin3^{RNAi} (50634)	0	0	1.9	2.5	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-bin3^{RNAi} (50634)	96.4***	41.8***	65.5	3.6	3.6	49***	Increased ectopic venation
nub-GAL4/UAS-Spt5^{RNAi} (106814)*; UAS-GFP/+*	0	0	1.1	0	0	0	
nub-GAL4/UAS-Spt5^{RNAi} (106814)*; UAS-gro^{RNAi}/+*	86.4***	0	100	0	0	0	
nub-GAL4/UAS-Spt5^{RNAi} (4075)*; UAS-GFP/+*	0	0	0	0	0	0	
nub-GAL4/UAS-Spt5^{RNAi} (4075)*; UAS-gro^{RNAi}/+*	54.5	0	100	0	0	20.5	Pigmentation defects, wing folding
nub-GAL4/+; UAS-GFP/ UAS-Cdk12^{RNAi} (35163)	0	0	7.8	0	0	0	

nub-GAL4/+; UAS-gro^{RNAi} / UAS-Cdk12^{RNAi} (35163)	42.9	1.2	78.6	0	0	9.5	
*nub-GAL4/+; UAS-GFP/ UAS-Cdk9^{RNAi} (*35323)	LETHAL	LETHAL	LETHAL	LETHAL	LETHAL	LETHAL	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Cdk9^{RNAi}(35323)	LETHAL	LETHAL	LETHAL	LETHAL	LETHAL	LETHAL	
nub-GAL4/+; UAS-GFP/ UAS-CycT^{RNAi} (35168)	0	0	14.3	0	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-CycT^{RNAi} (35168)	42.3	11.5	53.8	0	0	11.6	
nub-GAL4/+; UAS-GFP/ UAS-Nelf-E^{RNAi} (32835)	0	0	0	7.9	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Nelf-E^{RNAi} (32835)	6.25	0	65.6	9.4	0	21.2	
nub-GAL4/+; UAS-GFP/ UAS-Nelf-A^{RNAi} (32897)	0	0	0	0	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Nelf-A^{RNAi} (32897)	35	3.3	35	0	0	35***	Increased ectopic venation, wing folding
nub-GAL4/+; UAS-GFP/ UAS-Pho^{RNAi} (39529)	0	0	51.2***	6.7	0	0	Wing inflation defects
nub-GAL4/+; UAS-gro^{RNAi} /UAS-Pho^{RNAi} (39529)	23.5	0	100	5.1	0	17.2	Increase in ectopic venation, wing folding

Table 3: Table showing the quantification of different features of the wing for single and double RNAi knockdown wings of the indicated genotypes.

Six different features of the wing were quantified: total or partial loss of anterior crossvein (ACV), total or partial loss of posterior crossvein (PCV), ectopic venation above L2, notching, wing blistering, and strong vein thickening. Strong vein thickening was considered when the wing phenotype belonged either to Class 3 or Class 4 according to the

classification system shown in Figure 12 A. The last column in the table mentions additional effects observed. Fisher's exact tests were used for pairwise statistical comparisons. For each feature, the single knockdowns were compared to the control (*nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+),* and double knockdowns were compared to the single knockdowns of the genes in the double. The presence of three asterisks by the number indicates a p-value ≤ 0.0001. Cells highlighted in yellow indicate the features of the double knockdown wings that were significantly increased compared to their single knockdowns.

3.1.2 An RNAi-based genetic assay in *Drosophila* eyes to test for interaction with *gro*

The RNAi-based genetic assay in *Drosophila* wings already revealed that Gro cooperates with factors that are key Pol II pausing factors during wing development, supporting a model in which Gro mediates repression by promoting Pol II pausing *in vivo*. To investigate whether Gro acts with these factors in other tissues, we designed a very similar RNAi assay in the *Drosophila* eye.

In this assay, I used the *eyeless-GAL4* driver that expresses GAL4 under the *eyeless* promoter, which is expressed in early and late third instar eye discs, before the stage of furrow initiation (Hazelett et al., 1998). The design of this genetic assay is similar to the one described in the previous section, and it is summarized in Figure 19.

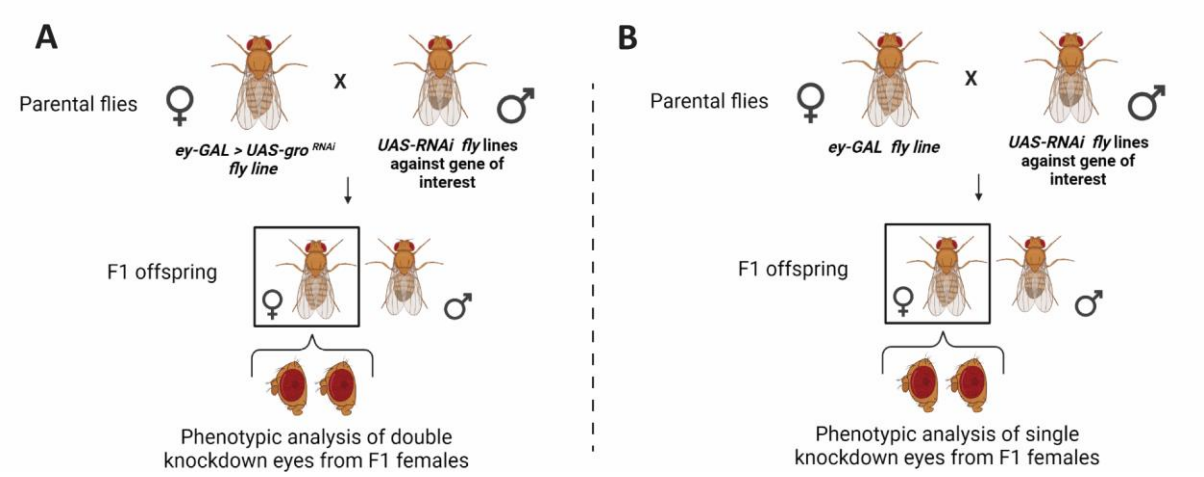


Figure 19: Schematic representation of the genetic assay designed to generate single and double knockdown *Drosophila* eyes.
(A) Females from the *ey-GAL4 > UAS-gro^{RNAi}* line (genotype: *ey-GAL4/CyO; UAS-groRNAi/TM3, Ser*) were crossed over several *UAS-RNAi* lines against the genes of interest to generate double knockdown flies in the developing eye. Subsequently, the F1 offspring flies were collected and eyes from females were mounted for imaging. **(B)** Females from the *ey-GAL4* were crossed over the same *UAS-RNAi* lines to generate single knockdown in the developing eye, subsequently, the F1 offspring were collected and eyes from females were mounted for imaging. Illustration created with images from BioRender.com.

3.1.2.1 Characterization of *gro* downregulation and overexpression in the developing eye

I first performed a preliminary analysis of the eye phenotype caused by the RNAi downregulation of *gro* in the developing eye using two different GAL4 drivers: *GMR-GAL4* and *eyeless-GAL4. eyeless-GAL4* (*ey-GAL4*) drives the expression of UAS transgenes since early larval eye development (Hazelett et al., 1998) whereas *GMR-GAL4* expresses later during eye development. Specifically, GMR-GAL4 expresses exclusively in cells behind the morphogenetic furrow from mid-late larval eye discs (Freeman, 1996).

The expression of one copy *UAS-gro^{RNAi}* transgene under the *GMR-GAL4* driver showed no effect in the adult eye phenotype compared to the control. Since *GMR-GAL4* only drives *gro* downregulation during late eye development, the lack of effects on the adult eye phenotype may be due to a defective *gro* downregulation due to residual Gro protein expressed earlier in eye development.

In contrast, the heterozygous downregulation of *gro* under the *ey-GAL4* driver resulted in weak effects on the adult eye, characterized by small regions of necrosis localized in the posterior side of the eye as well as weak ommatidia disorderliness (black dots in Figure 20 A b). In addition, an increase in the adult eye area was also observed (Figure 20 B). The *gro* knockdown phenotype was very consistent among the eyes scored (at least 50), thus we decided to proceed using the *ey-GAL4* driver.

I next made a fly line in which the downregulation of *gro* is driven in the developing eye by *ey-GAL4* (genotype: *ey-GAL4/CyO; UAS-gro^{RNAi}/TM3, Ser,* refer to *ey-GAL4 > UAS-gro^{RNAi}*). I crossed flies from this stock over the *UAS-RNAi* lines against the genes of interest to yield double knockdown flies (scheme in Figure 19 A). In this assay, single knockdowns were yielded by crossing the *ey-GAL4* over the same RNAi lines (Figure 19 B). I performed an additional control to test if the different total number of *UAS-RNAi* transgenes contained in the single and double knockdowns could affect the severity of the eye phenotype. Thus, I crossed flies from the *ey-GAL4 > UAS-gro^{RNAi}* over flies from a mock RNAi line (*UAS-GFP^{RNAi}*) to yield a progeny of flies in which *gro* was downregulated in the developing eye containing an extra mock UAS site. The adult eye phenotype (Figure 20 A d) was consistent with the shown in

Figure 20 A b, showing black dots in the posterior side of the eye, and an increase of the eye area (Figure 20 B). Thus, we could conclude that the number of UAS sites does not affect significatively to the efficiency of the expression of the RNAi transgenes in this specific case.

The homozygous *gro* knockdown in the developing eye resulted in high lethality, and the few flies that did hatch showed enhanced necrosis and damage in the posterior part of the eye and general ommatidia disorderliness (Figure 20 A e). This lethality is most likely a consequence of *gro* knockdown in the embryonic central nervous system, and in the larval brain, as *eyeless* is also expressed in these tissues (Quiring et al., 1994).

Conversely, overexpression of *gro* resulted in no eye in approximately 37% of the eyes, which indicates that Gro plays a crucial role in the formation of the *Drosophil*a eye. The other 63% of the eyes scored showed a severe reduction in the eye area but no alternations on the ommatidia (Figure 20 A c, and Figure 20 B), which is consistent with previous observations (Choi et al., 2005). As expected *gro* overexpression phenotype was partially rescued when *gro* knockdown and overexpression are simultaneously driven by *ey-GAL4* in the developing wing (Figure 20 A f and B).

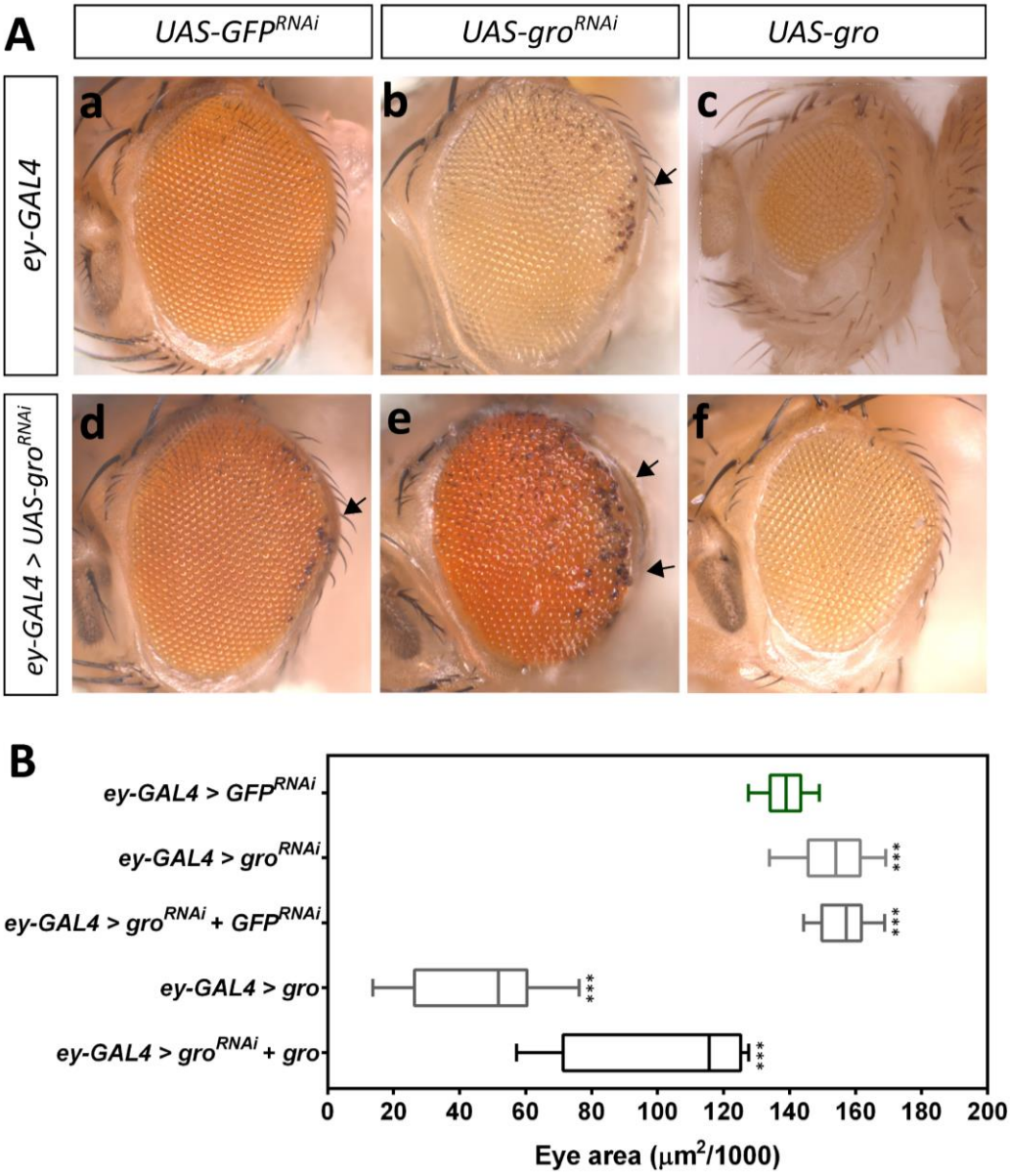


Figure 20: RNAi downregulation of *gro* in the developing eye increases eye area and cell death, whereas *gro* overexpression results in a severe reduction in the eye area.

(A) Representative examples of *Drosophila* adult female eyes raised at 25°C expressing the indicated transgenes. **(a)** Control adult eye phenotype. **(b)** single *gro* RNAi knockdown (genotype: *ey-GAL4/+; +/UAS-gro^{RNAi}*) showing an increase in the eye area compared to the control, weak ommatidia disorderliness, and weak cell death in the posterior part of the eye (arrow). **(c)** overexpression of *gro* showing a severe reduction in the eye size compared to the control **(d)** single *gro* RNAi knockdown (genotype: *ey-GAL4/ UAS-GFP^{RNAi}; +/UAS-gro^{RNAi}*) showing an increase in the eye area compared to the control, weak ommatidia disorderliness, and weak cell death in the posterior part of the eye (arrow). **(e)** double *gro* RNAi knockdown showing damage, moderate ommatidia disorderliness, and severe cell death in the posterior part of the eye (arrows). **(f)** Simultaneous *gro* knockdown and overexpression showed a partial rescue of the eye size. **(B)** Box and whisker plot of eye area (μm^2/1000); the Y-axis indicates the *UAS^{RNAi}* transgenes expressed in each of the phenotypes under the *ey-GAL4* driver. Genotypes from the top to the bottom are *ey-GAL4/UAS-GFP^{RNAi}; +/+*, *ey-GAL4/+; UAS-gro^{RNAi}/+*, *ey-GAL4/UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+*, *ey-GAL4/UAS-gro; +/+*, and *ey-GAL4/UAS-gro; UAS-gro^{RNAi}/+*. T-test was used for pairwise statistical comparison between the eye area measurements of the indicated genotypes and the control (green) (*ey-GAL4 > UAS- GFP^{RNAi}*) ***p<0.0001.

3.1.2.2 Genetic interactions tested between *gro* and key pausing genes in developing eye

After the characterization of the eye phenotype caused by *gro* knockdown, I proceed with the genetic assay. Thus, I crossed flies from the *ey-GAL4* and *ey-GAL4>UAS-gro^{RNAi}* lines over the *UAS-RNAi* lines against the genes of interest to yield single and double knockdowns. However, this genetic assay was not as extensive as the one performed in the wing.

The single downregulation of *Trl*, *Nelf-B*, *Nelf-D*, *Larp7*, and *bin3* (Figure 21 A b-f) during eye development did not show any obvious effect on either the eye morphology or the eye area (only a decrease in the eye area was observed in the single knockdown of *bin3*) (Figure 21 B). The double knockdowns of *gro* and those genes showed high variability in the severity of the effects on the adult eye morphology among eyes with the same genotype. Thus, to simplify the phenotypic analysis we decided to select the most severe eye phenotype observed in each genotype and compare them with the most severe *gro* knockdown phenotype (Figure 21 A g). Specifically, I focused the phenotypic analysis on the posterior part of the adult eye since it was the area most affected in the single *gro* knockdown.

The double knockdown of *gro* and *Trl* (Figure 21 A h) did not show any increase in the severity of the eye phenotype compared to the single knockdown of *gro*, thus those genes do not seem to interact in the developing eye. This result contrasts with the cooperation observed between *gro* and *Trl* during wing development, suggesting that the interaction between those genes may be tissue dependent.

An increase in the severity of the eye phenotype was observed in the double knockdowns of *gro* and *Nelf-B*, *Nelf-D*, *Larp-7*, and *Bin3* (Figure 21 A i-l) compared to the single knockdown phenotype of these genes. The most severe phenotype observed was caused by the simultaneous downregulation of *gro* and *Nelf-D*, which lead to fly lethality, and the few flies that hatched showed strong changes in the eye morphology (Figure 21 A j). The simultaneous knockdowns of *gro* and *Nelf-B* (Figure 21 i), *gro* and *Larp7* (Figure 21 A k), and *gro* and *Bin3* (Figure 21 A l) showed similar phenotypes, characterized by an increase in cell death in the posterior part of the eye.

Taken together these results show that *gro* interacts with the Pol II pausing factors *Nelf, Larp7,* and *bin3* in the developmental pathways that give rise to the *Drosophila* eye. These results agree with the observations in the wing assay, and with the model of Gro promoting promoter-proximal Pol II pausing.

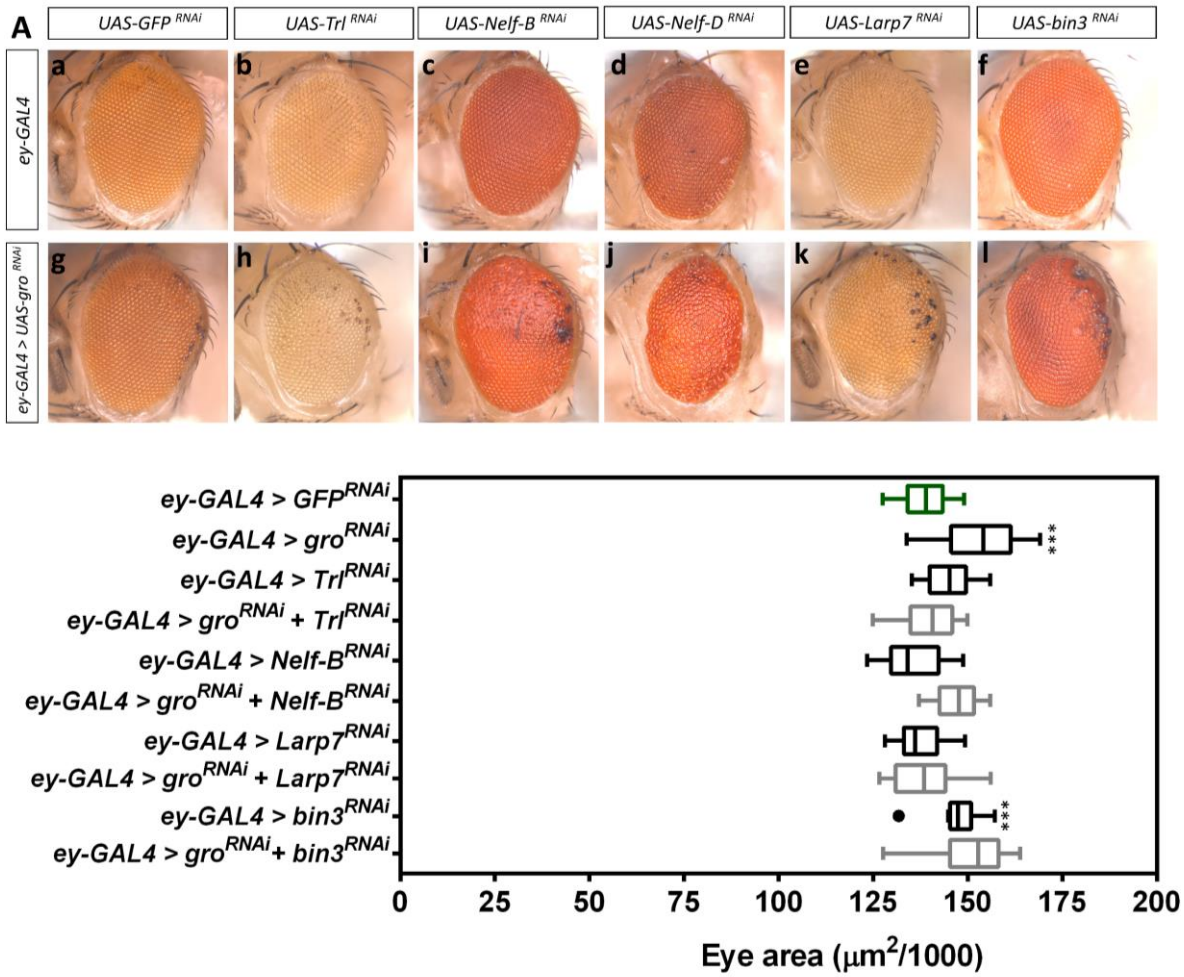


Figure 21: Double RNAi knockdowns of *gro* and the pausing factors *Nelf-B*, *Nelf-D*, *Larp7,* and *bin3* in the developing eye results in increased cell death, and ommatidia disorderliness.

(A) Representative examples of *Drosophila* adult female eyes raised at 25°C expressing the indicated transgenes. **(a)** Control adult eye. **(b-f)** single RNAi knockdown of *Trl, Nelf-B, Nelf-D, Larp7, and Bin3* showing no change in the eye phenotype compared to the control, only weak ommatidia disorderliness was observed in the single downregulation of *Nelf-D*. **(h)** single *gro* RNAi knockdown showing an increase in the eye area compared to the control, weak ommatidia disorderliness, and weak cell death in the posterior part of the eye. **(i)** double RNAi knockdown of *gro* and *Trl* showing weak ommatidia disorderliness, and weak cell death in the posterior part of the eye. **(j)** double RNAi knockdown of *gro* and *Nelf-B* showing severe ommatidia disorderliness, and moderate cell death in the posterior part of the eye. **(k)** double RNAi knockdown of *gro* and *Nelf-D* showing severe ommatidia disorderliness and change in the eye morphology. **(l)** double RNAi knockdown of *gro* and *Larp7* showing weak ommatidia disorderliness and severe cell death in the posterior part of the eye. **(m)** double RNAi knockdown of *gro* and *Bin3* eye phenotype showing weak ommatidia disorderliness and severe cell death in the posterior part of the eye. **(B)** Box and whisker plot of eye area (μm^2/1000); the Y-axis indicates the *UAS^{RNAi}*

transgenes expressed in each of the phenotypes under the *ey-GAL4* driver. Genotypes from the top to the bottom: *ey-GAL4/UAS-GFP^{RNAi}; +/+, ey-GAL4/+; UAS-gro^{RNAi}/+, ey-GAL4/+; UAS-Trl^{RNAi}/+, ey-GAL4/+; UAS-gro^{RNAi}/UAS-Trl^{RNAi}, ey-GAL4/+; +/UAS-Nelf-B^{RNAi}, ey-GAL4/+; UAS-gro^{RNAi}/UAS-Nelf-B^{RNAi}, ey-GAL4/UAS-Larp7^{RNAi}; +/+, ey-GAL4/UAS-Larp7^{RNAi}; UAS-gro^{RNAi}/+, ey-GAL4/+; UAS-bin3^{RNAi}/+, ey-GAL4/+; UAS-gro^{RNAi}/UAS-bin3^{RNAi}*. At least 10 eyes of each genotype were used for eye area measurements. T-test was used for pairwise statistical comparisons. Eye measurements of single knockdowns of the indicated genotypes were compared to the control (green) (*ey-GAL4 > UAS-GFP^{RNAi}*). For the statistical analysis, eye areas of double knockdowns were compared to the single knockdown of the genes in the double. ***p<0.0001, *p<0.01.

Moreover, this genetic assay enabled us to test the possible biological interaction between *gro* and *Cdk9*, which I could not explore in the wing as the RNAi knockdown of *Cdk9* in the developing wing is lethal (see section 3.1.1.4 of the results).

The single downregulation of *Cdk9* in the developing eye resulted in a severe reduction of the eye size (Figure 22 A c and B). Interestingly, when *Cdk9* was simultaneously downregulated with *gro* there was a partial rescue of the eye size, and no other obvious effect (Figure 22 f and B). Thus, Gro seems to partially counter the drastic effect of Cdk9 on growth, which suggests that Gro and Cdk9 might have opposite effects on the pathways that direct the growth of the adult eye.

The single downregulation of *Myc* in the developing eye also resulted in a significant reduction of the eye size (Figure 22 A b and B), which was clearly rescued in the double knockdown of *gro* and *Myc* (Figure 22 A e and B). Thus, *gro* seems to have a dominant epistatic effect over *Myc* on growth in the developing eye. However, the double knockdown also showed an increase in cell death in the posterior part of the eye (arrows in Figure 22 A e) compared to the single downregulation of *gro*, which suggests that both genes also cooperate.

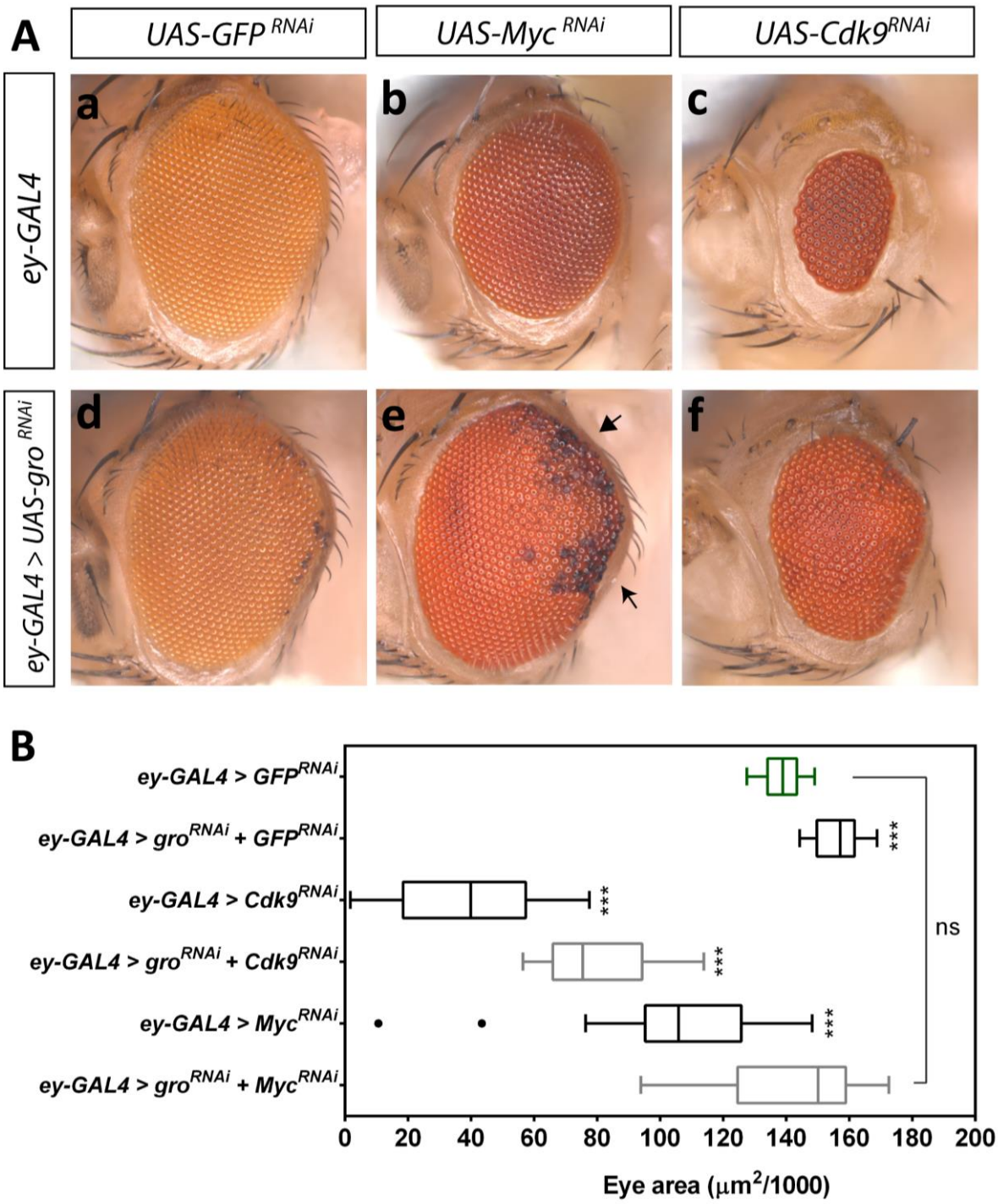


Figure 22: Single RNAi downregulation of *Cdk9* and *Myc* in the developing eye results in a severe reduction of the eye area, which is partially rescued by the simultaneous downregulation of these genes with *gro*.

(A) Representative examples of *Drosophila* adult female eyes raised at 25°C expressing the indicated transgenes. **(a)** Control adult eye phenotype. **(b)** single *Myc* RNAi knockdown showing a reduction in the eye area compared to the control and weak ommatidia disorderliness. **(c)** *Cdk9* RNAi knockdown showing a severe reduction in the eye size compared to the control and moderate ommatidia disorderliness. **(d)** single *gro* RNAi knockdown showing an increase in the eye area compared to the control, weak ommatidia disorderliness, and weak cell death in the posterior part of the eye. **(e)** *gro* and *Myc* double RNAi knockdown showing moderate ommatidia

disorderliness, and severe cell death in the posterior part of the eye. **(f)** *gro* and *Cdk9* double RNAi knockdown showing a reduction in the eye size, and moderate ommatidia disorderliness. **(B)** Box and whisker plot of eye area (µm^2/1000); the Y-axis indicates the *UAS^{RNAi}* transgenes expressed in each of the phenotypes under the *ey-GAL4* driver. Full genotypes from the top to the bottom: *ey-GAL4/UAS-GFP^{RNAi}; +/+, ey-GAL4/UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+, ey-GAL4/+; UAS-Cdk9^{RNAi}/+, ey-GAL4/+; UAS-gro^{RNAi}/ UAS-Cdk9^{RNAi}, ey-GAL4/+; UAS-Myc^{RNAi}/+, ey-GAL4/+; UAS-gro^{RNAi}/UAS-Myc^{RNAi}*. At least 10 eyes of each genotype were used for the eye are measurements. T-test was used for pairwise statistical comparisons. Eye measurements of single knockdowns of the indicated genotypes were compared to the control (green) (*ey-GAL4 > UAS-GFP^{RNAi}*). For the statistical analysis, eye areas of double knockdowns were compared to the single knockdown of the genes in the double. ***p<0.0001, ns: not significant.

3.1.3 Discussion

The work presented in this chapter revealed a strong biological interaction between *gro* and *Trl* (the gene encoding GAF) in the developing wing. However, no obvious interaction between these factors was detected in a similar genetic assay in the developing eye, indicating that Gro and GAF cooperation might depend on the biological context. Importantly, these genetic assays showed a strong biological interaction between *gro* and the key pausing factor *Nelf* in both the developing wing and eye. In addition, genetic interactions were detected between *gro* and *bin3* and *gro* and *Larp7*, two of the components of the 7sk snRNP in the proliferating eye and wing. Nelf, bin3, and Larp7 are key Pol II pausing factors that are highly conserved across evolution, in contrast to GAF, which is involved in many other functions other than Pol II pausing in *Drosophila*. In consequence, we can speculate that Gro and GAF cooperation is dependent on the developmental context, whereas Nelf, bin3, and Larp7 may be necessary for Gro-mediated repression in all tissues.

On the other side, the genetic assay in the wing showed cooperation between *gro* and *Myc*, which was unexpected since Gro is known to antagonize the positive function of Myc on gene expression at Gro-Myc target genes (Orian et al., 2007). Cooperation between *gro* and *Myc* was also observed during ommatidia formation during eye development. However, these factors also appear to antagonize each other to regulate growth in the proliferating eye, which is in agreement with Orian and colleagues' observations (Orian et al., 2007). Given the pleiotropic roles of Myc in growth, and proliferation [reviewed in (Baluapuri et al., 2020)], it would be reasonable to expect that Gro and Myc may interact in different ways to regulate different processes.

In addition, the eye genetic assay indicated that Gro and Cdk9 antagonize each other during eye growth. Cdk9 is a kinase that forms part of the P-TEFb complex, which is essential for the release of paused Pol II. This result agrees with a model in which Gro promotes pausing by blocking the action of factors that promote the release of paused Pol II, as is the case of P-TEFb.

Altogether, these results indicate that Gro cooperates *in vivo* with factors involved in Pol II pausing to regulate gene expression, which is in agreement with the model of Gro acting by promoting promoter-proximal pausing.

3.2 Bioinformatic analysis of Gro binding profile in three different *Drosophila* cell lines and co-localization with key pausing factors.

Section 3.1 of the results discovered that *gro* genetically interacts with genes encoding key pausing factors, supporting the model in which Gro mediates transcriptional repression by promoting Pol II pausing. To gain an understanding of the molecular basis underlying these biological interactions, I performed a genome-wide analysis of recruitment of Gro and factors involved in Pol II pausing. Published sequencing data from chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments were used. To this aim, the ChIP-seq method was first described in 2007 on the rise of next-generation sequencing technologies, and it allows the identification of DNA binding sites of specific proteins (Barski et al., 2007; Mikkelsen et al., 2007). In a typical ChIP-seq experiment, first, the DNA-protein complexes are cross-linked, and then the DNA in the samples are fragmented. Subsequently, the DNA-protein complexes of interest are immunoprecipitated by using specific antibodies, and finally, the DNA from the complexes is extracted and sequenced. ChIP-seq presents several advantages compared to previous methods (such as ChIP-ChIP) including single base-pair resolution, improved coverage, fewer artefacts and larger dynamic range [reviewed in (Park, 2009)]. This technique has been used to capture DNA targets for TFs, histone modifications, and histone variants, which have played essential roles in the understanding of transcriptional regulation in different organisms.

In this context, Kaul and colleagues used the ChIP-seq technique to unveil several features of Gro binding (see section 1.2.3 of the Introduction), including the high overlap between Gro and GAF binding sites in *Drosophila* Kc167 cells (Kaul et al., 2014). In this thesis, I further explored the colocalization between the ChIP-seq binding sites of these proteins by comparing their overlap in Kc167 and BG3 *Drosophila* cells.

In addition, I analysed the co-localization of peaks of Gro binding and peaks of binding of key factors involved in Pol II pausing in Kc167 cells. Several ChIP-seq datasets for Gro, GAF, Nelf-E, Cdk9, Myc and Pol II isoforms that were available in the databases (see Table 8 of the Appendix) were used for this purpose. For this analysis, I first determined the overlap

between the binding sites of Gro with peaks of GAF, Nelf-E, Cdk9, and Myc to examine whether Gro colocalizes with them individually. Subsequently, to obtain a general view of the colocalization pattern between Gro and the tested factors, their binding sites were analysed collectively.

Prior to the colocalization studies, and to gain understanding into Gro genome-wide distribution and binding, I performed a comparison of Gro ChIP-seq peaks in three different *Drosophila* cell lines: S2R+, Kc167, and BG3. The Kc167 and the S2R+ are both haemocyte-like embryonic cell lines, whereas the BG3 derives from the larval Central Nervous System (CNS) (Cherbas et al., 2011; De Obaldia et al., 2013). Therefore, comparison among Gro peaks in those cell types might reveal differences in Gro binding between the embryonic and the larval neural genome. The Gro ChIP-seq experiments in these cell lines were done by Aamna Kaul, a previous member of my lab, and the bioinformatics analysis of this data in Kc167 and S2R+ has been already published (Kaul et al., 2014). However, the Gro ChIP-seq datasets in BG3 cells had not been previously analysed.

3.2.1 Bioinformatic analysis of Gro ChIP-seq datasets in BG3 cells

3.2.1.1 Gro peaks frequently span less than 1 Kb and are located at open chromatin regions in BG3 cells

First, to obtain a set of high confidence Gro-bound regions in BG3 cells, I processed the sequencing data from two biological replicates of Gro ChIP-seq experiments performed by Aamna Kaul. To do so, the software Bowtie2 (Langmead and Salzberg, 2012) was used to align the sequencing reads to the *Drosophila* reference genome. Then, a set of Gro binding sites for each replicate was identified by the MACS software (Zhang et al., 2008). Lastly, a set of 1252 high confidence Gro peaks (typically with an FDR of less than 10%) was obtained by selecting the peaks found in both replicates (see Materials and Methods). The set of high confidence binding sites in BG3 cells was then analysed. Several features were examined, including the breadth of the peaks, the gene ontology, the sequence motives, and the chromatin accessibility.

Previous studies of Gro ChIP-seq data in the Kc167, and S2R+ cell types, as well as in embryos reported that Gro binds to the genome in discrete peaks (Chambers et al., 2017; Kaul et al., 2014) rather than spreading along broad regions of chromatin, as it was traditionally thought. That analysis in *Drosophila* cells also revealed that Gro was associated with DNase I hypersensitivity sites (DHSs) in Kc167 cells (Kaul et al., 2014). Thus, I analysed whether these characteristics of Gro binding to the genome were also observed in BG3 cells.

The analysis of the breadth of Gro peaks using showed that in BG3 cells Gro binds to the genome in discrete peaks with a mean width of 278 bp, and a median width of 224 bp. The histogram in Figure 23 A shows the distribution of Gro peaks according to their width, it is noteworthy that only 7 of the 1252 total peaks spanned more than 1 Kb (not plotted in the histogram).

To examine the chromatin environment of Gro bound regions, we made use of the DHSs in BG3 cells from modEncode. DHSs are DNA regions that exhibit hypersensitivity to cleavage by DNase I endonucleases and indicate the presence of open chromatin regions. DNase-seq studies, in which DNase I digestion is combined with sequencing have been widely used to unveil the chromatin accessibility in different cell types or tissues [reviewed by (Madrigal and Krajewski, 2012)]. The visualization of the DNase-seq signals at Gro binding sites showed a clear enrichment of DHSs at the genomic locations where Gro binds (Figure 23 B), which indicates that Gro is frequently present in open chromatin regions.

These observations are in agreement with previous studies that analysed Gro peaks in Kc167 and S2R+ cells (Kaul et al., 2014), and in embryos (Chambers et al., 2017) and indicate that are general features of Gro binding.

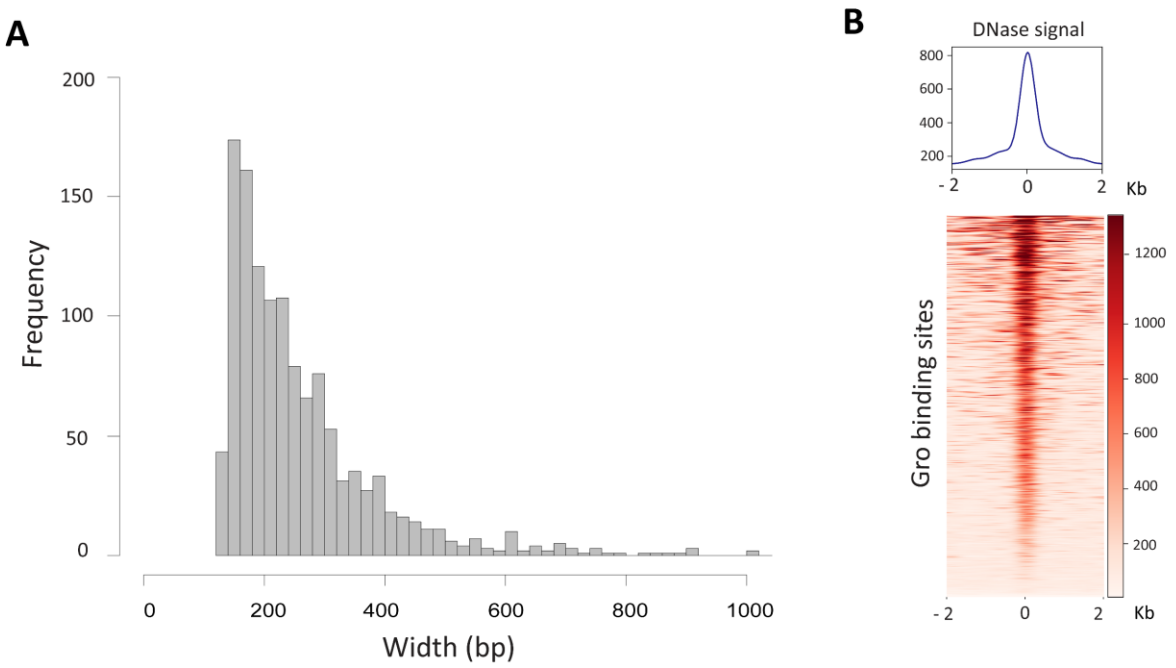


Figure 23: Gro is frequently recruited to the genome in discrete peaks at open chromatin regions in *Drosophila* BG3 cells.

(A) Histogram showing the frequency of the width peaks of Gro binding sites (in base pairs). **(B)** Average profile plot and heatmap showing the DNase-seq signal (retrieved from modENCODE) at Gro binding sites in BG3 cells. The average plot and the heatmap are centred at the centre of Gro peaks and extended 2 Kb upstream and downstream of that reference point.

3.2.1.2 Gro binding sites are enriched in GAF motifs, among others, in BG3 cells

Next, I determined the sequence-specific TF binding sites associated with the set of high confidence Gro peaks in BG3 cells. The CentriMo tool from the MEME suite (Bailey and Machanick, 2012) was used to identify known motifs from the databases that show significant enrichment at Gro bound regions. This analysis resulted in no single consensus motif at Gro binding sites as expected, since Gro is a co-factor recruited by many different DNA sequence-specific TFs. Instead, it displayed binding motifs for several *Drosophila* TFs, among which the five with lowest p-values were: GAF, crooked legs (crol), odd paired (opa), longitudinal lacking (lola), and lethal (3)neo38 [l(3)neo38] (Figure 24 A). The binding motifs of the known Gro interactors Eyeless and Brinker were also detected in the list (with p-values of 7.6×10^{-6} and 3.5×10^{-2}, respectively). Although neither of the motifs detected with the lowest p-values [GAF,

crol, opa, lola, and lethal (3)neo8] belonged to TFs that are known Gro interactors, the binding motifs of GAF and l(3)neo38 also ranked among the top with the lowest p-value in CentriMo analysis of Gro bound regions in S2R+ and Kc167 cells (Kaul et al., 2014). This indicates that Gro might work directly or indirectly with the TFs associated with GAF and l(3)neo38 motifs in both the embryonic and the CNS-derived cell types.

To explore if the TFs associated with the motifs found at Gro bound regions in BG3 cells are specific to the larval CNS, I used the FlyAtlas 2 (www.flyatlas2.org) to check their expression in different larval tissues. The FlyAtlas 2 database contains the expression data of *Drosophila* genes in different adult and larval tissues based on RNA-seq data (Leader et al., 2018). This analysis showed that the TFs crol, lola, and l(3)neo38 are highly enriched in the larvae CNS compared to other larval tissues (Figure 24 B). In addition, opa was found to be exclusively expressed in the CNS during larval development (Figure 24 B). GAF is also enriched in the larval CNS; however, it is similar or even more enriched in other larval tissues, thus this TFs does not seem to be as specific of the larval CNS as the others (Figure 24 B). Interestingly, the motifs for crol, lola and opa TFs did not rank among the most significant motifs in CentriMo analysis of Gro bound regions in the embryo-derived cell types Kc167 and S2R+ (Kaul et al., 2014). Taken together these results suggest that the TFs crol, lola and opa might recruit Gro to specific tissues. It is noteworthy that crol, lola, l(3)neo38, and opa motifs have been found at BG3-specific TAD borders in a recent study (Chathoth and Zabet, 2019).

In addition, a functional enrichment analysis was performed on the annotated genes that were nearest to Gro binding sites in BG3 cells using the DAVID gene ontology tool (Huang et al., 2007). Figure 24 C shows that the over-represented terms with the most significant p-values at Gro-bound genes were: “transcription”, “wing morphogenesis”, “open tracheal system development”, and “dendrite development”. Those biological terms are consistent with the biological role of Gro as a transcriptional co-repressor in developmental pathways.

In summary, the general features of Gro binding sites in BG3 cells are very similar to the previously published observations of Gro peaks in the embryo-derived cells Kc167 and S2R+ (Kaul et al., 2014). Specifically, peaks of Gro recruitment share the following features among the three cell types: the peaks typically span less than 1 Kb, overlap with open chromatin regions, and are enriched for the DNA binding sequence motifs of GAF and l(3)neo38. Hence, these are common features of Gro binding in several cell types.

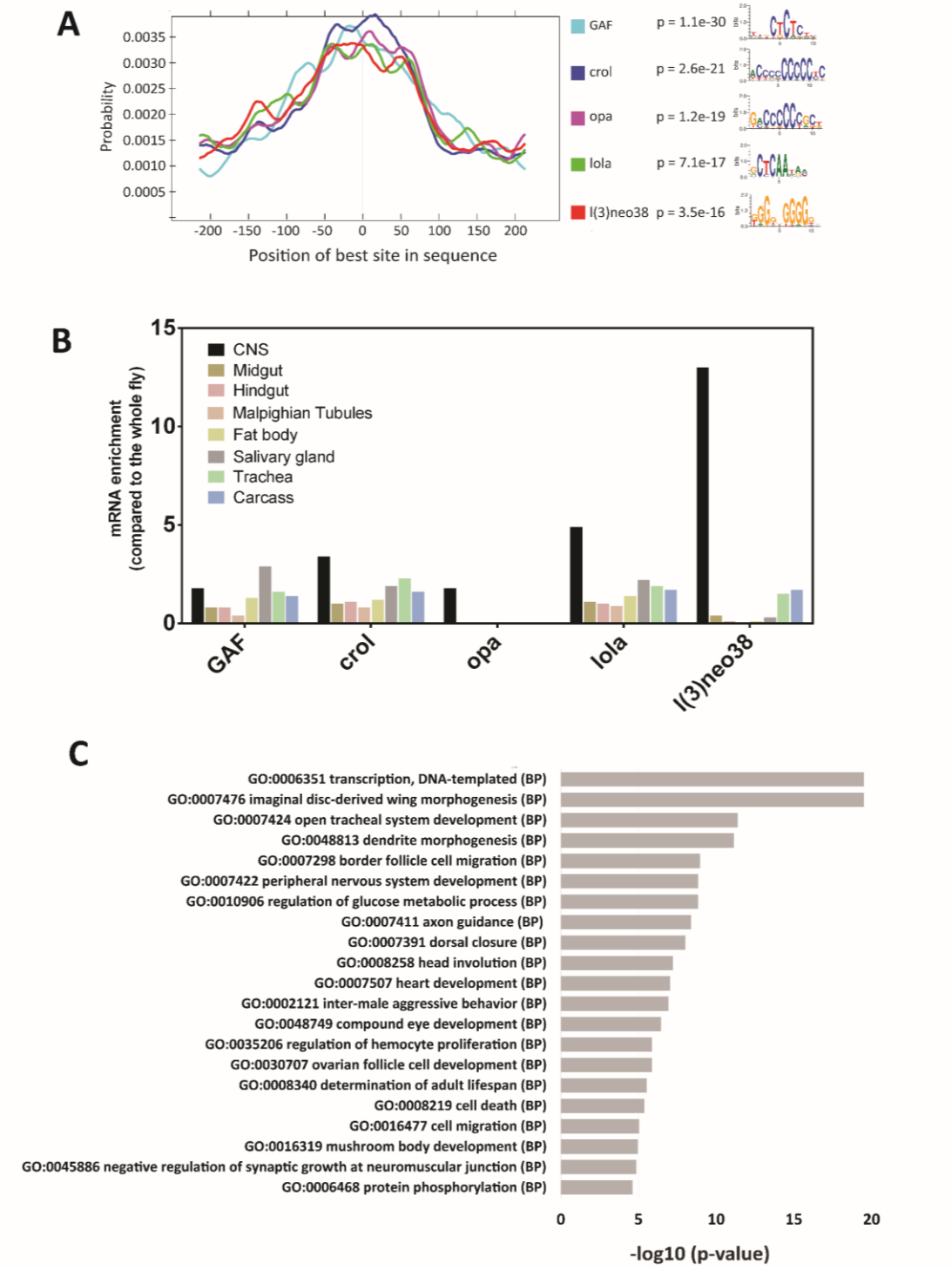


Figure 24: Functional characterization of Gro binding in BG3 cells.

(A) Graph showing the top-scoring five results from the CentriMo analysis at Gro binding sites in BG3 cells: GAF,crol, opa, lola, and l(3)neo38. Each curve represents the probability of the binding location for the named factors at each position of Gro ChIP-seq peak regions (500 bp). The legend shows the motifs and their central enrichment p-value **(B)** Data from Flyatlas 2 (Leader et al., 2018) showing a high enrichment of crol, opa, lola, and l(3)neo38 in the larval central nervous system (CNS). Expression is shown as a ratio of mRNA enrichment for each gene in each tissue to the average mRNA enrichment for all the tissues. **(C)** DAVID Gene Ontology analysis of genes associated with Gro binding sites in BG3 cells. Terms plotted were selected by taking the most significant term in a cluster (p-value < 10^{-5}) and the most significant unclustered terms.

3.2.2 Comparison of Gro binding profiles in Kc167, S2R+ and BG3 *Drosophila* cells

3.2.2.1 Genomic annotation and chromatin environment of Gro binding sites in Kc167, S2R+ and BG3 *Drosophila* cells.

The set of Gro ChIP-seq high confidence peaks generated in BG3 cells was subsequently used to explore the genomic distribution of Gro binding sites and compare it with the Gro genomic distribution in Kc167 and S2R+ cells. For this analysis, the published sets of Gro ChIP-seq high confidence peaks in Kc167 and the S2R+ cells were retrieved from the databases (Kaul et al., 2014). The ChIPseeker software was used to annotate and visualize the genomic regions where Gro peaks are located in the three cell lines (Yu et al., 2015). This software annotates TFs peaks into the following genomic regions: Promoter (defined -250 to 250 bp from the TSSs for the studies shown in this thesis), 5' UTR, 3' UTR, Exon, Intron, Downstream (of the gene end) and Intergenic. This analysis revealed a similar pattern of Gro recruitment in the three cell lines (Figure 25 A), with most of Gro peaks located either at promoters or at introns. However, subtle differences among cell lines were observed; whereas in Kc167 cells approximately 50% of Gro peaks mapped at promoters, in BG3 and S2R+ this percentage was around 30-35%.

Next, the overlap among Gro peaks in the three cell types was determined using the Intervene software (Khan and Mathelier, 2017). This analysis showed a larger number of overlaps between Gro peaks in Kc167 and S2R+ (344 peaks) than between either of these and the BG3 dataset (136 and 113 peaks, respectively) (Figure 25 B). This was expected since both the Kc167 and the S2R+ are cell lines derived from the embryo, whereas the BG3 is derived from the larval CNS. Importantly, there were 175 Gro peaks common to the three cell lines that included peaks associated with the *bona fide* Gro targets *E(spl)mβ-HLH*, *E(spl)m3-HLH*.

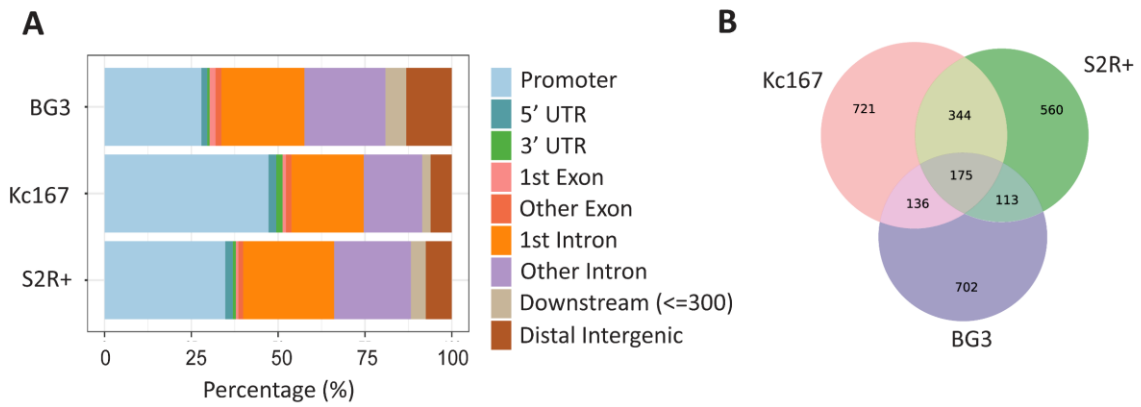


Figure 25: Genomic annotation and overlap among Gro peaks in the Kc167, S2R+ and BG3 cell types. **(A)** Bar graph showing the percentage of Gro peaks within each genomic annotation in the Kc167, S2R+ and BG3 cell lines. Promoters were defined in a region of ±250 around TSSs. **(B)** Venn diagram visualizing the overlap among Gro peaks in Kc167, S2R+ and BG3 cell lines. A minimum and reciprocal fraction of overlap of 5% among peaks was required to consider it an overlap.

I next explored the chromatin environment of Gro binding sites in Kc167 and BG3 cells. Using the chromatin states maps of BG3 and Kc167 cells that were generated by Sarah Bray's lab. They divided the genome of these cell lines into 11 chromatin states that were associated with different histone modifications and DHSs (Figure 26 A) (Skalska et al., 2015).

The analysis of the chromatin states at Gro binding sites showed almost all peaks located at active chromatin regions in the two cell types (Figure 26 B and C). Specifically, most of those peaks were located within "enhancer" chromatin (also known as red chromatin; 80% in BG3 and 64% in Kc167), and in less proportion within the "active TSS" chromatin (orange chromatin; 9% in BG3 and 20% in Kc167) (Figure 26 B and C). Both the enhancer and the active TSS chromatin states are marked by the chromatin modifications H3K56ac and H3K27ac, as well as DHS. However, the enhancer chromatin is enriched for H3K4me1, whereas the active TSS is enriched for H3K4me3 (Figure 26 A) (Skalska et al., 2015). Most of the remainder Gro peaks were located at the competent chromatin (green; 5% in BG3 and 10% in Kc167) and active intron (purple; 2% in BG3 and 3% in Kc167). A small number of peaks mapped within heterochromatin (light grey; 2% in BG3 and 1% in Kc167), "elongation" (dark yellow; 1% in BG3), "basal" (black; 1% in BG3), and "TSS proximal" (light yellow; 1%) chromatin states. As an example, the chromatin states and Gro binding profiles in BG3 and Kc167 cells were displayed at the *bona fide* Gro target genes *E(spl)mβ-HLH* and *E(spl)m3-HLH* (Figure 26 D)

using the IGV software (Thorvaldsdottir et al., 2013). At these loci, Gro binds within regions of red chromatin that are surrounded by competent (green) and polycomb (blue) chromatin in both cell lines (Figure 26 D).

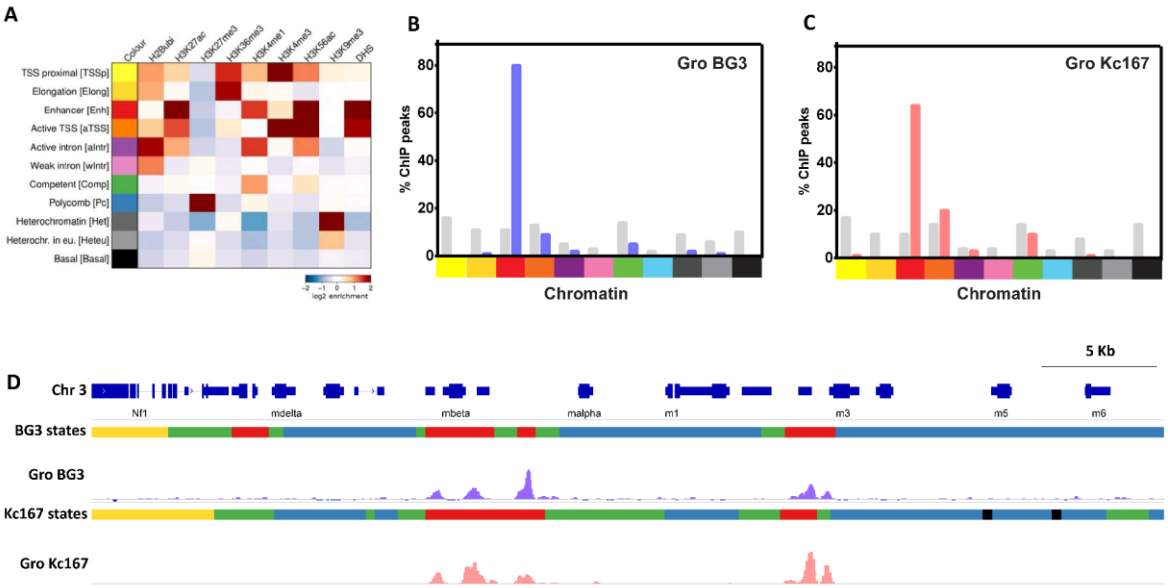


Figure 26: Chromatin states at Gro bound regions in Kc167 and BG3 cells.
(A) Heatmap modified from Skalska et al. 2015 that shows the summary of the 11 chromatin signatures and their enrichment in histone modifications. The heatmap ranges from brown (highly enriched) to blue (depleted). **(B-C)** Distribution of Gro binding sites according to their chromatin type in BG3 (B) and Kc167 (C). Chromatin states are coloured according to the scheme in (A). Gray bars indicate the proportion of each chromatin signature in the total *Drosophila* genome for BG3 (B), and Kc167 (C). **(D)** Gro binding profiles in BG3 and Kc167 cells along with the *E(spl)-C* locus aligned with the chromatin map colour coded as in (A).

In summary, this data shows the enrichment of Gro peaks within the enhancer and active TSS chromatin states in both cell types, adding increasing evidence of Gro present in open chromatin regions. Furthermore, Gro binding sites showed a very similar association with chromatin states to the binding profile of Su(H) (Skalska et al., 2015). This similarity was expected considering that Gro is known to interact with Su(H) via the adaptor protein Hairless to maintain repression of Notch target genes in the absence of Notch activation (Nagel et al., 2005).

3.2.3 Comparison of binding profiles of Gro and pausing factors

3.2.3.1 Gro and GAF binding sites co-localize in Kc167 and BG3 cells

As explained in the Introduction, Kaul and colleagues showed that 82% of Gro peaks at the TSSs and 45% not at the TSSs colocalize with GAF in Kc167 cells (Figure 27 A) (Kaul et al., 2014). To confirm the genome-wide co-localization of Gro and GAF binding sites in Kc167 cells I tested the overlap between those factors using a more recently published set of GAF ChIP-seq binding sites (7535 peaks) (Cubeñas-Potts et al., 2017), and a new generation of bioinformatics tools (see Materials and Methods). First Gro peaks were annotated using ChIPseeker (Yu et al., 2015) and divided into two groups: Gro peaks proximal to TSSs regions (or promoter regions; defined as ± 250 from the TSSs), and not at the TSSs regions (includes peaks at the remaining genomic locations). Subsequently, BedTools (Quinlan, 2014) was used to intersect and visualize the overlap between Gro peaks at and not at the TSSs regions and GAF binding sites. The result was then plotted in a Venn diagram, which shows that 89% of Gro peaks at the TSSs and 69% not at the TSSs overlap with GAF (Figure 27 B). The observed percentages are consistent with the overlap between those factors observed by Kaul and colleagues (Figure 27 A) (Kaul et al., 2014). Additionally, the high similarity among these results validated the bioinformatics workflow, which was subsequently used to determine the overlap of sets of peaks in this chapter.

To extend the co-localization study between Gro and GAF I explored the overlap between those factors in BG3 cells, using a GAF ChIP-ChIP dataset in BG3 cells from modENCODE. This analysis revealed that 49% of Gro peaks at the TSSs and 56% not at the TSSs overlap with GAF in BG3 cells (Figure 27 C). The high overlap between Gro and GAF observed in both Kc167 and BG3 cells suggest that the genome-wide co-localization of Gro and GAF is not unique to a specific cell type. This is also consistent with the observation of GAGA binding motifs located at Gro-bound regions in all three cell types (Kaul et al., 2014) (see also section 3.2.1.2 of the results).

Although there is a significant number of Gro peaks at TSSs regions that overlap with the binding sites of GAF in both Kc167 and BG3 cells, differences between the percentages of overlap were observed between the two cell lines. Whereas 89% of the Gro peaks at TSSs regions overlap with GAF in Kc167 cells, this percentage is 49% in BG3 cells (Figure 27 B and C). To examine further the Gro and GAF overlapping peaks, the percentage of those peaks within each genomic region was plotted for Kc167 (Figure 27 D) and BG3 (Figure 27 E) cells. In Kc167 cells, 54% of the overlapping peaks are located at promoter regions, and 39% of the remaining peaks are at intronic and distal intergenic regions (Figure 27 D). In contrast, in BG3, most of the overlapping peaks (60%) are distributed within intronic and distal intergenic regions, and 31 % at promoters (Figure 27 E). This is consistent with the genomic distribution of Gro peaks, as there are more Gro peaks in distal intergenic regions and introns in BG3 than in Kc167 (see section 3.2.2.1 of the results).

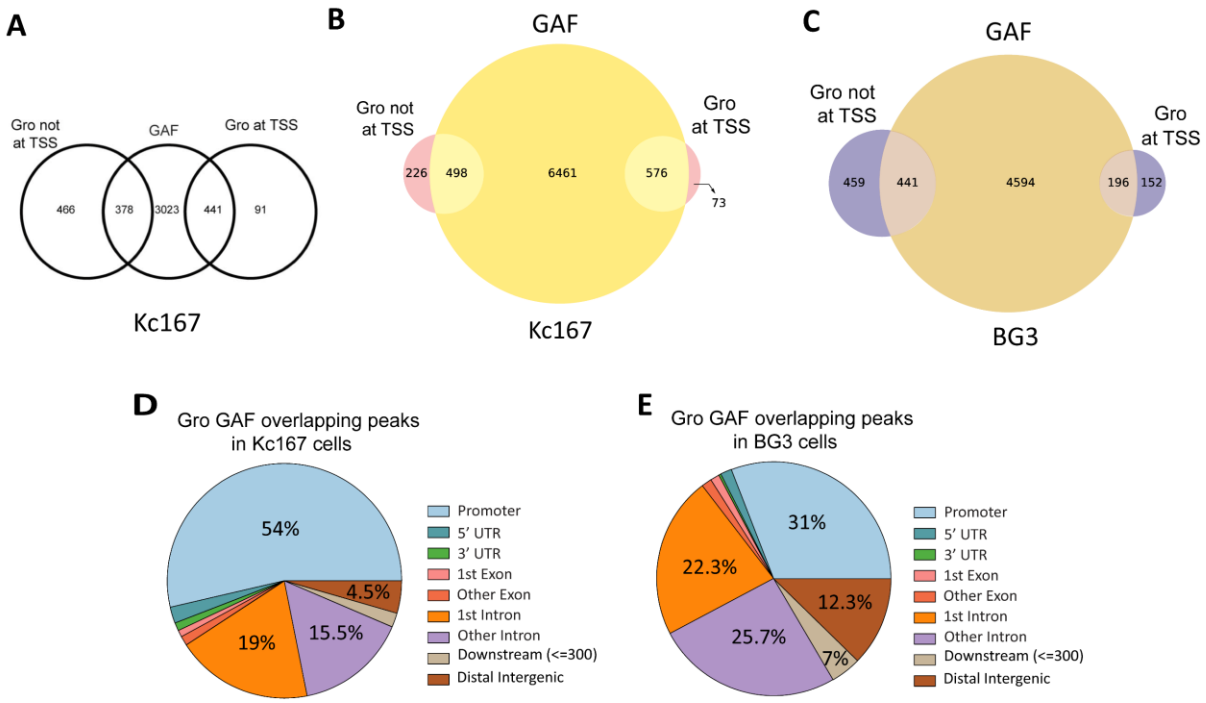


Figure 27: Gro binding sites overlap with peaks of GAF in Kc167 and BG3 cells.

(A) Image adapted from Kaul et al. 2014 showing the overlap between Gro and GAGA Factor (GAF) binding in Kc167 cells. **(B)** Venn diagram illustrating the overlap between Gro peaks at and not at the TSSs and GAF binding sites in Kc167 cells, and in **(C)** in BG3 cells. **(D)** Pie chart showing the percentage of Gro, and GAF overlapping peaks within each genomic location in Kc167 cells, and **(E)** in BG3 cells. Promoters were defined in a region of ±250 around TSSs. A minimum and reciprocal fraction of 5% of overlap among peaks was required to consider overlap.

3.2.3.2 Gro peaks co-localize with both GAF and Nelf at TSSs in Kc167 cells

In the previous section, it was shown the overlap between Gro and GAF binding through the genome in *Drosophila* cells, however, GAF regulates gene expression by multiple mechanisms besides Pol II pausing. To explore if the genome-wide co-localization between Gro and GAF is directly related to Pol II pausing, we checked if Gro and GAF binding sites overlap with the key pausing factor Nelf in Kc167 cells. As described in section 1.3 of the Introduction, Nelf is a protein complex that is key in the establishment of Pol II pausing, and it is known to recruit GAF (Li and Gilmour, 2013).

A set of high confidence Nelf-E binding sites (8865) was generated by selecting the common peaks between two sets of ChIP-seq Nelf-E peaks from the data generated by (Nazer et al., 2018). The distribution of Nelf-E peaks across the genome was visualized to obtain a general view of their genomic localisation, along with GAF and Gro. The bar plot in Figure 28 shows a very similar distribution of Gro and GAF binding sites, with most of the peaks located either at promoters or introns; a different pattern is observed in Nelf-E with peaks mostly present at promoters. This was expected since the role of Nelf-E in the establishment of the paused Pol II complex at TSSs is well-described, whereas GAF has a wide range of roles in transcriptional regulation.

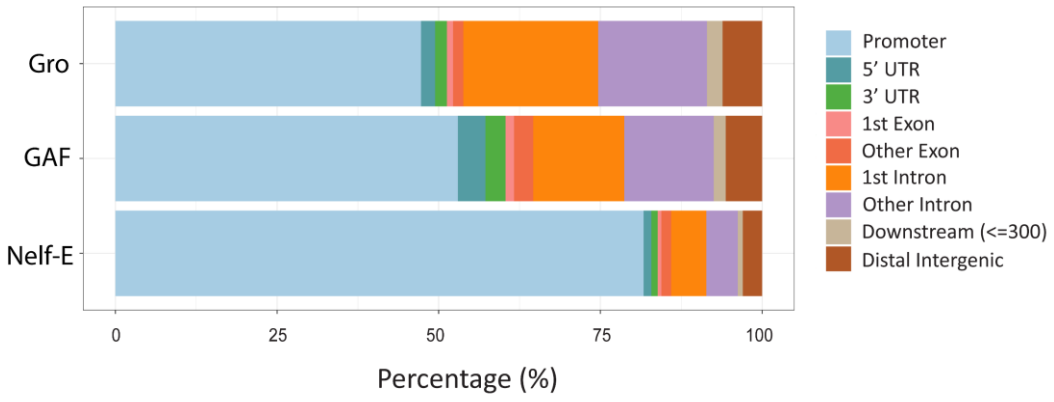


Figure 28: Genomic annotation of Gro, GAF, and Nelf-E binding in Kc167 cells.
Bar graph illustrating the percentage of peaks within each genomic location for Gro, GAF, and Nelf-E in Kc167 cells. Promoters were defined in a region of ±250 around TSSs.

Figure 29 A shows that 91% of Gro peaks at the TSSs and 41% not at the TSSs overlap with Nelf-E binding sites. Thus, the percentage of overlap at TSS between Gro and Nelf-E is very similar to the observed between Gro and GAF (89%), and it is consistent with the model of Gro promoting promoter-proximal pausing.

I next examined the overlap among the total ChIP-seq peaks of Gro, GAF, and Nelf-E. The results were plotted in a Venn diagram (Figure 29 B), which shows that 58% of Gro peaks overlap with both GAF and Nelf-E, 20% only with GAF, 6.5% only with Nelf-E, and 15% with neither GAF nor Nelf-E. Furthermore, the genomic annotation of the Gro, GAF, and Nelf-E overlapping binding sites showed that 67.5% of those peaks were located at promoter regions of genes (Figure 29 C). Taken together, these observations support the model in which Gro mediates promoter-proximal pausing by interacting with GAF and Nelf-E (either or indirectly) at the TSSs regions of genes.

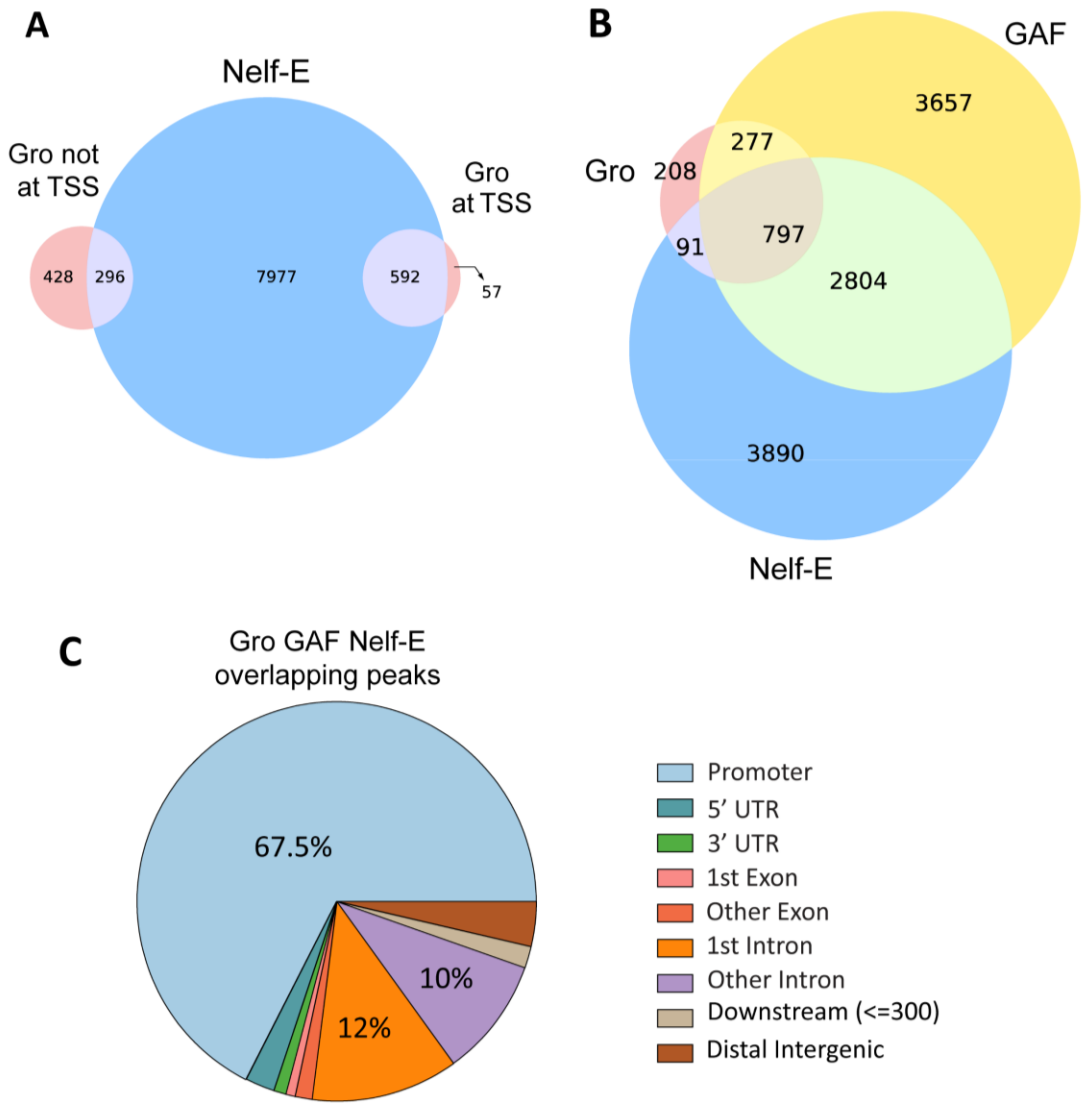


Figure 29: Gro binding sites overlap with peaks of GAF and Nelf-E in Kc167 cells.
(A) Venn diagram showing the overlap between Gro peaks at and not at TSSs regions and Nelf-E binding sites in Kc167 cells **(B)** Venn diagram showing the overlap among the total binding sites of Gro, GAF, and NELF-E. **(C)** Pie chart showing the percentage of Gro, GAF, and Nelf-E overlapping peaks within each genomic location in Kc167 cells. A minimum and a reciprocal fraction of 5% of overlap among peaks was required to consider an overlap. Promoters were defined in a region of ±250 around TSSs.

3.2.3.3 Cdk9 and Myc are not excluded from Gro peaks at TSSs regions in Kc167 cells

The genome-wide co-localization between Gro, GAF and Nelf at TSSs regions, together with the genetic interactions between those factors described in section 3.1 of the results strongly supports the model in which Gro promotes Pol II pausing. However, the precise molecular

mechanism by which Gro mediates this process is unknown. One possible mechanism is that the presence of Gro prevents the recruitment or activity of factors that release the paused Pol II. The recruitment of the P-TEFb complex (CycT and Cdk9) is required for the release of paused Pol II and the stimulation of productive elongation. P-TEFb is recruited to regulatory regions via several factors that include co-activators, chromatin-bound activators, RNA-bound activators, and DNA-bound activators (see section 1.3 of the Introduction). One of the DNA-binding TFs that is known to interact and recruit P-TEFb in mammals is Myc [reviewed (Baluapuri et al 2020, Rahl and Young 2014)]. In *Drosophila*, Gro has been shown to physically interact with Myc, and antagonize Myc's activating function on gene expression at genes that recruit both Myc and Gro (Orian et al., 2007). Therefore, Gro could mediate Pol II pausing by blocking the positive effects that P-TEFb and Myc have during the release of poised Pol II. To gain an understanding of this process, it was examined if Cdk9 and Myc are excluded or present at bound genomic locations bound by Gro in Kc167 cells.

Two sets of published Cdk9 ChIP-seq binding sites from Kc167 cells (Nazer et al., 2018) were used to generate a high confidence set of Cdk9 peaks (6017) that was subsequently analysed for overlap with Gro peaks. It is worth noting that the published datasets for Cdk9 binding sites contained fewer peaks than expected of such a ubiquitous factor, and only a relatively small number of high confidence peaks were detected.

Similarly, two sets of Myc ChIP-seq peaks in Kc167 from data published by (Yang et al., 2013) were used to generate a set of Myc high confidence binding sites. The genomic annotation of the sets of high confidence sets peaks of Gro, Cdk9 and Myc were visualized in a bar plot, showing a similar distribution among the three factors, with most of the binding sites present at promoter regions and introns (Figure 30).

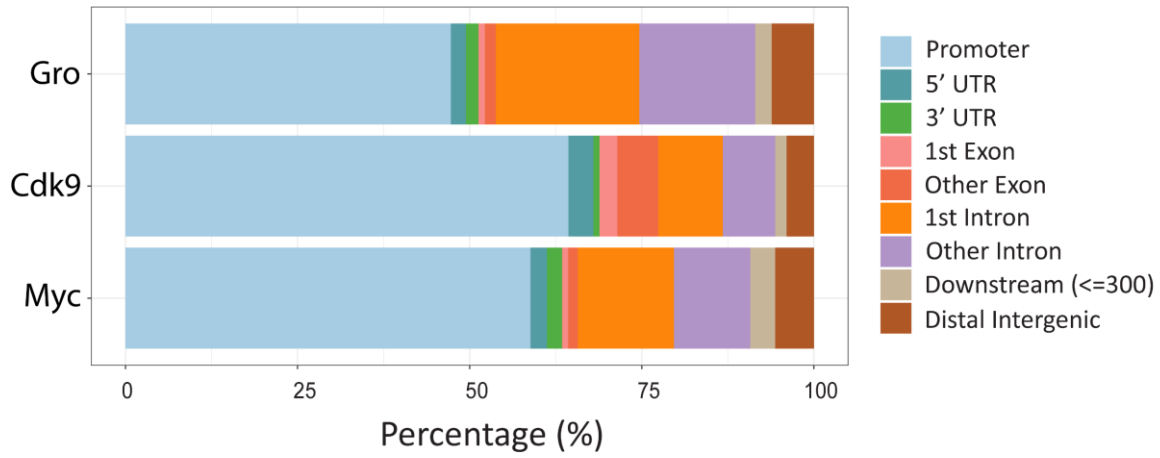


Figure 30: Genomic annotation of Gro, Cdk9, and Myc binding in Kc167 cells.
Bar graph illustrating the percentage of peaks within each genomic location for Gro, Cdk9, and Myc binding sites in Kc167 cells. Promoters were defined in a region of ±250 around TSSs.

Subsequently, the high confidence set of binding sites for Cdk9 and Myc were separately analysed for overlap with Gro peaks at the TSSs and not at the TSSs and visualized in a Venn diagram. The Venn diagram in Figure 31 A shows that 47% of Gro peaks not at the TSS regions and 75.5% at the TSS overlap with Cdk9, showing that Cdk9 is not excluded at genomic locations where Gro is bound in Kc167 cells. This data suggests that Gro does not obstruct the recruitment of P-TEFb to regulatory regions of genes. Moreover, 77% of Gro peaks at the TSSs and 84% not at the TSSs that overlap with Myc in Kc167 cells (Figure 21 B) indicating that Gro does not block Myc recruitment to the genome.

The intersection among the total ChIP-seq peaks of Gro, Cdk9 and Myc showed that 58% of Gro peaks overlap with both Cdk9 and Myc, 32.7% only with Myc and 2% only with Cdk9 (Figure 31 C). Moreover, the genomic annotation of the Gro, Cdk9 and Myc overlapping peaks showed that most of them (59%) are located at promoter regions of genes (Figure 31 D). These results support a model in which Gro promotes Pol II pausing by blocking the positive action of Myc and PTEF-b on the release of paused Pol II. Interestingly, there is a 32.7% of Gro peaks (450 peaks) that co-localize with Myc and not with Cdk9, thus the possibility that Gro interacts with Myc to regulate transcription in an additional manner that does not involve P-TEFb cannot be ruled out.

Taken together the results shown in this section reveal that Gro co-localizes genome-wide with key factors involved in the establishment and release of the stalled Pol II: GAF, Nelf, Cdk9 and Myc.

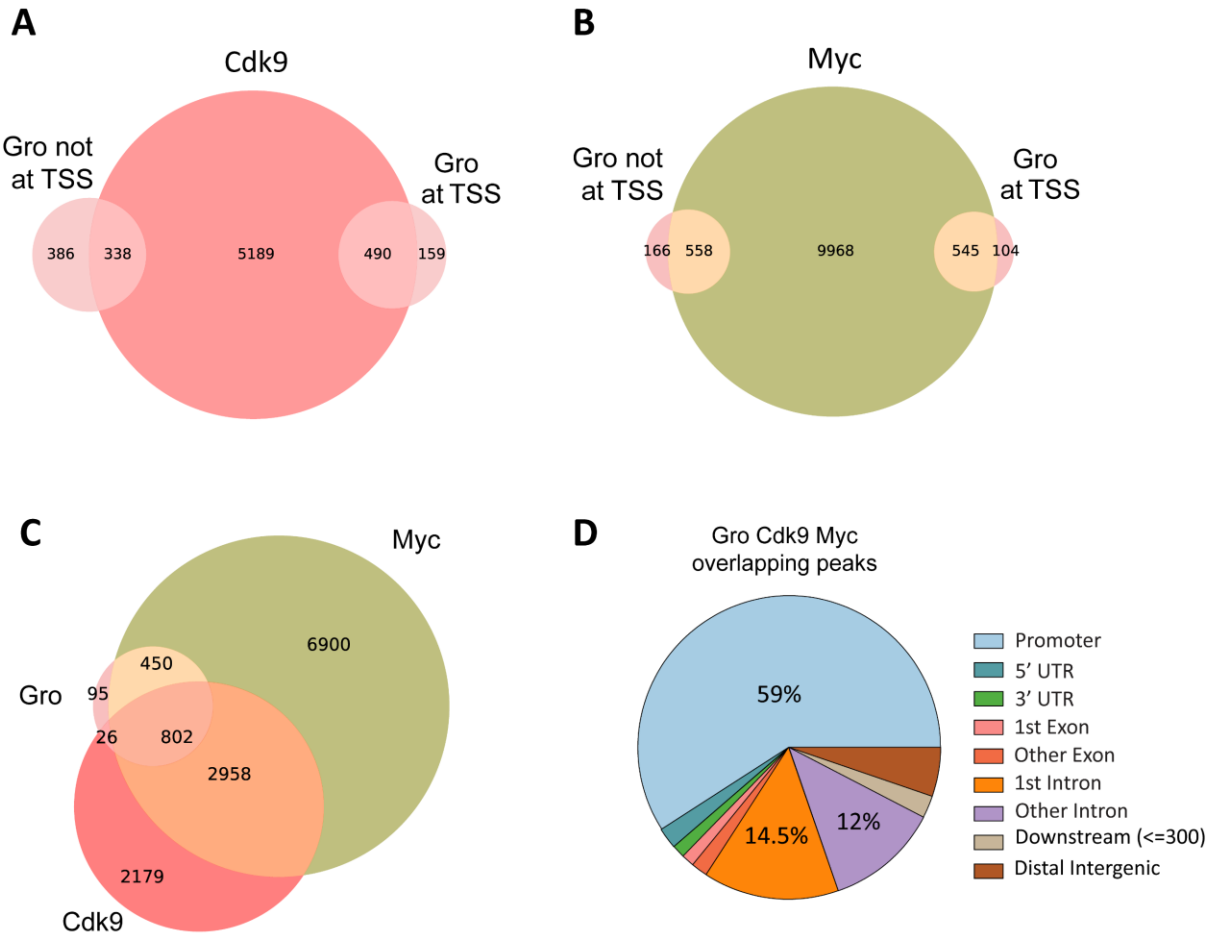


Figure 31: Gro binding sites overlap with peaks of Cdk9 and Myc in Kc167 cells.
(A) Venn diagram showing the overlap between Gro peaks at and not at TSSs regions and Cdk9 binding sites in Kc167 cells. **(B)** Venn diagram showing the overlap between Gro at and not at TSSs regions and Myc binding in Kc167 cells. **(C)** Venn diagram showing the overlap among the total binding sites of Gro, Cdk9, and Myc in Kc167 cells. **(D)** Pie chart showing the percentage of Gro, Cdk9, and Myc overlapping peaks within each genomic location in Kc167 cells. A minimum and reciprocal fraction of 5% of overlap among peaks were required to consider an overlap. Promoters were defined in a region of ±250 around TSSs.

3.2.3.4 Gro co-localizes with Pol II, GAF, Nelf, Cdk9, Myc at TSSs regions in Kc167 cells

To build up a picture of what factors co-localize together at Gro target sites overlaps between the high confidence peaks for Gro, GAF, Nelf-E, Cdk9, and Myc in Kc167 cells were visualized

in UpSet plots. ChIP-seq peaks for different isoforms of Pol II were added to this analysis, aiming to check whether Pol II was present or excluded at the genomic locations where all the tested factors overlap. To explore the hypo-phosphorylated Pol II, two sets of Pol II ChIP-seq binding sites of the unphosphorylated CTD form (8WG16 antibody) from the data generated by (Nazer et al., 2018) were used to generate a set of high confidence peaks. The same process was done with the set of Poll II ChIP-seq binding sites for the Serine 5 (S5) and Serine 2 (S2) phosphorylated CTD isoforms (Nazer et al., 2018).

Furthermore, in view that Gro peaks in Kc167 cells are mostly present at promoters and introns (Figure 22 A), we separately analysed the overlap between Gro and the tested factors at those genomic locations, aiming to detect any difference in the overlapping pattern. Thus, two sets of binding sites were analysed and visualized in two separated UpSet plots: one using the binding sites of all sets of ChIP-seq peaks that were annotated at promoters and the other at introns.

The UpSet plot in Figure 32 visualizes the overlap of factors at TSSs and reveals that most of Gro peaks fall into just three classes. In the largest class (317 peaks; 49% of Gro peaks at the TSSs) Gro peaks overlap with all the factors tested, including the different Pol II isoforms. The other two smaller classes of Gro peaks include overlaps with all the tested factors except either Nelf-E (69 peaks; 10.6% of Gro peaks at the TSSs) or GAF (59 peaks; 9% of Gro peaks at the TSSs). This result supports a model in which Gro promotes promoter-proximal pausing by blocking the positive action of the P-TEFb complex on transcriptional elongation.

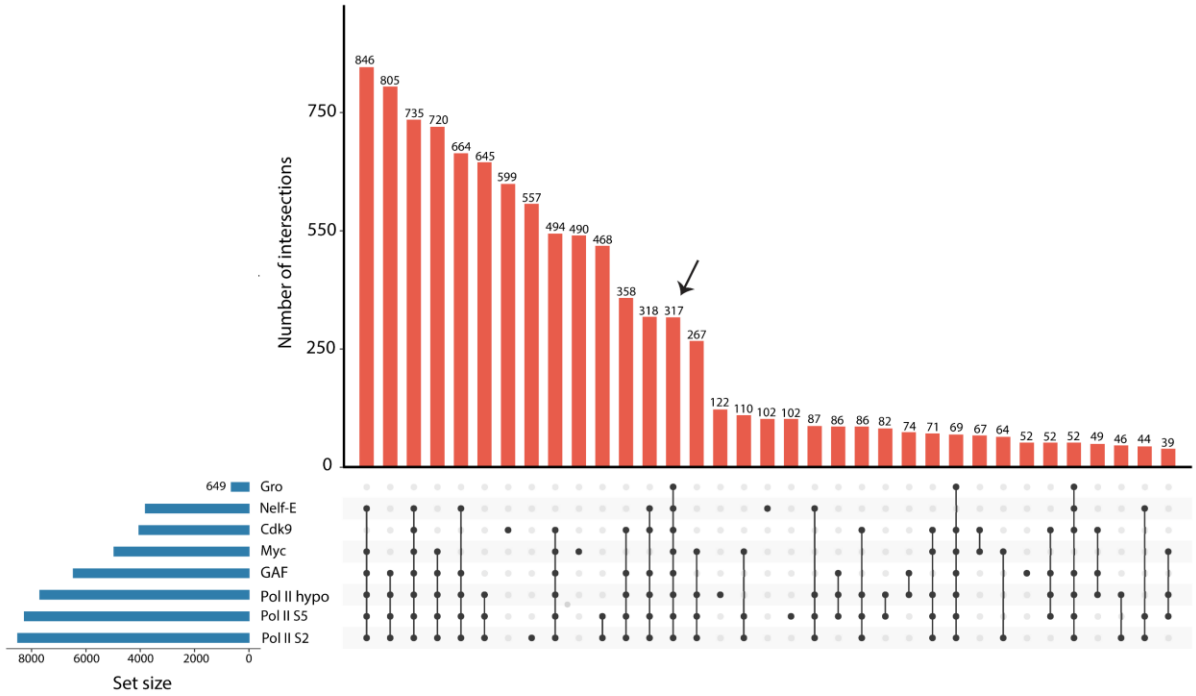


Figure 32: Gro binding sites frequently co-localize with Nelf-E, Cdk9, Myc, GAF, and Pol II at TSSs regions of genes in Kc167 cells.

UpSet plot showing the intersections among peaks at TSSs of the following data sets: Gro, Nelf-E, Cdk9, Myc, GAF, hypo-phosphorylated CTD of Pol II (Pol II hypo), Pol II with C-terminal domain (CTD) phosphorylated at serine 2 (Pol II S2), and Pol II with CTD phosphorylated at serine 5 (Pol II S5). The connected lines in the lower panel of the plot and the upper bar graph (in red) represent respectively the sets of peaks included in each intersection and its corresponding number of overlapping peaks. The horizontal left bar graph represents the total number of peaks in each data set. The largest intersection that includes Gro (317 peaks, point with an arrow) also includes GAF, NELF-E, Cdk9, Myc, and all Pol II isoforms.

Among the 317 binding sites where Gro overlaps with the tested factors (GAF, Nelf-E, Myc, Cdk9, and Pol II) at TSSs are included peaks at the *bona fide* Gro targets genes *E(spl)mβ-HLH* and *E(spl)m3-HLH.* The overlap among the binding profiles of those factors at the TSSs of the *E(spl)m3-HLH* and *E(spl)mβ-HLH* is illustrated in Figure 33 A and B, respectively. The other three Gro target genes included in the list of overlapping peaks at the TSSs are *argos* (*aos*), *cut*, and *naked (nkd)* (Figure 34 A, B, and C, respectively). It is noteworthy that in neither of the genes displayed (Figure 33 and 34) the Pol II ChIP-seq binding profiles show additional peaks downstream within the gene bodies, suggesting that Pol II may be poised at the TSSs. In terms of the chromatin states, the overlapping binding profiles at those genes are mainly within the enhancer (red) or active TSSs (orange) chromatin, however peaks at the Polycomb chromatin (blue) are not totally excluded (Figure 34 B). These examples are consistent with the genome-wide chromatin analysis at Gro binding sites in Kc167 cells (section 3.2.2.1 of the

results). Moreover, these examples showed that the overlapping peaks are frequently surrounded by competent (green), and Polycomb chromatin.

Interestingly, all the gene loci shown in Figures 33 and 34 displayed several Gro peaks. One Gro peak that overlaps with all the factors (yellow shaded), and additional Gro peaks either upstream or downstream of the TSSs which do not seem to overlap with Pol II but do overlap with some of the remaining tested factors (red shaded in Figure 33 and 34).

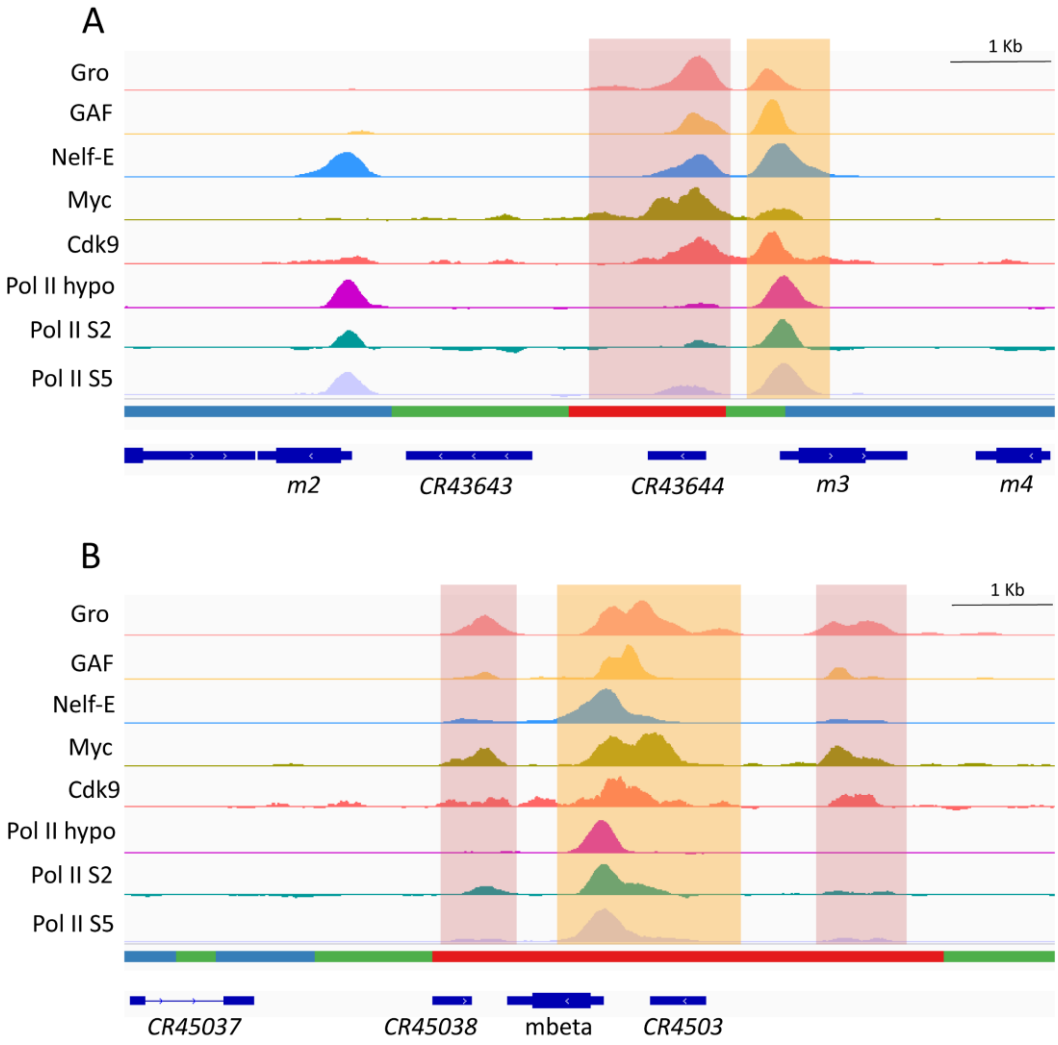


Figure 33: Visualization of binding profiles of the indicated ChIP-seq datasets, as well as the chromatin states, at the genes of the *E(spl)-C* locus in Kc167 cells.

The figure shows binding profiles for Gro, GAF, Nelf-E, Cdk9, Myc, hypo-phosphorylated CTD of Pol II (Pol II hypo), Pol II with C-terminal domain (CTD) phosphorylated at serine 2 (Pol II S2), and Pol II with CTD phosphorylated at serine 5 (Pol II S5) at the **(A)** *E(spl)m3-HLH* **(B)** *E(spl)mβ-HLH* genes. Gro peaks that overlap with all factors at TSSs are shaded in yellow, whereas Gro peaks that overlap only with some of the tested factors are shaded in red. Binding profiles are aligned with the chromatin map colour coded as in Figure 26 A.

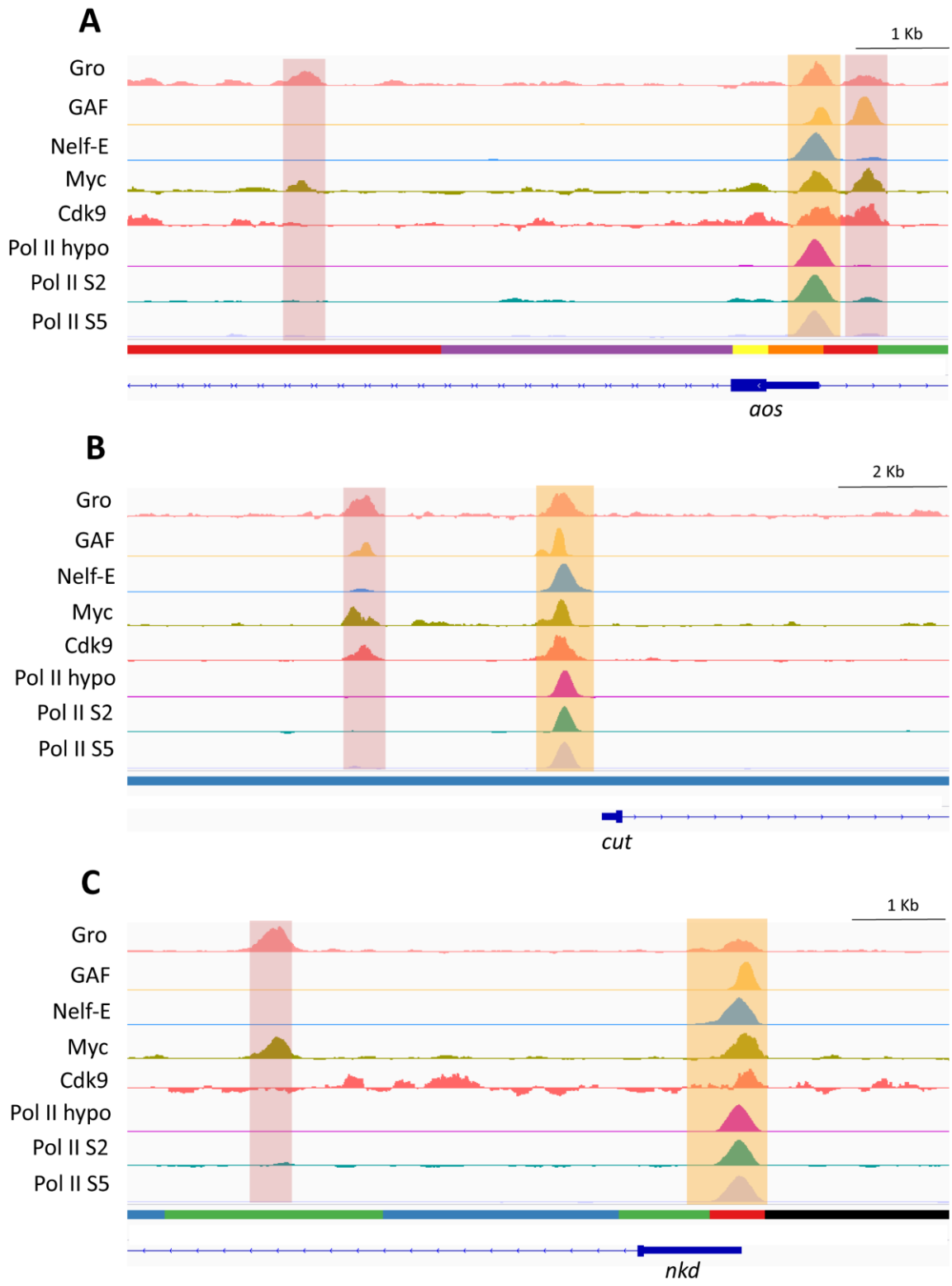


Figure 34: Visualization of binding profiles of the indicated ChIP-seq datasets, as well as the chromatin state, at the *argos* (*aos*), *cut* and *naked* (*nkd*) genes in Kc167 cells.

The figure shows binding profiles for Groucho (Gro), GAGA factor (GAF), Nelf-E, Cdk9, Myc, hypo-phosphorylated CTD of Pol II (Pol II hypo), Pol II with C-terminal domain (CTD) phosphorylated at serine 2 (Pol II S2), and Pol II with CTD phosphorylated at serine 5 (Pol II S5) at the **(A)** *aos* **(B)** *cut* and **(C)** *nkd* genes Gro peaks that overlap with all factors at TSSs are shaded in yellow, whereas Gro peaks that overlap only with some of the tested factors are shaded in red. Binding profiles are aligned with the chromatin map colour coded as in Figure 26 A.

There is a different pattern of factors co-localizing within intronic regions (Figure 35 A). There were fewer peaks of all tested factors within introns (set sizes in Figure 35 A) than at TSSs (see set sizes in Figure 32), which do not frequently overlap with others. This is expected since the factors tested are more frequently located at promoters. The number of Gro binding sites located introns (517 peaks) was similar to that present at TSSs (649 peaks). However, very few intronic Gro peaks overlap with the different Pol II isoforms (asterisks in Figure 35 A) in contrast to the observed at TSSs. To visualize the overlapping pattern of Gro with the tested factors at introns more clearly, a new UpSet plot was generated excluding the Pol II datasets (Figure 35 B). It showed that only 88 Gro peaks (arrow in Figure 35 B), overlap with all the tested factors (GAF, Cdk9, Nelf, and Myc). However almost all Gro peaks at introns overlap at least with one of the tested factors, and there were only 46 Gro peaks that did not overlap any of them (Figure 35 B). It is worth noting that the majority of Gro peaks at introns in the different classes of intronic Gro peaks include Myc, Cdk9 or both.

Two examples were Gro overlaps with Myc at introns of genes are observed at the *aos* and *nkd* genes in the above figure (red shaded in Figure 34 A and C).

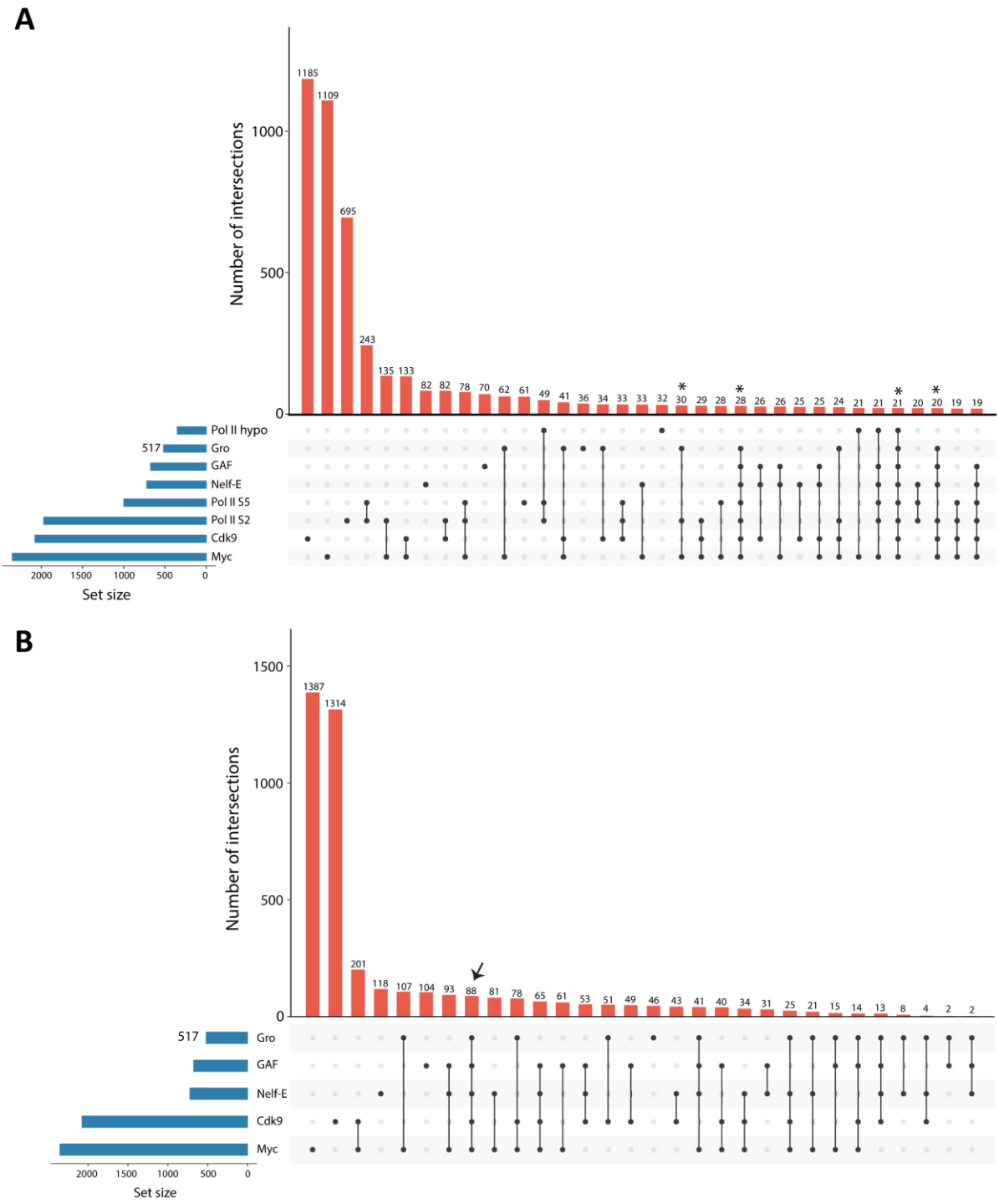


Figure 35: Gro binding sites do not co-localize with Pol II at Intronic regions of genes in Kc167 cells.
UpSet plot showing the intersections among peaks located at introns of the indicated data sets. **(A)** Intersection of high confidence peaks located at introns of the following data sets: Gro, Nelf-E, Cdk9, Myc, GAF, hypo-phosphorylated CTD of the Pol II (Pol II hypo), Pol II with C-terminal domain (CTD) phosphorylated at serine 2 (Pol II S2), and Pol II with CTD phosphorylated at serine 5 (Pol II S5). Asterisks show the intersections that include Gro and Pol II isoforms. **(B)** Intersection of high confidence peaks located at introns of the following data sets: Gro, GAF, Nelf-E, Cdk9, and Myc. The arrow shows the intersections where Gro overlaps with all the tested factors. A minimum and reciprocal fraction of 5% of overlap among peaks was required to consider overlap. Asterisks show the intersections where Gro overlaps with the rest of the factor.

3.2.4 Discussion

The analysis of the genomic distribution of Gro binding sites in the three cell lines analysed (S2R+, Kc167 and BG3) showed that Gro frequently binds to the promoter or intronic regions of genes. Considering that in metazoans, enhancers are often located in the first introns of genes (Arnold et al., 2013; Kharchenko et al., 2011), this finding indicates that Gro occupies enhancers, in addition to promoters to regulate gene expression. Profound analysis of Gro peaks in BG3 cells showed that GAGA binding site was the most significant motif found, which was also observed previously in the S2R+ and Kc167 embryo-derived cell lines (Kaul et al., 2015).

The chromatin states analysis in Kc167 and BG3 cells performed in this chapter revealed that a vast majority of Gro peaks are located at regions of enhancer chromatin in both cell lines. The enhancer chromatin is enriched in the chromatin marks H3K27ac, H3K56ac, H3K4me1, as well as in DHS. In contrast, previous analysis of H3K27ac ChIP-seq data in Kc167 cells revealed that H3K27ac is excluded at Gro peaks, although it is present in surrounding regions. The observation led the authors to suggest that those regions are nucleosome-free regions or regions where nucleosomes are modified (Kaul et al., 2014). The chromatin states analysis also showed a smaller fraction of Gro binding sites were located within regions of active chromatin, which is very similar to the enhancer chromatin but differs in the histone marks H3K4me3 and H3K4me1. Whereas the enhancer chromatin is rich in H3K4me1, the active chromatin is rich in H3K4me3. Previous work in my lab already observed H3K4me3 enrichment at Gro peaks, especially when Gro is bound at promoters (Kaul et al., 2014), which is consistent with this thesis's observations.

The overlap analysis among ChIP-seq binding sites of Gro and the different pausing factors in Kc167 cells showed that GAF, and Nelf co-occupy genomic regions bound by Gro that are mostly located at promoters. This is consistent with the model of Gro acting with those factors to promote Pol II pausing. In addition, the results revealed that Cdk9 and Myc are not excluded from Gro bound regions, indicating that Gro may block the positive action of those factors on the release of poised Pol II. Finally, it was shown that Pol II co-occupies genomic regions where Gro overlaps with the factors involved in Pol II pausing, in agreement with

previous observations that reported that Pol II binds to genomic locations occupied by Gro (Kaul et al., 2014).

Taken together these results support a model in which Gro acts together with GAF and Nelf to promote Pol II pausing at regulatory regions characterized by open chromatin. In addition, they suggest that Gro may mediate this process by blocking the action of factors that promote transcriptional elongation, such as P-TEFb.

3.3 Investigating Gro interacting partners using protein-protein interaction approaches

Most biological processes, including transcription, are dependent on protein complexes rather than the action of individual proteins. Thus, the identification of the proteins and protein complexes that interact with Gro partners is a key step to understanding the molecular mechanisms underlying Gro function. To date, several proteins that interact with Gro have been identified from two-hybrid screens and affinity purification techniques (see Table 1 in the Introduction). However, the majority of these are sequence-specific DNA-binding TFs that recruit Gro to the genome, while relatively few proteins that assist Gro to invoke repression have yet been identified.

Despite the relevance of studying Gro partners, very few protein-protein interaction screens have been carried using Gro as bait. Furthermore, those screens have not used the full-length Gro protein, but individual regions, such as the WD domain (Goldstein et al., 2005), or the Gro central region (Kwong et al., 2015). One of the objectives of this work was to carry out an unbiased protein-protein interaction screen to find novel proteins that interact with Gro, intending to gain understanding into the molecular mechanisms through which Gro repress transcription. In this direction, I first carried out a yeast two-hybrid assay (Y2H), which uses *Saccharomyces cerevisiae* as a host to test physical interactions between pairs of proteins.

The Y2H has been the most used high-throughput method *in vivo* to detect binary protein-protein interactions since it was first described in the 90s (Fields and Song, 1989). This technique exploits the bifunctional nature of the yeast GAL4 TF, which can activate transcription if its DNA-binding domain (DBD) is physically associated with the activation domain (AD). In the Y2H, the GAL4 DBD domain is fused to a protein of interest (the bait), while another protein is fused to the AD domain (the prey). Thus, if the bait and the prey physically interact, the GAL4 TF is reconstructed, activating the expression of specific reporter genes, which allows the identification of physical interaction between these proteins. However, this method requires the expression of the proteins of interest in the yeast nucleus, which has inherent limitations. The screen Y2H might not be suitable for proteins that are natively expressed in another cell compartment (such as cytosolic or membrane associate

proteins), which do not fold properly in the nucleus environment. Additionally, protein-protein interactions that require post-translational modifications not available in yeast are also unlikely to be detected using this method [reviewed in (Brückner et al., 2009)].

The Y2H approach using the full-length Gro protein as the bait was not successful in finding protein-protein interactions, even though Gro is a nuclear protein, and there were no apparent technical problems during the assay (see section 3.3.1 of this chapter).

Thus, we decided to move on to another approach and perform an enzyme-catalysed proximity labelling assay. In this approach, usually, an enzyme is fused to a protein of interest (bait), which when expressed in culture cells catalyses the addition of a biotin covalent tag to proximal proteins in the neighbourhood. Subsequently, the biotinylated proteins can be isolated and identified by mass spectrometry (MS) (workflow summarized in Figure 36). This approach allows the capture of protein complexes, in contrast to the Y2H technique, in which only direct interactions between two proteins are detected. Additionally, proximity labelling techniques address some of the challenges of similar approaches to detect protein complexes, such as affinity purification assays coupled to MS (AP-MS). Among the advantages of proximity labelling assays is the identification of transient interactors with lower affinity, and the detection of protein-protein interactions in the native environment of living cells. There are three main enzymes used for proximity labelling: biotin ligase (BioID, BioID2, TurboID, miniTurbo), horseradish peroxidase (HRP), and engineered ascorbate peroxidase (APEX, APEX2).

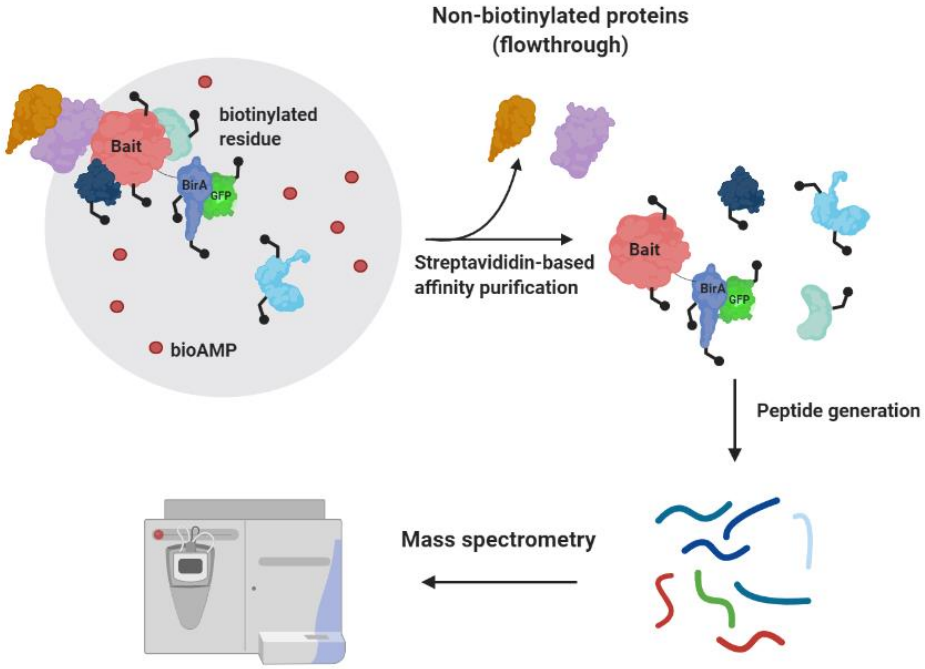


Figure 36: Scheme illustrating the workflow in proximity labelling.

The protein of interest (bait) is fused to a promiscuous form of the *E.coli* biotin ligase (BirA) (the fused protein may also include a tag, such as GFP). BirA transforms biotin into biotinoyl-5'-AMP (bioAMP) (red dots in the illustration). bioAMP can react with the primary amines of proximal proteins (labelling radious ~10 nm) (black dots). Distal proteins are not labelled, whether they interact with the bait or not (yellow and purple proteins in the illustration). Subsequently, cells are lysed, and biotinylated proteins are purified from the extracts using streptavidin beads. The isolated biotinylated proteins are then digested and identified by mass spectrometry. Illustration created with BioRender.com.

Specifically, I applied a TurboID -based labelling approach to detect Gro interacting partners in *Drosophila* culture cells. TurboID and miniTurbo are two promiscuous mutants of the *Escherichia coli* biotin ligase that were generated using yeast display directed evolution (Branon et al., 2018). TurboID and miniTurbo offer some advantages over other enzymes, including a faster biotinylation time. They can generate enough biotinylated material for proteomic analysis after 10 minutes of labelling, compared to BioID and BioID2 that needed between 18 and 24 hours. This increased catalytic efficiency of TurboID and miniTurbo facilitates the application of proximity labelling for studying dynamic processes that occur on a timescale of minutes, and hence to study rapid developmental processes in living organisms, such as worms or flies. In addition, this method is simple and non-toxic to living cells, only biotin needs to be supplied to initiate tagging. This is an advantage over APEX labelling that

require the use of the chemical compound hydrogen peroxide, which is toxic for cells (Branon et al., 2018).

Since TurboID and miniTurbo were reported in 2018, they have been successfully used to profile networks of protein-protein interactions in different organisms and tissues, such as HEK cells (Branon et al., 2018), *S. pombe* (Larochelle et al., 2019), and *Arabidopsis* (Zhang et al., 2019). Furthermore, Branon and colleges proved the capacity of TurboID and miniTurbo to effectively biotinylated proteins when expressed in the model organisms *Drosophila* and *C. elegans.* However, this approach was not reported in *Drosophila* cell lines at the time where this work was done, hence I first explored and optimized the expression of the Gro bait fused to TurboID and miniTurbo in *Drosophil*a Kc167 cells.

3.3.1 Yeast two-hybrid using full-length Gro as bait

I first performed a yeast two-hybrid (Y2H) screen to identify novel proteins that interact with *Drosophila* Gro. This technique allows the detection of direct interactions between proteins in living yeast cells, and it is based on the separation of the two modules of the yeast GAL4 TF: the DNA-binding domain (DBD), and the activation domain (AD). In the Y2H technique, the proteins of interest are separately fused to the DBD and the AD domains, which will act as the bait and the prey, respectively. If there is an interaction between the bait and the prey, the functional GAL4 TF is reconstituted, recruiting Pol II, and inducing the transcription of different reporter genes. The expression of these genes then leads to specific yeast phenotypes, enable them to grow on a selective medium (auxotrophic selection), and/or activate a colour reaction. This Y2H approach and the *Drosophila* library used in assay had been used successfully in the lab previously when Pho was recovered as it interacted with the C-terminal domain of Spt5 (Harvey et al., 2013).

For this Y2H screen, the full-length Gro protein fused to the GAL4 DBD was used as the bait. As pray, we used a normalized *Drosophila* cDNA library from Clontech, which is constructed from mRNA isolated in equal quantities from embryos (~20 hr), larvae, and adults. This cDNA library encodes fragments of *Drosophila* proteins fused to the GAL4 activation domain. Four reporter genes were used to select proteins that interact physically with Gro: *AUR1-C*, *ADE2*, *HIS3*, and *MEL1* (see Material and Methods).

Prior to the screen, I did control experiments to test the bait for autoactivation and toxicity. I was able to determine that the expression of the Gro bait did not activate the transcription of reporter genes by itself and is not toxic when expressed in yeast (Figure 37 A). Thus, I proceeded to screen the *Drosophila* protein library with Gro as the bait, following the Matchmaker® Gold Yeast Two-Hybrid System User Manual.

The quality controls carried out during the Y2H screen did not reveal any obvious technical problems with the screening process. We screened 4.7 million colonies from the cDNA library (screening fewer than 1 million colonies may result in an inability to detect positive interactions), with a mating efficiency of 4.2 % (>2% is a good mating efficiency). Moreover, the positive control provided by the kit (Clontech), as well as the positive controls formed by the expression of the bait together with the WRPW domain (Figure 37 B) and Hairy (Figure 37 D) resulted in the activation of all reporter genes.

However, we recovered relatively few putative positive colonies from the screen (< 200) and we did not recover any proteins previously known to interact with Groucho. We did recover multiple "positive" colonies expressing fragments of the Vdup1 protein.

Thus, the full-length Vdup1 protein was retested to explore whether it is a true Gro interactor or if truncated Vdup1 was a false positive; the yeast strain containing the Gro bait and the full-length Vdup1 was not able to activate any of the reporter genes (Figure 37 F), therefore we concluded that Vdup1 is a false positive. The possible biological interaction between *gro* and *Vdup1* was additionally tested in the wing genetic assay and no interaction was detected, as expected (Figure 57 in the Appendix, and section 3.1.1.6).

In conclusion, the Y2H approach was not fruitful for the identification of Gro interacting partners, despite the controls indicating that there were no technical problems in the performance of the screen. One of the possible reasons that may explain why the screen was not successful is the repressive nature of Gro. In this context, it should be considered that the WD domain is highly conserved between *Drosophila* Gro and its ortholog in *Saccharomyces cerevisae* Tup1 (Flores-Saaib and Courey, 2000). Furthermore, the N-terminal domains of Gro and Tup1 have both similar functions, promoting the corepressor self-association (Courey and Jia, 2001). Thus, we think that the Gro bait could somehow have interacted with the yeast Pol II machinery and prevented the transcription of the reporter genes.

Consequently, we abandoned this approach, and we pursued an alternative method to detect novel proteins that interact with Gro.

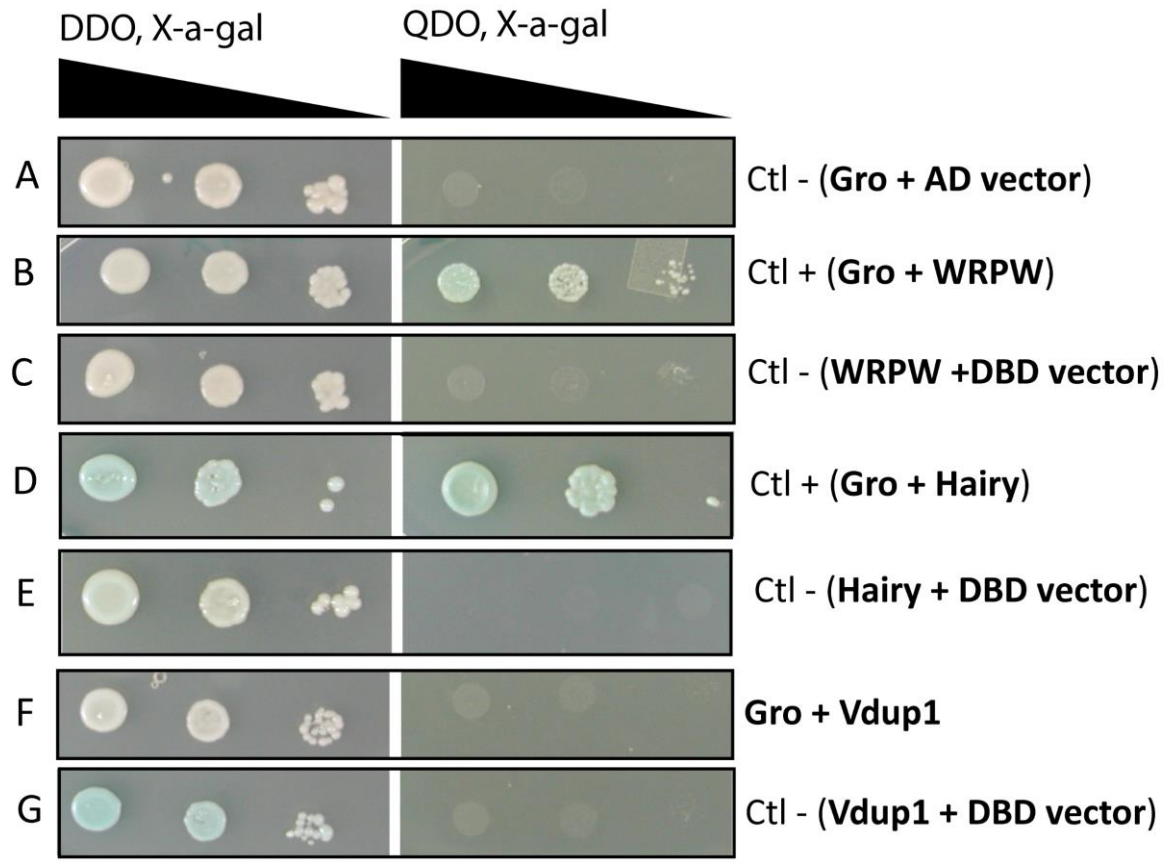


Figure 37: Gro and the Vitamin D3 up-regulated protein 1 (Vdup1) do not interact physically.
The image shows yeast strains expressing the indicated constructs. Double dropout medium containing X-α-gal (DDO, X-a-gal) was used to test the expression of the plasmids. Protein-protein interactions were tested using X-α-gal in quadruple dropout medium (QDO, X-a-gal). **(A)** yeast strain co-expressing the full-length Gro and an empty vector [pGADT7; containing an activation domain (AD)] as a negative control to demonstrate that Gro does not activate reporter gene expression by itself **(B)** yeast strain co-expressing the full-length Gro and the WRPW domain as a positive control of the assay. **(C)** yeast strain co-expressing the WRPW domain and an empty vector [pGBKT7; containing a DNA-binding domain (DBD)] as a negative control. **(D)** yeast strain co-expressing the full-length Gro and Hairy (a well-known Gro interacting partner) as a positive control. **(E)** yeast strain co-expressing Hairy and an empty vector [pGBKT7; containing a DNA-binding domain (DBD)] as a negative control. **(F)** yeast strain co-expressing the full-length Gro and the full-length Vdup1 does not grow on QDO medium showing that there is not a physical interaction between Gro and Vdup1. **(G)** yeast strain co-expressing the full-length Vdup1 and an empty [pGBKT7; containing a DNA-binding domain (DBD)] as a negative control. Colonies were grown for 3 days at 30°C. 20-fold serial dilutions were performed prior to colony plating to ensure growth on the selective medium is dependent solely on *HIS3*, *ADE2* and *MEL1* reporter gene expression.

3.3.1.1 Y2H and co-immunoprecipitation assays indicate that Gro does not physically interact with GAF

Despite the result that the Y2H did not detect any new Gro partners, the controls worked properly, showing activation of all reporter genes in the yeast strains co-expressing Gro and the WRPW domain (Figure 37 B), as well as Gro and Hairy (Figure 37 D). Given the availability of the Y2H reagents in our lab, we decided to use this method to test if there is a physical direct interaction between Gro and GAF. In the previous result chapters of this work, it was demonstrated the strong biological interaction between Gro and GAF, as well as the large genome-wide co-localization between their binding sites. Thus, we next tested if Gro physically interacts with the two major GAF isoforms: the one formed by 519 aa (GAF-519), and the formed by 581 aa (GAF-581). Both isoforms share a common N-terminal region, but they differ in the length of the glutamine-rich C-terminal domain (Benyajati, 1997).

The yeast strain co-expressing the full-length proteins Gro and GAF-519 was not able to activate the expression of any of the reporter genes (Figure 38 D), and hence no physical interaction was detected between Gro and GAF-519. Regarding the other GAF isoform, the yeast strain co-expressing Gro and GAF-581 did not grow in the medium that selects for the expression of the plasmids [double dropout medium (DDO)] (Figure 38 F). Therefore, the presence of both plasmids in that yeast strain was not proven, and hence we could not conclude if Gro and GAF-581 physically interact or not. It is worth noting that yeast transfected in parallel with individual plasmids encoding Gro and GAF-581 as negative controls did grow in DDO medium (Figure 38 B and G, respectively), showing the proper expression of those plasmids individually. Thus, it remains unknown the reason why the yeast strain co-expressing both plasmids did not survive in the selective DDO medium. We can speculate that the Gro together with GAF-581 could have formed a repressive complex that turned off the expression of some essential genes that resulted toxic to yeast.

Therefore, the Y2H experiment indicated that Gro and GAF-519 do not physically interact, however we could not conclude anything about the interaction between Gro and GAF-581. Additional approaches, such as co-immunoprecipitation assays, need to be taken to test the physical interactions between these proteins.

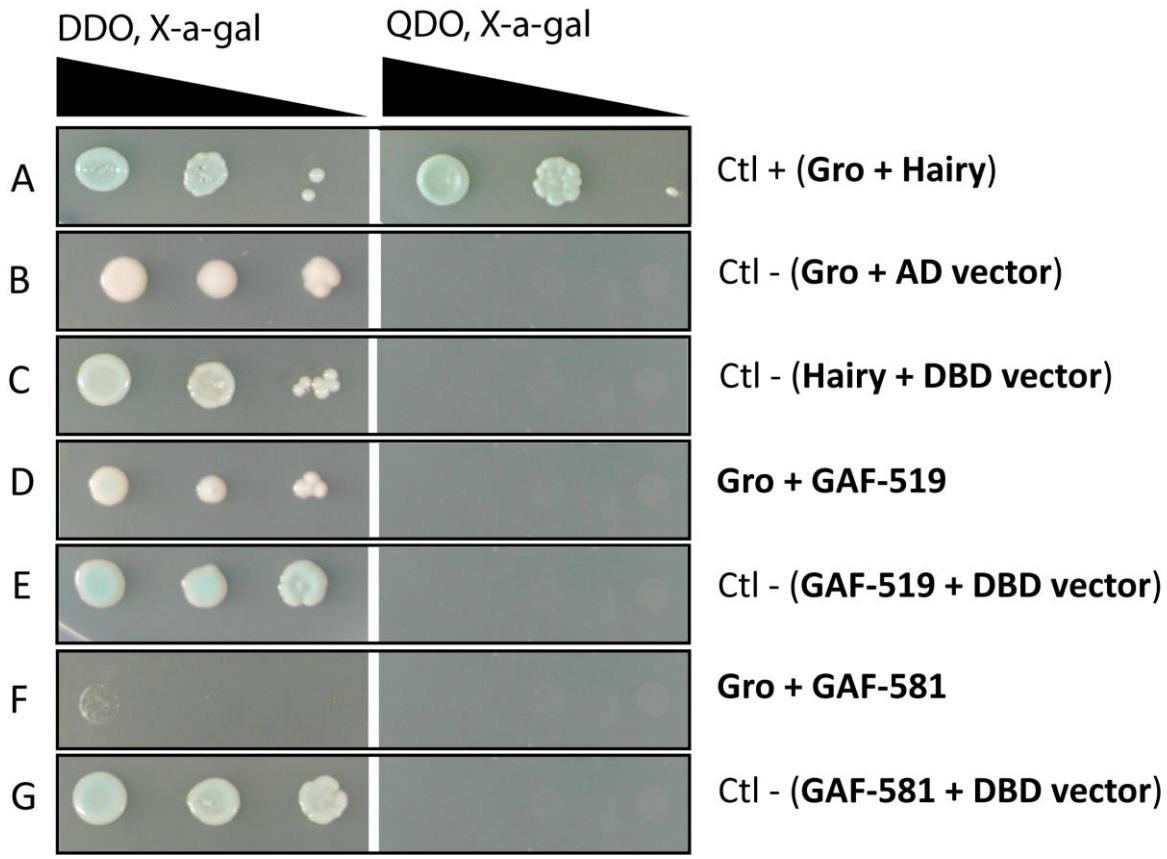


Figure 38: Gro and the GAGA factor isoform, GAF-519, do not interact physically in a yeast 2-hybrid experiment.

The image shows yeast strains expressing the indicated constructs. Double dropout medium containing X-α-gal (DDO, X-a-gal) was used to test the expression of the plasmids. Protein-protein interactions were tested using X-α-gal in quadruple dropout medium (QDO, X-a-gal). **(A)** yeast strain co-expressing the full-length Gro and Hairy (a well-known Gro interacting partner) as a positive control of the assay. **(B)** yeast strain co-expressing the full-length Gro and an empty vector [pGADT7; containing an activation domain (AD)] as a negative control to demonstrate that Gro does not activate reporter gene expression by itself **(C)** yeast strain co-expressing Hairy and an empty vector [pGBKT7; containing a DNA-binding domain (DBD)] as a negative control. **(D)** yeast strain co-expressing the full-length Gro and the full-length GAF-519 did not grow in QDO medium showing there is no physical interaction between Gro and GAF-519 **(E)** yeast strain co-expressing GAF-519 and an empty vector [pGBKT7, containing a DNA-binding domain (DBD)] as a negative control. **(F)** yeast strain co-expressing the full-length Gro and the full-length GAF-581 did not grow either in DDO or QDO medium. **(G)** yeast strain co-expressing GAF-581 and an empty vector [pGBKT7; containing a DNA-binding domain (DBD)] as a negative control. Colonies were grown for 3 days at 30^{o}C. 20-fold serial dilutions were performed prior to colony plating to ensure growth on the selective medium is dependent solely on *HIS3*, *ADE2* and *MEL1* reporter gene expression.

We next decided to test the physical interaction between Gro and the GAF isoforms by an alternative method, using co-immunoprecipitation assays (co-IPs) in S2R+ *Drosophila* cells. Co-IPs are one of the standard methods used to identify or confirm protein-protein interactions and allow the isolation of protein complexes from a solution by using a specific

antibody. Thus, in contrast to the Y2H method, co-IPs identify not only proteins that interact directly, but also the close indirect interactions through protein complexes.

To facilitate the immunoprecipitations, the full-length Gro, GAF-519, and GAF-581 proteins were tagged at the N-terminal with two different tags: three repetitions of Flag (3xFlag), and six repetitions of Myc (6xMyc). The WRPW domain was also tagged with 3xFlag and 6xMyc to use it as a positive control of the assay (see Materials and Methods).

To test for proteins that co-immunoprecipitated with Gro, different combinations of the tagged proteins were overexpressed in S2 *Drosophila* cells, which was followed by immunoprecipitation (IP), and western blot (WB) to detect the proteins of interest. First, the 6xMyc-tagged Gro was overexpressed in pairwise combinations with the following 3xFlag-tagged proteins: GAF-519, GAF-581 and WRPW. Figure 39 A shows that the 6xMyc-tagged Gro co-immunoprecipitated when overexpressed with the 3xFlag-tagged WRPW domain (IP was performed with an anti-Flag antibody) (arrowhead in lane 1), showing the proper function of the positive control. In contrast, the 6xMyc-tagged Gro did not co-immunoprecipitated when overexpressed with the 3xFlag-tagged GAF-519 protein (Figure 39 A lane 2), indicating that Gro might not physically interact with GAF-519. In the case of the simultaneous overexpression of 6xMyc-tagged Gro and the 3xFlag-tagged GAF-581, we were unable to solve if they physically interact since GAF-581 expression was not detected in the input (asterisk in Figure 39 A lane 3).

The opposite combination of tags was additionally tested (3xFlag-tagged Gro with the 6xMyc-tagged proteins: GAF-519, GAF-581, and WRPW). Figure 39 B shows that neither of the 6x Myc-tagged proteins GAF-581 and GAF-519 co-immunoprecipitated with the 3xFlag-tagged Gro (IP was performed with an anti-Flag antibody). However, in the positive control, even if the WRPW domain did co-immunoprecipitated with Gro, the WB signal was very weak (arrowhead in Figure 39 B lane 1). Thus, we cannot rule out the possibility that there is a physical interaction between Gro and GAF-581.

In summary, the Y2H and co-IP results indicated that there is no physical interaction between Gro and GAF-519 since this was not detected by either of the assays. This is consistent with a previous study in which Gro was not detected as one of the top GAF interactors (p-value = 0.02) in an AP-MS screen that used GAF-519 as a bait (Lomaev et al., 2017). However, a

physical interaction between Gro and the other GAF isoform, GAF-581, cannot be excluded since I could not test it due to the technical problems mentioned previously. If I had more time, I would have optimized the conditions for the co-IPs to test this interaction. Furthermore, I would have tested the possible physical interaction between Gro and the other candidate, Nelf, since both the genetic assay and the bioinformatics indicated a strong interaction and co-localization between those proteins.

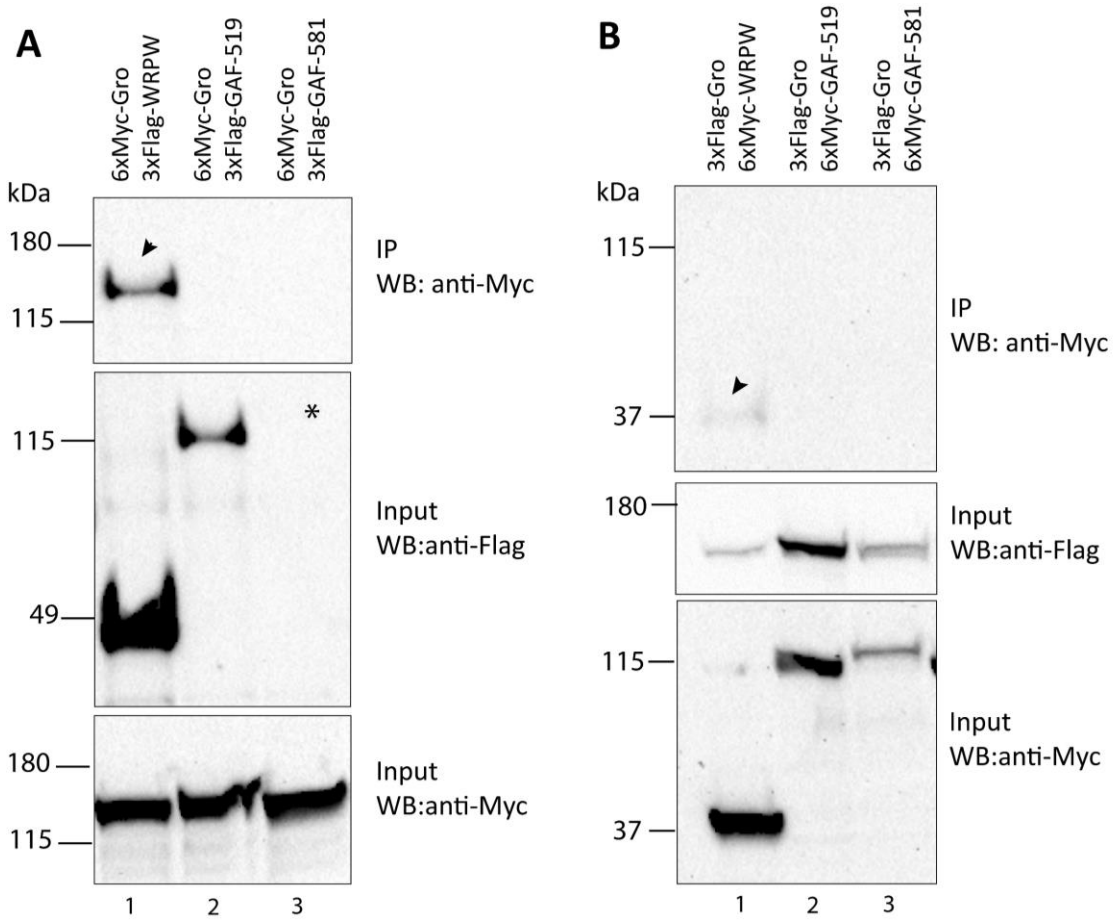


Figure 39: Gro and the GAGA factor isoform, GAF-519, do not physically interact in co-immunoprecipitation assays.

The total protein lysates were immunoprecipitated using an anti-Flag antibody. The immunoprecipitated (IP), as well as the total protein lysates (Inputs), were analysed by western blotting (WB). Images show western WB the indicated antibodies for IP and input samples from cells that co-expressed the following proteins **(A)** (1) 6xMyc-tagged Gro (6xMyc-Gro), and 3xFlag-tagged WRPW (3xFlag-WRPW). (2) 6xMyc-Gro, and 3xFlag-tagged GAF-519 (3xFlag-GAF-519). (3) 6xMyc-Gro, and 3xFlag-tagged GAF-581 (3xFlag-GAF-581). **(B)** (1) 3xFlag-tagged Gro (3xFlag-Gro), and 6xMyc-tagged WRPW (6xMyc-WRPW). (2) 3xFlag-Gro, and 6xMyc-tagged GAF-519 (6xMyc-GAF-519). (3) 3xFlag-Gro, and 6xMyc-tagged GAF-581 (6xMyc-GAF-581). Arrowheads show the co-immunoprecipitation of the positive controls. Asterisk in A points to the lack of the band corresponding 3xFlag-GAF-581 protein. S2R+ *Drosophila* cells were transiently transfected and total protein extracts were collected three days after transfection.

3.3.2 Turbo-ID proximity labelling assay using the full-length Gro as bait

As explained in the Introduction of this chapter, a TurboID -based labelling approach (Branon et al., 2018) was used as an alternative to the Y2H to perform an unbiased screen for Gro interacting partners. Since the use of the TurboID biotin ligase had not been previously reported in *Drosophila* cells at the time when this experiment was performed, we first explored the feasibility of the system in *Drosophil*a cell lines and the characterization of the fusion proteins. To do that, I generated the bait and the control constructs that I used transient transfection of *Drosophila* Kc167 cells.

We chose to carry out this screen in Kc167 cells as there was published ChIP-seq Gro data in this cell line (Kaul et al., 2014), which could be used to validate the putative interactions that came up from the TurboID-based labelling assay. Moreover, this cell line has been extensively profiled for the genomic distribution of DNA binding proteins and histone modification marks, providing a great data resource.

3.3.2.1 Design and expression of the TurboID or miniTurbo bait and control fusion proteins

In the first place, a preliminary study was performed to test which biotin ligase, TurboID or miniTurbo, was best suited for this experiment. Thus, we designed and generated two Gro fusion proteins (baits), using the cDNA encoding the full-length Gro. In each construct, Gro was tagged at the N-terminal with either TurboID or miniTurbo, and GFP (will be referred to as GFP-TbID-Gro and GFP-miniTb-Gro). We decided to tag Gro at the N-terminal as Gro tagged at this protein end previously shown to have the same activity as the endogenous Gro in *Drosophila* Kc167 cells (Kaul et al., 2014). The GFP tag was added to the fusion protein to facilitate immunocytochemistry assays, as well as the detection of the protein expression in living cells. In addition, a short sequence encoding the repetition of 10 Glycine-Serine aa (G/S) was added to act as a linker between the biotin ligases (Turbo or miniTurbo) and Gro, providing the bait with additional flexibility to label adjacent proteins. In parallel, two control fusion proteins were generated containing either the TurboID or miniTurbo fused to a nuclear localisation signal (NLS) and the GFP tag (will be referred to as GFP-TbID -NLS and GFP-miniTb-

NLS, respectively). These controls were designed to express the biotin ligases in the same subcellular localization as the bait (the nucleus) where they will promiscuously biotinylate proximal proteins, providing an estimation of background. An additional control was used in the first batch of control experiments to mimic endogenous biotinylation in *Drosophila* cells, in this control the N-terminal of Gro was fused to GFP (will be referred to as GFP-Gro). Figure 40 A shows a schematic illustration of the bait (Figure 40 A d and e) and the control fusion proteins designed for in this experiment (Figure 40 A a, b and c).

The bait and the control constructs were generated by genetic fusion following a designed strategy that involved several cloning technologies: TOPO, Gateway, and Gibson. Gateway cloning is based on the site-specific recombination mediated by the lambda integrase and allows the transference of inserts between different cloning vectors in a single reaction (Hartley, 2000). This reaction requires an entry vector with the insert of interest, and a destination vector to generate the desired expression construct. One of the advantages of this technique is that it allows to easily make several constructs, containing different tags or promoters.

Thus, the first step of the strategy was to clone the cDNA sequences encoding *TurboID*, and *miniTurbo* (abbreviated as *TbID* or *miniTb*) into entry vectors by a topoisomerase-catalysed reaction (pENTR/D-TOPO cloning) (Patel, 2009), generating the following plasmids: *pENTR/D-TbID* and *pENTR/D-miniTb*. Two NLS-tagged versions of these biotin ligases were also cloned in parallel to make the control constructs (*pENTR/D-TbID-NLS* and *pENTR/D-miniTb-NLS*).

Gibson assembly cloning (Gibson et al., 2009) was subsequently used to join the DNA sequence of *gro* to the *TurboID* and *miniTurbo* sequences in the entry pENTR/D vectors. A pENTR/D plasmid with the full-length cDNA *gro* sequence was already available in the lab (referred to as *pENTR/D-gro*), thus we made use of this. The Gibson cloning technique has been extensively used and allows the fusion of multiple overlapping DNA fragments in a single reaction due to the activity of three different enzymes (Gibson et al., 2009). To apply this technique, appropriate primers containing homologous overlapping ends (see Materials and Methods and Table 7 in the Appendix) were designed and used to PCR amplify DNA fragments from the entry plasmids. It should be noted that an overhang with a DNA sequence encoding a G/S linker was added at the necessary primers to generate bait constructs with a linker

between Gro and the biotin ligases. Therefore, a first Gibson reaction was performed with two overlapping PCR products encoding *pENTR/D-gro* and *TbID,* to achieve a *pENTR/D-TbID -gro* plasmid. A parallel Gibson reaction was performed with two overlapping PCR products encoding *pENTR/D-gro* and *miniTb* to generate the *pENTR/D-miniTb-gro* plasmid.

Once all the necessary entry plasmids were generated, I proceed to make the final bait and control plasmids by performing Gateway reactions. The inserts of interest were transferred from the entry vectors into a destination vector used for transient expression in *Drosophila* cells, which is designed to express proteins fused to an GFP tag in the N-terminal (*pAGW*). To summarize, the final bait constructs used in this experiment were *pA-GFP-TbID-gro,* and *pA-GFP-miniTbID-gro*. And the control constructs were *pA-GFP-TbID-NLS* and *pA-GFP-miniTbID-NLS.*

I then commenced the preliminary analysis to compare the expression and function of TurboID and miniTurbo biotin ligases in *Drosophila* Kc167 cells. To check the expression and tagging of the bait and the control fusion proteins, total protein lysates were extracted three days after transient transfection and analysed by western blotting. Both Gro baits were expressed successfully at the expected size (GFP-TbID-Gro and GFP-miniTb-Gro, Figure 40 B, lanes 4 and 8, respectively), as well as the control fusion proteins (GFP-TbID-NLS and GFP-miniTb-NLS, Figure 40 B, lanes 3 and 7, respectively).

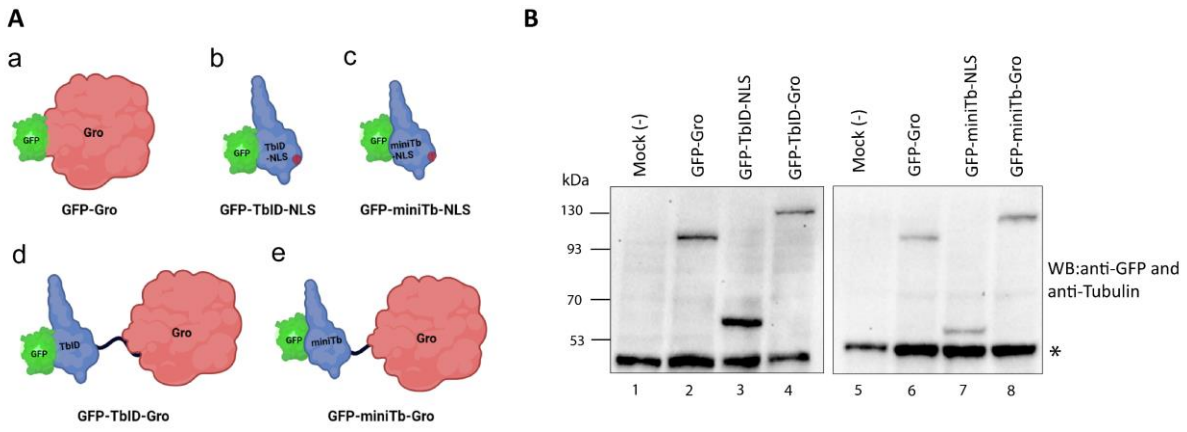


Figure 40: Schematic representation of the designed fusion proteins, and their expression in Kc167 *Drosophila* cells.

(A) Schematic illustration of the designed fusion proteins used as controls: **(a)** GFP-Gro **(b)** GFP-TbID-NLS, **(c)** GFP-miniTb-NLS, and as baits: **(d)** GFP-TbID-NLS-Gro, and **(e)** GFP-miniTb-Gro. Magenta dot in b and c represent the NLS. **(B)** Western blotting (WB) analysis of total cell lysates from untransfected cells (lane 1 and 5), and cells expressing Gro-GFP (lanes 2 and 6), GFP-TbID-NLS (lane 3), GFP-TbID-Gro (lane 4), GFP-miniTb-NLS (lane 7), and GFP-miniTb-Gro (lane 8). Western blotting analysis was performed using anti-GFP and anti-tubulin antibodies.

The asterisk on the right indicates bands belonging to tubulin. Kc167 cells were transfected by transient transfection. Protein extracts were collected three days after transient transfection. Illustration in A created with BioRender.com.

Next, I examined if the biotin ligases TurboID and miniTurbo were functional when bound to Gro in the bait fusion proteins, as well as the controls. It is known that tagging a bait with a promiscuous biotin ligase induces extensive self-biotinylation of the bait (Branon et al., 2018; Roux et al., 2012). Thus, I analysed whether the self-biotinylation of the bait occurred by streptavidin blot. It is important to note that the initial experiments were performed using two conditions in parallel, with and without the addition of exogenous biotin. However, no difference was detected between the two conditions, indicating that there is enough biotin in *Drosophila* cell culture for the biotin ligases to function, in contrast to what it was reported in other cell types (Branon et al., 2018). In consequence, we proceed to perform the experiments without adding exogenous biotin.

Figure 41 shows a streptavidin blot from total protein lysates extracted three days after transient transfection, showing the self-biotinylation of the GFP-TbID-Gro bait (arrowhead in lane 4). There were additional bands (black background in Figure 41 lane 4) that were not observed in the controls which did not express the TurboID biotin ligase (lanes 1 and 2). Importantly, those bands belong to proteins biotinylated by the bait, that are potential Gro interactors. Very similar results were observed with the GFP-miniTb-Gro bait (Figure 41, lane 8). Regarding the control fusion proteins, the self-biotinylated GFP-TbID-NLS protein was detected in the streptavidin blot (arrowhead in Figure 41 lane 3), together with additional biotinylated proteins. However, we were not able to detect the self-biotinylated GFP-miniTb-NLS protein (Figure 41, lane 7), and hence we were not sure the miniTurbo biotin ligase fused to NLS was functioning properly. Since the TurboID control construct seemed to work better, we decided to use the TurboID bait (GFP-TbID-Gro) and control (GFP-TbID-NLS) in the following experiments. It should be noted that there were two bands present in all protein extracts (asterisks in Figure 41); including the control without biotin ligase (GFP-Gro). Those bands are endogenously biotinylated proteins that are likely to be *Drosophila* carboxylases, as they were also found in similar studies in *Drosophila* (Ramirez et al., 2015).

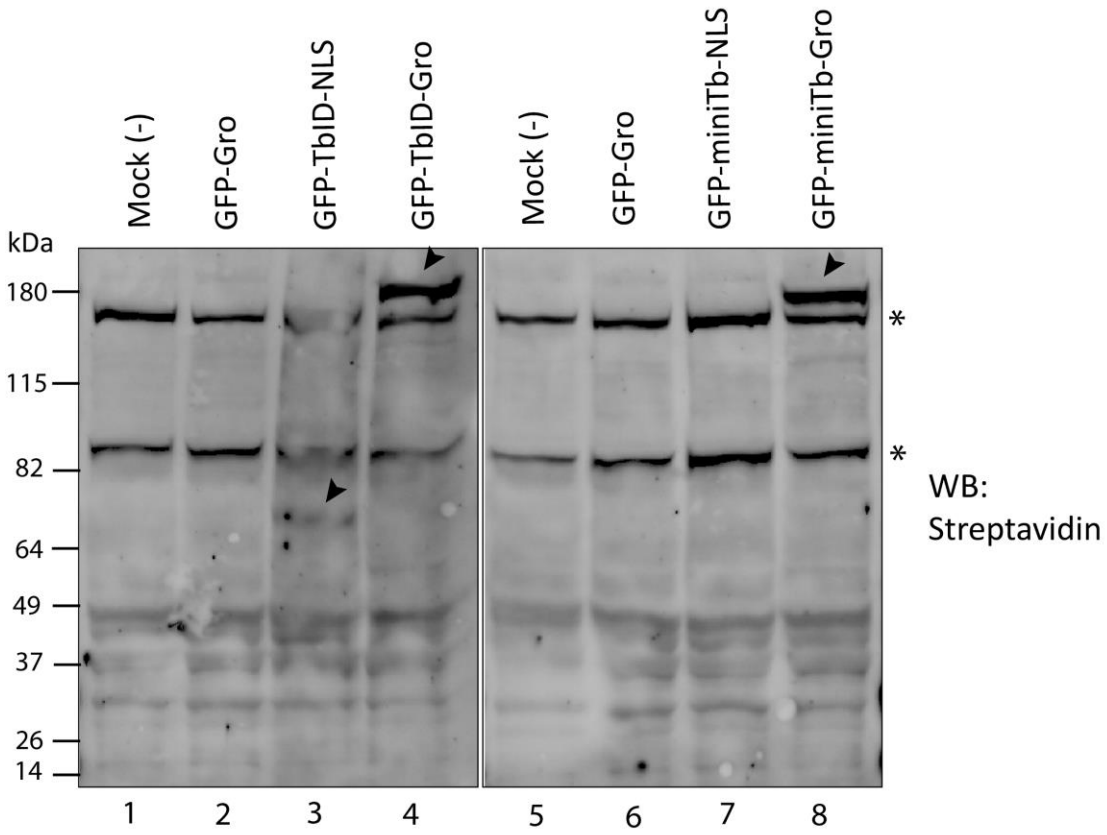


Figure 41: Self-biotinylation of the bait (GFP-TbID-Gro) and the control (GFP-TbID-NLS) fusion proteins by the TurboID and miniTurbo biotin ligases.

Western blotting analysis of total protein lysates from untransfected cells (lane 1 and 5), cells expressing GFP-Gro (lane 2 and 6), GFP-TbID-NLS (lane 3), GFP-TbID-Gro (lane 4), GFP-miniTb-NLS (lane 7), and GFP-miniTb-Gro (lane 8). Western blotting (WB) analysis was performed using a fluorescently labelled streptavidin. Asterisks (right) indicate the presence of endogenously biotinylated proteins. Arrowheads indicate self-biotinylated GFP-TbID-Gro (lane 4), GFP-miniTb-Gro (lane 8), and GFP-TbID-NLS (lane 3). Kc167 cells were transfected by transient transfection and incubated in absence of exogenous biotin. Protein extracts were collected three days after transient transfection.

3.3.2.2 Optimisation of the expression and harvesting of biotinylated proteins using TurboID bait in *Drosophila* cells

As a result of the presence of biotin in the *Drosophila* media, we could not control the labelling time window by adding exogenous biotin, as it would have been desirable. Thus, we proceed to optimize the expression conditions of the bait and control proteins, aiming to minimize the background due to unspecific biotinylation. Since the fusion proteins were expressed by transient transfections, we decided to explore the time point after transfection when the self-biotinylated bait starts to be detected, as well as other biotinylated proteins. Thus, I carried

out a time-course experiment, in which the total protein extracts were collected at different time points after transient transfection: 24 h, 48 h, and 72 h.

This experiment showed that the self-biotinylated bait (arrowheads in Figure 42) starts to be detected by streptavidin blot 48 hours after transfection, and the signal increases over time, observing a stronger signal 72 h after transfection. We decided to use the 48 h point for the experiment and hence extract and harvest the biotinylated proteins at this time point after transfection for the following experiments.

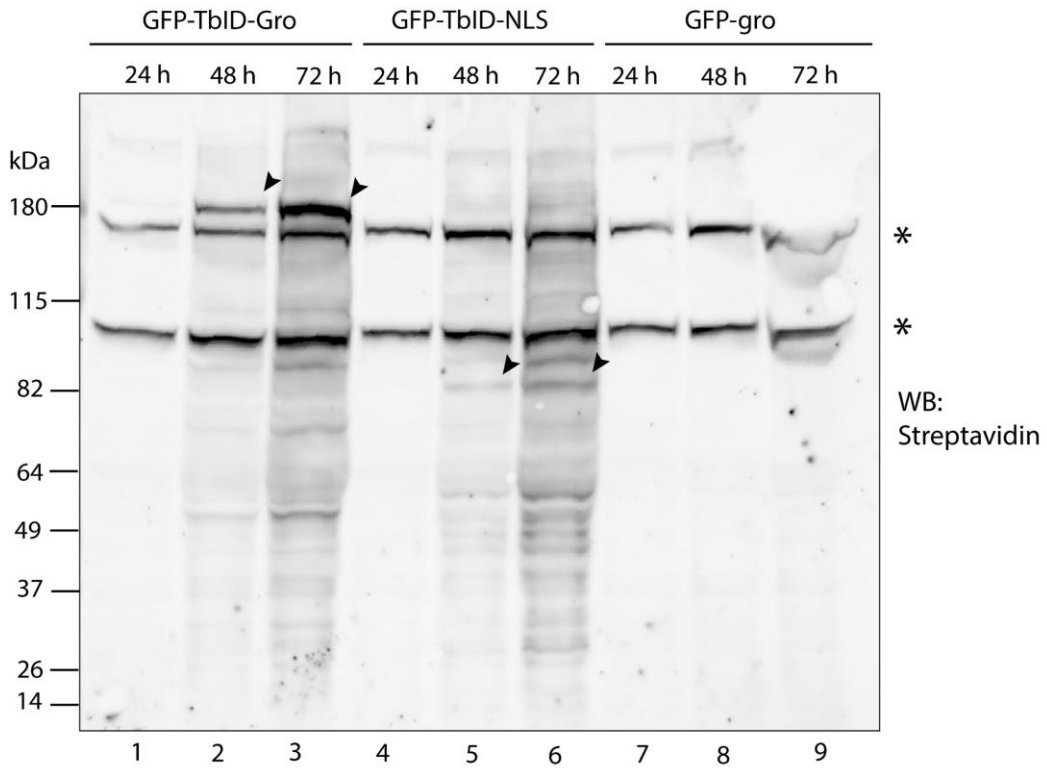


Figure 42: Biotinylated proteins in total protein lysates collected at different time points cells expressing the bait (GFP-TbID-Gro) and control (GFP-TbID-NLS) fusion proteins.

Western blotting analysis of total protein extracts from cells expressing GFP-TbID-Gro (lanes 1-3), GFP-TbID-NLS (lanes 4-6), and GFP-Gro (lanes 7-9) that were collected at the indicated time points. Kc167 cells were transfected by transient transfection, incubated in absence of exogenous biotin, and protein extracts were collected 24, 48 and 72 hours after transfection. Western blotting (WB) analysis was performed using a fluorescently labelled streptavidin. Asterisks (right) indicate the presence of endogenously biotinylated proteins. Arrowheads indicate self-biotinylated bait (lanes 2 and 3) and self-biotinylated GFP-TbID-NLS (lanes 5 and 6).

Subsequently, it was tested if the biotinylated proteins could be harvested by using streptavidin-coated beads, a procedure that is essential for the successful harvest of biotinylated proteins. Figure 44 A shows that the self-biotinylated GFP-TbID-Gro bait was

specifically recovered in the SA pulldown (lane 9), as it was the self-biotinylated control GFP-TbID-NLS (Figure 44 A, lane 6). In addition, the capture of the self-biotinylated bait and control proteins was notably efficient since there is a clear enrichment of those proteins after the streptavidin pulldown compared to the starting protein extract (Figure 44, compare lane 8 with 9, and lane 5 with 6).

I next checked the intracellular location of the bait and the control fusion proteins by immunostaining assays. The endogenous Gro protein is known to be broadly expressed in the nucleus, where it acts as a corepressor. Gro is imported to the nucleus as it carries a nuclear localization sequence, which is present in its CcN domain. Therefore, the fusion protein (GFP-TbID-Gro) is expected to be localized in the nucleus of cells to ensure the original function of Gro. Previous studies in *Drosophila* Kc167 cells in my lab demonstrated that the genome localization of Gro function was not altered by tagging it with GFP in the N-terminal (Kaul et al., 2014).

To explore the intracellular expression of the bait, I performed immunofluorescence microscopy in *Drosophila* Kc167 cells. In parallel, it was also checked the subcellular location of the control fusion protein. The resulting microscopy images showed that the bait (GFP-TbID-Gro) was expressed inside and outside the nucleus (Figure 43 F), and hence showed a different localization to the expected for endogenous Gro in Kc167 cells (Apidianakis et al., 2001). Differently, the GFP-TbID-NLS fusion protein showed a more specific localization in the nucleus (Figure 43 J).

We can hypothesize that a cause of the mislocalization of the bait could be a consequence of its overexpression (Roux et al., 2018). Nevertheless, the localization of endogenous Gro in *Drosophila* Kc167 has not been extensively explored and it would be interesting to characterize it better since the human Gro ortholog TLE1 have been shown to localize in the cytoplasm of cells following the activation of the EGF pathway (Zahavi et al., 2017).

Secondly, the immunofluorescence assay showed that the TurboID biotin ligase is functional when is fused to Gro (Figure 43 G), in agreement with the streptavidin blot results (Figure 40 and 41). There are biotinylated proteins in both, the cytoplasm and the nucleus of cells expressing the bait (Figure 43 G). In contrast, untransfected cells showed much fewer biotinylated proteins that were mostly located at the cytoplasm (Figure 43C), which is

consistent with the presence of endogenously biotinylated proteins previously observed in the streptavidin blot (asterisks in figures 41 and 42). In cells expressing the control proteins, GFP-TbID-NLS (Figure 43 K) a high signal of biotinylated proteins was detected in the nucleus of cells as expected.

Altogether the immunohistochemistry assays indicated that the TurboID biotin ligase acts efficiently biotinylating proximal proteins in both the bait and the control constructs in *Drosophila* culture cells. However, the presence of the bait in both the nucleus and the cytoplasm led us to consider the use of nuclear extracts instead of total lysates to reduce the cytoplasmic background.

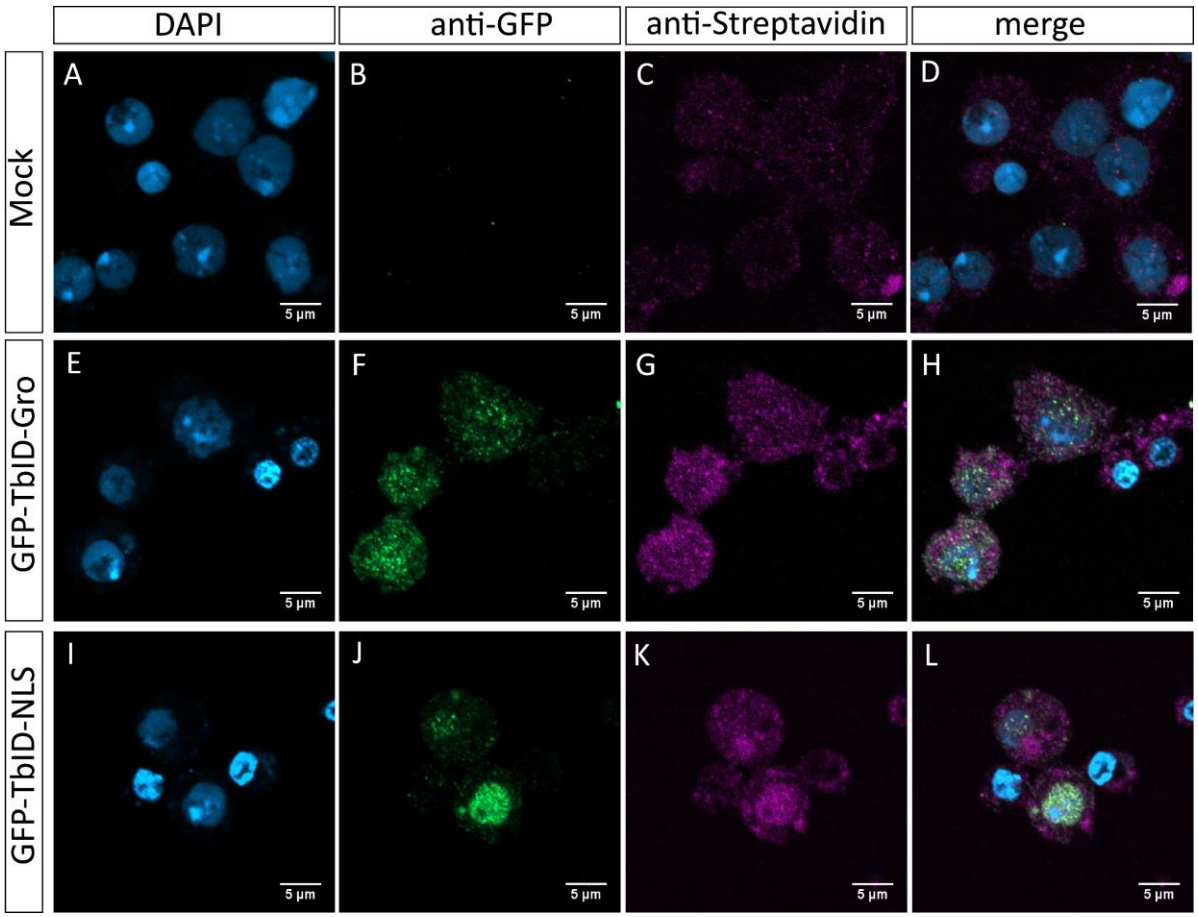


Figure 43: The bait fusion protein (GFP-TbID-Gro) is expressed in both the nucleus and the cytoplasm whereas the control fusion protein (GFP-TbID-NLS) is localized in the nucleus.

Intracellular location of the TurboID fusion proteins, and biotinylated proteins in **(A-D)** untransfected cells (mock), and in cells expressing the **(E-H)** bait; GFP-TbID-Gro, and the **(I-L)** control construct; GFP-TbID-NLS**.** GFP fusion proteins were detected with an anti-GFP antibody, biotinylated proteins were detected with fluorescently labelled streptavidin (violet), and DNA was labelled with DAPI (blue). Transiently transfected Kc167 cells were

transfected by transient transfection and incubated in absence of exogenous biotin. Protein extracts were collected 24 hours after transient transfection.

As a last step for the optimization process, we tested if it was viable to isolate the biotinylated proteins from nuclear protein extracts instead of the total lysates. This would allow us to reduce the protein background due to proteins in other subcellular localizations and simplify the posterior proteomics analysis. Thus, I validated a protocol in which the incubation and pulldown with the streptavidin beads To gain an understanding of this process were performed using the nuclear subcellular fraction instead of total lysates (see Materials and Methods). Figure 44 B shows the enrichment of the bait after the streptavidin pulldown from the nuclear fraction (lane 4) compared to the starting nuclear extract (lane 3). In addition, western blotting using an antibody against H3 (Figure 44 B) showed the presence of this protein in the total protein lysate (lane 1) and the nuclear fraction (lane 3) but not in the cytoplasmic fraction (lane 2), as expected, which validated the cellular fractionation protocol used.

Finally, silver staining (Figure 44 C) also revealed a general reduction of the number of proteins in the nuclear fraction (lane 3), and in the streptavidin pulldown (lane 4), compared to the total protein lysate (lane 1). Therefore, the isolation of the biotinylated proteins from the nuclear extracts seemed to provide cleaner samples and reduce the background generated by the cytoplasmic proteins.

Taken together all these data, we decided to carry out the final experiments to prepare the samples for MS analysis, using streptavidin pull down from nuclear fractions extracted two days after transient transfection. By following these conditions, we aimed to reduce the background of biotinylated proteins due to an excess exposition of proteins to the action of TurboID, as well as the presence of biotinylated proteins from the cytoplasm.

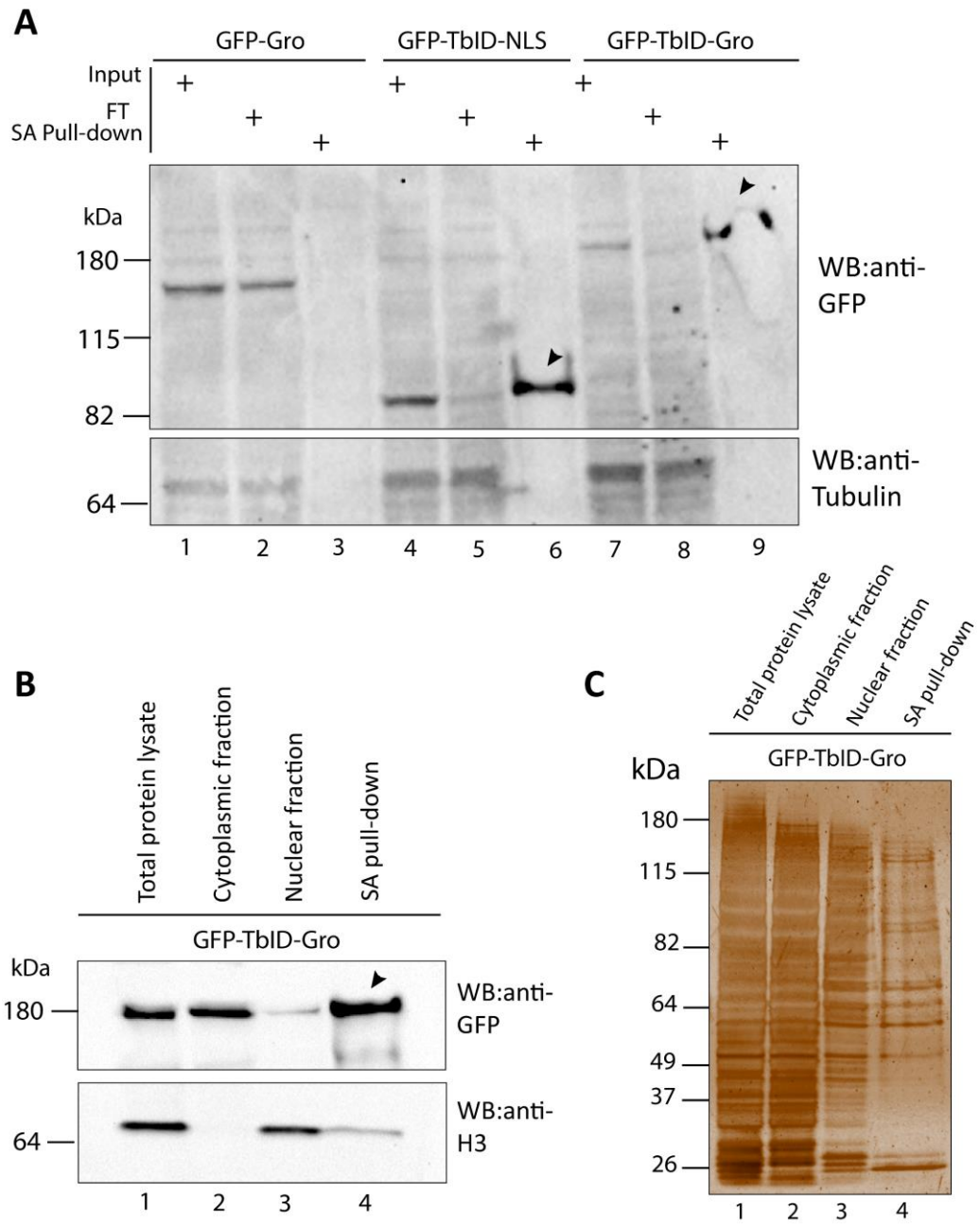


Figure 44: Comparison of the recovery of fusion proteins by streptavidin-coated beads from total protein lysates and from nuclear extracts.

(A) Western blotting analysis of total protein lysate, flow-through (FT), and biotinylated proteins recovered by streptavidin-coated beads (SA) (lanes 3, 6 and 9). Total protein lysates were prepared from cells expressing Gro-GFP (lanes 1-3), GFP-TbID-NLS (lanes 4-6) and GFP-TbID-Gro (lanes 7-9). Western blotting (WB) analysis was performed using anti-GFP (top) and anti-tubulin (bottom) antibodies. Arrowheads indicate self-biotinylated GFP-TbID-Gro (lane 9) and GFP-TbID-NLS (lane 6) recovered by SA-coated beads. **(B)** western blotting analysis of total cell lysate (lane 1), the cytoplasmic fraction (lane 2), nuclear fraction (lane 3), and biotinylated proteins recovered by SA-coated beads from the nuclear fraction (lane 4). WB was performed using anti-GFP (top) and anti-histone 3 (bottom) antibodies. **(C)** SDS-PAGE gel stained by silver staining showing total proteins in total cell lysate (lane 1), the cytoplasmic fraction (lane 2), nuclear fraction (lane 3), and biotinylated proteins recovered by streptavidin-coated beads from the nuclear fraction (lane 4). Transiently transfected Kc167 cells were incubated in absence of exogenous biotin, and protein extracts were collected 48 hours after transient transfection.

3.3.2.3 Production and pull-down of biotinylated proteins for LC-MS/MS analysis

Once the experimental conditions were optimized, we scaled up the protocol (see Materials and Methods) to generate enough biotinylated material to send out for MS analysis. Three biological replicates of two conditions were generated: the bait (GFP-TbID-Gro), and the control (GFP-TbID-NLS). To ensure that the replicates sent for MS were expressing the constructs, and containing enough material, a small sample of the bait and the control conditions was taken from each of the replicates to perform WB and silver staining analysis.

The western blot in Figure 45 A shows the enrichment of the GFP-TbID-Gro bait (lane 3), and the GFP-TbID-NLS control fusion proteins (lane 6) after streptavidin pull-down from nuclear extracts in the first replicate performed. This demonstrates the proper expression and isolation of the self-biotinylated bait and control fusion proteins. In addition, silver staining revealed the presence of a large number of biotinylated proteins after streptavidin pull-down in both the bait (Figure 45 B, lane 1) and the control (Figure 45 B, lane 3). It should be noted that there was also a background of endogenously biotinylated proteins recovered by streptavidin beads from untransfected cells (Figure 45 B, lane 2).

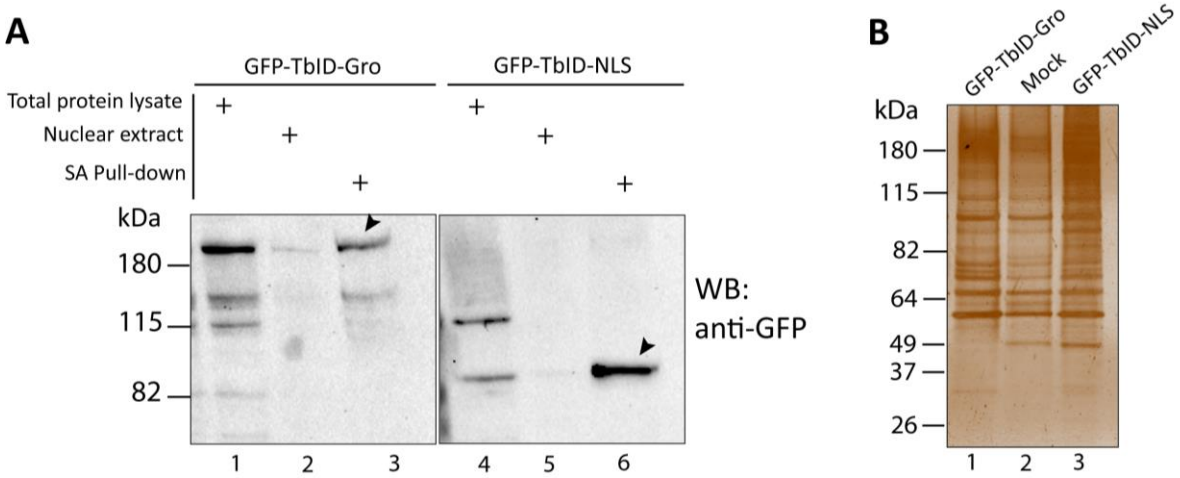


Figure 45: Quality controls for replicate 1 used for LC-MS/MS analysis.
(A) Western blotting analysis from total protein lysates, nuclear extracts, and biotinylated proteins recovered by streptavidin-coated beads (SA) from Kc167 cells expressing the GFP-TbID-Gro (lanes 1, 2 and 3), and GFP-TbID-NLS (lanes 4, 5 and 6) fusion proteins. Western blotting (WB) analysis was performed using an anti-GFP antibody. Arrowheads indicate self-biotinylated bait (lane 3) and GFP-TbID-NLS (lane 6) recovered by SA-coated beads. **(B)** SDS-PAGE gel stained by silver staining showing total biotinylated proteins recovered by streptavidin-coated beads from the nuclear fractions from untransfected cells (Mock; lane 2), cells transfected with GFP-TbID-Gro

(lane 1), and with GFP-TbID-NLS (lane 3). Transiently transfected Kc167 cells were incubated in absence of exogenous biotin, and protein extracts were collected 48 hours after transient transfection.

After the western blotting and the silver staining analysis performed on replicate one demonstrated that both, the experimental conditions and scale-up were successful, we decided to proceed and perform two more biological replicates. A subsequent identical analysis was carried out on replicate two and three, which are shown in Figure 46. In both replicates, the western blotting analysis revealed the rescue of the self-biotinylated bait (Figure 46 A, lanes 1 and 4), and control fusion proteins (Figure 46 A, lanes 3 and 6) after streptavidin pull-down. Silver staining in Figure 46 B also showed a large number of biotinylated proteins isolated after SA pull down for the bait (lanes 1 and 4) and the control in both replicates (lanes 3 and 5). It should be noted that the western blot of replicate three showed a higher signal in the self-biotinylated control (Figure 46 A, lane 6) than in the bait (Figure 46 B lane 4), indicating that the amount of the control fusion protein was larger than the bait in this replicate. In addition, the silver staining on replicate three also showed that amount of total biotinylated proteins recovered by streptavidin beads was also higher in the control sample than in the bait (Figure 46 B, lanes 4 and 5, respectively). Despite this difference observed in the expression and activity between the bait and the control fusion proteins in replicate 3 we decided to proceed and send this replicate for MS analysis. We considered generating a fourth replicate, however, due to a lack of time we decided to proceed with what we had and sent these three replicates for mass-spectrometry analysis.

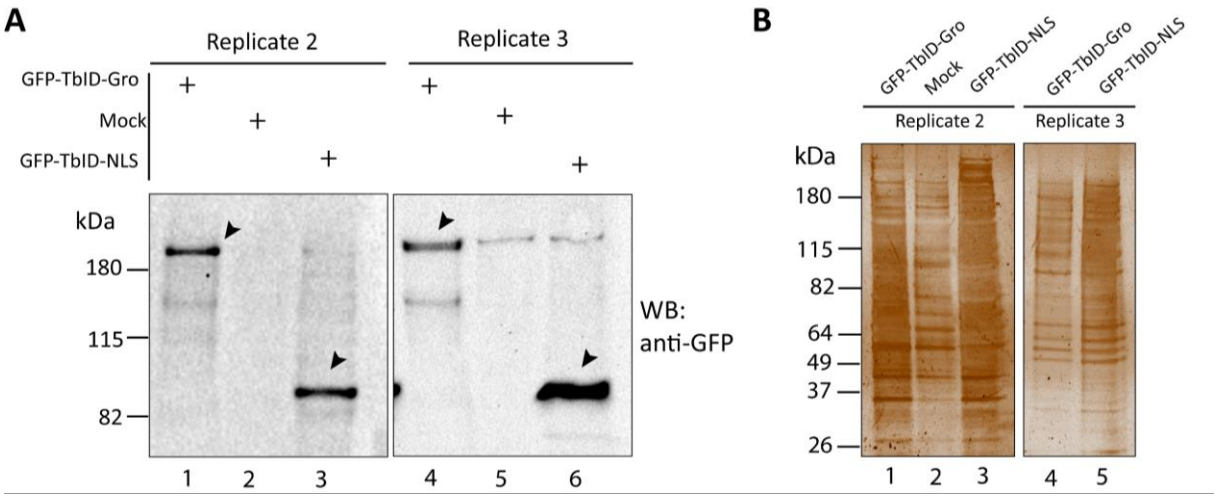


Figure 46: Quality controls for replicates 2 and 3 used for LC-MS/MS analysis.

(A) Western blotting analysis of biotinylated proteins recovered by streptavidin-coated beads (SA) from nuclear extracts of Kc167 cells expressing the GFP-TbID-Gro (lanes 1, and 4), and GFP-TbID-NLS (lanes 3 and 6) fusion

proteins, as well as untransfected cells (Mock; lanes 2 and 5). Panel on the right shows the samples from replicate 3 and panel on the left replicate 2. Western blot analysis was performed using anti-GFP antibody. Arrowheads indicate self-biotinylated bait and GFP-TbID-NLS fusion proteins recovered by SA-coated beads. **(B)** SDS-PAGE gel stained by silver staining showing total biotinylated proteins recovered by SA-coated beads from the nuclear fractions from untransfected cells (Mock; lane 2), cells transfected with GFP-TbID-Gro (lane 1 and 4), and with GFP-TbID-NLS (lane 3 and 5). Panel on the right shows the samples from replicate 3 (mock sample is not shown due to the sample being lost) and panel on the left replicate 2. Transfected Kc167 cells were incubated in absence of exogenous biotin, and protein extracts were collected 48 hours after transient transfection.

3.3.2.4 Analysis of proteomics data

The biological replicates for the control group (GFP-TbID-NLS), and the bait group (GFP-TbID-Gro) were sent to the Advanced Mass Spectrometry Facility at the University of Birmingham (https://www.birmingham.ac.uk/facilities/advanced-mass-spectrometry/index.aspx). The samples were digested into peptides that were separated and identified by LC-MS/MS. The raw data were subsequently analysed by the facility using the MaxQuant proteomics software, a widely used label-free quantification (LFQ) software (Cox et al., 2014). The output files from MaxQuant ("proteinGroups.txt") that contain information on the identified protein groups were then sent to me for further analysis.

I employed the Perseus software (Tyanova et al., 2016) to analyse the proteomics data as described in Materials and Methods. The number of proteins identified in the bait replicates (GFP-TbID-Gro) were compared among each other and visualized using a Venn diagram, which showed a high overlap among them (Figure 47 B). Venn diagrams showing the overlap between the proteins detected in the bait and the control in each replicate are shown in Figure 58 in the Appendix. Replicate 3 contained a higher number of identified proteins that did not overlap with the other two replicates (424) (Figure 47 B). The number of proteins identified in the three control samples (GFP-TbID-NLS) also showed a high overlap among them (Figure 47 A).

In a first approach to visualize the spread of the data for each sample, the LFQ intensities of the control and the bait replicates were plotted in logarithmic histograms, which showed a normal distribution (Figure 47 C and D, respectively). Importantly, LFQ intensity values were observed for Gro in replicates 1 and 2 of the bait but it was not observed in any of the control replicates (red in Figure 47 D compared to C). We expected to find biotinylated Gro in cells

expressing GFP-TbID-Gro due to self-biotinylation of the bait, but also as Gro is known to interact closely with itself to form tetramers (Jennings et al., 2008; Song et al., 2004).

A large number of missing values were observed in all datasets, which is common in label-free proteomics but can compromise the completeness of data (Albrecht et al., 2010; Karpievitch et al., 2009). To avoid mathematical difficulties, missing values are frequently replaced with random numbers, which are drawn from the dataset's empirical distribution, a process known as "imputation" (Tyanova et al., 2016). Thus, I proceed to impute the missing values in the Perseus platform by using that method, however, this approach failed due to the large number of missing values imputed, which converted the datasets into a bimodal distribution (see histograms in Figure 59 in the Appendix).

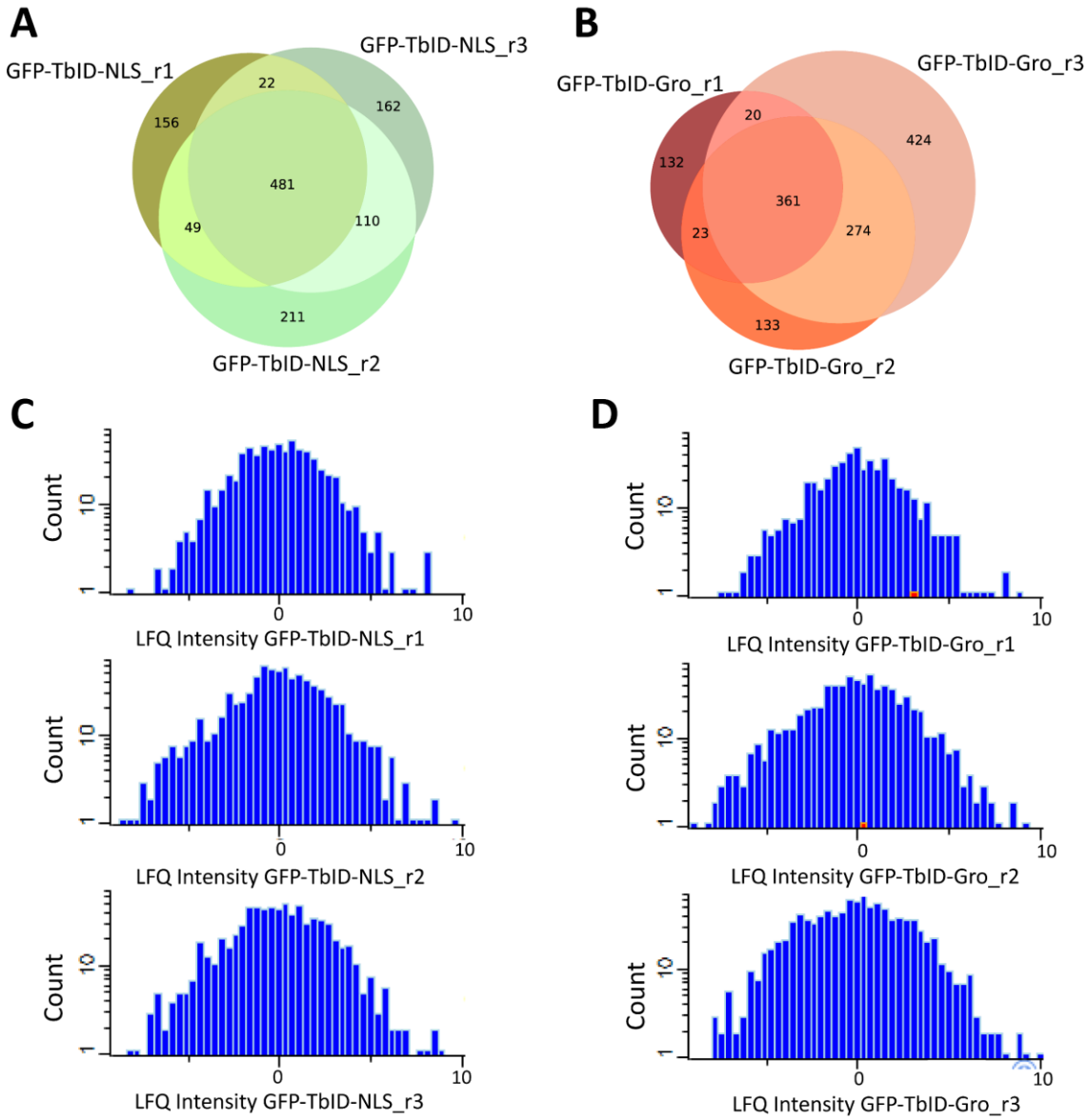


Figure 47: Reproducibility of the proteomics datasets among replicates, and distribution of LFQ intensity values

(A-B) Venn diagrams illustrating the overlap among the proteins identified in the three control replicates (GFP-TbID-NLS) and in the three bait (GFP-TbID-Gro) replicates. **(C-D)** Histograms showing the logarithmic distribution of the LFQ intensity datasets (blue) for each of the three control replicates (GFP-TbID-NLS) and the bait replicates (GFP-TbID-Gro). Data were normalized by subtracting the median. The contribution of Gro in each dataset is coloured in red.

As an alternative, I imputed the missing values by a constant number, which was one unit lower than the lowest LFQ intensity found among all datasets analysed. This approach worked a bit better than the first one as it did not change the distribution of the datasets to binomial distribution, and the imputed LFQ intensity values for Gro were very low (see histograms in Figure 60 in the Appendix). Thus, the missing values were imputed by using that approach and the resulting datasets were subsequently used to assess whether there were significant

differences in the proteome of the bait compared to the control. These significant changes were visualized in a volcano plot in which the Log_2 of the mean fold change (Log_2 LFQ intensity Bait – Log2 LFQ intensity Control) for each protein was plotted against the Log_{10} of the p-value (calculated by T-tests). There was no protein detected with significant enrichment in the bait proteome compared to the control proteome, either using a cut-off threshold of 0.05 or 0.1 (brown and blue dashed lines in Figure 48 A, respectively). However, the volcano plot showed that Gro protein was highly enriched in the bait proteome, which indicates that the assay worked properly. There were other proteins in the volcano plot that showed similar enrichment and p-values to Gro and hence were further analysed. These proteins were either membrane, cytoplasmic or nuclear proteins. As Gro is a nuclear protein, the analysis only focused on the three nuclear proteins (highlighted in blue in Figure 48 A): SR family splicing factor SC35 (SC35), TAR DNA-binding protein-43 homolog (TBPH), and noisette (noi). Interestingly, these three proteins are involved in mRNA alternative splicing (Bradley et al., 2015; Meyer et al., 1998; Romano et al., 2021). This result is reminiscent of the findings from (Kwong et al., 2015), who recovered splicing factors when they used the central region of Gro to pull down interacting proteins.

As no significant change was detected in the bait proteome compared to the control, I next proceed to analyse the average of the LFQ intensity values calculated from the replicates for the control and bait groups. First, the overlap between the proteins detected in the bait and the control groups were plotted in a Venn diagram to assess the number of overlaps (Figure 48 B). There were 1726 total proteins identified, of which 910 overlapped between both conditions, 391 were only detected in the control group, and 425 in the bait group.

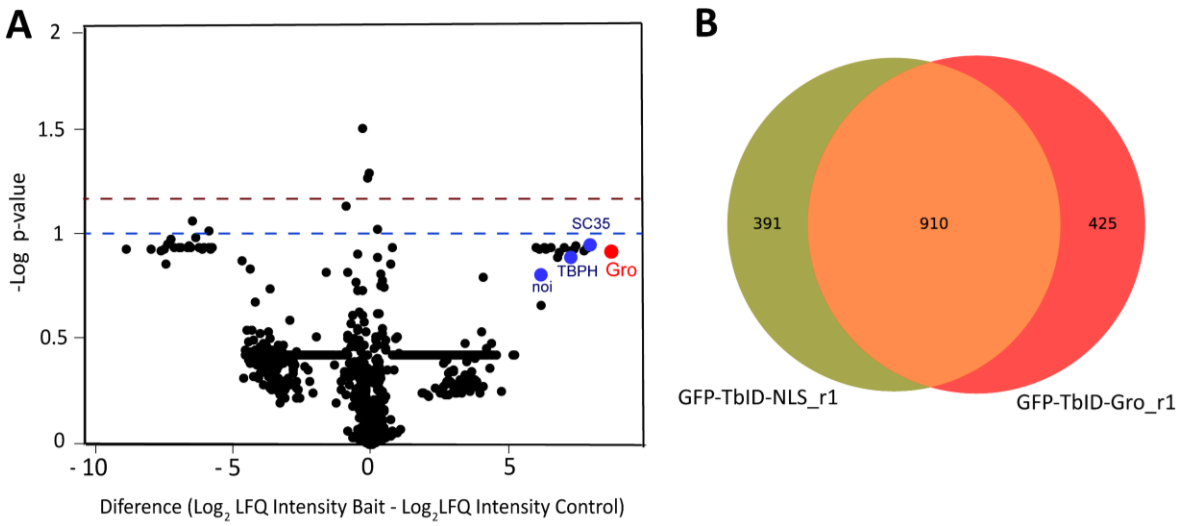


Figure 48: Fold change of proteins identified and overlap between the total number of proteins identified in the bait (GFP-TbID-Gro) and control (GFP-TbID-NLS) groups.

(A) Volcano plot showing the Log_2 of the mean fold change for each protein versus the $-log_{10}$ of P-values obtained from pairwise calculations performed using T-tests. Gro protein is highlighted in red, and nuclear proteins close to Gro in the plot are highlighted in blue. Brown and blue dashed lines mark a p-value = 0.05, and 0.1, respectively. **(B)** Venn diagram illustrating the overlap between the total proteins identified in the control (GFP-TbID-NLS) and bait (GFP-TbID-Gro) groups.

Next, the 425 proteins that were exclusively detected in the bait group were ranked according to their total spectral count, which is defined as the total number of spectra that map to peptides of a given protein. Spectral counting is frequently used as a semiquantitative measure of protein abundance in label-free proteomics [reviewed in (Lundgren et al., 2010)]. Importantly, from the 425 proteins found exclusively in the bait group, Gro ranked the first [Total Spectral Count (TSC) = 17], which was used as a positive control. Although not many Gro interactors were detected among these 425 proteins, two were: Alhambra (TSC=2), which is a TF required for maintaining *eve* expression and normal growth (Bahri et al., 2001), and the snRNP-U1-C protein; a member of the U1 snRNP splicing complex. Interestingly, the TF Lola, which motif was found in BG3 cells (section 3.2.1.2 of the results) was also detected in the list (TSC=5). The full list of proteins detected in the bait proteome and not in the control proteome is shown Table 10 in the Appendix.

The list was next filtered by the total number of spectral counts (cut-off TSC ≥ 6) to analyse the top hits of the list. 54 of the 425 proteins had a TSC greater or equal than 6 from which 20 were expressed in the nucleus (list shown in Table 4) (proteins expressed in other parts of the cell excluded from this table). Although not previously identified Gro interactors were present in the filtered list (aside from Gro), it is of note that six of the proteins were involved

in alternative splicing and mRNA 3'-end processing: SC35, Su(f), Ago2, Nipped-A, IntS11, and BCAS2. Moreover, SC35 reported a similar value of enrichment and p-value to Gro in the volcano plot shown in Figure 48.

Protein name	Total Spectral Counts (TSC)
Groucho	15
Ribosomal RNA processing 40 (Rrp40)	10
SR family splicing factor SC35 (SC35)	9
Cdc23	8
Heat-shock-protein-70 (Hsp70)	8
Arginine methyltransferase 1 (Art1)	7
CG8878	7
suppressor of forked [Su(f)]	7
CG11123	7
BCAS2	7
Twine (twe)	6
Argonaute 2 (Ago2)	6
CG3919	6
flightless I (flil)	6
Nipped-A	6
Integrator 11 (IntS11)	6
Retinoblastoma binding protein 5 (Rbbp5)	6
Dribble (dbe)	6
CG13928	6

Table 4: Proteins expressed in the nucleus with total spectral counts equal to or greater than 6.
Table showing proteins that were exclusively identified in the GFP-TbID-Gro sample after filtering by cellular location (only proteins expressed in the nucleus were selected), and by the total number of spectral counts (cut-off TSC ≥ 6)

In summary, despite Gro being successfully detected in two of the bait replicates and absent from the three control replicates, no significant differences were found in the bait proteome compared to the control by the statistical analysis described in this section. This may be a result of the imputation of a large number of missing values, which can alter significantly the results of the proteomics analysis (Medo et al., 2019). There are alternative approaches that are also used to identify significant changes between proteomes, which do not require the imputation of missing values, including the Significance Analysis of INTeractome (SAINT) (Choi et al., 2011), and ProtRank (Medo et al., 2019). Thus, If I had more time, I would have applied these approaches as an alternative approach.

3.3.3 Discussion

This chapter described the two high-throughput approaches that I used to detect novel Gro interacting partners. In both methods, the full-length Gro protein was used as a bait to detect proteins that interact with Gro directly (the Y2H and the proximity labelling assay), or indirectly (only proximity labelling assay). In the case of the Y2H, no protein-protein interactions with Gro were found, despite the observation that the controls gave the expected results, and no technical issues were detected with the screening process. The causes behind the unfruitful Y2H screen remain unknown, however, one possible cause that might explain it is that the repressive nature of Gro had interfered with the yeast transcriptional machinery. Another possibility is that there was an unidentified problem with the *Drosophila* library used. It is worth noting that the C-terminal domain of the *Drosophila* Gro protein was used as bait in a previous Y2H study, which successfully detected novel interactors (Goldstein et al., 2005).

On the other hand, the TurboID-proximity labelling assay also resulted unsuccessful in the detection of statistically significant Gro interacting partners. However, the controls performed during the performance of the assay, as well as the posterior proteomics analysis of the MS data indicated that the assay worked properly. Indeed, much higher levels of Gro protein were detected in the bait proteome than in the control. Furthermore, the proteomics analysis revealed three nuclear proteins (SC35, TBPH, and noi) with similar expression profiles to Gro. These proteins are involved in mRNA alternative splicing, and it would be interesting to validate and explore further their relationship with Gro in future studies.

Several causes might have influenced the proximity labelling assay leading to limited and not significant results, including the expression of a large bait fusion protein (GFP-TbID-Gro) which could have altered Gro function or targeting. Another limitation of this assay is the large radius of biotinylation [estimated to be approximately 10 nm (Kim et al., 2014)] that may have led to the biotinylation of proteins that do not work within the bait radius. It is also possible that the lack of accessible primary amines for biotinylation in the candidate interactions may have generated false negatives (Roux et al., 2018). In the future, experiments using a split-TurboID integrated into the genome to form a stable cell line may overcome these limitations. Split-TurboID is based on the splitting of the biotin ligase into two inactive fragments that are

brought together by a drug, which would reduce the non-specific biotinylation of proteins found outside of the nucleus by the promiscuous biotin ligase (Cho et al., 2020)

3.4 Investigating the role of the Gro in Notch-dependent tumorigenesis

Previous work has revealed a close association between Gro and GAF recruitment to the genome [(Kaul et al., 2014), and section 3.2.3.1 of the results], and genetic interactions between *gro* and *Trl* indicate they operate together *in vivo* (see section 3.1.1.2 of the results). GAF associates with GAGA motifs, which is involved with multiple mechanisms that influence gene expression, including regulation of promoter-proximal pausing (Fuda et al., 2015), and chromatin remodelling (Shimojima, 2003; Tsukiyama et al., 1995). These motifs are commonly found in Gro ChIP-seq peaks in Kc167, S2R+, and BG3 cell lines [(Kaul et al., 2015), and section 3.2.1 of this work].

In addition to GAF, there is another TF that binds to GAGA motifs; Pipsqueak (Psq) (Lehmann et al., 1998). Psq physically binds to GAF, and they have been found to co-localize at many loci on polytene interphase chromosomes (Schwendemann and Lehmann, 2002). Furthermore, genetic studies indicate that these factors cooperate to regulate the transcription of homeotic genes (Schwendemann and Lehmann, 2002). Importantly, it has been shown that the overexpression of Psq in combination with Notch signalling promotes aberrant tissue growth and tumorigenesis (Ferres-Marco et al., 2006)

As described in section 1.2.1 of the Introduction, the role of Gro has been well described in the context of Notch signalling. This pathway is essential during development and in adult life and has pleiotropic roles in mammals and flies, that include promoting cell proliferation and regulating cell fate via lateral inhibition, among other mechanisms [reviewed in (Sjöqvist and Andersson, 2019)]. Notch signalling can have oncogenic or tumour suppressor potential in distinct human cancers, or even both roles within the same tumour type [reviewed in (Andersson et al., 2011)]. In *Drosophila* Notch plays similar dual oncogenic and tumour suppressor roles [reviewed in (Dominguez, 2014; Lobry et al., 2014)].

Psq encodes different isoforms containing or lacking the BTB protein-protein binding domain (referred to as BTB-Psq and non-BTB-Psq, respectively). Previously, a study from Maria Dominguez's lab showed that the overexpression of the Notch ligand *Delta* in combination

with the deregulation of the TFs *Longitudinal lacking* (*Lola*) and *Psq* led to the formation of tumours in the *Drosophila* eye (phenotype that was called '*eyeful*'). However, no tumours were developed by the single overexpression of *Delta* or the single deregulation of *Lola* and/or *Psq*. The authors also found a strong reduction in the tumour phenotype when the dose of Rpd3, Su(var)3-9, or E(z) was reduced in half, showing the crosstalk between the Notch pathway and different epigenetic regulators (Ferres-Marco et al., 2006). At the time, it was unclear whether tumorigenesis was promoted by BTB-Psq or non-BTB-Psq, or both. Recently, it has been shown that distinct Psq isoforms had distinct oncogenic activities and bind to distinct DNA motifs genome-wide (Gutierrez-Perez et al., 2019). Specifically, the authors found that the overexpression of *Delta* together with the BTB-containing *Psq* isoform promoted overgrowth and tumorigenesis in the developing eye, whereas the co-expression of *Delta* with non-BTB Psq did not (Gutierrez-Perez et al., 2019). However, the molecular mechanism that links Psq to the epigenetic regulators and Notch signalling is not fully understood.

The frequent presence of Gro at GAGA motifs, as well as the strong molecular and genetic interaction observed between Gro and GAF motivated us to further investigate and test the possible relationship between Gro and Psq. The work shown in this section was performed in Maria Domínguez's lab (Neuroscience Institute of Alicante), thanks to an EMBO short-term fellowship. For this project, I used genetic assays *in vivo*, in addition to bioinformatics analysis of ChIP-seq data.

For the *in vivo* assays, the developing *Drosophila* eye was used as a model to study growth and tumour formation. It is well-established that Notch signalling is crucial in the process of eye development to promote cell proliferation and differentiation. The fly eye starts to develop from the eye-antennal imaginal disc of the second instar larvae. During this stage of eye development, the EGF and the Notch pathways act antagonistically promoting antenna and eye fate, respectively. Later in the second instar larval stage, the organizer midline cells express the Notch receptor, which establishes the dorso-ventral (D-V) polarity of the eye disc. Subsequently, Notch is activated by its ligands Delta and Serrate on the dorsal and ventral halves, respectively. At this stage, the glycosyltransferase Fringe is expressed on the ventral side, where suppresses Notch activation by Serrate everywhere except for the dorso-ventral boundary. In the late third instar stage, the differentiation wave, known as morphogenetic

furrow (MF) crosses the eye disc from the posterior to the anterior. The MF patterns proliferating and undifferentiated precursor cells into photoreceptors and accessory cells, a process that is precisely regulated by Notch signalling. This is followed by several differentiation and organization stages that eventually give rise to the adult eye composed of ~800 ommatidia [reviewed by (Cagan, 2009)].

I first tested the dysregulation of Gro levels in combination with a fly line that overexpresses *Delta* in the proliferating eye. This experiment aimed to understand better the molecular mechanisms underlying the strong effects of *gro* knockdown and overexpression on growth that are shown in section 3.1.2.1 of this work. The *eyeful* tumour model in the developing *Drosophila* eye (Ferres-Marco et al., 2006) was additionally used to explore the role of Gro in Notch-dependent tumorigenesis. This well-established *Drosophila* tumour model induces tumorigenesis in the eye by ectopically expressing the Notch ligand Delta and the P-element insertion *GS88A8*, which deregulates the epigenetic regulators *pipsqueak (Psq)* and *lola* (Ferres-Marco et al., 2006).

For the bioinformatics analysis, I used ChIP-seq datasets of BTB-Psq and all Psq isoforms (Gutierrez-Perez et al., 2019) to test their co-localization with Gro. Additionally, it was tested the co-localization of Gro with different proteins known to overlap with either BTB-Psq or non-BTB-Psq through the genome.

3.4.1 Gro function is essential in the Notch pathway to control growth during eye development

I first explored the relationship between Gro and the Notch ligand Delta during *Drosophila* eye development. To this aim, I used *ey-GAL4* to drive the overexpression of *Delta* together with the RNAi downregulation and overexpression of *gro*. The effects of the single *gro* RNAi knockdown (Figure 49 A b) and overexpression (Figure 49 A c) in the proliferating eye have been already described in section 3.1.2.1 of the results. The single overexpression of *Delta* during eye development is known to induce eye overgrowth that results in a 'large eye' phenotype (Ferres-Marco et al., 2006). As expected, *Delta* overexpression driven by *ey-GAL4* resulted in larger eye imaginal discs (Figure 49 B b), which gave rise to larger eyes (Figure 49 A d), compared to the control (Figure 49 B a and Figure 49 A a).

The simultaneous overexpression of *Delta* and RNAi downregulation of *gro* (*ey-GAL4 > Dl > gro*^{RNAi}) led to an enhancement of the 'large eye' phenotype in all the eyes scored (Figure 49 A e), that was also observed in eye imaginal discs (Figure 49 B c). Importantly, it was observed that in some of the double mutant flies the overgrowth in the eye lead to tumours that were evidenced by the appearance of tissue folds in the adult eye (Figure 50 B). In consequence, we decided to quantify the number of adult eyes that developed tumours *versus* the number of eyes that did not. To simply the quantification, the adult eye phenotypes were classified into three categories: eyes with no tumours (Figure 50 A), with tumours (Figure 50 B), and loss of eye (Figure 50 C). It was observed that around 7.6 % of the double mutant eyes (*ey-GAL4 > Dl > gro*^{RNAi}) scored developed eye tumours (asterisk in Figure 50 B). In contrast, none of the eyes scored of the single *gro* knockdown or *Delta* overexpression showed tumours (Figure 50).

Simultaneous *gro* and *Delta* overexpression in the proliferating eye (*ey-GAL4 > Dl > gro*) resulted in an adult eye smaller than the control (Figure 49 A f), and hence in an obvious rescue of the 'large eye' phenotype. In addition, quantification of eyes with this genotype showed that only 8.7% presented eye loss compared to the 37% eye loss observed in the single overexpression *gro* eyes (Figure 50 C). Thus, the phenotypic effects on eye growth caused by the single overexpression of *gro* and *Delta* seem to cancel each other out when expressed concurrently, indicating that these genes play opposing roles during eye development.

Together this data highlights the essential function of *Drosophila* Gro in the context of the Notch pathway for normal growth during eye development. In addition, it suggests that Gro plays a role in Notch-dependent tumorigenesis in the *Drosophila* eye.

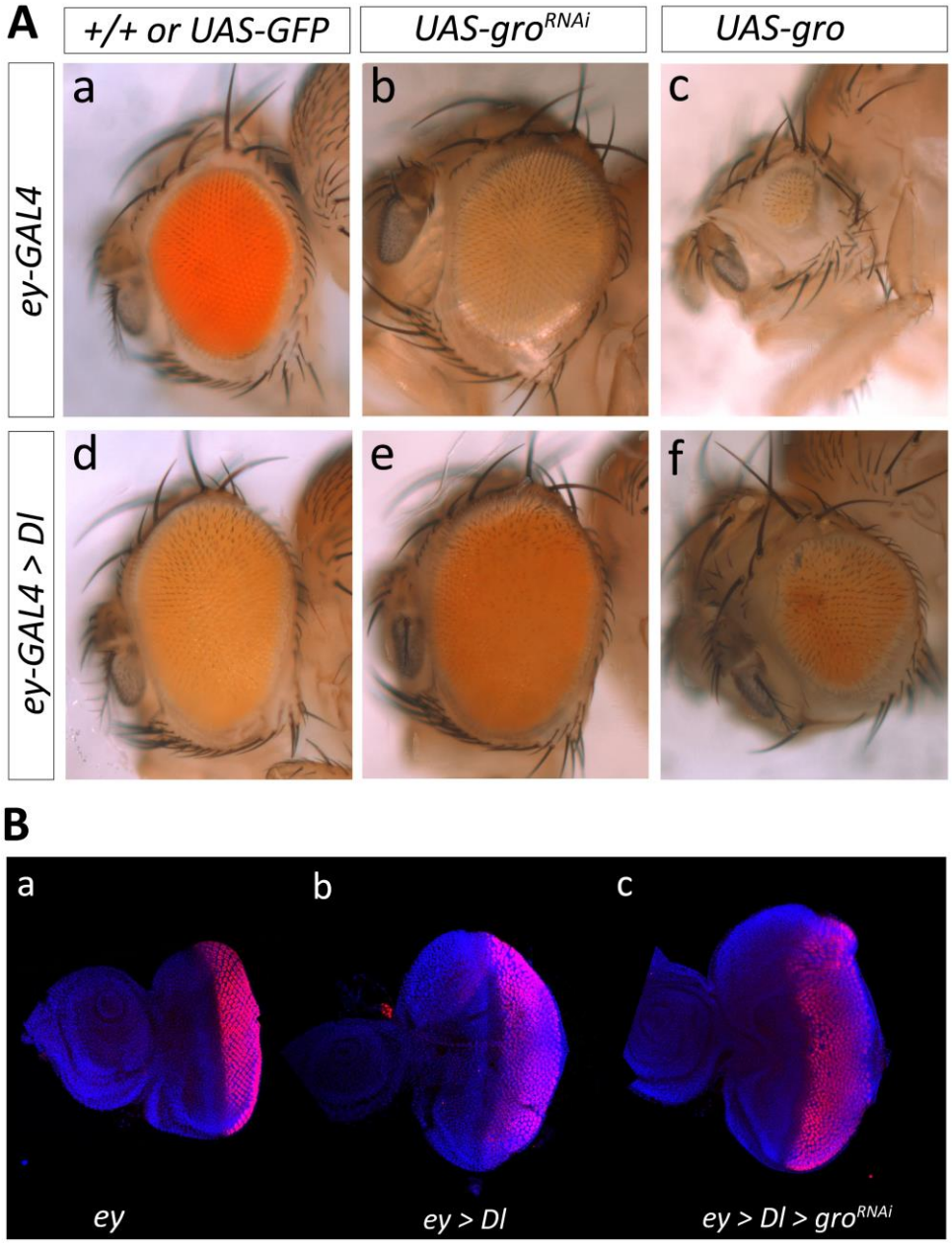
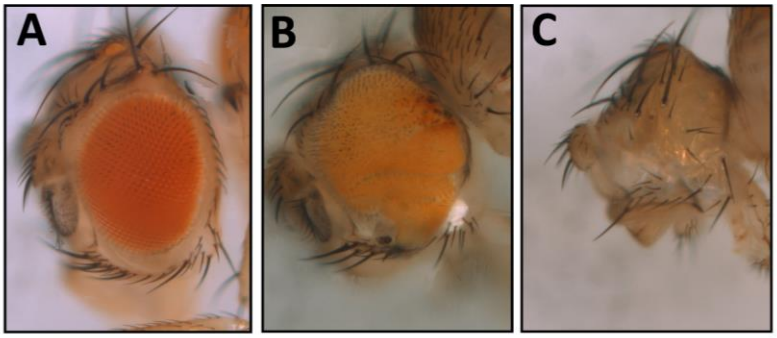


Figure 49: Simultaneous overexpression of *Delta* and RNAi downregulation of *gro* during eye development leads to enhancement of the 'large eye' phenotype, whereas simultaneous overexpression of *Delta* and *gro* results in a smaller eye.

(A) Representative images of *Drosophila* adult female eyes from flies raised at 26.5°C expressing the indicated transgenes. **(a)** Control adult eye **(b)** single *gro* RNAi knockdown **(c)** single *gro* overexpression showing a severe reduction in eye size **(d)** single *Delta* overexpression showing a 'large eye' phenotype **(e)** simultaneous overexpression of *Delta* and RNAi knockdown of *gro* showing an enhancement of the 'large eye' phenotype **(f)** simultaneous overexpression of *Delta* and *gro* showing a reduced eye size. Genotypes from (a) to (d): *ey-GAL4/UAS-GFP; +/+, ey-GAL4/+; UAS-gro^{RNAi}/+, ey-GAL4/UAS-gro; +/+, ey-GAL4 UAS-Dl/+; +/+, ey-GAL4 UAS-Dl /+; UAS-gro^{RNAi}/+, ey-GAL4 UAS-Dl /UAS-gro; +/+*. **(B)** Confocal images of third larvae eye imaginal discs stained with DAPI (blue) and Elav (red). Elav marks the photoreceptor neurons in the eye. **(a)** control eye imaginal disc **(b)** imaginal disc overexpressing *Delta* **(c)** imaginal disc overexpressing *Delta* and knocking down *gro*.



	No tumour (%)	Tumour (%)	Eye loss (%)
ey-GAL4 > GFP	100	0	0
ey-GAL4 > Dl	100	0	0
ey-GAL4 > gro^{RNAi}	100	0	0
ey-GAL4 > Dl > gro^{RNAi}	92.4	7.6 *	0
ey-GAL4 > gro	63	0	37
ey-GAL4 > Dl > gro	91	0.4	8.6

Figure 50: Simultaneous overexpression of *Delta* and RNAi downregulation of *gro* during eye development leads to tumour formation in a small percentage of eyes.

Representative images of *Drosophila* adult female eyes from flies raised at 26.5^{o} to establish the classification system into three different categories: **(A)** No tumour eyes; includes wild type eyes, as well as eyes showing weak undergrowth or overgrowth **(B)** tumour eyes showing epithelial folds and **(C)** eye loss. Table in the bottom panel shows the percentages of eyes within each category for each of the indicated genotypes. Genotypes from top to bottom are *ey-GAL4/UAS-GFP; +/+, ey-GAL4, UAS-Dl; +/+, ey-GAL4/+; UAS-gro^{RNAi} /+*, *ey-GAL4, UAS-Dl;* UAS-gro^{RNAi} */+, ey-GAL4/UAS-gro; +/+,* and *ey-GAL4, UAS-Dl /UAS-gro; +/+.* More than 150 eyes of each genotype were scored. Statistics were performed using Fisher's exact tests for pairwise comparison between the number of eyes presenting tumours in each genotype with respect to the control (*ey-GAL4 > GFP*). Asterisk was used to indicate a significance of $p < 0.001$.

3.4.2 Gro acts suppressing Notch-associated tumorigenesis during eye development

I next explored the effect of the dysregulation of Gro expression in the context of tumorigenesis caused by the crosstalk between Notch, and epigenetics factors. I used the model of tumorigenesis in *Drosophila* eye named *eyeful*, in which the overexpression of *Delta*, in combination with the GS88A8 insertion (http://flybase.org/reports/FBti0074325.html) that deregulates *psq* and *lola* (genotype: *ey-GAL4 > Dl, GS88A8*), led to the formation of metastatic tumours (see Introduction of this chapter). The *eyeful* phenotype is characterized by tissue overgrowth and folds in the adult eye. In addition, it induces less frequent metastasis of differentiated photoreceptor cells in other parts of the body (Ferres-Marco et al., 2006).

Thus, I tested how *gro* knockdown and overexpression affected tumorigenesis induced by *eyeful.* The introduction of the *UAS-gro^{RNAi}* transgene in the *eyeful* background (*ey-GAL4 > Dl, GS88A8*) caused a clear increase in tissue growth and folds (Figure 51 E) compared to the *eyeful* phenotype (Figure 51 C). In addition, many bunches of bristles were observed around the eye (arrows in Figure 51 E), indicating loss of lateral inhibition mediated by Notch signalling. The enhanced severity in the phenotype was also evidenced in the increased deformation of eye imaginal discs and mislocalization of proneural photoreceptor cells (Figure 51 F) compared to *eyeful* eye imaginal discs (Figure 51 D). Additionally, the quantification of eyes (following the same classification system as in the previous section) revealed an increase of 26% in eye tumour formation in *ey-GAL4 > Dl, GS88A8 > gro^{RNAi}* flies compared to *ey-GAL4 > Dl, GS88A8* flies (Figure 52 A; visualized in graph in Figure 52 B). It is important to note that a clear increase in lethality in crosses yielding *ey-GAL4 > Dl, GS88A8 > gro^{RNAi}* flies was observed compared to the *ey-GAL4 > Dl, GS88A8*. However, due to lack of time, I could not quantify it by viability assays.

In contrast, the introduction of the *UAS-gro* transgene in the *eyeful* background resulted in opposite effects, showing the rescue of the tumour eye phenotype (Figure 51 G and H). The quantification of eyes showed that only 4.2% of the *ey-GAL4 > Dl, GS88A8 > gro* eyes scored developed eye tumours, compared to the 70.6% observed in *ey-GAL4 > Dl, GS88A8* flies (Figure 52 A; visualized in graph in Figure 52 B).

It is important to note that when the fly lines overexpressing and knocking down *gro* were crossed over the *ey-GAL4 > GS88A8* line (in the absence of ectopic *Dl*) there was no obvious overgrowth or tumours in the eyes in the progeny. This suggests that the effect of Gro in overgrowth and tumour formation observed in this section is a consequence of the role of Gro in the Notch pathway, and not of the interaction between *gro* and *psq* and/or *lola*.

Taken together this data indicates that Gro plays a key role in regulating Notch-associated tumorigenesis during *Drosophila* eye development.

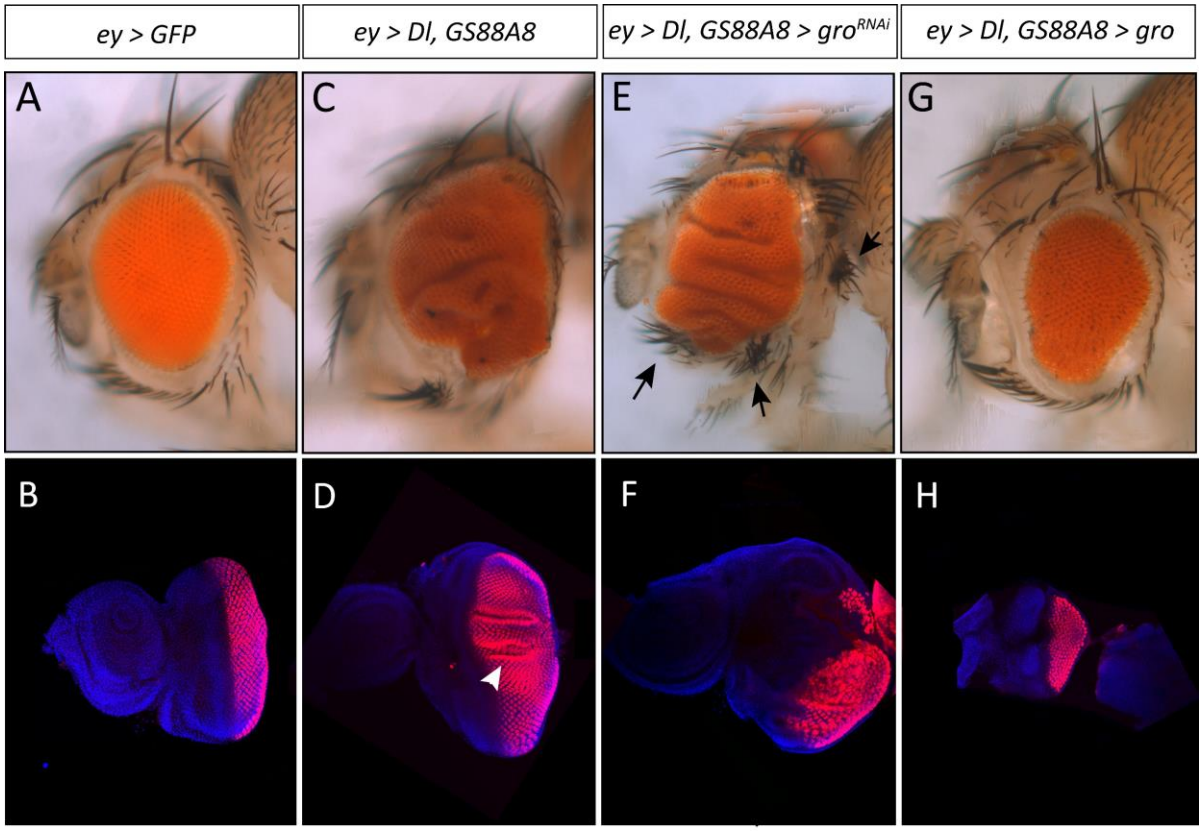


Figure 51: RNAi downregulation and overexpression of *gro* during eye development in the *eyeful* background results in increased severity and rescue of the 'eyeful' phenotype, respectively.

Representative examples of *Drosophila* adult female eyes, and third instar eye imaginal discs of flies raised at 26.5ºC expressing the indicated transgenes: **(A-B)** control adult eye and eye imaginal disc. **(C-D)** Overexpressing *Delta* and the *GS88A8* insertion. The adult eye shows the 'eyeful' phenotype with tissue folds, and the eye imaginal disc shows also foldings (white arrowhead). **(E-F)** Downregulation of *gro* in the *eyeful* background shows an enhancement of the eyeful phenotype and the formation of extra bunches of bristles (black arrows) in the adult eye. The imaginal disc shows deformity and mislocalization of proneural photoreceptor cells (red) compared to the control. **(G-H)** Overexpression of *gro* in the *eyeful* background shows rescue of the eyeful phenotype and results in a smaller eye compared to the control in the adult eye. The eye imaginal disc shows a reduced size. Genotypes in (A-B): *ey-GAL4/UAS-GFP; +/+* (C-D) *ey-GAL4, UAS-Dl, UAS-GS88A8/+; +/+* (E-F) *ey-GAL4, UAS-Dl, UAS-GS88A8/+; UAS-gro^{RNAi} /+* (G-H) *ey-GAL4, UAS-Dl, UAS-GS88A8/ UAS-gro; +/+.* Confocal images of third larvae eye imaginal disc were stained with DAPI (blue) and Elav (red). Elav marks the photoreceptor neurons in the eye.

A

	No tumour (%)	Tumour (%)	Eye loss (%)
ey-GAL4 > GFP	100	0	0
ey-GAL4 > Dl, GS88A8	27.2	70.6	2.2
ey-GAL4 > Dl, GS88A8 > gro^{RNAi}	3.4	96.6 *	0
ey-GAL4 > Dl, GS88A8 > gro	67.8	4.2 *	28

B

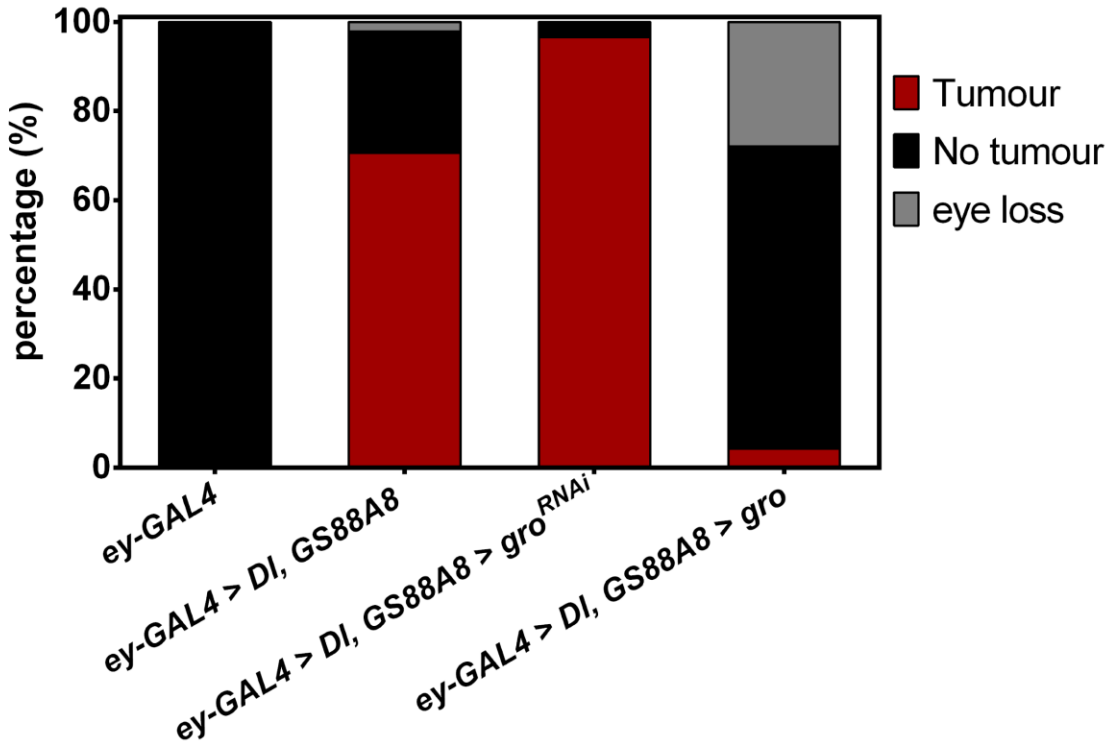


Figure 52: RNAi downregulation and overexpression of *gro* during eye development in the *eyeful* background results in a significant enhancement and rescue of the eye tumour frequency, respectively.

(A) Table showing the quantification of eye phenotypes of the indicated genotypes into the three categories defined in Figure 50: Tumour, No tumour, and eye loss. More than 150 eyes of each genotype were scored, except for *ey-GAL4 > Dl, GS88A8 > gro ^{RNAi}* in which 88 eyes. Statistics were performed using Fisher's exact test comparing the number of eyes presenting tumours in each genotype with respect to *eyeful* eyes (*ey-GAL4 > Dl, GS88A8*). Asterisks were used to indicate a significance of $p < 0.001$. **(B)** Graph illustrating the percentages shown in the table in (A).

3.4.3 Gro co-localizes preferentially with non-BTB-Psq isoform genome-wide

I aimed to establish whether Gro co-localizes with Psq genome-wide and in case it does if this co-localization is specific for the Psq containing or lacking BTB isoforms. Bioinformatic analysis of Gro and Psq binding sites was performed to address this aim. Two Psq ChIP-seq datasets in Kc167 cells (Gutierrez-Perez et al., 2019) were used. One dataset was generated using an antibody that specifically recognized Psq-BTB isoforms (referred to Psq^{BTB}), whereas the other used an antibody that recognized all Psq isoforms (BTB and non-BTB isoforms; referred to Psq^{Total}) (Gutierrez-Perez et al., 2019).

The distribution of ChIP-seq signals for Gro, GAF, Psq^{Total}, and Psq^{BTB} (the two first used as controls) were plotted in heatmaps surrounding Gro binding sites across 2 Kb regions (columns in Figure 53 A). The heatmaps colours range from red to blue and represent increasing amounts of proteins 2 Kb upstream and downstream from the centre of Gro binding sites. In the case of Gro and GAF this distribution is displayed as a blue line in the centre of the heat map (columns 1 and 2 in Figure 53 A). The analysis revealed a high ChIP-seq signal of Psq^{Total} at Gro-bound regions, whereas the Psq^{BTB} signal was weaker (columns 3 and 4 in Figure 53 A, respectively). These results showed that Gro co-localizes with Psq through the genome. Additionally, the analysis of the overlap between Gro peaks and the binding sites of Psq^{Total} and Psq^{BTB} showed that many Gro peaks overlapped only with Psq^{Total} peaks (926), (Figure 61 in the Appendix). There were fewer Gro peaks that overlapped with both Psq^{Total} and Psq^{BTB} (272). These data suggests that Gro might preferentially co-localize with non-BTB Psq, but overlaps with BTB Psq are not excluded.

The overlap between Gro and Psq is illustrated at the Gro targets genes *E(spl)m3-HLH* and *E(spl)mβ-HLH*, in which is observed that Gro binding sites overlap with GAF and Psq^{Total} peaks but not with Psq^{BTB} (Figure 53 B and C).

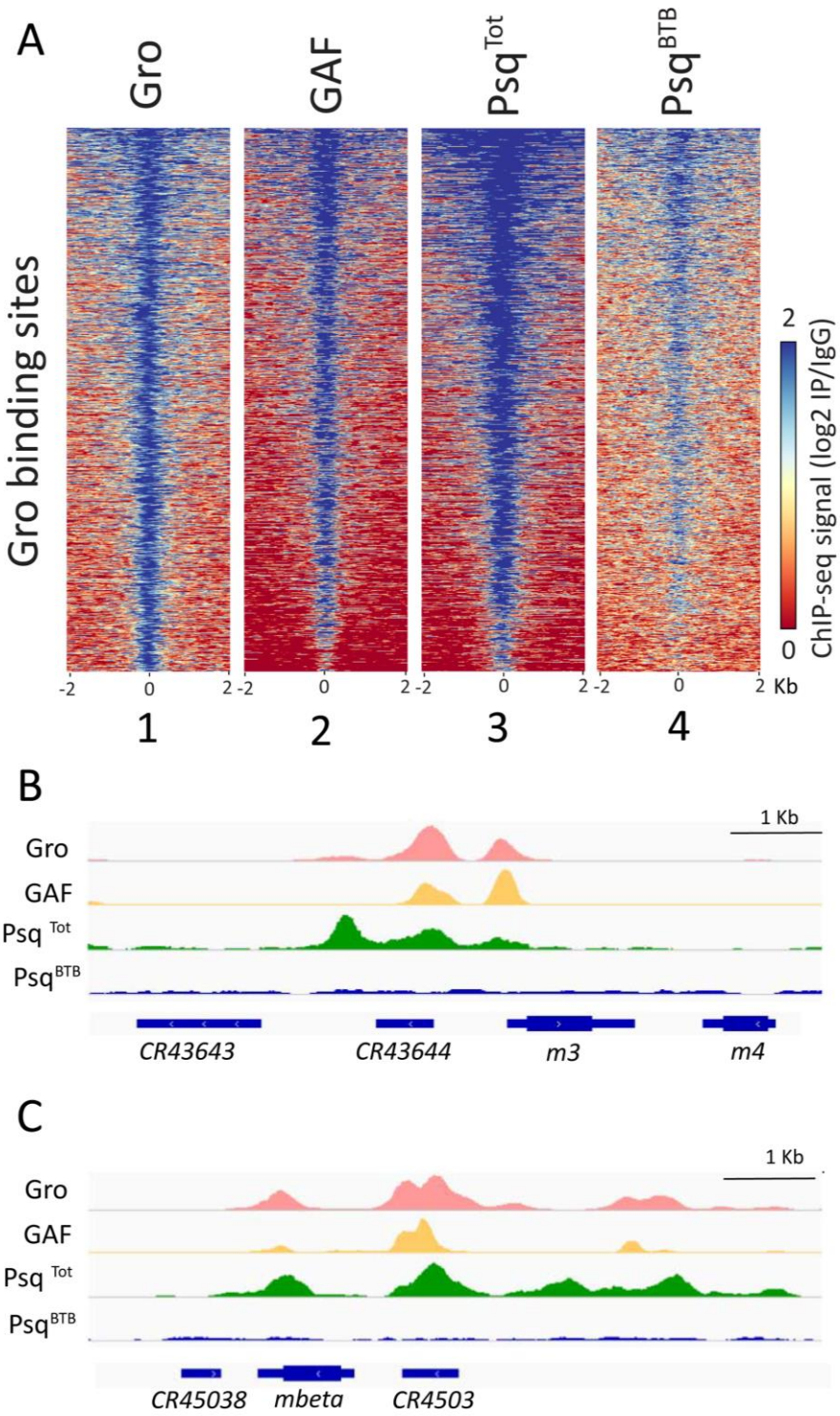


Figure 53: Gro co-localizes with Psq^{Total} genome-wide in Kc167 cells, which can be visualized at *E(spl)m3-HLH* and *E(spl)mβ-HLH* genes.

(A) Heatmaps plotting ChIP-seq signals of (from left to right): Gro, GAF, Psq^{Total}, and Psq^{BTB} 2 Kb upstream and downstream from the centre of Gro peaks. **(B-C)** IGV tracks showing Gro, GAF, Psq ^{Total} and Psq^{BTB} binding sites at the (B) *E(spl)m3-HLH* and (C) *E(spl)mβ-HLH* genes.

3.4.4 Gro co-localizes with CBP and ISWI genome-wide

Psq isoforms containing and lacking the BTB domain are thought to have distinct functions *in vivo* as they bind to different genomic regions and co-localize with different sets of proteins genome-wide. Non-BTB Psq isoforms are known to co-localize with GAF, Polycomb (Pc), the imitation SWI (ISWI), and the CREB-binding protein (CBP) at enhancers and promoters enriched in the histone modification H3K27ac (Gutierrez-Perez et al., 2019).

ISWI is an ATP-dependent nucleosome remodelling factor that is known to bind physically to GAF and it has been implicated in facilitating the release of paused Pol II (Judd et al., 2021; Tsukiyama and Wu, 1995). CBP is a histone acetyltransferase that plays essential role in enhancer activity by controlling H3K27 acetylation [reviewed in (Holmqvist and Mannervik, 2013)]. Furthermore, a recent study in *Drosophil*a S2 cells revealed an additional role of CBP in the regulation of Pol II activity at promoters (Boija et al., 2017).

In contrast, BTB-Psq co-localizes with the architectural proteins Su(Hw), CP190 and Mod(mdg4)2.2 (Gutierrez-Perez et al., 2019).

This work already demonstrated that Gro and GAF work together to regulate gene expression. In addition, the results shown in the previous section indicate that Gro co-localizes with Psq through the genome. As GAF preferentially co-localizes with non-BTB Psq isoforms, we hypothesized that Gro may co-localize with CBP, ISWI, and Pc, and not with Su(Hw), CP190, and Mod(mdg4). To test this, ChIP-seq signals for these factors were plotted in heatmaps surrounding Gro binding sites across 2 Kb regions, as in the previous section. The results showed strong ChIP-seq signals for CBP and ISWI at Gro binding sites, indicating a co-localization of Gro with these factors (Figure 54; columns 4 and 5). A less strong signal was observed for Pc (Figure 54, column 6) and, weak signals were observed for Su(Hw), CP190, and Mod(mdg4) (Figure 54, columns 7, 8 and 9).

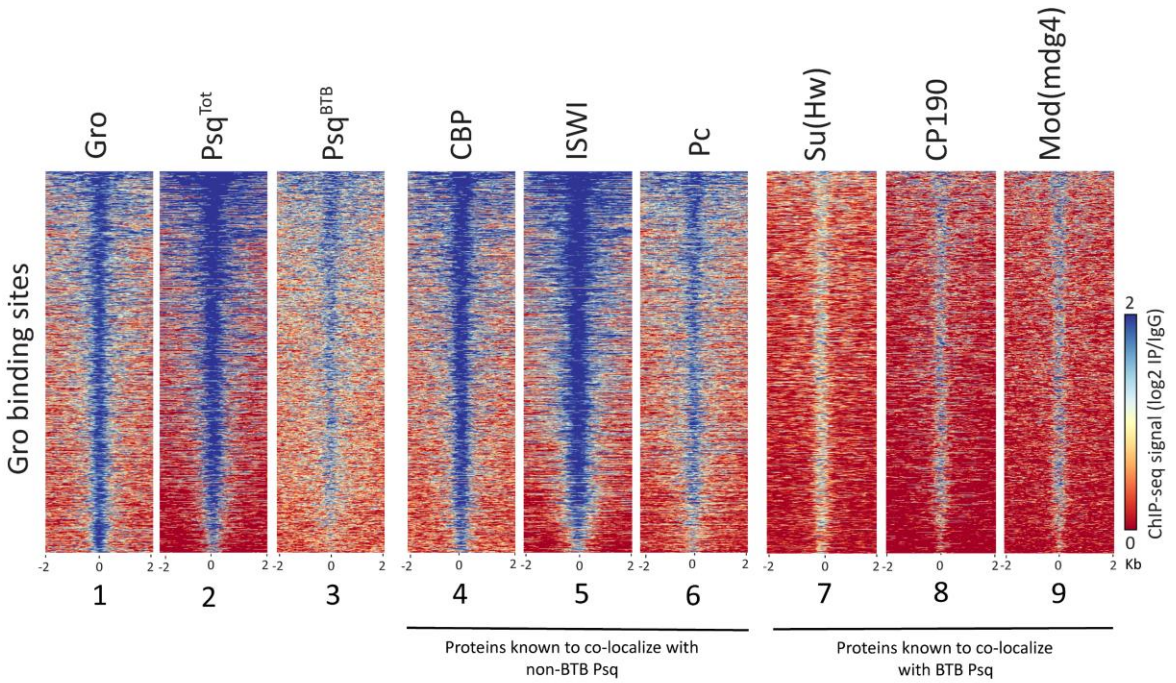


Figure 54: Gro co-localizes with Psq^{Total}, CBP, and ISWI genome-wide in Kc167 cells.

Heatmaps plotting ChIP-seq signals of (from left to right): Gro, Psq^{Total}, Psq^{BTB}, CBP, ISWI, Pc, Su(Hw), CP190, and Mod(mdg4) 2 Kb upstream and downstream from the centre of Gro peaks. CBP, ISWI and Pc have been shown to co-localize with non-BTB, whereas CTCF and Su(Hw) have been shown to co-localize with BTB Psq (Gutierrez-Perez et al., 2019).

The co-localization of Gro with these proteins can be illustrated at the following Notch target genes: *E(spl)m3-HLH*, *E(spl)mβ-HLH*, and *argos* (Figure 55). These three genes were previously used in this work (Figure 33 and 34) to illustrate Gro overlap with pausing factors and the paused Pol II at TSSs. Gro overlaps with GAF, Psq^{Tot}, CBP, and ISWI at the *E(spl)m3-HLH*, *E(spl)mβ-HLH*, and *argos* genes, (Figure 55 A and B), whereas no overlap was detected with Psq^{Tot}, Pc, Su(Hw), CP190, and Mod(mdg4).

Taken together this analysis revealed that Gro co-localizes with CBP and ISWI, as well as with GAF and Psq^{Tot}, which indicates that Gro might preferentially co-localize with non-BTB Psq. Furthermore, the finding of the strong co-localization of Gro with CBP and ISWI has not been previously reported and contributes to our understanding of the model in which Gro promotes promoter-proximal pausing.

It is noteworthy to mention that a preliminary assay to study whether the *psq* isoforms genetically interact with *gro* in the developing eye was performed, however, due to problems with viability this assay could not be completed. In the future, it would be interesting to assay

the relationship between *gro* and each *psq* isoforms *in vivo* to gain insight into Gro-mediated transcriptional repression.

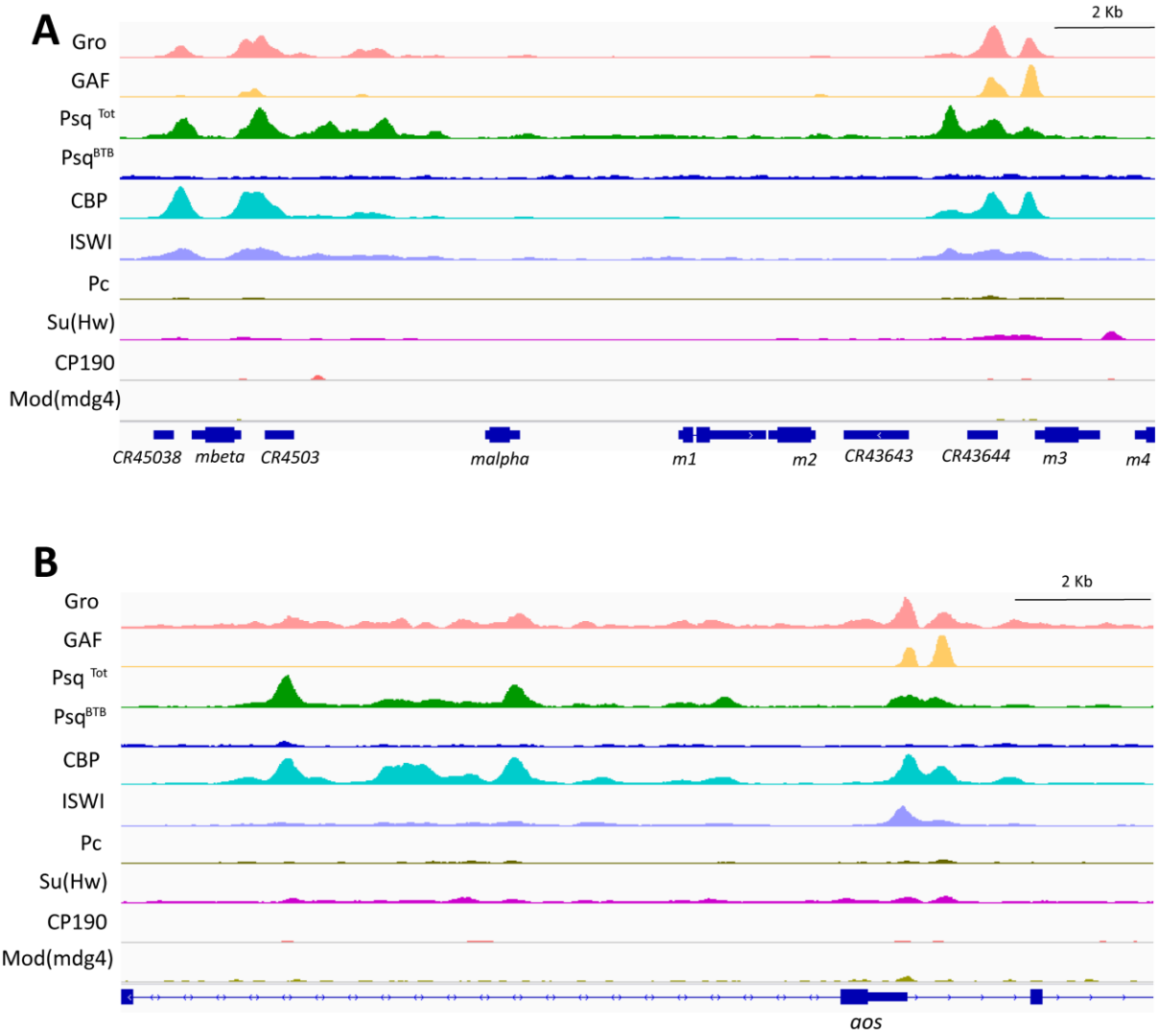


Figure 55: Visualization of binding profiles of the indicated ChIP-seq datasets, at the *Enhancer of split gene complex* [*E(spl)-C*], *argos* (*aos*), and *cut* genes in Kc167.

Figure showing ChIP-seq binding profiles for Gro, GAF, Psq^{Total}, Psq^{BTB}, CBP, ISWI, Pc, Su(Hw), CP190, and Mod(mdg4) at the **(A)** *[E(spl)-C]* locus **(B)** *aos* gene and **(C)** *cut* gene.

3.4.5 Discussion

The results of this chapter revealed a high occupancy of Psq at Gro binding sites in Kc167 cells. It is important to consider that Psq and GAF are both TFs that bind to GAGA sequences, which is the most frequent motif found at Gro binding sites in S2R+, Kc167 (Kaul et al., 2014), and BG3 cells (data from this thesis). It is also known that Psq and GAF physically bind to each

other, and broadly co-localize at many loci (Lehmann et al., 1998; Schwendemann and Lehmann, 2002). Recent work demonstrated that different Psq isoforms play diverse roles during development and tumorigenesis, however, the precise mechanisms underlying the coordinated function of GAF and Psq during development are unknown. It is known that non-BTB-Psq binds together with GAF, ISWI, Pc, and CBP to enhancer regions that are rich in H3K27ac, and form enhancer-promoter loops that regulate transcription in response to hormone stimuli (Gutierrez-Perez et al., 2019). My results indicated a high occupancy of ISWI and CBP proteins at Gro binding sites, whereas the presence of Pc at those sites was lower.

Altogether these results suggest that Gro might co-localize with non-BTB-Psq, CBP and ISWI in regions of open chromatin. A recent study in *Drosophila* cells demonstrated that the histone acetyltransferase CBP is required for the recruitment of Pol II to promoters and facilitates transcription through the +1 nucleosome by histone acetylation (Boija et al., 2017). Importantly, they found that promoters that are co-occupied by high levels of CBP, GAF and Pol II at open chromatin regions belong to highly paused genes, whereas low CBP and GAF genes contain substantially less Pol II are weakly paused (Boija et al., 2017). The genome-wide co-localization observed between Gro and CBP was unexpected and indicates that Gro may play a role in blocking the action of the co-activator CBP at genes with high levels of GAF and CBP. Nevertheless, further experiments would be necessary to explore the molecular and functional relationship among those factors.

4 Conclusions and Future directions

The genetic assays performed in this work (section 3.1 of the results) revealed that Gro cooperates with the key pausing factors GAF, Nelf, Bin3, and Larp7 during development, as well as antagonising the function of Myc and Cdk9 on growth in the proliferating eye. These observations provide the first biological evidence of Gro interacting with factors involved in promoter-proximal pausing and support a model in which Gro promotes the establishment of Pol II.

The antagonism observed between Gro and Myc observed during eye development agrees with a previous study that showed that those factors antagonize each other's functions during neurogenesis (Orian et al., 2007). However, my results also revealed unexpected cooperation between Gro and Myc in the developmental processes that specify the vein and intervein tissue in the wing and the ommatidia in the eye. The cooperation between these factors suggests that Gro may act with Myc in different manners depending on the developmental process and opens new questions about how Gro and Myc interact with each other to regulate gene expression. Myc is a TF that regulate genes involved in different processes such as cell proliferation, differentiation, apoptosis, and metabolism, and it has also been shown to promote pause release in mammalian cells (Rahl et al., 2010). Given the pleiotropic roles of Myc in transcriptional regulation, it would not be surprising that Myc and Gro interact in different manners to regulate gene expression at different target genes. In future studies, the interaction between Myc and Gro during development could be explored by using the CRISPR/Cas9 system to generate double mutant flies for those genes in different tissues (Koreman et al., 2021). A set of experiments could then be designed to study those mutants in different stages of development, including immunolabelling of embryos and imaginal discs.

In agreement with the *in vivo* observations, the bioinformatic analysis (section 3.2 of the results) showed a large genome-wide overlap among the binding sites of Gro, GAF, Nelf, Myc, Cdk9, and Pol II at promoters of genes in embryo-derived Kc167 *Drosophila* cells. A previous study already showed that Pol II is present at Gro-bound regions in Kc167 cells, as well as the overlap between the binding sites of Gro and GAF at promoters (Kaul et al., 2014). Thus, my findings confirmed that GAF and Pol II binds to genomic regions bound by Gro and revealed a genome-wide overlap among the binding sites of Gro, Nelf, GAF, Cdk9, Myc, and Pol II at promoters. Altogether, these observations suggest that Gro, in the presence of the pausing

factors Nelf and GAF acts by blocking the action of factors that promote pausing releases, such as Myc or the P-TEFb complex.

Interestingly, Gro peaks also showed a large and unexpected co-localization with the co-activator CBP (section 3.4 of the results), which is a histone acetyltransferase that has been recently shown to co-localize with GAF at promoters of highly paused genes (Boija et al., 2017). The overlap observed between Gro and the CBP adds support to the hypothesis that Gro acts via blocking the action of factors that promote the release of Pol II at promoters. Acetylation of components of the P-TEFb complex (Cdk9 and CycT in *Drosophila*) plays an important role in regulating P-TEFb activity (Cho et al., 2009; Fu et al., 2007; Sabò et al., 2008). In mammalian cells, it has been demonstrated that CBP acetylates Cdk9, enhancing its kinase activity to phosphorylate the Pol II C-terminal domain (Fu et al., 2007). The study by Fu et al., also revealed that deacetylation of Cdk9 by the N-CoR/HDAC3 co-repressor complex inhibits P-TEFb-mediated transcriptional elongation and N-CoR co-immunoprecipitated with HEXIM (member of the 7sk sNRP complex). These findings led the authors to suggest a model in which the acetylation of Cdk9 plays a regulatory role in partitioning P-TEFb into an active or inactive complex. We can speculate then that Gro may act via a similar mechanism, and hence regulate Pol II pausing by inhibiting the activity of the P-TEFb complex, in a process that may or may not involve the recruitment of histone deacetylases. In future studies, different approaches could shed light on the relationship between Gro and the components of P-TEFb complex. Experiments based on the PRO-seq (Precision Run-On Sequencing) technique could enable to test the levels of transcription in conditions where both *gro* and *Cdk9* (or *CycT*) are downregulated *in vitro*. In addition, the quantification of mRNA production at Gro target genes in flies created by CRISPR/Cas9 expressing mutant versions of Cdk9 protein with deletions at the residues targeted by acetylation may also be informative.

This work also detected the presence of ISWI at Gro binding sites in Kc167 cells (section 3.4 of the results), which is consistent with previous work that used the bacterially produced GST-tagged Q-domain of Gro as bait to pull down proteins from *Drosophila* embryo nuclear extract (Kwong et al., 2015). This result was expected as ISWI and SWI/SNF families of remodelling complexes are key for the pioneer activity of GAF to establish optimal chromatin architecture for transcription at target promoters (Judd et al., 2021).

The analysis of the genomic location of Gro peaks in neuronal BG3 cells, and in the embryo-derived cell lines S2R+ and Kc167 revealed that almost all Gro-bound regions are located either at promoters or introns in the three cell lines (section 3.2 of the results). This analysis also showed that in BG3 cells Gro frequently binds to open chromatin regions, in agreement with the observed previously in Kc167 cells (Kaul et al., 2014). Furthermore, a vast majority of Gro-bound regions in both Kc167 and BG3 cells were located within the enhancer chromatin, which is a type of active chromatin associated with enhancers and characterized by the canonical modifications H3K4me1, H3K27ac, and H3K56ac [described in (Skalska et al., 2015)]. Interestingly, a recent study from the Furlong lab revealed that enhancers, as well as promoters, are frequently accessible during early *Drosophila* embryonic development, even in lineages where they are not active (Reddington et al., 2020). It is during later embryogenesis when cells begin to differentiate, when opened enhancers become progressively closed (Reddington et al., 2020). The authors also found that DHS display a range of variation in chromatin accessibility during embryogenesis, which is correlated with different TFs occupancy. DHSs with dynamic accessibility located distal to TSSs associate with tissue-specific TFs that are important for embryonic development, including Twist, Snail, and the Gro interactors Dichaete and Knirps. In contrast, TSS-distal DHSs with constitutive accessibility are ubiquitously expressed (such as ISWI, E(bx), E2F). A very similar pattern was found in the analysis of DHSs proximal to TSS, with the exception of GAF, which was found at DHSs with dynamic accessibility proximal to promoters but not distal (Reddington et al., 2020). This study is also consistent with Gro acting at regulatory regions characterized by open chromatin with dynamic accessibility at regions proximal and distal to TSSs.

It is also important to consider the essential role that the pioneer TFs (such as the *Drosophila* GAF and Zelda) play during early embryogenesis. These factors occupy enhancers, where they render transcriptional competent chromatin that is susceptible to the binding of additional TFs, which ultimately activate enhancers at later developmental stages (Falo-Sanjuan et al., 2019; Kim et al., 2018; Liber et al., 2010). In the context of the Notch pathway, recent work from Sarah Bray's lab indicated that the tissue-specific TFs Twist and Dorsal act by priming Notch-responsive enhancers, sensitizing them for the binding of the Notch intracellular domain (Falo-Sanjuan et al., 2019). The authors proposed that the action of these factors leads to the synchronized and sustained transcriptional activity generated upon Notch

activation. They also hypothesized that the recruitment of the repressor complex to primed enhancers could poise Notch-responsive enhancers and set a threshold for its activation. Consistent with this model, Gro has been demonstrated to mediate the repression of Notch target genes in the absence of Notch signalling in *Drosophila* (Barolo et al., 2002; Kaul et al., 2014; Nagel et al., 2005). In *Drosophila* this repression requires a complex formed by Hairless, Su(H), that also includes the co-repressors Gro and CtBP (Barolo et al., 2002; Morel et al., 2001; Nagel et al., 2005). Similarly, Gro also maintains repression of Wnt target genes in the absence of Wnt signal by interacting with the Tcf/Lef transcription factors (Brantjes, 2001; Cavallo et al., 1998). Thus, Falo-SanJuan's study is consistent with a model in which Gro keeps target genes in a poised state at regulatory regions that have been previously opened by pioneer TFs.

Considering these recent revelations on enhancer activity during *Drosophila* development, we can hypothesise that Gro is recruited to regulatory regions of target genes that have been previously primed by tissue-specific pioneer factors. This is based on three findings of this thesis: (1) Most Gro binding sites localize either at promoters or introns (putative enhancers) of genes at regions of open chromatin in different *Drosophila* cell lines. (2) The strong genetic interaction observed between *gro* and *Trl* during wing development, but the inability to detect this interaction in the eye. (3) Gro co-occupies genomic regions bound by Psq in *Drosophila* Kc167 cells.

Once Gro is recruited to these regulatory regions by specific TFs, Gro will act to repress target genes by inhibiting the action of P-TEFb and hence blocking the release of the poised Pol II.

Additionally, the genetic assay shown in section 3.4 of the results revealed a crucial role of Gro in the regulation of normal growth, as well as Notch-dependent tumorigenesis during fly eye development. A previous study already showed that Gro is key to maintaining Notch target genes repressed in the absence of pathway activation in the developing eye (Nagel and Preiss, 2011). The authors showed that Hairless-mediated repression of Notch signalling during early eye development is dependent on Gro recruitment, and results in the inhibition of growth as it affects cell division (Nagel and Preiss, 2011). Therefore, my findings on Gro effects on growth in the *Drosophila* proliferating eye are consistent with Nagel and Preiss's revelations. Besides effects on growth, the downregulation of *gro* in the developing eye showed effects in cell death (section 3.1 of the results). Although the specific functions of Gro

during each stage of eye development remain unexplored, it is known that Gro interacts with factors that are key in eye development. One of these factors is Optix, which expresses in front of the morphogenetic furrow, and is key in the formation of the compound eye (Anderson et al., 2012). As the *Drosophila* eye developmental is a very complex process, further studies would be required to decipher the function of Gro during each stage of eye development.

Taking together all the results of this thesis, I propose the following mechanism to explain Gro-mediated repression via promoting promoter-proximal pausing:

1. During early *Drosophila* development, pioneer factors, such as GAF, enable the opening of chromatin at regulatory regions, allowing the binding of Pol II, GTFs, and transcriptional activators (including CBP).
2. Pol II then transcribes approximately 30-60 bp downstream of the TSS and then, the Nelf and DSIF complexes, which have also been recruited to the promoter regions, establish the paused Pol II.
3. At the same time, the P-TEFb complex is recruited to those promoter regions to release the paused Pol II.
4. However, at Gro target genes, specific TFs recruit Gro to the promoter regions of target genes where it inhibits the action of the P-TEFb complex, possibly through stabilising the 7SK/PTEF-b complex, impeding the release of the paused Pol II (Figure 56 A).

   At his stage, the maintenance or the release of paused Pol II would depend on the competition between factors that promote pausing or release at the regulatory regions.
5. The expression of high levels of activators by the activation of specific developmental cues enables the recruitment of the P-TEFb complex to their target genes, overcoming Gro-mediated repression and leading to the phosphorylation of Pol II, NELF, and DSIF. This causes the dissociation of Nelf from the complex and converts DSIF in positive transcription elongation, inducing transcriptional elongation. In addition, several TFs can localize at enhancers and increase transcription of genes through loop formation with promoters (Figure 56 B).

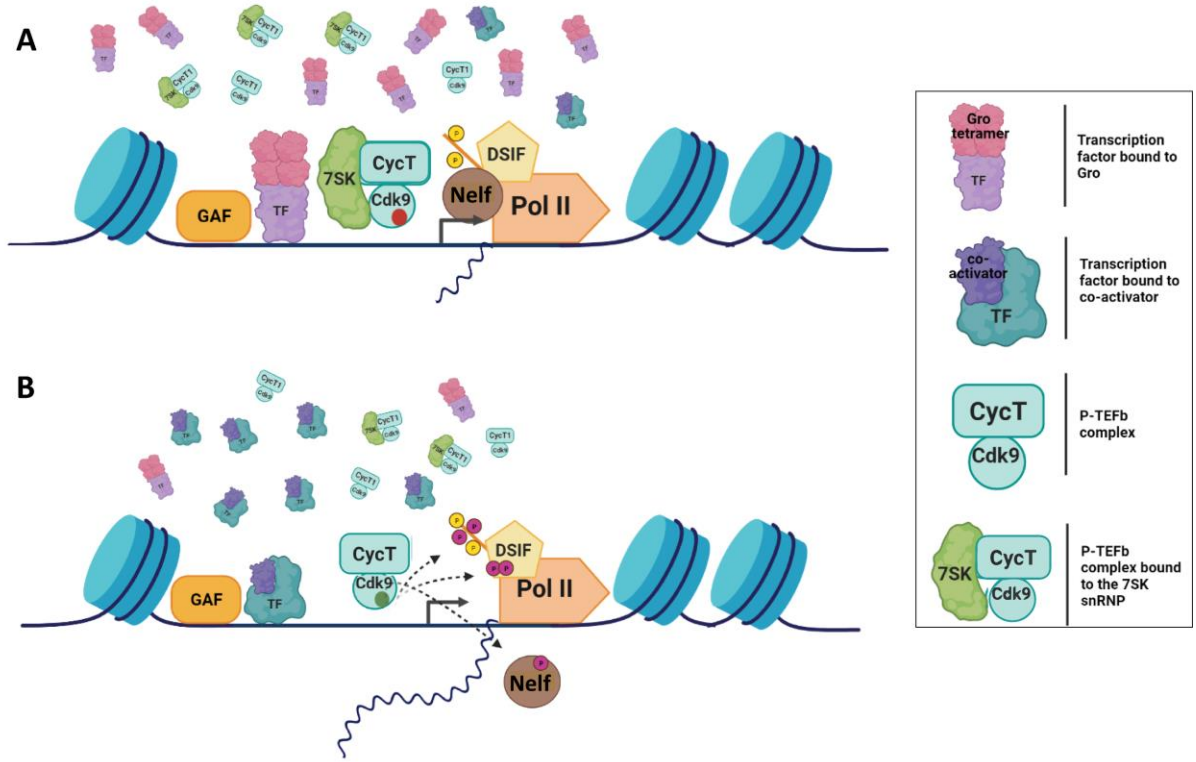


Figure 56: Schematic illustration of the proposed mechanism by which Gro promotes Pol II pausing. **(A)** DNA-binding transcription factors (TFs) recruit Gro to accessible promoter DNA sequences that have been previously opened by the action of GAF. At those regulatory regions, Gro mediates transcriptional repression by blocking the action of the P-TEFb complex (CycT and Cdk9 in *Drosophila*), either directly or indirectly. Inactive P-TEFb complex bound to the 7SK snRNP complex is represented by a red dot in Cdk9. **(B)** Upon the activation of signalling pathways, there is an increased number of activators and co-activators that are recruited to promoter regions, which subsequently recruit the active form of PTEF-b (not bound to the 7SK snRNP complex). The active P-TEFb complex is represented by a green dot in Cdk9. Then, Cdk9 phosphorylates (magenta dots) the C-terminal of Pol II, DSIF (composed of Spt4 and Spt5), and Nelf. Phosphorated Nelf is released from the complex, and DSIF is converted into a positive transcription elongation factor that travels with Pol II through DNA to produce RNA transcripts. TSS in the image is indicated with an arrow. Nascent RNA is shown in blue. Created with BioRender.com.

Therefore, the model of Gro-mediated repression proposed here is based on the competition between the protein levels of activators and repressors, which will determine whether the target genes are paused or active. This model is also supported by another study that showed that the Notch activator complex Su(H)-N^{ICD} induces transcriptional activation at the *E(spl)-C* locus by recruiting factors (including co-activators) that increase the residence time of the activator complex at their targets (Gomez-Lamarca et al., 2018).

Importantly, the analysis of the overlap between Gro peaks and the factors involved in Pol II pausing performed in this thesis showed a similar pattern in promoters and introns, except for Pol II, which overlapped the rest of the tested factors at promoters but not at introns (section 3.2 of the results). This suggests that Gro acts by maintaining both enhancers and promoters in a poised state, in the absence of the activation of developmental pathways, in a process that may involve chromatin looping. In future studies, this model could be tested by varying the levels of activators that "compete" with Gro to test whether they can overcome Gro-mediated repression in cell culture or different *Drosophila* tissues during development.

5 Materials & methods

5.1 Fly husbandry

The conditional knockdown of specific genes was achieved with the GAL4/UAS system (Brand and Perrimon 1993), using *nubbin-GAL4*, *GMR-GAL4* and *Eyeless-GAL4* as drivers, and the UAS or UAS-RNAi transgenic lines. The following UAS-RNAi and UAS stocks were obtained from Bloomington Drosophila Stock Centre (BDSC); stock numbers indicated in brackets: *UAS-Nelf-D*^{RNAi} (38934), *UAS-Nelf-B*^{RNAi} (34847), *UAS-Nelf-A*^{RNAi} (32897), *UAS-Nelf-E*^{RNAi} (32835), *UAS-Bin3*^{RNAi} (50634), and *UAS-GFP*^{RNAi} (9331), *UAS-GFP* (4775), *UAS-Myc*^{RNAi} (35784), *UAS-CycT*^{RNAi} (35168), *UAS-Cdk9*^{RNAi} (35323), *UAS-Cdk12*^{RNAi} (35163), and *UAS-CycT*^{RNAi}(35168). UAS-RNAi stocks used were also obtained from Vienna Drosophila Resource Center (VDRC) (stock numbers indicated in brackets): *UAS-gro*^{RNAi} (6316), *UAS-Trl*^{RNAi} (41095), *UAS-Larp7*^{RNAi} (31635), *UAS-Hexim*^{RNAi} (34632 and 100500), *UAS-Vdup1*^{RNAi} (15203),and *UAS-Spt5*^{RNAi} (4075 and 106814), *UAS-Pho* ^{RNAi} (39529), Other UAS stocks used in this thesis were from the laboratory of Dr Barbara Jennings: *Spt5*^{W049} (Jennings et al., 2004) and Dr Maria Domínguez: *UAS-Dl*, *UAS-GS88A8* (Ferres-Marco et al., 2006; Vallejo et al., 2011). *UAS-gro* (Orian et al., 2007) was a gift from Dr Ze'ev Paroush.

Endogenous alleles were obtained from BDSC: *Trl*^{ex15} (64191), *Trl*^{R67} (58475), *Spt5*^{MGE} (8352), *Hexim*^{E03375} (18151), *Hexim*^{EY14285} (20799). *Nelf*^{A22} (Wang et al., 2010) and *Nelf-E* (Wang et al., 2010) were a gift from Dr. Peter Gergen.

All the flies were raised on standard cornmeal food and maintained either at 25^{o}C or at 18^{o}C in an incubator. Crosses were set up either at 25^{o}C or 26.5^{o}C on standard cornmeal media with the addition of dry yeast. Parents were cleared from the vial 3 days after the cross was set up, and progeny were collected one week after the pupas started to hatch.

5.1.1 Image capture of adult fly wings

Flies were collected in a microcentrifuge tube containing 100% ethanol. Wings were then removed with forceps in a watch glass containing 100% ethanol, and transferred to a watch glass containing 100% Isopropanol. Next wings were placed on slides in a drop of Euparal and covered with a glass coverslip for light microscopy imaging. These wings were then imaged

using a Zeiss AxioZoom.V16 microscope mounted with an objective Apo Z 0.5x/0.125 FWD 114 MM and Axiocam 506 colour (D) camera.

5.1.2 Image and processing capture of adult fly eyes

Two different methods and microscopes were used. For eyes shown in section 3.1:

2-to 3-old flies were collected and immobilized by freezing at -80^{o}C. On the day of imaging, flies were thawed, and heads were removed with forceps and mounted on double side tape. Eyes were immediately imaged using a Zeiss AxioZoom.V16 microscope mounted with an objective Apo Z 1.0x/0.25 FWD 60 MM and Axiocam 506 colour (D) camera. Images shown in figures are maximum projections of consecutive optical z-sections. Eye size area was measured using ImageJ software (Schneider et al., 2012).
For eyes shown in section 3.4 [processed in Neuroscience Institute (Alicante, Spain)]:

2-to 3-old flies were collected and immobilized in 100% ethanol. Eyes were immediately imaged using an optical microscope Zeiss Axiophot, using a MicroPublisher 5.0 camera (QImaging) and the QCapture software (Qimaging). Pictures were taken using a 5x objective with 1.5x zoom. Images shown in figures are composite of 5 to 10 images of the same sample focused on different focal planes, which were 3D coupled using the software AutoMontage Essentials v5.0.

5.1.3 Immunostaining of eye imaginal discs

Eye-antennal imaginal discs from third instar *Drosophila* larvae were collected and dissected in PBS 1x (maxim period of 30 minutes each round of dissection). Discs were fixed in 4% paraformaldehyde (PFA) for 30 minutes at room temperate and with agitation. Fixation was followed by three washes of 5 minutes with PBS 1x. Blocking was performed for 30 minutes at room temperature using PBT-BSA (PBS 1x buffer, 0.3% Triton and 1% BSA). The primary antibody anti-Elav [from the Developmental Studies Hybridoma Bank (DSHB)] was diluted with PTB-BSA at a 1:100 dilution and incubated overnight at 4^{o}C. After three washes with PBS, 1x the tissues were incubated with the secondary antibody anti-rat AlexaFluor-555 (Molecular Probes) at a 1:1000 dilution for 1 hour at room temperature. Discs were mounted in

Vectashield with DAPI (Vector Laboratories) and the images capture on the Olympus FV1200 Laser Scanning Microscope.

5.2 Bioinformatics analysis of ChIP-seq datasets

Sequences were aligned to the dm6.26 reference genome using bowtie2 v2.3.3 (Langmead and Salzberg, 2012) with the default parameters and specifying the options "--no-unal" and "-local" to remove unmapped reads and trim adapters. Multimapping reads were then removed using the "view" program of SAMtoolstools v1.7 (Li et al., 2009) with the parameter "-q 20". The program "markdup" was also used to remove PCR duplicates from mapped reads, with the parameter "-r".

For the analysis of Gro ChIP-seq data in BG3 (dataset generated by Aamna Kaul in my lab), two biological replicates were used to obtain a high confidence set of peaks. Peaks were called using the MACS2 v2.1.1.2 software (Gaspar, 2018) providing input and IP samples simultaneously with default parameters. Peaks present in both biological replicates (with an FDR less than 10% in at least one of the samples and a p-value less than 0.001 in the other) were selected by using BEDTools intersect (Quinlan, 2014) with a minimum overlap of 80%.

ChIP-seq peaks for the rest of the factors analysed in this work were retrieved from public databases. Peaks from the databases that were aligned to the dm5 version of the *Drosophila* genome were converted to dm6 using the coordinate converter from Flybase (Larkin et al., 2021). Where ChIP-seq peaks from two replicates were provided, BEDTools intersect (Quinlan, 2014) was used to obtain a single set of high confidence peaks using a minimum overlap between 50 and 80%.

The Bioconductor library ChIPseeker (Yu et al., 2015) was used to annotate the peaks by identifying nearby genes and to visualize their genomic locations. For annotations in this thesis, promoters were defined -250 to 250 bp from the TSSs.

The breadth of Gro peaks in BG3 cells was analysed and plotted using the Bioconductor library ChIPpeakAnno v3.22.2 (Zhu et al., 2010). For optimal visualization, the histogram was divided into 500 bins and the two Gro peaks that were greater than 4000 bp were excluded from the histogram.

Open chromatin at Gro bound regions in BG3 cells was analyzed using a DNA-seq data set in BG3 from modEncode (accession number: 5528). DNA-seq signal at Gro binding sites was plotted in an average plot and a heatmap using computeMatrix and plotHeatmap from deepTools v3.4.3 (Ramírez et al., 2014). The plots were centred at the centre of Gro peaks and extended 2 Kb upstream and downstream of the reference point.

CentriMo v5.4.1 (Bailey and Machanick, 2012) software from the MEME suite was used to identify sequence motifs that were enriched in 500 bp sequences centred on the Gro binding peak summits in BG3 identified by MACS2. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 (Huang et al., 2007) was used for the functional annotation of genes present at Gro binding sites or nearby identified by ChIPseeker.

The chromatin states files for the *Drosophila* genome in Kc167 and BG3 cells files were provided by the Dr Sarah Bray laboratory (Skalska et al., 2015). They were first converted from dm5 to dm6.26 using the coordinate converter from the Flybase (Larkin et al., 2021). To analyse the chromatin states at Gro binding sites in those cell lines, the BEDTools intersect was used with a minimum overlap of 50%. Percentage of Gro peaks within each chromatin state (or colour) were then plotted using GraphPad Prism v6.

BEDTools intersect or Intervene (Khan and Mathelier, 2017) were used for the analysis of overlapping binding sites between Gro and the rest of the factors. A minimum and reciprocal fraction of 5% of overlap was required to consider it an overlap. The Venn diagrams shown in the figures were created using the matplotlib-venn package (Hunter, 2007)Upset plots were generated directly from the Intervene module UpSet.

BigWig files were generated for the visualization of the genome tracks of ChIP-seq signal in the Integrate genomics viewer (Thorvaldsdottir et al., 2013). BigWig files were generated from BAM files using bamCoverage from deepTools v3.4.3 using a bin size of 1, and normalization by Reads Per Kilobase per Million mapped reads (RPKM).

Heatmaps in Chapter 3.4 were generated using bigWig files as input. First, BAM files were converted into bigWig using bamCoverage from deepTools v3.4.3, using a bin size of 1, and normalization by Reads Per Kilobase per Million mapped reads (RPKM). Then, bigwigCompare from deepTools v3.4.3 was used to calculate the log2 ratio between the enriched versus the input (IgG ChIP-seq), using a bin size of 1. The output files were used to create the heatmaps

using deepTools v3.4.3 (computeMatrix and plotHeatmap). Heatmaps were centred at genomic locations bound by Gro and extended 2 Kb upstream and downstream of the reference point.

5.3 Yeast 2-hybrid

The cDNA of *Drosophila gro* (Jennings et al., 2006) was used as a bait of the assay. The sequence was amplified by PCR and cloned in frame into the vector pGBKT7 (Clontech) using the In-Fusion cloning kit (Takara), following the manufacturer's protocol. The pGBKT7 is a yeast expression vector designed to express proteins fused to the GAL4 DNA binding domain (DNA-BD). The full coding region of *hairy* (from the *Drosophila* Genomics Resource Center [DGRC], clone RE40955) was also amplified and cloned in frame into the pGADT7 plasmid (Clontech) for use as control prey. The pGADT7 is a yeast expression vector designed to express proteins fused to the activation domain (AD). The sequence encoding the WRPW domain fused to the Glutathione-S-transferase (GST) (Jennings et al., 2006) was also cloned into the pGADT7 vector to use it as an additional control. The bait and the two prey plasmids were sequenced (Eurofins genomics) to check the inserts. The bait plasmid was transformed in the yeast strain Y2HGold (Clontech) using the Yeastmarker™ Yeast Transformation System 2 (Clontech).

As prey, we used the normalized Clontech Universal *Drosophila* Mate & Plate Library, which is constructed from mRNA isolated in equal quantities from embryos (~20 hr), larvae, and adults and it is provided transformed into the yeast strain Y187.

Interactions are based on the yeast reporter genes *AUR1-C*, *ADE2*, *HIS3*, and *MEL1. AUR1-C* is a dominant mutant version of the *AUR1* gene that encodes the enzyme inositol phosphoryl ceramide synthase. AUR1-C is expressed in Y2HGold Yeast Strain in response to protein-protein interactions that bring the GAL4 AD and GAL4 DNA-BD domains in proximity. In *Saccharomyces cerevisiae*, its expression confers strong resistance to the otherwise highly toxic drug Aureobasidin A. The *HIS3* and *ADE2* genes are also expressed when the GAL4 AD and GAL4 DNA-BD domains are in proximity. The expression of these genes is necessary for the yeast strain Y2HGold to grow, as this strain is unable to grow on a medium that does not

contain histidine and adenine. *MEL-1* encodes a-galactosidase, an enzyme occurring naturally in many yeast strains. As a result of the Y2H protein-protein interactions, a-galactosidase (MEL1) is expressed and secreted by the yeast cells. Yeast colonies that express *Mel1* turn blue in the presence of the chromogenic substrate X-alpha-Gal

The yeast 2-hybrid screen was performed using the Matchmaker Gold Yeast Two-Hybrid System from Clontech following the manufacturer's protocols. The haploid yeast strains Y187 and Y2HGold were mated with each other to form diploid cells co-expressing the bait and the prey plasmids. To calculate the mating efficiency, 100 ul of the mated culture (1/10, 1/100, 1/1000 and 1/10000 dilutions) was plated on plates of SD/-Trp [synthetically defined (SD) medium that includes every essential amino acid except for Tryptophane], SD/-Leu (lacks leucine), and SD/-Trp/-Leu [lacks both Trp and Leu; referred to as double Dropout medium (SD/DDO)] and incubated for 3-5 days at 30 ^{o}C. The remaining culture was plated on SD/DDO/X-a-gal/A plates [SD/DDO medium that is supplemented with the antibiotic Aureobasidin A (Takarabio), and X-alpha-Gal (takarabio)] and incubated for 3-5 days at 30 ^{o}C. The blue colonies that appeared on SD/DDO/X-a-gal/A agar plates were picked and streaked on the high stringent plates SD/-Ade/-His/-Trp/-Leu/X-a-Gal/A [plates lacking Trp, Leu, histidine, and adenine, and supplemented with X-alpha-Gal; referred to as Quadruple Dropout (SD/QDO/X-a-gal)] and incubated at 30°C for 3–5 days.

Following that procedure, 4.7 million interactions were screened. The inserts from the prey plasmids present in the positive colonies were amplified by PCR and sequenced. They were identified by sequence comparison with the BLAST tool of Flybase (Larkin et al., 2021).

The physical interaction between the bait and the Vitamin D3 up-regulated protein 1 was tested as a fragment of the sequence encoding that protein was contained in the prey plasmids detected in most of the positive colonies. The complete coding sequence of *Vdup1* (clone RE30730 from DGRC) was PCR amplified and cloned into the pGADT7 vector using In-fusion cloning. The sequence of the resulting plasmid was checked by sequencing. The bait and the prey plasmid encoding *Vdup1* sequence were co-transformed into the yeast strain Y2HGold and plated into SD/DDO/X-a-gal/ and SD/QDO/X-a-gal/ plates.

Additionally, the physical interaction between the bait and the two GAF isoforms (GAF-519 and GAF-581) was tested. To do so, the coding sequence of *Trl-519* and *Trl-581* (clones

LD41963 and AT11176 from DGRC, respectively) were PCR amplified and cloned into the pGADT7 vector using In-fusion cloning. The sequence of the resulting insert was checked by sequencing. The bait and the prey plasmids were co-transformed into the yeast strain Y2HGold and plated into SD/DDO/X-a-gal/ and SD/QDO/X-a-gal/ plates.

5.4 Co-immunoprecipitation assays

The cDNA sequences of *gro*, *Trl-519*, *Trl-581* (same as the used for the yeast 2-hybrid assay), as well as the sequence encoding the GST-tagged WRPW domain, were PCR amplified and cloned in frame into the entry plasmid pENTR/D-TOPO vector (Thermofisher). Then, the coding regions were transferred from the entry vector into two different destination vectors (pAFW and pAMW from Drosophila Gateway Vector Collection distributed by the DGRC) in different LR reactions using the Gateway LR Clonase II Enzyme mix (Thermofisher). pAFW is an expression vector designed to express proteins fused to a 3xFLAG tag in the N-terminal under the act5c promoter. The pAMW is very similar but expresses proteins fused to a 6xMYC tag in the N-terminal instead. The sequence of the inserts in the final plasmids were sequenced to check that the sequences of interest were in frame with the sequence of the tag. The following plasmids were generated by the described process: *pAFW-3xFLAG-gro*, *pAMW-6xMYC-gro, pAFW-3xFLAG-GST-WRPW*, *pAMW-6xMYC-GST-WRP, pAFW-3xFLAG-Trl-519, pAMW-6xMYC-Trl-519, pAFW-3xFLAG-Trl-581, pAMW-6xMYC-Trl-581*.

Different combinations of those plasmids were expressed in *Drosophila* S2R+ cells (from DGRC) to test for physical protein-protein interactions. S2R+ cells were maintained in Schneider's Drosophila Medium (Thermofisher) supplemented with 10% inactivated fetal bovine serum (FBS) (Gibco) and 1% penicillin/streptomycin (Sigma) at 25°C without CO_2. On the day of transfection, S2R+ cells growing were first washed in PBS 1x, collected by centrifugation (5 min at 1.100 r.p.m), and re-suspended in fresh medium (supplemented with FBS and antibiotics) at a density of $3x10^6$ cells/ml. Then, 1.5 ml of cell suspension was plated in each well in 6-well plates. Cells were then transfected using Fugene HD (Promega) in a reagent:DNA ratio of 8:2 following the manufacture's protocol and incubated for 3 days in normal growth conditions. Cells were then collected using a cell scraper, pelleted by centrifugation, and washed three times in cold PBS 1x. They were resuspended in cold

standard lysis buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% IGEPAL, 1 mM EGTA) supplemented with protease inhibitors [2 mM of Pefabloc (Merck), and 1x of cOmplete Protease Inhibitor Cocktail (thermofisher] and incubated for 15 min on ice. Then, cells were spun at 4°C maximum speed for 20 min, and the supernatant was collected. At this point, supernatants were either directly used for immunoprecipitation or frozen at -80^{o}C for later use.

For the immunoprecipitation protocol, Dynabeads magnetic beads (Thermo fisher) were first resuspended by vortexing for approximately 30 seconds. 50 ul of Dynabeads (1.5 mg) were added to each microcentrifuge tube. Then, the tubes were placed on the magnet to remove the buffer. Beads were then washed three times in IP buffer (same composition as the used for cell lysis described above). To bind the antibodies to the beads, beads were resuspended in IP buffer and incubated with the antibody (typically 1-10 ug) at room temperature for 15 minutes with rotation. Either one of the following antibodies was used with the dilutions indicated in brackets: anti-Myc (10:200) [(A-14): sc-789 from Santa Cruz], or anti-FLAG (1:200) [(R) M2 from sigma]. The tubes were then placed in the magnet to remove the supernatant. The complex formed by the beads and the antibodies (Dynabeads-Ab) was washed gently in IP buffer. The cell lysate to analyse was then mixed with the beads and IP buffer (1:2 lysate: buffer ratio) and incubated with rotation for 20 min at 4°C to allow the binding of the antigen to the Dynabeads-Ab complex. After the incubation, the tubes were placed again in the magnet to remove the supernatant. The complex formed by the beads, the antibody, and the antigen (Dynabeads-Ab-Ag) was washed four times in IP buffer, and in the last wash, it was transferred into clean tubes to avoid the co-elution of proteins bound to the tube wall. To elute and denature the proteins bound to the antibody, 1x NuPAGE LDS sample buffer (Thermo fisher) was added to the Dynabeads-Ab-Ag complex and incubated for 10 min. at 70°C. Lastly, the tube was placed on the magnet and the supernatant containing eluted antibody and antigen to a clean tube for posterior western blot analysis.

5.5 Turbo-ID proximity labelling assay in *Drosophila* Kc167 cells

5.5.1 Cloning of the bait and control constructs

First, the sequences encoding *TurboID-NLS,* and *TurboID* (plasmid #107171 from addgene), as well as *miniTurbo-NLS,* and *miniTurbo* (plasmid # 107172 from addgene) were PCR amplified and cloned into the pENTR/D-TOPO vector. The following plasmids were generated: *pENTR/D-TbID, pENTR/D-miniTb*, *pENTR/D-TbID-NLS* and *pENTR/D-miniTb-NLS.* The *pENTR/D-gro* plasmid was already available in the lab and encoded the cDNA of *gro* (Jennings et al., 2008). Then the coding regions of *pENTR/D-TbID* and *pENTR/D-miniTb* were PCR amplified with the designed primers (see primer list in 6.1), as well the full *pENTR/D-gro* plasmid. Subsequently, the PCR amplified products encoding *TbID* and *pENTR/D-gro* were joined using NEBuilder HiFi DNA Assembly (New England Biolabs) to generate the *pENTR/D-TbID-gro* plasmid. In a parallel, the PCR amplified products encoding *miniTb* and *pENTR/D-gro* were also joined using NEBuilder HiFi DNA Assembly to generate the *pENTR/D-miniTb-gro* plasmid.

Finally, the inserts of interest were transferred from the entry vectors (*pENTR/D-TbID-NLS*, *pENTR/D-miniTb-NLS, pENTR/D-TbID-gro,* and *pENTR/D-miniTb-gro*) into the pAGW destination vector (from Drosophila Gateway Vector Collection distributed by the DGRC) that is used for transient expression in *Drosophila* cells. pAFW is an expression vector designed to express proteins fused to an eGFP tag in the N-terminal under the act5c promoter. To summarize, the final bait constructs used in this experiment were *pA-GFP-TbID-gro,* and *pA-GFP-miniTbID -gro*. And the control constructs were *pA-GFP-TbID-NLS* and *pA-GFP-miniTbID-NLS.*

5.5.2 Pull-down of biotinylated proteins

For the proximity labelling assay *Drosophila* Kc167 cells (from DGRC) were maintained in Schneider's Drosophila Medium (Thermofisher) supplemented with 10% inactivated fetal bovine serum (FBS) (Gibco) and 1% penicillin/streptomycin (Sigma) at 25°C without CO_2. Transient transfection of Kc167 cells was performed as explained in section 5.4. Between 1

and 3 days after transfections, cells were pelleted and washed three times in ice-cold PBS 1x. Then, cells were lysed to prepare nuclear extracts following a protocol adapted from (Wysocka et al., 2001). Pelleted cells were resuspended in buffer A (10 mM HEPES [pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM dithiothreitol, and 1x cOmplete EDTA-free Protease Inhibitor Cocktail), then Triton X-100 was added (0.1% final concentration), and cells were incubated on ice for 8 min. Nuclear fraction was pelleted by centrifugation (5 min. at 3000 r.p.m, 4°C), washed once in buffer A, and resuspended in buffer A. Benzonase (Merck) was then added (250 units) and incubated with agitation for 1 hour at 4°C. Samples were then sonicated on ice in three sessions of 10 seconds at 10% amplitude to share DNA into smaller fragments. Subsequently, the stock of Streptavidin-Sepharose beads (GE Healthcare) was gently resuspended by tapping. The number of beads necessary for the experiment was washed once in buffer A, and centrifuged (2 min. at 1000 g) to collect the beads and remove the supernatant. Then, the corresponding quantity of streptavidin beads (10 ul of beads for each well of 6-well plates) was added to the tubes with the nuclear extracts and incubated in rotation overnight at 4°C. On the following day, the tubes were spined (5 min. at 1000 g) to collect the beads. At this point, the beads were treated differently depending on whether they were destined for western blotting analysis or mass spectrometry.

For western blot analysis, the beads were washed 6 times in buffer A. In each of the washes, the beads were rotated for 8 min, collected by centrifugation, and resuspended again in a fresh buffer. In the last wash, the beads were transferred into a fresh tube, they were then collected by centrifugation, and the supernatant was removed. Finally, the beads were resuspended in 1x NuPAGE LDS sample buffer with 5% β-mercaptoethanol and 2mM biotin and heated to 98°C for five minutes to separate the biotinylated proteins from the streptavidin beads and denature them.

For mass spectrometry, the beads were first washed 5 times in buffer A, and two more times in 20 mM ammonium bicarbonate. In each of the washes, the SA-beads were rotated for 8 min, collected by centrifugation, and resuspended again in a fresh buffer. In the last wash, the beads were transferred into a fresh tube, they were then collected by centrifugation, and the supernatant was removed. Finally, the beads were frozen at -80°C and sent to the mass spectrometry facility at the University of Birmingham shipped on dry ice.

5.5.3 LC-MS/MS analysis of biotinylated proteins

Liquid chromatography tandem mass spectrometry (LC-MS) is a powerful tool to detect molecules in a sample. In brief, the molecules in the sample are first separated by size using liquid chromatography, and the eluted molecules are directed to the mass spectrometer (MS). Then, the MS nebulizes the molecules using an ionization source, generating charged ions. Subsequently, the charged particles are sorted through a set of magnets, separating the charged particles by their unique mass/change (m/z) ratios. The m/z rations and their relative abundance are sent to a data system to generate a mass spectrum. For this specific analysis, we used label-free quantification (LFQ) which is a reliable and less costly alternative to labelled quantification (Neilson et al., 2011).

The samples were digested with trypsin to release peptides from the beads, which were suitable for mass spectrometry. The samples were processed using the facility's standard protocols, and the mass spectrometer QExactive HF(QEHF) (Thermo Fisher Scientific). Label-free quantification was performed by the University of Birmingham mass spectrometry facility using the MaxLFQ algorithm of the MaxQuant software (Cox et al., 2014; Cox and Mann, 2008).

5.5.4 Proteomic analysis using Perseus

The output files from MaxQuant ("proteinGroups.txt"), which contained the LFQ intensities, were further analysed using the Perseus software (Tyanova et al., 2016). The data were first filtered to eliminate the contaminants, the reverse sequences, and the proteins only identified with modified peptides. Then, the LFQ intensity values were Log_2 transformed, and replicates were grouped into control condition (three replicates), and bait condition (three replicates). Next, the missing LFQ intensity values were imputed by a constant number, which was the lowest Log_2 LFQ value found among all the samples minus 1. The T-test (p-value cut-off value =0.-5) was used to compare the Log_2 LFQ intensities for individual proteins between the bait and the control groups. The volcano plot and the histograms were generated by Perseus and the Venn diagram by Matplotlib-venn (Hunter, 2007) using the data provided by Perseus.

5.6 Western blotting

Protein extracts were denatured in 1x NuPAGE LDS sample buffer with 5% β-mercaptoethanol by heating at 98°C for five minutes. They were then electrophoresed on Invitrogen NuPAGE 4 to 12%, Bis-Tris polyacrylamide gels (Invitrogen), and transferred to an Amersham Hybond ECL nitrocellulose membrane (GE Healthcare). For the co-immunoprecipitation assays, the membranes were blocked for 30 min at room temperature in a milk blocking solution (5% powdered milk, 0.1% Tween-20, in PBS), whereas for proximity-labelling assays, they were blocked in a BSA blocking solution (1% BSA, 0.2% Triton X-100 in PBS). Membranes were then incubated with the primary antibodies diluted in the blocking solution overnight at 4°C or at room temperature for 1 hour. The following primary antibodies were used; dilutions indicated in brackets: anti-β-tubulin (1:2000; DSHB), anti-GFP (1:2000; a290 from Abcam), anti-Myc [1:200; (A-14): sc-789 from Santa Cruz], anti-FLAG (1:250; M2 from sigma), anti-H3 (1:5000; from Millipore). Subsequently, membranes were washed three times for 15 min with PBT (0.1% Tween-20, in PBS). Then, the membranes were incubated with the HRP-conjugated antibodies [anti-Rabbit IgG (1:2000) from fisher scientific, and anti-Mouse IgG (1:2000) from fisherscientific], or with Streptavidin AlexaFluor-680 conjugate (Life Technologies, S21378) diluted in PBT for 1 hour at room temperature. This was followed by three 15 min washes in PBS, and incubation with the chemiluminescent substrate Pierce ECL (Thermo Fisher). For the incubated with Streptavidin Alexa Fluor 580, membranes were washed in PBS only for 5 min. Proteins were detected using ChemiDoc Imaging System.

The Pierce Silver Stain Kit (Thermo Fischer) was used for protein detection in polyacrylamide gels after electrophoresis following the manufacture's manual.

5.7 Immunostaining of *Drosophila* cells

The immunostaining protocol was adapted from (Vrettos et al., 2017). Kc167 cells were seeded in 6-well plates containing poly-L-lysine-coated slides on the bottom and transiently transfected with the plasmids of interest (as described in section 5.4) for two days. After that

time, the medium was removed, and cells on the slides were washed once with PBS gently, and then fixed with 2% formaldehyde in PBS for 15 min. After three rinses in PBS, cells were incubated with PBT (0.2% Triton-X, 0.5% BSA in PBS) for 30 min. Then, they were incubated with the primary antibody anti-GFP (1:1000; a290 from Abcam) diluted in PBT overnight at 4°C. After three 5 min washes in PBS, cells were incubated for 1 h at room temperature with anti-rabbit AlexaFluor-468 (Molecular Probes) diluted in PBS (1:1000), DAPI (1:1000), and Streptavidin AlexaFluor-680 conjugate (1:1000). After three 5 min washes in PBS, cells were mounted using Invitrogen ProLong Gold Antifade Mountant (molecular probes). Kc167 cells were imaged on Zeiss LSM 880 AxioObserver microscope.

6 Appendix

6.1 Lists of oligonucleotides

Name	Target sequence	Primer sequence
pGBKT7-Gro	*gro*	Fw:5'- CATGGAGGCCGAATTCATGTATCCCTCACCGGTG-3' Rv:5'-GCAGGTCGACGGATCCTTAATAAATAACTTCGTAGACAG-3'
pGADT7-Hairy	*hairy*	Fw: 5'- GGAGGCCAGTGAATTCATGGTTACCGGCGTAACAGC -3' Rv: 5'- CGAGCTCGATGGATCCCTACCAGGGCCGCCAGGG -3'
pGADT7-full_Vdup1	*Vdup1*	Fw:5'-GGAGGCCAGTGAATTCATGCCGCGCAAGTTGCTTAAATTC-3' Rv: 5'-CGAGCTCGATGGATCCCTATGCCTCGATAGGCTTCTCGTG-3'
pGADT7-Trl-581	*Trl-581*	Fw: 5'-GGAGGCCAGTGAATTCATGTCGCTGCCAATGAATTCGC-3' Rv: 5'-CGAGCTCGATGGATCCTCAGAGCGTCTGTTGTGTTTGC-3'
pGADT7-Trl-519	*Trl-519*	Fw: 5'-GGAGGCCAGTGAATTCATGTCGCTGCCAATGAATTCGC-3' Rv: 5'-CGAGCTCGATGGATCCCTACTGCGGCTGCGGCTGTTGC-3'
pGADT7-GST-WRPW	*GST-WRPW*	Fw: 5'-GGAGGCCAGTGAATTCATGTCCCCTATACTAGGTTATTGG-3' Rv: 5'-CGAGCTCGATGGATCCCCAGGGCCGCCAGGGGGATCC-3'

Table 5: List of oligonucleotides used to clone the indicated genes into vectors appropriated for Y2H.

Name	Target sequence	Primer sequence
pENTR/D-Gro	*gro*	Fw:5'- CACCATGGAGGCCGAATTATGTATCCCTCACCGGTGC -3' Rv:5'- CAGGTCGACGGATCTTAATAAATAACTTCGTAGACAG -3'
pENTR/D-GST-WRPW	*gst-WRPW*	Fw:5'- CACCATGTCCCCTATACTAGGTTAT -3' Rv: 5'- CCAGGGCCGCCAGGGG -3'
pENTR/D-Trl-581	*Trl-581*	Fw: 5'- CACCATGTCGCTGCCAATG -3' Rv: 5'- TCA GAGCGTCTG TTGTGT -3'
pENTR/D-Trl-519	*Trl-519*	Fw: 5'-CACCATGTCGCTGCCAATG -3' Fw: 5'-CTGCGGCTGCGGCTG -3'

Table 6: List of oligonucleotides used to clone the indicated genes into the entry vector *pENTR/D.*

Name	Target sequence	Primer sequence
pENTR/D-TbID	*TurboID*	Fw: 5'-C ACC AAA GAC AAT ACT GTG CCT CTG AAG CTG ATC-3' Rv: 5'- TCA CAC CTT CCT CTT CTT CTT GGG GTC CAC - 3'
pENTR/D-miniTb	*miniTurbo*	Fw: 5'-C ACC ATC CCG CTG CTG AAC GCT AAA CAG ATT CTG-3' Rv: 5'- TCA CAC CTT CCT CTT CTT CTT GGG GTC CAC - 3'
G/S_linker_TbID	*TurboID*	Fw:5'- GCC CCC TTC ACC AAA GAC AAT ACT GTG CCT CTG AAG CTG ATC -3' Rv: 5'- TGA GCC CGA TGA GCC CGA TGA GCC CTT TTC GGC AGA CCG CAG ACT GAT -3'
pENTR/D-Gro_linker	*pENTR/D-Gro* vector	Fw: 5'- GGC TCA TCG GGC TCA TCG GGC TCA ATG TAT CCC TCA CCG GTG CGC CAC -3' Rv: 5'- AGT ATT GTC TTT GGT GAA GGG GGC GGC CGC GGA GCC TGC TTT -3'
G/S_linker_miniTb	*miniTurbo*	Fw: 5'-GCC CCC TTC ACC ATC CCG CTG CTG AAC GCT AAA CAG ATT CTG-3' Rv: 5'- TGA GCC CGA TGA GCC CGA TGA GCC CTT TTC GGC AGA CCG CAG ACT GAT -3'

Table 7: List of oligonucleotides used to clone the indicated genes into the entry vector *pENTR/D*, and to produce PCR products for Gibson assembly.
Underlined sequences indicate the G/S liker added in the overhang.

6.2 ChIP-seq datasets from public databases used for the bioinformatic analysis

Deposited data	GSE accession number	Reference
Gro ChIP-seq Kc167 cells	Array Express E-MTAB-2316	(Kaul et al., 2014)
Gro ChIP-seq BG3 cells	This study	
Gro ChIP-seq S2R+ cells	Array Express E-MTAB-2316	(Kaul et al., 2014)
GAF ChIP-seq Kc167 cells	GSE80700	(Cubeñas-Potts et al., 2017)
Nelf-E ChIP-seq Kc167 cells	GSE116889	(Nazer et al., 2018)
Cdk9 ChIP-seq Kc167 cells	GSE116889	(Nazer et al., 2018)
Pol II isoforms ChIP-seq Kc167 cells	GSE116889	(Nazer et al., 2018)
Myc ChIP-seq Kc167 cells	GSE39521	(Yang et al., 2013)
DNA-seq BG3 cells	GSE41351	(Lubelsky et al., 2014)
Psq^{BTB} ChIP-seq Kc167 cells	GSE118047	(Gutierrez-Perez et al., 2019)
Psq^{Total} ChIP-seq Kc167 cells	GSE118047	(Gutierrez-Perez et al., 2019)
ISWI ChIP-seq Kc167 cells	GSE118047	(Gutierrez-Perez et al., 2019)
CP190 ChIP-seq Kc167 cells	GSE63518	(Li et al., 2015)
Su(Hw) ChIP-seq Kc167 cells	GSE30740	(Wood et al., 2011)
Pc ChIP-seq Kc167 cells	GSE63518	(Li et al., 2015)
CBP ChIP-seq Kc167 cells	GSE63518	(Li et al., 2015)
Mod(mdg4) ChIP-seq Kc167 cells	GSE36393	(Van Bortle et al., 2012)

Table 8: List of ChIP-seq datasets used for the bioinformatics analysis.

6.3 Supplementary material for section 3.1

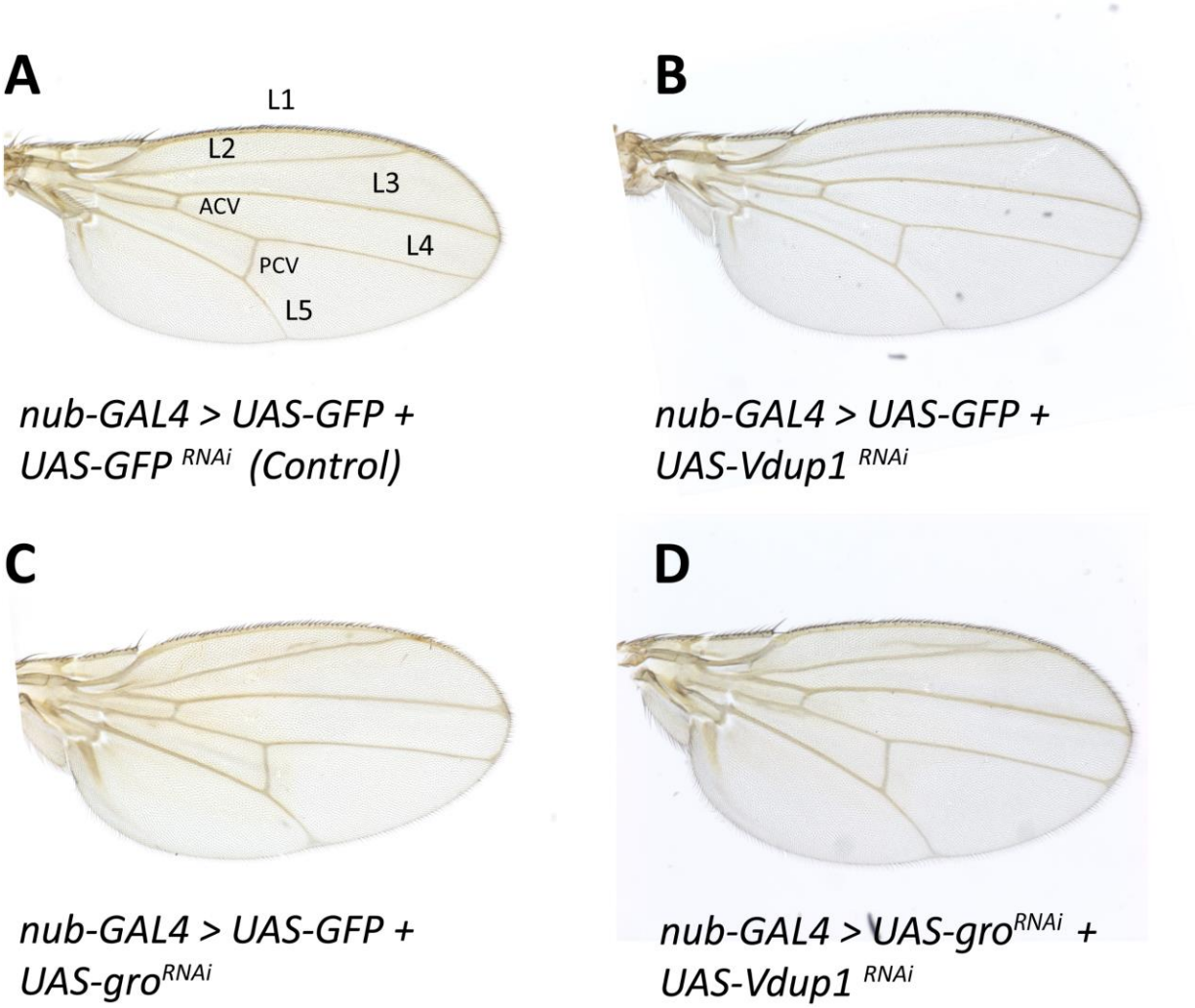


Figure 57 The double RNAi knockdown of *gro*, and *Vdup1* in the developing wing results in no change in the wing vein pattern compared to the single knockdown of these genes.

Representative examples of *Drosophila* adult female wings were raised at 25°C of the indicated genotypes. **(A)** Control wing showing key structures: anterior crossvein (ACV), posterior crossvein (PCV), and longitudinal veins L1 to L5. **(B)** *Vdup1* RNAi knockdown showing no obvious changes in the wing vein pattern. **(C)** *gro* RNAi knockdown showing ectopic venation above L2, and weak longitudinal vein spreading. **(D)** *gro* and *Vdup1* double knockdown showing no obvious change compared to the single *gro* knockdown.

	Total or Partial loss of ACV (%)	Total or Partial loss of PCV (%)	Ectopic venation above L2 (%)	Notching (%)	Wing blistering (%)	Strong vein thickening (%)	Other effects observed (%)
nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+	0	0	7.2		0	0	
nub-GAL4/+; UAS-GFP/ UAS-gro^{RNAi}	9.1	0	100***	0	0	0	
nub-GAL4/ UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+	26.7***	0	81.3***	0	0	0	
nub-GAL4/+; UAS-gro^{RNAi} /UAS-gro^{RNAi}	95.7***	15	100	0	64***	100***	
nub-Gal4/+; UAS-GFP/Trl^{ex15}	0	0	0		0	0	
nubGal4/+; UAS-gro^{RNAi} /Trl^{ex15}	59.2	0	100		0	7.5	
nub-Gal4/Spt4; UAS-GFP/+	0	0	0		0	0	
nub-Gal4/ Spt4; UAS-gro^{RNAi}/+	32.6	0	40.4***		0		
nub-Gal4/ Spt5 (W049); UAS-GFP/+	0	0	0		0	0	
nub-Gal4/ Spt5 (W049); UAS-gro^{RNAi} /+	33.3	0	14.6***		0	0	
nub-Gal4/ Spt5 (MGE); UAS-GFP/+	0	0	2.7		0	0	

nub-Gal4/ Spt5 (MGE); UAS-gro^{RNAi} /+	29.8	0	74.5		0	0	
nub-Gal4/+; UAS-GFP/NELFA^{22}	0	0	0		0	0	
nub-Gal4/+; UAS-gro^{RNAi} /NELFA^{22}	33	0	54.7		0	38***	Increase in ectopic venation, wing folding
nub-Gal4/+; UAS-GFP/Hexim^{E03375}	0	0	2.1		0	0	
nub-Gal4/+; UAS-gro^{RNAi}/ Hexim^{E03375}	10.9	0	98			0	
nub-Gal4/+; UAS-GFP/Hexim^{EY14285}	0	0	0		0	0	
nub-Gal4/+; UAS-gro^{RNAi} /Hexim ^{EY14285}	3.6	0	100		0	0	
nub-Gal4/+; UAS-GFP/Trl^{R67}	0	0	0		0	0	
nubGal4/+; UAS-gro^{RNAi} / Trl^{R67}	40	15	65		20	10	Increase in ectopic venation, wing inflation defects
nub-Gal4/+; UAS-GFP/Nelf-E	0	0	0	0	0	0	
nub-Gal4/+; UAS-gro^{RNAi} /Nelf-E	18.8	0	31.2***	12.5	0	0	

Table 9: Table showing the quantification of wing phenotypes for the indicated genotypes, in which endogenous mutations were in heterozygosity with RNAi *gro* knockdown.

Six different features of the wing were quantified: total or partial loss of anterior crossvein (ACV), total or partial loss of posterior crossvein (PCV), ectopic venation above L2, notching, wing blistering, and strong vein thickening. Strong vein thickening was considered when the wing phenotype belonged either to Class 3 or Class 4 according to the classification system shown in Figure 12 A. The last column in the table mentions additional effects observed. Fisher's exact tests were used for pairwise statistical comparisons in each feature. Single knockdown phenotypes and endogenous mutations in heterozygosis were compared to the control (*nub-GAL4/UAS-GFP^{RNAi}; UAS-GFP/+).* Endogenous lines in heterozygosity with RNAi *gro* knockdown were compared to control heterozygotic endogenous mutations as well as the *gro* knockdowns (including the two single *gro* knockdown genotypes: *nub-GAL4/+; UAS-GFP/ UAS-gro^{RNAi}*, and *nub-GAL4/UAS-GFP^{RNAi}; UAS-gro^{RNAi}/+*). The presence of three asterisks by the number indicates a p-value ≤0.0001. Cells highlighted in yellow indicate the features of the *gro* knockdown in heterozygosis with the endogenous mutations that were significantly increased compared to their single knockdowns and endogenous mutations. Cells highlighted in green indicate the features of the *gro* knockdown in heterozygosis with the endogenous mutations that were significantly reduced compared to their single knockdowns and endogenous mutations.

6.4 Supplementary material for Chapter 3.3

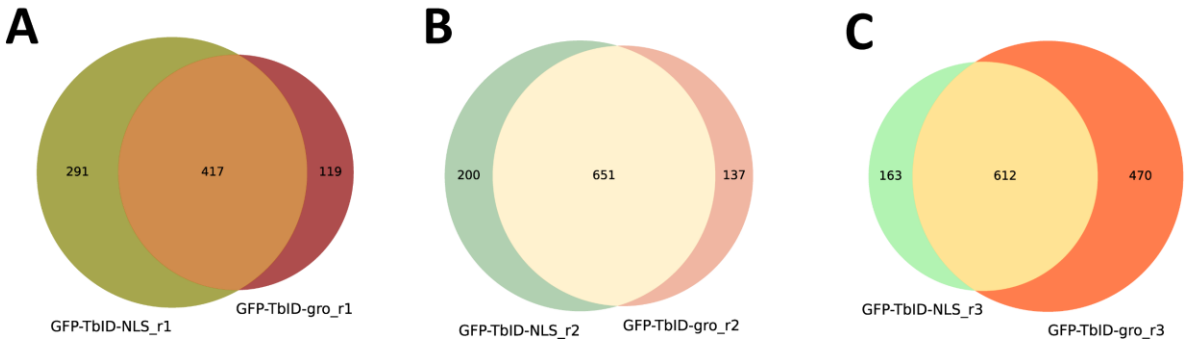


Figure 58: Overlap between the proteins identified in the control group (GFP-TbID-NLS) and the bait group (GFP-TbID-Gro) in the replicates 1, 2 and 3.

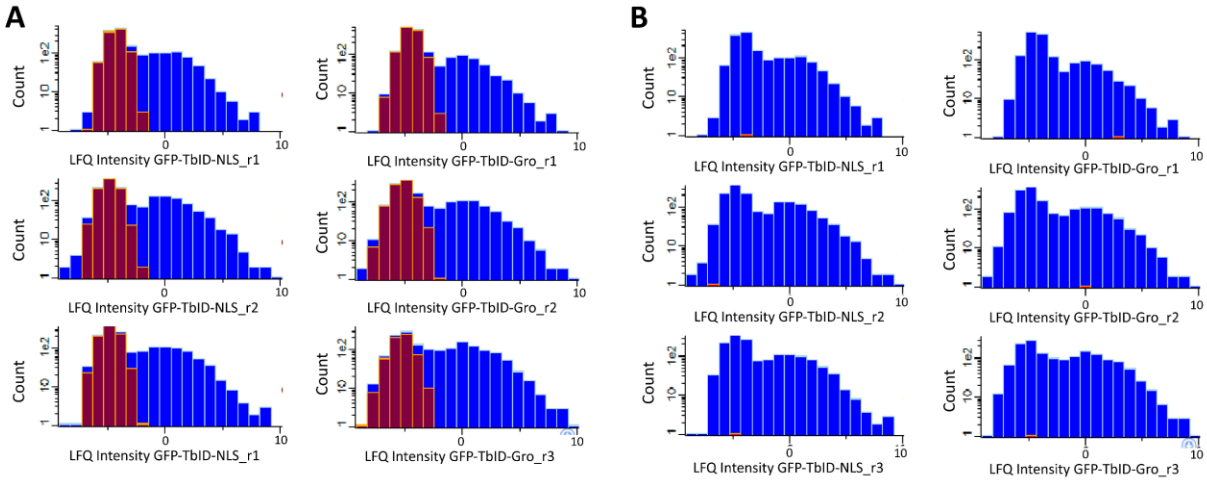


Figure 59: Histograms showing the distribution of the LFQ intensity values after imputation was drawn from a normal distribution.

(A) Histograms showing the logarithmic distribution of the LFQ intensity datasets for each of the three control replicates (GFP-TbID-NLS) and the bait replicates (GFP-TbID-Gro). The contribution of non-imputed and imputed values are coloured in blue and red, respectively. Data were normalized by subtracting the median. **(B)** Same histogram than in A but showing the contribution of Gro in red.

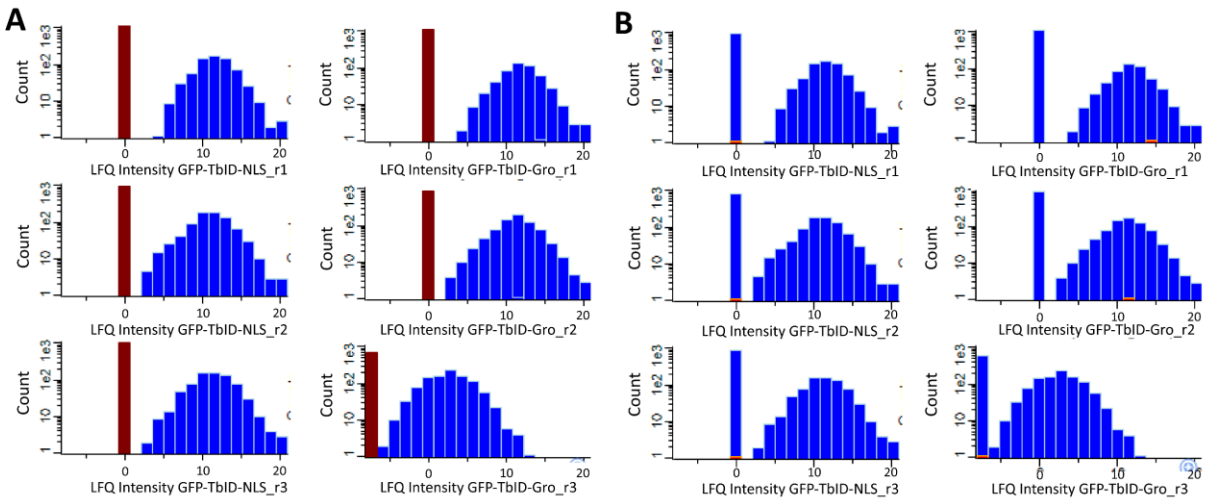


Figure 60: Histograms showing the distribution of the LFQ intensity values after imputation by a constant number.

(A) Histograms showing the logarithmic distribution of the LFQ intensity datasets for each of the three control replicates (GFP-TbID-NLS) and the bait replicates (GFP-TbID-Gro). The contribution of non-imputed and imputed values are coloured in blue and red, respectively. Data were normalized by subtracting the median. **(B)** Same histogram than in A but showing the contribution of Gro in red.

Gene names	TSC
gro;gro-RB	17
Aats-glupro	15
msn	11
Nipped-A	10
Trap1	9
SC35-RB;SC35	9
AnxB11	9
CG31688-RC;Cdc23;cdc23-RA	8
Hsp70B	8
NTPase	8
R	8
Kr-h2	8
CG6554;Art1	7
dare;dare-RA	7
CG8878-RB;CG8878;BcDNA.LD23371	7
Eip55E	7
su(f)	7
Vamp7	7
BCAS2	7
colt	7
BcDNA.GH02220	7
ORCT2;Orct2;ORCT;Orct	6
RhoGAP92B;RhoGAP92B-RA	6
Cdep	6
Cg25C-RC;Cg25C	6
Grip84	6
AGO2	6
tral	6
Rack1	6
CG3887	6
CG3919	6
GstD1	6
fliI	6
ND-15	6
Rrp40	6
CG4972-RA;CG4972	6
IntS11	6
l(2)37Bb	6
Rbbp5	6
Clp	6
dbe	6
sun	6
ND-B12	6
CG3016	6
Mvl-RA;Mvl;Mvl-RB	5
CG1647-RA;CG1647	5
MRG15	5
cg;cg-RF	5
l(2)k09913	5
atl	5
spt4	5
mrj;CG8448	5
CG17912	5
CG32441	5
CG9915;CG9915-RB	5
CG30178-RA	5
Trp1	5
Pitslre	5
lolal	5
g	5
RnrS	5
egl	5
Cp190	5
GstT1	5
Tpc1	5
Patr-1	5
Cpsf100	5
rod	5
mRpS22	5
ebd2	5
CG12018	5
faf	4
ced-6;ced-6-RC	4
CG3499-RC	4
CG4662	4
jar;jar-RG	4
nct	4
bgcn	4
Rpb4	4
CG11985	4
RpL39	4
yrt	4
Psi	4
CstF-64	4
CG1474;Es2-RA;Es2	4
CG13277;LSm7	4

nrv1	4
ECSIT-RA;ECSIT	4
Vha44	4
Rpt4;Rpt4-RA	4
Fim;Fim-RD	4
eRF1;eRF1-RC	4
kuk;kuk-RB	4
CG9662-RA;CG9662	4
CG1307	4
CG7800-RA;CG7800	4
Hr78;Hr78-RB	4
Ranbp16	4
mud	4
Wee1;wee	4
smid	4
Rcd-1	4
Cnx99A;Cnx14D	4
E2f2	4
tra2	4
mts	4
Arp3	4
Arf51F	4
pelo	4
CG31729;CG31729-RA;CG6263	4
CG10171-RB	4
Tsp42Ed	4
ClC-b	4
mRpS4	4
DMAP1	4
p24-2;eca	4

CG3838	4
CG2970	4
Mms19	4
ND-49	4
GLE1	4
Syf2	4
Gfat2	4
jnj	4
Nup58	4
Nup93-2	4
Cyp6d5	4
Pvf2	4
mRpS34	4
mRpL21	4
CG14230	4
cher	3
CalpA	3
Act57B;Act57B-RA;Act87E;Act88F-RA;Act88F;Act79B	3
Rpt2	3
snRNP-U1-C	3
CG18190-RB;Eb1-RF;Eb1;Eb1-RA	3
ORF1	3
Rb97D	3
Cyp1	3
CG3817-RA;CG3817	3
asp-RA;asp	3
CG7824-RC;CG7824-RA	3
Stim-RB;Stim-RA;Stim	3
CG10616-RA;CG10616	3

Su(var)3-7;Su(var)3-7-RA	3
l(2)05714;l(2)05714-RB	3
CG14450-RA	3
cid-RA;cid	3
DNApol-alpha60-RA;DNApol-alpha60	3
Kap-alpha1	3
ken	3
PCNA	3
Rala	3
pnt	3
Imp	3
trmt10c	3
ND-24	3
rictor	3
COX5B	3
CG12170	3
Sirt4	3
HSPC300	3
Prp3	3
CG9330;CG9330-RA	3
CTCF	3
TM9SF4	3
Scox	3
rev7	3
COX8	3
Pex16	3
Cwc25	3
MSBP	3
Upf1	3

Crag	3
ND-B16.6	3
MED18	3
stwl	2
jvl;CG3563-RA	2
Amph	2
Uba1	2
Alhambra;Alh	2
f-cup	2
Synd	2
Mes-4	2
tws-RB;tws	2
awd	2
Akt1	2
EG:87B1.3;bcn92	2
Ssb-c31a	2
CG11799;fd68A	2
ort	2
koi	2
cdc2rk;cdc2rk-RA	2
ana3	2
gag;pol	2
pbl	2
CG8549;CG8549-RA	2
CG11981;Prosbeta3	2
CG3305-RA;Lamp1	2
Lrpprc2;CG14786-RA	2
raptor-RA;raptor	2
l(2)not;l(2)not-RA;l(2)not2	2
CG30403;CG30403-RB	2

CG14200-RA	2
Drat;CG1600	2
Syb-RA;Syb	2
spin-RA;spin	2
CG9674;CG9674-RB	2
bab2	2
CG18292-RA	2
Taf6	2
D19A	2
Mkk4	2
eIF2B-beta	2
EG:100G10.7	2
Gapdh1	2
Cat	2
Vha16-1	2
Sep-01	2
Pdi	2
CG6997;CG32708-RA	2
Cdc6;cdc6	2
nrv2	2
Slh	2
ScpX	2
CG14411;CG14411-RA	2
whd	2
Sec61beta	2
Nup50	2
Mys45A	2
Mad1	2
CG8036	2
HBS1	2

CG10494	2
HspB8;CG14207-RA	2
Ced-12	2
GalNAc-T2	2
mRpS9	2
l(2)09851	2
CG6236-RA;CG6236	2
eIF2B-alpha	2
Rbf2	2
Su(z)12	2
Rab5	2
Sec13	2
meso18E	2
Chrac-16	2
PH4alphaEFB	2
mfrn	2
CG7477	2
CG5270	2
mtTFB2	2
BEST:CK01296	2
Bka	2
borr	2
Trs23	2
CG1218	2
BoYb	2
Rpp30	2
CG3098	2
CHMP2B	2
Liprin-beta	2
Tdrd3	2

rogdi	2
sec3-RA;sec3	2
CG8944	2
mRpS25	2
CtsB1	2
CG12006	2
TAF1C-like	2
CG9752	2
Hmg-2	2
IntS4	2
IntS6	2
epsilonCOP	2
unk	1
bbc;CG6016-RA	1
Rcd1	1
Xe7	1
Ada2b	1
Vha26	1
Rga	1
CG13298	1
CG8311;CG8311-RA	1
CG6967	1
CG18067;CG18067-RA	1
ens;ens-RA	1
CG7504-RC	1
Ptp61F;Ptp61F-RA	1
CG17159-RB	1
ATP8;mt:ATPase8	1
KrT95D-RA;KrT95D	1
Vap-33B;CG33523-RD	1

exo70-RA;exo70	1
Prosalpha1	1
tsu-RA;tsu	1
NK7.1;NK7.1-RA	1
Sac1	1
Aats-pro	1
VhaSFD	1
nxf2	1
CG12010-RB	1
Trip1	1
pav	1
cm	1
AnxB10	1
Dhod	1
Prosalpha2	1
Arp1	1
Cdc27	1
mei-9	1
iPLA2-VIA;CG6718	1
UQCR-6.4	1
Vps35	1
l(2)35Be	1
CG8728-RA	1
mRpS29	1
gag	1
CG13730	1
shrb	1
Srp54k	1
mRpL15	1
dom	1

lace	1
Aac11	1
	1
Surf6	1
Hmt-1	1
mRpL9	1
Sas	1
Blm	1
mRpL19	1
Zif	1
nesd	1
CG10639-RA	1
Vps24	1
mad2	1
Sec63	1
nudC	1
p24-1	1
babos	1
BcDNA.LD23876	1
CG9603	0
p53	0
Taf12	0
didum;didum-RC	0
sno	0
CG17691-RA;CG17691;CG17691-RE	0
Taf8-RA;Taf8	0
Hsc70Cb	0
thoc6	0
E(z)	0
Hop	0

AP-1gamma	0	ND-51	0	CG17904	0
CG18787	0	CG9302-RA;CG9302	0	MFS18	0
HP4	0	Arfip	0		

Table 10: List of proteins detected in exclusively in the bait proteome.

Table is sorted from largest to smaller total spectral count (TSC).

6.5 Supplementary material for section 3.4

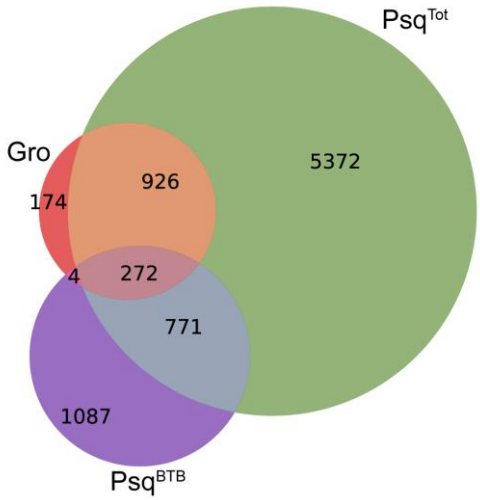


Figure 61: Overlap among ChIP-seq of Gro, Psq^{Total}, and Psq^{BTB} in Kc167 cells.

7 References

Agarwal, M., Kumar, P., Mathew, S.J., 2015. The Groucho/Transducin-like enhancer of split protein family in animal development: TLE Protein Family in Animal Development. IUBMB Life 67, 472–481. https://doi.org/10.1002/iub.1395

Ahringer, J., Gasser, S.M., 2018. Repressive Chromatin in *Caenorhabditis elegans* : Establishment, Composition, and Function. Genetics 208, 491–511. https://doi.org/10.1534/genetics.117.300386

Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H., Roignant, J.-Y., 2019. Promoter-proximal pausing mediated by the exon junction complex regulates splicing. Nat. Commun. 10, 521. https://doi.org/10.1038/s41467-019-08381-0

Albrecht, D., Kniemeyer, O., Brakhage, A.A., Guthke, R., 2010. Missing values in gel-based proteomics. PROTEOMICS 10, 1202–1211. https://doi.org/10.1002/pmic.200800576

Aldaz, S., Escudero, L.M., 2010. Imaginal discs. Curr. Biol. 20, R429–R431. https://doi.org/10.1016/j.cub.2010.03.010

Anderson, A.M., Weasner, B.M., Weasner, B.P., Kumar, J.P., 2012. Dual transcriptional activities of SIX proteins define their roles in normal and ectopic eye development. Development 139, 991–1000. https://doi.org/10.1242/dev.077255

Andersson, E.R., Sandberg, R., Lendahl, U., 2011. Notch signaling: simplicity in design, versatility in function. Dev. Camb. Engl. 138, 3593–3612. https://doi.org/10.1242/dev.063610

Apidianakis, Y., Grbavec, D., Stifani, S., Delidakis, C., 2001. Groucho mediates a Ci-independent mechanism of hedgehog repression in the anterior wing pouch. Dev. Camb. Engl. 128, 4361–4370.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., Stark, A., 2013. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science 339, 1074–1077. https://doi.org/10.1126/science.1232542

Aronson, B.D., Fisher, A.L., Blechman, K., Caudy, M., Gergen, J.P., 1997. Groucho-dependent and -independent repression activities of Runt domain proteins. Mol. Cell. Biol. 17, 5581–5587. https://doi.org/10.1128/MCB.17.9.5581

Bahri, S.M., Chia, W., Yang, X., 2001. The Drosophila homolog of human AF10/AF17 leukemia fusion genes (Dalf) encodes a zinc finger/leucine zipper nuclear protein required in the nervous

system for maintaining EVE expression and normal growth. Mech. Dev. 100, 291–301. https://doi.org/10.1016/S0925-4773(00)00539-6

Bailey, T.L., Machanick, P., 2012. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. 40, e128–e128. https://doi.org/10.1093/nar/gks433

Baluapuri, A., Wolf, E., Eilers, M., 2020. Target gene-independent functions of MYC oncoproteins. Nat. Rev. Mol. Cell Biol. 21, 255–267. https://doi.org/10.1038/s41580-020-0215-2

Barolo, S., 1997. hairy mediates dominant repression in the Drosophila embryo. EMBO J. 16, 2883–2891. https://doi.org/10.1093/emboj/16.10.2883

Barolo, S., Stone, T., Bang, A.G., Posakony, J.W., 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. Genes Dev. 16, 1964–1976. https://doi.org/10.1101/gad.987402

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. Cell 129, 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Beira, J.V., Paro, R., 2016. The legacy of Drosophila imaginal discs. Chromosoma 125, 573–592. https://doi.org/10.1007/s00412-016-0595-4

Beisel, C., Paro, R., 2011. Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. 12, 123–135. https://doi.org/10.1038/nrg2932

Benyajati, C., 1997. Multiple isoforms of GAGA factor, a critical component of chromatin structure. Nucleic Acids Res. 25, 3345–3353. https://doi.org/10.1093/nar/25.16.3345

Blanch, M., Piñeyro, D., Bernués, J., 2015. New insights for *Drosophila* GAGA factor in larvae. R. Soc. Open Sci. 2, 150011. https://doi.org/10.1098/rsos.150011

Boettiger, A.N., Levine, M., 2009. Synchronous and Stochastic Patterns of Gene Activation in the *Drosophila* Embryo. Science 325, 471–473. https://doi.org/10.1126/science.1173976

Boija, A., Mahat, D.B., Zare, A., Holmqvist, P.-H., Philip, P., Meyers, D.J., Cole, P.A., Lis, J.T., Stenberg, P., Mannervik, M., 2017. CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. Mol. Cell 68, 491-503.e5. https://doi.org/10.1016/j.molcel.2017.09.031

Bothma, J.P., Magliocco, J., Levine, M., 2011. The Snail Repressor Inhibits Release, Not Elongation, of Paused Pol II in the Drosophila Embryo. Curr. Biol. 21, 1571–1577. https://doi.org/10.1016/j.cub.2011.08.019

Bradley, T., Cook, M.E., Blanchette, M., 2015. SR proteins control a complex network of RNA-processing events. RNA 21, 75–92. https://doi.org/10.1261/rna.043893.113

Brand, A.H., Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Dev. Camb. Engl. 118, 401–415.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., Ting, A.Y., 2018. Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. 36, 880–887. https://doi.org/10.1038/nbt.4201

Brantjes, H., 2001. All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res. 29, 1410–1419. https://doi.org/10.1093/nar/29.7.1410

Bray, S.J., 2016. Notch signalling in context. Nat. Rev. Mol. Cell Biol. 17, 722–735. https://doi.org/10.1038/nrm.2016.94

Brückner, A., Polge, C., Lentze, N., Auerbach, D., Schlattner, U., 2009. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. Int. J. Mol. Sci. 10, 2763–2788. https://doi.org/10.3390/ijms10062763

Bucher, G., Klingler, M., 2004. Divergent segmentation mechanism in the short germ insect *Tribolium* revealed by *giant* expression and function. Development 131, 1729–1740. https://doi.org/10.1242/dev.01073

Buscarlet, M., Stifani, S., 2007. The 'Marx' of Groucho on development and disease. Trends Cell Biol. 17, 353–361. https://doi.org/10.1016/j.tcb.2007.07.002

Cagan, R., 2009. Chapter 5 Principles of Drosophila Eye Differentiation, in: Current Topics in Developmental Biology. Elsevier, pp. 115–135. https://doi.org/10.1016/S0070-2153(09)89005-4

Cai, Y., 2003. Groucho suppresses Pax2 transactivation by inhibition of JNK-mediated phosphorylation. EMBO J. 22, 5522–5529. https://doi.org/10.1093/emboj/cdg536

Carvalho, L.R., Brinkmeier, M.L., Castinetti, F., Ellsworth, B.S., Camper, S.A., 2010. Corepressors TLE1 and TLE3 Interact with HESX1 and PROP1. Mol. Endocrinol. 24, 754–765. https://doi.org/10.1210/me.2008-0359

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., Bejsovec, A., 1998. Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature 395, 604–608. https://doi.org/10.1038/26982

Chambers, M., Turki-Judeh, W., Kim, M.W., Chen, K., Gallaher, S.D., Courey, A.J., 2017. Mechanisms of Groucho-mediated repression revealed by genome-wide analysis of Groucho binding and activity. BMC Genomics 18, 215. https://doi.org/10.1186/s12864-017-3589-6

Chathoth, K.T., Zabet, N.R., 2019. Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. Genome Res. 29, 613–625. https://doi.org/10.1101/gr.246710.118

Chen, F.X., Smith, E.R., Shilatifard, A., 2018. Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 19, 464–478. https://doi.org/10.1038/s41580-018-0010-5

Chen, G., Courey, A.J., 2000. Groucho/TLE family proteins and transcriptional repression. Gene 249, 1–16. https://doi.org/10.1016/S0378-1119(00)00161-X

Chen, G., Fernandez, J., Mische, S., Courey, A.J., 1999. A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in Drosophila development. Genes Dev. 13, 2218–2230. https://doi.org/10.1101/gad.13.17.2218

Chen, G., Nguyen, P.H., Courey, A.J., 1998. A Role for Groucho Tetramerization in Transcriptional Repression. Mol. Cell. Biol. 18, 7259–7268. https://doi.org/10.1128/MCB.18.12.725

9

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., Samsonova, A., Choi, J.-H., Roberts, J., Davis, C.A., Tang, H., van Baren, M.J., Ghosh, S., Dobin, A., Bell, K., Lin, W., Langton, L., Duff, M.O., Tenney, A.E., Zaleski, C., Brent, M.R., Hoskins, R.A., Kaufman, T.C., Andrews, J., Graveley, B.R., Perrimon, N., Celniker, S.E., Gingeras, T.R., Cherbas, P., 2011. The transcriptional diversity of 25 Drosophila cell lines. Genome Res. 21, 301–314. https://doi.org/10.1101/gr.112961.110

Chetverina, D., Erokhin, M., Schedl, P., 2021. GAGA factor: a multifunctional pioneering chromatin protein. Cell. Mol. Life Sci. 78, 4125–4141. https://doi.org/10.1007/s00018-021-03776-z

Cho, K.F., Branon, T.C., Rajeev, S., Svinkina, T., Udeshi, N.D., Thoudam, T., Kwak, C., Rhee, H.-W., Lee, I.-K., Carr, S.A., Ting, A.Y., 2020. Split-TurboID enables contact-dependent proximity labeling in cells. Proc. Natl. Acad. Sci. 117, 12143–12154. https://doi.org/10.1073/pnas.1919528117

Cho, S., Schroeder, S., Kaehlcke, K., Kwon, H.-S., Pedal, A., Herker, E., Schnoelzer, M., Ott, M., 2009. Acetylation of cyclin T1 regulates the equilibrium between active and inactive P-TEFb in cells. EMBO J. 28, 1407–1417. https://doi.org/10.1038/emboj.2009.99

Chodaparambil, J.V., Pate, K.T., Hepler, M.R.D., Tsai, B.P., Muthurajan, U.M., Luger, K., Waterman, M.L., Weis, W.I., 2014. Molecular functions of the TLE tetramerization domain in Wnt target gene repression. EMBO J. 33, 719–731. https://doi.org/10.1002/embj.201387188

Choi, C.Y., Kim, Y.H., Kim, Y.-O., Park, S.J., Kim, E.-A., Riemenschneider, W., Gajewski, K., Schulz, R.A., Kim, Y., 2005. Phosphorylation by the DHIPK2 Protein Kinase Modulates the Corepressor Activity of Groucho. J. Biol. Chem. 280, 21427–21436. https://doi.org/10.1074/jbc.M500496200

Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.-C., Nesvizhskii, A.I., 2011. SAINT: probabilistic scoring of affinity purification–mass spectrometry data. Nat. Methods 8, 70–73. https://doi.org/10.1038/nmeth.1541

Ciarapica, R., Methot, L., Tang, Y., Lo, R., Dali, R., Buscarlet, M., Locatelli, F., del Sal, G., Rota, R., Stifani, S., 2014. Prolyl isomerase Pin1 and protein kinase HIPK2 cooperate to promote cortical neurogenesis by suppressing Groucho/TLE:Hes1-mediated inhibition of neuronal differentiation. Cell Death Differ. 21, 321–332. https://doi.org/10.1038/cdd.2013.160

Cifuentes, F.J., Garcia-Bellido, A., 1997. Proximo-distal specification in the wing disc of Drosophila by the nubbin gene. Proc. Natl. Acad. Sci. 94, 11405–11410. https://doi.org/10.1073/pnas.94.21.11405

Clark, E., 2017. Dynamic patterning by the Drosophila pair-rule network reconciles long-germ and short-germ segmentation. PLOS Biol. 15, e2002439. https://doi.org/10.1371/journal.pbio.2002439

Copley, R.R., 2005. The EH1 motif in metazoan transcription factors. BMC Genomics 6, 169. https://doi.org/10.1186/1471-2164-6-169

Core, L., Adelman, K., 2019. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. 33, 960–982. https://doi.org/10.1101/gad.325142.119

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., Lis, J.T., 2012. Defining the Status of RNA Polymerase at Promoters. Cell Rep. 2, 1025–1035. https://doi.org/10.1016/j.celrep.2012.08.034

Courey, A.J., Jia, S., 2001. Transcriptional repression: the long and the short of it. Genes Dev. 15, 2786–2796. https://doi.org/10.1101/gad.939601

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., Mann, M., 2014. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol. Cell. Proteomics 13, 2513–2526. https://doi.org/10.1074/mcp.M113.031591

Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372. https://doi.org/10.1038/nbt.1511

Crozatier, M., Glise, B., Vincent, A., 2004. Patterns in evolution: veins of the Drosophila wing. Trends Genet. 20, 498–505. https://doi.org/10.1016/j.tig.2004.07.013

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., Corces, V.G., 2017. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res. 45, 1714–1730. https://doi.org/10.1093/nar/gkw1114

Dasen, J.S., 2001. Temporal regulation of a paired-like homeodomain repressor/TLE corepressor complex and a related activator is required for pituitary organogenesis. Genes Dev. 15, 3193–3207. https://doi.org/10.1101/gad.932601

de Celis, J.F., Ruiz-Gómez, M., 1995. groucho and hedgehog regulate engrailed expression in the anterior compartment of the Drosophila wing. Dev. Camb. Engl. 121, 3467–3476.

De Obaldia, M.E., Bell, J.J., Wang, X., Harly, C., Yashiro-Ohtani, Y., DeLong, J.H., Zlotoff, D.A., Sultana, D.A., Pear, W.S., Bhandoola, A., 2013. T cell development requires constraint of the myeloid regulator C/EBP-α by the Notch target and transcriptional repressor Hes1. Nat. Immunol. 14, 1277–1284. https://doi.org/10.1038/ni.2760

de Pretis, S., Kress, T.R., Morelli, M.J., Sabò, A., Locarno, C., Verrecchia, A., Doni, M., Campaner, S., Amati, B., Pelizzola, M., 2017. Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. Genome Res. 27, 1658–1664. https://doi.org/10.1101/gr.226035.117

Dege, C., Hagman, J., 2014. Mi-2/NuRD chromatin remodeling complexes regulate B and T-lymphocyte development and function. Immunol. Rev. 261, 126–140. https://doi.org/10.1111/imr.12209

Delidakis, C., Monastirioti, M., Magadi, S.S., 2014. E(spl): genetic, developmental, and evolutionary aspects of a group of invertebrate Hes proteins with close ties to Notch signaling. Curr. Top. Dev. Biol. 110, 217–262. https://doi.org/10.1016/B978-0-12-405943-6.00006-3

Diaz-Benjumea, F.J., Hafen, E., 1994. The sevenless signalling cassette mediates Drosophila EGF receptor function during epidermal development. Dev. Camb. Engl. 120, 569–578.

Dominguez, M., 2014. Oncogenic programmes and Notch activity: An 'organized crime'? Semin. Cell Dev. Biol. 28, 78–85. https://doi.org/10.1016/j.semcdb.2014.04.012

Doyle, M.J., Loomis, Z.L., Sussel, L., 2007. Nkx2.2-repressor activity is sufficient to specify α-cells and a small number of β-cells in the pancreatic islet. Development 134, 515–523. https://doi.org/10.1242/dev.02763

Dubnicoff, T., Valentine, S.A., Chen, G., Shi, T., Lengyel, J.A., Paroush, Z. 'e., Courey, A.J., 1997. Conversion of Dorsal from an activator to a repressor by the global corepressor Groucho. Genes Dev. 11, 2952–2957. https://doi.org/10.1101/gad.11.22.2952

Eberhard, D., Jiménez, G., Heavey, B., Busslinger, M., 2000. Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. EMBO J. 19, 2292–2303. https://doi.org/10.1093/emboj/19.10.2292

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., Paro, R., 2011. Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res. 21, 216–226. https://doi.org/10.1101/gr.114348.110

Falo-Sanjuan, J., Lammers, N.C., Garcia, H.G., Bray, S.J., 2019. Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling. Dev. Cell 50, 411-425.e8. https://doi.org/10.1016/j.devcel.2019.07.002

Fant, C.B., Levandowski, C.B., Gupta, K., Maas, Z.L., Moir, J., Rubin, J.D., Sawyer, A., Esbin, M.N., Rimel, J.K., Luyties, O., Marr, M.T., Berger, I., Dowell, R.D., Taatjes, D.J., 2020. TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. Mol. Cell 78, 785-793.e8. https://doi.org/10.1016/j.molcel.2020.03.008

Ferres-Marco, D., Gutierrez-Garcia, I., Vallejo, D.M., Bolivar, J., Gutierrez-Aviño, F.J., Dominguez, M., 2006. Epigenetic silencers and Notch collaborate to promote malignant tumours by Rb silencing. Nature 439, 430–436. https://doi.org/10.1038/nature04376

Fields, S., Song, O., 1989. A novel genetic system to detect protein–protein interactions. Nature 340, 245–246. https://doi.org/10.1038/340245a0

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., van Steensel, B., 2010. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell 143, 212–224. https://doi.org/10.1016/j.cell.2010.09.009

Fisher, A.L., Ohsako, S., 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. https://doi.org/10.1128/MCB.16.6.2670

Flatt, T., 2020. Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*. Genetics 214, 3–48. https://doi.org/10.1534/genetics.119.300160

Flores-Saaib, R.D., Courey, A.J., 2000. Regulation of Dorso/Ventral Patterning in the Drosophila Embryo by Multiple Dorsal-interacting Proteins. Cell Biochem. Biophys. 33, 1–17. https://doi.org/10.1385/CBB:33:1:1

Freeman, M., 1996. Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. Cell 87, 651–660. https://doi.org/10.1016/s0092-8674(00)81385-9

Fu, J., Yoon, H.-G., Qin, J., Wong, J., 2007. Regulation of P-TEFb Elongation Complex Activity by CDK9 Acetylation. Mol. Cell. Biol. 27, 4641–4651. https://doi.org/10.1128/MCB.00857-06

Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A., Lis, J.T., 2015. GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. PLOS Genet. 11, e1005108. https://doi.org/10.1371/journal.pgen.1005108

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A.S., Gaudenz, K., De Kumar, B., Krumlauf, R., Zeitlinger, J., 2012. Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression. Cell Rep. 2, 1670–1683. https://doi.org/10.1016/j.celrep.2012.11.024

Gaertner, B., Zeitlinger, J., 2014. RNA polymerase II pausing during development. Development 141, 1179–1183. https://doi.org/10.1242/dev.088492

Gaskill, M.M., Gibson, T.J., Larson, E.D., Harrison, M.M., 2021. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. eLife 10, e66668. https://doi.org/10.7554/eLife.66668

Gaspar, J.M., 2018. Improved peak-calling with MACS2 (preprint). Bioinformatics. https://doi.org/10.1101/496521

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., Furlong, E.E.M., 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512, 96–100. https://doi.org/10.1038/nature13417

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gilbert SF. Developmental Biology. 6th edition. Sunderland (MA): Sinauer Associates; 2000. The Origins of Anterior-Posterior Polarity. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10039/, n.d. Developmental Biology. 6th edition. Sunderland (MA).

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., Adelman, K., 2008. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. 22, 1921–1933. https://doi.org/10.1101/gad.1643208

Goldstein, R.E., Cook, O., Dinur, T., Pisanté, A., Karandikar, U.C., Bidwai, A., Paroush, Z., 2005. An eh1-Like Motif in Odd-skipped Mediates Recruitment of Groucho and Repression In Vivo. Mol. Cell. Biol. 25, 10711–10720. https://doi.org/10.1128/MCB.25.24.10711-10720.2005

Goldstein, R.E., Jiménez, G., Cook, O., Gur, D., Paroush, Z., 1999. Huckebein repressor activity in Drosophila terminal patterning is mediated by Groucho. Dev. Camb. Engl. 126, 3747–3755.

Gomez-Lamarca, M.J., Falo-Sanjuan, J., Stojnic, R., Abdul Rehman, S., Muresan, L., Jones, M.L., Pillidge, Z., Cerda-Moya, G., Yuan, Z., Baloul, S., Valenti, P., Bystricky, K., Payre, F., O'Holleran, K., Kovall, R., Bray, S.J., 2018. Activation of the Notch Signaling Pathway In Vivo Elicits Changes in CSL Nuclear Dynamics. Dev. Cell 44, 611-623.e7. https://doi.org/10.1016/j.devcel.2018.01.020

Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., Young, R.A., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88. https://doi.org/10.1016/j.cell.2007.05.042

Gutierrez-Perez, I., Rowley, M.J., Lyu, X., Valadez-Graham, V., Vallejo, D.M., Ballesta-Illan, E., Lopez-Atalaya, J.P., Kremsky, I., Caparros, E., Corces, V.G., Dominguez, M., 2019. Ecdysone-Induced 3D Chromatin Reorganization Involves Active Enhancers Bound by Pipsqueak and Polycomb. Cell Rep. 28, 2715-2727.e5. https://doi.org/10.1016/j.celrep.2019.07.096

Hales, K.G., Korey, C.A., Larracuente, A.M., Roberts, D.M., 2015. Genetics on the Fly: A Primer on the *Drosophila* Model System. Genetics 201, 815–842. https://doi.org/10.1534/genetics.115.183392

Hamaratoglu, F., Affolter, M., Pyrowolakis, G., 2014. Dpp/BMP signaling in flies: From molecules to biology. Semin. Cell Dev. Biol. 32, 128–136. https://doi.org/10.1016/j.semcdb.2014.04.036

Harrison, M.M., Li, X.-Y., Kaplan, T., Botchan, M.R., Eisen, M.B., 2011. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. PLoS Genet. 7, e1002266. https://doi.org/10.1371/journal.pgen.1002266

Hartley, D.A., Preiss, A., Artavanis-Tsakonas, S., 1988. A deduced gene product from the Drosophila neurogenic locus, Enhancer of split, shows homology to mammalian G-protein β subunit. Cell 55, 785–795. https://doi.org/10.1016/0092-8674(88)90134-1

Hartley, J.L., 2000. DNA Cloning Using In Vitro Site-Specific Recombination. Genome Res. 10, 1788–1795. https://doi.org/10.1101/gr.143000

Harvey, R., Schuster, E., Jennings, B.H., 2013. Pleiohomeotic Interacts with the Core Transcription Elongation Factor Spt5 to Regulate Gene Expression in Drosophila. PLoS ONE 8, e70184. https://doi.org/10.1371/journal.pone.0070184

Hasson, P., 2001. Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. EMBO J. 20, 5725–5736. https://doi.org/10.1093/emboj/20.20.5725

Hasson, P., Egoz, N., Winkler, C., Volohonsky, G., Jia, S., Dinur, T., Volk, T., Courey, A.J., Paroush, Z., 2005. EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. Nat. Genet. 37, 101–105. https://doi.org/10.1038/ng1486

Hasson, P., 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. https://doi.org/10.1038/sj.bjc.6603019

Hazelett, D.J., Bourouis, M., Walldorf, U., Treisman, J.E., 1998. decapentaplegic and wingless are regulated by eyes absent and eyegone and interact to direct the pattern of retinal differentiation in the eye disc. Dev. Camb. Engl. 125, 3741–3751.

Hendrix, D.A., Hong, J.-W., Zeitlinger, J., Rokhsar, D.S., Levine, M.S., 2008. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc. Natl. Acad. Sci. 105, 7762–7767. https://doi.org/10.1073/pnas.0802406105

Holmqvist, P.-H., Mannervik, M., 2013. Genomic occupancy of the transcriptional co-activators p300 and CBP. Transcription 4, 18–23. https://doi.org/10.4161/trns.22601

Hoy, M.A., 2003. Genetic Systems, Genome Evolution, and Genetic Control of Embryonic Development in Insects, in: Insect Molecular Genetics. Elsevier, pp. 76–126. https://doi.org/10.1016/B978-012357031-4/50023-6

Huang, D., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8, R183. https://doi.org/10.1186/gb-2007-8-9-r183

Huang, D.-H., Chang, Y.-L., Yang, C.-C., Pan, I.-C., King, B., 2002. *pipsqueak* Encodes a Factor Essential for Sequence-Specific Targeting of a Polycomb Group Protein Complex. Mol. Cell. Biol. 22, 6261–6271. https://doi.org/10.1128/MCB.22.17.6261-6271.2002

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Imai, Y., Kurokawa, M., Tanaka, K., Friedman, A.D., Ogawa, S., Mitani, K., Yazaki, Y., Hirai, H., 1998. TLE, the Human Homolog of Groucho, Interacts with AML1 and Acts as a Repressor of AML1-Induced Transactivation. Biochem. Biophys. Res. Commun. 252, 582–589. https://doi.org/10.1006/bbrc.1998.9705

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I.I., Paciello, G., González-Blas, C.B., Koldere, D., Aibar, S., Halder, G., Aerts, S., 2018. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat. Genet. 50, 1011–1020. https://doi.org/10.1038/s41588-018-0140-x

Janssens, H., Crombach, A., Richard Wotton, K., Cicin-Sain, D., Surkova, S., Lu Lim, C., Samsonova, M., Akam, M., Jaeger, J., 2013. Lack of tailless leads to an increase in expression variability in Drosophila embryos. Dev. Biol. 377, 305–317. https://doi.org/10.1016/j.ydbio.2013.01.010

Javed, A., Guo, B., Hiebert, S., Choi, J.Y., Green, J., Zhao, S.C., Osborne, M.A., Stifani, S., Stein, J.L., Lian, J.B., van Wijnen, A.J., Stein, G.S., 2000. Groucho/TLE/R-esp proteins associate with the nuclear matrix and repress RUNX (CBF(alpha)/AML/PEBP2(alpha)) dependent activation of tissue-specific gene transcription. J. Cell Sci. 113 (Pt 12), 2221–2231.

Jennings, B.H., 2013. Pausing for thought: Disrupting the early transcription elongation checkpoint leads to developmental defects and tumourigenesis: Problems & Paradigms. BioEssays 35, 553–560. https://doi.org/10.1002/bies.201200179

Jennings, B.H., 2011. Drosophila – a versatile model in biology & medicine. Mater. Today 14, 190–195. https://doi.org/10.1016/S1369-7021(11)70113-4

Jennings, B.H., Ish-Horowicz, D., 2008. The Groucho/TLE/Grg family of transcriptional co-repressors. Genome Biol. 9, 205. https://doi.org/10.1186/gb-2008-9-1-205

Jennings, B.H., Pickles, L.M., Wainwright, S.M., Roe, S.M., Pearl, L.H., Ish-Horowicz, D., 2006. Molecular Recognition of Transcriptional Repressor Motifs by the WD Domain of the Groucho/TLE Corepressor. Mol. Cell 22, 645–655. https://doi.org/10.1016/j.molcel.2006.04.024

Jennings, B.H., Shah, S., Yamaguchi, Y., Seki, M., Phillips, R.G., Handa, H., Ish-Horowicz, D., 2004. Locus-Specific Requirements for Spt5 in Transcriptional Activation and Repression in Drosophila. Curr. Biol. 14, 1680–1684. https://doi.org/10.1016/j.cub.2004.08.066

Jennings, B.H., Wainwright, S.M., Ish-Horowicz, D., 2008. Differential *in vivo* requirements for oligomerization during Groucho-mediated repression. EMBO Rep. 9, 76–83. https://doi.org/10.1038/sj.embor.7401122

Jimenez, G., Paroush, Z. ’e., Ish-Horowicz, D., 1997. Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev. 11, 3072–3082. https://doi.org/10.1101/gad.11.22.3072

Judd, J., Duarte, F.M., Lis, J.T., 2021. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. Genes Dev. 35, 147–156. https://doi.org/10.1101/gad.341768.120

Kaltenbrun, E., Greco, T.M., Slagle, C.E., Kennedy, L.M., Li, T., Cristea, I.M., Conlon, F.L., 2013. A Gro/TLE-NuRD Corepressor Complex Facilitates Tbx20-Dependent Transcriptional Repression. J. Proteome Res. 12, 5395–5409. https://doi.org/10.1021/pr400818c

Kao, S.-Y., Calman, A.F., Luciw, P.A., Peterlin, B.M., 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature 330, 489–493. https://doi.org/10.1038/330489a0

Karpievitch, Y., Stanley, J., Taverner, T., Huang, J., Adkins, J.N., Ansong, C., Heffron, F., Metz, T.O., Qian, W.-J., Yoon, H., Smith, R.D., Dabney, A.R., 2009. A statistical framework for protein quantitation in bottom-up MS-based proteomics. Bioinformatics 25, 2028–2034. https://doi.org/10.1093/bioinformatics/btp362

Kaul, A., Schuster, E., 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. https://doi.org/10.1371/journal.pgen.1004595

Kaul, A.K., Schuster, E.F., Jennings, B.H., 2015. Recent insights into Groucho co-repressor recruitment and function. Transcription 6, 7–11. https://doi.org/10.1080/21541264.2014.1000709

Kenyon, K.L., Yang-Zhou, D., Cai, C.Q., Tran, S., Clouser, C., Decene, G., Ranade, S., Pignoni, F., 2005. Partner specificity is essential for proper function of the SIX-type homeodomain proteins Sine oculis and Optix during fly eye development. Dev. Biol. 286, 158–168. https://doi.org/10.1016/j.ydbio.2005.07.017

Khan, A., Mathelier, A., 2017. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18, 287. https://doi.org/10.1186/s12859-017-1708-7

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., Xi, R., Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.K., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C.R., Kuroda, M.I., Pirrotta, V., Karpen, G.H., Park, P.J., 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471, 480–485. https://doi.org/10.1038/nature09725

Kim, D.I., Kc, B., Zhu, W., Motamedchaboki, K., Doye, V., Roux, K.J., 2014. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. 111, E2453–E2461. https://doi.org/10.1073/pnas.1406459111

Kim, H.S., Tan, Y., Ma, W., Merkurjev, D., Destici, E., Ma, Q., Suter, T., Ohgi, K., Friedman, M., Skowronska-Krawczyk, D., Rosenfeld, M.G., 2018. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. Nature 556, 510–514. https://doi.org/10.1038/s41586-018-0048-8

Koreman, G.T., Xu, Y., Hu, Q., Zhang, Z., Allen, S.E., Wolfner, M.F., Wang, B., Han, C., 2021. Upgraded CRISPR/Cas9 tools for tissue-specific mutagenesis in *Drosophila*. Proc. Natl. Acad. Sci. 118, e2014255118. https://doi.org/10.1073/pnas.2014255118

Kwong, P.N., Chambers, M., Vashisht, A.A., Turki-Judeh, W., Yau, T.Y., Wohlschlegel, J.A., Courey, A.J., 2015. The Central Region of the Drosophila Co-repressor Groucho as a Regulatory Hub. J. Biol. Chem. 290, 30119–30130. https://doi.org/10.1074/jbc.M115.681171

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., Levine, M.S., 2013. Paused Pol II Coordinates Tissue Morphogenesis in the Drosophila Embryo. Cell 153, 976–987. https://doi.org/10.1016/j.cell.2013.04.045

Lai, E.C., 2004. Notch signaling: control of cell communication and cell fate. Dev. Camb. Engl. 131, 965–973. https://doi.org/10.1242/dev.01074

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Larkin, A., Marygold, S.J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P.V., Goodman, J.L., Gramates, L.S., Millburn, G., Strelets, V.B., Tabone, C.J., Thurmond, J., FlyBase Consortium, Perrimon, N., Gelbart, S.R., Agapite, J., Broll, K., Crosby, M., dos Santos, G., Falls, K., Gramates, L.S., Jenkins, V., Longden, I., Matthews, B., Sutherland, C., Tabone, C.J., Zhou, P., Zytkovicz, M., Brown, N., Antonazzo, G., Attrill, H., Garapati, P., Larkin, A., Marygold, S., McLachlan, A., Millburn, G., Pilgrim, C., Ozturk-Colak, A., Trovisco, V., Kaufman, T., Calvi, B., Goodman, J., Strelets, V., Thurmond, J., Cripps, R., Lovato, T., 2021. FlyBase: updates to the *Drosophila melanogaster* knowledge base. Nucleic Acids Res. 49, D899–D907. https://doi.org/10.1093/nar/gkaa1026

Larochelle, M., Bergeron, D., Arcand, B., Bachand, F., 2019. Proximity-dependent biotinylation by TurboID to identify protein-protein interaction networks in yeast. J. Cell Sci. jcs.232249. https://doi.org/10.1242/jcs.232249

Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., Dow, J.A.T., 2018. FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res. 46, D809–D815. https://doi.org/10.1093/nar/gkx976

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., Gilmour, D.S., 2008. NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. Mol. Cell. Biol. 28, 3290–3300. https://doi.org/10.1128/MCB.02224-07

Lehmann, M., Siegmund, T., Lintermann, K.-G., Korge, G., 1998. The Pipsqueak Protein of *Drosophila melanogaster* Binds to GAGA Sequences through a Novel DNA-binding Domain. J. Biol. Chem. 273, 28504–28509. https://doi.org/10.1074/jbc.273.43.28504

Lehmann, R., Nüsslein-Volhard, C., 1991. The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. Dev. Camb. Engl. 112, 679–691.

Levanon, D., Goldstein, R.E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z., Groner, Y., 1998. Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc. Natl. Acad. Sci. 95, 11590–11595. https://doi.org/10.1073/pnas.95.20.11590

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Gilmour, D.S., 2013. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. 32, 1829–1841. https://doi.org/10.1038/emboj.2013.111

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H.Q., Ong, C.-T., Cubeñas-Potts, C., Hu, M., Lei, E.P., Bosco, G., Qin, Z.S., Corces, V.G., 2015. Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Mol. Cell 58, 216–231. https://doi.org/10.1016/j.molcel.2015.02.023

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456, 400–403. https://doi.org/10.1038/nature07388

Liber, D., Domaschenz, R., Holmqvist, P.-H., Mazzarella, L., Georgiou, A., Leleu, M., Fisher, A.G., Labosky, P.A., Dillon, N., 2010. Epigenetic Priming of a Pre-B Cell-Specific Enhancer through Binding of Sox2 and Foxd3 at the ESC Stage. Cell Stem Cell 7, 114–126. https://doi.org/10.1016/j.stem.2010.05.020

Linderson, Y., Eberhard, D., Malin, S., Johansson, A., Busslinger, M., Pettersson, S., 2004. Corecruitment of the Grg4 repressor by PU.1 is critical for Pax5-mediated repression of B-cell-specific genes. EMBO Rep. 5, 291–296. https://doi.org/10.1038/sj.embor.7400089

Lindsley D L & Grell E H. Genetic variations of Drosophila melanogaster.Washington, DC: Carnegie Institution, 1968., n.d. 1.

Liu, K., Shen, D., Shen, J., Gao, S.M., Li, B., Wong, C., Feng, W., Song, Y., 2017. The Super Elongation Complex Drives Neural Stem Cell Fate Commitment. Dev. Cell 40, 537-551.e6. https://doi.org/10.1016/j.devcel.2017.02.022

Liu, X., Gogate, A.A., Tastemel, M., Malladi, V.S., Yao, H., Nguyen, K., Huang, L.J.-S., Bai, X., 2017. Dynamic change of transcription pausing through modulating NELF protein stability regulates granulocytic differentiation. Blood Adv. 1, 1358–1367. https://doi.org/10.1182/bloodadvances.2017008383

Lobry, C., Oh, P., Mansour, M.R., Look, A.T., Aifantis, I., 2014. Notch signaling: switching an oncogene to a tumor suppressor. Blood 123, 2451–2459. https://doi.org/10.1182/blood-2013-08-355818

Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A.V., Moresco, J.J., Blokhina, T., Wolle, D., Aoki, T., Ryabykh, V., Yates, J.R., Shidlovskii, Y.V., Georgiev, P., Schedl, P., Chetverina, D., 2017. The GAGA factor regulatory network: Identification of GAGA factor associated proteins. PLOS ONE 12, e0173602. https://doi.org/10.1371/journal.pone.0173602

Lubelsky, Y., Prinz, J.A., DeNapoli, L., Li, Y., Belsky, J.A., MacAlpine, D.M., 2014. DNA replication and transcription programs respond to the same chromatin cues. Genome Res. 24, 1102–1114. https://doi.org/10.1101/gr.160010.113

Lundgren, D.H., Hwang, S.-I., Wu, L., Han, D.K., 2010. Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics 7, 39–53. https://doi.org/10.1586/epr.09.69

Madrigal, P., Krajewski, P., 2012. Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. Front. Genet. 3. https://doi.org/10.3389/fgene.2012.00230

Manika Pal, B., 2017. GAGA Factor Expedites Development in Drosophila. Open J. Biol. Sci. 004–011. https://doi.org/10.17352/ojbs.000009

Mannewik, M., Nibu, Y., Levine, H.Z.M., 1999. Transcriptional Coregulators in Development 284, 4.

Marshall, N.F., Price, D.H., 1995. Purification of P-TEFb, a Transcription Factor Required for the Transition into Productive Elongation. J. Biol. Chem. 270, 12335–12338. https://doi.org/10.1074/jbc.270.21.12335

Martinez, C.A., Arnosti, D.N., 2008. Spreading of a Corepressor Linked to Action of Long-Range Repressor Hairy. Mol. Cell. Biol. 28, 2792–2802. https://doi.org/10.1128/MCB.01203-07

McDaniel, S.L., Gibson, T.J., Schulz, K.N., Fernandez Garcia, M., Nevil, M., Jain, S.U., Lewis, P.W., Zaret, K.S., Harrison, M.M., 2019. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. Mol. Cell 74, 185-195.e4. https://doi.org/10.1016/j.molcel.2019.01.014

McNamara, R.P., Bacon, C.W., D'Orso, I., 2016. Transcription elongation control by the 7SK snRNP complex: Releasing the pause. Cell Cycle 15, 2115–2123. https://doi.org/10.1080/15384101.2016.1181241

Medo, M., Aebersold, D.M., Medová, M., 2019. ProtRank: bypassing the imputation of missing values in differential expression analysis of proteomic data. BMC Bioinformatics 20, 563. https://doi.org/10.1186/s12859-019-3144-3

Meyer, V., Oliver, B., Pauli, D., 1998. Multiple Developmental Requirements of Noisette, the *Drosophila* Homolog of the U2 snRNP-Associated Polypeptide SF3a ^{60}. Mol. Cell. Biol. 18, 1835–1843. https://doi.org/10.1128/MCB.18.4.1835

Mieszczanek, J., de la Roche, M., Bienz, M., 2008. A role of Pygopus as an anti-repressor in facilitating Wnt-dependent transcription. Proc. Natl. Acad. Sci. 105, 19324–19329. https://doi.org/10.1073/pnas.0806098105

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560. https://doi.org/10.1038/nature06008

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., Lis, J.T., 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 25, 742–754. https://doi.org/10.1101/gad.2005511

Missra, A., Gilmour, D.S., 2010. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. Proc. Natl. Acad. Sci. 107, 11301–11306. https://doi.org/10.1073/pnas.1000681107

Morel, V., Lecourtois, M., Massiani, O., Maier, D., Preiss, A., Schweisguth, F., 2001. Transcriptional repression by Suppressor of Hairless involves the binding of a Hairless-dCtBP complex in Drosophila. Curr. Biol. 11, 789–792. https://doi.org/10.1016/S0960-9822(01)00224-X

Morisalo, D., Anderson, K.V., 1995. Signaling Pathways that Establish the Dorsal-Ventral Pattern of the *Drosophila* Embryo. Annu. Rev. Genet. 29, 371–399. https://doi.org/10.1146/annurev.ge.29.120195.002103

Muhr, J., Andersson, E., Persson, M., Jessell, T.M., Ericson, J., 2001. Groucho-Mediated Transcriptional Repression Establishes Progenitor Cell Pattern and Neuronal Fate in the Ventral Neural Tube. Cell 104, 861–873. https://doi.org/10.1016/S0092-8674(01)00283-5

Mummery-Widmer, J.L., Yamazaki, M., Stoeger, T., Novatchkova, M., Bhalerao, S., Chen, D., Dietzl, G., Dickson, B.J., Knoblich, J.A., 2009. Genome-wide analysis of Notch signalling in Drosophila by transgenic RNAi. Nature 458, 987–992. https://doi.org/10.1038/nature07936

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., Adelman, K., 2007. RNA polymerase is poised for activation across the genome. Nat. Genet. 39, 1507–1511. https://doi.org/10.1038/ng.2007.21

Nagarkar, S., Wasnik, R., Govada, P., Cohen, S., Shashidhara, L.S., 2020. Promoter Proximal Pausing Limits Tumorous Growth Induced by the Yki Transcription Factor in *Drosophila*. Genetics 216, 67–77. https://doi.org/10.1534/genetics.120.303419

Nagel, A.C., Krejci, A., Tenin, G., Bravo-Patiño, A., Bray, S., Maier, D., 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. https://doi.org/10.1128/MCB.25.23.10433-10441.2005

Nagel, A.C., Preiss, A., 2011. Fine tuning of Notch signaling by differential co-repressor recruitment during eye development of Drosophila. Hereditas 148, 77–84. https://doi.org/10.1111/j.1601-5223.2011.02221.x

Nazer, E., Dale, R.K., Palmer, C., Lei, E.P., 2018. Argonaute2 attenuates active transcription by limiting RNA Polymerase II elongation in Drosophila melanogaster. Sci. Rep. 8, 15685. https://doi.org/10.1038/s41598-018-34115-1

Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C., Haynes, P.A., 2011. Less label, more free: Approaches in label-free quantitative mass spectrometry. PROTEOMICS 11, 535–553. https://doi.org/10.1002/pmic.201000553

Nevil, M., Gibson, T.J., Bartolutti, C., Iyengar, A., Harrison, M.M., 2020. Establishment of chromatin accessibility by the conserved transcription factor Grainy head is developmentally regulated. Development dev.185009. https://doi.org/10.1242/dev.185009

Ngian, Z., Lin, W., Ong, C., 2021. NELF-A controls *Drosophila* healthspan by regulating heat-shock protein-mediated cellular protection and heterochromatin maintenance. Aging Cell 20. https://doi.org/10.1111/acel.13348

Nguyen, D., Fayol, O., Buisine, N., Lecorre, P., Uguen, P., 2016. Functional Interaction between HEXIM and Hedgehog Signaling during Drosophila Wing Development. PLOS ONE 11, e0155438. https://doi.org/10.1371/journal.pone.0155438

Nguyen, D., Krueger, B.J., Sedore, S.C., Brogie, J.E., Rogers, J.T., Rajendra, T.K., Saunders, A., Matera, A.G., Lis, J.T., Uguen, P., Price, D.H., 2012. The Drosophila 7SK snRNP and the essential role of dHEXIM in development. Nucleic Acids Res. 40, 5283–5297. https://doi.org/10.1093/nar/gks191

Nibu, Y., 2001. Local action of long-range repressors in the Drosophila embryo. EMBO J. 20, 2246–2253. https://doi.org/10.1093/emboj/20.9.2246

Nien, C.-Y., Liang, H.-L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R., Rushlow, C., 2011. Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo. PLoS Genet. 7, e1002339. https://doi.org/10.1371/journal.pgen.1002339

Nishinaka, Y., Masutani, H., Oka, S., Matsuo, Y., Yamaguchi, Y., Nishio, K., Ishii, Y., Yodoi, J., 2004. Importin α1 (Rch1) Mediates Nuclear Translocation of Thioredoxin-binding Protein-2/Vitamin D3-up-regulated Protein 1. J. Biol. Chem. 279, 37559–37565. https://doi.org/10.1074/jbc.M405473200

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801. https://doi.org/10.1038/287795a0

Nuthall, H.N., Joachim, K., 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. https://doi.org/10.1128/MCB.24.19.8395-8407.2004

Oberacker, T., Bajorat, J., Ziola, S., Schroeder, A., Röth, D., Kastl, L., Edgar, B.A., Wagner, W., Gülow, K., Krammer, P.H., 2018. Enhanced expression of thioredoxin-interacting-protein regulates oxidative DNA damage and aging. FEBS Lett. 592, 2297–2307. https://doi.org/10.1002/1873-3468.13156

Orian, A., Delrow, J.J., Rosales Nieves, A.E., Abed, M., Metzger, D., Paroush, Z., Eisenman, R.N., Parkhurst, S.M., 2007. A Myc Groucho complex integrates EGF and Notch signaling to regulate neural development. Proc. Natl. Acad. Sci. 104, 15771–15776. https://doi.org/10.1073/pnas.0707418104

Ott, M., Geyer, M., Zhou, Q., 2011. The Control of HIV Transcription: Keeping RNA Polymerase II on Track. Cell Host Microbe 10, 426–435. https://doi.org/10.1016/j.chom.2011.11.002

Paparidis, N.F. dos S., Durvale, M.C., Canduri, F., 2017. The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. Mol. Biosyst. 13, 246–276. https://doi.org/10.1039/C6MB00387G

Papizan, J.B., Singer, R.A., Tschen, S.-I., Dhawan, S., Friel, J.M., Hipkens, S.B., Magnuson, M.A., Bhushan, A., Sussel, L., 2011. Nkx2.2 repressor complex regulates islet -cell specification and prevents -to- -cell reprogramming. Genes Dev. 25, 2291–2305. https://doi.org/10.1101/gad.173039.111

Park, P.J., 2009. ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 10, 669–680. https://doi.org/10.1038/nrg2641

Paroush, Z., 1994. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. Cell 79, 805–815. https://doi.org/10.1016/0092-8674(94)90070-1

Patel, B., 2009. Simple, Fast, and Efficient Cloning of PCR Products with TOPO ® Cloning Vectors. BioTechniques 46, 476. https://doi.org/10.2144/000113158

Peng, J., Marshall, N.F., Price, D.H., 1998. Identification of a Cyclin Subunit Required for the Function of *Drosophila* P-TEFb. J. Biol. Chem. 273, 13855–13860. https://doi.org/10.1074/jbc.273.22.13855

Pereira, A., Paro, R., 2017. Pho dynamically interacts with Spt5 to facilitate transcriptional switches at the hsp70 locus. Epigenetics Chromatin 10, 57. https://doi.org/10.1186/s13072-017-0166-9

Pickles, L.M., Roe, S.M., Hemingway, E.J., Stifani, S., Pearl, L.H., 2002. Crystal Structure of the C-Terminal WD40 Repeat Domain of the Human Groucho/TLE1 Transcriptional Corepressor. Structure 10, 751–761. https://doi.org/10.1016/S0969-2126(02)00768-2

Pinto, M., 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. https://doi.org/10.1074/jbc.271.51.33026

Plet, A., Eick, D., Blanchard, J.M., 1995. Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. Oncogene 10, 319–328.

Quinlan, A.R., 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinforma. 47. https://doi.org/10.1002/0471250953.bi1112s47

Quiring, R., Walldorf, U., Kloter, U., Gehring, W., 1994. Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science 265, 785–789. https://doi.org/10.1126/science.7914031

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., Young, R.A., 2010. c-Myc Regulates Transcriptional Pause Release. Cell 141, 432–445. https://doi.org/10.1016/j.cell.2010.03.030

Rahl, P.B., Young, R.A., 2014. MYC and Transcription Elongation. Cold Spring Harb. Perspect. Med. 4, a020990–a020990. https://doi.org/10.1101/cshperspect.a020990

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., Manke, T., 2014. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191. https://doi.org/10.1093/nar/gku365

Ramirez, J., Martinez, A., Lectez, B., Lee, S.Y., Franco, M., Barrio, R., Dittmar, G., Mayor, U., 2015. Proteomic Analysis of the Ubiquitin Landscape in the Drosophila Embryonic Nervous System and the Adult Photoreceptor Cells. PLOS ONE 10, e0139083. https://doi.org/10.1371/journal.pone.0139083

Reddington, J.P., Garfield, D.A., Sigalova, O.M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R.R., Degner, J.F., Ohler, U., Furlong, E.E.M., 2020. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. Dev. Cell 55, 648-664.e9. https://doi.org/10.1016/j.devcel.2020.10.009

Romano, G., Holodkov, N., Klima, R., Feiguin, F., 2021. TDP-43 regulates GAD1 mRNA splicing and GABA signaling in Drosophila CNS. Sci. Rep. 11, 18761. https://doi.org/10.1038/s41598-021-98241-z

Roth, M., Bonev, B., Lindsay, J., Lea, R., Panagiotaki, N., Houart, C., Papalopulu, N., 2010. FoxG1 and TLE2 act cooperatively to regulate ventral telencephalon formation. Development 137, 1553–1562. https://doi.org/10.1242/dev.044909

Rougvie, A.E., Lis, J.T., n.d. The RNA Polymerase II Molecule at the 5' End of the Uninduced hspTOGene of D. melanogaster Is Transcriptionally Engaged 10.

Roux, K.J., Kim, D.I., Burke, B., May, D.G., 2018. BioID: A Screen for Protein-Protein Interactions. Curr. Protoc. Protein Sci. 91. https://doi.org/10.1002/cpps.51

Roux, K.J., Kim, D.I., Raida, M., Burke, B., 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196, 801–810. https://doi.org/10.1083/jcb.201112098

Sabò, A., Lusic, M., Cereseto, A., Giacca, M., 2008. Acetylation of Conserved Lysines in the Catalytic Core of Cyclin-Dependent Kinase 9 Inhibits Kinase Activity and Regulates Transcription. Mol. Cell. Biol. 28, 2201–2212. https://doi.org/10.1128/MCB.01557-07

Sainsbury, S., Bernecky, C., Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 129–143. https://doi.org/10.1038/nrm3952

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089

Schwendemann, A., Lehmann, M., 2002. Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. Proc. Natl. Acad. Sci. 99, 12883–12888. https://doi.org/10.1073/pnas.202341499

Sekiya, T., 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. https://doi.org/10.1016/j.molcel.2007.10.002

Shao, W., Alcantara, S.G.-M., Zeitlinger, J., 2019. Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. eLife 8, e41461. https://doi.org/10.7554/eLife.41461

Sharma, M., Brantley, J.G., Vassmer, D., Chaturvedi, G., Baas, J., Vanden Heuvel, G.B., 2009. The homeodomain protein Cux1 interacts with Grg4 to repress p27 kip1 expression during kidney development. Gene 439, 87–94. https://doi.org/10.1016/j.gene.2009.03.014

Shimmi, O., Matsuda, S., Hatakeyama, M., 2014. Insights into the molecular mechanisms underlying diversified wing venation among insects. Proc. R. Soc. B Biol. Sci. 281, 20140264. https://doi.org/10.1098/rspb.2014.0264

Shimojima, T., 2003. Drosophila FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. Genes Dev. 17, 1605–1616. https://doi.org/10.1101/gad.1086803

Sjöqvist, M., Andersson, E.R., 2019. Do as I say, Not(ch) as I do: Lateral control of cell fate. Dev. Biol. 447, 58–70. https://doi.org/10.1016/j.ydbio.2017.09.032

Skalska, L., Stojnic, R., Li, J., Fischer, B., Cerda-Moya, G., Sakai, H., Tajbakhsh, S., Russell, S., Adryan, B., Bray, S.J., 2015. Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. EMBO J. 34, 1889–1904. https://doi.org/10.15252/embj.201489923

Song, H., Hasson, P., Paroush, Z., Courey, A.J., 2004. Groucho oligomerization is required for repression in vivo. Mol. Cell. Biol. 24, 4341–4350. https://doi.org/10.1128/mcb.24.10.4341-4350.2004

St Johnston, D., Driever, W., Berleth, T., Richstein, S., Nüsslein-Volhard, C., 1989. Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. Dev. Camb. Engl. 107 Suppl, 13–19.

Stifani, S., Blaumueller, C.M., Redhead, N.J., Hill, R.E., Artavanis-Tsakonas, S., 1992. Human homologs of a Drosophila Enhancer of split gene product define a novel family of nuclear proteins. Nat. Genet. 2, 343. https://doi.org/10.1038/ng1292-343a

Strobl, L.J., Eick, D., 1992. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. EMBO J. 11, 3307–3314. https://doi.org/10.1002/j.1460-2075.1992.tb05409.x

Swingler, T.E., Bess, K.L., Yao, J., Stifani, S., Jayaraman, P.-S., 2004. 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. 279, 34938–34947. https://doi.org/10.1074/jbc.M404488200

Tadros, W., Lipshitz, H.D., 2009. The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042. https://doi.org/10.1242/dev.033183

Tettey, T.T., Gao, X., Shao, W., Li, H., Story, B.A., Chitsazan, A.D., Glaser, R.L., Goode, Z.H., Seidel, C.W., Conaway, R.C., Zeitlinger, J., Blanchette, M., Conaway, J.W., 2019. A Role for FACT in RNA Polymerase II Promoter-Proximal Pausing. Cell Rep. 27, 3770-3779.e7. https://doi.org/10.1016/j.celrep.2019.05.099

Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192. https://doi.org/10.1093/bib/bbs017

Tolkunova, E.N., Fujioka, M., Kobayashi, M., Deka, D., Jaynes, J.B., 1998. Two Distinct Types of Repression Domain in Engrailed: One Interacts with the Groucho Corepressor and Is Preferentially Active on Integrated Target Genes. Mol. Cell. Biol. 18, 2804–2814. https://doi.org/10.1128/MCB.18.5.2804

Tsukiyama, T., Daniel, C., Tamkun, J., Wu, C., 1995. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell 83, 1021–1026. https://doi.org/10.1016/0092-8674(95)90217-1

Tsukiyama, T., Wu, C., 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell 83, 1011–1020. https://doi.org/10.1016/0092-8674(95)90216-3

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J., 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods 13, 731–740. https://doi.org/10.1038/nmeth.3901

Uv, A.E., Harrison, E.J., Bray, S.J., 1997. Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol. Cell. Biol. 17, 6727–6735. https://doi.org/10.1128/MCB.17.11.6727

Vallejo, D.M., Caparros, E., Dominguez, M., 2011. Targeting Notch signalling by the conserved miR-8/200 microRNA family in development and cancer cells: Notch pathway tuned by microRNAs miR-8/200c. EMBO J. 30, 756–769. https://doi.org/10.1038/emboj.2010.358

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., Corces, V.G., 2012. Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res. 22, 2176–2187. https://doi.org/10.1101/gr.136788.111

Villanueva, C.J., Waki, H., Godio, C., Nielsen, R., Chou, W.-L., Vargas, L., Wroblewski, K., Schmedt, C., Chao, L.C., Boyadjian, R., Mandrup, S., Hevener, A., Saez, E., Tontonoz, P., 2011. TLE3 Is a Dual-Function Transcriptional Coregulator of Adipogenesis. Cell Metab. 13, 413–427. https://doi.org/10.1016/j.cmet.2011.02.014

Vissers, J.H.A., Manning, S.A., Kulkarni, A., Harvey, K.F., 2016. A Drosophila RNAi library modulates Hippo pathway-dependent tissue growth. Nat. Commun. 7, 10368. https://doi.org/10.1038/ncomms10368

Vrettos, N., Maragkakis, M., Alexiou, P., Mourelatos, Z., 2017. Kc167, a widely used Drosophila cell line, contains an active primary piRNA pathway. RNA N. Y. N 23, 108–118. https://doi.org/10.1261/rna.059139.116

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., Speck, N.A., 1996. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc. Natl. Acad. Sci. 93, 3444–3449. https://doi.org/10.1073/pnas.93.8.3444

Wang, X., Hang, S., Prazak, L., Gergen, J.P., 2010. NELF Potentiates Gene Transcription in the Drosophila Embryo. PLoS ONE 5, e11498. https://doi.org/10.1371/journal.pone.0011498

Wang, X., Lee, C., Gilmour, D.S., Gergen, J.P., 2007. Transcription elongation controls cell fate specification in the Drosophila embryo. Genes Dev. 21, 1031–1036. https://doi.org/10.1101/gad.1521207

Wheeler, J.C., VanderZwan, C., Xu, X., Swantek, D., Tracey, W.D., Gergen, J.P., 2002. Distinct in vivo requirements for establishment versus maintenance of transcriptional repression. Nat. Genet. 32, 206–210. https://doi.org/10.1038/ng942

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., Adelman, K., 2015. Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. Mol. Cell 58, 311–322. https://doi.org/10.1016/j.molcel.2015.02.003

Wingender, E., Schoeps, T., Haubrock, M., Dönitz, J., 2015. TFClass: a classification of human transcription factors and their rodent orthologs. Nucleic Acids Res. 43, D97–D102. https://doi.org/10.1093/nar/gku1064

Winkler, C.J., Ponce, A., Courey, A.J., 2010. Groucho-Mediated Repression May Result from a Histone Deacetylase-Dependent Increase in Nucleosome Density. PLoS ONE 5, e10166. https://doi.org/10.1371/journal.pone.0010166

Wood, A.M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B.C., Jones, K.C., Corces, V.G., 2011. Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. Mol. Cell 44, 29–38. https://doi.org/10.1016/j.molcel.2011.07.035

Wysocka, J., Reilly, P.T., Herr, W., 2001. Loss of HCF-1–Chromatin Association Precedes Temperature-Induced Growth Arrest of tsBN67 Cells. Mol. Cell. Biol. 21, 3820–3829. https://doi.org/10.1128/MCB.21.11.3820-3829.2001

Xu, J., Liu, H., Park, J.-S., Lan, Y., Jiang, R., 2014. Osr1 acts downstream of and interacts synergistically with Six2 to maintain nephron progenitor cells during kidney organogenesis. Development 141, 1442–1452. https://doi.org/10.1242/dev.103283

Yadav, D., Ghosh, K., Basu, S., Roeder, R.G., Biswas, D., 2019. Multivalent Role of Human TFIID in Recruiting Elongation Components at the Promoter-Proximal Region for Transcriptional Control. Cell Rep. 26, 1303-1317.e7. https://doi.org/10.1016/j.celrep.2019.01.012

Yang, J., Sung, E., Donlin-Asp, P.G., Corces, V.G., 2013. A subset of Drosophila Myc sites remain associated with mitotic chromosomes colocalized with insulator proteins. Nat. Commun. 4, 1464. https://doi.org/10.1038/ncomms2469

Yao, J., Lai, E., Stifani, S., 2001. The Winged-Helix Protein Brain Factor 1 Interacts with Groucho and Hes Proteins To Repress Transcription. Mol. Cell. Biol. 21, 1962–1972. https://doi.org/10.1128/MCB.21.6.1962-1972.2001

Yao, J., Liu, Y., Lo, R., Tretjakoff, I., Peterson, A., Stifani, S., 2000. Disrupted development of the cerebral hemispheres in transgenic mice expressing the mammalian Groucho homologue Transducin-like-Enhancer of split 1 in postmitotic neurons. Mech. Dev. 93, 105–115. https://doi.org/10.1016/S0925-4773(00)00278-1

Yu, G., Wang, L.-G., He, Q.-Y., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Yuan, D., Yang, X., Yuan, Z., Zhao, Y., Guo, J., 2017. TLE1 function and therapeutic potential in cancer. Oncotarget 8, 15971–15976. https://doi.org/10.18632/oncotarget.13278

Zahavi, T., Maimon, A., Kushnir, T., Lange, R., Berger, E., Kornspan, D., Grossman, R., Anzi, S., Shaulian, E., Karni, R., Nechushtan, H., Paroush, Z., 2017. Ras–Erk signaling induces phosphorylation of human TLE1 and downregulates its repressor function. Oncogene 36, 3729–3739. https://doi.org/10.1038/onc.2016.517

Zaret, K.S., Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 25, 2227–2241. https://doi.org/10.1101/gad.176826.111

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., Young, R.A., 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. 39, 1512–1516. https://doi.org/10.1038/ng.2007.26

Zeng, X., Han, L., Singh, S.R., Liu, H., Neumüller, R.A., Yan, D., Hu, Y., Liu, Y., Liu, W., Lin, X., Hou, S.X., 2015. Genome-wide RNAi Screen Identifies Networks Involved in Intestinal Stem Cell Regulation in Drosophila. Cell Rep. 10, 1226–1238. https://doi.org/10.1016/j.celrep.2015.01.051

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Y., Song, G., Lal, N.K., Nagalakshmi, U., Li, Y., Zheng, W., Huang, P., Branon, T.C., Ting, A.Y., Walley, J.W., Dinesh-Kumar, S.P., 2019. TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. Nat. Commun. 10, 3252. https://doi.org/10.1038/s41467-019-11202-z

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., Green, M.R., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237. https://doi.org/10.1186/1471-2105-11-237