# Chapter 1

# Introduction:

# Groucho – A Multifunctional Regulator of Drosophila Development

## Introduction

*The Groucho/TLE family of corepressors are ubiquitous regulators of animal development*

The Groucho/TLE (Gro) family of of corepressors play crucial roles in the interpretation and integration of multiple spatial, temporal, and signaling inputs during development in higher eukaryotes. Groucho, the sole *Drosophila melanogaster* member of this protein family, was first discovered in the context of a slight hypomorphic allele which resulted in the formation of extra supraorbital bristles reminiscent of the bushy eyebrows of Groucho Marx {Lindsley, 1968 #3055}. Subsequent research on Gro in *Drosophila* has served to characterize this factor’s central importance to developmental gene regulation in response to multiple developmental programs and signaling pathways. As a corepressor, Groucho has no documented ability to bind DNA directly in a sequence-specific manner, instead relying on recruitment to genomic loci through interactions with a diverse array of transcriptional repressors. Through its interactions with these repressors, it is essential fie almost all aspects of embryonic and imaginal *Drosophila* development {Paroush, 1994 #3090}. In humans, Gro/TLE family proteins are involved in such processes as organ development, adipogenesis, neurogenesis, hematopoiesis, and osteogenesis {Bajoghli, 2005 #3068}{Villanueva, 2011 #1659}{Javed, 2000 #3070}{Metzger, 2012 #2956}.

Groucho consists of five domains, two of which are highly conserved throughout higher eukaryotes {Chen, 2000 #3056}{Turki-Judeh, 2012 #2385}. A great body of work has arisen documenting the contributions of each domain to the overall function and regulation of Groucho. While much of this work has focused on the N- and C- terminal domains, as they are more conserved and more sensitive to point mutagenesis {Jennings, 2006 #3059}{Jennings, 2007 #2990}, the central domains of Groucho have been investigated for their roles in Groucho activity through interaction with a number of regulatory targets, including protein kinases, histones, and histone modifying enzymes {Turki-Judeh, 2012 #2385}.

Homologs of Groucho with similar roles in developmental decision making have been identified throughout metazoans (Fig. 1.1) {Paroush, 1994 #3090}. Homologs have been identified and characterized to various extents in rats {Schmidt, 1993 #3126}, nematodes {Pflugrad, 1997 #3125}, frogs {Choudhury, 1997 #3121}, zebrafish {Wulbeck, 1997 #3127}, mice {Mallo, 1993 #3123}, and humans {Stifani, 1992 #3065}. While the *Drosophila* and *C. elegans* genomes each encode single Gro family genes, the mouse, chick, and human genomes each encode four members, while zebrafish and medaka each encode six members {Li, 2000 #3128}. The full-length human Gro orthologs are termed transducin-like Enhancer of Split 1-4 (TLE1-4) {Miyasaka, 1993 #3124}, are expressed combinatorially during cell differentiation and have non-redundant roles during development {Stifani, 1992 #3065;Yao, 1998 #3069}.

Mammalian genomes additionally encode two truncated Gro homologs, *Amino Enhancer of Splt (AES)*, which is homologous to the two N-terminal domains of Groucho {Gasperowicz, 2005 #3129}, and *Tle6/Grg6,* which possesses a poorly conserved N-terminal region and a C-terminal WD-repeat domain {Dang, 2001 #3130}. Both factors are thought to antagonize the activity of full-length TLE family members. AES may function by directly binding to TLE proteins through Q-domain interactions {Brantjes, 2001 #3058} or by interacting with a subset of TLE-dependent repressors {Muhr, 2001 #3132}. Similarly, TLE6/Grg6 has been shown to interact with repressors to block recruitment of full-length TLE family proteins and thereby alleviate repression {Marcal, 2005 #3131}. More distantly related Gro homologs have been identified in yeast (Tup1) and plants (TOPLESS) {Courey, 2001 #3139}{Smith, 2000 #3169}{Lee, 2012 #2961}.

*The domain architecture of Groucho/TLE family proteins*

The N-terminal Q (glutamine rich) domain is one of the two highly conserved domains and is responsible for the formation of tetramers and potentially higher-order oligomers of Gro {Chen, 1998 #267}. Additionally, the Q-domain mediates a subset of interactions with transcriptional repressors, including the Tcf/Lef family of proteins {Brantjes, 2001 #3058}. The structure of the Q-domain of TLE1, a human homologue of Gro, was recently solved, revealing that the domain forms a dimer of dimers consisting of two coiled-coils interdigitated in a head-to-head complex {Chodaparambil, 2014 #3057}. The resulting structure provides an elegant explanation of the mechanics of tetramerization, and corroborates the large frictional coefficient measured in hydrodynamic studies of the purified Q-domain, as the predicted structure is thin and rod-like {Kuo, 2011 #506}.

The ability of the Q domain to direct the formation of high-order oligomers has been proposed to mediate the spreading of Gro along chromatin allowing for the establishment of large transcriptionally silent domains. This might explain the documented ability of Gro to direct long-range repression in which entire loci are organized into transcriptionally silent states. In support of this idea, assays involving Grg3, a mouse homolog of Gro, on *in vitro* chromatin arrays, showed that oligomerization mediated through the Q-domain is not required for recruitment of Gro to chromatin, but is required for subsequent aggregation of chromatinized fragments into a form that was resistant to transcription {Sekiya, 2007 #1658}.

Contrary to the idea that the Q domain could mediate spreading, chromatin immunoprecipitation (ChIP) assays in cell culture revealed that oligomerization-deficient mutants of *Drosophila* Gro exhibited similar median peak widths to wild-type Gro {Kaul, 2014 #2204}. The interpretation of this result is somewhat complicated by the fact that binding data was generated from two *Drosophila* cell lines depleted of endogenous Groucho via RNAi and overexpressing either GFP-tagged wild-type or oligomerization-deficient Groucho. The authors showed a significant reduction of endogenous Gro that nonetheless remained detectable by immunoblot. Thus, it remains a possibility that low levels of endogenous Groucho were contributing to peak formation or spreading in both contexts.

Regardless of the role of oligomerization in the definition of the size of Groucho binding domains, loss of oligomerization does result in significant differences in the recruitment patterns of overexpressed wild-type and oligomerization-deficient mutants. Of the approximately 3000 distinct Groucho binding sites identified in Kc167 cells expressing wild-type or oligomerization-deficient Gro samples, 48% are unique to a single condition {Kaul, 2014 #2204}. Loss of oligomerization potential therefore, while preserving many aspects of wild-type Groucho binding patterns, does disrupt Groucho association with chromatin in some contexts, the nature of which remains unexplained.

The WD-domain is the second conserved domain of Gro and comprises the C-terminal XXX amino acids of the protein of the protein. The WD-domain consists of a seven-bladed β-propeller domain, and is responsible for the majority of Groucho interactions with DNA-binding repressors (Table 1-1) {Pickles, 2002 #3060}. The majority of these interactions are mediated through binding of the WD-domain to short peptide motifs {Jennings, 2006 #3059}, which are recognized by the central pore of the propeller domain. Several such peptide motifs have been identified in Groucho-interacting proteins. The majority of these peptide motifs fall into one of two categories. C-terminal WRPW/Y recognition sequences have been found in Hairy/Enhancer of split (HES) and Runt family transcription factors {Jimenez, 1997 #3075;Fisher, 1996 #3133;Paroush, 1994 #3090; Aronson, 1997 #3095;Canon, 2003 #3174}. And the engrailed homology domain-1 (eh1) motif is an internal peptide motif with the consensus sequence FxIxxIL that is found in Engrailed, Dorsal, Odd-skipped, and Goosecoid, among others {Copley, 2005 #3134}{Dubnicoff, 1997 #2366}{Jiménez, 1997 #264}{Jimenez, 1999 #3092}{Smith, 1996 #3135}{Tolkunova, 1998 #3136}. The WD domain binds to these motifs with differing affinities. These differences are utilized in controlling the recruitment of Groucho to specific factors. For example, the affinity of Groucho for binding the eh1-like motif of Dorsal is relatively weak {Flores-Saaib, 2000 #656}, necessitating the assistance of additional factors in facilitating a stable interaction between the two proteins. This weak affinity of the Dorsal/Groucho interaction is crucial to allowing Dorsal to function as a bifunctional transcription factor, as mutation of this motif to a higher-affinity sequence abolishes Dorsal’s ability to activate genes in the embryo, due to constitutive recruitment of Groucho {Ratnaparkhi, 2006 #3108}.

The WD-repeat domain may be involved in additional protein interactions. Studies of Grg3, a mouse Gro/TLE family member, have shown that the WD domain is critical for binding to histone arrays *in vitro* as well as condensation of these arrays{Sekiya, 2007 #1658}. The observation that the Q domain is also capable of strong interaction with K20 methylated H4 tails suggests multiple levels of interaction between Gro/TLE proteins and histones, and may contribute to the protein’s ability to associate with histones both locally, at its recruitment site, and distantly, through association with non-contiguous stretches of chromatin {Chodaparambil, 2014 #2275}.

The central region of Groucho is divided into three domains, the GP, CcN, and SP domains. The GP domain binds to a histone deacetylase (HDAC1/Rpd3), which is involved with some but not all Groucho-repressive activity {Chen, 1999 #3061}. The CcN domain is involved in Groucho regulation, containing multiple Ck2 and Cdc2 phosphorylation sites {Nuthall, 2002 #3062}. The SP domain contains multiple sites phosphorylated in response to MAPK signaling, resulting in down-regulation of Groucho activity via nuclear export {Hasson, 2005 #3064}. There is evidence that the central regions of Groucho are intrinsically disordered {Turki-Judeh, 2012 #2966}, which has emerged as a common strategy among eukaryotic proteins to facilitate participation in diverse protein-protein interactions, expose signaling motifs, and/or accept posttranslational modifications {Dunker, 2008 #3091}.

*Groucho integrates multiple signaling pathways to generate specific cellular responses and fates*

In *Drosophila,* Groucho’s roles in responses to signaling pathways are well documented. The factor participates in Ras/MAPK, Notch, Decapentapletic (Dpp/BMP), and Wingless/Wnt signaling, among others. Groucho activity is down-regulated via the Ras/MAPK pathway in response to signals initiated at multiple receptor tyrosine kinases (RTKs) such as EGFR, FGFR, and Torso {Hasson, 2005 #3064;Cinnamon, 2008 #242}. The resulting relief of Groucho-mediated repression is critical to the cellular response to RTK signaling and is thought to precipitate in cellular memory, whereby the attenuation of Groucho activity persists after loss of signaling {Cinnamon, 2008 #242} {Helman, 2011 #2938}.

In the absence of Notch signaling, Groucho represses *E(spl)* complex genes through interactions with Hairy, which is itself associated with Su(H), a sequence-specific transcription factor that targets Notch-responsive genes {Delidakis, 1991 #3082}. Recruitment of a Notch ligand to Notch transmembrane receptors activates the pathway, leading to proteolytic cleavage of the receptor and subsequent release of the Notch Intracellular Domain (Notch ICD). The Notch ICD rapidly enters the nucleus, where it displaces Hairy binding at Su(H) sites, relieving Groucho repression and initiating expression of *E(spl)* genes. Groucho then interacts with newly expressed E(spl) family proteins to repress a number of proneural genes {Preiss, 1988 #3083} {Wurmbach, 1999 #3084}. This repressive activity is alleviated by MAPK signaling, which results in the phosphorylation of Gro, negatively affecting its ability to repress these proneural genes in cooperation with E(spl) members {Andersson, 2011 #3168}. The partial or complete negation of Notch signaling through the activation of the MAPK pathway thus represents a Groucho-mediated point of crosstalk between the two pathways {Hasson, 2005 #3064}.

Groucho is also critical to signaling via Decapentaplegic (dpp), a *Drosophila* TGF-β homolog whose diffusion over long distances is essential to patterning during embryogenesis and later during appendage development {Upadhyai, 2013 #2339}. The Dpp morphogen is expressed dorsally in the embryo and is required for the definition of cell-fate along the dorsal-ventral axis {Ferguson, 1992 #3088}. Groucho, through interaction with Dorsal, represses the ventral expression of *dpp*, meaning that Gro is involved in both the spatiotemporal definition and interpretation of dpp signaling{Schwyter, 1995 #3038}*.* In the absence of Dppsignaling, Brinker (Brk) represses a subset of dpp target genes through two independent repressive mechanisms, one involving dCtBP (a short-range corepressor), and the other involving Gro {Hasson, 2001 #3033}. Upon activation of Dpp signaling, Brinker becomes repressed by Schnurri in dorsal regions of the embryo, while continuing to be expressed in ventrolateral regions {Marty, 2000 #3089}.

Finally, Groucho participates in Wingless/Wnt signaling, through interactions with Tcf/Lef family proteins, to regulate cell-fate choice {Cavallo, 1998 #3071}{Roose, 1999 #3086}. In unstimulated cells, Groucho assists in repressing Tcf/Lef target genes through interactions with the Q-domain {Clevers, 2006 #3085}. Upon Wnt activation, nuclear beta-catenin (Armadillo) concentration increases, which binds to Tcf, releasing Groucho and leading to gene activation. In this context, Groucho is essential in guarding against spurious activation of Wnt target genes in unstimulated cells {Daniels, 2005 #3087}.

While there are hundreds of cell types in the adult fly, far fewer developmental signaling pathways have been documented {Perrimon, 2012 #3175}. To generate this cellular complexity, informational content from multiple extracellular signals must be interpreted within each cell’s specific spatial and temporal context {Hsueh, 2009 #3177}. Even with this ability to simultaneously respond to multiple signals, the high number of discrete transcriptional states required during development necessitates that these signals are integrated non-additively {Housden, 2014 #3176}. Factors that participate in multiple signaling pathways, such as Groucho, are a necessary component of a non-additive response. Groucho therefore presents a convenient node through which a cell can process limited combinations of inputs to produce a larger number of outcomes.

*Groucho is an essential component of the embryonic axial patterning network*

It is primarily through the spatially and temporally controlled regulation of gene transcription that Groucho becomes fundamental to embryonic patterning. Many early embryonic patterning proteins can be divided into effectors of the dorsal-ventral and anterior-posterior programs, though these processes are complex and highly interconnected {Jaeger, 2012 #3103}, requiring the coordinated regulation of dozens of transcriptional activators, repressors, and co-regulators {Mannervik, 2014 #2280}. Definition of the dorsal-ventral axis, which is critical to germ layer development, is carried out by the maternally-contributed gradient of nuclear Dorsal along this axis {Roth, 1989 #1112}. Dorsal is a sequence-specific transcription factor, and the strength, spacing, and grouping of Dorsal binding sites, along with the distribution of adjacent binding sites for other interacting factors modulate Dorsal binding and cofactor recruitment in order to correctly interpret the Dorsal gradient {Zeitlinger, 2007 #3025}.

On the ventral side of the embryo, high concentrations of nuclear Dorsal initiate transcriptional programs that determine the mesoderm {Gonzalez-Crespo, 1993 #3043}. In ventrolateral regions, modest Dorsal concentrations help direct a neuroectodermal fate {Ip, 1992 #3042}. Dorsal also acts as a repressor of dorsal ectodermal genes and, by keeping them off in ventral and ventrolateral region, it restricts there expression to the dorsal ectodermal primordium. {Zeitlinger, 2007 #3025}Groucho is required for this repression and plays a critical role in switching Dorsal from an activator to a repressor {Dubnicoff, 1997 #2366}.

In addition to its roles in dorsal/ventral patterning, Groucho has multiple roles in anterior/posterior pattern formation. For example, it is required for repression by numerous segmentation gene products such as Hairy, Runt, and Engrailed., {Levine, 2008 #3104}Groucho also required for the patterning of the anterior and posterior terminal domains by the Torso RTK through its interaction with Capicua {Ajuria, 2011 #2947}, a process regulated by Ras/MAPK signaling {Chen, 2009 #3073} {Paroush, 1997 #3074}. Capicua recruits Gro to the *tailless* and *huckebein* throughout the embryo maintaining these genes in an off state. Torso RTK then activates Ras/MAPK signaling at the termini leading to the phosphorylation and consequent inactivation of both Capicua and Gro at the embryonic termini allowing the expression of *tll* and hkb as required for specification of terminal fate {Winkler, 2010 #2964}.

*Groucho is capable of both short- and long-range repression*

Transcriptional repressors in *Drosophila* can be classified as acting as either short- or long-range repressors dependent on their ability to counteract the regulatory potential of local (within ~100 bp) or distant (thousands of bp away or more) activating elements or promoters {Gray, 1994 #3138} {Gray, 1996 #3137}. Some repressors are specific for one type of repression, while others can adopt a short- or long-range repressive activity through association with multiple corepressors operating via distinct mechanisms of repression {Courey, 2001 #3139}. Groucho was originally considered a long-range co-repressor recruited exclusively by long-range repressors such as Hairy and Dorsal {Cai, 1996 #3140} {Dubnicoff, 1997 #2366}. CtBP, in contrast, is a well-studied corepressor capable of short-range repression when recruited by such short-range repressors as Kruppel, Giant, and Snail {Nibu, 1998 #3142;Nibu, 2001 #3141}.

Evidence that Groucho could oligomerize and potentially crosslink non-contiguous regions of chromatin provides a mechanistic explanation for its ability to quench distant regulatory elements. More recently, it was found that in some contexts Groucho behaves as a short-range corepressor. Groucho appears to be recruited by Knirps, a short-range repressor capable of interacting with CtBP, to repress the expression of *even-skipped* {Payankaulam, 2009 #2955}*.* Sloppy-paired 1 (Slp1), a Groucho-interacting repressor, is involved in the short-range repression of regulatory elements controlling the expression of multiple pair-rule genes {Andrioli, 2004 #3143}. If Groucho is in fact commonly utilized as both a short- and long-range repressor, this sheds light on the observation that Groucho oligomerization is required in a context-dependent manner *in vivo* {Jennings, 2007 #2990}, suggesting a mechanism whereby Groucho oligomerization is necessary for long-range repression but dispensable for short-range. Likely the classification of repressors as short- and long-range actors, while a useful abstraction when classifying repressors, masks much of the complexity of repressive activity that would be provided by a thorough understanding of repressive mechanisms.

*The mechanism of Groucho-mediated repression*

While a great deal is known about the developmental participation and interactors of Gro, details of the mechanism by which Gro achieves repression have remained elusive. Multiple models have been proposed to explain Groucho’s ability to fully and reversibly initiate and maintain both short- and long-range repression, yet a full picture, able to account for all observations of Groucho behavior, has yet to emerge.

Much of the speculation surrounding Groucho activity centers on the possibility that oligomeric structures *in vivo*, how these structures interact with chromatin, and what relevance these structures have to repression. Early evidence showed that Groucho tetramerizes *in vitro* via the Q-domain {Chen, 1998 #267} {Song, 2004 #1161}. Groucho was found to be associated with chromatin over 2 kb away from its recruitment site, leading to the hypothesis that Groucho spreads from its recruitment site, analogous to the spreading activity of Sir family corepressors {Pirrotta, 2005 #3106}. Experiments on a mouse Gro homolog showed that while tetramerization is not required for recruitment to chromatin, it is necessary for the aggregation of nucleosomal arrays in vitro {Sekiya, 2007 #1658}. Monomeric forms of the protein successfully bind to and increase the density of dinucleosomes *in vitro* {Sekiya, 2007 #1658}*.*  *In vivo,* the loss of tetramerization is lethal but does not entirely abolish Gro-mediated repression {Jennings, 2007 #2990}. More recent evidence in cell culture has shown that Gro binds in discrete peaks less than 1kb in width, though longer stretches of binding do occur {Kaul, 2014 #2204}.

Gro preferentially associates with histone tails and can do so without the involvement of additional DNA-binding interacting factors {Flores-Saaib, 2000 #656} {Sekiya, 2007 #1658}. Additionally, Gro associates with a histone deacetylase, HDAC1/Rpd3 {Chen, 1999 #3105}. This association accounts for some but not all of Groucho’s repressive ability *in vivo*, where Groucho binding is associated with decreased acetylation of the tails of histones H3 and H4, as well as increased nucleosome density{Winkler, 2010 #2964}. Colocalization of Gro and Rpd3 is prevalent in Kc167 cells (a cell line derived from *Drosophila* embryos), with over half of Groucho binding sites found to overlap Rpd3 binding {Kaul, 2014 #2204}.

Given the many gaps in our knowledge regarding the mechanisms of Gro-mediated repression, I carried out a genome-wide analysis of Gro function in hopes of filling in some of these gaps. Experiments described in Chapter 2, employing a combination of Gro-ChIP-seq on staged wild-type embryos and RNA-seq on staged embryos expressing different levels of Gro show that Groucho associates with chromatin in discrete < 1 kilobase peaks, often clustered closely upstream or within regulated genes. This data was used to generate a set of high-confidence Groucho targets at multiple developmental stages. Experiments described in Chapter 3, employing nascent-seq on staged wild-type embryos show that Groucho-regulated genes are enriched for promoter-proximal paused polymerase, suggesting a possible role for PolII stalling in Groucho-mediated gene repression. Chapter 4, is a published paper in which we identified the Gro interactome as a way of illuminating mechanisms of Gro-mediated repression.

**Figure 1-1. Groucho/TLE family proteins are partially conserved throughout metazoans.** The Gro/TLE family of corepressors are typified by five domains defined based on function and sequence. Domain-wise homology to the *D. melanogaster* Groucho is indicated by percentages, when significant. Two domains, the N-terminal Q domain and the C-terminal WD-repeat domain are well conserved while the central region, consisting of the GP, CcN, and SP domains shares little sequence homology between species. The Q domain is involved in association with repressor and the formation of homo-oligomeric Groucho complexes. The WD domain is additionally involved in repressor association. The central region is predicted to be intrinsically disordered and serves as a scaffold for a number of protein interactions, notably with Rpd3, a histone deacetylase involved in some aspects of Groucho-mediated repression. The central regions also serve as a regulatory region of Groucho via being target for multiple post-translational modifications.

**Fig. 1-1ch1_introduction.figures/ch1_introduction.figures.split.1.pdf**

**Table 1-1. Groucho interacts with numerous transcriptional co-regulators.** Groucho interacting partners are involved in diverse and interconnected aspects of *Drosophila* development, including anterior-posterior and dorsal-ventral patterning and the interpretation of extra-cellular signaling. These interactions are largely mediated through interactions between short peptide motifs and the Gro WD-domain. Some factors, notably Dorsal, require the participation of additional factors to facilitate Gro recruitment.

**Table 1-1**

|  |  |  |
| --- | --- | --- |
| **Interacting Protein** | **Biological Role** | **Citation** |
| Capicua | RTK signaling; embryonic terminal gene expression | {Jimenez, 2000 #3093} |
| Huckebein | Embryonic terminal gene expression | {Goldstein, 1999 #3094} |
| Hairy | Segmentation/ Anterior-posterior patterning | {Paroush, 1994 #3090} |
| Runt | Segmentation/ Anterior-posterior patterning | {Aronson, 1997 #3095} |
| Even-skipped | Segmentation/ Anterior-posterior patterning | {Kobayashi, 2001 #3076} |
| Odd-skipped | Segmentation/ Anterior-posterior patterning | {Goldstein, 2005 #3096} |
| Sloppy-paired 1 | Segmentation/ Anterior-posterior patterning | {Andrioli, 2004 #3097} |
| Engrailed | Segmentation/ Anterior-posterior patterning | {Jimenez, 1997 #3075} |
| Knirps | Segmentation/ Anterior-posterior patterning | {Payankaulam, 2009 #2955} |
| Goosecoid | Segmentation/ Anterior-posterior patterning | {Jimenez, 1999 #3092} |
| Dorsal | Dorsal-ventral patterning | {Dubnicoff, 1997 #2366} |
| Brinker | Dorsal-ventral patterning | {Zhang, 2001 #3099} |
| Ind | Dorsal-ventral patterning | {Von Ohlen, 2007 #3101} |
| Vnd | Dorsal-ventral patterning | {Cowden, 2003 #3102} |
| Su(H) | Notch signaling | {Barolo, 2002 #3072} |

## References