## Dissertation Title

### List of Figures

### List of Tables

### Acknowledgments

### Vita

### Selected Publications

### Abstract of the Dissertation

### Chapter X: Introduction: Groucho and *Drosophila* development

* Drosophila development
  + Brief overview of early embryonic stages and patterning events
  + Relevance of Drosophila development on humans
* Transcription factors
  + Overview of role of major activators/repressors in early development
  + Overview of role of major coactivators/corepressors
* Groucho
  + Discovery & early characterization
  + Domain structure
  + Homology
  + Pathway participation
  + Regulation
  + Potential regulatory mechanism(s)
  + Known/suspected target genes
  + Importance to development/signal transduction
  + Importance to disease states (developmental abnormalities and cancer)
* DNA topological considerations

### Chapter X: Groucho activity in the developing embryo

#### Abstract

#### Introduction

* Groucho and development (short section)
  + Roles throughout Drosophila development
  + Ubiquitously expressed
  + Roles in tissue specification in response to multiple extracellular signaling pathways (RTK signaling being most prominent)
  + Groucho as a decision node, integrating signal information combinatorially and in a context-dependent fashion, as it's activity is dependent on the presence or absence of one or more interacting partners
* RNA-seq for transcriptome analysis
  + Allows for the analysis of Groucho regulatory activity in a developmental context
  + Development is a rapid and fluid process, so necessary and instructive to analyze groucho activity at multiple timepoints throughout the process
  + Interested in looking at Groucho activity at the initiation of zygotic transcription onwards, though Grouch is maternally deposited and so ubiquitously present in the embryo before this time
* Short reiteration of major developmental events w/ Gro involvement at each of the three timepoints analyzed

#### Materials & Methods

* Fly strains/culturing/collection
  + All fly lines were grown according to established procedures (look up how exactly to cite this)
* RNA isolation/library generation/sequencing
  + RNA isolated with Qiagen RNA extraction kit (look up name)
  + RNA quality assessed by Agilent Bioanalyzer
  + Stranded RNA-seq libraries generated using Illumina TruSeq transcriptome generation kits (look up name of kit I suppose, or ask Sean)
    - Libraries were barcoded for multiplex sequencing
  + Sequencing performed at UCLA High-Throughput sequencing core (name) on \_\_\_\_ (machine? Illumina HiSeq something probably)
* Data analysis methodologies
  + QC/filtering
    - Quality assessment performed with FastQC (no available publication)
  + Demultiplexing performed by in-house scripts (github link?)
  + Mapping
    - Mapping performed using Tophat2 v2.0.9 {Kim, 2013, r08998}
    - Using Drosophila assembly BDGP?/dm3 (Illumina iGenomes)
    - Transcript accessions (dm3) provided by FlyBase (Illumina iGenomes)
  + Assignment of mapped fragments to genes performed by HTSeq
  + Differential expression analysis w/ DESeq2 {Love, 2014 #3031}
    - Setup of model and DE paramaters (show some diagnostic plots?)
  + Differential expression analysis w/ MaSigPro {Nueda, 2014 #3032}

#### Results

* Sequencing overview
  + Table: Total reads/qc pass reads/mapped reads/reads by feature (intergenic/intronic/exon/etc)
  + Replicate similarity
    - Heatmap of spearman coefficients
    - PCA analysis
    - Characteristic MA plot
* Perturbation of Groucho levels results in the misregulation of thousands of genes
  + Wild-type and mutant expression profile of Groucho shows ubiquitious maternal contribution in the early embryo, and steep fall-off at later timepoints
    - Overexpression shows higher expression at early timepoints, and a slight recovery towards wt levels aftewards
    - Gro LoF shows constitutively low levels of Groucho expression
    - Is Gro zygotically expressed at all?
    - According to modENCODE data, gro transcript levels are highest at embryo deposition, and steadily decrease, reaching a low point during larval L3 stage. Gro levels are slightly elevated during late L3 and white prepupae stages, before dropping once more to approximately early/mid larval levels
    - Our Gro profile is consistent with these observations
    - Figure: Groucho expression timecourse (wt and mut)
  + Drosophila Gro is a dedicated repressor
    - Introduce concept of Gro homologs that serve primarily as activators
    - Evidence from LoF and OE regulated genes
* Groucho target genes are enriched for developmentally-regulated transcription factors
* Groucho recruitment is more dynamic during early embryonic stages (consistent with highest levels of Gro at these stages)
* Groucho localizes in distinctive patterns adjacent to, and overlapping, predicted target genes
  + Capicua-regulated genes
  + Brinker-regulated genes
  + Zelda interaction
* Groucho binds at transcription start sites
* Groucho spreading does not appear to be an obligate characteristic of repression
  + Groucho peaks exhibit a mean width of approximately 500 bp
  + Groucho widths slightly increase in size at later timepoints, though whether this is the result of biological variability or slight deviations in chip-seq library fraction efficiencies is uncertain. In either case, the effect is slight.
  + Clustering of Gro binding regions does appear to be prevalent. Over half of all Groucho-associated genes exhibit two or more distinct Groucho binding sites
* At all timepoints, about half of Gro binding is either upstream or overlapping the start site of the nearest gene feature, consistent with Groucho's association with enhancer regions, which are typically, but certainly not always upstream of their target gene
  + Approximately 40% of all Gro binding sites are contained with gene bodies, with 35 to 50% of those sites falling within an intron. Of those, 40% are found within the first intron.
  + In Drosophila, the first intron is typically longer, and most relevant to gene regulation.
  + It is unknown whether Groucho is being actively recruited to these intronic regions, or if looping from an enhancer region is responsible for depositing Gro within these areas.
  + Motif analysis of binding sequences within introns reveals few conserved sequences, possibly indicating the lack of a sequence-specific DNA binding factor being responsible for recruiting Gro at these positions
    - Despite this, recruitment of Gro to introns seems highly specific (is being directed by some factor or process), as it often occurs with 100-200 bps of the TSS, in discreet peaks or clusters of peaks
    - Several examples of genes with alternative TSS's exhibiting Groucho peaks in introns adjacent to each start site (ex. dpp)
* Recruitment of Groucho to intronic regions correlates with repression of those genes, but does not appear to be an obligate feature of repression
  + Multiple known or suspected Gro targets lack introns (zen, tolloid, twist, wntD, tll, sna)
* Some genomic areas exhibit widespread Groucho binding, especially around Gro targets in areas with low gene density
  + Ex. snail; 21 distinct Groucho bindng sites spread over a 50kb region
  + Ex. odd skipped; 17 distinct sites over 30kb
* Groucho involved in active, reversible repression is enriched for upstream and within genes
* Groucho regulated genes are enriched for paused PolII, and depleted for actively transcribing PolII
  + Stronger enrichment with genes with Gro intronic binding

#### Discussion

### Chapter X: Nascent-seq

#### Abstract

#### Introduction

* Introduction to nascent-seq
  + Transcription rate determination
    - Importance to non-steady state transcript level analysis
  + Fractionation methodology
* Very brief introduction to deep sequencing technologies

#### Materials & Methods

* Flies
  + Strains
  + Culturing
  + Collection
* Isolation of chromatin-associated RNA
  + Embryo fractionation
  + RNA isolation
  + Validation of fractionation (anti-H3/anti-tubulin immunoblot)
  + Depletion of rRNA
    - Sequences of depletion oligomers
  + Validation of mRNA quality (Bioanalyzer)
* Library preparation/QC
* Sequencing
* Data processing
  + QC/pre-processing
  + Mapping
  + Transcript assignment and expression level normalization

#### Results

* Validation of chromatin-associated RNA isolation
  + anti-H3/anti-tubulin immunoblot
  + Bioanalyzer profiles/pseudogels
  + Schematic of rRNA depletion procedure(?)
* Overview of data sets
  + Reads per sample
  + Intra- and inter-replicate similarity analyses
* Primary differences with total mRNA seq experiments

#### Discussion

#### Figures & Tables

#### References

### Appendices

#### Appendix X: Groucho-regulated gene data

#### Appendix X: Groucho binding regions by timepoint