### Chapter 3: Investigating the dynamics of the embryonic transcriptome

#### Abstract

#### Introduction

*Drosophila* development involves the coordinated expression of a vast number of genes under strict temporal and spatial control {Brown, 2014 #3144}. Transcript levels in the embryo are therefore highly dynamic, undergoing rapid shifts in level dependent on altered rates of accumulation, processing, and degradation. In excess of ~7,000 unique transcripts, arising from at least 3,000 distinct genes are expressed from the earliest onset of zygotic transcription {Graveley, 2011 #3044} {De Renzis, 2007 #3149}. Of these, at least 1,000 are expressed in a spatially constricted manner {Tomancak, 2002 #1150}. Regulatory systems go to great lengths to minimize even relatively slight stochastic changes in expression, which can nevertheless have a negative effect on viability {Perry, 2010 #385}. Additional processes influencing mRNA abundance are also tightly controlled, with regulated degradation pathways resulting in significant and transcript-specific differences in the kinetics of mRNA decay in the embryo {Thomsen, 2010 #1350}. Ideally, accurate and quantitative measurements of protein abundance could be utilized to investigate the mechanics of development, but lacking such a technique, mRNA abundance is often substituted as a measurement providing a strong correlation with protein abundance {Fu, 2009 #3150}.

A complicating factor in the interpretation of transcriptome data from the early *Drosophila* embryo arises from the significant effects of maternally-contributed mRNAs and the timeline of activation of the zygotic genome. This latter process, the maternal-to-zygotic transition (MZT) is a common feature of animal development and encompasses a cascade of processes whereby widespread alterations to the chromatin landscape are engendered by the activity of pioneering transcription factors {Tadros, 2009 #3146} {Li, 2014 #3151}. In *Drosophila,* the zinc-finger TF Zelda is a well-studied example of such a pioneering factor {Liang, 2008 #3153} {Harrison, 2011 #3152} {Xu, 2014 #1928}. In most animals, the MZT process consists of two distinct stages. An initial “minor wave” of activation becomes significant at ~1.5 hours post-fertilization in *Drosophila*, followed by a more rapid and synchronous “major wave” at ~2.5 hours {Pritchard, 1996 #3147}. During this time the majority of the maternally-contributed transcriptome is destabilized and undergoes coordinated degradation {Tadros, 2007 #3155;Benoit, 2009 #3154}.

As mRNA abundance is a complex process, determined by the integrative inputs of the rates of transcription, processing, and degradation, measuring the rates of transcription in a temporally-discriminate manner becomes challenging in an evolving system such as the embryo. A number of techniques have been proposed to address this question, one popular technique being the direct sequencing of RNA populations enriched for nascent and chromatin-associated RNAs (nascent-seq). Nascent RNA-seq, or nascent-seq, has been shown to be an effective strategy for identifying the actively transcribed genes in a cell or tissue, as well as quantitating the relative transcriptional rate of these genes. This is accomplished through the use of existing deep-sequencing platforms to specifically sequence the fraction of RNA that is chromatin-associated, and therefore enriched for transcripts undergoing active elongation. By using this method in *Drosophila melanogaster* embryos, we seek to obtain a timeline of transcriptional activation and repression to a high degree of temporal accuracy, which will aid us in identifying genes regulated by Groucho as well as the timeframes over which this regulatory ability is exercised.

Nascent-seq has been successfully applied to track the transcriptional changes in a number of biological contexts, including macrophages(Bhatt et al., 2012), where it was utilized to obtain a timeline of transcriptional changes following induction of an immune response, as well as in adult *D. melanogaster* tissues to analyze the prevalence of cotranscriptional splicing {Khodor, 2011 #2081}, as well as circadian transcript cycling{Rodriguez, 2013 #1782}, in which the authors saw significant differences in total mRNA and nascent mRNA levels between non-overlapping ninety minute embryo collections. We have adopted the method to developing embryos, using an established protocol for embryo nuclei isolation {Nechaev, 2010 #3006} followed by isolation of a chromatin-associated fraction from these nuclei. Purification of RNA from the chromatin fraction yields a RNA pool significantly enriched for nascent RNA.

Integrating this data with whole RNA-seq data will additionally aid in eliminating false-positives from our derived list of Grouch-regulated genes. As Groucho’s ability to repress transcription is regulated both spatially and temporally throughout development, discreet measurements of transcription over time will allow us to more accurately describe and understand Groucho’s multitude roles in fly development.

#### Materials & Methods

#### Results

*RNA from fractionated embryos exhibits multiple characteristics of nascent transcripts*

Total RNA was extracted from chromatin isolated from *D. melanogaster* embryos collected over three time spans in early development and subjected to high-throughput sequencing. This chromatin-associated RNA is expected to be enriched for nascent transcripts, as well as additional RNA species which associate with chromatin in structural, catalytic, or regulatory capacities {Cernilogar, 2011 #3156}. Isolated RNA was affinity-depleted for polyadenylated RNA in order to further minimize the contribution of mature mRNA from analysis. The level of enrichment for nascent transcript was validated and quantified through various measures. The efficiency of chromatin isolation was confirmed through analysis of protein compartmental markers (Fig. 3-1A/B), confirming that the sequenced RNA was derived from a embryonic fraction enriched for histones and deficient for a cytoplasmic marker. Sequencing reads obtained from mature transcripts ideally map uniformly across genes, though this is heavily dependent on the quality of the RNA utilized for library generation. Non-uniformity generally manifests as a 3’ bias in mappable reads, as partially fragmented mRNA is purified by affinity selection to polyadenylation sites {Roberts, 2011 #3157}. Chromatin-associated RNA exhibits a significant 5’ bias at each developmental stage, and is partially depleted at the 3’ end (Fig. 3-2). The large enrichment of reads arising from the initial 15% of gene bodies may be indicative of promoter-proximal paused polymerase. The sharp decrease in read occupancy near the 3’ terminus may result from frequent polymerase pausing in terminal exons {Carrillo Oesterreich, 2010 #3159}. Pausing in terminal exons is thought to promote recognition of polyadenylation sites and transcriptional termination {Gromak, 2006 #3158}.

Chromatin-associated RNA is enriched for intronic sequence when compared to poly(A)+ libraries prepared from the same developmental stages (Fig. 3-3). Our data indicates that on average, 13% of poly(A)+ RNA-seq reads map to constitutive introns compared to 35% of chromatin-associated RNA reads. While 60-70% of gene sequence in *D. melanogaster* is annotated as intronic, the large majority of introns are believed to be cotranscriptionally spliced, with only 16% of introns exhibiting little or no splicing {Khodor, 2011 #2081} {Wuarin, 1994 #259}. Therefore, an intronic content of between 13 and 60% should be expected for a library enriched for pre-mRNA.

#### Discussion

#### References

#### Figures and Tables