

## Evolution of Transcriptional *cis*-Regulatory Modules in the Fly Genital Disc

### Introduction and Motivation:

While numerous papers have demonstrated the power of comparative genomics to identify *cis*-regulatory modules (CRMs) in metazoan genomes<sup>1,2,3</sup>, there has been significantly less investigation into how evolution shapes the architecture of these regions to modulate their functionality. It is known that there exist orthologous pairs of CRMs that have had significant shuffling of binding sites for the regulating transcription factors (TFs), yet have largely conserved function<sup>4</sup>. Conversely, within a given species, one can find many examples of enhancers that are bound by the same collection of TFs, yet drive expression in fundamentally different ways<sup>5</sup>. One can easily imagine that this would also be the case when looking between species—that there would exist orthologous pairs of CRMs that, because of slight modifications, drive expression differently.

It is believed that there is a flexible grammar to the “*cis*-regulatory code,” where some alterations in the number and spacing of binding sites for the regulating transcription factors have little or no effect on regulatory function (where “regulatory function” is taken to mean the spatio-temporal pattern in which the CRM acts to enhance or repress transcription), while other classes of alterations have a strong effect<sup>5,6,7</sup>. A current limitation to the systematic examination of CRM evolution and grammar, however, is the lack of a sufficiently large collection of examples. One would ideally like a group of CRMs with the following three properties: 1) It would be a *large* collection of pairs of orthologous CRMs from two distinct species; 2) all CRMs-pairs in the set would be bound by (roughly) the same collection of TFs, and one would know what these TFs are; 3) approximately half of the CRM-pairs would be functionally equivalent and approximately half of the CRM-pairs would be functionally distinct (in order to have examples of each type), despite all being bound by the same set of TFs. This third criterion is especially difficult to satisfy, since the organisms under consideration would have to be sufficiently close to have their genomes aligned (otherwise one would not be able to identify orthologous CRMs), but such evolutionary proximity could result in largely identical mechanisms of regulation. Therefore, in order to have a sufficient number

of both functionally equivalent and functionally distinct CRMs, one must look in a biological system where selection has driven two recently divergent species to generate differences.

We propose genital disc development in *Drosophila* as a candidate model system. It is known that the external genitalia of the species *D. melanogaster* and *D. pseudoobscura* have greatly different morphology despite their close evolutionary proximity, and it has been postulated that selection has driven this divergence in order to prevent unfruitful inter-species copulation<sup>8,9</sup>. Recently, there has been tremendous progress in understanding the transcriptional regulatory networks underlying patterning of the genital disc, and the relevant pathways are ancient and largely conserved between the two species. It is known that the TF *doublesex* (*dsx*), which initiates the regulatory program responsible for sex-specific (i.e. male vs. female) morphology, acts in conjunction with the hox proteins during the third instar larvae stage at the three most posterior abdominal segments (A8-A10). Additionally, it is known that the signaling proteins *wingless* (*wg*), *decapentaplegic* (*dpp*), and *hedgehog* (*hh*) set up morphogenic gradients, and hence the TFs downstream of these signaling pathways (*dTCF*, *mad*, and *ci*, respectively) are also involved in regulating the transcriptional networks underlying this patterning. Although several enhancers regulated by the TFs mentioned above have been found in *D. melanogaster*, in order to truly address CRM grammar and evolution, a much larger set is needed (as is a thorough understanding of the orthologs of these CRMs in *D. pseudoobscura*).<sup>10</sup>

We propose a high-throughput strategy for generating and validating such a large set of enhancers. Our approach is both computational and experimental, but does not require the development of any new technologies or methods. The approach is hierarchical, proceeding from general to specific, where the most parallel experiments are performed first, and the most laborious and gene-specific experiments are performed last. During this initial set of experiments, we do not intend to study the discovered CRMs in great depth (i.e. we do not intend to do mutagenesis studies to footprint all binding sites); rather, our goal is to develop a large collection of candidate genital disc CRMs that have sufficient supporting experimental evidence to be of use for high-level questions on the evolution of transcriptional regulation. We do hope, however, that this candidate list will be examined more closely by our group and others in later experiments, and that it will thus be a resource to the greater community.

### **Aim 1: Perform expression arrays on *D. melanogaster***

We will begin by performing expression arrays on *D. melanogaster* females at the third instar larvae stage (a crucial patterning stage in genital development)<sup>10</sup>. In order to enrich for genes that are specifically up- and down-regulated in the genital disc, we will use flies that express green-fluorescent protein (GFP) under the control of GAL4, driven by an *abdB* promoter<sup>11</sup>; at this stage of development *abdB* is expressed exclusively in the A8 segment which is the progenitor of the female genital disc (we have chosen to use females, as there is not currently a driver for A9, the male progenitor segment)<sup>12</sup>. We then intend to perform FACS on the flies to isolate a pure population of female genital disc cells. We will perform Affymetrix expression arrays on these sorted cells, and will also perform expression arrays on a heterogeneous population of female third instar larvae fly cells (i.e., non-sorted) in order to have a reference population to determine fold-change. This technique of performing arrays on sorted fly cells has been successfully applied by several groups on other fly tissues.<sup>13,14</sup>

### **Aim 2: Computational Search for Candidate Enhancers**

Several groups have shown that it is possible to find developmental enhancers for *D. melanogaster in silico*, provided that one begins with a regulatory model that is composed of candidate TFs with known binding specificities.<sup>15,16,17</sup> The TFs in our model (the *hox* genes, *dTCF*, *ci*, *mad*, *dsx*) are all well-characterized, and we intend to modify the program of (16) to simultaneously take multiple TFs as input (this program generates a quantitative score measuring the likelihood that a given region of sequence is an enhancer). We will run this program on the regions adjacent (i.e., upstream, downstream and intronic) to *D. melanogaster* genes that exhibited significant up- or down- regulation in the genital disc from the expression arrays in Aim 1, with the intention of generating a large list of candidate CRMs. Once such a list of *D. melanogaster* CRMs has been generated, we will then remove all those candidates that do not have *D. pseudoobscura* orthologs (where “orthology” is defined as at least 70% identity over at least 100 bp, as determined by the UCSC BLAT alignments<sup>18</sup>). This will give a list of candidate pairs of genital disc CRMs common to *D. melanogaster* and *pseudoobscura*.

### **Aim 3: Chromatin Immunoprecipitation of Candidate Enhancers**

We expect to identify a large collection of candidate enhancers from Aims 1 and 2 (indeed, this is a central goal of the project). In order to get a broad understanding of which of these candidate CRMs are most likely to be functional, we will perform chromatin immunoprecipitations (ChIPs) on each region in both *D. melanogaster* and *D. pseudoobscura*, followed by quantitative PCR. Reliable polyclonal antibodies for all *D. melanogaster* TFs under consideration currently exist and, because they are polyclonal, they are likely to also work for *D. pseudoobscura*. Although one would ideally like to perform genome-level location analyses comparable to those done in yeast<sup>19</sup>, whole-genome arrays do not yet exist for *D. melanogaster*, and are unlikely to exist in the near future for *D. pseudoobscura*. Fortunately, the quantitative PCRs can be multiplexed, and the number of regions inspected will not be so numerous as to make this set of experiments overly laborious (if more than one hundred regions are obtained during Aims 1 and 2, then we will restrict to one-hundred randomly chosen regions).

**Aim 4: *In Situ* hybridizations on candidate genes in *D. melanogaster* and *D. pseudoobscura***

For genes adjacent to all candidate regions from Aim 3 with significant binding, we intend to perform *in situ* hybridizations in both *D. melanogaster* and *D. pseudoobscura* in order to reveal patterning differences. Note that this step is a necessity, as many of the genes are likely to have different patterns of expression between the species, yet not have differences in *absolute* levels of expression (thus, the expression analysis of Aim 1 is not sufficient to reveal differences in expression between the two species). Once these *in situs* have been performed, we will use a recently developed software package that is capable of taking *in situ* images and measuring differences in expression between embryos<sup>20</sup>. We also intend to perform *in situ* hybridizations for all TFs inspected, to control for the fact that the differences in expression could be due to *trans*-, rather than *cis*-, evolution. Indeed, we hope that it will be a mixture of both types of change, in order to get a better understanding of which mechanisms are more commonly used.

- 
- <sup>1</sup> Wasserman, W.W., *et al.* Human-mouse genome comparisons to locate regulatory sites. *Nature Genetics*. **26**, 225-228 (2000).
- <sup>2</sup> Krivan, W & Wasserman, W.W. A predictive model for regulatory sequences directing liver-specific transcription. *Genome Research*. **11**, 1559-66 (2001)
- <sup>3</sup> Kellis, M., *et al.* Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*. **423**, 241-254 (2003).
- <sup>4</sup> Ludwig, M.Z., *et al.* Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*. **403**, 564-567 (2000).
- <sup>5</sup> Senger, K, *et al.* Immunity Regulatory DNAs Share Common Organizational Features in *Drosophila*. *Mol. Cell*. **13**, 19-32 (2004).
- <sup>6</sup> Erives, A & Levine M. Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc. Natl. Acad. Sci.* **101**, 3851-3856 (2004).
- <sup>7</sup> Levine, M & Tjian R. Transcription regulation and animal diversity. *Nature*. **424**, 147-151 (2003).
- <sup>8</sup> Marin, I & Baker, B. The evolutionary dynamics of sex determination. *Science*. **281**, 1990-94 (1998).
- <sup>9</sup> Pitnick, S, *et al.* Ejaculate-female coevolution in *Drosophila mojavensis*. *Proc. R. Soc. Lond. B. Biol. Sci.* **270**, 1507-12 (2003)
- <sup>10</sup> Christiansen, AE, *et al.* Sex comes in from the cold: the integration of sex and pattern. *Trends in Genet.* **18(10)**, 510-516 (2002).
- <sup>11</sup> Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*. **118**, 401-415 (1993).
- <sup>12</sup> Estrada, B Casares, F, Sanchez-Herrero E. Development of the genitalia in *Drosophila melanogaster*. *Differentiation*. **71**, 299-310 (2003).
- <sup>13</sup> Bryant, Z *et al.* Characterization of differentially expressed genes in purified *Drosophila* follicle cells: Toward a general strategy for cell type-specific developmental analysis. *Proc. Natl. Acad. Sci.* **96**, 5559-64 (1999).
- <sup>14</sup> Jasper, H. A genomic switch at the transition from cell proliferation to terminal differentiation in the *Drosophila* Eye. *Develop. Cell* **3**, 511-521 (2002).
- <sup>15</sup> Halfon, M *et al.* Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Research*. **12**, 1019-1028, 2002
- <sup>16</sup> Rebeiz, M, Reeves, NL and Posakony, JW. SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. *Proc. Natl. Acad. Sci.* **99**, 9888-93 (2002)
- <sup>17</sup> Berman, B *et al.* Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci.* **99**, 757-762.
- <sup>18</sup> [www.genome.ucsc.edu](http://www.genome.ucsc.edu)
- <sup>19</sup> Lee, T *et al.* Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*. **298**, 799-804 (2002).
- <sup>20</sup> Peng, H. & Myers E.W. Comparing *in situ* mRNA expression patterns of *Drosophila* embryos. *Proc. Eighth Annual Intl. Conf. on Res. in Comp. Mol. Bio. (RECOMB)*. 157-166 (2004)