The thesis of Jizhong Zou was reviewed and approved* by the following:

Pamela J. Mitchell  
Associate Professor of Biochemistry and Molecular Biology  
Thesis Advisor  
Chair of Committee

Zhi-Chun Lai  
Associate Professor of Biology

Ross C. Hardison  
T. Ming Chu Professor of Biochemistry and Molecular Biology

David S. Gilmour  
Associate Professor of Molecular and Cell Biology

Kyung-An Han  
Assistant Professor of Biology

Robert A. Schlegel  
Professor of Biochemistry and Molecular Biology  
Head of the Department of Biochemistry and Molecular Biology

*Signatures are on file in the Graduate School
ABSTRACT

Transcription factor AP-2 family genes have important roles in limb, orofacial and nervous system development in vertebrates and *Drosophila*. Null mutations in *dAP-2*, the sole AP-2 gene in *Drosophila*, cause severe leg shortening, loss of tarsal joints, and defects in proboscis, antenna, and central nervous system development. *Drosophila* is an excellent model system for genetic and bioinformatic analysis of regulatory gene networks within which AP-2 transcription factors function in developing tissues.

This doctoral thesis describes research I have carried out using molecular genetic, transgenic, and comparative genomics approaches to identify *cis*-elements in *dAP-2* that mediate its expression in developing leg and antenna (i.e., leg and antennal imaginal discs) of *Drosophila*. Previous studies showed that *dAP-2* is expressed along the proximodistal (PD) axis of the developing leg in a segmental pattern representing future leg joints, and that *dAP-2* expression in presumptive tarsal joints (4 distal-most joints) was dependent on Notch signaling. I have extended these findings by showing that *dAP-2* is expressed in leg and antennal discs in nine and eight PD-axis domains, respectively; and that, in both disc types, the four distal-most *dAP-2* expression domains require Notch signaling, while the remaining more proximal domains do not. I have identified two enhancers on *dAP-2* genomic fragments BXE and E6 that activate reporter gene expression in distal (BXE) or proximal regions (E6) of leg and antennal discs in patterns coincident with endogenous *dAP-2* expression. Both enhancers function preferentially with the *dAP-2* exon 1a promoter. BXE contains multiple sub-fragments that act synergistically to mediate Notch-dependent activation and positive autoregulation by *dAP-2* itself. Two essential binding sites for Suppressor of Hairless (Su(H), the transcription factor which transduces target gene effects of Notch signaling) and one essential binding site for *dAP-2* were identified by phylogenetic footprinting, electrophoresis mobility shift assay (EMSA), site-directed mutagenesis and *in vivo* reporter gene analysis. These findings prove that *dAP-2* is a direct target of Notch signaling in presumptive tarsal joints. Genome wide searches for other genes that conserve sequences found in BXE and E6 may identify gene batteries with roles in morphological diversification of proximal versus distal regions of the leg and antenna.
# TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................. vii  
LIST OF TABLES .................................................................................................................... viii  
ACKNOWLEDGMENTS ........................................................................................................... ix  

## Chapter 1  Introduction ........................................................................................................... 1  
1-1  *Cis*-elements for transcriptional regulation of metazoan genes ............................... 1  
1-1.1 Core promoter .............................................................................................................. 3  
1-1.2 Enhancers .................................................................................................................... 4  
1-1.3 Identification of *cis*-elements in developmentally important gene ....................... 6  
1-1.3.1 *in vivo* approaches with transgenic animals ...................................................... 6  
1-1.3.2 *in vitro* approaches ............................................................................................ 7  
1-1.4 Comparative genomics ............................................................................................ 9  
1-1.4.1 Evolutionarily conserved genes .......................................................................... 9  
1-1.4.2 Promoter prediction ........................................................................................... 11  
1-1.4.3 Enhancer prediction ........................................................................................... 12  
1-1.4.4 Phylogenetic footprinting .................................................................................. 13  
1-1.4.5 Phylogenetic shadowing .................................................................................... 14  
1-1.4.6 Comparative genomics programs ...................................................................... 15  
1-1.4.7 Identification of regulatory motifs for co-regulated genes ................................. 16  
1-1.4.8 Consideration of phylogenetic footprinting applications ................................ 17  
1-2  *Drosophila* leg development ...................................................................................... 18  
1-2.1 Patterning the proximodistal axis of the leg ............................................................ 20  
1-2.2 Notch signaling in leg joint development and leg elongation ................................. 23  
1-2.3 *dAP-2* in *Drosophila* development ..................................................................... 26  

## Chapter 2  Transcriptional regulation of *Drosophila dAP-2* in leg and antennal discs ...................................................... 29  
2-1  Introduction .................................................................................................................. 29  
2-2  Materials and Methods ............................................................................................... 33  
2-2.1 Reporter constructs ................................................................................................. 33  
2-2.2 GAL4 driver constructs for *dAP-2* mutant rescue .............................................. 35  
2-2.3 Mutagenesis of Su(H) binding sites and *dAP-2* binding sites ............................ 36  
2-2.4 Rapid small scale isolation of *Drosophila* genomic DNA .................................. 36  
2-2.5 Alkaline lysis plasmid minipreps .......................................................................... 37  
2-2.6 RNA preparation and RT-PCR ............................................................................. 38  
2-2.7 Embryo injection and identification of transformants .......................................... 39  
2-2.8 Fly stocks ................................................................................................................ 40  
2-2.9 Genetic mosaic (clonal) analysis .......................................................................... 41  
2-2.10 β-gal staining and light microscopy ...................................................................... 41  
2-2.11 Antibodies and immunostaining ......................................................................... 42  
2-2.12 Electrophoresis mobility shift assay (EMSA) ...................................................... 43  
2-2.13 DIG-labeling of RNA probes .............................................................................. 44
2-2.14 RNA in situ hybridization ................................................................. 45
2-2.15 Mounting adult fly legs .................................................................... 46

2-3 Results .................................................................................................. 47
2-3.1 Expression of dAP-2 in Drosophila legs ......................................... 47
2-3.2 Expression of dAP-2 in Drosophila antenna .................................. 54
2-3.3 Notch signaling-dependent regulation of dAP-2 along PD axis of limb .............................................................. 59
2-3.4 Identification of cis-elements for expression in proximal and distal leg regions using a taulacZ reporter with hsp70 promoter........ 63
2-3.5 Prediction of candidate cis-elements in dAP-2 genomic regions using phylogenetic footprinting ............................................ 67
2-3.6 Preferential use of dAP-2 exon 1a promoter in leg discs .......... 73
2-3.7 Dissection of BXE enhancer for distal limb expression using Pelican vectors ........................................................................ 77
2-3.8 BXE enhancer is direct target of Notch signaling ............................ 82
2-3.9 Identification of essential AP-2 binding site in BXE that mediates autoregulation .............................................................. 88
2-3.10 Model for dAP-2 regulation in developing leg and antenna ....... 93

2-4 Discussion .......................................................................................... 96
2-4.1 Evolutionarily conserved function and regulation of dAP-2 .......... 96
2-4.2 Regulatory mechanisms of dAP-2 in appendage development .... 102

Chapter 3 Analysis of CG10440, a BTB/POZ protein and candidate co-repressor for dAP-2 ................................................................. 104

3-1 Introduction .......................................................................................... 104
3-1.1 Target genes of dAP-2 ................................................................... 104
3-1.2 CG10440 ...................................................................................... 107

3-2 Materials and Methods....................................................................... 108
3-2.1 Fly stocks ....................................................................................... 108
3-2.2 Construction of pUAS-CG10440HA .............................................. 108
3-2.3 Construction of pUAS-CG10440Wi .................................................. 108
3-2.4 Immunostaining ............................................................................ 109
3-2.5 RT-PCR ....................................................................................... 109
3-2.6 RNA in situ hybridization ............................................................ 110
3-2.7 Genomic PCR detection of P-element insertion .......................... 110

3-3 Results ................................................................................................ 112
3-3.1 Expression and subcellular localization of CG10440 ................. 112
3-3.2 Overexpression of CG10440 affects leg development ................. 114
3-3.3 CG10440 negatively regulates dAP-2 target genes ...................... 117
3-3.4 P-element mediated mutagenesis of CG10440 ............................ 119
3-3.5 Knockdown of CG10440 by RNA interference ........................... 122

3-4 Discussion ........................................................................................... 124

Bibliography ............................................................................................. 127

Appendix A List of transgenic lines ......................................................... 144
Appendix B Regulation of \textit{dAP-2} by Bowl...............................................................146
Appendix C Regulation of \textit{dAP-2} by Distalless .......................................................150
## LIST OF FIGURES

| Fig. 2-1 | Expression of wild-type *dAP-2* in leg imaginal disc | 50 |
| Fig. 2-2 | Comparison of *Distalless* and *dAP-2* expression domains | 51 |
| Fig. 2-3 | Comparison of *bib-lacZ* and *dAP-2* expression domains | 52 |
| Fig. 2-4 | Summary of *dAP-2* expression in *Drosophila* leg | 53 |
| Fig. 2-5 | Expression of wild-type *dAP-2* in antenna disc | 57 |
| Fig. 2-6 | Summary of *dAP-2* expression in *Drosophila* antenna | 58 |
| Fig. 2-7 | Notch dependent regulation of *dAP-2* expression in leg disc is restricted to distal parts (presumptive tarsal joints) | 61 |
| Fig. 2-8 | Notch dependent regulation of *dAP-2* expression in antenna disc is restricted to distal parts | 62 |
| Fig. 2-9 | Dissection of 20kb *dAP-2* genomic region using *tauZ* vector | 65 |
| Fig. 2-10 | Rescue of leg shortening and tarsal joints by B12-GAL4 driven *dAP-2* exon 1a cDNA | 66 |
| Fig. 2-11 | VistaBrowser alignment of *dAP-2* genes of 8 *Drosophila* species | 70 |
| Fig. 2-12 | Comparison of genome alignment tools | 71 |
| Fig. 2-13 | Differential use of two alternative first exons for *dAP-2* | 75 |
| Fig. 2-14 | *dAP-2* enhancers for leg expression prefer exon 1a promoter | 76 |
| Fig. 2-15 | Dissection scheme for analysis of distal leg specific enhancer | 80 |
| Fig. 2-16 | Enhancer activities of B6, BXE and various subfragments in leg and antennal discs | 81 |
| Fig. 2-17 | Notch signaling is required only for tarsal leg specific enhancer activities | 85 |
| Fig. 2-18 | Identification of Su(H) binding sites in BXE enhancer | 86 |
| Fig. 2-19 | Electrophoresis mobility shift assay of Su(H) binding sites | 87 |
| Fig. 2-20 | Distal leg specific autoregulation of *dAP-2* | 91 |
| Fig. 2-21 | Identification of *dAP-2* binding site in BXE mediating autoregulation | 92 |
| Fig. 2-22 | Model for differential regulation of *dAP-2* | 95 |
| Fig. 2-23 | Conservation between AP-2 family genes | 101 |
| Fig. 3-1 | Expression and subcellular localization of CG10440 | 113 |
| Fig. 3-2 | *dAP-2* and CG10440 do not regulate each other | 115 |
| Fig. 3-3 | Leg defects caused by CG10440 overexpression | 116 |
| Fig. 3-4 | Candidate dAP-2 target genes are negatively regulated by CG10440 over-expression | 118 |
| Fig. 3-5 | Scheme for P-element mediated mutagenesis | 121 |
| Fig. 3-6 | CG10440 RNAi knock-down | 123 |
| Fig. B1 | Repression of *dAP-2* by Bowl | 147 |
| Fig. C1 | Regulation of *dAP-2* by Distalless | 150 |
LIST OF TABLES

Table A1. List of transgenic lines ................................................................................144
ACKNOWLEDGEMENTS

I would like to thank in the first place my advisor, Dr. Pamela J. Mitchell, for her encouragement and guidance through my research. I specially appreciate the opportunity I was given to learn so much about animal development using *Drosophila* and vertebrate models with assistance of comparative genomics.

I want to thank my laboratory colleagues, Youngwook Ahn for helping me with numerous micro-injections, Il Minn for his supporting work in protein assays, and all other Mitchell lab members for their help and discussions improving my work.

Thanks to my committee members, Dr. Zhi-Chun Lai for his suggestions in *Drosophila* genetics, Dr. Ross C. Hardison for his guidance in comparative genomics, Dr. David S. Gilmour and Dr. Kyung-An Han for their advice on my thesis and future career.

I also want to thank my wife, my parents and my brother for their constant love and enormous support. You made this journey a much better and worthwhile experience. I could not go this far and complete my Ph.D. study without your love. This work is dedicated to you.
Chapter 1

Introduction

1-1  *Cis*-elements for transcriptional regulation of metazoan genes

The control of gene expression is a fundamental question in studies of animal development. From the single fertilized egg to the multicellular adult with its many different tissues and organs, development proceeds by complex changes in gene expression patterns. Regulation of gene expression can occur at multiple levels, including RNA processing, protein translation and protein stability, but the primary control in many cases is at the level of RNA transcription.

The genomes of metazoan species vary in total size (C value) over a range of several orders of magnitude. For the past few decades, it was widely believed that the large amount of intergenic DNA found in genomes of higher organisms was merely “junk DNA” with no important or obvious functions. However, this view has changed considerably in recent years based on progress in molecular genetic analysis of gene structure and function in invertebrate and vertebrate animal model systems. It is now apparent, based on comparative genomics and genetic analyses, that evolution of higher animal species did not involve large increases in the total number of genes, but rather has been largely dependent on increases in the amount of regulatory DNA associated with genes and a concomitant increase in regulatory complexity (reviewed by Levine and Tjian, 2003). For example, a comparison of the genomes of the fruitfly *Drosophila melanogaster* (180 Mb) and the nematode *Caenorhabditis elegans* encoding ~14,000 and
~19,000 genes respectively, revealed that, in both species, genes with developmentally complex functions typically contained significantly larger amounts of intergenic DNA than housekeeping genes (Nelson et al., 2004). Furthermore, gene loci with high regulatory complexity were significantly larger in size in Drosophila than in C. elegans, correlating with increased complexity of the body plan and nervous system and larger number of cell types in Drosophila relative to C. elegans.

Several basic types of cis-regulatory elements are involved in the regulation of mRNA transcription by RNA polymerase II in eukaryotic cells. These include the initiator element, the core promoter, enhancers, silencers, and insulators. These elements serve as binding sites for sequence-specific binding factors that interpret and convey regulatory information to the factors and cofactors that make up the transcriptional machinery required for RNA synthesis. The core promoter is located within about 100 base pairs around the transcription initiation site, and contains various combinations of cis-elements that include the TATA box, the initiator element (INR) and the downstream promoter element (DPE). The core promoter is essential for recruiting the transcription factor IID (TFIID) which serves as the foundation for the assembly of the transcription initiation complex (reviewed by Smale and Kadonaga, 2003).

The core promoter is necessary but generally insufficient to direct significant levels of transcription in vivo on its own. Efficient, high level transcription initiation requires additional elements termed enhancers. Enhancers studied to date are typically several hundred nucleotides in length, and activate transcription initiation from a linked core promoter in a signal-dependent and cell type-specific manner. The activity of enhancers can be antagonized by other cis-regulatory elements termed silencers. In
addition, some boundary-defining sequences called insulators can prevent enhancers from associating inappropriately with neighboring promoters (Capelson and Corces, 2004).

A typical Drosophila gene containing 2-3kb of coding DNA sequence has cis-regulatory elements typically scattered over a 10kb genomic region. In vertebrates, the number and complexity of cis-regulatory elements for a single gene is typically much greater (reviewed by Levine and Tjian, 2003). The organization of these cis-regulatory elements are determining factors as to when, where, and how often the genes with which they function are transcribed.

1-1.1 Core promoter

The core promoter, the most extensively studied cis-regulatory element, consists of various combinations of sub-elements (mentioned earlier) not all of which are found in every gene. A database analysis of 205 core promoters from Drosophila genes showed 29% contain a TATA box (consensus TATAA/TAAG/A) and no DPE, 26% have a DPE but no TATA box, 14% contain both, and 31% do not appear to contain either (Kutach and Kadonaga, 2000). A larger study of 1941 Drosophila gene promoters showed 33% had a TATA box (Ohler et al., 2002). In addition to the evolutionarily conserved TATA-binding protein (TBP) present in all eukaryotes, tissue-specific TBP-related factors (TRFs) (e.g., TRF1 and TRF2) have been identified in Drosophila and other species (reviewed by Davidson, 2003). Studies suggest that these factors, including the evolutionarily conserved TBF2 for which a binding site has not been identified, function as core promoter-selectivity factors responsible for coordinating transcription of specific subsets of genes (Holmes and Tjian, 2000; Rabenstein et al., 1999). DPE consensus
sequences may be species specific. DPE elements identified to date in *Drosophila* genes are located at approximately +28 to +32, have the consensus sequence A/GGA/TC/TG/A/C, and are often, but not exclusively, found in TATA-less promoters (Kutach and Kadonaga, 2000).

The general factors TFIIB, TFIIE, TFIIF, and TFIIH, together with TBP, are required to assemble the transcription initiation complex for eukaryotic RNA polymerase II (reviewed by Smale and Kadonaga, 2003). TBP and multiple TBF-associated factors (TAFs) comprise the complex called TFIID. TAFs are responsible for productive interactions between the core promoter and enhancers for tissue-specific regulation. Various studies suggest that these interactions stabilize the basal transcription machinery complex for RNA polymerase II at the transcription initiation site. Studies have shown that INR-binding TAF$_{II}150$ and TAF$_{II}250$ form trimeric complex with TATA-binding TBP to stabilize TFIID. TAFs are used to link the INR to TATA or DPE or both. In TATA-less, DPE-dependent promoters, cooperation between INR and DPE is mediated by DPE-binding TAF$_{II}40$ and TAF$_{II}60$ and INR-binding TAFs. Some tissue-specific TAFs also contribute to core promoter diversity by facilitating the formation of unique TFIID complexes.

**1-1.2 Enhancers**

Although the basal transcription complex can be customized for certain cell types or even certain genes, its own activity is not enough to define temporal and tissue-specific expression of gene in metazoan. Enhancers are the major determinants of differential transcription. Enhancers are typically several hundred base pairs long, contain
clusters of transcription factor binding sites to enable the formation of multi-subunit transcription complexes and may integrate signaling inputs received by the cell (reviewed by Levine and Tjian, 2003). Enhancers are operationally defined as being able to function at a distance from the promoter (typically greater than 1 kb away and, in some cases, up to 100 kb or more away) and either upstream or downstream of the promoter, and even from within introns. Based on several detailed studies of selected enhancers, enhancer activity is mediated by DNA binding transcription factors and a host of coactivator and corepressor proteins, which bind in complex arrays to the constellation of DNA sequence motifs that define the particular enhancer. The enhancer/transcription factor complex, also called enhanceosome, facilitates selection of the promoter, especially when there are multiple promoters for the gene, and controls the magnitude of gene expression (Merika and Thanos, 2001).

Cooperation between enhancers and promoters can be mediated by transcription factors bound to them, or by cofactors such as CRSP, TRAP, and ARC/DRIP that serve bridges between activators and the basal transcription complex (reviewed by Levine and Tjian, 2003). Compared to basal transcription complex units, the co-factors are usually structurally less conserved between species and some can undergo conformational change (Taatjes et al., 2002).

Long-range enhancer-promoter interactions are commonly seen in complex genetic loci, such as *Drosophila Antp* complex where T1 enhancer bypasses the neighboring *fushi tarazu* (*ftz*) gene and interacts with the distant *Sex combs reduced* (*Scr*) promoter (Calhoun and Levine, 2003). The T1-*Scr* interaction requires a tethering protein to bind on proximal-promoter elements. While many enhancers may bind factors that
allow direct interaction with the targeted transcription initiation complex, some enhancers may be tethered to the core promoter by interactions with transcription factors bound to proximal promoter element(s) located adjacent to the core promoter.

Altogether, transcriptional \textit{cis}-elements and their interacting proteins create a great deal of regulatory diversity. The complexity of developmental mechanisms and cell differentiation in higher organisms correlates with the number of \textit{cis}-regulatory elements. To study the gene regulation, multiple approaches have been used to identify and characterize cell type-specific enhancers.

\section*{1-1.3 Identification of \textit{cis}-elements in developmentally important gene}

\subsection*{1-1.3.1 \textit{In vivo} approaches with transgenic animals}

Potential enhancer elements for developmentally regulated and cell type specific gene expression can be tested functionally by reporter gene analysis in transgenic animals or in transfected cultured cells. In the former, reporter gene DNA is microinjected into embryos and transgenic animals are generated for analysis of reporter expression in tissues at developmental stages of interest. Experiments with cultured cells, while less labor- and time-intensive than transgenic analysis, provide limited insights given the highly artificial context. The earliest definition of enhancers included the idea that they should be able to function with heterologous promoters since all promoters essentially recruit the same basal transcription machinery. For this reason, enhancers are most frequently tested by fusion upstream of downstream of a reporter gene with a generic basal promoter such the thymidine kinase promoter for mammalian studies or the \textit{hsp70} promoter in \textit{Drosophila} studies. However, several studies have identified examples of
selective compatibility between different enhancer-promoter combinations (Li and Noll, 1994). Such findings are less surprising now in view of the growing evidence for differences in core promoters with regard to cis-elements used and tissue specificity.

Given the above, the ideal promoter to use in reporter constructs designed to test enhancer function is the cognate promoter (if known). Popularly used reporter genes include the E. coli lacZ gene (encoding beta-galactosidase), the jellyfish green fluorescent protein (GFP) gene, and the bacterial gene for chloramphenicol acetyltransferase (CAT). An advantage of using GFP is that animals and their GFP-expressing cells can be viewed live under the fluorescent microscope.

In Drosophila, enhancers can be identified by enhancer trap methods. An enhancer trap is a P-element construct like the transgenic construct mentioned above, except that it lacks an enhancer. When randomly integrated into the genome, the reporter gene will be expressed under the instruction of any enhancer(s) in adjacent flanking DNA that is within proper range. The genomic sequence around enhancer trap construct then can be “rescued” and cloned for further analysis (O’Kane and Gehring, 1987).

Other genetics tools are also used to study cis-regulatory element in vivo. One gain-of-function technique uses a tissue-specific enhancer to drive ectopic expression of the yeast GAL4 transcription factor, which in turn activates another gene’s expression under the control of GAL4-binding UAS enhancer. Loss-of-function analysis uses deletions or mutations in the enhancer element to see if there is alteration of its activity.

1-1.3.2 in vitro approaches
*Cis*-regulatory elements can also be studied by protein binding assays such as the electrophoresis mobility shift assay (EMSA) and DNase I footprinting assay. In EMSA, a double-stranded DNA probe representing the *cis*-element of interest is incubated with a purified transcription factor or nuclear extract of interest; specific binding of transcription factor to the probe will resist non-specific DNA competition and will result in a DNA-protein complex that, upon electrophoresis in a non-denaturing acrylamide gel, will show a decreased electrophoretic mobility relative to the free DNA probe.

In the DNase I footprinting assay, a double-stranded DNA fragment end-labeled with $^{32}\text{P}$ on one strand is incubated with purified protein or a mixture of proteins. Specific binding of protein to a recognition site on the probe will be resistant to non-specific DNA competition and, furthermore, will protect the binding site from digestion by DNase I. In order to visualize the binding site, diluted DNase I is added to the DNA-protein mixture; digestion time is limited so that, on average, each probe molecule is only nicked once at random locations in regions where it is unprotected by bound protein. The reaction is stopped; and the DNA is extracted, precipitated, and electrophoresed on a denaturing acrylamide gel in order to resolve the DNA fragments into a ladder of progressively smaller fragments. The gel is exposed to film. Upon development, protected regions can be visualized as interruptions (footprints) where bound protein had prevented DNase I nicking. These experiments can detect binding but *in vivo* experiments are generally required to test functions and importance of particular binding sites with regard to the tissue and developmental stage of interest.

The chromatin immunoprecipitation (ChIP) assay examines the transcription factor-*cis* regulatory element interaction in living cells. The protein is cross-linked with
DNA when cells are grown under the desired experimental conditions. After cell lysis, chromatin is mechanically sheared and then immunoprecipitated (IP) with protein using antibodies. Cross-links are reversed and only DNA is purified. Finally, DNA can be amplified by particular PCR primers to recover the *cis*-regulatory elements.

To expand the target DNA of ChIP assay from defined primers to all DNA enriched in IP, the microarray is used in ChIP-chip assay. Instead of amplifying DNA by defined primers, Chip-chip assay labels IP-enriched DNA with one fluorophore (Cy3) by ligation-mediated PCR (LM-PCR). Another pool of non-enriched DNA is labeled with another fluorophore (Cy5). Both of the labeled DNAs are hybridized to a microarray of genomic DNA probes and the genome-wide IP-enriched *cis*-regulatory elements can be analyzed (Horak and Snyder, 2002; Sikder and Kodadek, 2005).

1-1.4. Comparative genomics

1-1.4.1. Evolutionarily conserved genes

The identification of *cis*-regulatory elements can be facilitated by using comparative genomics. Based on the principle that essential regulatory regions of orthologous genes in related species are evolutionarily constrained by purifying selection, a search for islands of conserved sequences can be used to identify candidate *cis*-regulatory regions for further structural and functional analysis.

Comparative genomics has been used to guide annotation of protein-coding sequences in that functional protein-coding sequences maintain higher similarity during evolution. This has been demonstrated by many gene prediction programs such as
ROSETTA, SGP1, SGP2, TWINSCAN and DOUBLESCAN (Batzoglou et al., 2000; Korf et al., 2001; Meyer and Durbin, 2002; Parra et al., 2003; Wiehe et al., 2001).

It is easy to understand that some genes, such as housekeeping genes, encode the proteins critical for cell division, differentiation, proliferation and normal function. The very basic machinery used by every cell has to be maintained through evolution. Although there are always great diversities for animal developmental strategy, expression patterns of conserved genes are sometimes also conserved during embryogenesis, such as body axes and homologous structures. This means that the regulatory patterns of gene expression can be under positive selection sufficient to preserve high similarity between diverged species. For example, comparative analysis of human and chicken genomes has been shown to be useful in identifying conserved coding regions as well as functionally important non-coding sequence elements (Hillier et al., 2004). Similarly, comparison of genomic sequences from mouse and the puffer fish *Fugu rubripes* separated by about 430 million years of evolutionary distance has revealed high sequence similarity between orthologs of developmentally important genes such as HOXD4 which involves not only conservation between coding exons but also between *cis*-regulatory elements of these genes (Aparicio et al., 1995). Most recently, a genome wide comparison of human and puffer fish genomes revealed, surprising, that the most highly conserved non-coding DNA sequences shared between these species are actually highly clustered in the genome. Close inspection revealed that the common defining feature for these clusters of CNR’s was their proximity, in each case, to a gene encoding a developmentally regulated transcription factor. The study identified a set of 46 genes, termed “trans-dev” genes, all of encoding transcription factors expressed during embryonic development (Woolfe et al.,
2005). These genes included *TFAP2A* (human ortholog of *dAP-2*) as well as multiple representatives from HOX, PAX, DLX, and SOX gene families to name a few. Based on this solid placement of the AP-2 family into such a select group of developmentally important genes, a study of *cis*-elements for regulation of *Drosophila* transcription factor AP-2 would seem to be a worthwhile endeavor.

Entire genome sequences are now available for a large number of organisms (Dickmeis and Muller, 2005). The rapidly growing field of bioinformatics will undoubtedly continue to provide new algorithms and programs for comparative genomics studies to facilitate future studies of *cis*-regulatory regions in genes.

### 1-1.4.2 Promoter prediction

Among the various *cis*-regulatory elements, clusters of which are also known as *cis*-regulatory modules (CRMs), core promoters are the best studied due to their defined position relative to the transcription start site. The prediction of core promoters can be done with computer programs such as PromoterInspector which examines the sequence similarity to known core promoter elements and McPromoter which takes into account additional features such the joint consideration of DNA sequence as well as profiles of physical properties like CpG content (Rombauts et al., 2003).

Trinklein et al. aligned full-length cDNA clones from the Mammalian Gene Collection to the human genome rough draft sequence to estimate the start sites of more than 10,000 human transcripts and design putative promoters upstream. Among them, 152 randomly selected promoters were tested in luciferase-based transfection assay in cell lines and 90% showed significant transcription activity (Trinklein et al., 2003).
Another study comparing mouse and human promoter sequences found conserved block structure located within 500bp from the INR and transcription factors and brain-specific genes have shorter conserved blocks than other genes (Suzuki et al., 2004).

1-1.4.3 Enhancer prediction

Prediction of another cis-regulatory elements – enhancers and silencers – is becoming the main targets of comparative genomics. Enhancers and silencers usually contain clusters of transcription factor binding sites (TFBS), such as in *Drosophila even-skipped* gene stripe 2 enhancer. These transcription factors synergistically or independently activate or repress gene expression (Small et al., 1992). The high level of sequence conservation can also help the detection of enhancers. However, the prediction of enhancers has much more complicated factors, such as species-specific sequence change due to stabilizing selection, flexibility of enhancer distance from the coding sequence, and changes in CpG island density in mammals (Miller et al., 2004).

Current models of using interspecies alignments for predicting clustered enhancers evaluate the alignments either for their quality, with higher scores for more slowly changing regions, or for characteristic patterns. Like gene prediction algorithms, the aim is to find patterns in the alignment that are characteristic of known CRMs but are rarely seen in neutrally evolving DNA. One approach uses alignments with a set of known CRMs as positive training set and alignments with ancestral repeats as negative training set, then compute the log likelihood that any alignment has patterns of alignment symbols more characteristic of CRMs than neural DNA. This computed regulatory
potential has shown improvement in separation between regulatory and neutral sequences and would aid the identification of putative regulatory regions (Elnitski et al., 2003).

Identification of enhancers can be based on transcription factor binding profile. *In vitro* and *in vivo* analysis for TFBS were used to construct consensus sequences or the more flexible positional weight matrices, which include information about the frequencies of different nucleotides in different positions of the binding site (Hertz and Stormo, 1999). There are large databases of these transcription factor binding profiles available, such as TRANSFAC or JASPAR (Sandelin et al., 2004; Wingender et al., 1996). Studies have been done to identify genome-wide targets of transcription factors, such as Dorsal and Suppressor of Hairless (Su(H)) in *Drosophila* (Markstein et al., 2002; Rebeiz et al., 2002).

Programs using consensus sequence such as flyenhancer (www.flyenhance.org), or weight matrix such as Target Explorer (http://trantor.bioc.columbia.edu/Target_Explorer/) are available online (Markstein et al., 2002; Sosinsky et al., 2003).

1-1.4.4 Phylogenetic footprinting

“Phylogenetic footprinting” is a term inspired by the wet-lab technique of DNAse footprinting. The reasoning behind the approach is that, just like coding sequences, regulatory elements are functionally important and are under evolutionary selection, so they should have evolved much more slowly than other non-coding sequences. Since phylogenetic footprinting was applied on *cis*-regulatory element prediction, it has been found that non-coding sequences that are conserved among evolutionarily distant species.
predominately clustered around a subset of genes. These *trans-dev* genes, including *TFAP2A* as one of the top 46 genes, are known to be highly constrained, with a conserved pattern of expression among species across extreme evolutionary distances (Woolfe et al., 2005).

It is reasonable to think that some of the *cis*-regulatory elements that control these genes are similarly constrained to precisely preserve the expression levels and patterns of genes that are crucial for basic embryonic development (Boffelli et al., 2004). For example, in a sequence comparison of human and zebrafish *Shh* loci, apart from exons, there were only a few sequence “islands” showing significantly higher conservation. These islands were either the promoter region or proved to include enhancers ar-A and ar-C by functional assays. When the conserved ar-C was introduced into transgenic animals, it directed floor plate and notochord expression in both mouse and zebrafish (Muller et al., 2002).

Another study comparing *Sox2* loci between chicken and mammals revealed an array of more than 20 scattered conservation blocks. Interestingly, the functionally defined neural enhancers all fell into one of these blocks. Similar examination of *Pax6* loci also confirmed these conserved sequence blocks match the coordinates of functionally determined enhancer (Uchikawa et al., 2004).

### 1-1.4.5 Phylogenetic shadowing

A variant of phylogenetic footprinting, phylogenetic shadowing, was used to identify functional CRM in conserved sequences specific for primates. Due to many accumulated sequence changes in closely related primates, pairwise comparisons could
could not reveal differences between functional and non-functional CRMs. Phylogenetic shadowing uses collective sequence variation to separate slow evolving sequences, which often correspond to functional sequences, from fast evolving sequences (Boffelli et al., 2003). In the subsequent deletion analysis, nine out of ten conserved regions were found to be functional CRMs, whereas 7 non-conserved regions were likely non-functional.

1-1.4.6 Comparative genomics programs

Many programs have been developed for identification of CRMs with the assistance of phylogenetic footprinting. They reflect different alignment algorithms, selection of TFBS database, setting and training data tailored for different organismal models. Two examples are Consite (www.phylofoot.org) developed by Karolinska Institute and VISTA (http://pipeline.lbl.gov/cgi-bin/gateway2) developed by Lawrence Berkeley National Laboratory. Consite, one of the programs that include CRM database, genomic alignment and visualization in a single package, has many unique features. One distinguishing feature is that it uses a refined database, JASPAR, which was made by searching the literature extensively in order to assemble 108 high-quality TFBS profiles. Compared to most programs using TRANSFAC database enriched with vertebrate TFBS, Consites’ TFBS database contains more refined weight matrices for Drosophila transcription factors (Lenhard et al., 2003).

VISTA uses the most updated genomic sequences of many species from NCBI and global alignment algorithms such as AVID and LAGAN. The visualization of whole-genome alignments of multiple species via VISTABrowser is very user-friendly.
Prediction of CRM using TRANSFAC database can be done through rVista on UCSC Browser (Couronne et al., 2003).

rVista is also used by other programs for CRM prediction. zPicture sends BLASTZ alignments used by Pipmaker to rVista, and GALA also automatically forwards genome scan to rVista (Giardine et al., 2003; Ovcharenko et al., 2004; Schwartz et al., 2000). However, rVista program linked to UCSC Browser has problems using user-defined consensus sequence, including failure to search both DNA strands and to specify exact matches to the consensus sequence. Although most of these programs use vertebrate sequence as training data, some program, like Target Explorer, considers parameters specifically for *Drosophila*.

The major source for alignments, including multiple alignments, is the UCSC Genome Browser (Kent et al., 2002). Conservation scores and regulatory potential scores based on the alignments are highly effective.

1-1.4.7 Identification of regulatory motifs for co-regulated genes

Another relatively new comparative genomics method to identify specific sequence motifs based on comparisons of known co-regulated genes as a means of discovering regulatory motifs for unknown (or unidentified) transcription factors. Grad et al. recently proposed an algorithm to assess binding site conservation independently of knowledge about the transcription factors involved (Grad et al., 2004). This algorithm identifies clusters of ungapped conserved subsequences and then searches for the most similar clusters within a set of co-regulated genes. Genome-wide search of *Drosophila* blastodermal enhancers search indicated 74% of the hits were in proximity to genes
showing expression in blastoderm. Another novel method uses statistical method that can measure the frequency of short runs of alignment columns in regulatory regions compared with neutral DNA, creating a “regulatory potential score” for such sequences (Kolbe et al., 2004). It is reasonable to assume that many if not most cases of co-regulated genes are regulated at the transcriptional as opposed to post-transcriptional level. Therefore, co-regulated genes should contain shared cis-regulatory motifs which will be statistically overrepresented within the intergenic regions of co-regulated genes compared with their appearance within unrelated genes.

AlignACE, a motif discovery program, was used to align the cluster elements to discover regulatory motif. When Halfon et al. filtered these motif and only retained those conserved between D. melanogaster and D. pseudoobscura, 4 of 5 binding sites in the model enhancer were found in other functionally related genes. Ten more motifs were also identified including a functional POU/OCT1 transcriptional repressor binding site (Halfon et al., 2002).

1-1.4.8 Consideration of phylogenetic footprinting application

Although integration of phylogenetic footprinting appears to be helpful in many studies, its application needs careful consideration. First, evolutionary distance sometimes makes the choosing of proper orthologs difficult. Although comparison between extremely distant species may identify a few ultraconserved enhancers like those in the DACH gene, distant comparisons usually return low or no similarity. Conversely, if compared species are too close, there will be too much background “noise” that hinders the identification of true CRMs (Boffelli et al., 2004; Lenhard et al., 2003).
Second, evidence has shown that enrichment for regulatory elements in conserved sequences is frequently located close to developmentally regulated genes, especially genes for early development (Dickmeis and Muller, 2005). Although this is consistent with the idea that proper regulation of developmental genes is critical and highly conserved, we need to consider that the application of phylogenetic footprinting may not be suitable for all kinds of genes.

Finally, prediction of CRMs from comparative genomics needs confirmation and feedback from in vivo or in vitro experiments. Producing transgenic animals on a high-throughput scale for cis-regulatory element analysis is tedious (and in the case of transgenic mice, costly), but such experiments generate the only real proof of prediction for comparative genomic studies. Cheaper high-throughput analysis of cis-regulatory elements may be possible in zebrafish or chick embryos. For example, investigators in one study were able to scan a 50kb region of chicken Sox2 locus for enhancer activity using embryo electroporation (Uchikawa et al., 2004). A number of enhancers identified by functional analysis exactly correspond to extragenic sequence blocks that are conspicuously conserved between the chicken and mammalian genomes and embedded in sequences with a wide range of sequence conservation between human and mice.

1-2  *Drosophila* leg development

Molecular genetic analysis of *Drosophila* leg development has made steady progress in the past ten years in defining essential regulatory genes that function in specification and development of the limb and its constituent segments (reviewed by
Kojima, 2004). Although separated by 600 million years of evolution, the *Drosophila* leg and vertebrate limb still share a number of important similarities in use of conserved signaling pathways and orthologous transcription factors that set up positional information specifying the anteroposterior (AP), dorsoventral (DV) and proximodistal (PD) axes of the growing limb. For example, in both species, Hedgehog (Hh) signaling and Ci/Gli transcription factors regulate limb AP axis polarity, Wingless (Wg)/Wnt and Dpp/BMP signaling and LIM homeodomain factors affect dorsal-ventral patterning, Hth/Meis and Exd/Pbx homeodomain transcription factors specify proximal limb regions (limb parts closer to main body axis), while Distalless (Dll)/Dlx family homeodomain genes and Notch, Wg/Wnt and Dpp/BMP signals are important in specifying distal limb regions (Beermann et al., 2001; Capdevila and Johnson, 2000; Irvine and Vogt, 1997; Mercader et al., 1999; Panganiban and Rubenstein, 2002; Rauskolb and Irvine, 1999; Schoppmeier and Damen, 2001; Wu and Rao, 1999).

In *Drosophila*, precursor cells for the adult legs, antennae, proboscis, wings, and other structures are specified during embryogenesis and invaginate from specific sites in embryonic ectoderm to form small epithelial sacs which grow during larval stages into anlagen called the imaginal discs. Precursors for the 6 leg imaginal discs arise together in ventrolateral regions of the three thoracic segments during embryogenesis (Cohen, 1990). The C$_2$H$_2$ zinc-finger transcription factors buttonhead (btd) and *Drosophila* Sp1 are expressed in very early leg primordia under the control of Dpp and Wg, and mediate the activation of Distalless (Dll) expression by Wg signaling to establish the future “distal organizer” of the nascent limb (Estella et al., 2003). Dll and its orthologs in other animal
species have central, ancient roles in development of all types of body appendages (Panganiban and Rubenstein, 2002).

The leg disc anlaga initiate their first PD subdivision in late-embryogenesis with the activation of Homothorax/Extradenticle (Hth/Exd) expression in cells in the presumptive proximal region of the leg surrounding the central domain of Dll-expressing cells which are precursors for future distal parts of the leg (Gonzalez-Crespo et al., 1998; Kubota et al., 2003). Antennal imaginal discs also show this same early PD-axis subdivision. However, while antennal discs maintain this “ground state”, in leg discs, the homeodomain protein Antennapedia (Antp) prevents the default antennal fate by repressing Hth expression in intermediate PD-axis regions and restricting Hth activity to a narrower proximal region than in antennal discs. Antp loss-of-function mutations cause leg-to-antenna transformations in Drosophila and other arthropods, indicating that Antp has an evolutionarily conserved function in determining leg versus antennal identity (Casares and Mann, 2001; Cummins et al., 2003).

1-2.1 Patterning the proximodistal axis of the leg

The adult leg of Drosophila consists of 10 segments (coxa, trochanter, femur, tibia, tarsals 1-5, and pretarsus) separated by flexible joints. During larval development, the imaginal disc cells show increased rates of proliferation relative to earlier stages. In a manner somewhat reminiscent of embryonic segmentation, the proximodistal axis of the leg is progressively specified by sets of regulatory genes with increasingly fine expression PD-axis domains. During the larval second instar stage, the nuclear protein Dachshund (Dac) is induced by Wg and Dpp to be expressed in intermediate regions of
the leg discs. The distal and proximal borders of the Dac expression domain overlap with the distal Dll and proximal Hth expression domains (Mardon et al., 1994). Dac expression in the presumptive femur segment represses Hth expression in the presumptive coxa and is itself repressed by zinc-finger protein Teashirt (Tsh) expressed in the proximal part of the leg (Wu and Cohen, 2000). Dll and Dac repress each other to separate medial and distal leg segments (Dong et al., 2001). More importantly, the expression domains of Hth, Dac and Dll roughly correspond to the regions of the leg affected by their absence. Dll, Dac and Hth have sometimes been referred to as “leg gap genes” based on their resemblance in function and expression to the gap gene class of embryonic segmentation genes (Rauskolb, 2001). During the larval third instar stage, Dac expression expands both distally and proximally, so that it overlaps with the distal Dll expression domain in the presumptive tibia and with a late-arising proximal domain of Dll expression in the presumptive trochanter and proximal femur.

Thus, specification of several subregions within proximal, intermediate and distal leg regions is accomplished by interactions among Dll, Dac and Hth. Subdivisions of tarsus and pretarsus are also achieved stepwise, under the regulation of Dll, EGFR and Notch signaling. Three genes, spineless (ss) encoding a homolog of the dioxin receptor, rotund (rn) encoding a zinc-finger protein, and bric-a-brac (bab) encoding a BTB domain transcription factor, are expressed in the distal edges of tarsal 1-4. Activated by Dll, ss and rn are only transiently expressed from the end of 2nd instar to mid 3rd instar stage (Couso and Bishop, 1998; Duncan et al., 1998; St Pierre et al., 2002). Bab instead is expressed consistently and its expression level forms an increase gradient from tarsal 1 to tarsal 4, plus a weak expression in tarsal 5. All of the three genes are required for the
formation of tarsal 1-4 based on observation that in the loss-of-function mutant of each one, fusion of corresponding tarsal segments happens (Couderc et al., 2002; de Celis Ibeas and Bray, 2003; Godt et al., 1993).

Another four genes, Bar and aristaeless (al) encoding homeodomain proteins, apterous (ap) and Lim1 encoding LIM-homeodomain proteins, are expressed from early to mid third instar and required for formation of tarsal 4, 5 and pretarsus. Bar expression starts with tarsal 3-5 and later only keeps increasing gradient from tarsal 4-5 at late 3rd instar stage. The weak expression of Bar plus tarsal 4-specific ap specify the identity of tarsal 4. The strong expression of Bar at tarsal 5 determines tarsal 5 identity. Al and Lim1 expression in pretarsus together specify this segment (reviewed by Kojima, 2004).

The differential expressions of tarsal segment specific genes are not independent and many of them were found under the regulation of EGFR signaling recently (Campbell, 2002; Galindo et al., 2002). Secretion of EGFR ligands such as vein (vn), and processing and activation of these ligands by rhomboid (rho) from the distal end of leg disc result in the distal-to-proximal gradient of EGFR signaling, which differentially activates al, Lim1 and Bar in pretarsus and tarsal 5 segments, as well as represses rn and bab in proximal tarsal segments. Bar was also found required for ap expression in tarsal 4, as well as determining the proximal limit of al and Lim1 where tarsal 5/pretarsus boundary was drawn (Kojima et al., 2000; Pueyo et al., 2000; Tsuji et al., 2000).

Another model of gene interaction determining tarsal segments development came from studies of the gene bab. Brother of odd with entrails limited (bowl), a Notch-target gene encoding a C2H2 zinc-finger protein, was found only expressed outside the proximal and distal boundaries of bab expressing domain and negatively regulates bab
expression. The proposed model was that as tarsal cells proliferate, the *bowl* expression is only retained at the boundary and lost in the intervening cells where *bab* is derepressed. Activation of *bab* in turn repress *dac* at the proximal tarsal boundary and *Bar* at the distal tarsal boundary (de Celis Ibeas and Bray, 2003). The other three odd-skipped family members downstream of Notch, *odd-skipped (odd)*, *sister of odd and bowl (sob)* and *drumstick (drm)* were found to induce invagination in the leg disc epithelial cells and morphological changes in the adult legs, but not affect leg segmentation (Hao et al., 2003).

1-2.2 Notch signaling in leg joint development and leg elongation

The Notch signaling pathway was first discovered in studies of *Drosophila* neurogenesis. Through ligand-receptor interaction, this cell:cell signaling pathway is highly conserved from worms to mammals and plays an important role in cell fate decisions in multiple tissues during animal development (Artavanis-Tsakonas et al., 1999; Bray, 1998; Lewis, 1998).

Notch family receptors (Notch 1-4 in vertebrates) on the signal receiving cells are activated by Type I transmembrane ligands on the signal sending cells. These ligands, known as DSL, include Delta (Dl) and Serrate (Ser) in *Drosophila*, Jagged (Serrate homolog) in vertebrate and Lag-2 in *C. elegans*. The ligand activation recently was linked to internalizations by ubiquitin-mediated endocytosis. *Neuralized (neur)* in *Drosophila* and *mind bomb (mib)* in Zebrafish both encode E3 ubiquitin ligase that promotes endocytosis. Internalization of ligands physically pulls Notch extracellular domain, which binds to ligands’ extracellular domain via EGF repeats, from the surface of
receiving cells and promotes proteolysis of Notch receptor (Kadesch, 2004). During the maturation of Notch within the secretory pathway, a furin-like convertase cuts at Notch S1 site and therefore separates extracellular domain from transmembrane and interacellullar domains. Upon activation by ligands, mature Notch receptor is cleaved by extracellular metalloproteases of the ADAM/TACE/Kuzbanian family at S2 site. After release of extracellular domain, the truncated form of Notch is activated and subject to further proteolysis by presenilin-containing gamma-secretase complex at S3 and S4 sites within transmembrane domain (Mumm and Kopan, 2000; Schweisguth, 2004). Recent study reported that the cleavage by gamma-secretase requires monounibquitation of Notch intracellular domain and endocytosis (Kanwar and Fortini, 2004).

The intracellular domain of Notch (N\text{\textsuperscript{icd}} or N\text{\textsuperscript{intra}}) is then freed to enter nucleus and bind to transcription factors LAG-1 in \textit{C. elegans}, Suppressor of Hairless (Su(H)) in \textit{Drosophila}, and CBF-1 or RBPJ\kappa in mammals, collectively known as CSL. The binding switches CSL from repressor to activator with a transcriptional activation complex including cofactors Mastermind and histone acetyltransferase p300. Activated Notch targets include HES (Hairy/Enhancer of split) bHLH (basic helix-loop-helix) repressors in \textit{Drosophila} and mammals.

Notch signaling has been involved in different ways to determine cell fate. The classic model is lateral inhibition in \textit{Drosophila} sensory organ precursor (SOP) development. When one cell in a proneural cluster was picked as SOP, it expresses Notch ligand that activates Notch receptor in all the neighboring cells and inhibits them from adopting neural fate. Later, the division of SOP also requires Notch signaling to specify cell lineage. Asymmetric segregation of membrane protein Neur and Numb into pIIb cell
upregulates DI activity and inhibits Notch activity in this cell, and promotes Notch activity in pIIa cell.

Another usage of Notch signaling is demarcating boundaries in *Drosophila* wing and leg development and vertebrate somitogenesis (Bishop et al., 1999; de Celis et al., 1998; Rauskolb and Irvine, 1999). *Drosophila* leg joint formation requires Notch signaling in the distal edge of each segment activated by ligands expressed proximal to it. However, Notch signaling will not be activated on the proximal side of the ligand expressing cells because Fringe modifies glycosylation on specific Notch EGF-repeats and inhibits ligand-receptor interaction on the proximal side. Some Notch targets, such as Bowl and dAP-2, are found required for joint formation. dAP-2 is also cell non-autonomously required for interjoint cell survival, different from Notch’s growth promoting function (Kerber et al., 2001). The vertebrate homologs of Hairy and Fringe, HES and Lunatic fringe (L-fringe), are regulated by both Notch activation and negative feedback loop, generating oscillating expression pattern required for somitogenesis (Bessho and Kageyama, 2003).

In *Drosophila*, the function of Notch is almost always carried out through transcription factor Su(H). Therefore, several direct Notch targets have been found *in vivo* or *in vitro* by identifying Su(H) binding sites, which often enrich in clusters in *cis*-regulatory elements (Bailey and Posakony, 1995). Genome-wide computational approaches were recently used to identify a number of new direct targets containing clusters of Su(H) binding sites (Rebeiz et al., 2002). To activate cell-specific expression of target gene, Notch is often working cooperatively with other signaling pathways or transcription factors. For example, synergy between Notch and BMP receptor signaling is
required for activation of Herp2 genes in endothelial cells (Itoh et al., 2004). Notch activation and Dorsal/Twist together stimulate single-minded (sim) gene expression in single row of cells in the Drosophila mesectoderm (Morel and Schweisguth, 2000). Interaction between Su(H) and Daughterless mediated by specific cis-element code is necessary for activation of E(Spl)m8 (Cave et al., 2005).

1-2.3 dAP-2 in Drosophila development

AP-2 alpha was first identified as an activator of SV40 early and human metallothionein IIA promoters. The transcription factor AP-2 family now is known to include five members conserved in humans and mice (AP-2 alpha, AP-2 beta, AP-2 gamma, AP-2 delta, and AP-2 epsilon) (Feng and Williams, 2003; Hilger-Eversheim et al., 2000; Mitchell et al., 1987; Zhao et al., 2001). The homologs of AP-2 alpha so far have been studied in various species including mouse, chicken, zebrafish, Xenopus and Drosophila (Epperlein et al., 2000; Knight et al., 2004; Knight et al., 2003; Luo et al., 2002; Meulemans and Bronner-Fraser, 2002; Mitchell et al., 1991; Monge and Mitchell, 1998; Shen et al., 1997). These proteins have substantially diverged N-terminal halves which may differ with regard to the interacting proteins, and share highly conserved helix-span-helix DNA binding and dimerization domain in C-terminal halves (Braganca et al., 2002; Williams and Tjian, 1991).

Analysis of the in vivo expression of mouse AP-2 family genes indicated that they have relatively restricted, cell type-specific expression patterns in organs and tissues of adult mice (Schorle et al., 1996; Williams et al., 1988). During mouse embryogenesis, AP-2 family genes show overlapping expression patterns in neural crest cells, facial and
limb mesenchyme, with neural tissues being major sites of expression (Chazaud et al., 1996; Mitchell et al., 1991; Moser et al., 1995; Moser et al., 1997). Null mutants of three murine AP-2 genes (AP-2 alpha, AP-2 beta, and AP-2 gamma) are embryonic or perinatal lethal, due to defects in cranial neural crest cells, kidney distal tubule epithelia, and placental development respectively (Schorle et al., 1996; Zhang et al., 1996). The evolutionary conservation of AP-2 function was also demonstrated by the fact that both mouse and zebrafish tfap2a are required for Hoxa2 expression in development of cranial neural crest cells in the second branchial arch (Knight et al., 2003; Maconochie et al., 1999). The human Char Syndrome, caused by a similar Arg to Cys dominate mutation in AP-2 beta DNA binding domain also found in Drosophila dAP-2 alleles, shows heart defects requiring neonatal surgery and facial and hand anomalies which resemble orofacial structures/proboscis and limb phenotypes in mouse and Drosophila mutants (Monge et al., 2001; Satoda et al., 2000). As an essential transcription factor involved in cell growth, differentiation and apoptosis, AP-2 have a broad range of regulatory relationship with genes in normal tissue and cancer progression (Hilger-Eversheim et al., 2000).

In Drosophila, dAP-2 is the sole member of AP-2 family genes, carrying the essential functions derived from the common ancestor and shared with vertebrate homologs. Both genomic structure and protein sequences of dAP-2 are slightly more similar to AP-2 alpha, featuring two alternative first exons and amino acid identity in conserved C- and N-terminal functional domains (Bauer et al., 1998; Monge and Mitchell, 1998). dAP-2 is expressed in maxillary segment and protocerebrum during embryogenesis, as well as central nervous system, labial, leg and antenna imaginal discs
during third instar larval stage (Monge and Mitchell, 1998). EMS-induced mutagenesis generated dAP-2 null and hypermorphic alleles, which gave the mutant phenotype including shortened leg segments and missing joints, reduced proboscis and impaired locomotion (Monge et al., 2001). The colocalization of dAP-2 with Notch target genes in concentric ring expression pattern in leg imaginal disc, and its dependence on Su(H), indicate dAP-2 responds downstream to Notch signaling pathway in presumptive joints. In addition to the cell-autonomous role in joint formation, dAP-2 is also required cell non-autonomously for inter-joint cell survival (Kerber et al., 2001).

Although important for animal development, the transcriptional regulation of AP-2 family genes was only studied on cis-regulatory elements of human AP-2 alpha. Conserved among vertebrates, the proximal promoter of human AP-2 alpha contains binding sites for AP-2, NF-1 and octamer proteins but lacks a TATA box motif. Protein assay and cell transfection analysis indicated octamer and initiator elements are critical for promoter activity (Creaser et al., 1996). Enhancer analysis of human AP-2 alpha using transgenic mice identified discrete cis-elements including a face and limb-specific enhancer within fifth intron, and suggested potential autoregulation (Zhang and Williams, 2003).
Chapter 2

Transcriptional regulation of *Drosophila dAP-2* in leg and antennal discs

2-1 Introduction

The *dAP-2* gene, encoding *Drosophila* transcription factor AP-2, is activated along the proximal-distal (PD) axis of the developing leg in a segmental pattern that is Notch-signaling dependent. Prior to this thesis research, it was unclear whether this pattern of expression was directly or indirectly activated by Notch signaling; and cis-elements for transcriptional regulation of *dAP-2* by Notch or any other signaling pathways and transcription factors were unknown. Notch signaling has important roles in cell fate choice, stem cell survival, and cell differentiation in many tissues during development. However, surprisingly few direct target genes for Notch/Su(H) have been identified and the known targets do not account for all functions of Notch. Much remains to be understood about how Notch signaling is cell type-specifically interpreted. For example, although dAP-2 is activated by Notch signaling in the leg, it is not activated in many other sites where Notch signaling is active (sites in the eye and wing disc, to name just a few). Thus, a systematic analysis of *dAP-2* cis-regulatory regions could provide many new insights, not least of which could be the identification of regulatory motifs that mediate a context-dependent response to Notch signaling.
In Chapter 2 of this thesis, I report the identification of multiple cis-regulatory elements that are responsible for different portions of the highly specific expression pattern of \(dAP-2\) in developing legs and antennae. Most of the enhancer elements that activate the 9 domains of \(dAP-2\) expression along the PD-axis of the developing leg are present on two genomic fragments, BXE and E6. Together these two fragments can direct expression of a \(dAP-2\) cDNA sufficient to rescue joint development and leg outgrowth in \(dAP-2\) null mutant flies.

In early experiments presented in Chapter 2, I established that Notch signaling is only required for \(dAP-2\) expression in the presumptive distal parts of the leg and antenna, while expression in the more proximal parts of the leg and antenna is activated without a requirement for Notch signaling. I also found that the precise location of \(dAP-2\) expressing cells at leg segment boundaries/future joints differs along the PD-axis of the leg. At presumptive tarsal/tarsal joints (in the distal part of the leg), \(dAP-2\) expressing cells are located at the distal ends of tarsal segments (i.e., on the proximal side of the boundary/joint). However, there is a gradual shift in location of \(dAP-2\)-expressing cells in presumptive joints of more proximal leg segments. For the 3 most proximal leg segments (coxa, trochanter and femur), the AP-2 expressing cells are at the proximal ends of each leg segment (i.e., on the distal side of the boundary/joint). As I will show, these differences also correlate with a fundamental separation of the regulatory regions in \(dAP-2\) that control it’s expression in distal versus proximal leg segments.

Much of Chapter 2 concerns the \(dAP-2\) enhancer, BXE, located 5 kb upstream of exon 1a. \(dAP-2\) has two alternative promoters; of these, only the exon 1a promoter is active in the leg and antennal imaginal discs. I showed that BXE functions most
effectively in combination with this promoter. BXE is sufficient to activate reporter gene expression in the four \textit{dAP-2} domains that mark presumptive tarsal joints, as well as in the 4 distal-most rings of dAP-2 in the antennal disc. Phylogenetic footprinting to compare the \textit{dAP-2} genes of \textit{D. melanogaster} and other \textit{Drosophila} species showed that BXE contains multiple peaks of extended DNA sequence conservation. Further deletional analysis revealed that some of these regions are expendable for leg and antennal expression, while others are not. Three essential subdomains that function synergistically to mediate Notch-responsive induction and dAP-2-dependent autoregulation were identified in BXE. In these regions, I identified two essential binding sites for Su(H), the transcription factor through which Notch activates target genes. One of these is present in an extended peak of high conservation while the second site is in a poorly conserved region. Both sequences bound bacterially-expressed, purified Su(H) protein with high affinity in \textit{in vitro} DNA binding bandshift experiments. Mutation of either site in the context of BXE-reporter transgenes abolished the ability of BXE to drive expression in leg discs \textit{in vivo}. The same mutations also abolished in vitro DNA binding of Su(H). Genetic mosaic analysis confirmed Su(H) was required cell-autonomously for BXE-lacZ expression in leg and antennal discs. In contrast, another reporter gene E6-lacZ which drives expression in multiple dAP-2 domain in proximal leg regions was unaffected by loss of Su(H) in these regions.

I discovered that both endogenous \textit{dAP-2} expression and \textit{BXE-lacZ} expression are attenuated in future tarsal regions of leg discs from \textit{dAP-2} null mutant larvae. This loss of expression was not likely to be caused by cell death of joint cells because expression of several other tarsal joint markers was retained in the mutant leg discs. Two conserved
consensus AP-2 binding sites were identified within BXE. Mutagenesis of each of these revealed that one of the two sites is essential for BXE activity supporting the idea that dAP-2 positively autoregulates its own transcription. These data strongly suggest that dAP-2 expression in presumptive tarsal joints is activated by Notch signaling and maintained by subsequent autoregulation.

Taken together, my research data suggest that dAP-2 is regulated by different mechanisms in distal versus proximal regions of the leg and antenna, and that these differences involve a physical separation of enhancer elements used to carry out these regulations.
2-2 Material and Methods

2-2.1 Reporter constructs

Two kinds of reporter constructs, tauZ vectors based on pREPtauZ backbone and Pelican lacZ or GFP vectors based on pPelican or pStinger backbone, were used. pREPtauZ is based on P-element construct pKB256, which was designed for heat-shock induction of a particular cDNA in *Drosophila* (made by K. Basler). The heat-shock inducible promoter of hsp70 gene was amplified by PCR from –250 to +90 to make pKB256 and previously served well to integrate the activity of heterologous enhancer. The tau-β-galactosidase fusion protein (Callahan and Thomas, 1994) followed by the SV40 3’-UTR was used to replace tubulin trailer in pKB256 and thus obtained pREPtauZ vector.

pPelican or pStinger vectors (Barolo et al., 2000) are promoter-less vectors with lacZ or nuclear GFP gene and gypsy insulator sequence to reduce position effect. A heat-shock basal promoter (also called hspTATA) is present in pHPelican (hlacZ) and pHStinger (hGFP). Minimal promoter of dAP-2 1a and full length 5’-UTR was amplified by PCR from –207 to +425 and digested by EcoRI/BglII to clone into pPelican for the base of most reporter vectors (alacZ). Another dAP-2 1a promoter sequence from –473 to +31 was amplified by PCR and digested by BamHI/BglII and used in a few Pelican vectors (APel). dAP-2 1b promoter sequence from –190 to +154 was amplified by PCR and digested by BamHI/BglII and used in blacZ vectors.

*Cis*-regulatory elements were either digested using available enzyme or amplified by PCR. When PCR was used, the fragment was first cloned into pGEM T-easy vector.
(Promega) and sequenced to rule out undesired mutation, then cut out and put into P-element transgenic vectors.

The following primers were used in genomic PCR to amplify B6 fragment and dAP-2 promoters:

B6-5’: 5’-CAGAACTCACAATCCAGG –3’
B6-3’: 5’-ACTCTCCCCGATTGAAGG-3’
alacZp(alp)-5’: 5’-ACTGGATCCTCACACCCATACAGACG –3’
alacZp(alp)-3’: 5’-CGGAGATCTGTTGATCCAAATTAGACGG-3’
blacZp(1bp)-5’: 5’-GGATCCAAATCGGCGACAATCTCTC-3’
blacZp(1bp)-3’: 5’-AGATCTGTTCTGGGCGACAAGTC-3’
Apelp(d2p)-5’: 5’- GTGGGATCCGTATCGACATCTCG-3’
Apelp(d2p)-3’: 5’- GGTAGATCTTTTGCAGCCACCAGGAGTAG-3’

BXE, BX, XE, B6EB fragment were digested using restriction enzymes from B6.

Deletions of B6 and BXE were constructed by PCR from B6 using following primers:

BX0.8-3’: 5’-TCAGATCTGCAGCGTACATAAATTGCGAGC-3’
BX1.4-3’: 5’-TTGTCTAGAGTTCTCGTTGCGCCAGG-3’
BX0-5’: 5’-GACAGATCTCCAGAATCCAGAACCGAG-3’
BX0.5-3’: 5’-CGATCTAGAAGCGATACTCAAACGAGAG-3’
BX1.0-5’: 5’-TGTACGCTGCAGATCTGAAAACGCACACACC-3’
BX1.5-5’: 5’-TCGAGATCTGCTGATGATTCGCATTCGTG-3’
BX0.5-3’: 5’-CGATCTAGAAGCGATACTCAAACGAGAG-3’
BX1-3’: 5’- CCTTCTAGATATTTTCTACGGCTGTGCTC-3’
BX1.5-3’: 5’- GGATCTAGAATGCGAATCATCAGCGAC-3’
BX2-3’: 5’- GCATCTAGAAGCAGTTTTAGAGCAAGG-3’
XE0-5’: 5’-GCCGAATTCCCTTCTGCTTAAAGTTACG-3’
XE0.6-5’: 5’-GCCGAATTCACAAGTCTGGGCAGTAGTC-3’
XE1.2-5’: 5’-CACGAATTCACCTGCACCTCCACTCCACG-3’
XE0.6-3’: 5’-CGAGAATTCATGACTACTGCCCAGACTG-3’
XE1.2-3’: 5’-GCTGAATTCCAGTGCTGAGTGGAAGC-3’
XE-1.7-3’: 5’-CCGGAATTCAGTAAGCGTCGCAATCCTC-3’

2-2.2 GAL4 driver constructs for dAP-2 mutant rescue

GAL4 construct was used for overexpression and rescue experiments. A promoter-less pGATB vector (Brand and Perrimon, 1993) was modified to add a EcoRI-SmalI-BgIII-BamHI linker after KpnI of the multiple cloning site. The new vector, pGATEB, was used to add exon 1a promoter and desired enhancer fragment such as B6 and XE. The enhancer-promoter-GAL4 fragment was then digested by KpnI/SpeI or KpnI/XbaI and ligated with KpnI/SpeI digested pPelican for transformation.

B12-GAL4 driver was generated by fusing ~12kb BamHI genomic fragment around exon1a, ~150bp flanking sequence of exon2 splicing acceptor and gal4 ORF together. Exon2 splicing acceptor was first fused with pGATEB by BamHI/SphI digested PCR fragment using 5’-CTCGGATCCTTCCCACACACTTATCC-3’ and 5’-GTGCATGCTTCTGCTGAGGCGTCGCTTGCTGGGAAGAC-3’. Then B6 was introduced into BgIII/BamHI sites of pGATEB-Exon2, leaving single BamHI site at 3’-end of B6. KpnI/SpeI digested pPelican was ligated to similarly digested pGATEB-B6-Exon2 to provide the P-element backbone. Next, 3kb BamHI fragment within E6 was
amplified to generate BamHI/BglII sites to insert into single BamHI site of the vector. Finally, B4 fragment covering exon1a was ligated into the unique BamHI site left from last step. Thus, B12 contains genomic fragment from 5’-end of B6 to 5kb into E6, generating fused GAL4 protein with first 22 amino acid of dAP-2.

2-2.3 Mutagenesis of Su(H) binding sites and dAP-2 binding sites

Mutagenesis of Su(H) binding sites, BXEmS1 or BXEmS2, was carried out using mSbx and mSxe primers as described in EMSA. Mutagenesis of dAP-2 binding sites, BXEmA1 or BXEmA1, was carried out using the following primers:

mAbx(m1300)-5’: 5’-TATCTT<ctgcagt>ATTGGCGTGTGCACCATTCG-3’
mAbx(m1300)-3’: 5’-CCAATA<ctgcagt>AAGATATGATGGATGAGCC-3’
mAxe(m2500)-5’: 5’-GAGTTT<ctgcagt>GATCGCATTTTAAGACGCAT-3’
mAxe(m2500)-3’: 5’-GCGATC<ctgcagt>AAACTCGTGAATGCCGCAGG-3’

The mutagenesis was done following the protocol of Quick-change site-directed mutagenesis kit (Stratagene) or Mutagenesis by Overlap Extension (MOE). When MOE was used, 5’- or 3’- mutagenesis primer was paired with 3’- or 5’- primer of BX or XE fragment for the first-round PCR. Two first-round PCR products were purified and mixed at 1:1 molar ratio as templates for the second-round MOE PCR as following: 95°C x 2 min; 3 cycles 95°C x 30 sec, 37°C x 2 min, 72°C x 2 min; 13 cycles 95°C x 30 sec, 55°C x 1min, 72°C x 2 min; 72°C x 2 min. The final PCR product was cloned into pGEM-T easy vector (Promega) for sequencing and further cloning.

2-2.4 Rapid small scale isolation of Drosophila genomic DNA
Flies were collected in an eppendorf tube and left at –80°C for 5 min. 100μl Solution A (0.1M Tris.HCl pH9.0, 0.1M EDTA, 1% SDS) was added for 1-5 flies (200μl for 6-10 flies, 400μl for 50 flies). Flies were homogenized for more than 1 min on ice. The eppendorf tube was then incubated for 30 min at 70°C, followed by adding 14μl of KAc (8M) per 100μl of Solution A. The tube then was incubated for another 30 min on ice followed by spinning 15 min at RT. Then the supernatant was transferred to a new tube, without taking pellet at the bottom or lipid on the top. ½ volume of isopropanol was added and the tube was spun 5 min at RT. The supernatant was discarded and the pellet was washed with 70% ethanol. The pellet was dried 2 min using Speed-Vac and redissolved it in 5μl 1/10 TE (1xTE: 10mM Tris.HCl pH7.5, 1mM EDTA pH8.0) per fly.

2-2.5 Alkaline lysis plasmid minpreps

1.5 ml of overnight culture was spun in eppendorf tube for 15 sec at RT. Media was the discarded and the spin was repeat once. The rest of media was discarded and 100μl Solution I (2.25g glucose, 6.25 ml 1M Tris.HCl pH8.0, 5μl 0.5M EDTA pH8.0, 244ml ddH₂O) plus 1μl RNase A was added to the tube. Then the tube was vertexed vigorously until pellet completely dissolved. 200μl Solution II (2ml 10% SDS, 400μl 10M NaOH, 17.6 ml ddH₂O) was added to the tube followed by inverting gently to mix. 150μl Solution III (73.6g KOAc, 28.6 ml glacial acetic acid, ddH₂O to 250 ml) was then added and the tube again was inverted gently to mix. The sample was then incubated on ice for 5 min and spun for 10 min at RT. The supernatant was transferred to a clean tube and was added with 2 volumes of 100% ethanol. The mixed sample then was incubated at –20°C for 30 min, followed by spinning down for 15 min at 4°C and the pellet was
washed with 70% ethanol. The pellet was finally dried by Speed-Vac and dissolved with desired amount of TE or water.

2-2.6 RNA preparation and RT-PCR

Embryos from stage 0-17 (0-21 hours after egg laying), leg discs and CNS dissected from 3rd instar wandering larvae were used for RNA preparation. Resuspend 10mg of tissue (50μl embryo, or discs/CNS from 20 larvae) in 100μl Trizol and homogenize the tissue with plastic pestle or pipette. Add 20μl of chloroform and shake vigorously by hand. Stand for 5 min and spin for 15 min at 4°C. Transfer the aqueous phase into a new tube (about 50μl). Add equal volume of PCI (phenol: chloroform: isopropanol=25:24:1), mix well and spin 10 min at RT. Transfer the upper phase into a new tube and add equal volume of isopropanol and mix well. Stand 10 min at RT and spin 10 min at 4°C. Wash with 70% ethanol and speed-vac dry. Dissolve in 30-50μl DEPC-treated water.

1 μg of total RNA and oligo(dT)18 primer were used in 20μl RT reaction (42°C x 50 min) using SuperScript II RNase H Reverse Transcriptase kit (Invitrogen) as described in its manual. Positive control GAPDH was used to normalize RT-PCR condition. Then cDNA corresponding to 20ng total RNA from RT product was used for semi-quantitative RT-PCR. (28 cycles of PCR with 94°C x 30 sec, 56°C x 30 sec, and 72°C x 1 min). GAPDH primers (5’- CAGCCATCACAGTGATTCC-3’ and 5’- TTCCCGTGAGCTTACCCTTG-3’), exon 1a 5’-primer (5’-AGTTGCCTTAAACCAGTGCC-3’), exon 1b 5’-primer
ACTTGTGATTGTGATCCGCC-3’) and dAP-2 3’-primer spanning intron 2 (5’-GACGGCTCCATGGTTCTGTTGCTG-3’) were used.

1-2.7 Embryo injection and identification of transformants

The transgenic vectors carrying promoters and enhancers were purified with Qiagen Midiprep Kit (Qiagen) and dissolved into ddH2O to 300-400ng/μl. DNA has its best quality when freshly made. About 200-400 pairs of Δ2-3,Sb (99B)/TM6,Ubx flies were put into a cage under 25°C several days before injection and provided with fresh grape juice plate plus yeast paste every day. In the morning of injection day, food was changed every 45min to make fly accommodate the injection cycle. The injection was carried out under 18oC to slow down the development of embryo. Needles were made by needle-puller pulling BOROSILICATE with filament (Sutter instrument co. catalog #: BF100-78-10), then were back-loaded with 1-2μl DNA solution by SGE 10μl microsyringe. Embryos collected every 45 min were washed by water and then dechorionated by 50% bleach for 1 min. The well-dechorionated embryos will float when washed by water. The dechorionated embryos then were lined up in a column on the agar stage with same orientation. To transfer embryos onto a coverslip, a glue made by dissolving double-sided tape into heptane was drawn a line in the middle of the coverslip. The embryos would stick to the glue when they were gently touched by the coverslip. The embryos on the coverslip were desiccated in the warm desiccant chamber for 1 min and then covered by halocarbon oil (HC-700:HC-27=4:1) ready for injection. The back-loaded needle needs to be broken using the edge of a coverslip protected by oil. Slowly move the needle towards the edge of the coverslip while pumping the solution inside the
needle until air bubble/liquid droplet coming out freely. Put embryos on the microscope stage with posterior end facing the needle mounted on the micromanipulator attached on the microscope. To penetrate the vitelline membrane, slowly move the embryo towards the needle. After needle enters the embryo, move it close to the posterior end where the pole cells will form, pump a little DNA out and quickly withdraw the needle. When the injection is done, store the embryo inside a humid grape juice plate to let embryo develop for 2 days.

The hatched larvae were collected and later become G0 flies. Each G0 fly will be crossed with 3 yw flies. Progeny (G1) flies with red-eye marker were positive transformants carry p[w+⁷] transgene. Single male red-eye G1 fly then was crossed with Adv/CyO to identify if the insertion was on X chromosome. If yes, crossing with first chromosome balancer FM7C is needed. Otherwise, crossing between red-eye CyO flies are needed. If the insertion is on second chromosome, the line will be established. Otherwise, additional crossing with third chromosome balancer TM3,Sb/TM6B is needed. Finally the transgenic allele will be either in homozygous or balanced with a balancer chromosome.

2-2.8 Fly stocks

Flies were grown on standard medium and crosses were performed at 25°C unless indicated otherwise. bib (big brain)-lacZ (de Celis et al., 1998), hth-lacZ, andDllem212-Gal4,UAS-GFP (Courtesy of Gerard Campbell) were used as molecular marker for leg imaginal disc in immunostaining. Su(H)D47 null allele (Morel and Schweisguth, 2000), w hsFLP1.22, P[arm-lacZ] M(2)z FRT40A/CyO (Courtesy of Antonio Garcia-Bellido) and
w; P[ubi-GFP] FRT40/CyO (www.flybase.org) were used for mosaic analysis. UAS-N\textsuperscript{intra} (Struhl and Adachi, 1998) were crossed with p339-Gal4 (Halder et al., 1995) driver to direct ectopic Notch expression in leg imaginal discs. yw was used as dAP-2 wild type fly. hsFLP1.22; FRT42 P[arm-lacZ] M(2)60E/CyO (Courtesy of Gerald Campbell) and FRT42 Dll\textsuperscript{SAl}/CyO (Courtesy of Stephen Cohen) were used to generate Dll mutant clone. UAS-bowl1.1 is a gift from Sarah Bray lab. dpp-Gal4 UAS-GFP/TM3, Sb is from Antonio Garcia-Bellido lab, UAS-N\textsuperscript{RNAi} is from flybase. Eutaeg Yeo made UAS-dAP2a4 rescue line using dAP-2 1a cDNA driven by UAS enhancer. dAP-2 null alleles dAP-2\textsuperscript{2}, dAP-2\textsuperscript{15}, dAP-2\textsuperscript{13} and deficiency line Df(3L)1118 were used to generate dAP-2 null background for autoregulation (Monge et al., 2001). UAS-lacZ was obtained from flybase.

2-2.9 Genetic mosaic (clonal) analysis

Su(H) null mutant clones was generated in background of wild type dAP-2 or enhancer-driven lacZ genes using FLP/FRT system (Xu and Rubin, 1993). Clones in w hsFLP1.22; P[arm-lacZ] M(2)z FRT40A/FRT40A Su(H)\textsuperscript{d47} were induced at 48-72hr AEL (after egg laying) by heat shock at 33°C for 1 hour. Clones in w hsFLP1.22; P[ubi-GFP] FRT40A/FRT40 Su(H)\textsuperscript{d47}; B6-lacZ/+ , w hsFLP1.22; P[ubi-GFP] FRT40A/FRT40 Su(H)\textsuperscript{d47}; BXE-alacZ+ and w hsFLP1.22; P[arm-lacZ] M(2)z FRT40A/FRT40A Su(H)\textsuperscript{d47}; EB\textsuperscript{1.7R}-GFP/+ were induced at 36-60hr AEL by heat shock at 38°C for 1hr. Dll\textsuperscript{SAl} null mutant clones were induced by 33°C heat-shock at 48-72hr AEL or 84-108hr AEL in hsFLP1.22; FRT42 P[arm-lacZ] M(2)60E/FRT42 Dll\textsuperscript{SAl} larvae.

2-2.10 β-gal staining and light microscopy
Tissues dissected from 3\textsuperscript{rd} instar wandering larvae or prepupae were fixed with 4% formaldehyde in PBST (PBS, 0.1% Triton X-100) for 5-10 minutes, washed with PBST for 15 minutes, then stained overnight with staining buffer (PBS, 1mM MgCl\textsubscript{2}, 0.3% Triton X-100, 3mM K\textsubscript{a}[Fe(II)(CN)\textsubscript{6}]\textsubscript{3}\textsubscript{H}\textsubscript{2}O, 3mM K\textsubscript{3}[Fe(III)(CN)\textsubscript{6}], 1mg/ml X-gal). Nikon E300 microscope, Optronics digital camera and MagnaFire SP software were used to capture images.

\textbf{2-2.11 Antibodies and immunostaining}

Rabbit anti-dAP-2 antibody was used to detect endogenous dAP-2 expression by immunostaining of whole-mount larval tissues (Monge et al., 2001). Monoclonal anti-N\textsuperscript{intra} (C17.9C6), anti-Dac, and anti-Antp were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biological Sciences, Iowa City, IA 52242. Mouse anti-β-galactosidase (Promega), rabbit anti-β-galactosidase (gift from Zhi-Chun Lai) and mouse anti-GFP (BD Biosciences) were used to visualize the reporter gene expression. Rat anti-bab2 was a gift from Frank A. Laski (Couderc et al., 2002). Cy3-conjugated anti-rabbit, Cy3-conjugated anti-rat (Jackson ImmunoResearch Laboratories), Alex488-conjugated anti-rabbit and Alex488-conjugated anti-mouse (Invitrogen) were used as secondary antibodies. Immunostaining was carried out on dissected imaginal discs from 3\textsuperscript{rd} instar wandering larvae or prepupae. Tissues were first fixed with 4% formaldehyde in PEM (0.1M PIPES, 2mM EGTA, and 1mM MgSO\textsubscript{4}) buffer for 30 minutes, then incubated with primary antibody in TBST (50mM Tris.HCl pH7.4, 150mM NaCl, 0.1% Triton X-100, 2% NGS) at 4\textdegree C overnight. On the second day, tissues were washed with
TBST for 20 min x 3. After secondary antibody incubation in TBST for 2hr and washing for 20 min x 3, imaginal discs are mounted with 90% glycerol and analyzed with Olympus FluoView 300 confocal microscope and Fluoview software.

2-2.12 Electrophoresis mobility shift assay (EMSA)

pGEX-Su(H) (courtesy of James W. Posakony) was used to make bacteria-translated Su(H)-GST fusion protein as described (Singson et al., 1994). Double-stranded oligonucleotides probes sequences are as follows:

E(Spl)m4S1 (positive control)

5’- CACAAAAACACCGAGTGTGGGAAACTACAAGGATA -3’
3’- GTTTTGTGGCTCACACCCTTTGATTTCTATGC -5’

BXES1 (anti-sense strand)

5’- AAAACGGATTTTCTATGTGAGAAAATTATCAGTGC –3’
3’- TTGCCTAAAGATACACTCTTTTAATAGTCACGCA –5’

BXES2 (anti-sense strand)

5’- AATTAGATTGTAGCATGGGAAATAATGTTGAAAG –3’
3’- AATCTAACATCGTACCCTTTAATAGTCACGCA –5’

BXEmS1 (anti-sense strand)

5’- AAAACGGATTTTCTATG	ctcGAAAATTATCAGTGC –3’
3’- TTGCCTAAAGATAC
gagCTTTTAATAGTCACGCA –5’

BXEmS2 (anti-sense strand)

5’- AATTAGATTGTAGCAT	ctcAAATAATGTTGAAAG –3’
3’- AATCTAACATCGT
gagTTATTACAACCTTTCCA –5’
Annealed double-stranded oligos were purified by PAGE and labeled by T4 polynucleotide kinase and \([\gamma^32P]\) ATP, then purified by G25-Sephadex. 400ng GST-Su(H) protein (20-fold as much as the probe in molar ratio) and 200 fM radiolabeled probes were mixed in binding buffer (25 mM K\(^+\) HEPES, pH 7.5, 50 mM KCl, 5mM MgCl\(_2\), 0.1mM EDTA, 1mM DTT, 1mg/ml poly[d(I, C)] ) and incubated at room temperature for 30 min. Concentration of the unlabeled competitors were 50 times as those of labeled probes. Electrophoresis was performed with 4% polyacrylamide / 0.5 x TBE gel, running 2.5 hours at 4°C at 200V.

2-2.13 DIG-labeling of RNA probes

Riboprobes representing \(dAP-2\) exon 1a antisense strand (481bp) and exon 1b antisense strand (260bp) were prepared using DIG-RNA Labeling kit (Roche Diagnostics). The cDNA of each exon was cloned into pGEM T-easy vector in the orientation that antisense strand is on the same strand of T7 primer. The vector was then linearized by SalI near SP6 primer site on the pGEM T-easy. 1\(\mu\)g of purified linearized DNA, 2\(\mu\)l of 10x DIG DNA Labeling Mix (10mM each of ATP, CTP, and GTP; 6.5mM UTP, 3.5mM DIG-UTP), 2\(\mu\)l 10X concentrated transcription buffer, 2\(\mu\)l T7 RNA polymerase, and add ddH\(_2\)O to 20\(\mu\)l. Incubate the reaction 2 hr at 37°C. Add 0.8\(\mu\)l 0.5M EDTA (ph8.0) to stop the reaction. Add 2.5\(\mu\)l 4M LiCl and 75\(\mu\)l cold ethanol. Let the precipitant form for more than 30 min at –80°C. Spin 15 min at 4°C and wash the pellet with 70% ethanol. Dissolve the pellet with 50\(\mu\)l DEPC-treated water. Check OD with spectrophotometer and run a small aliquot on agarose gel.
2-2.14 RNA in situ hybridization

The in situ hybridization protocol was applied to dissected and fixed yw or dAP-2\textsuperscript{2}/dAP-2\textsuperscript{2} leg imaginal discs and embryos (Tautz and Pfeifle, 1989). Embryos were washed by 50% bleach to dechorionate for 1 min. After washing off residue bleach, add 1.5ml fixative (0.1M Hepes pH6.9, 2mN MgSO\textsubscript{4}, 1mM EGTA), 0.5ml 16% formaldehyde, and 8ml heptane, rotate 20 min. Embryo should stay at the interphase, remove lower phase (fixative) and some heptane on top. Add 10ml methanol and shake 30 sec to devitellinize. The embryos should sink to the bottom if well devitellinized, replace haptane and add methanol to repeat if needed. The embryos can be kept in methanol for months at refrigerator. Wash the embryos 5 min x 3 in PTw (1x PBS, 0.1% tween 20), postfix in PTw plus 4% formaldehyde for 15 min, and wash in PTw 5 min x 3 again. The embryos now are ready for hybridization. If the dissected imaginal discs are used, leave the discs attached on the cuticle and fix it directly in PBS, 0.5% Triton X-100 plus 4% formaldehyde on ice for 15 min and at RT for 20 min. Then wash in PTw 5 min x 3 and ready for hybridization. Embryos and discs are first prehybridize with 1:1 PTw : pre-hyb buffer (50% formamide, 5x SSC, 0.1% Tween 20) for 10 min and hybridization buffer (pre-hyb + 100μg/ml sonicated salmon sperm DNA, 50μg/ml heparin) for 1 hr at 60°C. If needed, tissues can be stored in hybridization buffer at -20°C for a few days. If proceed, dilute the probe in hybridization buffer to 0.5ng/μl and heat at 80°C for 10, cool on ice for 5 min. Remove old hybridization buffer and add 100-200μl probe, hybridize at 60°C overnight. The next day, probe is taken out and can be reused for several times. Wash with serial dilution (5:0, 4:1, 3:2, 2:3, 1:4, 0:5) of hybridization:PTw for 10 min each. The tissue is then blocked by PTw plus 1% NGS for 10 min x 2, incubate with AP-
conjugated anti-DIG-Fab fragment antibody at 1:2000 in PTw for 1hr. Wash with PTw for 20 min x 3 and Staining buffer (100mM NaCl, 50mM MgCl₂, 100mM Tris.HCl pH9.5, 0.1% Tween 20) 5 min x 3. Finally, the signal is detected by 1ml staining buffer plus 4.5μl NBT and 3.5μl BCIP substrate (Roche Diagnostics) for 30 min to overnight. The reaction can be stopped by cold PTw and tissue is mounted in 90% glycerol.

2-2.15 Mounting adult fly legs

Legs from newly eclosed adult flies were plucked off at the coxa level. Immerse them into 70% ethanol for at least 1 min, then transfer to 95% ethanol x 1 min, 100% ethanol x 1 min twice, xylene x 1min twice. When the tissues inside leg is cleared out, transfer leg with some xylene into a drop of Permount on slide. Position the leg well and cover with coverslip.
2-3 Results

2-3.1 Expression of dAP-2 in *Drosophila* legs

*Drosophila* adult legs develop from their primordia, 6 leg imaginal discs representing 3 pairs of legs, in larval stage. The development of leg discs requires progressively expressed leg “gap genes”, *Dll, dac* and *hth* along proximodistal axis. The formation of leg joint and interjoint growth require receptors and ligands of Notch signaling expressed in the distal edge cells (also called proximal joint cells) and their proximally adjacent cells of each leg segment. These genes’ temporal and spatial expressions in the evert ing leg discs serve as good molecular markers to identify dAP-2 expression pattern to the details of subdomain within leg-segment. dAP-2 expression has been detected at the proximal boundary of the expression domain of *Dll* at the mid-third instar larval stage (Kerber et al., 2001). By the late-third instar stage, dAP-2 transcripts and protein are expressed in concentric rings that signify a segmental pattern along the proximodistal axis of the leg (Fig. 2-1). The four distal-most rings correspond to presumptive tarsal/tarsal joints based on co-expression of dAP-2 and Notch target *E(Spl)* and *bib* in these locations (Kerber et al., 2001). In order to better define locations of dAP-2 expressing cells in proximal regions of the developing leg, I carried out double-staining analysis with antibodies against dAP-2 and other markers with known proximal leg expression domains in the third instar larvae and prepupae. A *Dll-GAL4* driven *UAS-GFP* reporter transgene was used to identify proximal and distal domains of *Dll* expression in leg imaginal discs. The distal (main) *Dll* domain extends from the pretarsus to the middle
of the presumptive tibia, while the proximal Dll domain extends from the distal trochanter to the proximal femur (reviewed by Kojima, 2004; Wu and Cohen, 1999).

As a consequence of growth of the leg disc in late third instar larva, distal and proximal regions of the leg disc are often not visible in the same focal plane. The five distal-most rings of dAP-2 expressing cells (presumptive tarsal/tarsal and tibia/tarsal boundaries) are all contained within the distal Dll expression domain (Fig. 2-2B-B’’). Moving proximally, the next ring of dAP-2 expression (arrows in Fig. 2-2) is located between the distal and proximal Dll domains, and corresponds approximately to the presumptive femur/tibia boundary. In the proximal portion of the leg disc, three rings of dAP-2 expression are evident. The inner-most of these coincides with the distal part of the proximal Dll domain, and thus corresponds to the presumptive trochanter/femur boundary (Fig. 2-2A-A’’). A bib-lacZ transgene was used to assess co-expression of dAP-2 and bib, a Notch target gene activated at the distal end of every presumptive leg segment (de Celis et al., 1998). Notably, expression of bib and dAP-2 overlapped in the presumptive tarsal joints but not in more proximal leg regions (Fig. 2-3). With the exception of the proximal-most dAP-2 ring, the proximal dAP-2 rings sit just distally to the bib-lacZ rings. Combined with results from Dll domain, endogenous dAP-2 appears to be expressed in the proximal domain of proximal and intermediate leg segments including coxa, trochanter, femur and tibia (Fig. 2-3A-A’’). In the distal leg of dAP-2 expressing domains, bib-lacZ well overlaps with dAP-2 in four tarsal joints and partially overlaps with dAP-2 tibia/tarsal1 domain (Fig. 2-3B-B’’). Overall, the endogenous dAP-2 expression pattern consists of 9 concentric rings during 3rd instar larva /prepupa stage. Proximal and intermediate rings with several rows of dAP-2 expressing cells are located
at the proximal end of those segments. Tarsal rings with 1-2 rows of cells are located at the distal end of tarsal segments (Fig. 2-4). This differential expression patterns between proximal and distal leg joints indicate dAP-2 might be associated with different transcription factor or signaling pathway that have distinct functional domains.
**Fig. 2-1 Expression of wild-type dAP-2 in leg imaginal disc.** (A) Composite confocal image along Z-axis. dAP-2 expression in late third instar larval stage detected by antibody against C-terminal 11 amino acid peptide shows multiple concentric rings. (B) By prepupa (white pupa) stage, 9 concentric rings can be clearly counted in the evverting leg disc. Tarsal rings (stripes) are best viewed in this stage. (C) Wild-type adult male leg consists of 10 leg segments, coxa (co), trochanter (tr), femur (fe), tibia (ti), tarsal1 (t1) to tarsal5 (t5) and pretarsus (pt/pr).
Fig. 2-2 Comparison of Distalless and dAP-2 expression domains. (A-A’’) Dll proximal domain overlaps with the third proximal dAP-2 ring and extends more proximally. (B-B’’) Dll distal (main) domain covers five distal dAP-2 rings. Only one dAP-2 ring (arrows in A and B) at femur/tibia joint is located between Dll proximal and distal domain.
Fig. 2-3 Comparison of *bib-lacZ* and *dAP-2* expression domains. (A-A’’) Bib-lacZ expression is located proximal to dAP-2 rings in the proximal leg segments focal plane, indicating dAP-2 is expressed at the proximal ends of proximal leg segments. (B-B’’) dAP-2 expression in prepupa stage expands to multiple rows of cells. The proximal tibia dAP-2 expression domain (arrows in A and B), distal to bib-lacZ expression domain, is occupying most of tibia segment. Tarsal dAP-2 expression mostly overlaps with bib-lacZ tarsal domain, while extends a few rows of cells distally.
Fig. 2-4 Summary of dAP-2 expression in Drosophila leg. dAP-2 is expressed in the proximal end of coxa, trochanter, femur, tibia and tarsal 1 segments. The proximal tibia domain (arrow) is the only dAP-2 expression domain located between the proximal and distal (main) Dll expression domains. The bib expression domain located at the distal end of each leg segment only overlaps with dAP-2 domains in distal tarsal 1 to tarsal 4. dAP-2 is not expressed at tarsal5/pretarsus boundary.
2-3.2 Expression of dAP-2 in Drosophila antenna

The *Drosophila* antenna and leg are considered to be homologous structures because they can be interconverted through the alteration of homeotic genes. For example, ectopic expression of Antennapedia (Antp) or other homeotic genes can transform antenna to leg. By contrast, the antenna appears to develop without any input from homeotic genes (Yao et al., 1999). The identity of antenna requires its own selector genes, such as *spineless (ss)*, *distal antenna (dan)* and *distal antenna related (danr)*.

The morphological and developmental similarities between antenna and leg partially attributed to the utilization of similar sets of developmental genes. In the primordia of both leg and antenna, gradients of secreted factors encoded by *dpp* and *wg* regulate formation of PD axis. *dpp* and *wg* are activated similarly by Hedgehog (Hh), possess similar relative expression patterns and exhibit similar mutual antagonism in both appendage primordia (Dong et al., 2001). Hth, Dac and Dll are expressed in proximal, intermediate, and distal segments of both leg and antenna. However, their expression level and function have crucial difference between two appendages. Hth and Dll expression overlap extensively in the antenna. This overlap has been linked to antenna identity based on the observation that co-expression of Hth and Dll can induce formation of antennal structures in the proximal regions of wing and leg discs (Casares and Mann, 2000; Dong et al., 2000). Dac expression levels have also been indicated to differentiate the functional medial segment in leg from the counterpart in antenna (Dong et al., 2001).

To examine if function of dAP-2 is also conserved between two homologous structures. I carried out double staining of dAP-2 with molecular markers present in both appendage primordia. Like in leg disc, wild-type dAP-2 is also expressed in multiple
concentric ring pattern, which everts out along proximodistal axis from the beginning of prepual stage. Two proximal rings, instead of three, are seen on the proximal antenna focal plane, and another six distal rings are observed on the distal focal plane together with part of most proximal ring. Dll has single expression domain that covers the second proximal dAP-2 ring to the distal tip of antenna disc (Fig. 2-5A-B). Dac overlaps with the medial four dAP-2 rings, leaving the most distal two rings and the most proximal two rings (Fig. 2-5D). The expression of Hth extends from the proximal end of antenna disc to the third distal dAP-2 antenna ring (Fig. 2-5C). The expression of Notch target bib, which is also in concentric ring pattern, overlaps with dAP-2 rings well in the distal antenna segments, while moves proximally to the dAP-2 proximal domains (Fig. 2-5E). This differential expression pattern strongly resembles its counterpart in leg disc, indicating conserved functions between two homologous structures.

Based on the homologous segments determined by transformation phenotypes, the antenna a1 segment corresponds to leg coxa; the antenna a2 corresponds to leg trochanter; the antenna a3 corresponds to leg femur, tibia and tarsal 1; the a4 and a5 collectively correspond to leg tarsal 2 to 4; and the arista of antenna corresponds to leg tarsal 5 and pretarsus. Although leg gap genes have different patterns between leg and antenna, dAP-2 gene shows conserved multiple repeated expression domains along PD axis (Fig. 2-6). Even a3 of antenna is not divided into multiple segments morphologically, dAP-2 still preserves segmented expression domains. Like dAP-2, bib-Z expression in antenna disc resembles its counterpart in leg disc. These observation indicated dAP-2 might be regulated differently by its upstream regulatory genes in the antenna disc comparing that in the leg disc, but it may preserve evolutionarily fundamental functions, such as
promoting cell survival/growth that both Notch signaling and dAP-2 have been implicated in the leg development. The conserved functions of dAP-2 also indicated dAP-2 does not serve as a selector gene for appendage identity.
Fig. 2-5 Expression of wild-type dAP-2 in antenna disc. dAP-2 expression (red) in antenna disc is compared to other molecular markers (green): proximal (A-A’’) and distal (B-B’’) Dll domains, Hth (C-C’’) domain, Dac (D-D’’) domain, bib (E-E’’) domains.
Fig. 2-6 Summary of *dAP-2* expression in *Drosophila* antenna. (A) Schematics of adult *Drosophila* antenna (top) and leg (bottom). Arrows indicate homologous domains (Dong et al., 2001; Postlethwait and Schneiderman, 1971). (B) *dAP-2* expression domains comparison between leg and antenna.
2-3.3 Notch signaling-dependent regulation of \( dAP-2 \) along PD axis of limb

Notch signaling has been shown to regulate \( dAP-2 \) expression in \textit{Drosophila} leg joints (Kerber et al., 2001). However, previous studies had not addressed if this regulation is applicable to every \( dAP-2 \) expression domain. Since I found that \( dAP-2 \) is expressed in different sub-domains of leg segment along proximal/distal axis and co-localizes with Notch target \( bib \) only in tarsal joints, I then examined if \( dAP-2 \) responds to Notch signaling in the same fashion for every leg segment.

I used a genetic strategy similar to that of Kerber et al. (2001) to stimulate or block Notch signaling in different leg region. \textit{p339-Gal4} driver was used to drive ectopic \( N^{\text{intra}} \), a constitutively active form of Notch, in a broad proximal area in the leg disc. This failed to induce ectopic \( dAP-2 \) expression (Fig. 2-7A-A’’). Heat-shock induced FRT/FLP-mediated recombination was used to generate \( Su(H)^{-} \) mutant clones during early larval development. Surprisingly, I found \( Su(H)^{-} \) clones in intermediate and proximal leg segments did not abolish \( dAP-2 \) expression while clones in tarsal segments did (latter confirms findings of Kerber et al, 2001). Notably, these experiments showed that Notch-independent expression of \( dAP-2 \) extends to the tibia/tarsal 1 joint. Thus, of 9 PD-axis domains of \( dAP-2 \) expression in the leg disc, only the 4 distal-most domains are regulated by Notch signaling (Fig. 2-7B-B’’). Although \( Su(H) \) is the major transducer mediating almost all Notch signaling activity, it also has Notch-independent functions, such as promoting late development of sensory organ precursors (SOP) (Barolo et al., 2000; Koelzer and Klein, 2003) To confirm that loss of \( dAP-2 \) expression in \( Su(H) \) null clones is due to loss of Notch signaling, I used RNA interference (RNAi) technology to down-regulate Notch expression (Presente et al., 2002). Dpp-GAL4 drove expression of
Notch RNAi along PD axis anterior to the AP boundary during leg development. By third instar larval stage, dAP-2 expression was also seen downregulated in the tarsal segments while unaffected in the intermediate and proximal joints (Fig. 2-7C-C’’).

I also examined Su(H)-/- clones in the antenna disc where dAP-2 is expressed in multiple concentric rings. Interestingly, I noticed that in the antenna disc, as in the leg disc, dAP-2 expression in proximal rings did not require Su(H) while expression in distal rings did (Fig.2-8A-A’’). Similarly, down-regulation of Notch signaling by Dpp-GAL4 driven Notch RNAi also affected distal antennal rings of dAP-2 expression but had no effect on proximal rings (Fig. 2-8B-B’’). These data showed that regulation of dAP-2 by Notch signaling was conserved between these homologous appendages and was specifically associated with dAP-2 expression in distal versus proximal regions of the developing appendage.
Fig. 2-7 Notch dependent regulation of dAP-2 expression in leg disc is restricted to distal parts (presumptive tarsal joints). (A-A’’) Overexpression of Notch intracellular domain (Notchintra) in proximal regions of leg disc using the p339-GAL4 driver does not affect dAP-2 expression (arrows). (B-B’’) Su(H)^{-/-} clones affect dAP-2 expression only in distal leg segments (arrows) but not intermediate and proximal ones (arrowheads). (C-C’’) Notch loss-of-function using Dpp-GAL4 driven UAS-NotchRNAi,UAS-GFP shows loss of dAP-2 expression in tarsal ring (arrow) but not in tibia/tarsal joint (arrowhead).
Fig. 2-8 Notch-dependent regulation of dAP-2 expression in antenna disc is restricted to distal parts. (A-A’’) dAP-2 expression in the distal antenna, presumptive a3/a4 boundary (arrow) is affected by loss of Su(H), whereas more proximal dAP-2 rings are not affected (arrowheads). (B-B’’) Notch RNAi expression, inferred by Dpp-GAL4 driven GFP, down-regulates dAP-2 expression in distal (arrow) but not proximal (arrowhead) antenna.
2-3.4 Identification of cis-elements for expression in proximal and distal leg regions using a tau lacZ reporter with hsp70 promoter

The differential response of dAP-2 leg domains to Notch signaling prompted the question that how dAP-2 expression is regulated in Drosophila appendages. The identification and analysis of cis-regulatory elements using P-element mediated transgenic flies become one of my major approaches. At first, I used tauZ P-element vector with heat-shock inducible promoter for its ability to integrate the activity of a heterologous enhancer and potential to visualize the axons of targeted neurons in the central nervous system where dAP-2 is also expressed. About 20kb genomic DNA region surrounding dAP-2 gene, which has two alternative exon 1s and another 7 exons, was dissected by restriction enzyme digestion or PCR amplification (Fig. 2-9A). B6d, a promoter-less 5.3kb fragment located upstream of exon 1a, shows strong distal leg activity. E6 fragment downstream of exon 1a drove strong expression in proximal leg segments. B4 fragment partially overlaps with E6 and shows weak proximal leg activity. BC fragment upstream exon 1b only shows activity in a cluster of femur cells, possibly belonging to femur chordotonal organ. SE downstream of exon 1b has no leg activity (Fig. 2-9B-F). Double immunofluorescent staining showed that the distal activity of B6d co-localizes with endogenous dAP-2 tarsal pattern, and proximal activity of E6 overlaps with proximal coxa and proximal femur dAP-2 rings (Fig. 2-9G-H). This first-step dissection revealed multiple dAP-2 expression domains are separable and discrete cis-regulatory elements are responsible for leg segment-specific activities.

Then I combined proximal and distal activities from two cis-elements into B12 fragment and tested if it is enough to represent the wild-type dAP-2 activity. (Fig. 2-9A).
The β-gal staining of UAS-lacZ driven by B12-GAL4 shows a multiple-ring pattern like that of endogenous dAP-2 (Fig. 2-10A). Immunostaining showed a perfect co-localization between dAP-2 and β-galactosidase in proximal and intermediate leg segments, while a slightly ectopic activity of B12 (Fig. 2-10B-C). The B12-GAL4 was then used to drive dAP-1a cDNA (UAS-dAP2a4) to rescue the dAP-2 null mutant. Without driver, UAS-dAP2a4; dAP-2^{23}/dAP-2^{15} null mutant has severely shorten leg, and adult fly rarely ecloses. The leg defects, loss of joint and leg shortening, both can be successfully rescued by B12-GAL4 driven UAS-dAP2a4 in the background of dAP-2^{23}/dAP-2^{15}. Most rescued flies can climb out of pupa case, although being able to support their bodies. They were also trying to move their legs, although seemed uncoordinated. This suggested the spatial and temporal expression of dAP2 1a cDNA rescued by B12G4 in leg is similar to the endogenous pattern (Fig. 2-10E-H). Therefore, overexpression by B12G4 did not generate leg shortening caused by misexpression. The uncoordinated movement of leg may be due to the B12 activity in the central nervous system (CNS), which only overlaps with a small part of endogenous pattern (Fig. 2-10D)
Fig. 2-9 Dissection of 20kb dAP-2 genomic region using tauZ vector. (A) Dissection scheme. (B) B6d-tauZ, (C) B4-tauZ, (D) E6-tauZ, (E) BC-tauZ, and (F) SE-tauZ leg expression shown by β-gal staining. (G-G’’) B6d overlaps with dAP-2 tarsal joint rings. (H-H’’) E6 overlaps with dAP-2 proximal coxa and proximal femur rings.
Fig. 2-10 Rescue of leg shortening and tarsal joints by B12-GAL4 driven dAP-2 exon 1a cDNA. (A) β-gal staining of B12-GAL4/UAS-lacZ. (B-D) immunostaining against dAP-2 (red) and β-galactosidase (green) shows B12-GAL4/UAS-lacZ in distal, proximal leg and CNS. (E) First leg of UAS-dAP2a4/++;dAP-2^2/+ has wild-type phenotype. (F) UAS-dAP2a4/++; dAP-2^2/dAP-2^15 has severely shortened leg and loss of joints. (G) B12-GAL4/UAS-dAP2a4; dAP-2^2/+ has normal leg. (H-H’) Leg length and joints are rescued in B12-GAL4/UAS-dAP2a4; dAP-2^2/dAP-2^15. Details in tarsal joints (arrows) in shown in (H’).
2-3.5 Prediction of candidate cis-elements in dAP-2 genomic regions using phylogenetic footprinting

As I continued the dissection and identification of cis-regulatory elements, I faced the challenge of making large number of transgenic vectors and fly lines. In order to efficiently choose genomic regions most likely containing cis-regulatory elements, I sought help from comparative genomics approaches.

With the advances in genome sequencing data and computational tools, researchers aligned human and mouse genomes and have hypothesized, based on the extent of conservation, which only about 5% of human genome is under the purifying selection since the primate-rodent separation. Beside 1.5% of coding exons and 1% untranslated regions, about 2.5% of human genome fulfills other functions such as regulating gene expression. The cis-regulatory modules (CRMs) containing clusters of transcription factor binding sites, therefore, are most likely located within evolutionarily conserved genomic region under the selective pressure.

Since the initial release of Drosophila pseudoobscura genome sequence data in 2003, genome sequences for 6 other Drosophila species, 1 mosquito (Anopheles gambiae) and 1 honeybee (Apis mellifera) have been released and are available for alignment. Choosing species that have reasonable evolutionary distance for comparison is the first important step. The phylogenetic tree shows the Drosophila species, from the closest to the farthest evolutionary distance to Drosophila melanogaster, are D. simulans, D. yahuba, D. erecta, D. ananassae, D. pseudoobscura, D. mojavensis, and D. virilis (Fig. 2-11A). From the result of VistaBrowser (http://pipeline.lbl.gov), developed by Lawrence Berkeley National Laboratory, one can see the conservation peaks using the default
setting of 70% similarity over 100bp windows. Too many predicted conserved regions can not help us select the most important ones. Too few peaks will leave us less chance to detect the conserved \textit{cis}-regulatory elements that diverge relatively faster (Dickmeis and Muller, 2005). \textit{D. pseudoobscura} and \textit{D. melanogaster}, diverged about 25 million years ago, appear to have a reasonable evolutionary distance for a preliminary consideration of essential conserved regulatory elements (Fig. 2-11B). In addition to VistaBrowser, a variety of other alignment algorithms and visualization tools have been developed for sequences comparison. UCSC Genome browser (http://genome.ucsc.edu) combined both genome annotation and comparison. It shows not only Vista tracks based on global alignment tool SLAGAN, but also multiZ alignment based on local alignment tool BLASTZ and phastCons conservation scores. The BLASTZ alignment of genome sequences can be also visualized by Pipmaker (http://bio.cse.psu.edu) or zPicture (http://zpicture.dcode.org). While Pipmaker and zPiture can also use user-uploaded sequence files, VistaBrowser and UCSC genome browser rely on automatically updated genome assemblies and users can easily persue any gene of interest at any degree of resolution. To predict \textit{cis}-regulatory element, conservation level of dAP-2 genome sequence between \textit{D. melanogaster} and \textit{D. pseudoobscura} were examined. VistraBrowser and zPicture showed clustered conservation peaks in very similar pattern. PhastCons scores from UCSC genome browser, on the other hand, showed too many conserved peaks. Previously identified B6d fragment contains multiple conserved peaks packed within 5' -end 3.7kb BXE fragment, which was predicted to represent essential distal leg activity. The promoter of exon 1a was predicted to be located within the major peak
upstream of transcription initiation site (Fig. 2-12). Since I prefer the VistaBrowser control of visualization, it was used most frequently in my research.
Fig. 2-11 VistaBrowser alignment of dAP-2 genes of 8 Drosophila species. (A) Phylogenetic tree of Drosophila species (available genome sequences are highlighted by red color, modified from www.flybase.org). (B) 7 species’ genome sequences each pairwise aligned with D. melanogaster sequence around dAP-2 gene (chr3L: 21,515,000-21,548,000). From top to bottom is D. simulans, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura (red box), D. mojavensis, and D. virilis.
**Fig. 2-12 Comparison of genome alignment tools.** (A) UCSC genome browser with conservation scores. (B) VistaBrowser using SLAGAN. (C) zPicture using BLASTZ. All sequences are using *Drosophila melanogaster* *dAP-2* gene (chr3L: 21,515000-21,548,000) as reference sequence. BXE element (green box) and exon 1a promoter (a, orange box) are conserved between species.
2-3.6 Preferential use of dAP-2 exon 1a promoter in leg discs

With the assist from comparative genomics, I was ready to carry out further dissection to isolate smaller segment-specific enhancers. However, using tauZ vectors has a couple of disadvantages. The independent lines based on same reporter construct showed a lot of variation in expression pattern due to position effect. The heat-shock inducible promoter of hsp70 can not be used in the genetic experiments requiring heat-shock, such as mosaic analysis. Therefore, I decided to switch the backbone vector to the Pelican vectors for my transgenic analysis. These vectors use insulator sequences to flank the transgene and its regulatory sequences to reduce position effects dependent on chromosome insertion site. Various derivatives of the basic Pelican vector allow the user to use either basal hsp70 promoter or to insert a promoter of their own choice (Barolo et al., 2000).

Many transcription factors are known to have diversified functions in dynamic developmental stages by using alternative promoters and multiple splice isoforms. As the only member of AP-2 family genes in Drosophila, dAP-2 is expressed in multiple cell types during fly development. Multiple first coding exons have been identified for both Drosophila and vertebrate AP-2 family members (Bauer et al., 1998; Knight et al., 2005; Meier et al., 1995; Monge and Mitchell, 1998). I carried out in situ hybridization to compare relative use of exon 1a and exon 1b in dAP-2 transcripts in embryos and in larval imaginal discs and CNS (Fig. 2-13A-B). The data showed that the use of these two exons differs significantly both during embryogenesis and larval stages. At embryonic stage 14, exon 1a is strongly expressed in cells at multiple locations in the lateral protocerebrum while exon 1b is only found in an area of cells located in the medial
protocerebrum (Fig. 2-13C-D). Semi-quantitative RT-PCR was also used to measure the abundance of exon 1a or exon 1b-specific dAP-2 transcripts in embryo and larval RNA samples. The results corroborated the in situ data by showing that exon 1a is 10 folds more abundant than exon 1b in dAP-2 RNA during late larval development in both leg imaginal discs and CNS. In embryonic stage, both transcripts are made at similar low level, which is consistent with the fact that dAP-2 is only expressed in small group of cells at this stage. (Fig. 2-13E). The choice of activating specific transcription often indicates preference of certain combination of developmentally regulated enhancers and endogenous promoters (Boll and Noll, 2002; Butler and Kadonaga, 2001). To find out whether the two dAP-2 differ in compatibility with particular dAP-2 enhancers, I inserted cis-regulatory BXE into three different Pelican vectors using hsp70 basal promoter (h), predicted dAP-2 1a promoter (a) or dAP-2 1b promoter (b) respectively (Fig. 2-14A). The results confirmed these vectors gave less variation in expression pattern among independent lines. Cis-regulatory element BXE was able to drive strong lacZ expression in tarsal leg specific pattern when paired with cognate dAP-2 1a promoter, which has no background signal alone. BXE showed very weak expression in discrete cells on the distal leg rings when working together with hsp70 basal promoter or dAP-2 1b promoter (Fig. 2-14B-E). This suggested cell type-specific expression of dAP-2 is defined by the cooperation between enhancers and promoters. Therefore, all of my transgenic fly lines for the following enhancer analysis were generated using Pelican vector and dAP-2 1a promoter.
**Fig. 2-13 Differential use of two alternative first exons for dAP-2.** In situ hybridization with exon 1a and exon 1b specific RNA probes showed that exon 1a (A) predominates over exon 1b (B) in dAP-2 RNA in leg imaginal discs. In embryos, exon 1a-specific transcripts are detected in the lateral protocerebrum (arrow, C), maxillary segment (arrowhead) and ventral nerve cord (VNC) (not visible at this focal plane). In contrast, exon 1b transcripts are present in medial protocerebrum (arrow, D), but not in other parts of protocerebrum, maxillary segment or VNC. (E) RT-PCR results confirm the RNA in situ data, and show that exon 1a usage predominates over exon 1b usage in dAP-2 transcripts in larval leg discs (L) and CNS (C). In RNA from 0-24h embryos (E), exons 1a and 1b are fairly equally represented. GAPDH serves as an internal control for RNA input quantity.
**Fig. 2-14 dAP-2 enhancers for leg expression prefer exon 1a promoter.** (A) BXE fragment was used to construct Pelican vectors using different promoters: 1a promoter (a), 1b promoter (b) and heat-shock basal promoter (h). (B-E) β-gal staining of leg imaginal discs from 3rd instar larvae of the following transgenic lines: (B) alacZ report with dAP-2 1a promoter, in absence of inserted enhancer. (C-E) Compared to 1b and heat-shock basal promoter, 1a promoter is more compatible with BXE enhancer, showing strong distal leg specific expression.
2-3.7 Dissection of BXE enhancer for distal limb expression using Pelican vectors

The identification of BXE fragment responsible for tarsal leg-specific activity opened more questions for us: Can we identify sequences regulated for Notch-dependent expression in tarsal joints? Can we find separate enhancers for each tarsal joint? Are enhancers exclusively located within evolutionarily conserved regions? Whether leg-specific cis-regulatory elements work independently or co-operate with each other? I then carried out further dissection to analyze the necessity and sufficiency of smaller cis-regulatory elements for tarsal leg specific activity (Fig. 2-15A).

As shown earlier in the VISTA plot, dAP-2 exon 1a promoter (-207~ +424) contains major conserved peaks. Although this proximal segment can not activate transcription in vivo on its own, it is able to cooperate effectively with the BXE enhancer to drive report gene expression in presumptive tarsal joints. Higher similarity in DNA sequences around promoter region indicates transcription initiation mechanism is conserved between two Drosophila species. Several discrete conserved regions were investigated in the context of B6, a fragment that adds 200bp to the 3’-end of B6d to include the endogenous 1a promoter. B6 itself was able to drive not only tarsal joints specific expression like BXE, but also a weak proximal coxa expression domain. Both domains overlap with dAP-2 endogenous pattern, as shown by co-localization with dAP-2 protein (Fig. 2-15B-D). Conservation peaks #2, #3, and #4-6, originally identified by Consite alignment (http://www.phylofoot.org) before I started using VistaBrowser, were deleted respectively (Fig. 2-15A). Compared to B6-lacZ pattern in leg imaginal disc, deletion of peak #2 (dBX200-lacZ) and #3 (dBX600-lacZ) did not affect the tarsal leg-specific activity, while deletion of peaks #4-6 (dXE680-lacZ) reduced the distal leg signal
With no significantly conserved peaks, \textit{B6EB-lacZ} shows negative signal and \textit{BXE-lacZ} resembles \textit{B6-lacZ} pattern (Fig. 2-16 E,F). When splitting BXE into two fragments, BX and XE, I found neither of them showed strong complete ring pattern in distal leg region (Fig. 2-16G,H). This suggested both BX and XE contain essential enhancers elements and they are working co-operatively to define a strong tarsal joint specific activity.

Since BXE represents full tarsal activity of B6, another round of deletions was carried out in the context of the smaller BXE fragment. Although XE500 does not contain any conservation peak between \textit{D. melanogaster} and \textit{D. pseudoobscura} by default VistaBrowser setting, deletion of it or BX500 (\textit{dBX500-alacZ} or \textit{dXE500-alacZ}) significantly reduced tarsal leg-specific expression (Fig. 2-16I,J).

Then I combined critical fragments, BX500, XE500 and XE680, to see if they can sufficiently induce strong tarsal joint expression. XE680 alone (\textit{XE680-alacZ}), or the combination of BX500 and XE500 (\textit{BXE2S-alacZ}) only made very weak expression in distal leg segments (Fig. 2-16K,L). Only when I put all three fragments together in \textit{BXE2SA-alacZ}, the strong concentric ring pattern was restored (Fig. 2-16M).

Another interesting finding in the dissection was the antenna expression patterns of these transgenes greatly resembled their leg counterparts (Fig. 2-16A’-M’). The deletion of distal leg essential fragment BX500, XE500 or XE680 also resulted the loss of distal antenna expression. This suggested regulation of dAP-2 expression in distal antenna and leg development is evolutionarily conserved, which can also explain the conserved response to Notch signaling in antenna and leg discs.
I also tried to narrow down cis-acting element to even smaller fragment, but failed to get any strong leg-specific pattern (data not shown). No enhancer was found in my dissections to be responsible for only one of four tarsal rings. Three smaller cis-regulatory elements, BX500, XE500, and XE680, appeared to be required for specifying tarsal leg activity. Synergy from them was sufficient to induce strong tarsal expression pattern. Not only conserved, but also non-conserved regions were proved to contain cis-regulatory elements.
**Fig. 2-15 Dissection scheme for analysis of distal leg specific enhancers.** (A) At the top is VISTA plot for alignment between *D. melanogaster* and *D. pseudoobscura*, with conservation peaks #1~6 marked under the curve. Start codon (ATG), restricted enzyme sites and ends of each cloning fragments from *D. melanogaster* *dAP-2* are noted under VISTA plot by numbers in parentheses referring to distances in base pairs from the transcription start sites (+1). Enhancers were cloned into Pelican vector using intrinsic promoter, cloned *dAP-2* 1a promoters (orange) alp (-207 to +424). Name of dissected cis-element is located right to each construct diagram. *B6-lacZ* expression co-localizes with *dAP-2* tarsal rings (B-B’) and proximal coxa domain (arrows, C-C’). *BXE-alacZ* expression co-localizes with *dAP-2* tarsal domains (D-D’).
Fig. 2-16 Enhancer activities of B6, BXE and various subfragments in leg and antennal discs. Compared to strong tarsal activity of B6 (A), deletion in dBX200-lacZ (B) or dBX600-lacZ (C) did not affect tarsal leg-specific expression. Although BXE (E) retains B6 pattern without 3’-fragment B6EB (F), neither XE (G) nor BX (H) alone from BXE can restore this pattern. Enhancer fragments XE680, BX500 and XE500, deletion of which causing loss of lacZ expression (D, I, J), were found to be required for tarsal leg activity. XE680 only (K), or combination of BX500 and XE500 (L) could not restore strong tarsal leg expression. Only when three enhancer elements were put together, strong tarsal leg expression can be restored (M). (A’-M’) Corresponding antenna activities of these enhancers resemble their counterparts in the leg.
2-3.8 **BXE enhancer is direct target of Notch signaling**

The transgenic analysis in the previous section identified the *dAP-2* BXE fragment as an enhancer that activates reporter gene expression in presumptive tarsal joints in a pattern coincident with endogenous dAP-2 expression. I next asked whether the B6 and BXE reporter transgenes could recapitulate the Notch-dependency shown by endogenous dAP-2 expression in the distal leg joints. *Su(H)-/-* clones were induced in early larvae carrying *B6-lacZ* or *BXE-alacZ* transgenes and double immuno-staining of imaginal discs was carried out at late larval and early pupal stages to assess affects of loss of Notch signaling on reporter gene expression. The experiments confirmed that both reporter genes required SuH for expression in presumptive tarsal joints (Fig. 2-17A-A”, B-B”). As a control, the *EB*-lacZ reporter (Ahn et al, in preparation) was also examined with respect to Su(H) dependency. The *EB* genomic fragment (a subfragment of E6) contains two enhancer elements, PrF and PrC, that drive reporter gene expression in the presumptive proximal femur and proximal coxa domains of endogenous *dAP-2* expression, respectively. As I showed earlier, these domains of *dAP-2* expression in the proximal part of the developing leg are not affected by loss of Su(H). As predicted, the *EB*-lacZ reporter expression was not abolished in *Su(H)-/-* clones (Fig. 2-17C-C”, D).

The DNA binding transcription factor Su(H) and its mammalian homolog RBPJk have been studied extensively by several laboratories. The DNA binding sequence specificity of Su(H) has been well documented and examples of both high and low affinity binding sites for Su(H) in Notch-signaling dependent target genes such as those in the E(Spl) complex have been noted. Su(H) can mediate Notch-dependent activation, in
pairs or clusters to activate expression of Notch target genes (Bailey and Posakony, 1995; Nellesen et al., 1999). I searched B6 fragment for Su(H) consensus sequences derived from in vivo and in vitro experiments using several transcription factor binding site (TFBS) prediction programs. Many programs make prediction based on weight matrix of transcription factor binding sequence, either user-defined or imported from database such as TRANSFAC. For the last couple of years, there has been a trend to integrate phylogenetic footprinting into the prediction. My earlier trial with Consite returned several potential binding sites later all proved dispensable for Notch regulation (data not shown). Recently zPicture combined visualization of BLASTZ alignment and rVista TFBS prediction, using either TRANSFAC database or user-defined consensus sequences. Using zPicture and default setting for TRANSFAC, only one Su(H) binding site (BXES1) located in conserved BX500 region was predicted in B6 (Fig. 2-18A). When providing low-affinity consensus sequence defined by Posakony lab (Bailey and Posakony, 1995; Nellesen et al., 1999), I obtained two more sites: BXES2 located in the XE500 and B6EBS3 in the 1a promoter (Fig. 2-18A"). Deletion of regions containing BXES1 or BXES2 resulted in significant reduction in BXE activity (Fig. 2-16I,J). Site-directed mutagenesis (BXEmS1 or BXEmS2) targeting three critical nucleotides within BXES1 or BXES2 also showed similar effects (Fig. 2-18B-E). EMSA analysis showed BXES1 and BXES2 oligos were able to specifically bind bacteria-expressed Su(H) protein. Although BXES2 matches the low-affinity but not the high-affinity Su(H) consensus sequence, both BXES1 and BXES2 showed high-affinity binding of Su(H) similar to the positive control sequence from E(Spl)m4. Oligonucleotides containing mutated Su(H) binding sites also lost their binding ability in vitro (Fig. 2-19). The third potential Su(H) binding
site B6EBS3, found within B6EB fragment, showed almost no binding to Su(H) protein in EMSA analysis (data not shown). The negative signal from B6EBS3-containing 1a promoter-only transgene indicated this site alone was not sufficient to drive any expression. These results are consistent with \textit{in vivo} deletion data that B6EB is not required for tarsal leg expression.

Taken together, I have showed that Notch signaling is required for dAP-2 expression only in tarsal joints and dispensable for proximal leg segments. The regulation by Notch was through direct binding of Su(H) transcription factor on two high-affinity sites located within tarsal leg-specific enhancer BXE.
Fig. 2-17 Notch signaling is required only for tarsal leg specific enhancer activities.
(A-B) Su(H)-/- null mutant clones were induced at 36-60 hr AEL by 38°C heat–shock for 1 hour and analyzed for the effects on tarsal leg specific enhancer activities at white pupa stage. B6-lacZ (A-A’’) or BXE-alplacZ (B-B’’) expression (red) were lost in the Su(H)-/- clones at presumptive tarsal joints (arrows). Mutant clones (white line) are marked by absence of GFP staining (green). (C-C’’) Su(H)-/- null mutant clones were induced at 48-72 hr AEL by 33°C heat-shock for 1 hr with EB^{1.7R}-GFP (green) reporter construct. Su(H)-/- clone is marked by loss of β-galactosidase (red). (D) wild-type EB^{1.7R}-GFP shows two proximal rings (arrowheads), which was not affected in Su(H)-/- clones (C-C’’).
**Fig. 2-18 Identification of Su(H) binding sites in BXE enhancer.** Prediction of Su(H) binding sites BXES1, BXES2 and B6EBS3, by zPicture/rVista using TRANSFAC database (A) or RTGRGAR consensus sequence (A’). (B) BXES1 (-5200) and BXES2 (-3143) were mutated in the context of BXE enhancer (-5200, ATGTGAGAA → ATGCTCGAA; -3143, CATGGGAA → CATCTCAA). Compared to BXE pattern, mutagenesis of Su(H) binding sites lost tarsal-specific expression (C-E).
Fig. 2-19 Electrophoresis mobility shift assay of Su(H) binding sites. (A) Alignment of predicted Su(H) binding sites. Su(H) high-affinity and low-affinity sequences examined in vitro are summarized (Bailey and Posakony, 1995; Nellesen et al., 1999). Y=C/T, R=A/G. E(Spl)m4S1 is used as control. (B) Radiolabeled wild-type DNA probes containing predicted Su(H) sites and surrounding nucleotides bound efficiently in vitro to the GST-Su(H) protein. Mutated versions of oligos are unable to bind Su(H). 20X (molar ratio) bacterially expressed GST-Su(H) protein was used to mix with 200fmol probes. 50X unlabeled probes were used as competitor and competed effectively.
Identification of essential AP-2 binding site in BXE that mediates autoregulation

We noted consensus dAP-2 binding sites in BXE when searching for Su(H) sites. Therefore, I examined the endogenous dAP-2 and leg segment specific enhancer activities in dAP-2/- leg discs to assess possible dAP-2 autoregulation. During 3rd instar larval stage, dAP-2 1a was preferentially expressed in leg imaginal disc as shown by RT-PCR and in situ hybridization (Fig. 2-13). When I examined dAP-2 1a transcription in the background of dAP-2 null mutant dAP-2^2/dAP-2^2, I found distal leg transcripts level was significantly reduced during both larval and pupal stages (Fig. 2-20A-D). dAP-2^2/3 null allele was then used to observe change in endogenous dAP-2 protein level, because the mis-sense mutation in dAP-2^2/3 preserved C-terminal epitope recognized by dAP-2 antibody. As expected, dAP-2 protein level was also reduced in distal leg segments (Fig. 2-20E,F). Then I combined reporter gene driven by distal leg segment-specific enhancer elements, BXE-alacZ, with dAP-2^2/dAP-2^2/3 null background. I found tarsal leg enhancer activity decreased greatly when dAP-2 expression was lost (Fig. 2-20G,H). On the contrary, reporter gene driven by proximal leg specific enhancer dissected from E6, EB^Apel-lacZ, was not affected by dAP-2 null background (Fig. 2-20I,J). Previous study has suggested that in dAP-2 null mutant leg was severely shortened with most of cell death happening in the interjoint region (Kerber et al., 2001). To rule out the possibility that missing distal leg expression was due to cell death, I examined cell joint marker bric-a-brac2 (bab2) and found the originally discrete expression of bab2 merged into continuous distal expression pattern (Fig. 2-20K,L). It confirmed joint cells were surviving and therefore loss of BXE-lacZ expression in tarsal joints was not due to cell
death. My results indicated *Drosophila dAP-2* was under the autoregulation through tarsal, but not proximal leg segment-specific *cis*-regulatory elements. The autoregulation, like Notch signaling pathway, has differential effects on specific leg segments along proximal/distal axis.

I further searched potential dAP-2 binding sites on tarsal-specific enhancer BXE. Although dAP-2 consensus sequence has not been systematically identified, binding site of its closet homology in human AP-2 family, AP-2α, has been well studied by SELEX method (Mohibullah et al., 1999). Using zPicture/rVista, I predicted 3 potential dAP-2 binding sites matching SELEX sequence HGCCBNVRGS (data not shown). One of them is located in the BX600 (peak#3) fragment and therefore has been proved unnecessary for tarsal leg activity (Fig2-15, 2-16). The other two, BXEA1 and BXEA2, were mutated in the context of BXE fragment (Fig. 2-21A,B). Only mutagenesis of BXEA2 site resulted in reduction, but not total loss of tarsal-specific expression (Fig. 2-21C-E). The data revealed the direct autoregulation site for dAP-2 tarsal activity, and suggested without autoregulation, other transcription factors or signaling pathways were still able to drive weak tarsal leg expression. Therefore, the autoregulation might be a cooperative or supplementary regulatory mechanism.

Studies from transcriptional regulation of human *AP-2* gene showed multiple transcriptional factor binding sites, including multiple AP-2 sites conserved among different mammalian species, were found in upstream *cis*-regulatory sequences. These AP-2 sites were able to bind AP-2 protein *in vitro* and increase the responsiveness to exogenous AP-2 activation. The expression of a human AP-2α: lacZ transgene is expanded in the forelimb buds of AP-2α null mice. These results indicated the potential
autoregulation mechanism for human AP-2 (Creaser et al., 1996). This study indicated potential negative autoregulation in human $AP-2\alpha$ gene.
Fig. 2-20 Distal leg specific autoregulation of dAP-2. dAP-2 1a transcript is predominately expressed in the larval (A) and prepupal (C) leg. In the dAP-22 null mutants (B,D), distal leg expression reduced (arrows). dAP-2213 mis-sense mutant (F) preserved C-terminal epitope for antibody recognition. Loss of protein level was also seen in distal leg (arrow in F) compared to wild type pattern (E). BXE-alacZ tarsal-specific expression (G) was also affected in BXE-alacZ; dAP-22/dAP-215 mutant (H). However, proximal specific expression from EBAlp-lacZ (I) is not affected in EBAlp-lacZ; dAP-22/Df1118 mutant (J). The distal expressions in (I) and (J) are non-specific. Joint marker bab2 in wild type (K) is expressed in multiple discrete rings. In contrast, bab2 expression in distal part of leg disc is continuous in dAP-2 mutant (L) background, suggesting dAP-2 may normally represses bab2 expression directly or indirectly in presumptive tarsal joints.
Fig. 2-21 Identification of dAP-2 binding site in BXE mediating autoregulation (A) Two potential dAP-2 binding sites conserved between two *Drosophila* species, one located between peak #2 and #3 and another within peak #4, were mutated (-4096 and -2892). (B) Alignment of SELEX consensus sequence, BXEA1, BXEA2, and their mutated sites. (C) BXE-alacZ. (D) BXEmA1 (E) BXEmA2.
2-3.10 Model for \textit{dAP-2} regulation in developing leg and antenna

Based on differential expression pattern and regulation by Notch signaling and autoregulation of \textit{dAP-2} along proximodistal axes of leg and antenna, I have proposed a model to explain some aspects of \textit{dAP-2} regulation in \textit{Drosophila} leg (Fig. 2-22).

In the distal leg and antenna segments, the \textit{dAP-2} is activated and maintained by transcription factors binding BXE enhancer element. For example, at the boundary of distal tarsal 1 and proximal tarsal 2 segments, Notch signaling is turned on distally to the ligand-expressing cells and repressed by Fng proximally. Recent study reported that the EGFR signaling also represses Notch activation in the cells proximally adjacent to ligand-expressing cells signaling (Galindo et al., 2005). \textit{dAP-2} is directly activated by Notch signaling via BXE enhancer, which was shown by requirement of Su(H) binding sites identified within BXE for \textit{in vitro} protein binding ability and \textit{in vivo} BXE activity (Fig. 2-18, 2-19). Thereafter, \textit{dAP-2} protein also directly binds to BXE to maintain further \textit{dAP-2} expression (Fig. 2-22). Loss of \textit{dAP-2} function or mutagenesis of \textit{dAP-2} binding sites also significantly reduce \textit{dAP-2} expression and BXE enhancer activity (Fig. 2-20, 2-21).

In more proximal leg and antenna segment, Notch signaling is not required for \textit{dAP-2} activation based on the observations that \textit{dAP-2} expression is not affected by loss of Su(H). This is supported by the finding that endogenous \textit{dAP-2} expression domain in proximal leg and antenna segments does not overlap with that of Notch target gene (Fig. 2-3). The identification of E6 fragment downstream of BXE suggested proximal leg and antenna \textit{dAP-2} expression is regulated by different \textit{cis}-regulatory elements. Dissection of E6 revealed that multiple enhancers within EB, a sub-fragment of E6, are dedicated to
proximal femur and proximal trochanter activities, and directly regulated by Hox/Exd (Ahn et al., unpublished data). I further showed that Notch activation and dAP-2 autoregulation are not required for proximal dAP-2 expression and EB enhancer activity (Fig. 2-17, 2-20).
Fig. 2-22 Model for differential regulation of dAP-2. (A) Regulation of dAP-2 expression at distal tarsal joints. For example, at tarsal 1/proximal tarsal 2 (dT1/pT2) joint, dAP-2 is activated by Notch in dT1 cells. Notch signaling proximal to ligand (Ser)-expressing cells is inhibited by both Fng and EGFR signaling. Direct binding of Su(H) on BXE enhancer activates dAP-2, which in turn also binds directly on BXE and maintains further dAP-2 expression. (B) Hox/Exd directly regulates dAP-2 activity in proximal leg segments via multiple enhancers within EB (Ahn et al., unpublished data). Notch activation and dAP-2 autoregulation are not required for proximal dAP-2 expression or EB enhancer activity.
2-4 Discussion

2-4.1 Evolutionarily conserved function and regulation of dAP-2

I have reported here that dAP-2 gene shares conserved function of AP-2 family genes and is involved in Drosophila appendage development. The regulation of dAP-2 expression is controlled by multiple signaling pathways and transcription factors via discrete and synergistic leg-specific enhancers that are evolutionarily conserved.

The conservation of dAP-2 function with other AP-2 family members is first reflected by the similarity of its coding sequence. As the sole member of AP-2 family genes in Drosophila, dAP-2 has the conserved N-terminal transcriptional activation domain and C-terminal DNA binding and dimerization domain as vertebrate homologs have. The protein sequence comparison showed dAP-2 is slightly more similarity to mammalian AP-2 alpha than to other mammalian family members, and less closely related to C. elegans AP-2 protein (Wilson et al., 1994). This suggests the ancestral AP-2 gene diverged after the separation of vertebrate and invertebrate species and that gene duplications occurred in vertebrate evolution. Still, we can see resemblance in expression pattern and functions between dAP-2 and its vertebrate homologs. dAP-2 is expressed in the embryonic and larval CNS, and in appendage primordia including labial, antenna and leg discs. Mouse AP-2 alpha is expressed in the developing brain, spinal cord, facial primordia and limb buds. Therefore, the loss of function of corresponding genes displays parallel phenotypes. Drosophila dAP-2 null mutant have severely shortened legs and proboscis and hypomorphic mutants show locomotor defects that correlate with a
disruption of multiple brain central complex structures. *AP-2 alpha* knockout mice showed craniofacial defects such as failure of maxilla and mandible growth, and missing radius lone and first digit of limb (Monge et al., 2001; Monge and Mitchell, 1998; Schorle et al., 1996; Zhang et al., 1996). One of human *AP-2 beta* dominate negative allele, resulting in Char syndrome with missing knuckle in the small finger and facial anomaly, shares the same mutation with *dAP-2* in *Drosophila* (Satoda et al., 2000). My data showing that *dAP-2* mutant lost distal dAP-2 expression due to autoregulation indicated that the mis-sense mutation, which has been shown to abolish protein-binding ability in human *AP-2 beta* gene, is likely to have same effect in *Drosophila* dAP-2 protein.

In addition to protein sequence conservation, the exon structure of *dAP-2* gene is also more similar to mouse *AP-2 alpha* gene, with two alternative first exons and two conserved exon/intro boundaries within the DNA binding domain. However, the comparison between *Drosophila* and mouse genomic sequence surrounding AP-2 gene revealed no significant alignment except exon5 and exon 6 coding for the most highly conserved motifs of DNA binding/dimerization domains (Fig. 2-23A). Therefore, in order to utilize comparative genomic in predicting conserved *cis*-regulatory element in non-coding sequences, choosing reasonably diverged species is important. As I have shown, *D. pseudoobscura* and *D. ananassae* are better choices. Prediction using more closely related *D. simulans* or *D. yakuba* potentially includes all *cis*-regulatory elements, but increases false positive results. *D. mojavensis* an *D. virilis* has less conserved peaks with *D. melanogaster*, but some of the important *cis*-regulatory elements could be missed (Fig. 2-11B). The multi-species comparison also appears helpful for identify CRMs of
vertebrate AP-2 family genes (Fig. 2-23B). Pairwise alignment of human AP-2 alpha against mouse, rat, frog, dog, cow and opossum respectively by VistaBrowser showed conserved non-coding sequences are mostly located within introns and downstream of AP-2 alpha coding sequence. If using human vs. dog comparison, only intron 1, 2, 3, 5 and 5kb downstream genomic sequence contain major conservation peaks. Interestingly, in vivo study in human AP-2 alpha gene cis-acting element identified facial mesenchyme and limb bud specific enhancers within intron 5, and potential trigeminal ganglia and spinal cord specific enhancers downstream of AP-2 alpha (Zhang and Williams, 2003). The coincidence of functional enhancers and conserved non-coding sequences between remotely related vertebrates suggests these conservation peaks contain essential and functional CRMs.

My dissection work has both gained suggestions from and provided additional information to the results of comparative genomics. Non-coding sequence around dAP-2 genes is consists of islands of conserved regions. Two separable fragments B6 and E6 are responsible for distal and proximal leg segment-specific expression of dAP-2 (Fig. 2-9). The combination of two fragments can restore cell-specific expression of dAP-2 and function required for both joint formation and interjoint cell survival (Fig. 2-10). BXE, the conserved region within B6, was predicted by conservation between two Drosophila species and later proved representing same B6 tarsal-specific activities (Fig. 2-15). Further dissection following phylogenetic footprinting guide successfully identified essential motifs, BX500 and XE680, within BXE CRMs. Strikingly, functional transcription factor binding sites, for Su(H) and dAP-2, were found in each of two motifs. These studies illustrated functional genomics and comparative genomics can assist each
other very well. However, I also realized that cis-regulatory elements are changing under stabilizing selection. As in the case of even-skipped (eve) stripe 2 enhancer analysis, only 3 out of 17 transcription factor binding sites are completely conserved among 6 Drosophila species. Bicoid (bcd) binding site 3 and hunchback (hb) binding site 1 do not have corresponding sites in D. pseudoobscura and D. picticornis. Large majority of the nucleotide substitutions in the binding sites happens only once in the phylogeny, suggesting the change is likely to be either adaptive substitutions or slightly deleterious mutation fixed by genetic drift (Ludwig et al., 1998). For dAP-2 tarsal leg-specific enhancers among seven Drosophila species, we can see the highly conserved peak 1 within BX500 motif is inverted exclusively in D. pseudoobscura, while the Su(H) binding site BXES1 and surrounding sequence are still well conserved. BXES2 in XE600 is missing from the counterpart position in D. pseudoobscura. However, searching for Su(H) binding sites over the counterpart of B6 in D. pseudoobscura returns 8 potential sites, comparing to 3 sites in D. melanogaster. I propose that BXES2 might be adaptive response to Notch signaling in D. melanogaster, as many positional variations of Su(H) binding sites for other Notch target genes can be seen between these two species (data not shown). Comparison of dAP-2 sites in XE600 (BXEA2) between multiple species also illustrated minor substitutions while keeping at least 8/9 consensus sequence intact (data not shown). Although changes of nucleotide sequences were observed in different species, I would expect to see unchanged expression pattern by BXE/B6 enhancer element. In fact, counterpart of XE fragment in D. pseudoobscura also contains distal leg specific enhancer activities (data not shown). As suggested in eve stripe 2 enhancer, the multiplicity of binding sites may be important in cooperative binding and in assuring
robust performance (Ludwig et al., 1998). My finding that synergy among BXES1, BXES1 and BXEA2 was required for tarsal specific enhancer activity also supported this idea.

The cooperation between enhancer and promoter is also presented in my study. *dAP-2* leg specific enhancer prefers its own TATA-less promoter than *hsp70TATA* basal promoter (Fig. 2-14). My results showed that the two alternative endogenous promoters for *dAP-2* are used in different cell types and at different levels, suggesting selective cooperation between enhancers and promoters (Fig. 2-13).
Fig. 2-23 Conservation between AP-2 family genes. (A) zPicture alignment of mouse AP-2 alpha vs. Drosophila dAP-2 (used as reference sequence) shows only exons 5-6 have significant conservation in DNA sequence. (B) VistaBrowser pairwise alignments of human AP-2 alpha vs its homologs in mouse, rat, frog, dog, cow and opossum respectively. Green box indicates conservation in intron 5, containing potential enhancer for limb bud progress zone (LPZ) and frontonasal prominence (F). Orange box indicates conservation downstream of 3’-end of coding sequence, containing potential enhancer for trigeminal ganglia (TG) and spinal cord (SC) (Zhang and Williams, 2003).
2-4.2 Regulatory mechanisms of dAP-2 in appendage development

In this section, I present data supporting the idea that conserved aspects of dAP-2 function and regulation are found not only between species, but also between homologous structures. In the dissection of distal specific cis-regulatory elements, I found striking resemblance of expression patterns between leg and antenna discs (Fig. 2-16). Although they have evolved into morphologically and functionally distinct appendages, leg and antenna both express wild-type dAP-2 in multiple-ring pattern, located at the presumptive segment boundaries. Bib, another Notch signaling target, also preserves the multiple-ring pattern in both leg and antenna. More interestingly, for both appendages the two genes are expressed in overlapping distal segment domains while sitting adjacent to each other on two sides of proximal boundaries (Fig. 2-3, 2-5). This suggests Notch signaling is conserved and probably involved in regulating dAP-2 in both leg and antenna joint/segment boundary development. My genetic experiment confirmed dAP-2 is downstream of Notch signaling in both leg and antenna. This regulation is only present in tarsal leg and distal antenna segments but absent in intermediate and proximal segments (Fig. 2-7, 2-8). It indicates not only the regulation of dAP-2 by Notch signaling is conserved between leg and antenna, but also other signaling pathways and transcription factors are required to direct dAP-2 proximal segment-specific expression. The identification of direct Su(H) binding sites within tarsal-specific enhancer BXE provided additional support to this idea. When these sites were mutated, both leg and antenna lost distal segment specific activities (Fig. 2-18).

The direct regulation of tarsal-specific enhancer activity by Notch signaling is also proved by in vivo Su(H) mosaic analysis and in vitro EMSA. I have also showed that
Su(H) regulates dAP-2 expression by direct binding to tarsal enhancer while proximal leg enhancer EB is not responding to Notch regulation (Fig. 2-17, 2-19).

Following identification of Su(H) binding sites, I found the combination of Notch-responding cis-regulatory element only was not sufficient to restore strong tarsal expression. This indicates for tarsal-specific regulation Notch signaling does not provide enough cell-type specific activation. In searching for other potential regulations of dAP-2 expression, I found dAP-2 is subject to autoregulation, which has been suggested in human AP-2 alpha in vivo and in vitro analysis. The interesting feature of dAP-2 autoregulation is that, like Notch regulation on dAP-2, it also appears to be distal leg-specific. dAP-2 transcripts, protein and enhancer activity levels were all reduced in the tarsal segments, while proximal enhancer EB was not affected in dAP-2 null mutant (Fig. 2-20). The autoregulation was later also identified as direct regulation by loss of transgene activity when dAP-2 binding sites were mutated (Fig. 2-21). Since dAP-2 is first turned on by Notch signaling and loss of enhancer activity in dAP-2 mutant is not complete, I proposed that autoregulation is complementary to Notch activation to maintain the strong tarsal expression.

Our studies of dAP-2 and studies in other laboratories of genes such as Drosophila eve, fitz, and poxneuro, and mouse Pax6, have shown that developmentally important genes are regulated by a surprising number of separable, independent enhancers each responsible for different spatiotemporal aspects of the gene’s expression pattern during development (Boll and Noll, 2002; Morgan, 2004; Small et al., 1992).
Chapter 3

Analysis of CG10440, a BTB/POZ protein and candidate co-repressor for dAP-2

3-1 Introduction

3-1.1 Target of genes of dAP-2

Since AP-2 family members are essential for development of various tissues and structures during animal embryogenesis, an identification of downstream target genes for dAP-2 would help us understand how AP-2 genes carry out their functions in different developmental contexts. Specific cells. Based largely on analysis with tissue cultured cells, it has been shown that vertebrate AP-2 genes are associated with the regulation of a wide range of target genes, including E-cadherin and bcl-2 (Decary et al., 2002), Hoxa2 (Maconochie et al., 1999), matrix metalloproteinases 9 gene (Ma et al., 2004), a number of genes encoding neural cell type-associated proteins (e.g., dopamine beta-hydroxylase) (Kim et al., 2001), and hormone biosynthetic enzyme genes such as cholesterol side-chain-cleavage cytochrome P450 (Ben-Zimra et al., 2002).

A great number of AP-2 target genes identified in cell culture studies are related to cell differentiation, proliferation, apoptosis and tumorigenesis, where AP-2 genes also play important roles. Several studies have suggested that AP-2 alpha is a tumor suppressor gene. For one, AP-2 has been shown to block the activity of c-myc, a
transcription factor that induces cellular proliferation and apoptosis, by direct binding competition to overlapping AP-2 and Myc binding sites on the cis-regulatory elements of myc-target genes. AP-2 also inhibits cancer cell growth through direct activation of p21 expression (Hilger-Eversheim et al., 2000). Support for a possible role of AP-2 as a tumor suppressor also comes from the studies of human malignant melanomas, where loss of AP-2 expression results in inhibition of proto-oncogene c-Kit and overexpression of adhesion molecule MCAM/MUC18. On the other hand, overexpression of AP-2 alpha and AP-2 gamma are correlated with overexpression of oncogenic gene c-erbB2 in mammary carcinomas. Functional binding sites for AP-2 genes were found in the c-erbB2 promoter and transfection analysis showed the activation of c-cerbB2 by AP-2. More importantly, transgenic animal overexpressing AP-2 gamma showed increases in c-erbB2 and other proliferation markers and repression of differentiation markers, as well as induction of apoptosis promoting gene IGFBP-5 (Jager et al., 2003). The oncogenic function of AP-2 gamma can be suppressed by Wwox tumor suppressor protein via protein-protein interaction and subcellular redistribution (Aqeilan et al., 2004).

Genome-wide searching of AP-2 alpha target genes has been carried out using a combination of suppression-subtractive hybridization and high throughput differential screening (Pfisterer et al., 2002). By comparing expressed genes in a single head of an E8.75 knockout and control mouse, Pfisterer et al. identified a set of genes repressed by AP-2 alpha and retard cellular proliferation and induce differentiation and apoptosis.

Our lab has initiated genome-wide identification of Drosophila dAP-2 leg-specific target genes by microarray analysis (MacIver et al., unpublished data). RNA transcripts profiles of 3rd instar larval leg imaginal discs from wild-type and dAP-2 null
mutant were used with cDNA microarray containing 12,000 *Drosophila* genes, and the data were subjected to a rigorous statistical analysis. 20 genes were identified as differentially expressed, and 16 of them were verified by RT-PCR. Among them, 4 genes, a cytochrome p450 (Cyp4p2), anachronism (ana), odorant binding protein 99A (Obp99A) and photorepair (phr), are downregulated with more than 3 folds change. In situ hybridization of Cyp4p2, ana and Obp99A confirmed their leg disc-specific expression were significantly reduced, and also revealed additional changes in their larval CNS pattern. Ana and Obp99A are expressed in restricted regions of the leg disc corresponding to discrete parts of the dAP-2 ring pattern; in contrast, Cyp4p2 is expressed ubiquitously at low levels. Expression of *Obp99A* in optical lobe is upregulated in dAP-2 mutant, while no CNS *ana* expression change is observed. Cyp4p2’s low-level expression in CNS is totally lost in the dAP-2 mutant.

*ana* encodes a secreted glycoprotein expressed by CNS glial cells and is required to delay neuroblast proliferation in early larvae (Ebens et al., 1993). Mutations in *ana* are viable and cause defects in optic lobe organization and in larval olfactoary behavior (Park et al., 1997). Obp99A belongs to class 4 of a large family of odorant binding proteins, whose generalized functions are involved in presenting insoluble odorant molecules to odorant receptors in olfactory sensilla and to possibly act as a sink for odorants or to protect odorants from degradation (Galindo and Smith, 2001; Graham and Davies, 2002; Hekmat-Scafe et al., 2002). Based on the overlapping expression pattern of these two genes with that of *dAP-2*, as well as potential binding sites found near their promoter sequences, *ana* and *Obp99A* are likely to be direct targets of dAP-2. Cyp4p2 has been classified as being involved in steroid hormone metabolism by comparison to other
known cytochromes, and no experimental data about its function is available (Tijet et al.,
2001). Its ubiquitous expression pattern comparing to dAP-2 ring pattern makes it likely
to be an indirect target of dAP-2.

3-1.2 CG10440

Several studies have identified protein interaction partners for mammalian AP-2
including SV40 T antigen, retinoblastoma proteins (Batsche et al., 1998; Mitchell et al.,
1987). A potentially interesting interaction partner for dAP-2, called CG10440, was
identified in a genome-scale yeast two-hybrid protein-interaction screen as the only
protein showing a high confidence interaction with dAP-2 (Giot et al., 2003). CG10440
encodes a 338 amino acid protein which is predicted to contain a protein dimerization
domain belonging to a subfamily of BTB/POZ (Broad-complex, tramtrack, and bric-a-
brac/ poxvirus and zinc finger) domain that mediates homomeric or heteromeric
dimerization.

Many BTB/POZ proteins are also zinc-finger transcription factors, including, for
example, Broad-complex, a protein involved in ecdysone-induced regulatory cascade,
and Tramtrack, a transcriptional repressor involved in regulation of embryonic
segmentation genes and eye development (Zollman et al., 1994). Another BTB domain
containing protein, bab, is a transcription factor required for Drosophila tarsal joint
development and tarsal segment outgrowth. Besides potential binding to dAP-2 and itself,
CG10440 also interacted with moderate confidence to the protein lesswright (lwr), a
protein with features of a SUMO-conjugating enzyme (Fly GRID interaction data).
3-2 Material and Methods

3-2.1 Fly stocks

P-element insertion line \( P\{w^{+}\}^{G{T}{T}} \) and deficiency line \( Df(2R)Egfr18/CyO \) used for P-element mediated mutagenesis experiment are from flybase (www.flybase.org). \( Dll^{m212-} \)-\( GAL4 \) is a gift from Gerald Campbell; \( btd-GAL4 \) is from Gines Morata; \( ptc-GAL4, tubP-GAL4, \) and \( UAS-lacZ \) are from flybase; \( B12-GAL4 \) is made with 12kb dAP-2 leg-specific enhancer. \( dAP-2^{2} \) and \( dAP-2^{15} \) mutants, \( UAS-dAP2a4 \) stocks are from our lab stocks (details in Chapter 2).

3-2.2 Construction of \( pUAS-CG10440HA \)

CG10440 coding sequence cDNA was reverse transcribed by 18-mer oligo(dT) primer from adult total RNA and amplified by PCR using 5’-primer 5’-CAAACGAAAAATGCGATATGGAC-3’ and 3’-primer 5’-TCATGCGTAATCCGGAACGTCGTAGGGGTACATTGGAACGCGTCGCGCCAG-3’ to attach hemagglutinin (HA) epitope to the C-terminal end of expressed protein. The PCR product was first cloned into pGEM-T easy vector (Promega) for sequencing confirmation, then subcloned using EcoRI site into pUAST vector for embryo injection.

3-2.3 Construction of \( pUAS-CG10440^{Wi} \)

pWIZ vector (Lee and Carthew, 2003) is a derivative of pUAST expression vector with \( hsp70TATA \) promoter and \( Drosophila white \) gene intron2 spacer for quick and high-efficiency RNAi construct generation. A 695bp cDNA fragment covering part of exons
4-5 was amplified by RT-PCR using 5’-primer 5’-CCTTCTAGAACCGTCTCGCAACGATC-3’ and 3’-primer 5’-TCCTCTAGAGCTCGTCGAGCAGAGC-3’, then digested by XbaI at both ends and cloned into Avr II digested pWIZ upstream of white intron 2. The orientation of first insertion was determined by colony-PCR using hsp70TATA promoter primer (5’-GAGCGCCGAGTATAAATAGAGCGCGCTTCTCTAC-3’) and one of insertion cloning primers. Then the same insertion fragment digested by XbaI was cloned into the intermediate vector from last step, which was then digested by NheI downstream of white intron 2. The orientation of the second insertion was confirmed by SphI digestion to ensure opposite to that of the first insertion. Two kinds of CG10440 RNAi constructs, which are oriented as head-head and tail-tail, therefore were obtained for embryo injection.

3-2.4 Immunostaining

1:1000 dilution of mouse anti-HA antibody (gift from Micheal Teng lab), 1:500 dilution of mouse anti-GFP (Invitrogen), 1:1000 dilution of mouse anti-β-galactosidase antibody (Promega) and 1:800 dilution of rabbit anti-dAP-2 antibody were used in immunostaining.

3-2.5 RT-PCR

Leg imaginal discs were collected from 10-20 3rd instar wandering larvae to extract total RNA using Trizol protocol. 1µg total RNA was used for reverse transcription using 18-mer oligo(dT) primer. cDNA from 20ng total RNA was used to
amplify by PCR (94°C x 30sec, 56°C x 30 sec, 72°C x 1 min) using the following primers for target genes (28 cycles for GAPDH, ana, and dAP-2; 26 cycles for the rest):

GAPDH-5’: 5’CAGCCATCACAGTCACTGGATTCC-3’
GAPDH-3’: 5’-TTCCGGTGAGCTTACCGTTG-3’
CG10440-5’: 5’-TCTGATTGCGAGGATTTC-3’
CG10440-3’: 5’-GAGCCTATTCACATTGGGAACG-3’
Ana-5’: 5’-CCACGCAGAGGAAGAAACAGC-3’
Ana-3’: 5’-CAAACCGATGAAGCTAAACG-3’
Obp99A-5’: 5’-GCCTCCGCCGACTATGTTG-3’
Obp99A-3’: 5’-TCCCTGGCTGTTTTGTTGCCAC-3’
Cyp4p2-5’: 5’-AGGAACTGTGTACCAAGAGA-3’
Cyp4p2-3’ 5’-ACGACGATTAGCATGGAGAG-3’
dAP-2-5’: 5’-AGTTGCCTTAACCAGTGCC-3’
dAP-2-3’: 5’-GAGGGCTCCATGGTTCTTGCTG-3’

3-2.6 RNA in situ hybridization:

CG10440, ana, and Obp99A cDNA fragments amplified using primers above were cloned into pGEM-T easy vector in the orientation that antisense strand was on the same strand of T7 primer. Riboprobes for these three fragments were made using DIG-RNA Labeling Kit and in situ hybridization were carried out following protocols described in Chapter 2. The riboprobe for CG10440 was 508bp.

3-2.7 Genomic PCR detection of P-element insertion
P-element primer Pp31 (5’-CGACGGGACCACCTATGTTATTTCATCATG-3’) was paired with one of the following primers in CG10440 promoter or coding region (number in parenthesis is the position with regard to transcription initiation site):

P1-5’(-865): 5’-CGGTTCTGGTTCTGCTTCTGC-3’
P2-3’(+1164): 5’-AATCGGACTCGTTCTCGTCTGC-3’
P3-5’ (+8320): 5’-GGCGATTTGGTCTTGTCCACG-3’
P4-3’ (+10302): 5’-GAGCCTATCACATTGGAACG-3’

Genomic DNA preparation and PCR protocols are described in Chapter 2.
3-3 Results

3-3.1 Expression and subcellular localization of CG10440

In situ hybridization showed *CG10440* is expressed in the maxillary segment, protocerebrum and ventral nerve cord during embryogenesis (BDGP in situ gene expression database). The embryonic expression pattern of CG10440 greatly resembles that of *dAP-2* at the same developmental stage (Fig. 3-1 A,B). CG10440 is expressed in third instar larval leg imaginal discs in a weak ubiquitous pattern together with a stronger concentric ring pattern reminiscent of *dAP-2* expression (Fig. 3-1C-D). The similar expression patterns indicate that *dAP-2* and *CG10440* could possibly regulate each other or work cooperatively to regulate shared downstream target genes.

The prediction based on conserved N-terminal motif of CG10440 protein indicates it belongs a subfamily of BTB/POZ proteins that also include many nuclear transcription factors. CG10440 was also predicted to interact with lwr, a nuclear SUMO-conjugating enzyme. I used PSORT program to predict the subcellular localization and obtained probability of 47.8% being nuclear protein and 17.4% probability being cytoplasmic protein (Nakai and Horton, 1999). Since antibodies against CG10440 are not available, I attached an HA epitope to CG10440 so that the over-expressed protein could be detected with anti-HA. Assuming the HA-tag does not affect folding and subcellular localization of CG10440 protein, the data suggest that CG10440-HA protein was distributed in both cytoplasm and nucleus, with stronger signal near the plasma membrane and weaker signal in the nucleus (Fig. 3-1E-E’).
Fig. 3-1 Expression and subcellular localization of CG10440. (A-D) RNA in situ hybridization to detect CG10440 transcripts (A, C) and dAP-2 exon 1a-specific transcripts (B, D). A) CG10440 expression pattern during embryogenesis (BDGP gene expression database) is highly similar to that of dAP-2 (B). Embryos are stage 13-14. mx, maxillary; pro, protocerebrum. (C, D) RNA expression pattern for CG10440 (C) in leg imaginal disc resembles that of dAP-2 (D). (D) dAP-2 1a RNA transcripts detected 481bp dAP-2 exon 1a riboprobe in leg imaginal disc. (E-E’). double staining of CG10440HA (green) overexpression driven by ptc-GAL4 and dAP-2 (red) in leg imaginal disc. CG10440 overexpression detected by HA-tag was seen not only overlapping nuclear protein dAP-2, but also concentrated near plasma membrane.
### 3.3.2 Overexpression of CG10440 affects leg development

Since there was subcellular colocalization between dAP-2 and CG10440 expression in leg disc, I investigated if dAP-2 and CG10440 could regulate each other's expression in leg discs. Genetic experiment showed CG10440 transcription still existed, though the ring pattern was fused, in dAP-2 null background (Fig. 3-2A-B). On the other hand, overexpression of CG10440 by ptc-GAL4 did not generate ectopic dAP-2 expression (Fig. 3-2C-D). However, when I used strong GAL4 driver such as Dl<sup>em212</sup>-GAL4 and tubP-GAL4 to ectopically induce CG10440, I found the number of eclosed adult flies sharply decreased. Most adults died inside pharate pupae with shortened, nonfunctional leg (Fig. 3-3). Even some of them escaped, they were not able to walk or fly and died within a few days without getting enough food. The leg-shortening phenotypes included fusion of joints and undergrowth of interjoint region, and were correlated with GAL4 driver expression domain and strength. Dl<sup>em212</sup>-GAL4 is the strongest driver we have and expressed mainly in distal leg segments, therefore it resulted in extremely shortened leg when driving ectopic expression of CG10440. tubP-GAL4 is less stronger and expressed ubiquitously, while buttonhead (btd)-GAL4 is even weaker and expressed in all leg segments except tarsal 1 segment.
Fig. 3-2 *dAP-2* and *CG10440* do not regulate each other. (A, B) RNA *in situ* hybridization to detect *CG10440* transcripts in wild-type leg discs (A) and *dAP-2*/*dAP-2* null mutant leg discs. (C, D). Double immunostaining to detect dAP-2 protein (red) and β-galactosidase (green, indicating where *UAS-lacZ* expression is driven by *ptc-GAL4*) in wild-type leg discs (C) and in leg discs where expression of UAS-CG10440-HA is driven by *ptc-GAL4* and monitored by anti-HA immunostaining (green). (D-D’’). In the ptc:CG10440-HA domain, no significant changes in dAP-2 expression are detected in either presumptive joints or interjoint regions.
**Fig. 3-3 Leg defects caused by CG10440 overexpression.** (A) First leg of driver-only fly was used as wild-type control. (B-D) First legs of flies overexpressing CG10440 by Dll\textsuperscript{em212}-GAL4 (B), tubP-GAL4 (C), and btd-GAL4 (D) showed defects including fusion of joints and shortening of interjoint regions correlated to the driver expressing domain and strength.
3-3.3. CG10440 negatively regulates dAP-2 target genes

Although dAP-2 and CG10440 do not appear to regulate each other’s expression in larval tissues, the leg defects caused by CG10440 overexpression are reminiscent of those seen in dAP-2 hypomorphic mutants. Therefore, it seemed reasonable to ask whether CG10440 over-expression had any affects on candidate target dAP-2 genes.

I chose three genes, *ana*, *Obp99A*, and *Cyp4p2*, shown in our laboratory to be significantly down-regulated in dAP-2 mutant leg discs relative to wildtype in a microarray analysis (MacIver et al, in preparation). CG10440 was over-expressed in larvae using the UAS-GAL2 system. UAS-CG10440 flies were crossed to two different GAL4 driver lines. Leg discs were dissected from 3rd instar larvae and total RNA was prepared. Transcript levels were determined by semi-quantitative RT-PCR to assess effects of CG10440 over-expression on candidate dAP-2 target genes. I found all three candidate target genes showed decreased transcript levels when the *Dllem212-GAL4* driver was used and even stronger reduction for *tubP-GAL4* (Fig. 3-4A). In situ hybridization was used to assess whether the target genes were downregulated cell-autonomously. In wild-type leg discs, *ana* and *Obp99A* are each expressed in two incomplete rings. *Dllem212-GAL4* driven over-expression of CG10440 only affected *ana* and *Obp99A* in part of their expressing domains suggesting that one of the two rings of expression for each of these genes is within expressed the Dll expression domain, while the other is not. In addition, *ana* and *Obp99A* expression in distal antenna was also down-regulated by CG10440 overexpression (Fig. 3-4.B-I). My results confirmed that the negative regulation of dAP-2 target genes by CG10440 was cell-autonomous.
Fig. 3-4 Candidate dAP-2 target genes are negatively regulated by CG10440 over-expression. (A). RT-PCR of leg disc total RNA showing effects of GAL4-driver mediated over-expression of UAS-CG10440 on transcript levels of GAPDH, dAP-2, CG10440 and three candidate dAP-2 target genes ana, Obp99A and Cyp4p2. Leg discs were dissected from late third instar larvae carrying either driver only (yw = yw; tubP-GAL4 wild-type control larvae) or carrying UAS-CG10440 transgene together with either Dllem212-GAL4 (dll) or tubP-GAL4 (tub) driver transgenes directing restricted or ubiquitous GAL4 expression, respectively. Transcript levels for dAP-2 and housekeeping gene GAPDH (positive control) showed no change; levels for CG10440 increased in response to specific drivers; levels for three candidate dAP-2 target genes decreased in driver-dependent manner. (B-E) RNA in situ hybridization to detect ana transcripts in leg (B,C) and antennal discs (D,E) from wild-type (wt) or Dllem212-GAL4; UAS-CG10440 larvae (Dll:CG10440). (F-I) RNA in situ with Obp99A probe. Arrows (B-I) mark the missing distal domains of ana and Obp99A expression caused by CG10440 over-expression in Dll-expression domains.
3-3.4. P-element mediated mutagenesis of CG10440

To test if CG10440 was necessary for *Drosophila* leg development, I carried out loss-of-function experiments. Since no CG10440 mutants were available, I attempted to generate CG10440 mutant alleles by P-element transposition. The closest P-element insertions are about 40-50kb away from CG10440. P-element insertions in the adjacent EGFR gene were not chosen for the screen because their analysis would be complicated by the fact that EGFR is known to be essential for leg development (Campbell, 2002). A better candidate, P\{w+\}^{GT1}^{BG01317}, from the BDGP gene disruption project, was located about 50kb downstream of CG10440 in *cricklet (clt)*. This line is homozygous viable and had no known mutant phenotype, and pGT1 is a gene-trapping P-element that can be used as a GAL4 driver if integrated downstream an exon (Lukacsovich et al., 2001).

The local transposition was induced by transposase gene from Δ2-3(99B),Sb line and selected by darker red-eye color indicating multiple copies of P-element insertions (Fig. 3-5). 500 darker red-eye color G2 flies were used to cross single non-Sb male with multiple Adv/CyO females. The progenies showing variation in eye-color were discarded due to the possibility of interchromosomal transposition or recombination of two insertions. To confirm the insertion did disrupt CG10440 gene, genomic PCR was used to detect P-element inserted within ~±1kb region flanking promoter or coding sequences. A pool of 10 flies from 10 vials was firstly used to expedite the PCR procedure. Only one pool showed positive 900bp PCR product with the primer 1kb downstream of promoter (data not shown). It indicated the insertion was located within the first exon, which was consistent with the observation that most local-transposition insertions were located within the promoter regions (Timakov et al., 2002). Unfortunately, later we could not
detect further positive insertion from the stabilized flies in any single vial. We speculated that the positive insertion we had once might be a rare case and reversion happened later. More crossings could be set up to repeat this experiment.
Fig. 3-5 Scheme for P-element mediated mutagenesis. The crossing scheme of P-element transposition to generate CG10440 mutant is shown above. The proposed new insertion P\{w+\}* in CG10440 gene is shown below with primers for genomic PCR indicated.
3-3.5 Knockdown of *CG10440* by RNA interference

The second approach to knock down *CG10440* expression was using RNA interference. During last several years, RNA interference was used as a convenient method to provide gene knock-down in plants and animals. When combined with GAL/UAS system, cell type-specific knock-down can be easily achieved in transgenic animal. Recently, the efficiency of RNAi has been improved by inserting a functional splicing intron between two inverted repeats and using quick cloning P-element vector pWIZ for *Drosophila* RNAi (Lee and Carthew, 2003; Reichhart et al., 2002). I chose a 695bp dsRNA fragment targeting coding sequence to make sure high efficiency of gene-silencing. Both head-head and tail-tail constructs were made to prevent cryptic splicing from disrupting hairpin RNA. Multiple transgenic lines were obtained using either construct and showed uniform phenotype, confirming the correct loopless hairpin RNA was generated.

RT-PCR data confirmed that the *CG10440* expression was knocked down by RNAi constructs (Fig. 3-6A). Adult flies with induced RNAi expression had difficulties in eclosing, due to necrotic joint defect (Fig. 3-6 B-E). Only distal joint necrosis was observed when using distal leg-specific driver *Dll^{em212}-GAL4*, and *tubP-GAL4* could induce necrosis in every leg joint. The RNAi effect was not always penetrate in every leg or every joint, but more severe when using stronger driver or higher temperature. Although no severe leg shortening was observed in CG10440 RNAi induction, the necrotic joint was reminiscent of one of dAP-2 gain-of-function phenotype (Monge et al., 2001).
Fig. 3-6 CG10440 RNAi knock-down. (A) RT-PCR using 3rd instar larval leg disc total RNA showed CG10440 transcription was knocked down by tubP-GAL4 driven pUAS-CG10440^{W} (tub:40^{W}). yw flies were used as wild-type. GAPDH was used as control. (B,D) wild-type first leg (B) and second leg (D). (C, E) necrotic joints were frequently seen in distal leg of Dll^{pm212}: CG10440^{W} flies. Arrows in first leg (C) and second leg (E) indicate necrotic joints.
3-4 Discussion

I have studied the expression and subcellular localization of CG10440, the sole Drosophila protein shown to interact with dAP-2 in a yeast two-hybrid protein-interaction screen (Giot et al., 2003). The highly similar expression patterns of CG10440 and dAP-2 during embryogenesis suggested they might have cooperation in regulation and function (Fig. 3-1). The subcellular localization of HA-tagged CG10440 revealed that the CG10440 might have both plasma membrane and nuclear distribution. The weak nuclear localization was consistent with its predicted interaction with several nuclear proteins, such as dAP-2 and lwr. It is possible that CG10440 was mostly “docked” on the cytoplasmic side of plasma membrane, and upon activation or cleavage, small amount of entire protein or at least C-terminal part was released and entered nucleus. This kind of expression and activation have been observed in many membrane proteins, such as Notch receptor.

I have shown that CG10440 and dAP-2 did not seem to regulate the expression of each other (Fig. 3-2). Instead, I surprisingly found the overexpression of CG10440 resulted in leg defects like dAP-2 mutant does, and expressions of dAP-2 target genes downregulated in dAP-2 mutant were also reduced (Fig. 3-3, 3-4). Considering the in vitro yeast two-hybrid interaction data, I proposed that CG10440 might work as a co-repressor of dAP-2 when activated. Some direct target genes, such as ana and Obp99A, are temporally and tissue-specifically activated by dAP-2 with co-activator, as these genes are expressed in a subset of dAP-2 expressing cells. At certain developmental stage, CG10440 is activated and enters nucleus to replace co-activator or switch activating
complex to repressing complex as a co-repressor, therefore shuts down the expressions of these direct target genes. Indirect target gene, such as Cyp4p2, can be also downregulated via an intermediate target gene.

The other possible scenario is that CG10440 could regulate the stability of dAP-2. Although I have not seen dAP-2 expression pattern and transcript level change in CG10440 overexpression, the resolution of leg disc antibody staining may not be enough to detect protein level change in nucleus. Instead of changing subcellular localization like AP-2 gamma by protein-protein interaction with Wwox, dAP-2 protein could possibly bind CG10440, which was also predicted to interact with SUMO-conjugating enzyme lwr, to go through protein modification or degradation (Aqeilan et al., 2004; Lehembre et al., 2000). It would be interesting to do co-immunoprecipitation (co-IP) using cultured cell lines expressing dAP-2, CG10440 and lwr to see if the protein complex precipitated includes these proteins. The co-IP experiment using leg imaginal disc overexpressing CG10440HA also could be an in vivo confirmation of dAP-2-CG10440 interaction.

Either by affecting dAP-2 target genes only, or by affecting dAP-2 protein stability, CG10440 loss-of-function and gain-of-function data both showed influence on leg development. Without leg shortening phenotype as in dAP-2 gain-of-function experiment, loss of CG10440 only resulted in necrotic joint. The reason that dAP-2 ectopic expression causes shortened leg is misexpressed dAP-2 in the interjoint region disrupts the boundary between dAP-2 expressing and non-expressing cells (Monge et al., 2001). The necrosis phenotype in the leg joints shared by dAP-2 gain-of-function and CG10440 loss-of-function is not understood yet. It is possible that upregulated dAP-2 function by loss of CG10440 promotes cell proliferation and/or cell-death. More
interestingly, searching for CG10440 homologs, along with families of K+-channel proteins, also returns human tumor necrosis factor alpha-induced protein 1 (TNFAIP1) with 27% identities and 65% positives. The TNFAIP1 is a membrane ligand induced by tumor necrosis factor alpha to promote angiogenesis, a hallmark of chronic inflammation (Pandey et al., 1995).

As I have obtained CG10440 RNAi lines, tests on dAP-2 target genes can be done to reveal how CG10440 interacts with dAP-2 and affects *Drosophila* leg development.
Bibliography


## Appendix A

**List of transgenic lines**
(by the order of their appearance in the thesis)

<table>
<thead>
<tr>
<th>Stock name</th>
<th>#</th>
<th>Expression pattern /Phenotype in leg</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4-taulacZ</td>
<td>4</td>
<td>1 proximal coxa ring</td>
<td>Strong in dorsal half</td>
</tr>
<tr>
<td>B12-GAL4</td>
<td>5</td>
<td>Proximal and distal rings co-localized with dAP-2</td>
<td>Media-lateral brain signal</td>
</tr>
<tr>
<td>alacZ</td>
<td>10</td>
<td>-</td>
<td>a.k.a. ALP (-207~+424 of dAP-2 1a)</td>
</tr>
<tr>
<td>BXE-alacZ</td>
<td>13</td>
<td>4 strong tarsal rings</td>
<td></td>
</tr>
<tr>
<td>BXE-blacZ</td>
<td>5</td>
<td>Very weak spotty distal exp.</td>
<td>(-190~+154 of dAP-2 1b)</td>
</tr>
<tr>
<td>BXE-hlacZ</td>
<td>4</td>
<td>Very weak spotty distal exp.</td>
<td>hsp70TATA promoter</td>
</tr>
<tr>
<td>B6-lacZ</td>
<td>11</td>
<td>4 strong tarsal rings + 1 weak proximal coxa ring</td>
<td>Media-lateral brain signal</td>
</tr>
<tr>
<td>dBX200-lacZ</td>
<td>9</td>
<td>4 strong tarsal rings</td>
<td>Deletion of conserved peak #2 in the context of B6</td>
</tr>
<tr>
<td>dBX600-lacZ</td>
<td>9</td>
<td>4 strong tarsal rings</td>
<td>Deletion of conserved peak #3 in the context of B6</td>
</tr>
<tr>
<td>dXE680-lacZ</td>
<td>6</td>
<td>Very weak distal exp. + 1 incomplete proximal ring,</td>
<td>Deletion of conserved peak #4-6 in the context of B6</td>
</tr>
<tr>
<td>dBX500-alacZ</td>
<td>8</td>
<td>Very weak most distal exp.</td>
<td>Deletion of conserved peak #1 in the context of BXE</td>
</tr>
<tr>
<td>dXE500-alacZ</td>
<td>4</td>
<td>Very weak distal exp.</td>
<td>Deletion of 5’ 500bp of XE in the context of BXE</td>
</tr>
<tr>
<td>XE680-alacZ</td>
<td>24</td>
<td>Weak diffused distal exp.</td>
<td></td>
</tr>
<tr>
<td>BXE2S-alacZ</td>
<td>14</td>
<td>Weak incomplete distal rings</td>
<td>a.k.a. sBE1.0</td>
</tr>
<tr>
<td>BXE2SA-alacZ</td>
<td>9</td>
<td>4 strong distal rings</td>
<td>a.k.a. sBE1.7</td>
</tr>
<tr>
<td>BXEmS1-alacZ</td>
<td>8</td>
<td>Very weak distal exp.</td>
<td></td>
</tr>
<tr>
<td>BXEmS2-alacZ</td>
<td>5</td>
<td>Very weak distal exp.</td>
<td></td>
</tr>
<tr>
<td>BXEmA1-alacZ</td>
<td>6</td>
<td>Strong tarsal rings</td>
<td></td>
</tr>
<tr>
<td>BXEmA2-alacZ</td>
<td>5</td>
<td>Weak tarsal rings</td>
<td></td>
</tr>
<tr>
<td>UAS-CG10440HA</td>
<td>9</td>
<td>Leg shortening, loss of joints</td>
<td></td>
</tr>
<tr>
<td>UAS-CG10440W1</td>
<td>10</td>
<td>Necrotic joints</td>
<td></td>
</tr>
</tbody>
</table>
APel | 5 | Diffused distal exp. | (-473~+31 of \textit{dAP-2 1a})
| | | | data not shown in the thesis

Table A1. List of transgenic lines
Appendix B

Regulation of dAP-2 by Bowl

In search of additional regulators of dAP-2 expression, I examined genetic relationships between dAP-2 and several other genes with roles in regulating distal leg development and joint formation. One gene of special interest was bowl, encoding a C2H2 zinc finger repressor shown to function downstream of Notch and upstream of bab2. In the leg disc, the two domains of bowl expression flank the proximal and distal boundaries of bab2 expression suggesting that Bowl restricts bab2 expression to presumptive tarsals 2-4 where it is required for joint formation and segment elongation (de Celis Ibeas and Bray, 2003; Hao et al., 2003).

Like dAP-2, bowl is expressed in a Notch-signaling dependent manner in the developing leg and is required for formation of tarsal joints and elongation of tarsal segments. In contrast to dAP-2, which is expressed in or directly adjacent to all leg segment boundaries excluding the t5/pretarsus boundary, bowl is expressed at the latter boundary and at all proximal leg segment boundaries, but is excluded from tarsal:tarsal segment boundaries. Using the UAS/GAL4 system, I ectopically expressed Bowl using the Dll-GAL4 driver and found that mis-expression of Bowl in the distal leg region caused partial disruption of dAP-2 expression in tarsal joints (Fig. B1-A-A’’) and partial loss of joints in adult legs (Fig. B1-B). de Celis and colleagues have reported that even when bowl mRNA is strongly over-expressed, very little Bowl protein is detected in late-
third-instar/prepupal tarsal leg discs, suggesting that Bowl is not efficiently translated or is unstable in these leg regions (de Celis Ibeas and Bray, 2003). This may explain why dAP-2 was only partially disrupted. My data suggest that Bowl may repress \textit{dAP-2} expression in selected cells at joint/interjoint boundaries in the tarsus.

**Fig. B1 Repression of \textit{dAP-2} by Bowl.** (A-A’’). Double immunostaining of \textit{Dll-GAL4,UAS-GFP;UAS-bowl} pupal leg disc to detect dAP-2 (red) and \textit{Dll}-driven GFP (green, to infer where Bowl is being ectopically expressed). Disruptions in \textit{dAP-2} rings (arrows) in the distal region of leg disc suggest Bowl could be a repressor of \textit{dAP-2} (cell autonomy cannot be absolutely determined here as Bowl protein does not accumulate in all cells where it is ectopically expressed; and Bowl antibody was not available). (B) Ectopic Bowl results in partial loss of joints in adult tarsus. This correlates with the disrupted \textit{dAP-2} expression seen in presumptive tarsal joints in A. Ectopic sex combs are a sign of distal expansion of tarsal 1 fate caused by ectopic expression of \textit{bab2} (de Celis Ibeas and Bray, 2003).
Appendix C

Regulation of dAP-2 by Distalless

Beginning during the 2nd instar larval stage, proximal and distal leg segments are first distinguished by differential expression of proximal leg-specific hth and distal leg-specific Dll. Loss of Dll expression in distal Dll expression domain results in sorting-out/separation of the mutant cells from the leg disc epithelia. Similarly, ectopic expression of Dll in the proximal region of the leg disc results in sorting out of Dll-positive cells from the hth domain (Wu and Cohen, 1999). Dll activates distal-specific target genes such as bric-a-brac (bab) and spineless (ss), and also represses activation of proximal-specific genes such as Antp and dac in the distal domain of the leg disc. The importance of Dll for distal development of appendages prompted me to examine whether it played a role in regulating dAP-2 expression.

I found that when Dll null mutant clones were induced in 2nd instar larva, dAP-2 was ectopically induced in clones located in the distal part of the leg disc (future tarsals) where Dll function is known to be required, but not in Dll- clones in the tibia (where Dll is not expressed in any case). Despite this difference, cells in Dll- clones in the tibia still sorted out from the adjacent wildtype epithelia similarly to how Dll- clones in tarsal regions (Fig. C1-A-A’’’). Ectopic expression of Antp was seen to coincide with ectopic dAP-2 expression in Dll- tarsal clones (Fig. C1-B-B’’’). Dll- clones in the proximal femur dAP-2 domain did not affect dAP-2 expression there, but did while adjacent
trochanter clones induced ectopic dAP-2 expression (Fig. C1-D-D’’’). When Dll clones was induced in mid 3\textsuperscript{rd} instar stage, no influence on dAP-2 expression was observed (Fig. C1-C-C’’’). These results suggested a spatial and temporal regulation of dAP-2 by Dll. In tarsal segments, Dll represses Antp through ss, which is transiently expressed in distal leg during 2\textsuperscript{nd} instar stage and required for tarsal development (Duncan et al., 1998; Emerald and Cohen, 2004). 

Dll-/− tarsal clones induced at 2\textsuperscript{nd} instar stage expressed ectopic dAP-2 as well as Antp, while later Dll-/− clone induction failed to affect dAP-2 expression. This suggests Dll regulated dAP-2 indirectly via a transient mediator, possibly ss which in turn repressed Antp. Dll mutant clones were also frequently seen sorted out to form vesicles (Campbell and Tomlinson, 1998; Wu and Cohen, 1999). Ectopic dAP-2 expression was only seen in some of these cell fate transformation within endogenous Dll domains (arrows in Fig. C1-B’’’’,D’’’’). In the femur and tibia, Dac is the major repressor of Antp, and Dll is absent in these regions. Dll-/− cells in the distal tibia sorted-out (arrowhead in Fig. C1-A), but did not induce dAP-2 expression. Perhaps this is due to Dac down-regulation of Antp. Dll-/− clones did not affect dAP-2 expression in the proximal femur dAP-2 domain, which overlaps with the Dll proximal domain. However, Dll-/− clones in the trochanter were seen to induce ectopic dAP-2 expression, suggesting there might be a distal trochanter-specific repressor of dAP-2, whose expression depends on Dll and inhibits Antp activation in distal trochanter. This repressor, however, is not activated by Dll or even not expressed in proximal femur cells. On the other hand, my data indicates that the ectopic dAP-2 induction in Dll-/− clones is not likely due to de-repression of Ser as others have suggested (Rauskolb, 2001), because activation of dAP-2 in these clones was cell-autonomous.
Fig. C1 Regulation of *dAP-2* by Distalless. (A-A’’’*) *dAP-2* expression was ectopically induced on inside perimeter of *Dll-/-* clones (marked by loss of beta-galactosidase/green and outlined in white or indicated by white arrow) in tarsal regions, but not in clones in tibia (arrowhead). Cells in Dll mutant clones sort out from the rest of the disc epithelia (visible in DIC panels on right). (B-B’’’*) *Dll-/-* clone in tarsus induced *Antp* (green) and *dAP-2* (red) expression as well as cell sorting-out (arrows). (C-C’’’*) *Dll-/-* clones induced relatively late, at mid 3rd instar larval stage, did not affect *dAP-2* expression. (D-D’’’*) *dAP-2* expression domain in the proximal femur was not affected by 3 small *Dll-/-* clones, but a clone in the ventral trochanter caused ectopic *dAP-2* expression (arrow).
VITA

Jizhong Zou
(jxz181@psu.edu)

EDUCATIONAL BACKGROUND
Ph. D. in Biochemistry, Microbiology and Molecular Biology...............December 2005
The Pennsylvania State University, University Park, PA 16802, USA
Thesis: Cis-elements for Notch-dependent expression and autoregulation of Drosophila AP-2 in developing leg and antenna

Master of Science in Cell Biology and Genetics .............................................July 1999
Peking University, Beijing, China
Thesis: Establishment of C57BL/6J ES cell line by conditioned-medium from rat heart cells expressing LIF factor

Bachelor of Science in Genetics .................................................................July 1996
Nankai University, Tianjin, China
Thesis: Immortalization of cultured rat heart cells

PUBLICATIONS AND PRESENTATIONS


