University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations | Graduate School

2004

IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN DROSOPHILA

Jason Scott Rawlings
University of Kentucky, jrawling@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Rawlings, Jason Scott, "IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN DROSOPHILA" (2004). *University of Kentucky Doctoral Dissertations*. 268.
https://uknowledge.uky.edu/gradschool_diss/268

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Jason Scott Rawlings

The Graduate School

University of Kentucky

2004

IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN *DROSOPHILA*

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Jason Scott Rawlings

Lexington, Kentucky

Director: Dr. Douglas Harrison, Associate Professor of Biology

Lexington, Kentucky

2004

Copyright © Jason Scott Rawlings 2004

ABSTRACT OF DISSERTATION

IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN *DROSOPHILA*

The JAK/STAT pathway is but one of the signal transduction cascades responsible for proper development and homeostasis. Gain-of-function mutations of pathway components are causative agents of several leukemias, highlighting the necessity for proper regulation of signal transduction. *Drosophila* presents an attractive model to study JAK/STAT signaling because mutations in the pathway behave in an analogous manner. Furthermore, the *Drosophila* cascade is much simpler as only one of each component required for activation has been characterized; whereas in mammals, there are many ligands, receptors, 4 JAKs and 7 STATs.

Suppressors of Cytokine Signaling (SOCS) are one family of molecules which regulate JAK/STAT signaling via a negative feedback loop. All SOCS share a distinct modular domain architecture, which we exploited to locate three putative SOCS homologues within the *Drosophila* genome. I present the identification and initial characterization of one of these homologues, *Socs44A*. I show that *Socs44A* is not responsive to or dependent on JAK activity. However, I demonstrate that Socs44A is capable of downregulating JAK/STAT signaling in the developing wing but not in

oogenesis, indicating that its ability to regulate the pathway is tissue specific, a phenomenon observed in the mammalian model.

Signal transduction pathways are integrated at multiple levels. This interplay allows for combinatorial signaling, resulting in a higher order of complexity in the signals that can be received and interpreted by a cell. Well documented are the interactions between the JAK/STAT and the EGFR/MAPK pathways. In this work, I show that Socs44A can genetically interact with, and upregulate, the EGFR/MAPK pathway, analogous to a recent report involving SOCS-3.

Starting with the *Drosophila* genome sequence, I initiated a reverse genetic approach to studying the function of the *Socs44A* locus. During the course of this investigation, I designed and implemented a novel post-processor of the BLAST algorithm, called Multi-BLAST, which facilitates retrieval of multiple domain sequences from public databases. In what would have been the ultimate achievement of this study, I attempted two mutagenesis screens designed to isolate *Socs44A* loss-of-function alleles. Progress on these screens is reported.

KEYWORDS: JAK/STAT pathway, Suppressors of Cytokine Signaling, regulation of signal transduction, Socs44A, *Drosophila* development

Jason Scott Rawlings

July 8, 2004

IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN *DROSOPHILA*

By

Jason Scott Rawlings

Douglas A. Harrison
Director of Dissertation

Peter M. Mirabito
Director of Graduate Studies

Copyright © Jason Scott Rawlings 2004

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Dorctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

DISSERTATION

Jason Scott Rawlings

The Graduate School

University of Kentucky

2004

IDENTIFICATION AND CHARACTERIZATION OF SOCS44A IN *DROSOPHILA*

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences at the University of Kentucky

By
Jason Scott Rawlings

Lexington, Kentucky

Director: Dr. Douglas Harrison, Associate Professor of Biology

Lexington, Kentucky

2004

Copyright © Jason Scott Rawlings 2004

This work is dedicated to my wife Laura, and to my daughter, Chloe

ACKNOWLEDGEMENTS

Throughout my graduate career I benefited from many people, without whom this dissertation would not have come to fruition.

First, I would like to express my appreciation for my advisor, Dr. Douglas Harrison. He took me in, incredibly green behind the ears, and made me the scientist I am today. I thank him for giving me the freedom to fly and providing the net to catch me when I fell. I thank him for his insight, and for teaching me patience and perseverance above all else.

I'd like to thank the other members of committee: Drs. John Rawls, Rebecca Kellum, Sheldon Steiner, and Tae Ji. I greatly benefited from their support and I thank them for helping me stay focused throughout my dissertation research. I would also like to thank Dr. Kellum for the opportunity to rotate in her lab.

I would be remiss if I did not express my deepest gratitude for Dr. Chuck Staben who played numerous roles during my graduate career. Without his efforts, generosity, and kindness, graduation may not have been possible. I thank him for all that he has done for me over the years, for being a mentor, a colleague, and friend.

I want to thank the current members of the Harrison lab and wish them the best in their future endeavors: Travis Sexton, Liqun "Shiney" Wang, and Dr. Susan Harrison. I would also like to thank many past members of the lab: Dr. Gabriela Rennebeck, who was a second mentor to me, Dr. Rongwen Xi, Jennifer McGregor, Mike Ragland, Patty Keating Urso, Amy Gorman, and Dr. Kristin Rosler. I also want to thank Elizabeth Cobern, Amanda Tencza, the "Argonauts" and all others who helped with the EMS screen and Chris Saunders who provided statistical support of my project.

I want to recognize other members of the MCB faculty and staff for their support: Drs. Peter Mirabito, Brian Rymond, Robin Cooper, Randal Voss, Judy Lesnaw, Grace Jones, and Bruce O'Hara. I would also like to thank Bev Taulbee and Stella Matsuzak for all their hard work.

I also want to thank my "brothers in arms" the past and present grad students and others of the BGP for their support, advice, and friendship. Particularly, Brian Thompson, Brian Dillon, Dr. RamaKrishna Badugu, Dr. Young-Dong Yoo, Dr. Jay Dunn, Dr. CK Prakash, Sakshi Pandit, Andy Johnstone, Dr. Meg Hatch, Dr. Rick Holtzberg, Dr. Mark Leis, Tom Badgett, Dr. Srinivas Venkatram, Dr. Neetu Dembla-Rajpal, Albert Ribes-Zamora, Weixi Li, Venu Puram, Martin Striz, James Monaghan, Jeremiah Smith, Amy Samuels, and Sameera Desari.

Most importantly, I am grateful for the support of my family. I thank my parents who instilled in me discipline and determination, without which I could not complete this degree. Special thanks to my wife, Laura, for all her love, support, and encouragement. I also want to thank her for all she's done helping to raise our beautiful F_1, Chloe.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS .... **iii**

LIST OF TABLES .... **vii**

LIST OF FIGURES .... **viiii**

LIST OF FILES .... **ix**

CHAPTER I

BACKGROUND **1**

General signal transduction *1*
The JAK/STAT signal transduction cascade *2*
The JAK/STAT pathway in Drosophila melanogaster *4*
Regulation of signal transduction *5*
The Suppressors of Cytokine Signaling (SOCS) *6*

CHAPTER II

THE *DROSOPHILA* GENOME ENCODES THREE PUTATIVE SOCS HOMOLOGUES **15**

INTRODUCTION 15
RESULTS 16
BLAST analysis identified three putative SOCS genes in the fly genome *16*
Phylogenetic relationships between Drosophila and mammalian SOCS *16*
Cloning of Drosophila SOCS genes *17*
DISCUSSION 19

CHAPTER III

THE ROLE OF *DROSOPHILA* SOCS IN THE JAK/STAT CASCADE **25**

INTRODUCTION 25
RESULTS 27
Expression patterns of Drosophila SOCS *27*
Socs36E but not Socs44A is responsive to the JAK/STAT signal *28*
Socs44A regulation of JAK/STAT signaling in wing development *30*
DISCUSSION 34

CHAPTER IV

***SOCS44A* IS CAPABLE OF INTERACTING WITH THE EGFR/MAPK PATHWAY** **46**

INTRODUCTION 46

RESULTS 48
Interactions between Socs44A and other signaling cascades *48*
Socs44A interacts with the EGFR/MAPK pathway *49*
DISCUSSION 51

CHAPTER V

ATTEMPTS TO ISOLATE A LOSS-OF-FUNCTION *SOCS44A* MUTATION 61

INTRODUCTION 61
RESULTS 62
P-element excision *62*
EMS mutagenesis *64*
DISCUSSION 68
P-element excision mutagenesis *69*
EMS mutagenesis *71*
Recent developments *74*

CHAPTER VI

MULTI-BLAST: DATABASE QUERY FOR MULTI-DOMAIN PROTEINS OR CODING REGIONS 89

INTRODUCTION 89
METHODS AND IMPLEMENTATION 92
Domain definition *92*
Use of the BLAST suite of programs *93*
The Multi-BLAST program *93*
User interface *94*
Results page *95*
RESULTS 96
DISUCSSION 98
SPECIAL ACKNOWLEDGEMENTS 101

CHAPTER VII

CONCLUSIONS AND DISCUSSION 106

The Drosophila genome encodes three SOCS homologues *106*
Role of Socs44A in JAK/STAT signal transduction *108*
Role of Socs44A in EGFR/MAPK signal transduction *110*
Crosstalk between the JAK/STAT and EGFR/MAPK pathways *111*
The role of Socs44A in wing vein development *112*
Other roles for Socs44A in Drosophila development *114*
Closing remarks *116*

CHAPTER VIII

MATERIALS AND METHODS ... 120

Fly strains ... *120*
Cloning of Socs44A ... *120*
Sequencing of the Socs44A EST ... *121*
In situ hybridization ... *121*
Immunological staining of embryos ... *123*
Generation of ovarian clones ... *123*
Misexpression studies ... *124*
Image capture and processing ... *124*
RNAi construction ... *125*
EMS mutagenesis ... *125*
P-element excision mutagenesis ... *126*

REFERENCES ... 128

VITA ... 147

LIST OF TABLES

Table 1-1. The SOCS family of proteins: interactions, regulation, and phenotypes. 14
Table 3-1. Socs44A misexpression phenotypes.. 44
Table 3-2. Misexpression of Socs44A exacerbates the reduced viability of hop heteroallelic mutants. 45
Table 4-1. Socs44A misexpression assay data. 59
Table 5-1. Summary of P-element excision screen. 83
Table 5-2. Summary of EMS mutagenesis screen. 84
Table 6-1. Comparison of Multi-BLAST and BLASTp. 105

LIST OF FIGURES

Figure 1-1. Activation of JAK signaling and subsequent inhibition by SOCS proteins.. 11
Figure 1-2. Regulators of JAK/STAT signaling 12
Figure 1-3. Model for SOCS mediated proteosomal degradation of JAK....................... 13
Figure 2-1. Domain architecture of SOCS.. 21
Figure 2-2. Protein sequence comparison of Drosophila and mouse SOCS................... 22
Figure 2-3. Schematics of Drosophila SOCS cDNAs. 23
Figure 2-4. Nucleotide sequence of Socs44A... 24
Figure 3-1. Socs36E but not Socs44A is expressed similarly to upd in both embryogenesis and oogenesis. 38
Figure 3-2. Socs36E but not Socs44A expression is dependent on JAK activity............ 39
Figure 3-3. Wing misexpression phenotypes generated with the Socs44A transgene. ... 40
Figure 3-4. Socs44A is capable of downregulating the JAK/STAT cascade in the wing. 41
Figure 3-5. Endogenous Socs44A regulates JAK pathway activity. 42
Figure 3-6. Socs36E but not Socs44A regulates JAK activity during oogenesis. 43
Figure 4-1. Socs44A increases activity of EGFR signaling. 54
Figure 4-2. Concurrent misexpression of Socs44A rescues argos misexpression in a dose-dependent manner. 55
Figure 4-3. Phenotypic classes generated by misexpression of argos. 56
Figure 4-4. Socs44A deficiencies enhance argos misexpression phenotypes. 57
Figure 4-5. Socs44A misexpression rescues argos misexpression phenotypes. 58
Figure 5-1. P-element map of the Socs44A region.. 78
Figure 5-2. Scheme for imprecise P-element excision screen. 80
Figure 5-3. EMS mutagenesis scheme.. 81
Figure 5-4. Deficiency map of the 43F – 44B region. 82
Figure 6-1. Multi-BLAST user interface, input screen... 102
Figure 6-2. Multi-BLAST user interface, results page. 103
Figure 7-1. Model for Socs44A involvement in wing vein formation. 119

LIST OF FILES

Rawlings.pdf 12.1Mb

Chapter I

Background

General signal transduction

Cell to cell communication is essential for the orchestrated chain of events leading to proper development of multicellular organisms. Not only must each cell be cognizant of its location, but it must also have the competence to differentiate into the proper cell type. This is accomplished through the utilization of signal transduction cascades, which allow cells to send, receive and respond to extracellular signals. Each cell then must be able to correctly interpret and integrate all of the signals it receives in order to behave properly during the course of development.

Surprisingly, there are relatively few signal transduction cascades compared to the number of tasks they must perform during development and homeostasis. This disparity is overcome by the fact that each cascade can produce multiple cellular responses by modulating either the amplitude or duration of their signal. Higher levels of complexity can be achieved if the integration of multiple signals from multiple cascades is considered. Finally, the cellular context in which a signal is received also contributes to the complexity of signaling.

The largest class of signal transduction cascades are those that have a cell surface receptor, which include the Transforming Growth Factor β, Receptor Tyrosine Kinases, Notch, Wingless, and the JAK/STAT pathways (for reviews, see Dierick and Bejsovec, 1999; Nilson and Schupbach, 1999; Rawlings et al., 2004; Schweisguth, 2004; Shi and Massague, 2003). While these cascades are distinct in the particular signals they

recognize, they behave in a similar fashion. A cell surface receptor (or receptors) comes into contact with its ligand, resulting in activation of the pathway. The signal then may be propagated and/or amplified by intermediates and/or second messengers before reaching the nucleus, where target gene expression is either increased or decreased.

The JAK/STAT signal transduction cascade

In vertebrates, the JAK/STAT (Janus Kinase/Signal Transducer and Activator of Transcription) signal cascade is an integral intracellular response mechanism involved in many essential biological processes including cell proliferation, differentiation, and apoptosis (reviewed in Igaz et al., 2001; O'Shea et al., 2002). Furthermore, the JAK/STAT cascade is utilized in many tissues in multiple stages of development. Discovered about 15 years ago, this pathway has been shown to be responsive to a wide array of extracellular ligands, including interferons, interleukins, growth hormone, and erythropoietin (see reviews by Imada and Leonard, 2000; Rawlings et al., 2004; Schindler, 1999; Yeh and Pellegrini, 1999) in addition to some receptor tyrosine kinases, including epidermal growth factor receptor (EGFR) and platelet derived growth factor receptor (PDGFR). Due to its virtual omnipresence in development, abnormal JAK/STAT signaling often leads to deleterious, if not lethal effects, with the majority of these being defects in hematopoiesis (reviewed in Bowman et al., 2000; Coffer et al., 2000; Ward et al., 2000). Severe combined immunodeficiency (SCID) is the result of improper differentiation of lymphocytes. This differentiation is dependent on the JAK/STAT cascade; patients with SCID have been found to have mutations in either the shared interleukin receptor gamma chain or JAK-3 (Leonard, 1996). Conversely, the

fusion of JAK-2 to the TEL ETS transcription factor is but one clinical example that results in constitutive pathway activity. The consequence is either chronic myeloid leukemia (CML) or acute lymphoblastic leukemia (ALL), the most prevalent childhood form of the disease (Ho et al., 1999; Lacronique et al., 1997; Peeters et al., 1997). Additionally, these leukemias have been reported in patients where STATs are constitutively activated, although a direct causal link has not been established (Gouilleux-Gruart et al., 1996).

The JAK/STAT pathway is a simple cascade, in which extracellular signals can be rapidly transduced to the nucleus to elicit a cellular response in the form of altered transcription of target genes (Fig. 1-1). The pathway is initiated by ligand induced dimerization or multimerization of transmembrane receptors. While these receptors have no intrinsic catalytic activity, they are constitutively associated with a family of tyrosine kinases, the Janus Kinase (JAK) proteins. Dimerization of the receptors juxtaposes their associated JAK proteins, allowing the JAKs to transphosphorylate and activate each other. Activated JAKs also phosphorylate tyrosine residues on the receptors, providing docking sites for SH2 domain containing molecules, such as STAT (Signal Transducer and Activator of Transcription). STAT molecules are latent, cytosolic transcription factors, that when bound to the receptor, can be phosphorylated by the JAK proteins. Activated STATs can homo- or hetero-dimerize and translocate to the nucleus where they directly bind DNA and modulate transcription of target genes.

The JAK/STAT pathway in Drosophila melanogaster

The JAK/STAT pathway is conserved across many species, even insects; however, *Drosophila melanogaster* is the only invertebrate model organism that possesses a complete JAK/STAT pathway that functions similarly to what is found in mammals (Barillas-Mury et al., 1999; Hombria and Brown, 2002; Rawlings et al., 2004; Zeidler et al., 2000). Although *C. elegans* and *D. discoidium* both possess STAT-like molecules, they lack JAK homologues and they use STATs in a different manner than do higher eukaryotes (Dearolf, 1999; Kawata et al., 1997). Because of its relatively simplistic version of the JAK/STAT pathway, and the powerful tools for genetic manipulation that the species offers, we have chosen to study JAK/STAT signaling in *Drosophila.*

The JAK/STAT pathway was discovered in *Drosophila* through the molecular characterization of the *hopscotch* gene, which was recovered in a screen for maternal effect lethal mutations and shown to have defects in embryonic segmentation (Binari and Perrimon, 1994; Perrimon and Mahowald, 1986). Interestingly, the other two well-characterized components of the pathway in flies, *Stat92E* and *unpaired*, were identified based on the fact that they share this same embryonic segmentation defect (Harrison et al., 1998; Hou et al., 1996). The receptor for the cascade, *domeless,* was identified in two separate screens, one was looking for mutants with posterior spiracle defects (Brown et al., 2001), the other isolated the same gene in a screen for suppressors of *upd* misexpression in the eye (Chen et al., 2002a). While *hopscotch* and *Stat92E* both share specific sequence similarity to their mammalian counterparts, both *unpaired* and *domeless* are unlike any specific cytokine or cytokine receptor, respectively.

As in mammals, the *Drosophila* JAK/STAT pathway is involved in a broad array of biological processes. In addition to its roles in embryogenesis, the pathway is responsible for follicle cell fate decisions in oogenesis (McGregor et al., 2002; Xi et al., 2003), proliferation of larval hematopoietic lineages (Luo et al., 1997), the formation of larval melanotic tumors, and hypertrophy of larval lymph glands (Harrison et al., 1995). Furthermore, the JAK/STAT pathway has also been implicated in sex determination (Jinks et al., 2000), multiple processes in eye development (Luo et al., 1999; Zeidler et al., 1999), wing vein formation (Yan et al., 1996a), stem cell maintenance during spermatogenesis (Brawley and Matunis, 2004; Kiger et al., 2001; Tulina and Matunis, 2001), and patterning of other adult structures (Harrison et al., 1995; Zeidler et al., 2000).

Regulation of signal transduction

During development, cells must receive and respond to a variety of signals, many of which utilize the same signal cascade. Furthermore, a single signal can elicit a variety of responses; the cell must be able to correctly interpret each signal, or combination of signals, and produce the appropriate response. To achieve this, a signal cascade must be tightly regulated with mechanisms to both activate and deactivate the pathway. Negative regulation of a signal transduction pathway may serve to modulate amplification and thus change the response to a signal, or return the cell to a "reset" state such that it can respond to future signals.

Recent studies have revealed several protein families that regulate the JAK/STAT cascade in several different ways; they include SH2-Bβ, StIP, SHP, PIAS, and SOCS (Fig. 1-2). Both SH2-Bβ and its relative APS interact with JAK, but have opposing roles.

SH2-Bβ is a potent activator, while APS negatively regulates JAK activity (O'Brien et al., 2002). StIP (STAT interacting protein) is a WD40 repeat protein that preferentially interacts with unphosphorylated STAT in addition to activated JAK. It is believed that this protein acts as a scaffold in the recruitment of STAT to the receptor complex (Collum et al., 2000). The SHP (SH2 domain containing phosphatase) family represents a logical means to downregulate the phosphorylation-dependent JAK/STAT cascade since phosphorylation is a rapid and easily reversible protein modification. SHP-1 has been shown to bind to the tyrosine phosphorylated erythropoietin receptor and dephosphorylate JAK2 (Jiao et al., 1996). The PIAS (Protein Inhibitor of Activated STAT) family was identified in a two-hybrid screen for proteins that interact with STAT proteins. PIAS-1 serves as a mechanistic paradigm of this family of proteins; it binds to activated STAT-1, the resulting complex is unable to bind DNA. Interestingly, it seems as though each member of the PIAS family interacts with a specific STAT target (Chung et al., 1997; Liu et al., 1998).

The Suppressors of Cytokine Signaling (SOCS)

The largest and perhaps most versatile family of JAK/STAT negative regulators is the SOCS (suppressor of cytokine signaling) family, with eight mammalian members identified. Three different groups identified the SOCS family of genes concurrently, each giving a different name to their find. Starr and Hilton identified SOCS-1 by virtue of its ability to suppress interleukin-6 induced macrophage differentiation of M1 cells (Starr et al., 1997). Yoshimura's lab utilized a two-hybrid strategy using the kinase domain of JAK-2 as bait. They isolated a protein that they named JAB (JAK binding protein) and

illustrated that it reduces the tyrosine-kinase activity of JAKs and suppresses the tyrosine phosphorylation and activation of STATs (Endo et al., 1997). Finally, Kishimoto's group isolated a protein they named SSI-1 (STAT induced STAT Inhibitor) based upon its immunocrossreactivity to the SH2 domain of STAT-3 and showed that it inhibited the signaling of a variety of cytokines. Furthermore, they showed that SSI-1 transcription was induced by interleukin-4 and interleukin-6 (Naka et al., 1997). Two years prior to these discoveries, Yoshimura's group isolated CIS (Cytokine Inducible Suppressor), a novel gene that was able to inhibit signaling through the Epo and several interleukin receptors and showed that CIS was induced by several cytokines (Yoshimura et al., 1995). In the interest of clarity, we will refer to all of these mammalian genes as SOCS-1 to SOCS-7 and CIS (Table 1-1).

All SOCS proteins possess a similar structure; in addition to a central SH2 domain, these genes also contain a carboxy-terminal motif, termed the SOCS box, as well as an undefined amino terminal region of variable length. The SOCS box is approximately forty amino acids in length and has been found by sequence homology in twenty different proteins belonging to five structural classes (Hilton et al., 1998). Each class is delineated by the identity of the central domain, be it an SH2 domain (SOCS), SPRY domain (SSB), WD-40 repeat (WSB), ankyrin repeat (ASB) or undefined motif.

Current studies in mammals have shown that SOCS proteins may downregulate the JAK/STAT pathway in several ways (Fig. 1-1). All SOCS proteins contain an SH2 domain, which allows for interaction with phosphotyrosine activated components of the JAK/STAT pathway (for review of the SH2 domain in signaling, see Yaffe, 2002). Beyond this, explicit mechanisms of action must be implied from biochemical analyses.

SOCS-1 has been shown to associate directly with all four mammalian JAKs, acting as a pseudosubstrate, inhibiting their catalytic activity (Endo et al., 1997; Naka et al., 1997; Nicholson et al., 1999; Yasukawa et al., 1999). Unlike SOCS-1, CIS does not interact directly with JAKs; instead, CIS associates with the activated Epo receptor at one of two STAT-5 binding sites (Verdier et al., 1998) and competes with STAT-5 for binding to the receptor (Matsumoto et al., 1999). SOCS-3, unlike SOCS-1, does not inhibit JAK-1 or JAK-2 catalytic activity via *in vitro* kinase assay (Nicholson et al., 1999); and unlike CIS, it does not compete with STAT-5 (Ram and Waxman, 2000). However, it does possess a sequence resembling a JAK activation loop similar to SOCS-1 (Yasukawa et al., 1999) and it can bind JAK-1, albeit with a much lower affinity than SOCS-1 (Nicholson et al., 1999). Furthermore, SOCS-3 has been shown to bind to the gp130 cytokine receptor subunit (Nicholson et al., 2000), IL-2Rβ receptor (Cohney et al., 1999), and GH receptor (Hansen et al., 1999). Taken together, these observations suggest that SOCS-3 inhibits JAK/STAT signaling by a different mechanism than SOCS-1 or CIS. Finally, the carboxy-terminal SOCS box has been shown to interact with the elongin BC complex, an interaction that is inhibited by tyrosine phosphorylation of two residues within the SOCS domain (Haan et al., 2003). The Elongin BC complex also interacts with an E3 ubiquitin ligase (cullin-2) that could target bound proteins for proteosomal degradation (Fig. 1-3). . Thus, the SOCS proteins could act as adaptors whose role would be to couple activated JAK/STAT constituents to this elongin BC complex, ultimately leading to their ubiquitination and degradation (Kamura et al., 1998; Zhang et al., 1999).

Phenotypically, SOCS genes parallel what is seen in other JAK/STAT molecules (Table 1-1). Targeted knockouts of SOCS-1 produce runted mice that die at three weeks

of age, suffering from severe reduction in platelet and hematocrit cells, fatty degeneration of the liver and macrophage infiltration of major organs (Alexander et al., 1999; Marine et al., 1999b; Metcalf et al., 1999; Naka et al., 1998; Starr et al., 1998). This is reminiscent of the phenotype seen when IFNγ is administered to neonatal wild-type mice (Gresser et al., 1981). Transgenic mice that overexpress SOCS-3 do not survive to birth (Marine et al., 1999a), but have a phenotype that resembles that of *JAK-2*$^{-/-}$ mice (Neubauer et al., 1998; Parganas et al., 1998). Mice lacking SOCS-2 grew significantly larger than their wild-type littermates; a result from deregulated growth hormone (GH) and insulin-like growth factor-I (IGF-I) signaling (Metcalf et al., 2000).

Very little is known about the remaining four SOCS (SOCS-4 to SOCS-7). Recently, a knockout of SOCS-6 was created; however, these animals showed only mild growth retardation (Krebs et al., 2002). Biochemical analyses suggest that SOCS-6 may regulate insulin signaling (Krebs et al., 2002; Mooney et al., 2001). A single report indicates that overexpression of SOCS-5 disrupted Th2 cell differentiation in culture (Seki et al., 2002). However, a recently developed knockout of the gene had neither no visible phenotype save a mildly increased erythropoiesis, nor did its absence affect T cell differentiation (Brender et al., 2004). To date, there are no functional studies of either SOCS-4 or SOCS-7.

Data mining of the *Drosophila* genome has produced three putative SOCS homologues (Adams et al., 2000). Based on their inferred cytological location on polytene chromosomes, we named these homologues Socs16D, Socs36E, and Socs44A. Of these, Socs36E is the most studied; it shares similar embryonic and ovarian patterns of expression to Unpaired. Furthermore, it shares misexpression phenotypes consistent with

a role in negatively regulating the JAK/STAT pathway (Callus and Mathey-Prevot, 2002; Karsten et al., 2002 and Rennebeck, unpublished). This work will focus on the molecular and functional characterization of Socs44A. This study examines its roles in *Drosophila* development and its functions in JAK/STAT signaling. Furthermore, we will discuss the discovery of an interaction between Socs44A and EGFR signaling.

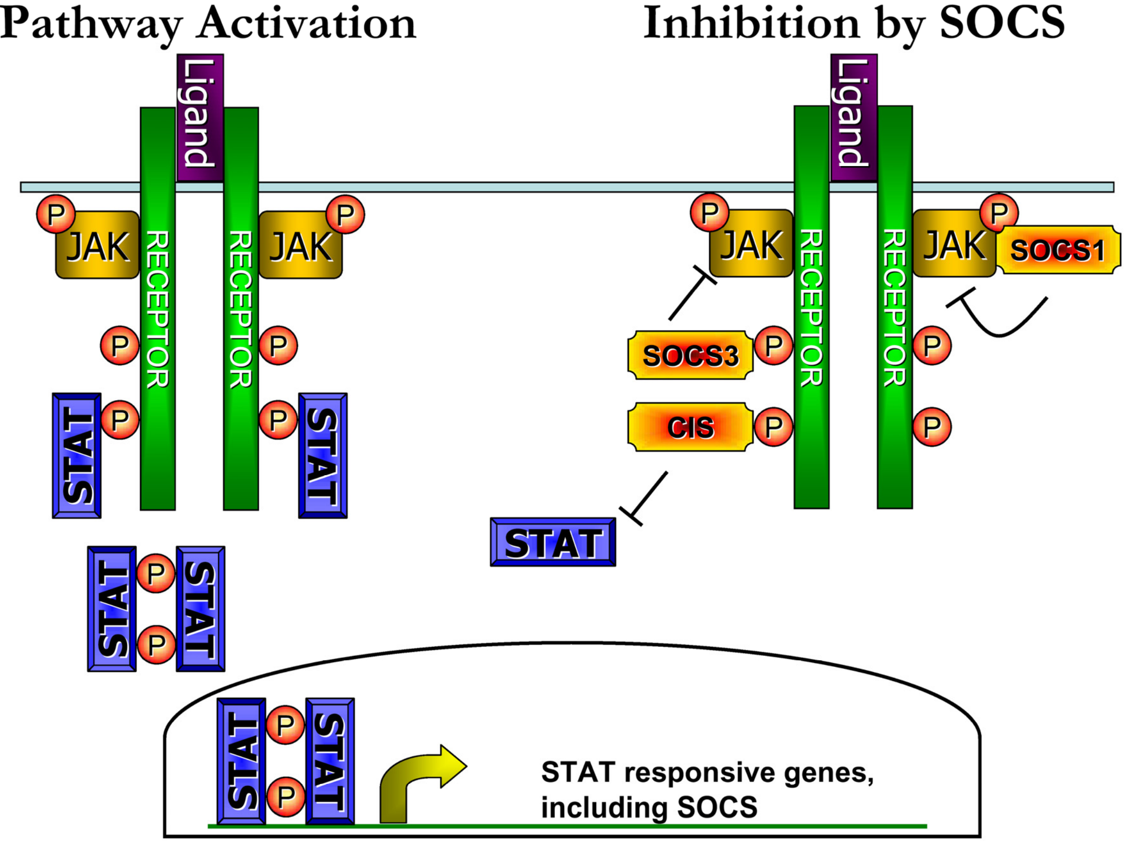


Figure 1-1. Activation of JAK signaling and subsequent inhibition by SOCS proteins. Ligand binding induces receptor dimerization and pathway activation propagated by tyrosine phosphorylation of Janus Kinase (JAK) and Signal Transducer and Activator of Transcription (STAT) intermediates. Activated STATs can dimerize, translocate to the nucleus and control transcription of target genes, including SOCS (Suppressor of Cytokine Signaling). SOCS can then act to inhibit JAK signaling by a variety of mechanisms. See text for more details.

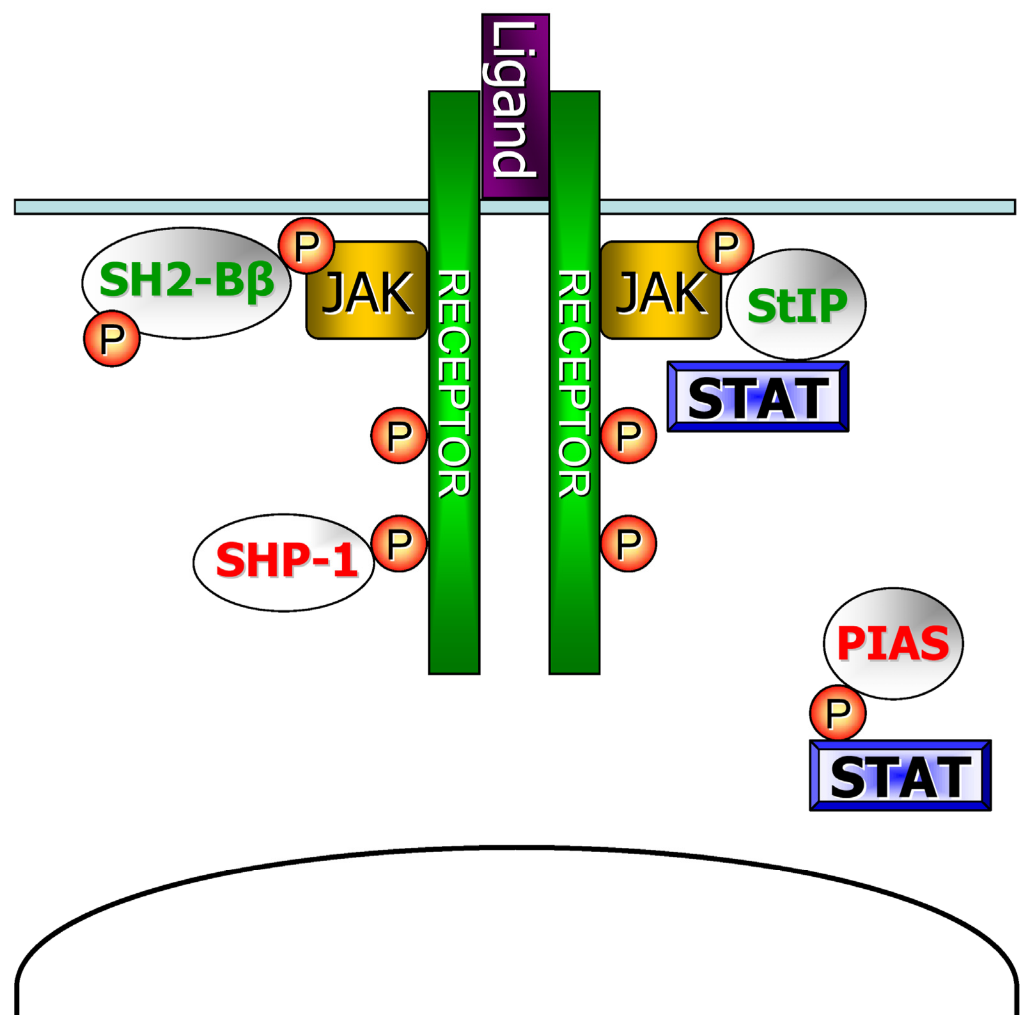


Figure 1-2. Regulators of JAK/STAT signaling. Positive regulators are labeled in green, while negative regulators are labeled in red. Both SH2-Bβ and its relative APS interact with JAK, but have opposing roles; SH2-Bβ is a potent activator, while APS negatively regulates JAK activity. StIP preferentially interacts with unphosphorylated STAT in addition to activated JAK, acting as a scaffold in the recruitment of STAT to the receptor. The SHP family represents a logical means to downregulate the phosphorylation-dependent JAK/STAT cascade since phosphorylation is a rapid and easily reversible protein modification. The PIAS family binds to activated STATs and this complex is unable to bind DNA.

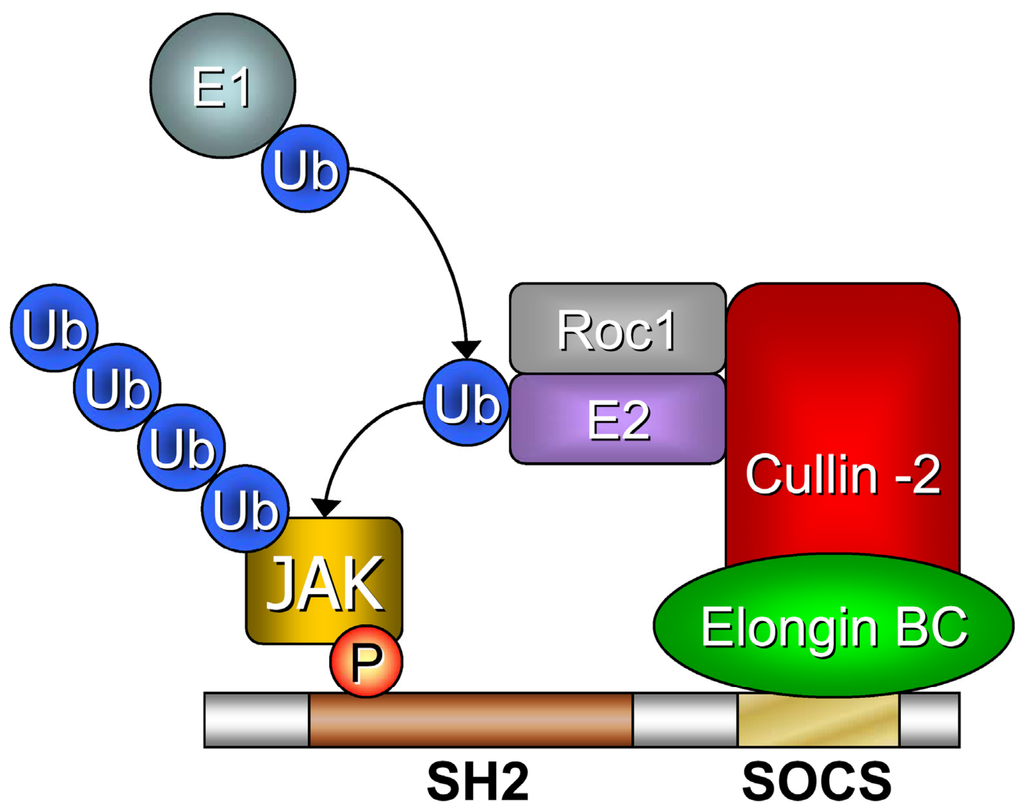


Figure 1-3. Model for SOCS mediated proteosomal degradation of JAK. The SH2 domain allows SOCS proteins to interact with activated components of the JAK/STAT pathway. The SOCS box interacts with the elongin BC complex, which in turn interacts with cullin-2, an E3 Ubiquitin ligase. Roc1 is a RING finger protein that recruits the E2 Ubiquitin conjugating enzyme to the ligase complex (Furukawa et al., 2003; Furukawa et al., 2002). The resulting polyubiquitination of the activated component labels it for degradation by the proteosome.

Table 1-1. The SOCS family of proteins: interactions, regulation, and phenotypes.

Name	Synonyms	Induced by	Regulates	Knockout phenotype
CIS		Interleukins 2, 3, 6, 9 IFN-α, TNF-α EPO, TSLP GH, prolactin, IGF-1	Interleukins 2, 3 EPO GH, prolactin	No abnormalities
SOCS-1	SSI-1, JAB	Interleukins 2, 4, 6, 7, 9, 13 IFN-α/β IFN-γ, LIF, TNF-α EPO, OSM, TPO, TSLP, G-CSF, GM-CSF GH, prolactin, IGF-1, insulin, CNTF LPS, CpG DNA	Interleukins 2, 4, 6, 7, 12, 15 IFN-α/β IFN-γ, LIF, TNF-α EPO, TPO, TSLP GH, prolactin, insulin, leptin LPS, CpG DNA	Neonatal lethality due to deregulated IFN-γ
SOCS-2	SSI-2, CIS2	Interleukin-6 IFN-α, IFN-γ, LIF GH, prolactin, IGF-1 insulin, CNTF, cadiotropin	Interleukin-6 GH, IGF-1	Gigantism due to deregulated GH signaling
SOCS-3	SSI-3, CIS3	Interleukins 1, 2, 6, 9, 10, 13 IFN-α, IFN-γ, LIF EPO, OSM, GM-CSF GH, prolactin, IGF-1, insulin, leptin, CNTF LPS, CpG DNA	Interleukins 2, 4, 5, 9, 11 IFN-α, IFN-γ, LIF EPO GH, prolactin, insulin, leptin	Midgestational lethality due to deregulated LIF and IL-6
SOCS-4	CIS7	Unknown	Unknown	No knockout reported
SOCS-5	CIS6	Interleukin-6	Interleukins 4, 6	Mildly increased erythropoiesis in spleen
SOCS-6	CIS4	Unknown	insulin	Mild growth retardation
SOCS-7	CIS5, NAP4	Unknown	Unknown	No knockout reported

Abbreviations: SOCS = Suppressor of Cytokine Signaling, CIS = Cytokine inducible SH2 containing protein, SSI = STAT induced STAT inhibitor, JAB = JAK binding protein, CNTF = ciliary neurotrophic factor, EPO = erythropoietin, G-CSF = granulocyte colony stimulating factor, GM-CSF = granulocyte-macrophage colony stimulating factor, GH = growth hormone, IFN = interferon, IGF-1 = insulin like growth factor, LIF = leukemia inhibitory factor, LPS = lipopolysaccharide, OSM = oncostatin, PKC = protein kinase C, TNF = tumor necrosis factor, TPO = thrombopoietin, TSLP = thymic stromal lymphopoietin

Chapter II

The *Drosophila* genome encodes three putative SOCS homologues

INTRODUCTION

The eight mammalian SOCS genes can be readily identified by their distinct modular arrangement of domains (Fig. 2-1). Specifically, each member of the family has a central SH2 domain followed closely by a carboxy-terminal SOCS domain. The ~70 amino acid SH2 domain permits interactions with phosphotyrosine residues of activated JAK/STAT components and the ~40 amino acid SOCS domain targets bound molecules for proteosomal degradation. Additionally, SOCS-1 and SOCS-3 share a third functional motif directly preceding the SH2 domain, termed the KIR (Kinase Inhibitory Region). This 24 amino acid motif resembles a JAK activation loop and is believed to allow these SOCS to act as a pseudosubstrate, inhibiting JAK catalytic activity (Narazaki et al., 1998; Nicholson et al., 1999; Yasukawa et al., 1999). The amino termini of mammalian SOCS proteins are divergent in both length and composition. However, SOCS proteins can be divided into two classes based on the length. SOCS-1 to SOCS-3 as well as CIS all are less than 300 amino acids in length, while the remaining SOCS (SOCS-4 to SOCS-7) are all longer than 400 amino acids.

RESULTS

BLAST analysis identified three putative SOCS genes in the fly genome

Using the modular domain architecture of mammalian SOCS as a guide, we searched the *Drosophila* genome for putative SOCS homologues. Specifically, the consensus SOCS sequence (Hilton et al., 1998) was used as a query in a tBLASTn search of the *Drosophila* genome. Open reading frames that contained a SOCS domain were further queried for an upstream SH2 domain. Three putative SOCS were identified using this approach and were named *Socs16D, Socs36E*, and *Socs44A* based upon their cytological position in the genome. All three *Drosophila* SOCS contain the expected arrangement of domains, in that they each have a SH2 domain followed by a SOCS domain located at the carboxy terminus of the locus. None of the *Drosophila* SOCS have sequence in common with the KIR of SOCS-1 and SOCS-3. Like their mammalian counterparts, their amino termini are divergent; however, all of them are at least 350 amino acids in length (Fig. 2-1).

Phylogenetic relationships between Drosophila and mammalian SOCS

A multiple sequence alignment based upon conceptual translations of the *Drosophila* SOCS and their mammalian counterparts reveals that there is high degree of similarity in the carboxy portions of SOCS proteins across taxa, however, the amino termini remain divergent. Socs36E is the most similar to mammalian SOCS, sharing 71% similarity to murine SOCS-5 and 68% similarity to murine SOCS-4 over the last 180 amino acids comprising the SOCS, SH2, and the immediate 30 amino acids preceding the SH2 domain. Socs16D is most similar to murine SOCS-6, sharing 44%

similarity. Socs44A is least similar to mammalian SOCS, but it does share 43% similarity to murine SOCS-6 (Fig 2-2). As expected, these similarities were found solely in the SH2 and SOCS domains, and in the case of Socs36E, the similarity extended into the 30 amino acid region preceding the SH2 domain. However, this sequence shared no similarity to the analogous region defined as the KIR for SOCS-1 and SOCS-3.

From this alignment, we can infer a phylogenetic relationship between these SOCS (Fig 2-2). It appears that there are three distinct clades of SOCS of which there were at least two present at the time of divergence of mammals and dipterans. These two clades include Socs36E which diverged into SOCS-4 and SOCS-5 and the tandem of Socs16D and Socs44A which diverged into SOCS-6 and SOCS-7. The third clade, which comprises CIS and SOCS-1 to SOCS-3, is not present in flies.

Cloning of Drosophila SOCS genes

Two distinct clones for *Socs36E* were recovered in an embryonic cDNA library screen (Fig 2-3 and G. Rennebeck, unpublished). The first clone (designated 2.1.1) contains the expected domain architecture, including complete SH2 and SOCS domains. The second clone (designated 6.3.1) lacks the SOCS domain as well as most of the SH2 domain. Another lab observed two distinct transcripts in a Northern analysis corresponding to *Socs36E*, one consistent with a full length clone (3.5kb), while the other was 2.3kb, smaller than the 2.75kb abbreviated clone recovered in our screen (Callus and Mathey-Prevot, 2002). Our full length clone indicates that the *Socs36E* locus is comprised of four exons and three introns. The 5' most intron is nearly 10kb in length

and contains two clusters of four putative Stat92E binding sites (Fig. 2-3, Yan et al., 1996b).

I PCR amplified an 1133bp fragment comprising the predicted start and stop codons for *Socs44A*. Subsequently, we acquired a larval/pupal derived EST corresponding to the *Socs44A* locus (LP02169). This EST was sequenced in its entirety (Sanger et al., 1977) and deposited into Genbank (Accession AF435923, Fig 2-4). Another EST isolated (GM29526) from ovaries is consistent with the larval/pupal clone. Both ESTs are consistent with the predicted start and stop codons for this locus. The larval/pupal EST contains a 74bp 5' UTR,, 2 exons measuring 557 and 472 nucleotides, and a 681bp 3' UTR (Fig 2-3). This EST is predicted to encode a 342 amino acid protein, approximately 38kD in size. Comparison of the EST sequence to genomic sequence reveals that the locus has but a single 62bp intron. Analysis of all genomic sequences 10kb upstream and downstream failed to detect the presence of any clusters of putative Stat92E binding sites.

The Berkeley *Drosophila* Genome Project (BDGP) recovered 2 ESTs corresponding to the third SOCS homologue, *Socs16D*: an embryonic EST and an EST derived from *Drosophila* S2 cells (Fig. 2-3, FlyBase Consortium, 2003). The S2 cell derived EST (SD11104) contains a larger amino terminus than the embryonic EST (LD08944). Sequence analysis of the embryonic EST clone was inconsistent with the sequence reported by the BDGP. Additionally, I was able to PCR amplify an 827bp segment of the *Socs16D* genomic locus which could be used to isolate a full-length cDNA from a library screen.

DISCUSSION

Based on sequence similarity to a known modular architecture, we identified three putative SOCS homologues in *Drosophila.* Each of these SOCS genes contains the expected arrangement of the SH2 and SOCS domains at their carboxy terminus and an amino terminus of variable length. While the amino termini of SOCS-1 and SOCS-3 are functionally dispensable, this did not extend to the other SOCS family members (Nicholson et al., 1999). Therefore, it is not surprising to find such disparity among SOCS across taxa in this regard.

Phylogenetic analysis revealed that there are perhaps three distinct clades of SOCS genes (Fig 2-2). The four most studied mammalian SOCS genes all belong to a single clade that has no *Drosophila* homologues. This absence may be reflective of this clade arising after the evolution of dipterans, or that this clade was lost in *Drosophila*. Examination the closely related *Drosophila pseudoobscura* genome as well as the recently completed *Anopheles gambiae* genome will be required to address this issue. Or, perhaps this is reflective of SOCS functions that are unique to chordates. Socs36E is most similar to SOCS-4 and SOCS-5 while Socs44A is most similar to SOCS-6, accounting for the other two observed SOCS clades. This would suggest that the functions of Socs36E and Socs44A may be representative of the lesser studied mammalian SOCS genes. The presence of the sole *C. elegans* SOCS homologue in the same clade as Socs44A and Socs16D could suggest that these arose from the “original” SOCS gene and may be representative of a more generalized SOCS.

It is sometimes possible to infer evolutionary decent by analysis of intron/exon structures; however, the *Drosophila* SOCS genes do not share similar intron/exon

structures. *Socs44A* is perhaps the simplest, containing 2 exons and a single, small 62bp intron. There is evidence that both *Socs36E* and *Socs16D* have multiple splice forms. In the case of *Socs36E*, this is supported by isolation of embryonic cDNAs and Northern analysis (Callus and Mathey-Prevot, 2002). ESTs from the BDGP indicate that *Socs16D* may have two splice variants, the only difference between them being the size of the amino terminus of the resulting protein. Both *Socs16D* and *Socs36E* are spliced within the SH2 domain; however, the location of this splice site is not conserved. Both splice junctions occur in a less conserved region of the SH2 domain near the carboxy end. Altogether, these observations are consistent with the divergence of three clades of SOCS proteins.

The general mechanism of SOCS regulation of signal transduction is based on a simple negative feedback loop. SOCS genes respond to JAK/STAT pathway activity and in turn act to downregulate the cascade. The possession of both an SH2 and SOCS domain indicates that *Drosophila* SOCS genes may at least be capable of downregulating the cascade in flies. Furthermore, the presence of two clusters of Stat92E binding sites in the 5' intron of *Socs36E* suggests that it may also be responsive to pathway activity, consistent with the negative feedback loop model. In contrast, the *Socs44A* and *Socs16D* loci and surrounding genomic regions do not contain a similar cluster of Stat92E binding sites, which suggests that they may not be responsive to the JAK/STAT cascade. Nevertheless, these findings provided sufficient motivation for the study of these genes as potential regulators of *Drosophila* JAK/STAT signaling.

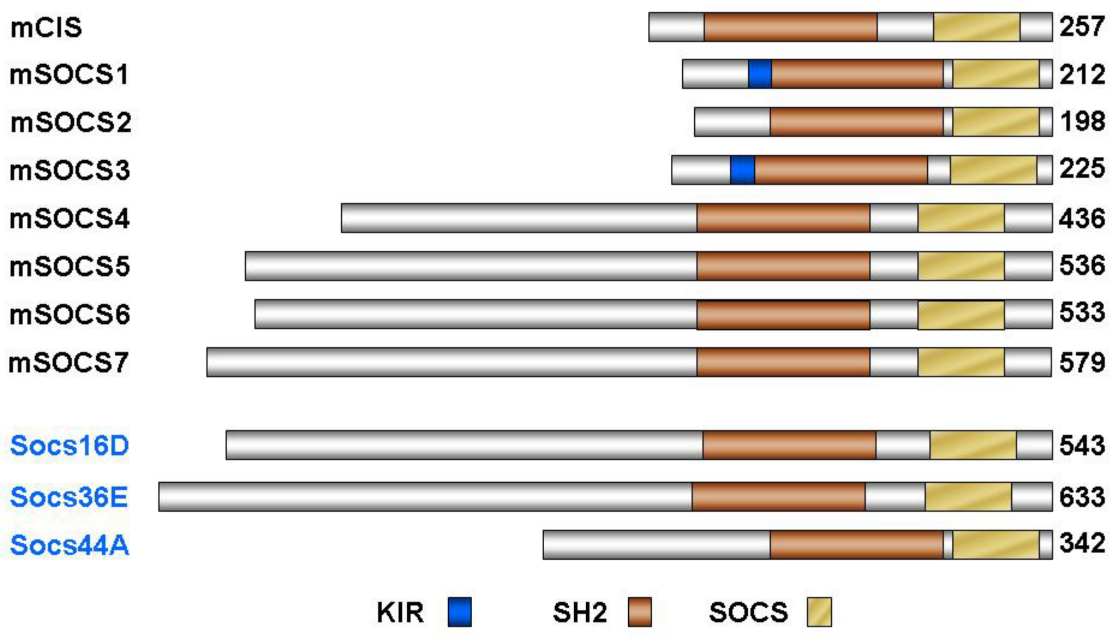


Figure 2-1. Domain architecture of SOCS. Eight mammalian SOCS have been identified and designated CIS and SOCS-1 to SOCS-7. Based on the modular domain architecture shared by these proteins, we identified three putative *Drosophila* SOCS homologues, designated Socs16D, Socs36E, and Socs44A based upon their cytological position in the fly genome. The proteins are depicted to scale and length of each SOCS is indicated to the right and the coloring scheme is defined below the figure.

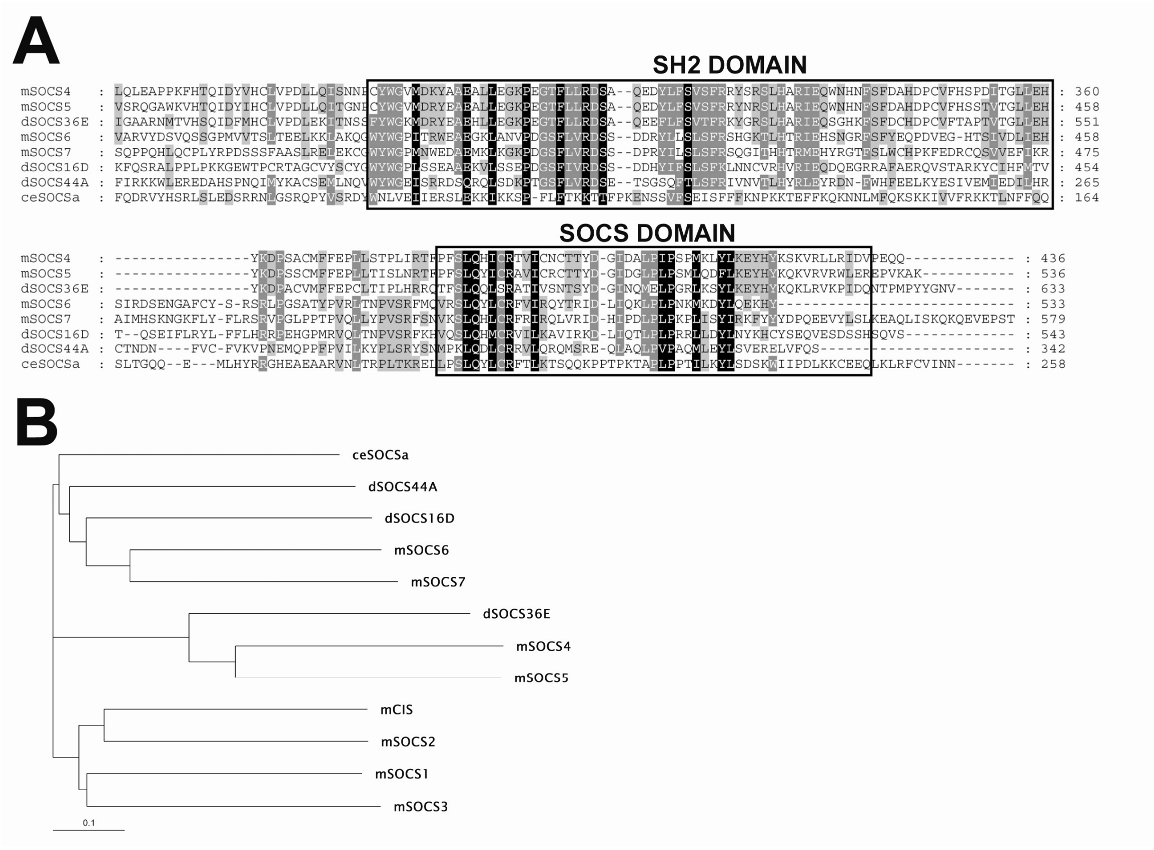


Figure 2-2. Protein sequence comparison of Drosophila and mouse SOCS. (A) The predicted protein sequences of fly and selected mouse SOCS genes are aligned and shaded to indicate similarities and identities. Fly Socs36E is most similar to mouse SOCS-5, while the remaining fly homologues, Socs16D and Socs44A, are more distantly related to mouse SOCS-6 and SOCS-7. (B) Phylogenetic analysis of SOCS genes indicates that there are perhaps three clades of SOCS. *Drosophila* Socs36E is located in the same clade as mouse SOCS-4 and SOCS-5, while *Drosophila* Socs16D and Socs44A share the same clade as mouse SOCS-6 and SOCS-7. The most studied mammalian SOCS (CIS and SOCS-1 to SOCS-3) occupy the third clade.

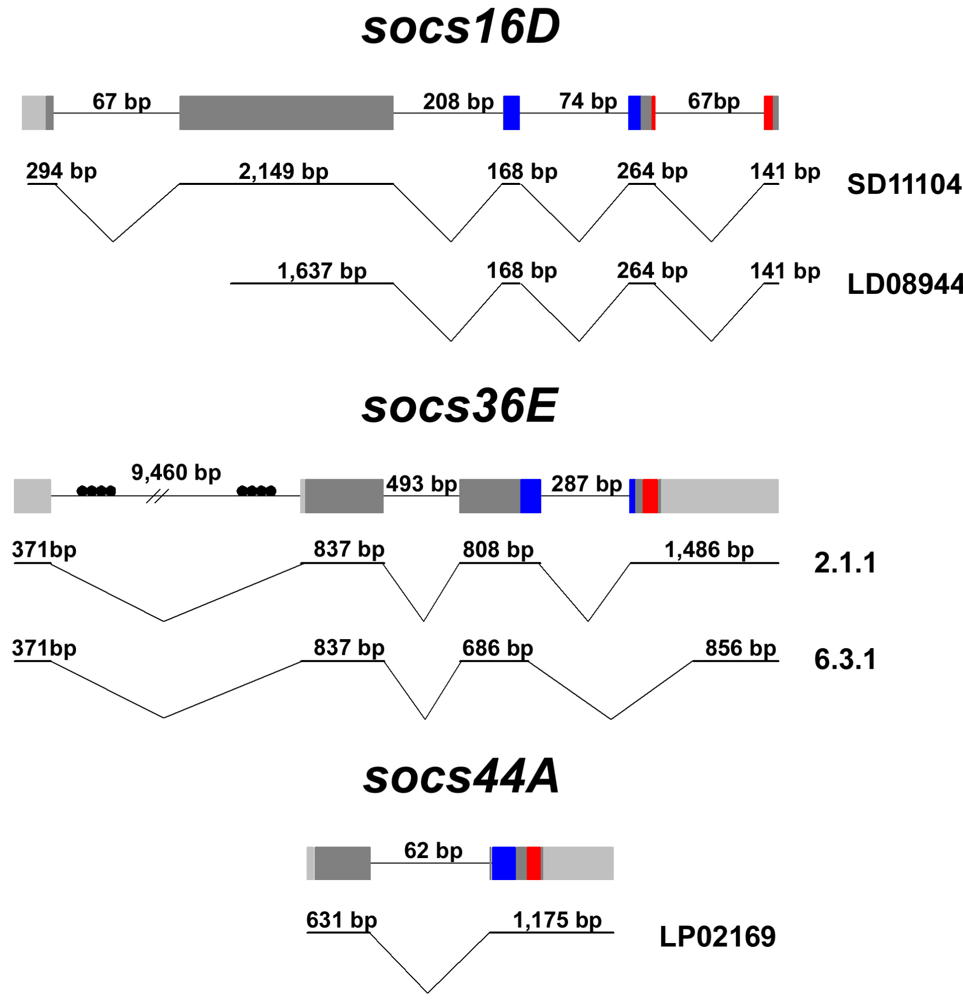


Figure 2-3. Schematics of Drosophila SOCS cDNAs. Transcribed regions are indicated by boxes. The SH2 domain is blue, the SOCS domain is red, while light gray indicates untranslated regions and dark gray indicates undefined regions of the polypeptide. Two potential Stat92E binding sites are indicated by closed circles. The EST or clone reference is indicated to the right. Further information about these clones can be found in the text.

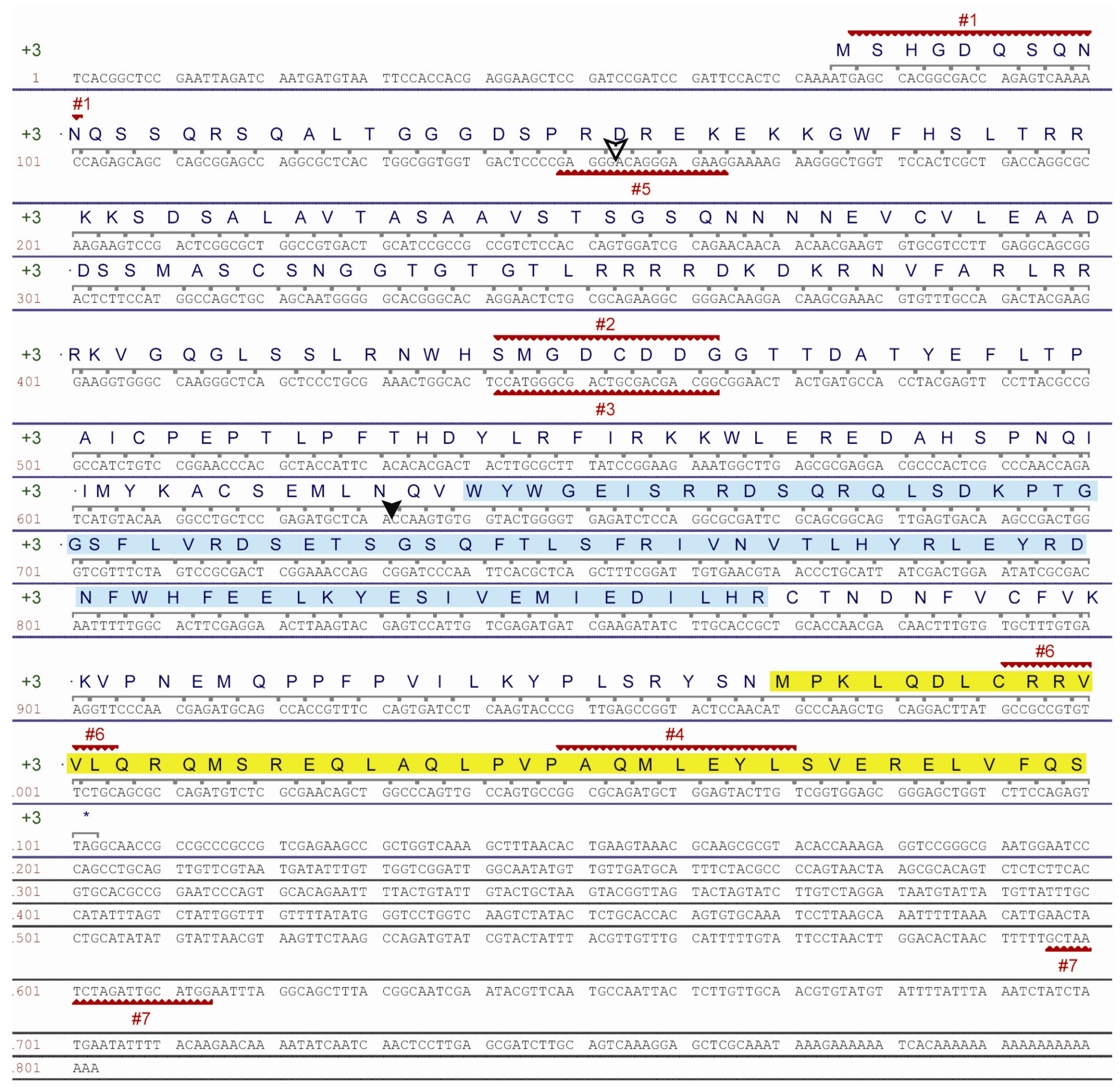


Figure 2-4. Nucleotide sequence of Socs44A. The sequence of the Socs44A EST LP02169 is shown. The conceptual translation of the Socs44A protein is indicated in one letter amino acid code above the corresponding nucleotide sequence. The SH2 domain is shaded blue and the SOCS domain is shaded yellow. Primers used in sequencing are indicated in red. The closed arrowhead indicates the location of the intron (see Fig 2-3). The open arrowhead indicates the location of the *PiggyBac* P-element insertion (see chapter five).

Chapter III

The role of *Drosophila* SOCS in the JAK/STAT cascade

INTRODUCTION

Regulation of signal transduction can serve two purposes. First, it can act to attenuate the amplitude or duration of a signal in order to elicit the appropriate response. Second, the signaling pathway must have a means to reset itself so that the cell can respond to future signals. These regulatory phenomena are required throughout development and homeostasis. A common mechanism for such regulation is the negative feedback loop, which allows a signaling cascade to regulate itself based upon activation of the pathway.

The four well-studied mammalian SOCS genes (CIS, SOCS-1 to SOCS-3) have been shown to operate in a classical negative feedback loop to regulate the JAK/STAT pathway. Cytokines and growth factors induce pathway activation resulting in the modulation of target gene transcription, including SOCS. The expression of CIS, as well as SOCS-1 to SOCS-3 can be induced by a variety of interleukins, interferons, growth hormones, and colony stimulating factors (reviewed in Fujimoto and Naka, 2003). However, the pattern of SOCS induction seems to be tissue or cell-line specific. For example, IFN-γ induces SOCS-1 and SOCS-3 expression in NIH-3T3 cells, but only SOCS-1 in M1 cells (Sakamoto et al., 1998). Likewise, growth hormone can induce CIS, SOCS-2, and SOCS-3 expression in the mouse liver, but only CIS and SOCS-2 are induced in the mammary gland (Davey et al., 1999). The general observation is that multiple SOCS are induced by a single cytokine in a tissue specific manner.

After pathway activation and induction of gene expression, SOCS can act to downregulate JAK/STAT signaling by a variety of mechanisms described previously. CIS, as well as SOCS-1 to SOCS-3 have been shown to inhibit signaling by an array of cytokines. However, there is no pattern to the relationship between the SOCS proteins and the cytokines they inhibit. Considering only the interleukins, CIS is induced by IL-2, IL-3, IL-6, and IL-9; however, it can only inhibit IL-2 and IL-3 signaling (reviewed in Fujimoto and Naka, 2003) Conversely, SOCS-5 is induced by IL-6 but can negatively regulate IL-4 in addition to IL-6 signaling (Heinrich et al., 2003; Nicholson et al., 1999).

Clearly the activity of SOCS is more complex than a "simple" negative feedback loop. The observations mentioned above indicate that there is considerable crosstalk among cytokines via SOCS. With numerous ligands and receptors for the mammalian JAK/STAT pathway, dissection of SOCS function is an arduous task at best, lending credence to the study of SOCS in the relatively simplistic *Drosophila* JAK/STAT system.

This chapter examines the roles of the *Drosophila* Socs36E and Socs44A in the JAK/STAT pathway. Specifically, we show that *Socs36E* expression is dependent on JAK activity and it is responsive to JAK pathway activation. These observations are consistent with a recently published report (Karsten et al., 2002). On the other hand, *Socs44A* failed to respond to JAK/STAT activation. Likewise, a reduction in pathway activity had no effect on *Socs44A* expression. Others showed that Socs36E could negatively regulate the JAK/STAT cascade during wing development (Callus and Mathey-Prevot, 2002). We extend this by illustrating that Socs36E can also negatively regulate the pathway during oogenesis. Socs44A appears to have tissue specificity in its regulation of the pathway. While it is able to regulate JAK/STAT signal transduction in

wing development, it does not regulate the pathway during follicular development in oogenesis.

RESULTS

Expression patterns of Drosophila SOCS

If *Socs36E* and *Socs44A* were responsive to JAK/STAT signaling, we would expect them to have similar patterns of expression to known components of the pathway. The role of the JAK/STAT pathway in proper segmentation of the embryo has been well established (Binari and Perrimon, 1994; Harrison et al., 1998; Hou et al., 1996; Perrimon and Mahowald, 1986; Yan et al., 1996b). It has been shown that the cascade coordinates expression of several of the pair-rule class of genes, including *even-skipped, fushi turazu,* and *runt* (Binari and Perrimon, 1994; Small et al., 1996). The ligand, *unpaired*, is expressed in a distinctive, dynamic pattern during embryogenesis (Harrison et al., 1998). Early on, transcripts are detected broadly in the embryo and later refined to seven and then fourteen stripes during gastrulation, resembling the classical pair-rule pattern of expression. At germ band retraction, expression is restricted to the tracheal pits, hindgut, and other structures. Whole mount *in situ* hybridization to mRNA in embryos (work done by G. Rennebeck) revealed that *Socs36E* expressed in a strikingly similar pattern throughout embryogenesis (Fig. 3-1). The principle difference is that the domain of *Socs36E* expression appeared to be broader than that of *unpaired*. In embryogenesis, expression of *Socs44A* is not similar to *unpaired* (Fig. 3-1). Its expression can be seen only in the dorsal trunk and lateral braches of the tracheal system of late stage embryos. No expression can be detected in earlier stages, to include the tracheal pits at germ band

retraction. It must be noted that tracheal staining is a common artifact in mRNA *in situ* hybridization. Therefore, it is possible that the observed expression pattern is not legitimate. However, we did not detect tracheal staining with sense mRNA control probes. Furthermore, the tracheal staining was consistent with both DNA and RNA probes generated from variety of polymerases. We also examined animals homozygous for the *Df(2R)CA53* deficiency which includes the *Socs44A* locus. These embryos died before the onset of tracheal morphogenesis.

Similar results were seen when expression of these SOCS genes was observed during oogenesis (Fig. 3-1). In the developing egg chamber, *unpaired* is expressed specifically in the two polar follicle cells found at each end of the egg chamber. *Socs36E* is also expressed at the poles (work done by G. Rennebeck), but again, its domain of expression is much broader than that of *unpaired.* In contrast, *Socs44A* was not expressed in the somatic epithelium; however, transcripts were detected in germline, accumulating in the oocyte by stage 8.

Socs36E but not Socs44A is responsive to the JAK/STAT signal

To more directly assess the role of JAK/STAT signaling on the expression of *Socs36E* and *Socs44A*, mRNA *in situ* hybridizations were performed on embryos that lacked JAK/STAT pathway activity. We presumed that a decrease in cascade activation would result in a decrease in SOCS expression. *Hopscotch,* the *Drosophila* JAK, is an essential gene that has both maternal and zygotic functions and is required for early embryogenesis. Therefore, we utilized the dominant female sterile technique (DFS) to generate females that fail to produce *hop* in the germline (Chou and Perrimon, 1992).

Mating these females to wild-type males produces two populations of embryos. Because *hopscotch* is X-linked, the resulting male embryos are null for *hop*, while females are paternally rescued because they receive one zygotic copy of *hop* from their fathers. It has been documented that this technique produces embryos with the expected segmentation defects associated with *hop* loss of function (Binari and Perrimon, 1994). *In situ* hybridization of *Socs36E* mRNA in these *hop* germline clone derived animals (work done by G. Rennebeck) revealed a marked reduction in *Socs36E* expression. Half of the embryos (presumably the males) showed no staining above basal levels, with the exception of a persistent mesodermal expression (Fig. 3-2). *In situ* hybridization of *Socs44A* mRNA to *hop* germline clone derived embryos did not result in reduction in the tracheal staining. Interestingly, these animals had vastly reduced tracheal tissue as marked by the *trachealess*-LacZ enhancer trap (Fig. 3-2). The little tracheal tissue that was present was disorganized and did not resemble normal tracheal morphology. The paternally rescued animals had near wild-type trachea with only minor migration defects present. This phenotype is consistent with previous data illustrating the role of the JAK receptor*, domeless,* in tracheal morphogenesis (Brown et al., 2001). The failure to reduce the tracheal staining in *hop* germline clone derived animals suggests that *Socs44A* expression is not affected by loss of JAK pathway activity. These data indicate that while *Socs36E* expression is dependent on JAK/STAT activity, it is not the case for *Socs44A*.

The above assays do not address whether *Drosophila* SOCS are responsive to pathway activation. Recent work has shown that overexpression of *unpaired* or other JAK/STAT components could induce the expression of *Socs36E* (Karsten et al., 2002 and G. Rennebeck, unpublished). However, misexpression of *upd* using a *paired-GAL4*

driver failed to induce embryonic *Socs44A* expression (data not shown). Taken together, these data indicate that *Socs44A* is not dependent upon or responsive to the JAK/STAT pathway during embryogenesis.

Socs44A regulation of JAK/STAT signaling in wing development

The JAK/STAT pathway has been implicated in the development of the wing vein pattern. The hypomorphic $stat^{HiJak}$ mutant exhibits ectopic wing vein material in the posterior compartment of the wing at the posterior crossvein (Yan et al., 1996a). A strikingly similar phenotype is seen in certain heteroallelic combinations of *hop* mutants (Fig. 3-4). If SOCS negatively regulate JAK/STAT signaling in flies, we would expect misexpression in the posterior to cause a similar phenotype. Indeed, misexpression of two copies of a *Socs36E* transgene using the *engrailed*-GAL4 driver produced a similar phenotype (Callus and Mathey-Prevot, 2002 and G. Rennebeck, unpublished).

It is possible that Socs44A could negatively regulate the JAK/STAT pathway even though it is unable to respond to its signal because these events are uncoupled. To address this issue, Socs44A was subcloned from the full length LP02169 EST into the pUAST vector and the resulting construct was used in the standard transformation technique to generate transgenic animals (Spradling, 1986). These animals express Socs44A under the control of the GAL4/UAS binary expression system (Brand and Perrimon, 1993). GAL4, a yeast transcription factor, is not expressed in *Drosophila*, nor does it induce expression of endogenous genes. However, GAL4 recognizes the upstream activating sequence (UAS) that precedes the subcloned Socs44A transgene. The UAS-Socs44A transgenic flies were then mated to an array of established lines that

express GAL4 under the control of a variety of enhancers providing both temporal and spatial control of misexpression (FlyBase Consortium, 2003; Harrison et al., 1995; Tracey et al., 2000).

Socs44A was misexpressed in a variety of tissues (Table 3-1). In most cases, misexpression of *Socs44A* was either lethal or had no discernable effect. However, misexpression of *Socs44A* in the wing yielded several phenotypes (Fig. 3-3). In all cases, misexpression in the wing caused either formation of ectopic wing vein material and/or alterations in the guidance of the lateral or cross veins. *Patched* encodes a receptor involved in hedgehog signaling and is expressed specifically along the anterior/posterior boundary in the wing imaginal disc (Capdevila et al., 1994). *Patched*-GAL driven misexpression of *Socs44A* along this boundary, which is located between the third and fourth lateral veins (L3 and L4), caused both of these veins to pinch together at the anterior crossvein. Misexpression using the GAL-T6 driver also caused this pinching phenomenon and also resulted in ectopic wing vein material extruding from the anterior side of L4. GAL-T113 misexpression resulted in the formation of one or more ectopic crossveins between L2 and L3. Misexpression of *Socs44A* using the *engrailed*-GAL4 driver also resulted ectopic wing vein formation around the posterior crossvein (Fig. 3-4). This ectopic material can be seen as a thickening of the PCV at the distal ends where it intersects L4 and L5. Additional ectopic material is also seen beneath L5 and at the most distal portions of L4 and L5. Finally, these wings also exhibit an exaggerated arching of L3.

While misexpression of *Socs44A* results in phenotypes similar to and consistent with JAK/STAT mutants, it does not directly address the capacity of Socs44A to

downregulate the pathway. The *engrailed*-GAL4 driven *Socs44A* wing phenotype is fully penetrant. The addition of a single copy of the *hop*^{c111} null allele enhances this phenotype (Fig. 3-4), which would be expected of a negative regulator of JAK/STAT signaling. Misexpression of *hop* using the *engrailed*-GAL driver results in complete loss of L5 and the most of PCV as well as ectopic wing venation at the anterior crossvein (ACV). Concurrent misexpression of *Socs44A* in this background rescues the wing vein defects, restoring the PCV and L5 (Fig. 3-4). The resulting wing has ectopic wing vein material at the distal part of L5 and occasionally along L3 and arching of L3 is also present in these wings, phenotypes attributed to misexpression of *Socs44A* alone. Taken together, these data suggest that Socs44A is capable of downregulating the JAK/STAT pathway during development of the wing.

Socs44A appears capable of downregulating the JAK/STAT pathway during wing development, but is that reflective of the endogenous activity *in vivo?* The PCV wing vein spur associated with a specific heteroallelic combination of *hop* is 98% penetrant (n=89). Furthermore, these flies show a reduction in viability. It would be expected that misexpression of *Socs44A* would enhance these phenotypes. Misexpression of *Socs44A* in three different transheterozygous combinations of *hop* always resulted in lethality (Table 3-2). Exacerbation of the reduced viability caused by the *hop* mutations is consistent with a role in downregulating the JAK/STAT cascade. These data indicate that ectopic expression of *Socs44A* has the capacity to downregulate endogenous *hop* in the developing wing.

In a reciprocal assay, we would expect that a reduction in Socs44A would rescue the *hop* transheterozygous wing spur phenotype. Currently, a loss of function allele of

Socs44A has not been isolated. In lieu of this, there are several deficiencies that remove the Socs44A locus. CA53 and NCX10 are deficiencies that lack *Socs44A* by virtue of their failure to complement lethal P-element insertions that flank either side of the locus. A single copy of either of these deficiencies when placed in a *hop*$^{M38/msv}$ background partially rescued the PCV wing spur phenotype, reducing the penetrance by as much as 52% (Fig. 3-5). These observations suggest that endogenous Socs44A downregulates *hop* in wing development. Unexpectedly, these animals often had misguided PCVs that failed to connect to L4 and occasional perturbations of the ACV.

Socs36E but not Socs44A regulates JAK/STAT signaling in oogenesis

As mentioned previously, mammalian SOCS responsiveness and subsequent pathway regulation is context specific (reviewed in Johnston and O'Shea, 2003). We utilized the *Drosophila* ovary to test if there is tissue specificity to Socs44A regulation of the cascade. Recent work has established that the JAK/STAT pathway is responsible for patterning the somatic follicle cells of the vitellarium (Beccari et al., 2002; McGregor et al., 2002; Silver and Montell, 2001; Xi et al., 2003). Specifically, JAK activity is essential for identification of the posterior terminal follicle cells (Xi et al., 2003). These cells are molecularly defined by the expression of the ETS domain transcription factor, *pointed* (Gonzalez-Reyes and St Johnston, 1998; Morimoto et al., 1996; Roth et al., 1995). Expression of the pointed protein is detected as a gradient emanating from the posterior pole of developing egg chambers. Clones of cells in the posterior that lack *hop* activity fail to express the reporter, *pnt*-LacZ (Xi et al., 2003) establishing this as an assay for JAK pathway activity.

Clones of cells misexpressing either *Socs36E* or *Socs44A* in the posterior of developing egg chambers were generated using the FLP-OUT cassette technique (Neufeld et al., 1998; Struhl and Basler, 1993). Clones of cells misexpressing *Socs36E* at high levels in the posterior cells of the developing egg chamber exhibited a loss of the *pnt*-LacZ marker (Work done by R. Xi, Fig. 3-6). However, when *Socs44A* was misexpressed in a similar fashion, there was no reduction of *pnt*-LacZ expression (Fig. 3-6). These data suggest a functional role for *Socs36E* but not *Socs44A* in the regulation of JAK signaling in oogenesis. These data also suggest that context specificity of SOCS regulation of the JAK pathway is conserved across taxa.

DISCUSSION

In this chapter, the role of Socs44A in JAK/STAT signaling was examined. It has been shown that the four well-studied mammalian SOCS (CIS and SOCS-1 to SOCS-3) respond to and downregulate the JAK/STAT cascade in a classical negative feedback loop. However, it should be noted that SOCS-4, SOCS-6, and SOCS-7 have not been shown to be induced by any cytokine or growth factor. Furthermore, SOCS-4 and SOCS-7 have not been shown to inhibit JAK/STAT signaling *in vitro* or *in vivo* (reviewed in Fujimoto and Naka, 2003).

Socs36E appears to be able to respond to and downregulate the JAK/STAT pathway in flies, indicating that this mechanism of JAK/STAT regulation is conserved. First, Socs36E is expressed in a dynamic pattern strikingly similar to the ligand, *unpaired.* We observed that in both embryos and ovaries, the domain of *Socs36E* expression exceeded that of *unpaired.* This observation is consistent with the model that

Socs36E is responsive to JAK/STAT signaling. Unpaired is a secreted molecule which could induce pathway activation in neighboring cells, leading to *Socs36E* expression in those cells, creating a broader domain of *Socs36E* expression. Although diffusion of Upd has not been directly observed, Upd expressed in follicle cells of the ovary leads to JAK activation in neighboring cells (Xi et al., 2003) which would presumably lead to expression of *Socs36E* in those cells. Second, ours and other labs have shown that *Socs36E* expression can be induced by ectopic activation of the JAK/STAT pathway (Karsten et al., 2002 and G. Rennebeck, unpublished). Finally, reduction of cascade activity has a direct effect on *Socs36E* expression (work done by G. Rennebeck, Fig. 3-2). Interestingly though, in *hop* null animals, mesodermal expression of *Socs36E* persisted. This suggests that there may be alternative pathways that activate *Socs36E* transcription in those tissues.

Socs44A mRNA expression was detected in both ovarian and embryonic tissues in patterns inconsistent with JAK pathway activation. During oogenesis, an accumulation of transcripts was detected in the germline during the period when maternal RNAs are loaded into the oocyte. Further expression was not detected until late stages of embryogenesis, when it was restricted to the trachea. However, the expression of *unpaired* in the tracheal pits (Fig. 3-1) coupled with studies implicating a role for JAK/STAT signaling in tracheal morphogenesis (Brown et al., 2001) support our findings. Furthermore, the existence of both embryonic and ovarian ESTs supports the observed temporal expression patterns.

Through misexpression, it was shown that Socs44A has the capacity to downregulate ectopic activated Hop. Additionally, the loss-of-function mutation *hop*^{c111} enhances the *Socs44A* misexpression phenotype. Consistent with these data, misexpression of *Socs44A* exacerbates the reduced viability associated with transheterozygous *hop* mutants. Finally, deficiencies for Socs44A rescue *hop* heteroallelic phenotypes. Taken together, these data strongly suggest that endogenous Socs44A downregulates JAK pathway activity *in vivo*. In contrast, misexpression of *Socs44A* had no effect on expression of a marker for JAK pathway activity during oogenesis. However, it must be noted that both the EGFR/MAPK and JAK/STAT pathways work synergistically to pattern the posterior terminal follicle cells in oogenesis (Gonzalez-Reyes and St Johnston, 1998; Roth et al., 1995; Xi et al., 2003). Roles of Socs44A in EGFR/MAPK signaling will be examined in chapter four; this issue will be revisited there. Nonetheless, our observations indicate that there may be context specificity to SOCS activity in *Drosophila*, a phenomenon that has been observed in the study of mammalian SOCS.

When we attempted to rescue the wing spur phenotype associated with certain heteroallelic combinations of *hop* by introducing a single copy of *Socs44A*-containing deficiency, we observed unexpected wing phenotypes. While the penetrance of the *hop* heteroallelic phenotype was reduced, we observed wings with misguided PCVs and occasional perturbations of the ACV. Both of these defects appeared to be the result of failure of the crossvein to migrate properly. The development and placement of crossveins is believed to be the result of signaling from the Jun Kinase, BMP, EGFR/MAPK, and Notch pathways (reviewed in Marcus, 2001). Although the role of

JAK/STAT signaling in wing venation has not been determined, *hop* (Fig. 3-5) and *Stat92E* (Yan et al., 1996a) mutant phenotypes suggest that this pathway is somehow involved in this process.

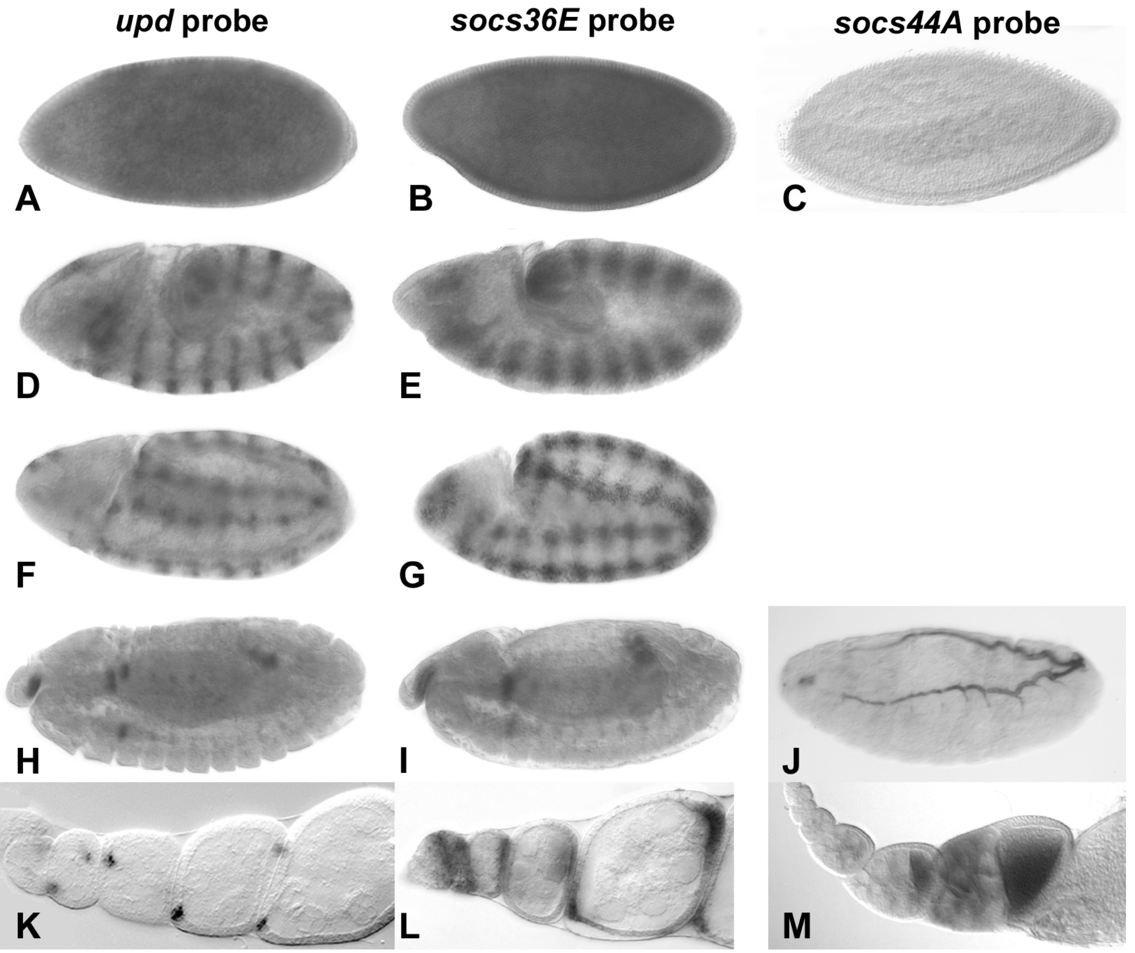


Figure 3-1. Socs36E but not Socs44A is expressed similarly to upd in both embryogenesis and oogenesis. The embryonic expression patterns of *upd* (first column) are remarkably similar to the patterns of expression of *Socs36E* (second column), while *Socs44A* expression patterns (third column) are distinct. At each stage during embryonic development, the dynamic pattern of *upd* expression (A, D, F, and H) is nearly matched by *Socs36E* expression (B, E, G, and I). *Socs44A* expression is not detected until later stages, where it is restricted to the trachea (C and J). In the ovary, *upd* is expressed specifically in the polar follicle cells at each end of the chamber (K). *Socs36E* expression encompasses the entire anterior and posterior polar caps of post-germarium stage chambers. See text for more details. *Upd* and *Socs36E in situs* done by G. Rennebeck.

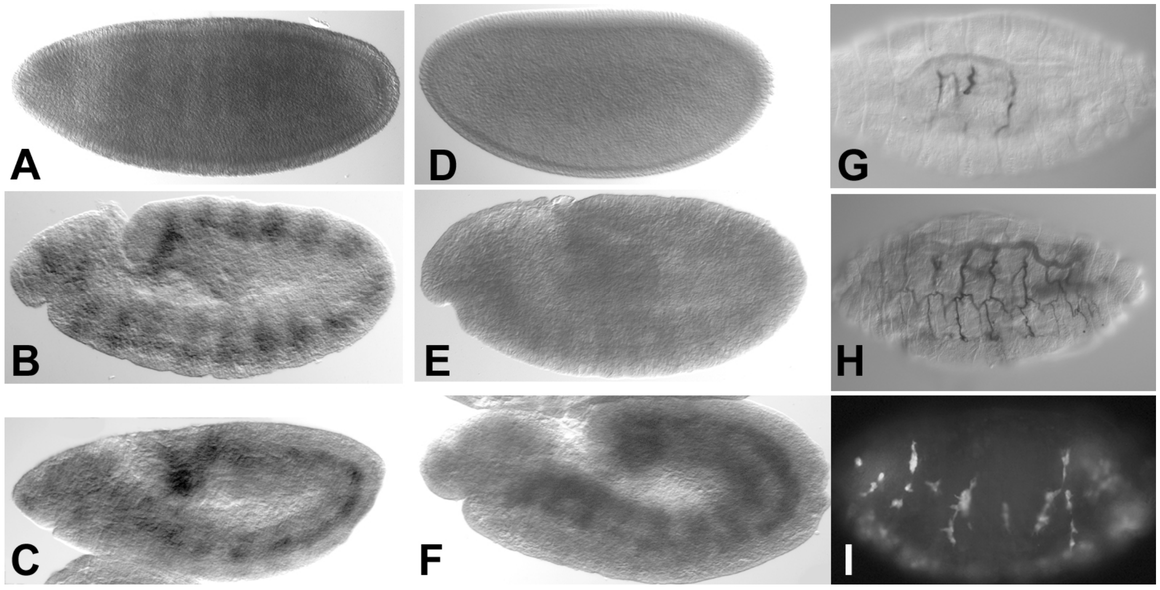


Figure 3-2. Socs36E but not Socs44A expression is dependent on JAK activity. As compared with wild-type at various embryonic stages (A, B, and C), germline clone derived embryos from hop^{c111} mothers display dramatically reduced or eliminated expression of *Socs36E* (D, E, and F). Only a stripe of mesodermal staining in germ band extended mutant embryos (F) remains at near normal intensity. Expression of *Socs44A* persists in hop^{c111} germline clone derived embryos (G and H). However, the tracheal morphology is drastically reduced in these animals (G) and somewhat reduced in animals that are paternally rescued (H), evidenced by the *trachealess-LacZ* enhancer trap (I). Panels A - F are the work of G. Rennebeck.

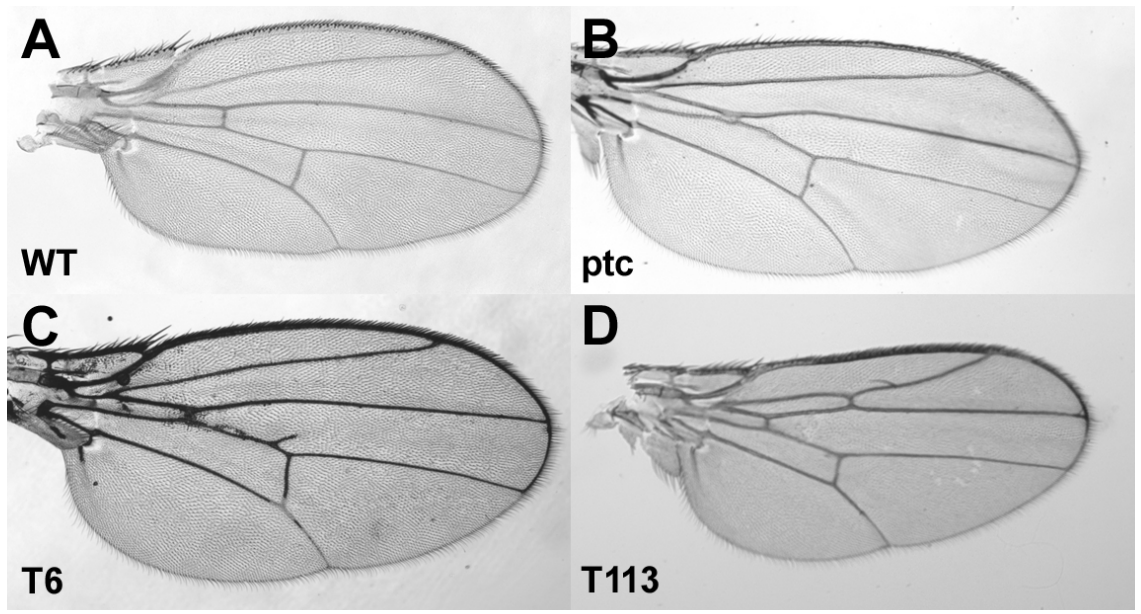


Figure 3-3. Wing misexpression phenotypes generated with the Socs44A transgene. Anterior is up and posterior is down for all wing figures. A Socs44A transgene was misexpressed using the GAL4/UAS binary system (Brand and Perrimon, 1993) altering the wing vein pattern seen in wild-type wings (A). When misexpressed along the A/P axis with a *patched* GAL4 driver, the L3 and L4 pinched together at the ACV (B). Misexpression with other wing drivers resulted in ectopic wing vein formation (C) or extra crossveins (D).

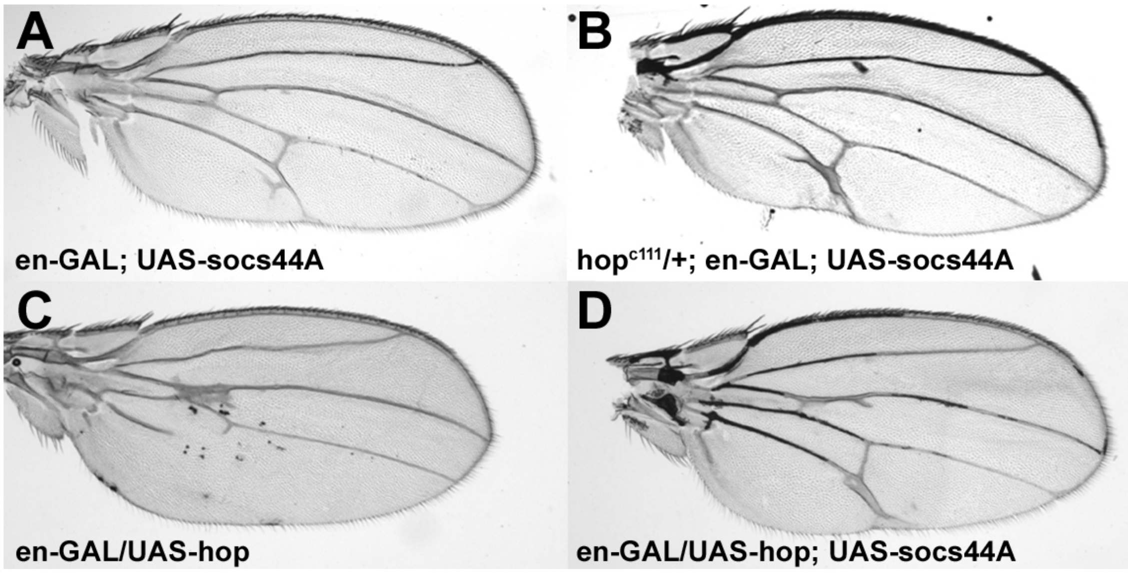
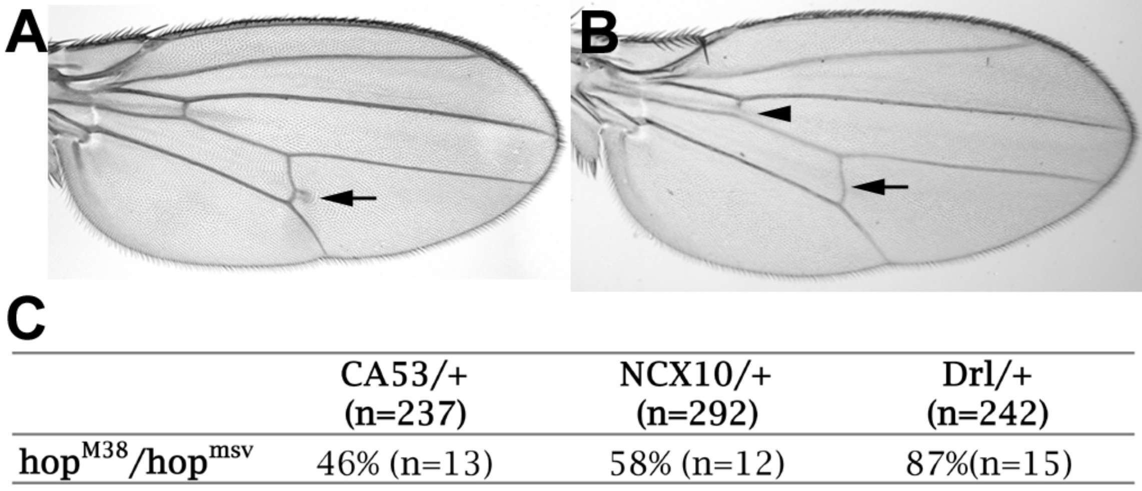


Figure 3-4. Socs44A is capable of downregulating the JAK/STAT cascade in the wing. *Engrailed*-GAL4 misexpression of *Socs44A* in the posterior half of the wing results in ectopic wing venation (A) that is enhanced by the addition of a single copy of the hop^{c111} null allele (B). *Engrailed*-GAL4 misexpression of *hopscotch* results in a loss of L5 and most of the PCV and ectopic wing venation near the ACV. This phenotype is rescued by the concurrent misexpression of *Socs44A* (D). The PCV and L5 are restored and the amount of ectopic wing vein is reduced.



	CA53/+ (n=237)	NCX10/+ (n=292)	Drl/+ (n=242)
hop^{M38}/hop^{msv}	46% (n=13)	58% (n=12)	87%(n=15)

Figure 3-5. Endogenous Socs44A regulates JAK pathway activity. The *hop*^{msv}/*hop*^{M38} transheterozygote has a wing spur phenotype (A) that is 98% penetrant (n=89). A single copy of a deficiency (*Df(2R)CA53* or *Df(2R)NCX10*) that includes the *Socs44A* locus can rescue this phenotype resulting in reduced (B, arrowhead) or misguided (B, arrow) crossveins. Furthermore, the penetrance of the spur phenotype is dramatically reduced (C). This phenomenon was not seen with a local deficiency that does not include *Socs44A* (C, *Df(2R)Drl*^{rv18}).

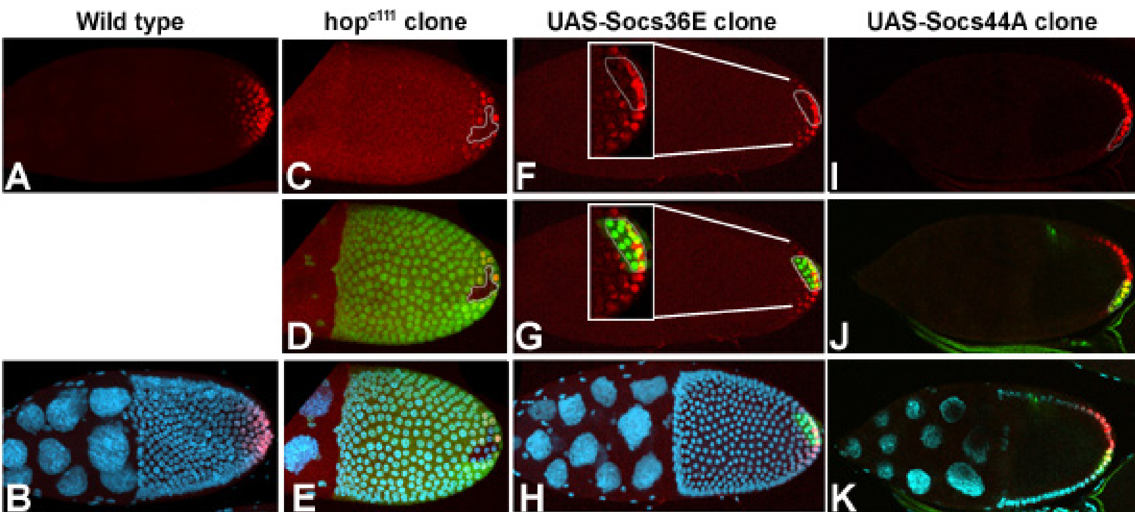


Figure 3-6. Socs36E but not Socs44A regulates JAK activity during oogenesis. In wild-type ovaries (first column), *pnt*-lacZ (red) is expressed in a gradient in the posterior terminal cells. Cells that lack *hop* activity (marked by a lack of green, see outline), also fail to express *pnt*-lacZ (second column). Similarly, UAS-Socs36E expressed in clones (marked by presence of green, see outline), have reduced or missing *pnt*-LacZ expression (third column, see insets). Conversely, UAS-Socs44A expressed in clones (marked by presence of green, see outline), had no effect on *pnt*-LacZ expression (third column). DAPI nuclear staining is shown in blue. Panels A – H are the work of R. Xi.

Table 3-1. Socs44A misexpression phenotypes.

GAL4 driver	Expression pattern	Phenotype
{ey.H-GAL}3-8	eye	wild-type
{ey.H-GAL}4-8	eye	wild-type
ap gal	wing, scut, ovary, discs, brain	ectopic wing vein, reduced scutelum
btl-GAL	Tr	lethal
GAL-71B	discs	wild-type
GAL-c135	LG, SG, FB, Tr, Malp, OL, gut, brain	wild-type
GAL-c136a	unknown	wild-type
GAL-c263	brain	wild-type
GAL-c272	unknown	wild-type
GAL-c273	LG, SG, FB, Tr, Malp, OL, cut, discs, gut, brain	wild-type
GAL-c355	LG, SG, FB, Tr, Malp, OL, cut, discs, gut, brain	wild-type
GAL-c591	SG, FB, Tr, Malp, discs, gut	wild-type
GAL-c617	unknown	reduced viability*
GAL-c704a	MB, OL	wild-type
GAL-c718	unknown	wild-type
GAL-c729	LG, SG, FB, Tr, Malp, discs, gut	lethal
GAL-c833	unknown	slight enlargement of eye*
GAL-e16E (en)	FB, discs, cut, gut	ectopic wing vein
GAL-e22C	embryo(ect), disc, brain, ovary	reduced viability
GAL-e33C	LG, Tr, brain, OL, gut, Malp, VNC	lethal
GAL-T113	discs	extra crossvein between L2 and L3
GAL-T155	LG, SG, FB, Tr, OL, cut, discs, gut, brain	wild-type
GAL-T59	LG, Tr, OL, FB, SG, discs, gut, brain	wild-type
GAL-T6	wings	ectopic wing vein
h-GAL (1J3)	embryo(PR), wings	ectopic wing vein*
NGT	embryo	wild-type
ptc-GAL4	wing, brain, gut, embryo(PR)	ectopic wing vein, pinched ACV

The UAS-Socs44A transgene was crossed to an array of GAL4 drivers (FlyBase Consortium, 2003; Harrison et al., 1995). ACV, anterior crossvein; cut, cuticle; discs, imaginal discs; FB, fat bodies; LG, lymph gland; Malp, Malpighian tubules; MB, mushroom body; OL, optic lobes; PR, pair-rule pattern; SG, salivary glands; Tr, trachea; VNC, ventral nerve cord. Asterisks indicate phenotypes scored from crosses performed at 29^{0}C all other crosses performed at 25^{0}C.

Table 3-2. Misexpression of Socs44A exacerbates the reduced viability of hop heteroallelic mutants.

	Genotype	hop^{M38} (n=213)	hop^{GA32} (n=332)	hop^{M75} (n=172)
a	hop^{x}/FM7; en-GAL; TM3	33	52	21
b	hop^{x}/FM7; en-GAL; UAS-socs44A	25	33	28
c	hop^{x}/hop^{msv}; en-GAL; TM3	11	20	13
d	hop^{x}/hop^{msv}; en-GAL; UAS-socs44A	0 (E=8.33)	0 (E=12.69)	0 (E=17.33)

Misexpression of *Socs44A* in a range of *hop* heteroallelic mutants resulted in lethality. For each mutant combination, n represents the total number of progeny scored in the cross. The number of progeny of each genotype is indicated. E represents the expected number of progeny of that genotype if *Socs44A* misexpression were to have no effect on viability, where E = b/a x c.

Chapter IV

Socs44A is capable of interacting with the EGFR/MAPK pathway

INTRODUCTION

The number of developmental and homeostatic programs far surpasses the number of signal transduction pathways that control them. One method of compensation for this deficit is the crosstalk among signaling cascades. Crosstalk permits a higher order of complexity resulting in a greater number of possible responses without increasing the number of cascades that generate them. Combinatorial control of gene expression through crosstalk may serve to attenuate transcription levels and possibly contribute to the robustness to changes in gene dosage that are inherently compensated for by signaling cascades (Eldar et al., 2002).

There is considerable interplay between the JAK/STAT and several other cascades (Heinrich et al., 2003; Rane and Reddy, 2000; Shuai, 2000). This is not surprising because certain species lack some of the components required for activation and propagation of the JAK/STAT signal. For example, the slime mold *Dictyostelium* possesses three STAT molecules but lacks any other component. The nematode, *C. elegans*, also possesses a STAT molecule, as well as a predicted protein with high similarity to SOCS, but also lacks a JAK or other components (see reviews by Dearolf, 1999; Rawlings et al., 2004). These observations imply that the STATs can be activated by some other mechanism than by JAK phosphorylation. In fact, it has been shown that STAT and SMAD (a TGF-β pathway molecule) can interface through p300 (Iwamoto et

al., 2002; Moustakas, 2002). The phosphotyrosine docking sites on the JAK/STAT receptor are inviting targets for interaction with other signal cascades as well. For example, p85 interacts with the phosphorylated receptor. This molecule also interacts with IRS (insulin receptor substrate) bridging JAK/STAT with the phosphoinositide 3-kinase (PI3K) pathway (Foster et al., 2003).

Of particular interest is the EGFR/MAPK pathway, which has a bidirectional interface with JAK/STAT on several levels (reviewed in Rawlings et al., 2004). The protein tyrosine phosphatase SHP-2 (SH2 containing protein) interacts with the JAK/STAT receptor gp130 as well as GRB2, an adaptor in the EGFR/MAPK cascade (Qu, 2002). Also, it has been shown that JAK can activate MAPK through the Pyk2 intermediary (Takaoka et al., 1999). Several MAPKs can contribute to the transcriptional ability of STATs via serine phosphorylation, independent of JAK activation (David et al., 1996; Decker and Kovarik, 2000; Leaman et al., 1996). A recent report directly links murine SOCS-3 to EGFR signaling. Specifically, SOCS-3 binds to and degrades p120 RasGAP, the small GTPase responsible for inactivating Ras. The result of this interaction is an upregulation of the EGFR/MAPK pathway (Cacalano et al., 2001).

The interplay between JAK/STAT and EGFR/MAPK is also reflected in the *Drosophila* model. In the follicular epithelium of the *Drosophila* ovary, JAK/STAT pathway pre-patterns the termini and is required for EGFR competence in the posterior of developing egg chambers (McGregor et al., 2002; Xi et al., 2003). Likewise, the MAPK pathway is required for JAK driven differentiation of hematopoietic cells because a direct interaction between hopscotch and D-Raf is required for the blood cell alterations caused

by the hypermorphic *hop*$^{Tum\text{-}l}$ and *hop*^{T42} alleles (Luo et al., 2002). Furthermore, Socs36E has been reported to weakly suppress EGFR signaling in the wings (Callus and Mathey-Prevot, 2002).

In this chapter, potential interactions between Socs44A and other signaling cascades is explored. Strong evidence of a genetic interaction between Socs44A and the EGFR/MAPK (Epidermal growth factor receptor/Mitogen-activated protein kinase) pathway is presented. Specifically, Socs44A acts to upregulate the EGFR/MAPK cascade, analogous to what was reported for mouse SOCS-3.

RESULTS

Interactions between Socs44A and other signaling cascades

As described previously, misexpression of *Socs44A* using the *engrailed*-GAL4 driver resulted in a fully penetrant phenotype marked by the presence of ectopic wing vein material in the posterior compartment of the wing. To test possible dominant interactions between Socs44A and other signaling pathways, flies were generated that were heterozygous for mutations or carried GAL4/UAS compatible transgenes in an *engrailed*-GAL4; UAS-*Socs44A* genetic background to determine if they had any effect on the Socs44A misexpression phenotype. Members of the TGF-β, dpp (decapentaplegic), and hedgehog pathways were tested (Table 4-1), none of which showed any detectable modulation of the Socs44A misexpression phenotype. An interaction between Socs44A and the EGFR/MAPK pathway was detected and will be described in the next section.

Socs44A interacts with the EGFR/MAPK pathway

Engrailed-GAL4 driven misexpression phenotypes of Socs44A were suppressed in the background of heterozygous mutations for EGFR/MAPK components, including: *Ras85D, Son of sevenless (Sos),* and *Egfr* (Table 4-1 and Fig. 4-1). Consistent with these observations, reduction in the dosage of *argos* enhanced the *Socs44A* misexpression phenotype. *Argos* is an extracellular ligand that possesses an EGF motif and acts as an inhibitor of EGFR/MAPK signaling in the developing wing and other tissues (Sawamoto et al., 1996; Sawamoto et al., 1994; Schweitzer et al., 1995). Concurrent misexpression of *Socs44A* and *argos* also showed a phenotypic interaction consistent with the previous data. Misexpression of two copies of an *argos* transgene under the *engrailed-GAL4* driver resulted in wings lacking the 4th lateral vein (L4) as well as both cross-veins (Fig. 4-2). Concurrent misexpression of a single copy of the *Socs44A* transgene in this background was able to rescue this phenotype, restoring the posterior crossvein and both the most proximal and distal portions of L4. The resulting wing phenotype mimicked that seen when only a single copy of *argos* was used in the misexpression assay or what is seen in heteroallelic *Egfr* mutants. Finally, concurrent misexpression of a single copy of the *argos* and *Socs44A* transgenes produced a nearly wild-type wing. These data indicate that *Socs44A* expression is able to suppress *argos* misexpression phenotypes in a dose-dependent manner. It should be noted that concurrent misexpression of the gene encoding GFP did not have similar effects in these assays (Fig. 4-5), indicating that the phenotypes produced were not merely a consequence of titrating GAL4.

To address whether endogenous Socs44A activity had a similar effect on EGFR pathway activity, we assayed the effects of a *Socs44A* deficiency in the *argos* misexpression background. *Engrailed-GAL4* misexpression of *argos* produces a range of phenotypic classes in which parts or all of L4 and/or the PCV are missing (Fig. 4-3). Addition of a single copy of a *Socs44A* deficiency shifted the distribution of phenotypes to the more severe classes compared to a deficiency that does not include the *Socs44A* locus (Fig. 4-4). Taken with the previous data, these observations suggest that *Socs44A* does indeed upregulate EGFR signaling in the *Drosophila* wing.

Not all proteins that interact with various MAPK pathways showed interactions with Socs44A. For example, *sevenless*, the receptor used in eye development (Hafen et al., 1994) had no effect on the *Socs44A* misexpression phenotype. *Sprouty* is an intracellular inhibitor of the EGFR/MAPK pathway that possesses a mutant phenotype similar to the *engrailed*-GAL4 driven misexpression phenotype of *Socs44A* (Casci et al., 1999). In my hands, *sprouty* heterozygotes exhibited a PCV wing spur nearly identical to $hop^{M38/msv}$ transheterozygotes. Addition of a deficiency that includes the *Socs44A* locus had no effect on the penetrance of this phenotype. *Veinlet* is an allele of *rhomboid*, a transmembrane protein that interacts with EGFR and promotes signaling during wing development. Homozygous *veinlet* mutants exhibit wings that lack the distal portions of all of the lateral veins (Guichard et al., 1999; Sturtevant et al., 1993). Removal of one copy of *Socs44A* via a deficiency had no effect on this phenotype. These observations do not contradict my previous findings because these assays tested for dominant interactions.

DISCUSSION

Crosstalk is a common theme amongst signal transduction cascades. This phenomenon is likely to serve multiple purposes. Among them are to allow a cell to integrate multiple signals into a single response, or to contribute to the compensation for changes in gene dosage required to maintain the fidelity of developmental and homoeostatic processes dependent on signal transduction mechanisms.

While others have shown that Socs36E genetically interacts with and downregulates the EGFR pathway (Callus and Mathey-Prevot, 2002), it is illustrated here that Socs44A acts in the opposite manner. Socs44A is able to rescue misexpression of the EGFR negative regulator *argos* in a dose-dependent manner. Furthermore, mutations in EGFR pathway components rescue Socs44A misexpression phenotypes. Finally, a loss of endogenous Socs44A activity enhanced the *argos* misexpression phenotype. Taken together, these data indicate that Socs44A is upregulating the EGFR pathway. This finding is consistent with a recent report describing an interaction between SOCS-3 and the p120 RasGAP (Cacalano et al., 2001). In this model, SOCS-3 degrades p120 RasGAP, the GTPase-activating protein responsible for inactivating Ras, resulting in an upregulation of the EGFR/MAPK pathway. It is possible that Socs44A is acting in an analogous manner; biochemical analyses are required to address this hypothesis.

Interactions were not detected between Socs44A and several known EGFR/MAPK pathway proteins including: *sevenless, sprouty,* and *veinlet.* In addition, interactions with other cascades were also not observed. Members of the Notch, Dpp, Hedgehog, and TGF-β pathways were tested, none of which showed an interaction with Socs44A in misexpression assays. This does not rule out interactions between Socs44A

and EGFR/MAPK or those other pathways. The assay may not have been sensitive enough or the location of misexpression may have been inappropriate to observe interactions. In light of the fact that Socs44A regulation of JAK signaling seems to be context specific, it is possible the interactions between Socs44A and other cascades may exist and not be detected by this particular misexpression assay. Finally, the use of animals heterozygous for mutations may not have sufficiently reduced the function of the genes tested to alter the *Socs44A* misexpression phenotype. Clearly, this assay was designed to detect dominant interactions; therefore, negative data cannot be interpreted as a lack of an interaction.

In the previous chapter, I showed that ectopic expression of *Socs44A* in clones of cells at the posterior terminus of the developing egg chamber had no effect on the posterior cell fate marker, pointed, suggesting that Socs44A does not downregulate JAK/STAT signaling in oogenesis. Recent work in our lab showed that the JAK/STAT pathway is responsible for establishing a cell fate pre-pattern exhibiting a symmetrical mirror image during oogenesis (Xi et al., 2003). This symmetry is broken by Gurken, expressed from the oocyte positioned in the posterior of the egg chamber, which activates EGFR/MAPK signaling, resulting in the posterior cells adopting the appropriate terminal cell fate (Gonzalez-Reyes and St Johnston, 1998; Roth et al., 1995). Thus, in order for cells to adopt their proper terminal cell fates, both the JAK/STAT and EGFR/MAPK pathways must be activated. This is evidenced by the failure of terminal cell fate differentiation via ectopic EGFR induction in main body cells, where JAK is not active; whereas success is seen in induction of EGFR induction in anterior cells, where JAK is active (Xi et al., 2003).

One could imagine that ectopic Socs44A has no effect on pointed expression because its downregulation of JAK/STAT and upregulation of EGFR/MAPK cancels itself out. This theory is dependent on incomplete reduction of JAK/STAT signaling by Socs44A because complete abolishment of pathway activity would result in loss of pointed expression, regardless of its effects on EGFR/MAPK signaling. If indeed, it was incomplete reduction of JAK/STAT signaling, but still sufficient to pre-pattern the terminal cells, we would expect to see cell autonomous ectopic pointed expression in clonal cells, driven by Socs44A upregulation of EGFR/MAPK signaling. This was not the case, as the gradient of pointed expression from the posterior terminus was unaffected by induction of clones ectopicly expressing *Socs44A*. Therefore, we can conclude that Socs44A is not capable of affecting JAK/STAT or EGFR/MAPK signaling in patterning the terminal follicle cells of the ovary, at least to the sensitivity of the assay used. This does not rule out roles in patterning other cell fates within the ovary. Furthermore, this suggests that Socs44A regulation of EGFR/MAPK signaling might also be tissue specific. Further clonal analyses will be required to validate the issues raised here.

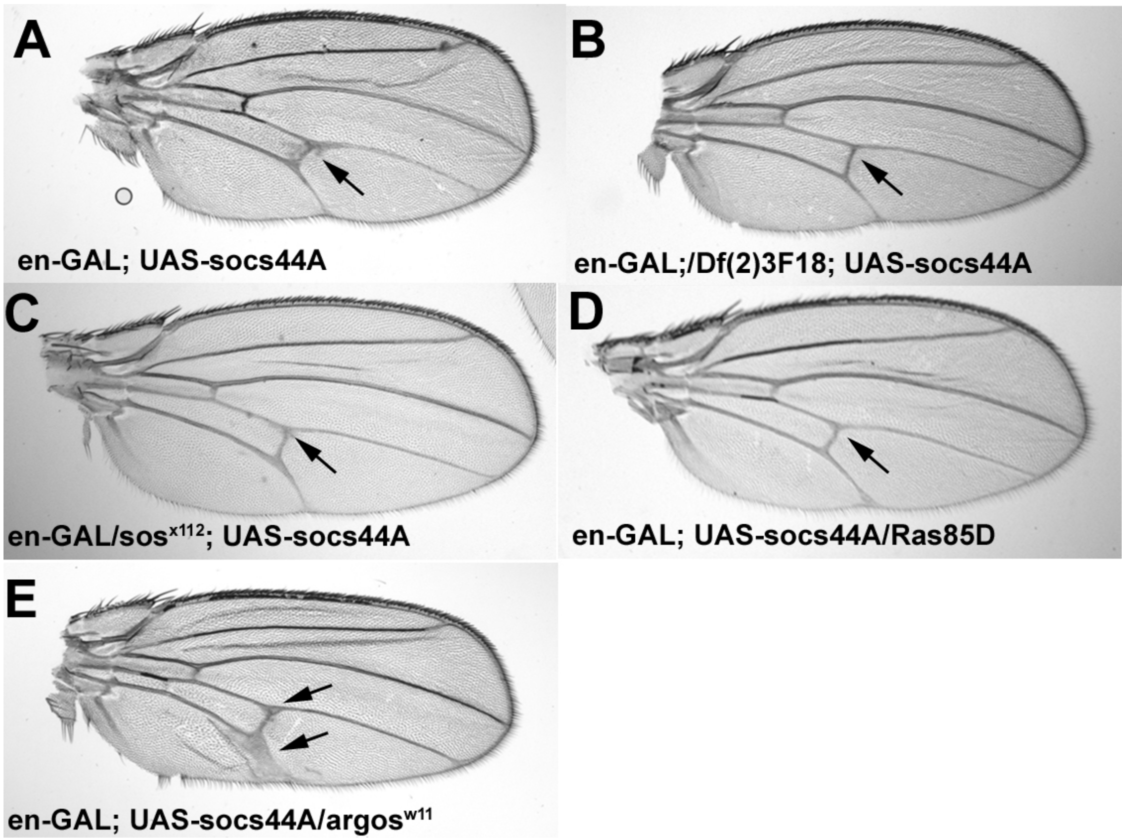


Figure 4-1. Socs44A increases activity of EGFR signaling. The ectopic wing vein phenotype of *Socs44A* misexpression (A, arrow) is rescued by reduction of *Egfr* (B), *Sos* (C) or *Ras85D* (D), all positive effectors of EGFR signaling. In contrast, reduction of *argos*, a negative regulator of EGFR signaling, enhances the phenotype (E, arrows).

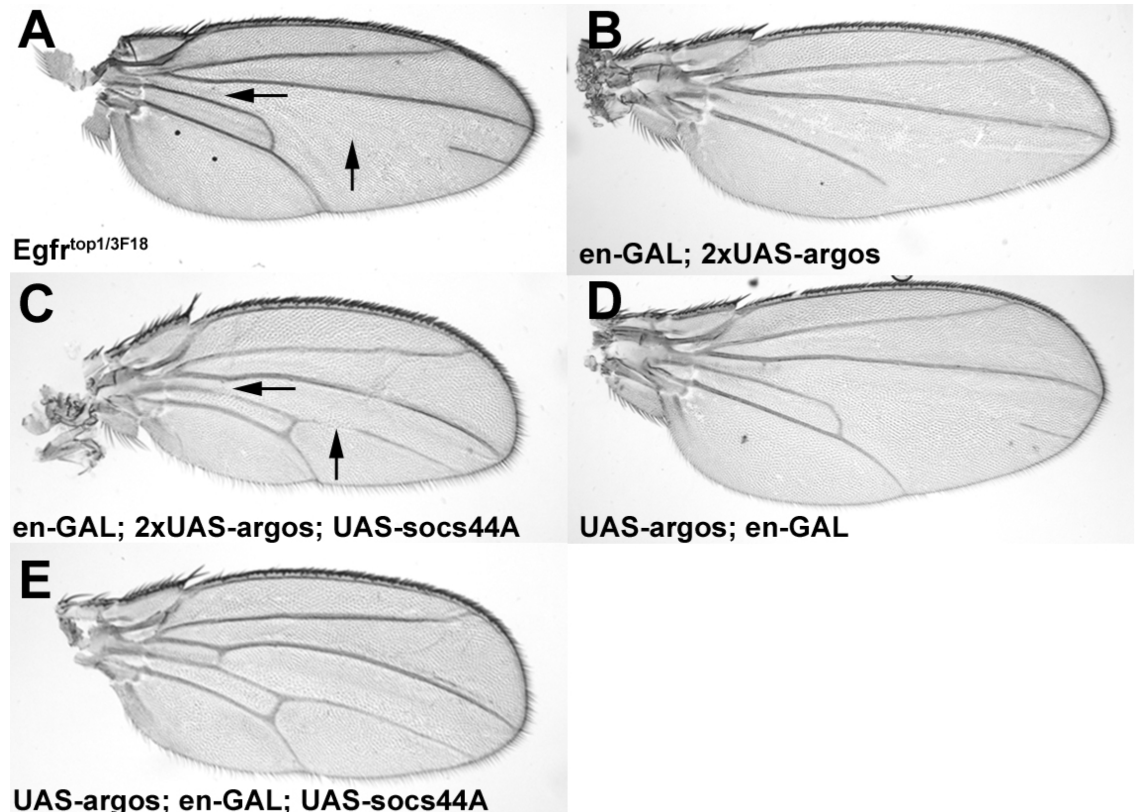


Figure 4-2. Concurrent misexpression of Socs44A rescues argos misexpression in a dose-dependent manner. Heteroallelic combinations of EGFR mutations result in loss of the ACV and the medial portion of L4 (A). Misexpression using two copies of the *argos* transgene results in loss of both crossveins and L4 (B, arrows). Concurrent misexpression of *Socs44A* rescues this phenotype, resulting in a wing similar to the EGFR double mutant (compare C and A, arrows). When a single copy of the *argos* transgene is used (D), the *Socs44A* misexpression rescue results in a nearly wild-type wing (E). All UAS constructs are driven in the posterior of the wing using one copy of *engrailed*-GAL (en-GAL).

Representative picture	Class Description
	Class I: Missing L4 past presumptive ACV AND missing all of the PCV
	Class II: has L4 past presumptive ACV OR has any portion of PCV
	Class III: has PCV that curves to parallel with L5
	Class IV: same as class III AND has contiguous L4 to the PCV
	Class V: same as class IV AND has distal portions of L4 present, similar to EGFR transheterozygous mutant (See Fig. 4-2)
	Class VI: complete PCV and complete L4 that runs contiguously from proximal to distal portion of the wing

Figure 4-3. Phenotypic classes generated by misexpression of argos. Misexpression of *argos* via the *engrailed*-GAL4 driver produces a range of phenotypic classes, categorized by the presence of the posterior crossvein (PCV) and L4. Left column contains representative pictures of each class. Right column, corresponding description used to score wings into the phenotypic classes. It should be noted that Class V and Class VI phenotypes were only observed in animals where *Socs44A* and *argos* are concurrently misexpressed.

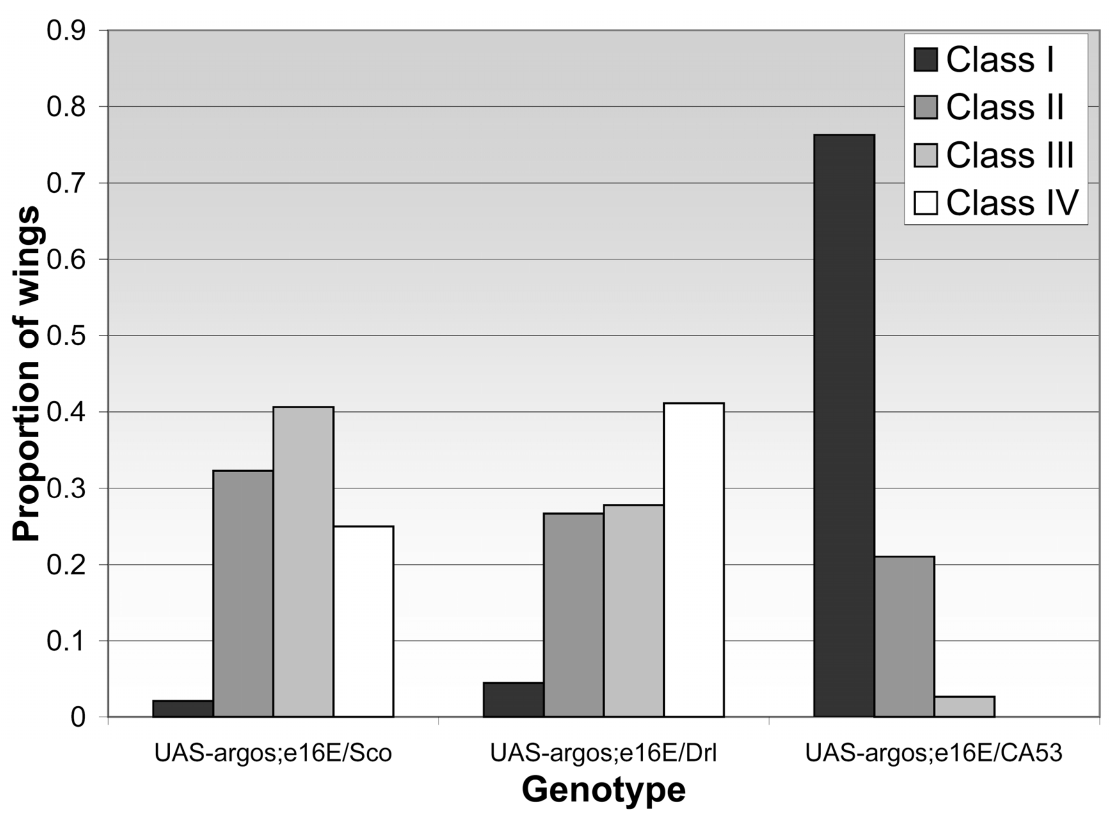


Figure 4-4. Socs44A deficiencies enhance argos misexpression phenotypes. A deficiency (*Df(2R)CA53*, CA53) that includes the *Socs44A* locus shifted the distribution of phenotypes to the more severe classes compared to a local deficiency that does not include *Socs44A* (*Df(2R)Drl*^{rv18}, Drl) or no deficiency at all (*Scutoid*, Sco). See Fig 4-3 for class definitions.

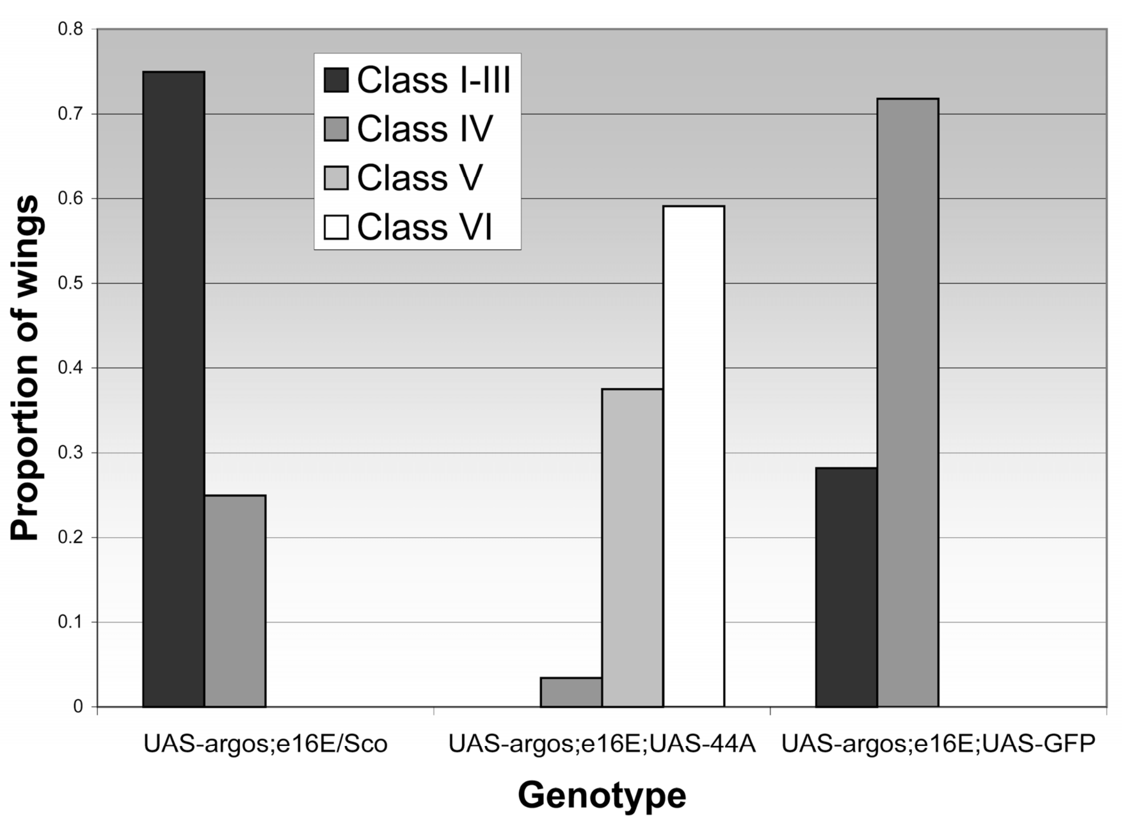


Figure 4-5. Socs44A misexpression rescues argos misexpression phenotypes. Concurrent misexpression of *Socs44A* (UAS-44A) rescues misexpression of *argos,* shifting the distribution of phenotypes, producing Class V and Class VI wing phenotypes (compare UAS-44A to Sco). Concurrent misexpression of GFP (UAS-GFP) did not produce class V or class VI wing phenotypes. Misexpression of UAS constructs in the posterior of the wing was achieved using the *engrailed*-GAL4 driver. For class definitions, see Fig. 4-3.

Table 4-1. Socs44A misexpression assay data.

Pathway	Heterozygous Mutant/transgene	Effect on Socs44A misexpression phenotype
dpp	dpp^{e87}	no change
	sax^{4}	no change
	Mad^{k00237}	suppress
Hh	hh^{2}	no change
	smo^{3}	no change
JAK/STAT	hop^{msv}	enhance
	hop^{c111}	enhance
	UAS-hop	socs rescue
	$stat92E^{06346}$	enhance
	UAS-Socs36E	no change
	Df(2R)NCX10	no change
	$dome^{468}$	enhance
	UAS-Socs44A	enhance
MAPK	Sos^{X122}	suppress
	Df(2R) 3F18	suppress
	UAS-DNDER	ND - lethal
	$Egfr^{k05115}$	no change
	$argos^{W11}$	enhance
	UAS-argos	socs rescue
	sev^{14}	no change
	$Ras85D^{e1B}$	suppress
	$Rho1^{k02107b}$	no change
	UAS-λtop	ND - lethal
	top^{18A}	no change
	$Gap1^{B2}$	no change
	$vap^{KG02382}$	no change
Notch	$UAS\text{-}N^{WT}$	ND - lethal
	$UAS\text{-}N^{intra}$	ND - lethal
TGF	tkv^{7}	no change
	put^{135}	suppress

Mutants and transgenics were placed in the *engrailed*-GAL4; UAS-*Socs44A* misexpression background. Wings were analyzed for suppression or enhancement of the *Socs44A* misexpression phenotype. Several transgenes were lethal when misexpressed with *engrailed*-GAL4; therefore interactions could not be determined (ND-lethal). Concurrent misexpression of *Socs44A* rescued the phenotypes associated with *engrailed*-GAL4 misexpression of *hop* (UAS-hop) and *argos* (UAS-*argos*).

Chapter V

Attempts to isolate a loss-of-function *Socs44A* mutation

INTRODUCTION

Evidence presented in previous chapters suggests that Socs44A interacts with multiple signal transduction cascades. However, these studies utilized misexpression analysis of Socs44A function. While ectopic expression demonstrates the capacity of Socs44A, it may also overestimate its functions, as elevated levels of protein or inappropriate location of expression could result in non-specific interactions. Misexpression studies of SOCS in mammals overestimated their physiological roles, as was the case for SOCS-3 and SOCS-5. Initial misexpression studies of SOCS-3 indicated that it might be a pleiotropic inhibitor of cytokine signaling as it is capable of regulating multiple cytokines, including IL-10 (Alexander, 2002). However, STAT3 activation induced by IL-10 was not enhanced in *Socs3*$^{-/-}$ macrophages. Misexpression of SOCS-5 indicated that it might be involved in T cell development and differentiation (Seki et al., 2002); however, analysis of a recently generated SOCS-5 knockout indicated that it was dispensable for these functions (Brender et al., 2004). Misexpression could also potentially lead to a qualitative error in the misjudgment of Socs44A function, as was the case for murine SOCS-2. When expressed at low levels, SOCS-2 inhibits growth hormone signaling, but at higher levels, it actually stimulates GH signaling (Favre et al., 1999). Finally, we have already demonstrated context specificity for Socs44A action that is consistent with the mammalian model. Therefore, the interpretation of expression of *Socs44A* out of context must be guarded.

In the absence of an available mutant, we utilized several overlapping deficiencies to study the loss-of-function of *Socs44A*. These deficiencies behaved consistently with our expectations based on misexpression analyses. Nevertheless, these deficiencies are quite large, removing multiple genes in addition to *Socs44A* (Table 5-1) that theoretically could be responsible for the interactions we observed. Analyses utilizing a loss-of-function allele of *Socs44A* would not be plagued with these issues. It would allow for the analyses of the endogenous activity of the protein in its appropriate context without the ambiguity associated with the deletion of multiple genes. Therefore, to validate findings presented in this work, several avenues designed to isolate a *Socs44A* loss-of-function mutation were pursued. This chapter will discuss progress on these avenues.

RESULTS

P-element excision

The *Drosophila* P-elements are perhaps the best-studied eukaryotic transposons. They are a widely exploited tool used by geneticists to manipulate the *Drosophila* genome (Engels, 1997). The *Drosophila* community has developed several strategies for targeted manipulation of the fly genome using a variety of P-elements (Cooley et al., 1988; Parks et al., 2004; Sentry and Kaiser, 1992; Spradling et al., 1999; Thibault et al., 2004; Timakov et al., 2002; Tower et al., 1993). These strategies are largely made possible by the engineering of P-elements that lack the enzyme transposase, required for excision and integration of the element (Beall and Rio, 1997; Beall and Rio, 1998; Doll et al., 1989; Kaufman et al., 1989; Rio and Rubin, 1988). P-element induced mutagenesis has allowed for the isolation of several components of the *Drosophila* JAK/STAT

cascade. *Marelle*^{6346} (a synonym for *stat92E*) was recovered in a P-element mobilization screen for autosomal zygotic lethal mutations exhibiting specific maternal effect phenotypes (Hou et al., 1996). The *Drosophila* JAK/STAT receptor, *domeless*, was identified in two separate screens of P-element insertions. The first was designed to identify novel factors in posterior spiracle development (Brown et al., 2001). The second screen sought modifiers of *GMR*-GAL4 driven misexpression of *upd* (Chen et al., 2002a).

Drosophila P-elements utilize cut-and-paste transposition generating a double strand break in the chromosome (Engels et al., 1990; Gloor et al., 1991; Kaufman and Rio, 1992) resulting in a variety of excision products (Gloor et al., 2000; O'Brochta et al., 1991). Precise excisions involve removal of the P-element without affecting adjoining DNA, although a P-element footprint is left behind (Engels et al., 1990; Nassif and Engels, 1993). Imprecise or "dirty" excisions occur when the P-element removes flanking DNA sequences when it is excised, resulting in a deletion surrounding the original P-element locus (Gloor et al., 2000).

There are several characterized P-element insertions in the vicinity of *Socs44A* on chromosome 2R (Fig. 5-1). We attempted to imprecisely excise one of these P-elements, in hopes of deleting the *Socs44A* locus, generating a null allele. One of these elements, KG03963, is homozygous viable and is located approximately 5 kbp upstream of the 5' end of the *Socs44A* locus. This element contains the mini-*white* gene which permits the rescue of *white*$^{-}$ mutants by tracking the P-element via the red eye color it imparts on the animal (Roseman et al., 1995). This element was excised (see scheme, Fig. 5-2) using Δ2-3 (Robertson et al., 1988) as the source of transposase. Of 109 independent excision events recovered, 102 remained viable, while 7 were lethal as homozygotes. The lethal

excision events were then subjected to complementation analysis with lethal P-element insertions that flank either side of *Socs44A* (Fig 5-1). All seven lethal excisions complemented both P-elements, indicating that those loci remained intact in these animals (Table 5-1). The lethal excision events were balanced over the P{Act-GFP}CyO balancer, which ubiquitously expresses GFP under the Actin promoter. This allowed recovery of DNA from homozygous embryos and larvae; DNA was collected from homozygous adults (marked by lack of the CyO marker) for the viable excision events. All 109 lines were then tested by PCR for the presence of the *Socs44A* locus (Table 5-1). Specifically, the 5' portion of *Socs44A* proximal to the KG03963 insertion site was PCR amplified from animals homozygous for the excision. As a positive control, a portion of the *unpaired* locus was simultaneously amplified for each line. For the lethal excisions, selection of homozygous animals was confirmed by failure to PCR amplify GFP. Furthermore, the intergenic region 5' of *Socs44A* was also amplified from animals presumed to have imprecise excision events (Fig. 5-1). These data indicate that all of the excisions recovered failed to delete *Socs44A*.

EMS mutagenesis

Since its introduction in 1968, Ethyl methanesulphonate (EMS) has been the most commonly used chemical mutagen in *Drosophila* genetics (Lewis and Bacher, 1968). EMS is an alkylating agent that produces G/C to A/T transition class mutations (Pastink et al., 1991). In theory, the standard 25mM dose of EMS will lead to a 1 in 1000 mutation rate for the average gene (Greenspan, 1997). However, there is considerable variation to this estimate based largely on the size of the coding region and the number of

critical amino acids in the gene target. Nonetheless, EMS mutagenesis screens have successfully produced mutant alleles of several of the known *Drosophila* JAK/STAT components. Mutations in *hopscotch* were recovered in a screen for maternal effect lethal genes exhibiting segmentation defects (Binari and Perrimon, 1994; Perrimon and Mahowald, 1986). The *stat92E*^{HiJak} allele was isolated in a mutagenesis screen for dominant autosomal suppressors of the hyperactive *hop*$^{Tum-l}$ allele (Hanratty and Dearolf, 1993; Yan et al., 1996a). *Upd* alleles were isolated in an elaborate screen involving a duplication of polytene segment 17 of the X chromosome placed on the Y chromosome which allowed for the recovery and complementation analyses of zygotic lethal mutants in that segment of the X chromosome (Eberl et al., 1992).

Our lab, in collaboration with several others, embarked upon a large scale F_3 EMS screen designed to isolate mutations of several genes located on the second chromosome: *Socs36E, Socs44A, StIP,* and *Ric*. *Socs36E* and *Socs44A* have already been described in detail within this text. StIP (Stat interacting protein) is a WD-40 repeat containing protein that has been shown to interact with both JAK and STAT in mammalian systems. Biochemical analyses suggest that it may serve as an adaptor molecule that facilitates STAT recruitment to the receptor complex (Collum et al., 2000). Biochemical analysis of Ric indicates that it is a small GTPase that binds Calmodulin (FlyBase Consortium, 2003; Wes et al., 1996). Ric has not been associated with JAK/STAT signaling in *Drosophila*.

Starting with a line carrying an isogenized second chromosome containing *cinnabar (cn)*, *brown (bw),* and *speck (sp)* markers, 600 males were exposed to 25mM EMS per the standard protocol (Ashburner, 1989). These males were then mated *en*

masse to *y w; Sco/CyO* females. Of the progeny collected, 7,477 males were individually mated again to three *y w; Sco/*CyO females. From these crosses, 3,986 balanced stocks were recovered, each carrying a mutagenized second chromosome (the complete scheme can be found in Fig. 5-3). Approximately 62% of these stocks were homozygous viable for the mutagenized chromosome, assayed for by the presence of flies with wild-type wings in the CyO balanced stocks. From this we can extrapolate that each mutagenized stock carries 1.5 recessive lethal mutations on the second chromosome (Ashburner, 1989) for a total of 5,979 lethal hits. Although the precise number of essential genes in the *Drosophila* genome is not known, it has been estimated that approximately 3,600 genes are essential (Miklos and Rubin, 1996). The second chromosome accounts for approximately 40% of the entire *Drosophila* genome (Adams et al., 2000), thus containing approximately 1,440 essential genes. To carry this line of reasoning further, we would then expect that our screen hit each lethal gene on the second chromosome on average 4.15 times, suggesting that we approached saturation.

While similar strategies were employed to isolate mutations in all four genes, I will focus on our efforts to obtain loss-of-function alleles of *Socs44A*. There are several deficiencies located in the vicinity of *Socs44A* (Fig. 5-4). These deficiencies were tested by complementation with molecularly defined P-elements that flank either side of *Socs44A* to determine which deficiencies uncovered the *Socs44A* locus. All 3,986 balanced stocks were crossed to the *Df(2R)CA53* deficiency, which contains the *Socs44A* locus, and tested for complementation. Those lines that failed to complement *Df(2R)CA53* were submitted to secondary screens against another deficiency that takes out the *Socs44A* locus (*Df(2R)NCX10*) as well as deficiency that does not take out

Socs44A (*Df(2R)Drl*^{rv18}). Of the 3,986 lines, 42 lines failed to complement *Df(2R)CA53* and *Df(2R)NCX10* and successfully complemented *Df(2R)Drl*^{rv18} and were retained as candidate *Socs44A* mutants.

Of the 42 candidate mutants, 4 were viable as homozygotes while the remainder were lethal. None of the candidates exhibited dominant phenotypes. When crossed to a *Socs44A* containing deficiency, four of the candidates displayed wing vein defects. Three lines were marked by a loss in the most distal portions of L5, while one line had weak ectopic wing veins that did not seem to be localized to a particular portion of the wing. The remaining 38 lines were lethal when crossed to the *Socs44A* deficiency. All 42 of the candidate mutants were crossed pair-wise in a complementation analysis to determine the number of genes represented in the collection. The results of this analysis were not easily interpreted as inconsistent complementation behavior was observed. These behaviors may have been the result of semilethality, intergenic complementation, or some other unknown phenomenon. Nonetheless, we estimate that there are no more than 23 complementation groups present in the *Socs44A* candidate collection (Table 5-2).

As detailed in the previous chapter, Socs44A interacts with the EGFR/MAPK pathway. I developed an assay to test this interaction based on modification of the distribution of phenotypes seen when *argos* is misexpressed using the *engrailed*-GAL4 driver. Deficiencies for Socs44A shifted the distribution to the more severe phenotypic classes (Fig. 4-4). Loss-of-function mutants generated in the EMS screen would be expected to behave in a similar fashion. Representatives from each of the 23 complementation groups were subjected to this *argos* misexpression interaction assay. The distribution of phenotypes generated by each line was then compared to *Df(2R)CA53*

as well as *Sco* using Fisher's Exact Test (Armitage and Berry, 1994) and are presented in Table 5-2 (statistical analyses courtesy of Chris Saunders). Of particular interest is candidate 7267, of which we did not observe anything that would lead us to believe that there is a statistical difference between its behavior and that of the *Socs44A* deficiency *Df(2R)CA53* (p = 1). Several other candidates showed interactions by this assay and are identified in Table 5-2.

In addition to the *argos* interaction assay, representatives for several complementation groups were selected for sequencing (Table 5-2). These lines were balanced over the P{Act-GFP}CyO chromosome. Embryos and larvae that were homozygous for the mutagenized chromosome were selected by virtue of their lack of GFP fluorescence. The *Socs44A* locus was then PCR amplified from genomic DNA isolated from these animals. Sequence analysis from these lines revealed no molecular lesions within the amplified region when compared to the sequence of the original isogenic *cn bw sp* chromosome.

DISCUSSION

This chapter chronicles efforts to obtain a loss-of-function allele of *Socs44A*. I had hoped that the isolation of a *Socs44A* hypomorph would allow me to validate my previous findings implicating roles for *Socs44A* in JAK/STAT and EGFR/MAPK signaling. Furthermore, this reagent would be indispensable in determining the roles of *Socs44A* in *Drosophila* development. Unfortunately, there are no guarantees when undertaking random mutagenesis screens that attempt to target a specific gene for mutation. I attempted both a P-element excision as well as a saturating F_3 EMS screen

against *Socs44A* deficiencies. Although not all of the candidates from the EMS screen have been analyzed completely, recent developments in the *Drosophila* community indicate it is impossible that our strategy would produce the desired mutation. These implications will also be discussed.

P-element excision mutagenesis

I attempted to excise a viable P-element located approximately 5kb away from the *Socs44A* locus. There are several variations of P-elements in use by *Drosophila* geneticists, each of these elements exhibits different behavior in regards to the frequency of excision and the class of the resulting molecular lesion produced by excision. Furthermore, the location of the P-element insertion may also affect its behavior in these regards (Gloor et al., 2000; Handler, 2001; Spradling et al., 1999). Of the 215 attempted excisions, 109 independent excision events were recovered (excision evidenced by loss of the *white* eye marker), indicating that 51% of the males gave rise to some progeny with excision events. Of these events, 102 remained viable, while in the other 7, excision resulted in homozygous lethality. It is likely that the vast majority of the 102 viable lines represent precise excision events, as this phenomenon is often reported as the most common P-element excision event (Gloor et al., 2000; Sentry and Kaiser, 1992). I can presume that the seven lethal lines represent P-element excision events that were not precise because the original KG03963 insertion is homozygous viable. PCR amplification of genomic DNA from these animals revealed that none of these events excised the *Socs44A* locus or the adjacent 5' region. I can assume these excisions were simply not large enough to remove *Socs44A*.

Several labs have conducted imprecise excision screens using P as well as other transposable elements and there is no consistency in the size or the rates in which deletions are recovered. Johnson-Schlitz and Engels (1993) excised a P-element from the *white* locus, generating 65 excision events. None of these deleted more then 905bp of genomic DNA surrounding the insertion. In stark contrast, Mohr and Gelbart (2002) excised *P{wHy}* from multiple locations on the second chromosome. This element is a hybrid between P and *hobo* type transposons. They produced a total of 66 deletions (of 138 recombinants), of which 53 were greater than 10kb. In another excision screen, Preston, Sved, and Engels (1996) excised *P{CaSpeR}Cp1*^{50C}, generating 243 recombinant lines. Of these 21 were short deletions less than 650bp while 24 were much larger; the largest measuring over 100kb by cytology. In this study, Preston et al noticed that the breakpoints of the deletions often correspond to P-element insertion site preferences. These include euchromatic regions often near the 5' end of genes and target octamers similar to sequence GGCCAGAC (Engels, 1989; Kelley et al., 1987; O'Hare and Rubin, 1983). This suggests that P-element mobilization behavior can be used as a guide for imprecise excision activity. Thus, we would expect that the majority of deletions would be small because P-elements preferentially mobilize to local sites within 2kb of the insertion site (Tower et al., 1993), although 2 out of 36 mobilizations in the Tower et. al. study moved as far as 128kb away.

Why did the P-element excision screen fail? The scope of the screen was well within that of those describe above, all of which generated deletions larger than the 5kb needed to remove *Socs44A*. However, it is known that P-elements tend to insert themselves into "hotspots" in the genome (Spradling et al., 1999), which may have some

bearing here because breakpoints apparently mimic insertion site preferences. Perhaps *Socs44A* is not in a "hotspot" for the particular P-element used. Also, it is known that there is directionality to the deletions caused by P-element excision (Preston et al., 1996). It is possible that the starting P-element was not oriented in the proper direction for efficient deletion of the *Socs44A* locus.

Although the P-excision screen failed to generate a *Socs44A* deletion, 7 lethal lines were recovered. Based on complementation analyses, the breakpoints of these presumed excisions fall between the *l(2)k16504 – Socs44A* interval (Fig. 5-1). This genomic region is approximately 55kb and contains 13 genes. Of these, five genes have been named while the others are predicted loci. One of these, CG11210, has been disrupted by a P-element insertion and is lethal (Bellen et al., in press). This predicted gene does not have any functionally characterized domains or motifs, nor has any analyses been published on the mutant. Four of the 13 genes have mutant alleles reported in FlyBase that are viable (FlyBase Consortium, 2003). The remaining seven genes include *sut1*, *sut2,* and five predicted genes. It is likely that the lethal excisions created contain a small deletion around the KG03963 insertion that includes CG11210 and/or one or more of these loci.

EMS mutagenesis

In a large collaborative effort, an F_3 EMS screen was undertaken to isolate mutations in several genes on the second chromosome, including *Socs44A.* A total of 3,986 balanced lines were generated, each containing a unique mutagenized second chromosome. Based on the assumption that the second chromosome contains

approximately 1,400 essential genes, our screen should have hit each of these an average of 4.15 times. This suggests that we should have isolated a mutation in *Socs44A* provided that such a mutant would produce a discernable phenotype when placed over a deficiency for the locus. This reasoning exposes a potentially fatal weakness in our screening strategy. If a *Socs44A* null animal has no visible phenotype and is not essential for viability, then it would be impossible to recover such an animal in our screen even though we approached saturation of the genome. Based on studies using *engrailed*-GAL4 misexpression of *Socs44A* we hoped that in lieu of lethality we might have detected a wing vein defect in potential *Socs44A* mutants. Indeed, we recovered several candidate mutants that lacked wing vein material in the posterior compartment near the PCV, consistent with our expectations. However, this phenotype is likely caused by another mutation within the *Df(2)CA53 - Df(2R)Drl*^{rv18} interval. The reasoning for this is provided in the next section.

These results are consistent with a multitude of mutagenic screens, some of which were designed to target JAK/STAT components directly. Many large scale EMS screens, including the Nobel winning work of Christiane Nüsslein-Volhard and Eric Wieschaus, that isolated most of essential the mutations that pattern the embryo did not recover a *Socs44A* allele (Nusslein-Volhard, 1979; Nusslein-Volhard and Wieschaus, 1980; Nusslein-Volhard et al., 1984). Other screens for female sterility (Schupbach and Wieschaus, 1989; Schupbach and Wieschaus, 1991) or maternal effect mutants, one of which isolated *hopscotch* (Nusslein-Volhard et al., 1987; Perrimon et al., 1989; Schupbach and Wieschaus, 1986) also failed to isolate alleles of *Socs44A*. Erica Bach spearheaded a screen for dominant modifiers of a GMR-GAL driven *unpaired*

misexpression rough eye phenotype in an effort to uncover new interacting JAK/STAT genes, but also failed to isolate a *Socs44A* mutation (Bach et al., 2003). A similar screen for modifiers of GMR-GAL driven misexpression of *argos* also failed to isolate a *Socs44A* mutant (Taguchi et al., 2000). The Hou lab at the National Cancer Institute, which studies JAK/STAT signaling, recently attempted a P-element mobilization screen targeting the second chromosome. They report the isolation of 2,355 second chromosome insertions in 850 different genes, but they failed to produce an insertion in the *Socs44A* locus (Oh et al., 2003). Another P-element mutagenesis of the second chromosome isolated 2,711 independent insertions, none of which disrupted *Socs44A* (Torok et al., 1993).

All twelve of these screens had the potential to identify alleles of *Socs44A*, but none were successful. All twelve of these screens (this list is by no means exhaustive) employed different strategies, be they various forward genetic approaches, modifier screens, or P-element mobilizations/disruptions. Add to that list our labor intensive F_3 EMS screen of 3,986 mutagenized lines against *Socs44A* deficiencies and my local P-element excision screen.

Ultimately, an allele of *Socs44A* was found as part of a large scale transposon mediated gene disruption project (Thibault et al., 2004). In this endeavor, a modified lepidopteran transposon was successfully mobilized into the *Socs44A* locus. This approach was successful for several reasons. First, it did not depend on an observable phenotype; instead, transposon insertion sites were molecularly defined. Second, the use of the PiggyBac transposon eliminated the insertional bias associated with P-elements (Spradling et al., 1999). Although it has not been determined if PiggyBac has its own

preferential insertion sites, it is clear that PiggyBac does not share the same "hotspots" as P-elements. Finally, this disruption project benefited from vast financial resources and a large number of personnel, permitting a scale not possible in previous screens.

Although it is unlikely that any of the candidate mutants from the EMS screen are bona fide *Socs44A* alleles, I did recover several lines, 7267 in particular, that showed significant enhancement of the *argos* misexpression phenotype (Table 5-2). Based upon the failure to complement *Df(2R)CA53* and success in complementing *Df(2R)Drl*^{rv18}, these lines contain mutations within the 44A3 to 44B6 cytological region. This region encompasses approximately 120 kb of DNA and houses 22 genes in addition to *Socs44A* (Lewis et al., 2002). Of these, 8 have been studied but none of these are known to be involved in JAK/STAT or EGFR/MAPK signal transduction. The remaining 14 are predicted genes that have not been studied (FlyBase Consortium, 2003). Interestingly, one of the unknown genes is predicted to be an ARF GTPase activating protein (GAP) based on sequence similarity. If this gene is involved in EGFR/MAPK signaling, we could postulate that our EMS induced candidate mutations are at that locus and not *Socs44A*.

Recent developments

With the completion of the *Drosophila* genome (Adams et al., 2000), there has been increased effort in the generation of community-wide tools for genetic analysis of *Drosophila*. Among these is the large scale P-element mobilization mutagenesis spearheaded by several prominent labs in the fly community designed to disrupt each and every gene in the genome (Bellen et al., in press). A similar large scale effort, conducted

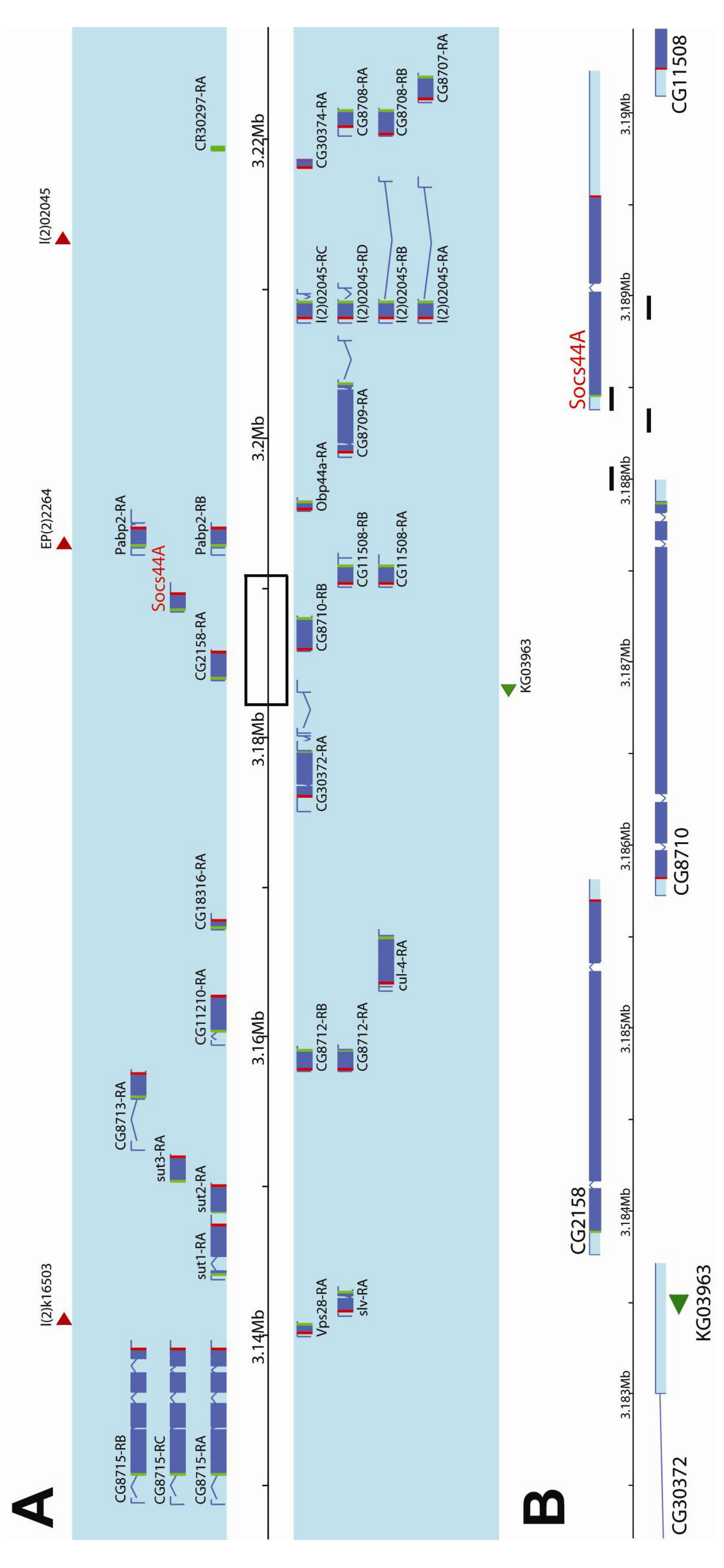
by Exelixis using a modified form of the lepidopteran *PiggyBac* transposon (Handler and Harrell, 1999) successfully disrupted the *Socs44A* locus (Thibault et al., 2004). This mutant was made available to the *Drosophila* community only few months prior to this writing. The transposon insertion is located at the 5' end of the locus and disrupts the codon that encodes D28 (Fig. 2-4); suggesting that the allele is null. These animals are homozygous viable as adults and do not exhibit any discernable phenotype.

There is a possibility that this disruption of the *Socs44A* locus is not equivalent to a null allele. It is possible that a constitutive promoter within the transposon (Thibault et al., 2004) might be facilitating the production of a functional *Socs44A* product, making this a hypomorphic allele. Mutational analyses of mammalian SOCS genes discovered that the 5' end upstream of the KIR (see Fig. 2-1) are dispensable for function with regard to SOCS-1 and SOCS-3 (Nicholson et al., 1999). Therefore, deletion and/or alteration of the amino terminus of the protein may not have an effect on the function of the protein. Yet animals heterozygous for the transposon insertion and the *Df(2R)CA53* deficiency were also viable and presented no phenotypes (J. Rawlings, unpublished). Another possibility is that the transposon insertion is a gain-of-function allele because the *PiggyBac* transposon contains a UAS element which, although highly unlikely, could be generating a weak ectopic expression even in the absence of a GAL4 driver. However, when crossed to the *engrailed*-GAL4 driver, no ectopic wing veins (or any other visible phenotype) were observed (data not shown). Furthermore, the modified *PiggyBac* element also contains "splice-trap" and transcriptional silencing elements (Thibault et al., 2004), which would presumably prevent any ectopic or leaky expression of the disrupted gene

Taken as it is, the *PiggyBac* transposon disruption of *Socs44A* has several important ramifications. First, it is not likely that the *PiggyBac* insertion is causing leaky or ectopic expression of *Socs44A* and given the composition of the transposon and its point of insertion, the allele is most likely null. Therefore, it is likely that *Socs44A* is not an essential gene that has a straightforward adult phenotype. Second, the EMS screen was predicated on the ability to screen for mutations based on complementation to a deficiency, meaning that we had to be able to observe a phenotype in the mutant. Since the *Socs44A* disruption had no phenotype, it is highly unlikely that the candidate mutants generated by the EMS screen are *Socs44A* mutants. Finally, even if the *PiggyBac* insertion was not null, the most straightforward next step would be to conduct an imprecise excision of this element, which would guarantee a null allele. However, the *PiggyBac* transposon exhibits a remarkable efficiency in precise excisions (Thibault et al., 2004), making this strategy moot.

Nevertheless, this transposon insertion merits further scrutiny. It may have an adult phenotype not detected by direct visual observations. Furthermore, it could have transient phenotypes present in earlier development. One possibility is that it could affect somatic follicle cell populations in oogenesis, as this process is governed by multiple signaling pathways including JAK/STAT and EGFR/MAPK. These pathways in particular act to pattern various cell populations at the termini of developing egg chambers (Beccari et al., 2002; Gonzalez-Reyes et al., 1995; McGregor et al., 2002; Roth et al., 1995; Silver and Montell, 2001; Xi et al., 2003). Since these distinct populations present specific cell surface markers, we could examine *Socs44A* mutant ovaries for changes in the distribution and/or differentiation of these cell populations. Another

possibility is that Socs44A could have an effect on larval hematopoiesis, which also depends upon JAK/STAT signaling (reviewed in Evans et al., 2003; Luo and Dearolf, 2001; Zeidler et al., 2000). As in oogenesis, the specific larval blood cell populations also present markers (Lebestky et al., 2000) that could be exploited to determine if a *Socs44A* mutant has an effect on hematopoiesis. Although crystal cell derived melanotic tumors associated with dominant gain-of-function *hop* alleles were not observed (Hanratty and Dearolf, 1993; Harrison et al., 1995; Luo et al., 1995; Luo et al., 1997), it is possible that Socs44A has roles in the development and differentiation of other blood cell types. Examination of the distribution of blood cell types in *Socs44A* mutants would address this issue.



A
CG8715-RB
CG8715-RC
CG8715-RA
l(2)k16503
sut1-RA
sut2-RA
sut3-RA
CG8713-RA
CG11210-RA
CG18316-RA
CG2158-RA
Socs44A
Pabp2-RA
Pabp2-RB
EP(2)2264
l(2)02045
CR30297-RA
3.14Mb
3.16Mb
3.18Mb
3.2Mb
3.22Mb
Vps28-RA
slv-RA
CG8712-RB
CG8712-RA
cul-4-RA
CG30372-RA
CG8710-RB
CG11508-RB
CG11508-RA
Obp44a-RA
CG8709-RA
l(2)02045-RC
l(2)02045-RD
l(2)02045-RB
l(2)02045-RA
CG30374-RA
CG8708-RA
CG8708-RB
CG8707-RA
KG03963
B
CG30372
3.183Mb
KG03963
CG2158
3.184Mb
3.185Mb
3.186Mb
CG8710
3.187Mb
3.188Mb
Socs44A
3.189Mb
3.19Mb
CG11508

Figure 5-1. P-element map of the Socs44A region. (A) Genetic map of the *Socs44A* region illustrating relative positions of genes and P-element insertions (marked by arrowheads above and below the blue boxes). *Socs44A* (shown in red) is flanked on either side by the lethal P-elements (red arrowheads) l(2)k16503, EP(2)2264, and l(2)02045. The viable P-element insertion (green arrowhead) KG03963 was used in a P-element excision screen designed to make a small deficiency at the *Socs44A* locus (see text for more details). Each tick mark represents 10kb of DNA. (B) Enlargement of the boxed region shown in A. PCR primers used for amplification of *Socs44A* and upstream regions are indicated as bars above and below the scale bar. Each tick mark represents 500bp of DNA.

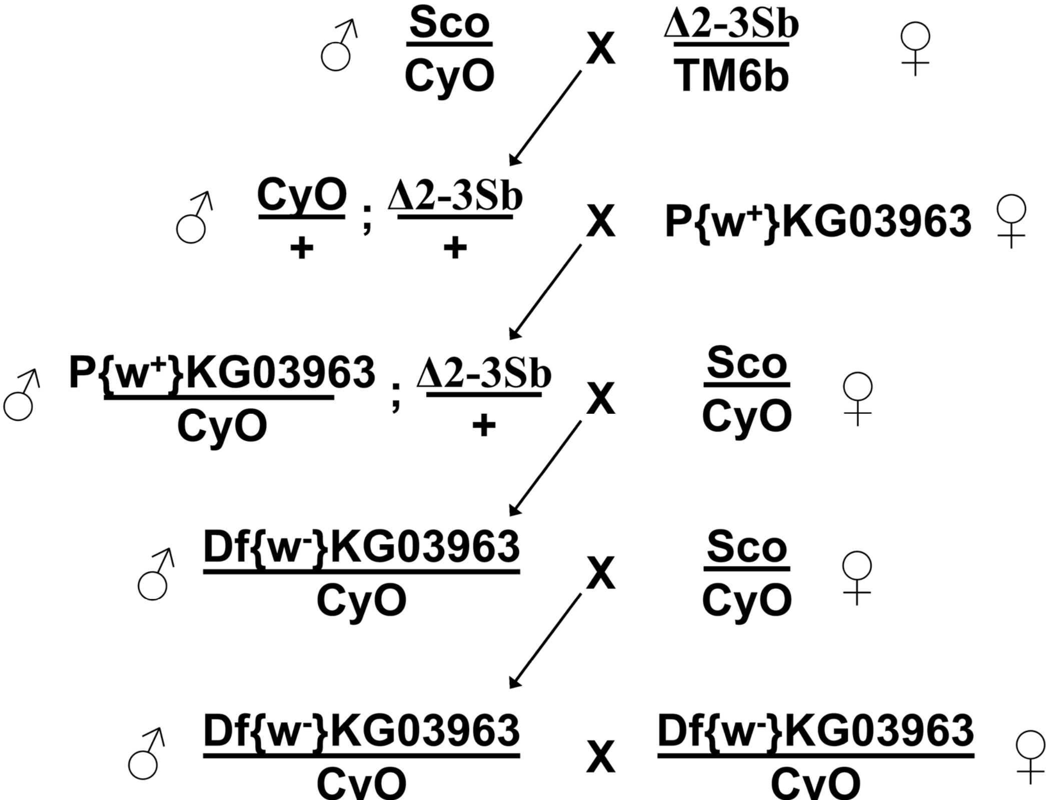


Figure 5-2. Scheme for imprecise P-element excision screen. The stable source of transposase, Δ2-3, was used to excise the viable KG03963 P-element. The element contains the *mini-white* gene which confers red eye color. The actual P-element excisions occurred in the germline of the 3rd generation. 109 excision events were selected for by the loss of eye color and balanced over the CyO balancer chromosome to establish individual stable stocks. Only the relevant chromosomes are shown and all crosses were done in a *w-* background.

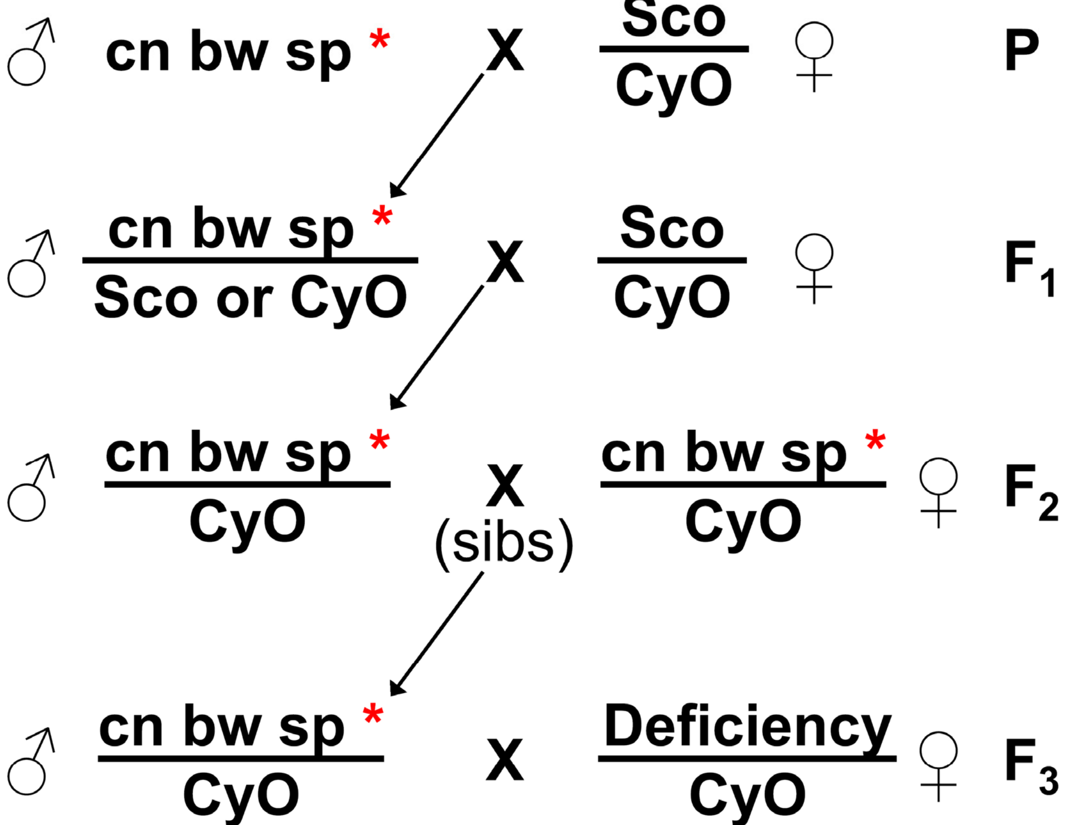


Figure 5-3. EMS mutagenesis scheme. 600 isogenized *cn bw sp* males were mutagenized with EMS (denoted by red asterisk). The scheme illustrates crosses to generate balanced stocks, each possessing a unique mutagenized chromosome. Each of 3986 balanced stocks that were recovered were screened in the F_3 generation using a deficiency for the gene of interest.

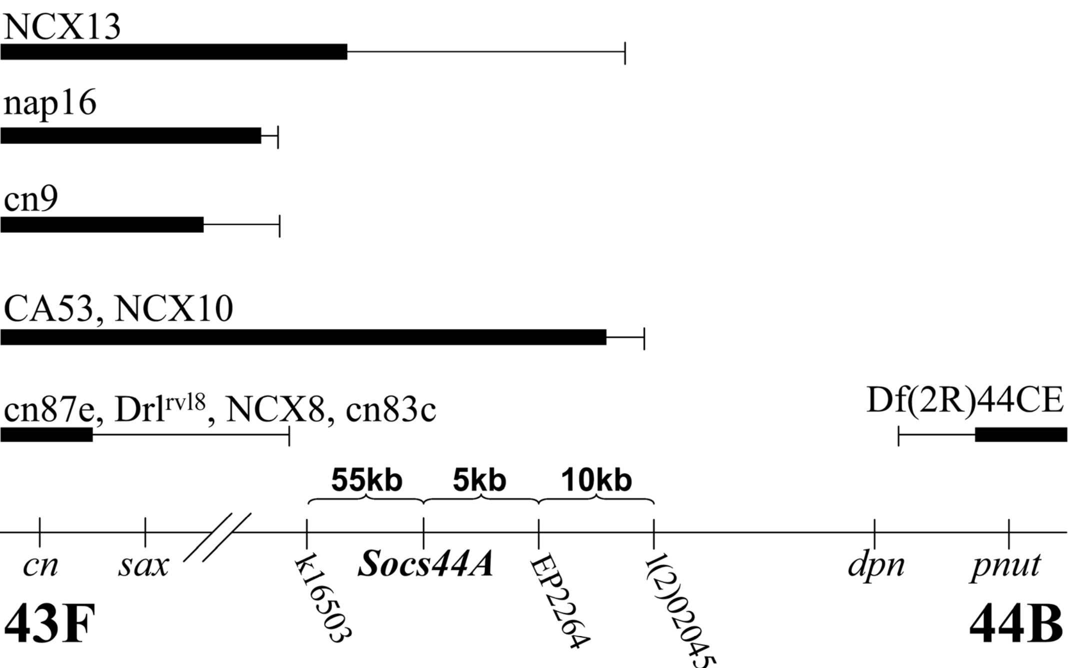


Figure 5-4. Deficiency map of the 43F – 44B region. Breakpoints of several deficiencies in the vicinity of *Socs44A* were mapped by complementation to several lethal P-elements that flank the *Socs44A* locus. Think bars represent the extent of the deficiency as mapped by complementation. The thin bars represent the theoretical limit of the breakpoint. Both *Df(2R)CA53* and *Df(2R)NCX10* deficiencies were found to include *Socs44A* and were used to screen for potential mutants. The *Df(2R)Drl^{rv18}* deficiency does not include *Socs44A* and was used as a means to pare down candidate mutants (see text for details).

Table 5-1. Summary of P-element excision screen.

Line	Complementation analyses k16503	EP2264	PCR Socs44A?
Viables			
102 lines	ND	ND	Yes
Lethals			
33	Yes	Yes	Yes
94	Yes	Yes	Yes
121	Yes	Yes	Yes
163	Yes	Yes	Yes
204	Yes	Yes	Yes
206	Yes	Yes	Yes
212	Yes	Yes	Yes

The viable KG0369 P-element was excised, generating 109 independent excision events, of which 7 were lethal as homozygotes. All lethals complemented the nearby lethal P-elements k16503 and EP2264 (for relative positions see Fig. 5-1) delimiting the outermost boundary of the presumed imprecise excision. The *Socs44A* locus was PCR amplified from homozygous animals from all 109 lines indicating that all failed to excise *Socs44A*. ND = not determined.

Table 5-2. Summary of EMS mutagenesis screen.

Group	members	Argos assay	p = Sco	p = Df	Sequenced
Control	Sco				
Control	CA53/NCX10				
1	265, 680, 1824, 2589, 2933*, 3930, 4504, 6792, 7355*		0.2671	<0.0001	265
2	**565*, 3569*, 3782**		**0.0011**	**0.0488**	**3782**
3	1014* , 3501, 3926		0.0026	<0.0001	3926
4	2014*		<0.0001	<0.0001	2014
5	1547*, 2608, 3840		0.2937	<0.0001	1547
6	1938, X094*		0.3714	<0.0001	NO

Table 5-2, continued.

Group	members	Argos assay	p = Sco	p = Df	Sequenced
Control	Sco				
Control	CA53/NCX10				
7	3488, 4339*, 6681*		0.0229	<0.0001	6681
8	**4613***		**0.0157**	**0.0019**	**NO**
9	1679*		0.1175	<0.0001	1679
10	**7267***		**<0.0001**	**1**	**NO**
11	1080*		0.5356	<0.0001	NO
12	754*		0.2618	<0.0001	754

Table 5-2, continued.

Group	members	Argos assay	p = Sco	p = Df	Sequenced
Control	Sco				
Control	CA53/NCX10				
13	1997*		0.1717	<0.0001	13
14	**2004***		**0.0043**	**0.2235**	**NO**
15	2008		<0.0001	<0.0001	NO
16	4094*		<0.0001	<0.0001	16
17	4959*		0.4616	<0.0001	17
18	**705***		**0.0012**	**0.0015**	**NO**

Table 5-2, continued.

Group	members	Argos assay	p = Sco	p = Df	Sequenced
Control	Sco				
Control	CA53				
19	5809*		0.0083	<0.0001	19
20	42*		<0.0001	0.0015	42
21	**4761***		**<0.0001**	**0.0367**	**NO**
22	5952*		0.3615	<0.0001	5952
23	**3504***		**0.0388**	**0.0148**	**NO**

Table 5-2, continued.

We isolated no fewer than 23 complementation groups (column one) in an EMS screen for alleles of *Socs44A*. Second column lists the members of each complementation group. An asterisk indicates which line(s) of that complementation group were subjected to the *argos* interaction assay. Third column shows that distribution of wing phenotypes in the *argos* interaction assay. The x-axis, from left to right, is class I – class IV. The y-axis is proportion of wings in each class; each gridline is 10%. Fisher's Exact Test (Armitage and Berry, 1994) was used to compare this distribution to that of *Scutoid* (Sco, fourth column) and *Df(2R)CA53* or *Df(2R)NCX10* (Df, fifth column). The p value indicates the probability that the candidate mutant phenotype distribution is similar to that of *Sco* or the *Socs44A* deficiency. A p = 1 indicates we did not observe anything that would lead us to believe that there is a difference in the distribution of phenotypes between the candidate and *Sco* or the deficiency. Those lines that behaved like expected *Socs44A* hypomorphs are indicated in red. Finally, the last column indicates which members of the complementation group, if any, were sequenced.

Chapter VI

Multi-BLAST: database query for multi-domain proteins or coding regions

INTRODUCTION

The reverse genetic investigation of the three *Drosophila* SOCS proteins was predicated on not only their possession of both an SH2 and SOCS domain, but also the specific arrangement of those domains within their amino acid sequence. In fact, this modular arrangement of domains was instrumental in data mining the *Drosophila* genome for potential SOCS homologues. The consensus SOCS domain (Hilton et al., 1998) was used in a tBLASTn query of the *Drosophila* genome. The results of this search were queried again for the presence of an SH2 domain in upstream sequences. This arduous task inspired the creation of Multi-BLAST, a novel post-processor of the popular BLAST suite of algorithms, designed for the efficient retrieval of multiple domain containing proteins from public databases. This chapter describes the design and implementation of Multi-BLAST. Multi-BLAST is available at: http://genome.kbrin.uky.edu/multiblast/

Protein structure and function has proven to be surprisingly modular; the domains that a protein possesses confer its functions. By arranging domains in different combinations and orientations, large numbers of polypeptides, each with unique functions, are produced. This phenomenon occurs naturally via genome rearrangements during evolution. Additionally, alternative splicing can create proteins differing in their modular domain arrangement from a single primary transcript. Although there are a

multitude of algorithms designed for the identification and retrieval of sequences by similarity, the approach to protein and genome analysis that seemed lacking was a simple method to identify proteins or coding regions based upon a specific modular design. *In silico* efforts to analyze protein sequences based on modular architectures have produced large databases of protein domains such as SMART-A (Letunic et al., 2004; Schultz et al., 2000; Schultz et al., 1998) and Pfam (Bateman et al., 2000; Bateman et al., 2004), and methods to effectively find these individual domains.

Perhaps the most widely used method to identify similar protein sequences or coding regions is the BLAST suite of programs (Altschul et al., 1997). BLAST and related programs attempt to find overall similarity by dynamically building local alignments between a query and database sequences. This is accomplished by assigning a positive score to an alignment based on matches and similarities while assigning a penalty to mismatches and gaps. BLAST has limitations, especially when searching for proteins with multiple domains. First, BLAST discriminates against smaller domains because the BLAST score is length dependent, the longer the alignment, the better the score. Second, BLAST will stop building an alignment if it encounters a large enough span of dissimilarity to make continuing such an alignment cost prohibitive. For example, if one conducts a BLAST search against the nr protein database for proteins that contain both an SH2 domain and a SOCS domain by using mouse SOCS-1 (Accession AAD53324) as a query, the results are predominantly records that contain only the SH2 domain, primarily because the SH2 domain is ~70 amino acids long, while the SOCS domain is only ~40 amino acids. The long match to the SH2 domain is sufficient to generate a high BLAST score, while matches to the SOCS domain generate lower scores.

Although the desired sequences that contain both domains are in the long list of hits, sorting this list is not trivial. Motif searching using tools like PHI-BLAST (Zhang et al., 1998), Pfam, or the SMART database is a very powerful way to identify lists of domain-containing sequences. Nevertheless, identifying particular modular arrangements within proteins on that BLAST hit list is not trivial with the tools available. Additionally, these resources do not address the fact that the relative location of domains within a protein may be important to its function. Furthermore, when using motif searches like SMART-A, the user is confined to using the domain definitions provided, preventing one from inputting novel sequences into a query.

NCBI recognizes the need to identify proteins based on modular architectures; in that vein, they created CDART (Geer et al., 2002). This program accepts a protein sequence as a query and displays proteins with similar architectures. For CDART to be useful it must know the definitions of the domains of interest and be able to recognize those domains within the query sequence; one cannot search for novel domains and/or architectures. Additionally, CDART defines similar protein architectures as those proteins that have one or more similar domains to the query; the distance between domains is not considered. Finally, CDART does not accept nucleotide queries, nor can it perform searches against translated databases.

We designed a novel post-processor of the BLAST suite of programs, called Multi-BLAST, which allows the user to query protein or nucleic acid databases for multiple domain containing proteins or coding regions in which the domains are separated by a user-defined distance. A set of CGI, JavaScript, and Perl programs collect input from the user and conducts independent BLAST searches with user-defined

parameters for each input domain. Then the program compares the results to identify the sequences matching the modular design, and presents both the independent and combined domain search results to the user. The user interface couples this algorithm with the domain definitions of over 600 protein domains offered through the SMART-A, Pfam, and NCBI Conserved Domain databases. This communication presents the features of Multi-BLAST and illustrates how it improves upon the BLAST suite in identifying multiple domain containing proteins in sequence databases.

METHODS AND IMPLEMENTATION

Domain definition

In order to utilize this modular domain concept in the Multi-BLAST program, we first needed a domain definition. When classifying a protein domain, one compares it with sequences of known domains. Various algorithms like Hidden Markov Models (Eddy, 1998) have been used to classify protein sequences into domain families and to include these in massive alignments that determine a domain's consensus sequence resulting in databases such as Pfam and SMART. Our method capitalizes upon these databases and upon the flexibility of user input BLAST queries. Through the feature termed CD (conserved domain) Search, NCBI maintains current information regarding the domain definitions offered by these two databases. We therefore dynamically linked Multi-BLAST to the CD-search service, allowing Multi-BLAST to offer an extensive list of domains (either consensus or representative sequences) for use. Additionally, this dynamic linking also allows Multi-BLAST to offer direct links to the associated information regarding domains selected by the user. We realize that although the domain

definitions offered are extensive, they are nonetheless incomplete and that representative, rather than consensus, sequences are available for some domain definitions. Therefore, the Multi-BLAST program also allows the user to input novel sequences in addition to, or in lieu of, the predefined domain sequences offered. Furthermore, Multi-BLAST provides methods for modification of domain definitions if needed. Taken together, we believe that Multi-BLAST offers an unprecedented level of user control and flexibility in designing a database query.

Use of the BLAST suite of programs

Rather than design a novel core database search algorithm, we used the BLAST suite of programs for several reasons. First, the BLAST algorithms can effectively identify each domain queried since the sequence similarity within a domain is likely to be continuous. Second, the BLAST suite of programs is freely available and relatively easy to install and implement. Third, the BLAST suite of programs is well recognized, widely used, and accepted in the scientific community. Finally, the BLAST suite of programs allows Multi-BLAST the opportunity to accept nucleotide and protein queries and conduct searches against nucleotide, protein, and translated databases, providing the user with the utmost flexibility.

The Multi-BLAST program

The core concept of the Multi-BLAST program is quite simple. Instead of creating a single BLAST query containing multiple domains of interest, Multi-BLAST conducts independent, simultaneous BLAST searches for each domain entered. Multi-

BLAST currently accepts two domains in a single query. Each domain that is entered, either predefined or user-defined, is BLASTed according to the parameters set by the user. Multi-BLAST then compares the results for each search and compiles a list of records that appear in both search results. The BLAST-generated alignments are then compared between matching records against a set of range variables entered by the user. By entering a set of range values, the user has control over the relative position of each domain within positive results. The end result is that Multi-BLAST is flexible, efficient and easy to use.

User interface

On the surface, the Multi-BLAST user interface appears similar to the BLAST suite of programs (Fig. 6-1). The user can select various BLAST programs, databases, and several advanced BLAST options including expect threshold, scoring matrix, and word size. Beneath this, the interface is divided into two sections, one for each domain to be queried. For each domain, the user has the option of using domain definitions from the Conserved Domain Database (CDD) at NCBI (Marchler-Bauer et al., 2003) consisting of SMART, Pfam and other NCBI annotated domains. The SMART domains are organized based upon function or cellular location: signaling, extracellular, nuclear, or others. The Pfam and NCBI domains are organized alphabetically. All domain lists include a brief description of each domain; furthermore, each domain is dynamically linked to the CDD at NCBI, allowing instant access to associated information on each domain (descriptions, alignments, references etc). Next, a set of range values must be entered. The range values serve two purposes. First, for queries against nucleotide

databases this prevents ludicrous matches to large multi-gene records (e.g. *Drosophila* chromosomes). Second, this feature allows the user some control over the modular design of desired records retrieved by Multi-BLAST. Finally, the user has the option to have results sent via e-mail or to run the query interactively. The current version of this service runs on a BLAST server at the University of Kentucky with databases updated monthly. The simple code could be installed on essentially any BLAST server.

Results page

The results page generated by Multi-BLAST is divided into three parts (Fig. 6-2). First, Multi-BLAST gives a summary of the query, including the program, database, search parameters, and domains used. Second, Multi-BLAST offers links to the results of each individual domain BLAST search. These two features are useful for troubleshooting when no results are achieved from a given query. Finally, the user is presented with a list of descriptions and alignments of records that matched the search criteria for both domains. BLAST scores and Expect values are offered for each domain; an asterisk denotes multiple matches within the user specified range to a given domain, with the best BLAST score and E value reported. For example, the *Drosophila* dock protein contains three tandem SH3 domains and an SH2 domain. All of the SH3 domains that fall within the user specified range are shown in the alignment section, ranked by BLAST score with the best score and E value reported in the descriptions section along with an asterisk.

RESULTS

To test the ability and efficiency of Multi-BLAST, we compared the results when using both BLASTp and Multi-BLAST when querying NCBI databases for modular protein sequences. For each standard BLASTp search, a protein containing two known domains was used as a query; the Multi-BLAST query contained the appropriate domain representations provided by Pfam and/or SMART using a reasonable distance variable. In both types of searches, all parameters were set to default with the exception of the number of records to display in the results, which was set to maximum in all searches.

Multi-BLAST was extremely efficient in recovering protein sequence records containing both query domains (Table 1). All relevant two-domain proteins were recovered in every case. In at least one case (AAF00543 query of the nr database), Multi-BLAST recovered protein sequences that contain both domains but were not recovered in the standard BLASTp search. This was probably due to the fact that BLASTp search results depended on the particular query protein sequence in its entirety, but that Multi-BLAST results depended only on the domain definitions.

Generally, using a known homologue in a BLASTp search is the most efficient means to collect sequences that have a match to the domain architecture of the query. This strategy does not work as well when domains are not spaced exactly like the query or when the relative positions of domains differ from what is in the query sequence. Furthermore, in most cases, similarity to one domain is sufficient for a record to be included in the results. The end result is that along with desired records that contain both domains, the results page contains records that have only one of the query domains. For example, consider the SH2 and SH3 cell signaling domains found in most organisms,

including Drosophila melanogaster. To access all Drosophila protein sequences at NCBI containing both of these domains using a BLASTp strategy, one might use the Drosophila GRB2 homologue (DRK) as a query. DRK is an adaptor protein in the Ras/Raf signaling cascade that contains an SH3-SH2-SH3 modular architecture of these domains (Olivier et al., 1993). By using this protein, the chances of obtaining records with both domains in both possible orientations are optimized. The BLASTp search yielded 139 hits; upon closer inspection, only twelve of these hits actually contained both SH2 and SH3 domains. Furthermore, these twelve hits are not the first twelve presented. Indeed, one must look through the first 28 hits to find all twelve bona fide SH2 and SH3 domain-containing protein sequences. When a Multi-BLAST search is conducted against the Drosophila genome using the SH2 and SH3 domain representations from either SMART or Pfam with default parameters, the correct twelve (and only those twelve) results are displayed (Table 6-1). Similar results can be obtained using less promiscuous domain combinations. SOCS proteins (suppressor of cytokine signaling) are involved in the negative regulation of the JAK/STAT signal cascade. In addition to an SH2 domain, all SOCS proteins contain a novel carboxy-terminal SOCS domain. Eight mammalian SOCS proteins have been identified (Rawlings et al., 2004). To identify putative Drosophila SOCS homologues, one can use murine SOCS-1 as a query for a BLASTp against the Drosophila genome. This search yields fifteen hits, only two of which contain both SH2 and SOCS domains. A similar Multi-BLAST search using those domains yields just the two correct hits found using BLASTp. These examples, as well as similar searches using other domain combinations against different databases are illustrated in Table 6-1.

DISUCSSION

Multi-BLAST is a novel post-processor of the BLAST suite of programs designed to enhance its ability to find specific modular arrangements of protein domains within a protein or nucleic acid database. A common approach to searching databases for particular proteins using BLAST programs is to input a known homologue of the protein of interest. This approach is not efficient in acquiring desired records of specific modular proteins from protein databases. Many factors contribute to the appearance of "false positives", proteins without the intended domain architecture, in the results of BLAST queries. Similarity to one domain is often sufficient to garner a favorable BLAST score and it is not trivial to sort through the results to retrieve only the records that fit the intended modular query. Also, BLAST does not take into account the orientation of domains or distance between domains, both of which may be relevant to a particular search. In both of these instances, BLAST considers each portion as a separate database "hit." For example, if one uses the Drosophila VAV protein as a BLASTp query for other Drosophila proteins containing both an SH2 and SH3 domain, obvious matches such as Src49A and RasGAP receive low BLAST scores (30 for both) and Expect values (4.1 and 5.1, respectively) simply because the SH2 and SH3 domains of these proteins are in the opposite orientation as in the query, with the BLAST score reflecting the match to the best portion, not both domains. Furthermore, proteins like the Drosophila SL protein which contain the SH2 and SH3 domains in the proper orientation are "false negatives" in that they receive poor BLAST scores because the distance between domains differs from the query. In the case of SL, which appears in the middle of the results list (#30 out of 97 against Drosophila database; #472 out of 1172 against nr database), the

BLAST score (35, Expect 0.083) is based only on its SH2 domain. Also, domains that tend to be repeated within a protein sequence, such as Cadherin repeats, can cause problems in obtaining desired records since the repeated elements are enough to generate a high enough BLAST score making other domains present in the query insignificant (see Cadherin and Laminin G query, Table 6-1). Multi-BLAST takes all of these issues into account giving the user the utmost flexibility and control in designing database queries. By conducting independent BLAST searches, each domain is given equal weight in the results. The use of a range value allows the user to gain further control over which records are included in the results and also prevents multi-gene containing records such as cosmids or Drosophila chromosomes from being included in the results.

The Multi-BLAST program takes advantage of three publicly available lists of domain definitions (SMART-A, Pfam, and NCBI). It is important to remember that many of these domain definitions are representative sequences, not consensus sequences. One can access the Conserved Domain Database (CDD) information on each domain selected. From here, one can access the alignments used to generate the representative sequences. Also, when a protein domain is selected, the sequence appears in the query box. This gives the user flexibility to make any modifications to the domain sequence before the search is executed. Finally, Multi-BLAST gives the user the opportunity to enter novel or undefined sequences in lieu of or in addition to the domain definitions provided.

This communication illustrates the use of Multi-BLAST in searching protein sequence databases with protein queries; however, Multi-BLAST can use the entire suite of BLAST programs, with the exception of PHI/PSI-BLAST. Although PHI/PSI-BLAST

offers distinct advantages over the traditional suite of BLAST programs; these programs were not included in Multi-BLAST because of their limitation to protein queries and the need for iterative searching of both domains. The basic comparison algorithm could be extended to include these programs, however.

Multi-BLAST has several limitations, some of which will be examined here. First, it only accepts two domains in a query. It may be beneficial to add a third domain to the query, especially when searching large databases with one or more promiscuous domains. However, this would dramatically increase the complexity and speed of the algorithm because domain order would have to be considered. Second, Multi-BLAST cannot execute Boolean (AND, OR, NOT) type searches with the input domains. This feature would allow for more complex queries making new types of sequence searches possible. Third, Multi-BLAST could be improved by the inclusion of more of the standard BLAST parameters and formatting options, most notably, "limit by Entrez query." This command allows results to be limited to subsets of the search database. All of these improvements represent logical extensions of the present version of Multi-BLAST and could be implemented into the program.

The uses of Multi-BLAST extend past simple sequence retrieval. With the tBLASTn option, it is possible to use Multi-BLAST in genefinding, although we did not test this specifically. One can also use Multi-BLAST to search for genetic elements that are arranged in tandem (e.g. promoter elements), giving the Multi-BLAST program additional utility. By far, the most promising and ambitious application of Multi-BLAST lies in genome analysis. Evolutionary changes in genomes are, at least in part, accomplished through sequence change including the creation of new and the

rearrangement of existing functional domains. By using Multi-BLAST to query for every possible combination of domains one could address questions genomic scale. For example, Multi-BLAST could be used to determine which domain combinations are unique to a particular organism or conserved within a group of organisms. By comparing multiple genomes, it may be possible to decipher the core proteome of life. However it is utilized, Multi-BLAST offers an efficient means to identify specific domain architectures within the ever-growing databases of sequence information.

SPECIAL ACKNOWLEDGEMENTS

Although the ideas and work presented in this chapter are largely mine, I must acknowledge that this project was done in collaboration with Albert Ribes-Zamora and Rongwen Xi. Furthermore, all of the coding of the Multi-BLAST algorithm was done by Venu Puram and Weixi Li under my supervision. This project was advised by Dr. Chuck Staben.

Figure 6-1. Multi-BLAST user interface, input screen. Multi-BLAST user interface on which an example of the drop-down domain selection menu has been activated for domain 1. After selecting the domain, its sequence is displayed in the text box and can be modified by the user.

Multi-BLAST

Below is the Blast query which you submitted on 12/26/2001, 13:00

PROGRAM : blastp
DATABASE : drosoph.aa
Max Hits: 1000
Expect: 10
Word Size: 3
Matrix: BLOSUM62
Domain1 Name(S): pfam00054
Domain2 Name(S): pfam00028
Domain1 Sequence:
FRTTEPSGLLLYGGTNDDRDFLALELRGGRLEVHYDLGSGPASVTSGDRLNDGKWHSVELVRNGRKGTLSV

Domain2 Sequence:
YSASVPENAPVGTEVLTVTATDADLGPNGRIFYSILGGGPGGWFRIDPDTGDLSTTKPLDRESIGEYELTVLAT
RANGE : 0 - 500

RESULT :

You can view domain1 blast result here
You can view domain2 blast result here
You can view the final results here

```
RESULT
                                                             Score1    E1   Score2    E2
Sequences found in both results                              (bits)  Value  (bits)  Value
gb|AAF58763.2| (AE003828) stan gene product [Dros... 71  1e-13     86* 2e-18
gb|AAF53635.1| (AE003656) CadN gene product [Dros... 60  3e-10     67* 1e-12
gb|AAF51036.1| (AE003577) ft gene product [Drosop... 55  6e-09     77  7e-16
gb|AAF53636.1| (AE003656) CG7527 gene product [Dr... 55  8e-09     53* 2e-08
gb|AAF49078.1| (AE003515) CG7749 gene product [Dr... 49  4e-07     75* 4e-15

Note: * indicates multiple matches
---------------------------------------------------

>gb|AAF58763.2|AAF58763.2 (AE003828) stan gene product [Drosophila melanogaster]
          Length = 3606

 Score = 70.9 bits (172), Expect = 1e-13
 Identities = 53/158 (33%), Positives = 73/158 (45%), Gaps = 26/158 (16%)

Query: 1    FRTTEPSGLLLYGGT-NDDRDFLALELRGGRLEVHYDLG--SGPASVTSGDRLNDGKWHS 57
            F T + +GLLLY G  N+  DF+ALE+  G +   + LG  S   SV    +++DGKWH
Sbjct: 1584 FATVQENGLLLYNGRYNELHDFIALEIHEGHVSFSFSLGDHSERISVIQEAKVSDGKWHQ 1643

Query: 58   VELVRNGRKGTLSVDGEESVDGTSPSGPD---------------------ITPEFLDLDT 96
            VE+V   R  TL +D  ++    S    D                         FLDL
Sbjct: 1644 VEVVYLNRSVTLVLDNCDTAIALSGQLGDRWSCANRTTLKLDKRCSLLTETCHRFLDLTG 1703

Query: 97   PLYVGGLPELAVKRLLLVISTSFKGCIRDVSINGKPLD 134
            PL VGGLP +       V +  F GCI D+ I+ + +D
Sbjct: 1704 PLQVGGLPRIPAH--FPVTNRDFVGCISDLRIDDRFVD 1739

>gb|AAF58763.2|AAF58763.2 (AE003828) stan gene product [Drosophila melanogaster]
          Length = 3606

 Score = 80.1 bits (196), Expect = 1e-16
 Identities = 44/90 (48%), Positives = 55/90 (60%), Gaps = 1/90 (1%)

Query: 1    YSASVPENAPVGTEVLTVTATDADLGPNGRIFYSILGGG-PGGWFRIDPDTGDLSTTKPL 59
            Y AS+ E+A VGT V+ V A+D D+G NGRI Y +       G F IDP +G + T K L
Sbjct: 1012 YQASILEDALVGTSVIQVAASDPDVGLNGRIKYLLSDRDIEDGSFVIDPTSGTIRTNKGL 1071

Query: 60   DRESIGEYELTVLATDSGGPPLSGTTTVTI 89
            DRES+  + LT +A D G PPLS T  V I
Sbjct: 1072 DRESVAVFHLTAIAVDKGSPPLSSTVEVQI 1101


 Score = 46.2 bits (108), Expect = 2e-06
 Identities = 35/92 (38%), Positives = 44/92 (47%), Gaps = 8/92 (8%)

Query: 5    VPENAPVGTEVLTVTATDADLGPNGRIFYSILGGGPGGWF----RIDPDTGDLSTTKPLD 60
            VPEN+PVG+ V  + A D D G N  + YSI+GG     F    R   +   L T   LD
Sbjct: 1122 VPENSPVGSVVGEIHAHDPDEGVNAVVHYSIIGGDDSNAFSLVTRPGSERAQLLTMTELD 1181

Query: 61   RESIGE-YELTVLATDSGGPPLSGTTTVTITV 91
             ES  + +EL V A     PPL     + I V
Sbjct: 1182 YESTRKRFELVVRA---ASPPLRNDAHIEILV 1210


---------------------------------------------------
```

Figure 6-2. Multi-BLAST user interface, results page. Top is a summary of the user query, including the program, database, query sequences and range value. Beneath the summary is a list of records that match the query along with BLAST scores and Expect values. An asterisk by the BLAST score indicates that multiple domains within the sequence matched the domain query; only the highest BLAST score is reported. Beneath the results summary are the alignments of domains within the first record to the domains used in the query.

Table 6-1. Comparison of Multi-BLAST and BLASTp.

Domain combination	Protein Query Name	Protein Query Accession	Database	BLASTp total hits	BLASTp two domain hits	Multi-BLAST
S_TKc and FHA	yeast Dun1	NP_010182	yeast	129	3	3
LamG1 and CA2	Drosophila CG77496	AAF49078	Drosophila	82	6	6
SH2 and SH3	Drosophila DRK	Q08012	Drosophila	139	12	12
SH2 and RhoGEF	human VAV3	XP_010540	Drosophila	77	1	1
SH2 and SOCS	mouse SOCS-1	AAD53324	nr	229	34	34
SH2 and TyrKc	Src42A Drosoph.	AAF57295	Drosophila	330	9	9
SH2 and TyrKc	FER Canis familiaris	AAF00543	Drosophila	272	9	9
SH2 and TyrKc	FER Canis familiaris	AAF00543	nr	503	261	299

Similar searches using Multi-BLAST and BLASTp were conducted and results containing the domains of interest are reported. S_TKc, Serine threonine kinase catalytic domain; FHA, Fork-head associated domain; LamG1, Laminin G domain; CA2, Cadherin repeats; SH2, Src homology 2; SH3, Src homology 3; RhoGEF, Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases; SOCS, SOCS box; TyrKc, Tyrosine kinase catalytic domain.

Chapter VII

Conclusions and Discussion

In this work, I described the identification and characterization of *Socs44A*, a presumed negative regulator of the JAK/STAT signal transduction pathway. I showed that although *Socs44A* is not responsive to JAK/STAT pathway activity, it is capable of downregulating the pathway in a tissue specific manner. Furthermore, I illustrated a positive role for *Socs44A* in regulating the EGFR/MAPK pathway. This work also chronicled the ongoing efforts to isolate a loss-of-function mutation in the *Socs44A* locus. Finally, I presented a novel BLAST post-processor called Multi-BLAST that efficiently retrieves multiple domain containing sequences from public databases.

The Drosophila genome encodes three SOCS homologues

Drosophila melanogaster presents an attractive model for the study of the JAK/STAT pathway because its pathway is much simpler than its mammalian counterparts. While the mammalian version of the pathway consists of numerous ligands and receptors, four JAKs and seven STATs, the *Drosophila* version possesses but one characterized homologue of each component required for pathway activation. Furthermore, the *Drosophila* genome contains only one PIAS (Protein Inhibitor of Activated STAT) homologue, *zimp*, compared to four mammalian homologues (Betz et al., 2001; Hari et al., 2001; Mohr and Boswell, 1999). PIAS proteins negatively regulate the pathway by interacting with and degrading STAT dimers, preventing their translocation to the nucleus (reviewed in Wormald and Hilton, 2004). In this work I describe the characterization of one of three *Drosophila* SOCS homologues that were

identified on the basis of a shared modular domain architecture with the eight known vertebrate SOCS. The presence of three fly SOCS genes raises an interesting question. If the *Drosophila* JAK/STAT pathway is so much simpler than its mammalian counterpart, why does the fly genome encode three SOCS?

One explanation is that the three *Drosophila* SOCS are redundant in function. We know that this is not entirely true because *Socs36E* and *Socs44A* appear to have differential functions in oogenesis. Furthermore, I show that *Socs44A* has an opposing role to *Socs36E* in EGFR/MAPK signal regulation in the wing. Also, the mRNA expression patterns of *Socs36E* and *Socs44A* do not overlap in either embryogenesis or oogenesis. Finally, it appears that *Socs44A* lacks the capacity to respond to JAK/STAT signaling. Taken together, these observations suggest that *Socs36E* and *Socs44A* are not functionally redundant. The third *Drosophila* SOCS homologue, *Socs16D* has not been functionally characterized. It is possible that it is redundant with one or both of the other *Drosophila* SOCS. Once mutants for all three of these genes are isolated and characterized, potential redundancies among these homologues can be addressed.

Another explanation for multiple SOCS genes in the *Drosophila* genome is that these genes are more generalized regulators of signal transduction, not specific to JAK/STAT. The possession of an SH2 domain would allow SOCS molecules to theoretically interface with any signal transduction pathway propagated by tyrosine phosphorylation. In vertebrates, various SOCS proteins have been shown to interact with Vav, the Rho family guanine nucleotide exchange factor (De Sepulveda et al., 2000); RasGAP, the Ras family GTPase activating protein (Cacalano et al., 2001); FAK, the focal adhesion kinase (Liu et al., 2003); c-kit, a tyrosine kinase receptor (Bayle et al.,

2004); and the Pim family of serine/threonine kinases (Chen et al., 2002b; Peltola et al., 2004). This pleiotropy extends to *Drosophila* as we found that Socs44A genetically interacts with and upregulates the EGFR/MAPK pathway. Our findings indicate that Socs44A operates in an opposite fashion to Socs36E which was found to weakly suppress EGFR/MAPK signaling (Callus and Mathey-Prevot, 2002).

Yet a third possibility is that, like their mammalian counterparts, *Drosophila* SOCS are context specific. In mammals, this can be seen in the tissue specificity in which SOCS genes can be induced. When growth hormone is injected into the murine liver, CIS, SOCS-2 and SOCS-3 expression is induced; however, only CIS and SOCS-2 expression is induced in mammary glands (Davey et al., 1999). Unlike *Socs36E,* the pattern of *Socs44A* expression in the embryo was not altered upon misexpression of *unpaired*. It may be possible that *Socs44A* transcription could be induced by *unpaired* in other tissues.

Role of Socs44A in JAK/STAT signal transduction

The hallmark of mammalian SOCS is their ability to both respond to and downregulate the JAK/STAT cascade, forming a classical negative feedback loop. In this work, I show that Socs44A expression is not responsive to or dependent upon the JAK/STAT pathway. Furthermore, in the absence of a characterized *Socs44A* mutant, the data presented herein are comprehensive lines of evidence that clearly suggest that *Socs44A* has a role in downregulating JAK/STAT activity in the wing. One explanation for the aberrant negative feedback loop participation is that the ability to respond to pathway activation arose after the evolution of *Socs44A*. All of the mammalian SOCS

have been shown to be responsive pathway activation, with the exception of SOCS-4, SOCS-6, and SOCS-7. Interestingly, Socs44A occupies the same phylogenetic clade as SOCS-6 and SOCS-7 (Fig. 2-2), consistent with this hypothesis. The third *Drosophila* SOCS homologue, Socs16D, also belongs to this clade. It would be interesting to see if it is responsive to JAK/STAT pathway activation. Another explanation for these observations is that at some point *Socs44A* lost the capacity to respond to pathway activation.

In addition to *Drosophila melanogaster*, JAK/STAT components can be found in several invertebrate model organisms; however, these models do not contain a complete pathway (Rawlings et al., 2004). The slime mold *D. discoideum* has three STATs but no other components and the nematode *C. elegans* has both a STAT and a SOCS homologue, but no JAKs, receptors, or ligands. The mere presence of a SOCS in *C. elegans* suggests that the "original" function of SOCS proteins may not have been to downregulate JAK/STAT signaling because the *C. elegans* genome does not contain any of the known targets for SOCS-mediated downregulation of the JAK/STAT pathway. If not JAK/STAT signaling, perhaps the "original" function of SOCS is to regulate another cascade and interactions with JAK/STAT components arose with the divergence of mammals. Phylogenetic analysis of known SOCS proteins is consistent with this hypothesis, as all four of the mammalian SOCS genes that have been shown to both respond to and downregulate the JAK/STAT pathway fall into a single clade not populated by SOCS from either *Drosophila* or *C. elegans* (Fig. 2-2).

Role of Socs44A in EGFR/MAPK signal transduction

The JAK/STAT and EGFR/MAPK pathways are integrated at multiple levels resulting in combinatorial control of signal transduction. Theoretically, this phenomenon allows for increased complexity in the interpretation of extracellular signals. It has been shown that each pathway can activate the other independent of the endogenous ligand. EGFR can mediate the activation of STAT independent of JAK activation (David et al., 1996; Leaman et al., 1996). Additionally, there are several phosphorylation sites on STAT not used by the canonical JAK/STAT mechanism. Studies have shown that STATs can be phosphorylated at a serine residue at the C-terminus by several MAPKs, potentiating STAT activity (Decker and Kovarik, 2000). Likewise, IL-6 and GH can activate MAPK via JAK-2, presumably by activating the Shc and GRB2 adaptor proteins (Giordano et al., 1997; Yamauchi et al., 1997).

In this work, I discovered and investigated the interactions between Socs44A and the EGFR/MAPK pathway, illustrating that *Socs44A* can act to upregulate the EGFR/MAPK pathway. A recent study showed that SOCS-3 behaves in a similar manner, achieving this task through targeted degradation of RasGAP (Cacalano et al., 2001). This interaction requires the phosphorylation of SOCS-3 on a C-terminal tyrosine within the SOCS domain. Socs44A does not contain a tyrosine at this position, instead contains a glutamic acid which could mimic a constitutively phosphorylated state. Biochemical analyses of Socs44A will be required to address this possibility. Interestingly, it has been shown that Socs36E also interacts with the EGFR/MAPK pathway; however, it appears to have an opposite role. Misexpression of Socs36E was

able to weakly suppress wing vein phenotypes associated with several EGFR/MAPK pathway mutants (Callus and Mathey-Prevot, 2002).

Crosstalk between the JAK/STAT and EGFR/MAPK pathways

As mentioned above, the complexity required for the countless number of signals that must be generated during development and homeostasis can be achieved, in part, through crosstalk among signal transduction cascades. Combinatorial control of gene expression may serve to attenuate levels of transcription or perhaps interplay provides a means of redundancy, contributing to the robustness of these signaling pathways. There is evidence that the JAK/STAT cascade can interface with several other signaling mechanisms. This is not surprising as phosphorylation-mediated activation is a common theme amongst many pathways (for review, see Graves and Krebs, 1999). Of particular interest is the EGFR/MAPK (Epidermal Growth Factor/Mitogen Activated Protein Kinase) pathway. Like JAK/STAT, this cascade relies on phosphorylation for signal propagation. Therefore, it is not entirely surprising that EGFR can activate STAT, independent of JAK activation (David et al., 1996; Leaman et al., 1996). Furthermore, STATs may be activated by serine phosphorylation that is also mediated by several MAPKs (Decker and Kovarik, 2000). Conversely, IL-6 and Growth hormone can activate the EGFR/MAPK pathway in addition to JAK/STAT (Giordano et al., 1997; Yamauchi et al., 1998; Yamauchi et al., 1997). Also, it has been shown that JAK can activate MAPK through the Pyk2 intermediary (Takaoka et al., 1999). The interplay between EGFR/MAPK and JAK/STAT may not be limited to vertebrates. Recent work in our lab demonstrated the combinatorial requirement of these cascades in patterning the

follicular epithelium during oogenesis (McGregor et al., 2002; Xi et al., 2003), however, the precise mechanism of this control has yet to be elucidated. Perhaps this is accomplished through direct interaction between these pathways. If so, then both Socs44A and Socs36E are attractive as mediators of this interaction. Although *Socs44A* is not expressed in the ovarian follicle cells, its transcripts were seen accumulating in the oocyte, where Gurken initiates the EGFR/MAPK pathway.

The role of Socs44A in wing vein development

Developmental biology, when reduced to its simplest form, is the patterning of a field of cells. The *Drosophila* wing provides an excellent model to investigate this process. It consists of specific wing vein pattern dispersed in a field of epithelial cells that is two cell layers thick. The wing veins provide structural rigidity and also serve as conduits for sensory axons, hemocites, and tracheal cells (reviewed in De Celis, 2003). The wing veins consist of six longitudinal veins, two of which (L1 and L6) do not reach the distal margin. Additionally, there are two transverse veins that that intersect longitudials. The anterior crossvein (ACV) connects L3 and L4 and the posterior crossvein connects L4 and L5.

The development of the longitudinal veins takes place in progressive steps that establish vein location and identity. The Hedgehog, Notch, Wingless, EGFR/MAPK and Dpp pathways all work synergistically in this process that begins in the imaginal wing disc with the establishment of provein and intervein cell-fate decisions along with establishing the anterior-posterior boundary (Crozatier et al., 2002; Williams et al., 1993; Zecca et al., 1995). It is known that Hedgehog is responsible for patterning L3, L4, and

the ACV (Methot and Basler, 1999; Mullor et al., 1997). EGFR/MAPK signaling is responsible for patterning L4 (Garcia-Bellido et al., 1994; Schnepp et al., 1996). Finally, Dpp regulates *spalt* gene complex expression which is responsible for positioning of L2 and L5 (de Celis and Barrio, 2000). Interestingly, it is not known what directs the patterning and positioning of the posterior crossvein.

I have provided evidence suggesting that endogenous Socs44A acts to simultaneously downregulate JAK/STAT and upregulate EGFR/MAPK pathway activity in the wing. It is known that the EGFR/MAPK pathway is involved in the specification of vein and intervein cell fates. Heteroallelic *Egfr* mutants lack the central portion of L4 (Fig. 4-2). Other EGFR/MAPK pathway mutants share consistent phenotypes (Guichard et al., 1999; Sawamoto et al., 1994). While much is known about the roles of EGFR/MAPK in specifying vein and intervein fates, the roles of JAK/STAT signaling in wing vein development have not been determined. Nonetheless, mutations in either *Stat92E* or *hop* exhibit a relatively subtle wing vein defect, marked by a small amount of ectopic vein material protruding longitudinally from the posterior crossvein (Yan et al., 1996a and Fig. 3-5).

A simple model for the role of Socs44A in wing vein development can be generated based upon findings in this study (Fig. 7-1). In this model, Socs44A acts to upregulate EGFR/MAPK activity, as evidenced by ectopic wing vein phenotypes caused by *Socs44A* misexpression that can be modulated by EGFR/MAPK pathway mutants. One can speculate that targeting RasGAP for proteosomal degradation is the *modus operandi* of Socs44A upregulation of EGFR/MAPK signaling, leading to ectopic specification of vein cell fate in an otherwise intervein territory. Biochemical analyses

will allow scrutiny of this hypothesis. Once the roles of JAK signaling in wing vein development have been elucidated, Socs44A regulation of that pathway in this context can be addressed.

Other roles for Socs44A in Drosophila development

The pleiotropy of the relatively few signaling pathways accounts for their ability to govern the multitude of developmental and homeostatic processes. As a likely regulator of both JAK/STAT and EGFR/MAPK signaling, it would not be surprising to find that Socs44A is involved in many contexts. Tracheal morphogenesis and hematopoiesis stand out as two developmental processes that are attractive models for the continued study of the roles of Socs44A in *Drosophila* development.

Tracheal development (for reviews, see Manning and Krasnow, 1993; Metzger and Krasnow, 1999; Shilo et al., 1997) begins with the assignment of tracheal cell fates to ten segmental clusters of cells on each side of the embryo approximately four hours after fertilization. These cells will then give rise to the tracheal pits in which *unpaired* expression can ultimately be detected (Harrison et al., 1998 and Fig. 3-1). The remainder of tracheal morphogenesis occurs solely by cell migration leading to the formation of a large tree-like structure. Over 50 genes have been found to play roles in tracheal development (Metzger and Krasnow, 1999). Mutants of these genes exhibit varying phenotypes ranging from migration defects to faulty cell fate decisions (for review, see Zelzer and Shilo, 2000).

The JAK/STAT pathway is involved in tracheal development. In embryos that lack *hopscotch*, the trachea is reduced, and what tissue remains is grossly malformed,

failing to migrate properly (Fig. 3-2). Furthermore, the elusive JAK/STAT pathway receptor, *domeless*, was first isolated based upon its posterior spiracle defect (Brown et al., 2001). Subtle tracheal defects were also observed in the $Stat^{HiJak}$ mutant (Yan et al., 1996a). I detected expression of *Socs44A* in the trachea of late stage embryos; however, its expression does not overlap that of *unpaired*, which is expressed at an earlier stage. Misexpression of *Socs44A* using the *breathless*-GAL4 driver resulted in lethality (Table 3-1). *Breathless* is an FGF homologue required for the proper migration of tracheal cells (Yan et al., 1996a). In addition to FGF the TGF-β and EGFR/MAPK pathways are involved in the migration of tracheal cells, helping to determine the size and spacing of primary branches {reviewed in \Metzger, 1999 #280}. The coordination of all of the signaling cascades involved in patterning the trachea is poorly understood. Socs44A could have a role in the coordination of JAK/STAT and EGFR/MAPK signaling during the patterning of the trachea. In light of its expression pattern in embryogenesis and the implication of JAK/STAT and EGFR/MAPK signaling in tracheal development, the study of Socs44A in tracheal development should merit further consideration.

Perhaps hematopoiesis is the most studied developmental process involving JAK/STAT signal transduction. Abnormal levels of pathway activity have been directly implicated as causative agents in hematopoietic disorders including several leukemias and immune disorders (Aringer et al., 1999; Bowman et al., 2000; Coffer et al., 2000; Lacronique et al., 1997; Leonard, 1996; Ward et al., 2000). Recently, the JAK/STAT cascade has also been implicated in the progression of rheumatoid arthritis (Ivashkiv and Hu, 2003; Naka and Kishimoto, 2002; Naka et al., 2002). As negative regulators of cytokine signaling, SOCS are inviting targets for gene based therapies. Indeed, SOCS

involvement in the immune response has been well documented (see review by Kubo et al., 2003). It has been proposed that SOCS-3 could be used in such a fashion to combat inflammatory arthritis (Rottapel, 2001; Shouda et al., 2001).

The *Drosophila* hematopoietic system is primitive compared to its mammalian counterpart; however, several mechanisms in the hematopoietic process are conserved (Evans et al., 2003; Lanot et al., 2001; Lebestky et al., 2000). Perturbations in JAK/STAT signaling have analogous effects to what is seen in mammals. The gain-of-function alleles, $hop^{Tum\text{-}l}$ and hop^{T42} are well-characterized. Animals carrying either of these dominant mutations exhibit reduced viability, enlargement of the lymph gland and develop melanotic masses (Hanratty and Dearolf, 1993; Harrison et al., 1995; Luo et al., 1995; Luo et al., 1997). These observations make hematopoiesis an attractive target for study. Preliminary data involving ectopic misexpression, RNA interference (RNAi), and the use of deficiencies indicate that Socs36E may suppress the $hop^{Tum\text{-}l}$ phenotypes (G. Rennebeck and J. Rawlings, unpublished). Clearly, JAK/STAT signaling is paramount for proper development of hematopoietic lineages in both mammals and flies and perhaps Socs36E and/or Socs44A may be regulating this process.

Closing remarks

I describe in this work the identification and initial characterization of a SOCS homologue in *Drosophila*. Based on its domain architecture, we believed that Socs44A would participate in a negative feedback loop to regulate JAK/STAT signaling. Therefore, we initiated a reverse genetic study that revealed important information about this gene and the process of its characterization. First, I provide the first example of a

SOCS gene, vertebrate or invertebrate, that is not regulated by JAK/STAT activity (at least in the tissues we tested). Although it is not known if SOCS-4, SOCS-6, and SOCS-7 are induced by cytokines, no one has shown that these molecules are not transcriptionally regulated by the pathway. This important find illustrates the relevance of negative data and the simplicity of the *Drosophila* JAK/STAT model, which allows this claim to be made. Second, this work is the first to demonstrate differential tissue specificity of SOCS activity in *Drosophila*, a phenomenon seen in the mammalian model. This bolsters the validity of the insect as a model to study JAK/STAT regulation via SOCS. Because the rest of the JAK/STAT pathway is much simpler in *Drosophila*, future study in the dipteran could elucidate the mechanisms underlying this complexity. Third, I discovered that Socs44A upregulates EGFR/MAPK signaling in wing development in contrast to Socs36E which weakly suppressed it (Callus and Mathey-Prevot, 2002). This is the first evidence of SOCS having direct opposing roles in any developmental process. Fourth, our futile attempts at obtaining a Socs44A loss-of-function allele uncovered weaknesses in the reverse genetic approach to the investigation of gene function, exposing the principle deficiency in *Drosophila melanogaster* as a genetic system, specifically, the lack of efficient, rapid, targeted gene disruption. Although a method for gene targeting has been recently developed (Rong and Golic, 2001; Rong et al., 2002), its efficiency and efficacy is largely untested. Fortunately, efforts from the *Drosophila* community, namely Exelixis (Thibault et al., 2004) and the BDGP (Bellen et al., in press) are addressing this flaw by attempting to disrupt every gene in the genome through transposon mobilization.

It is clear that SOCS play pivotal roles in the regulation of at least two signal transduction cascades. I have shown that this regulation extends to the lower eukaryote, *Drosophila melanogaster*. For years, *Drosophila* has been a pioneering model for genetic investigation of developmental processes and my work regarding *Socs44A* illustrates the utility of the fly as a model system. Clearly, this work has presented as many, if not more, questions than it has answered. Is *Socs44A* responsive to JAK/STAT signaling in other tissues? What are the precise roles of *Socs44A* and JAK/STAT signaling in wing development? Is *Socs44A* involved in tracheal morphogenesis and hematopoiesis? How? Most importantly, is our understanding of *Socs44A* and its roles in signal transduction applicable to the human condition? Time will tell.

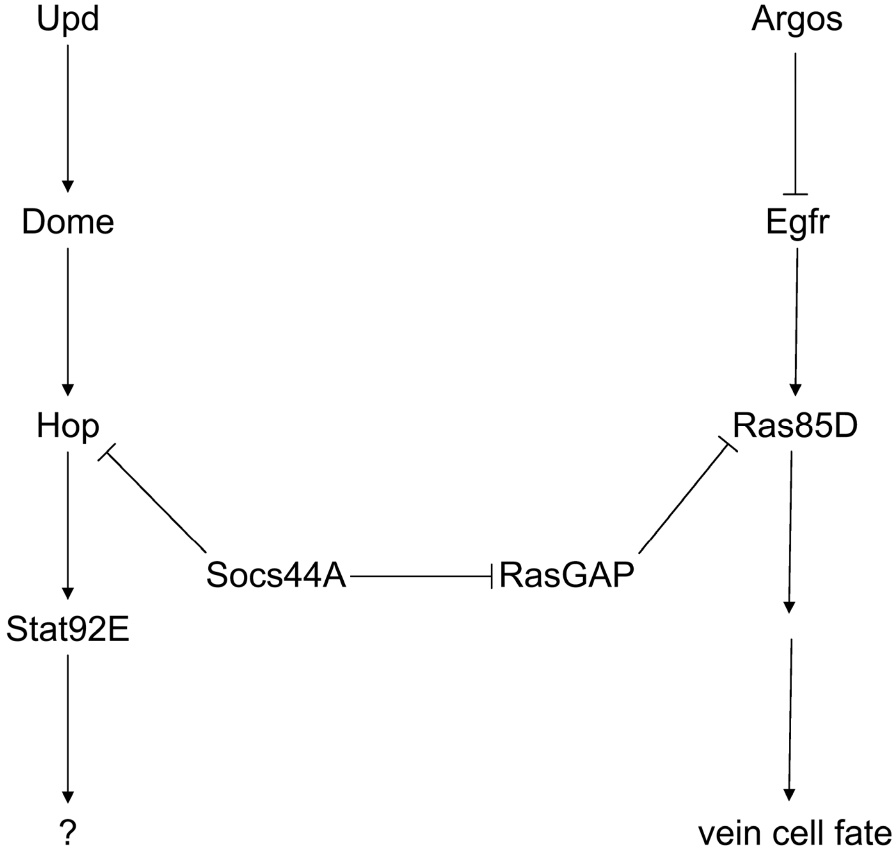


Figure 7-1. Model for Socs44A involvement in wing vein formation. Based on findings presented in this work, we can postulate that the ectopic wing venation caused by *Socs44A* misexpression is the result of ectopic upregulation of EGFR/MAPK signaling. This may be caused by Socs44A-mediated proteosomal degradation of RasGAP. Based on distinct wing phenotypes exhibited in *hop* and *stat92E* mutants, the JAK/STAT pathway also participates in patterning the wing veins, although its specific role has yet to be determined.

Chapter VIII

Materials and Methods

Fly strains

Fly strains were raised under standard conditions at 25^{0}C unless otherwise stated. *Sty* alleles and *Egfr*^{top1} were obtained from M. Freeman. *Btl*-GAL4 was obtained from M. Krasnow. *Pnt*-LacZ was obtained from T. Schupbach. *[Act5C>y>GAL4][UAS-GFP.S65T]* was obtained from M. Zeidler. A Bgl I/Xho I fragment from LP02169 (described below) containing the entire *Socs44A* EST sequence was subcloned into the pUAST vector (Brand and Perrimon, 1993) in order to ectopically express *Socs44A* in the GAL4/UAS system. The resulting construct was used in standard germline transformation (Spradling, 1986). Three independent insertions recovered and balanced. All three behaved similarly when misexpressed with GAL drivers (Table 3-1). The remaining GAL4 drivers (Table 3-1); mutants and recombinants used in *Socs44A* interaction studies (Table 4-1); the *Df(2R)CA53, Df(2R)Drl*^{rv18}*,* and *Df(2R)NCX10* deficiencies; and the *KG03963, EP(2)2264,* and *l(2)02045,* and *l(2)k16503* P-elements are all described in FlyBase.

Cloning of Socs44A

An 1133bp fragment of genomic DNA corresponding to the *Socs44A* locus was PCR amplified using the following primers: '5 – GAG CCA CGG CGA CCA GAG TCA AAA A – 3' and 5' – CAA GTA CTC CAG CAT CTG CGC C – 3'. The resulting product, pBS-Socs44Agenomic, was cloned into pBlueScript II KS+. *Socs44A* was also

subcloned from the LP02169 EST clone (described below) as a Bam HI/Xho I fragment and placed into pBluescript II KS+, making pBS-Socs44A. These constructs were used in the *in situ* hybridizations to embryos, DNA sequencing and RNAi construction described below.

Sequencing of the Socs44A EST

The Sanger/dideoxy method (Sanger et al., 1977) of nucleotide determination was used to sequence the entire LP02169 EST sequence corresponding to *Socs44A*. The following primers (obtained from IDT) were used (see Fig. 2-4):

#1: '5 – GAG CCA CGG CGA CCA GAG TCA AAA A – 3'

#2: '5 – CCA TGG GCG ACT GCG ACG ACG G – 3'

#3: '5 – CCG TCG TCG CAG TCG CCC ATG G – 3'

#4: 5' – CAA GTA CTC CAG CAT CTG CGC C – 3'

#5: 5' – CTT CTC CCT GTC CCT C – 3'

#6: 5' – GCC GCC GTG TTC TGC – 3'

#7: 5' – CCA TGC AAT CTA GAT TAG C – 3'

Additionally, standard primers corresponding to the T3 and T7 promoter sequences of pBlueScript II were used. The sequence of this *Socs44A* EST has been deposited in Genbank (Accession AF439523).

In situ hybridization

In situ hybridization of embryos and ovaries was performed essentially as described (Harrison et al., 1998; Wilkie et al., 1999). Sense and anti-sense digoxigenin

probes from the previously described *Socs44A* cDNAs were generated. Probe for *Socs36E* was generated from the 5' end of the 2.1.1 cDNA. Embryos aged 0-24 hours were collected, dechorionated in 50% bleach, fixed in 3.7% formaldehyde and dehydrated in methanol and stored until ready for use. Embryos were rehydrated in PBT, fixed again in 3.7% formaldehyde, washed 5 x 5' in PBT. Embryos were then moved into hybridization solution (50% deionized formamide, 5X SSC, 50ug/ml heparin, 100ug/ml tRNA, and 0.1% Tween) and allowed to prehybridize for one hour at 70^0C. Appropriate probes diluted between 1:200 and 1:50 and then hybridized overnight at 70^0C. The embryos were then washed for 20 minutes in prewarmed hybridization solution at 70^0C followed by a 20 minute wash in prewarmed hybridization/PBT (1:1) solution. Next, embryos were washed 4 x 20' in PBT at 70^0C followed by 3 x 20' washes in PBT at room temperature. Embryos were then washed 3 x 5' in pH 9 solution and then developed in NBT and X-phosphate. The reaction was stopped by 6 x 5' washes in PBT + EDTA. Embryos were mounted with 70% glycerol in PBT. All washes were done on a standard microfuge tube rotator and incubations were done on a standard heating block.

Germline clone hop^{c111} null allele animals were generated using the dominant female sterile technique (Chou and Perrimon, 1992). Embryos misexpressing *upd* in the seven stripe pair-rule pattern were generated by crossing females carrying a UAS-*upd* transgene with males heterozygous for *paired*-GAL4. Embryos were then collected and hybridized as described.

Immunological staining of embryos

Germline clone *hop*^{c111} mutants in a *trh*^{10512} enhancer trap background (Isaac and Andrew, 1996) were generated. These animals (genotype *hop*^{c111}*/hop*^{c111}*; trachealess-LacZ/+*) were stained with anti-β-gal antibody to visualize developing trachea (Patel, 1994). Briefly, embryos were dechorionated in 50% bleach and fixed in 3.7% formaldehyde in heptane, washed in PBT, and blocked for 1 hour in 5% BSA in PBT. Embryos were then incubated in a 1:1000 dilution of rabbit anti-β-gal overnight at 4^{0}C. Following washes in PBT, embryos were then incubated in a 1:500 dilution of FITC anti-rabbit secondary antibody for 2 hours at room temperature. Embryos were mounted in 70% Glycerol, 2.5% DABCO in PBT.

Generation of ovarian clones

Ovarian clones of the *hop*^{c111} null allele were generated by mitotic recombination mediated by hsFLP as previously described (McGregor et al., 2002). Misexpression clones of *Socs36E* and *Socs44A* were generated using a GAL4 flip-out cassette, also controlled by hsFLP (Ito et al., 1997). Genotype of those animals was *w [hsFLP]1; [Act5C>y>GAL4][UAS-GFP.S65T]/[UAS-Socs36E]11.2; pnt-LacZ* and *w [hsFLP]1; [Act5C>y>GAL4][UAS-GFP.S65T]/+; [UAS-Socs44A]11D/ pnt-LacZ,* respectively. For each, ovaries were fixed and stained with anti-β-gal and anti-GFP as previously described (McGregor et al., 2002).

Misexpression studies

Flies carrying the *[UAS-Socs36E]* or *[UAS-Socs44A]* transgenes were mated to various GAL4 drivers and compared to wild-type flies for any abnormal phenotypes. For e16E, T113, T6, and *ptc*-GAL, wings were dissected and mounted in Hoyer's medium (Ashburner, 1989). Two copies of the *[UAS-Socs36E]* transgene were required to generate the indicated wing phenotype when crossed to GAL-e16E (*engrailed*). For interaction assays, flies of genotype *y w; GAL-e16E/CyO; [UAS-Socs44A]11D/TM3* were crossed to the mutant or recombinant listed in Table 4-1. In the *argos* interaction assay of EMS mutants, flies of genotype *[UAS-argos]/Y; GAL-e16E/[UAS-argos]* were crossed to each candidate *Socs44A* mutant in addition to *y w; CA53/CyO, y w; NCX10/CyO,* and *y w;* Drl^{rv18}*/CyO.* Female progeny possessing the X chromosome UAS-*argos* insertion, the *GAL-e16E* driver, and the mutant/recombinant variable were examined. Each wing on each animal was classified according to Fig. 4-3). Fisher's Exact Test was performed (courtesy of C. Saunders) on the data to determine statistical significance.

Image capture and processing

All *in situ* hybridization and wing images (Nomarski, DIC, and epifluorescence) were acquired using a Spot Camera (Diagnostic Instruments) on a Nikon E800 microscope. A Leica TCS-SP laser scanning confocal microscope was used to capture all confocal micrographs. All images were then exported to Adobe Photoshop for manipulation and annotation.

RNAi construction

pBS-Socs44A (described above) was cut with Xho I and Stu I, filled in with Klenow and relegated to form pBS-Socs44A-5'-Stu. A 600bp Bam HI/Eco RI fragment of pYES was subcloned into pBS-Socs44A-5'-Stu to form pBS-44A-spacer. A Sma I/Xba I fragment of the EST clone, LP02169, was subcloned into pBS-44A-spacer via an Xba/ blunted Bam HI ligation to create pBS-44A-hp6. An Xba I/Stu I fragment was removed from this construct, the resulting DNA was blunted and relegated to form pBS-44A-hp6.1. The entire symmetrical *Socs44A* inverted repeat was then isolated as a Kpn I/Not I fragment and subcloned into pUASp (Rorth, 1998) to create the final construct pUASp-Socs44A-RNAi.

EMS mutagenesis

Six hundred isogenized males of genotype *y w; cn bw sp* were fed 25mM EMS in a 1% sucrose solution per the standard protocol (Lewis and Bacher, 1968). Mutagenized males were then mated in groups of 8 to approximately 25 females of genotype *y w; Sco/CyO*. 7,477 individual male progeny of genotype *y w; cn bw sp*/CyO* or *y w; cn bw sp*/Sco* were then mated to three females of genotype *y w; Sco/CyO* (asterisk indicates mutagenized chromosome). From each cross, siblings of genotype *y w; Sco/CyO* were mated to establish a balanced stock. We recovered 3,986 balanced stocks. Each stock was then mated to the *CA53* deficiency and tested for complementation. Those stocks that failed to complement (were lethal or exhibited a phenotype) were recrossed to *CA53* for verification. Verified stocks were crossed to *NCX10* and *Drl*^{rv18} deficiencies; stocks that complemented *Drl*^{rv18} but failed to complement *NCX10* were retained as candidate

Socs44A mutants. Each of these stocks, a total of 43, were crossed pair-wise in complementation analyses to determine number of genes represented in the collection. Representatives of several of these groups were selected for sequencing. These lines were balanced over the P{Act-GFP}CyO chromosome, permitting the collection of DNA from homozygous animals based on their failure to fluoresce. *Socs44A* was PCR amplified, sequenced, and scanned for molecular lesions by comparing this sequence to the original, unmutagenized *cn bw sp* chromosome. Representatives of all of the groups were subjected to the *argos* interaction assay described above.

P-element excision mutagenesis

The *KG03963* viable P-element was excised in the germline of males of genotype *[KG03963]/CyO; Δ2-3,Sb/+*. These males were crossed individually to *y w; Sco/CyO* females and the progeny examined for excision based on loss of red eye color associated with the P-element. Progeny of genotype *[KG03963](w^-)/CyO* were mated to *y w; Sco/P{Act-GFP}CyO* to establish 109 balanced stocks, each representing an independent excision event. For each line, DNA was isolated from homozygous animals. In the case of 7 lethal stocks, this was accomplished by selecting embryos or larvae that failed to fluoresce. This process was verified by failure to amplify GFP from the DNA collected. For all 109 stocks, a 700bp fragment of the *upd* locus (positive control) and a 400bp fragment of the proximal portion of *Socs44A* were simultaneously amplified by PCR. In parallel, a reaction containing all PCR reagents except template DNA was used as a negative control. Finally, DNA from *y w; Sco/P{Act-GFP}CyO* flies was used as a *positive* control for *upd*, *Socs44A,* and *GFP* amplification.

Socs44A primers:

#1: ‘5 – GAG CCA CGG CGA CCA GAG TCA AAA A – 3’

#3: ‘5 – CCG TCG TCG CAG TCG CCC ATG G – 3’

Socs44A upstream primers:

S44-5’-1: 5’ – CTA GAG ATG GAA GTA TGT TCG – 3’

S44-5’-2: 5’ – GTG ATC AAT GCC TAA CTT TA – 3’

Unpaired primers:

Upd-54: 5’ CTG CAC ACT GAT TTC GAT ACG GAC CGC GG – 3’

Upd-798R: 5’ – GAT CCC AGC GGA TCT GCT GGC GCC – 3’

Bioinformatic analyses

All sequence alignments were done using AlignX (Informax Inc.) using ClustalX parameters. The Phylogenetic trees were produced from the alignments using the Neighbor-Joining method. All BLAST analyses were done at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) using the default parameters. All BLASTp searches in the Multi-BLAST comparison also used default options, with the exception that the number of results to display was set at maximum (1000) for each search.

References

Adams, M. D. Celniker, S. E. Holt, R. A. Evans, C. A. Gocayne, J. D. Amanatides, P. G. Scherer, S. E. Li, P. W. Hoskins, R. A. Galle, R. F. et al. (2000). The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-95.

Alexander, W. S. (2002). Suppressors of cytokine signalling (SOCS) in the immune system. *Nat Rev Immunol* **2**, 410-6.

Alexander, W. S., Starr, R., Fenner, J. E., Scott, C. L., Handman, E., Sprigg, N. S., Corbin, J. E., Cornish, A. L., Darwiche, R., Owczarek, C. M. et al. (1999). SOCS1 is a critical inhibitor of interferon gamma signaling and prevents the potentially fatal neonatal actions of this cytokine. *Cell* **98**, 597-608.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.

Aringer, M., Cheng, A., Nelson, J. W., Chen, M., Sudarshan, C., Zhou, Y. J. and O'Shea, J. J. (1999). Janus kinases and their role in growth and disease. *Life Sci* **64**, 2173-86.

Armitage, P. and Berry, G. (1994). Statistical methods in medical research. Oxford ; Boston: Blackwell Scientific Publications.

Ashburner, M. (1989). Drosophila: A Laboratory Handbook and Manual. Two volumes. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Bach, E. A., Vincent, S., Zeidler, M. P. and Perrimon, N. (2003). A sensitized genetic screen to identify novel regulators and components of the Drosophila janus kinase/signal transducer and activator of transcription pathway. *Genetics* **165**, 1149-66.

Barillas-Mury, C., Han, Y. S., Seeley, D. and Kafatos, F. C. (1999). Anopheles gambiae Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. *Embo J* **18**, 959-67.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. and Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res* **28**, 263-6.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. et al. (2004). The Pfam protein families database. *Nucleic Acids Res* **32 Database issue**, D138-41.

Bayle, J., Letard, S., Frank, R., Dubreuil, P. and De Sepulveda, P. (2004). Suppressor of cytokine signaling 6 associates with KIT and regulates KIT receptor signaling. *J Biol Chem* **279**, 12249-59.

Beall, E. L. and Rio, D. C. (1997). Drosophila P-element transposase is a novel site-specific endonuclease. *Genes Dev* **11**, 2137-51.

Beall, E. L. and Rio, D. C. (1998). Transposase makes critical contacts with, and is stimulated by, single-stranded DNA at the P element termini in vitro. *Embo J* **17**, 2122-36.

Beccari, S., Teixeira, L. and Rorth, P. (2002). The JAK/STAT pathway is required for border cell migration during Drosophila oogenesis. *Mech Dev* **111**, 115-23.

Bellen, H. J., Levis, R. W., Liao, G., He, Y., Carlson, J. W., Tsang, G., Evans-Holm, M., Hiesinger, P. R., Shultze, K. L., Rubin, G. M. et al. (in press). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics*.

Betz, A., Lampen, N., Martinek, S., Young, M. W. and Darnell, J. E., Jr. (2001). A Drosophila PIAS homologue negatively regulates stat92E. *Proc Natl Acad Sci U S A* **98**, 9563-8.

Binari, R. and Perrimon, N. (1994). Stripe-specific regulation of pair-rule genes by hopscotch, a putative Jak family tyrosine kinase in Drosophila. *Genes Dev* **8**, 300-12.

Bowman, T., Garcia, R., Turkson, J. and Jove, R. (2000). STATs in oncogenesis. *Oncogene* **19**, 2474-88.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-15.

Brawley, C. and Matunis, E. (2004). Regeneration of male germline stem cells by spermatogonial dedifferentiation in vivo. *Science* **304**, 1331-4.

Brender, C., Columbus, R., Metcalf, D., Handman, E., Starr, R., Huntington, N., Tarlinton, D., Odum, N., Nicholson, S. E., Nicola, N. A. et al. (2004). SOCS5 Is Expressed in Primary B and T Lymphoid Cells but Is Dispensable for Lymphocyte Production and Function. *Mol. Cell. Biol.* **24**, 6094-6103.

Brown, S., Hu, N. and Hombria, J. C. (2001). Identification of the first invertebrate interleukin JAK/STAT receptor, the Drosophila gene domeless. *Curr Biol* **11**, 1700-5.

Cacalano, N. A., Sanden, D. and Johnston, J. A. (2001). Tyrosine-phosphorylated SOCS-3 inhibits STAT activation but binds to p120 RasGAP and activates Ras. *Nat Cell Biol* **3**, 460-5.

Callus, B. A. and Mathey-Prevot, B. (2002). SOCS36E, a novel Drosophila SOCS protein, suppresses JAK/STAT and EGF-R signalling in the imaginal wing disc. *Oncogene* **21**, 4812-21.

Capdevila, J., Estrada, M. P., Sanchez-Herrero, E. and Guerrero, I. (1994). The Drosophila segment polarity gene patched interacts with decapentaplegic in wing development. *Embo J* **13**, 71-82.

Casci, T., Vinos, J. and Freeman, M. (1999). Sprouty, an intracellular inhibitor of Ras signaling. *Cell* **96**, 655-65.

Chen, H. W., Chen, X., Oh, S. W., Marinissen, M. J., Gutkind, J. S. and Hou, S. X. (2002a). mom identifies a receptor for the Drosophila JAK/STAT signal transduction pathway and encodes a protein distantly related to the mammalian cytokine receptor family. *Genes Dev* **16**, 388-98.

Chen, X. P., Losman, J. A., Cowan, S., Donahue, E., Fay, S., Vuong, B. Q., Nawijn, M. C., Capece, D., Cohan, V. L. and Rothman, P. (2002b). Pim serine/threonine kinases regulate the stability of Socs-1 protein. *Proc Natl Acad Sci U S A* **99**, 2175-80.

Chou, T. B. and Perrimon, N. (1992). Use of a yeast site-specific recombinase to produce female germline chimeras in Drosophila. *Genetics* **131**, 643-53.

Chung, C. D., Liao, J., Liu, B., Rao, X., Jay, P., Berta, P. and Shuai, K. (1997). Specific inhibition of Stat3 signal transduction by PIAS3. *Science* **278**, 1803-5.

Coffer, P. J., Koenderman, L. and de Groot, R. P. (2000). The role of STATs in myeloid differentiation and leukemia. *Oncogene* **19**, 2511-22.

Cohney, S. J., Sanden, D., Cacalano, N. A., Yoshimura, A., Mui, A., Migone, T. S. and Johnston, J. A. (1999). SOCS-3 is tyrosine phosphorylated in response to interleukin-2 and suppresses STAT5 phosphorylation and lymphocyte proliferation. *Mol Cell Biol* **19**, 4980-8.

Collum, R. G., Brutsaert, S., Lee, G. and Schindler, C. (2000). A Stat3-interacting protein (StIP1) regulates cytokine signal transduction. *Proc Natl Acad Sci U S A* **97**, 10120-5.

Cooley, L., Kelley, R. and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P elements. *Science* **239**, 1121-8.

Crozatier, M., Glise, B. and Vincent, A. (2002). Connecting Hh, Dpp and EGF signalling in patterning of the Drosophila wing; the pivotal role of collier/knot in the AP organiser. *Development* **129**, 4261-9.

Davey, H. W., McLachlan, M. J., Wilkins, R. J., Hilton, D. J. and Adams, T. E. (1999). STAT5b mediates the GH-induced expression of SOCS-2 and SOCS-3 mRNA in the liver. *Mol Cell Endocrinol* **158**, 111-6.

David, M., Wong, L., Flavell, R., Thompson, S. A., Wells, A., Larner, A. C. and Johnson, G. R. (1996). STAT activation by epidermal growth factor (EGF) and amphiregulin. Requirement for the EGF receptor kinase but not for tyrosine phosphorylation sites or JAK1. *J Biol Chem* **271**, 9185-8.

De Celis, J. F. (2003). Pattern formation in the Drosophila wing: The development of the veins. *Bioessays* **25**, 443-51.

de Celis, J. F. and Barrio, R. (2000). Function of the spalt/spalt-related gene complex in positioning the veins in the Drosophila wing. *Mech Dev* **91**, 31-41.

De Sepulveda, P., Ilangumaran, S. and Rottapel, R. (2000). Suppressor of cytokine signaling-1 inhibits VAV function through protein degradation. *J Biol Chem* **275**, 14005-8.

Dearolf, C. R. (1999). JAKs and STATs in invertebrate model organisms. *Cell Mol Life Sci* **55**, 1578-84.

Decker, T. and Kovarik, P. (2000). Serine phosphorylation of STATs. *Oncogene* **19**, 2628-37.

Dierick, H. and Bejsovec, A. (1999). Cellular mechanisms of wingless/Wnt signal transduction. *Curr Top Dev Biol* **43**, 153-90.

Doll, R. F., Kaufman, P. D., Misra, S. and Rio, D. C. (1989). Molecular biology of Drosophila P-element transposition. *Prog Nucleic Acid Res Mol Biol* **36**, 47-57.

Eberl, D. F., Perkins, L. A., Engelstein, M., Hilliker, A. J. and Perrimon, N. (1992). Genetic and developmental analysis of polytene section 17 of the X chromosome of Drosophila melanogaster. *Genetics* **130**, 569-83.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755-63.

Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B. Z. and Barkai, N. (2002). Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature* **419**, 304-8.

Endo, T. A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H. et al. (1997). A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* **387**, 921-4.

Engels, W. R. (1989). P elements in Drosophila. In *Berg, Howe, 1989*, (ed., pp. 437--484.

Engels, W. R. (1997). Invasions of P elements. *Genetics* **145**, 11-5.

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. and Sved, J. (1990). High-frequency P element loss in Drosophila is homolog dependent. *Cell* **62**, 515-25.

Evans, C. J., Hartenstein, V. and Banerjee, U. (2003). Thicker than blood: conserved mechanisms in Drosophila and vertebrate hematopoiesis. *Dev Cell* **5**, 673-90.

Favre, H., Benhamou, A., Finidori, J., Kelly, P. A. and Edery, M. (1999). Dual effects of suppressor of cytokine signaling (SOCS-2) on growth hormone signal transduction. *FEBS Lett* **453**, 63-6.

FlyBase Consortium. (2003). The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Res* **31**, 172-5.

Foster, F. M., Traer, C. J., Abraham, S. M. and Fry, M. J. (2003). The phosphoinositide (PI) 3-kinase family. *J Cell Sci* **116**, 3037-40.

Fujimoto, M. and Naka, T. (2003). Regulation of cytokine signaling by SOCS family molecules. *Trends Immunol* **24**, 659-66.

Furukawa, M., He, Y. J., Borchers, C. and Xiong, Y. (2003). Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat Cell Biol* **5**, 1001-7.

Furukawa, M., Ohta, T. and Xiong, Y. (2002). Activation of UBC5 ubiquitin-conjugating enzyme by the RING finger of ROC1 and assembly of active ubiquitin ligases by all cullins. *J Biol Chem* **277**, 15758-65.

Garcia-Bellido, A., Cortes, F. and Milan, M. (1994). Cell interactions in the control of size in Drosophila wings. *Proc Natl Acad Sci U S A* **91**, 10222-6.

Geer, L. Y., Domrachev, M., Lipman, D. J. and Bryant, S. H. (2002). CDART: protein homology by domain architecture. *Genome Res* **12**, 1619-23.

Giordano, V., De Falco, G., Chiari, R., Quinto, I., Pelicci, P. G., Bartholomew, L., Delmastro, P., Gadina, M. and Scala, G. (1997). Shc mediates IL-6 signaling by interacting with gp130 and Jak2 kinase. *J Immunol* **158**, 4097-103.

Gloor, G. B., Moretti, J., Mouyal, J. and Keeler, K. J. (2000). Distinct P-element excision products in somatic and germline cells of Drosophila melanogaster. *Genetics* **155**, 1821-30.

Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R. and Engels, W. R. (1991). Targeted gene replacement in Drosophila via P element-induced gap repair. *Science* **253**, 1110-7.

Gonzalez-Reyes, A., Elliott, H. and St Johnston, D. (1995). Polarization of both major body axes in Drosophila by gurken-torpedo signalling. *Nature* **375**, 654-8.

Gonzalez-Reyes, A. and St Johnston, D. (1998). Patterning of the follicle cell epithelium along the anterior-posterior axis during Drosophila oogenesis. *Development* **125**, 2837-46.

Gouilleux-Gruart, V., Gouilleux, F., Desaint, C., Claisse, J. F., Capiod, J. C., Delobel, J., Weber-Nordt, R., Dusanter-Fourt, I., Dreyfus, F., Groner, B. et al. (1996). STAT-related transcription factors are constitutively activated in peripheral blood cells from acute leukemia patients. *Blood* **87**, 1692-7.

Graves, J. D. and Krebs, E. G. (1999). Protein phosphorylation and signal transduction. *Pharmacol Ther* **82**, 111-21.

Greenspan, R. J. (1997). Fly pushing: the theory and practice of Drosophila genetics. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Gresser, I., Aguet, M., Morel-Maroger, L., Woodrow, D., Puvion-Dutilleul, F., Guillon, J. C. and Maury, C. (1981). Electrophoretically pure mouse interferon inhibits growth, induces liver and kidney lesions, and kills suckling mice. *Am J Pathol* **102**, 396-402.

Guichard, A., Biehs, B., Sturtevant, M. A., Wickline, L., Chacko, J., Howard, K. and Bier, E. (1999). rhomboid and Star interact synergistically to promote EGFR/MAPK signaling during Drosophila wing vein development. *Development* **126**, 2663-76.

Haan, S., Ferguson, P., Sommer, U., Hiremath, M., McVicar, D. W., Heinrich, P. C., Johnston, J. A. and Cacalano, N. A. (2003). Tyrosine phosphorylation disrupts elongin interaction and accelerates SOCS3 degradation. *J Biol Chem* **278**, 31972-9.

Hafen, E., Dickson, B., Brunner, D. and Raabe, T. (1994). Genetic dissection of signal transduction mediated by the sevenless receptor tyrosine kinase in Drosophila. *Prog Neurobiol* **42**, 287-92.

Handler, A. M. (2001). A current perspective on insect gene transformation. *Insect Biochem Mol Biol* **31**, 111-28.

Handler, A. M. and Harrell, R. A., 2nd. (1999). Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect Mol Biol* **8**, 449-57.

Hanratty, W. P. and Dearolf, C. R. (1993). The Drosophila Tumorous-lethal hematopoietic oncogene is a dominant mutation in the hopscotch locus. *Mol Gen Genet* **238**, 33-7.

Hansen, J. A., Lindberg, K., Hilton, D. J., Nielsen, J. H. and Billestrup, N. (1999). Mechanism of inhibition of growth hormone receptor signaling by suppressor of cytokine signaling proteins. *Mol Endocrinol* **13**, 1832-43.

Hari, K. L., Cook, K. R. and Karpen, G. H. (2001). The Drosophila Su(var)2-10 locus regulates chromosome structure and function and encodes a member of the PIAS protein family. *Genes Dev* **15**, 1334-48.

Harrison, D. A., Binari, R., Nahreini, T. S., Gilman, M. and Perrimon, N. (1995). Activation of a Drosophila Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. *Embo J* **14**, 2857-65.

Harrison, D. A., McCoon, P. E., Binari, R., Gilman, M. and Perrimon, N. (1998). Drosophila unpaired encodes a secreted protein that activates the JAK signaling pathway. *Genes Dev* **12**, 3252-63.

Heinrich, P. C., Behrmann, I., Haan, S., Hermanns, H. M., Muller-Newen, G. and Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J* **374**, 1-20.

Hilton, D. J., Richardson, R. T., Alexander, W. S., Viney, E. M., Willson, T. A., Sprigg, N. S., Starr, R., Nicholson, S. E., Metcalf, D. and Nicola, N. A. (1998). Twenty proteins containing a C-terminal SOCS box form five structural classes. *Proc Natl Acad Sci U S A* **95**, 114-9.

Ho, J. M., Beattie, B. K., Squire, J. A., Frank, D. A. and Barber, D. L. (1999). Fusion of the ets transcription factor TEL to Jak2 results in constitutive Jak-Stat signaling. *Blood* **93**, 4354-64.

Hombria, J. C. and Brown, S. (2002). The fertile field of Drosophila Jak/STAT signalling. *Curr Biol* **12**, R569-75.

Hou, X. S., Melnick, M. B. and Perrimon, N. (1996). Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes a protein similar to the mammalian STATs. *Cell* **84**, 411-9.

Igaz, P., Toth, S. and Falus, A. (2001). Biological and clinical significance of the JAK-STAT pathway; lessons from knockout mice. *Inflamm Res* **50**, 435-41.

Imada, K. and Leonard, W. J. (2000). The Jak-STAT pathway. *Mol Immunol* **37**, 1-11.

Isaac, D. D. and Andrew, D. J. (1996). Tubulogenesis in Drosophila: a requirement for the trachealess gene product. *Genes Dev* **10**, 103-17.

Ito, K., Awano, W., Suzuki, K., Hiromi, Y. and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124**, 761-71.

Ivashkiv, L. B. and Hu, X. (2003). The JAK/STAT pathway in rheumatoid arthritis: pathogenic or protective? *Arthritis Rheum* **48**, 2092-6.

Iwamoto, T., Oshima, K., Seng, T., Feng, X., Oo, M. L., Hamaguchi, M. and Matsuda, S. (2002). STAT and SMAD signaling in cancer. *Histol Histopathol* **17**, 887-95.

Jiao, H., Berrada, K., Yang, W., Tabrizi, M., Platanias, L. C. and Yi, T. (1996). Direct association with and dephosphorylation of Jak2 kinase by the SH2-domain-containing protein tyrosine phosphatase SHP-1. *Mol Cell Biol* **16**, 6985-92.

Jinks, T. M., Polydorides, A. D., Calhoun, G. and Schedl, P. (2000). The JAK/STAT signaling pathway is required for the initial choice of sexual identity in Drosophila melanogaster. *Mol Cell* **5**, 581-7.

Johnson-Schlitz, D. M. and Engels, W. R. (1993). P-element-induced interallelic gene conversion of insertions and deletions in Drosophila melanogaster. *Mol Cell Biol* **13**, 7006-18.

Johnston, J. A. and O'Shea, J. J. (2003). Matching SOCS with function. *Nat Immunol* **4**, 507-9.

Kamura, T., Sato, S., Haque, D., Liu, L., Kaelin, W. G., Jr., Conaway, R. C. and Conaway, J. W. (1998). The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. *Genes Dev* **12**, 3872-81.

Karsten, P., Hader, S. and Zeidler, M. P. (2002). Cloning and expression of Drosophila SOCS36E and its potential regulation by the JAK/STAT pathway. *Mech Dev* **117**, 343-6.

Kaufman, P. D., Doll, R. F. and Rio, D. C. (1989). Drosophila P element transposase recognizes internal P element DNA sequences. *Cell* **59**, 359-71.

Kaufman, P. D. and Rio, D. C. (1992). P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. *Cell* **69**, 27-39.

Kawata, T., Shevchenko, A., Fukuzawa, M., Jermyn, K. A., Totty, N. F., Zhukovskaya, N. V., Sterling, A. E., Mann, M. and Williams, J. G. (1997). SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. *Cell* **89**, 909-16.

Kelley, M. R., Kidd, S., Berg, R. L. and Young, M. W. (1987). Restriction of P-element insertions at the Notch locus of Drosophila melanogaster. *Mol Cell Biol* **7**, 1545-8.

Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. and Fuller, M. T. (2001). Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**, 2542-5.

Krebs, D. L., Uren, R. T., Metcalf, D., Rakar, S., Zhang, J. G., Starr, R., De Souza, D. P., Hanzinikolas, K., Eyles, J., Connolly, L. M. et al. (2002). SOCS-6 binds to insulin receptor substrate 4, and mice lacking the SOCS-6 gene exhibit mild growth retardation. *Mol Cell Biol* **22**, 4567-78.

Kubo, M., Hanada, T. and Yoshimura, A. (2003). Suppressors of cytokine signaling and immunity. *Nat Immunol* **4**, 1169-76.

Lacronique, V., Boureux, A., Valle, V. D., Poirel, H., Quang, C. T., Mauchauffe, M., Berthou, C., Lessard, M., Berger, R., Ghysdael, J. et al. (1997). A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**, 1309-12.

Lanot, R., Zachary, D., Holder, F. and Meister, M. (2001). Postembryonic hematopoiesis in Drosophila. *Dev Biol* **230**, 243-57.

Leaman, D. W., Pisharody, S., Flickinger, T. W., Commane, M. A., Schlessinger, J., Kerr, I. M., Levy, D. E. and Stark, G. R. (1996). Roles of JAKs in activation of STATs and stimulation of c-fos gene expression by epidermal growth factor. *Mol Cell Biol* **16**, 369-75.

Lebestky, T., Chang, T., Hartenstein, V. and Banerjee, U. (2000). Specification of Drosophila hematopoietic lineage by conserved transcription factors. *Science* **288**, 146-9.

Leonard, W. J. (1996). Dysfunctional cytokine receptor signaling in severe combined immunodeficiency. *J Investig Med* **44**, 304-11.

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P. and Bork, P. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32 Database issue**, D142-4.

Lewis, E. B. and Bacher, F. (1968). Methods of feeding ethyl methane sulfonate (EMS) to Drosophila males. *D. I. S.* **43**, 193.

Lewis, S. E., Searle, S. M., Harris, N., Gibson, M., Lyer, V., Richter, J., Wiel, C., Bayraktaroglir, L., Birney, E., Crosby, M. A. et al. (2002). Apollo: a sequence annotation editor. *Genome Biol* **3**, RESEARCH0082.

Liu, B., Liao, J., Rao, X., Kushner, S. A., Chung, C. D., Chang, D. D. and Shuai, K. (1998). Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci U S A* **95**, 10626-31.

Liu, E., Cote, J. F. and Vuori, K. (2003). Negative regulation of FAK signaling by SOCS proteins. *Embo J* **22**, 5036-46.

Luo, H., Asha, H., Kockel, L., Parke, T., Mlodzik, M. and Dearolf, C. R. (1999). The Drosophila Jak kinase hopscotch is required for multiple developmental processes in the eye. *Dev Biol* **213**, 432-41.

Luo, H. and Dearolf, C. R. (2001). The JAK/STAT pathway and Drosophila development. *Bioessays* **23**, 1138-47.

Luo, H., Hanratty, W. P. and Dearolf, C. R. (1995). An amino acid substitution in the Drosophila hopTum-l Jak kinase causes leukemia-like hematopoietic defects. *Embo J* **14**, 1412-20.

Luo, H., Rose, P., Barber, D., Hanratty, W. P., Lee, S., Roberts, T. M., D'Andrea, A. D. and Dearolf, C. R. (1997). Mutation in the Jak kinase JH2 domain hyperactivates Drosophila and mammalian Jak-Stat pathways. *Mol Cell Biol* **17**, 1562-71.

Luo, H., Rose, P. E., Roberts, T. M. and Dearolf, C. R. (2002). The Hopscotch Jak kinase requires the Raf pathway to promote blood cell activation and differentiation in Drosophila. *Mol Genet Genomics* **267**, 57-63.

Manning, G. and Krasnow, M. A. (1993). Development of the Drosophila tracheal system. In *Bate, Martinez Arias, 1993*, (ed., pp. 609--685.

Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J. et al. (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **31**, 383-7.

Marcus, J. M. (2001). The development and evolution of crossveins in insect wings. *J Anat* **199**, 211-6.

Marine, J. C., McKay, C., Wang, D., Topham, D. J., Parganas, E., Nakajima, H., Pendeville, H., Yasukawa, H., Sasaki, A., Yoshimura, A. et al. (1999a). SOCS3 is essential in the regulation of fetal liver erythropoiesis. *Cell* **98**, 617-27.

Marine, J. C., Topham, D. J., McKay, C., Wang, D., Parganas, E., Stravopodis, D., Yoshimura, A. and Ihle, J. N. (1999b). SOCS1 deficiency causes a lymphocyte-dependent perinatal lethality. *Cell* **98**, 609-16.

Matsumoto, A., Seki, Y., Kubo, M., Ohtsuka, S., Suzuki, A., Hayashi, I., Tsuji, K., Nakahata, T., Okabe, M., Yamada, S. et al. (1999). Suppression of STAT5 functions in liver, mammary glands, and T cells in cytokine-inducible SH2-containing protein 1 transgenic mice. *Mol Cell Biol* **19**, 6396-407.

McGregor, J. R., Xi, R. and Harrison, D. A. (2002). JAK signaling is somatically required for follicle cell differentiation in Drosophila. *Development* **129**, 705-17.

Metcalf, D., Alexander, W. S., Elefanty, A. G., Nicola, N. A., Hilton, D. J., Starr, R., Mifsud, S. and Di Rago, L. (1999). Aberrant hematopoiesis in mice with inactivation of the gene encoding SOCS-1. *Leukemia* **13**, 926-34.

Metcalf, D., Greenhalgh, C. J., Viney, E., Willson, T. A., Starr, R., Nicola, N. A., Hilton, D. J. and Alexander, W. S. (2000). Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* **405**, 1069-73.

Methot, N. and Basler, K. (1999). Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819-31.

Metzger, R. J. and Krasnow, M. A. (1999). Genetic control of branching morphogenesis. *Science* **284**, 1635-9.

Miklos, G. L. and Rubin, G. M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**, 521-9.

Mohr, S. E. and Boswell, R. E. (1999). Zimp encodes a homologue of mouse Miz1 and PIAS3 and is an essential gene in Drosophila melanogaster. *Gene* **229**, 109-16.

Mohr, S. E. and Gelbart, W. M. (2002). Using the P[wHy] hybrid transposable element to disrupt genes in region 54D-55B in Drosophila melanogaster. *Genetics* **162**, 165-76.

Mooney, R. A., Senn, J., Cameron, S., Inamdar, N., Boivin, L. M., Shang, Y. and Furlanetto, R. W. (2001). Suppressors of cytokine signaling-1 and -6 associate with and inhibit the insulin receptor. A potential mechanism for cytokine-mediated insulin resistance. *J Biol Chem* **276**, 25889-93.

Morimoto, A. M., Jordan, K. C., Tietze, K., Britton, J. S., O'Neill, E. M. and Ruohola-Baker, H. (1996). Pointed, an ETS domain transcription factor, negatively regulates the EGF receptor pathway in Drosophila oogenesis. *Development* **122**, 3745-54.

Moustakas, A. (2002). Smad signalling network. *J Cell Sci* **115**, 3355-6.

Mullor, J. L., Calleja, M., Capdevila, J. and Guerrero, I. (1997). Hedgehog activity, independent of decapentaplegic, participates in wing disc patterning. *Development* **124**, 1227-37.

Naka, T. and Kishimoto, T. (2002). Joint disease caused by defective gp130-mediated STAT signaling. *Arthritis Res* **4**, 154-6.

Naka, T., Matsumoto, T., Narazaki, M., Fujimoto, M., Morita, Y., Ohsawa, Y., Saito, H., Nagasawa, T., Uchiyama, Y. and Kishimoto, T. (1998). Accelerated apoptosis of lymphocytes by augmented induction of Bax in SSI-1 (STAT-induced STAT inhibitor-1) deficient mice. *Proc Natl Acad Sci U S A* **95**, 15577-82.

Naka, T., Narazaki, M., Hirata, M., Matsumoto, T., Minamoto, S., Aono, A., Nishimoto, N., Kajita, T., Taga, T., Yoshizaki, K. et al. (1997). Structure and function of a new STAT-induced STAT inhibitor. *Nature* **387**, 924-9.

Naka, T., Nishimoto, N. and Kishimoto, T. (2002). The paradigm of IL-6: from basic science to medicine. *Arthritis Res* **4 Suppl 3**, S233-42.

Narazaki, M., Fujimoto, M., Matsumoto, T., Morita, Y., Saito, H., Kajita, T., Yoshizaki, K., Naka, T. and Kishimoto, T. (1998). Three distinct domains of SSI-1/SOCS-1/JAB protein are required for its suppression of interleukin 6 signaling. *Proc Natl Acad Sci U S A* **95**, 13130-4.

Nassif, N. and Engels, W. (1993). DNA homology requirements for mitotic gap repair in Drosophila. *Proc Natl Acad Sci U S A* **90**, 1262-6.

Neubauer, H., Cumano, A., Muller, M., Wu, H., Huffstadt, U. and Pfeffer, K. (1998). Jak2 deficiency defines an essential developmental checkpoint in definitive hematopoiesis. *Cell* **93**, 397-409.

Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. and Edgar, B. A. (1998). Coordination of growth and cell division in the Drosophila wing. *Cell* **93**, 1183-93.

Nicholson, S. E., De Souza, D., Fabri, L. J., Corbin, J., Willson, T. A., Zhang, J. G., Silva, A., Asimakis, M., Farley, A., Nash, A. D. et al. (2000). Suppressor of cytokine signaling-3 preferentially binds to the SHP-2-binding site on the shared cytokine receptor subunit gp130. *Proc Natl Acad Sci U S A* **97**, 6493-8.

Nicholson, S. E., Willson, T. A., Farley, A., Starr, R., Zhang, J. G., Baca, M., Alexander, W. S., Metcalf, D., Hilton, D. J. and Nicola, N. A. (1999). Mutational analyses of the SOCS proteins suggest a dual domain requirement but distinct mechanisms for inhibition of LIF and IL-6 signal transduction. *Embo J* **18**, 375-85.

Nilson, L. A. and Schupbach, T. (1999). EGF receptor signaling in Drosophila oogenesis. *Curr Top Dev Biol* **44**, 203-43.

Nusslein-Volhard, C. (1979). Maternal effect mutations that affect the spatial coordinates of the embryo of Drosophila melanogaster. In *Subtelny, Konigsberg, 1979*, (ed., pp. 185--211.

Nusslein-Volhard, C., Frohnhofer, H. G. and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. *Science* **238**, 1675-81.

Nusslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795-801.

Nusslein-Volhard, C., Wieschaus, E. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. *I. Zygotic loci on the second chromosome.* **193**, 267--282.

O'Brien, K. B., O'Shea, J. J. and Carter-Su, C. (2002). SH2-B family members differentially regulate JAK family tyrosine kinases. *J Biol Chem* **277**, 8673-81.

O'Brochta, D. A., Gomez, S. P. and Handler, A. M. (1991). P element excision in Drosophila melanogaster and related drosophilids. *Mol Gen Genet* **225**, 387-94.

Oh, S. W., Kingsley, T., Shin, H. H., Zheng, Z., Chen, H. W., Chen, X., Wang, H., Ruan, P., Moody, M. and Hou, S. X. (2003). A P-element insertion screen identified mutations in 455 novel essential genes in Drosophila. *Genetics* **163**, 195-201.

O'Hare, K. and Rubin, G. M. (1983). Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. *Cell* **34**, 25-35.

Olivier, J. P., Raabe, T., Henkemeyer, M., Dickson, B., Mbamalu, G., Margolis, B., Schlessinger, J., Hafen, E. and Pawson, T. (1993). A Drosophila SH2-SH3 adaptor protein implicated in coupling the sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. *Cell* **73**, 179-91.

O'Shea, J. J., Gadina, M. and Schreiber, R. D. (2002). Cytokine signaling in 2002: new surprises in the Jak/Stat pathway. *Cell* **109 Suppl**, S121-31.

Parganas, E., Wang, D., Stravopodis, D., Topham, D. J., Marine, J. C., Teglund, S., Vanin, E. F., Bodner, S., Colamonici, O. R., van Deursen, J. M. et al. (1998). Jak2 is essential for signaling through a variety of cytokine receptors. *Cell* **93**, 385-95.

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E. et al. (2004). Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nat Genet* **36**, 288-92.

Pastink, A., Heemskerk, E., Nivard, M. J., van Vliet, C. J. and Vogel, E. W. (1991). Mutational specificity of ethyl methanesulfonate in excision-repair-proficient and -deficient strains of Drosophila melanogaster. *Mol Gen Genet* **229**, 213-8.

Patel, N. H. (1994). Imaging neuronal subsets and other cell types in whole-mount Drosophila embryos and larvae using antibody probes. *Methods Cell Biol* **44**, 445-87.

Peeters, P., Raynaud, S. D., Cools, J., Wlodarska, I., Grosgeorge, J., Philip, P., Monpoux, F., Van Rompaey, L., Baens, M., Van den Berghe, H. et al. (1997). Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood* **90**, 2535-40.

Peltola, K. J., Paukku, K., Aho, T. L., Ruuska, M., Silvennoinen, O. and Koskinen, P. J. (2004). Pim-1 kinase inhibits STAT5-dependent transcription via its interactions with SOCS1 and SOCS3. *Blood* **103**, 3744-50.

Perrimon, N., Engstrom, L. and Mahowald, A. P. (1989). Zygotic lethals with specific maternal effect phenotypes in Drosophila melanogaster. I. Loci on the X chromosome. *Genetics* **121**, 333-52.

Perrimon, N. and Mahowald, A. P. (1986). l(1)hopscotch, A larval-pupal zygotic lethal with a specific maternal effect on segmentation in Drosophila. *Dev Biol* **118**, 28-41.

Preston, C. R., Sved, J. A. and Engels, W. R. (1996). Flanking duplications and deletions associated with P-induced male recombination in Drosophila. *Genetics* **144**, 1623-38.

Qu, C. K. (2002). Role of the SHP-2 tyrosine phosphatase in cytokine-induced signaling and cellular response. *Biochim Biophys Acta* **1592**, 297-301.

Ram, P. A. and Waxman, D. J. (2000). Role of the cytokine-inducible SH2 protein CIS in desensitization of STAT5b signaling by continuous growth hormone. *J Biol Chem* **275**, 39487-96.

Rane, S. G. and Reddy, E. P. (2000). Janus kinases: components of multiple signaling pathways. *Oncogene* **19**, 5662-79.

Rawlings, J. S., Rosler, K. M. and Harrison, D. A. (2004). The JAK/STAT signaling pathway. *J Cell Sci* **117**, 1281-3.

Rio, D. C. and Rubin, G. M. (1988). Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the P transposable element. *Proc Natl Acad Sci U S A* **85**, 8929-33.

Robertson, H. M., Preston, C. R., Phillis, R. W., Johnson-Schlitz, D. M., Benz, W. K. and Engels, W. R. (1988). A stable genomic source of P element transposase in Drosophila melanogaster. *Genetics* **118**, 461-70.

Rong, Y. S. and Golic, K. G. (2001). A targeted gene knockout in drosophila. *Genetics* **157**, 1307-12.

Rong, Y. S., Titen, S. W., Xie, H. B., Golic, M. M., Bastiani, M., Bandyopadhyay, P., Olivera, B. M., Brodsky, M., Rubin, G. M. and Golic, K. G. (2002). Targeted mutagenesis by homologous recombination in D. melanogaster. *Genes Dev* **16**, 1568-81.

Rorth, P. (1998). Gal4 in the Drosophila female germline. *Mech Dev* **78**, 113-8.

Roseman, R. R., Johnson, E. A., Rodesch, C. K., Bjerke, M., Nagoshi, R. N. and Geyer, P. K. (1995). A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. *Genetics* **141**, 1061-74.

Roth, S., Neuman-Silberberg, F. S., Barcelo, G. and Schupbach, T. (1995). cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. *Cell* **81**, 967-78.

Rottapel, R. (2001). Putting the brakes on arthritis: can suppressors of cytokine signaling (SOCS) suppress rheumatoid arthritis? *J Clin Invest* **108**, 1745-7.

Sakamoto, H., Yasukawa, H., Masuhara, M., Tanimura, S., Sasaki, A., Yuge, K., Ohtsubo, M., Ohtsuka, A., Fujita, T., Ohta, T. et al. (1998). A Janus kinase inhibitor, JAB, is an interferon-gamma-inducible gene and confers resistance to interferons. *Blood* **92**, 1668-76.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7.

Sawamoto, K., Okabe, M., Tanimura, T., Mikoshiba, K., Nishida, Y. and Okano, H. (1996). The Drosophila secreted protein Argos regulates signal transduction in the Ras/MAPK pathway. *Dev Biol* **178**, 13-22.

Sawamoto, K., Okano, H., Kobayakawa, Y., Hayashi, S., Mikoshiba, K. and Tanimura, T. (1994). The function of argos in regulating cell fate decisions during Drosophila eye and wing vein development. *Dev Biol* **164**, 267-76.

Schindler, C. (1999). Cytokines and JAK-STAT signaling. *Exp Cell Res* **253**, 7-14.

Schnepp, B., Grumbling, G., Donaldson, T. and Simcox, A. (1996). Vein is a novel component in the Drosophila epidermal growth factor receptor pathway with similarity to the neuregulins. *Genes Dev* **10**, 2302-13.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**, 231-4.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857-64.

Schupbach, T. and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *III. Zygotic loci on the X-chromosome and fourth chromosome.* **195**, 302--317.

Schupbach, T. and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. *Genetics* **121**, 101-17.

Schupbach, T. and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119-36.

Schweisguth, F. (2004). Notch signaling activity. *Curr Biol* **14**, R129-38.

Schweitzer, R., Howes, R., Smith, R., Shilo, B. Z. and Freeman, M. (1995). Inhibition of Drosophila EGF receptor activation by the secreted protein Argos. *Nature* **376**, 699-702.

Seki, Y., Hayashi, K., Matsumoto, A., Seki, N., Tsukada, J., Ransom, J., Naka, T., Kishimoto, T., Yoshimura, A. and Kubo, M. (2002). Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proc Natl Acad Sci U S A* **99**, 13003-8.

Sentry, J. W. and Kaiser, K. (1992). P element transposition and targeted manipulation of the Drosophila genome. *Trends Genet* **8**, 329-31.

Shi, Y. and Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700.

Shilo, B. Z., Gabay, L., Glazer, L., Reichman-Fried, M., Wappner, P., Wilk, R. and Zelzer, E. (1997). Branching morphogenesis in the Drosophila tracheal system. *Cold Spring Harb Symp Quant Biol* **62**, 241-7.

Shouda, T., Yoshida, T., Hanada, T., Wakioka, T., Oishi, M., Miyoshi, K., Komiya, S., Kosai, K., Hanakawa, Y., Hashimoto, K. et al. (2001). Induction of the cytokine signal regulator SOCS3/CIS3 as a therapeutic strategy for treating inflammatory arthritis. *J Clin Invest* **108**, 1781-8.

Shuai, K. (2000). Modulation of STAT signaling by STAT-interacting proteins. *Oncogene* **19**, 2638-44.

Silver, D. L. and Montell, D. J. (2001). Paracrine signaling through the JAK/STAT pathway activates invasive behavior of ovarian epithelial cells in Drosophila. *Cell* **107**, 831-41.

Small, S., Blair, A. and Levine, M. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Dev Biol* **175**, 314-24.

Spradling, A. C. (1986). P element-mediated transformation. In *Roberts, 1986*, (ed., pp. 175--198.

Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverty, T., Mozden, N., Misra, S. and Rubin, G. M. (1999). The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135-77.

Starr, R., Metcalf, D., Elefanty, A. G., Brysha, M., Willson, T. A., Nicola, N. A., Hilton, D. J. and Alexander, W. S. (1998). Liver degeneration and lymphoid deficiencies in mice lacking suppressor of cytokine signaling-1. *Proc Natl Acad Sci U S A* **95**, 14395-9.

Starr, R., Willson, T. A., Viney, E. M., Murray, L. J., Rayner, J. R., Jenkins, B. J., Gonda, T. J., Alexander, W. S., Metcalf, D., Nicola, N. A. et al. (1997). A family of cytokine-inducible inhibitors of signalling. *Nature* **387**, 917-21.

Struhl, G. and Basler, K. (1993). Organizing activity of wingless protein in Drosophila. *Cell* **72**, 527-40.

Sturtevant, M. A., Roark, M. and Bier, E. (1993). The Drosophila rhomboid gene mediates the localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway. *Genes Dev* **7**, 961-73.

Taguchi, A., Sawamoto, K. and Okano, H. (2000). Mutations Modulating the Argos-Regulated Signaling Pathway in Drosophila Eye Development. *Genetics* **154**, 1639-1648.

Takaoka, A., Tanaka, N., Mitani, Y., Miyazaki, T., Fujii, H., Sato, M., Kovarik, P., Decker, T., Schlessinger, J. and Taniguchi, T. (1999). Protein tyrosine kinase Pyk2 mediates the Jak-dependent activation of MAPK and Stat1 in IFN-gamma, but not IFN-alpha, signaling. *Embo J* **18**, 2480-8.

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L. et al. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat Genet* **36**, 283-7.

Timakov, B., Liu, X., Turgut, I. and Zhang, P. (2002). Timing and targeting of P-element local transposition in the male germline cells of Drosophila melanogaster. *Genetics* **160**, 1011-22.

Torok, T., Tick, G., Alvarado, M. and Kiss, I. (1993). P-lacW insertional mutagenesis on the second chromosome of Drosophila melanogaster: isolation of lethals with different overgrowth phenotypes. *Genetics* **135**, 71--80.

Tower, J., Karpen, G. H., Craig, N. and Spradling, A. C. (1993). Preferential transposition of Drosophila P elements to nearby chromosomal sites. *Genetics* **133**, 347-59.

Tracey, W. D., Jr., Ning, X., Klingler, M., Kramer, S. G. and Gergen, J. P. (2000). Quantitative analysis of gene function in the Drosophila embryo. *Genetics* **154**, 273-84.

Tulina, N. and Matunis, E. (2001). Control of stem cell self-renewal in Drosophila spermatogenesis by JAK-STAT signaling. *Science* **294**, 2546-9.

Verdier, F., Chretien, S., Muller, O., Varlet, P., Yoshimura, A., Gisselbrecht, S., Lacombe, C. and Mayeux, P. (1998). Proteasomes regulate erythropoietin receptor and signal transducer and activator of transcription 5 (STAT5) activation. Possible involvement of the ubiquitinated Cis protein. *J Biol Chem* **273**, 28185-90.

Ward, A. C., Touw, I. and Yoshimura, A. (2000). The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood* **95**, 19-29.

Wes, P. D., Yu, M. and Montell, C. (1996). RIC, a calmodulin-binding Ras-like GTPase. *Embo J* **15**, 5839-48.

Wilkie, G. S., Shermoen, A. W., O'Farrell, P. H. and Davis, I. (1999). Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei. *Curr Biol* **9**, 1263-6.

Williams, J. A., Paddock, S. W. and Carroll, S. B. (1993). Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. *Development* **117**, 571-84.

Wormald, S. and Hilton, D. J. (2004). Inhibitors of cytokine signal transduction. *J Biol Chem* **279**, 821-4.

Xi, R., McGregor, J. R. and Harrison, D. A. (2003). A gradient of JAK pathway activity patterns the anterior-posterior axis of the follicular epithelium. *Dev Cell* **4**, 167-77.

Yaffe, M. B. (2002). Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol* **3**, 177-86.

Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., Sekine, N., Wada, M., Honjo, M., Takahashi, M., Takahashi, T., Hirai, H. et al. (1998). Growth hormone-induced tyrosine phosphorylation of EGF receptor as an essential element leading to MAP kinase activation and gene expression. *Endocr J* **45 Suppl**, S27-31.

Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., Sekine, N., Wada, M., Honjo, M., Takahashi, M., Takahashi, T., Hirai, H. et al. (1997). Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. *Nature* **390**, 91-6.

Yan, R., Luo, H., Darnell, J. E., Jr. and Dearolf, C. R. (1996a). A JAK-STAT pathway regulates wing vein formation in Drosophila. *Proc Natl Acad Sci U S A* **93**, 5842-7.

Yan, R., Small, S., Desplan, C., Dearolf, C. R. and Darnell, J. E., Jr. (1996b). Identification of a Stat gene that functions in Drosophila development. *Cell* **84**, 421-30.

Yasukawa, H., Misawa, H., Sakamoto, H., Masuhara, M., Sasaki, A., Wakioka, T., Ohtsuka, S., Imaizumi, T., Matsuda, T., Ihle, J. N. et al. (1999). The JAK-binding protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop. *Embo J* **18**, 1309-20.

Yeh, T. C. and Pellegrini, S. (1999). The Janus kinase family of protein tyrosine kinases and their role in signaling. *Cell Mol Life Sci* **55**, 1523-34.

Yoshimura, A., Ohkubo, T., Kiguchi, T., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Hara, T. and Miyajima, A. (1995). A novel cytokine-inducible gene CIS encodes an SH2-containing protein that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin receptors. *Embo J* **14**, 2816-26.

Zecca, M., Basler, K. and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the Drosophila wing. *Development* **121**, 2265-78.

Zeidler, M. P., Bach, E. A. and Perrimon, N. (2000). The roles of the Drosophila JAK/STAT pathway. *Oncogene* **19**, 2598-606.

Zeidler, M. P., Perrimon, N. and Strutt, D. I. (1999). Polarity determination in the Drosophila eye: a novel role for unpaired and JAK/STAT signaling. *Genes Dev* **13**, 1342-53.

Zelzer, E. and Shilo, B. Z. (2000). Cell fate choices in Drosophila tracheal morphogenesis. *Bioessays* **22**, 219-26.

Zhang, J. G., Farley, A., Nicholson, S. E., Willson, T. A., Zugaro, L. M., Simpson, R. J., Moritz, R. L., Cary, D., Richardson, R., Hausmann, G. et al. (1999). The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proc Natl Acad Sci U S A* **96**, 2071-6.

Zhang, Z., Schaffer, A. A., Miller, W., Madden, T. L., Lipman, D. J., Koonin, E. V. and Altschul, S. F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res* **26**, 3986-90.

Vita

DATE OF BIRTH

January 13, 1974

PLACE OF BIRTH

San Antonio, Texas

EDUCATION

Bachelor of Science, *cum laude* (1996)
Furman University

EMPLOYMENT

Part-time Instructor (2004)
Biology Program, Transylvania University

Teaching/ Research Assistant (1998-2003)
Department of Biology, University of Kentucky

AWARDS AND CERTIFICATES

Vanderbilt University Kennedy Center Postdoctoral fellowship (2004, declined)
Vanderbilt University Neurogenomics Training Grant Postdoctoral fellowship (2004, declined)
Graduate School Academic Year non-service fellowship (2001, 1998)
University of Kentucky Bioinformatics Certificate (2000)
Commandant's List, U.S. Army Officer Basic Course (2000)
Gertrude Flora Ribble Graduate Research Scholarship (1999)
Daughters of the American Revolution Medal (1996)
Four-year Army ROTC Scholarship (1992)

PUBLICATIONS

J.S. Rawlings, K.M. Rosler, and D.A. Harrison. (2004) The JAK/STAT Signaling Pathway. *Journal of Cell Science* **117**, 1281-1283.

J.S. Rawlings, G. Rennebeck, S.M.W. Harrision, R. Xi, and D.A. Harrison. (submitted) Two Suppressors of Cytokine Signaling (SOCS) differentially regulate JAK and EGFR pathway activities.