Identification and analysis of JAK/STAT pathway target genes in *Drosophila melanogaster*

Doctoral Thesis

In partial fulfilment of the requirements
for the degree "Doctor rerum naturalium (Dr. rer. nat.)"
in the Molecular Biology Program
at the Georg August University Göttingen,
Faculty of Biology

submitted by

SAMIRA BINA

Born in

Tehran

2009

AFFIDAVIT

Here I declare that my doctoral thesis entitled "Identification and analysis of JAK/STAT pathway target genes in *Drosophila melanogaster*" has been written independently with no other sources and aids than quoted.

Samira Bina

Sheffield, March 2009

LIST OF PUBLICATIONS

Publication

Bina, S., V. M. Wright, M. P. Zeidler (2009). Transcriptional targets of *Drosophila* JAK/STAT pathway signalling as effectors of haematopoietic tumour formation. (In preparation).

Bina, S., M. Zeidler (2009). JAK/STAT pathway signalling in *Drosophila melanogaster*. JAK-STAT pathway in disease. A. Stephanou, Austin, Landes Bioscience.

Selected Presentations

Bina, S., M. Zeidler (2008). Identification of direct target genes of the JAK/STAT signalling pathway. 10th International EMBL PhD Student Symposium (Heidelberg, Germany).

Bina, S., M. Zeidler (2008). Identification of direct target genes of the JAK/STAT signalling pathway. 49th Annual *Drosophila* Research Conference (San Diego, USA).

Bina, S., M. Zeidler (2006). Identification of JAK/STAT pathway target genes induced by Upd and Upd-like ligands in a microarray-based assay. 47th Annual *Drosophila* Research Conference (Houston, USA).

CONTENTS

TABLE OF CONTENTS

CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

The JAK/STAT signalling cascade is one of the most conserved pathways in the animal kingdom. Aside from its developmental roles, the JAK/STAT pathway is needed for stem cell maintenance, and its mis-regulation is associated with various malignancies, ranging from solid tumours to myeloproliferative disorders. Using *Drosophila* as a genetically tractable model organism has enabled the identification of the main JAK/STAT pathway components.

This work was aimed at identifying downstream effectors of the *Drosophila* JAK/STAT pathway which are be responsible for the generation and progression of blood tumours. Gene expression profiling identified 1197 loci that are regulated either directly or indirectly at different time points by the main pathway ligand UPD. Bioinformatic analysis of the 1197 genes showed a temporally dynamic distribution of functional categories relevant to immunity. Furthermore promoter analysis illustrated differential distribution of high and low affinity binding sites of the JAK/STAT pathway transcription factor among up-regulated genes. The significance of ten UPD-regulated genes in mediating tumourigenesis in flies was also validated *in vivo*. Among these are genes that are important in cell polarity, a function that has recently emerged as being important for polarised JAK/STAT signalling in epithelial cells.

Although the focus of this work was mainly on validating the role of JAK/STAT target genes in tumourigenesis, gene expression profiling has generated a non-exhaustive list of candidate genes that can also be used for studies of other roles that are mediated by the JAK/STAT pathway, such as cell movement and stem cell maintenance.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Signalling pathways in development

All multicellular organisms develop from a single fertilised egg which then forms the embryo ultimately giving rise to the adult. Although Mendel showed the relationship between development and inheritance, it was not clear if the nucleus or the cytoplasm of the fertilised egg could be attributed to this link. While geneticists believed that developmental processes were dependent on gene expression, their theory was dismissed by embryologists until they could prove how identical chromosomes within the nucleus of every cell could produce different cell types. It was not until the 1930s that Gluecksohn-Schoenheimer and Waddington showed that mutations of some genes affect early development (Gluecksohn-Schoenheimer 1938; Waddington 1939). Finally in the 1950s Briggs and King developed the somatic nuclear transfer technique and demonstrated that a complete tadpole can be generated by transferring the blastula cell nuclei into an enucleated frog egg (Briggs and King 1952). The same result was, however, not achieved using somatic cell nuclei, suggesting that although the content of the nuclei were identical, they acquired different abilities to direct development. In the early 1950s the paradigm of differential gene expression was born, a hypothesis, which could explain how despite having identical genetic material in all somatic cells, different cell types can be formed. Beermann showed that different regions of polytene chromosomes in different cells "puffed" out to enable mRNA transcription (Beermann 1952).

We now know that the combinatorial effects of transcription factors result in tissue-specific gene expression pattern (Xu et al. 2000). Often transcription factors are present in the cell, but need to be activated by signalling pathways. Signal transduction can be transmitted by direct contact between adjacent cells, a process called juxtacrine interaction. Alternatively, signalling molecules generated in distant cells can diffuse to and induce changes in the receiving cell. This event is called paracrine interaction and the diffusible molecules called paracrine factors. Combinatorial interplay of only a small number of signalling pathways induced by paracrine factors is needed to generate different tissues within an organism, a phenomenon, which can be encompassed under the term "pleiotropy" (Brivanlou and Darnell 2002).

The major signal transduction pathways can be broadly described as variations of a common linear scheme. Signalling molecules (ligands) must first be synthesised and released by the signalling cell and transported to the target cell. Binding of the ligand to its receptor induces a cascade of protein modifications which eventually leads to activation of transcription factor(s). The interaction of activated transcription factors with the basal transcription machinery leads to expression of downstream target genes which ultimately bring about changes in cellular metabolism, function and development. Signalling is eventually terminated at various levels of the pathway via regulatory feedback loop mechanisms.

Over 2000 transcription factors are known in humans and are normally divided into groups based on common structural elements, or their mode of activation and their role in developmental and cellular contexts (Brivanlou and Darnell 2002) (Table 1). Constitutive transcription factors, as their name suggests, are present in all cells at all times and are required for the transcription of genes that need to be ubiquitously expressed. Regulatory transcription factors on the other hand, are only present when a cell needs to undergo developmental changes, or they are activated when the cells receives a signal. Regulatory factors are divided into two categories, the developmental and the signal-dependent transcription factors. Developmental factors enter the nucleus as soon as they are translated and do not require any posttranslational modification. It is the combination of these factors that directs cell differentiation, rather than tissue-specific restriction of individual factors. Signal-dependent transcription factors are present in most cells in their latent form until a signalling cascade within the cells triggers their activation. Signal-dependent transcription factors are divided into three groups. Firstly, the steroid receptors which are all transcription factors themselves. Their ligands are lipophilic and diffuse across the cell membrane in order to bind to the receptor, which in turn dimerises with another identical receptor and translocates into the nucleus. Secondly, transcription factors that are activated by internal signals have recently been recognised, with the famous p53 being in this group (Levine 1997). Thirdly, transcription factors that are activated by cell surface receptors include those that are constitutively localised in the cell nucleus (resident nuclear factors), and those that are present in the cytoplasm in a latent state (latent cytoplasmic factors). The major signalling pathways (including the JAK/STAT) that signal via a cell-membrane anchored receptor, and deliver the transcription factor to the nucleus, are depicted in Table 2. These signalling pathways use diverse posttranslational modification processes to

activate the transcription factors. Although many signalling components of the pathways are conserved throughout the animal kingdom, they are not found in plants or fungi, suggesting a specificity for animal evolution.

The repeated use of all the above-described signalling pathways in a spatially and temporally controlled manner is a key event in development of multi-cellular organisms.

Classification of transcription factors based on their cellular role

Group	Example
CONSTITUTIVE	Sp1, CCAAT, NF1
REGULATORY	
Developmental (active as soon as translated)	GATA, HNFs, MyoD, Biocoid, Hox
Signal dependent (not active unless signal given to cell)	
Steroid receptor superfamily	estrogen, testosterone, retinoids
Internal signals	SREBP, p53, STEROID
Cell surface receptor-ligand	
Resident nuclear factors	ETS, CREBs, FOS-JUN, SRF, ATMs
Latent cytoplasmic factors	STATs, SMADs, CI/GLI, CATENINS, NFAT

Table 1.   Classification of transcription factors based on their cellular role. More than 2000 transcription factors are encoded in the human genome and are typically classified according to common structural domains. Alternatively, one can group factors based on their cellular role, as depicted here.

Cell surface receptors which activate latent cytoplasmic transcription factors

Cytokine receptors	RTKs	TGFβ receptors	Hh receptors	Wnt receptors	Notch receptor
Ligand	RTK	TGFβ Superfamily ligand	Hedghog	Wnt	Delta/Delta-like
Receptor	GNRP	Receptor type II	Patched	Frizzled	Notch
JAK	RAF	Receptor type I	Smoothened	Disheveled	Notch cleavage
STAT	MEK	Smad activation	Ci activation	GSK-3	NICD/nuclear TFs
STAT dimerisation	ERK	Transcription	Transcription	β-catenin	Transcription
Transcription	Transcription factor			Trascription	
	Transcription				

Table 2.   Cell surface receptors which activate latent cytoplasmic transcription factors.

Cytokine receptors are non-covalently associated to JAK kinases and are activated by a number of ligands. Activated JAKs phosphorylate and thereby activate STAT transcription factors. Receptor tyrosine kinases contain a cytosolic tyrosine kinase domain, which upon receptor dimerisation autophosphorylate the receptor. An adaptor protein links the phosphorylated RTK to cytosolic MAP kinases, which eventually translocate to the nucleus. TGFβ receptors contain a cytosolic serine/threonine kinase domain, which activates Smad transcription factors by phosphorylation. Binding of Hedgehog to its receptor Patched releases Ci from a cytosolic complex and prevents its cleavage allowing it to act as a transcriptional activator. Binding of a Wnt ligand to the seven transmembrane receptor complex enables the release of the transcription factor β-catenin from a multi-protein complex in the cytosol. Binding of the transmembrane ligand Delta to the Notch receptor triggers the proteolysis of the intracellular domain of Notch (NICD) that then accumulates in the nucleus.

The JAK/STAT signalling pathway

The conserved canonical (JAK-dependent) pathway

The canonical JAK/STAT signalling is initiated by binding of an extracellular ligand to a cell-surface cytokine receptor, thereby triggering a conformational change (Figure 1). The intracellular domains of many cytokine receptors are non-covalently associated with tyrosine kinases of the JAK family. Conformational change in the intracellular domain of the receptor leads to the activation of JAKs. These then tyrosine-phosphorylate each other and the receptor leading to recruitment of the inactive cytoplasmic STATs via their SRC homology 2 (SH2) domain. Once STATs are phosphorylated on a single tyrosine residue, they dimerise in a parallel fashion. Dimerised STATs enter the nucleus and in concert with other co-activators and chromatin remodelling factors, as well as the basal transcriptional machinery, initiate gene expression. In some cases full transcriptional STAT activity requires an additional serine phosphorylation (Decker and Kovarik 2000) but other post-translational modifications such as arginine methylation (Mowen et al. 2001; Meissner et al. 2004) and lysine acetylation (Wang et al. 2005; Yuan et al. 2005) have also been reported. Finally, STATs are inactivated by dephosphorylation and exported from the nucleus (Haspel and Darnell 1999; Begitt et al. 2000; McBride et al. 2000).

The JAK/STAT pathway is evolutionary conserved throughout the animal kingdom and has been examined in various model organisms such as the slime mold *Dictyosteliym discoideum*, the fruit fly *Drosophila melanogaster* and the zebrafish *Danio rerio*, as well as in mammals including mouse, rat and human (Table 3) (Hou et al. 2002; Rawlings et al. 2004). In mammals the JAK/STAT pathway components include four kinases JAK1-3 and TYK2 and seven STAT transcription factors STAT1-6 (including two isoforms of STAT5 called STAT5a and STAT5b). As shown in Table 3, STAT molecules are the only components of the JAK/STAT signalling pathway which are present in all model organisms. Over 40 peptides and numerous receptors have been reported to activate the mammalian JAK/STAT pathway (Schindler and Plumlee 2008). Regulators of the pathway include PIAS (Protein Inhibitors of Activated Stat) family members which inhibit STAT activity possibly by SUMOylation (Ungureanu et al. 2003). SOCS (Suppressor Of Cytokine Signalling) proteins negatively regulate the pathway by binding to and inhibiting the activity of the receptor (Yamamoto et al. 2003) or JAK (Yasukawa et al. 1999), or alternatively, they target JAK for degradation (Ali et al. 2003).
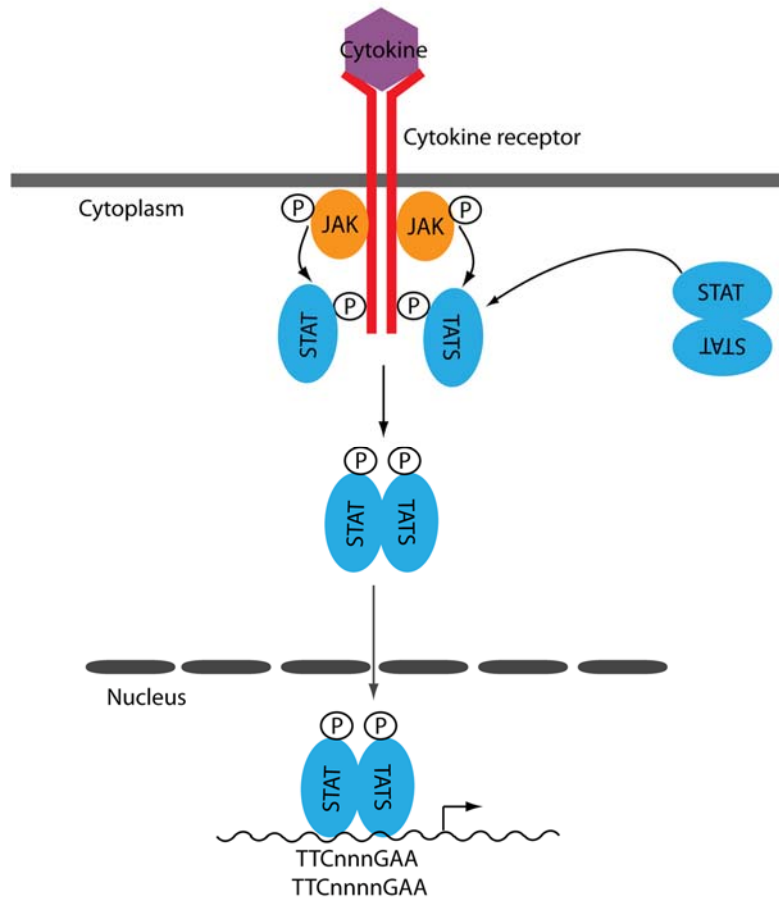
Figure 1.   The canonical JAK/STAT pathway.
Binding of a cytokine to a cell-surface cytokine receptor triggers receptor dimerisation/oligomerisation. The receptor-associated tyrosine kinases of the JAK family then auto-phosphorylate and phosphorylate the receptor. This leads to recruitment of STATs via their SH2 domains which interact with the phosphorylated receptor. STATs then become tyrosine-phosphorylated by the JAKs and form "parallel" dimers, wich enter the nucleus and bind to a palindromic consensus binding site and activate target genes. Once STATs dissociate from the DNA, they are dephosphorylated by nuclear tyrosine phosphatases.

Conserved JAK/STAT signalling components

	Mammals	Drosophila	C. elegans	Dictyostelium	Zebrafish
Ligand	Many cytokines	UPD, UPD2, UPD3	ND	cAMP, DIF	Many cytokines
Receptor	Many cytokine receptors	DOME, CG14255	ND	Serpentine receptor	Many cytokine receptors
Tyrosine kinase	JAK1-3, TYK2	HOP	ND	ND	JAK1-2
STAT	STAT1-6	STAT92E	STAT5	Dd-STATa, b, c, d	STAT1, 3, 4, 5
PIAS	PIAS1, 3, xα, xβ, γ	DPIAS	ND	ND	PIAS2, 4
SOCS	CIS, SOCS1-7	SOCS16D, SOCS36E, SOCS44E	ND	ND	SOCS3, 7

Table 3.   Conserved JAK/STAT signalling components.
JAK/STAT pathway components have been most extensively studied in mammalian models as well as in *Drosophila*. STAT-like molecules have also been found in *C. elegans*, *Dictyostelium* and zebrafish. ND: not determined. Modified after Hou et al. 2002.

Drosophila JAK/STAT pathway

One of the first mutations of the *Drosophila* JAK/STAT pathway components were the *outstretched* alleles which were later given the name *unpaired* (*upd*) due to segmentation defects of the larval cuticle in loss-of-function alleles (Müller 1930; Wieschaus et al. 1984). Later, sequence homology searches identified two further *upd*-like genes, namely *upd2* and *upd3*, which are located adjacent to *upd* on the X chromosome (Hombria and Brown 2002). UPD is the major signalling molecule activating the JAK/STAT pathway *in vivo*. Although *in vivo* analysis of *upd2* showed its role as a semi-redundant ligand, flies carrying its null mutation are fully rescued by *upd* whereas *upd2* can only partially rescue phenotypes which arise from loss-of-function of *upd* (Gilbert et al. 2005; Hombria et al. 2005). While UPD and UPD2 are both secreted ligands and have several putative N-glycosylation sites, they display different biophysical characteristics, with UPD strongly associating with the extra-cellular matrix (ECM), whereas UPD2 is more diffusible (Hombria et al. 2005). The limited diffusion ability of UPD may not only help it bind to the receptor but may also be necessary for correct patterning events *in vivo* (Zeidler et al. 1999). To date no clear homologues of *upd* have been identified beyond the *Drosophilae*, though it bears some similarity with the vertebrate Leptin (Boulay et al. 2003). *upd3* is expressed in haemocytes of adult flies in response to septic injury and signals to the fat body to produce anti-microbial peptides (Agaisse et al. 2003). The role of *upd3* in other tissues has not been established as yet.

The receptor DOMLESS (DOME) (Brown et al. 2001) - later also named MOM (Chen et al. 2002) - was given its name on the basis of the dome-shaped posterior spiracles in the larva of *dome* mutants. Its extracellular domain is most similar to the vertebrate cytokine class I receptor. *In vivo* studies have confirmed the necessity of DOME homo-dimerisation for JAK/STAT signalling, prior to binding of the ligand (Brown et al. 2003). This is in agreement with the cytokine class I receptor IL-2R and EPO-R, which also preassemble in the absence of cytokine (Damjanovich et al. 1997). In the *Drosophila* genome, homology searches of *dome* have also identified *CG14225* which bears structural similarities to the vertebrate IL-6Rα receptor (Hombria and Brown 2002). Although no functional studies on CG14225 have been published as yet, one can speculate the existence of CG14225/DOME hetero-dimer formation *in vivo*.

The *Drosophila* JAK kinase HOPSCOTCH (HOP) is most similar to mammalian JAK1 and 2 (Stahl et al. 1994). Mutations in *hop* give segmentation phenotypes similar to that of the *upd* mutation, but also reduce the size of imaginal discs, testis, ovaries and hindgut (Perrimon and Mahowald 1986). A temperature-sensitive dominant gain-of-function allele of *hop*, termed *hop^{Tuml}* (Hanratty and Dearolf 1993; Luo et al. 1995) results in over-proliferation of *Drosophila* blood cells, as well as differentiation of blood cells into lamellocytes. Both processes cause the formation of black melanotic tumours. Similarly, a V617F mutation in the pseudokinase domain of human JAK2 results in constitutive kinase activity. It is carried in over 80% of MPD (myeloproliferative disease) patients due to over-proliferation of erythrocytes (James et al. 2005; Staerk et al. 2005). HOP not only activates STAT92E but it has also been shown to physically interact and activate dRaf (Luo et al. 2002). *draf* is a STAT92E target gene and encodes a component of the Ras/Raf/Erk pathway which itself has been shown to play a major role in blood cell proliferation (Asha et al. 2003).

The *Drosophila* JAK/STAT pathway transcription factor STAT92E is most similar to mammalian STAT3 and 5. It was identified in a genetic mutagenesis approach where *stat92E* mutants produced similar segmentation defects to *upd* and *hop* mutants (Hou et al. 1996). By analogy to the mammalian field, it is generally assumed that STAT92E only translocates to the nucleus upon its phosphorylation by HOP. However, more recent *in vivo* data show that unphosphorylated STAT92E is also able to bind to DNA and thereby affect the surrounding chromatin structure (Shi et al. 2006; Shi et al. 2008). Interestingly, while STAT92E is able to exert both proliferative and anti-proliferative effects *in vivo*, in mammals this function is attributed to different STAT molecules (Mukherjee et al. 2005). In addition, an N-terminally truncated splice variant of *stat92E*, termed ΔNSTAT92E, functions as a negative regulator of JAK/STAT signalling (Henriksen et al. 2002).

STAT structure and binding sites

As mentioned earlier, STAT is the only JAK/STAT pathway component that is found in all model organisms (Table 3), suggesting its important role during animal development. The domains which make up STAT were first described by sequence comparisons and mutagenesis studies (Darnell 1997). Although no crystal structure of a whole STAT molecule has been resolved, the 3D core structure (amino acid ~130-712) of dimeric STAT1 or STAT3 bound to DNA has been solved (Figure 2A) (Chen et al. 1998).
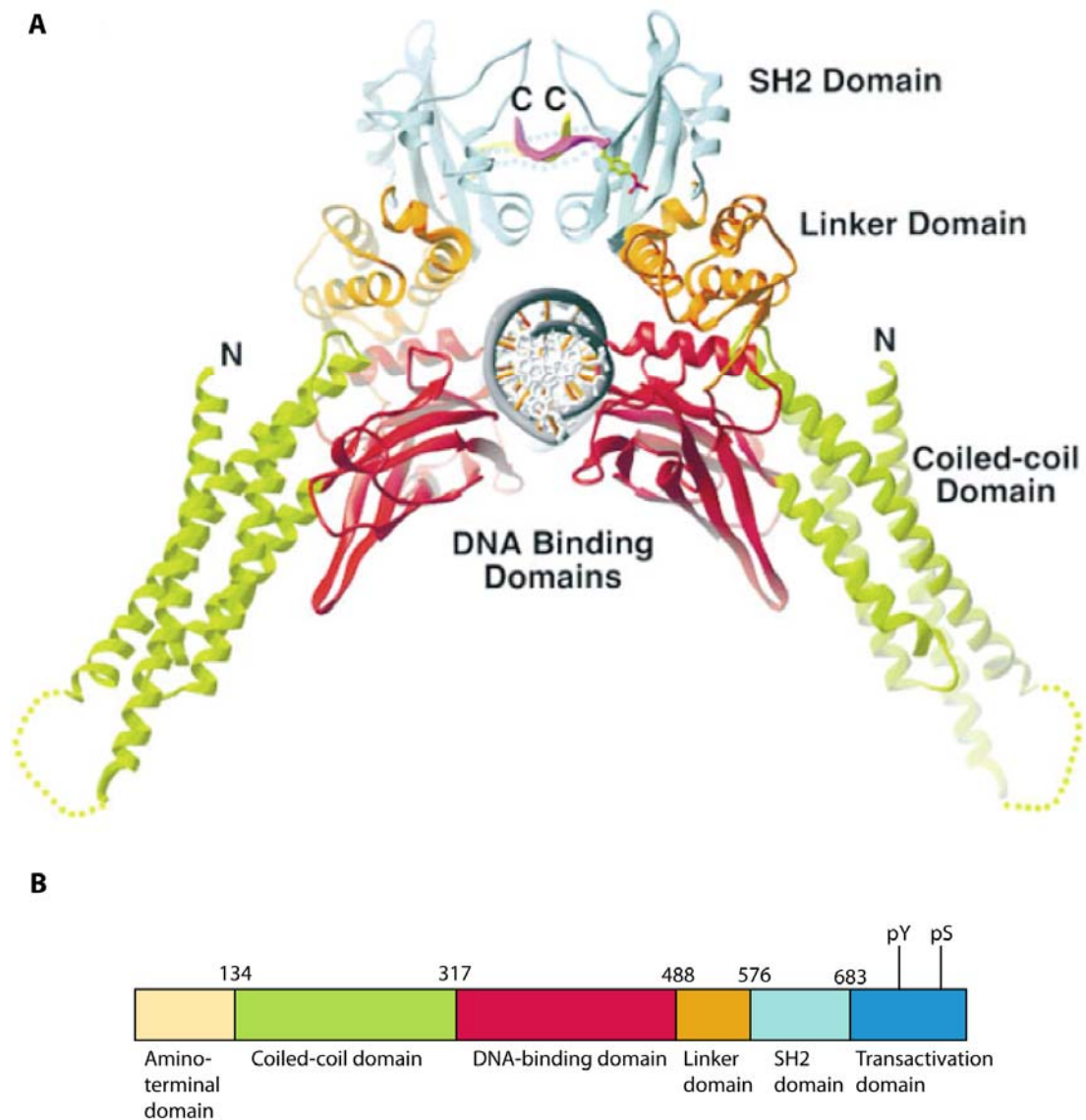
Figure 2.   Structure of a STAT1:1 DNA complex and STAT domains.
A) The core structure which is formed by amino acids ~ 130-712 show binding of STAT1 dimer to DNA. The structures of the amino-terminal domain and the carboxy-terminal transactivation domains are unknown. B) STAT domains. STAT, signal transducer and activator of transcription. SH2, Src-homology-2 domain. The critical tyrosine residue lies in the transactivation domain, as does the serine residue whose phosphorylation is often required for full STAT transcriptional activity. Adapted after Levy et al. 2002.

The crystal structure shows a reciprocal P-tyrosine-SH2 interaction of a parallel STAT dimer with the DNA cutting through the long axis of the two monomers. The structure includes an amino-terminal domain (N-terminus), a coiled-coiled domain followed by a DNA-binding domain, a SH2 domain, and finally, several transcriptional activation domains (TAD) at the carboxy-terminus (Figure 2B) (Levy and Darnell 2002).

The N-terminal domain is important for stable STAT dimerisation or tetramerisation, especially for STAT binding to promoters that contain tandem low-affinity DNA binding sites (Vinkemeier et al. 1996). In case of STAT4, this domain is further important for cytokine induced tyrosine phosphorylation. Furthermore, this domain facilitates a required reorientation of a STAT dimer from its active and phosphorylated "parallel" form to an "anti-parallel" dimer, which is then ready for inactivation by dephosphorylation (Mertens et al. 2006).

The coiled-coil domain represents an exposed surface that allows for numerous protein-protein interactions, either at the promoter site or when STAT is in solution (Horvath 2000).

The DNA binding domain structurally resembles the immunoglobulin-like DNA binding domain of NFκB and p53 (Chen et al. 1998). It is followed by a linker domain (LD) which is important for stability of DNA binding (Yang et al. 2002).

In all STATs the SH2 domain (which is necessary for recruitment to the receptor and dimerisation) is followed by several transcriptional activation domains (TADs) which carry the critical tyrosine residue. They vary greatly in their sequences between the different STATs. Apart from containing the absolutely required tyrosine phosphorylation site, serine phosphorylation in the TAD is also needed for full transcriptional activity of STAT1, STAT3 and STAT4 (Decker and Kovarik 2000). Transcriptional activity of all STATs depends on the TAD domains which bind co-activators. In fact, the increased transcriptional activity underlying serine phosphorylation is due to the selective recruitment of additional co-activators, including the minichromosome maintenance 5 (MCM5) protein (Zhang et al. 1998).

All STATs have been shown to form homo-dimers after being phosphorylated, with the exception of STAT2:2 that only forms weak dimers. Known STAT heterodimers include STAT1:2 and STAT1:3 as well as weak STAT2:3 dimers (Horvath 2000).

The first STAT DNA binding sites were found in genes activated by IFN-α, and termed interferon-stimulated response elements (ISREs), and were later renamed as GAS elements (Reich and Darnell 1989). Subsequent analysis using a pool of random oligonucleotides identified a palindromic sequence of TTCCNGGAA (N being any nucleotide) that is bound by STAT1 & 3. This sequence was similar to the GAS consensus site ($TT(N)_5AA$). More in depth analysis revealed the DNA binding specificity of different STAT proteins (Ehret et al. 2001; Brierley et al. 2006). With the core palindromic sequence being $TTC(N)_xGAA$, STAT6 preferably binds to N_4 spacing whereas all other STATs bind N_3 spacing, although N_2 sites can also be bound by STAT1 with low affinity.

In vitro and *in vivo* assays also showed the preference of the *Drosophila* STAT for N_3 sites (Figure 3A) and further validated its *in vivo* significance in the enhancer region of the pair rule gene *evenskipped* (*eve*) (Yan et al. 1996). Interestingly *Drosophila* as well as *C. elegans* STATs are also able to activate the IFN-γ N_3 GAS sites suggesting that other spacers (N_2 and N_4) evolved after the vertebrate STAT radiation. However, recent studies have also identified two low affinity N_4 STAT92E binding sites in the first intron region of the *Drosophila* JAK/STAT receptor *domeless* (Rivas et al. 2008). These sites are able to activate transcription in luciferase reporter assays as well as *in vivo* (Figure 3B). Nevertheless, STAT92E shows higher preference to the N_3 binding site, potentially explaining why only this site was originally identified by *in vitro* assays.

The non-canonical (JAK-independent) pathway

Recent data suggest that some classical descriptions of the JAK/STAT pathway have to be re-adjusted (Sehgal 2008). For example, non-phosphorylated STATs have generally been considered as inactive latent forms that reside in the cytoplasm as free monomers. More recent reports have, however, described non-phosphorylated STAT species as dimers and tetramers in the cytoplasm. Also, a new conformational model, suggesting a switch from an antiparallel to a parallel dimer upon tyrosine-phosphorylation, has been proposed. Also interesting is the fact that the N-terminal domain is needed for this dimer rearrangement back to its antiparallel form, allowing efficient dephosphorylation of the STAT molecule and thereby inactivation of the pathway. Furthermore it was long believed that only phosphorylated STAT proteins are able to enter the nucleus. This idea has now been challenged by results showing that unphosphorlyated STATs constitutively shuttle in and out of the nucleus.
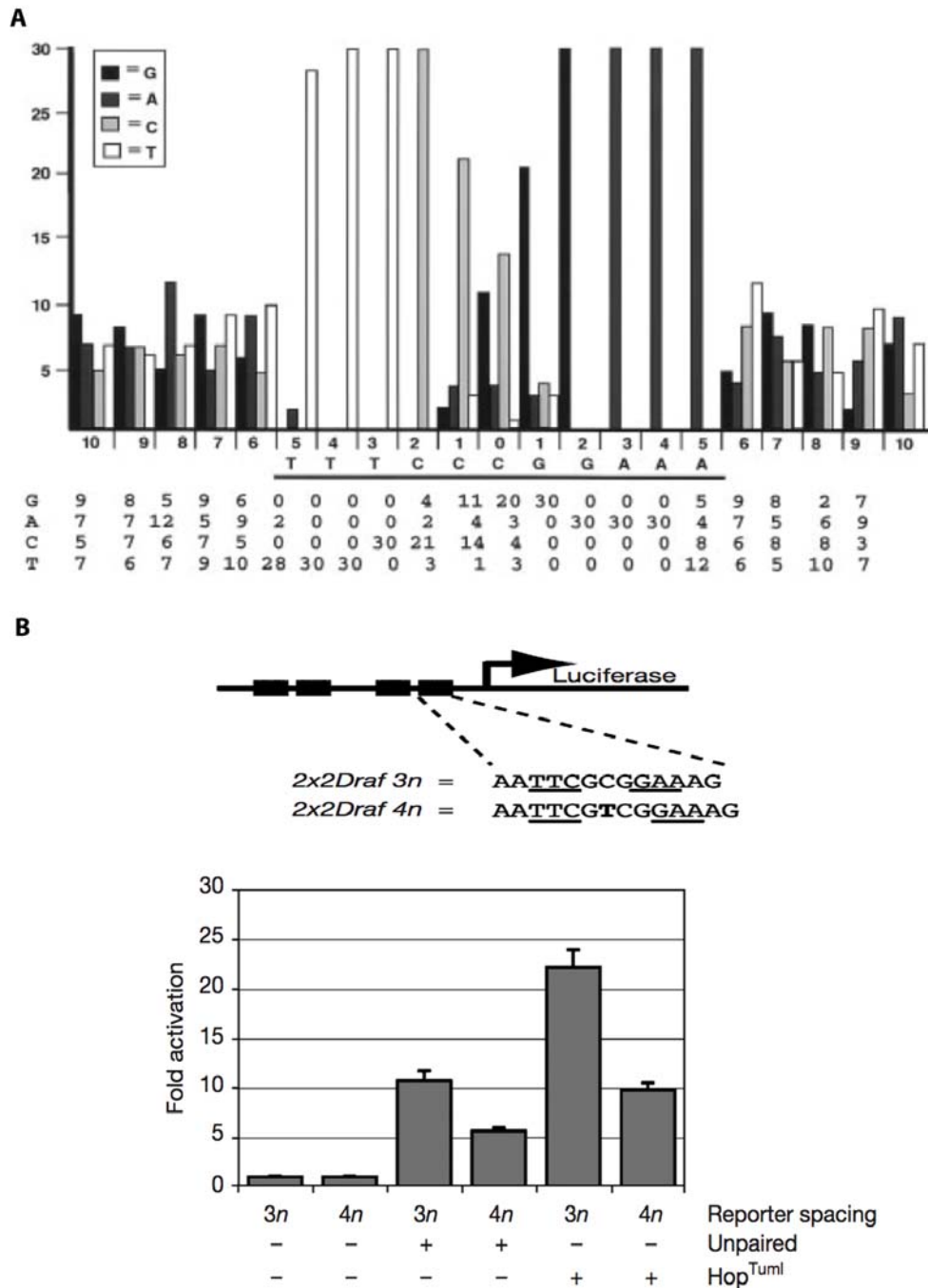
Figure 3.   STAT92E consensus binding sites.
A) random oligonucleotides were allowed to bind to nuclear extracts that contain the activated transcription factor STAT92E. The complex was precipitated and the oligonucleotides enriched by several repetition of this step. The DNA was then sequenced. The frequency of each nucleotide in each position is given below the corresponding bars. The most frequently selected sequences in the core region are displayed and underlined. From Yan et al. 1996. B) The first identified N_3 binding site $TTC(N)_3GAA$ and a newly discovered N_4 binding site $TTC(N)_4GAA$ were tagged to a luciferase reporter and transfected in Kc_{167} cells. JAK/STAT stimulation of Kc_{167} cells with the ligand Unpaired or the gain-of-function JAK kinas Hop^{Tuml} activates both N_3 and N_4 binding sites. From Rivas et al. 2008.

The differential behaviour of GFP-tagged STATs, particularly of the unphosphorylated form, in entering the nucleus is also becoming apparent, questioning the validity of earlier STAT translocation studies. In addition, some unphosphorylated STAT forms are also capable of initiating transcription, albeit their target genes being different from those activated by phosphorylated STATs.

The above data support recent findings in the *Drosophila* field, where the unphosphorylated STAT92E form has been shown to directly control heterochromatin stability (Shi et al. 2006; Shi et al. 2008). Its translocation to the nucleus does not require the activity of any of the upstream JAK/STAT components and thereby presents a form of non-canonical JAK/STAT signalling. Biochemical and microscopical studies suggest an essential physical interaction of unphosphorylated STAT92E with the heterochromatin component HP1 (heterochromatinprotein 1) (Figure 4). This interaction is required for maintaining heterochromatin stability. STAT92E phosphorylation disrupts this interaction and causes disassociation of STAT92E and HP1 from the DNA and thereby destabilisation of heterochromatin. Disruption of the heterochromatin state leads to a global change of gene transcription, with the effected genes not necessarily being direct STAT92E targets. The *Dictyostelium* STAT has been shown to act as a suppressor by binding to suppressor elements that regulate a set of differentiation genes (Kawata et al. 1997; Mohanty et al. 1999). Given this suppressor activity on gene expression we can assume an ancient role of STAT as a negative regulator of transcription.

Negative regulators of Drosophila JAK/STAT pathway

Homologues of the mammalian JAK/STAT regulators SOCS and PIAS are also functionally conserved in *Drosophila* (Hombria and Brown 2002; Arbouzova and Zeidler 2006). Although the role of PIAS as a negative regulator of the JAK/STAT pathway was first established with the *Drosophila* homolog dPIAS, it is not clear if these effects are caused by SUMOylation. Genetic studies have shown that dPIAS is allelic to the *su(var)2-10* locus. SU(VAR)2-10 is thought to be involved in heterchromatinisation with its mutant showing defects in chromatin segregation.

There are three *socs* genes in the *Drosophila* genome, *socs16D*, *socs36E* and *socs44A*, of which only *socs36E* is a direct target of the pathway and a negative regulator of JAK/STAT pathway (Callus and Mathey-Prevot 2002). SOCS36E is most similar to the mammalian SOCS5 with 29.7% identity. *socs16D* and *socs44A* are less well characterised.

The protein tyrosine phosphotase, PTP61F, is also transcriptionally induced by STAT92E and negatively regulates the JAK/STAT pathway (Baeg et al. 2005; Müller et al. 2005) (Figure 4A). JAK/STAT pathway target genes can also be regulated by a site-specific transcriptional repressor called KEN (Arbouzova et al. 2006). KEN recognises STAT92E-binding sites containing GAAA (i.e. sites with an additional A at the end of the STAT92E binding site TTC(N)GAA) and in concert with NURF (nucleosome remodelling factor) down-regulates a specific subset of pathway target genes (Figure 4B) (Kwon et al. 2008). NURF is a component of an ISWI-containing chromatin-remodelling complex which alters chromatin structure and gene transcription. *nurf* mutants develop melanotic tumours similar to those observed in *hop*^{Tuml} mutants (Badenhorst et al. 2002)
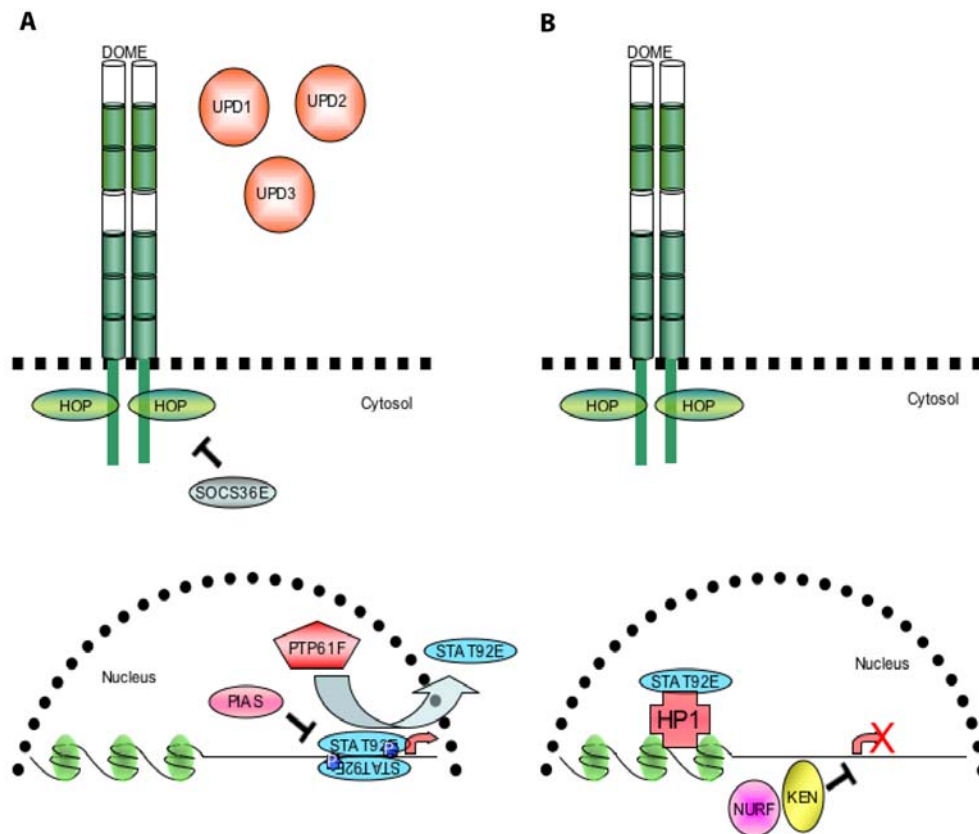
Figure 4.   The non-canonical JAK/STAT pathway.
A) In the canonical JAK/STAT signalling cascade the transcriptional activity of STAT92E is terminated by a negative feedback loop of the phosphotase PTP61F or by PIAS. The exact mechanisms of these regulations remain to be determined. B) Unphosphorylated STAT92E can also globally repress gene transcription via physical interaction with the heterochromatin protein HP1. This process of heterochromatinisation is not dependent on the activity of upstream JAK/STAT signalling components. The interaction of the JAK/STAT pathway regulator Ken with the nucleosome remodelling factor NURF also suppresses the transcription of a specific subset of pathway target genes. From Brown and Zeidler 2008.

Roles of *Drosophila* JAK/STAT pathway in Development and Disease

The developmental roles of the *Drosophila* JAK/STAT pathway can be broadly divided into embryonic, larval and adult stages. In the embryo the role of the JAK/STAT pathway starts in the sex determination process, where the ratio information of X chromosome(s) to autosomes is supported by X-linked signal elements (XSE) including the JAK/STAT pathway ligand UPD. All JAK/STAT elements also take part in regulating the segmentation process of the *Drosophila* embryo, which is distinct from that induced by the 'classical' maternal, gap, pair-rule and segment polarity genes. The JAK/STAT pathway also takes part in the development of the tracheal system which supplies oxygen to the fly's organs. Its target genes are crucial for tracheal cell movement and elongation, as well as cell invagination processes. The importance of the pathway in cellular movement and rearrangement also comes into play during the *Drosophila* gut development. Localised JAK/STAT signalling is required for polarised cell rearrangements within the hindgut tissue, enabling its elongation process during embryogenesis (Johansen et al. 2003).

In the larva the JAK/STAT signalling pathway mainly manifests itself in cellular proliferation processes during eye and wing imaginal disc, as well as blood cell development (haematopoiesis). In the eye imaginal discs, *upd* loss-of-function leads to a small eye phenotype, whereas the converse effect is observed during ectopic pathway activation (Bach et al. 2003; Mukherjee et al. 2006). The pathway plays a similar role in the developing wing imaginal disc, although the situation here is more complex. Whereas at early larval stages the requirement of STAT92E for cell proliferation has been demonstrated, at later stages, STAT92E was found to exert anti-proliferative activity, an effect that has been attributed to non-canonical JAK/STAT signalling (Mukherjee et al. 2005). The role of the JAK/STAT pathway in haematopoiesis will be introduced in more detail in the next section, due to its importance for the validation of STAT92E target genes.

Recently, the *Drosophila* model has become increasingly appreciated for studies of stem cells and the microenvironment (niche) that is required for their maintenance (Jones and Wagers 2008). In the *Drosophila* adult male, tightly packed cells forming the stem cell niche (also called hub cells) at the tip of the testis produce the ligand UPD. The ligand maintains stem cell identity of cells adjacent to the niche, whereas cells that are displaced from the signalling source after asymmetric division begin to differentiate (Kiger et al. 2001; Tulina and Matunis 2001). While in the male cell autonomous JAK/STAT signalling

is required for stem cell maintenance, in the female ovary the pathway is needed for maintenance of Dpp signalling which in turn is required for stem cell maintenance (Lopez-Onieva et al. 2008). UPD, the JAK kinase HOP and STAT92E are also required for epithelial cell migration in the developing ovary of the female adult. In light of the fact that STAT3 and 5 are constitutively active in a number of invasive human cancers, investigation of the role of JAK/STAT pathway in cell migration is important for future research. JAK/STAT signalling is also involved in multipotent stem cells in the malpighian tubules (*Drosophila* kidney) (Singh et al. 2007), as well as in maintenance of haematopoietic stem cells in the lymph gland, as highlighted in the next section.

Haematopoiesis

The *Drosophila* blood cell development (haematopoiesis) occurs in two waves, the embryonic and the larval haematopoiesis, which is paralleled by the vertebrate primitive and definitive haematopoiesis. As yet, the importance of the JAK/STAT pathway during embryonic haematopoiesis has not been established, but the relevance of other signalling pathways at this stage has been described elsewhere (Crozatier and Meister 2007). The role of the JAK/STAT pathway in *Drosophila* blood cells, termed haemocytes, is manifested in two ways, namely in larval haematopoiesis and in immunity. Haemocytes exert a cellular function by engulfing pathogens or apoptotic cell debris generated during metamorphosis, and they mediate signalling to the fat body in order to induce innate immune response in case of an infection. In healthy flies, 90-95% of haemocytes constitute the macrophage-like plasmatocytes. The rest of the haemocyte population is represented by crystal cells. Lamellocytes, as a third type of blood cell, are only induced in larval stages upon infestation by parasitic wasp eggs. Their role is to encapsulate particles that are too large for plasmatocytes to engulf.

The lymph gland is the only haematopoietic organ in larval *Drosophila*. Its development starts during embryogenesis and completes at late larval stages. At the onset of metamorphosis the lymph gland bursts, releasing its content into the hemolymph. The lymph gland consists of several lobes located along the dorsal vessel with the most anterior one being the crucial site of blood development (Figure 5A). This lobe is structured into three regions, the posterior signalling centre (PSC), the medullary zone (MZ) and the cortical zone (CZ). Localised UPD3 signalling from the PSC maintains the un-differentiated state of pro-haemocytes which reside in the MZ. The MZ is marked by

expression of the JAK/STAT pathway receptor *dome.* (Figure 5A). As haemocytes move towards the periphery, they begin to differentiate into diverse blood cell types which eventually reside in the CZ.

Inappropriate pathway activity caused by a temperature sensitive gain-of-function mutation of the *hop* kinase, named *hop^{Tuml}*, can result in significant haematopoietic phenotypes (Hanratty and Dearolf 1993; Luo et al. 1995; Luo et al. 2002). At low temperatures, HOP^{Tuml} is only moderately active resulting in overproliferation of plasmatocytes. At higher temperatures, the fully active HOP^{Tuml} induces lamellocyte formation, as well as increased haemocyte proliferation, which results in encapsulation of self-tissue and thereby formation of black melanotic tumours (Figure 5B) (Luo et al. 1995). This phenotype can be reduced by removal of one copy of the transcription factor *stat92E*. The genes activated by HOP^{Tuml} in the lymph gland to produce these phenotypes are however not known, and in *hop* mutants, lamellocyte differentiation is impaired. This is possibly due to the role of JAK/STAT activity in maintaining a pool of pro-haemocyte stem cells in the MZ that can be used for production of lamellocytes upon immune challenge (Krzemien et al. 2007).
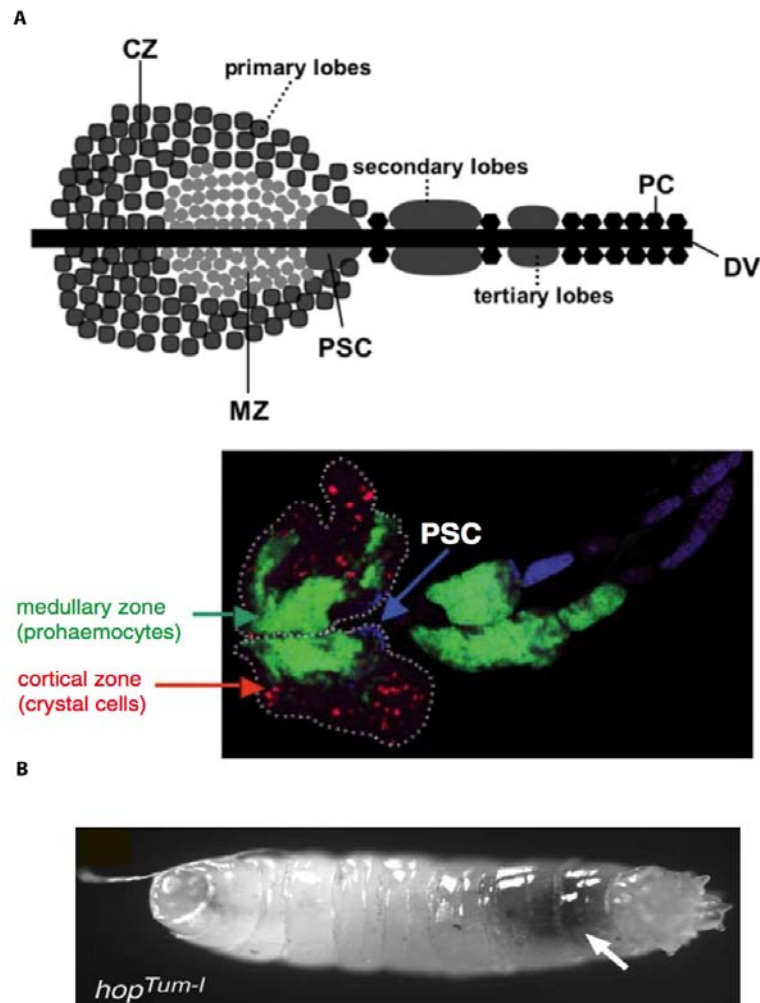
Figure 5. **The *Drosophila* haematopoietic organ and Hop^{Tuml}-induced tumour.**

A) The lymph gland is the only larval *Drosophila* haematopoietic organ and consists of several lobes and several pericardial cells (PC) which are situated along the dorsal vessel (DV). The most anterior (primary) lobe constitutes the posterior signalling center (PSC) which signals to the medullary zone (MZ) that contains pro-haemocytes. Differentiated haemocytes are situated in the cortical zone (CZ). The MZ can be visualised by the *dome-Gal4/UAS-GFP* reporter system. Secretion of one of the JAK/STAT ligands (probably UPD3) from the PSC maintains the pro-haemocyte identity of cells in the MZ. B) Constitutive JAK/STAT pathway activation by the gain-of-function *hop^{Tuml}* mutation induces haemocyte overproliferation which results in the formation of black melanotic tumours (arrow). From Bina and Zeidler 2009. Confocal image from Crozatier and Meister 2007.

Identification of JAK/STAT pathway target genes in *Drosophila*

While most *Drosophila* JAK/STAT pathway components have been known for some time, the identification of pathway regulators has been less straightforward. Two forward genetic interaction screens were performed in order to identify *in vivo* JAK/STAT pathway modifiers of the *GMR-Upd*-induced overgrown eye phenotype, which is the result of ectopic Upd expression in the eye (Bach et al. 2003; Mukherjee et al. 2006). However, genetic screens are inherently non-saturating. Since the completion of the *Drosophila* genome in 2000 (Adams et al. 2000), reverse genetic screens such as genome-wide RNAi knock-down assays allowed researchers to query the entire genome for pathway modulators. Two such RNAi screens aimed for the genome-wide identification of JAK/STAT pathway regulators, and complemented the list of regulators from previous genetic screens (Baeg et al. 2005; Müller et al. 2005). Although powerful in elucidating the functional aspects of the JAK/STAT pathway, none of the above-mentioned screens were designed to identify direct target genes, which may ultimately mediate JAK/STAT related diseases as well as developmental processes.

Numerous studies have been undertaken to identify and study mammalian STAT target genes (Murray 2007; Hennighausen and Robinson 2008). Of specific interest for this work are the STAT5 and STAT3 induced expression profiles, as these STATs are not only the closest homologues of the *Drosophila* STAT, but their constitutive activation is also observed in the majority of leukaemias and many solid tumours. The general role of cytokine signalling in leukaemia has been reviewed in (Van Etten 2007). It was long believed that the diverse outcome in STAT5a & b signalling is attributable to the differential expression of receptors. Recent data, however, suggest that specific post-translational modifications, along with the interaction of STAT molecules with cell specific co-activators as well as SOCS proteins, contribute to the different STAT5 induced target genes and phenotypes (Clark et al. 2005). Cross-talk between different STAT molecules further adds to the complexity of cytokine signalling. For example, loss of STAT5 signalling in liver leads to ectopic activation of STAT1 & 3 due to increased growth hormone levels (Cui et al. 2007). This would suggest that removal of one STAT member leads to inappropriate activation of another STAT member, which does not necessarily compensate the loss, but rather activates a different set of genes leading to different consequences. More intriguing is a study performed in macrophages where

despite the fact that IL-10 or IL-6 both activated the same JAK1-STAT3 pathway, the transcriptional output through their respective receptors (IL-10R and IL-6R) are very distinct (Murray 2007). Studying the mammalian JAK/STAT pathway is therefore very complex where, for example, the mechanism by which a receptor activates one of the seven STAT members is still not known. The genetic redundancies of mammalian JAK/STAT components, varying post-translational modifications of these components, differential interaction of STAT members with co-activators, as well as the diversity of transcriptional outputs, further add to the challenge of dissecting the pathway. In contrast, *Drosophila* only posses three JAK/STAT pathway ligands, a single receptor, a single JAK, and a single transcription factor. However, despite its simplicity, the JAK/STAT pathway itself, as well as its role in diverse biological processes, are highly conserved. The identification and analysis of pathway target genes using *Drosophila* as a model system should therefore be more straightforward.

Technologies aimed at quantifying the transcriptome can be broadly divided in hybridisation or sequence-based approaches, both of which enable simultaneous identification of gene expression. Hybridisation-based approaches include cDNA arrays, short oligonucleotide arrays (GeneChip®) (Hoheisel 2006) and the more recent tilling arrays which are used for detection of differentially spliced isoforms (Clark et al. 2002). More specific transcription factor-DNA interaction can further be studied by ChIP-on-chip analysis which is based on the co-immunoprecipitation of chromosomal DNA bound by a DNA-binding factor (Ren et al. 2000). In contrast to microarray, sequence-based approaches directly determine cDNA sequences. The techniques include the low-throughput method of Sanger sequencing of cDNA or EST libraries, tag-based methods like serial analysis of gene expression (SAGE) or cap-analysis of gene expression (CAGE), massive parallel signature sequencing (MPSS) and a more recent approach, called RNA-seq which uses the deep-sequencing technology (Wang et al. 2009). Despite the variety of sequence-based technologies which have gained popularity in other model organisms, the SAGE system is one of the few approaches that has been used in the *Drosophila* system. However, the SAGE approach not only relies on expensive Sanger sequencing technology, but a large proportion of the short tags cannot be mapped to the respective genome (Harbers and Carninci 2005).

Microarrays on the other hand have been more widely used in *Drosophila*. cDNA arrays use PCR products amplified from cDNA clones which are then spotted on various

surfaces/platforms. They can be manufactured in house or purchased from commercial suppliers. Oligonucleotide arrays on the other hand, which were first developed by Affymetrix (GeneChip®), provide a faster alternative, especially when researchers have access to the required workstations. GeneChip® arrays come with standardised protocols, and the availability of a variety of software packages provide more flexibility for data analysis. GeneChip® arrays are also the more commonly used platforms in the *Drosophila* field because of their excellent coverage of *Drosophila* transcripts, and the fact that data generated from different labs using GeneChip® arrays can be more easily compared, shared and analysed.

Given the above-mentioned advantages of GeneChip® arrays, I made use of this system in order to dissect down-stream target genes regulated by the *Drosophila* JAK/STAT pathway. JAK/STAT signalling was induced in *Drosophila* cell lines in a time course using the pathway ligands UPD and UPD2. The transcriptome of JAK/STAT activated cells was examined using GeneChip® arrays and the data obtained from the arrays was analysed using a novel probe level analysis approach. Focusing on transcript profiling results generated by the main pathway ligand UPD, the expression of a representative subset of genes was confirmed by real-time PCR. Furthermore, genes whose expression behaved similarly over time were grouped together using hierarchical clustering methods. Functional and promoter analysis were performed in order to dissect the underlying cause of the differential gene expression over time. The biological role of some putative STAT92E target genes was further validated, focusing on their importance in haematopoiesis. Tissue specific modulation of gene expression can be achieved using the yeast-derived GAL4/UAS system (Brand and Perrimon 1993). The *gal4* gene encodes a transcriptional activator that can be fused to a tissue specific promoter. GAL4 binds to the *UAS* (*upstream activating sequence*) sequence present in the promoter region controlling the expression of the gene of interest. In animals which carry both the GAL4 and the UAS constructs, the expression of the gene of interest can be modulated by GAL4 (Figure 6). Ultimately, dissecting the role of JAK/STAT pathway target genes in haematopoiesis will help to design better therapeutic strategies for human cancers.
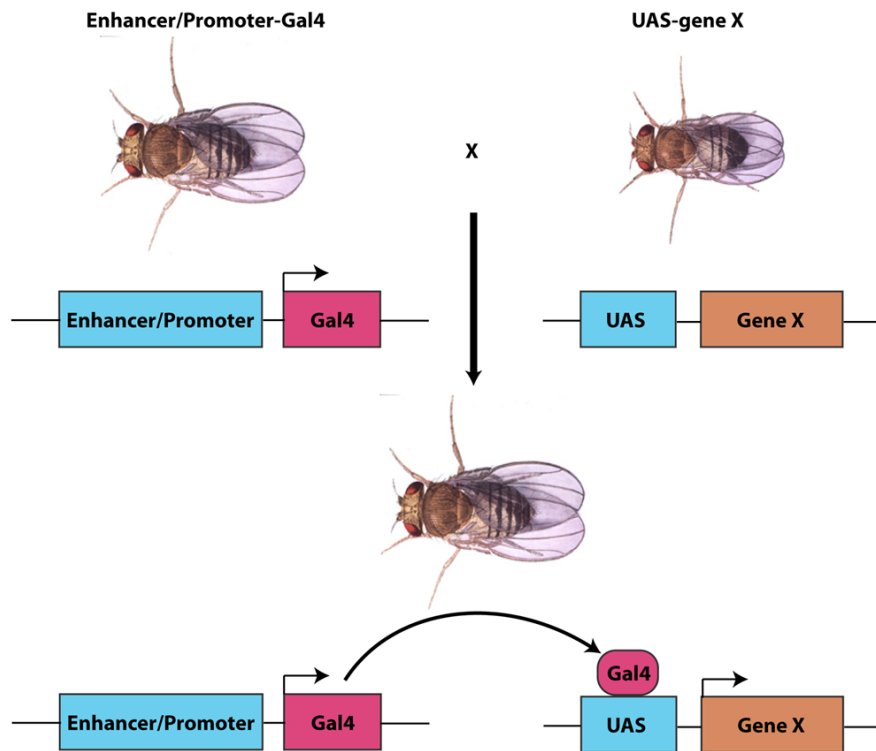
Figure 6.   The GAL4/UAS system.
The *UAS* sequence is inserted in front of a gene of interest (gene X) and is bound by the transcription factor GAL4 whose expression is under the control of a tissue-specific enhancer or promoter. In the progeny which carries both transgenes, gene X is expressed in tissues where the enhancer element is active.

MATERIALS AND METHODS

Analysis and manipulation of nucleic acids

Quantification of nucleic acid concentration

Undiluted nucleic acid was quantified using the NanoDrop spectophotometer (NanoDrop Technologies) according to their guidelines or subjected to electrophoresis using DNA or RNA ladders along side.

Polymerase chain reaction (PCR)

PCR reactions were carried out according to standard protocols (Ausubel et al. 1999). The annealing temperature and elongation time were adjusted based on the melting temperature of the primer and the length of the PCR product respectively. Taq polymerase (Roche) was used for all PCR reactions, except for cloning purposes where Pfu polymerase (Stratagene) was used to ensure high fidelity amplification. Typical PCR cycling parameters were: pre-denaturation 5min at 95˚C, 30 cycles: denaturation 30sec at 95˚C, annealing 30sec at appropriate temperature for a primer pair, elongation at 72˚C with the time depending on product length and the polymerase used, completion of synthesis for 10min at 72˚C. The PCR reaction was carried out in an MJ Research PTC-200 Petier Thermal Cycler. A typical PCR reaction contained:

1x PCR reaction buffer (supplied by manufacturer)

0.5mM dNTP

0.2μM primer

DNA from 1pg (plasmid) to 1μg (genomic DNA)

1U Taq or 2.5U Pfu polymerase

RNA isolation and cDNA synthesis

Typically $1\text{-}5 \times 10^6$ Kc_{167} cells were used to isolate RNA according to the TRIzol protocol (Invitrogen). For microarray purposes the TRIzol protocol was only carried out until completion of the phase separation. Thereafter, the RNA from the aqueous phase was isolated using the RNeasy® Mini Kit (Qiagen). 1μg of total RNA (and usually random hexamers) was subjected to cDNA synthesis using the ImProm-II™ Reverse Transcription

Kit (Promega), the Verso™ RT-PCR kit (ABgene®) or GeneChip® One-Cycle Target Labeling kit (Affymetrix).

Primer design for real time PCR

All primers were designed using the web-based program Primer3 (Rozen and Skaletsky 2000) and OligoAnalyzer 3.1 (Integrated DNA Technologies). All oligos were blasted against the *Drosophila* genome to ensure specific binding to target gene.

Particularly, for real-time PCR reactions, the primers were checked for absence of homo-, or hetero-dimer formation. In addition, primers were typically 18-20 nucleotides long and designed to yield of a product size of 100-250 base pairs, with an annealing temperature of 60˚C. See Table 4 for a list of all primers used in this work.

Real time PCR

SYBR Green RT-PCR was carried out as previously described (Vandesompele et al. 2002) using the SYBR® Green JumpStart™ Taq ReadyMix™ kit (Sigma). Briefly, total RNA was treated with RNase-free DNase according to the manufacturer's protocol (Promega). Treated RNA samples were purified using the RNeasy Mini Kit (Qiagen). 1μg RNA was used for first-strand cDNA synthesis using random hexamers and subsequently 1:10 diluted with dH_2O. 5μl of diluted cDNA was used in a 25μl RT-PCR amplification mixture according to Sigma's guidelines. The PCR reaction was carried out using the Biorad iCycler, data collected with MyIQ™ single color Real-Time PCR Detection system and analysed with iQ™5 Optical System Software. All primers were designed to yield a product size of 100-250 base pairs. Primer-dimer formation was ruled out by performing melt curve analysis (Vandesompele et al. 2002). All PCR efficiencies were above 80%. The housekeeping gene Rpl32 was used for normalisation as previously described (Dostert et al. 2005). Data analysis was performed using the delta-delta Ct method (Livak and Schmittgen 2001).

A common PCR program was used for all reactions:

Cycle 1:   95˚C   3min
Cycle2: (40X) 95˚C   30sec
                60˚C   30sec
                72˚C   30sec
                Perform data collection and real-time analysis

Cycle3:   95˚C   30sec

Cycle4:   55˚C   30sec

Cycle5: (40X) 55˚C   10sec

                Increase temperature after cycle 2 by 1˚C.

                Perform melt curve data collection and analysis.

TOPO cloning

PCR reaction was carried out using Taq polymerase (Roche) and the product was ligated into pCR®II-TOPO® vector or pCR®2.1-TOPO® vector (Invitrogen) according to the manufacturer's protocol. 2µl of the ligation product was added to One Shot Chemically competent cells (supplied by Invitrogen) and transformation was carried out as described below.

Restriction digest of DNA

All restriction digestions were performed using NEB enzymes. In order to linearise plasmids, 5µg DNA was subjected to 2h of digestion at 37˚C in a 50µl reaction as suggested by the manufacturer for complete digestion. The linearised plasmid was isolated by conventional ethanol precipitation including glycogen at -20˚C for 2h. The mixture was centrifuged and the precipitate was resuspended in 30-50µl H_2O.

DNA vectors

All DNA vectors used in this work have been published in Müller et al. 2005 and Hombria et al. 2005.

Transformation of bacterial cells

Chemically competent cells were thawed on ice, incubated with appropriate volumes of DNA for 30min on ice and heat-shocked for 30sec at 42˚C. 250µl SOC medium was added to cells, which were allowed to express the antibiotic for 30-60min at 37˚C on a shaker. Transformed cells were plated on LB agar plates containing the required antibiotics (and X-gal and IPTG for blue/white screening if needed) and incubated over night at 37˚C.

Genomic DNA isolation from a single fly

A single adult fly was mashed for 5-10sec with a pipette tip containing 50µl of squishing buffer. The mixture was incubated at 37˚C for 30min and the enzyme deactivated by heating to 95˚C for 2min. 1µl of the mixture was used for PCR analysis.

Squishing buffer: 10mM Tris.HCl pH8.2, 1mM EDTA, 25mM NaCl, 200µg/ml Proteinase K

Large-scale preparation of Drosophila genomic DNA

50 frozen flies were ground in 400µl DNA extraction buffer using a plastic Eppendorf homogeniser and further 400µl DNA extraction buffer were added to rinse the pestil. The homogenate was incubated at 65°C for 30min. 120µl of 8M K-OAc was added followed by 30min incubation on ice and 5min centrifugation at 13k RPM in a microcentrifuge. 750µl supernatant was transferred into a fresh tube and precipitated with 750µl 100% ethanol for 5min at room temperature. Following centrifugation the pellet was washed with 70% EtOH and resuspended in 400µl TE buffer. RNA was removed by adding RNAse A at a final concentration of 2µg/ml and incubating at 37°C for 30min. The enzyme was removed from the mixture by adding $1/10^{th}$ vol/vol of StrataClean (Stratagene) and centrifuging it at 13k RPM. The supernatant was carefully transferred to a fresh tube and precipitated by conventional techniques.

DNA extraction buffer: 0.1M NaCl, 0.2M sucrose, 0.1M Tris.HCl pH9, 50mM EDTA, 0.5% SDS

Drosophila cell culture

Cell maintenance and storage

Drosophila Kc_{167} cells (Echalier and Ohanessian 1970) were grown at 25°C, at subconfluent densities, in Schneider's medium (Invitrogen) supplemented with 10% heat inactivated fetal calf serum (Sigma) and penicillin-streptavidin (0.1mg/ml).

For permenant storage, $8x10^7$ Kc_{167} cells were centrifuged for 5min at 2000rpm (Hermile) and resuspended in freezing Schneider's medium supplemented with 10% DMSO at a final concentration of $2x10^7$ cells/ml. 0.5ml aliquots in sterile cryovials were placed in a Dewar flask and placed at -80°C for three days allowing cells to freeze slowly. The frozen vials were transferred to liquid nitrogen for long-term storage. When required, cells were thawed at room temperature, transferred into a flask and the DMSO medium was replaced with Schneider's medium as soon as possible.

List of primers

Name	Forward/Reverse	Target	Sequence (5' -> 3')
AttD_F_2	F	AttD	GGAGTAAGGGTCGGTGATGA
AttD_R_2	R	AttD	CGATCGGCTATGACTGTGAA
Baz_F_1	F	baz	GGCACCTATCAGCGGAATAA
Baz_R_1	R	baz	AAACTGGGCATTAGCACTGG
CG10262_F_2	F	CG10262	ATGGACAACTCCCACGTTTC
CG10262_R_2	R	CG10262	CTGTCAACGGCCTTAATGGT
CG13559_F_1	F	CG13559	ATCTGTCCAATGTGCCATGA
CG13559_R_1	R	CG13559	TGAAACAGTCGAGGATGCAG
CG13912_F_1	F	CG13912	AAGCCGGAATACGATCACAG
CG13912_R_1	R	CG13912	GATACCGGAAACGCTGACTC
CG4804_F_1	F	CG4804	TTCTTCAAAAGCTGCGCATA
CG4804_R_1	R	CG4804	CAGCTCGGAAGAATTTCTGG
CG6014_F_1	F	CG6014	AAAACGATCGCAGGTACCAC
CG6014_R_1	R	CG6014	CTGTGCTGGCTGGTAGTTCA
CG9317_F_1	F	CG9317	CCCCAATATGCGCTTAAAGA
CG9317_R_1	R	CG9317	CACAACAGCCGACAGAAAGA
ptr_F_1	F	ptr	TTGGACGAGCTGACTCCTTT
ptr_R_1	R	ptr	TCGGCCAGACTGTTCTTTCT
Rpl32_F_paper*	F	Rpl32	GACGCTTCAAGGGACAGTATCTG
Rpl32_R_paper*	R	Rpl32	AAACGCGGTTCTGCATGAG
Socs36e_F_1	F	socs36E	AAGTGCACACTGTCGAATGG
Socs36e_R_1	R	socs36E	TTCCCCGTTTTCACGTTATC
TotA_F_1	F	TotA	TGAGGAACGGGAGAGTATCG
TotA_R_1	R	TotA	GCCCTTCACACCTGGAGATA
alphaTub_forward	F	alphaTub84B	GAATTTTCCTTGTCGCGTGT
alphaTub_reverse	R	alphaTub84B	GGTCACCAGAGGGAAGTGAA

* Dostert et al. 2005

Table 4.   List of primers

Transfections and generation of conditioned media

For Luciferase reporter assays, Kc_{167} cells were transfected in a 6-well plate with 0.5µg *6x2DrafLuc* and 0.6µg *pAct-RL* (Müller et al. 2005) according to the Effectene (Qiagen) protocol. Cells were incubated 4-7h or overnight at 25°C. Reporter cells were diluted with Schneider's medium such that 50µl, containing 30-50 000 reporter cells, could be split into a 96-well plate. Luciferase read-out was performed 3-4 days later as described below.

For generation of conditioned media $5x10^6$ Kc_{167} cells were transfected in a 6-well plate with 2µg of *pAC5.1*, *pAC5.1-Upd-GFP* (Müller et al. 2005) or *pAC5.1-Upd2-GFP* (Hombria et al. 2005) plasmid, 20µl Enhancer, 16µl Effectene and incubated 2-3 days at 25°C. Transfected cells were expanded into 100mm dishes and further into 100ml spinner flask whenever a minimum concentration of $1x10^6$ cells/ml was achieved. Heparin was added at an end-concentration of 50µg/ml. The conditioned medium was sterile filtered through 0.2µm filters (Nalgen), aliquoted, and stored at -80°C.

Dual luciferase assay

The Dual-Luciferase® Reporter Assay (Promega) was used to measure Firefly and *Renilla* luciferase activity on Mithras LB 940 Luminometer (Berthold Technologies). After removing the Schneider's medium, cells were lysed in a 96-well plate in 20µl of 1X Passive Lysis Buffer (PLB). 15µl PLB lysate were transferred into a white 96-well plate (CELLSTAR®) and 50µl LARII was added. Depending on the desired detection level Firefly activity was measured for 0.1 to 3sec. 50µl of Stop&Glo® was then added to the same well in order to measure *Renilla* Luciferase activity for 0.1 to 3sec. Relative reporter activity was estimated as the ratio between the Firefly Luciferase and *Renilla* Luciferase read out.

Transcript profiling using GeneChip® Drosophila Genome 2.0 Array

$5x10^6$ Kc_{167} cells in 3ml serum-free Scheider's medium were divided in a 6-well plate 16h prior to JAK/STAT stimulation. For all three time points and conditioned media treatment, two biological samples were prepared, making a total of 18 samples (3 time points x 3 treatments x 2 biological replicates). 1.5ml serum-free medium was removed followed by addition of 1.5ml UPD-, UPD2- or Mock-conditioned media. 30min later, all media was removed and replaced by 2.5ml serum-free medium. Total RNA was isolated 2h, 4h or 10h after first addition of conditioned media according to the protocol suggested by

Affymetrix. The RNA was labelled according to GeneChip® One-Cycle Target Labeling manual and hybridised to GeneChip® *Drosophila* Genome 2.0 Arrays. The Fluidics Station 450/250 was used to wash and stain the probe arrays which were subsequently scanned using the GeneChip® Scanner 3000. Prior to conducting array analysis, the quality of the array image (.dat files) was assessed following the guidelines of the GeneChip® Expression Analysis manual.

Drosophila genetics

Fly stocks and husbandry

Flies were maintained on complex cornflour-soyflour-molasse medium at 18°C (unless otherwise specified) with 50-70% humidity and an approximately 12h / 12h light / dark cycle. A list of flies used in this study is in Table 5.

Ectopic expression using GAL4/UAS system

The GAL4/UAS system is based on using the yeast transcription factor GAL4 and its recognition site *uas* (Brand and Perrimon 1993). For this study a GAL4 driver line was used in which the *gal4* gene is inserted downstream of a tissue specific promoter that is responsible for induction of the JAK/STAT receptor *dome* (Bourbon et al. 2002). hop^{Tuml},*dome-Gal4* recombinant virgins (see Figure 19B) were crossed to UAS-transgenic males listed in Table 5. hop^{Tuml} is an oncogenic allele which induces haematopoietic tumours. This are manifested as black melanotic masses in the abdomen (Figure 19C). For leukaemia and haemocyte count assays, crosses were set up at 20°C (unless otherwise specified) for 24h and the embryos were kept at this temperature for further 24h before being transferred to 29°C for full *GAL4/UAS* and hop^{Tuml} activation. For each cross up to 200 progenies were scored for modulation of the tumour size as described in the results section. For statistical analysis unpaired t-test was performed.

Haemocyte counts

Wandering third instar larvae of the appropriate genotype were washed in water, dried and bled by tearing the larva with two pairs of forceps in a 15µl drop of serum-free Schneider's medium placed on a hemocytometer for immediate counting. For each cross 6-18 larvae were used. For statistical analysis unpaired t-test was performed.

List of fly stocks

Name	Genotype	Transformant ID	Reference
wildtype lines			
OreR	wildtype		Lindsley and Zimm 1992
white	w^{1118}		Lindsley and Zimm 1992
Gal4 driver and UAS lines			
UAS-baz	w, P{UAS-baz.K}		Kuchinke et al. 1998
UAS-CG10746	w, P{UAS-CG10764, w^+}	15515/15516	Dietzl et al. 2007
UAS-CG13559	w, P{UAS-CG13559, w^+}	6643/6644	Dietzl et al. 2007
UAS-CG15211	w, P{UAS-CG15211, w^+}	35704	Dietzl et al. 2007
UAS-CG15678	w, P{UAS-CG15678, w^+}	23784	Dietzl et al. 2007
UAS-CG1572	w, P{UAS-CG1572, w^+}	23220/23221	Dietzl et al. 2007
UAS-CG31158	w, P{UAS-CG31158, w^+}	42321	Dietzl et al. 2007
UAS-CG3829	w, P{UAS-CG3829, w^+}	42870/42872	Dietzl et al. 2007
UAS-CG4080	w, P{UAS-CG4080, w^+}	9026	Dietzl et al. 2007
UAS-CG4804	w, P{UAS-CG4804, w^+}	38018	Dietzl et al. 2007
UAS-CG6014	w, P{UAS-CG6014, w^+}	31067	Dietzl et al. 2007
UAS-Ga73B	w, P{UAS-Ga73B, w^+}	17054/17056	Dietzl et al. 2007
UAS-Idgf1	w, P{UAS-Idgf1, w^+}	12414/12416	Dietzl et al. 2007
UAS-l(2)gl	w, P{UAS-l(2)gl, w^+}	51247/51249	Dietzl et al. 2007
UAS-mfas	w, P{UAS-mfas, w^+}	37888/37889	Dietzl et al. 2007
UAS-net	w, P{UAS-net, w^+}	41069	Dietzl et al. 2007
UAS-pxb	w, P{UAS-pxb, w^+}	24932/24933	Dietzl et al. 2007
UAS-Rh4	w, P{UAS-Rh4, w^+}	46919/46920	Dietzl et al. 2007
UAS-stat92E	w, P{UAS-stat92E, w^+}	43866/43867	Dietzl et al. 2007
UAS-TotA	w, P{UAS-TotA, w^+}	14415/14416	Dietzl et al. 2007
UAS-wnt4	w, P{UAS-wnt4, w^+}	38010	Dietzl et al. 2007
dome-Gal4	P{GawB}dome^{PG5}		Bourbon et al 2002
hop^{Tuml}, dome-Gal4	y,w,hop^{Tuml},P{GawB}dome^{PG5}/FM6		this study
hop^{Tuml}	y,w,hop^{Tuml}/FM7		Hanratty and Dearolf 1993

Table 5.   **List of fly stocks**

Embryo collection and fixation

Adult flies were transferred in cages covered with apple juice agar plate and baker's yeast and allowed to lay eggs for 20-24hrs at 25˚C. Embryos were transferred to a sieve and thoroughly washed with dH_2O to remove yeast and dechorionated in 50% bleach for 3-5min. The embryos were again washed with dH_2O and transferred to scintillation vials containing 2ml fixative and 8ml heptane. Embryos were fixed for 20min with vigorous shaking. The fixative was removed from the bottom layer and 8ml MeOH added to the vial followed by intensive shaking for 30sec to remove the vitelline membrane. The upper and interphase phase were removed and sunk embryos washed several times with MeOH and stored in MeOH at -20˚C for long-term storage.

Fixative: 10% paraformaldehyde in PBS, 50mM EGTA, pH 7 (adjusted with 1M NaOH). Fixative was aliquoted and stored at -80˚C.

Computational analyses

PUMA

Raw microarray data was obtained after scanning the GeneChip® arrays and analysed (in collaboration with Dr. Marta Milo in Sheffield University) according to the manual of the open source PUMA package (Liu et al. 2006) that has been released through Bioconductor (Gentleman et al. 2004). The R software environment is used to perform PUMA analysis (R-Development-Core-Team 2004).

Differential fold change cut off

The filtering of genes using intensity-dependent Z-scores has been described previously (Quackenbush 2002; Yang et al. 2002). Briefly, PUMA analysis calculates the fold changes (Log2 ratios) of signal intensities (calculated as an average from the replicates) between UPD (or UPD2) arrays compared to Mock arrays at each time point. In addition, I calculated the mean signal intensities of the arrays to be compared. These mean signal intensities were then sectioned into groups using a sliding window size of 0.5 (Figure 12). For each group the mean and standard deviation of the distribution of Log2 ratios were calculated, which enabled me to calculate an intensity dependent Z-score that can be associated to each Log2 ratio. A Z-score measures the number of standard deviations a particular data point (Log2 ratio) is away from the mean. Log2 ratios (and their associated

genes) that are associated with a Z-value of more than 2 (for up-regulated genes) or less than –2 (for down-regulated genes) were filtered (Figure 12).

Hierarchical clustering

Hierarchical clustering was performed as previously described (Eisen et al. 1998) using the average-linkage method and the data was visualised using the TreeView tool.

Functional analyses of UPD regulated genes

For biological interpretation of UPD-regulated genes, genes were analysed using the web-based GoMiner[TM] (Zeeberg et al. 2003) and the GOstat (Beissbarth and Speed 2004) tool. For GoMiner analysis, genes were examined using the gene ontology (GO) evidence level 2, and the fly base (fb) data source. All other settings were kept as default. For GOstat analysis all settings were kept as default except for the selection of the fb data source. To determine putative *Drosophila* protein homologues in other species, BLASTP searches were performed against protein databases from *Homo sapiens* (RefSeq release 31) as well as *Mus musculus* (RefSeq release 31).

Promoter analysis

The web-based program RSAT (Thomas-Chollier et al. 2008) was used in order to retrieve the 3Kb upstream region of *Drosophila* genes and the DNA-pattern tool used to search for STAT92E consensus binding sites (TTCNNNGAA or TTCNNNNGAA) (Figure 3A). The number of binding sites for individual genes was counted manually. For statistical analysis, unpaired t-test was performed using the GraphPad Prism software.

RESULTS

Transcript profiling of JAK/STAT activated *Drosophila* cells

Introduction

The JAK/STAT pathway is a fundamental developmental signalling pathway which we now know to be relevant to cancer development and progression. It is therefore not surprising that numerous attempts have been made to identify the downstream target genes which lead to disease. As outlined in the introduction, over 40 ligands have been shown to activate the mammalian JAK/STAT pathway, but cross-activation of alternative pathways and transcription factors induced by those ligands are possible (van Boxel-Dezaire et al. 2006). This drawback led to the construction of constitutively active STATs, which do not require any ligand for their activation. However, not only does this system prevent time course analysis of pathway activation, but there is also inconsistency in the overlap of STAT target genes in different mammalian cell lines. Furthermore, it is unclear if in the *in vivo* situation these gain-of-function (GOF) STAT molecules can mimic the same response as their normal STAT counterparts would do, when they are activated. For example, whereas a specific GOF mutation in the mammalian STAT5A resulted in its constitutive activity in tissue culture experiments, the corresponding amino acid mutation in the *Drosophila* STAT92E was shown to have a dominant negative effect in tissue culture and *in vivo* (Karsten et al. 2006). It is therefore advantageous to not only study signalling pathways, but also validate results obtained in the mammalian system in less complex systems, especially as *in vivo* studies can be carried out quicker.

The dissection of JAK/STAT target genes in *Drosophila* cell lines should prove to be more straightforward, for the following reasons: 1) addition of the ligands UPD, UPD2 or UPD3, activate a single transcription factor STAT92E. Given that all *Drosophila* JAK/STAT pathway components, except the ligands, are non-redundant, we would expect a common transcriptional output after pathway activation. However, this hypothesis has never been tested before in the *Drosophila* field, which led us to the idea of activating the JAK/STAT pathway in *Drosophila* cell lines by adding different ligands and examining the transcriptional outcome in a time course dependent manner. 2) from the three widely used *Drosophila* cell lines, namely S2R+, S2 and Kc_{167}, the latter have been shown to express all JAK/STAT pathway components and would therefore need little intervention in

order to activate the pathway (Müller et al. 2005). 3) Kc$_{167}$ cells have also been validated as a viable model for genome-wide RNAi screening in order to find JAK/STAT pathway modulators (Müller et al. 2005).

Generation of conditioned media

upd and *upd2* encode semi-redundant ligands which are expressed in overlapping patterns during *Drosophila* embryogenesis (Gilbert et al. 2005; Hombria et al. 2005). *upd3* however, is only expressed at later embryonic stages in the gonad and its larval role seems to be restricted to immune signalling (Agaisse et al. 2003). At the time of designing the screen, only the full cDNA sequences of *upd* and *upd2* but not *upd3* were known. Therefore, tissue culture experiments were only performed with the two available ligands. While both UPD and UPD2 are secreted, UPD strongly associates with the extracellular matrix (ECM) and in contrast to UPD2, needs addition of heparin in order to diffuse into and condition the medium (Harrison et al. 1998). Conditioned media were generated from *pAct-UpdGFP* (or *pAct-Upd2GFP*) transfected Kc$_{167}$ cells grown in multi-well plates (Figure 7A) and tested for their ability to activate luciferase-based reporter cells (Figure 7A-C). A *6x2xDrafLuc* reporter was previously generated and contains 12 STAT92E binding sites which are responsive to pathway activation by UPD and UPD2 (Figure 7B) (Hombria et al. 2005; Müller et al. 2005). In order to keep technical variation low during transcript profiling experiments, large batches of conditioned media were generated by growing *upd* or *upd2* transfected Kc$_{167}$ cells in 50ml suspension using spinner flasks. This media was used for all subsequent experiments. UPD- and UPD2-Conditioned media generated by this means showed comparable pathway activation in reporter cells (Figure 7D) which enabled me to use the media in equal amounts in order to stimulate JAK/STAT signalling in Kc$_{167}$ cells.

Design of time course series

In order to identify the dynamics of the transcriptional response in Kc$_{167}$ cells after JAK/STAT stimulation, the expression pattern of *socs36E* was analysed. *socs36E* is a direct STAT92E target gene and a negative regulator of the pathway and its role in Kc$_{167}$ cells has been demonstrated previously (Callus and Mathey-Prevot 2002; Baeg et al. 2005). The expression pattern of *socs36E* was analysed by adding UPD- or UPD2-conditioned media to Kc$_{167}$ cells for 30min, which were then replaced by serum free media. The RNA of JAK/STAT activated cells was isolated at different time points after initial
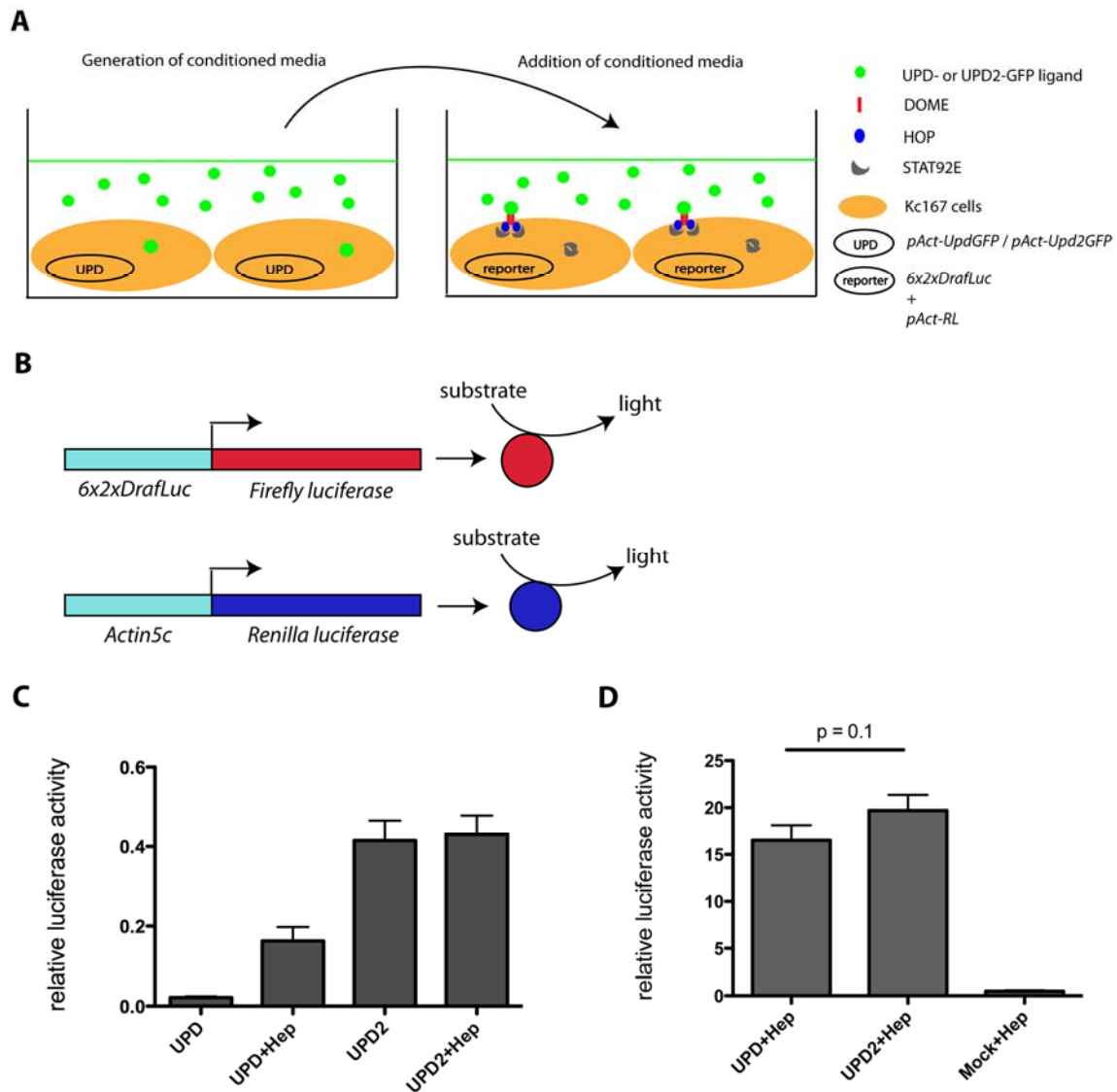
Figure 7. **Generation and activity of conditioned media.**
A) Generation of conditioned media by transfection of Kc$_{167}$ cells with *Upd-*, *Upd2-GFP* or just "empty" (Mock) expression plasmid. Both ligands are secreted into the medium, which is then used to activate the JAK/STAT pathway in reporter Kc$_{167}$ cells. B) Luciferase reporter system. Firefly luciferase expression is under the control of 12 STAT92E binding sites (*6x2Draf*) which are responsive to pathway activation by UPD or UPD2. Constitutively expressed Renilla luciferase acts as a control for cell viability. C) The addition of heparin to UPD conditioned media allows the ligand to diffuse more readily into the medium resulting in higher JAK/STAT activation in reporter cells. D) UPD- or UPD2-conditioned media generated from Kc$_{167}$ cells grown in suspension induce comparable JAK/STAT activity in reporter cells. These media were used for transcript profiling experiments.

addition of conditioned media and subjected to reverse transcription and quantitative PCR analysis (Figure 8A). Although *socs36E* transcript levels were already detectable at 0h, there was a clear RNA enrichment after 2h, which decreased 10h after pathway activation. These results could also be reproduced by real-time PCR analysis (Figure 8B).

For transcript profiling assays, serum-starved Kc_{167} cells were pulse-activated with UPD-, UPD2- or Mock-conditioned media for 30min. The medium was then replaced by serum-free media, and the RNA isolated 2h, 4h or 10h after initial addition of conditioned media. For each time point and treatment, biological duplicates were generated resulting in a total of 18 samples that were prepared for hybridisation to GeneChip® *Drosophila* Genome 2.0 arrays (Figure 8C). The arrays were scanned using the Affymetrix workstation and the raw data was analysed as outlined below. Figure 9 summarises the data analyses performed, starting from the raw microarray data to downstream *in silico* analyses.

Microarray data analysis

Program to analyse transcript profiling data

One of the unique properties of the Affymetrix platforms (GeneChip® arrays) is that each gene is represented by approximately 14 pairs of oligonucleotides, composed of a perfect match (PM) and a mismatch (MM) probe. While a PM probe has the correct complementary sequence to its corresponding transcript, the MM probe is the same, except for one base in the middle which is changed to the complementary one. MM probes serve to measure unspecific binding of targets and thus serve to estimate background intensities. The GeneChip® *Drosophila* Genome 2.0 Array consists of over 500,000 data points (oligonucleotides) which measure the expression of 18,500 transcripts and variants. Since such microarrays have been available, a lot of work has been invested into optimising the data analysis of the vast amount of information generated by them.

The information stored in each probe has proven to be extremely precious, especially for fold change estimation of low abundance target genes. Moreover, one of the greatest obstacles in DNA microarray analysis is how to summarise the intensities of the 14 probe pairs, representing the abundance of one gene transcript, into a single estimate. This is not a trivial task, as each probe pair (PM & MM) has a different sequence and therefore different hybridisation constants, hence the signals from different probes are not necessarily the same. To further add to the complexity, several labs have shown specific
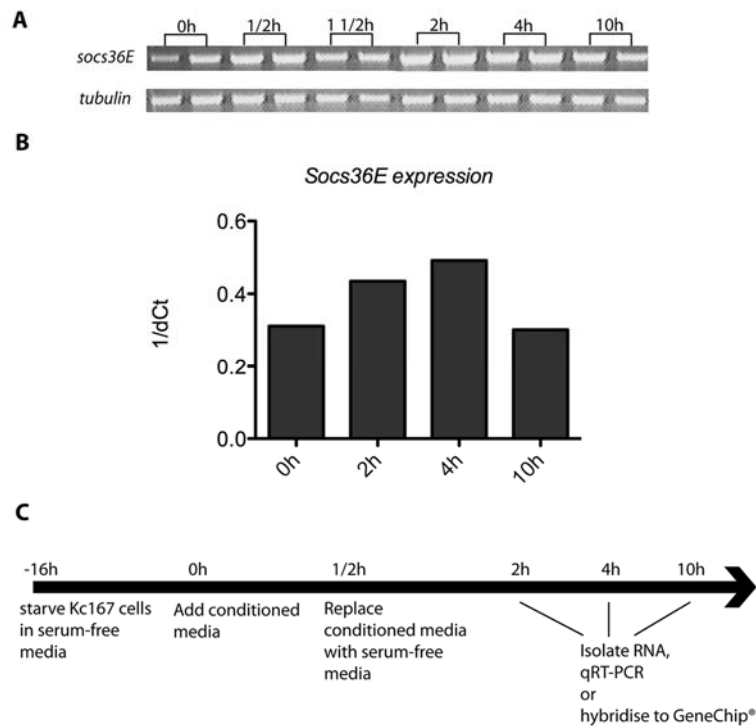
Figure 8.   Expressional levels of *socs36E* and flowchart of transcript profiling.

A) semi-quantitative RT-PCR analysis of *socs36E* and *tubulin* (housekeeping control gene) expression after serum-starved Kc_{167} cells have been pulse activated with UPD- or UPD2-conditioned media for 30min. Left and right band at each time point correspond to UPD and UPD2 stimulation respectively. Background transcript level of *socs36E* is already detected in un-stimulated cells (0h). B) Real time PCR quantification of *socs36E* expression under identical conditions as in A. C) Flowchart of transcript profiling of JAK/STAT-stimulated cells
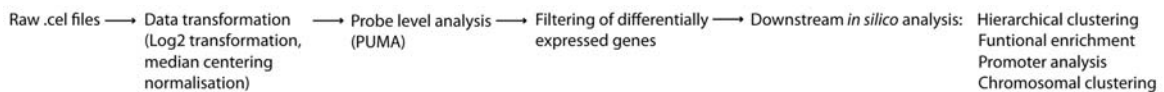
Figure 9.   Summary of the analysis of GeneChip® *Drosophila* Genome 2.0 arrays.

Raw .cel files were obtained after scanning the chips and were subjected to analysis using the PUMA program. First, signal intensities were Log2 transformed and normalised by median centering. Probe set analysis uses the information of all probes, which represent a particular gene, in order to estimate the enrichment of the gene and assigns a confidence level to this enrichment. Z-score analysis was used in order to filter differentially expressed genes of UPD- or UPD2-treated cells. These genes were then examined for their temporal expression pattern (hierarchical clustering) and their functional enrichment. Furthermore the promoter region of differentially expressed genes were examined for transcription factor binding sites. The proximity of co-expressed genes on chromosomal level was also determined (Chromosomal clustering).

target binding to MM probes (Naef et al. 2002) and in addition, both PM and MM intensities increase with target concentration (Irizarry et al. 2003). Most widely used methods summarise probe intensities into a single data point, discarding useful information about measurement error. Recently, probabilistic methods, such as PUMA, have been developed which estimate the gene expression levels by integrating the information of each probe (Liu et al. 2006). PUMA is able to estimate the confidence level associated with each measurement, and propagates this information into further downstream analysis. Given the statistical power of PUMA and the opportunity of close collaboration with the bioinformaticians who developed this algorithm, I decided to use this method in order to analyse the raw microarray data.

Data normalisation

In order to identify differential gene expression patterns, the signals of a particular gene within the "Upd" array and the "Mock" array need to be compared resulting in a signal ratio value. However, before such comparison is made, one has to perform a number of transformations and adjustments to the raw data (Quackenbush 2002). To demonstrate the necessity of such adjustments, a two-fold enrichment results in a signal ratio of 2, whereas a down-regulation by the same factor results in a ratio of 0.5. Using absolute numbers in order to calculate fold changes have the disadvantage of treating up- and down-regulated genes differently, where down-regulated genes are squashed between 1 and 0. Therefore, one of the first transformations performed in microarray data analysis, is Log2 transformation of the signals. This transformation produces a symmetrical spectrum of ratios, where a two-fold difference in either direction would result in Log2 ratios of 1 for up-regulation or -1 for down-regulation or 0 for no change. For the rest of this chapter, all values are Log2 transformed unless otherwise stated.

When performing the transcript profiling experiments, great effort was invested to minimise variation between array treatments, although systematic bias can never be completely avoided. There are a number of sources for technical variations, such as pipetting error, difference in labelling, sensitivity of signal detection and so on, which ultimately lead to differences in the overall signal intensities between arrays. Normalisation of the signal intensities is therefore a crucial step in microarray data analysis (Quackenbush 2002). Another question is, whether to normalise the microarray data across the entire set of arrays or to form groups of arrays based on time point or treatment, and

normalise within these individual groups. I have chosen to normalise within groups of treatments (namely UPD, UPD2 and Mock). Box plots visualise the distribution of signal intensities across different arrays, and demonstrate the impact of normalisation on overall intensity (Figure 10). Here, individual intensities were scaled such, that the median intensities are the same within arrays that are grouped together (median centring normalisation). It is these normalised signal intensities that are subsequently used for performing fold change analysis of genes between different arrays.

Although Log2 transformation of signals is an essential step in microarray analysis, it results in intensity dependence of fold changes (Log2 ratios) in which a deviation from zero is observed at low signal intensities (Quackenbush 2002). This phenomenon is best visualised by plotting mean signal intensities (x-axis called *A* for *add*) as a function of Log2 ratios (y-axis called *M* for *minus*), Figure 11. Log2 ratio values above 0 represent up-regulation of genes in samples treated with UPD or UPD2, whereas values below 0 indicate down-regulation of genes after stimulation as compared to mock samples. Interestingly, the shape of an *MA* plot is characteristic for the samples being compared. For closely related samples, the *MA* plot shows a 'tadpole' shape with a broader distribution of the Log2 ratios at low signal intensities, while *MA* plots generated from data of very different samples have a characteristic 'diamond' shape with wider Log2 ratio distribution at intermediate expression levels. The shape of the *MA* plots in Figure 11 fulfil the criteria that the transcript profiling results of UPD- or UPD2-treated cells originate from closely related samples.

MA plots are also helpful in visualisation of the behaviour of differentially regulated genes in a time course manner and under different treatments. 2h and 4h after stimulation of Kc_{167} cells with UPD results in comparable distribution of up- and down- regulated genes across all signal intensities (Figure 11A). 10h after treatment, the pattern of up-regulation is comparable to those at 2h and 4h, but there is an additional group of down-regulated genes, especially at medium signal intensities. The suppression of genes after JAK/STAT stimulation is surprising because in *Drosophila*, JAK/STAT signalling has only been described as an activating pathway.

Interestingly, UPD2 stimulation of Kc_{167} cells leads to a very different transcriptional response, compared to UPD treatment (Figure 11B). The striking absence of down-regulated genes is in stark contrast to the pattern of UPD stimulation. Given that the

activity of the UPD2 conditioned media, to stimulate JAK/STAT signalling, has been validated (Figure 7) and given the consistency of the *MA* patterns across all three time points, it seems very likely that the transcriptional response observed is real. As described earlier, UPD is the main ligand responsible for JAK/STAT signalling during *Drosophila* development. Whereas UPD2 has been shown to act semi-redundantly, its mutants are fully viable and it can only partially rescue the phenotypes of JAK/STAT pathway mutant. Therefore, I decided to focus on transcript profiling results generated by UPD stimulation for further downstream analysis.
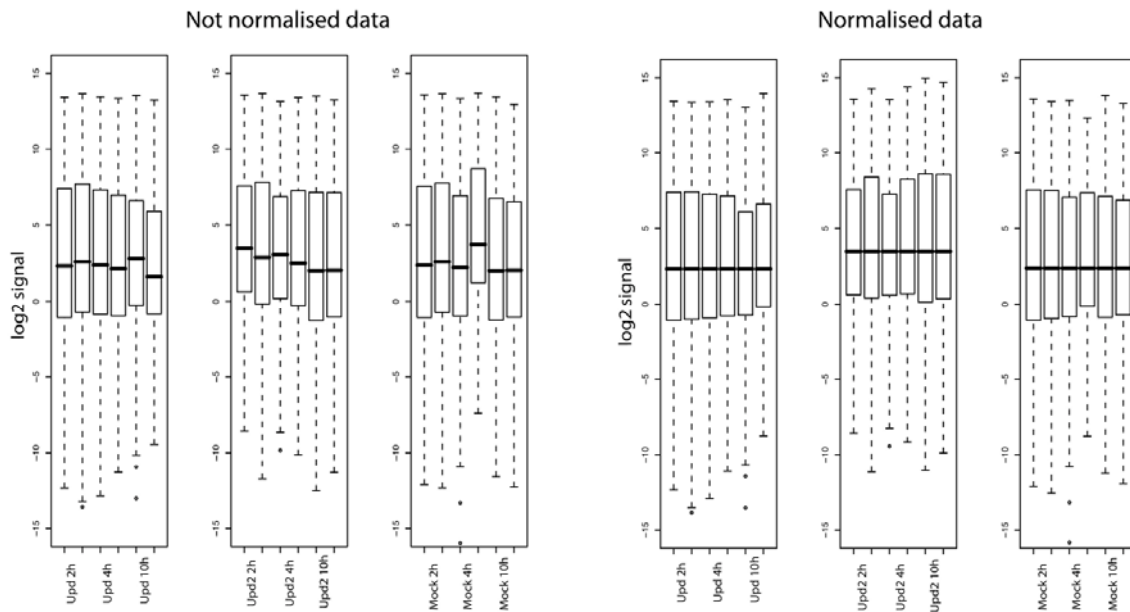
Figure 10.   Box plots.
The spread of probe signals within each array is visualised by box plots, with the thick horizontal lines representing the median signal intensity. Data generated from the GeneChip® arrays were grouped based on the conditioned media treatments (UPD, UPD2 or Mock) and the signals within each group normalised. Whereas the median intensities vary before normalisation, they are equalised after normalisation enabling a more accurate comparison between different arrays.
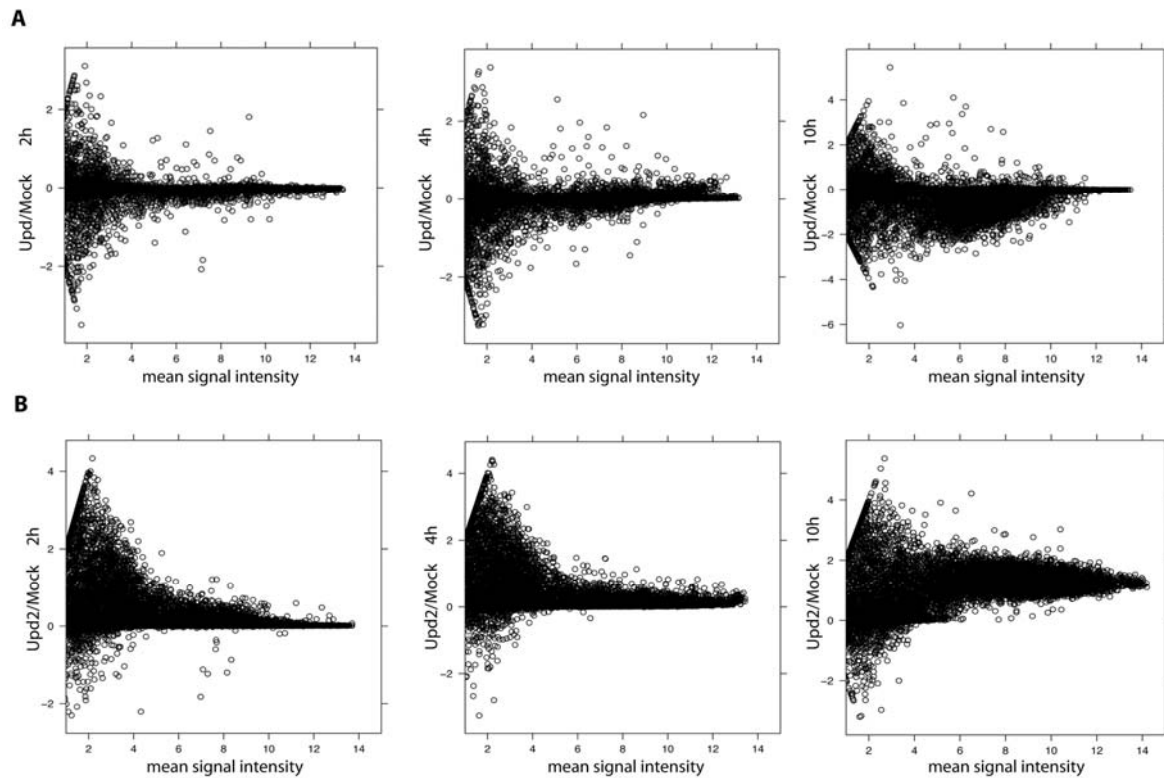
Figure 11. *MA* **plots.**
Mean signal intensities (x axis = *A*) are plotted against Log2 signal ratios (y axis = *M*).
Points above the horizontal line 0 correspond to up-regulated genes and those below 0
correspond to down-regulated genes in UPD- (A) or UPD2- (B) stimulated Kc$_{167}$ cells. The
transcriptional output of UPD- or UPD2-treated cells vary greatly, although both ligands
signal through a common linear signalling cascade.

Differential fold change cut off

Initial microarray analysis often used a uniform fold change cut off to define differentially expressed genes. The *MA* plots in Figure 11, however, suggest that this approach does not accurately reflect the structure of the data. For example, at low expression levels one might misidentify genes as significant, but at higher intensities, more subtle modulation of gene expression might not be identified at all. Therefore, great attention must be paid to the choice of fold change cut off. An alternative, as suggested in the literature, is to focus on the local structure of the data in order to measure the statistical significance of fold changes (Log2 ratios) (Yang et al. 2002). For the UPD *MA* plots, a sliding window of fixed width was used to divide the data set into sections spanning all signal intensities. For each section the mean of the population of the Log2 ratios (y-axis) and the standard deviation surrounding each Log2 ratio was calculated. Using this information, an intensity-dependent Z-score was calculated for each Log2 ratio measurement. A Z-score measures the number of standard deviations (SD) a particular point (here the Log2 ratio) is away from the mean of the population, and hence gives a measure of the significance of the data point. Differentially expressed genes associated with a Z-value that lies outside the 2-SD limit (red lines in Figure 12) were selected and represent a confidence level of 95% (i.e. $p<0.05$) (Figure 12). This procedure results in more stringent fold change cut offs at lower intensities, while at higher intensities, smaller yet significant changes are identified. Figure 13 quantitatively illustrates the regulation of these differentially regulated genes over time. From a total of 1197 UPD-regulated genes, the majority is either up- or down-regulated at only one time point. The expression of 35, 21 and 7 genes overlap between 2h and 4h, 4h and 10h, 2h and 10h respectively (Figure 13B). 11 genes are up-regulated across all three time points, including *socs36E*, a known STAT92E target gene, as well as *CG13559*, *Gα73B* and *CG4793*, which have been validated in this work, as mediators of tumourigenesis. *net*, *CG4804* and *CG10764* were also examined, but did not show any interaction with the JAK/STAT-mediated tumour formation. A list of all UPD-regulated genes is included in the Appendix.
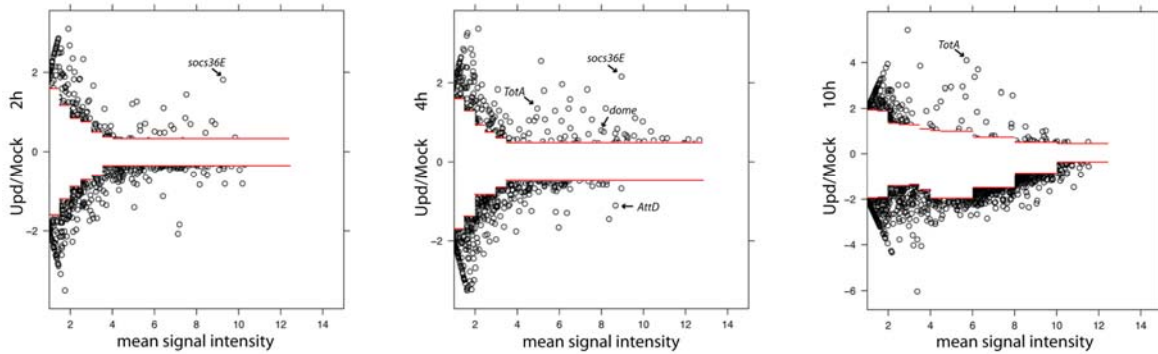
Figure 12. **Z-score analysis of differentially expressed genes regulated by UPD.**

For each Log2 ratio a Z-score is calculated enabling more stringent fold change cut offs at low mean signal intensities. At higher intensities, where less error is expected, lower fold change cut offs can be imposed on genes. *TotA*, *socs36E*, *dome* and *AttD* are known STAT92E target genes, which were also identified in this screen.
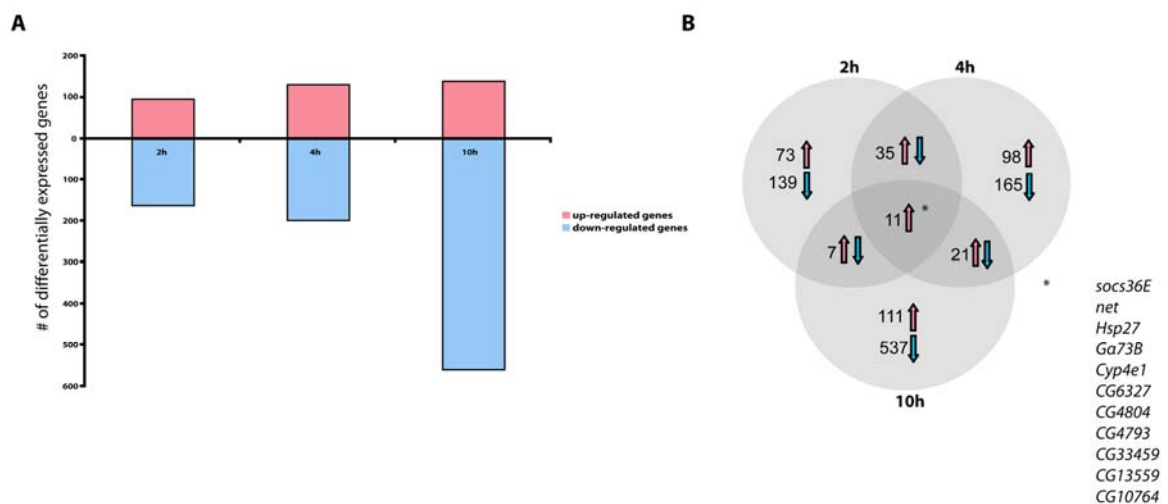
Figure 13. **Quantification of differentially expressed genes regulated by UPD.**

A) The number of up-regulated genes upon UPD signalling only increase slightly over time. However, a large proportion of genes are down-regulated particularly 10h after JAK/STAT pathway activation. B) Venn diagram illustrating the proportion of genes that intersect at least two out of three time points (dark grey area) and those that are uniquely regulated at only one time point. 11 genes that are up-regulated by UPD intersect at all three time points. These include the known pathway target gene *socs36E*, as well as *CG13559*, *Gα73B* and *CG4793*, which have been validated in their role in mediating tumourigenesis.

In silico and *in vitro* validation of microarray data

Real-time PCR validation of microarray data

When performing microarray analysis, unspecific hybridisation of transcripts to probe sets may be incorrectly detected as "true" signals, generating false positive results. In order to gain confidence in the data set, 11 genes were randomly selected for analysis of their expression pattern over time, by real-time PCR validation (Figure 14). The expression patterns of all genes predicted to be up-regulated by microarray analysis were also fully reproducible by RT-PCR analysis (Figure 14A). In addition, there was a strong correlation between microarray and PCR results in the temporal expression pattern of *TotA* and *CG6014*. The overall down-regulation of the selected genes could also be reproduced by RT-PCR analysis. In case of *ptr*, *CG13912* and *CG10262*, PCR analysis detected lower gene abundancy in UPD-treated samples at all time points, whereas microarray analysis was not able to detect some of these changes. It should be noted that for down-regulated genes the signal intensities of the probe sets were very low, making microarray data analysis more prone to error. Real-time PCR is likely to be more sensitive in detecting weakly expressed genes.

By successfully reproducing some of the microarray results, using an independent method, I could not only confirm the expected up-regulation of genes, but also underline the novel finding that some genes are also down-regulated after JAK/STAT pathway stimulation.

Hierarchical clustering

Given that the JAK/STAT transcript profiling in Kc_{167} cells was performed in a time-course dependent manner, it would be interesting to find similar expression pattern of groups of genes over time. Different methods have been developed, particularly for the field of transcript profiling, in order to visualise large-scale data. These techniques involve unsupervised methods such as hierarchical clustering and *k*-means clustering (Eisen et al. 1998) and methods such as self-organised maps (SOM), which reduce high-dimensional expression data sets into lower-dimensional space based on local similarity (Gerstein and Jansen 2000). The aim of hierarchical clustering is to group a set of genes which behave similarly in their expression pattern to the same cluster, whereas genes that behave differently fall in another cluster. The interplay among genes is represented by a tree, whose branch lengths mirror the degree of similarity between clustered genes. Next to this

tree a gene expression heatmap is displayed. A heatmap is a graphical representation of the data (in this case Log2 ratios) by quantitatively representing each data point with a colour. The 1197 differentially expressed genes depicted in Figure 12 & 13 were subjected to hierarchical clustering (Figure 15). For visualisation purposes, genes with Log2 ratios of 0 (no change) are black and up- and down-regulated genes are pink and blue respectively, with increasing colour intensities representing the increasing intensities of Log2 ratios. As seen in Figure 15 the data set can be divided into six major clusters, which highlight early (cluster 1 & 4), intermediate (cluster 2 & 5) and late (cluster 3 & 6) response genes. All six clusters mainly contain genes that are uniquely regulated at only one time point, an observation that is expected, given that only 74 genes intersect between two out of three time points and 11 genes are up-regulated at all time points (Figure 13B). Whereas the majority of genes are either up- or down-regulated, a small group of genes behave differently over time. For example, cluster 5 contains a small group of genes which are initially up-regulated at 2h but subsequently down-regulated at 4h.

Functional enrichment of clustered genes

Although computational analysis is a potentially powerful approach, biological knowledge needs to be applied to interpret these results fully. The reasons for the expression pattern of genes behaving similarly over time can be due to common *cis*-regulatory elements or close proximity of genes on the genome. In addition, co-evolution of *cis*-regulatory elements may have enabled the co-expression of genes which are functionally related. In fact, one of the most compelling validation of a clustering algorithm is to demonstrate a biological rational underlying the grouping of genes.

Several reports have shown that gene expression clusters can relate to common functional categories (Eisen et al. 1998; Marcotte et al. 1999; Niehrs and Pollet 1999). Three main sources of biological data are used to assign pathway and functional information to genes, these are the Gene Ontology project (GO), the Kyoto Encyclopaedia of Genes and Genomes (KEGG) and GenMAPP. In the past, researchers examined the list of genes manually, using public databases to extract relevant biological meanings. This is a very time-consuming process. Batch processing of large data sets was first introduced in 2002 and has since been improved to produce online tools (Khatri et al. 2002; Zeeberg et al. 2003). I have chosen the freely available web-based programs GoMiner[TM] (Zeeberg et al. 2003) and GOstat (Beissbarth and Speed 2004) to analyse the function of UPD-regulated
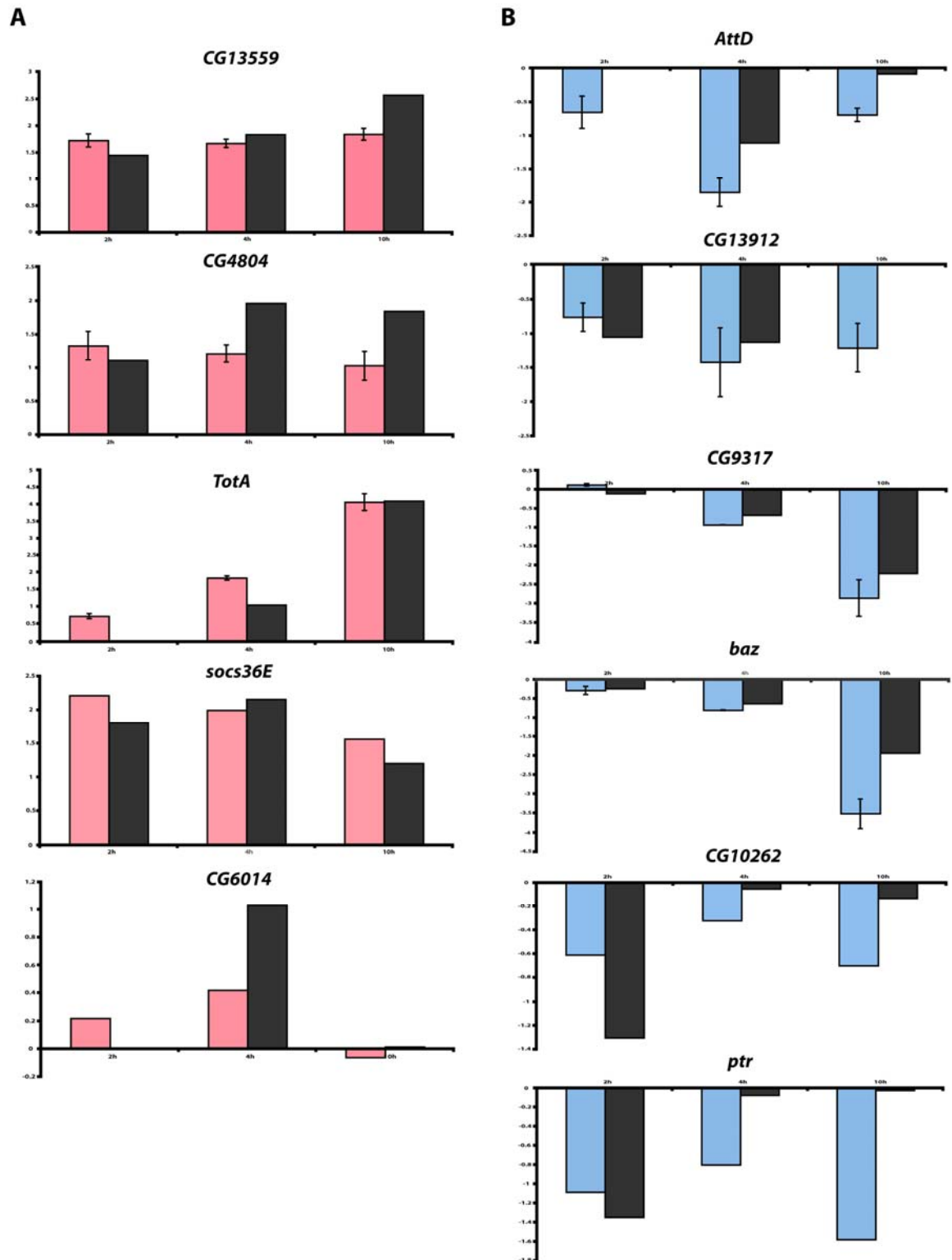
Figure 14.   Real-time PCR validation of putative STAT92E target genes.
Red and blue bars correspond to Log2 ratios of up- and down-regulated genes from real time PCR experiments respectively. Grey bars show the Log2 ratio results from the microarray data. Error bars represent standard error of the mean (SEM) from three independent PCR experiments. Where no error bars are indicated, the PCR experiment was performed once.
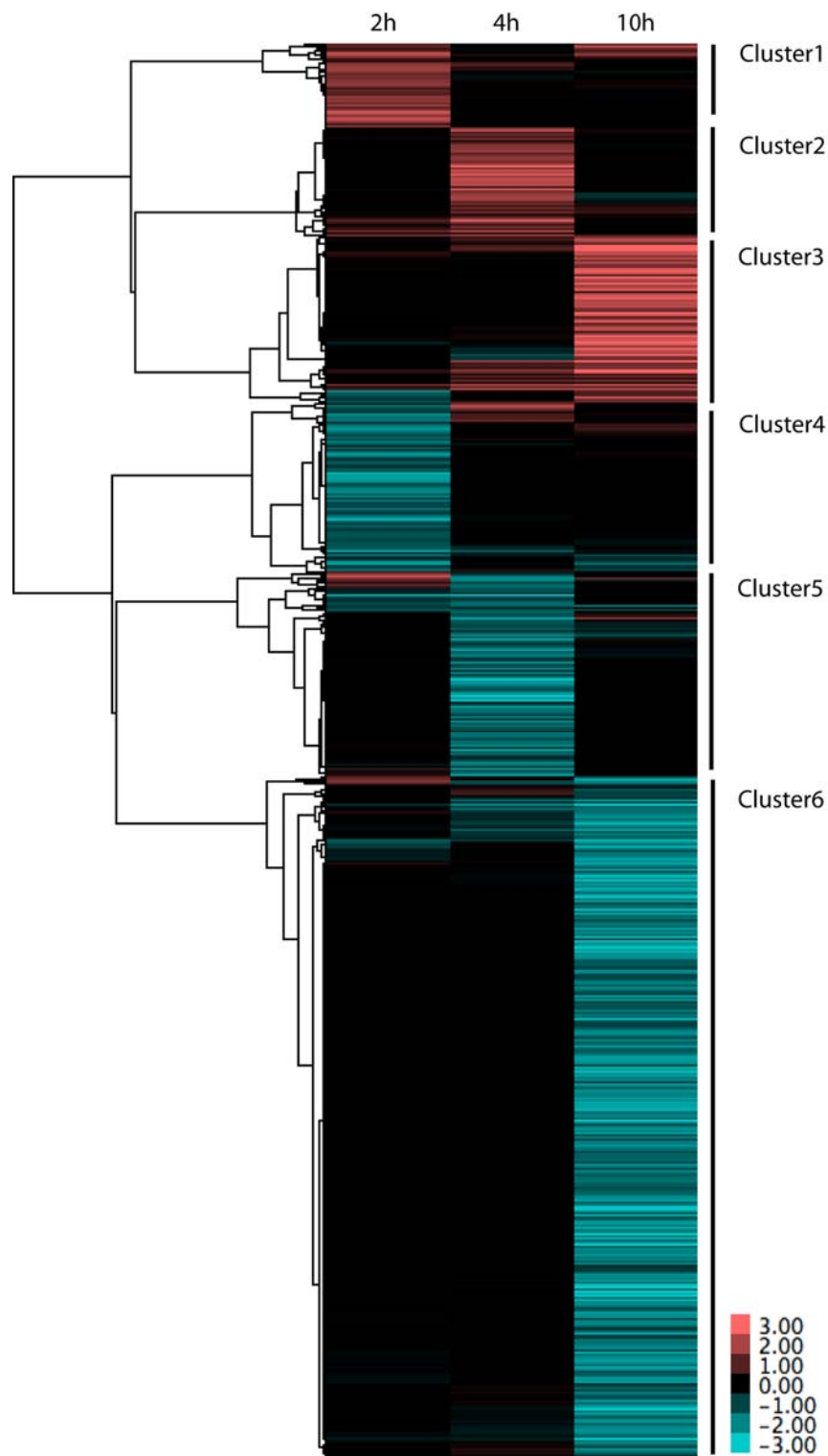
Figure 15.   Hierarchical clustering.
Co-expressed genes were clustered into groups. Horizontal lines represent the Log2 ratios (expression pattern) of individual genes at 2h, 4h and 10h after UPD stimulation. The colour bar represents the different Log2 ratio intensities. The data set can grossly be divided into six groups of early (cluster 1 & 4), intermediate (cluster 2 & 5) and late (cluster 3 & 6) response genes.

genes. The most important reason for the choice of GoMiner[TM] and GOstat is that both programs calculate the enrichment of a GO category with respect to what would have been expected by chance alone. They provide information on the statistical reliability of the enrichment by assigning a p-value to the occurrence of each GO category. Furthermore, the algorithms used in GoMiner[TM] examine the distribution of p-values that would occur by chance, thereby specifying a false discovery rate (FDR) for each category. FDR reflects the chance of a category to be a false positive and generally it is recommended to focus on categories which have been assigned with an FDR of less than 10%.

The functional enrichment of all 1197 filtered genes, which include up- as well as down-regulated genes, was assessed using GoMiner[TM] and the results are summarised in Figure 16 in form of a table (A) and a 3-D bar graph (B), visualising significantly enriched GO categories. Genes involved in gastrulation are highly enriched at early time points, whereas immune response (= defense response) genes are mainly enriched at 4h, and genes involved in protein modification processes are enriched at the latest time point. Whereas the graphs depicted in Figure 16 help identifying significant GO categories in the entire data set at different time points, it would be interesting to see if these results are also reflected in the structure of the hierarchically clustered genes from Figure 15. For this, each cluster was individually analysed for its functional enrichment using GOstat. As shown in Figure 17, many significantly enriched GOs identified by GoMiner[TM] also associate with individual expression clusters, such as proteolysis, immune response, cell morphogenesis and protein modification processes. Other functions such as gastrulation or segmentation were not enriched in any of the six expression clusters.

Promoter analysis of UPD-regulated genes

Genes can have similar expression profiles because they are regulated by a common *cis*-regulatory element in their promoter region. As apposed to many other transcript profiling assays, I had the advantage of knowing the DNA bindings site of the JAK/STAT transcription factor STAT92E. The N_3 STAT92E DNA binding site was first identified by *in vitro* SELEX approaches (Yan et al. 1996) and subsequently validated in a number of target genes *in vivo* (Small et al. 1996; Kwon et al. 2000). More recently, a lower affinity N_4 STAT92E binding site has been identified *in vivo* (Rivas et al. 2008) and has allowed us to screen for potential discrimination in the number of N_3 or N_4 binding sites, either between up- and down-regulated genes, or in a time course manner.

A

Significantly enriched molecular and biological functions

Function	Enrichment		
	2h	4h	10h
1) Defense response; GO: 0006952	2.4	5.8	0.3
2) Mesoderm development; GO: 000748	3.9	2.9	1.0
3) Segmentation; GO:0035282	3.8	2.4	1.2
4) Proteolysis; GO: 0006508	2.2	1.9	1.1
5) Dephosphorylation; GO: 0016311	3.3	1.9	2.1
6) Gastrulation; GO: 0007369	5.2	1.9	1.7
7) Imaginal disc development; GO: 00074444	1.7	1.1	1.4
8) Signal transduction; GO: 0007166	2.5	1.1	1.0
9) Cell projection morphogenesis, GO: 0048858	0.6	1.1	1.9
10) Regulation of transcription; GO: 0045449	2.4	0.9	1.6
11) Post-transl. protein modification; GO: 0043687	1.1	0.5	1.4

* Numbers correspond to those in the bar chart

B
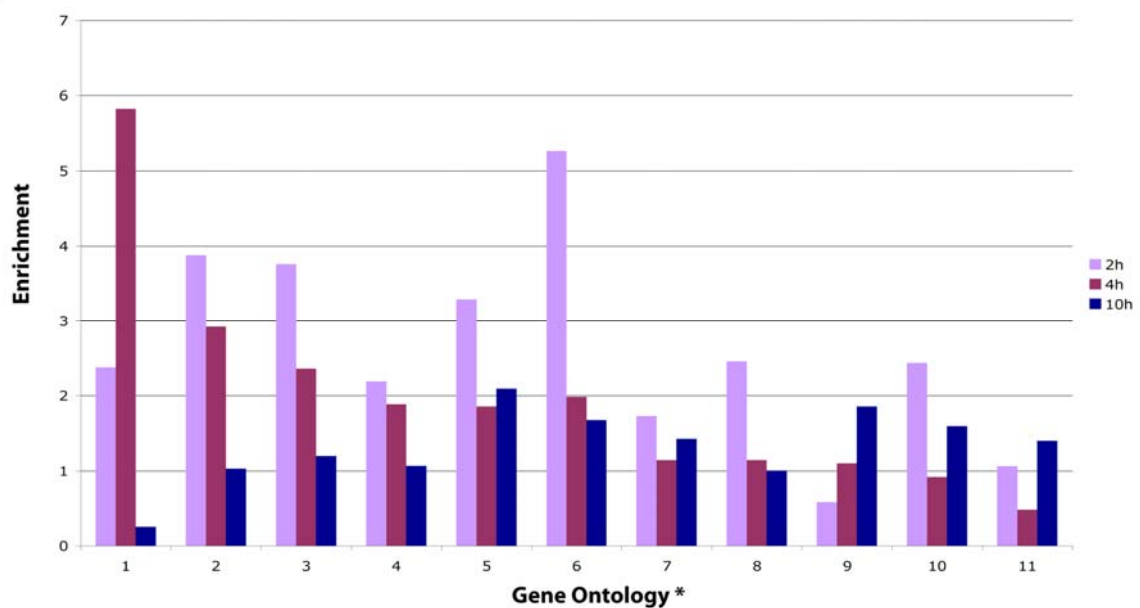
Figure 16. Functional enrichment of UPD-regulated genes
represented in form of a table (A) and a chart (B). The 1197 UPD-regulated genes were examined for enrichment (y axis) of functional categories (x axis) at the three time points 2h, 4h and 10h.
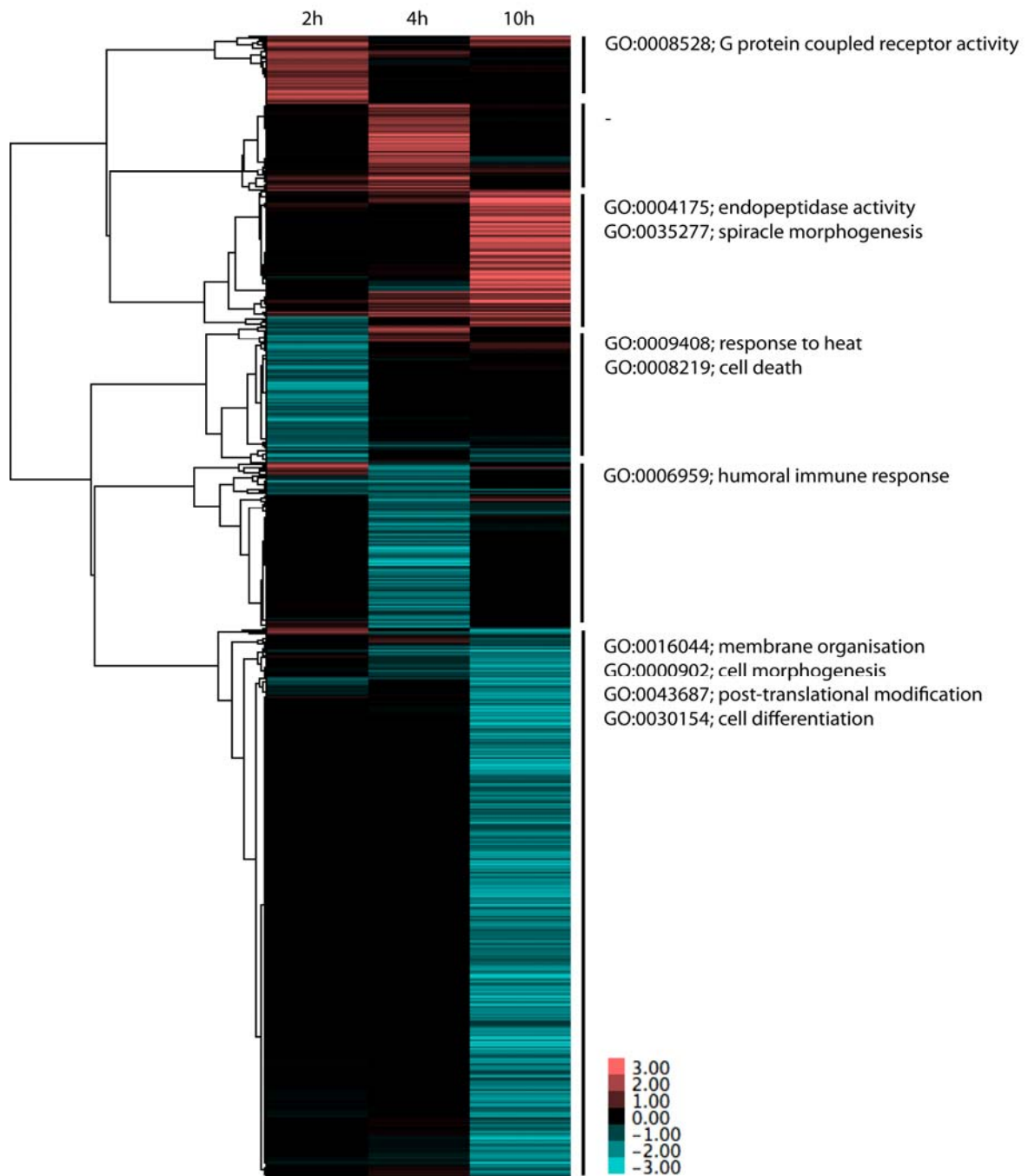
Figure 17.   Functional enrichment of co-expressed genes.
Groups of co-expressed genes were re-examined for functional enrichment. Some gene ontology categories that were previously enriched in the entire data set (Figure 16) do not appear to be enriched with individual clusters, but other categories relevant to immunity (endopeptidase activity and humoral immune response) are enriched.

In order to quantify the potential enrichment of STAT92E binding sites in promoters of putative JAK/STAT target genes (Figure 12 & 13), the 3Kb upstream region of their transcriptional start was retrieved using the RSAT web-based tool (Thomas-Chollier et al. 2008). As a control dataset, 3Kb upstream regions of a random set of *Drosophila* genes were also retrieved. First, all STAT92E binding sites (N_3 & N_4) were quantified in these regions (Figure 18A). Comparison between up- and down-regulated genes shows that only up-regulated genes contain more STAT92E sites in their promoter region, with genes at the 2h time point having significantly enriched binding sites compared to random genes. This result is expected as the JAK/STAT pathway ends with an activating transcription factor. We would also expect that STAT92E binding sites are more enriched in promoters of early-response genes, whereas later, secondary-response genes may have been induced, which were not necessarily directly activated by STAT92E. Having established that up-regulated genes are more likely to be true STAT92E targets, we were also interested in a potential discrepancy in the number of N_3 and N_4 binding sites at different time points (Figure 18B). Whereas the high affinity N_3 site is enriched in promoters of up-regulated genes at all time points, low affinity N_4 sites are only enriched in promoters of early-response genes, indicating that a strong JAK/STAT stimulation is necessary in order to activate genes that contain N_4 binding sites.

The chances of target genes being directly regulated by a transcription factor are higher when their promoter regions contain multiple transcription factor bindings sites within a 500bp window (Markstein and Levine 2002; Ochoa-Espinosa et al. 2005). Such sites are often termed clustered binding sites. As the six base pair long palindromic STAT92E sequence is likely to be found by chance, the filtering of target genes based on clustering of STAT92E binding sites is helpful. Given that the results in Figure 18A show that only up-regulated genes contain enriched STAT92E binding sites, I focused on this set of genes. A total of 30 up-regulated genes (~10% of all up-regulated genes) contain clustered STAT92E binding sites in a 500bp window (Figure 18C). It would be interesting to see which of these clustered DNA sites are also evolutionary conserved in other *Drosophilae* or even vertebrae species.

Although the significance of the STAT92E binding sites remain to be validated *in vivo*, their enrichment in up-regulated genes gives further confidence in the validity of the transcript profiling assay.
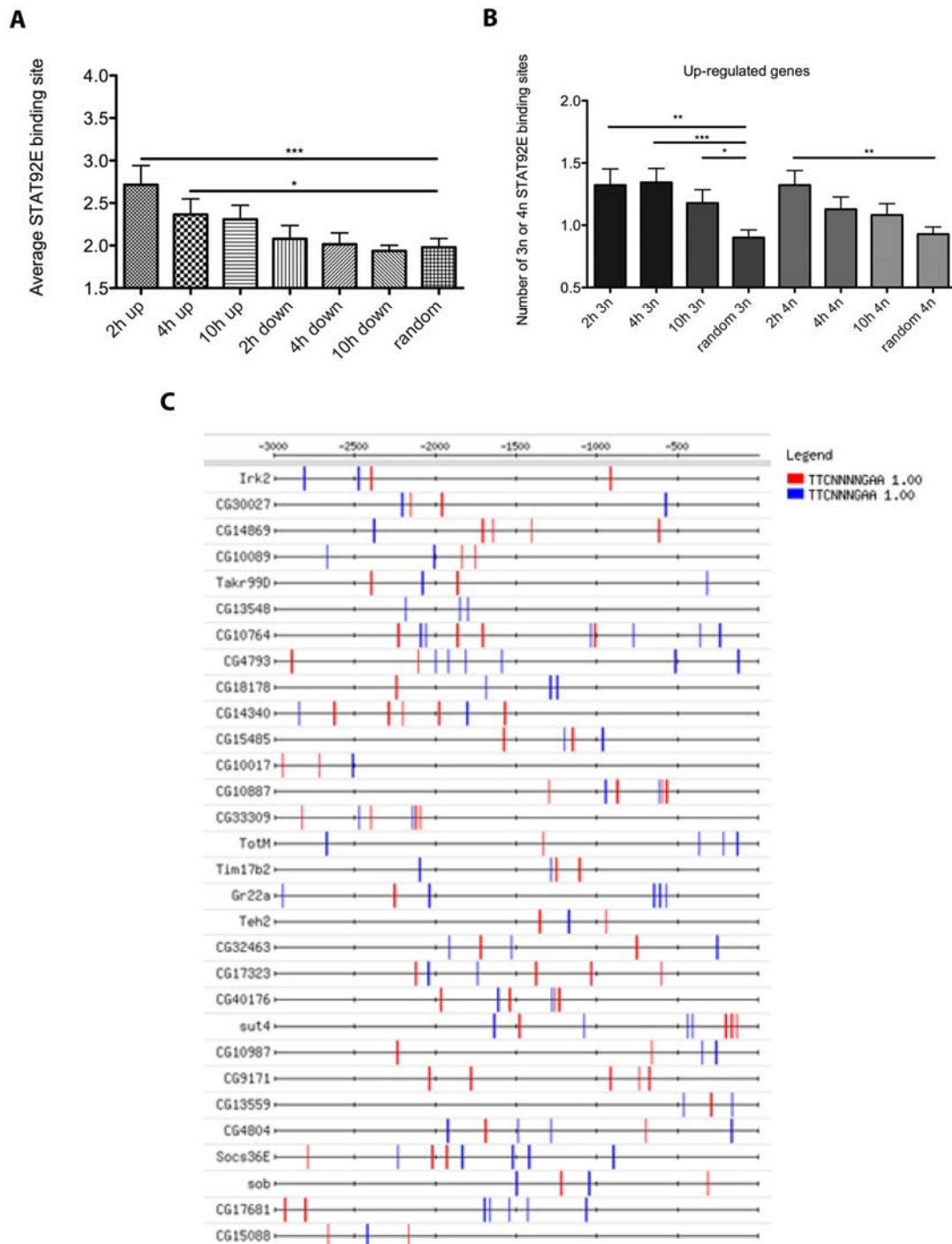
Figure 18.   Quantification of STAT92E binding sites.
A) Average number of STAT92E binding sites in the upstream region of UPD-regulated genes compared to a random set of *Drosophila* genes. STAT92E binding sites are only enriched in up-regulated genes. B) Average number of N_3 or N_4 STAT92E binding sites in up-regulated genes compared to those in random *Drosophila* genes. While high affinity N_3 sites are enriched in promoter regions of early to late response genes, the low affinity N_4 sites are mainly enriched in promoters of early response genes. ***$=p< 0.0005$, **$=p<0.005$, *$=p<0.05$ C) Up-regulated genes which contain at least three N_3 or N_4 binding sites in a 500 bp window.

Distribution of co-expressed genes in the Drosophila genome

The possibility of co-expressed genes being physically situated next or close to each other at the chromosomal level was first highlighted in the yeast (Cho et al. 1998; Cohen et al. 2000). Elegant statistical analysis could show that some adjacent genes that are transcriptionally co-expressed, are similar in their function, or alternatively, contain a common *cis*-regulatory element regulating their co-expression. Striking analogy could also be drawn for some co-regulated *Drosophila* immune response (Khush and Lemaitre 2000), circadian (McDonald and Rosbash 2001) and sex-biased genes (Parisi et al. 2004), which are also located in clusters on the genome. Interestingly, Spellman and Rubin used a dataset of 80 experimental conditions using microarrays in order to search for groups of adjacent genes with similar expression pattern (Spellman and Rubin 2002). The dataset came from studies of aging, DNA damage response, immune response, resistance to DDT and embryonic development. They concluded that 20% of assayed genes fall into groups of 10-30 genes which are expressed similarly across numerous experimental conditions. However, they could not show any correlation between the grouping of these genes with any known chromosomal structures, or with functional similarity, which is in stark contrast to previously published data.

Given that only up-regulated genes contained enriched STAT92E binding sites in my data set, it was interesting to find the reason for the down-regulation of genes (clusters 4-6 in Figure 15). One possibility is the close proximity of genes on chromosomal level, which would allow concomitant regulation by chromosomal modification, ultimately leading to heterochromatin formation. Or alternatively, several genes can be regulated by a common repressor element. In order to address these questions, firstly, the distribution of co-expressed genes (i.e. clustered genes from Figure 15) along the four *Drosophila* chromosomes was examined (Table 6). Statistical analysis (chi-square test, $p<0.05$) suggests a non-random distribution of co-expressed genes from Figure 15 along the chromosome. I then examined whether a minimum of three co-expressed genes are situated in a 25Kb window in the genome (Table 7). Whereas none of the up-regulated genes fulfilled this grouping criteria, some down-regulated genes (cluster 4-6 from Figure 15) did. Considering the number of down-regulated genes in cluster 6, statistical analysis needs to be performed in order to rule out the occurrence of these grouped genes by chance. Furthermore, although some proximal genes are functionally related, the majority are not.

Observed and expected frequency of distribution of target genes in *Drosophila* genome

Chromosome	No Genes	%	Observed (Expected) frequency of clusters:					
			1	2	3	4	5	6
X	2251	16.1	10(13	14(17)	10(24)	17(26)	29(33)	136(103)
2L	2571	18.4	10(13)	25(16)	31(24)	25(26)	38(32)	85(102)
2R	2709	19.4	9(13)	16(17)	28(25)	23(28)	40(34)	109(107)
3L	2666	19.0	15(13)	16(17)	23(25)	32(27)	26(33)	107(105)
3R	3400	24.3	22(16)	18(21)	38(31)	42(34)	39(42)	116(131)
4	81	0.6	0(0)	0(1)	1(1)	3(1)	3(1)	1(4)
Unmapped	320	2.3	2(0)	0(0)	0(0)	0(0)	0(0)	1(0)

Table 6.   Distribution of UPD-regulated genes in the *Drosophila* genome.
The distribution of co-expressed UPD-regulated genes in the *Drosophila* genome compared to their expected distribution is shown. Cluster numbers refer to those from Figure 15.

Group of genes in a 25Kb window size

Cluster 4	Common Function	Cytogenetic map
Hsp70Ba, HspBbb, Hsp70Bb, Hsp70Bc	response to heat	Chr3R (87B)

Cluster 5		
AttC, IM10, CG33470	antibacterial humoral response	Chr2R (50A)
Dro, AttA, AttB	antibacterial humoral response	Chr2R (51C)
CG32599, mud, CG13404	-	ChrX (12E)

Cluster 6		
CG13949, aru, PNUTS	-	Chr2L (21E)
b, tam, sop2	-	Chr 2L (34D)
TpplI, Nrk, CG33137, CG17724, seq	-	Chr2R (49F)
fat-spondin, CG30460, sply	-	Chr2R (53F)
Mrtf, CG9977, Atg2	-	Chr3L (62F)
CG32066, simj, CG33493	-	Chr3L (76E)
CG6199, CG7600, CG18490	-	Chr3L (68B)
CG10140, CG10732, CG10738	-	Chr3L (70B)
Mi-2, Su(Tpl), Su(z)12	regulation of transcription	Chr3L (76D)
CG7518, CtBP, ry	-	Chr3R (87D)
amon, CG6420, CG31079	-	Chr3R (97C)
CG14516, Rpn2, CG11897	proteolysis	Chr3R (98F)
CG12065, Crag, Caf1-180	-	ChrX (7F)
CG9941, mus101, CG10990	-	ChrX (12B)
shtd, CG6299, CG6294, CG6340	-	ChrX (13C)
CG8134, CG9281, CG8184	-	ChrX (13E)
Pvf1, CG7282, CG7280, CG32544, CG32542	hemocyte migration	ChrX (17E)
e(y)3, Ranbp21, CG1415	-	ChrX (18D)
CG11940, CG11943, CG15618	-	ChrX (18F)
HERC2, cactin, mal	-	ChrX (19C)

Table 7.   Proximity of co-expressed genes on chromosomal level.
Genes within each cluster (from Figure 15) were examined for proximity of at least three genes in a 25Kb window size. Genes may be situated close to each other and co-expressed due to their common biological function or to a common DNA regulatory element (to be determined). If a minimum of two genes had a common gene ontology classification, the entire group was assigned as having a common function.

It would be interesting to examine the existence of a common repressor element which may be responsible for the down-regulation of grouped genes.

Selection of genes for *in vivo* validation

In order to verify the *in vivo* significance of JAK/STAT pathway target genes, the 1197 UPD-regulated genes were systematically screened for their function (Gene Ontology, Figure 16 & 17), number of STAT92E binding sites in their promoter region (Figure 18), homology to mammalian proteins and overlaps with other published microarray data, which mostly focused on *Drosophila* immunity (Asha et al. 2003; Wertheim et al. 2005; Kambris et al. 2006; Kwon et al. 2008). Literature analysis identified over 100 up- and down-regulated genes which are likely to play a role in immune response. Combining the results gained from the above analyses, enabled the generation of a list of 21 genes which were selected for validation in their role during *Drosophila* haematopoiesis (Table 8). In addition, the majority of the selected genes were homologous to known human disease-causing genes (Chien et al. 2002).

Role of JAK/STAT target genes in haematopoiesis and tumour formation

The JAK/STAT pathway has pleiotropic roles during *Drosophila* development. This includes the control of cell proliferation in the eye and wing, germline stem cell maintenance, as well as haematopoiesis and immunity (see Introduction). Notably, mis-regulation of JAK/STAT signalling in haemocytes has been shown to result in "fly leukaemia" (Hanratty and Ryerse 1981; Luo et al. 1995).

Since the transcript profiling data were obtained using haemocyte-like Kc_{167} cells, I chose the fly leukaemia model to study the *in vivo* role of putative STAT92E-regulated genes (Table 8). In order to do so, I used the hop^{Tuml} mutation in combination with the GAL4/UAS system. The hop^{Tuml} mutation is a temperature-sensitive dominant gain-of-function allele of *hop* and causes over-proliferation of haemocytes, mis-differentiation of lamellocytes and as a result, formation of melanotic tumours (Figure 5B). The GAL4/UAS system was used to modify the expression of putative STAT92E target genes in tissues where the oncogenic hop^{Tuml} allele is active. This is achieved by using a *gal4* construct that is activate in tissues where the JAK/STAT pathway receptor *dome* is expressed. Several haemocyte-specific GAL4 driver lines have been previously described, but the majority are constrained to already differentiated blood cells (Jung et al. 2005). By contrast, the *dome-*

Putative STAT92E target genes chosen for *in vivo* RNAi assay

Gene name *Drosophila melanogaster*	Fold change Log(2) 2h	4h	10h	Gene ontology Biological function	Molecular function	STAT92E binding sites N3	N4	Gene name *Homo sapiens*
baz [1]	-0.2	-0.6	-1.9	asymmetric protein localization	protein kinase C binding	0	1	PAR3
CG10764	0.5	1.3	2.5	proteolysis	serine-endopeptidase	6	4	F9
CG13559 [5]	1.5	1.8	2.6	-	-	2	1	LITAF
CG15211 [1]	0	0.6	0	-	-	1	0	CMTM4
CG15678	0.8	0.1	0.7	negative regulation of innate immune respomse	-	1	0	-
CG1572	0	0.6	0.6	-	-	2	1	-
CG31158	0.7	0.7	0	regulation of ARF protein signal transduction	ARF guanyl-nucleotide exchange factor	1	0	PSD3
CG3829	0.7	1.4	0.1	defense response	scavenger receptor activity	2	0	SRB1
CG4080	0.7	1.1	0	-	protein binding	1	1	MARCH2
CG4793 [3,5]	0.7	1.6	3.7	proteolysis	serine-endopeptidase	6	2	CTRB2
CG4804	1.1	2.0	1.8	-	serine-endopeptidase	4	2	PAI-2
CG6014 [1]	0	1.0	0	-	-	2	0	-
Gα73B	1.2	2.6	2.9	mesoderm development	GTPase activity	2	0	GNAL
Idgf1 [3]	0	0.9	2.9	imaginal disc development	imaginal disc growth factor activity	0	0	CHIT1
l(2)gl	0	1.3	2.4	protein localization	myosin binding	1	0	LLG1
mfas	0	0.8	1.0	axonegenesis	-	2	0	TGFBI
mthl3	0.1	0.8	0.2	response to stress	G-protein coupled receptor activity	6	4	LPHN1
net	2.3	1.2	2.2	imaginal disc-derived wing vein morphogenesis	transcription factor activity	2	2	ATOH8
pxb	-0.7	-0.8	0	smoothend signalling pathway	-	0	0	-
TotA [2]	0	1.1	4.1	response to stress	-	2	2	-
Wnt4 [4]	2.2	-1.5	0	cell migragtion	signal transducer activivty	0	2	-

1: Asha et al. 2003. Analysis of Ras-induced overproliferation in *Drosophila* hemocytes. Genetics 163(1): 203-15
2: Agaisse et al. 2003. Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury. Dev Cell 5(3):441-50
3: Kwon et. al 2008. The nucleosome remodeling factor (NURF) regulates genes involved in *Drosophila* innate immunity. Dev Biol 316(2): 538-47
4: Terry et al. 2006. Novel regulators revealed by profiling *Drosophila* testis stem cells withing their niche. Dev Biol 294(1):246-57
5: Wertheim et al. 2005. Genome-wide gene expression in response to parasitoid attack in *Drosophila*. Genome Biol 6(11): R94

Table 8. **Summary of genes selected for *in vivo* validation.**

Gal4 driver is expressed in the medullary zone of the lymph gland where haemocyte progenitors are maintained and where the JAK/STAT pathway is active (Figure 5A) (Bourbon et al. 2002; Jung et al. 2005; Krzemien et al. 2007; Mandal et al. 2007). Interfering with haematopoiesis at an early stage before the onset of haemocyte proliferation or differentiation should result in a more robust phenotype affecting tumour formation.

Leukaemia assay using genotype tumour index

The crossing scheme for the haematopoiesis screen is outlined in Fig. 19A. hop^{Tuml},*dome-Gal4* recombinant flies were generated as shown in Figure 19B. For screening purposes, hop^{Tuml},*dome-Gal4* females were crossed to males carrying a target-gene specific UAS-construct. The progeny of this cross was screened for modulation of the tumour phenotype. Only female flies could be used, as hemizygous hop^{Tuml},*dome-Gal4* recombinant males did not survive.

With the exception of *baz* and *pxb*, all analysed candidate genes (Table 8) were up-regulated after JAK/STAT pathway activation. Therefore, to address their potential *in vivo* function during Hop^{Tuml}-induced tumourigenesis, gene expression was knocked down using a gene-specific RNAi construct (Dietzl et al. 2007). RNAi transgenes contain a gene fragment that is cloned as an inverted repeat which upon its expression forms a long double-stranded 'hairpin' RNA that mediates knock down of its target RNA (Fortier and Belote 2000; Lam and Thummel 2000). In contrast, *baz* expression, which is down-regulated by JAK/STAT signalling, was analysed by overexpression of a *baz* UAS-construct (Kuchinke et al. 1998; Wodarz et al. 1999). Unfortunately, no UAS-overexpression construct was available for *pxb*.

In order to establish a reliable protocol for tumour detection and quantification, a primary screen was conducted. For this, hop^{Tuml},*dome-Gal4* females were crossed to *UAS*-target-gene-RNAi males at 25°C. Females were allowed to lay eggs for 24 hours, before the embryos were transferred to 29°C, the restrictive temperature for full hop^{Tuml} and *GAL4/UAS* activation. Wild type w^{1118} (negative control) and *UAS-stat92ERNAi* (positive control) males were used as controls for hop^{Tuml} induced tumours and their reduction when knocking down the transcription factor STAT92E, respectively. For each cross 50-200 offsprings were scored one day after eclosure. The tumour size was variable (Figure 19C), which made it necessary to assign a score for each size. The scores ranged from 0 to 4,

with 0 being no tumour and score 1 was defined as a tumour with a diameter equal to the width of an abdominal segment. The tumour index (TI) was then calculated as:

TI = ∑(tumour size x n)/N

n: number of occurrence of particular tumour size

N: total number of flies of particular genotype

With this scoring scheme, tumour sizes were recorded for all flies rather than counting each tumour for individual flies. Knock-down of six out of 14 target genes and the positive control s*tat92E* reduced the tumour, when compared to the w^{1118} wild-type control cross (Figure 20). These genes were selected for re-examination in two further independent experiments under the same conditions as described above. Average TI results from the three experiments show that knock-down of *CG13559*, *Gα73B*, *Idgf1* and *CG4793* consistently reduced tumour growth, whereas knock-down of *TotA* and *CG10764* did not rescue the tumour phenotype (Figure 21). Here, statistical analysis was not performed, as p-values gained from a data set with low number of experimental replicates (in this case three) cannot be relied on.

A

$hop^{Tuml},P\{GawB\}dome^{PG5}/FM6$ X $UAS\text{-}targetgeneRNAi/TM3\ (Cyo)\ or\ UAS\text{-}Baz$

$hop^{Tuml},P\{GawB\}dome^{PG5}/+\ ;\ UAS\text{-}targetgeneRNAi/+$

B

hop^{Tuml}/Y X $P\{GawB\}dome^{PG5}/FM7$

$FM6/Y$ X $hop^{Tuml}/P\{GawB\}dome^{PG5}$

$hop^{Tuml},P\{GawB\}dome^{PG5}/FM6$

C
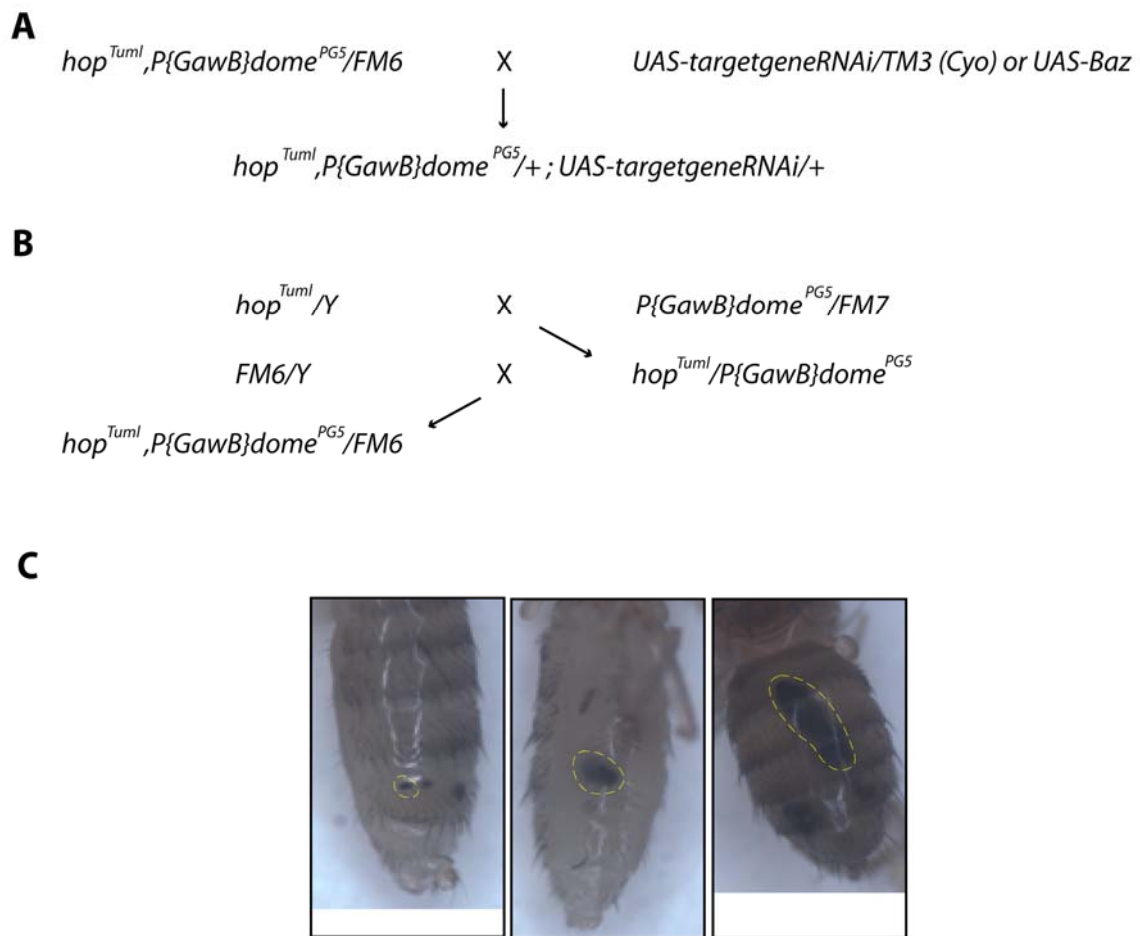
Figure 19.   Experimental set up of Leukaemia assay.
A) Crossing scheme for generating hop^{Tuml},*dome-Gal4*(i.e.*P{GawB}domePG5*) recombinants. B) Crossing scheme of tumourigenesis screen. Female recombinants from A) were crossed to males carrying a UAS-transgene and their progeny was examined for tumour sizes according to the scheme in C) Examples of different tumour sizes (circled in yellow) that are caused by the hop^{Tuml} gain-of-function mutation.
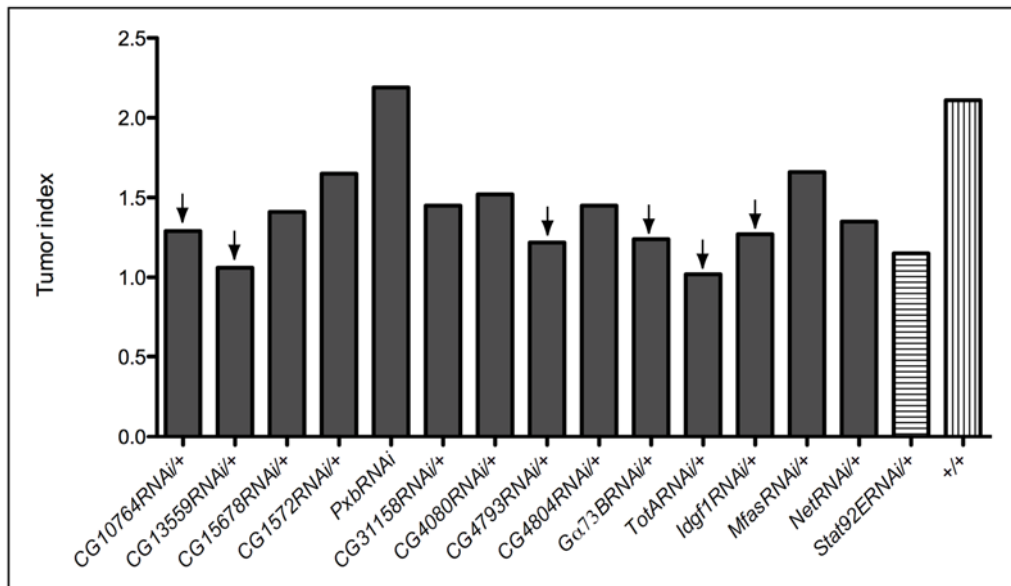
Figure 20.   Leukaemia assay, Genotype I.
hop^{Tuml},*dome-Gal4* females were crossed to *UAS-RNAi* males. Knock down of six out of 14 tested genes (arrows) along with knock-down of *stat92E* as a positive control reduced tumour size when compared to the control *wt* (+/+) cross.
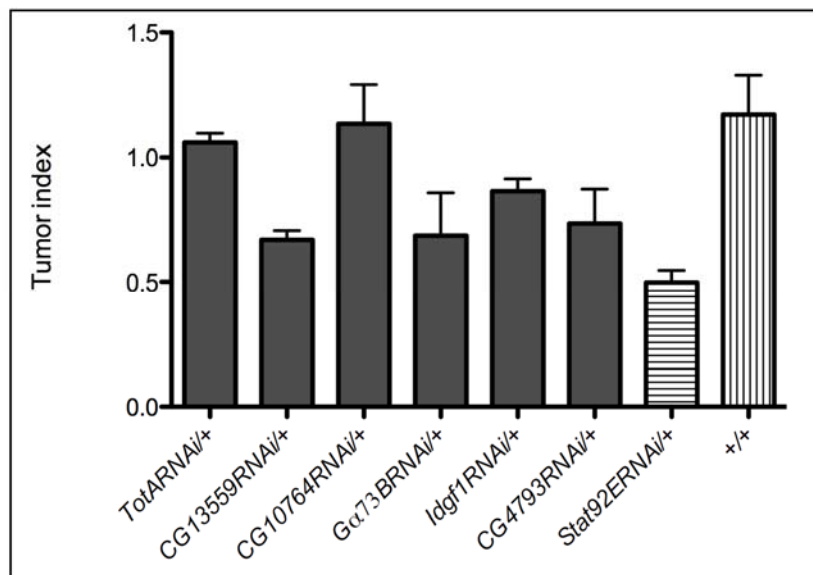
Figure 21.   Leukaemia assay, Genotype I.
The six genes from Figure 20 which reduced tumour growth were re-examined in two further experiments using identical screening conditions. The average TI of all three screens is depicted here. Knock-down of *CG13559*, *Gα73B*, *Idgf1* and *CG4793* consistently reduced tumour growth, whereas knock-down of *TotA* and *CG10764* failed to do so. Statistical test was not performed, as the number of replicates was low (three screens).

Leukaemia assay using single fly tumour index

In order to be able to perform high quality statistical tests, I next to assigned a tumour index to each individual fly (rather than an entire genotype), thereby increasing the number of replicates. RNAi lines targeting *CG13559, Gα73B, Idgf1, CG4793* were selected to test the new scoring system. In addition, although *TotA* did not consistently reduce tumour size, it was chosen again due to its known significance in immune response (Agaisse et al. 2003). For each cross 30-100 offsprings were analysed. Figure 22 shows the average tumour size per fly. T-test was performed in order to determine the significance of the tumour reduction. Knock-down of *CG13559, Gα73B, CG4793* and *TotA* (although less significant) reduced average tumour size, whereas knock-down of *Idgf1* failed to do so when compared to control cross.

Having established the single fly tumour count as a valid method for studying the role of JAK/STAT pathway target genes in a fly leukaemia-like model, I used this system to examine the remaining untested genes as well as those from Figure 22. 50-200 females were counted per cross, with exception of *CG3829* lines, where only 30 offsprings were obtained. All tested genes could be validated in their role as positive regulators of tumourigenesis, as their knock-down reduced tumour size significantly (Figure 23). Most notably, *baz* over-expression reduces tumour growth even more than knock-down of the transcription factor *stat92E*.

All experiments described so far were performed at temperatures (25-29°C) in which the hop^{Tuml} allele is highly active, resulting in high levels of blood cell proliferation. However, excess blood cell proliferation may be detrimental to fly development, with the possibility that only less severely affected flies survive (escaper flies). This would lead to a result biased towards more healthy escaper flies with fewer tumours.

In order to test this, I carried out new *RNAi* experiments at lower temperatures where crosses were set up at 20°C. Females were allowed to lay eggs for 24h. The embryos were maintained at this temperature for a further 24h or 48h (resulting in 48hr or 72hr at 20°C respectively) before being transferred to 29°C. Adult offsprings were scored one day after eclosure. For this experiment *UAS-Rh4-RNAi* was included as an additional control to address general effects of *RNAi* overexpression on the formation of tumours. *rh4* is an eye-specific gene that is thought not to be involved in haematopoiesis. As shown in Figure 24, RNAi-mediated knock-down of *rh4* does not affect tumour growth when compared to
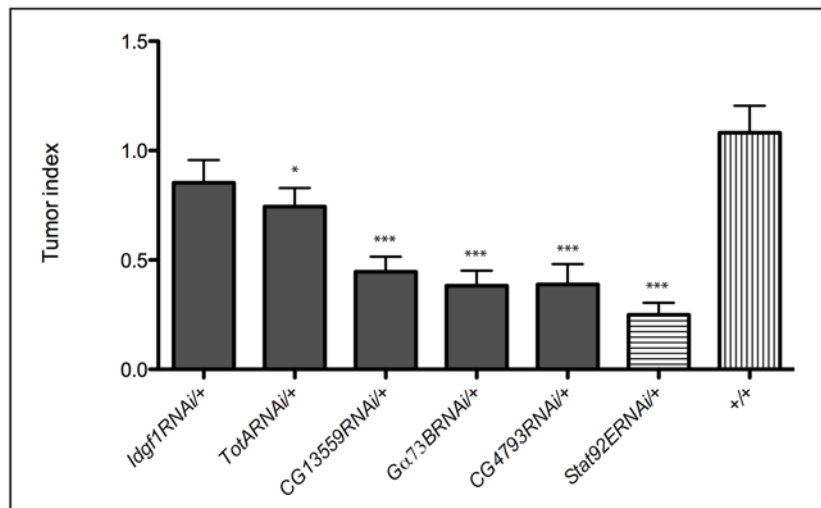
Figure 22.   Leukaemia assay, Single fly I.
Hits from Figure 21 along with *TotA*, were selected for a modified screen where individual offsprings were treated as replicates in order to increase the power of statistical analysis. 30-100 offsprings were scored for tumour size in each cross. Knock-down of *CG13559*, *Gα73B* and *CG4793* yielded the most consistent tumour reduction compared to control. The data was statistically analysed by t-test ***=p< 0.0005, **=p<0.005, *=p<0.05.
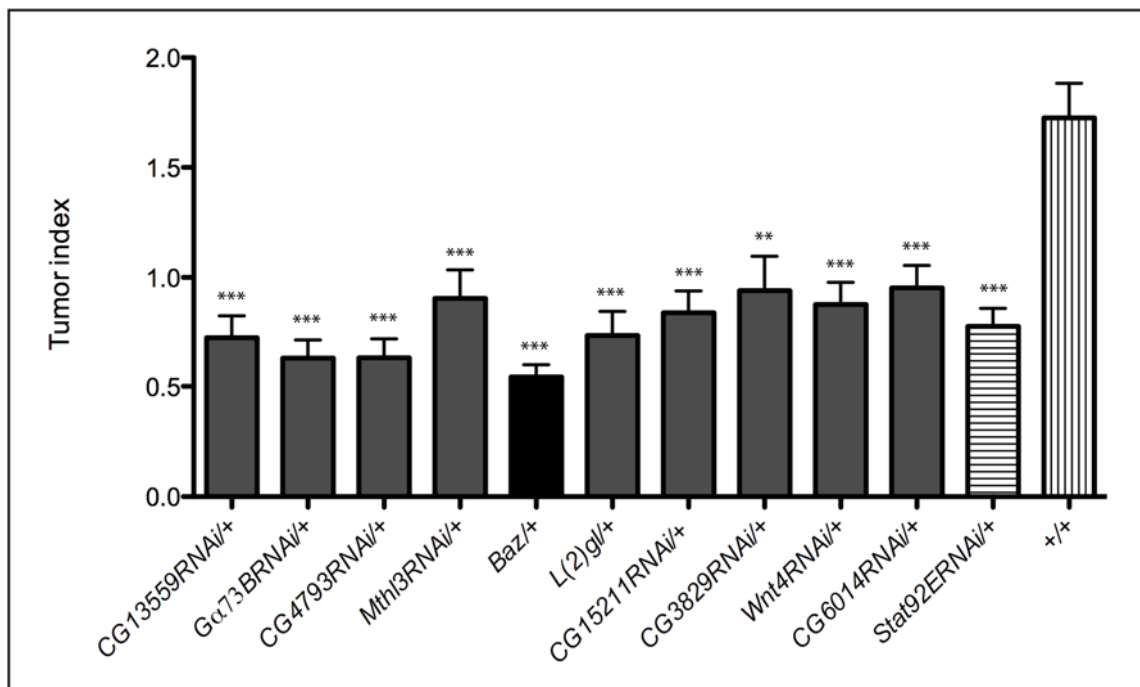
Figure 23.   Leukaemia assay, Single fly II.
Genes from previous screen and a new set of untested RNAi lines as well as a *UAS-Baz* overexpression line were examined for their ability to reduce tumour growth using the same conditions as in Figure 22. *Baz* overexpression resulted in a comparable tumour reduction phenotype as *stat92E* knock-down. T-test analysis was performed on 50-200 offsprings per cross. ***=p< 0.0005, **=p<0.005, *=p<0.05

wild-type controls. This result demonstrates that the reduction of tumour growth as seen in many candidate genes is due to a gene-specific knock-down rather than a general RNAi effect. Consistent with previous experiments, the RNAi-mediated knock-down of most candidate genes as well as over-expression of *baz* results in reduction of tumour size. Furthermore, incubation of embryos for 48h or 72h at 20°C before being shifted to 29°C did not result in obvious differences in the average tumour size.

In a final experiment, I combined the single fly counting method with an initial 48h incubation of embryos at low temperature (Figure 25). Statistical analysis shows that knock-down of *CG13559*, *Gα73B*, *mthl3* and over-expression of *baz* rescue the leukaemia phenotype. These results suggest that CG13559, Gα73B, Mthl3 act as positive regulators and Baz as a negative regulator of blood development. In contrast to the previous screens however, knock-down of *CG4793*, *CG6014*, *CG15211*, *CG3829* and *l(2)gl* did not significantly reduce tumour size in this experiment. A summary of the results from all leukaemia assays is depicted in Table 9.

JAK/STAT target genes which reduce haemocyte proliferation

As described earlier, melanotic tumours observed in flies carrying the oncogenic hop^{Tuml} allele are a result of over-proliferation of blood cells (haemocytes) which engulf self-tissue. In addition, at high temperatures, the more strongly signalling hop^{Tuml} mutation drives haemocyte differentiation into lamellocytes, a process which is also observed during immune response to septic injury (Agaisse and Perrimon 2004). Lamellocytes are almost never found in healthy flies.

In order to establish that the tumour reduction, by knock-down of putative STAT92E targets, is due to reduced haemocyte numbers, I examined the blood content in the hemolymph of late 3rd instar larvae. As in the previous experiments, crosses were set up at 20°C. Embryos were kept for 48h at 20°C before being transferred to 29°C. For each experiment the average haemocyte number of 6-20 late wandering 3rd instar larvae were examined using a hemocytometer (Figure 26). Knock-down of *CG13559* and *Gα73B* and over-expression of *Baz* reduced the total number of haemocytes. These results are consistent with the reduction in tumour size observed in previous experiments. Interestingly, knock-down of *CG6014* and *Wnt4* also reduced total blood count, although these genes did not consistently affect tumour growth (Table 9).
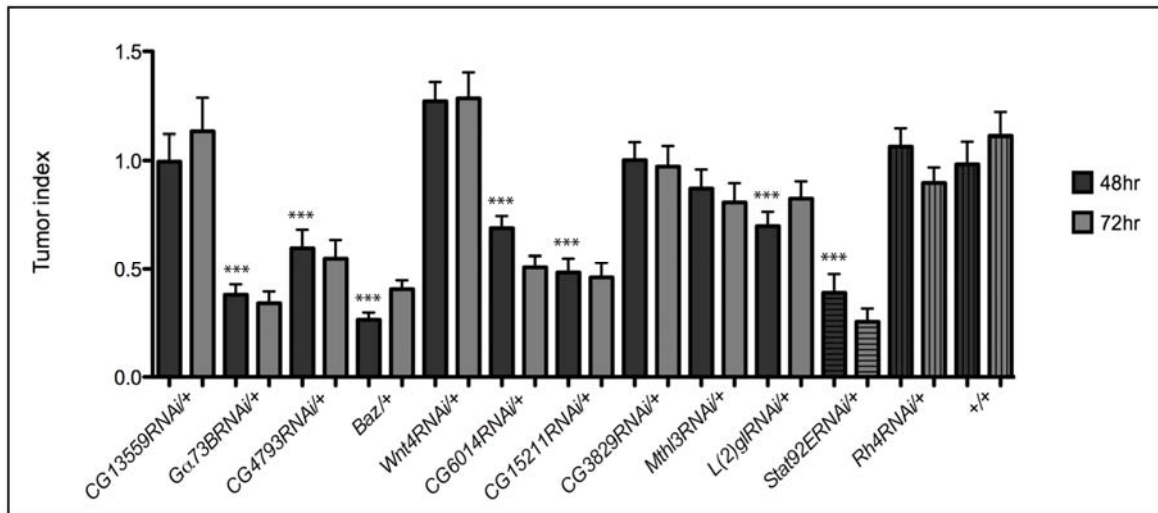
Figure 24. Leukaemia assay, Single fly III with incubations at varying temperatures.

Overactivation of the temperature-sensitive hop^{Tuml} allele may lead to loss of flies (due to large tumour formation) before they reach eclosure. Therefore offspring embryos were kept at low temperature (20°C) for 48h or 72h before being shifted to 29°C for full hop^{Tuml} and *GAL4/UAS* activity. Comparison of the results between the 48h or 72h experiments show that there is no clear difference in the tumour size of the progenies. T-test analyses of tumour sizes were performed on the data generated from the 48h experiments. Results from RNAi crosses were compared to the *rh4* RNAi knock-down results. *rh4* knock-down was used to asses the general effect of RNAi on tumour size. Results obtained from the *Baz*-overexpression experiment were compared to the *wt* (+/+) control cross. ***=p< 0.0005, **=p<0.005, *=p<0.05
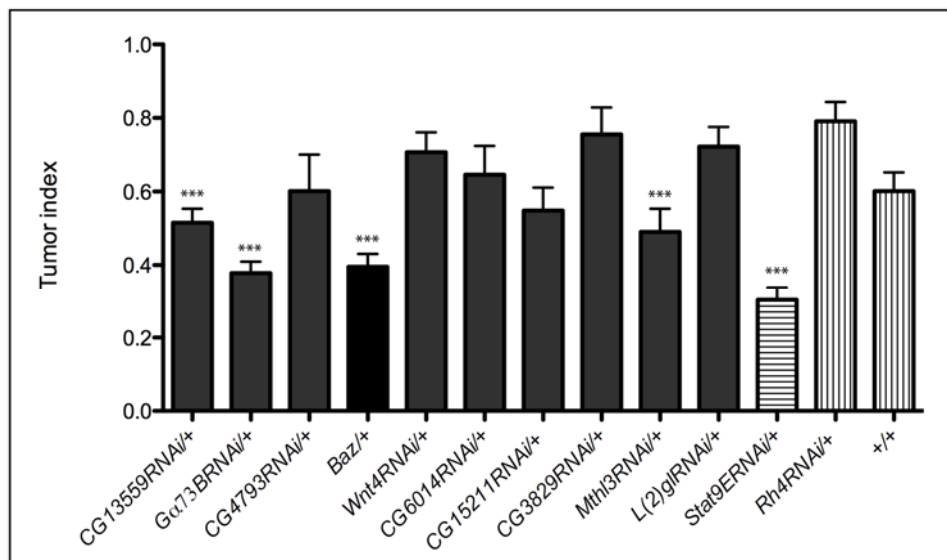
Figure 25. Leukaemia assay, Single fly IV with incubation at varying temperatures.

A final leukaemia assay was performed with the lines that yielded the most consistent results. Here the singly fly counting system was combined with the 48h incubation at 20°C. ***=p< 0.0005, **=p<0.005, *=p<0.05

Summary of leukaemia assays

Gene	Genotype I Figure 20	Genotype II Figure 21	Single fly I Figure 22	Single fly II Figure 23	Single fly III Figure 24	Single fly IV Figure 25
stat92E	+	+	+	+	+	+
CG13559	+	+	+	+	-	+
Ga73B	+	+	+	+	+	+
CG4793	+	+	+	+	+	-
Idgf1	+	+	-			
TotA	+	-	+			
CG17064	+	-				
CG15211				+	+	-
CG3829				+	-	-
Wnt4				+	-	-
CG6014				+	+	-
mthl3				+	-	+
baz				+	+	+
l(2)gl				+	+	-
CG15678	-					
CG1572	-					
pxb	-					
CG31158	-					
CG4080	-					
CG4804	-					
mfas	-					
net	-					

Table 9.   Summary of all leukaemia screens.
+: knock-down of gene reduced tumour size. -: knock-down of gene did not reduce tumour size

A

B

hop^{Tuml}, dome-Gal4/+; +/+   hop^{Tuml}, dome-Gal4/+; UAS-Baz/+
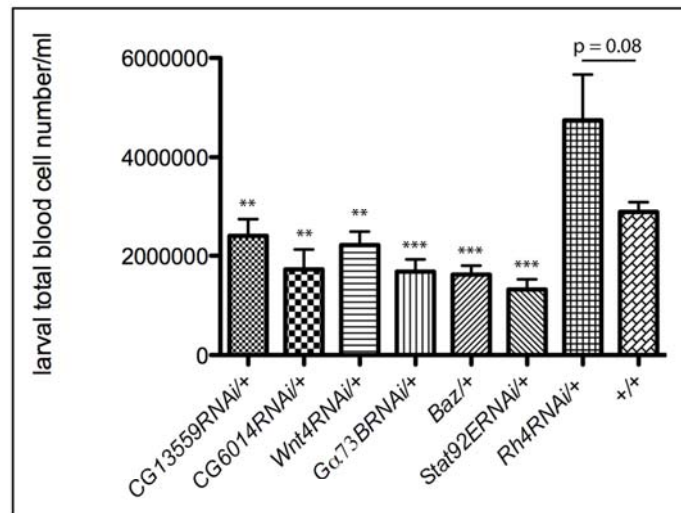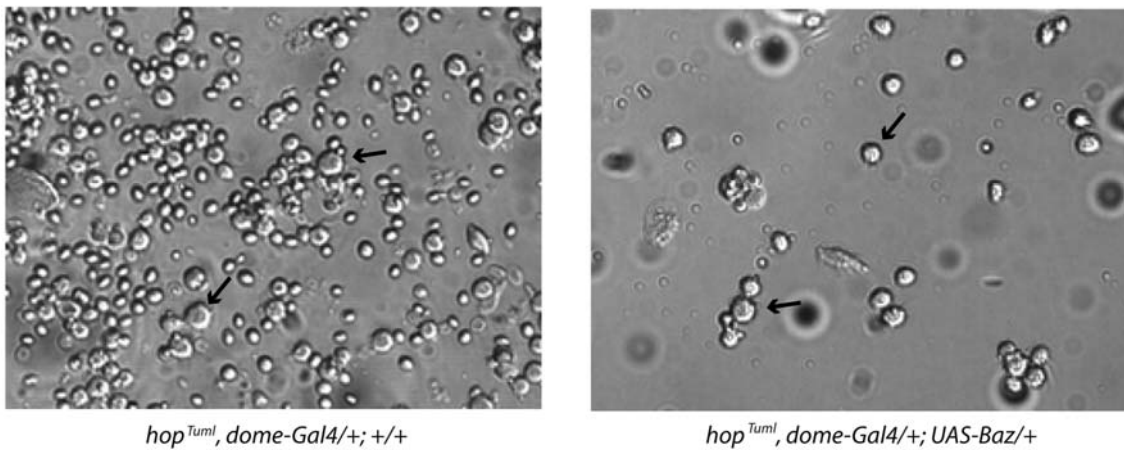
Figure 26.   Haemocyte counts.
A) *hop^{Tuml},dome-Gal4* females were crossed to *UAS-RNAi* or *UAS-Baz* lines. The offspring 3rd instar larvae were examined for total blood cell number. *Rh4RNAi* was used as control for knock-down experiments, whereas wt (+/+) cross was used as control for *UAS-Baz* overexpression. B) Example of enriched plasmatocytes (arrows) in *hop^{Tuml},dome-Gal4* flies. Baz over-expression in these flies reduced the haemocyte proliferation.

Taken together, transcript profiling of JAK/STAT activated *Drosophila* cells led to the identification of novel JAK/STAT target genes. Functional analysis of selected candidate genes *in vivo*, by means of RNAi-mediated knock-down or GAL4/UAS mediated-overexpression, demonstrates these genes to be involved in haematopoiesis and tumour formation in the fly. *Gα73B* was consistently identified as a positive mediator of JAK/STAT signalling that affects both JAK/STAT dependent tumour formation and blood cell proliferation. Although less consistent, eight further genes were also validated in their role in mediating JAK/STAT-induced tumourigenesis (Table 9). In addition, BAZ was shown to be a negative regulator of blood cell development and JAK/STAT signalling.

Processes of haemocytes overproliferation and inappropriate lamellocyte differentiation in *hop^{Tuml}* mutant flies can be tracked by morphological changes within the different zones of the lymph gland, where pro-haemocytes (medullary zone) and differentiated haemocytes (cortical zone) are maintained (Sorrentino et al. 2002). Future experiments should therefore examine the morphology of lymph glands in *hop^{Tuml}* mutants in which JAK/STAT target gene expression has been modulated.

DISCUSSION

Analysis of knock-out and gain-of-function mutations of JAK/STAT pathway components in mammalian and *Drosophila* models has shed light on the causes of disease pathology and progression that are associated with the mis-regulation of the pathway. Recently, genome-wide RNAi screens and classical genetic screens have identified several additional loci which regulate the cascade. Furthermore, the crystal structures of several STATs in various conformations have been solved and will hopefully enable more efficient design of pharmacological inhibitors. However, a major question remains regarding how the JAK/STAT pathway regulates gene expression, and more intriguingly, which of these targets contribute to diseases.

Design of screen

In order to identify JAK/STAT pathway target genes I used *Drosophila* haemocyte-like Kc$_{167}$ cells treated with conditioned media containing UPD, UPD2 or no ligand (Mock) which were generated in large batches and used for all experiments. This ensured minimal technical variation and enabled better comparison across all experiments. Furthermore, rather than activating the JAK/STAT pathway continuously by ectopic ligand expression, a 30min induction of the pathway was envisaged, in order to mimic a more physiological transitory stimulus. The chosen time frame for pathway stimulation is in accordance with previous studies of STAT3-dependent enhanceosome assembly on the promoter, and time course analysis of gene transcription (Lerner et al. 2003). Tyrosine-phosphorylated STAT3 was shown to accumulate in the nucleus within 7.5min, which almost disappeared by 45min. The gene expression, however, was elevated at 45min and required 90-120min to achieve maximal signal. In this study, the UPD- or UPD2-induced expression pattern of *socs36E*, a known JAK/STAT pathway target gene, was assessed. Results similar to those for STAT3-induced gene expression were obtained, whereby *socs36E* signal was elevated 30min after first addition of conditioned media, reaching its maximum between 2-4h before declining by 10h. Based on these results, transcript profiling of UPD, UPD2 or Mock treated cells was performed at 2h, 4h and 10h time points. Considering the increase of *socs36E* RNA level after 30min, it is possible that some direct STAT92E target genes may also be regulated very early and their transcription may have returned by 2h. Such a set of genes may therefore not have been detected in this particular transcriptome analysis. However, *socs36E* is also known to be an early response target gene and its expression

remained elevated at 2h and even 10h after pathway stimulation. Nevertheless, in retrospect, a screen with higher time resolution starting at an earlier time point would have ensured the detection of very early response genes which are most likely to be directly regulated by STAT92E.

UPD or UPD2 stimulation results in different transcriptional output

The RNA expression levels of UPD or UPD2 regulated genes were assessed using Affymetrix *Drosophila* GeneChip® 2.0 arrays and the data were analysed by the PUMA tool. In contrast to most publicly available programs, PUMA integrates signals generated from all probe sets that represent a specific gene and uses the inherent signal variation to estimate the confidence with which a given level of gene expression can be assessed. Due to this, PUMA can more accurately predict the presence of low abundance target genes.

Comparison of the transcriptome regulated by UPD or UPD2 shows that despite both ligands signalling via a linear cascade containing non-redundant components, their gene expression output is strikingly different. Whereas addition of UPD results in both up- and down-regulation of genes, UPD2 stimulation almost exclusively resulted in the up-regulation of gene transcription. A similar observation of differential downstream readout has been observed in macrophages that were stimulated with IL-10 or IL-6, despite the fact that both ligands activate an identical JAK1-STAT3 pathway. A distinction between the different signals is that, while the IL-6 receptor (IL-6R) is negatively regulated by SOCS3, the IL-10R is not, resulting in a prolonged IL-6 stimulation compared to IL-10 (Yasukawa et al. 2003). As a consequence, the temporal activation profile of STAT3 signal could potentially be the factor determining the gene expression output. A similar scenario could also be true for UPD and UPD2, where a stronger interaction of one ligand with the receptor results in a more prolonged STAT92E activation, ultimately leading to a different transcriptional output. We now know that in contrast to UPD2, the highly glycosylated UPD ligand associates strongly to the ECM, which may change its affinity for the DOME receptor (Gilbert et al. 2005; Hombria et al. 2005). Supporting evidence from the mammalian field suggest that differences in critical amino acids between different subtypes of IFNs can result in different ligand-receptor interactions and downstream responses, although the IFNs subtypes activate the same cell-surface receptor (van Boxel-Dezaire et al. 2006). The identification of such critical amino acid differences, possibly

responsible for the distinct responses mediated by UPD or UPD2, is an interesting subject that needs to be pursued.

JAK/STAT stimulation down-regulates a large number of genes

Given that UPD is the main signalling molecule required for the known JAK/STAT associated developmental roles, downstream analyses were focused on target genes regulated by UPD. In order to ensure the validity of the microarray data, I confirmed the up- and down-regulation, as well as the temporal expression pattern, of a sub-set of target genes by real-time PCR. Furthermore, the dynamics of the UPD-regulated transcriptome was examined using hierarchical clustering methods, which groups genes based on their common expression pattern over time. The expression of most genes appeared to be tightly regulated such that few genes overlapped between more than one time point. As previously mentioned, the temporal gap between each RNA isolation step of JAK/STAT activated cells was rather large such that genes with highly variable expression pattern may have been missed with the experimental design used. Another explanation could be that proteins which elicit the main biological functions of JAK/STAT signalling are tightly regulated at the transcriptional level. Such a regulation is expected for genes whose mis-regulation could lead to cytotoxic and pathological effects (Kim et al. 2007).

UPD stimulation resulted in a strong down-regulation of genes at all three time points, particularly at 10h (Figures 12 & 13). This finding was unexpected, as the phosphorylated active form of *Drosophila* STAT has previously only been known to act as a transcriptional activator and inducer of cellular proliferation (Arbouzova and Zeidler 2006). This work has for the first time revealed the ability of the *Drosophila* JAK/STAT pathway to repress target gene expression. In the mammalian field, the suppressor activity of the STAT1 transcription factor is well known. Transcriptionally active STAT1 induced by interferons (IFN) is required to restrain cellular growth (Bromberg et al. 1996) with STAT1-null mice displaying enhanced tumourigenesis (Kaplan et al. 1998). The IFNγ-mediated anti-proliferative effect of STAT1 is partly due to down-regulation of growth promoting cell cycle genes (c-*myc*, *cyclin D*, c*yclinA*) and *matrix metalloproteinases* (*MMP*s), which facilitate tumour invasion (Ramana et al. 2000). Interestingly, phosphorylation of both the tyrosine and the serine residues of STAT1 is required to suppress c-*myc* transcription. The transcriptional suppression of many STAT1 target genes often requires the presence of GAS elements in their promoter region. GAS elements are

however not needed for the *cyclinA* suppression, suggesting that its suppression involves binding of other repressors to silencer elements. To date, co-repressors that interact with STAT1 or their consensus binding sites have not been identified. Although the *Drosophila* STAT bears less similarity to STAT1 on the sequence level, functionally it seems to be able to mediate similar inhibitory roles.

The unphosphorylated form of STAT92E has recently been shown to bind HP1 and thereby stabilise the formation of heterochromatin (Shi et al. 2006; Shi et al. 2008). This mode of indirect gene repression represents a non-canonical function of STAT92E and is actively prevented by upstream JAK/STAT pathway activation. As a consequence, it is unlikely to account for the down-regulated genes in UPD stimulated cells. The anti-proliferative role of STAT92E has been shown in *Drosophila* wing discs, specifically at late larval stages (Mukherjee et al. 2005). However, it is not clear if this function can be attributed to the down-regulation of growth-promoting genes, or in fact, to its newly identified non-canonical function. Finally, the *Dictyostelium* STAT has been shown to act as a repressor protein by directly binding to repressor elements present in the promoters of genes that regulate cell differentiation (Kawata et al. 1997; Mohanty et al. 1999). Given its transcriptional repressor role, it is conceivable that in higher organisms this ancient function has been taken over by co-repressors, which bind to specific repressor *cis*-elements in the promoter of STAT target genes, possibly in concert with STAT molecules themselves. Considering that the repressor role of activated STAT has been repeatedly reported in various model organisms, future experiments should attempt to reveal the conserved *cis*-regulatory elements in the promoter region of genes repressed following activation of the (*Drosophila*) JAK/STAT pathway.

The JAK/STAT pathway controls the expression of humoral immune response genes

In order to bring biological relevance to the transcripts that are regulated by UPD-induced JAK/STAT pathway, genes that were co-clustered into groups, due to their common expression pattern, were examined for their biological and molecular function. The most enriched biological function was the gene ontology of defense response. This category encompasses genes that play a role in inflammation, innate immunity as well as general defense response to foreign bodies or the occurrence of an injury. It is well established that the JAK/STAT pathway plays an important and conserved role in immune responses

(Shuai and Liu 2003; Agaisse and Perrimon 2004). In *Drosophila*, induction of JAK/STAT signalling in the fat body (the functional equivalent to the human liver) by UPD3 secreting haemocytes causes secretion of antimicrobial peptides upon septic injury. These peptides elicit the main humoral immune response. There are seven classes of antimicrobial peptides (Attacin, Cecropin, Defensin, Diptericin, Drosocin, Drosomycin and Metchnikowin) and the level of their respective transcripts is also regulated by the NF-κB-like transcriptional regulators Dif and Relish. Four of these classes (Attacin, Cecropin, Drosocin and Metchnikowin) are, however, down-regulated in Kc$_{167}$ cells 4h after UPD stimulation. This result is consistent with recent data showing that STAT92E is found in a repressosome complex that binds to conserved *cis*-elements and inhibits NF-κB-mediated transcriptional activation of a subset of immune effector genes (Kim et al. 2007). Taken together, these data suggest that the situation may be more complex, with JAK/STAT pathway regulating the synthesis of antimicrobial peptides. It is speculated that this may be a feedback response to suppress the cytotoxic effects that can be caused by high levels of humoral response. It should be noted that this assumption is based on results gained from tissue culture experiments. It would be interesting to further investigate the relevance of these findings in the *in vivo* scenario, particularly in the *Drosophila* fat body, where these peptides are produced. Intriguingly, in the mammalian system, the negative regulation of inflammation by STAT3 is controlled indirectly (Murray 2005).

Two members of the *tot* family, namely *totA* and *totM* were up-regulated upon JAK/STAT stimulation in Kc$_{167}$ cells. TOT family members do not share similarities with known proteins in the database and therefore no function can be assigned to these proteins. Nevertheless, their members were found to be released from the fat body upon septic injury and their expression has previously been shown to be controlled by the JAK/STAT pathway *in vivo* (Agaisse and Perrimon 2004).

Another gene ontology class which has repeatedly been suggested to be involved in immune response (Wertheim et al. 2005; Schlenke et al. 2007), is the functional category of proteolytic genes. In this work, proteolytic genes are particularly up-regulated 4h after JAK/STAT stimulation (Figure 17, cluster 2). Activation of the proteolytic cascade is ultimately required for coagulation and melanisation processes as well as synthesis of humoral peptides.

Clustering of genes based on similarity of their expression pattern is assumed to reflect similar biological function. However, clustering does not prove such relationships but

rather serves as a helpful guide for further research. Ultimately, the validity of the observations made in this section must be tested *in vivo*, particularly focusing on their actual cellular role rather than predicted molecular functions by gene ontology.

Differential STAT92E binding sites in JAK/STAT target genes

Quantification of the number of STAT92E binding sites regulating JAK/STAT pathway target genes revealed that promoters of up-regulated genes show enriched numbers of STAT92E binding sites when compared to a random set of *Drosophila* genes (Figure 18A). At the time of designing the screen, evidence generated within the lab (in collaboration with James Castelli-Hombria, Seville) showed that STAT92E not only binds to the originally identified N_3 consensus site (TTCNNNGAA) but it also shows weaker affinity to N_4 sites that have been found in the first intron region of the JAK/STAT pathway receptor *dome* (Rivas et al. 2008). Given that this site was not identified earlier via the stringent *in vitro* SELEX method (Yan et al. 1996) it was assumed that stronger JAK/STAT signalling may be required to activate genes that contain N_4 sites in their promoter regions. This could be achieved by modulating the concentration of the ligand UPD. In a temporal context, genes enriched with N_4 sites could be expressed with the first wave of early response genes, when the STAT activity is highest.

Focusing on up-regulated genes, the high affinity N_3 binding site was enriched in their promoter region at all three time points when compared to randomly selected *Drosophila* genes. N_4 binding sites, however, were only enriched in promoters of genes induced at 2h and were less enriched in promoters of later induced genes (Figure 8). This observation supports the hypothesis that high concentration of activated STAT92E is required to bind to the low affinity N_4 binding site in order to induce low affinity target genes whereas high affinity target genes containing N_3 binding site require less STAT92E activity in their promoter region. In light of the current knowledge, one could imagine an *in vivo* scenario where promoters containing low affinity N_4 sites are only expressed in regions adjacent to a localised source of ligand where high pathway activity is ensured. Conversely, genes with promoters that contain N_3 sites can be expressed further away from UPD producing cells. The ability of a ligand to induce a gradient of pathway induction leading to differential gene expression is a key definition of a morphogen. In order to address this, one could examine the expression pattern of JAK/STAT target genes in tissues where a localised source of *upd* expression is known (for example in wing discs or hindgut). In

addition, by performing *in silico* sequence comparison of the identified N_4 binding sites and their flanking regions, one could compute an ideal STAT92E N_4 binding site. This sequence would represent a valuable tool with which the activity of the JAK/STAT pathway can be detected *in vivo* and in cell-based reporter assays.

The analysis of enhancers in various model organisms suggest that transcription factor binding sites are clustered in the genome in 300bp to 1kb regions (Markstein and Levine 2002; Ochoa-Espinosa et al. 2005). Simple searches for clusters of STAT92E binding sites in the putative promoter regions of up-regulated genes have generated a list of genes which contain such clusters and are more likely to be direct targets of STAT92E (Figure 18C). Interrogating this list with sophisticated comparative genomics tools that study the conservation of such clustered sites across different species will enable a more reliable prediction of direct STAT92E target genes.

In addition to promoter analyses, the distribution of co-regulated genes in the *Drosophila* genome was assessed. Although most genes were fairly equally distributed in the genome, the X-chromosome contained fewer up-regulated genes at 10h (Table 6, cluster 3), but it was more enriched in genes that are down-regulated at the same time point (Table 6, cluster 6). This observation could be associated to the sex-specific role of UPD involved in the sex determination of *Drosophila* and ultimately the control of dosage compensation of the X chromosome (Jinks et al. 2000; Sefton et al. 2000).

In vivo validation of JAK/STAT target genes

The significance of pathway target genes in their role *in vivo* is the ultimate approach of demonstrating the validity of the transcript profiling assay performed. Due to the ability of the microarray technique to asses genome-wide gene expression, a large amount of data is generated. In order to reduce the list of candidate target genes to a number that is feasible for *in vivo* validation, the 1197 UPD-regulated genes were systematically screened for their predicted function as well as their profile of STAT92E binding sites in their promoter region. In addition, the list of genes was compared to previously published microarray data that focused on the identification of genes that play a role during haematopoiesis and immune response (Agaisse et al. 2003; Asha et al. 2003; Wertheim et al. 2005; Kwon et al. 2008). As a result, a total of 21 genes were selected for examination of their role in modulating the *hop^{Tuml}*-induced tumour phenotype (Table 8). The majority of the selected genes have also predicted homology with genes that are associated with human diseases

(Chien et al. 2002). Therefore, it would be interesting to test their functionality in human JAK/STAT signalling. Human homologues of *Drosophila* JAK/STAT pathway regulators have previously been validated in this laboratory using HeLa-based assays (Müller et al. 2005).

For this study, the role of the JAK/STAT pathway during haematopoiesis was used as a tool to validate pathway target genes. However, the target gene list is also a potentially valuable resource for other researchers investigating JAK/STAT signalling in other contexts. One example might be studies on its role during cell rearrangement and elongation processes that are involved in the hindgut development. In order to efficiently select differentially regulated genes for validation in this tissue, one can query the list of UPD-regulated genes with publicly available *in situ* hybridisation data and focus on the study of genes that are expressed in the hindgut.

In vivo RNAi as a tool to investigate the role of JAK/STAT pathway target genes in haematopoiesis

The GAL4/UAS system was used to modulate gene expression in tissues that express the JAK/STAT pathway receptor *dome*. The *dome-gal4* reporter has previously been shown to activate *uas-gfp* in the medullary zone of the lymph gland where non-differentiated haemocytes are maintained and where the JAK/STAT pathway is active during haemocyte development. Overexpression of *baz* and knock-down of several (Table 9) genes reduced hop^{Tuml} induced tumours in adult flies as well as circulating blood cell numbers in the hemolymph of larvae, suggesting a role for these genes in haemocyte proliferation and/or differentiation. Specifically for RNAi constructs whose expression did not modulate the hop^{Tuml} phenotype, either the target genes do not modulate tumourigenesis, or the RNAi knock-down was not effective. However, it is difficult to interpret negative results from *in vivo* RNAi experiments. The mechanisms of genes whose depletion has an effect on tumourigenesis is potentially more tractable and can further be dissected in detail using alternative haemocyte drivers that are specific to the various blood cell types. Markers that have been shown to be restricted to the cortical zone, due to their expression in mature haemocytes, and for which GAL4 overexpression constructs exist, include Collagens and Hemolectin (Jung et al. 2005). The *upd3-gal4* and *dorothy-gal4* reporters are almost exclusively expressed in the posterior signalling centre (PSC) (Jung et al. 2005). However, the PSC itself does not respond to JAK/STAT signalling since it lacks the expression of the

receptor *dome*. In addition, PSC cells do not contribute to haemocytes. Therefore, the use of PSC specific driver lines is less likely to shed light on the role of JAK/STAT pathway target genes in haematopoiesis.

A potential concern with RNAi knock-down in the lymph gland is the general or haemocyte specific requirement of the affected genes for cell viability. If this were the case, reduction in the lymph gland size due to reduced haemocyte survival would reduce tumour growth in hop^{Tuml} flies, an effect which could no longer be solely attributed to the role of JAK/STAT target gene in tumourigenesis. However, the *dome-gal4* reporter is also active and expressed in multiple tissues throughout embryogenesis and larval life. Given that knock-down of the selected putative STAT92E target genes did not seem to affect the viability of the flies, it seems unlikely that the transcripts are required for general cell viability. Furthermore, the examination of genes for their requirement in Kc_{167} cell viability (Sims et al. 2006) indicated that their knock-down did not effect growth. In order to rule out their lymph gland-specific necessity, the apoptosis rate of haemocytes within the lymph gland from flies where the genes have been knocked down by *dome-gal4* should be examined.

The practicality and the speed with which *in vivo* RNAi assays can be performed makes this system particularly attractive when compared to classical genetic screens. However, there are several caveats that must be considered, including off-target effects (due to non-specific knockdown of genes with homology to the dsRNA), incomplete knock-down and variability of transgene expression. Recent software tools have been developed in order to identify off-target effects (Arziman et al. 2005). The majority of dsRNA sequences that were used in this assay only targeted the transcript of interest. For those that targeted additional gene transcripts, the number of predicted off-targets was limited to two. In order to examine the degree of transcript knock-down, one can perform quantitative PCR analysis. However, it is ultimately the protein half-live which determines the efficiency of RNAi. It should also be noted that for some genes tested here, more than one RNAi line was tested in the haematopoiesis experiments. In some cases only one line was able to modulate the tumour phenotype. Such variability is common in RNAi experiments (H. Strutt, personal communication) and is most likely due to the different integration sites of the *UAS-RNAi* constructs which influence the knock-down efficiency. Recent site-specific integration methods have been developed and could remedy this problem in the future (Bischof et al. 2007; Ni et al. 2008).

JAK/STAT target genes regulating cell-cell affinity

The importance of the JAK/STAT pathway in pathological diseases that involve tumour formation and metastatic migration is recognised in the vertebrate and invertebrate field. Cell rearrangements, including detachment from neighbouring cells, migratory behaviour and re-attachment all facilitate tumour metastasis. The movement of cells ultimately depends on changes in cell polarity which is in turn controlled by signalling pathways. The regulation of cell polarity is particularly important in epithelial tissues. Carcinomas (malignant tumours of epithelial origin) are characterised by major disorganisation of cell architecture. It is as yet not known whether loss of epithelial organisation is indeed the cause of tumourigenesis. During *Drosophila* development several morphological processes (i.e. border cell migration during oogenesis, hindgut elongation, spiracle cell elongation and germ cell migration) require the components of the JAK/STAT signalling cascade (Hombria and Sotillos 2008). However, the downstream STAT92E target genes involved in these cellular movements have not been reported yet.

In this work, WNT4, LGL and BAZ were examined for their role in haematopoiesis and have previously been linked to cell polarity (Bilder 2004; Macara 2004; Nelson and Nusse 2004). BAZ and LGL control apical-basal polarisation in epithelial cells and their localisation is mutually exclusive whereby BAZ is at the apical domain of the epithelial cell and LGL at the basal-lateral domain. Jointly with other complexes BAZ directs the maturation of epithelial cell polarity. Epithelial cells are attached to each other by adherens junctions. In *Drosophila*, adherens junctions are constituted of E-cadherin whose accumulation requires BAZ (Harris and Peifer 2004). Furthermore, the medullary zone (MZ) of the lymph gland (a mesodermally derived tissue) has been defined on the basis of its high levels of *E-cadherin* expression (Jung et al. 2005), suggesting that adherens junctions may be important for MZ maintenance. In mammals, E-cadherin has been shown to be a tumour suppressor whose loss enables the transition from benign lesion to invasive, metastatic cancer (Jeanes et al. 2008). Interestingly, in epithelial cells that form the hindgut in *Drosophila*, STAT92E has been shown to localise apically, a process which does not depend on JAK/STAT pathway activity but rather on the presence of BAZ. In addition, BAZ-dependent apical localisation of STAT92E increases the efficiency of STAT92E translocation to the activated receptor DOME (Sotillos et al. 2008). *baz* expression has not been examined in the compact mesodermal cells that form the medullary zone (MZ) of the lymph gland. However, *E-cadherin* is strongly expressed in the MZ, though it is not known

whether it is required in the formation of adherens junctions in this tissue and whether this process is dependent on the presence of BAZ (as is the case in epithelial cells). Given the interaction of *baz* overexpression with the *hop^{Tuml}* induced tumour phenotype, it is tempting to speculate that this effect is due to the reduction of haemocyte proliferation in the MZ. The interplay between BAZ and STAT92E as well as E-cadherin in mesodermal cells specifically in the lymph gland is therefore an interesting issue that needs to be addressed.

LGL has been shown to be required for localisation of adherens junctions in order to organise the epithelial architecture in embryos. Wing discs of *lgl* mutants lack proper epithelial structure and are massively over grown, underling the tumour suppressor role of LGL (Woods and Bryant 1989; Manfruelli et al. 1996; Bilder et al. 2000). This data is in contrast to the results obtained in this study where knockdown of *lgl* inhibited the growth of melanotic tumours in the mesoderm-derived lymph gland. The basis of this interaction remains to be determined.

Pathways that are used by WNT proteins are known to facilitate cell movement and planar cell polarity. However, no ligand has yet been conclusively associated with these processes (Nelson and Nusse 2004). Strikingly, the key component of the canonical signalling pathway β-Catenin also functions with the cadherin complex to control cell migration. Recent data show that the canonical Wnt pathway activated by WNT4 is needed for proper migration of early stage salivary gland cells (Harris and Beckendorf 2007). To date, the requirement of the JAK/STAT pathway in salivary gland tissues has not been shown. However, it has been shown that the JAK/STAT pathway receptor DOME is pre-dimerised in this tissue (Brown et al. 2003), a pre-requisite for pathway signalling that is also found in tissues known to require JAK/STAT pathway, such as the trachea, posterior spiracles and the hindgut. *wnt4* has also been shown to be up-regulated in stem cells of the testis where the JAK/STAT pathway has been ectopically activated (Terry et al. 2006). UPD signalling from the niche is required to maintain the stem cell identity of the surrounding cells, and the fact that *wnt4* was found to be up-regulated in these cells as well as in Kc$_{167}$ cells activated by UPD, again suggests that *wnt4* is a downstream pathway target gene. Given the detection of *wnt4* in this screen as a putative JAK/STAT target gene, the fact that its knock down is sufficient to reduce haemocyte proliferation and that it is required for migratory processes of salivary gland cells, it would be interesting to elucidate its interaction with the JAK/STAT pathway in cell rearrangement processes.

Although the validation of JAK/STAT pathway target genes (this work) and regulators (Müller et al. 2005; Kwon et al. 2008; Shi et al. 2008) in the mesodermally derived haemocytes and lymph gland has served as a valuable tool, the majority of potentially JAK/STAT-mediated migratory roles occur in ectodermally derived tissues (border cells, hindgut, trachea, posterior spiracle). As such, the relationship between the roles of JAK/STAT signalling in these two germ layers represents an important direction for future research.

JAK/STAT target genes that are less well studied in their biological function

Knock-down of *Gα73B*, *CG4793*, *CG13559* and *Mthl3* reduced tumour growth in *hop*^{Tuml} adult flies. Furthermore, knock-down of the first three genes also reduced haemocyte proliferation in 3rd instar *hop*^{Tuml} larvae.

CG4793 is situated in a 30Kb intronic region of *beat-Ib* and contains nine putative STAT92E binding sites in the 3Kb region upstream of its transcriptional start site. It is a member of the serine proteases (SPs) and serine protease homologues (SPHs), which constitute the second largest family of genes in the *Drosophila* genome (Ross et al. 2003). During immune response, the Toll signalling cascade is required to fight against fungi and gram-positive bacterial infections. The Toll receptor is activated by binding of a cleaved form of the cytokine-like molecule Spatzle (Spz) which is thought to be processed by SPs. A large-scale *in vivo* RNAi screen has been performed aiming at identifying SPs which could be responsible for the cleavage of Spz (Kambris et al. 2006). CG4793 was among the 75 SP genes which were examined, however, the depletion of its transcript by RNAi did not impair the immune response to bacterial infection, suggesting that it does not play a role in the Toll-mediated immunity. CG4793 has also been shown to be a common target gene in *hop*^{Tuml} as well as *NURF* mutant flies (Kwon et al. 2008). As pointed out in the introduction, NURF is part of a chromatin remodelling complex which alters chromatin structure and regulates transcription. NURF physically interacts with the JAK/STAT signalling modulator KEN in order to repress the transcription of a subset of STAT92E target genes. *NURF* mutants induce melanotic tumours (Badenhorst et al. 2002). Given that CG4793 transcript levels were increased in UPD-stimulated Kc$_{167}$ cells as well as in *hop*^{Tuml} and *NURF* mutants, it is highly likely to be a direct JAK/STAT pathway target genes, and its role in JAK/STAT-mediated immune response should be further investigated.

mthl3 encodes an atypical seven-transmembrane domain (7TM) protein which is a member of the large superfamily of G protein-coupled receptors (GPCRs) (Brody and Cravchik 2000). Furthermore, it is one of the 10 *methuselah-like* genes which are related to *mth* in the *Drosophila* genome. *mth* mutants show increased life span and enhanced resistance to stress (Lin et al. 1998). One of the *mth-like* genes, namely *mthl8*, has previously been identified as a modulator of the JAK/STAT pathway in the eye and has been suggested as a putative pathway target gene (Mukherjee et al. 2006). It is worth noting that *mthl3* and *CG10764* (both of which were up-regulated in UPD-treated cells and have been tested for their role in haematopoiesis) are situated next to each other in the genome. There are 10 putative STAT92E binding sites within this small region, suggesting that these sites are likely to induce their concomitant expression. The ability of these STAT92E binding sites to activated reporters such as Renilla luciferase in tissue culture experiments should be tested.

Gα73B is a member of the Gα protein family, which along with other other heterotrimeric (β and γ) G proteins mediate G-protein signalling. As yet, no functional role has been assigned to Gα73b, but it was shown to be primarily expressed in the mesoderm during embryonic, larval and early pupal stages and at low levels in adult flies (Quan et al. 1993).

SUMMARY AND CONCLUSIONS

During the development of all multicellular organisms the repeated used of a limited number of signalling cascades is required to generate an adult from a fertilised egg. The Mis-regulation of such pathways, including the JAK/STAT, however, can lead to disease. Therefore, an understanding of the consequences of aberrant signalling represents the foundation upon which applied research can develop compounds for treatment of many human diseases. Overactivation of the JAK/STAT signalling cascade results in the development of solid tumours such as breast cancer, head and neck cancer, melanoma as well as blood tumours including acute and chronic leukaemia and lymphomas. Although mutations of JAK/STAT pathway components underlying these phenotypes are known, the downstream effectors responsible for cellular overproliferation are not. In this study a *Drosophila* cell line which resembles blood cells, has been used as a tool in order to identify JAK/STAT pathway target genes that are regulated by the pathway ligands UPD and UPD2. Although both ligands signal through a common linear signalling cascade, their transcriptional output was unexpectedly different, an observation, which needs to be addressed in future experiments. Focusing on candidate target genes downstream of the main pathway ligand UPD, genes were examined for their temporal expression pattern, their common functionality as well as *cis*-regulatory elements in their promoter region which may be responsible for concomitant regulation of co-expressed genes. Novel putative pathway target genes which have previously not been described during haematopoiesis (a process which requires JAK/STAT signalling) have been validated in their role in "fly leukaemia" as well as blood cell proliferation. A significant proportion of these genes have homology with genes that are associated with human diseases.

Through the identification and analysis of JAK/STAT pathway target genes in *Drosophila*, this work has generated a wealth of data that can be used to get one step closer to understanding the downstream mechanisms of the JAK/STAT pathway and their role in mediating human diseases.

REFERENCES

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, et al. (2000). The genome sequence of *Drosophila melanogaster*. Science 287(5461): 2185-95.

Agaisse, H. and N. Perrimon (2004). The roles of JAK/STAT signaling in *Drosophila* immune responses. Immunol Rev 198: 72-82.

Agaisse, H., U. M. Petersen, M. Boutros, B. Mathey-Prevot and N. Perrimon (2003). Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury. Dev Cell 5(3): 441-50.

Ali, S., Z. Nouhi and N. Chughtai (2003). SHP-2 regulates SOCS-1-mediated Janus kinase-2 ubiquitination/degradation downstream of the prolactin receptor. J Biol Chem 278(52): 52021-31.

Arbouzova, N. I., E. A. Bach and M. P. Zeidler (2006). Ken & barbie selectively regulates the expression of a subset of Jak/STAT pathway target genes. Curr Biol 16(1): 80-8.

Arbouzova, N. I. and M. P. Zeidler (2006). JAK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. Development 133(14): 2605-16.

Arziman, Z., T. Horn and M. Boutros (2005). E-RNAi: a web application to design optimized RNAi constructs. Nucleic Acids Res 33(Web Server issue): W582-8.

Asha, H., I. Nagy, G. Kovacs, D. Stetson, I. Ando and C. R. Dearolf (2003). Analysis of Ras-induced overproliferation in *Drosophila* hemocytes. Genetics 163(1): 203-15.

Ausubel, F. M., R. Brent, R. T. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl, eds. (1999). Current Protocols in Molecular Biology. John Wiley & Sons, Inc.

Bach, E. A., S. Vincent, M. P. Zeidler and N. Perrimon (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(3): 1149-66.

Badenhorst, P., M. Voas, I. Rebay and C. Wu (2002). Biological functions of the ISWI chromatin remodeling complex NURF. Genes Dev 16(24): 3186-98.

Baeg, G. H., R. Zhou and N. Perrimon (2005). Genome-wide RNAi analysis of JAK/STAT signaling components in *Drosophila*. Genes Dev 19(16): 1861-70.

Beermann, W. (1952). [Chromomore constancy and specific modifications of the chromosome structure in development and organ differentiation of Chironomus tentans.]. Chromosoma 5(2): UNKNOWN.

Begitt, A., T. Meyer, M. van Rossum and U. Vinkemeier (2000). Nucleocytoplasmic translocation of Stat1 is regulated by a leucine-rich export signal in the coiled-coil domain. Proc Natl Acad Sci U S A 97(19): 10418-23.

Beissbarth, T. and T. P. Speed (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics 20(9): 1464-5.

Bilder, D. (2004). Epithelial polarity and proliferation control: links from the *Drosophila* neoplastic tumor suppressors. Genes Dev 18(16): 1909-25.

Bilder, D., M. Li and N. Perrimon (2000). Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. Science 289(5476): 113-6.

Bina, S., M. Zeidler (2009). JAK/STAT pathway signalling in *Drosophila melanogaster*. JAK-STAT pathway in disease. A. Stephanou. Austin, Landes Bioscience.

Bischof, J., R. K. Maeda, M. Hediger, F. Karch and K. Basler (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104(9): 3312-7.

Boulay, J. L., J. J. O'Shea and W. E. Paul (2003). Molecular phylogeny within type I cytokines and their cognate receptors. Immunity 19(2): 159-63.

Bourbon, H. M., G. Gonzy-Treboul, F. Peronnet, M. F. Alin, C. Ardourel, C. Benassayag, D. Cribbs, J. Deutsch, P. Ferrer, M. Haenlin, et al. (2002). A P-insertion screen identifying novel X-linked essential genes in *Drosophila*. Mech Dev 110(1-2): 71-83.

Brand, A. H. and N. Perrimon (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118(2): 401-15.

Brierley, M. M., K. L. Marchington, I. Jurisica and E. N. Fish (2006). Identification of GAS-dependent interferon-sensitive target genes whose transcription is STAT2-dependent but ISGF3-independent. Febs J 273(7): 1569-81.

Briggs, R. and T. J. King (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc Natl Acad Sci U S A 38(5): 455-63.

Brivanlou, A. H. and J. E. Darnell, Jr. (2002). Signal transduction and the control of gene expression. Science 295(5556): 813-8.

Brody, T. and A. Cravchik (2000). *Drosophila melanogaster* G protein-coupled receptors. J Cell Biol 150(2): F83-8.

Bromberg, J. F., C. M. Horvath, Z. Wen, R. D. Schreiber and J. E. Darnell, Jr. (1996). Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. Proc Natl Acad Sci U S A 93(15): 7673-8.

Brown, S., N. Hu and J. C. Hombria (2001). Identification of the first invertebrate interleukin JAK/STAT receptor, the *Drosophila* gene domeless. Curr Biol 11(21): 1700-5.

Brown, S., N. Hu and J. C. Hombria (2003). Novel level of signalling control in the JAK/STAT pathway revealed by in situ visualisation of protein-protein interaction during *Drosophila* development. Development 130(14): 3077-84.

Brown, S. and M. P. Zeidler (2008). Unphosphorylated STATs go nuclear. Curr Opin Genet Dev 18(5): 455-60.

Callus, B. A. and B. Mathey-Prevot (2002). SOCS36E, a novel *Drosophila* SOCS protein, suppresses JAK/STAT and EGF-R signalling in the imaginal wing disc. Oncogene 21(31): 4812-21.

Chen, H. W., X. Chen, S. W. Oh, M. J. Marinissen, J. S. Gutkind and S. X. Hou (2002). 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(3): 388-98.

Chen, X., U. Vinkemeier, Y. Zhao, D. Jeruzalmi, J. E. Darnell, Jr. and J. Kuriyan (1998). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell 93(5): 827-39.

Chien, S., L. T. Reiter, E. Bier and M. Gribskov (2002). Homophila: human disease gene cognates in *Drosophila*. Nucleic Acids Res 30(1): 149-51.

Cho, R. J., M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart, et al. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2(1): 65-73.

Clark, D. E., C. C. Williams, T. T. Duplessis, K. L. Moring, A. R. Notwick, W. Long, W. S. Lane, I. Beuvink, N. E. Hynes and F. E. Jones (2005). ERBB4/HER4 potentiates STAT5A transcriptional activity by regulating novel STAT5A serine phosphorylation events. J Biol Chem 280(25): 24175-80.

Clark, T. A., C. W. Sugnet and M. Ares, Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296(5569): 907-10.

Cohen, B. A., R. D. Mitra, J. D. Hughes and G. M. Church (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet 26(2): 183-6.

Crozatier, M. and M. Meister (2007). *Drosophila* haematopoiesis. Cell Microbiol 9(5): 1117-26.

Cui, Y., A. Hosui, R. Sun, K. Shen, O. Gavrilova, W. Chen, M. C. Cam, B. Gao, G. W. Robinson and L. Hennighausen (2007). Loss of signal transducer and activator of transcription 5 leads to hepatosteatosis and impaired liver regeneration. Hepatology 46(2): 504-13.

Damjanovich, S., L. Bene, J. Matko, A. Alileche, C. K. Goldman, S. Sharrow and T. A. Waldmann (1997). Preassembly of interleukin 2 (IL-2) receptor subunits on resting Kit 225 K6 T cells and their modulation by IL-2, IL-7, and IL-15: a fluorescence resonance energy transfer study. Proc Natl Acad Sci U S A 94(24): 13134-9.

Darnell, J. E., Jr. (1997). STATs and gene regulation. Science 277(5332): 1630-5.

Decker, T. and P. Kovarik (2000). Serine phosphorylation of STATs. Oncogene 19(21): 2628-37.

Dietzl, G., D. Chen, F. Schnorrer, K. C. Su, Y. Barinova, M. Fellner, B. Gasser, K. Kinsey, S. Oppel, S. Scheiblauer, et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 448(7150): 151-6.

Dostert, C., E. Jouanguy, P. Irving, L. Troxler, D. Galiana-Arnoux, C. Hetru, J. A. Hoffmann and J. L. Imler (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *drosophila*. Nat Immunol 6(9): 946-53.

Echalier, G. and A. Ohanessian (1970). In vitro culture of *Drosophila melanogaster* embryonic cells. In Vitro 6(3): 162-72.

Ehret, G. B., P. Reichenbach, U. Schindler, C. M. Horvath, S. Fritz, M. Nabholz and P. Bucher (2001). DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. J Biol Chem 276(9): 6675-88.

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-8.

Fortier, E. and J. M. Belote (2000). Temperature-dependent gene silencing by an expressed inverted repeat in *Drosophila*. Genesis 26(4): 240-4.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10): R80.

Gerstein, M. and R. Jansen (2000). The current excitement in bioinformatics-analysis of whole-genome expression data: how does it relate to protein structure and function? Curr Opin Struct Biol 10(5): 574-84.

Gilbert, M. M., B. K. Weaver, J. P. Gergen and N. C. Reich (2005). A novel functional activator of the *Drosophila* JAK/STAT pathway, unpaired2, is revealed by an *in vivo* reporter of pathway activation. Mech Dev 122(7-8): 939-48.

Gluecksohn-Schoenheimer, S. (1938). The Development of Two Tailless Mutants in the House Mouse. Genetics 23(6): 573-84.

Hanratty, W. P. and C. R. Dearolf (1993). The *Drosophila* Tumorous-lethal hematopoietic oncogene is a dominant mutation in the hopscotch locus. Mol Gen Genet 238(1-2): 33-7.

Hanratty, W. P. and J. S. Ryerse (1981). A genetic melanotic neoplasm of *Drosophila melanogaster*. Dev Biol 83(2): 238-49.

Harbers, M. and P. Carninci (2005). Tag-based approaches for transcriptome research and genome annotation. Nat Methods 2(7): 495-502.

Harris, K. E. and S. K. Beckendorf (2007). Different Wnt signals act through the Frizzled and RYK receptors during *Drosophila* salivary gland migration. Development 134(11): 2017-25.

Harris, T. J. and M. Peifer (2004). Adherens junction-dependent and -independent steps in the establishment of epithelial cell polarity in *Drosophila*. J Cell Biol 167(1): 135-47.

Harrison, D. A., P. E. McCoon, R. Binari, M. Gilman and N. Perrimon (1998). *Drosophila* unpaired encodes a secreted protein that activates the JAK signaling pathway. Genes Dev 12(20): 3252-63.

Haspel, R. L. and J. E. Darnell, Jr. (1999). A nuclear protein tyrosine phosphatase is required for the inactivation of Stat1. Proc Natl Acad Sci U S A 96(18): 10188-93.

Hennighausen, L. and G. W. Robinson (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. Genes Dev 22(6): 711-21.

Henriksen, M. A., A. Betz, M. V. Fuccillo and J. E. Darnell, Jr. (2002). Negative regulation of STAT92E by an N-terminally truncated STAT protein derived from an alternative promoter site. Genes Dev 16(18): 2379-89.

Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 7(3): 200-10.

Hombria, J. C. and S. Brown (2002). The fertile field of *Drosophila* Jak/STAT signalling. Curr Biol 12(16): R569-75.

Hombria, J. C., S. Brown, S. Hader and M. P. Zeidler (2005). Characterisation of Upd2, a *Drosophila* JAK/STAT pathway ligand. Dev Biol 288(2): 420-33.

Hombria, J. C. and S. Sotillos (2008). Disclosing JAK/STAT links to cell adhesion and cell polarity. Semin Cell Dev Biol 19(4): 370-8.

Horvath, C. M. (2000). STAT proteins and transcriptional responses to extracellular signals. Trends Biochem Sci 25(10): 496-502.

Hou, S. X., Z. Zheng, X. Chen and N. Perrimon (2002). The Jak/STAT pathway in model organisms: emerging roles in cell movement. Dev Cell 3(6): 765-78.

Hou, X. S., M. B. Melnick and N. Perrimon (1996). Marelle acts downstream of the *Drosophila* HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell 84(3): 411-9.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf and T. P. Speed (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2): 249-64.

James, C., V. Ugo, J. P. Le Couedic, J. Staerk, F. Delhommeau, C. Lacout, L. Garcon, H. Raslova, R. Berger, A. Bennaceur-Griscelli, et al. (2005). A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 434(7037): 1144-8.

Jeanes, A., C. J. Gottardi and A. S. Yap (2008). Cadherins and cancer: how does cadherin dysfunction promote tumor progression? Oncogene 27(55): 6920-9.

Jinks, T. M., A. D. Polydorides, G. Calhoun and P. Schedl (2000). The JAK/STAT signaling pathway is required for the initial choice of sexual identity in *Drosophila melanogaster*. Mol Cell 5(3): 581-7.

Johansen, K. A., D. D. Iwaki and J. A. Lengyel (2003). Localized JAK/STAT signaling is required for oriented cell rearrangement in a tubular epithelium. Development 130(1): 135-45.

Jones, D. L. and A. J. Wagers (2008). No place like home: anatomy and function of the stem cell niche. Nat Rev Mol Cell Biol 9(1): 11-21.

Jung, S. H., C. J. Evans, C. Uemura and U. Banerjee (2005). The *Drosophila* lymph gland as a developmental model of hematopoiesis. Development 132(11): 2521-33.

Kambris, Z., S. Brun, I. H. Jang, H. J. Nam, Y. Romeo, K. Takahashi, W. J. Lee, R. Ueda and B. Lemaitre (2006). *Drosophila* immunity: a large-scale *in vivo* RNAi screen identifies five serine proteases required for Toll activation. Curr Biol 16(8): 808-13.

Kaplan, D. H., V. Shankaran, A. S. Dighe, E. Stockert, M. Aguet, L. J. Old and R. D. Schreiber (1998). Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. Proc Natl Acad Sci U S A 95(13): 7556-61.

Karsten, P., I. Plischke, N. Perrimon and M. P. Zeidler (2006). Mutational analysis reveals separable DNA binding and trans-activation of *Drosophila* STAT92E. Cell Signal 18(6): 819-29.

Kawata, T., A. Shevchenko, M. Fukuzawa, K. A. Jermyn, N. F. Totty, N. V. Zhukovskaya, A. E. Sterling, M. Mann and J. G. Williams (1997). SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in *Dictyostelium*. Cell 89(6): 909-16.

Khatri, P., S. Draghici, G. C. Ostermeier and S. A. Krawetz (2002). Profiling gene expression using onto-express. Genomics 79(2): 266-70.

Khush, R. S. and B. Lemaitre (2000). Genes that fight infection: what the *Drosophila* genome says about animal immunity. Trends Genet 16(10): 442-9.

Kiger, A. A., D. L. Jones, C. Schulz, M. B. Rogers and M. T. Fuller (2001). Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. Science 294(5551): 2542-5.

Kim, L. K., U. Y. Choi, H. S. Cho, J. S. Lee, W. B. Lee, J. Kim, K. Jeong, J. Shim, J. Kim-Ha and Y. J. Kim (2007). Down-regulation of NF-kappaB target genes by the AP-1 and STAT complex during the innate immune response in *Drosophila*. PLoS Biol 5(9): e238.

Krzemien, J., L. Dubois, R. Makki, M. Meister, A. Vincent and M. Crozatier (2007). Control of blood cell homeostasis in *Drosophila* larvae by the posterior signalling centre. Nature 446(7133): 325-8.

Kuchinke, U., F. Grawe and E. Knust (1998). Control of spindle orientation in *Drosophila* by the Par-3-related PDZ-domain protein Bazooka. Curr Biol 8(25): 1357-65.

Kwon, E. J., H. S. Park, Y. S. Kim, E. J. Oh, Y. Nishida, A. Matsukage, M. A. Yoo and M. Yamaguchi (2000). Transcriptional regulation of the *Drosophila* raf proto-oncogene by *Drosophila* STAT during development and in immune response. J Biol Chem 275(26): 19824-30.

Kwon, S. Y., H. Xiao, B. P. Glover, R. Tjian, C. Wu and P. Badenhorst (2008). The nucleosome remodeling factor (NURF) regulates genes involved in *Drosophila* innate immunity. Dev Biol 316(2): 538-47.

Lam, G. and C. S. Thummel (2000). Inducible expression of double-stranded RNA directs specific genetic interference in *Drosophila*. Curr Biol 10(16): 957-63.

Lerner, L., M. A. Henriksen, X. Zhang and J. E. Darnell, Jr. (2003). STAT3-dependent enhanceosome assembly and disassembly: synergy with GR for full transcriptional increase of the alpha 2-macroglobulin gene. Genes Dev 17(20): 2564-77.

Levine, A. J. (1997). p53, the cellular gatekeeper for growth and division. Cell 88(3): 323-31.

Levy, D. E. and J. E. Darnell, Jr. (2002). Stats: transcriptional control and biological impact. Nat Rev Mol Cell Biol 3(9): 651-62.

Lin, Y. J., L. Seroude and S. Benzer (1998). Extended life-span and stress resistance in the *Drosophila* mutant methuselah. Science 282(5390): 943-6.

Liu, X., M. Milo, N. D. Lawrence and M. Rattray (2006). Probe-level measurement error improves accuracy in detecting differential gene expression. Bioinformatics 22(17): 2107-13.

Livak K. J. and T. D. Schmittgen (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25(4): 402-8.

Lopez-Onieva, L., A. Fernandez-Minan and A. Gonzalez-Reyes (2008). Jak/Stat signalling in niche support cells regulates dpp transcription to control germline stem cell maintenance in the *Drosophila* ovary. Development 135(3): 533-40.

Luo, H., W. P. Hanratty and C. R. Dearolf (1995). An amino acid substitution in the *Drosophila* hopTum-l Jak kinase causes leukemia-like hematopoietic defects. Embo J 14(7): 1412-20.

Luo, H., P. E. Rose, T. M. Roberts and C. R. Dearolf (2002). The Hopscotch Jak kinase requires the Raf pathway to promote blood cell activation and differentiation in *Drosophila*. Mol Genet Genomics 267(1): 57-63.

Macara, I. G. (2004). Parsing the polarity code. Nat Rev Mol Cell Biol 5(3): 220-31.

Mandal, L., J. A. Martinez-Agosto, C. J. Evans, V. Hartenstein and U. Banerjee (2007). A Hedgehog- and Antennapedia-dependent niche maintains *Drosophila* haematopoietic precursors. Nature 446(7133): 320-4.

Manfruelli, P., N. Arquier, W. P. Hanratty and M. Semeriva (1996). The tumor suppressor gene, lethal(2)giant larvae (1(2)g1), is required for cell shape change of epithelial cells during *Drosophila* development. Development 122(7): 2283-94.

Marcotte, E. M., M. Pellegrini, M. J. Thompson, T. O. Yeates and D. Eisenberg (1999). A combined algorithm for genome-wide prediction of protein function. Nature 402(6757): 83-6.

Markstein, M. and M. Levine (2002). Decoding cis-regulatory DNAs in the *Drosophila* genome. Curr Opin Genet Dev 12(5): 601-6.

McBride, K. M., C. McDonald and N. C. Reich (2000). Nuclear export signal located within theDNA-binding domain of the STAT1transcription factor. Embo J 19(22): 6196-206.

McDonald, M. J. and M. Rosbash (2001). Microarray analysis and organization of circadian gene expression in *Drosophila*. Cell 107(5): 567-78.

Meissner, T., E. Krause, I. Lodige and U. Vinkemeier (2004). Arginine methylation of STAT1: a reassessment. Cell 119(5): 587-9; discussion 589-590.

Mertens, C., M. Zhong, R. Krishnaraj, W. Zou, X. Chen and J. E. Darnell, Jr. (2006). Dephosphorylation of phosphotyrosine on STAT1 dimers requires extensive spatial reorientation of the monomers facilitated by the N-terminal domain. Genes Dev 20(24): 3372-81.

Mohanty, S., K. A. Jermyn, A. Early, T. Kawata, L. Aubry, A. Ceccarelli, P. Schaap, J. G. Williams and R. A. Firtel (1999). Evidence that the *Dictyostelium* Dd-STATa protein is a repressor that regulates commitment to stalk cell differentiation and is also required for efficient chemotaxis. Development 126(15): 3391-405.

Mowen, K. A., J. Tang, W. Zhu, B. T. Schurter, K. Shuai, H. R. Herschman and M. David (2001). Arginine methylation of STAT1 modulates IFNalpha/beta-induced transcription. Cell 104(5): 731-41.

Mukherjee, T., J. C. Hombria and M. P. Zeidler (2005). Opposing roles for *Drosophila* JAK/STAT signalling during cellular proliferation. Oncogene 24(15): 2503-11.

Mukherjee, T., U. Schafer and M. P. Zeidler (2006). Identification of *Drosophila* genes modulating janus kinase/signal transducer and activator of transcription signal transduction. Genetics 172(3): 1683-97.

Müller H. J. (1930). Types of visible variations induced by X-rays in *Drosophila*. J Genet 22(3): 299-334.

Müller, P., D. Kuttenkeuler, V. Gesellchen, M. P. Zeidler and M. Boutros (2005). Identification of JAK/STAT signalling components by genome-wide RNA interference. Nature 436(7052): 871-5.

Murray, P. J. (2005). The primary mechanism of the IL-10-regulated antiinflammatory response is to selectively inhibit transcription. Proc Natl Acad Sci U S A 102(24): 8686-91.

Murray, P. J. (2007). The JAK-STAT signaling pathway: input and output integration. J Immunol 178(5): 2623-9.

Naef, F., D. A. Lim, N. Patil and M. Magnasco (2002). DNA hybridization to mismatched templates: a chip study. Phys Rev E Stat Nonlin Soft Matter Phys 65(4 Pt 1): 040902.

Nelson, W. J. and R. Nusse (2004). Convergence of Wnt, beta-catenin, and cadherin pathways. Science 303(5663): 1483-7.

Ni, J. Q., M. Markstein, R. Binari, B. Pfeiffer, L. P. Liu, C. Villalta, M. Booker, L. Perkins and N. Perrimon (2008). Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. Nat Methods 5(1): 49-51.

Niehrs, C. and N. Pollet (1999). Synexpression groups in eukaryotes. Nature 402(6761): 483-7.

Ochoa-Espinosa, A., G. Yucel, L. Kaplan, A. Pare, N. Pura, A. Oberstein, D. Papatsenko and S. Small (2005). The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. Proc Natl Acad Sci U S A 102(14): 4960-5.

Parisi, M., R. Nuttall, P. Edwards, J. Minor, D. Naiman, J. Lu, M. Doctolero, M. Vainer, C. Chan, J. Malley, et al. (2004). A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. Genome Biol 5(6): R40.

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

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat Genet 32 Suppl: 496-501.

Quan, F., W. J. Wolfgang and M. Forte (1993). A *Drosophila* G-protein alpha subunit, Gf alpha, expressed in a spatially and temporally restricted pattern during *Drosophila* development. Proc Natl Acad Sci U S A 90(9): 4236-40.

Ramana, C. V., M. Chatterjee-Kishore, H. Nguyen and G. R. Stark (2000). Complex roles of Stat1 in regulating gene expression. Oncogene 19(21): 2619-27.

Rawlings, J. S., K. M. Rosler and D. A. Harrison (2004). The JAK/STAT signaling pathway. J Cell Sci 117(Pt 8): 1281-3.

Reich, N. C. and J. E. Darnell, Jr. (1989). Differential binding of interferon-induced factors to an oligonucleotide that mediates transcriptional activation. Nucleic Acids Res 17(9): 3415-24.

Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, et al. (2000). Genome-wide location and function of DNA binding proteins. Science 290(5500): 2306-9.

Rivas, M. L., L. Cobreros, M. P. Zeidler and J. C. Hombria (2008). Plasticity of *Drosophila* Stat DNA binding shows an evolutionary basis for Stat transcription factor preferences. EMBO Rep 9(11): 1114-20.

Ross, J., H. Jiang, M. R. Kanost and Y. Wang (2003). Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene 304: 117-31.

Rozen, S. and H. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-86.

Schindler, C. and C. Plumlee (2008). Inteferons pen the JAK-STAT pathway. Semin Cell Dev Biol 19(4): 311-8.

Schlenke, T. A., J. Morales, S. Govind and A. G. Clark (2007). Contrasting infection strategies in generalist and specialist wasp parasitoids of *Drosophila melanogaster*. PLoS Pathog 3(10): 1486-501.

Sefton, L., J. R. Timmer, Y. Zhang, F. Beranger and T. W. Cline (2000). An extracellular activator of the *Drosophila* JAK/STAT pathway is a sex-determination signal element. Nature 405(6789): 970-3.

Sehgal, P. B. (2008). Paradigm shifts in the cell biology of STAT signaling. Semin Cell Dev Biol 19(4): 329-40.

Shi, S., H. C. Calhoun, F. Xia, J. Li, L. Le and W. X. Li (2006). JAK signaling globally counteracts heterochromatic gene silencing. Nat Genet 38(9): 1071-6.

Shi, S., K. Larson, D. Guo, S. J. Lim, P. Dutta, S. J. Yan and W. X. Li (2008). *Drosophila* STAT is required for directly maintaining HP1 localization and heterochromatin stability. Nat Cell Biol 10(4): 489-96.

Shuai, K. and B. Liu (2003). Regulation of JAK-STAT signalling in the immune system. Nat Rev Immunol 3(11): 900-11.

Sims, D., B. Bursteinas, Q. Gao, M. Zvelebil and B. Baum (2006). FLIGHT: database and tools for the integration and cross-correlation of large-scale RNAi phenotypic datasets. Nucleic Acids Res 34(Database issue): D479-83.

Singh, S. R., W. Liu and S. X. Hou (2007). The adult *Drosophila* malpighian tubules are maintained by multipotent stem cells. Cell Stem Cell 1(2): 191-203.

Small, S., A. Blair and M. Levine (1996). Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. Dev Biol 175(2): 314-24.

Sorrentino, R. P., Y. Carton and S. Govind (2002). Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. Dev Biol 243(1): 65-80.

Sotillos, S., M. T. Diaz-Meco, J. Moscat and J. Castelli-Gair Hombria (2008). Polarized subcellular localization of Jak/STAT components is required for efficient signaling. Curr Biol 18(8): 624-9.

Spellman, P. T. and G. M. Rubin (2002). Evidence for large domains of similarly expressed genes in the *Drosophila* genome. J Biol 1(1): 5.

Staerk, J., A. Kallin, J. B. Demoulin, W. Vainchenker and S. N. Constantinescu (2005). JAK1 and Tyk2 activation by the homologous polycythemia vera JAK2 V617F mutation: cross-talk with IGF1 receptor. J Biol Chem 280(51): 41893-9.

Stahl, N., T. G. Boulton, T. Farruggella, N. Y. Ip, S. Davis, B. A. Witthuhn, F. W. Quelle, O. Silvennoinen, G. Barbieri, S. Pellegrini, et al. (1994). Association and activation of Jak-Tyk kinases by CNTF-LIF-OSM-IL-6 beta receptor components. Science 263(5143): 92-5.

Terry, N. A., N. Tulina, E. Matunis and S. DiNardo (2006). Novel regulators revealed by profiling *Drosophila* testis stem cells within their niche. Dev Biol 294(1): 246-57.

Thomas-Chollier, M., O. Sand, J. V. Turatsinze, R. Janky, M. Defrance, E. Vervisch, S. Brohee and J. van Helden (2008). RSAT: regulatory sequence analysis tools. Nucleic Acids Res 36(Web Server issue): W119-27.

Tulina, N. and E. Matunis (2001). Control of stem cell self-renewal in *Drosophila* spermatogenesis by JAK-STAT signaling. Science 294(5551): 2546-9.

Ungureanu, D., S. Vanhatupa, N. Kotaja, J. Yang, S. Aittomaki, O. A. Janne, J. J. Palvimo and O. Silvennoinen (2003). PIAS proteins promote SUMO-1 conjugation to STAT1. Blood 102(9): 3311-3.

van Boxel-Dezaire, A. H., M. R. Rani and G. R. Stark (2006). Complex modulation of cell type-specific signaling in response to type I interferons. Immunity 25(3): 361-72.

Van Etten, R. A. (2007). Aberrant cytokine signaling in leukemia. Oncogene 26(47): 6738-49.

Vandesompele, J., A. De Paepe and F. Speleman (2002). Elimination of primer-dimer artifacts and genomic coamplification using a two-step SYBR green I real-time RT-PCR. Anal Biochem 303(1): 95-8.

Vinkemeier, U., S. L. Cohen, I. Moarefi, B. T. Chait, J. Kuriyan and J. E. Darnell, Jr. (1996). DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. Embo J 15(20): 5616-26.

Waddington, C. H. (1939). Preliminary Notes on the Development of the Wings in Normal and Mutant Strains of *Drosophila*. Proc Natl Acad Sci U S A 25(7): 299-307.

Wang, R., P. Cherukuri and J. Luo (2005). Activation of Stat3 sequence-specific DNA binding and transcription by p300/CREB-binding protein-mediated acetylation. J Biol Chem 280(12): 11528-34.

Wang, Z., M. Gerstein and M. Snyder (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1): 57-63.

Wertheim, B., A. R. Kraaijeveld, E. Schuster, E. Blanc, M. Hopkins, S. D. Pletcher, M. R. Strand, L. Partridge and H. C. Godfray (2005). Genome-wide gene expression in response to parasitoid attack in *Drosophila*. Genome Biol 6(11): R94.

Wieschaus E., C. Nusslein-Volhard, G. Jurgens (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. III. Zygotic loci on the X-chromosome and fourth chromosome. Roux Arch Dev Biol 193: 296-307.

Wodarz, A., A. Ramrath, U. Kuchinke and E. Knust (1999). Bazooka provides an apical cue for Inscuteable localization in *Drosophila* neuroblasts. Nature 402(6761): 544-7.

Woods, D. F. and P. J. Bryant (1989). Molecular cloning of the lethal(1)discs large-1 oncogene of *Drosophila*. Dev Biol 134(1): 222-35.

Xu, C., R. C. Kauffmann, J. Zhang, S. Kladny and R. W. Carthew (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the *Drosophila* eye. Cell 103(1): 87-97.

Yamamoto, K., M. Yamaguchi, N. Miyasaka and O. Miura (2003). SOCS-3 inhibits IL-12-induced STAT4 activation by binding through its SH2 domain to the STAT4 docking site in the IL-12 receptor beta2 subunit. Biochem Biophys Res Commun 310(4): 1188-93.

Yan, R., S. Small, C. Desplan, C. R. Dearolf and J. E. Darnell, Jr. (1996). Identification of a Stat gene that functions in *Drosophila* development. Cell 84(3): 421-30.

Yang, E., M. A. Henriksen, O. Schaefer, N. Zakharova and J. E. Darnell, Jr. (2002). Dissociation time from DNA determines transcriptional function in a STAT1 linker mutant. J Biol Chem 277(16): 13455-62.

Yang, I. V., E. Chen, J. P. Hasseman, W. Liang, B. C. Frank, S. Wang, V. Sharov, A. I. Saeed, J. White, J. Li, et al. (2002). Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol 3(11): research0062.

Yasukawa, H., H. Misawa, H. Sakamoto, M. Masuhara, A. Sasaki, T. Wakioka, S. Ohtsuka, T. Imaizumi, T. Matsuda, J. N. Ihle, et al. (1999). The JAK-binding

protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop. Embo J 18(5): 1309-20.

Yasukawa, H., M. Ohishi, H. Mori, M. Murakami, T. Chinen, D. Aki, T. Hanada, K. Takeda, S. Akira, M. Hoshijima, et al. (2003). IL-6 induces an anti-inflammatory response in the absence of SOCS3 in macrophages. Nat Immunol 4(6): 551-6.

Yuan, Z. L., Y. J. Guan, D. Chatterjee and Y. E. Chin (2005). Stat3 dimerization regulated by reversible acetylation of a single lysine residue. Science 307(5707): 269-73.

Zeeberg, B. R., W. Feng, G. Wang, M. D. Wang, A. T. Fojo, M. Sunshine, S. Narasimhan, D. W. Kane, W. C. Reinhold, S. Lababidi, et al. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 4(4): R28.

Zeidler, M. P., N. Perrimon and D. I. Strutt (1999). Polarity determination in the *Drosophila* eye: a novel role for unpaired and JAK/STAT signaling. Genes Dev 13(10): 1342-53.

Zhang, J. J., Y. Zhao, B. T. Chait, W. W. Lathem, M. Ritzi, R. Knippers and J. E. Darnell, Jr. (1998). Ser727-dependent recruitment of MCM5 by Stat1alpha in IFN-gamma-induced transcriptional activation. Embo J 17(23): 6963-71.

CURRICULUM VITAE

Personal Data

Full name   Samira Bina
Date of birth   May 12, 1982
Place of birth   Tehran, Iran

Education

2005-present   **PhD Thesis in Dr. Martin P. Zeidler's laboratory,** Max Planck
Institute for biophysical chemistry, Göttingen, Germany and
University of Sheffield, Sheffield, UK

2004-2005   **International MSc/PhD Program Molecular Biology,** Georg-
August University Göttingen, Germany

2001-2004   **BSc in Molecular Biotechnology,** University of Heidelberg,
Germany

Awards

2008   travel award for symposium participation (EMBL PhD Student
Symposium)
2008   travel award for conference participation (International Max
Planck Research School)
2005   stipend from the International Max Planck Research School

Publications

Bina, S., V. M. Wright, M. P. Zeidler (2009). Transcriptional targets of *Drosophila* JAK/STAT pathway signalling as effectors of haematopoietic tumour formation. (In preparation).

Bina, S., M. Zeidler (2009). JAK/STAT pathway signalling in *Drosophila melanogaster*. JAK-STAT pathway in disease. A. Stephanou, Austin, Landes Bioscience.

Selected Presentations

Bina, S., M. Zeidler (2008). Identification of direct target genes of the JAK/STAT signalling pathway. 10th International EMBL PhD Student Symposium (Heidelberg, Germany).

Bina, S., M. Zeidler (2008). Identification of direct target genes of the JAK/STAT signalling pathway. 49th Annual *Drosophila* Research Conference (San Diego, USA).

Bina, S., M. Zeidler (2006). Identification of JAK/STAT pathway target genes induced by Upd and Upd-like ligands in a microarray-based assay. 47th Annual *Drosophila* Research Conference (Houston, USA).

APPENDIX

List of UPD-regulated genes