Identification of Heat Shock Factor Binding Sites in the
_Drosophila_ Genome

by

Sarah Elizabeth Gonsalves

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Cell & Systems Biology
University of Toronto

© Copyright by Sarah Elizabeth Gonsalves 2012
Identification of Heat Shock Factor Binding Sites in the Drosophila Genome

Sarah Elizabeth Gonsalves

Doctor of Philosophy

Cell & Systems Biology
University of Toronto

2012

Abstract

The heat shock response (HSR) is a highly conserved mechanism that enables organisms to survive environmental and pathophysiological stress. In Drosophila, the HSR is regulated by a single transcription factor, heat shock factor (HSF). During stress, HSF trimerizes and binds to over 200 loci on Drosophila polytene chromosomes with only nine mapping to major heat shock (HS) inducible gene loci. The function of HSF binding to the other sites in the genome is currently unknown. Some of these sites may contain yet unidentified “minor” HS genes.

Interestingly, the binding of HSF also coincides with puff regression at some sites. Two such sites contain the major developmentally regulated genes Eip74 and Eip75: key regulators in the response to 20-hydroxyecdysone (20E), the main hormone responsible for the temporal coordination of post-embryonic development in Drosophila. Previous work in our and other labs indicates that the regression of non-HS puffs during the HSR is dependent on the presence of functional HSF.

Using chromatin immunoprecipitation (ChIP) followed by hybridization to genome tiling arrays (Chip), I have identified 434 regions in the Drosophila Kc cell genome that are bound by HSF
during HS, and have determined that 57% of these sites are located within the transcribed regions of genes. By examining the transcriptional response to HS in Kc cells and third instar larvae using expression microarrays, I found that only about 10% of all genes within 1250 bp of an HSF binding site are transcriptionally regulated by HS and many genes whose transcript levels change during HS do not appear to be near an HSF binding site. Furthermore, genes with an HSF binding site within their introns are significantly enriched (modified Fisher Exact p-value between 2.0x10^{-3} and 1.5x10^{-6}) in gene ontology terms related to developmental processes and reproduction.

Using expression microarray technology, I characterized the transcriptional response to 20E and its structural analog ponasterone A. I have identified multiple HSF binding sites within Eip74 and Eip75, and show that induction of the HSR correlates with repression of these genes and all other 20E-inducible genes. Taken together, this work provides a basis for further investigation into the role of HSF binding to sites not associated with HS genes and its possible function as a repressor of gene transcription during conditions of stress and as a regulator of developmental genes under stress and non-stress conditions.
Acknowledgments

Thanks to Tim for your guidance and support and for having the confidence in me to work with the technologies that made this project possible. You have been a great teacher and mentor. Also to the other members of my committee, Dorothea and Henry, thanks for your help and advice. Thanks to Maria, Anna, and Kristin for keeping me sane. I owe you all a debit of gratitude as this was no easy feat! Thanks for listening to my many trials and tribulations. You have been the best lab mates I could have asked for. Thanks to Scott and Tony for getting me started in the lab. Your great lab habits have stuck with me to this day. Thanks Dahlia, Mandy, Micki, Mo, and all other past and present members of the Westwood lab and CDMC for your help throughout the years. Thanks to my parents, for your unconditional support and understanding and for providing me with the belief that I could be anything I wanted to be in life. Finally, thanks to Zak, without whom this would not have been possible. No words could ever describe how much it means to me that you have always been and continue to be there for me throughout everything.
# Table of Contents

Acknowledgments ........................................................................................................ iv

Table of Contents ........................................................................................................ v

List of Tables ................................................................................................................ ix

List of Figures ................................................................................................................ x

List of Appendices ......................................................................................................... xii

List of Abbreviations .................................................................................................... xiii

Chapter 1 General Introduction .................................................................................... 1

1.1 The heat shock response (HSR) ............................................................................. 1

1.2 Transcription during the HSR .............................................................................. 2

  1.2.1 Discovery of the heat shock proteins (HSPs) ................................................. 3

  1.2.2 Cloning of the HS genes .............................................................................. 8

  1.2.3 HS-induced gene repression ......................................................................... 8

1.3 Heat shock factor (HSF) and the heat shock element (HSE) ................................. 9

  1.3.1 Transcriptional regulation ............................................................................ 10

  1.3.2 Structure and regulation of HSF ................................................................... 15

  1.3.3 Heat shock elements (HSEs) ....................................................................... 18

  1.3.4 Transcriptional regulation by HSF ............................................................... 19

  1.3.5 Other roles of HSF ..................................................................................... 21

1.4 Genomic studies on the HSR .............................................................................. 24

  1.4.1 Identification of novel HS genes .................................................................. 24

  1.4.2 HSF binding sites ....................................................................................... 28

1.5 The ecdysone response ......................................................................................... 30

1.6 Aims of this thesis: ............................................................................................. 33
1.6.1 Aim I: Map Drosophila HSF binding sites genome-wide and identify novel heat-inducible genes .......................................................... 33

1.6.2 Aim II: Identify genes transcribed in Kc167 cells and salivary glands in response to 20E stimulation and examine how HS affects their expression. .......34

Chapter 2 Whole-Genome Analysis Reveals That Active Heat Shock Factor Binding Sites Are Mostly Associated with Non-Heat Shock Genes in Drosophila melanogaster .............36

2.1 Abstract ................................................................................ 36

2.2 Introduction .............................................................................. 37

2.3 Methods .................................................................................. 39

2.3.1 Cell culture and heat-shock treatments .......................................................... 39

2.3.2 ChIP .................................................................................. 39

2.3.3 End-point and qPCR ........................................................................ 41

2.3.4 Ligation-Mediated PCR ................................................................. 41

2.3.5 Indirect labeling of amplified chromatin and hybridization to genome-tiling arrays ........................................................................ 42

2.3.6 HSF binding site identification ...................................................... 43

2.3.7 PWM matching ....................................................................... 43

2.3.8 Larval heat shock treatment ............................................................. 44

2.3.9 RNA extraction cDNA synthesis, and labeling and hybridization to expression microarrays .......................................................... 44

2.3.10 Expression microarray data extraction and analysis .......................... 45

2.3.11 Functional Enrichment of Gene Lists ............................................. 45

2.4 Results ..................................................................................... 46

2.4.1 Identification of HSF binding sites in the Drosophila genome ................................. 46

2.4.2 Heat Shock Elements (HSEs) ................................................................ 53

2.4.3 Genes associated with HSF binding events ....................................... 56

2.4.4 Transcriptional profile of Kc cell and 3rd instar larvae in response to heat shock .......................................................... 61
Chapter 3 Genome-wide examination of the transcriptional response to ecdysteroids 20-hydroxyecdysone and ponasterone A in *Drosophila melanogaster* ............................................. 75

3.1 Abstract ........................................................................................................................................... 75

3.2 Introduction ...................................................................................................................................... 76

3.3 Methods .......................................................................................................................................... 78

3.3.1 Hormone and HS treatments ....................................................................................................... 78

3.3.2 RNA isolation and hybridization to cDNA arrays ...................................................................... 79

3.3.3 Microarray data acquisition, normalization and analysis ............................................................. 80

3.3.4 Quantitative RT-PCR .................................................................................................................. 80

3.3.5 Analysis of functional classes ...................................................................................................... 81

3.4 Results and Discussion .................................................................................................................. 82

3.4.1 Identification of 20-hydroxyecdysone (20E)-responsive genes in Kc167 cells ......................... 82

3.4.2 Identification of primary 20E-response genes ............................................................................. 83

3.4.3 Transcriptional response of Kc167 cells to an increased concentration of 20E and to its structural analog, ponasterone A (PoA) ......................................................... 87

3.4.4 Transcriptional response of salivary glands to 20E ................................................................. 90

3.4.5 Tissue specific differences in ecdysone receptor (EcR) isoform induction by 20E .................... 97

3.4.6 Heat shock represses the ecdysone-response ............................................................................. 100

3.5 Conclusions ..................................................................................................................................... 104

Chapter 4 General Discussion ........................................................................................................... 106

4.1 HSF binds to many sites not associated with HS-induced gene transcription ....................... 106

4.2 Possible explanations of HSF binding to sites that are apparently not transcriptionally active during HS ............................................................................................................. 107

4.3 Possible roles for HSF apart from being a transcriptional activator during the HSR ....... 111
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4 Perspectives</td>
<td>119</td>
</tr>
<tr>
<td>References</td>
<td>121</td>
</tr>
<tr>
<td>Appendix A</td>
<td>159</td>
</tr>
<tr>
<td>Appendix B</td>
<td>160</td>
</tr>
<tr>
<td>Copyright Acknowledgement</td>
<td>161</td>
</tr>
</tbody>
</table>
List of Tables

Table 2-1 Chromatin segments bound by HSF exhibiting a 30 or greater fold enrichment over whole cell extract (WCE). ................................................................. 52

Table 2-2 Genes regulated by heat shock. ........................................................................ 64

Table 2-3 Genes associated with HSF that are regulated by HS in either Kc cells, 3rd instar larvae or both. ............................................................... 67

Table 2-4 Occurrence of BEAF motifs and binding sites in HSF-bound promoters........... 69

Table 3-1 Primary and secondary 20E-regulated genes identified in Kc167 cells .......... 84

Table 3-2 Genes that respond similarly to 0.5 μM 20E, 10 μM 20E, or 0.0625 μM PoA. .... 89

Table 3-3 Annotation enrichment of genes regulated by either 0.5μM or 10μM 20E or 0.0625μM PoA ........................................................................................................ 92

Table 3-4 Annotation enrichment of genes that respond to 20E in either Kc167 cells or salivary glands .................................................................... 96

Table 4-1 Preliminary RNA-seq results for HS RNA. .................................................... 110
List of Figures

Figure 1-1 Representative arrangement of DNA sequence elements involved in the regulation of gene transcription. ................................................................. 11

Figure 1-2 Model for the transcriptional regulation of genes............................................. 13

Figure 1-3 Functional organization and activation of Drosophila HSF .............................. 16

Figure 1-4 The Ashburner model for gene regulation by ecdysone (20E)....................... 32

Figure 2-1 Confirmation of HSF binding to select regions............................................. 47

Figure 2-2 Overlap of HSF binding sites identified by ChIP-chip of Kc cells and immunostaining of polytene chromosomes of 3rd instar larvae.......................... 49

Figure 2-3 Comparison of HSF binding under HS and non-HS (RT) conditions.............. 51

Figure 2-4 Representation of select genomic regions bound by HSF........................... 54

Figure 2-5 The 5 bp motif nGAAAn arranged as direct inverted repeats is enriched in HSF-bound segments. ........................................................................................................ 55

Figure 2-6 Breakdown of HSF-binding site by location................................................. 57

Figure 2-7 Example of an intronic HSF-binding site....................................................... 57

Figure 2-8 Heat map summary of select categories from DAVID functional enrichment analysis. ........................................................................................................ 59

Figure 2-9 The BEAF cis regulatory motif is enriched in promoters, but not introns, bound by HSF.................................................................................................... 60

Figure 2-10 Many genes associated with HSF binding are not induced by HS in Kc cells or 3rd instar larvae............................................................................... 65

Figure 3-1 Effect of PoA and increased 20E concentration on gene regulation.............. 91
Figure 3-2 Overlap of genes regulated by 20E in salivary glands and whole organ culture...... 95

Figure 3-3 The transcriptional responses of Kc167 cells and salivary glands to 20E exhibit very little overlap ......................................................................................................................... 95

Figure 3-4 The transcription of EcR isoforms are differentially regulated in Kc cells and salivary glands ........................................................................................................................................................................ 98

Figure 3-5 Heat shock prevents transcription of ecdysone-regulated genes in Kc cells. ........ 101

Figure 3-6 HSF binds to introns in the three major ecdysone-inducible genes, Eip75B, Eip74EF, and br. .................................................................................................................................................................................................. 102

Figure 4-1 Preliminary evidence of possible HS-induced transcription near sites bound by HSF in Eip75 bound ...................................................................................................................................................................... 112

Figure 4-2 Possible models of genes repression by HSF ................................................................. 115

Figure 4-3 HSF binds within introns (and possibly promoters) of ecdysone-inducible genes during HS, coincident with their repression ........................................................................................................ 117
List of Appendices

Appendix A..................................................................................................................................................159

Appendix B..................................................................................................................................................160
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>20E</td>
<td>20-hydroxyecdysone</td>
</tr>
<tr>
<td>ChIP-chip</td>
<td>chromatin immunoprecipitation and microarray</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>chromatin immunoprecipitation and next generation sequencing</td>
</tr>
<tr>
<td>CHX</td>
<td>cycloheximide</td>
</tr>
<tr>
<td>GO</td>
<td>gene ontology</td>
</tr>
<tr>
<td>HS</td>
<td>heat shock (36.5°C)</td>
</tr>
<tr>
<td>HSE</td>
<td>heat shock element</td>
</tr>
<tr>
<td>HSF</td>
<td>heat shock factor</td>
</tr>
<tr>
<td>HSP</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>HSR</td>
<td>heat shock response</td>
</tr>
<tr>
<td>IL-6</td>
<td>Interleukin 6</td>
</tr>
<tr>
<td>IL-β</td>
<td>Interleukin beta</td>
</tr>
<tr>
<td>IP</td>
<td>immunoprecipitation</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>ncRNA</td>
<td>non-coding RNA</td>
</tr>
<tr>
<td>PoA</td>
<td>ponasterone A</td>
</tr>
<tr>
<td>qRTPCR</td>
<td>quantitative reverse transcriptase PCR</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>next generation sequencing of cDNA representing the transcriptome</td>
</tr>
<tr>
<td>RNPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature (22°C)</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td>TNF-α</td>
<td>tumor necrosis factor alpha</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
<tr>
<td>WCE</td>
<td>whole cell extract</td>
</tr>
</tbody>
</table>
Chapter 1
General Introduction

1.1 The heat shock response (HSR)

Every organism studied to date, from bacteria to man, has the capacity to cope with environmental stress. What is perceived as stress will differ from organism to organism and largely depends on the environment to which the organism is adapted. However the fundamental necessity of a requirement for dealing with stress is universal. As the requirement for coping with stress is universal, so too is the solution: the heat shock response (HSR).

The HSR provides a means of dealing with cellular stresses. Its success is illustrated by its high conservation throughout evolution. The HSR was first documented in 1962 by Ferruccio Ritossa who observed a novel yet highly reproducible puffing pattern in the salivary glands of Drosophila busckii resulting from exposure to thermal shock and an inhibitor of respiratory metabolism, 2,4-Dintrophemol (DNP) (Ritossa 1962). Unbeknownst at the time, these observations would spawn a field of study that included one of the most conserved molecular responses, some of the most highly conserved genes in evolution, and would eventually provide the field of molecular biology with an ideal system to examine the molecular mechanisms of transcription regulation among other things. Since Ritossa’s original observation, other environmental and pathophysiological stressors that lead to induction of the response have been identified such as ionic stress, heavy metals, amino acid analogs, cold stress, proteasome inhibitors, oxidative respiration inhibitors, fever, inflammation, ischemia, oxidative damage, neural injury and ageing. Regardless of the causative nature of the stress, activation of the HSR results in both the rapid up regulation of transcription and translation of a sub-set of genes, as well as the concurrent repression of transcription and translation of the majority of other genes and transcripts. Much work has been done towards understanding the consequences of HSR activation both at the cellular and gross organism levels and towards identifying and understanding the main function of the principle regulator of the response, the heat shock factor (HSF) as well as the main players called to action during the response, the heat shock proteins (HSPs). As we now approach the 50th anniversary of the discovery of the HSR a lot has been learned but still, much remains to be worked out.
1.2 Transcription during the HSR

The alteration in the puffing pattern observed by Ritossa in response to heat-shock included both the appearance of a small number of new puffs as well as the disappearance of the majority of previously formed puffs and suggested that certain changes in environmental conditions can have a profound effect on the transcriptional profile of organisms (Ritossa 1962; Ritossa 1964). Confirmation that HS puffs are sites of nascent transcription and that the regressing puffs are no longer transcriptionally active was originally provided by monitoring $[^3]$H uridine incorporation into nascent transcripts followed by autoradiography of polytene chromosomes (Ritossa 1964; Berendes 1968; Tissières et al. 1974; Belyaeva and Zhimulev 1976; Bonner and Pardue 1976). In addition to changes in the transcriptional activity at these sites, an increase in the dry mass of heat shock puff loci after temperature treatment was observed suggesting an accumulation of non-histone proteins at these sites (Holt 1970; Holt 1971). Indirect immunofluorescence later provided evidence that the protein accumulating at these loci following heat treatment was RNA polymerase II (RNPII) (Plagens et al. 1976; Jamrich et al. 1977; Greenleaf et al. 1978). These studies also revealed the simultaneous depletion of RNPII levels at the regressing non-HS puff sites following heat treatment by a general reduction in the fluorescence intensity due to RNPII at previously labeled sites (Jamrich et al. 1977; Greenleaf et al. 1978).

Acrylamide gel electrophoresis of labeled nascent RNA provided the first indication of the number of newly transcribed species following activation of the HSR (Spradling et al. 1977; Moran et al. 1978) and subsequent analysis of the in vitro translated polypeptides by trypsin digestion revealed some of them to be transcripts of major HSPs (Moran et al. 1978). Although the major species of RNA isolated from heat shocked cells in these and other studies hybridize to HS puffs, there are still some species of less abundant RNA that hybridized to some extent to at least 25 chromosomal sites that do not puff during HS suggesting that transcription of these genes somehow escapes the global repression seen for most non-HS genes (Spradling et al. 1977; Ashburner and Bonner 1979; Stevens 1999). It is possible that some of these genes serve essential functions, such as the histone genes that are continually synthesized during HS (Spradling et al. 1975; Spradling et al. 1977; Westwood et al. 1991), while others may be minor heat inducible genes perhaps serving similar functions or accessory functions to the proteins coded for by the major HS genes (Westwood et al. 1991).
1.2.1 Discovery of the heat shock proteins (HSPs)

HSPs were first described by Tissières and colleagues in 1974 who showed that the protein synthesis profiles of *Drosophila* salivary glands and other tissues changed drastically following HS in a manner reminiscent of the changes observed in the puffing pattern of polytene chromosomes. Heat shocked tissues produced a discrete set of new proteins, as seen by labeling cells with $^{35}$S-methionine followed by SDS gel electrophoresis and autoradiography, while the production of previously synthesized proteins was reduced or disappeared altogether (Tissières et al. 1974). The change in protein synthesis profile following HS is so dramatic that one third of all nascent polypeptides were resolved into only six bands on an acrylamide gel. Within the next few years, numerous investigators used this experimental approach to show that HSPs with similar molecular weights were synthesized during HS in essentially all eukaryotic organisms.

The major HSPs are grouped based on molecular weight and amino acid sequence homology (Feige and Polla 1994) and together constitute a family of molecular chaperones with roles in maintaining protein homeostasis and protein folding both under normal conditions and in the presence of stress. One reason for their strong induction during stress is related to their mode of action; unlike other enzymes, chaperones are required in stoichiometric ratios to substrates (Richter et al. 2010). In *Drosophila* the major HSPs consist of the HSP70 gene family, the HSP83 gene, and the small HSP gene family.

The HSP70 family is the largest and best studied of the chaperone families and the most conserved protein family throughout evolution (Hunt and Morimoto 1985; Lindquist and Craig 1988; Gupta and Singh 1994; Daugaard et al. 2007). Roughly 50% amino acid identity is shared between Eukaryotic and Prokaryotic HSP70s (named DnaK in prokaryotes) (Daugaard et al. 2007). Every Eukaryote studied to date has multiple *Hsp70* genes (Daugaard et al. 2007) and has both constitutively expressed and heat-inducible forms (Lindquist and Craig 1988). Under normal conditions, members of this family aid in the maturation of nascent polypeptides (Teter et al. 1999; Thulasiraman et al. 1999), protein translocation (Chirico et al. 1988; Deshaies et al. 1988), and the uncoating of clathrin-coated vesicles (Chappell et al. 1986). Under stressful conditions, they help prevent aggregate formation, aid in the refolding denatured proteins (Mayer and Bukau 2005), and target misfolded proteins for degradation (Nover and Scharf 2005).
1997; Young et al. 2004). All of these processes have in common ATP-dependent protein binding and release.

HSP70 proteins consist of a central peptide-binding domain (PBD), made up of a beta-sandwich sub-domain and an alpha helical lid segment, a highly conserved amino-terminal ATPase domain, and a carboxy-terminal region that regulates substrate release (Zhu et al. 1996; Daugaard et al. 2007; Stetler et al. 2010). A typical folding cycle initiates when the PBD beta-sandwich sub-domain recognizes and binds to exposed hydrophobic amino acid residues of an unfolded peptide (Rüdiger et al. 1997). HSP70 co-factors such as DnaJ (member of HSP40/J-domain protein family, the most common group of HSP70 co-factors) can aid in this process by binding to and delivering unfolded peptides to HSP70 and by stimulating HSP70’s ATPase activity, in the case of J-domain containing proteins, through interaction with their J-domains (Young et al. 2003; Mayer and Bukau 2005; Kampinga and Craig 2010; Richter et al. 2010).

In Drosophila melanogaster, there are seven constitutively expressed family members collectively known as the Heat Shock Cognates (Hsc70): Hsc70-1, Hsc70-2, Hsc70-3, Hsc70-4, Hsc70-5, Hsc70-6, and Hsc70Cb (Craig et al. 1983; Palter et al. 1986). The different forms of Hsc70 genes exhibit differing spatial and temporal expression profile, are developmentally regulated, and unlike other non-heat shock-inducible genes, escape repression during heat shock (Craig et al. 1983; Palter et al. 1986; Elefant and Palter 1999). Although they are transcribed during heat, levels of Hsc70 transcripts are not greatly increased, however, as seen for their heat-inducible counterparts, the Hsp70 genes. This is likely due to Hsc70’s already high levels in non-shock cells (Palter et al. 1986). In fact, Hsc70 proteins are so abundant that their basal levels in non-stressed cells exceed the maximum levels of heat-inducible HSP70 family members following induction by the stress response (Palter et al. 1986).

Heat-inducible members of the D. melanogaster HSP70 family are represented by five or six HSP70 proteins depending on strain: HSP70Aa, HSP70Ab, HSP70Ba, HSP70Bb, HSP70Bbb, and HSP70Bc plus the high homologous heat-inducible HSP68 protein (Neidhardt et al. 1984; Bettencourt and Feder 2002; Maside et al. 2002). Genes for the HSP70 proteins are clustered at two loci, 87A and 87B. The 87A locus contains Hsp70Aa and Hsp70Ab separated by roughly 1.7 kb of DNA and arranged as inverted repeats (Goldschmidt-Clermont 1980; Ish-Horowicz and Pinchin 1980). The 87B locus contains the other four genes with Hsp70Bb, Hsp70Bbb, and
Hsp70Bc occurring in tandem and separated from the divergently orientated Hsp70Ba by 35kb of DNA (Goldschmidt-Clermont 1980; Ish-Horowicz and Pinchin 1980; Lakhotia and Prasanth 2002). Interestingly, this 35kb of intervening sequence contains a heat-inducible, non-protein coding gene with a highly repetitive sequence known as the alpha-gamma repeat; the function of which has yet to be elucidated (Lindquist 1986; Lakhotia and Prasanth 2002). Hsp70 genes originating from the 87A and 87B loci exhibit differences both in their heat-shock-inducibility and post-transcriptional metabolism despite their highly homologous sequences (Lakhotia and Prasanth 2002). Unlike the Hsc70s, heat-inducible HSP70 members are typically not found, or found only in minute amounts, under non-shock conditions in Drosophila (Velazquez et al. 1983). Those Hsp70 transcripts that are found under non-shock conditions are usually either not expressed and/or rapidly degraded owing to the AU-rich 3' UTR of Hsp70 transcripts (Petersen and Lindquist 1989; Yost et al. 1990). During cellular stress, however, heat-inducible members of the HSP70 family are among the most strongly induced HSPs (Feige and Polla 1994; Lakhotia and Prasanth 2002). For example, Drosophila Hsp70 transcript levels have been observed to increase 1000 fold from a ten degree shift in temperature from 25°C to 35°C (Velazquez et al. 1983).

Drosophila HSP83 is a member of the second most highly conserved HSP family, the HSP90 family (Lindquist 1986). It shares 40% amino acid identity with the Escherichia coli HSP90 family member, HtpG and over 50% amino acid identity with the most distantly related eukaryotic HSP90 family member (Lindquist and Craig 1988). A single gene for Drosophila HSP83 is located at the 63BC locus (Holmgren et al. 1979; Lindquist 1986). In addition to being stress induced, Hsp83 is also developmentally regulated (Nathan and Lindquist 1997). HSP83 protein is present at high levels under non-shock conditions in all developmental stages in the insect and is one of the most abundant cytosolic proteins constituting approximately 1-2% of total protein (Csermely et al. 1998; Zhao et al. 2002). It is also the only major Drosophila HSP gene to contain an intron (Hackett and Lis 1983; Blackman and Meselson 1986). The expression profile of Hsp83 also differs from other HSPs. Firstly, Hsp83 is induced by a wider range of temperatures than other Hsps and is induced most strongly at moderate HS temperatures and most poorly by severe HS temperatures (Ashburner and Bonner 1979; Lindquist 1980). Secondly, Hsp83 transcripts are not correctly processed at high HS temperatures owing to a

Given the sequence conservation between the members of the HSP90 family, structural and functional characteristics of *Drosophila* HSP83 are largely inferred from studies of other eukaryotic HSP90 proteins including recently solved crystal structure of yeast HSP90 in complex with an ATP analogue and a co-chaperone (Ali et al. 2006). These studies have revealed that HSP90 family members function as homodimers (Carbajal et al. 1990) and that each protomer consists of three domains: an ATP-binding domain located at the N-terminus (N), a domain that regulates the N-terminal ATPase activity in the middle of the protein (M), and a dimerization domain located at the C-terminus (C) (Pearl and Prodromou 2006). The C-terminus also contains a highly conserved EEVD sequence that mediates interaction with tetratricopeptide repeat (TPR) containing co-chaperones (Scheufler et al. 2000). Some of these co-chaperones such as HSP70-HSP90 organizing protein (HOP) create a bridge between HSP90 and HSP70 to promote the transfer of partially folded substrates from HSP70 to HSP90 (Scheufler et al. 2000). Substrate binding in HSP90 is thought to occur at the C and M domains (Whitesell and Lindquist 2005). Unlike most other chaperones, HSP90 appears to be highly selective in substrate binding interacting primarily with proteins that have roles in cell cycle control or signal transduction for example, steroid hormone receptors (Picard 2002). This is likely because the major function of HSP90 involves the regulation of these substrate proteins by holding them in either active or inactivates states (Pratt and Toft 2003). For example, *Drosophila* Hsp83 is required as part of multi-chaperone containing complex (MCH) that also contains Hsc70 for ecdysone receptor activity *in vivo* (Arbeitman and Hogness 2000). In the absence of MCH, purified EcR/USP possesses ligand binding ability but lacks DNA binding ability which can be gained through the addition of purified components of the MCH specifically Hsp83, Hsc70, Hip, Hop, FKBP52, and p23 (Arbeitman and Hogness 2000). Its role in regulating proteins involved in cell cycle control make it an attractive target for the development of drugs for the treatment of disease such as cancer. Still to be worked out is the nature of its substrate spectrum during the heat shock response and whether or not it differs from its preferred clients during normal conditions (Richter et al. 2010).

In contrast to the other major HSP families, the small HSPs (sHSPs) have little conservation between family members. In *Drosophila*, this group consists of HSP22, HSP23, HSP26, HSP27.
All four are located at the 67B locus within a 12 kb span of DNA (Corces et al. 1980; Southgate et al. 1983). Beyond the alpha crystalline domain that defines this group, sHSPs can be wildly variable in their sequence composition (Ingolia and Craig 1982). The sHSPs also do not functionally depend on ATP (Haslbeck et al. 2005; McHaourab et al. 2009). This is likely because the function of the sHSPs is not to fold nascent or intermediate transcripts like HSP70 and HSP83 respectively but rather to hold unfolded proteins for later re-folding by other chaperones such as HSP70 and in doing so, prevent them from forming unfavorable associations and aggregates in the interim (Ehrnsperger et al. 1997; Lee et al. 1997; Lee and Vierling 2000; Mogk et al. 2003; Haslbeck et al. 2005; McHaourab et al. 2009). During heat shock, the sHSPs of yeast are responsible for keeping roughly 1/3 of the cytosolic protein in a soluble state (Haslbeck et al. 2004).

In *Drosophila*, expression of the sHSP genes occurs both in response to stress, as well as under normal conditions at various stages during development (Sirotkin and Davidson 1982; Zimmerman et al. 1983; Mason et al. 1984; Marin et al. 1993). In response to stress, induction of these genes is coordinate whereas under normal conditions, expression of each sHSP is independently regulated (Michaud et al. 1997). Developmental regulation of sHSP genes, like the HSP83 gene, occurs via a different DNA element than the one responsible for their heat-induced transcription (Cohen and Meselson 1985; Mestril et al. 1986; Riddihough and Pelham 1986). In some instances, sHSP expression during development is regulated by the steroid hormone 20-hydroxyecdysone (Ireland et al. 1982; Thomas and Lengyel 1986). Expression of these HSPs in the absence of heat-shock is thought to confer thermal tolerance and protection from oxidative stress (Berger and Woodward 1983; Rollet et al. 1992).

In addition to stimulating the transcription of genes coding for HSPs, activation of the HSR also leads to the induction of at least two ncRNAs in *Drosophila*: the alpha-gamma element, mentioned earlier in the section dealing with the HSP70 gene family, and Hsrω. The *D. melanogaster* Hsrω gene is found at the chromosomal locus 93D, the site of one of the largest HS-induced puffs (Bonner and Pardue 1976). The Hsrω gene spans over 10 kb DNA over 5 kb of which is tandem repeats (Lakhotia 2011). Two primary transcripts are produced from this gene: Hsrω-RB (formally hsrω-n) that spans the entire gene and Hsrω-RC (formally hsrω-pre-c) that is 1.9 kb and lacks the stretch of tandem repeats. Splicing of the single 700 bp intron from Hsrω-RB and Hsrω-RC produces hsrω-n2, which remains in the nucleus and Hsrω-RA
(formally hsrω-c), which is exported to cytoplasm (Lakhotia 2011; Mallik and Lakhotia 2011). The function of Hsrω is still largely unknown but it appears to be involved in directing the localization, stability, or activity of a number of proteins during both development and the HSR (Lakhotia 2011).

1.2.2 Cloning of the HS genes

_Drosophila_ HSP genes were among the first eukaryotic genes ever cloned (Schedl et al. 1978; Holmgren et al. 1979; Corces et al. 1980; Craig and McCarthy 1980; Voellmy et al. 1981). The first published report of isolation of HSP coding DNA was that of two segments coding for HSP70 identified from a collection of approximately 11,000 clones containing fragments of fly embryonic DNA by colony hybridization probed with an _in vitro_ labeled poly A HS RNA and shown to code for HSP70 by _in vitro_ translation and _in situ_ hybridization to polytene chromosomes (Schedl et al. 1978). Since during HS, non-HSP messages are also present in the cell and consequently contaminate HS RNA extracts, a differential hybridization procedure was utilized to subtract out non-HSP messages. Holmgren and coworkers later performed a similar screen of 23,000 clones, also created from fly embryonic DNA (Livak et al. 1978), that led to the identification of sequences coding for HSP83 and HSP68 (Holmgren et al. 1979). A few months after that, the identification of cloned sequences for the small hsps following similar experimental procedures came from two separate reports (Corces et al. 1980; Craig and McCarthy 1980).

1.2.3 HS-induced gene repression

Not all non-HS genes show the same repression profile in response to stress, but in general, a ten degree increase in temperature results in a 25-50% decline in synthesis in 90% of all proteins and an additional increase of four more degrees further reduces the synthesis of these proteins by 90-92% (Lindquist 1980). The reduction in the synthesis of non-HS proteins during HS is at least in part due to a HS-induced block in mRNA processing and translation (Ashburner and Bonner 1979; Lindquist 1980; Neidhardt et al. 1984; Yost and Lindquist 1986). HS also leads to repression at the level of gene transcription; HS causes RNPII to disengage from DNA (Hieda et al. 2005) as well as a reduction in the levels of RNPII associated with non-HS genes/puffs (Jamrich et al. 1977; Gilmour and Lis 1985) and the relocation of RNPII to sites of HS puffs (Jamrich et al. 1977; Greenleaf et al. 1978).
How HS triggers these events, however, is currently not well understood. Recently evidence for the role of nucleosomes in this process has been mounting. Histone deacetylases, HDAC1 and HDAC, through association with mammalian HSF1 have been implicated in HS-induced genome-wide gene silencing (Fritah et al. 2009) and the nucleosome remodeling complex SWI/SNF has been shown to play a role in the HS-induced repression of genes during HS in yeast (Shivaswamy and Iyer 2008). Furthermore, HS was shown to cause a genome-wide decline in RNPII elongation and nucleosome turnover (Teves and Henikoff 2011). Another possibility entails regulation by a contemporary set of transcriptional regulators- non-coding RNAs (ncRNAs). In mammals, two ncRNAs, B2 and Alu, transcribed from short interspersed elements (SINEs) by RNA polymerase III (Kramerov and Vassetzky 2005), are induced during HS (Liu et al. 1995; Li et al. 1999) and are able to prevent RNPII from engaging with DNA during closed complex formation (Yakovchuk et al. 2009) thereby functioning as repressors of transcription (Allen et al. 2004; Mariner et al. 2008).

Interestingly, the ability of HS to repress gene transcription comes at the expense of other transcriptional programs that might be active concurrently. For example, in mammals, several cytokine genes are induced in response to exposure to lipopolysaccharides and the expression of these genes are repressed during HS (Sun et al. 2005; Shi et al. 2006). Likewise, in Drosophila, HS effectively blocks the cascade of gene transcription activated by the steroid hormone 20-hydroxyecdyone (Ashburner 1970; Westwood et al. 1991)(Chapter 3). The repression of these sets of genes will be discussed in greater detail later on in this chapter.

1.3 Heat shock factor (HSF) and the heat shock element (HSE)

Regulation of the HSR is largely accomplished at the level of gene transcription. The trans-acting factor responsible for transcription regulation during the HSR was first identified in D. melanogaster cells through DNA-protein interaction studies and has since come to be known as heat shock factor (HSF) (Parker and Topol 1984; Wu 1984). Since this discovery, only a single HSF has been identified in yeast, Drosophila, and other invertebrates while multiple HSFs have been characterized in plants and in vertebrates (HSF1, HSF2, HSF3, and HSF4) with various roles in addition to the transcriptional regulation of the HSR (Wu 1995; Nover et al. 2001; Pirkkala et al. 2001; Åkerfelt et al. 2007). In animal species with multiple HSFs, HSF1 is the
functional homologue of the single HSF found in yeast and *Drosophila* (Wu 1995; Pirkkala et al. 2001). The focus of this section will be on *Drosophila* HSF and its homologues while the other HSFs will be covered in a later section “Other roles of HSF”.

### 1.3.1 Transcriptional regulation

Eukaryotic gene transcription requires the coordinated action of the general transcription machinery (comprised of RNPII and a set of six transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIF) collectively known as the general transcription factors (GTFs)), sequence specific transcription factors (such as HSF) and coactivators (such as Mediator). DNA sequence elements within and around genes serve as binding sites for the general transcription machinery and sequence specific transcription factors and consequently, determine when, where and to what extent a given gene is transcriptionally active. These DNA elements are referred to the core promoter, proximal promoter, and enhancer elements. Figure 1-1 provides an overview of their typical arrangement with respect to the gene they regulate. The core promoter, usually centered on the gene’s TSS, spans a region roughly 60 bp in length that can be further subdivided into elements that promote the binding of specific GTFs such as the TFIIB recognition element (BRE), the TATA box, the Initiator element (Inr), and the downstream core promoter element (DPE) (Figure 1-1) (Maston et al. 2006). It is not necessary for all of these elements to be present in a given core promoter and it is thought that the presence/absence of a particular element relates to the gene’s regulation pattern. For example, TATA boxes tend to be absent in the core promoters of “housekeeping” genes, i.e. genes that are constantly expressed (Yang et al. 2007). The proximal promoter, usually located within a few hundred base pairs upstream of the core promoter, contains binding sites for sequence specific transcription factors such as the HSE described in the previous section. Sequence specific transcription factors that bind to these sites may interact directly with the GTFs or with coactivators and are thought to be integral in bringing the general transcription machinery to the core promoter. Enhancers, like proximal promoters, are bound by sequence specific transcription factors. Unlike proximal promoter elements, however, enhancers may be located many kilobases upstream of the gene they regulate, within introns, or downstream of the gene (de Laat and Grosveld 2003; Zhao and Dean 2005; Maston et al. 2006). Enhancers are not believed to be capable of functioning independently of proximal promoters and it is believed that coactivators form a bridge between the two and bring the enhancer in the vicinity of the general transcription machinery. Whether
Figure 1-1 Representative arrangement of DNA sequence elements involved in the regulation of gene transcription.

Image adapted from (Smale and Kadonaga 2003; Carey et al. 2009).
an enhancer cooperates with a particular proximal promoter depends on several factors; the activators that bind each element must be compatible, as does the coactivator that bridges the activators, and insulator elements should not be present between the enhancer and the proximal promoter.

The packaging of DNA into chromatin provides an additional level of regulation by limiting access to the core promoters of several genes (Figure 1-2 (1)). Transcriptional activators/repressors may exploit this situation by directly or indirectly recruiting chromatin-remodeling enzymes to alter the chromatin state of the gene it regulates, and subsequently, the ability of RNPII and the GTFs to access to the gene. For example, certain human and yeast transcriptional activators directly recruit the SWI/SNF remodeling complex (Peterson and Workman 2000), while others recruit histone acetyltransferases (HATs) to acetylate histone tails and stimulate the recruitment of SWI/SNF and other bromodomain-containing remodeling complexes (Figure 1-2 (2)) (Larschan and Winston 2001).

Initiation of the transcription cycle occurs with the assembly of the general transcription machinery into a preinitiation complex (PIC) at the core promoter (Figure 1-2 (3)). There is some debate surrounding the manner of PIC formation in vivo, but in vitro it has been shown to occur in a step-wise manner. First the TATA box is bound by the TATA binding protein (TBP), a subunit of TFIID. This interaction is subsequently stabilized by TFIIA, which has the added functionality of preventing association of transcription inhibitors and by TFIIIB, which binds to the BRE and helps RNPII locate the TSS. RNPII together with TFIIIF then assembles over the TSS. Next TFIIIE joins the complex and serves as a docking station for TFIIH, the final component of the PIC, which binds thereafter (Figure 1-2 (3)). Following PIC assembly, the TFIIH helicase subunit, XPB, melts 11-15 bp of DNA at the TSS (Kim et al. 2000; Hahn 2004) to form a transcription bubble known as the open complex (Kornberg 2007). At this point transcription is initiated (Figure 1-2 (4)). It is common for RNPII to undergo several rounds of abortive transcription producing RNA products three to ten base pairs in length before escaping the promoter. Promoter escape requires dissociation of TFIIIB from the PIC and causes RNPII to break contacts with the promoter-bound factors of the PIC. Retention of the majority of the PIC factors at the promoter provides a scaffold that facilitates subsequent rounds of transcription. After the synthesis of 20-50 base of RNA, RNPII pauses (Figure 1-2 (5)). In Drosophila, this pause is mediated, at least in part, by the SPT4-SPT5 complex and NELF (negative elongation
Figure 1-2 Model for the transcriptional regulation of genes.

1) DNA packaged into nucleosomes (Nuc., orange decagons). 2) An activator (Act., red rectangle), binds to proximal promoter element and recruits chromatin-remodeling enzymes (green ovals) to clear the promoter. 3) Additional activators bind (green rectangle) and recruits coactivators such as mediator (Med., purple circle) and TFIID (light blue oval) which in turn recruit RNA polymerase (RNPII, dark blue oval) and GTFs (light blue ovals) to form the PIC. 4) THIIF melts DNA at the TSS to form the open complex (open oval) and phosphorylates (open red circles) the carboxy-terminal domain of RNPII (CTD, dark blue line attached to RNPII). 5) TFIIB dissociates from the PIC and RNPII escapes the promoter and transcribes 20-50 bp (orange line) before pausing. SPT4-SPT5 (SPT, green diamond) and NELF (red hexagon) mediate the pause. Medator and GTFs TFIIA, TFIID, TFIIE, TFIIH remain at the promoter to form a scaffold that facilitates reinitiation. 6) Activators, either directly or indirectly, recruit P-TEFb. P-TEFb phosphorylates NELF, SPT, and the CTD of RNPII (purple open circles). Phosphorylated NELF dissociates from the paused complex and the complex enters into productive elongation. Image adapted from (Hahn 2004; Koch et al. 2008; Carey et al. 2009; Fuda et al. 2009).
factor) (Andrulis et al. 2000; Wu et al. 2003). Release from the pause into productive elongation is facilitated by P-TEFb (positive transcription elongation factor b) (Peterlin and Price 2006), which is recruited either directly or indirectly by the activator to phosphorylate SPT5, NELF, and the carboxy terminal domain (CTD) of RNPII (Figure 1-2 (5)) (Lis et al. 2000; Renner et al. 2001). Phosphorylation of SPT5 causes it to stimulate transcription, while phosphorylation of NELF causes it to disengage from the elongating complex (Figure 1-2 (6)).

1.3.2 Structure and regulation of HSF

The structure of HSF contains several motifs and is highly conserved (Figure 1-3A). Drosophila, human and mouse HSFs are constitutively expressed and in the absence of stress, maintained as a monomer in an inactive conformation by a coiled-coil structure that forms between the arrays of hydrophobic heptad repeats that make up HSF’s amino-terminal trimerization domain and carboxy-terminal trimerization suppression domain (Rabindran et al. 1993; Zuo et al. 1994) (Figure 1-3B). As a monomer, HSF can associate with chromatin only loosely and non-specifically (Westwood et al. 1991; Baler et al. 1993; Rabindran et al. 1993; Sarge et al. 1993; Westwood and Wu 1993; Kim et al. 1994; Sistonen et al. 1994). Chemical cross-linking assays and analysis of various heteromeric forms of HSF arising from the association of long and short forms of HSF revealed that high affinity DNA binding is achieved by HSF through the formation of homotrimers (Perisic et al. 1989; Sorger and Nelson 1989). Such species of HSF form quantitatively with exposure to cellular stress which relieves the repressive interactions that maintain HSF as a monomer and thereby allow the formation of coiled-coil structures between trimerization domains of individual HSF molecules (Westwood et al. 1991; Baler et al. 1993; Sarge et al. 1993; Westwood and Wu 1993; Sistonen et al. 1994) (Figure 1-3B). Within seconds following HS, HSF homotrimers can already be detected at the promoter of Drosophila hsp70 and within two minutes, HSF levels are at saturated (Boehm et al. 2003; Yao et al. 2006).

Connected to the amino-terminal side of the trimerization domain by a variable length linker region is the most highly conserved HSF domain, a winged helix-turn-helix DNA-binding domain (DBD) (Vuister et al. 1994; Åkerfelt et al. 2010) (Figure 1-3A). HSF contacts DNA via its DBD at the major groove and the phosphate backbone (Wu 1995). The winged region in the
Figure 1-3 Functional organization and activation of *Drosophila* HSF.

(A) HSF has several highly conserved domains. N and C stand for amino and carboxyl termini respectively and numbers refer to the amino acid position. (B) Model of HSF activation and deactivation. See text for details. Figure adapted from Ho and Westwood 2002, Figure 2.
DBD aids in HSF trimer formation and/or cooperative binding between trimers by providing an interface for interaction between HSF molecules (Figure 1-3). In the case of yeast HSF, the DBD has also been found to have influence over transactivation and may mediate interactions with secondary factors (Bulman et al. 2001).

For both mammalian and Drosophila HSF, DNA-binding and transcriptional activation can be separable events under certain conditions (Wu 1995; Winegard et al. 1996). An acidic transactivation domain is located at the carboxy-terminal for all HSFs except for yeast HSF that has an additional transactivation domain at the amino terminus (Green et al. 1995; Newton et al. 1996). Control of the transactivation domain resides in a regulatory domain located between the transactivation and trimerization domains (Green et al. 1995; Shi et al. 1995; Zuo et al. 1995; Wisniewski et al. 1996; Åkerfelt et al. 2010) (Figure 1-3A). The exact function of the regulatory domain is still under much investigation, however, including the possible role of post-translational modifications (PTMs) as there are several amino acids that are known targets for PTMs in this domain (Åkerfelt et al. 2010). Along these lines, yeast and Drosophila HSF and mammalian HSF1 have been reported to undergo stress-inducible phosphorylation (Larson et al. 1988; Sorger and Pelham 1988; Sarge et al. 1993; Fritsch and Wu 1999) and that the transcription of hsp70 temporally correlates with HSF1 phosphorylation as opposed to DNA binding (Kline and Morimoto 1997). Furthermore, sodium salicylate induces DNA binding for both mammalian HSF1 and Drosophila HSF but does not induce phosphorylation of either HSF nor activate hsp70 transcription (Jurivich et al. 1992; Jurivich et al. 1995; Winegard et al. 1996). Attempts to identify residues important in the regulation of transactivation of HSF1 revealed that at least twelve serine residues are phosphorylated either constitutively or in response to stress (Chu et al. 1996; Cotto et al. 1996; Knauf et al. 1996; Kline and Morimoto 1997; Xia et al. 1998; Holmberg et al. 2001; Guettouche et al. 2005). The constitutive phosphorylation of at least two of these residues appear to have a repressive effect on HSF1 in the absence of stress (Chu et al. 1996; Knauf et al. 1996; Kline and Morimoto 1997) while the stress-inducible phosphorylation of at least two others seems to have a positive effect on HSF1 mediated transcription (Holmberg et al. 2001; Guettouche et al. 2005). Other than these residues, at least one other highly conserved residue in the regulatory domain is subject to stress-inducible sumoylation (Hong et al. 2001; Hietakangas et al. 2003). This modification appears to be involved in the repression of HSF1 transactivation (Hietakangas et al. 2006). In
addition to phosphorylation and sumoylation, HSF1 is also subject to acetylation that negatively regulates DNA-binding (Westerheide et al. 2009). Although great progress has been made towards identifying PTMs of HSF, much still remains to be done to fully elucidate the mechanisms by which they alter HSF activity and how different PTMs work in concert. Several lines of evidence suggest that HSF activity is subject to regulation by interacting proteins including HSPs (Figure 1-3B). In brief, HSP90 has been found to interact with monomeric HSF1 to repress trimer formation and DNA binding (Ali et al. 1998; Zou et al. 1998). Additionally, both HSP90 and HSP70 in conjugation with their respective cochaperones associate with the regulatory and activation domain of trimeric HSF1, respectively, to keep it inactive (Baler et al. 1992; Duina et al. 1998; Shi et al. 1998; Zou et al. 1998; Bharadwaj et al. 1999; Guo et al. 2001). However, isolation and biochemical characterization of monomeric Drosophila HSF from unshocked cells indicates that it does not form stable stoichiometric complexes with HSPs or cochaperones (Westwood and Wu 1993; Orosz et al. 1996). Apart from feedback inhibition provided by HSPs, HSF1 is thought to be subject to positive regulation by both the HSP70 cochaperone CHIP (C-terminus of HSP70-interacting protein) (Dai et al. 2003) and the nuclear protein DAXX (Boellmann et al. 2004). HSF1 has also been reported to interact with a ribonucleoprotein complex composed of the elongation factor eEF1A and a ncRNA called HSR1 that promotes trimer formation upon stress (Shamovsky et al. 2006).

1.3.3 Heat shock elements (HSEs)

Promoters of HSP genes contain cis regulatory elements known as Heat shock elements (HSEs) that mediate HSF DNA binding and transcriptional activity (Sakurai and Enoki 2010). HSEs are composed of alternating inverted repeats of the sequence nGAAn varying in size from a minimum of two repeats to over six (Amin et al. 1988; Xiao and Lis 1988). Usually however, at least three repeats are present as the binding affinity is significantly reduced with only two (Perisic et al. 1989). This is likely because the nature of HSF’s interaction with the HSE is such that the DBD of each subunit of trimeric HSF contacts an individual repeat (Perisic et al. 1989; Sakurai and Enoki 2010). When greater numbers of repeats are present, HSF trimers bind cooperatively (Xiao et al. 1991). This type of interaction makes it possible to get high affinity binding to HSEs composed of repeats other than multiples of three. For example, an HSE composed of four repeats will bind two HSF trimers thereby making contact with only four of a
possible six HSF subunits, presumably leaving one subunit from each trimer unbound (Perisic et al. 1989; Yamamoto et al. 2009; Sakurai and Enoki 2010). Systematic evolution of ligands by exponential enrichment (SELEX), a non-biased approach to identifying DNA-binding sites from a random population, confirmed previous findings of HSF’s preference for alternating inverted arrays of the sequence nGAAn and the preference for adenine in the first position and pyrimidine in the fifth position (Amin et al. 1988; Xiao and Lis 1988; Perisic et al. 1989; Cunniff and Morgan 1993; Fernandes et al. 1994; Kroeger and Morimoto 1994). Additionally, the SELEX approach revealed that the average number of repeats was four to five (Kroeger and Morimoto 1994).

1.3.4 Transcriptional regulation by HSF

HSF affects gene transcription through interaction with transcriptional coactivators and chromatin-modifying complexes as well as by direct or indirect recruitment of elongation factors and chaperones. For example, following HS, the transcriptional coactivator Mediator is rapidly recruited to sites coincident with HSF binding in Drosophila and through direct interaction with HSF, creates a link to the basal transcription machinery to relay activation signals (Park et al. 2001; Kim et al. 2004). Another transcriptional coactivator, activating signal cointegrator (ASC)-2, has been shown to interact strongly with mammalian HSF1 in a GST pull-down assay and to potentiate HSF1-mediated transactivation in a luciferase reporter assay (Hong et al. 2004).

In terms of chromatin-modifying complexes, the ATPase subunit of the nucleosome remodeling complex SWI/SNF has been shown to interact with the transactivation domain of human HSF1 to facilitate chromatin remodeling that promotes stimulation of elongation by RNPII at HSF bound sites (Sullivan et al. 2001; Corey et al. 2003). And although brahma, the ATPase subunit of the related complex in Drosophila, does not seem to interact with Drosophila HS loci (Armstrong et al. 2002), the TAC1 chromatin-modifying complex is recruited to several HS gene loci on Drosophila polytene chromosomes following HS where it appears to have a role in promoting high levels of gene expression (Smith et al. 2004). Furthermore, the ATP-dependent chromatin modifying complex NURF, in combination with GAGA factor and ATP, facilitates nucleosome alternation at the Drosophila Hsp70 promoter (Tsukiyama and Wu 1995) and flies deficient in NURF-301, which together with ISWI provides the chromatin remodeling
functionality of NURF, exhibit impaired HSF binding and Hsp26 and Hsp70 activation (Badenhorst et al. 2002).

GAGA factor (GAF) and paused RNPII also help to establish an open chromatin state at gene promoters prior to HS that is thought to facilitate HSF binding to certain HSEs following HS in *Drosophila* (Lee et al. 1992; Tsukiyama et al. 1994; Shopland et al. 1995; Wall et al. 1995; Wilkins and Lis 1997; Leibovitch et al. 2002; Guertin and Lis 2010). In addition, both factors are thought to be important for HS-induced transcription of certain genes. Recently however, it was reported that although GAF appears to be present at many sites that coincide with HS-induced genes, it is not necessary for HSF activation (Guertin and Lis 2010). This same study also suggests, based on a sample of 16 HSF bound genes, that there appears to be no significant correlation between the presence of paused RNPII and HS gene activation (Guertin and Lis 2010).

Boundary element-associated factor (BEAF) is another factor that associates with promoter regions and helps to keep many of them in a state permissive for high levels of transcription (Jiang et al. 2009). Comparison of BEAF and HSF binding sites on a genome wide scale suggest a strong correlation between the two (Guertin and Lis 2010)(Chapter 2). Interestingly, BEAF and HSF binding were more likely to coincide at promoters of genes that were not transcriptionally active suggesting that BEAF may have a role in blocking HSF-induced activation of certain genes consistent with its role as a insulator protein (Bushey et al. 2009; Jiang et al. 2009).

In *Drosophila*, Poly (ADP-) ribose polymerase (PARP) is required for puff formation at HS gene loci in response to HS (as well as at developmental and immunity gene loci in response to hormones and infection respectively) presumably by facilitating the removal of local chromatin proteins from the gene to make it accessible (Tulin and Spradling 2003; Zobeck et al. 2010) and either catalytic inhibition or RNAi (interference) knockdown of PARP can prevent the HS-induced loss of chromatin structure at Hsp70 (Petesch and Lis 2008). Recently, it has been shown that the nucleosome remodeler dMi-2 acts as a co-activator for the full transcriptional activation of HS genes and is recruited to active HS genes in a PARP dependent manner (Murawska et al. 2011). Taken together, these findings suggest a stimulatory role for PARP in the transcription of HS-induced genes in *Drosophila*. In contrast, murine PARP-1 has recently
been shown to have a repressive effect on Hsp70 gene expression as the HSR in murine fibroblasts is augmented by knockdown of PARP-1 (Aneja et al. 2011). Reasons for the apparently differing responses to PARP inhibition between insects and mammals are currently unknown.

The elongation factors Spt5 and Spt6 and the protein complex FACT (Facilitates Chromatin Transcription) are all rapidly recruited to the hsp70 locus on Drosophila polytene chromosomes following HS and track with the elongating RNPII complex through the length of the gene (Saunders et al. 2003; Ardehali et al. 2009). Furthermore, recruitment of P-TEFb, which is responsible for release of paused RNPII into productive elongation (Peterlin and Price 2006), is also influenced by the HS-induced binding of HSF to HSEs in the Hsp70 promoter (Lis et al. 2000; Boehm et al. 2003). Similarly, Menin (multiple endocrine neoplasia type I) is recruited to the HSP70 promoter following HS and although it is not required for the initiation of the stress response, it appears to function as a positive regulator for sustaining the expression of hsp genes in Drosophila (Papaconstantinou et al. 2005; Papaconstantinou et al. 2010). The exact mechanism of recruitment to HS genes upon HS for both P-TEFb and Menin is not currently known, although in neither case does it appear to involve direct contact between these factors and HSF (Lis et al. 2000; Boehm et al. 2003; Papaconstantinou et al. 2005).

### 1.3.5 Other roles of HSF

Over the years, evidence for additional roles of HSF beyond the regulation of the HSR has been mounting. Yeast HSF is an essential protein for growth and viability in the absence of stress (Sorger and Pelham 1988; Wiederrecht et al. 1988; Gallo et al. 1993), and while it is possible that it has a role in regulating the basal expression of HSP genes (Sorger 1991), the determination that nearly 3% of the yeast genome may be subject to transcriptional regulation by yeast HSF (Hahn et al. 2004) raises the possibility that additional functions may be involved. Insight into an additional role of Drosophila HSF in development has been provided by a temperature sensitive mutation of the Drosophila HSF gene that prevents it from acquiring DNA binding ability at heat shock temperatures (Jedlicka et al. 1997). This mutation, known as the hsfs4 mutation, has revealed that unlike in yeast, HSF is not required for general growth and viability under non-heat shock conditions in adult Drosophila; however, it is required for both
oogenesis and early larval development. Although the exact role of HSF in this context is not understood, it does not appear to be related to the regulation of HSP gene expression.

Like *Drosophila* HSF, mammalian HSF1 also plays a role in development that, for the most part, also does not seem to involve control of basal HSP expression (Xiao et al. 1999; Christians et al. 2000). For example, mouse HSF1 has been shown to play a critical regulatory role in extra-embryonic development as *Hsf1* null mice exhibit defects in the chorioallantoic placenta coupled to increased prenatal lethality (Xiao et al. 1999). *Hsf1* deficient mice that manage to develop to adults experience growth retardation and female infertility. The latter defect seems to arise from a requirement for maternally contributed HSF1 which is essential for normal progression of meiosis (Metchat et al. 2009) and for the control of early post-fertilization embryo development and cannot be compensated by wild type HSF1 contribution from spermatozoa (Christians et al. 2000). Recently, a study examining *Hsf1* null mouse oocytes identified a network of meiotic genes regulated by HSF1 in both the adult and embryonic phases of female meiosis providing a molecular basis for the observed phenotypes (Le Masson et al. 2011). HSF1’s role in reproduction is not limited to females, however, as aberrant activation of HSF1 leads to infertility in male mice due to apoptotic cell death in the pachytene spermatocytes in the testis (Nakai et al. 2000; Izu et al. 2004). It appears that at least in the context of female reproduction, HSF1’s role does involve the regulation of one chaperone, HSP90α (Metchat et al. 2009). Another place where HSF1 does appear to function in the control of the basal expression of certain HSPs is in the mouse heart where the basal levels of HSP25, αβ-cystallin, and HSP70 are reduced in HSF1 null mice compared to their wild type counterparts (Yan et al. 2002). Similarly, HSP27 and αβ-crystallin levels are altered in mouse brains that lack HSF1 and the brains of these mice exhibit major morphological abnormalities including enlarged ventricles and reduced white matter in addition to astrogliosis and neurodegeneration (Xiao et al. 1999; Santos and Saraiva 2004; Homma et al. 2007). The requirement for HSF1 in adult mice for the maintenance of olfactory neurogenesis is also related to necessity for controlled expression of HSP genes as well as cytokine genes (Takaki et al. 2006).

*Hsf1* knockout mice have also led to the identification of roles for HSF1 in the inflammation and immune responses, where HSF1 is required for full lipopolyliposaccharide (LPS)-mediated induction of interleukin-6 (IL-6) (Inouye et al. 2004) as well as for the suppression of pro-inflammatory cytokine tumor necrosis factor α (TNF-α) (Xiao et al. 1999). The requirement for
HSF1 in the induction of IL-6 seems to be related to the ability of the HSF1 to bind to the IL-6 promoter and maintain it a partially open state that facilitates binding of an activator or repressor (Inouye et al. 2007). The suppression of TNF-α is particularly interesting as it points to a role for HSF1 as a negative regulator in addition to its well-established role as transcriptional activator. A direct link between the repression of TNF-α and HSF1 was provided by the identification of a high affinity HSF1 binding site in the 5’ UTR encoding region of the TNF-α gene that confers the repressive effects (Singh et al. 2002). The HSF1 binding site in this case is not the complete cognate HSE but does consist of multiple repeats of the 5 bp sequence nGAAAn. Recently, HSF1 has also been implicated in the negative regulation of LPS-induced expression of IL-6 by stimulating transcription of its repressor, activating transcription factor 3 (ATF3) (Takii et al. 2010). In yet another interesting twist of HSF1 function, HSF1 was found to bind to and antagonize the activity of nuclear factor of interleukin 6 (NF-IL6) to inhibit the LPS-induced transcription in interleukin β (IL-β) (Cahill et al. 1996; Xie 2002) again highlighting its ability to negatively regulate gene expression however, in a very different way.

As mentioned above, in addition to HSF1, mammalian genomes encode three other members of the HSF family, HSF2, HSF3, and HSF4. The existence of multiple HSFs suggests that they may have specialized roles and/or respond to specific stimuli. Originally, HSF1 was thought to be the sole regulator of the HSR but recently the others members of the HSF family have been found to play a role in the HSR; HSF2 as a modulator of HSF1-mediated expression of major HSPs (He et al. 2003; Ostling et al. 2007) and of non-classical HS genes (Sandqvist et al. 2009); and HSF3 and HSF4 also as activators of non-classical HS genes, HSF3 independently from HSF1 (Fujimoto et al. 2010), and HSF4 by facilitating HSF1 binding through chromatin modification (Fujimoto et al. 2008). It is important to note that these additional family members are not able, however, to entirely compensate for HSF1 in the HSR, as disruption of the HSF1 gene prevents heat-induced expression of HSP25 and HSP70 (Zhang et al. 2002) and acquisition of thermotolerance required for protection against heat-induced apoptosis (McMillan et al. 1998; Zhang et al. 2002). Additionally, mammalian HSF2 and HSF4 have further roles unrelated to the regulation of the HSR (Åkerfelt et al. 2010).

Like HSF1, HSF2 also plays a role in brain development and fertility; HSF2 null mice have enlarged ventricles and females have a reduced number of ovarian follicles and produce abnormal eggs while males have small testes, low sperm count, and abnormalities in
semineferous tubules, sperm head shape, and spermatocyte meiosis (Kallio et al. 2002; Wang et al. 2003; Akerfelt et al. 2008) and in the background of an HSF1 deficiency, are infertile resulting from a block in spermatogenesis (Wang et al. 2004). The role of HSF2 in this context might be even more complex and interrelated with other factors, however, as a related study on HSF2 null mice found that HSF2 was dispensable for normal development and fertility (McMillan et al. 2002).

HSF4 null mice have revealed a requirement for HSF4 in postnatal lens development since HSF4 deficient mice exhibit abnormal lens fiber cells and develop cataracts soon after birth (Fujimoto et al. 2004). Mutations in HSF4 have also been associated with human cataracts (Bu et al. 2002). It is likely that the single HSF gene encoded by genomes of lower eukaryotes may fulfill some of the additional functions covered by the multiple HSFs in higher eukaryotes and this may be at least part of the reason for HSF being an essential gene in these species.

1.4 Genomic studies on the HSR

1.4.1 Identification of novel HS genes

The studies reviewed above in the section titled “Transcription and translation during the HSR” were successful in identifying the major genes transcribed during the HSR however it appeared that there was still a subset of genes that eluded identification by the classical methods employed. The popularization of technologies allowing for the analysis of transcription on a genome-wide scale opened up a new avenue for the exploration of transcription during the HSR and provided a means to seek out these genes. The creation of extensive cDNA libraries as well as full genome sequencing made possible the use of microarray technology to simultaneously monitor the expression of thousands of genes for analysis of transcriptional response on a much larger scale than previously possible. Many studies have since been performed looking either directly at the HS-induced gene transcription or at transcription resulting from the other forms of stress. In this section the earliest studies addressing HS-induced transcription in non-\emph{Drosophila} model organisms and human cells lines will be introduced first followed by an introduction of major microarray studies examining HS-induced transcription in \emph{Drosophila}.

The first study to examine the transcription of thousands of genes in parallel in response to HS was performed with microarrays representing just over 35% of known yeast ORFs and was
successful in indentifying HS-induced changes in transcription of both known HS-responsive genes such as *HSP12* and *HSP26* (members of the shsp family) and novel HS-responsive ORFs and genes such as *YLR194, KIN2*, and *HXT6* (Lashkari et al. 1997) providing support both of the existence of novel HS-inducible genes and for the utility of microarrays in their identification.

In addition to a study on *Drosophila* embryos (Leemans et al. 2000), which will be discussed later on in this section, one of the earliest studies to look at HS-induced transcription in metazoans on a global scale was performed on the nematode worm *Caenorhabditis elegans* (GuhaThakurta et al. 2002). In this study, the transcriptional response to HS of approximately 12,000 genes, or roughly 60% of the protein coding genes in the *C. elegans* genome, was monitored by microarray. Twenty-eight genes were identified that exhibited a twofold or greater increase in transcription when the response to a one hour HS and a two hour HS followed by a two hour recovery period at non-HS temperatures were averaged. Using an average fold change for the two treatments, the most strongly induced genes were two members of the hsp70 family and one member of the shsp family. The three next most highly induced genes had no known function and showed a 9-13 fold induction in the treatment that included the recovery period. Interestingly, after the one-hour HS treatment, the gene with the highest fold change next to the hsp family members encodes a serine/threonine kinase pointing to a common theme when considered with the tyrosine kinase *shark* that showed the strongest induction among non-hsp genes in *Drosophila* embryos (Leemans et al. 2000). Among the remaining genes are two more members of the hsp70 family and a gene for GST (GuhaThakurta et al. 2002).

Trinklein and colleagues assayed 176 putative human heat-responsive genes by RT-PCR for their transcriptional response in K562 human erythroleukemia cells to a 1 hr heat treatment and found that 93 exhibited a 2-fold or greater increase in transcript level including known hsp genes encoding hsp27, hsp40, hsp60, hsp70, hsp105, hsp110 (Trinklein et al. 2004). Roughly one third of these genes encoded proteins of unknown function and the majority had not previously been shown to be heat-inducible.

Other studies have compared the transcriptional responses to a variety of other stresses to HS in yeast (Gasch et al. 2000; Causton et al. 2001), and in cultured human cells (Murray et al. 2004). In yeast, a large portion of the genome (at least 10-15%) exhibits a common response to the environmental stresses examined including HS while smaller subsets of genes are activated in a
stress-specific manner (Gasch et al. 2000; Causton et al. 2001). Many of the genes activated in the response to HS have functions in protein folding and transport although many more had undefined functions at the time of publication. Interestingly, the response to HS is highly similar to the response to progression into a stationary growth phase with the induction of genes involved in respiration and alternative carbon source regulation (Gasch et al. 2000). Gasch and colleagues propose that a change in ATP:AMP ratios brought about by the sudden activation of numerous cellular chaperones may cause the change in central energy metabolism gene expression similar to the situation in mammalian cells explaining the commonalities between the responses. In higher eukaryotes, however, while there is still a common set of stress responsive genes, most genes respond in a stress specific fashion (Murray et al. 2004). The common set of stress responsive genes tend to be involved in processes specific to higher eukaryotes such as cell to cell communication.

The first study to look at the global transcriptional response to HS in Drosophila utilized oligonucleotide arrays representing 1,519 genes (or roughly 10% of all currently known genes) in D. melanogaster to assay stage 10-17 embryos and identified 74 genes (or about 4.8% of genes tested) in 12 functional classes whose relative expression level was changed by at least 1.5 fold over non-hs levels in response to a 25 minute 36°C heat shock. Among the changed genes, 36 were induced, including genes for known hsps with the largest changes in expression being observed for the shsps (Leemans et al. 2000). Apart from hsp genes, only two other genes showed an increase in transcription greater than 3-fold; Shark, a putative tyrosine kinase (Shishido et al. 1991) and anon-23Da, which is thought to have a role in lipid metabolism as well as bristle and wing disc development (Wisotzkey et al. 2003; Bronstein et al. 2010). The remaining hs-induced genes showed only modest 1.5-3 fold increases in expression (Leemans et al. 2000). The majority down regulated genes also only showed moderate changes of 1.5-3 fold in response to hs with the exception for two genes that were repressed by greater than 3 fold: mus210, which is predicted to have a role in DNA damage repair (FlyBase 1992); and anon-X, which is involved in the regulation of the Notch signaling pathway, vacuolar acidification (Yan et al. 2009), and tracheal system development (Metzstein et al. 2007).

Examination of the HSR in adult D. melanogaster using cDNA microarrays that represented approximately half of the known Drosophila genes, identified 92 genes significantly regulated by a 60 minute HS followed by 30 minutes of recovery, 80 of which were upregulated (Neal et
An extension of this study looking at nearly all the remaining known genes to the same treatment as well as the response to a 30 minute HS without recovery in both D. melanogaster adults and third instar larvae identified 687 genes with altered expression profiles, of which approximately one third were downregulated (Neal 2005). Functional analysis of genes responding at each developmental stage revealed no major differences, however, relatively few genes (41), all of which were upregulated, are common to the two treatments and developmental stages examined. The most highly induced of this core set of genes code for: known HSPs (HSP83, HSP70Ab, HSP68, DnaJ-1, and the sHSPs); Hsrω; another chaperone, CG5001; a kinesin motor protein, nod; and a protein of unknown function, CG16782. A time course analysis of the HSR in the Drosophila SL2 cell line revealed that many genes from this core set are also consistently upregulated at various time points over the 1.5 hr interval examined. One additional gene induced to levels comparable to the major HSP genes was identified, Ssl, which codes for a component of the casein kinase II complex. However, overall, far fewer genes (131 total) were HS-responsive across all times examined in the cell line compared to the whole organism.

The first whole genome study in Drosophila to examine the transcriptional effect of HS and recovery from HS identified 1222 differentially expressed genes in eight time points examined in adult female flies: early -upregulated, early-downregulated, and late-upregulated (Sørensen et al. 2005). Early upregulated genes, of which 265 were identified, were enriched in functions also enriched among genes that responded to other forms of stress: chaperone activity, response to stress, and glutathione transferase activity, as well as a couple of functions that appear to be unique to HS: protein kinase activity, and gluconeogenesis. Early downregulated genes, of which 508 were identified almost doubling the number of early upregulated genes, were enriched in functions primarily related to metabolism, with the most enriched functions being monosaccharide transporter, lysosome, cytochrome P450 activity, response to biotic stimulus, eye pigmentation and with many having catalytic, hydrolase, and/or peptidase activity. In general, genes that are induced exhibit a greater change in levels than those that are repressed. The transcriptional profiles for stress responsive genes elucidated by this study provided support for Ritossa’s findings that the primary transcriptional response to HS is relatively short lived and the time course revealed that the majority of the primary response genes return to normal within four hours (Ritossa 1962; Ritossa 1964; Sørensen et al. 2005). A comparison of genes regulated
by HS to genes regulated by other cellular stresses in *Drosophila* from three other microarray studies looking at aging (Landis et al. 2004), oxidative stress (Girardot et al. 2004; Landis et al. 2004), endoplasmic reticulum stress (Girardot et al. 2004), and inbreeding (Kristensen et al. 2005) revealed a greater degree of overlap between genes repressed by different stressors than between genes that were induced (Sørensen et al. 2005) suggesting that a general repression of genes not required for stress survival is the most basic common link at least in *Drosophila*.

1.4.2 HSF binding sites

The genome-wide binding distribution of *Drosophila* HSF during HS and non-HS conditions was initially examined by immunostaining of polytene chromosomes isolated from wandering third instar larvae (Westwood et al. 1991). The observations made in this study were largely unexpected; that is, in addition to binding loci known to be transcriptionally active during the HSR, HSF also bound to many loci that were inactive or repressed during the HSR. Furthermore, roughly 20% of the non-HSP gene containing loci bound by HSF exhibited staining intensities comparable to that seen for several of the HS puff sites. These observations provided an important overview of HSF binding and shed light on the possibility of HSF regulating the expression of yet unidentified HS responsive genes, or serving an unforeseen role as a transcriptional repressor of non-HSP genes during the HSR.

With the adoption of microarray technology, and more recently next generation sequencing, to the identification of transcription factor binding sites through techniques such as ChIP-chip and ChIP-seq, several studies have attempted to provide a higher resolution map of HSF binding sites on a genome-wide scale in yeast (Hahn et al. 2004), human gene promoters (Trinklein et al. 2004), and *Drosophila* embryos (Birch-Machin et al. 2005) and S2 cells (Guertin and Lis 2010).

The first whole-genome analysis of HSF binding sites during HS was performed in yeast and identified genes with a diversity of functions representing almost 3% of the genome (Hahn et al. 2004). Two thirds of the sites identified are upstream of open reading frames and the majority of genes downstream of these sites are heat-inducible. In accord with previous studies that suggest yeast HSF is constitutively bound to DNA in the absence of stress (Sorger et al. 1987; Jakobsen and Pelham 1988), a subset of the more strongly bound promoters were found to be bound equally well under HS and non-HS conditions (Hahn et al. 2004). Many target promoters, including some prominent target loci, however, were found to instead display heat-inducible
binding more akin to the situation observed for the HSF of higher eukaryotes underlining the importance of the global view.

Trinklein and colleagues examined the relationship between heat-induced DNA binding of human HSF1 and heat-induced gene expression for human genes that are either stress inducible, contain an HSE like sequence in their promoters, or that are preferentially bound by HSF1 in its promoter region (Trinklein et al. 2004). While this study only examined the above mentioned properties in a total of 437 genes, they found that like Drosophila HSF (Westwood et al. 1991), human HSF1 bound many sites that did not appear to be transcriptionally active under the conditions tested (Trinklein et al. 2004). The authors also discovered 24 human genes upregulated by heat but whose promoters were not bound by HSF1 and did not show significant heat-inducible expression in independent promoter assays. In addition, they also found that HSF1 binds to roughly only half of the promoters containing HSEs that were tested.

The first study to use a ChIP-chip based approach to map HSF binding sites in Drosophila embryos identified 141 genes with significant ChIP enrichment, including several HSP genes and other predicted chaperone-encoding genes, from cDNA based arrays representing roughly 5,000 genes or approximately 30% of the Drosophila genome (Birch-Machin et al. 2005). Based on the total number of genes examined, the authors speculated that by extrapolation of the number of bound targets to the scale of the full genome, approximately 3% of all genes are targets of HSF, a proportion consistent with the finding for yeast HSF (Hahn et al. 2004). Birch-Machin et al (2005) also found that roughly 50% of the most significantly bound genes indentified in their study map to cytological bands identified as containing HSF binding sites by immunostaining (Westwood et al. 1991). Examination of the expression of genes bound by HSF revealed that only 16 were induced by HS including known HSP genes and two were repressed providing support for previous observations that very few bound targets are transcriptionally regulated by HSF under the conditions under investigation (Westwood et al. 1991; Trinklein et al. 2004; Birch-Machin et al. 2005). Among HSF bound genes, enriched functional categories included chaperone or HS protein activity, metabolism, and protein modification or degradation, (Birch-Machin et al. 2005). In a parallel approach the authors examined HSF binding distribution at even higher resolution on a genome tiling array covering 2.9 Mb of chromosome arm 2L, the regulatory regions of some early segmentation genes, Ubx bound genes, and five HSP gene loci. Distribution of HSF across the five loci with HSP genes showed good
correlation with known HSEs, and these regions showed the highest enrichment among all regions represented on the tiling array, proving the utility of tiling arrays for high resolution analysis of *Drosophila* HSF binding sites.

Recently, *Drosophila* HSF binding has been characterized in S2 cells by ChIP-seq (Guertin and Lis 2010) representing one of the first full genome-wide studies to map HSF in high resolution (the other will be presented in Chapter II). Using the ChIP-seq approach, HSF was detected at 464 sites after HS of which only 20 sites showed detectable HSF binding prior to HS (albeit to a much lesser extent) providing confirmation of previous reports that the majority of sites bound by *Drosophila* HSF are inducibly bound after HS (Westwood et al. 1991). While the majority of sites bound by HSF were found to contain HSEs, less than 15% of all motifs in the genome significantly matching HSEs (in this case with p-values ≤ 5x10^{-6}) are bound by HSF during HS, paralleling the findings with human HSF1 that the presence of an HSEs alone is not sufficient to predict HSF binding (Trinklein et al. 2004). Interestingly, HSEs bound by HSF after HS are enriched both within genes and within gene promoters (in the case defined as -500 bp of a transcription start site) while HSEs that are not bound by HSF after HS are not enriched in these regions, suggesting that HSF has a preference for binding HSEs located in genes and in gene promoters (Guertin and Lis 2010). Although the transcript accumulation of only 16 genes was assessed by RT-qPCR, the binding of HSF to a gene promoter was not found to be an indicator of whether or not that gene would be transcriptionally induced, again a finding remarkably similar to the situation of human HSF1 (Trinklein et al. 2004). Comparison of HSF binding sites to the positioning of nucleosomes and other transcription factors revealed that the propensity of HSF to bind particular HSEs is not merely a function of the DNA being nucleosome-free but rather, is correlated with the presence of marks of active chromatin and factors that are associated with active transcription (Guertin and Lis 2010).

### 1.5 The ecdysone response

20-hydroxyecdysone (ecdysone, 20E) is a prototypical steroid hormone that coordinates development in *Drosophila* (Kozlova and Thummel 2000; Riddiford et al. 2000; Spindler et al. 2009). It is produced in response to prothoracicotropic hormone (PPTH) as well as nutritional inputs from dietary yeast ergosterol or plant sterols (Gilbert et al. 2002). Increases in 20E titer triggers larval molts, puparium formation, the onset of metamorphosis as well as differentiation
of adult structures during metamorphosis by controlling transcription in various target tissues
(Kozlova and Thummel 2000; Spindler et al. 2009).

Like the HSR, 20E’s action on gene transcription was first suggested by the effects it had on the
puffing pattern of polytene chromosomes, in this case, from *Chirononus tenans* salivary glands
(Clever and Karlson 1960; Clever 1961). Later, through a series eloquent studies on *Drosophila*
polytene chromosomes, where the effects of 20E concentration, 20E treatment duration and
RNA and protein synthesis inhibitors were examined, Ashburner was able to deduce that 20E
triggers a cascade of gene activation (Ashburner 1971; Ashburner 1972; Ashburner 1972;
regulatory cascade was found to consist of a small set of genes that were directly activated by
20E (primary response genes) and a large set of genes that was subsequently activated by
products of the primary response genes (secondary response genes) (Figure 1-4).

Direct activation of primary response genes by 20E is accomplished though interaction with a
heterodimeric nuclear receptor complex composed of the ecdysone receptor (EcR) and
Ultraspiracle (Usp) (Yao et al. 1993) (Figure 1-4). The EcR is one of 18 nuclear hormone
receptors in *Drosophila* and one of the best-studied nuclear hormone receptors over all (King-
Jones and Thummel 2005; Gáliková et al. 2011). There are five isoforms of EcR and the
expression of each may vary between different tissues (Robinow et al. 1993; Talbot et al. 1993;
Schubiger et al. 1998; Cherbas et al. 2003). Typically, tissues expressing the isoforms show
similar responses to 20E. Usp, another one of the 18 *Drosophila* nuclear hormone receptors, is
structurally and functionally similar to the vertebrate retinoid X receptor (RXR) (Oro et al.
1990; Koelle et al. 1991; Yao et al. 1993). The EcR/Usp complex interacts with DNA at
ecdysone response elements (EcREs), 15-bp imperfect palindromes (5’-RG(GT)TCANTGA(CA)CY-3’)
in the upstream regions of primary response genes that have
high homology to vertebrate hormone response elements (Cherbas et al. 1991).

Among the primary response genes, three have been studied in great detail: *Eip75B, Eip74EF,*
and *br* (Thummel 1996). These three genes are particularly large (>60 Kb), encode site-specific
DNA-binding proteins, and contain multiple promoters for the ecdysone-induced synthesis of
various mRNA isoforms. *Eip75B*, like *EcR*, encodes members of the steroid receptor
Figure 1-4 The Ashburner model for gene regulation by ecdysone (20E).

20E binds to and activates the ecdysone receptor complex composed of EcR/USP. The active receptor complex binds to and directly activates early response genes while inhibiting the expression of late response genes. The protein products of some early response genes activate late response gene and inhibit the expression of early response genes as part of a negative feedback loop that helps regulate the duration of the response.
superfamily that contain a zinc finger (Segraves and Hogness 1990). While br also encodes a set of zinc finger containing proteins (DiBello et al. 1991), Eip74EF encodes an ets-related protein (Burtis et al. 1990). Interestingly, all three of these genes have HSF binding sites within their massive introns (Guertin and Lis 2010; Gonsalves et al. 2011). Although HSF occupies these sites during the HSR, these genes are not induced but rather are repressed during HS (Ashburner 1970; Westwood et al. 1991; Gonsalves et al. 2011).

Many secondary response genes are presumed to code for proteins with direct impact on development. The first secondary response gene to be characterized at the molecular level, Eig71Eg, was found to encode a family of secreted polypeptides believed to protect the animal from infection during metamorphosis, thus, satisfying the presumed notion that late genes serve effector functions (Wright et al. 1996). Subsequently, at least two other secondary response genes have been characterized molecularly and found to produce transcripts that more closely resemble the large, complex transcripts of the primary response genes rather than that of Eig71Eg (DiBello et al. 1991; Stowers et al. 1999; Stowers et al. 2000). These observations suggest that not all secondary response genes encode simple effector functions (Thummel 2002).

1.6 Aims of this thesis:

The work presented in this thesis aims to:

- Map Drosophila HSF binding sites genome-wide and identify novel heat-inducible genes.

- Identify genes transcribed in Kc167 cells and salivary glands in response to 20E stimulation and examine how HS affects their expression.

1.6.1 Aim I: Map Drosophila HSF binding sites genome-wide and identify novel heat-inducible genes.

During the HSR, Drosophila HSF has been observed to bind over 200 loci on polytene chromosomes, yet only a very small fraction of these sites appear to be transcriptionally active. Even fewer have been formally identified as HS-inducible genes to date. What then is the purpose of this potent transcriptional activator binding to these additional sites? Some of these sites may contain essential genes such as the histone genes (Spradling et al. 1975; Spradling et
al. 1977; Westwood et al. 1991) that escape the global repression typically seen for of the majority of non-HS genes during HSR in *Drosophila* (Jamrich et al. 1977). Others may be sites of novel HS genes that have thus far evaded detection (Westwood et al. 1991). It has also been suggested that HSF might act as direct repressor of specific non-hs genes (Westwood et al. 1991). Of course, identification of such genes and their roles will further our understanding of the HSR. Given the large number of sites bound by HSF, however, it is also possible that HSF binding to these sites serves a yet unidentified purpose. If other roles for HSF do exist, determining them would greatly benefit, not only our understanding of the HSR, but also our understanding of transcription in general.

A first step towards establishing the purpose of HSF binding to these additional sites is to obtain both a high-resolution and global view of HSF binding sites. This has been made possible in the last five years or so with the production of high-resolution genome-tiling microarrays (Chip) and more recently with next generation sequencing (seq). Coupling these powerful new technologies with classical chromatin immunoprecipitation (ChIP) techniques can provide us with a detailed map of HSF binding genome-wide.

In Chapter 2, I identify HSF binding sites in the *Drosophila* Kc cell genome using ChIP-chip. I also examine the transcriptional profiles in heat shocked Kc cells, wild type 3rd instar larvae and *hsf* mutant larvae. I analyze the surrounding sequences for the segments bound by HSF for known transcription factor binding sites. I also compare the transcript profiles with the mapped HSF binding sites in an attempt to understand the relationship between the binding of HSF and genes whose transcript levels are regulated by heat.

1.6.2 **Aim II: Identify genes transcribed in Kc167 cells and salivary glands in response to 20E stimulation and examine how HS affects their expression.**

A number of sites bound by HSF are involved in the response to the developmental hormone ecdysone (20E), but do not exhibit HS-inducible transcription in larval salivary glands (Stevens 1999). Instead, genes at these sites appear to be transcriptionally repressed during the HSR. The exact mechanism responsible for the repression of these and other non-HSP genes during the HSR is currently not fully understood. In mammalian systems, HSF1, the functional homolog to *Drosophila* HSF, has been shown to be responsible for the repression of TNF-α by binding to its
5’UTR (Singh et al. 2002) and IL-6 by stimulating transcription of its repressor (Takii et al. 2010) raising the possibility that HSF binding might also be at least partly responsible for the observed transcriptional repression during the HSR. In support of this possibility, we have found that regression of puffs on polytene chromosomes during the HSR relies on the presence of functional HSF capable of binding DNA at HS temperatures (Stevens 1999).

Prior to the availability of tools such as microarrays that allow the monitoring of the activity of thousands of genes in parallel, only a subset of the genes involved in the 20E response had been identified. A first step in establishing a role for HSF in the direct repression of 20E-response genes therefore is to gain a comprehensive view of the transcriptional events triggered by 20E which can be accomplished by microarray analysis. Once established, this will provide a basis for determining how genes such as those in the 20E regulatory cascade are affected by the simultaneous stimulation of HSF binding and HS-gene induction.

In Chapter 3, I examine the transcriptional profile of Kc cells and the salivary glands of third instar larvae exposed to 20E. For Kc cells, I determine which genes are the early ecdysone inducible genes and which are the late genes. I also examine the effect of hs on the induction of the 20E-regulated genes and show that several of the major early 20E genes have HSF binding sites contained within their genes. These results provide insight into the possible role of HSF in the negative regulation of the 20E response.
Chapter 2
Whole-Genome Analysis Reveals That Active Heat Shock Factor Binding Sites Are Mostly Associated with Non-Heat Shock Genes in *Drosophila melanogaster*

**Please note:** The contents of this chapter are published in PLoS ONE (see Gonsalves et al. 2011). Permission to reproduce this material here is provided under the Creative Commons Attribution License. The following people were all contributing authors: Sarah E. Gonsalves, Alan M. Moses, Zak Razak, Francios Robert, and J. Timothy Westwood and their contributions are as follows: Conceived and designed the experiments: SEG AMM JTW. Performed the experiments: SEG. Analyzed the data: SEG ZR. Contributed reagents/ materials/analysis tools: SEG JTW. Wrote the paper: SEG JTW. Provided advice on experimental design: AMM FR.

2.1 Abstract

During heat shock (HS) and other stresses, HS gene transcription in eukaryotes is up-regulated by the transcription factor heat shock factor (HSF). While the identities of the major HS genes have been known for more than 30 years, it has been suspected that HSF binds to numerous other genes and potentially regulates their transcription. In this study, we have used a chromatin immunoprecipitation and microarray (ChiP-chip) approach to identify 434 regions in the *Drosophila* genome that are bound by HSF. We have also performed a transcript analysis of heat shocked Kc167 cells and third instar larvae and compared them to HSF binding sites. The heat-induced transcription profiles were quite different between cells and larvae and surprisingly only about 10% of the genes associated with HSF binding sites show changed transcription. There were also genes that showed changes in transcript levels that did not appear to correlate with HSF binding sites. Analysis of the locations of the HSF binding sites revealed that 57% were contained within genes with approximately 2/3rds of these sites being in introns. We also found that the insulator protein, BEAF, has enriched binding prior to HS to promoters of genes that are bound by HSF upon HS but that are not transcriptionally induced during HS. When the genes associated with HSF binding sites in promoters were analyzed for gene ontology terms, categories such as stress response and transferase activity were enriched whereas analysis of genes having HSF binding sites in introns identified those categories plus ones related to developmental processes and reproduction. These results suggest that *Drosophila* HSF may be
regulating many genes besides the known HS genes and that some of these genes may be regulated during non-stress conditions.

2.2 Introduction

More than four decades ago Ritossa described a phenomenon where specific loci on the polytene chromosomes from third instar larvae of *Drosophila* decondensed or “puffed” when the larvae were exposed to heat or other forms of stress such as oxidative stress, inhibitors of respiration and certain metals (Ritossa 1962). These puffs represented heat-induced sites of gene transcription and the genes residing there became known as the heat shock (HS) genes and their protein products the heat shock proteins (HSPs). The stress induced molecular and cellular events collectively became known as the heat shock response and is highly conserved in all organisms. During normal and stressed conditions, HSPs and their cognate proteins (HSCs) have essential functions in helping proteins fold properly, acting as protein chaperones during protein synthesis, processing, and degradation as well as the translocation of proteins across intracellular membranes (Parsell and Lindquist 1993; Young et al. 2004). HSPs are also known to have direct and important positive functions in a number of disease conditions and pathophysiological states including immunity against infection, ischemia, neural injury, and neural degenerative diseases (Barral et al. 2004).

Heat shock gene regulation in eukaryotes occurs at the transcriptional and post-transcriptional levels. Stress induced HS gene transcription is governed by the protein factor called Heat Shock Factor (HSF). HSF recognizes and binds to a specific DNA sequence in the promoter of HS genes known as the HS element (HSE) (Amin et al. 1988; Xiao and Lis 1988; Perisic et al. 1989) for a review of HSEs see (Fernandes et al. 1994). Single genes for HSF have been cloned from yeast, fruit flies (*Drosophila*), and frogs, and multiple homologous but distinct HSF genes have been cloned in chickens, mice, and humans. The HSF that is primarily involved in responding to heat and other stress agents has been designated HSF1 in most species with multiple HSFs (for reviews of HSF see (Wu 1995; Cotto and Morimoto 1999; Pirkkala et al. 2001; Ho and Westwood 2002; Voellmy 2004)). HSF is present in cells at all times and is activated to its transcriptionally competent form upon stress. In the metazoans studied thus far, binding of HSF or HSF1 to HSEs is low to virtually nonexistent in unshocked cells and upon
HS or other stresses, HSF converts from a monomer to a trimeric form that binds to the HSEs with high affinity.

HSF is an essential gene in those species that have a single HSF gene (e.g. yeast and Drosophila) even under non-stress conditions. In the case of Drosophila, death was found to occur between the first and second larval instar stages in null mutants suggesting a critical role for HSF even under non-stress conditions (Jedlicka et al. 1997). In addition, the same study found that HSF was required for oogenesis. Furthermore, mice lacking HSF1 can live to adulthood but have a severely compromised stress response and display several other defects including prenatal lethality, growth retardation and female infertility (Xiao et al. 1999). Hsf1−/− female mice also produce defective oocytes that, when fertilized, do not develop very far into embryogenesis (Christians et al. 2000). Mammalian HSF1 and HSF4 play important roles in lens and olfactory epithelium development (Fujimoto et al. 2004; Takaki et al. 2006) and a mutation in HSF4 is associated with heritable cataract formation in humans (Bu et al. 2002). Hsf2−/− mice show embryonic brain defects that persist with adults displaying enlarged ventricles and a decrease in hippocampus size and striatum and cortex width (Kallio et al. 2002; Wang et al. 2003). Moreover, both HSF1 and HSF2 play roles in sperm development in mice (Kallio et al. 2002; Wang et al. 2003; Salmand et al. 2008).

There have been a few genome-wide screens using DNA microarrays to characterize the eukaryotic transcriptional response to HS in C. elegans (GuhaThakurta et al. 2002), human cell lines (Murray et al. 2004; Trinklein et al. 2004), Drosophila embryos (Leemans et al. 2000), and Drosophila adults (Sorensen et al. 2007). In addition to standard expression microarray experiments, others have used chromatin immunoprecipitation coupled with microarrays (ChIP-chip) to find HSF binding sites: in yeast, using probes in intergenic and coding regions (Hahn et al. 2004); in human tissue culture cells for HSF1, using a custom 768 element promoter array (Trinklein et al. 2004); in Drosophila embryos using a 5400 element cDNA array and 3000 element tiling array (Birch-Machin et al. 2005); and in mouse testis for HSF2 using a 26,000 promoter tiling array (Akerfelt et al. 2008). There has also been a recent study that has examined the binding sites for HSF in Drosophila S2 cells using ChIP and next generation DNA sequencing (ChIP-seq) (Guertin and Lis 2010).
When the polytene chromosomes from heat-shocked *Drosophila* 3rd instar larvae were stained with anti-HSF antibodies, HSF was found to be localized to more than 200 loci (Westwood et al. 1991). Given that only nine well documented HS gene loci existed at the time, the authors proposed that HSF had additional genomic targets besides the well-known major HS genes, perhaps stimulating lesser known HSP and HSC genes, other “novel” heat-induced genes. In addition, it was hypothesized that *Drosophila* HSF might also play a role in the transcriptional repression of certain other genes that are known to be repressed during HS. Supporting this hypothesis, HSF1 in human cells has been shown to be a repressor of cytokine genes (Cahill et al. 1996). In this study we have identified more precisely, using ChIP-chip analysis with genome-tiling arrays, more than 430 HSF binding sites in the *Drosophila* genome. We have also performed transcription analysis of heat shocked Kc167 cells and 3rd instar larvae in an attempt to correlate HSF binding events with induction of gene transcription.

### 2.3 Methods

#### 2.3.1 Cell culture and heat-shock treatments

*Drosophila* Kc167 cells (Echalier and Ohanessian 1970), obtained from the *Drosophila* Genomic Resource Center (Indiana University, Bloomington) were grown to confluence in Schneider’s media (Invitrogen) supplemented with 5% heat-inactivated FBS (Sigma) and 20 µg/ml gentamicin (Sigma) in tissue culture flasks at 22°C. Prior to heat-treatment, cells were transferred to Erlenmeyer flasks and aerated for 4 hrs at 22°C by gentle shaking (~180 rpm). Following aeration, half of the cells were heat-shocked by submersing the flask in a 36.5°C circulating water bath (Neslab) for 30 minutes. The remaining cells were maintained as room temperature controls. For the cycloheximide experiments, 118 µM cycloheximide was added 10 minutes prior to initiating the heat shock treatment (Zimarino et al. 1990) which was otherwise carried out as stated above.

#### 2.3.2 ChIP

Cell cross-linking, lysis and chromatin shearing were all performed essentially as reported by Weinmann and Farnham (Weinmann and Farnham 2002). In brief, flasks with ~1x10^8 cells in 20 ml were heat-shocked or left at room temperature for 30 minutes as described in section 2.3.1 and then cross-linked by adding 540 µl of 37% formaldehyde to the cell culture medium.
(Schneider’s media (Invitrogen) lacking serum) to a final concentration of 1%. Cross-linking was carried out for 10 minutes at RT. To stop the cross-linking reaction, 125 mM glycine was added for 5 min at RT. Fixed cells were then washed 2X with 20 ml PBS (pH 7.4 + 0.5 mM PMSF) and lysed by pipetting in 10 ml cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 1 mM PMSF, 1 µl/ml aprotinin, 5 µl/ml leupeptin). After a 10 min incubation on ice, nuclei were pelleted (5,000 rpm, 5 min at 4°C) and lysed by pipetting in 4 ml nuclei lysis buffer (50 mM Tris-HCl pH 8.1, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1 µl/ml aprotinin, 5 µl/ml leupeptin). Following a 10 min incubation on ice, chromatin was sonicated in 15 ml conical bottom tubes with an Ultrasonic Cell Disruptor (model W-220F) on setting 5, 10x on ice with 10 second pulses and 30 seconds rest on ice with brief vortexing in between pulses to produce chromatin fragments with an average size of ~700 bp. Intact nuclei were removed by pelleting (15,000 rpm, 10 min at 4°C) and the supernatant diluted 2X with IP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 1 mM PMSF, 1 µl/ml aprotinin, 5 µl/ml leupeptin). Dynabeads protein G magnetic beads (DMB; Invitrogen Cat. No. 100.03D) were prepared for use by pre-blocking with freshly made PBS/BSA (5 mg/ml BSA in PBS) and washing 2X and re-suspending in dialysis buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.20% sarkosyl). Chromatin was pre-cleared with pre-blocked DMB (25 µl of beads/10⁸ cells) and a 100 µl sample of total input chromatin was retained for later use. Pre-cleared chromatin was then incubated in 13 ml round bottom tubes over night on a rotating platform at 4°C with (ChIP) or without (mock IP) polyclonal anti-HSF antibody (943) (Westwood et al. 1991) at a dilution of 1:7500. Immunocomplexes were isolated with pre-blocked DMB (50 µl of beads per 10⁸ cells; 3 hr incubation at 4°C), transferred to 1.5 ml microfuge tubes and washed 2X with 1.5 ml dialysis buffer then 4X with 1.5 ml IP wash buffer (100 mM Tris-HCl pH 9.0, 500 mM LiCl, 1% NP-40, 1% deoxycholic acid). Immunocomplexes were eluted from the beads by adding 150 µl IP elution buffer (50 mM NaHCO₃, 1% SDS) for 10 min at 65°C, vortexing for 15 min at room temperature and pelleting the beads by centrifugation. The elution step was performed twice and the eluents pooled. 1 µg of RNase per 10⁷ cells was added to the immunocomplex containing supernatant followed by NaCl to a final concentration of 300 mM. To reverse the cross-links, samples were incubated at 67°C for 5 hr. To precipitate DNA, samples were incubated at -20°C in 2.5 volumes of EtOH overnight. Precipitated DNA was re-suspended in TE and diluted with proteinase K buffer (50 mM Tris-HCl pH 7.5, 25 mM EDTA, 1.25% SDS). 5 µg of proteinase K was added and samples were
incubated for 2 hr at 45°C to digest unwanted protein. DNA was purified with a Purelink PCR cleanup kit (Invitrogen).

2.3.3 End-point and qPCR

Endpoint PCR was performed on DNA from HS anti-HSF ChIP, mock ChIP and WCE as well as RT anti-HSF ChIP, mock ChIPs and WCE with primers designed against the Hsp26 promoter and on DNA from unamplified and amplified HS anti-HSF and mock ChIPs with primers designed against regions upstream of the following genes: Hsro, Hsp70Ab, Hsp83, CG11267, stv, CG5290, Tom34, DnaJ-1, CG10077, Taf7, and GstD2. Cycling conditions used were as follows: 95°C for 3’ followed by 35 cycles of: 95°C for 30”, 58°C for 45” and 72°C for 30”. Quantitative PCR was performed on the same DNA samples with the same primers plus additional primers designed to amplify the region of chromosomal DNA 1200 bp downstream of the Hsp26 promoter. qPCR reactions were performed with Brilliant SYBR Green (Stratagene) in a MX4000 lightcycler (Stratagene) under the same conditions as above (but for 40 cycles instead of 35). See Additional File A-1 for the sequences of the primers used.

2.3.4 Ligation-Mediated PCR

The immunoprecipitated sheared chromatin was repaired as described by Ren and colleagues (Ren et al. 2000). Briefly, 50 ng of ChIP/input DNA was repaired with 3U T4 DNA polymerase (NEB M0203S) in 110 µl total reaction volume for 60 minutes at 37°C. Reactions were cleaned up with Invitrogen Purelink PCR cleanup columns according to manufacture’s protocols and eluting in 30 µl Sigma water. Linker DNA used in ligation-mediated PCR (LM-PCR) was prepared according to Birch-Machin and colleagues (Birch-Machin et al. 2005) as was ligation of the repaired DNA to the linker and PCR amplification of the ligated chromatin. Briefly, the blunt end linker mix used in the ligation reaction was prepared by mixing linker 1 (5’ Phos-AGA AGC TTG AAT TCG AGC AGT CAG) and linker 2 (5’ CTG CTC GAA TTC AAG CTT CT) and incubating in a thermocycler for 2 minutes at 94°C, 5 minutes at 70°C, 5 minutes at 55°C and then allowing the mix to cool down to room temperature by turning off the thermocycler. Ligation of the blunt end linker to repaired ChIP/input DNA was carried out by incubating 1 µl of 1 mM linker mix and 5 ng of repaired DNA with 1 U T4 ligase (Invitrogen 15224-017) in a 10 µl total reaction volume overnight at 4°C. PCR amplification of the ligated
DNA was performed using 4 µM of linker 2 as a primer under the following cycling conditions: one cycle of 92°C for 2 minutes, 20 cycles of 92°C for 20 seconds, 43°C for 30 seconds, 53°C for 30 seconds, 72°C for 1 minute, 1 cycle of 72°C for 5 minutes. The reaction was cleaned up with Invitrogen Purelink PCR cleanup columns according to manufacturer’s protocols and eluting in 30 µl sigma water.

2.3.5 Indirect labeling of amplified chromatin and hybridization to genome-tiling arrays

A second round of PCR similar to that performed for LM-PCR was used to incorporate amino-allyl modified nucleotides into the amplified material. Following amplification, DNA clean-up, fluorescent dye conjugation and probe clean-up and precipitation was performed as described on the Canadian Drosophila Microarray Centre web site (www.flyarrays.com). Briefly, to conjugate the Alexa dyes, amino allyl-modified DNA was re-suspended in 0.3 M NaHCO₃, combine with 2 µl of dye (either Alexa 647 (A32757) or Alexa 555 (A32756)) and incubated for 1 hour at room temperature. To stop the conjugation reaction and remove unbound dye, the DNA was diluted with water and run through Invitrogen Purelink PCR cleanup columns according to manufacture’s protocols but performing the wash step 3X with 80% EtOH and eluting 3X in 50 µl sigma water. Labeled DNA from the ChIP or mock IP samples was mixed with an equal volume of differentially labeled input DNA and precipitated with isopropanol. Precipitated probe DNA was re-suspended in 120 µl and mixed with control nucleic acids (750 ng salmon sperm DNA, 40 µg yeast tRNA, 10 µg human cot-1 DNA in a total volume of 27.5 µl) and then added to 352.5 µl of hybridization buffer (50 mM Na-MES pH 6.9, 500 mM NaCl, 6 mM EDTA, 0.5% ultrapure sarcosine, 30% ultrapure formamide) heated to 95°C for 3 min and then incubated at 40°C for 15 min. Labeled DNA from HS samples was hybridized to Agilent 2X 244 k genome-tiling arrays (catalog number G4495A) and labeled DNA from RT samples was hybridized to Agilent 1X 1 million probe arrays (custom designed to contain the same probes as the 2X 244k arrays). Both arrays contain the same 60 mer probes, approximately 475,000 in total (i.e a probe every 233 nucleotides). Hybridizations were performed according to manufacturer’s directions with Agilent hybridization chambers (catalog number G2534A) and a total hybridization volume of 500 µl. After 20 hours of hybridization at 20 rpm, the slides were washed for 5 min with 6x SSPE, 0.005% ultrapure N-lauroylsarcosine, again for 5 min, 0.6x SSPE, dipped in acetonitrile and washed for 30s in Agilent’s Wash III. Dried slides were
scanned with Agilent’s microarray scanner (catalog # G2505-60502) and the resulting images were quantified with Agilent’s Feature Extraction software version 10.7.3.1.

2.3.6 HSF binding site identification

Data from Feature Extraction was normalized with Agilent’s ChIP Analytics software version 1.3. Blank subtraction normalization, inter-array median normalization and intra-array (dye-bias) median normalization were all applied. Probes were mapped to release 4.2 of the Drosophila genome. The Whitehead Error Model v1.0 and Whitehead Per-Array Neighborhood Model v1.0 were used with the default settings for error modeling and for peak detection/evaluation, respectively with a false discovery rate of 11%.

2.3.7 PWM matching

We used the pattern matching program Patser (Hertz and Stormo 1999) to find matches to four PWMs from TRANSFAC representing canonical (M00165) and non-canonical (M00163, M00164, M00166) HSEs in the sequence flanking each identified bound peak (+/-2500bp). For each matrix, we counted the number of matches with p-values below $1 \times 10^{-4}$ and $1 \times 10^{-5}$ in 50 bp windows relative to the segment peak and generated a frequency histogram. To determine if there is a significant difference in the number of matches in the region immediately surrounding the peak (-500bp to +500bp) relative to the local background (-1000bp to -550bp and +550bp to +1000bp) we performed a $X^2$ test. We repeated these steps to identify matches to all other unique PWMs from TRANSFAC and from the Drosophila DNase I Footprint Database this time counting matches to each matrix in HSF-bound promoters and HSF-bound introns with Bonferroni corrected p-values <5.6x10^{-2}. To control for base composition bias of the test sequence, for any matrix exhibiting an enrichment of binding sites in the peak region relative to the local background comparable to the enrichment seen for the canonical HSE, we repeated the test with a scrambled version of the matrix and then threw out any matrix still showing enrichment. For any remaining matrices, we performed a $X^2$ test to determine if there is a significant difference in the number of sites matching the matrix in the HSF-bound promoters relative to HSF-bound introns.
2.3.8 Larval heat shock treatment

Late third instar larvae (*dp cn bw* and *Hsf* cn *bw*) (Jedlicka et al. 1997) were selected by the blue gut method as previously described (Andres and Thummel 1994) and transferred to 2 ml screw cap tubes containing a strip of moist blotting paper with no more than 20 larvae per tube. Larvae were allowed to acclimatize for 1 hour at RT with loose lids and then either submerged in a 36.5°C circulating water bath (Neslab) for 30 min or kept at RT (22°C) for the same amount of time. Following treatment, larvae were snap frozen in liquid nitrogen.

2.3.9 RNA extraction, cDNA synthesis, and labeling and hybridization to expression microarrays

Total RNA was extracted from both cells and larvae using TRIzol reagent (Invitrogen) according to the manufacture’s protocol. Quality and quantity of RNA was verified by measuring the absorbance and the A260/A280 ratios were always above 1.8. cDNA synthesis, labeling and hybridization to *Drosophila* 385k (catalog #A4351001-00-01 design ID 4608) or 4X 72k (catalog #A4509001-00-01 design ID 080915_DM_TW) NimbleGen expression microarrays (Roche) was carried out as described in the manufacture’s protocol with the exception that HS and RT samples were differentially labeled and hybridized to a single array. Briefly, 30 µg of total RNA was reverse transcribed with SuperScript II (Invitrogen). Second strand cDNA synthesis was performed using T4 DNA polymerase (Invitrogen). RNase A was used to remove residual RNA template and a phenol:chloroform extraction was performed to clean up the reaction. 500 ng of double stranded cDNA was labeled with Cy-dyes using klenow (3’ → 5’ exo) (Roche). After isopropanol precipitation of the labeled cDNA, 2 µg of Cy5 labeled cDNA generated from experimental samples was mixed with 6 µg of Cy3 labeled cDNA from the respective untreated control sample and hybridized to the 385k or 72k NimbleGen expression microarrays (Roche). For each treatment, three independent biological replicates were performed. For the ecdysone plus heat shock experiments, a cDNA based microarray was used and the microarray experiments and analysis were carried out following the methods of Neal and co-workers (Neal et al. 2003). Briefly, 80 µg of total RNA from cells was reverse-transcribed with SuperScriptII (Invitrogen) in the presence of cyanine (Cy) dye coupled nucleotides (Perkin Elmer). After isopropanol precipitation of the labeled cDNA, Cy5 labeled cDNA generated from experimental samples was mixed with Cy3 labeled cDNA from the
respective untreated control sample and hybridized to the 12k_v1 cDNA microarray from the Canadian *Drosophila* Microarray Centre (CDMC).

### 2.3.10 Expression microarray data extraction and analysis

Images acquired after scanning slides with GenePix 4000B microarray scanner (Molecular Devices). PMT gain settings for each channel were set between 550 and 750 such that the normalized count ratio of the two channels was ~1.0 and the number of normalized counts at the 65,000 intensity level was ~1e-5. Scanned images were quantified and RMA normalized by container with background correction with NimbleScan version 2.4.27 (Roche). ArrayStar version 3.0 (DNASTAR) was used to analyze the resulting data files and identify genes with an average fold change across all biological replicates of 2 fold or greater and FDR corrected p-values less than 0.01. Log-converted expression ratios were clustered in the microarray data analysis software MeV (Saeed et al. 2003; Saeed et al. 2006) using the Manhattan Distance Metric and average linkage method.

### 2.3.11 Functional Enrichment of Gene Lists

Functional enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery Bioinformatics Resources (DAVID) (Dennis et al. 2003; Huang et al. 2009). Lists of Flybase Gene Identifiers from genes that were differentially expressed during HS in either cells or larvae and from genes that were bound by HSF in either the promoter regions or intronic regions were input into the functional annotation clustering tool and functional annotation chart tool. For genes with HSF binding sites in their intronic regions, annotations were compared to the pool of annotations found for all genes with introns, whereas all other lists were compared to the entire genome. Highly related groups of enriched annotations with modified Fisher Exact p-values (p-values) less than 0.01 were identified from each of the 4 gene lists and the corresponding p-values from the DAVID analysis for those annotation terms from each of the lists were compared to each other using a heat map.
2.4 Results

2.4.1 Identification of HSF binding sites in the *Drosophila* genome

We performed ChIP-chip analysis on heat-shocked Kc167 (Echalier and Ohanessian 1970) to identify HSF binding sites across the *Drosophila* genome (data available in GEO under GSE19744). HSF binding should reach a maximum level following a 30-minute heat-shock (HS) at 36.5°C (Wu et al. 1994; Shopland and Lis 1996) so we conducted our heat-shock treatment under this condition. We fixed both heat-shocked and non-shocked cells with formaldehyde to preserve protein-DNA interactions and then immunoprecipitated HSF bound chromatin complexes with an anti-HSF antibody generated by Westwood et al., (Westwood et al. 1991). This antibody is specific for HSF and has been used to visualize HSF binding sites at over 200 loci in *Drosophila* polytene chromosomes by indirect immunofluorescence (Westwood et al. 1991). As a control for non-specific binding and for HSF binding under non-shock conditions, we also performed a mock ChIP without antibody and an anti-HSF ChIP at room temperature ((RT), 22°C) respectively. We confirmed that our ChIP had successfully enriched for HSF-bound chromatin in HS cells but not mock treated or RT cells by measuring the relative abundance of *Hsp26* promoter DNA by PCR and qPCR (Figure 2-1). Furthermore, we confirmed that our ChIP conditions were specific enough that we did not get enrichment of the sequence 1200bp downstream of the *Hsp26* promoter (Figure 2-1B). Following confirmation, we amplified fragments from two independently produced HS anti-HSF ChIP, RT anti-HSF ChIP, and mock ChIP samples by ligation mediated-PCR, labeled them with fluorescent dyes, and hybridized them to genome tiling arrays (Agilent Technologies) as described in the Methods section.

Agilent’s Feature Extraction software quantified images of the arrays and Agilent’s ChIP Analytics software identified probes corresponding to regions or segments of chromatin that were bound by HSF. We compared the anti-HSF and mock ChIP segments to determine if any of the anti-HSF ChIP segments were non-specific. Only two ChIP segments exhibited any degree of overlap to mock ChIP segments: the first (chr3R:11,071,788-11,074,349) only partially overlaps with a mock ChIP segment and exhibits a much stronger min P[Xbar] on the ChIP array (1.16x10^{-11} vs 6.36x10^{-4}). Furthermore, this segment is contained in one of the loci
Figure 2-1 Confirmation of HSF binding to select regions.

(A) Enrichment of the region upstream of the Hsp26 gene by HSF ChIP following heat shock (HS; 36.5°C, lane 1) relative to whole cell extract (WCE; lane 4), HSF ChIP at room temperature (RT; 22°C, lane 3) and no antibody mock ChIP following HS (lane 2) by end-point PCR. (B) qPCR confirmation of enrichment of the same region as in (A) by HSF ChIP following HS (light grey) relative to HSF ChIP at RT (dark grey) and of a region 1200 bp downstream of the Hsp26 gene. (C) PCR amplification of select regions associated with the genes indicated on both an anti-HSF IP enriched sample (left column) and mock IP sample (right column).
(88E) bound by HSF on polytene chromosomes (Westwood et al. 1991) and so was retained as part of the dataset. The second segment (chr3L:18,124,038-18,125,011) completely overlaps with a mock ChIP segment and has a comparable min P[Xbar] to that of the mock ChIP segment, therefore, this segment was omitted from further analysis.

In total we identified 434 HSF bound chromatin segments including regions associated with all but one of the known major heat-inducible genes. A selection of targets was confirmed by PCR (Figure 2-1C). No HSF binding site was detected upstream of any of the Hsp70B genes due to the absence of probes on the tiling array in this highly repetitive region of genome (Figure 2-2B,C). We next examined regions bound by HSF at RT (non-HS conditions) and found that 81% coincided with segments bound by HSF under HS conditions, however, in every case, the level of HSF binding is substantially less at RT (i.e. 5-fold lower on average) (Figure 2-3 and Additional File A-2 data also available in GEO under GSE22335). The most highly enriched HSF binding site at RT is located upstream of Hsp83 in one of the only regions specifically occupied by HSF under non-HS conditions (Wu et al. 1994). With the possible exception of this site, the observed weak HSF binding in the RT samples may reflect transient HSF binding, HSF binding in a subset of cells, and/or is the result of the induction of a mild HS response brought on by the initial harvesting and fixation of the cells. Table 2-1 lists all sites bound by HSF during HS treatment with a greater than 30-fold enrichment over whole cell extract (WCE) including all the sites associated with the major heat-inducible genes. Where applicable, the staining intensity observed on polytene chromosomes by Westwood et al.,(Westwood et al. 1991) is indicated. Twenty-nine of our top 40 sites map to 24 loci that overlap with HSF-bound loci in polytene chromosomes. When we consider our entire set of HSF-bound sites the overlap with the polytene data is significant (p-value from X2 test = 2.5x10-10; Figure 2-2A). Of those 73 loci identified by Westwood et al.,(Westwood et al. 1991) that did not overlap directly with our ChIP-chip data, we found that 54 are within one cytological band of at least one HSF binding site (data not shown). Such an offset is within the estimated error rate associated with computing cytological locations based on sequence location (Flybase Reference Manual G, section G.5.1 (Tweedie et al. 2009)). If we consider these 54 sites offset by one cytological band, together with the 108 that directly overlap, then 90% of the HSF-bound loci identified on polytene chromosomes are covered in our ChIP-chip data. A X2 test on the independence of the two datasets taking into account the offset indicates that there is overlap between them (p-value from X2 test = 4.5x10-6).
Figure 2-2 Overlap of HSF binding sites indentified by ChIP-chip of Kc cells and immunostaining of polytene chromosomes of 3rd instar larvae.

(A) The 434 HSF-bound sites indentified by ChIP-chip analysis were mapped to 265 unique cytlocations for this comparison. (B) Zoomed-in view of the region of chromosome 3R where the Hsp70B genes are located highlighting both the absence of tiling-array probes and the repetitive and/or low complexity sequence in this region as indicated by the RepeatMasker track (bottom). (C) Expanded view of (B) to show the location of the nearest HSF-bound (green circle). This image, generated using the UCSC Genome Browser, illustrates the chromosome region represented in bp (according to release 4.2 of the Drosophila genome) as indicated at the top. Genes are depicted as blue boxes with the thick and thin parts representing exons and introns respectively. Arrows (either blue or white) within the gene indicate the direction of transcription. Vertical black lines show the location of each probe on the Agilent 2 x 244k tiling arrays (Agilent Technologies) in the depicted region.
A. HSF-bound loci identified indirect immunostaining of *Drosophila* polytene chromosomes (Westwood et al., 1991)

- 73
- 108
- 156

HSF-bound loci identified by ChIP-chip in this study

B. Chromosome 3R Tiling array probes

C. Chromosome 3R HSF-bound segment Tiling array probes
Figure 2-3 Comparison of HSF binding under HS and non-HS (RT) conditions.

(A) Scatter plot of the HSF binding ratio of representative probes from segments bound by HSF during both HS and RT (blue diamonds), during HS only (red squares), or at RT only (green triangles). Although most of the sites bound by HSF at RT overlap with sites bound during HS, levels of HSF binding are greatly diminished at RT (blue diamonds). As expected the only region to be strongly bound by HSF at RT is the region upstream of Hsp83.

(B) Level of HSF binding during HS (top) and at RT (bottom) to chromosome 3R. Each bar represents a probe from the tiling array that is part of an HSF bound segment and its height indicates its fold enrichment relative to WCE.
### Table 2-1 Chromatin segments bound by HSF exhibiting a 30 or greater fold enrichment over whole cell extract (WCE).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Location of the HSF binding site</th>
<th>Fold</th>
<th>Polytene staining intensity</th>
<th>Closest gene(s)</th>
<th>Position of peak relative to gene TSS (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3R:17121946..17122005</td>
<td>96</td>
<td>1.10E-07</td>
<td>98</td>
<td>93D1</td>
<td>Hsru</td>
</tr>
<tr>
<td>2L:13165433..13165492</td>
<td>76</td>
<td>3.70E-11</td>
<td>-161</td>
<td>34A</td>
<td>Sir2 &amp; DnaJ-H</td>
</tr>
<tr>
<td>3R:19983079..19983127</td>
<td>75</td>
<td>6.80E-08</td>
<td>25</td>
<td>95D</td>
<td>Hsp68</td>
</tr>
<tr>
<td>3L:93051600..93051707</td>
<td>75</td>
<td>8.00E-08</td>
<td>0</td>
<td>67B</td>
<td>Hsp26 &amp; Hsp67Ba</td>
</tr>
<tr>
<td>3R:7783512..7783571</td>
<td>73</td>
<td>2.00E-10</td>
<td>-447</td>
<td>87A</td>
<td>Hsp70Ab, Hsp70Aa</td>
</tr>
<tr>
<td>2L:2258186..2258245</td>
<td>73</td>
<td>3.20E-09</td>
<td>74</td>
<td>21B</td>
<td>Hop</td>
</tr>
<tr>
<td>3L:3176334..3176591</td>
<td>71</td>
<td>7.60E-08</td>
<td>-102</td>
<td>63B</td>
<td>Hsp83</td>
</tr>
<tr>
<td>3L:9305629..9305688</td>
<td>70</td>
<td>6.70E-08</td>
<td>-3</td>
<td>67B</td>
<td>Hsp27 &amp; Hsp23</td>
</tr>
<tr>
<td>3R:11056966..11056925</td>
<td>65</td>
<td>3.60E-12</td>
<td>22</td>
<td>88E</td>
<td>Hsc70-4</td>
</tr>
<tr>
<td>2L:22008546..22008605</td>
<td>60</td>
<td>1.60E-05</td>
<td>-51</td>
<td>79B</td>
<td>CG7133</td>
</tr>
<tr>
<td>3L:9347665..9347744</td>
<td>58</td>
<td>1.90E-04</td>
<td>-142</td>
<td>67B</td>
<td>Hsp22 &amp; Hsp67Bb</td>
</tr>
<tr>
<td>X:20846288..20846587</td>
<td>54</td>
<td>1.20E-06</td>
<td>-123</td>
<td>19E</td>
<td>Ntf-2</td>
</tr>
<tr>
<td>X:5722522..5725281</td>
<td>52</td>
<td>1.00E-07</td>
<td>-62</td>
<td>5C</td>
<td>CG16721</td>
</tr>
<tr>
<td>3R:27045260..27045409</td>
<td>49</td>
<td>1.70E-07</td>
<td>NA</td>
<td>100B</td>
<td>CG1746</td>
</tr>
<tr>
<td>3L:16745454..16745513</td>
<td>48</td>
<td>2.10E-07</td>
<td>133</td>
<td>73C</td>
<td>CG9705</td>
</tr>
<tr>
<td>2L:22342021..22342080</td>
<td>46</td>
<td>2.20E-11</td>
<td>135</td>
<td>40F</td>
<td>CG17018</td>
</tr>
<tr>
<td>X:10954037..10954096</td>
<td>42</td>
<td>2.30E-07</td>
<td>73</td>
<td>10A</td>
<td>Hsp60</td>
</tr>
<tr>
<td>3L:12906240..12906299</td>
<td>38</td>
<td>2.00E-07</td>
<td>88</td>
<td>69F</td>
<td>CG11267</td>
</tr>
<tr>
<td>2L:6966386..6966445</td>
<td>38</td>
<td>9.80E-12</td>
<td>-25</td>
<td>27C</td>
<td>smt3</td>
</tr>
<tr>
<td>3R:12473487..12473539</td>
<td>37</td>
<td>7.30E-12</td>
<td>15</td>
<td>89D</td>
<td>Cctgamma</td>
</tr>
<tr>
<td>3L:13454266..13454312</td>
<td>36</td>
<td>5.50E-07</td>
<td>3</td>
<td>70B</td>
<td>stv</td>
</tr>
<tr>
<td>3L:3866074..3866133</td>
<td>36</td>
<td>2.40E-07</td>
<td>32</td>
<td>63F</td>
<td>Ubi-p63E</td>
</tr>
<tr>
<td>2L:5000964..5000923</td>
<td>36</td>
<td>2.40E-08</td>
<td>-16</td>
<td>25C</td>
<td>Rtnl1</td>
</tr>
<tr>
<td>3R:9205991..9206650</td>
<td>36</td>
<td>2.60E-11</td>
<td>107</td>
<td>87E</td>
<td>Dro2</td>
</tr>
<tr>
<td>2L:12406460..12406519</td>
<td>35</td>
<td>2.40E-11</td>
<td>218</td>
<td>33B</td>
<td>CG6770</td>
</tr>
<tr>
<td>3R:11072836..11072985</td>
<td>35</td>
<td>1.20E-11</td>
<td>17</td>
<td>88E</td>
<td>Hsc70-4</td>
</tr>
<tr>
<td>3L:7839477..7839536</td>
<td>35</td>
<td>6.10E-07</td>
<td>-6</td>
<td>66A</td>
<td>Pdp1</td>
</tr>
<tr>
<td>3L:1787705..17877764</td>
<td>34</td>
<td>4.30E-07</td>
<td>-30</td>
<td>75A</td>
<td>CG5290</td>
</tr>
<tr>
<td>3L:8419259..8419281</td>
<td>34</td>
<td>1.00E-06</td>
<td>186</td>
<td>66D</td>
<td>CG6776</td>
</tr>
<tr>
<td>X:11208411..11208540</td>
<td>34</td>
<td>2.50E-06</td>
<td>48</td>
<td>10B</td>
<td>CG11750</td>
</tr>
<tr>
<td>X:2503300..2503359</td>
<td>33</td>
<td>1.60E-07</td>
<td>6</td>
<td>3A</td>
<td>sgg</td>
</tr>
<tr>
<td>X:6499617..6499676</td>
<td>33</td>
<td>1.30E-07</td>
<td>106</td>
<td>6C</td>
<td>CG3226</td>
</tr>
<tr>
<td>2L:22150737..22150796</td>
<td>32</td>
<td>2.40E-11</td>
<td>NA</td>
<td>40F</td>
<td>CG1832</td>
</tr>
<tr>
<td>U:5800960..5801039</td>
<td>32</td>
<td>1.20E-07</td>
<td>40</td>
<td>53F</td>
<td>CAP</td>
</tr>
<tr>
<td>3R:2950969..2950918</td>
<td>32</td>
<td>6.10E-07</td>
<td>13</td>
<td>99C</td>
<td>kay</td>
</tr>
<tr>
<td>3R:3859929..3859318</td>
<td>31</td>
<td>4.50E-11</td>
<td>89</td>
<td>84E</td>
<td>Tom34 &amp; CG11035</td>
</tr>
<tr>
<td>3L:112000..112059</td>
<td>31</td>
<td>2.40E-07</td>
<td>16</td>
<td>61B</td>
<td>Pk61C</td>
</tr>
<tr>
<td>3L:8578674..8578823</td>
<td>31</td>
<td>3.10E-07</td>
<td>86</td>
<td>66D</td>
<td>2</td>
</tr>
<tr>
<td>3L:1400278..14002737</td>
<td>30</td>
<td>1.30E-07</td>
<td>171</td>
<td>70C</td>
<td>Hsc70Cb</td>
</tr>
<tr>
<td>2L:16517060..16517119</td>
<td>30</td>
<td>8.20E-11</td>
<td>NA</td>
<td>36A</td>
<td>CG5953</td>
</tr>
<tr>
<td>3L:19838272..19838331</td>
<td>30</td>
<td>1.80E-06</td>
<td>-31</td>
<td>76D</td>
<td>Su(Tpl)</td>
</tr>
</tbody>
</table>

1 While no staining of HSF was noted on polytene chromosomes at 93D, a staining intensity of 1.5 was observed at 93C.
2.4.2 Heat Shock Elements (HSEs)

The HSF bound segments identified in our analysis span several oligonucleotide probes from the tiling array and average 1400bp in length. Thus, for each HSF bound segment we assigned a “peak” as the center of probe in the segment with the lowest P[Xbar] (i.e. the probe with the lowest probability that the observed difference between ChIP and WCE signals is due to non-biological causes). An example of how a peak identified in this way compares to the HSF-bound segments and probes identified by ChIP Analytics is presented in Figure 2-4. Figure 2-4A depicts the typical scenario where the HSF-bound probe with the lowest P[Xbar] in a given segment also exhibits the highest fold change relative to WCE in that segment and is approximately in the center of that segment. In the situation where clusters of HSF-bound sites are found over small distances, assigning a single peak to segments that may represent more than one binding site may result in an underestimation of the total number of true sites. This is the case for the HSP gene dense region on chromosome 3L where there is no peak for the HSF-binding site upstream of Hsp23 because it is incorporated into the neighboring segment due to its proximity (Figure 2-4B). Using Patser (Hertz and Stormo 1999), we scanned 2500bp of sequence flanking each peak to find matches to the position weight matrix (PWM) representing the canonical 15 bp HSE from TRANSFAC (Figure 2-5A). A $X^2$ test revealed a significant difference in the number of matches to this motif between the sequence around the peak and the flanking sequence (p-value $= 1 \times 10^{-59}$, Figure 2-5A) suggesting that assigning the peak as stated above was reasonable. This analysis also suggested that the most HSE rich region lies between -400bp to +300bp of each assigned peak (Figure 2-5A). The position of the HSE (p-value $< 1 \times 10^{-4}$) closest to the peak of each binding site is listed in Table 2-1 and in Additional File A-3.

Because the canonical HSE is composed of 3 repeats of the 5bp motif nGAAn we sought to determine if other orientations of this motif were enriched in the peak region relative to the local background. We used Patser to determine the number of matches to each PWM for all possible 3-way combinations of the 5bp motif represented in TRANSFAC and plotted a histogram to depict the distribution of matches (Figure 2-5B). Two of the alternate orientations are slightly enriched near the center of the fragments although not as strongly as the canonical motif (compare left columns of Figure 2-5A and Figure 2-5B: nGAAnnGAAAnnTTCn and nTTCnnGAAAnnGAAAn). When the stringency of what may be considered a “match” to the PWM was increased (i.e. by decreasing the Paster p-value from $< 1 \times 10^{-4}$ to $1 \times 10^{-5}$, right
Figure 2-4 Representation of select genomic regions bound by HSF.

This image, generated using the UCSC Genome Browser, illustrates the chromosome region represented in bp (according to release 4.2 of the Drosophila genome) as indicated at the top. Genes are depicted as blue boxes with the thick and thin parts representing exons and introns respectively. Arrows (either blue or white) within the gene indicate the direction of transcription. Large black and small grey boxes represent HSF-bound segments and probes respectively identified by Chip Analytics (Agilent Technologies). Darker grey shading is used to represent probes with higher fold-enrichment relative to WCE. The single black line above bound segments indicates the position assigned as the segment peak and is also the center of the probe with the lowest P[Xbar]. (A) A single HSF-bound segment is found in the region encompassing Hsp70A. The probe with the lowest P[Xbar] is located near the numerical center of the bound segment. (B) Three HSF-bound segments are found in the region encompassing the small HSP genes. It is likely that these segments represent more than 3 distinct HSF-binding sites, however, a single peak per segment has been assigned potentially resulting in an underestimation of the number of individual binding sites in this region.
Figure 2-5 The 5 bp motif nGAA\textit{n} arranged as direct inverted repeats is enriched in HSF-bound segments.

We used the pattern-matching program Patser (Hertz and Stormo 1999) to score the occurrences of 4 PWMs from TRANSFAC (M00165, M00164, M00167, M00166), depicted as sequence logos on the right side of the figure, in both the sequence bound by HSF and the local background (up to 2500 bp on either side of the peak of each HSF binding site). The histograms show the number of matches to each PWM in 50 bp windows centered on the peak of each HSF-bound segment. P-values at the top of each column of graphs indicate the cut-off used when considering a match by Patser. P-values at the top right of each histogram indicate the probability given by a $\chi^2$ test that the difference in the number of matches to the PWM found in the sequence at the peak and in the local background is due to chance. (A) The motif nGAA\textit{n}TTC\textit{n}GAA\textit{n} shows significant enrichment at the peak of the HSF-bound regions at both Patser p-value cut-offs (compare left and right columns) and occurs more frequently than any other orientation of this motif in HSF-bound regions (compare A and B). (B) Other orientations of the 5 bp core motif not as significantly enriched as the motif in (A), if at all, in the HSF-bound segments. What little enrichment is seen for the alternate orientations when the Patser p-value cut-off is set to $< 1 \times 10^{-4}$ (left column) is essentially lost when the Patser p-value cut-off is lowered to consider only highly probable matches (p-value $< 1 \times 10^{-5}$, right column).

Sequence logos were generated from TRANSFAC PWMs M00165, M00164, M00167, M00166 using the online web tool enoLOGOS (Workman et al. 2005).
columns of Figure 2-5B), the number of matches to these alternate motifs were however, essentially reduced to background levels, while the total number of matches to the canonical HSE was still significantly above background levels (p-value = 1 x 10^{-30}, right column of Figure 2-5A). Taken together, the result of this analysis suggests that the inverted repeat arrangement is strongly favored over all of the orientations examined.

### 2.4.3 Genes associated with HSF binding events

Because the previous analysis of HSF binding events on polytene chromosomes was not of sufficient resolution to determine which genes HSF associated with at heat shock temperatures on a global scale, we set out to determine what genes may be affected by HSF binding by identifying the genes closest to each HSF binding site. In a first attempt to identify genes that may be regulated by HSF binding, we identified the nearest transcription start site (TSS) to the peak of each HSF bound segment. The result of this analysis is included in Table 2-1 for the most strongly bound sites and in Additional File A-3 for these and all the remaining sites. Since we cannot rule out the possibility that HSF may be acting on more distant genes, we extended this analysis to identify all genes within a 2500 bp window centered on the peak of each HSF-bound segment and calculated the distance from peak to TSS for each of these genes. For any given binding site, there may be more than one gene within the 2500 bp window so Additional File A-3 lists all genes found within the window in order of proximity to the binding site.

In the course of this analysis, it became apparent that there were instances in which HSF was binding within the transcribed region of many genes. As such, we investigated the proportion of sites that were found within transcribed regions (intragenic) relative to those that were not (intergenic) (Figure 2-6). In total, 57% of all sites were located in the transcribed region of at least one gene with a preference for binding within introns (Figure 2-6). In contrast, only 41% of euchromatic sequence is intragenic. An example of HSF-binding within an intron is presented in Figure 2-7 for the transcription factor \textit{jumu}. Roughly 1/4 of sites in transcribed regions were, however, also located in the proximity of gene promoters (which we are considering to be a region surrounding 1250bp from the transcription start site). By our definition, 14% of the genome falls in gene promoters, however, of the 43% of sites that are intergenic, over half were found in promoter regions representing 27% of all HSF binding sites (Figure 2-6). It should be noted that the promoters of the major HSP genes account for less than 5% of all HSF bound
Figure 2-6  Breakdown of HSF-binding site by location.
Fewer than 1/3 of all HSF-binding sites are located in gene promoters with the promoters of major heat-inducible genes accounting for less than 5% of these sites (not shown). The majority of HSF binding sites are instead located in transcribed regions of the genome with introns accounting for the largest proportion of HSF targets. Fifteen percent of all sites are somewhat ambiguous in definition as they occur within transcribed regions but are also in the vicinity of a transcription start site (TSS).

Figure 2-7  Example of an intronic HSF-binding site.
As with Figure 2-4, this image was generated using the UCSC Genome Browser to illustrate the chromosome region represented in bp (according to release 4.2 of the Drosophila genome) as indicated at the top and the legend is the same as that in Figure 2-4.
promoters.

Given this distribution we were interested in determining if HSF is targeting a specific class of genes when binding to promoters so we used the online resource DAVID (Dennis et al. 2003) to assess enrichment in gene function among these genes. For this analysis we considered only those 27% of sites that were within promoter regions as we defined as 1250 bp from a transcription start site and otherwise not within the transcribed region of any gene. Not surprisingly, the most strongly enriched categories were related to the response to stress (Figure 2-8 first column). Also among the most highly enriched categories was glutathione transferase activity and TPR repeat.

To further this analysis, we investigated the enrichment in functional categories among genes that contained at least one HSF binding site within their transcribed region. In this case we considered only those 35% of sites that were within introns and greater than 1250 bp away from the transcription start site of any gene/isoform. This conservative estimate of the number of HSF binding sites found in introns still represents a 2-fold enrichment over the background distribution since only 17.1% of euchromatin is intronic. Our findings here were largely unexpected; there was a strong enrichment for genes involved in biological regulation and more specifically the regulation of transcription and metabolic processes as well as for genes involved in reproduction and development such as gamete generation and anatomical structure development (Figure 2-8 second column).

Because of the difference in functional classification of genes associated with HSF-bound promoters versus HSF-bound introns, we were interested in determining if any other transcription factor(s)/DNA binding protein(s) were associated with these sites. To identify possible candidates, we used Patser to scan HSF-bound promoters and introns for matches (Bonferroni corrected p-value < 5.6x10^-2) to PWMs representing 111 different DNA binding proteins from two databases, Transfac and the Drosophila DNase I Footprint Database. As expected, the PWM representing the 15 bp HSE composed of inverted repeats of nGAAn (Figure 2-5A) was enriched near the peak of both HSF-bound promoters and HSF-bound introns (Figure 2-9). Of the remaining PWMs, the PWM for BEAF was the only one enriched near the peak of HSF-bound promoters to also have a similar chi squared value (X² > 40) and show the
Figure 2-8 Heat map summary of select categories from DAVID functional enrichment analysis.
Genes that were associated with HSF-binding sites in either the promoter or intron regions (column 1 and 2 respectively) and genes that were differentially regulated by HS treatment in either Kc cells or 3rd instar larvae (columns 3 and 4 respectively) were analyzed for functional enrichment. The Functional Annotation Chart tool was used to obtain the p-values describing the probability that a functional term is enriched among genes in the groups examined by random chance, and the Functional Annotation Clustering tool was used to group similar annotation terms. Lower p-values indicating enrichment for the term on the right of the chart are colored in red, while p-values above 0.1 are indicated in maroon. Groups of similar annotation terms are indicated on the left of the chart.
The occurrence of 111 PWMs from TRANSFAC and the *Drosophila* DNase I Footprint Database in HSF-bound promoters and introns was scored with Patser (Hertz and Stormo 1999). The histograms show the number of matches (Bonferroni corrected p-value < 5.6x10^-2) to PWMs representing HSF and BEAF binding sites in 50bp windows centered on the peak of each HSF-bound segment (blue for promoters, red for introns). The PWM for DREF is very similar to the one for BEAF and gives the same result. P-values at the top right of each histogram indicate the probability given by a $X^2$ test that the difference in the number of matches to the PWM found in promoters and introns is due to chance.
same level of significance in a chi square test (Bonferroni corrected p-value < 2.3x10^{-8}) as the PWM representing the canonical HSE (Figure 2-9). Unlike the HSE PWM, however, this enrichment was only seen for those HSF-bound sites found in promoters; the occurrence of the BEAF PWM in HSF-bound segments located within introns was no different than the background (Figure 2-9). Consistent with this finding is a recent report that BEAF (boundary element associated factor) binding sites are enriched in 5’ UTRs and in the first 200 bp upstream of gene’s TSS (Bushey et al. 2009; Gurudatta and Corces 2009). Genes having both BEAF and HSF binding sites do not appear to be strongly enriched in any categories that differ from those enriched among all promoters except for a modest enrichment for genes with cell cycle annotation (p-value =0.0011; data not shown) which is consistent with the function of genes BEAF has been shown to regulate (Emberly et al. 2008).

2.4.4 Transcriptional profile of Kc cell and 3rd instar larvae in response to heat shock

Previous studies have predicted that heat activated HSF might be inducing the transcription of genes in addition to the well-known HS genes. After HS, it is known that RNA polymerase II (RNPII) relocalizes from several hundred discrete loci on polytene chromosomes to a far smaller number of loci with a large amount of RNPII accumulation at the HS puff sites (Jamrich et al. 1977; Greenleaf et al. 1978). RNPII can be seen at about 50 loci after a 20 minute HS and co-localizes with a subset of the approximately 200 observed HSF binding sites. In addition, nascent transcripts can be seen to co-localize with RNPII (J.P. Paraiso, M. Gibson and J.T. Westwood, unpublished results).

To determine if HSF binding had an effect on any of the genes with which it associated following heat stress in addition to the classical heat shock genes, we examined the transcriptional profile of Kc cells under the same conditions in which the binding sites were identified (30 minute HS at 36.5°C). RNA isolated from HS and untreated cells was reverse transcribed, labeled and hybridized to NimbleGen expression microarrays (data available in GEO under GSE19745). We identified 211 genes that showed at least a 2-fold change in expression due to heat shock with a FDR corrected p-value less than 0.01 (Additional File A-4). Not surprisingly, several major HSP genes were strongly induced including Hsp70, Hsp68, Hsp27, Hsp26, Hsp23, and Hsp22. In addition, seven other genes exhibited fold changes
comparable to the small HSPs: CG32850, CG12507, SP555, Gr63a, CG8086, CG7509, and Ir93a. Aside from CG7509, which is repressed in response to both oxidative stress and ER stress in Drosophila (Girardot et al. 2004), none of these genes have been associated with the stress response in Drosophila. In general, most genes we identified changing in response to heat shock in Kc cells were up-regulated, and showed only a modest change in transcript levels (less than 4-fold) (Additional File A-4).

DAVID analysis of genes differentially regulated in cells revealed an enrichment in many of the same categories enriched among genes whose promoters were associated with HSF following HS (Figure 2-8 third column). Also like the DAVID analysis on HSF-bound promoters, several genes with similar function to the major HSPs were identified. Interestingly, the terms transferase, transmembrane, and sensory transduction are enriched among genes regulated by HS in cells but not among genes that associated with HSF binding sites (Figure 2-8 compare column 1 and 2 to 3) indicating that there is a specific set of functionally related genes that are regulated by HS but that are not associated with HSF.

Because many novel heat responsive genes were identified in this genome-wide screen, we wanted to determine how far away the nearest HSF binding site was relative to transcription start sites of these genes. Additional File A-4 lists the distance from the TSS of each gene to the nearest HSF-binding site. Surprisingly, these genes exhibited mean and median distances of greater than 100 kb and 50 kb respectively. In some cases this may be explained by the lack of detection of a bona-fide HSF binding site by our approach since, although the Agilent genome-tiling array covers the entire 117 MB euchromatic genome, probes are lacking in areas with highly repetitive sequence or sequence with high homology to other regions. For example, there are no probes on the array covering the region ~40 kb upstream of Hsp70Bbb (the most highly induced gene in Kc cells) likely due to the highly repetitive nature of this sequence (Figure 2-2B,C). Instead the closest site to Hsp70Bbb we identified was greater than 100 kb upstream of its TSS. However, since this case is expected to be the exception rather than the rule, it is unlikely to be the cause of a lack of HSF binding to the promoters of the majority of the genes identified. To rule out the possibility that a secondary transcription factor transcribed in response to heat shock may be controlling the expression of some of these genes, we repeated the gene expression analysis in the presence of the translation inhibitor cycloheximide and found no significant affect on the transcription of any of these genes (data not shown).
Given that cell lines do not always provide an accurate picture of the biological response of whole organisms, and that several sites of gene transcription can be observed in addition to the HS puff sites on polytene chromosomes, we next examined the transcriptional response to heat shock in wandering 3rd instar larvae. As with the cells, the larvae were subjected to a 30 min heat shock at 36.5°C to match the conditions used for HSF binding site identification (data available in GEO under GSE19745). Overall, 237 genes exhibited a 2-fold or greater change in expression and a FDR corrected p-value of less than 0.01 (Additional File A-4). As in cells, the majority of genes show a modest change (less than 4-fold), are mostly up-regulated, and are mostly enriched in the same functional categories as promoter-bound genes (Figure 2-8, fourth column). Furthermore, the functional terms transferase and transmembrane are enriched among HS regulated genes in larvae suggesting that several HS-regulated genes not associated with HSF binding sites are still related in function (Figure 2-8 compare columns 3 and 4). Table 2-2A lists all genes exhibiting a 8-fold or greater induction in either cells or larvae. Comparison of all genes that were heat-responsive in cells and in larvae revealed few genes that were universally regulated by HS. Ninety-two percent of all stress-responsive genes identified were only affected in one system. The remaining 8% of genes that were affected in both systems include all of the major HSP genes (except Hsp83 and Hsp67Ba, which were only induced in larvae), DnaJ-1, and 22 other genes, of which 10 have been previously associated with at least one other stress in Drosophila (Table 2-2B). Functional enrichment analysis of HS responsive genes in larvae identified several non-classical HSPs predicted to have similar functions as the classical HSPs. Among them, we have identified HSF binding sites in the promoters of at least eight: CG11035, CG7130, CG7945, Droj2, PEK, Sir2, Tom34, and tra.

Because the transcriptional profile of larvae greatly differed from cells, we investigated whether genes responsive to heat shock in larvae were any more likely to be associated with HSF. The result was similar to cells; the mean distance from the nearest HSF binding site to the TSS was greater than 88 kb and the median distance was nearly 35 kb.

To further investigate possible HSF association with genes regulated during heat shock we compared the lists of genes exhibiting a 2-fold or greater change in expression in response to heat shock in either cells or larvae to the 471 genes that either contained an HSF binding site within their coding region or were located 1250 bp downstream of the peak of an HSF bound region (Figure 2-10B). Only nine genes were in common to all three lists: Hsp22, Hsp26,
Table 2-2 Genes regulated by heat shock.
(A) Genes induced by at least 8-fold in either Kc cells or 3rd instar larvae. (B) Genes exhibiting a 2- to 8-fold change in expression in both Kc cells and 3rd instar larvae.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Kc167 cells</th>
<th>3rd instar larvae</th>
<th>position of closest HSF binding site relative to TSS (bp)</th>
<th>cytology</th>
<th>staining intensity in polytene data (Westwood et al., 1991)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS vs RT fold change</td>
<td>FRD corrected p-value</td>
<td>HS vs RT fold change</td>
<td>FDR corrected p-value</td>
<td></td>
</tr>
<tr>
<td>Hsp70Bb</td>
<td>41.9</td>
<td>2.00E-10</td>
<td>103.3</td>
<td>1.80E-07</td>
<td>-124162</td>
</tr>
<tr>
<td>Hsp70Ba</td>
<td>33.3</td>
<td>1.60E-09</td>
<td>162.4</td>
<td>6.30E-10</td>
<td>-89431</td>
</tr>
<tr>
<td>Hsp70Bc</td>
<td>32.1</td>
<td>1.60E-09</td>
<td>140.2</td>
<td>8.30E-10</td>
<td>-130728</td>
</tr>
<tr>
<td>Hsp70Ab</td>
<td>31.7</td>
<td>3.30E-09</td>
<td>169.9</td>
<td>8.80E-10</td>
<td>-732</td>
</tr>
<tr>
<td>Hsp68</td>
<td>24.7</td>
<td>2.20E-10</td>
<td>106.8</td>
<td>4.20E-08</td>
<td>74</td>
</tr>
<tr>
<td>CG32850</td>
<td>16.6</td>
<td>6.90E-08</td>
<td>1.2</td>
<td>7.00E-01</td>
<td>716852</td>
</tr>
<tr>
<td>Hsp22</td>
<td>14.1</td>
<td>3.70E-09</td>
<td>227</td>
<td>3.30E-08</td>
<td>838</td>
</tr>
<tr>
<td>Hsp26</td>
<td>12.2</td>
<td>1.40E-09</td>
<td>28.5</td>
<td>1.10E-04</td>
<td>305</td>
</tr>
<tr>
<td>CG12507</td>
<td>10.9</td>
<td>1.20E-04</td>
<td>4.5</td>
<td>2.40E-03</td>
<td>-163722</td>
</tr>
<tr>
<td>SP555</td>
<td>10.8</td>
<td>3.70E-04</td>
<td>-1</td>
<td>9.70E-01</td>
<td>-28158</td>
</tr>
<tr>
<td>Gr63a</td>
<td>10.7</td>
<td>4.00E-08</td>
<td>15.4</td>
<td>2.00E-05</td>
<td>5969</td>
</tr>
<tr>
<td>CG8086</td>
<td>10.6</td>
<td>2.40E-06</td>
<td>1.5</td>
<td>4.60E-01</td>
<td>-87278</td>
</tr>
<tr>
<td>Hsp23</td>
<td>10.2</td>
<td>8.60E-08</td>
<td>8.3</td>
<td>5.90E-03</td>
<td>1829</td>
</tr>
<tr>
<td>CG7509</td>
<td>9.9</td>
<td>4.10E-08</td>
<td>6.7</td>
<td>1.20E-04</td>
<td>13407</td>
</tr>
<tr>
<td>Hsp27</td>
<td>8.8</td>
<td>7.60E-07</td>
<td>7.5</td>
<td>1.80E-04</td>
<td>-352</td>
</tr>
<tr>
<td>Ir93a</td>
<td>8.3</td>
<td>1.20E-04</td>
<td>1.5</td>
<td>2.00E-01</td>
<td>#N/A</td>
</tr>
<tr>
<td>DnaJ-1</td>
<td>5.3</td>
<td>3.30E-08</td>
<td>42</td>
<td>2.10E-05</td>
<td>-219</td>
</tr>
<tr>
<td>Cpi8</td>
<td>2.1</td>
<td>1.30E-03</td>
<td>55</td>
<td>3.10E-08</td>
<td>-5166</td>
</tr>
<tr>
<td>stv</td>
<td>3.9</td>
<td>3.70E-08</td>
<td>29.4</td>
<td>7.10E-10</td>
<td>-421</td>
</tr>
<tr>
<td>se</td>
<td>3.5</td>
<td>2.90E-03</td>
<td>18.7</td>
<td>4.90E-07</td>
<td>-1909</td>
</tr>
<tr>
<td>CG5290</td>
<td>1.0</td>
<td>8.60E-01</td>
<td>15.3</td>
<td>6.80E-08</td>
<td>-81</td>
</tr>
<tr>
<td>Hsp83</td>
<td>1.3</td>
<td>2.40E-03</td>
<td>15.1</td>
<td>2.70E-06</td>
<td>5</td>
</tr>
<tr>
<td>CG14961</td>
<td>3.2</td>
<td>7.90E-04</td>
<td>14.5</td>
<td>3.10E-05</td>
<td>-79865</td>
</tr>
<tr>
<td>Hsp67Bb</td>
<td>2.1</td>
<td>7.50E-04</td>
<td>14.1</td>
<td>1.60E-04</td>
<td>838</td>
</tr>
<tr>
<td>CG17352</td>
<td>2.6</td>
<td>8.60E-04</td>
<td>13.5</td>
<td>9.40E-06</td>
<td>4134</td>
</tr>
<tr>
<td>CG7130</td>
<td>1.0</td>
<td>9.10E-01</td>
<td>9.3</td>
<td>4.70E-04</td>
<td>-1021</td>
</tr>
<tr>
<td>CG3280</td>
<td>1.0</td>
<td>9.70E-01</td>
<td>8.9</td>
<td>2.90E-06</td>
<td>90028</td>
</tr>
<tr>
<td>Fdxh</td>
<td>1.2</td>
<td>1.10E-01</td>
<td>8.4</td>
<td>5.90E-06</td>
<td>2731</td>
</tr>
<tr>
<td>CG15199</td>
<td>-1.5</td>
<td>2.20E-01</td>
<td>8.3</td>
<td>1.00E-04</td>
<td>-32986</td>
</tr>
<tr>
<td>CG6785</td>
<td>1.5</td>
<td>1.80E-02</td>
<td>8.1</td>
<td>1.30E-07</td>
<td>1290</td>
</tr>
</tbody>
</table>

(B) Genes exhibiting a 2- to 8-fold change in expression in both Kc cells and 3rd instar larvae.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Kc167 cells</th>
<th>3rd instar larvae</th>
<th>position of closest HSF binding site relative to TSS (bp)</th>
<th>cytology</th>
<th>staining intensity in polytene data (Westwood et al., 1991)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS vs RT fold change</td>
<td>FRD corrected p-value</td>
<td>HS vs RT fold change</td>
<td>FDR corrected p-value</td>
<td></td>
</tr>
<tr>
<td>CG14937</td>
<td>2.5</td>
<td>3.60E-06</td>
<td>7</td>
<td>5.10E-06</td>
<td>218089</td>
</tr>
<tr>
<td>TwdlF</td>
<td>3</td>
<td>5.30E-04</td>
<td>5.7</td>
<td>1.10E-03</td>
<td>-31204</td>
</tr>
<tr>
<td>spz4</td>
<td>4.9</td>
<td>6.20E-05</td>
<td>5.3</td>
<td>1.60E-03</td>
<td>506289</td>
</tr>
<tr>
<td>CG13035</td>
<td>2.4</td>
<td>5.30E-04</td>
<td>5.1</td>
<td>1.50E-06</td>
<td>96856</td>
</tr>
<tr>
<td>CG11619</td>
<td>2.7</td>
<td>3.10E-04</td>
<td>4.3</td>
<td>3.10E-03</td>
<td>104557</td>
</tr>
<tr>
<td>Acp98AB</td>
<td>3</td>
<td>5.30E-04</td>
<td>4.2</td>
<td>3.40E-04</td>
<td>253666</td>
</tr>
<tr>
<td>CG32636</td>
<td>2.5</td>
<td>6.20E-04</td>
<td>3.9</td>
<td>4.60E-04</td>
<td>NA</td>
</tr>
<tr>
<td>CG8026</td>
<td>2</td>
<td>2.50E-04</td>
<td>3.8</td>
<td>4.90E-04</td>
<td>238705</td>
</tr>
<tr>
<td>Hdc</td>
<td>2.7</td>
<td>1.80E-03</td>
<td>3.6</td>
<td>7.20E-04</td>
<td>-119479</td>
</tr>
<tr>
<td>CG32715</td>
<td>2.7</td>
<td>4.60E-07</td>
<td>3.1</td>
<td>6.80E-03</td>
<td>NA</td>
</tr>
<tr>
<td>CG12990</td>
<td>7.4</td>
<td>2.00E-07</td>
<td>2.9</td>
<td>2.10E-03</td>
<td>-249793</td>
</tr>
<tr>
<td>CG5204</td>
<td>2.6</td>
<td>1.30E-03</td>
<td>2.6</td>
<td>6.80E-03</td>
<td>4937</td>
</tr>
<tr>
<td>CG5376</td>
<td>4</td>
<td>1.30E-03</td>
<td>2.2</td>
<td>9.60E-03</td>
<td>2152</td>
</tr>
<tr>
<td>Rbp1</td>
<td>-2.2</td>
<td>8.40E-03</td>
<td>-2</td>
<td>3.50E-03</td>
<td>-388811</td>
</tr>
</tbody>
</table>
Figure 2-10 Many genes associated with HSF binding are not induced by HS in Kc cells or 3rd instar larvae. (A) Venn diagram of the overlap between genes associated with an HSF-binding site and genes regulated by heat shock. (B) Hierarchical cluster of the HS transcriptional profile of all genes associated with an HSF-binding site identified in our ChIP-chip analysis and/or transcriptionally regulated by HS in cells and/or larvae (left). Sub-clusters highlight genes with similar expression profiles. For genes associated with an HSF-binding site, the log₂ fold enrichment of the associated binding site is indicated in the aligned heat map (right). Where there is no associated binding site, the fold enrichment is displayed as grey to indicate no data is available.
Hsp27, Hsp67Bb, Hsp68, Hsp70Ab, DnaJ-1, stv and CG32636. Not surprisingly, most of these genes are well-known heat-inducible genes. Starvin (stv), although not a classical HSP, is induced in response to several stresses including oxidative and ER stress, aging, starvation and HS (Zinke et al. 2002; Girardot et al. 2004; Landis et al. 2004; Harbison et al. 2005; Sorensen et al. 2005) and encodes a BAG-domain protein and is thought to be a HSP70-family co-chaperone (Coulson et al. 2005). In addition to these nine genes, a total of 40 differentially regulated genes are located within 1250bp of the peak of an HSF bound region. Thirty-nine of these are up regulated in response to HS in either cells or larvae but not both and one gene is down regulated in cells (Figure 2-10 and Table 2-3).

Because the FDR corrected p-value cutoff we applied in the identification of transcriptionally regulated genes of 0.01 is relatively stringent, it is possible that other genes associated with HSF-binding sites were transcriptionally regulated during HS but were not identified in our expression analysis. To investigate this possibility we generated a cluster of the expression profiles of all genes that were associated with an HSF binding site and/or were identified in our expression analysis of HS cells and larvae and aligned a heat map depicting the relative enrichment of the nearest HSF bound segment identified in our ChIP-chip analysis (where applicable; if no HSF-binding was located either within the coding region of the gene or <1250bp upstream of the gene’s TSS then the corresponding value in the heat map is grey indicating no data is available) (Figure 2-10B). This analysis revealed several things in support of our initial observations: First the majority of HS responsive genes exhibit an increase in transcript levels and are responsive in only one system even at relatively weak fold changes (Figure 2-10B clusters B and C). Second, the majority of HS-responsive genes, especially in cells, are not associated with HSF binding (Figure 2-10B clusters B, C, D and F). Finally, although there are a few HSF-associated genes that appear to undergo a small induction (less than 2-fold) in response to HS in larvae, the majority of HSF-associated genes are not transcriptionally responsive to HS in either cells or larvae (Figure 2-10 cluster E).

Of the 11 HSF-bound and HS-induced genes identified in Kc cells (Table 2-3) all are bound by HSF within 1250 bp of their annotated TSS (Additional File A-4). Given the enrichment of BEAF motifs in HSF-bound promoter segments (Figure 2-9) we were interested in determining if there is any correlation between the presence of a BEAF motif and the likelihood of that gene
Table 2-3 Genes associated with HSF that are regulated by HS in either Kc cells, 3rd instar larvae or both.

(see Figure 10).

<table>
<thead>
<tr>
<th>gene name</th>
<th>HS-regulated gene</th>
<th>HSF binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bound by HSF, induced in Kc cells and in 3rd instar larvae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DnaJ-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp67Bb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp70Ab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stv</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG32636*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG10924</td>
<td>1.5 1.70E-01 2.5 1.12E-04 2156 17 1.44E-10</td>
<td></td>
</tr>
<tr>
<td>CG10973</td>
<td>1.2 2.51E-02 2.8 8.58E-03 97 20 2.30E-07</td>
<td></td>
</tr>
<tr>
<td>CG11033</td>
<td>1.1 2.37E-01 2.1 9.46E-04 -740 28 1.46E-04</td>
<td></td>
</tr>
<tr>
<td>CG11035</td>
<td>-1.6 1.69E-02 5.3 9.67E-08 -620 31 4.50E-11</td>
<td></td>
</tr>
<tr>
<td>CG13472</td>
<td>-1.1 1.55E-01 3 3.24E-04 115 12 9.33E-06</td>
<td></td>
</tr>
<tr>
<td>CG14146</td>
<td>-1.1 1.32E-01 2.7 8.87E-04 410 7 1.35E-08</td>
<td></td>
</tr>
<tr>
<td>CG1553</td>
<td>-1.1 2.19E-01 4.4 8.86E-08 144 19 9.64E-11</td>
<td></td>
</tr>
<tr>
<td>CG1865</td>
<td>-1.1 5.19E-01 5.6 7.49E-04 909 16 1.17E-10</td>
<td></td>
</tr>
<tr>
<td>CG32103</td>
<td>-1 9.83E-01 2.3 3.60E-03 -472 5 7.73E-05</td>
<td></td>
</tr>
<tr>
<td>CG3226</td>
<td>1.1 2.38E-01 3 6.09E-05 -125 33 1.25E-07</td>
<td></td>
</tr>
<tr>
<td>CG5010</td>
<td>1.1 6.17E-01 3.4 1.27E-03 -376 26 1.70E-07</td>
<td></td>
</tr>
<tr>
<td>CG5290</td>
<td>-1 8.58E-01 15.3 6.80E-08 81 34 4.31E-07</td>
<td></td>
</tr>
<tr>
<td>CG5953</td>
<td>-1 8.89E-01 2.9 8.41E-04 2125 6 1.96E-06</td>
<td></td>
</tr>
<tr>
<td>CG6191</td>
<td>-1.3 1.47E-03 2.2 1.93E-03 1430 16 8.42E-11</td>
<td></td>
</tr>
<tr>
<td>CG6511</td>
<td>-1.1 2.96E-02 7.8 1.34E-06 36 28 2.24E-07</td>
<td></td>
</tr>
<tr>
<td>CG7945</td>
<td>1.1 7.28E-01 5.2 1.13E-05 345 17 9.31E-06</td>
<td></td>
</tr>
<tr>
<td>CG9153</td>
<td>-1.1 3.34E-01 2.6 6.02E-04 260 26 1.93E-07</td>
<td></td>
</tr>
<tr>
<td>Droj2</td>
<td>-1.1 1.50E-01 2.7 2.57E-04 823 36 2.57E-11</td>
<td></td>
</tr>
<tr>
<td>GstD10</td>
<td>1.6 5.04E-02 2.3 2.40E-05 -447 7 6.25E-07</td>
<td></td>
</tr>
<tr>
<td>Hop</td>
<td>-1.1 2.39E-01 2.9 1.22E-04 108 73 3.23E-09</td>
<td></td>
</tr>
<tr>
<td>Hsc70-3</td>
<td>1.1 2.39E-01 2.9 4.24E-04 1721 11 5.00E-05</td>
<td></td>
</tr>
<tr>
<td>Hsc70-4</td>
<td>-1.1 1.03E-01 2.1 3.16E-06 -123 65 3.64E-12</td>
<td></td>
</tr>
<tr>
<td>Hsc70-5</td>
<td>-1.1 8.90E-02 3.4 2.17E-05 236 18 1.05E-10</td>
<td></td>
</tr>
<tr>
<td>Hsc70Cb</td>
<td>-1.1 2.35E-01 2.7 1.47E-04 195 30 1.31E-07</td>
<td></td>
</tr>
<tr>
<td>HspB3</td>
<td>1.3 2.40E-03 15.1 2.69E-06 5 71 7.58E-08</td>
<td></td>
</tr>
<tr>
<td>PEK</td>
<td>-1.2 1.14E-01 3.7 3.83E-04 1048 6 5.22E-05</td>
<td></td>
</tr>
<tr>
<td>Pdk</td>
<td>-1.1 5.87E-01 2.6 6.91E-05 4359 10 4.05E-10</td>
<td></td>
</tr>
<tr>
<td>Sir2</td>
<td>-1 9.09E-01 2.8 9.63E-03 -87 76 3.73E-11</td>
<td></td>
</tr>
<tr>
<td>Taft</td>
<td>1.1 1.44E-01 7.3 6.86E-05 353 18 7.06E-11</td>
<td></td>
</tr>
<tr>
<td>Tom34</td>
<td>-1.1 3.90E-01 6.5 9.74E-05 263 31 4.50E-11</td>
<td></td>
</tr>
<tr>
<td>cn</td>
<td>-1.4 3.13E-01 2.2 4.62E-04 2174 17 1.25E-09</td>
<td></td>
</tr>
<tr>
<td>l(1)G0469</td>
<td>1.8 3.85E-03 3.8 6.01E-03 2448 21 5.05E-07</td>
<td></td>
</tr>
<tr>
<td>mbf1</td>
<td>1.2 5.24E-02 2.3 1.96E-04 -67 28 1.20E-06</td>
<td></td>
</tr>
<tr>
<td>pald</td>
<td>1.1 2.57E-01 3.3 4.90E-04 497 5 6.21E-05</td>
<td></td>
</tr>
<tr>
<td>sra</td>
<td>1 9.32E-01 3.3 1.99E-03 25 18 1.86E-10</td>
<td></td>
</tr>
<tr>
<td>tra</td>
<td>-1.1 2.42E-01 3.2 4.37E-03 248 5 1.10E-04</td>
<td></td>
</tr>
<tr>
<td>ttk</td>
<td>1.7 7.65E-03 2.3 3.18E-03 2452 13 7.08E-05</td>
<td></td>
</tr>
<tr>
<td>Bound by HSF, induced in Kc cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG10077</td>
<td>2.7 4.10E-03 7.8 1.38E-02 -104 22 4.39E-07</td>
<td></td>
</tr>
<tr>
<td>CG6770</td>
<td>3 1.25E-03 1.7 2.03E-01 -390 35 2.43E-11</td>
<td></td>
</tr>
<tr>
<td>Bound by HSF, repressed in Kc cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gstd2</td>
<td>-2.2 4.66E-03 -1.3 4.43E-01 -1110 6 1.23E-08</td>
<td></td>
</tr>
</tbody>
</table>
to be expressed following HS. BEAF-binding sites are enriched in 5’ UTRs and in the first 200 bp upstream of gene TSS (Bushey et al. 2009; Gurudatta and Corces 2009) so we wanted to take the analysis one step further and examine the relationship between the presence of a BEAF motif and the induction of the associated gene during HS. Of the 11 HSF-bound promoters associated with HS-induced genes in Kc cells, only 3 or 27% were found to contain a BEAF motif (p-value < 5x10^-4) (Table 2-4). Conversely, of the 104 HSF-binding sites located exclusively in promoters (ie. within 1250 bp) of genes that did not show a transcriptional change following HS in Kc cells, 69 or 66% were found to contain at least one BEAF motif. Given this difference in the distribution of BEAF motifs between induced and non-induced HSF-bound promoters, we were interested in determining if BEAF is preferentially bound to promoters of non-induced genes under non-HS condition prior to exposure to HS. To examine this possibility we compared HSF-bound chromatin segments to chromatin segments bound by BEAF in Kc167 cells under non-HS conditions (Bushey et al. 2009) (data available in GEO under GSE15661) and found that the proportion of HS-induced and non-induced HSF-bound promoters that were also bound by BEAF to be similar to the proportion containing at least one BEAF motif (Table 2-4) and that there is an enrichment for BEAF binding sites in non HS-induced HSF-bound gene promoters (Chi squared test; p-value =0.0198). Next we sought to determine if any other Drosophila insulators also co-localize with HSF binding sites or if this observation is specific for BEAF. We compared our HSF bound sites to the binding sites for two other insulators (dCTCF and Su(HW)) and an insulator associated protein, CP190, for which ChIP-chip data is available in Kc cells (Bushey et al. 2009). We did not see a significant enrichment for dCTCF or Su(HW) binding sites at HSF sites (data not shown) but we did see a large overlap with CP190 sites which is expected since CP190 does not bind DNA directly but does bind to insulators including BEAF (Bushey et al. 2009). This suggests that the enrichment of BEAF sites at HSF bound promoters is specific to BEAF and not a general feature of all insulators.

Because there were a large number of genes whose transcripts changed during HS but did not appear to be near an HSF binding site, we investigated whether the transcriptional changes were dependent on HSF. The transcriptional response to HS was measured in Hsf4 mutant 3rd instar larvae using the same approach described above (data available in GEO under GSE22332). These larvae have a temperature sensitive mutation in the HSF DNA binding domain which prevent HSF from binding to HSEs and inducing HSP gene transcription at non-permissive (i.e.
Table 2-4 Occurrence of BEAF motifs and binding sites in HSF-bound promoters.

<table>
<thead>
<tr>
<th>Description</th>
<th>Promoters of genes induced during hs in Kc cells</th>
<th>Promoters of genes not induced during hs in Kc cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of promoters with at least one BEAF motif (p-value &lt; 5x10^-4)</td>
<td>3</td>
<td>69</td>
</tr>
<tr>
<td>Number of promoters bound by BEAF under non-hs conditions (Bushey et al., 2009)</td>
<td>3</td>
<td>66</td>
</tr>
<tr>
<td>Number of promoters bound by BEAF under non-hs conditions with at least one BEAF motif</td>
<td>2</td>
<td>48</td>
</tr>
<tr>
<td>Total number of promoters</td>
<td>11</td>
<td>104</td>
</tr>
</tbody>
</table>
heat shock) temperatures (Jedlicka et al. 1997). Only 8 genes were up-regulated in the $Hsf^d$ larvae and the degree of induction was generally far less than what was seen in the wild type larvae (Additional File A-4). Thus, it would appear that the vast majority of transcript levels that change during heat shock are dependent on having functional HSF.

2.5 Discussion

HSF in metazoans is activated upon stress to trimerize and bind HSEs that are found in the promoters of HSP genes. HSF binding leads to the release of stalled RNA polymerase II as well as the recruitment of new polymerase complexes (Rougvie and Lis 1988). It has long been known that HSF binds to many other parts of the genome in addition to the well-known HS gene loci (Westwood et al. 1991).

In this study we attempted to determine all of the HSF binding sites in *Drosophila melanogaster* using ChIP-chip methodology on *Drosophila* genomic tiling arrays. In total we identified 434 HSF bound chromatin segments in heat shocked Kc167 cells with the transcription start site (TSS) of 270 genes mapping to within 1250 bp of an HSF binding site. A comparison of our binding data to an earlier study that utilized heat shocked *Drosophila* embryos and cDNA arrays for the ChIP-chip shows that approximately 33% of the HSF-bound genes in their study (i.e. 62 out of 188) correlate with our binding sites (for a comparison of the overlapping sites see Additional File A-3) (Birch-Machin et al. 2005). The binding sites we identified correlate quite well with those identified by another group that used ChIP-seq to uncover HSF binding sites in S2 cells (Guertin and Lis 2010). Two hundred and sixty-three of the 442 high confidence HSF binding sites found in that study coincided with HSF binding segments we found (Additional File A-3). If we look at our top 100 HSF bound segments, 90 were also identified as HSF binding sites by this group. Differences between the studies might be due to several factors-differences in the cell types, antibodies, array platform, and ChIP or sequence identification methods that were used as well as differences in the analyses that were performed. Others have noted variation in transcription factor binding sites (i.e. for NFκB) in ChIP-seq experiments for different biological individuals even though the same cell type and identical experimental procedures were utilized (Kasowski et al. 2010).
We attempted to correlate the HSF binding events with changes in gene transcription using standard expression microarray analysis of heat shocked Kc cells and Drosophila 3rd instar larvae. These experiments revealed a number of surprising results. First, the transcript profiles of heat shocked Kc cells and 3rd instar larvae were quite different (Figure 2-10). For example, even though each system resulted in more than 200 differentially expressed transcripts, only 33 or about 8% of these transcripts were in common to both systems and of these genes, only 9 had HSF bound segments within 1250 bp of a TSS. Second, only 49 (or about 11%) of the HSF binding sites were found to be within 1250 bp of a TSS for a differentially expressed gene in either system. This suggests that the majority of differentially expressed genes are either being regulated by HSF from a distance more than 1250 bp away or that the differential levels in this class of transcripts is being regulated by a different mechanism. We ruled out the possibility of these transcripts being regulated by an HSF-dependent newly synthesized transcription factor by repeating the experiment in the presence of cycloheximide which did not significantly alter the list of differentially expressed genes induced by heat. We also determined that the vast majority of transcript changes that occur during HS are dependent on having functional HSF since larvae that have a mutated HSF gene (Hsf4 larvae) show very few changes in transcript levels. For these few genes, differential levels of these transcripts might be regulated post-transcriptionally, a phenomenon reported for certain HSP genes in Hsf4 flies (Neal et al. 2006). How HSF might be regulating the other transcripts is still not clear and we cannot rule out that HSF may be interacting with other transcription/chromatin factors without binding to a nearby HSE. HSF could also be working at distances greater than 1250 bp since in three dimensional space, the binding of HSF may in fact be much closer to a TSS. We also cannot rule out that the ChIP-chip approach we used did not uncover all of the HSF binding sites in the genome.

We were also interested in determining if any other transcription factor(s)/DNA binding protein(s) were associated with these HSF sites. We did find that a number of HSF bound promoters (and not introns) also contained binding sites for the BEAF transcription factor (known to be important for insulating enhancers) and this finding is consistent with a recent report that BEAF-binding sites tend to be associated with the 5’ UTRs and regions immediately upstream of the transcription start site of genes (Bushey et al. 2009; Gurudatta and Corces 2009). Interestingly, we found an enrichment for BEAF binding under non-HS conditions (Bushey et al. 2009) to promoters of genes that are bound by HSF but that are not induced
during HS. Since a proportion of the BEAF-binding sites identified in Kc cells is cell type specific (Bushey et al. 2009; Gurudatta and Corces 2009), it is possible that BEAF may have a role preventing the induction of genes near select HSF-binding sites during HS in a cell specific manner.

A recent paper by Guertin and Lis (Guertin and Lis 2010) investigated the distribution of chromatin modifications and certain chromatin proteins at HSEs prior to the binding of HSF in *Drosophila*. Overall, they observed a correlation of histone acetylation, H3K4 trimethylation, RNA polymerase II and coactivators such as GAGA factor with HSEs that ultimately are bound by HSF after heat shock compared to HSEs that are not bound by HSF. These chromatin modifications and proteins are hallmarks of transcriptionally active chromatin and the authors argue that the modifications are requirements for HSF to bind to HSEs prior to transcriptional induction as opposed to a consequence of transcription (Guertin and Lis 2010). These authors also noted a large number of HSF binding sites that were also bound by BEAF prior to HS with a higher occurrence of the overlap taking place at promoters than within genes. Moreover, for the few HSF associated genes whose transcription were examined after HS, BEAF binding was more enriched at non-induced compared to induced genes (Guertin and Lis 2010).

What is the function of HSF binding to so many different places in the genome if it is not to regulate the heat shock genes during stress? There is the possibility that there are some genes that are being transcriptionally induced by HSF during HS in other developmental stages and/or tissue types. The different transcriptional response to heat by Kc cells and 3rd instar larvae lends support to this hypothesis. That is, the transcriptional response to an active transcription factor is likely dependent on the cellular/nuclear environment and/or chromatin state that exists in a given cell type. This could include the possibility that HSF might be acting as a specific transcriptional repressor of certain developmentally regulated genes whose puffs on polytene chromosome regress during heat shock (Westwood et al. 1991). A similar difference between binding events and transcriptional responses has also been seen for the Ecdysone receptor/Ultraspiracle nuclear hormone complex binding sites and the transcription profiles seen in Kc167 cells and the during *Drosophila* metamorphosis (Gauhar et al. 2009).

Another possibility is that HSEs appear in the genome with a certain frequency and have no biological consequence. It has been suggested that there are a large number of cis-regulatory
modules (CRMs) in the *Drosophila* genome that fall into this category (MacArthur et al. 2009). Natural selection would preserve those CRMs that are critical to transcriptional regulation but an organism could tolerate CRMs that had weak affinity for a given transcription factor that did not interfere with transcriptional regulation (Li et al. 2008; MacArthur et al. 2009). The existence of large numbers of transcription factor binding sites that have no apparent biological activity would appear to be a property of all eukaryotic organisms (Wunderlich and Mirny 2009).

Yet another possibility is that HSF has functions during non-HS conditions and that the ChIP-chip analysis is revealing many of those gene targets. Clearly HSF deficient *Drosophila* show developmental arrest (i.e. at 1st and 2nd instar) as well as defects in oogenesis under non-HS conditions (Jedlicka et al. 1997). In species such as *Drosophila* that have a single HSF, HSF may be performing numerous roles under both HS and non-HS conditions. It is also possible that during *Drosophila* development that other forms of stress are occurring that induce HSF transiently. As animals evolved, gene duplication and divergence resulted in multiple HSFs that distributed some of the important functions to specific and/or multiple HSFs. As previously mentioned, mice lacking HSF1 display growth retardation and female infertility due to defective oocytes (Xiao et al. 1999; Christians et al. 2000). Hsf1<sup>−/−</sup> oocytes exhibit a delay and blockage of meiotic maturation and this defect at least in part can be related to a decrease in Hsp90α transcript levels and HSP90 protein activity (Metchat et al. 2009). Hsf2<sup>−/−</sup> mice show embryonic brain defects with the defect in cerebral cortex formation being attributed to the reduced expression of an HSF2 regulated gene, *p35* (Chang et al. 2006). Both HSF1 and HSF2 have been shown to play roles in sperm development in mice with Hsf2<sup>−/−</sup> mice showing a more severe defect resulting in a reduced testis size and the disruption of spermatogenesis characterized by degenerating cells, the absence of differentiating spermatids and spermatocytes, vacuolization of the tubules and reduced sperm count (Kallio et al. 2002; Wang et al. 2003). A mouse HSF2 ChIP-chip study performed with testis found numerous promoters that bound HSF2 including almost 1/3<sup>rd</sup> of the 105 genes known to exist on the Y chromosome (Akerfelt et al. 2008). HSF2 was found to bind and regulate multi-copy genes in the male-specific region of the Y chromosome (MSYq) and HSF2 deficient mice had similar increases in sperm head defects as those with MSYq deletion mutations (Akerfelt et al. 2008).
It is interesting that the DAVID analysis of the HSF binding sites in *Drosophila* showed enrichment for a number of developmental processes as well as gamete formation including oogenesis (Figure 2-8). Also revealing was that this enrichment was only seen for the HSF binding sites found in the introns of genes whereas the analysis of HSF binding sites in promoters and introns and in the genes that showed transcriptional changes showed enrichment in GO categories such as response to stress and transferase activity.

### 2.6 Conclusions

Original observations of HSF binding to over 200 loci on *Drosophila* polytene chromosomes led to the speculation that HSF may be involved in the regulation of more genes than originally identified. With the advancements in techniques used to identify transcription factor binding sites and transcriptionally active genes since this initial identification, we set out to better understand the function of HSF binding to these sites. Indeed, we have come one step closer to understanding the functions of HSF outside of its well-established role of a transcriptional activator of the major HSP genes. We identify many genes that have not previously been associated with the heat shock response and find that the core set of major HSPs is well conserved as previously noted. We also find that HSF appears to have a previously undocumented preference for binding to sites located with the introns of developmentally regulated genes. In light of the fact that HSF does not appear to be transcriptionally activating the large majority of genes with which it associates, this novel finding provides a solid foundation for future studies of HSF function, particularly when examined in context of the poorly understood requirement for HSF in early development in *Drosophila* and female fertility in mammals.
Chapter 3
Genome-wide examination of the transcriptional response to ecdysteroids 20-hydroxyecdysone and ponasterone A in *Drosophila melanogaster*

**Please note:** The contents of this chapter are published in BMC Genomics (see Gonsalves et al. 2011) with the exception of section 3.4.6, which was modified from the original version published in PLoS ONE (see Gonsalves et al. 2011). Permission to reproduce this material here is provided under the Creative Commons Attribution License. The following people were all contributing authors to the BMC Genomics paper: Sarah E. Gonsalves, Scott J. Neal, Amy S. Kehoe, and J. Timothy Westwood and their contributions were as follows: SEG participated in the design of the study, carried out the hormone treatments, RNA isolations, microarray experiments, data analysis including analysis of functional classes, and qRTPCR experiments, and drafted the manuscript. SJN and ASK participated in the design of the study and participated in pilot experiments. JTW conceived of the study, participated in its design and helped to draft the manuscript. Please see chapter 2 title page for additional author and author contributions for section 3.4.6.

3.1 Abstract

The 20-hydroxyecdysone (20E) hierarchy of gene activation serves as an attractive model system for studying the mode of steroid hormone regulated gene expression and development. Many structural analogs of 20E exist in nature and among them the plant-derived ponasterone A (PoA) is the most potent. PoA has a higher affinity for the 20E nuclear receptor, composed of the ecyson receptor (EcR) and Ultraspiracle proteins, than 20E and a comparison of the genes regulated by these hormones has not been performed. Furthermore, in *Drosophila* different cell types elicit different morphological responses to 20E yet the cell type specificity of the 20E transcriptional response has not been examined on a genome-wide scale. We aim to characterize the transcriptional response to 20E and PoA in *Drosophila* Kc cells and to 20E in salivary glands and provide a robust comparison of genes involved in each response. Our genome-wide microarray analysis of Kc167 cells treated with 20E or PoA revealed that far more genes are regulated by PoA than by 20E (256 vs 148 respectively) and that there is very little overlap between the transcriptional responses to each hormone. Interestingly, genes induced by 20E relative to PoA are enriched in functions related to development. We also find that many genes regulated by 20E in Kc167 cells are not regulated by 20E in salivary glands of wandering 3rd instar larvae and we show that 20E-induced levels of EcR isoforms EcR-RA, ER-RC, and
EcR-RD/E differ between Kc cells and salivary glands suggesting a possible cause for the observed differences in 20E-regulated gene transcription between the two cell types. We report significant differences in the transcriptional responses of 20E and PoA, two steroid hormones that differ by only a single hydroxyl group. We also provide evidence that suggests that PoA induced death of non-adapted insects may be related to PoA regulating a different set of genes when compared to 20E. In addition, we reveal large differences between Kc cells and salivary glands with regard to their genome-wide transcriptional response to 20E and show that the level of induction of certain EcR isoforms differ between Kc cells and salivary glands. We hypothesize that the differences in the transcriptional response may in part be due to differences in the EcR isoforms present in different cell types.

3.2 Introduction

In Drosophila and other arthropods, pulses of the steroid hormone 20-hydroxyecdysone (20E) are responsible for the temporal coordination of larval molts and metamorphosis. Physiological responses during these events can be diverse; for example during metamorphosis obsolete larval tissues are destroyed and adult structures arise from imaginal disc cells. Remarkably, these actions are carried out in a coordinated, tissue specific manner. At the site of target tissues, 20E binds to its cognate nuclear receptor triggering a cascade of gene activation. Primary 20E-inducible genes, which are directly induced by the steroid-receptor complex, are the earliest genes in the cascade to be transcribed and are insensitive to protein synthesis inhibitors. In contrast, secondary 20E-inducible genes are expressed later and are dependent on the synthesis of primary-response genes. Early studies of the 20E cascade looking at the puffing patterns of polytene chromosomes of late 3rd instar larvae in response to 20E, predicted that primary-response genes would code for proteins that are responsible both for the induction of secondary response genes as well as for the inhibition of their own transcription (Ashburner 1974). Three of the most well characterized 20E-primary response genes, Eip74EF, Eip75B, and br fit this description perfectly. Furthermore, these three genes reside at chromosome cytolocations 74EF, 75B, and 2B5 respectively, which, along with approximately three other loci, exhibit rapid and dramatic puffing after exposure to 20E either naturally or artificially (Ashburner 1971; Ashburner 1972; Ashburner 1973). Likewise, the earliest characterized secondary response genes are found in a region (71E) that forms a distinct yet delayed puff upon 20E exposure
(Restifo and Guild 1986; Wright et al. 1996). Although the precise function of these genes has yet to be determined, based on their genomic sequence they are thought to encode effector proteins which is consistent with early predictions of secondary-response gene function (Wright et al. 1996). Since the original characterization of these 20E-response genes, however, many examples of primary and secondary response genes with diverse functions have emerged, and many of these 20E-inducible genes do not appear to be associated with any identifiable 20E-induced puffs (Guay and Guild 1991; Andres and Thummel 1992; Hurban and Thummel 1993; Andres and Thummel 1995; Fletcher and Thummel 1995; Hock et al. 2000; Beckstead et al. 2005) underling the importance identifying individual components of the cascade for a clearer picture of 20E action.

The transcription factor complex coordinating the entire 20E-hierarchy, the 20E nuclear receptor heterodimer is composed of the vertebrate retinoid X receptor homolog, Ultraspiracle (USP) and the ecdysone receptor (EcR). EcR isoform distribution differs between tissues and typically those tissues expressing different isoforms show different responses to ecdysone at metamorphosis (Robinow et al. 1993; Talbot et al. 1993; Schubiger et al. 1998; Cherbas et al. 2003). For example, immunolocalization experiments examining EcR isoform expression pattern at the onset of metamorphosis show that larval tissues destined to die during metamorphosis, such as the salivary gland, stain strongly for isoform B1 and weakly for isoform A whereas proliferating tissues, such as imaginal discs, show a reciprocal pattern staining strongly for isoform A and weakly for isoform B1 (Talbot et al. 1993). The most recent annotation of the Drosophila genome documents five EcR isoforms that differ in sequence at the amino terminus but share common DNA- and ligand- binding domains. The ligand-binding pocket of the 20E receptor shows remarkable flexibility enabling its activation by a variety of steroidal and non-steroidal 20E analogs/agonists (Billas et al. 2003). For example, the ligand-binding domain of the Drosophila EcR homolog in Heliothis virescens adopts different structures in the presence of different 20E analogs (Billas et al. 2003).

Of the insect and plant derived analogs characterized, ponsterone A (PoA) is the most potent agonist of the EcR receptor (Baker et al. 2000) with an affinity approximately eight times greater than 20E (Maroy et al. 1978). As such, PoA, which differs from 20E by the absence of a single hydroxyl group, has been used to work out numerous kinetic and physical parameters of the EcR (Sage et al. 1982). Interestingly, the natural role of this and other phytoecdysteroids is
still under debate, although most evidence disfavors a hormonal role in plants (Dinan 2001). Instead, it is predicted that phytoecdysteroids induce precocious molting and subsequently death in insects providing an effective defense against insect feeding (Dinan 2001). Given the increased binding affinity of PoA over 20E as well as its potential use in insect control, it is of importance to understand how this phytoecdysosterone affects transcription on a global scale.

The *Drosophila* Kc cell line is one of the most well documented 20E-responsive cell lines currently available and has been used in numerous studies examining the effects of 20E. In this study we use cDNA-based microarrays representing approximately 80% of the *Drosophila* genome to identify 20E-responsive genes in Kc167 cells. By exploiting the sensitivity of secondary-response gene transcription to protein synthesis, we are able to determine which of these genes are primary and which are secondary 20E-inducible genes. Furthermore, the transcriptional response to physiological levels of 20E is compared to that elicited under two other conditions: a 20-fold higher concentration of 20E or its plant derived structural analog, PoA. These analyses led to the identification of 35 genes that reacted similarly to all three treatments. In addition to examining the transcriptional response to 20E in Kc cells, we were also interested in examining the response in a natural target tissue of 20E signaling, salivary glands. Salivary glands have been critical in the elucidation of the 20E signaling hierarchy through the examination of puff patterns on polytene chromosomes. By focusing on the response of the salivary gland in the isolation from other larval tissues we were able to identify many 20E-responsive genes that were not detected when whole larvae were examined (Beckstead et al. 2005). Taken together, this work provides a detailed picture of the genes involved in the ecdysone hierarchy of gene transcription in Kc167 cells and salivary glands.

### 3.3 Methods

#### 3.3.1 Hormone and HS treatments

*Drosophila* Kc167 cells were obtained from *Drosophila* Genomic Resource Centre (Bloomington, IN) and were grown to confluence in Schneider’s medium (Invitrogen) supplemented with 5% heat-inactivated FBS (Sigma) and 20 µg/ml gentamicin (Sigma). Cells were passaged into a series of new flasks that were divided into two groups, experimental and control, and were allowed to recover for one hour. Experimental cells were treated with 0.5 µM 20-hydroxyecdysone (20E) (Sigma) for two hours or with one of the following for four hours:
0.0625 µM ponasterone A (Sigma), 0.5 µM 20E, 10 µM 20E, 0.5 µM 20E plus 100 µM cycloheximide (Sigma), 100 µM cycloheximide alone, or with 0.05% or 0.5% ethanol (solvent used to dissolve hormones). All treatments were performed at 22°C and three independent biological replicates were assessed. For the HS experiments, cells were treated with 0.5 µM 20-hydroxyecdysone (Sigma) for 2 hrs at room temperature either with or without a 15 minute pre-treatment with heat shock (36.5°C) or for 2 hrs at heat shock temperatures (36.5°C).

Late third instar larvae (dp cn bw cl) were selected by the blue gut method as previously described (Andres and Thummel 1994). Salivary glands were dissected from 10 larvae in physiological saline solution and cultured in a 10-well dish with no more than five glands per well in 120 µl of modified TB1 buffer (15 mM HEPES, pH 6.8, 80 mM KCl, 16 mM NaCl, 5 mM MgCl₂, 1% polyethylene glycol 6000) (Bonner 1981; Myohara and Okada 1987) for 1 hour to minimize the effects of any endogenous hormones that might mask the effects of 20E addition. One lobe from each pair of glands was then transferred to fresh TB1 containing 0.5 µM 20E (Sigma) while the sister lobe was transferred to fresh TB1 containing solvent (ethanol). As with the cells, all incubations were carried out for two or four hours and were performed at 22°C in triplicate.

3.3.2 RNA isolation and hybridization to cDNA arrays

TRIzol reagent (Invitrogen) was used to isolate total RNA from both cells and glands according to the manufacturer’s protocol. For each hormone treatment performed, RNA was isolated independently from three biological replicates. Quality and quantity of RNA was verified by spectrophotometry using a nanodrop and the A260/A280 ratios were greater than 1.8. 1 µg of total RNA extracted from glands was subjected to linear amplification using MessageAmp™ II aRNA Amplification Kit (Ambion). Labeling and hybridization to microarrays was carried out as described on the CDMC website (www.flyarrays.com and Neal et al. 2003). Briefly, 2 µg of amplified RNA from glands or 80 µg of total RNA from cells was reverse-transcribed with SuperScriptII (Invitrogen) in the presence of cyanine (Cy) dye coupled nucleotides (Perkin Elmer). After isopropanol precipitation of the labeled cDNA, Cy5 labeled cDNA generated from experimental samples was mixed with Cy3 labeled cDNA from the respective untreated control sample and hybridized to the 12k_v1 cDNA microarray from the Canadian Drosophila
Microarray Centre (CDMC). All downstream processing and hybridization steps were performed exactly as previously described (Neal et al. 2003; Neal and Westwood 2006).

3.3.3 Microarray data acquisition, normalization and analysis

A ScanArray 4000 laser scanner (Perkin Elmer) was used to acquire 16-bit TIFF images of the hybridized arrays that were subsequently analyzed with QuantArray v3.0 software (Perkin Elmer). Quantification data files and their associated images were loaded into GeneTraffic (GT) DUO (Iobion Informatics/Stratagene) where spots with raw intensities less than twice the average background or less than 128 fluorescence units were excluded from further analysis. Normalization and other analysis was performed following the guidelines outlined in Neal et al., 2003 and Neal and Westwood, 2006. Briefly, normalization of the data was performed in GT using the subgrid Lowess algorithm with a 20% smoothing factor for all experiments to correct for systematic differences in data collection. Normalized data from 3 replicate arrays was compiled and imported into Excel (Microsoft) for significance analysis. At this point, genes represented by less than 2/3 valid spots were removed. The Significance Analysis of Microarrays (SAM) package was used to identify genes whose expression significantly differs between samples with a false discovery rate (FDR) of less than 5% (Tusher et al. 2001). Lists of genes found significant by SAM that also showed a change of at least 1.5 fold a coefficient of variation ≤1 were generated in GT DUO for each of the treatments under investigation. Hierarchical clusters were generated using the Pearson uncentered distance metric in MeV (Saeed et al. 2003). The List Functions tool available on the CDMC web site (www.flyarrays.com) was used to compare lists of 20E-responsive genes identified in our study with those previously identified.

3.3.4 Quantitative RT-PCR

One microgram of total RNA isolated from hormone-treated cells or salivary glands was treated with DNase I (Fermentas) to remove contaminating genomic DNA and then reverse-transcribed with SuperScript II reverse-transcriptase (Invitrogen) from anchor oligo dT primers (Invitrogen). The resulting cDNA was treated with RNase H to remove the RNA component of the cDNA-RNA hybrids prior to PCR. The following primers were used for amplification: Act5Cf: GTG CCC ATC TAC GAG GGT TA, Act5Cr: GCC ATC TCC TGC TCA AAG TC, E75f: CTG CCA GTA TTT CCA GTC, E75r: GGA CAA TGT GGG ATA CCT, E74f: CTA
TTC ATG GGC GTT AGT, E74r: GAC AGT TGA AAG GTC ATT AG, Brr: ACA ACA ACA GCC CCG ACT T, Brr: GCT TGT CGC TGA TGG AGA TT, CG5346f: CGC TAG TTC AGG TGT ATC T, CG5346r: ACT TGT GCT CGC TAT ATC T, EcR-RAf: CAT AGG AGT CTT CAG TCT ACA, EcR-RAr: AGA TGG GGA TAG GGA TAC, EcR-RBf: CAT GGA TAC TTG TGG ATT AG, EcR-RBr: CTG GCA GTT GGT CTA TGT, EcR-RCf: TTG TGG ATT AGT AGC AGA AC, EcR-RCr: ACA CTT TCG CCT CAT GTA, EcR-RD/Ef: GCT ATA AAG ACA GGG AGA AC, EcR-RD/Er: GCA AAA TAT GGC TAG GTA AG, EcRf: GGA GAT TCT TGA CCT TAT GA, EcRr: TTT GTA AAC GCT GGT AGA C, USPf: GCG ATG AAA CTG GAG TAG, USPr: TGT AGG GTA TAA GGG ATA GAG. Triplicate qPCR reactions were performed with SYBR qPCR universal kit (KAPA) in a MX4000 qPCR instrument (Stratagene) under the following cycling parameters: 95°C for 10min followed by 40 cycles of 95° for 15s, 55°C for 25s and 72°C for 40s. A dissociation curve was plotted at the end of each run as a quality control for non-specific amplification products. For each gene the fold change ratio (relative to an untreated control) was normalized to Act5C mRNA level and calculated using the Pfaffl ($\Delta\Delta C_t$) method (Pfaffl 2001). The results presented are calculated from the mean fold change of two independent biological replicates.

### 3.3.5 Analysis of functional classes

Following the hierarchical cluster analysis, genes with similar expression profiles were examined to see if their products shared any functional annotations. The CDMC Lookup tool available at www.flyarrays.com was used to obtain their corresponding LocusLink ID (now called EntrezGene ID) for use in DAVID (Dennis et al. 2003; Huang et al. 2009). DAVID assesses functional annotation associated with groups of genes for enrichment over the background represented by all genes on the 12k array. Annotation terms/keywords with EASE scores of $\leq 0.05$ were taken as significantly enriched in a group of related genes and used to assign functional annotation to the group.
3.4 Results and Discussion

3.4.1 Identification of 20-hydroxyecdysone (20E)-responsive genes in Kc167 cells

The 20E-hierarchy of gene transcription serves as a good model for examining hormonal control of development. To gain insight into those genes that are part of the hierarchy, we assayed for 20E-induced changes in gene transcription across the genome. In the first part of this study, the transcriptional response to 0.5 µM 20E was examined in the *Drosophila* cell line, Kc167. Because 0.5 µM 20E causes the polytene chromosome puffs at the 74EF and 75B early 20E-inducible loci to reach their maximum size after 4 hours of organ culture (Ashburner 1973) and *in situ* hybridization using labeled RNA produced after 20E treatment shows puff specific RNAs also increase along with puff size (Bonner and Pardue 1977; Bonner and Pardue 1977) the transcriptional response was examined after a 4 hour treatment with 0.5 µM 20E. The transcriptional profile of cells exposed to 0.5 µM 20E for 2 hours was also examined to identify any genes that may be induced early on but turned off after 4 hours. To allow identification of 20E-responsive genes in a controlled extracellular environment, in the absence of endogenous hormones, these experiments were conducted *in vitro*. Kc167 cells were used because the Kc cell line is responsive to 20E both morphologically (Courgeon 1972; Courgeon 1972; Cherbas et al. 1980) and transcriptionally (Cherbas et al. 1981) and is commonly used for the study of 20E. To examine the transcriptional response to 20E on a genome-wide scale, RNA was isolated from both 20E-treated cells and untreated control cells, reverse-transcribed and labelled with Cy5 or Cy3 dye-coupled nucleotides respectively, and co-hybridized to cDNA microarrays. Following data acquisition and analysis (described in the Methods section), transcripts with at least a 1.5 fold change in abundance in response to 20E were identified. After a 2 hour treatment with 20E, 27 genes were induced (i.e. had increased transcript levels) (Additional File B-1). Two-thirds of these genes (18) including the known 20E-inducible genes *Eip75, Br, Eip28/29*, and *Eip55*, remained significantly up-regulated after 4 hours in the presence of 20E while 51 additional genes were induced and 77 genes were repressed (i.e. had decreased transcript levels) (Additional File B-1). Among the 20E-induced genes identified here in Kc cells, the most strongly induced after 4 hours, *Eip28/29*, was originally identified owing to its quick reaction to 20E in Kc cells (Savakis et al. 1980; Savakis et al. 1984) and has since been used to study aspects of the 20E response such as its tissue specificity (Andres and Cherbas 1992). Taken
together, the response observed here after a 2 and 4 hour 20E-treatment are consistent with two prior observations on the 20E-induced puffing patterns of larval salivary glands: first, that primary 20E-response genes are likely to still be detectable after a four-hour 20E-treatment and second, that the 20E-response of Kc167 is more robust after 4 hours of 20E exposure (Ashburner 1972) (Additional File B-1).

Of the 20E-inducible genes identified here, 30% were also identified in a recent study by Gauhar and colleagues that examined the transcriptional response of Kc cells following either a 1, 3, or 6 hr 20E treatment (Gauhar et al. 2009). However, if we apply more stringent fold-change criteria to the list of 20E-inducible genes we identified, the degree of overlap increases such that at a 2-fold cut-off, there is 50% overlap, and there is 100% overlap between ours and the Gauhar et al. lists if we select a 3-fold change in expression. Remaining differences in the identification of 20E responsive genes identified is likely due in part to the use of different culturing medium as we have previously found that media composition affects the transcriptional response of Kc cells to 20E (unpublished results). The use of a different microarray platform in the measurement of the transcriptional profile could also be a contributing factor.

3.4.2 Identification of primary 20E-response genes

Having identified genes involved in the 20E-response, we were interested in determining where these genes fit into the 20E-hierarchy of gene activation. In particular, we wanted to identify those genes comprising the primary response to the hormone and are thus likely to be directly induced by 20E. Based on earlier studies performed on third instar larvae salivary glands, maximum induction of primary-response genes by 20E is expected to occur after 4 hours of exposure to the hormone, however, some secondary-response gene transcription may have already begun at this point (Ashburner 1972; Ashburner 1974). One distinguishing characteristic of primary and secondary-response genes is their dependence on protein synthesis for transcription; only primary response genes are transcribed in the absence of protein synthesis (Clever 1964; Ashburner 1974). Thus, to identify primary response genes, cells were treated with 20E in conjunction with an inhibitor of protein synthesis, cycloheximide (Table 3-1). RNA isolated from treated cells was labelled and co-hybridized to cDNA arrays along with a differentially labelled untreated control. From this analysis we identified 149 genes that were
Table 3-1 Primary and secondary 20E-regulated genes identified in Kc167 cells

A two-class SAM analysis of triplicate data identified primary-response genes that responded to 20E in the presence of cycloheximide. The top 20 up-regulated primary-response genes and single down-regulated primary-response gene are listed (group E and G respectively). Secondary-response genes regulated by 20E only in the absence of cycloheximide were also identified. The top 10 up- and down-regulated secondary-response genes are listed (groups F and H respectively). The fold change of each gene following a 4 hour treatment with 0.5 µM 20E (20E fold difference) and, for primary-response genes, the fold change following a 4 hour treatment with 0.5 µM 20E and 100 µM cycloheximide (primary fold difference) is given. q-values determined by SAM represent the lowest false discovery rate at which that gene is considered significant. For a gene to be included as differentially expressed, it must have a q-value < 5% in that experimental condition. q-values > 5% are not shown. DAVID (http://david.abcc.ncifcrf.gov/) was used to identify enriched annotation categories among groups of genes that respond similarly to 20E. Some of the most enriched functional terms associated with each group (i.e. terms with modified Fisher Exact p-values ≤ 0.05; where the p-value gives the probability of that term being randomly associated with the genes in the group) are given in the right most column. A complete list of all genes in these groups and their functional annotation is in Additional File B-2.
<table>
<thead>
<tr>
<th>Group</th>
<th>Gene Symbol</th>
<th>20E Fold Difference</th>
<th>q value (%)</th>
<th>Primary fold difference</th>
<th>q value (%)</th>
<th>Select Functional Annotation Terms Enriched in Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>br</td>
<td>3.58</td>
<td>0.00</td>
<td>4.75</td>
<td>0.00</td>
<td>induction of programmed cell death by hormones</td>
</tr>
<tr>
<td></td>
<td>CG5346</td>
<td>3.10</td>
<td>0.00</td>
<td>2.91</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rrp46</td>
<td>2.45</td>
<td>0.00</td>
<td>2.68</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pect</td>
<td>2.89</td>
<td>0.00</td>
<td>2.49</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG14523</td>
<td>1.83</td>
<td>0.40</td>
<td>2.37</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyp9c1</td>
<td>1.87</td>
<td>1.13</td>
<td>2.35</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>granny-smith</td>
<td>1.79</td>
<td>0.00</td>
<td>2.21</td>
<td>0.00</td>
<td>catalytic activity</td>
</tr>
<tr>
<td></td>
<td>CG4825</td>
<td>2.10</td>
<td>0.00</td>
<td>2.14</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG17760</td>
<td>2.51</td>
<td>0.00</td>
<td>2.13</td>
<td>0.00</td>
<td>cytochrome P450, E-class, group I</td>
</tr>
<tr>
<td></td>
<td>Eip55E</td>
<td>1.92</td>
<td>0.00</td>
<td>2.03</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG15482</td>
<td>2.75</td>
<td>0.00</td>
<td>1.83</td>
<td>0.00</td>
<td>hydrolase activity</td>
</tr>
<tr>
<td></td>
<td>Eip75B</td>
<td>1.82</td>
<td>0.00</td>
<td>1.83</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nc</td>
<td>1.40</td>
<td>0.00</td>
<td>1.80</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>eater</td>
<td>1.97</td>
<td>0.00</td>
<td>1.76</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ldgf2</td>
<td>-1.06</td>
<td>1.74</td>
<td>0.00</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG11586</td>
<td>1.00</td>
<td>1.74</td>
<td>1.89</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PRL-1</td>
<td>2.10</td>
<td>0.00</td>
<td>1.70</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>l(2)09851</td>
<td>-1.08</td>
<td>1.70</td>
<td>0.00</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vri</td>
<td>1.52</td>
<td>0.84</td>
<td>1.68</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ImpL2</td>
<td>1.42</td>
<td>1.68</td>
<td>0.00</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Eip71CD</td>
<td>4.06</td>
<td>0.00</td>
<td></td>
<td>4.06</td>
<td>mitochondrial respiratory chain</td>
</tr>
<tr>
<td></td>
<td>CG11893</td>
<td>2.28</td>
<td>0.62</td>
<td></td>
<td>2.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG15711</td>
<td>2.20</td>
<td>0.00</td>
<td></td>
<td>2.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG30104</td>
<td>2.03</td>
<td>0.00</td>
<td></td>
<td>2.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG5059</td>
<td>2.00</td>
<td>0.34</td>
<td></td>
<td>2.00</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td></td>
<td>CG17819</td>
<td>2.00</td>
<td>0.72</td>
<td></td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG5104</td>
<td>1.99</td>
<td>0.00</td>
<td></td>
<td>1.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CG31633</td>
<td>1.89</td>
<td>2.76</td>
<td></td>
<td>1.89</td>
<td>cellular metabolic process</td>
</tr>
<tr>
<td></td>
<td>CG5694</td>
<td>1.87</td>
<td>4.45</td>
<td></td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obp8a</td>
<td>1.83</td>
<td>3.56</td>
<td></td>
<td>1.83</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Aldh</td>
<td>-1.32</td>
<td>3.56</td>
<td>-2.05</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>CG8801</td>
<td>-1.84</td>
<td>0.62</td>
<td></td>
<td>-1.84</td>
<td>nucleotide-binding</td>
</tr>
<tr>
<td></td>
<td>ncd</td>
<td>-1.92</td>
<td>0.38</td>
<td></td>
<td>-1.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hsp70Aa</td>
<td>-2.01</td>
<td>0.49</td>
<td></td>
<td>-2.01</td>
<td>kinase activity</td>
</tr>
<tr>
<td></td>
<td>Hsp68</td>
<td>-2.13</td>
<td>1.13</td>
<td></td>
<td>-2.13</td>
<td>glycolysis</td>
</tr>
<tr>
<td></td>
<td>Scp2</td>
<td>-2.13</td>
<td>4.45</td>
<td></td>
<td>-2.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gammaTub23C</td>
<td>-2.17</td>
<td>1.13</td>
<td></td>
<td>-2.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hsp60</td>
<td>-2.43</td>
<td>0.00</td>
<td></td>
<td>-2.43</td>
<td>cytoskeleton organization and biogenesis</td>
</tr>
<tr>
<td></td>
<td>Acon</td>
<td>-2.51</td>
<td>1.76</td>
<td></td>
<td>-2.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vago</td>
<td>-2.57</td>
<td>0.00</td>
<td></td>
<td>-2.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hsp70Ab</td>
<td>-2.91</td>
<td>1.13</td>
<td></td>
<td>-2.91</td>
<td></td>
</tr>
</tbody>
</table>
induced and 119 genes that were repressed by the 20E-cycloheximide combination (data not shown). To determine which of these genes were induced/repressed in response to 20E, and were not responding to exposure to cycloheximide alone, additional microarrays were run using RNA from Kc167 cells treated only with cycloheximide. From this experiment, 214 genes were identified whose transcription is altered due solely to the inhibition of protein synthesis (data not shown). A two-class SAM analysis identified 35 genes whose transcripts levels in the presence of cycloheximide differs significantly when 20E is present including the well characterized 20E-primary-response genes, Eip75B and br (Table 3-1 and Additional File B-2). Although Eip74EF was not among the genes on this list, we confirmed its induction by 20E by qRT-PCR (Additional File B-3) suggesting that the cDNA probe on the batch of arrays used for this study was likely of poor quality and unable to detect transcripts of this gene. Of the 35 primary-response genes identified, 15 were not induced by 20E alone. The induction of these 15 genes by 20E is likely masked in the absence of cycloheximide due to a negatively regulated auto-feedback loop where primary-response gene transcription is repressed by primary-response gene protein products (Ashburner 1974). Interestingly, the only gene transcript repressed as part of the primary response to 20E, CG3752, also evades detection unless protein synthesis is blocked (Table 3-1). The majority of 20E-responsive genes identified after a 4 hour treatment with 0.5 µM 20E behaved as secondary-response genes as their response to 20E was blocked in the presence of cycloheximide; 48 genes were induced and 77 were repressed as part of the secondary response (Table 3-1 and Additional File B-2).

In addition to differences in transcriptional response in the absence of protein synthesis, it is also expected that most primary-response genes encode regulators while secondary-response genes encode effectors (Ashburner 1974; Guay and Guild 1991; Fletcher and Thummel 1995; Wright et al. 1996). To determine if this is the case for the genes identified here as part of the primary and secondary response, we examined each group of genes for enrichment in functional annotation terms / keywords using the online resource DAVID. The results of this analysis are summarized in Table 3-1. Several of the primary-response gene transcripts include those that code for the well-known DNA binding transcription factors Eip75B and br as well as vrielle which is consistent with the model that primary-response proteins are required for the transcriptional induction of the secondary-response genes. “Induction of programmed cell death by hormones” is among the most highly enriched terms associated with primary-response genes
and comes as no surprise due to the well-established involvement of 20E-regulated genes such as br and Nc in cell death pathways during Drosophila development (Dorstyn et al. 1999; Jiang et al. 2000; Cakouros et al. 2002; Cakouros et al. 2004). However, we should point out that Kc cells do not undergo apoptotic cell death in response to ecdysone treatment and therefore other factors and/or conditions must need to be present in order to carry out this particular process.

“Catalytic activity” and “hydrolase activity” were both also among the most significantly enriched terms supporting the idea that primary-response genes are involved in a broad range of regulatory roles (Guay and Guild 1991; Andres and Thummel 1992; Hurban and Thummel 1993; Andres and Thummel 1995; Fletcher and Thummel 1995; Hock et al. 2000; Beckstead et al. 2005). For example, one of the genes belonging to the hydrolase class includes imaginal disc growth factor 2, a protein that when bound to its receptor initiates signal transduction cascades important to imaginal disc development. Some primary- and secondary-response genes include genes that code for proteins involved in cell movement and organization and/or are associated with the cytoskeleton. These include up-regulated transcripts for Roadblock, a gene coding for a dynein-associated protein and down-regulated transcripts for Spastin, and a gene coding for a microtubule severing protein. Transcript changes in genes of this sort are consistent with the cell movement and morphogenetic changes that occur during 20E-dependent developmental changes. Also of interest is that several of the secondary-response genes include genes that are involved in metabolic processes. More specifically, there seems to be an increase in transcripts for genes involved in mitochondrial respiration and a decrease in transcripts for genes involved in glycolysis, perhaps suggesting that in addition to the cellular organization and tissue changes that are induced by 20E, there is also a shift in how energy is being produced.

3.4.3 Transcriptional response of Kc167 cells to an increased concentration of 20E and to its structural analog, ponasterone A (PoA)

To examine the specificity of the 20E-response of Kc167 cells to both hormone concentration and ligand structure, two other conditions were tested: a 20-fold higher concentration of 20E, 10 μM, and the replacement of 20E with its structural analog, PoA. 20E concentration is known to affect at least three aspects of the 20E-response: the size of both early and late puffs (Ashburner 1973), the rate of early puff regression (Ashburner 1973), and the transcription of some early gene isoforms (Karim and Thummel 1992). PoA, often used as substitute for 20E, is a more
potent activator of the 20E-response. At least four different measures indicate that PoA activity is approximately eight fold higher than 20E activity: affinity for the ecdysone receptor (Maroy et al. 1978), morphological changes associated with the 20E-response (Cherbas et al. 1980), the level of induction of known 20E-inducible proteins (Savakis et al. 1980), and the level of transcription from an ecdysone response element (EcRE) (Baker et al. 2000). Thus, to minimize effects caused by the difference in receptor affinity and to make the comparison between analogs as similar as possible, our working concentration of PoA was eight fold less than that of 20E (ie. 0.0625 µM PoA versus 0.5 µM 20E).

Genes that responded to either of the two treatments were identified by microarray analysis as described above. Compared to the number of genes induced by a 4 hour treatment with 0.5 µM 20E, both 10 µM 20E and 0.0625 µM PoA stimulated the transcription of a greater number of genes: 85 and 115 respectively versus 69 induced by 0.5 µM 20E (Additional File B-4). Furthermore, PoA down-regulated substantially more genes than either 20E treatment. One-hundred and fifty genes were repressed after a 4 hour treatment with PoA - almost double the number repressed by 0.5 µM 20E (79) and nearly six fold more than the number repressed by 10 µM 20E (26) (Additional File B-4). It is possible that the differences in the response to PoA may be due to the binding and activation of additional nuclear receptors; currently it is only assumed to be the EcR/USP heterodimer because it has been shown to be the case in vitro (Dinan 2001).

Overall, 379 genes were identified that responded to at least one of the three hormone treatments examined (0.5 µM and 10 µM 20E and 0.0625 µM PoA; Additional File B-4). To determine which of these 379 hormone-responsive genes were sensitive to all three conditions tested and which are differentially regulated by either 10 µM 20E or PoA relative to 0.5 µM 20E, a multi-class SAM analysis was performed where the transcriptional response to 10 µM 20E and PoA was compared to the transcriptional response to 0.5 µM 20E for each gene. This analysis revealed two things about the transcriptional response to each of the hormone conditions tested. First, of the 379 genes that responded to at least one of the three treatments, 35 showed no significant difference in their transcriptional response to all three treatments and eight of these are primary 20E-inducible genes (Table 3-2). Second, PoA affected the transcription of a set of genes
Table 3-2 Genes that respond similarly to 0.5 μM 20E, 10 μM 20E, or 0.0625 μM PoA.

A multi-class SAM analysis revealed genes with similar transcriptional responses to the hormones and hormone concentrations studied. Genes identified as primary 20E-response genes (see Table 3-1) are indicated in bold. As with Table 3-1, the fold change and q-value of each gene following the treatment indicated is given. A missing fold change value means that gene was either not represented by at least 2/3 spots and/or did not exhibit a coefficient of variation of at least 1 for that treatment across replicate arrays.

<table>
<thead>
<tr>
<th>Gene Symbol or Clone ID</th>
<th>0.5 μM 20E Fold Difference</th>
<th>q value (%)</th>
<th>10 μM 20E Fold Difference</th>
<th>q value (%)</th>
<th>0.0625 μM PoA Fold Difference</th>
<th>q value (%)</th>
<th>GO Molecular Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eip71CD</td>
<td>4.07</td>
<td>0.00</td>
<td>3.58</td>
<td>0.00</td>
<td>3.51</td>
<td>0.00</td>
<td>oxidoreductase activity</td>
</tr>
<tr>
<td>CG5346</td>
<td>3.10</td>
<td>0.00</td>
<td>2.88</td>
<td>0.00</td>
<td>3.49</td>
<td>0.00</td>
<td>cation binding</td>
</tr>
<tr>
<td>Granny-Smith</td>
<td>1.79</td>
<td>0.40</td>
<td>1.79</td>
<td>0.00</td>
<td>1.94</td>
<td>0.00</td>
<td>cation binding</td>
</tr>
<tr>
<td>RhoGAP16F</td>
<td>2.88</td>
<td>0.00</td>
<td>2.91</td>
<td>0.00</td>
<td>2.49</td>
<td>0.00</td>
<td>transferase activity</td>
</tr>
<tr>
<td>PRL-1</td>
<td>2.10</td>
<td>0.00</td>
<td>2.12</td>
<td>0.00</td>
<td>1.58</td>
<td>0.00</td>
<td>hydrolase activity</td>
</tr>
<tr>
<td>CG4825</td>
<td>2.10</td>
<td>0.00</td>
<td>2.17</td>
<td>0.00</td>
<td>2.56</td>
<td>0.00</td>
<td>transferase activity</td>
</tr>
<tr>
<td>CG30104</td>
<td>2.03</td>
<td>0.00</td>
<td>2.40</td>
<td>0.00</td>
<td>2.39</td>
<td>0.00</td>
<td>hydrolase activity</td>
</tr>
<tr>
<td>Cyp4e2</td>
<td>1.97</td>
<td>0.00</td>
<td>2.52</td>
<td>0.00</td>
<td>2.20</td>
<td>0.00</td>
<td>receptor binding</td>
</tr>
<tr>
<td>CG14523</td>
<td>1.83</td>
<td>0.40</td>
<td>2.04</td>
<td>0.00</td>
<td>1.81</td>
<td>0.00</td>
<td>monooxygenase activity</td>
</tr>
<tr>
<td>LD33681</td>
<td>1.83</td>
<td>0.40</td>
<td>2.05</td>
<td>0.00</td>
<td>2.02</td>
<td>0.00</td>
<td>peptidase activity</td>
</tr>
<tr>
<td>CG7706</td>
<td>1.35</td>
<td>1.50</td>
<td>1.54</td>
<td>0.00</td>
<td>1.74</td>
<td>0.00</td>
<td>N/A</td>
</tr>
<tr>
<td>GstE1</td>
<td>1.05</td>
<td>-1.01</td>
<td>-1.52</td>
<td>0.00</td>
<td>transferase activity</td>
<td>N/A</td>
<td>ATP-dependent helicase activity</td>
</tr>
<tr>
<td>CG14739</td>
<td>-1.22</td>
<td>0.62</td>
<td>-1.53</td>
<td>0.63</td>
<td>ubiquitin conjugating activity</td>
<td>N/A</td>
<td>hydrolyase activity</td>
</tr>
<tr>
<td>CG10641</td>
<td>-1.49</td>
<td>0.00</td>
<td>-1.58</td>
<td>0.00</td>
<td>N/A</td>
<td>N/A</td>
<td>protein binding</td>
</tr>
<tr>
<td>CG13868</td>
<td>-1.35</td>
<td>0.49</td>
<td>-1.35</td>
<td>0.30</td>
<td>-1.58</td>
<td>0.00</td>
<td>protein binding</td>
</tr>
<tr>
<td>arm</td>
<td>-1.39</td>
<td>2.17</td>
<td>-1.58</td>
<td>0.00</td>
<td>-1.07</td>
<td>0.00</td>
<td>N/A</td>
</tr>
<tr>
<td>Rpn1</td>
<td>-1.41</td>
<td>-1.15</td>
<td>-1.54</td>
<td>0.49</td>
<td>N/A</td>
<td>N/A</td>
<td>alpha-catenin binding</td>
</tr>
<tr>
<td>elf3-S10</td>
<td>-1.47</td>
<td>0.00</td>
<td>-1.34</td>
<td>1.60</td>
<td>-1.58</td>
<td>0.00</td>
<td>N/A</td>
</tr>
<tr>
<td>Rhogap16F</td>
<td>-1.55</td>
<td>0.00</td>
<td>-1.36</td>
<td>0.00</td>
<td>-1.75</td>
<td>0.28</td>
<td>N/A</td>
</tr>
<tr>
<td>RH07164</td>
<td>-1.65</td>
<td>0.00</td>
<td>-1.34</td>
<td>0.00</td>
<td>-1.59</td>
<td>0.00</td>
<td>GTPase activator activity</td>
</tr>
</tbody>
</table>

Table 3-1 Analysis of genes identified as primary 20E-response genes following the indicated treatments. Genes identified as primary 20E-response genes are indicated in bold. A fold change of 1.0 indicates no change compared to the control. A fold change > 1.0 indicates induction, while a fold change < 1.0 indicates repression. A q-value of 0.05 indicates statistical significance.
distinct from those affected by both the 10 μM 20E and 0.5 μM 20E treatment (Figure 3-1), suggesting that PoA may not be acting solely through the ecdysone hierarchy. We confirmed the transcriptional response of three key 20E-inducible genes (*Eip74EF*, *Eip75B*, and *br*), the ecdysone-receptor complex, and a novel 20E-incudible gene (CG5346) by qRT-PCR. In all cases, the transcriptional response for these genes measured by qRT-PCR, although differing in magnitude when compared to the microarray analysis (i.e. showing on average a 5.58 fold more induction than the microarray data), had the same direction of change (Additional File B-3).

Given the observed similarities and differences in the ecdysteroid-response to both an increased 20E-concentration and the use of a structural analog, we wanted to identify functions associated with genes that responded similarly to all treatments tested. Using DAVID we found two categories enriched in similarly responding genes that were also enriched among primary-response genes: “catalytic activity” and “hydrolase activity” (Table 3-3). Conversely, genes that reacted significantly different to either 10 μM 20E, PoA or 0.5 μM 20E were enriched in many unique categories; the most strongly enriched functional terms are associated with genes down-regulated by PoA relative to 0.5 μM 20E (Table 3-3). Furthermore, other categories related to development are also strongly enriched in the list of genes that are down-regulated by PoA relative to 0.5 μM 20E such as “anatomical structure development”, “induction of programmed cell death”, and “instar larval or pupal development” (Table 3-3 and Figure 3-1). The apparent lack of activation of developmental genes by PoA, as indicated by this analysis, is suggestive of an alternate mode of action of PoA induced death in non-adapted insects more so than the simple induction of premature molting via the canonical ecdysone hierarchy. In fact, the results potentially suggest that PoA might interfere with normal 20E induced transcriptional changes and prevent the induction of some of the genes that are required for 20E dependent developmental processes.

### 3.4.4 Transcriptional response of salivary glands to 20E

We wanted to extend our analysis to a 20E-responsive larval tissue to see how it compares to the 20E transcriptional response of Kc167 cells. Salivary glands were chosen in this respect due to the relative ease of dissection and because the foundations of our understanding of the 20E response comes from the study of salivary glands. Cultured glands from 10 late third instar
Figure 3-1 Effect of PoA and increased 20E concentration on gene regulation

A multi-class SAM analysis identified groups of genes that are differentially regulated by either 10 µM 20E and/or 0.0625 µM PoA relative to 0.5 µM 20E from a list of genes exhibiting a +/- 1.5 fold change in response to at least one of the two treatments. Genes showing a transcriptional response to either 10 µM 20E or 0.0625 µM PoA that significantly differed from its response to 0.5 µM 20E were clustered based on fold difference in expression relative to its response to 0.5 µM 20E. Four sub-clusters are highlighted that contain groups of genes that respond similarly to at least one of the two altered treatments. Grey represents missing data.
Table 3-3 Annotation enrichment of genes regulated by either 0.5µM or 10µM 20E or 0.0625µM PoA

Expression differences are relative to the 0.5 µM 20E dataset (i.e. “induced by PoA” indicates genes that were more greatly induced by 0.0625 µM PoA than by 0.5 µM 20E). DAVID (http://david.abcc.ncifcrf.gov/) was used to identify enriched annotation categories among groups of genes that responded similarly to one or more of the hormone treatments examined in this study (Additional File B-4). The modified Fisher Exact p-value represents the probability that the number of genes associated with an annotation term in the sample list is random compared to the occurrence of the term in the all of the genes on the array used for the analysis (12kv1). (BP) biological process; (MF) molecular function; (CC) cellular component; (SP) Swiss Prot Keywords; (INT) Interpro Name; (UP) UniProt Sequence Feature.
<table>
<thead>
<tr>
<th>Group</th>
<th>Response</th>
<th># Genes in Group with Valid Identifier/ Total # in Group</th>
<th>Category</th>
<th>Term</th>
<th>p-value</th>
<th>Number of Genes from Group Annotated with Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Common to all three conditions tested</td>
<td>30/35</td>
<td>MF</td>
<td>catalytic activity</td>
<td>9.35E-03</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>positive regulation of translation</td>
<td>2.84E-02</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>protein metabolic process</td>
<td>2.84E-02</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>primary metabolic process</td>
<td>2.96E-02</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MF</td>
<td>hydrolase activity</td>
<td>3.81E-02</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Induced by both PoA &amp; 10 µM 20E</td>
<td>54/58</td>
<td>BP</td>
<td>response to temperature stimulus</td>
<td>4.16E-03</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SP</td>
<td>DNA binding</td>
<td>1.40E-02</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>lipid transport</td>
<td>2.87E-02</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Repressed by both PoA &amp; 10 µM 20E</td>
<td>46/54</td>
<td>CC</td>
<td>intracellular part</td>
<td>1.34E-02</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>protein folding</td>
<td>1.99E-02</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>generation of precursor metabolites and energy</td>
<td>2.94E-02</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>mitochondrial respiratory chain</td>
<td>3.63E-02</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Induced by PoA</td>
<td>87/89</td>
<td>CC</td>
<td>organelle part</td>
<td>6.17E-04</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>organelle membrane</td>
<td>3.47E-03</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>cytoplasm</td>
<td>4.16E-03</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>intracellular protein transport</td>
<td>1.08E-02</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SP</td>
<td>ribonucleoprotein</td>
<td>1.17E-02</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>cytoplasmic sequestering of transcription factor</td>
<td>2.55E-02</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Repressed by PoA</td>
<td>133/145</td>
<td>BP</td>
<td>developmental process</td>
<td>2.39E-06</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MF</td>
<td>nucleoside-triphosphatase activity</td>
<td>1.84E-05</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>cell differentiation</td>
<td>1.86E-05</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MF</td>
<td>hydrolase activity, acting on acid anhydrides</td>
<td>2.84E-05</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>gamete generation</td>
<td>3.68E-05</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>sexual reproduction</td>
<td>5.17E-05</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>anatomical structure development</td>
<td>6.51E-05</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>organelle organization and biogenesis</td>
<td>1.98E-04</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>neuron development</td>
<td>2.00E-04</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>system development</td>
<td>2.08E-04</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MF</td>
<td>helicase activity</td>
<td>9.13E-04</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>induction of programmed cell death</td>
<td>1.16E-03</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MF</td>
<td>ATPase activity</td>
<td>1.47E-03</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>instar larval or pupal development</td>
<td>2.10E-03</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>organ development</td>
<td>3.44E-03</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>immune system development</td>
<td>4.29E-03</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>DNA packaging</td>
<td>9.87E-03</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BP</td>
<td>response to ecdysone</td>
<td>3.73E-02</td>
<td>3</td>
</tr>
</tbody>
</table>
larvae staged by the blue gut method (see Methods section) were treated with 0.5 µM 20E for four hours. We chose to treat salivary glands prior to the natural 20E pulse rather than assay glands after the natural pulse to rule out potential confounding effects caused by other endogenous hormones and/or the presence of protein products of secondary-response genes within the 20E hierarchy. RNA was isolated from 20E treated glands and untreated control glands, amplified, and analyzed by microarray experiments as described in the Methods section. We identified 98 up-regulated genes including several known 20E-inducible genes such as, *Eip75B, Eip74EF, EcR, Nc*, and *ImpL2*, and 8 down-regulated genes (Additional File B-4). In a study by Beckstead and colleagues looking at 20E responsive genes in cultured larval organs 743 genes were identified as 20E responsive (Beckstead et al. 2005). Their treatments, however, utilized twice the concentration of 20E and lasted for an additional two hours. Nevertheless, we were still able to identify 52 genes in common, the majority of which showed similar changes in expression level (Figure 3-2 and Additional File B-5). By focusing only on salivary glands in this study we were able to identify 46 salivary gland specific 20E-inducible genes that were not identified in total organ preparations (Beckstead et al. 2005) (Figure 3-2 and Additional File B-6). A DAVID analysis reveals that these salivary gland specific genes are enriched for protein binding (data not shown).

Genes induced by 20E in salivary glands but not in Kc167 cells were enriched in Gene Ontology (GO) biological processes such as “salivary gland histolysis”, “programmed cell death” and various “developmental” processes as might be expected. “Cellularization” and “transporter activity” were also among the highest scoring hits. Conversely, genes that were up regulated only in Kc cells showed enrichment for “metabolic process” and “oxidoreductase activity”-categories that were also enriched among secondary response genes in Kc cells suggesting that the secondary response genes differ between Kc cells and salivary glands (Table 3-4). Comparison of 20E responsive genes from both systems allowed the identification of 10 genes whose response is conserved across both systems suggesting that they might occupy key positions within the hierarchy (Figure 3-3). Among the most highly induced of these genes are two genes that have not been traditionally associated with the 20E response in *Drosophila, CG6579* and *Rrp46* (Additional File B-4). CG6579 is currently un-annotated and Rrp46 is a component of the exosome that has been shown to relocate to developmental loci during periods
**Figure 3-2** Overlap of genes regulated by 20E in salivary glands and whole organ culture.

Venn diagram depicting overlap between number of genes identified as part of the 20E response in salivary glands and genes identified by Beckstead and co-workers (Beckstead et al. 2005) in whole organ culture as part of the primary or secondary response to 20E.

**Figure 3-3** The transcriptional responses of Kc167 cells and salivary glands to 20E exhibit very little overlap

Roughly equal numbers of genes are induced (black bars) and repressed (grey bars) in cells by 20E. In salivary glands, the response to 20E is largely stimulatory with nearly 10 times more genes being induced as repressed. Only a small number of genes are regulated by 20E in both cells and glands (common).
Table 3-4 Annotation enrichment of genes that respond to 20E in either Kc167 cells or salivary glands.

DAVID (http://david.abcc.ncifcrf.gov/) was used to identify enriched annotation categories among groups of genes that respond similarly to 20E. The modified Fisher Exact p-value represents the probability that the number of genes associated with an annotation term in the sample list is random compared to the occurrence of the term in all of the genes on the array used for the analysis (12kv1). (BP) biological process; (MF) molecular function; (CC) cellular component; (SP) Swiss Prot Keywords; (INT) Interpro Name; (UP) UniProt Sequence Feature.

<table>
<thead>
<tr>
<th>Response to 20E</th>
<th>Category</th>
<th>Term</th>
<th>p-value</th>
<th>Number in Category</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Induced in Kc cells</strong></td>
<td>56/69</td>
<td># Genes in Group with Valid Identifier/Total # in Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>chromoprotein</td>
<td>3.3E-03</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>peptidase M13</td>
<td>4.3E-03</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>electron transport</td>
<td>4.5E-03</td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>metabolic process</td>
<td>5.5E-03</td>
<td>31</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>oxidoreductase activity</td>
<td>1.0E-02</td>
<td>10</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>glutathione S-transferase</td>
<td>1.0E-02</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>generation of precursor metabolites and energy</td>
<td>1.6E-02</td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>iron ion binding</td>
<td>3.1E-02</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>cellularization</td>
<td>2.4E-04</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>salivary gland histolysis</td>
<td>4.9E-04</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>programmed cell death</td>
<td>2.8E-03</td>
<td>9</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>cellular developmental process</td>
<td>3.4E-03</td>
<td>18</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>exocrine system development</td>
<td>3.4E-03</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>germ-band shortening</td>
<td>5.7E-03</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>gland development</td>
<td>6.6E-03</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>sperm individualization</td>
<td>1.5E-02</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>UP</td>
<td>compositionally biased region:Poly-Ala</td>
<td>1.4E-02</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>kinase activity</td>
<td>1.4E-02</td>
<td>9</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>regulation of metabolic process</td>
<td>1.5E-02</td>
<td>15</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>ligand-dependent nuclear receptor activity</td>
<td>2.8E-02</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>cell differentiation</td>
<td>1.7E-02</td>
<td>16</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>female meiosis chromosome segregation</td>
<td>3.2E-02</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>cytoskeletal protein binding</td>
<td>3.0E-02</td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>embryonic development via syncytial blastoderm</td>
<td>4.3E-02</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>regulation of cellular metabolic process</td>
<td>4.8E-02</td>
<td>13</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>cell division/GTP binding protein</td>
<td>5.2E-02</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Induced in salivary glands</strong></td>
<td>87/98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>nucleotide-binding</td>
<td>3.2E-08</td>
<td>20</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>kinase activity</td>
<td>6.2E-04</td>
<td>11</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>chaperonin Cpn60</td>
<td>5.7E-03</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>unfolded protein binding</td>
<td>1.2E-02</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>glycolysis</td>
<td>1.5E-02</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>cytoskeleton organization and biogenesis</td>
<td>2.5E-02</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td><strong>Repressed in Kc cells</strong></td>
<td>71/77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Repressed in salivary glands</strong></td>
<td>9</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
of active transcription and is involved in mRNA processing (Andrulis et al. 2002). *Rrp46* showed the greatest correlation in expression profile to *Eip75B* highlighting the importance of gene transcription in the ec dysone response.

### 3.4.5 Tissue specific differences in ec dysone receptor (EcR) isoform induction by 20E

Some of the differences in the transcriptional response of Kc cells and salivary glands may be explained by differential chromatin states of inactive genes in the two systems. For example, salivary gland secretion protein (*sgs*) genes known be under the control of the hierarchy were expressed in gland cells but not in Kc cells most likely because of the difference in the two cell types. Differential gene regulation by 20E may also result from the expression of different *EcR* isoforms or combinations thereof (Robinow et al. 1993; Talbot et al. 1993; Schubiger et al. 1998; Mouillet et al. 2001). Typically, tissues with dissimilar developmental fates express different *EcR* isoforms (Robinow et al. 1993; Talbot et al. 1993; Schubiger et al. 1998). Our microarray results agree with previous findings that Kc cells express hemocyte and plasmatocyte marker genes *Pxn*, *ush*, and *Hml* (Cherbas et al. 2011) (data not shown) and that 20E treatment of Kc cells causes the induction of *Eip28/29* and *Eip55* (*Eip40*), both of which are 20E-regulated in the hematopoietic lymph gland (Andres and Cherbas 1992) providing further support for an embryonic hemocyte origin of Kc cells (Andres and Cherbas 1992; Cherbas et al. 2011). In contrast to larval salivary glands, embryonic hemocytes persist in the adult fly (Holz et al. 2003) suggesting that Kc cells are likely more similar to imaginal disc or adult tissues than larval tissues and placing them in a different metamorphic class than the salivary gland (Berger and Morganelli 1984).

Because *EcR* is a 20E-inducible gene, we first examined the overall level of induction of *EcR* by qRT-PCR using a primer designed against the 3′ region of the gene common to all isoforms (Figure 3-4). Interestingly, unlike many of the 20E-inducible genes we identified by microarray analysis, *EcR* was induced to a similar extent in PoA-treated Kc cells, 20E-treated Kc cells and 20E-treated salivary glands (Figure 3-4). We next examined the transcriptional response of each *EcR* isoform to PoA and 20E using primers that recognize unique regions of each isoform and found that there was no difference in isoform regulation in Kc cells treated with either hormone (Figure 3-4). Comparison of isoform induction in 20E-treated Kc cells to 20E-treated glands,
Figure 3-4 The transcription of EcR isoforms are differentially regulated in Kc cells and salivary glands. Relative induction of USP, EcR and five EcR isoforms in Kc cells and salivary glands by either 0.5 μM 20E or 0.0625 μM PoA as determined by qRT-PCR relative to untreated Kc cells or salivary glands, respectively.
however, revealed a tissue specific pattern in isoform regulation (Figure 3-4). At least two EcR isoforms (EcR-RA and EcR-RD and/or EcR-RE; where EcR-RD and EcR-RE were indistinguishable by qRT-PCR analysis and will be referred to hereafter as EcR-RD/E) were induced in Kc cells by both PoA and 20E but not induced at all in salivary glands treated with 20E (Figure 3-4). It has previously been shown that the proteins encoded by the EcR-RA, EcR-RD and EcR-RE isoforms show strong expression in both embryos and imaginal discs (Talbot et al. 1993) and EcR-A is thought to be responsible for adult differentiation (Schubiger et al. 2003) EcR-RC, on the other hand, although induced by both PoA and 20E in Kc cells, was more strongly induced by 20E in salivary glands (Figure 3-4).

Although differences in EcR isoform regulation in Kc cells and salivary glands by 20E may be linked to the variation in the observed transcriptional response to 20E, it is unlikely that the same explanation can account for the difference in the response observed in Kc cells treated with 20E and PoA, since qRT-PCR analysis reveals at best, subtle differences in EcR isoform regulation by 20E and PoA (Figure 3-4). It is tempting to speculate that the different EcR isoforms form 20E receptors with different target genes and thus contribute to differences in the 20E induced transcript profiles. Alternatively, the differences in transcript profiles may have very little to with which EcR isoforms are present and more to do with cell type differences in other proteins. For example, differences in the concentrations of particular chromatin modifying proteins could affect the accessibility of the various target genes differently in different cell types. Also, the relative amount of proteins that partner with EcR to make the 20E heterodimer nuclear receptor could affect the sites the 20E receptor binds to. While the classical partner for EcR has been thought to be Ultraspireacle (USP), it is now known that biologically active 20E receptors can be made without USP (Costantino et al. 2008). In mid-third instar larvae, the 20E dependent induction of the glue proteins requires EcR but not USP suggesting that EcR makes 20E receptors that consist of EcR and a yet to be characterized partner protein (Costantino et al. 2008). One approach to assess the relative contributions of each of the EcR isoforms would be to employ RNAi knockdown strategies of the different EcR isoforms in both Kc cells and larval tissues and examine what affect this has on the 20E induced transcript profiles.
3.4.6 Heat shock represses the ecdysone-response

It has been previously shown that the developmentally regulated chromosomal puffs like the 20E induced puffs regress during HS (Ashburner 1970; Westwood et al. 1991). Surprisingly, the loci where Eip74EF and Eip75B are located, 74EF and 75B respectively, are also occupied during HS by heat shock factor (HSF), the transcription factor that regulates transcription in response to heat stress (Westwood et al. 1991). Given that HSF-binding coincides with puff regression at these loci, we were interested in determining if HS would affect transcription of the 20E-response genes, so we examined the global transcriptional response of Kc cells to a 2 hr 20E treatment followed by both transient (15 min HS followed by 2hr 20E treatment at RT) and sustained exposure to HS (HS for the entire duration of the 2hr 20E treatment) (Figure 3-5A; data available in GEO under GSE23824). Our results support the previous observation that HS results in the repression of global gene expression including genes that are directly associated with HSF-binding sites such as Eip75B and Eip74EF (Westwood et al. 1991). The magnitude of repression is related to the duration of the stress; prolonged exposure to HS inhibits gene expression to a much greater degree than exposure to brief and transient HS treatment and 20E-induced gene transcription starts to recover to normal levels following removal of the stress with the primary 20E-response genes being the first to return to normal levels (Figure 3-5A).

Since the binding of HSF to 74EF and 75B occurs at the same time as the transcriptional repression of genes located there, we were interested in identifying the location of HSF binding site(s) in relation to these genes. To do this, we made use of the HSF binding data collected in chapter 2. Both Eip75B and Eip74EF as well as br (which is located on an entirely different chromosome at 2B) are very large in size spanning from 59 to over 100 kb and each encodes multiple isoforms with at least two distinct transcription start sites (for review see (Thummel 1990)). Within these three genes, we have found a total of six HSF-binding sites, four within Eip75B, one within Eip74EF and one within br (Figure 3-6). Interestingly, all but one of these intronic HSF-binding sites are located less than 5 kb upstream of the 5’ end of one or more isoforms. Given that introns in these genes that are bound by HSF are rather large (for example the first intron of Eip75B in which HSF occupies two sites, is over 60 kb in length and the intron bound by HSF in Eip74EF is over 20 kb) this distribution of HSF-binding sites appears to exhibit a strong bias for the extreme 3’ end of the introns close to the 5’ ends of alternate
Figure 3-5 Heat shock prevents transcription of ecdysone-regulated genes in Kc cells.

(A) Microarray analysis identifies 27 genes that are induced at least 1.5 fold in Kc cells by a 2 hr 0.5 µM ecdysone treatment under non-HS conditions (left most column). The number of genes that respond transcriptionally to ecdysone is greatly reduced when ecdysone is administered immediately following a brief, 15 min, HS. Even those genes that are still induced by ecdysone are induced to a lesser extent (middle column). When subjected to HS, however, the same genes that were induced by ecdysone under non-HS conditions, no longer respond transcriptionally to the same ecdysone treatment (ie. 2 hr 0.5 µM ecdysone) indicting that HS can repress ecdysone-inducible gene transcription. (B) Quantitative RT-PCR verification of the microarray results for the major ecdysone-inducible genes Eip75B, Eip74EF, and br yields the same conclusion. All three genes are induced when exposed to 2 hr 0.5 µM ecdysone treatment at RT (grey bars) but not during HS (black bars). Error bars represent standard error.
Figure 3-6 HSF binds to introns in the three major ecdysone-inducible genes, Eip75B, Eip74EF, and br.

This image, generated using the UCSC Genome Browser, illustrates the chromosome region represented in bp (according to release 4.2 of the Drosophila genome) as indicated at the top. Genes are depicted as blue boxes with the thick and thin parts representing exons and introns respectively. Blue arrows within the gene indicate the direction of transcription. Each orange bar represents a probe from the tiling array that is part of an HSF bound segment and its height indicates its fold enrichment relative to whole cell extract (WCE) (data from experiments presented in Chapter 2). We have identified six HSF-binding sites within the bodies of these three genes. Each site is located within an intron and overlaps with a binding sites for the ecdysone receptor complex (large black boxes) (Gauhar et al. 2009). Five out the six sites are also located within 5 kb of the transcription start site of one or more isoforms and the sixth site is still only 10 kb away from a transcription start site.
isoforms. In addition, all five HSF binding sites overlap with binding sites identified in Kc167 cells for the ecdysone receptor complex (Gauhar et al. 2009).

Given that HSF binding within the body of these genes is coincident with their repression it is possible that HSF may have a direct role in the repression of these genes. It is not clear, however, from the location of the HSF binding sites if HSF may be interfering with transcription initiation from alternate promoters in the vicinity of its binding site or if it may be interfering with transcription from upstream promoters by an obstruction to transcription elongation. The mammalian HSF homolog, HSF1, is known to cause repression of prointerleukin 1β and Tumor Necrosis Factor α (Cahill et al. 1996; Xie 2002), although in the present case it is not clear if the mode of action would be the same.

3.5 Conclusions

We have identified 35 primary 20E-response genes that are induced by 20E in the absence of protein synthesis in Drosophila Kc167 cells, one of the most widely used cell lines employed to study the ecdysone response. The primary 20E responsive genes were enriched for Gene Ontology (GO) terms such as induction of programmed cell death by hormones, catalytic activity, hydrolase activity, and cytochrome P450. We have also identified 125 secondary 20E-response genes (48 induced and 77 repressed). GO terms that are enriched in the secondary-response genes include an increase in transcripts for genes involved in mitochondrial respiration and a decrease in transcripts for genes involved in glycolysis suggesting that 20E induces shifts in cellular metabolism. Some primary- and secondary-response genes include genes that code for proteins involved in cell movement and organization and/or are cytoskeletal associated which is consistent with the cell movement and morphogenetic changes that occur during 20E-dependent developmental changes.

Comparison of the genome-wide transcriptional response to 20E to its plant derived structural analog ponasterone A (PoA) revealed a large difference in the transcriptional targets of these molecules. While these two compounds are structurally very similar, many more genes related to various aspects of development appear to be significantly induced by 20E than by PoA. More specifically, the most strongly enriched functional terms are associated with genes down-regulated (i.e. not induced) by PoA relative to 0.5 µM 20E. The genes not induced by PoA
include GO terms such as developmental process, cell differentiation, gamete generation, anatomical structure development, organelle organization and biogenesis, neuron development, and induction of programmed cell death. The apparent lack of activation of developmental genes by PoA suggests that PoA induced death in non-adapted insects may be due to more than just the simple induction of premature molting via the canonical ecdysone hierarchy.

We also compared the 20E response in Kc cells to that of a natural 20E target tissue where the function of 20E has been well described, the salivary glands of wandering 3rd instar larvae, and found little overlap in 20E-responsive genes. Genes induced by 20E in salivary glands but not in Kc167 cells were enriched in GO biological processes such as “salivary gland histolysis”, “programmed cell death” and various “developmental” processes as well as “cellularization” and “transporter activity”.

To help identify a potential mechanism to explain the difference in the transcriptional responses in Kc cells and salivary glands, we analyzed the 20E-induced transcription of the various EcR isoforms, a known 20E-inducible gene and a component of the nuclear receptor complex that binds 20E. We did find differences in the induction of EcR isoforms EcR-RA, ER-RC, and EcR-RD/E between Kc cells and salivary glands in response to 20E. This suggests that the relative amount of various EcR isoforms present in a cell is different in different cell types and possibly contributes to the transcriptional response a given tissue has to 20E.

Finally, we found that HS effectively blocked transcription of 20E-response genes and that the extent of repression is correlated to the duration of stress. Because we had previously observed HSF binding to loci known to contain primary-response genes (Westwood et al. 1991), we speculated that HSF may have a role in this repression and we sought to identify where HSF was binding relative to these genes. We found that HSF occupied six sites located in the three major primary-response genes, Eip75B, Eip74EF, and br and that these sites appeared to show strong bias for occurring in the 3’ end of introns. This observation lends support for a role in which HSF is a gene specific repressor of 20E-response genes much like its mammalian counterpart is for TNF-α (Singh et al. 2002), however, more work remains to be done before this can be determined conclusively.
Chapter 4
General Discussion

It has been almost 50 years since the initial observation of the effects of HS on gene transcription in polytene chromosomes. Since that time the HSR has served as a powerful model system for the study of numerous facets of gene transcription owing largely to the strong stimulatory effect that HSF binding at promoters of HSP genes has on their transcription. It is no wonder that the first report of *Drosophila* HSF binding to a vast number of sites on polytene chromosomes that did not appear to be transcriptionally active (Westwood et al. 1991) came as a surprise and sparked a search for possible explanations.

The work presented here set out to do just that; first by identifying HSF binding sites on a genome-wide scale at higher resolution than previously possible and correlating these binding sites to the transcriptional activity of nearby genes and binding sites for other transcription factors; and second by characterizing the transcriptional response to 20E, the hormone responsible for the induction of a set of genes that map to loci bound by HSF but do seem to be induced during HS, and then determining the effects that HS has on the transcription of 20E-response genes.

### 4.1 HSF binds to many sites not associated with HS-induced gene transcription

Original observations of HSF binding to over 200 loci on polytene chromosomes roughly 20 years ago suggested that HSF binds to many sites that are not transcriptionally active during the HSR (Westwood et al. 1991). With advancements in methods available for both the identification of transcription factor binding sites as well as for the monitoring of gene transcription in recent years, it is now possible to gain a more accurate picture of both the binding profile of HSF and the transcriptional response to HS to determine which and how many HSF bound sites are associated with the transcriptional activation of genes during HS. Indeed the findings presented in chapter 2 identify 434 regions bound by HSF in the *Drosophila* genome at around 250 bp resolution and 211 genes from Kc cells and 237 genes from 3rd instar larvae that show altered transcription during HS. Interestingly, these results were largely in support of the original suggestion that the vast majority of HSF bound sites are not
transcriptionally active during HS (Westwood et al. 1991) (Figure 2-10). Of the HS-inducible genes identified in Chapter 2, only 11 genes from Kc cells and 46 genes from 3rd instar larvae are located within 1250 bp of an HSF binding site. Of these genes, nine were in common between Kc cells and larvae and with the exception of one gene that has been withdrawn in more recent releases of the genome (CG32636), all have been previously shown to be HS-inducible and include the major HSP family members, *Hsp22, Hsp26, Hsp27, Hsp67Bb, Hsp68,* and *Hsp70Ab*; an HSP70 co-factor, *DnaJ-1*; and putative HSP70 family co-chaperone, *stv* (Table 2-3). A similar study examining the relationship between a subset of human HSF1 binding sites and HS-inducible genes identified 24 HS-inducible genes that showed no evidence of HSF1 binding in their promoter regions as well as 48 genes with HSF binding in their promoters that were not HS-induced (Trinklein et al. 2004) indicating that like other aspects of the HSR, these features are likely well conserved.

### 4.2 Possible explanations of HSF binding to sites that are apparently not transcriptionally active during HS

What it is the reason for such little association between HSF binding and HS-induced activation? One possibility could be my definition of association. Using the distance between HSF binding sites and the TSS of the major HSP genes as a standard, the closest HSF binding site is never more than 300 bp upstream (Nover 1991). Thus, my expectation was that all HS-induced genes would exhibit this same arrangement with respect to HSF binding site and gene TSS. However, it is possible that such a close association of HSF binding site and HSP TSS is a unique feature of these distinctively highly expressed genes. To account for this possibility, I expanded my definition to consider a gene associated with HSF if it was located within 1250 bp of an HSF binding site. However, perhaps the “true” norm for the genes that show lower levels of HS-induced transcription is for the HSF binding site to be located still further away. This would not be surprising since some enhancers have been identified as far as 100 kb away from their targets (Palstra 2009). The ability of enhancers to work at great distances is likely due to the three dimensional arrangement of chromatin and thus HSF may also associate with its target genes through the looping of DNA *in vivo*. A newly emerging area of study related to examining the role of chromosomal arrangement in gene regulation utilizes techniques such as 3C (Chromatin Conformation Capture) and RNA TRAP (RNA Tagging and Recovery of
Associated Proteins) to map regions of the chromatin that are in close proximity regardless of
their linear arrangement (Palstra 2009). It would be interesting to apply such techniques to map
HSF binding sites in three-dimensional space to see if any of the genes that appear to be
transcriptionally regulated during HS are actually in close proximity to HSF.

An indirect test of the necessity of HSF binding for the induction of non-classical HS genes is to
examine the HS-induced transcriptional response in the absence of HSF. If HSF binding is in
some way required for the induction of the non-classical HS genes, we would expect them not to
be induced in the absence of HSF. In chapter 2 I examined the transcriptional profile of hsf4
larvae which have a mutation in the DNA binding domain of HSF that prevents high specificity
DNA binding at HS temperatures (Jedlicka et al. 1997) and found that in general, HSF DNA
binding was required for HS-regulated transcription (Additional File A-4). Of course it is
possible that HSF directly induces the transcription of a gene or genes that in turn is/are
responsible for the regulation of the non-classical HS genes. In fact of the genes associated with
HSF binding sites, three that show mild to moderate induction (2.3 - 7.8 fold) in response to HS
in larvae, have documented roles in transcription regulation: mbf1, Sir2, and Taf7 (Table 2-3).
However, this cannot entirely explain all of the transcription changes seen as a result of HS
since no significant difference in HS-induced gene transcription was seen when the HS
experiments were repeated in the presence of the protein synthesis inhibitor cycloheximide (see
Chapter 2).

Another possibility is that the sites associated with HSF binding are transcriptionally induced by
a small amount, but even with the advanced technology utilized in the work described here, we
still are unable to easily identify these transcripts. One difficulty of studying rapid
transcriptional changes is the propensity for all but the most highly induced transcripts to evade
detection due to dilution by steady-state mRNA. During HS, this problem is potentially made
worse as HS is known to stabilize hsp70 mRNA (Petersen and Lindquist 1988) and so it is
possible that it may stabilize other transcripts as well. Measuring transcript abundance in the
hsf4 mutant after HS gives some indication of this as transcripts for genes such as Hsp83, a
known target of HSF, are found to increase during HS relative to non-shock conditions even in
the hsf4 mutant (Neal et al. 2006) (Additional File A-4). One potential solution is to inhibit
nascent transcription during HS with either actinomycin D or alpha-amanitin so transcripts
present prior to HS and are stabilized (or de-stabilized by HS) can be identified. Another
solution is to map transcriptionally active RNPII. In heat shocked polytene chromosomes, mapping hyperphosphorylated RNPII by immunofluorescence has revealed RNPII accumulation at approximately 50 sites, many of which co-localize with HSF, after a 20 minute HS (Paraiso 2002). Very recently, a higher resolution map of the distribution of paused RNPII during HS was generated by ChIP-seq in Drosophila S2 cells (Teves and Henikoff 2011). This study, like the one by Paraiso, showed a general re-location of RNPII away from sites of non-HSP genes, however, provided no indication about the association of RNPII with any specific non-HSP genes. It would be interesting to take a closer look at the individual genes associated with RNPII after HS and compare them against those genes identified in chapter 2 as bound by HSF and/or induced during HS.

Yet another possibility is the binding of HSF in some cases leads to the transcription of RNAs that would escape detection by the arrays utilized in Chapter 2. These would include ncRNAs and it is reasonable to suspect that HSF may be responsible for the regulation of non-protein coding genes as the most strongly enriched HSF bound site identified in chapter 2 is at 93D, a well known site of one the most strongly induced genes, a non-protein coding gene called hsrω. In addition to hsrω several other examples of HS-induced ncRNAs already exist such as Drosophila alpha-gamma element, and mammalian B2, and Alu (see chapter 1 for more details). Given the observed HS-induced re-distribution of RNPII to sites that do not map to any currently known HS-induced protein-coding gene, the possibility of HSF-mediated ncRNA synthesis cannot be ignored. To begin to address this possibility, I prepared cDNA libraries from HS and non-HS total RNA from Kc cells for next generation sequencing which provides a more complete picture of the transcriptome, including non-protein coding genes, than can be obtained from microarray analysis alone. At the time of writing this thesis, the results of next generation sequencing were still in a preliminary phase, thus no firm conclusion could be drawn. Nonetheless, preliminary analysis of two biological replicates of each treatment reveals 395 unique features that show a fold change of 2 or more and a p-value of 0.1 or less in response to a 30 min HS. The highest changing features were HSP genes, which is as expected, providing good validation of this approach. Table 4-1 A, B lists all unique features with highly significant p-values (≤ 0.01). A summary of the number of each type of feature identified in this analysis is given in Table 4-1C. Although not appearing in Table 4-1A because the stringent p-value cut off
Table 4-1 Preliminary RNA-seq results for HS RNA.

All unique features with a p-value < 0.01 and an average fold change ≥ 2 (A) or ≤ -2 (B). (C) Breakdown of the different features identified by RNA-seq as transcriptionally responsive to HS.

(A) | Name | Feature Type | HS vs RT Fold Change (biological replicate 1) | HS vs RT Fold Change (biological replicate 2) | Average Fold change (average across biological replicates) | p-value |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsp70Aa</td>
<td>CDS</td>
<td>39.0</td>
<td>54.1</td>
<td>45.9</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Hsp70Ab</td>
<td>CDS</td>
<td>39.0</td>
<td>54.1</td>
<td>45.9</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Hsp70Bb</td>
<td>CDS</td>
<td>38.7</td>
<td>31.6</td>
<td>35.0</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Hsp70Bbc</td>
<td>CDS</td>
<td>38.7</td>
<td>31.6</td>
<td>35.0</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Hsp70Bc</td>
<td>CDS</td>
<td>36.5</td>
<td>31.4</td>
<td>33.9</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Hsp70Ba</td>
<td>CDS</td>
<td>31.3</td>
<td>34.4</td>
<td>32.8</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>CG5018</td>
<td>mRNA</td>
<td>7.9</td>
<td>8.3</td>
<td>8.1</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>CG31715</td>
<td>CDS</td>
<td>6.9</td>
<td>7.5</td>
<td>7.2</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>CG6808</td>
<td>mRNA</td>
<td>6.5</td>
<td>7.9</td>
<td>7.2</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>CG6597</td>
<td>gene</td>
<td>6.2</td>
<td>8.1</td>
<td>7.1</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Hsp26</td>
<td>CDS</td>
<td>7.2</td>
<td>5.7</td>
<td>6.4</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>CG30493</td>
<td>mRNA</td>
<td>5.5</td>
<td>6.8</td>
<td>6.1</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>CG5004</td>
<td>mRNA</td>
<td>5.3</td>
<td>6.0</td>
<td>5.7</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Sec22</td>
<td>mRNA</td>
<td>4.7</td>
<td>3.6</td>
<td>4.1</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>CG15093</td>
<td>mRNA</td>
<td>3.7</td>
<td>3.8</td>
<td>3.7</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Aats-ala-m</td>
<td>mRNA</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>BHD</td>
<td>mRNA</td>
<td>2.2</td>
<td>2.3</td>
<td>2.2</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>CG31694</td>
<td>mRNA</td>
<td>2.3</td>
<td>2.0</td>
<td>2.2</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>CG7246</td>
<td>mRNA</td>
<td>2.0</td>
<td>2.3</td>
<td>2.1</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>CG18273</td>
<td>gene</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

(B) | Name | Feature Type | HS vs RT Fold Change (biological replicate 1) | HS vs RT Fold Change (biological replicate 2) | Average Fold change (average across biological replicates) | p-value |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ligatin</td>
<td>CDS</td>
<td>-2.6</td>
<td>-2.5</td>
<td>-2.5</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>KLHL18</td>
<td>CDS</td>
<td>-2.3</td>
<td>-2.6</td>
<td>-2.6</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>mm</td>
<td>CDS</td>
<td>-3.1</td>
<td>-2.5</td>
<td>-2.8</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>Noa36</td>
<td>CDS</td>
<td>-3.0</td>
<td>-2.7</td>
<td>-2.8</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>HDAC4</td>
<td>CDS</td>
<td>-2.8</td>
<td>-3.3</td>
<td>-3.0</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>CG4250</td>
<td>CDS</td>
<td>-2.9</td>
<td>-3.5</td>
<td>-3.2</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Cbp80</td>
<td>CDS</td>
<td>-4.8</td>
<td>-4.8</td>
<td>-4.8</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>bor</td>
<td>gene</td>
<td>-6.2</td>
<td>-4.3</td>
<td>-5.2</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>CG4747</td>
<td>CDS</td>
<td>-6.2</td>
<td>-4.5</td>
<td>-5.3</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>tweek</td>
<td>CDS</td>
<td>-7.2</td>
<td>-7.5</td>
<td>-7.4</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Cad87A</td>
<td>mRNA</td>
<td>-11.5</td>
<td>-7.6</td>
<td>-9.4</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>ref</td>
<td>CDS</td>
<td>-19.7</td>
<td>-25.7</td>
<td>-22.5</td>
<td>0.009</td>
<td></td>
</tr>
</tbody>
</table>

(C) | Feature Type | Induced | Repressed | TOTALS |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>22</td>
<td>133</td>
<td>155</td>
</tr>
<tr>
<td>mRNA</td>
<td>79</td>
<td>65</td>
<td>144</td>
</tr>
<tr>
<td>gene</td>
<td>39</td>
<td>48</td>
<td>87</td>
</tr>
<tr>
<td>mobile_element</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>ncRNA</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>rRNA</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TOTALS</td>
<td>148</td>
<td>246</td>
<td>394</td>
</tr>
</tbody>
</table>
was not met, HS-induced transcription of both *hsrω* and *alpha-gamma element* were detected by this method providing further support for the utility of this approach. *Alpha-gamma element* showed over 100 fold induction by HS in one sample while *hsrω* exhibited a change comparable to that of *Hsp26* (4.8 fold induced). Because in this preliminary stage of analysis, only those features that have been previously annotated were detected, it is possible that additional novel ncRNAs may have been missed. Interestingly, there is evidence of small peaks of transcription occurring during HS near each of the HSF binding sites located in the *Eip75* introns Figure 4-1. Other than a single nearby snoRNA, there has been no official annotation of an ncRNA in this area. Given that our current understanding of ncRNA is still in a primary stage, not to mention that a reliable method for ncRNA prediction has yet to be established, it would not be surprising if many ncRNAs eluded detection in this first pass analysis simply because they have yet to be annotated. Thus the next step in the analysis of this data would involve mining the data for all sites of transcription with a signal above a certain threshold and then testing those chosen for further investigation by qRT-PCR.

Of course another possibility that cannot be ignored is that HSF binds to sites not associated with HSP genes merely because they contain HSE like motifs, are accessible, and the binding of HSF to these sites has no deleterious effects. HSF is not unique in this respect as other transcription factors such as estrogen receptor and p53 have also been shown to bind gene promoters with no apparent transcription effect (Wei et al. 2006; Kininis et al. 2007).

### 4.3 Possible roles for HSF apart from being a transcriptional activator during the HSR

With HSF binding to so many genes that are not transcriptionally induced during the HS, could HSF be functioning in a yet uncharacterized role? Could HSF be a regulator of developmental genes? Or even a transcriptional repressor?

In Chapter 2 the location of HSF binding sites relative to gene regions was examined revealing that more than half of all HSF bound sites were located within transcribed regions of the genome (Figure 2-6). This was surprising as the typical configuration with respect to HSP genes is for HSF to bind in the upstream region (Nover 1991). As discussed in the previous section, it
Figure 4-1 Preliminary evidence of possible HS-induced transcription near sites bound by HSF in the Eip75 bound gene.

This image, generated using the UCSC Genome Browser, illustrates a portion of chromosome 3L (according to release 4.2 of the Drosophila genome) as indicated at the top of the figure. Genes are depicted as dark blue boxes with the thick and thin parts representing exons and introns respectively. Dark blue arrows within the gene indicate the direction of transcription. Each light blue bar represents a region detected by RNA-seq and its height represents the number of RPKM (reads per kilobase of exon model per million mapped reads) normalized counts and serves as a rough indication of enrichment. HSF bound segments are represented by large black boxes and the estimated center point of each segment is shown by a black line above the segment (data from experiments presented in Chapter 2; see Chapter 3 Figure 3-6 for more information about these sites). One HSF binding site directly overlaps with a region of transcription while two of the other three HSF binding sites in this gene are relatively close (2-3 kb) to other regions of transcription. Interestingly, no currently annotated ncRNA maps to these regions. Although the direction of transcription is not known, the two transcribed regions upstream of Eip75-RC are located upstream of the nearest HSF binding site suggested that they may be transcribed in the direction opposite to the Eip75 and therefore might represent anti-sense RNA.
is possible that this arrangement might be something specific for the strongly induced HSP genes. Further analysis of the functions enriched among genes with HSF binding sites specifically within their introns revealed a strong tendency for these genes to be involved in development and/or reproduction (Figure 2-8). Because HSF has been shown previously to be required for oogenesis and early larval development in *Drosophila* and HSF1 and HSF2 in mice have documented roles in brain development and fertility, it is not difficult to imagine that HSF might have a role in the regulation of these non-HSP genes under conditions not examined in the current work, for example, during development. Efforts are currently underway in our lab to test this possibility by examining the genome-wide transcriptional differences between late embryos and first instar larvae with wild type HSF and larvae with an amorphic mutation in the HSF gene (*hsf*) (i.e. behaves as a functional HSF null at non-permissive temperature). If the requirement for HSF during these stages in development is linked to the regulation of the genes with intronic HSF binding sites, the expectation is that the expression levels of these genes will differ between wild type and mutant larvae around this point.

If this is the case, it still does not explain the reason for HSF occupancy of these sites during the HSR. Is it possible that HSF occupies these sites simply because they are accessible? And if so, what other factors are involved in preventing HSF from activating these genes in this context?

As introduced in Chapter 1, one interesting yet poorly understood facet of the HSR is its ability to repress the transcription of nearly all genes that are not directly related to the HSR including genes involved in other transcriptional programs that are activated concurrently (Ashburner 1970; Westwood et al. 1991; Sun et al. 2005; Shi et al. 2006). Thus, it could be possible that the same mechanisms in place for the general repression of non-HS genes during HS also function to prevent HSF related induction of other genes it might control.

Interestingly, the repression of certain genes, such as the LPS-induced cytokine genes TNF-α, IL-6 and IL-β, have been directly linked to the activity of mammalian HSF1 (Cahill et al. 1996; Xiao et al. 1999; Singh et al. 2002; Xie 2002; Takii et al. 2010). For example, repression of TNF-α is conferred through high affinity binding of HSF1 to a site in its 5’ UTR where binding of HSF1 is thought to affect RNPII processivity (Singh et al. 2002). These observations raise an intriguing possibility that the binding of *Drosophila* HSF to developmental and reproductive
genes might function in the direct repression of these genes during HS. The results presented in Chapter 3 provide a basis for examining this possibility. Previously it had been shown that developmental puffs on polytene chromosomes regress during HS (Ashburner 1970; Westwood et al. 1991), however, the results of Chapter 3 are the first to show a direct link between HS and the repression of ecdysone-induced gene transcription (Figure 3-5). Furthermore, HSF was found to occupy multiple sites within all three major primary-response ecdysone genes, \textit{Eip74}, \textit{Eip75}, and \textit{br}, during the HSR placing it in contact with these genes during the period in which they are repressed (Figure 3-6).

If the binding of HSF to these ecdysone-inducible genes is in fact responsible for their repression, it would be expected that other inducers of HSF binding would also have the same effect. Likewise, HS in the absence of HSF binding should not be sufficient to promote the repression of these genes. The proline analog azetidine has been shown to have the same effect as HS with respect to HSF binding and repression of non-HS genes in immunostained polytene chromosomes of both wild type and \textit{hsf} \textsuperscript{4} mutant larvae (Stevens 1999) and in a genome-wide examination of the transcriptional response of Kc cells I have also found that it is able to prevent the transcription of primary-response ecdysone genes (data not shown). Additionally, I have found that the proteasome inhibitor MG132, which induces expression of HSPs (Kim et al. 2011), is also capable of repressing \textit{Eip75} to a similar extent as HS (data not shown). Furthermore, the necessity for HSF binding in repression of these genes was indicated by the lack of repression during HS in \textit{hsf} \textsuperscript{4} polytene chromosomes (Stevens 1999).

If HSF is in fact responsible for the HS-induced repression of these genes, what might be the mechanism? Observations of HS causing a general reduction in the amount of RNPII associated with non-HS genes coupled with the preferential recruitment of RNPII to sites of major HSPs (Jamrich et al. 1977; Greenleaf et al. 1978; Gilmour and Lis 1985; Stevens 1999; Paraiso 2002; Teves and Henikoff 2011) suggests a model in which HSF, by virtue of being such a strong transcriptional activator, is capable of sequestering all available RNPII to sites of major HSPs leaving little to none available for the transcription of non-HSP genes (Figure 4-2A). To test this model, the ability of HSF to block ecdysone-induced gene activation by binding to DNA without activating transcription could be examined, for example by using an inducer such as sodium salicylate which induces HSF binding but not \textit{Hsp70} transcription (Winegarden et al. 1996) or by studying a mutant form of HSF that lacks the transactivation domain.
Figure 4-2 Possible models of genes repression by HSF.
Despite the evidence in favor of the above model, it provides no explanation for the association of HSF with the repressed ecdysone-inducible genes. Thus a second model where HSF binds to and affects the transcription of a subset of genes is presented in Figure 4-2B. In this model, HSF might affect the processivity of RNPII in a manner similar to that for the lac repressor which binds downstream of the TSS and competes with the T7 RNA polymerase for interaction with the DNA and subsequently affects both initiation rate and enzyme processivity (Lopez et al. 1998). In the case of HSF, this competition might occur several kb downstream from the transcription initiation site if the goal is to block already elongating RNPII or just upstream of the TSS if the goal is to interfere with subsequent transcription initiation. Based solely on the location of HSF binding sites within the three major ecdysone-inducible genes (Figure 3-6) it is not possible to determine which of these mechanisms (if not both) might be at work. HSF binding sites within all three genes are located within introns of one or more isoforms but also show a strong bias for being positioned near the 5’ end of the TSS of one or more isoforms and also within the immediate vicinity of the binding site for the ecdysone receptor complex. To try to differentiate between these two possibilities, I measured levels of Eip75 transcript upstream of each HSF binding site present in ecdysone-treated HS Kc cells (Figure 4-3). The expectation was that if HSF was interfering only with transcription initiation, no transcripts from those isoforms that were associated with an HSF binding site would be detectable during HS (i.e. no transcript for Eip75-RA, Eip75-RC, or Eip75-RD should be detectable, but a transcript from Eip75-RB would be since it is ecdysone-inducible under the conditions tested and its promoter is not bound by HSF). If HSF was instead interfering only with elongation, a partial transcript representing the region upstream of the HSF binding site would be detectable during HS (i.e. the region of Eip75-RB 5’ of the HSF binding site). A caveat to this approach, however, is that such partial transcripts lacking poly A tails would be inherently unstable, and although this would be to the benefit of the repression mechanism, it makes them difficult to detect. Thus, the results of this experiment suggest that interference with initiation is probably not the sole mechanism, and at this time we are not able provide conclusive evidence of the inhibition of transcription elongation by HSF. It is still possible to test the ability of the HSF to interfere with RNPII processivity without knowing if RNPII initiation or elongation is affected by examining the effect of deleting or mutating HSF binding sites within these genes has on their repression during HS.
Figure 4-3 HSF binds within introns (and possibly promoters) of ecdysone-inducible genes during HS, coincident with their repression.

Using a cDNA microarray we identified 27 genes induced by 1.5 fold or more after a 2 hr treatment with 0.5 μM ecdysone (20E) in the cell line Kc167 in the absence of heat stress, including major primary-response genes, Eip75B and br. (A) Quantitative RT-PCR data showing the level of induction of each isoform of Eip75 after a 2 hr 20E treatment in the absence (blue bars) and presence (red bars) of HS. I found that exposing cells to HS temperatures (36.5°C) for the duration of the 20E treatment (2hr) prevents the expression of all isoforms of Eip75B. Using a cDNA based microarray, we have found that br and all secondary 20E-inducible genes we identified are also repressed under HS conditions (Figure 3-5). (B) Location analysis (ChIP on chip) indicates that HSF binds within intron sequences of three primary ecdysone-response genes including Eip75B shown here (for the other two see Figure 3-6). Orange bars above the gene sequence indicate regions where HSF bound probes from genome tiling arrays were identified. Solid blue bars represent coding sequence and connecting lines represent introns. The direction of transcription is indicated by the arrow heads. Like Figure 4-1, this image was generated using UCSC Genome Browser.
A third model of HSF regulated gene repression involves the recruitment of co-repressor proteins such as histone deacetylases and nucleosome remodeling complexes like SWI/SNF both of which have been implicated in HSF-mediated repression in conjunction with mammalian HSF1 (Fritah et al. 2009) and yeast HSF (Shivaswamy and Iyer 2008) respectively (Figure 4-2C). It would be interesting to see if in Drosophila, the repression of any HSF bound genes during HS is affected in mutants of different co-repressor proteins.

A fourth model is that there is heat-induced transcription of non-coding RNAs in the early ecdysone genes and these act to interfere with the transcription of and/or cause the degradation of Eip75 transcripts (Figure 4-2D). Recently, many diverse mechanisms of transcriptional regulation involving ncRNAs have come to light. For example, ncRNAs have been shown to affect chromatin remodeling, oligomeric state of activators, the activity of RNAPII, as well as serving as scaffolds for the formation of RNA-protein complexes and yield small interfering RNAs that facilitate transcript degradation by RNAi pathways (Goodrich and Kugel 2009; Wilusz et al. 2009). In Drosophila, TAC1 stimulated transcription of bxd ncRNAs located upstream of the Ubx gene has been shown to interfere with and prevent the transcription from the Ubx promoter (Petruk et al. 2006). Given that the putative ncRNA transcripts identified by RNA-seq reside upstream of the transcription start site of one or more isoforms of Eip75 inhibition by interference is an attractive model for future consideration. A first step towards testing this model involves identifying ncRNA transcripts produced during the HSR and RNA-seq provides the most effective way to do this de novo and without biases. The sequence of HS-induced ncRNA transcripts could then be examined computationally to determine if they are complementary to any coding regions of repressed transcripts that would give an indication as to the possible mechanisms that may be at work. If an RNAi repression mechanism is suspected, the ability of cells to repress transcription of a gene with complementary coding regions could be tested by examining the effect of transfecting cells with the ncRNA transcripts under conditions when the repressed transcripts would normally be produced such as with 20E treatment in the case of Eip75. We could also examine what effects that mutation(s) of genes necessary for RNAi have on the repression of these genes during heat shock. Other modes of action of ncRNA based repression such as the interference with downstream gene transcription could be tested by first determining if transcription of the ncRNA occurs in the vicinity of the promoter of the repressed genes. Where this appears to be the case, deleting or mutating the
ncRNA promoter but otherwise leaving the repressed gene intact could test for the possibility of the transcriptional interference by the ncRNA. Of course there is no requirement for any of these models to be mutually exclusive and so it would also be possible for some combination of all three to be at work and so would be necessary to test all the models.

A natural question that follows discussion of HSF function as a gene specific repressor is why certain genes would be specifically targeted for direct repression by HSF? In the case of the three primary-response ecdysone-inducible genes examined in this context one could imagine at least a couple of reasons. First, these genes are extraordinarily long taking several hours to transcribe (Thummel 1990) thus making them a liability to the rapid and efficient activation of the HSR if they cannot be rapidly inactivated. Second, these genes occupy principal positions in the transcription cascade controlled by ecdysone, making their inactivation an efficient way of repressing genes downstream in the cascade and thus suspend development until environmental conditions became more favourable.

4.4 Perspectives

The work presented here has expanded our understanding of the genes involved in two highly studied systems; the HSR, which not only serves as a model for the study of transcription regulation and but is also highly studied for its role in human disease; and the ecdysone response, which has long been studied as a model for steroid hormone action and nuclear receptor function. In addition, many insights have been gained into the binding site preferences of the master regulator of the HSR, HSF. Although the function of HSF binding to sites not obviously associated with HSPs still remains elusive, this work has provided many avenues for further exploration including the possible HS-induced transcription of previously unidentified ncRNAs, a possible role for HSF in the regulation of genes during development, as well as a possible role for HSF in the negative regulation of genes during HS.

With the recent identification of ncRNAs that can interfere with RNPII transcription in *Drosophila* and mammals, the existence of ncRNAs with similar roles during the HSR is an enticing model. Identification of such ncRNA in *Drosophila* will not only prove useful for understanding the HSR in *Drosophila* but consequently, because the HS system is such a
valuable tool for the study of the regulation of transcription in general, it should also prove useful for understanding the general mechanisms of transcription.

Likewise, with several roles for members of the mammalian HSF family in the regulation of development and reproduction emerging, it is not difficult to imagine that the more primitive fly HSF would be likely to encompass several of these roles in itself. Understanding the role of HSF in Drosophila development under non-HS conditions should provide a further model for the function of HSFs in mammalian development.

Finally, the possibility of HSF functioning as a transcriptional repressor in the specific repression of certain genes remains an exciting possibility that should not be ignored. If HSF can be shown to function both as an activator and repressor of transcription, it will serve as a powerful model system for understanding the genomic context that influences the ability of factors to act either positively or negatively on transcription.
References


HSF1 and confers protection against apoptosis and cellular stress." EMBO J 22(20): 5446-5458.


Appendix A


Additional File A-4. Complete list of genes exhibiting a 2 fold or greater change in response to heat shock in either Kc cells, wild type (dp), and Hsf mutant (Hsf') larvae. The position of the nearest HSF binding site to each gene is also given. (http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0015934.s003)
Appendix B

Additional File B-1. 20E-responsive genes in Kc167 cells after 2 and 4 hours of exposure to 0.5 µM 20E. Of the 27 genes induced by 0.5 µM 20E after 2 hours, one third are no longer up-regulated after 4 hours (group A) while the rest show similar levels of induction at both time points (group B). Many genes (51) are only significantly up-regulated after 4 hours of exposure to the hormone (group C). Likewise, genes that are repressed by 0.5 µM 20E are only detectable after 4 hours of treatment (group D). q-values determined by SAM represent the lowest false discover rate at which that gene is considered significant (see the Table 3-1 legend for details of q-values). A missing fold change value means that gene was either not represented by at least 2/3 spots and/or did not exhibit a coefficient of variation of at least 1 for that treatment across replicate arrays. (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s1.xls)

Additional File B-2. Extended version of Table 3-1. See Table 3-1 legend for details. (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s2.xlsx)

Additional File B-3. qRT-PCR confirmation of microarray data. Relative induction of selected genes identified by microarray analysis in Kc cells and salivary glands by either 0.5 µM 20E or 0.0625 µM PoA was confirmed by qRT-PCR relative to untreated Kc cells or salivary glands, respectively. (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s3.doc)

Additional File B-4. Genes identified as part of the transcriptional response to either 20E or PoA. Genes that exhibit a 1.5 fold change or greater to 0.0625 µM PoA in Kc167 cells or to either 10 µM 20E (20E high) or 0.5 µM 20E (20E) in Kc cells or to 0.5 µM 20E in salivary glands (glands) are listed. q-values determined by SAM represent the significance of fold change reported (see the Table 1 legend for details of q-values). (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s4.xls)

Additional File B-5. Genes that respond to 20E in salivary glands and in Drosophila organ culture. Genes identified as primary 20E-response genes by Beckstead and colleagues (Beckstead et al. 2005) are highlighted in grey. (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s5.doc)

Additional File B-6. Salivary gland-specific 20E responsive genes. These genes were identified as part of the response to 20E in salivary glands but not in whole organ culture (Beckstead et al 2005). (http://www.biomedcentral.com/content/supplementary/1471-2164-12-475-s6.doc)
Copyright Acknowledgement

The contents of chapter 2 and a modified version of section 3.4.6 of chapter 3 are published in PLoS ONE (see Gonsalves et al. 2011). Permission to reproduce this material here is provided under the Creative Commons Attribution License. The following people were all contributing authors: Sarah E. Gonsalves, Alan M. Moses, Zak Razak, Francios Robert, and J. Timothy Westwood and their contributions are as follows: Conceived and designed the experiments: SEG AMM JTW. Performed the experiments: SEG. Analyzed the data: SEG ZR. Contributed reagents/ materials/analysis tools: SEG JTW. Wrote the paper: SEG JTW. Provided advice on experimental design: AMM FR.

The contents of chapter 3 are published in BMC Genomics (see Gonsalves et al. 2011) with the exception of section 3.4.6, which was modified from the original version published in PLoS ONE (see above). Permission to reproduce this material here is provided under the Creative Commons Attribution License. The following people were all contributing authors to the BMC Genomics paper: Sarah E. Gonsalves, Scott J. Neal, Amy S. Kehoe, and J. Timothy Westwood and their contributions were as follows: SEG participated in the design of the study, carried out the hormone treatments, RNA isolations, microarray experiments, data analysis including analysis of functional classes, and qRTPCR experiments, and drafted the manuscript. SJN and ASK participated in the design of the study and participated in pilot experiments. JTW conceived of the study, participated in its design and helped to draft the manuscript.