UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Transcriptional stochasticity and the function of grainy head transcription factors in animals and fungi

Permalink

https://escholarship.org/uc/item/3jm2n4dr

Author

Paré, Adam Christopher

Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Transcriptional Stochasticity

and

The Function of Grainy head Transcription Factors in Animals and Fungi

A dissertation submitted in partial satisfaction of the requirements for
the degree of Doctor of Philosophy

in

Biology

by

Adam Christopher Paré

Committee in charge:

> Professor William McGinnis, Chair
> Professor Christopher Glass
> Professor James Kadonaga
> Professor Amy Pasquinelli
> Professor James Wilhelm

2011

The Dissertation of Adam Christopher Paré is approved, and it is acceptable

in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2011

iii

DEDICATION

To my parents,
who made it possible,

and to my wife,
who makes it bearable.

EPIGRAPH

Nothing in biology makes sense
except in the light of evolution.

Theodosius Dobzhansky

Science is the belief in the ignorance of experts.

Richard Feynman

This is how you do it:
you sit down at the keyboard
and you put one word after another until it's done.
It's that easy, and that hard.

Neil Gaiman

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

LIST OF TABLES

Chapter I

Chapter III

PREFACE

In much the same way that Evolution is the great underlying theory of biology, the field of Developmental Biology has become a common ground for many of the diverse sub-disciplines within "the study of life". Rightly so, as it is here that the probabilistic dance of molecules within cells is channeled into the reproducible (often exquisitely so) morphology and behavior of complex organisms. I believe it is from this vantage point that a biologist can most clearly see how their research fits into and expands our existing body of knowledge, and from here that one can most easily approach a true understanding of life on Earth.

The processes that govern the living world, and in particular the lives of multi-cellular organisms, can be thought of as occurring on three distinct but overlapping levels. On the smallest scale, DNA, RNA, protein, and other small molecules interact within cells to carry out the basic functions of life, as well as to steer cells towards specialized fates. One level abstracted from that, at the scale of whole organisms, gene networks are deployed within populations of cells in response to intrinsic and extrinsic cues, ultimately culminating in changes in cell-fate or organismal behavior; single cells develop into complex multi-cellular organisms, which live out their lives within a challenging, and often harsh, environment. Finally, at the highest level, we see natural selection acting on species over many generations, constantly shaping them towards ever more specialized (if not more complex) forms. It is at this last stage that we most clearly see the hand of evolution, although its fingerprints are present everywhere in biology, if one cares to look.

The three chapters of this dissertation, I believe, correspond well to the three levels of biological organization I have just described, and it is my wish that they will, at the very least, give the reader some appreciation of how these levels function and overlap during development. In regards to the disjointed nature of the topics presented, I hope this will not be interpreted as fickleness on the part of the author (to which I might admit some fault) but instead as a desire for a broader understanding of biology.

ACKNOWLEGEMENTS

Chapter I, in full, is a reprint of the material as it appears in **Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W.** (2009). Visualization of individual Scr mRNAs during embryogenesis yields evidence for transcriptional bursting. Curr. Biol. *19*, 2037-42. I was the primary investigator and author of this paper.

Chapter II, in part, is currently being prepared for submission for publication of the material. **Lemons, D., Paré, A., and McGinnis, W.** Drosophila Hox complex miRNAs do not have major effects on expression of evolutionarily conserved Hox gene targets (in preparation). I was the primary investigator and author of the material presented in this chapter.

Chapter III, in full, is currently being prepared for submission for publication of the material. **Paré, A., Kim, M., and McGinnis, W.** The homolog of the metazoan epidermal integrity regulator Grainy head is involved in cell-wall formation, defense, and virulence in the fungus Neurospora crassa (in preparation). I was the primary investigator and author of this material.

VITA

1981   Born, Queens, New York

2003   Bachelor of Science, Cornell University

2003-2005   Junior Research Scientist, New York University

2011   Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Clyde, D., Corado, M., Wu, X., Paré, A., Papatsenko, D., and Small, S. (2003). A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. Nature *426*, 849-53.

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

Oberstein, A.*, Paré, A.*, Kaplan, L., and Small, S. (2005). Site-specific transgenesis by Cre-mediated recombination in Drosophila. Nat. Methods *2*, 583-5.

Paré, A.*, Dean, D.*, and Ewer, J. (2008). Construction and characterization of deletions with defined endpoints in Drosophila using P-elements in trans. Genetics *181*, 53-63.

Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009). Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr. Biol. *19,* 2037-42.

Hsia, C., Paré, A., Hannon, M., Ronshaugen, M., and McGinnis, W. (2010) Silencing of an abdominal Hox gene during early development is correlated with limb development in a crustacean trunk. Evol. Dev. *12*, 131-43.

Paré, A., Kim, M., and McGinnis, W. (in preparation) The homolog of the metazoan epidermal integrity regulator Grainy head is involved in cell-wall formation, defense, and virulence in the fungus Neurospora crassa.

Lemons, D., Paré, A., and McGinnis, W. (in preparation) Drosophila Hox complex miRNAs do not have major effects on expression of evolutionarily conserved Hox genes.

ABSTRACT OF THE DISSERTATION

Transcriptional Stochasticity

and

The Function of Grainy head Transcription Factors in Animals and Fungi

by

Adam Christopher Paré

Doctor of Philosophy in Biology

University of California, San Diego, 2011

Professor William McGinnis, Chair

This dissertation is presented in three parts. In the first part, using a combination of simultaneous RNA and protein detection, high-resolution confocal microscopy, and image segmentation, I show that it is possible to resolve and count the number of mRNA transcripts for a given gene within single-cells of fixed *Drosophila* embryos. I used these methods to study the stochastic nature of transcription at the endogenous locus of the Hox gene *Sex combs reduced*, and I uncovered evidence for transcriptional bursting as well as divergent modes of transcription. This was the first time such analyses had been carried out in an intact metazoan organism.

In the second part, I present evidence that despite the presence of several well conserved putative binding sites in the 3'UTR of the Hox gene *Antennapedia*, the microRNA miR-iab-4-5p does not appear to play a large role in the regulation of

Antennapedia protein levels during embryogenesis, and at most has very subtle effects. This is important, because despite the fact that *Drosophila* Hox genes and Hox-cluster encoded microRNAs are very strongly predicted to interact *in silico*, most investigations of these interactions have uncovered only very subtle (if any) effects on protein levels. This suggests that our ability to predict miRNA target sites *in silico* is very lacking, or that, in general, microRNAs play very subtle roles during development.

In the third part, I investigate the function of the homolog of the Grainy head transcription factor in the fungus *Neurospora crassa*. In all animal model organisms in which they have been studied, Grainy head transcription factors play a conserved role in epidermal barrier formation and healing. I therefore thought it would be interesting to investigate the function of this transcription factor in fungi, organisms which lack an epidermis. Using microarray and phenotypic analyses I uncovered evidence that the Grainy head homolog in *Neurospora* plays a role in cell-wall formation, defense, and virulence. This points to an interesting connection between transcriptional control of physical-barrier formation in animals, and physical-barrier formation, defense, and virulence in fungi.

Chapter I

Visualization of individual *Scr* mRNAs during *Drosophila* embryogenesis yields

evidence for transcriptional bursting

SUMMARY

The detection and counting of transcripts within single cells using Fluorescent in situ Hybridization (FISH) has allowed researchers to ask quantitative questions about gene expression at the level of individual cells. This method is often preferable to quantitative RT-PCR, because it does not necessitate destruction of the cells being probed and maintains spatial information that may be of interest. Until now, studies using FISH at single molecule resolution have only been rigorously carried out in isolated cells (e.g., yeast cells or mammalian cell culture). Here, I describe the detection and counting of transcripts within single cells of fixed, whole-mount *Drosophila* embryos via a combination of FISH, immunohistochemistry, and image segmentation. My method takes advantage of inexpensive, long RNA probes detected with antibodies, and I present novel evidence to show that I can robustly detect single mRNA molecules. I use this method to characterize transcription at the endogenous locus of the Hox gene *Sex combs reduced*, by comparing a stably expressing group of cells to a group that only transiently expresses the gene. My data provide evidence for transcriptional bursting, as well for divergent "accumulation" and "maintenance" phases of gene activity at the *Scr* locus.

RESULTS AND DISCUSSION

In early *Drosophila* embryos, the limits of Hox expression domains along the anterior-posterior axis are set by parasegmental boundaries. Parasegments are repeating units of cellular organization that make up the body plan of early embryos, and the Hox gene *Sex combs reduced* (*Scr*) displays dynamic differences in expression between parasegments 2 and 3 (PS2 and PS3) (Figure 1). Cells in PS2, which give rise to the posterior mouthparts, stably express *Scr* from early embryogenesis onward. Cells in ventral PS3, which contribute to the first thoracic segment, display a transient burst of *Scr* transcription during midembryogenesis (Figures 1C-E) (Kuroiwa et al., 1985; Mahaffey and Kaufman, 1987; Martinez-Arias et al., 1987). Thus, this system offers a convenient way to compare *Scr* transcriptional dynamics between stably and transiently expressing groups of cells in the same embryo, as well the opportunity to shed light on the expression of a crucial developmental regulator.

Detection and counting of single *Scr* transcripts

At low-magnification, fluorescent signals from a probe directed against *Scr* mRNAs have a "speckled" appearance (Figure 2A). At high-magnification most cytoplasmic signals are resolvable as ellipsoids of roughly uniform size (~250-300 nm diameter in x and y, Figure 2B, arrows). Although I believed I was visualizing single mRNA molecules (Femino et al., 1998; Raj et al., 2006; Maamar et al., 2007; Raj et al., 2008; Zenklusen et al., 2008; Lu and Tsourkas, 2009; Fusco et al., 2003; Golding and Cox, 2004; Vargas et al., 2005), it was possible that they instead represented mRNA aggregates (e.g., P-bodies) (Lin et al., 2008). One method for demonstrating single

transcript Fluorescent in situ Hybridization (FISH) resolution is to show that a spatial

shift exists between signals from two different probes targeted to adjacent regions of an

mRNA (Femino et al., 1998; Raj et al., 2008; Lu and Tsourkas, 2009), which should not

be present if one is visualizing an aggregate of randomly oriented transcripts. Consistent

with this, I observed a randomly oriented spatial shift between signals from probes

directed against the coding region and the 3'UTR of *Scr* (Figure 3F). Another method to

demonstrate single RNA molecule detection is to show that the fluorescence emitted by

specific numbers of direct-labeled oligonucleotide probes bound at each locus is

reproducible and predictable (Femino et al., 1998; Raj et al., 2008; Vargas et al., 2005).

While the long RNA probes used in this study offer a large increase in signal-to-noise

ratios, compared to oligonucleotide probes, because they are indirectly labeled, the

fluorescence they emit is more variable (data not shown). Therefore, I developed a

different assay to test whether the punctate cytoplasmic signals represented single

transcripts. Single transcripts should contain only a single binding site for a unique probe

sequence. If two probes against the same sequence are labeled with different hapten tags

and simultaneously hybridized to embryos, there should be competition between the two

probes, and very low levels of association should be observed.

I tested for such competition using two unfragmented probes complementary to

the same 330 bp region of the *Scr* 3'UTR, labeled with either digoxigenin or

dinitrophenyl haptens (probes S1 and S2, respectively, Figure 2H). This experiment was

done as part of a triple hybridization, using a biotin-labeled coding region probe (ORF,

Figure 2H) as a marker for the adjacent *Scr* mRNA protein coding region (Figures 2D-

G). Randomly chosen S1 signals were almost always associated with an ORF signal

(79%; n = 100; Figures 2D, E, and I), but rarely with an S2 signal (17%; n = 100; Figures

2E, F, G, and I). A given S1 signal was only associated with both an S2 and an ORF

signal in a minority of cases (10%; n = 100; Figure 2I), which is strong evidence that

these locations contain only single binding sites for an S probe. Although the S1→S2

association statistics may seem high, rotating the S2 image stack 90° relative to the S1

image stack, and re-scoring the same 100 S1 signals (to simulate random association)

yields a nearly identical association level of 20% (Figure 2I), indicating that this

association can be explained by the chance overlap of numerous signals in a finite

volume. I therefore conclude that the large majority of the cytoplasmic *Scr* signals I

observe correspond to single mRNA molecules. More pairwise association data, as well

as antibody detection and probe binding efficiency data, are shown in Figures 3 and 4.

To group and count transcripts from single cells of the embryo, I used a

combination of RNA FISH, immunohistochemistry to detect cell boundaries, manual cell

segmentation, and automated transcript signal segmentation. Cells of interest were

manually segmented, with an anti-spectrin antibody (Pesacreta et al., 1989) used to stain

cell membranes as a guide. The cell segmentation process was accelerated using an

ImageJ plugin developed by W. Beaver that allows a user to quickly draw unique regions

of interest (ROIs) for each cell outline in a sequence of confocal image slices (Figures 5A

and B). These ROIs then defined the 3D boundaries used to group the FISH signals from

each cell (Figure 5C).

The punctate FISH signals themselves were automatically segmented and counted

using the Volocity 3D image analysis program (Figure 5D). Almost all FISH transcript

signal segmentations appeared correct upon visual inspection, and the algorithm yielded

transcript counts that were nearly identical (±6%) to those obtained by manual counting (Table 1). Given this variation, and the fact that more than one transcript will occasionally occupy the same volume, I believe this counting method yields transcript numbers that are within ±10% of the actual value.

Analysis of *Scr* transcription

Intense FISH signals representing sites of transcription in the nucleus are often detected with probes to upstream exons or introns of a gene (e.g., Figure 2B, arrowhead) (Shermoen and O'Farrell, 1991; Wilkie et al., 1999). The transcriptional activity of a gene can be roughly quantified by measuring the fluorescence intensity of these spots, which will vary according to the number of nascent transcripts associated with the locus (Femino et al., 1998; Chubb et al., 2006; Raj et al., 2006; Zenklusen et al., 2008; Boettiger and Levine, 2009). To characterize transcription at the *Scr* locus I counted cytoplasmic transcripts with an ORF probe (Figure 5E), and nascent transcript intensity with an *Scr* intron probe (Figures 1A and 5F).

I first examined several stably expressing cells from PS2, as well as several transiently expressing cells in PS3 (Figures 5E-G). To my surprise, cell groups from both PS2 and PS3 displayed a wide range of cellular transcript numbers (72-262 for PS2, and 3-14 for PS3; Figure 5G). In PS3 cells, the low number of cytoplasmic transcripts was consistent with undetectable levels of nuclear transcription in the same nuclei. However, in PS2, there was not a good correspondence between cytoplasmic and nascent transcript signals; an extreme example of this is shown for two nearby cells (Figure 5E, arrows, and

Figures 5H, I, and J). Although both cells contain over a hundred cytoplasmic mRNAs, one cell has two obvious sites of transcription, while the other has none.

To investigate this further, I carried out a more comprehensive analysis on three embryos during stages 10 and 11 of embryogenesis (Campos-Ortega and Hartenstein, 1997), and the results are shown in Figure 6. Embryos were chosen that were representative of different phases of *Scr* transcription: before, during, and after the transient period of *Scr* expression in PS3 (Figures 6A-F). Approximately 20 ventro-lateral ectodermal cells from both PS2 and PS3 were segmented, and all cells were located ~50 μm from the ventral midline (Figures 6A-C). Strongly expressing PS2 cells had an average of 94 mRNAs per cell, with the values exhibiting a large range from 33 to 177 mRNAs per cell (n = 58; Figure 6G). Differences in cell size were not responsible for this heterogeneity, because a similar distribution of values was seen after taking cell volume into account (Figure 6H). For each of the three stages examined, the average numbers of *Scr* transcripts per cell in PS2 were similar (100, 104, and 80). For PS3 cells, average *Scr* mRNA numbers were very low during stage 10 (5 transcripts), increased dramatically during early stage 11 (33 transcripts), and decreased during late stage 11 (12 transcripts) (Figure 6G). See Table 2 for data and statistical analyses.

To determine whether cells expressing other Hox genes produced similar numbers of transcripts, I also counted mRNAs for *Deformed* (*Dfd*) and *Ultrabithorax* (*Ubx*) in areas of abundant transcript accumulation during stage 11 of embryogenesis. Values for these two Hox genes were similar to those found for *Scr* in PS2, with *Dfd* having an average of 92 mRNAs per cell, and *Ubx* an average of 74 per cell (Figures 6G and H).

Graphs plotting number of *Scr* transcripts per cell and nascent transcription strength along the anterior-posterior axis are shown in Figure 6A'-C' (red and green lines, respectively). Surprisingly, for PS2 cells in the stage-10 embryo, the graphs were divergent (Figure 6A'). On the other hand, in stage-11 cells, transcript numbers and nascent transcription levels rose and fell largely in unison (Figures 6B' and C'). Figures 6I and 6J show scatter plots for the cell groups in PS2 and PS3, and nonparametric correlations were calculated for all cell groups. Consistent with the traces, stage-10 PS2 cells showed a significant negative correlation between cytoplasmic transcript numbers and nascent transcription ($r = -0.7$; $p < 0.05$), whereas early and late stage-11 PS2 cell groups both showed weak but significant positive correlations ($r = 0.47$ and 0.60; $p < 0.05$) (Figure 6I). On the other hand, PS3 cells had very significant positive correlations between cellular transcript numbers and nascent transcription for both early and late stage-11 cell groups ($r = 0.85$ and 0.67 respectively; $p < 0.001$) (Figure 6J; see Table 3 for correlation data). It is possible that the same mode of transcription occurring in PS3 during stage 11 may also be occurring in PS2 during the same period, although the positive correlations are not as striking because they are superimposed upon an existing pool of transcripts.

Recent data indicates that transcription is often not only stochastic (meaning transcription initiation is probabilistic) but also occurs in bursts, during which a gene will switch back and forth between prolonged active and inactive states (Golding et al., 2005; Chubb et al., 2006; Kaern et al., 2005; Raj et al., 2006; Voss et al., 2006; Zenklusen et al., 2008; Raj and van Oudenaarden, 2009). My observations of large variations in transcript numbers on a cell-by-cell basis, as well as the often poor correlation between nascent

transcription and cellular transcript numbers, indicate that transcriptional bursting is taking place at the *Scr* locus in PS2. One way to capture the relative intensity of these bursts is through the use of the Fano factor (FF) (Raj and van Oudenaarden, 2009), which is essentially a measurement of population heterogeneity. In this case, it is defined as the variance of the distribution of transcript numbers per cell divided by the mean. Even stochastically transcribing cell populations can have small FF values (<1) if most cells contain similar transcript numbers, but FF values larger than 1 are suggestive of transcriptional bursting. I observed FF values of 7.1, 8.4, and 16, for the three PS2 cell-groups (Table 2), which are intermediate to an observed FF value of ~4 in bacteria (Golding et al., 2005), and an FF value of >40 for a transgenic reporter gene in mammalian cells (Raj et al., 2006). Whether the heterogeneity I observe is due to intrinsic noisiness in *Scr* transcription or to high variations in activator and repressor input (extrinsic noise) (Elowitz et al., 2002; Swain et al., 2002; Raser and O'Shea, 2005; Raj and van Oudenaarden, 2008), is as yet unknown.

My observations also indicate that there may exist divergent "accumulation" and "maintenance" phases of *Scr* transcription, characterized by stage-11 PS3 cells and stage-10 PS2 cells, respectively. The *Scr* gene may begin transcribing in a stochastic, but still relatively constant manner, until a threshold number of transcripts are reached, after which it switches to a bursting mode of transcription to maintain the mRNA pool. The mechanisms whereby a cell might directly sense the concentration of a distinct mRNA species are unclear, although regulation of *Scr* transcription via downstream targets of the SCR protein could explain this phenomenon. The simultaneous RNA/protein detection procedures described in this paper should allow for more detailed studies in which

endogenous transcription factor concentrations can be correlated to target gene activity on a cell-by-cell basis.

In summary, I have characterized endogenous transcription of the *Scr* locus at single molecule resolution in *Drosophila* embryos and provided evidence for transcriptional bursting, as well as for two divergent modes of gene expression. To my knowledge, this is the first rigorous analysis of transcription using single-molecule FISH performed in a developing metazoan. Using FISH or live imaging in these kinds of studies is crucial, because biochemical methods that extract RNA from cell populations do not detect cell-to-cell variations. Carrying out these analyses at single-molecule resolution is similarly crucial, because metrics such as the FF are impossible to derive when using arbitrary whole cell fluorescence measurements (Raj and van Oudenaarden, 2009). Finally, single-molecule measurements are much more objective and should allow for the comparison of results between disparate experiments.

EXPERIMENTAL PROCEDURES

Simultaneous Fluorescent in situ Hybridization and protein detection

Haptenylated probes were created by in vitro transcription as described elsewhere (Kosman et al., 2004). The intronic probe was directly labeled with Alexa Fluor 555 dyes and was prepared by Invitrogen. Simultaneous RNA and protein detection was carried out via a modified standard FISH protocol (Kosman et al., 2004) with acetone used instead of Proteinase K permeabilization (Nagaso et al., 2001). Dechorionated embryos were fixed in 8% formaldehyde for 25 min, devitellinized by vigorous shaking in a 1:1 heptane:methanol mixture, washed with ethanol, rocked in a 1:1 ethanol:xylenes mixture for 30 min, washed with methanol, and then gradually rehydrated in a series of methanol:H_2O washes (3:1, 1:1, 1:3, and 0:1). Embryos were permeabilized in cold 80% acetone for 10 min at -20° C, and then were transferred into phosphate buffered saline plus 0.1% Tween (PBT). Embryos were then post-fixed in 5% formaldehyde in PBT for 25 min and washed with PBT. RNA probe hybridization and immunohistochemistry (including antibody combinations) were carried out as described elsewhere (Kosman et al., 2004). Spectrin (Pesacreta et al., 1989) and engrailed (Patel et al., 1989) antibodies were obtained from the Developmental Studies Hybridoma Bank (antibodies 3A9 and 4D9, concentrate) and were used at a 1:100 dilutions.

All images were collected with a Leica SP2 laser-scanning confocal microscope. Gain and offset were set to nonsaturating levels such that intensity data would span the entire dynamic range, and line averaging was set to 2. Stacks of at least one-cell thickness (~15 μm) were collected, and channels were shifted relative to one another to correct for

Z-axial chromatic aberration (which was measured independently with Tetraspeck fluorescent beads). All images were deconvolved using the AutoDeblur software program.

Cell segmentation, transcript counting, and nascent transcription quantification

A set of ImageJ (http://rsbweb.nih.gov/ij/) plug-ins was developed to allow us to manually segment confocal stacks (contact W. Beaver, wbeaver@cs.ucsd.edu). Transcript segmentation and counting was carried out using the image analysis program Volocity. First, transcripts were counted manually for several cells ($n = 4$), and this training set was used to tune the variables of the Volocity segmentation algorithm so that it predicted transcript numbers that were nearly identical to manual counts. I then used the algorithm to segment and count transcripts for the training set plus 8 more cells that were not part of the training set (Table 1). Overall the algorithm predicted values that were within ±6% of the manually derived values, and it was accurate over a wide range of values (20 - 153) without any obvious bias towards a certain range.

Volocity was also used to measure nascent transcript intensities as well as cell volumes. Both transcribing alleles were often distinguishable, although I could not rule out the possibility that cells containing solitary signals did not represent cases of overlapping alleles; therefore I simply summed the intron probe fluorescence from the entire nucleus.

ACKNOWLEDGMENTS

Chapter I, in full, is a reprint of the material as it appears in **Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W.** (2009). Visualization of individual Scr mRNAs during embryogenesis yields evidence for transcriptional bursting. Curr. Biol. *19*, 2037-42. The dissertation author was the primary investigator and author of this paper.
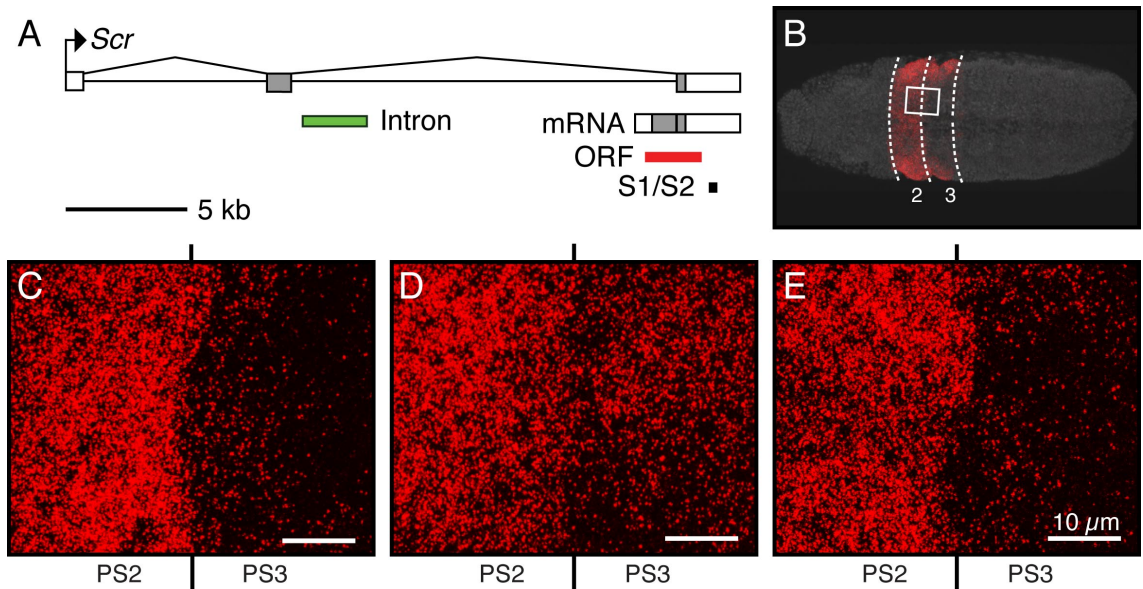
Figure 1. The *Scr* expression pattern during midembryogenesis

(A) The *Scr* genomic locus, mRNA, and the locations of in situ probes used in this study. **(B)** A ventral view of a stage-11 *Drosophila* embryo showing *Scr* mRNA (red) and nuclei stained with DAPI (gray). The boundaries of parasegments 2 and 3 (PS2 and PS3) are indicated with dashed lines. The white box highlights the approximate areas shown in (C-E). **(C-E)** Expanded views of the area marked in (B) showing the boundary between PS2 and PS3 for a stage-10 embryo (C), an early stage-11 embryo (D), and a late stage-11 embryo (E). Transcripts are detected via FISH with a probe specific to the coding region of *Scr* (ORF probe). A high accumulation of *Scr* transcripts in PS2 is maintained throughout the three stages, whereas PS3 cells only highly accumulate *Scr* during early stage 11.

Figure 2. Competition for binding sites demonstrates that punctate signals represent single mRNA transcripts, and not groups of transcripts

(A) A ventral view of *Scr* transcript expression during early stage 11. Parasegmental boundaries are indicated with dashed lines. A region with high transcript levels, and a region not expressing *Scr*, are marked with white boxes and are shown at high magnification in (B) and (C). **(B)** *Scr* FISH signals are punctate (arrow). Sites of nascent transcription appear as large, often irregularly shaped, nuclear signals (arrowhead). Nuclear boundaries are based on DAPI staining and are indicated with gray lines. **(C)** Areas outside the region of *Scr* expression sometimes contain very weak fluorescent signals (arrow), which are also seen with no probe controls. **(D-F)** Results from a triple-hybridization "competition assay". FISH was carried out using the *Scr* ORF probe (D) and two differentially labeled unfragmented probes (S1 and S2) both complementary to the same region of the 3'UTR (E and F). **(G)** A merge of (E) and (F) shows very little colocalization between the competing S1 and S2 probes. Most associated signals (arrows) can be attributed to sites of nascent transcription, where multiple RNAs are present in a small volume. **(H)** The *Scr* mRNA and the locations of FISH probes used in this assay. **(I)** A histogram summarizing the pairwise associations between signals in the three fluorescent channels. "Association" is defined as any overlap between signals in three dimensions. For example, the "S1→ ORF" bar refers to the percentage of time that an S1 signal overlaps with a signal from the ORF channel. "S1→ S2 (rotated)" refers to a control where an image stack from the S2 channel was rotated 90 degrees relative to the S1 channel, to simulate random association of signals. "S1 → ORF & S2" refers to cases of association in all three channels.
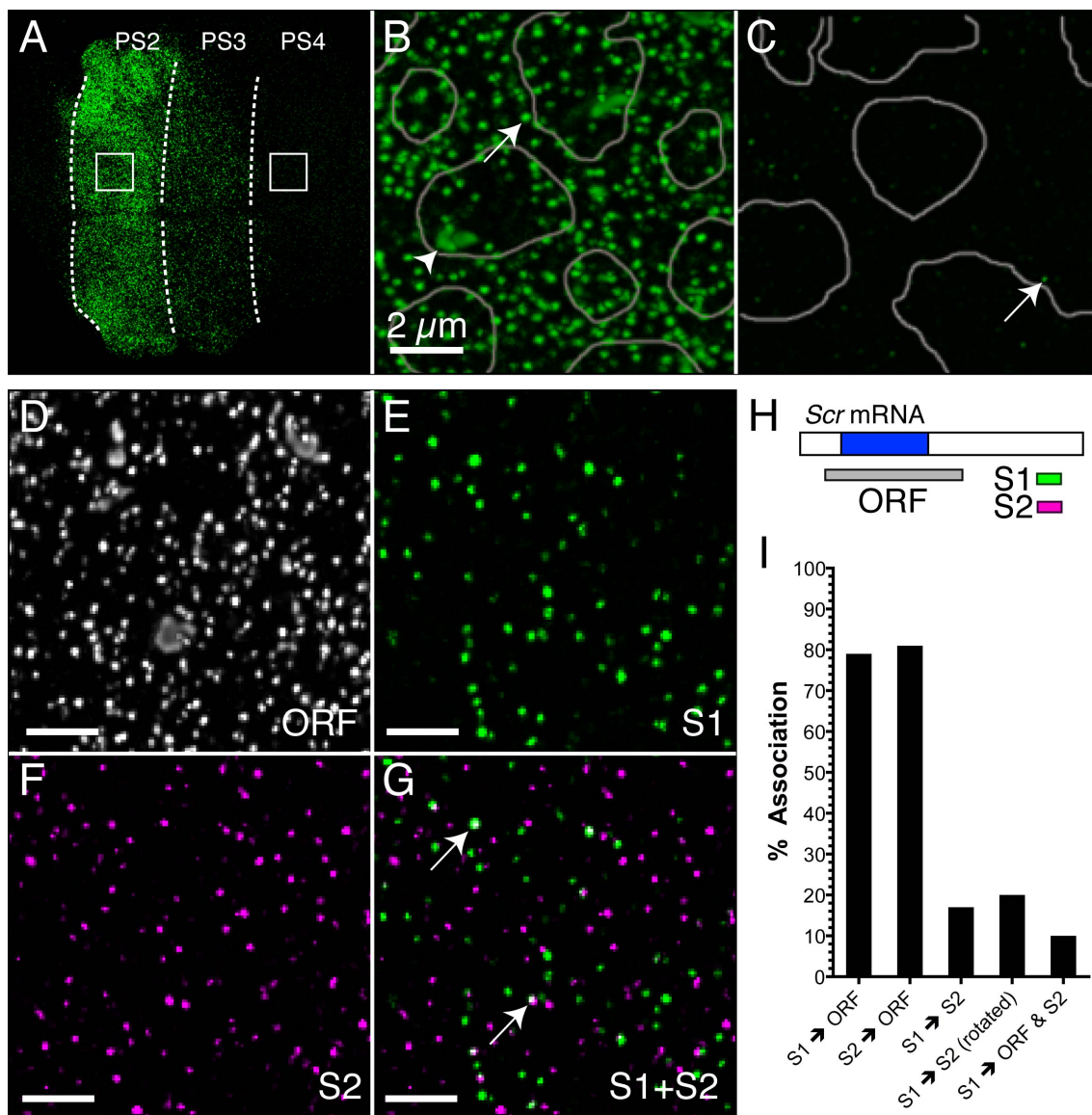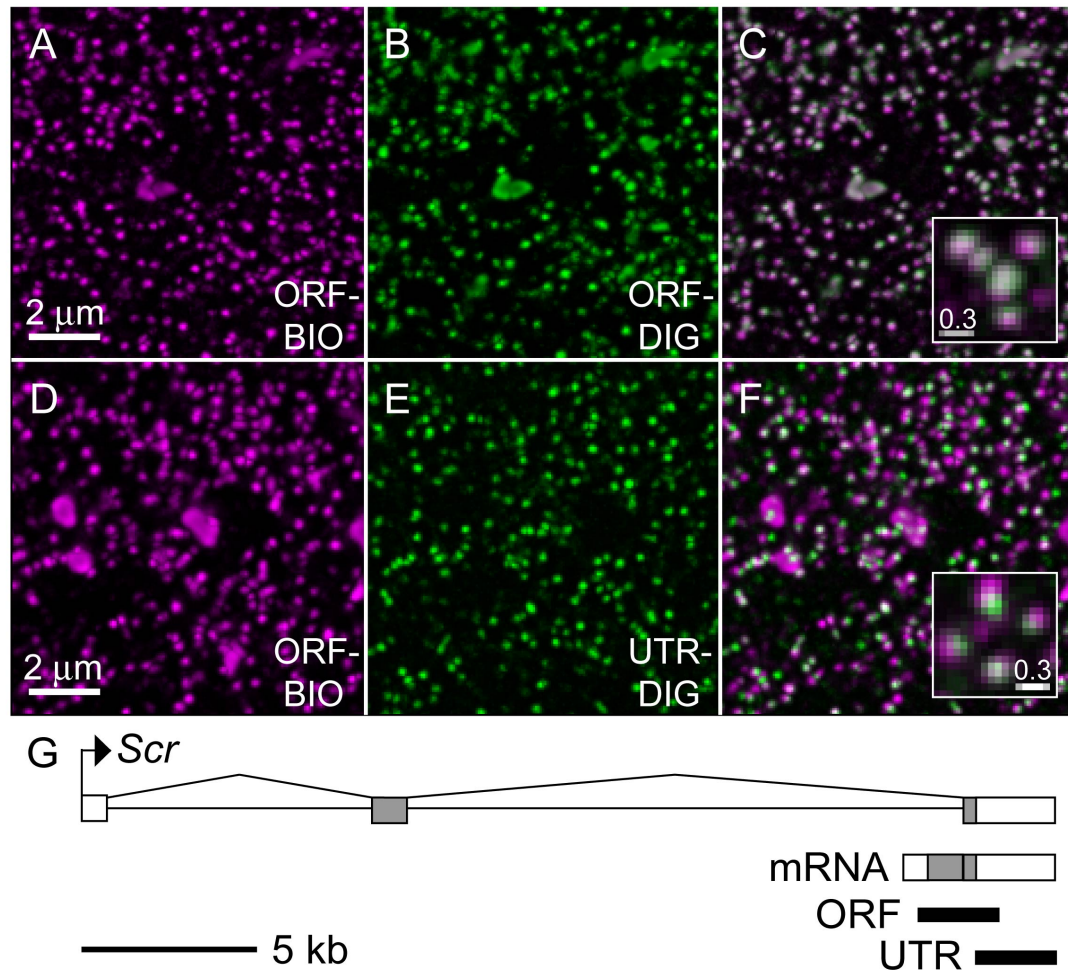
Figure 3. Probes are detected with high efficiency using antibodies, and RNA probes have a high hybridization efficiency

(A and B) An *Scr* open reading frame probe (ORF) dually labeled with two kinds of haptens (BIO and DIG) and detected with spectrally distinct antibody sets. **(C)** A merge of (A) and (B). The inset highlights the nearly perfect colocalization between channels. A perfect association rate was observed between the two channels (n = 100), demonstrating that if a probe is present in the sample it will be almost certainly be detectable. **(D)** *Scr* detected with an ORF probe. **(E)** The same cells in panel (D), showing *Scr* detected with a 3'UTR probe (UTR). **(F)** A merge of (D) and (E). The inset illustrates the slight offset between signals detected with adjacent probes, which was not seen for the dually labeled ORF probe. An association rate of 83% was observed between ORF and UTR signals (n = 100). This could be explained by a ~90% occupancy rate for both probes, or this reflects the partially degraded nature of the transcript pool in vivo. Similar association rates were observed between the ORF and S probes (see Figure 2). **(G)** Probes used in these assays.
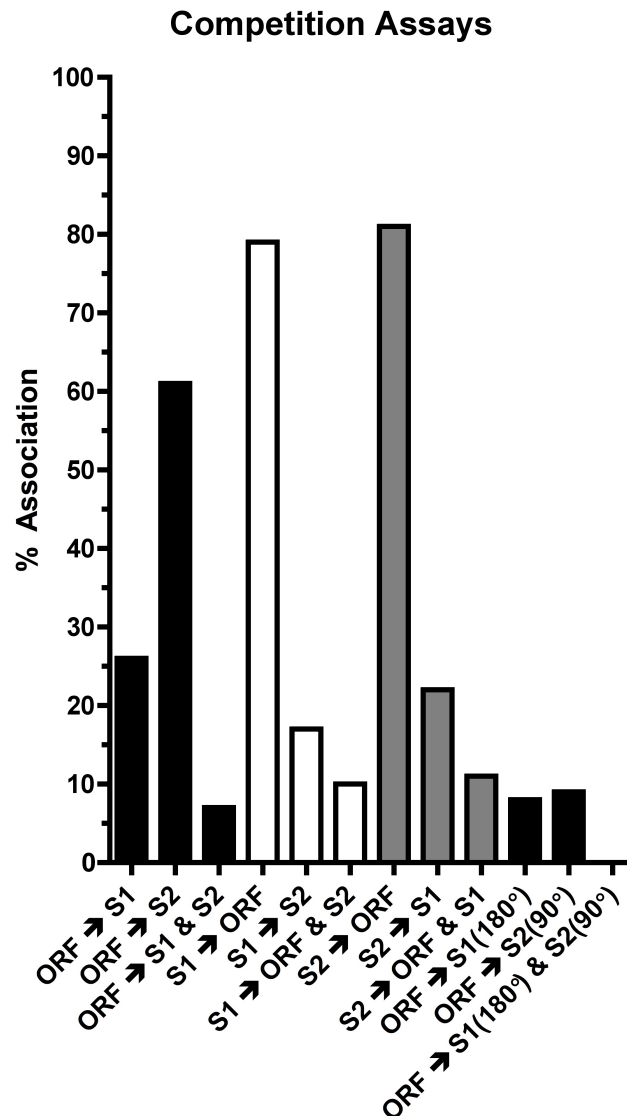
Figure 4. Pairwise associations of signals from the competition assay

A histogram summarizing the pairwise associations between signals in the competition assay. FISH was carried out with an ORF probe and two differentially labeled probes both specific to a short region of the 3'UTR (S1 and S2, Figure 2H). "Association" is defined as any overlap between signals in three dimensions. For example, the "S1 → ORF" bar refers to the percentage of time that a randomly chosen S1 signal overlaps with a signal from the ORF channel. "ORF → S1(180°)" refers to a control where the S1 channel was rotated 180 degrees relative to the ORF channel, to simulate random association of signals in a finite volume.

Figure 5. A combination of manual and automated image segmentation allows for counting transcripts within individual cells in complex tissues

(**A**) An image of *Scr*-expressing cells in a region of PS2 from a stage-11 embryo. Cell membranes are marked by spectrin staining (blue), nuclei are marked with DAPI (gray), and *Scr* transcripts are shown in red (ORF probe, Figure 1A). The segmented cell shown in (B-D) is indicated with a dashed line and arrow. (**B**) A surface rendering of the volume defined by manual segmentation for the cell highlighted in (A). (**C**) *Scr* probe signals from the segmented cell in (B). (**D**) A false-color rendering of the signals in (C), which were segmented into individual objects. (**E**) An image of *Scr*-expressing cells (ORF probe, Figure 1A) in PS2 and PS3 from a late stage-11 embryo. Cell membranes are marked by spectrin staining (blue), *Scr* transcripts are shown in red, and segmented cells are depicted as solid white objects. The pair of cells highlighted in (H-J) is indicated with arrows, and the boundary between PS2 and PS3 is marked with a dashed line. (**F**) Nascent transcription detected with an intronic probe (Intron probe, Figure 1A) in the same embryo as shown in (E). (**G**) A schematic showing *Scr* transcript numbers and relative nascent transcription strength for three groups of cells. Red numbers represent total transcripts per cell, and green numbers represent strength of nascent transcription (as the percentage of maximal intensity). Sites of nuclear transcription are represented as dots inside each cell. (**H-J**) A pair of neighboring cells with identical transcript concentrations exhibiting divergent transcriptional states; shown are *Scr* transcripts (H), nascent transcription (I), and a merge of (H) and (I) (in J).
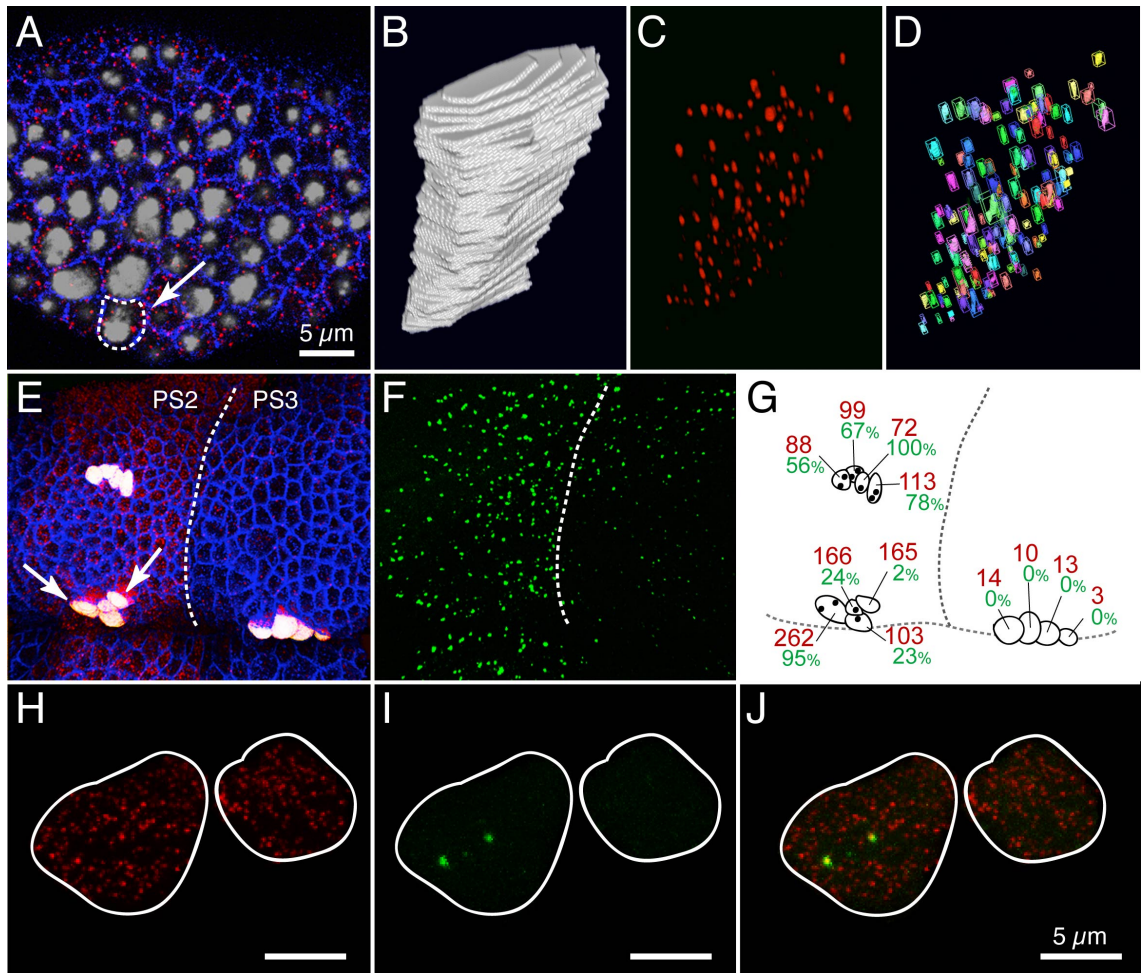
Figure 6. Analysis of *Scr* transcription at single-cell resolution

(A-C) *Scr* expression in three embryos: stage 10 (A), early stage 11 (B), and late stage 11 (C). *Scr* transcripts are shown in red, and cell membranes are marked in gray. Nuclei from the anterior compartment of each parasegment are stained for engrailed protein and appear as faint vertical gray stripes. The segmented cells analyzed in (G-J) are outlined in white. **(A'-C')** Graphs showing cellular transcript numbers (red lines) and relative nascent transcription strength (green line, as a percentage of the maximum value) in the outlined cells plotted against cell centroid position. **(D-F)** The same embryos depicted in (A-C) stained with an intronic probe, showing sites of nascent transcription (green). **(G)** Box plots summarizing transcripts per cell for various groups of cells. Boxes depict the median value and the middle two quartiles. Whiskers indicate the range of measurements, and the mean is shown as a dot inside each box. **(H)** A box plot summarizing transcripts per volume (for ease of comparison to (G), values are shown with the units "transcripts/250 μm^3", which is a typical cell volume) for various groups of cells. **(I and J)** Scatter diagrams plotting cellular transcript numbers against nascent transcription strength (as a percentage of maximum intensity) for PS2 and PS3 cells, respectively.

stage 10

early 11

late 11

Transcripts per Cell

Transcripts per Volume

PS2

PS3

Nascent Transcription (% of Maximum)

- 10
- early 11
- late 11
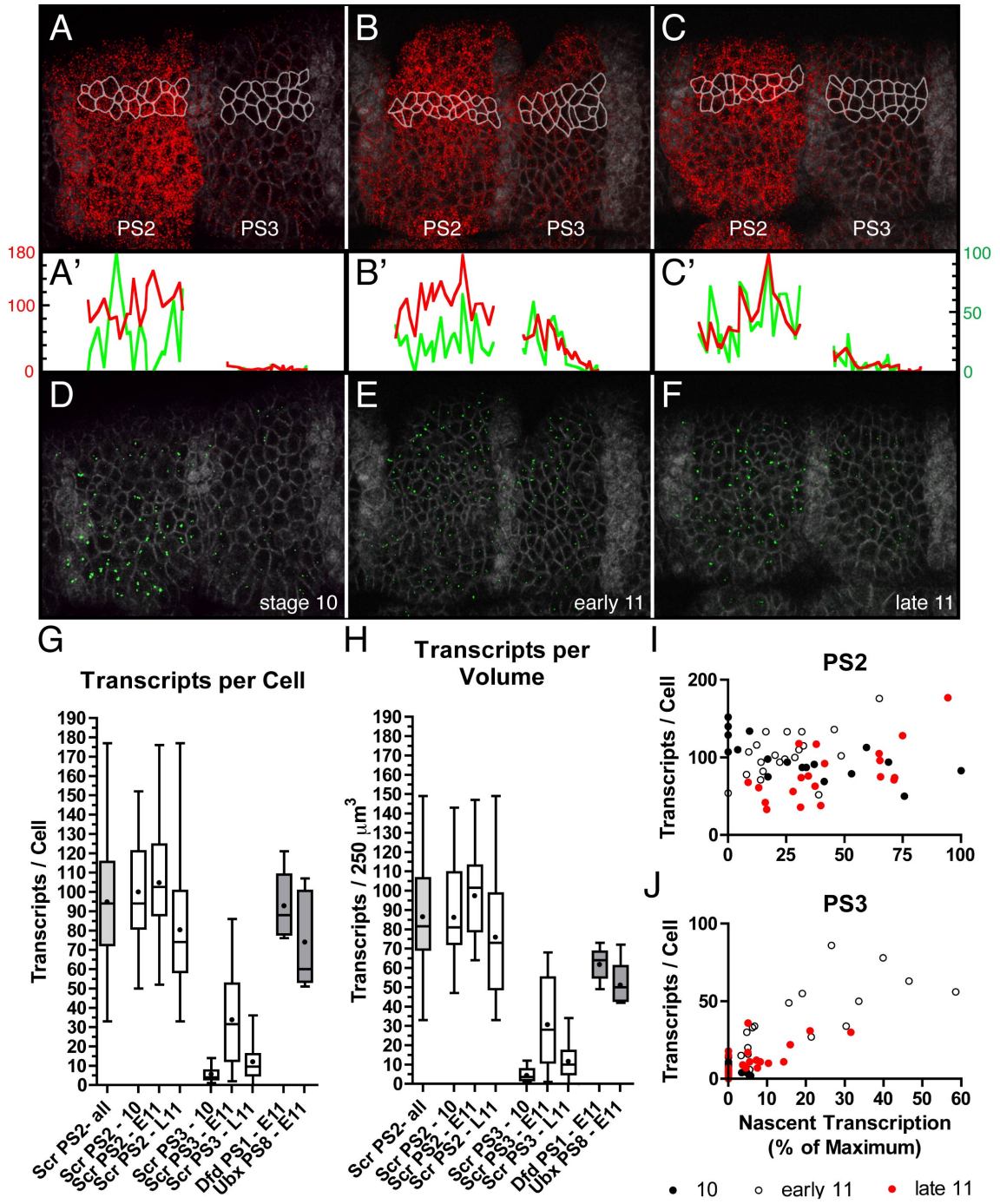
Table 1. Determination of transcript-segmentation algorithm error

Determination of error for the transcript-segmentation algorithm used in this paper. The transcript signals from 12 segmented cells were counted by hand. Four cells were chosen as a training set (gray rows), and the algorithm variables were tuned to output values that matched the manual counts as closely as possible. The algorithm was then used to count transcripts for all 12 cells and average percent error was calculated.

Cell	Manual Count	Segmentation Algorithm Count	% Error = $\frac{(SAC - MC)}{MC} * 100$
1	153	150	-2.0
2	106	96	-9.4
3	126	127	0.8
4	80	87	8.8
5	82	75	-8.5
6	125	133	6.4
7	112	113	0.9
8	144	139	-3.5
9	29	31	6.9
10	20	22	10.0
11	29	31	6.9
12	29	32	10.3
		Average | % Error | →	6.2

Table 2. Transcript number and transcript concentration statistics for various groups of cells

Abbreviations: SD - Standard Deviation, FF - Fano Factor (Variance/Mean), n - number of measurements. For ease of comparison with Cellular Transcript numbers, Transcript Concentration values are shown with the units "Transcripts / 250 μm^3", which is a typical cell volume. In other words, these values represent the number of transcripts that would be found in a given cell, if concentration were held constant but the volume of the cell were adjusted to 250 μm^3.

Transcripts per Cell

mRNA	Scr							Dfd	Ubx
Position	PS2	PS2	PS2	PS2	PS3	PS3	PS3	PS1	PS8
Stage	all	10	E11	L11	10	E11	L11	E11	E11
Mean	94.47	99.56	104.40	80.00	5.25	33.45	11.63	92.40	73.60
SD	32.38	26.49	29.69	35.89	3.45	25.34	9.76	17.90	25.19
Variance	1048.46	701.72	881.50	1288.09	11.90	642.12	95.26	320.41	634.54
FF	11.10	7.05	8.44	16.10	2.27	19.20	8.19	3.47	8.62
Median	94	94	102.5	74	4	31.5	9.5	88	60
Range	33-177	50-152	52-176	33-177	1-14	2-86	0-36	76-121	51-107
n	58	18	20	20	20	20	24	5	5

Transcript Concentration (Transcripts / 250 μm^3)

mRNA	Scr							Dfd	Ubx
Position	PS2	PS2	PS2	PS2	PS3	PS3	PS3	PS1	PS8
Stage	all	10	E11	L11	10	E11	L11	E11	E11
Mean	86.91	86.56	97.75	76.40	4.800	31.10	12.08	62.20	51.60
SD	28.12	26.96	22.89	30.95	3.205	23.06	9.806	8.643	11.95
Variance	790.73	726.84	523.95	957.90	10.27	531.76	96.16	74.70	142.80
FF	9.10	8.40	5.36	12.54	2.14	17.10	7.96	1.20	2.77
Median	81.50	81.00	101.5	73.00	3.500	28.00	10.00	64.00	50.00
Range	33-149	47-143	64-147	33-149	1-12	1-68	0-34	49-73	42-72
n	58	18	20	20	20	20	24	5	5

Table 3. Correlation analyses between cellular transcript numbers and strength of nascent transcription

The results of nonparametric correlation analyses between cellular transcript numbers and strength of nascent transcription for various groups of *Scr*-expressing cells.

	PS2 10	PS2 E11	PS2 L11	PS3 10	PS3 E11	PS3 L11
Number of XY Pairs	18	20	20	20	20	24
Spearman r	-0.7030	0.4655	0.5987	-0.1930	0.8463	0.6720
95% confidence interval	-0.8841 to -0.3382	0.01482 to 0.7590	0.1989 to 0.8277	-0.5947 to 0.2859	0.6371 to 0.9393	0.3574 to 0.8496
P value (two-tailed)	0.0011	0.0386	0.0053	0.4150	<0.0001	0.0003
P value summary	**	*	**	ns	***	***
Exact or approximate P value?	Gaussian Approximation					
Significant (alpha=0.5)	Yes	Yes	Yes	No	Yes	Yes

REFERENCES

Bengtsson, M., Ståhlberg, A., Rorsman, P., and Kubista, M. (2005). Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res. *15*, 1388-1392.

Boettiger, A., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471-473.

Campos-Ortega, J.A., and Hartenstein, V. (1997). The Embryonic Development of *Drosophila melanogaster*, Second Edition (Berlin, New York: Springer).

Chubb, J., Trcek, T., Shenoy, S., and Singer, R. (2006). Transcriptional pulsing of a developmental gene. Curr. Biol. *16*, 1018-1025.

Elowitz, M., Levine, A., Siggia, E., and Swain, P. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Femino, A., Fay, F., Fogarty, K., and Singer, R. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585-590.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S., Blanchard, J., Singer, R., and Bertrand, E. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr. Biol. *13*, 161-167.

Golding, I., and Cox, E.C. (2004). RNA dynamics in live Escherichia coli cells. Proc. Natl. Acad. Sci. USA *101*, 11310-11315.

Golding, I., Paulsson, J., Zawilski, S., and Cox, E. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025-1036.

Kaern, M., Elston, T., Blake, W., and Collins, J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. *6*, 451-464.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. Science *305*, 846.

Kuroiwa, A., Kloter, U., Baumgartner, P., and Gehring, W. (1985). Cloning of the homeotic Sex combs reduced gene in Drosophila and in situ localization of its transcripts. EMBO J. *4*, 3757-3764.

Lin, M., Jiao, X., Grima, D., Newbury, S., Kiledjian, M., and Chou, T. (2008). Drosophila processing bodies in oogenesis. Dev. Biol. *322*, 276-288.

Lu, J., and Tsourkas, A. (2009). Imaging individual microRNAs in single mammalian cells in situ. Nucleic Acids Res. *37*, e100.

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. Science *317*, 526-529.

Mahaffey, J., and Kaufman, T. (1987). Distribution of the Sex combs reduced gene products in Drosophila melanogaster. Genetics *117*, 51-60.

Martinez-Arias, A., Ingham, P., Scott, M., and Akam, M. (1987). The spatial and temporal deployment of Dfd and Scr transcripts throughout development of Drosophila. Development *100*, 673-683.

Nagaso, H., Murata, T., Day, N., and Yokoyama, K. (2001). Simultaneous detection of RNA and protein by in situ hybridization and immunological staining. J. Histochem. Cytochem. *49*, 1177-1182.

Patel, N., Martin-Blanco, E., Coleman, K., Poole, S., Ellis, M., Kornberg, T., and Goodman, C. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. Cell *58*, 955-968.

Pesacreta, T., Byers, T., Dubreuil, R., Kiehart, D., and Branton, D. (1989). Drosophila spectrin: the membrane skeleton during embryogenesis. J. Cell Biol. *108*, 1697-1709.

Raj, A., Peskin, C., Tranchina, D., Vargas, D., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

Raj, A., van den Bogaard, P., Rifkin, S., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877-879.

Raj, A., and van Oudenaarden, A. (2009). Single-molecule approaches to stochastic gene expression. Annu. Rev. Biophys. *38*, 255-270.

Raser, J., and O'Shea, E. (2005). Noise in gene expression: origins, consequences, and control. Science *309*, 2010-2013.

Shermoen, A., and O'Farrell, P. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303-310.

Swain, P., Elowitz, M., and Siggia, E. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. USA *99*, 12795-12800.

Vargas, D., Raj, A., Marras, S., Kramer, F., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. Proc. Natl. Acad. Sci. USA *102*, 17008-17013.

Voss, T., John, S., and Hager, G. (2006). Single-cell analysis of glucocorticoid receptor action reveals that stochastic post-chromatin association mechanisms regulate ligand-specific transcription. Mol. Endocrinol. *20*, 2641-2655.

Wagatsuma, A., Sadamoto, H., Kitahashi, T., Lukowiak, K., Urano, A., and Ito, E. (2005). Determination of the exact copy numbers of particular mRNAs in a single cell by quantitative real-time RT-PCR. J. Exp. Biol. *208*, 2389-2398.

Warren, L., Bryder, D., Weissman, I., and Quake, S. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc. Natl. Acad. Sci. USA *103*, 17807-17812.

Wilkie, G., Shermoen, A., O'Farrell, P., and Davis, I. (1999). Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei. Curr. Biol. *9*, 1263-1266.

Zenklusen, D., Larson, D., and Singer, R. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263-1271.

Chapter II

The *Drosophila* Hox complex miRNA miR-iab-4-5p does not have major effects on expression of the evolutionarily conserved Hox gene target *Antennapedia*

ABSTRACT

The discovery of microRNAs (miRNAs) has resulted in a major expansion of the number of molecules known to be involved in gene regulation. Elucidating the functions of animal miRNAs has posed a significant challenge, as their target interactions do not adhere to simple rules. Of the thousands of known animal miRNAs, very few miRNA:mRNA target interactions have been validated in their biological contexts. Here I present evidence that the miRNA miR-iab-4-5p does not appear to have any major effects on the expression of the Hox gene *Antennapedia* during embryogenesis, despite the presence of several high confidence miR-iab-4-5p target sites in its 3'UTR. It has also been previously shown that this miRNA appears completely dispensable for normal development in *Drosophila*. Taken together these observations are significant, and they indicate that many of the predicted miRNA:target interactions might not be biologically relevant, or at least not developmentally important. Similarly, the effects of most miRNAs might be so subtle that mutants only show detectable phenotypes in specific contexts, such as stressful environmental conditions or in a background containing multiple miRNA mutations, making detailed analyses very difficult.

INTRODUCTION

MicroRNAs (miRNAs) are a recently discovered class of biological molecules that have greatly expanded our knowledge concerning post-transcriptional regulation of gene expression. Since the discovery of these small RNA regulatory molecules, many of the proteins involved in their biogenesis, transport, and regulatory functions have been characterized (Winter et al., 2009). While much remains to be elucidated concerning the protein components of the miRNA pathway, the larger "black box" is that of miRNA:mRNA target interactions.

The first major discoveries in the field of miRNA research were made through careful analyses of genetic mutations in *C. elegans* with obvious phenotypes (Lee et al., 1993; Wightman and Ruvkin., 1993; Reinhart et al., 2000), and these studies gave rise to the first ideas concerning how miRNAs might interact with their target genes. Subsequent systematic mutational analyses of established miRNA target sites, as well as the use of synthetic miRNA targets, provided researchers with the first set of target pairing "rules" for animal miRNAs (Doench and Sharp, 2004; Kiriakidou et al., 2004; Kloosterman et al., 2004; Brennecke et al., 2005). These "rules" were crucial for the development of a number of computational algorithms which allow researchers to predict potential mRNA targets for a given miRNA, or conversely, potential miRNAs which target a specific mRNA (Rajewsky, 2006). However, due to the relative lack of functional data for validated miRNA:mRNA target interactions, these algorithms were largely based on small training sets in combination with the aforementioned target pairing "rules". In general, these algorithms produced lists of hundreds or even thousands of targets for a typical miRNA, of which a very small number have ever been experimentally validated.

The majority of experimental "validations" of computationally predicted miRNA:mRNA interactions are performed in cell culture (e.g., luciferase reporter assays) or utilize transgenic over-expression assays, which can indicate the potential for regulation, but do not necessarily indicate biological relevance. Also, there is a growing body of literature that adds another layer of complication, suggesting that, in general, miRNAs and their true mRNA targets are not expressed in the same cells. Instead most are expressed in complementary patterns, making in vivo validation of endogenous interactions very difficult (Stark et al., 2005, Sandberg et al., 2008).

Microarray data, both for miRNAs and their putative mRNA targets, have been used to support the validity of claims that there are hundreds of targets for each miRNA (Babak et al., 2004; Lim et al., 2005; Stark et al., 2005; Wang and Wang, 2006). Unfortunately, when mRNAs are seen to be up-regulated in miRNA mutants, or down-regulated after ectopic miRNA expression, this does not explicitly mean they are direct targets of the given miRNA. Even considering this, quantitative mass-spectrometry analysis of protein levels shows that a high proportion of predicted targets, even those which are conserved, are unaffected by changes in miRNA expression (Baek et al., 2008).

Recent studies have confounded matters even further. It appears that many (if not most) mRNAs that are coexpressed in a particular tissue with a given miRNA are generally found to be devoid of potential target sites for that miRNA (Farh et al., 2005; Stark et al., 2005; Sood et al., 2006). Systematic mutation of every miRNA in the *C. elegans* genome resulted in the surprising discovery that the vast majority had no significant phenotypes (Miska et al., 2007; Alvarez-Saavedra and Horvitz, 2010). Finally,

experiments using CLIP-seq/HITS-CLIP techniques have suggested that large proportions of predicted miRNA target sites in 3'UTRs are not bound by Argonaute proteins (Chi et al., 2009; Zisoulis et al., 2010).

The general lack of functional data, as well as a dearth of knowledge concerning concurrent expression of specific miRNAs and their putative targets at a cellular level, have led to a situation where most of the proposed miRNA:target interactions found in databases are conjectural. Additionally, many of the aforementioned studies revealed that target pairing "rules" are more complicated than previously assumed and that different miRNAs may have different sets of target pairing rules. Taken together, these data imply that computational prediction may only ever be a good indicator of the target landscape for a very limited number of miRNAs and targets.

Despite the ambiguity involved in computational predictions, there are indications that Hox genes are likely to be an important class of miRNA targets. Evidence for post-transcriptional regulation of Hox gene expression has been accumulating for some time. Discrepancies have been found between transcript and protein levels for the mouse Hox gene *Hoxb4* in the posterior neural tube (Brend et al., 2003), *Hoxc6* in the chick hindlimb (Nelson et al., 1996), and the *Sex combs reduced* ortholog in the first thoracic segment of *Porcellio scaber* (Abzhanov and Kaufman, 1999). It is currently unknown whether these discrepancies are due to miRNA regulation or to other mechanisms such as localized protein instability.

In contrast to the modest number of documented cases of post-transcriptional Hox regulation, a large fraction of Hox genes have been predicted *in silico* to be direct targets

of miRNAs in both vertebrates (Lewis et al., 2003) and invertebrates (Enright et al., 2003). One verified example of a miRNA:Hox interaction is the case of *Hoxb8* regulation by miR-196, which was shown in mice to cause endonucleolytic cleavage of the mRNA target site (Mansfield et al., 2004; Yekta et al., 2004). This target site is conserved in other vertebrate *Hoxb8* genes, and may indicate a conserved role in vertebrate axial patterning (Hornstein et al., 2005; McGlinn et al., 2009). On the basis of partial complementarity between miRNAs and 3'UTR sequences, *Drosophila melanogaster Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal A* (*abd-A*), and *Abdominal B* (*Abd-B*) transcripts have all been proposed as targets of miRNA regulation (Enright et al., 2003), and several studies have been published which tested a number of these predictions (Ronshaugen et al., 2005; Bender, 2008; Stark et al., 2008; Tyler et al., 2008; Woltering and Durston, 2008; McGlinn et al., 2009; Zhao et al., 2010).

Animal Hox complexes are relatively rich in miRNAs. For instance, most vertebrate and arthropod Hox complexes have at least 3 regions containing miRNA producing hairpins. miR-10 is the most highly conserved miRNA, not only in sequence but also in its genomic position in the complex between the Hox4 and Hox5 orthologs of most bilaterian animals. Other highly conserved miRNAs include the arthropod miR-iab-4 and miR-993 miRNAs, as well as the vertebrate miR-196 and miR-615 miRNAs, which reside in analogous positions in their respective Hox complexes, but which do not appear to be orthologous. Considering the theory that miRNAs are spawned by neighbouring genes which they can then go on to regulate (Allen et al., 2004), it is compelling to think that all these conserved Hox-cluster encoded miRNAs might be targeting nearby Hox genes, which, indeed, they have been strongly predicted to target.

In this study, I investigate one strongly predicted miRNA:Hox interaction - miR-iab-4-5p:*Antp*. I show that despite multiple well conserved putative target sites in the *Antp* 3'UTR, this predicted interaction does not appear to play a large role during embryonic development, and at most has a very subtle effect on ANTP protein levels.

RESULTS

Bithorax complex miRNAs are strongly predicted to interact with the *Antp* 3'UTR

In *Drosophila*, the bithorax complex is a large genomic region that is responsible for patterning most of the abdomen. In addition to three protein-coding Hox genes (*Ubx, adb-A, and Abd-B*), much of the intergenic regions are also transcribed, giving rise to numerous long noncoding RNAs. While the exact functions of these RNAs remain somewhat unclear, it is widely believed that early transcription through these regions is necessary to activate cryptic Hox-gene enhancers, and primes the complex for later epigenetic regulation (Bae et al., 2002; Drewell et al., 2002; Akbari et al., 2006).

It is now clear that at least two of these noncoding RNAs, *iab-4* and *iab-8*, contain hairpin precursors for three active miRNAs: miR-iab-4-5p, miR-iab-4-3p, and miR-iab-8-5p (which is sometimes referred to as miR-iab4AS or miR-iab4AS-5p). Interestingly, sense transcription of *iab-4* through the hairpin locus generates miR-iab-4-5p and miR-iab-4-3p in approximately abdominal segments A5-A7 (Bae et al., 2002; Ronshaugen et al., 2005; Stark et al. 2008), while transcription of *iab-8* in the opposite direction produces mir-iab-8-5p in abdominal segments A8-A9 (Bender, 2008; Stark et al., 2008) (Figures 7A and B). The hairpin-encoding sequence that generates these miRNAs can be found in the genomes of all sequenced arthropods in a conserved position between the Hox9 (*abd-A*) and Hox10 (*Abd-B*) orthologs (Figures 7A and 7B). Except for a few minor changes in non-essential nucleotides, the sequences of both arms of the hairpin are completely conserved (Figure 7B). These miRNAs have been shown several times to be present throughout embryogenesis via cloning and Northern analyses (Bender, 2008;

Stark et al., 2008; Tyler et al., 2008), and they have all been strongly predicted to interact

with the 3'UTR of one or more of the Hox genes. A number of recent studies have

presented evidence supporting several of these interactions: miR-iab-4-5p:*Ubx*

(Ronshaugen et al., 2005; Tyler et al., 2008; Thomsen et al., 2010), miR-iab-8-5p:*Ubx*

(Bender, 2008; Stark et al., 2008; Tyler et al., 2008), and miR-iab-8-5p:*abd-A* (Stark et

al., 2008; Tyler et al., 2008).

In addition to the aforementioned interactions, it has also been predicted that miR-

iab-4-5p and miR-iab-8-5p might also bind to the 3'UTR of the more distantly located

Hox gene *Antp* (Stark et al., 2008; Tyler et al., 2008). Through our own analyses (based

on predicted RNA duplex formation and evolutionary conservation) several conserved

potential binding sites were identified for miR-iab-4-5p and miR-iab-8-5p in the 3'UTR

of *Antp*. The most significant putative miR-iab-4-5p site is illustrated in Figures 7C and

7D. This site also corresponds to a miR-iab-4-5p:*Antp* site independently predicted by

Stark et al. (2008). While this site does not display canonical seed pairing (bases 2-7,

relative to the miRNA 5' end), it does display strong prototypical dual-helix pairing

(bases 3-9 and 13-20, relative to the miRNA 5' end), and similar miRNA:target duplexes

were predicted to form with the homologous sequences downstream of *Antp* orthologs

from many other insects (Figure 7C). Also, the specific nucleotides from these regions

that are predicted to interact with miR-iab-4-5p are extremely well conserved amongst

the *Drosophila* species, despite the relatively poor conservation of flanking regions

(Figure 7D).

We also found two other putative miR-iab-4-5p binding sites near the

aforementioned site, as well as three potential sites for miR-iab-8-5p (Figures 8C and D).

These were all predicted (by the RNA Hybrid program) to form regions of extensive secondary structure with their putative miRNA regulators, and all are conserved to varying degrees (data not shown). A fourth miR-iab-4-5p site which had been predicted elsewhere (Stark et al., 2008) is also shown, although this site displays only seed-pairing and is not well conserved (Figures 8C and D).

Putative-binding-site placement within the *Antp* 3'UTR

Interestingly, the putative miR-iab-4-5p and miR-iab-8-5p sites are all clustered near the distal end of the *Antp* 3'UTR (Figure 8C). Furthermore, it is known that *Antp* transcripts can be produced with two different 3'UTRs (Schneuwly et al., 1986). Alternative polyadenylation (polyA) signals can produce transcripts with either a short 3'UTR, which contains only the most weakly predicted miR-iab-4-5p binding site, or a longer 3'UTR, which contains the cluster of putative miR-iab-4-5p and miR-iab-8-5p binding sites (Figures 8C and D). While the mechanisms guiding polyA-site choice in the *Antp* 3'UTR are unknown at this time, this does raise the intriguing possibility that in addition to the more canonical ways a cell might use to modulate ANTP protein levels (e.g., down-regulating transcription or up-regulating a translational inhibitor) it could also express different isoforms of *Antp*, some of which would be "immune" to miRNA regulation. Indeed, this has been shown to be exactly the situation between miR-iab-4-5p and *Ubx* in the ventral nerve cord (VNC) of developing *Drosophila* embryos (Thomsen et al., 2010).

To determine the expression patterns of *Antp* transcripts containing these alternative 3'UTRs, I carried out in situ hybridization on wildtype embryos using a probe

specific to sequences from the short 3'UTR (which should be common to both isoforms) as well as a probe specific to only the longer 3'UTR (which should be "optional"). Both probes produced expression patterns which were very similar to each other (Figures 8A and B) and essentially identical to expression patterns observed using probes specific to the *Antp* coding region (data not shown). Expression is seen mainly in the ectoderm of thoracic segments T2 and T3, and in the VNC in thoracic segments T2 and T3, as well as in abdominal segments A1 through A7 (Figures 8A and B). Simultaneous, high-resolution Fluorescent in situ Hybridizations (FISH) using both probes were also carried out, although because both probes require Tyramide amplification to produce strong signals there was significant cross-talk between the channels. This made interpretation of the patterns difficult, however, in these stains it also appeared that usage of the short and long isoforms was spatially similar (data not shown).

These expression patterns are consistent with the observation that *Antp* 3'UTR usage does not appear to be linked to promoter choice, as transcripts from both the *Antp* P1 and P2 promoters (which do have different expression patterns) seem to contain both 3'UTRs in roughly equal proportions (Schneuwly et al., 1986). Therefore, I concluded that if down-regulation is occurring between miR-iab-4-5p or miR-iab-8-5p and *Antp* transcripts with the long 3'UTR, it would be occurring against a backdrop of "miRNA-immune" *Antp* transcripts, which would make interpretation of results difficult.

Coexpression of *Antp* and the *iab-4* primary transcript

To investigate any possible interactions between *Antp* and miR-iab-4-5p, I performed FISH on *Drosophila* embryos with probes against the *iab-4* and *Antp*

transcripts (Figures 9A-D). The *Antp* probe was specific to transcripts from the P1

promoter of the *Antp* gene (Bermingham et al., 1990). The *iab-4* probe was specific to

sequences from the long intron of *iab-4* (Ronshaugen et al., 2005), and as such it appears

as bright nuclear dots, representing sites of nascent transcription (Kosman et al., 2004).

Unfortunately, signals from this probe do not necessarily correspond to regions of mature

miR-iab-4-5p expression, as *iab-4* itself can be expressed in either a short or long form,

of which only the long form contains the miRNA hairpin region (Ronshaugen et al.,

2005). Attempts were made to stain for more "informative" regions of the *iab-4*

transcript, although neither probes against the hairpin region of *iab-4* nor Locked Nucleic

Acid (LNA) probes against the mature miRNA produced interpretable stains, presumably

due to strong secondary structure in the target regions (data not shown). However,

previous studies investigating the endogenous effects of miR-iab-4-5p do indicate that an

active form of miR-iab-4-5p is present in roughly the same areas in which its primary

transcript is produced (Bender, 2008; Thomsen et al., 2010).

During early embryogenesis *Antp* and *iab-4* are expressed in broad, partially

overlapping, gap-gene-like patterns throughout the future trunk and abdomen (Figure

9A). However, it is unlikely that coexpression at this stage is biologically relevant, as the

Antp transcription unit is quite long and does not have the necessary time to produce full-

length mRNAs during the short nuclear division cycles of the blastoderm. However, it is

also possible that miRNAs produced during this time might have effects later on during

development, if they remain active.

During midembryogenesis (e.g., stage 11) *Antp* and *iab-4* begin to resemble their

"mature" expression patterns (Figure 9B). *Antp* is strongly expressed in the ectoderm of

the thorax, and it is also beginning to be expressed in the neuroblasts of the VNC. *iab-4* is mainly expressed in the ectoderm of abdominal segments A5-A7, and also weakly in the neuroblasts of the VNC (in roughly the same segments). It is interesting to note that unlike the other Hox genes, the transcriptional expression pattern of *Antp* appears quite "noisy" at this time, and *Antp* transcripts are often seen in regions of the embryo that do not produce ANTP protein (Figures 9B and C). This "misexpression" is most often seen in regions of the posterior abdominal ectoderm, which is also happens to be where *iab-4* is expressed during this stage. Therefore, it is possible that one of the functions of miR-iab-4-5p is to repress translation of these aberrant *Antp* transcripts during midembryogenesis. Unfortunately, the rare and stochastic nature of these events precluded further investigation in vivo.

During germband retraction these patterns refine, with the major domain of coexpression corresponding to the VNC of abdominal segments A3-A7 (Figure 9D). Both genes also appeared to be coexpressed in the in the dorso-lateral ectoderm of the abdomen, although this expression is largely complementary on a cell-by-cell basis (data not shown).

During late-embryogenesis (i.e., stage 14 and later), ANTP protein levels correlate well with the *Antp* transcript pattern (data not shown), and again the main region of coexpression between ANTP and *iab-4* is in the VNC of abdominal segments A3-A7 (Figure 9E). Looking closely at the expression patterns in the VNC, I saw that while there are significant areas of overlap, in general, areas of high ANTP correspond to areas of low *iab-4* expression, and vice versa, consistent with the potential for negative regulation (Figure 9F).

ANTP levels are largely unchanged in a miR-iab-4/8 hairpin deletion background

To determine whether significant endogenous regulation of ANTP by miR-iab-4-5p occurs in vivo, I took advantage of a *Drosophila* strain in which the hairpin that produces the miR-iab-4 and miR-iab-8 miRNAs had been precisely deleted (Δ*miR-iab-4/8*) (Bender, 2008). Conveniently, as the rest of the primary transcript is still produced normally in these mutants, I could still stain for the *iab-4* precursor RNA to pinpoint areas that might now contain higher ANTP levels. Unfortunately, no obvious changes in ANTP protein levels were seen in the coexpressed regions throughout any stage of embryogenesis (compare Figures 9G-I to Figures 9J-L). Occasionally, Δ*miR-iab-4/8* embryos would be observed in which certain subsets of neuroblasts displayed higher ANTP levels in areas that would normally be expressing miR-iab-4-5p (Figures 9J and L, arrows) compared to similarly positioned neuroblasts in more anterior segments that never express *iab-4*. Image segmentation and subsequent protein-level determination of several of these neuroblasts did indicate slightly higher ANTP protein levels compared to their more anterior counterparts (data not shown), although the stochastic nature of these events, the difficulty involved in determining neuroblast identity, and the labor-intensive nature of the protein quantification precluded further analysis.

The long *Antp* 3'UTR also contains putative binding sites for miR-iab-8-5p (Figures 8C and D). However, the miR-iab-8 miRNAs are only produced in the most posterior abdominal segments A8 and A9 and do not overlap significantly with ANTP expression. Very weak *Antp* RNA and ANTP protein staining is visible in segments A8 and A9 (data not shown), indicating the potential for subtle regulation by the miR-iab-8

miRNAs, although no obvious expansion of the ANTP domain into segments A8 or A9 was observed in Δ*miR-iab-4/8* embryos, and this interaction was not investigated further.

Finally, to determine whether the function of miR-iab-4-5p was to buffer ANTP levels against environmental fluctuations, I subjected developing heterozygous and homozygous Δ*miR-iab-4/8* mutant embryos to several rounds of temperature shifts (Li et al., 2009) in the hopes that heat shock treatment might produce aberrant *Antp* transcripts outside regions of normal expression. Again, I observed no reproducible changes in ANTP protein levels, either between heterozygous and homozygous embryos, or between the anterior and posterior neuromeres in the VNC, even under these stressful developmental conditions (data not shown).

DISCUSSION

A growing body of circumstantial evidence and computational predictions consistently point to *Drosophila* Hox genes as one of the most significant classes of miRNA targets. However, this fact appears to be somewhat at odds with experimental validations of endogenous interactions, which indicate that in vivo miRNA regulation of Hox genes is usually quite subtle, when it is detectable at all (Bender 2008, Thomsen 2010). This leads to several questions: 1) If these predicted miRNA:target interactions are functional in vivo, why are they so difficult to observe? 2) If they are not functional, why have these target sites been so well conserved? 3) Finally, is continued investigation into the function of Hox-cluster encoded miRNAs in *Drosophila* compelling, considering there has only been a single reproducible morphological phenotype uncovered so far (Bender, 2008), despite a large amount of work by several labs?

The most likely scenario appears to be that the major conserved function of Hox-cluster encoded miRNAs is to subtly modulate Hox proteins levels within subregions of the central nervous system (CNS) (referred to here as the VNC, in the case of *Drosophila*). In addition to the few validated examples of endogenous interactions (Bender et al., 2008; Thomsen et al., 2010), there are numerous strong computational predictions for interactions between several coexpressed miRNAs and Hox genes in the VNC (Enright et al, 2003; Ronshaugen et al., 2005; Stark et al., 2008, Tyler et al., 2008). There is also compelling evolutionary data: miR-10 is conserved in both sequence, genomic position, and expression pattern in the CNS of *Drosophila* and zebrafish (Wienholds et al., 2005); members of another vertebrate Hox miRNA family, miR-196, are also expressed in the CNS of zebrafish (Wienholds et al., 2005); and miR-iab-4 and

miR-iab-8 (while not orthologous by sequence to miR-196, are orthologous by their chromosomal position in the Hox complex) are also expressed in the *Drosophila* VNC. My findings indicate miR-iab-4-5p does not appear to have a large effect on ANTP levels in the VNC during embryogenesis, although I did uncover some evidence for subtle regulation within certain neuroblasts. It is also important to note that subtle changes in protein levels in the CNS would likely lead to subtle neurological phenotypes, which are extremely difficult to pinpoint and characterize. Therefore, simply because they have not been uncovered in miRNA mutant backgrounds does not mean these phenotypes do not exist, or that they are not relevant in the wild.

Another scenario is that the major function of these miRNAs is to suppress "endogenous noise" within gene networks (i.e., to "impart robustness"), which in this case would mean suppressing Hox gene expression outside of their proper domains. One study indicates that, in general, *Drosophila* miRNAs and their most strongly predicted targets are almost always expressed in complementary patterns (Stark et al., 2005), precluding the possibility of dramatic phenotypes in miRNA mutants. Furthermore, most miRNA:Hox interaction "validations" only demonstrate that misexpression of miRNAs in regions where they are not normally found can result in homeotic transformations (Ronshaugen et al., 2005; Stark et al., 2008; Tyler et al., 2008), and not that these interactions play large roles during development. Consistent with this, miR-iab-4 miRNAs (which are extensively coexpressed with several proposed Hox targets in the VNC) appear completely dispensable for normal *Drosophila* development, and it is only upon loss of the miR-iab-8 miRNAs (which are only expressed in the most posterior segments of the VNC) that one sees visible adult phenotypes (male sterility and genital

malformation) (Bender, 2008). This is presumably due to the expansion of *Ubx*, *abd-A*, and possibly *Antp* expression-domains into the most posterior segments where they are not normally found. This theory is also consistent with the complementary expression patterns we have observed for the predicted miRNA:target pairs miR-10-5p:*Scr*, miR-10-3p:*Abd-B*, and miR-iab-8-5p:*Antp* (data not shown). I also report here the potential for miR-iab-4-5p having a "clean-up role" in repressing the translation of spuriously expressed *Antp* transcripts in the abdominal ectoderm during germband elongation (where ANTP protein is never observed), or, in other words, buffering the endogenous noise in *Antp* expression.

Another closely related scenario to the one just described is that the major function of miRNAs is to suppress "exogenous noise" in gene networks; in other words, noise caused by environmental variations and stress. This attractive (yet very difficult to demonstrate) hypothesis has been shown to be true in at least one case in *Drosophila* (Li et al., 2009). My own attempts to induce aberrant ANTP protein accumulation in Δ*miR-iab-4/8* mutants via temperature fluctuations were unsuccessful, although the stresses encountered by a developing embryo in the wild are certainly too numerous to exhaustively test in a laboratory context. Even if this were the sole function of miRNAs, one could imagine it might account for the observed conservation of both miRNAs and their target sites during evolution, especially considering the highly deleterious nature of Hox gene misexpression.

When considering miRNA regulation of Hox genes in the CNS, an added layer of complexity must also be considered. *Ubx, Antp, abd-A, and Abd-B* all appear to produce both short and long forms of their 3'UTRs in various sub-regions of their expression

patterns. This results in a mixture of miRNA-sensitive and -insensitive Hox transcripts, which can be used by cells to modulate Hox protein levels in the presence of static miRNA regulation (Thomsen et al., 2010). This also appears to be a mechanism utilized by tissues and cells in general (Sandberg et al., 2008). While very intriguing, this further complicates efforts to characterize endogenous miRNA:target interactions, as down-regulation may be masked by translation from miRNA insensitive transcripts, or buffered by alternative polyA-site usage.

Still, it is hard to believe these Hox-cluster miRNAs and their putative Hox target-sites could be so well conserved if they do not play fairly important roles during development. It is possible that the gain and loss of Hox miRNA target-sites, or changes in the spatio-temporal expression of Hox-targeting miRNAs, have played an important role in the evolution of Hox gene function during animal evolution. It will be exciting to discover the roles that these miRNA genes have been playing, some of which have been residents of animal Hox clusters for more than 500 million years.

MATERIALS AND METHODS

Fly stocks and embryo collection

Drosophila stocks were raised between 18°C and 25°C on cornmeal/molasses food, according to standard procedures. Δ*miR-iab-4/8* / *TM3,Ubx-LacZ* flies were a gift from W. Bender (Bender, 2008). Unless otherwise noted, all embryos shown are of the genotype *w*^{1118} (Bloomington Stock Center, http://flystocks.bio.indiana.edu).

Embryos were raised, collected, and fixed as reported elsewhere (Kosman et al., 2004; http://biology.ucsd.edu/~davek). Briefly, embryos were collected on apple juice plates and dechorionated in 50% bleach for 3 min. Embryos were then fixed in scintillation vials for 25 min with vigorous shaking, in a solution of 5 ml heptane (organic phase) and 5 ml fixation buffer (aqueous phase). Fixation buffer consisted of the following: 8% ultrapure, methanol-free formaldehyde (e.g., Polysciences #04018); 1X phosphate buffered saline (PBS); and 50 mM ethylene glycol tetraacetic acid (EGTA), pH 8.0. To devitellinize the embryos, the lower aqueous phase was removed, 8 ml of methanol were added, and embryos were shaken vigorously for 1 min. The upper organic and most of the lower methanol phases were then removed, and the embryos were washed 3 times with fresh methanol, and 5 times with ethanol. Fixed embryos were then stored in ethanol at -20°C.

Reagents and in situ hybridizations

Unless otherwise noted, antisense RNA probes were created by cloning appropriate PCR fragments into the pCR II vector (Invitrogen, K207040), and hapten-

tagged probes were transcribed in vitro, as described elsewhere (Kosman et al., 2004). Hapten tags included: biotin (BIO), digoxigenin (DIG), fluorescein (FITC), and dinitrophenyl (DNP). When probes are referred to as "unfragmented", this means that chemically fragmented probes did not produce high-quality stains, so unfragmented probes were used. Probes used in this study: *Antp*-distal-DIG (5'-CAAATGGCGTCA AAATCCATTGC-3' and 5'-TCCATTCATGCGATTAGTGTTC-3'), unfragmented; *Antp*-proximal-DIG (5'-GATCGACGGAGTCTACCCAC-3' and 5'-GCGCTAGG ATTGCTACAAAC-3'), unfragmented; *iab-4*-DIG and *iab-4*-DNP (5'-ACCACAAG AAGGAGCAGTCG-3' and 5'-GCACTCTCACCTACACGAATGC-3'); *Antp*-P1-BIO was made as described elsewhere (Bermingham, et al., 1990); *LacZ*-FITC was transcribed from the pBS-LacZ plasmid.

Unless otherwise noted primary antibodies were obtained from Roche, and used at 1:800 dilutions. Primary antibodies used in this study: mouse anti-ANTP (Developmental Studies Hybridoma Bank, 4C3 concentrate; 1:300 dilution); mouse anti-BIO; sheep anti-DIG; rabbit anti-DNP; guinea pig anti-FITC; Alkaline Phosphatase-conjugated sheep anti-DIG (sheep anti-DIG-AP); and Horseradish Peroxidase-conjugated sheep anti-DIG (sheep anti-DIG-HRP).

Alexa Fluor labeled secondary antibodies were obtained from Molecular Probes/Life Technologies, and are as follows: donkey anti-sheep Alexa647; donkey anti-mouse Alexa 488; donkey anti-rabbit Alexa 555; and goat anti-guinea-pig Alexa 594. All secondaries were used at 1:400 dilutions.

Nonfluorescent in situ hybridizations were carried out according to standard procedures, and were visualized using sheep anti-DIG-AP and the substrate NBT/BCIP

(Roche, 11681451001). Standard fluorescent in situ hybridizations were carried out as described elsewhere (Kosman et al., 2004; http://biology.ucsd.edu/~davek). Simultaneous RNA and protein detections were carried out via a modified standard FISH protocol with acetone used instead of Proteinase K for permeabilization, to preserve epitope integrity (Nagaso et al., 2001; Paré et al., 2009).

Detection schemes

To detect wildtype (w^{1118}) expression patterns of *Antp* and *iab-4* transcripts, the detection scheme was as follows: *iab-4*-DIG > sheep anti-DIG > donkey anti-sheep Alexa 647; *Antp*-P1-BIO > mouse anti-BIO > donkey anti-mouse Alexa 488.

To detect wildtype (w^{1118}) expression patterns of ANTP protein and *iab-4* transcripts the detection scheme was as follows: *iab-4*-DIG > sheep anti-DIG > donkey anti-sheep Alexa 647; mouse anti-ANTP > donkey anti-mouse Alexa 488.

To visualize ANTP protein levels in Δ*miR-iab-4/8* / *TM3,Ubx-LacZ* embryos, the detection scheme was as follows: *iab-4*-DNP > rabbit anti-DNP > donkey anti-rabbit Alexa 555; mouse anti-ANTP > donkey anti-mouse Alexa 488; *Antp*-distal-DIG > sheep anti-DIG-HRP > Cy5 Tyramide amplification; *LacZ*-FITC > guinea pig anti-FITC > goat anti-guinea-pig Alexa 594. *LacZ* staining was used to differentiate heterozygous and homozygous miRNA deletion embryos.

ACKNOWLEGEMENTS

I thank W. Bender for the generous gift of the miRNA-deletion *Drosophila* stock.

I also thank Derek Lemons for experimental advice and help with this chapter.

Chapter II, in part, is currently being prepared for submission for publication of

the material. **Lemons, D., Paré, A., and McGinnis, W.** Drosophila Hox complex

miRNAs do not have major effects on expression of evolutionarily conserved Hox gene

targets (in preparation). I was the primary investigator and author of the material

presented in this chapter.

Figure 7. miRNAs of the bithorax complex, and conservation of a putative miR-iab-4-5p target site in the *Antp* 3'UTR

(A) A simplified version of the *Drosophila* bithorax complex, showing the three protein-coding Hox genes (*Ubx, abd-A*, and *Abd-B*) and two noncoding transcripts (*iab-4* and *iab-8*) from which at least three active miRNAs are processed. Also shown is the upstream *Antp* gene. **(B)** The hairpin region of the *iab-4* primary transcript from which miR-iab-4-5p and miR-iab-4-3p are processed, and alignments of the homologous regions from selected arthropods. Blue lines indicate the sequences of the most common isoforms of each miRNA. Orange shading indicates regions of perfect conservation. Shown below are the two miRNAs (miR-iab-8-5p and miR-iab-8-3p) that are produced via antisense transcription through the same locus by the *iab-8* primary transcript. **(C)** Putative duplexes between miR-iab-4-5p and 3'UTR sequences from *Antp* orthologs in selected insects. Colored nucleotides indicate weaker G:U base pairings. **(D)** Putative miR-iab-4-5p target sites in the 3'UTRs of *Brachyceran Antp* orthologs are conserved within neighboring regions of relatively poor conservation. Shown is an alignment of sequences from the 3'UTRs (or 3' of the stop codon in putative 3'UTRs) of *Antp* orthologs in *Brachycerans*. Highlighted in green are nucleotides that are >90% conserved. Outlined in red are nucleotides that can pair with miR-iab-4-5p.

A

Antp // Ubx abd-A

5' iab-4 3'

3' iab-8 5'

Abd-B

B

miR-iab-4-5p   miR-iab-4-3p

5' ┣━━━━━━━┫ 3'   5' ┣━━━━━━━━━━━━┫ 3'

```
D.mel  ACGUAUACUGAAUGUAUCCUGAGUGUAU------CCUAUCCGGUAUACCUUCAGUAUACGUAA
D.vir  ACGUAUACUGAAUGUAUCCUGAGUGUAU------CCUAUCCGGUAUACCUUCAGUAUACGUAA
D.gri  ACGUAUACUGAAUGUAUCCUGAGUGUAU------CCUAUCCGGUAUACCUUCAGUAUACGUAA
G.mor  ACGUAUACUGAAUGUAUCCUGAGUGCUU------CCUAUCCGGUAUACCUUCAGUAUACGUAA
A.gam  ACGUAUACUGAAUGUAUCCUGAGUGCAU------CCUAUCCGGUAUACCUUCAGUAUACGUAA
A.aeg  ACGUAUACUGAAUGUAUCCUGAGUGCAU------UCUAUCCGGUAUACCUUCAGUAUACGUAA
C.pip  ACGUAUACUGAAUGUAUCCUGAGUGCAU------UCUAUCCGGUAUACCUUCAGUAUACGUAA
B.mor  ACGUAUACUGAAUGUAUCCUGAGUGCGCGA--UUCUUGUCCGGUAUACCUUCAGUAUACGUAA
T.cas  ACGUAUACUGAAUGUAUCCUGAGUGUUU-----UCUGUCCGGUAUACCUUCAGUAUACGUAA
A.mel  ACGUAUACUGAAUGUAUCCUGAGUGUAU-----UUCUGUCCGGUAUACCUUCAGUAUACGUAA
N.vit  ACGUAUACUGAAUGUAUCCUGAGUGUAU-----UUCUGUCCGGUAUACCUUCAGUAUACGUAA
A.pis  ACGUAUACUAAAUGUAUCCUGAGUGCAUUAUAAAACUGUCCGGUAUACCUUCAGUAUACGUAA
R.pro  ACGUAUACUGAAUGUAUCCUGAGUAAAUC----AACCAUCCGGUAUACCUUCAGUAUACGUAA
D.pul  ACGUAUACUGAAUGUAUCCUGAGCUUCA-----UUCAUUCCGGUAUACCUUCAGUAUACGUAA
I.sca  UCGUAUACUGAGUGUAUCCUGAGUGUUU-----AAUUUUCCGGUAUACCUUCAGUAUACGUAA
```

3' AUUUGCAUAUGACUUACAUAGG 5'   3' GCCAUAUGGUUGUCAUAUGCAUU 5'

┣━━━━━━━┫   ┣━━━━━━━━━━━┫

miR-iab-8-3p   miR-iab-8-5p

C

```
                    5' AA         A          CG 3'
D.mel Antp            AGGAUAUA    AGUAUAC
miR-iab-4-5p          UCCUAUGU    UCAUAUG
                    3' AG         AAG        CA 5'

                    5' AUU        G         AG 3'
M.des Antp            GGAUAUG     AGUAUAC
miR-iab-4-5p          CCUAUGU     UCAUAUG
                    3' AGU        AAG       CA 5'

                             U           U
B.mor Antp          5' UCAGGAU CA  AG UAUACGU 3'
miR-iab-4-5p        3' AGUCCUA GU  UC AUAUGCA 5'
                             U  AAG

                    5' AA        AAA       A 3'
T.cas Antp            AGGAUA     AAA      UAUACG
miR-iab-4-5p          UCCUAU     UG      AUAUGC
                    3' AG        AAGUC      A 5'

                              UUAAAA
P.hum Antp          5' UCAGGA AC        UAUACGU 3'
miR-iab-4-5p        3' AGUCCU UG        AUAUGCA 5'
                           A  UAAGUC
```

D

```
D.mel  CAAAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAAUUACAAAAAAAUAUGAAU
D.sec  CAAAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAAUUACAAAAAAAUAUGAAU
D.sim  CAAAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAAUUACAAAAAAAUACGAAU
D.yak  CAAAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAAUUAAAAAAAAAUAUCAAUU
D.ere  CAAAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAAUUAAAAAAAAAUAUCAAUUG
D.ana  UAUAAUUCGAAUUAGGUCGAAAAGGAUAUA--AAGUAUACCGAUAAAGCAAAAAUAUUAAUU
D.pse  UAAACUCGAAUUAGGUCGUAAAGGAUAUA--AAGUAUACCGAAAUCAACAAAACAAAGAAG
D.per  UAAACUCGAAUUAGGUCGUAAAGGAUAUA--AAGUAUACCGAAAUGCAACAAAACAAAAA
D.sub  UAAACUCGAGUUAGGUCGUAAAGGAUAUA--AAGUAUACCGAAAUGCAAAAAAACAAAACA
D.wil  AAAUGAAUGAAUUAGAUCUUAAAGGAUAUA--AAGUAUACCAACAAACCAAAAAAAAAAAAA
D.moj  AAGCACUUGAAUUAGUUCGUAAAGGAUAUA--AAGUAUACCGACAAAAACCAAUUGGCAACA
D.vir  AAGCACUCGAAUUAGGUCGUAAAGGAUAUA--AAGUAUACCGACAAACCAAUUGGCAACAUU
D.gri  AAGCACUUGAAUUAGGUCGUAAAGGAUAUA--AAGUAUACCGACAAAAACCAAUUGGCAAAC
M.des  AAUAAAAAAAAAACUUUCAACAUUGGAUAUG--GAGUAUACAGAAAACAAAAGAAAAAAAAA
                              ||||||||   ||||||||
miR-iab-4-5p           3' AGUCCUAUGUAAGUCAUAUGCA 5'
```

Figure 8. *Antp* has two different 3'UTRs, the longer of which contains several putative binding sites for both miR-iab-4-5p and miR-iab-8-5p

(A) A ventral view of a stage-14 embryo stained with a probe against the common proximal ("short") 3'UTR sequence. Staining is found mostly in the ventral nerve cord (roughly segments T2-A7) and in the ectoderm of thorax (roughly segments T2 and T3). **(B)** A lateral view of a stage-15 embryo stained with a probe against the optional distal ("long") 3'UTR sequence. The staining pattern is similar to that described in (A). **(C)** A schematic showing the location of putative miR-iab-4-5p (blue arrows) and miR-iab-8-5p (gray arrows) binding sites in both the short and long *Antp* 3'UTRs. Sites 1-3 and 1*-3* are ranked in order of estimated significance, based on miRNA:target secondary-structure predictions (RNA Hybrid) and evolutionary conservation. Site 4 is included due to a seed-site match (Stark et al., 2008), although it is not well conserved. **(D)** The sequence of the full "long" *Antp* 3'UTR. The two polyadenylation signals are outlined with red boxes. Putative miR-iab-4-5p sites 1-4 are outlined with blue boxes, and putative miR-iab-8-5p sites 1*-3* are outlined with dashed boxes.

A **Antp** (short 3'UTR)

B **Antp** (long 3'UTR)

Antp

Antp

C

4

AAAAAA...

4

1 3 2

500nt

AAAAAA...

1* 3* 2*

▼ putative miR-iab-4-5p site

▲ putative miR-iab-8-5p site
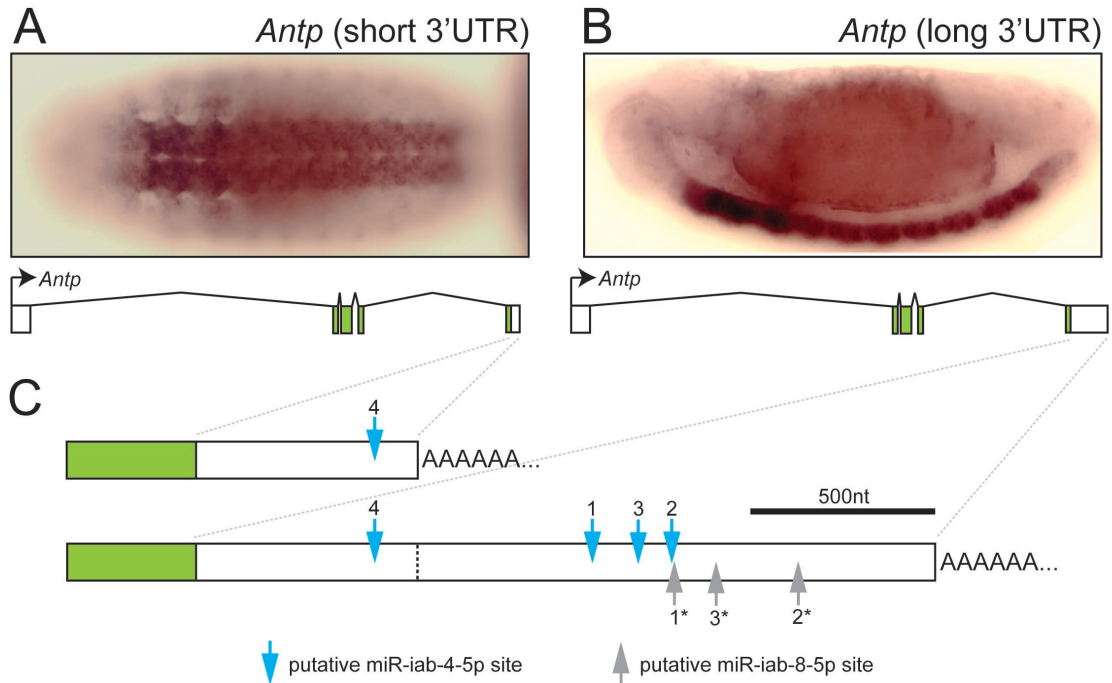
D

```
GATCGACGGAGTCTACCCACTTAAATGAAATTTCTATCTAAATACAATTTACGTTAGTTCGGAGAGCGCAAATGAATTTACTTCGATCCCAGAGGACTATCTAATAACTATC
CAATCCGTTGAACTTCGCGTGAACAAACCTAAACTAAACTAAACAAAGAGCAGAGCTGAGAACTCTACCTACAACTTAGTTAATTGTTATTATTTTCTACTTATTATTTAAT
TGTACACGAAAGGCAAGTGGGGAAAGCGAAATAAGATTAACGTAAAGATAGCGATTACGATAAAGATACAAGTAAAGCGTAAAACTCAAACAAAACCAACTCATGTGACCTC
AGATCTAAATAAGCTATATTTAACTATAATGCATATATATATACACATAAATATATGGATAACTATAAATGATACCAAGTAAAGCTAAAGGCAAGGAGTTATATATAAATAA
ATATATATGAAGCATATATAATGTAACATTAGATCTⒷACGCGTCATAAGTACTATACGAⒷTTAACTTATATATACACCCCAGCATAAACCCTAAAACTAAACCTAAACATTAAA
CTAAATCAATGTTTGTAGCAATCCTAGCGCAAAAATATAAⓇAATAAAATCCAATAAATAAAAATAAAⓇAAACAAATGGCGTCAAAATCCATTGCATGTTGGTTCATAAAACCTA
TACATTTTTCATAACTCTGAACATGATAACAGAAAACTTTGACCTAAGTGAATGTCGCACTTTTAGACAAAGAAATACCAAAACTACGAAAGAAGCGTTGCTTAAAGTGAAA
TTAACGTTTTACACATACAATAAGAGTAAAACCTATAAACTTGCAGATGCTTAACTATAAATAGAAGAACTCGCAAGAGATTGGCCAACTTAAAAAAAAAATAGATGTATAT
TTCTACGACAATTCAACTTTCAAAAACTCGCAAGTGGATATTTATGCTTAACCAATTTGAGAGTTCCCCTTTTCTTTGGCTTCAATACCCATTATATCTGTATTTTTATTTG
TTTGTAATTTCTTGTGCAATTTTTAGTTCTTGCAAAAAAACAAAATTCGAATTAGGTCGAⒷAAAGGATATAAAGTATACCGⒷAATTACAAAAAAAATATGAATTGGCAAGTAAGG
AGGAAAAATAAAAATTGTTAACAGAGAATCATATGTATAAATAATGAATTCTAACAAAACCCGTGTAAAGTAGTAAATⒷTGAAATGCATTATTATACGAGⒷAATGTAGGGAATC
CATTTTGAAAGAGCAACCGAACGATTTAAATATGAATATTTCCCTAACAAⒷCTATATTAATGTATGTGTⒷACGTAAⒷCTTAAATCATTTTCCACGCCATCTGTGGAAATCCATA
TACCAAAGTCAAAGGAACGAAGGAGAAAGAAAAAGGAGGACGGAATGGCAⒷAGTATCATATGTTTATATGTAGAⒷTATATATATTTAAACAAGCCTAATACAAAAC
ATATAACTTTATAGAGCGTTTCGTTTGTAAATTCCCCAGAAATCCCGTTTACCTCCCCAGCCCGAATCCCAATCCCAGGTGAAAAGACTTGTGATTTGCAATAGAACCGAAT
AGTCAAGAAAAAAAAACTTTACGAAGTATTGGCTAAGCAACATTGAGAGCAAAATⒷTCAACTCAATCCAGATACGTAACⒷTTTCGGCCTCATTCTGTAAGAAACTAATTATTTA
AGTTTCAATTATGAAAATTATTAAAAAAAAAAAACGAACAAAAAAAAAACGTAAATATAAAAAAGTAGAAATTCACAAGAAACAAAATTCGAAAGCGTTGATTAAAATATATAT
ATGTGAAACAAAAACAAAAAACATTAAACAAACCCAAACATTTGTCGCAAATTGGAACGAAATGTGGAAGCTAAATAATTTGTTGTAAATAAAATTTAAACTTTAGTGTAAA
GATGAGAACGGAGAATGCAGTGAAAGGGCTGCAAAGCGGAAAGCATAAAGAAAATTATAACTAAATTACAAATGAGATTTTGTTTTTATTTGTAATATTTATⓇAATAAAAATA
TAGTATTTAAAAATTTⓇATAACGAAGCGGTTATTATTAGGAAGGGTTTTTATTTCAGTGAATCCTATTTGTCTGGAGCTTCTAAGTGTGAAGGTAAGTATTTTAAATTTAATT
ATTTTGTATTTGAAGTAATCATTTTCTGTTTATCCAGGTATTTTTACTGAACACTAATCGCATGAATGGA
```

Figure 9. *iab-4* **is coexpressed with** *Antp* **throughout embryogenesis, but does not appear to play a large role in repressing translation of endogenous transcripts**

(A-D) *Antp* and *iab-4* are coexpressed throughout embryogenesis. Fluorescent in situ Hybridization was carried out using probes specific for transcripts from the *Antp* P1 promoter and the long intron of *iab-4*, which are shown in magenta and green, respectively. Embryos are oriented anterior towards the left and dorsal towards the top. **(A)** A blastoderm-stage embryo. Colored lines indicate the extents of the gap-gene-like expression patterns for each gene. **(B)** The trunk of a stage-11 embryo. Spots of *Antp* "misexpression" in the abdominal segments (indicated with a white dashed box) sometimes appear in regions of *iab-4* expression. **(C)** A close-up view of the area indicated in (B). Sites of nascent transcription of the *Antp* gene (arrows) are often observed in the abdominal ectoderm, within regions of *iab-4* expression. **(D)** A stage-14 embryo, illustrating the "mature" expression patterns of *Antp* and *iab-4* in late-stage embryos. *Antp* expression is found mainly in the ectoderm of the second and third thoracic segments, and in segments T2-A7 of the ventral nerve cord (VNC); *iab-4* expression is found mainly in the lateral ectoderm and in the VNC of abdominal segments A4-A7. **(E and F)** ANTP protein and *iab-4* are coexpressed in neuromeres A3-A7 in the developing VNC. ANTP protein and *iab-4* transcripts are shown in magenta and green, respectively. **(E)** The VNC of a stage-14 embryo. The extent of *iab-4* expression is indicated with a green line. **(F)** A close-up view displaying the largely complementary nature of ANTP and *iab-4* coexpression in the neuromeres. **(G-L)** ANTP expression appears largely unchanged in embryos homozygous for a *miR-iab-4/8* hairpin deletion when compared to heterozygous embryos. **(G)** ANTP expression in neuromeres A1-A4 of a stage-14 wildtype embryo (Δ*miR-iab4/8/TM3,Ubx-lacZ*). **(H)** *iab-4* expression in the same region as (G). **(I)** A merged image of (G) and (H). **(J)** ANTP expression in neuromeres A1-A4 of a stage-14 homozygous miRNA-deficient embryo (Δ*miR-iab-4/8*). **(K)** *iab-4* expression in the same region as (J). **(L)** A merged image of (J) and (K). Arrows indicate cells which appear to have higher levels of ANTP, compared to similar cells in more anterior neuromeres (compare left and right arrows).
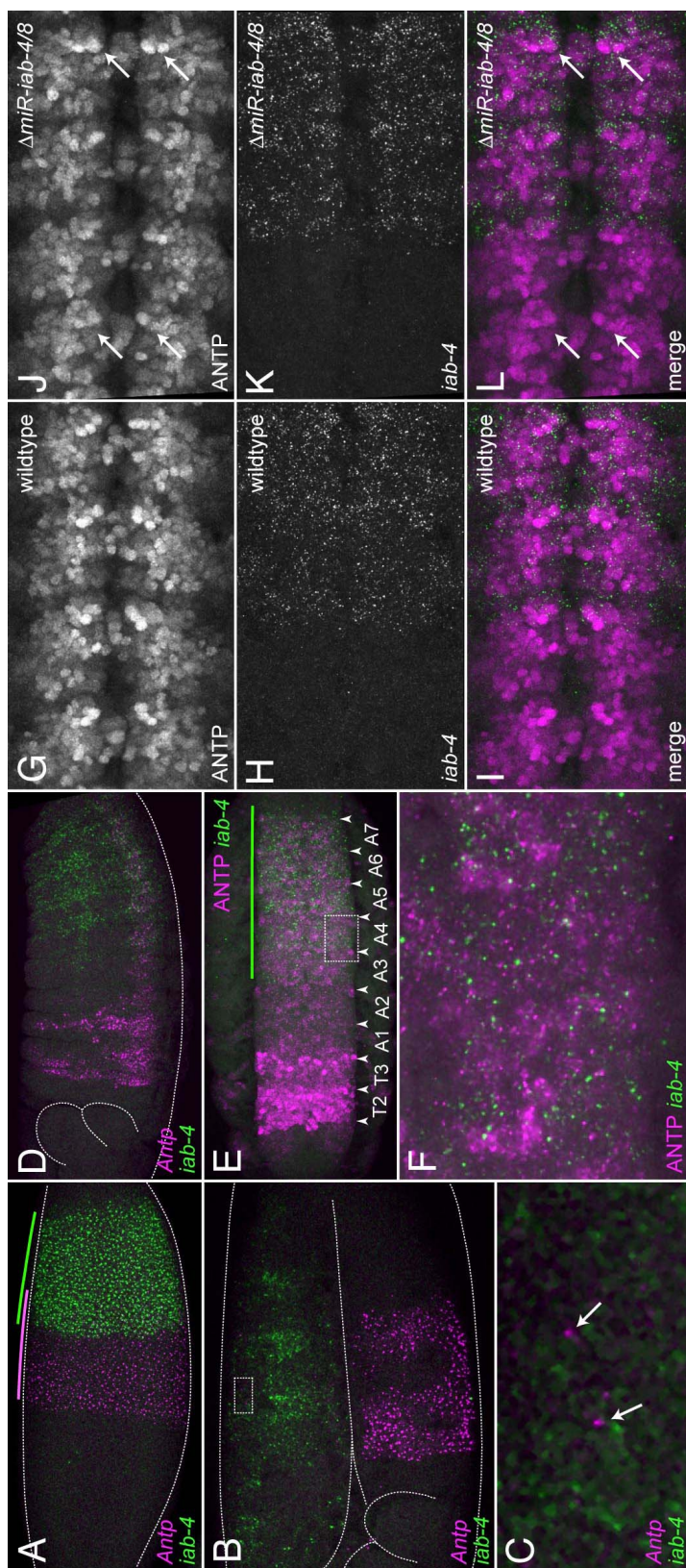
REFERENCES

Abzhanov, A., and Kaufman, T. (1999). Novel regulation of the homeotic gene Scr associated with a crustacean leg-to-maxilliped appendage transformation. Development *126*, 1121-8.

Akbari, O., Bousum, A., Bae, A., and Drewell, R. (2006). Unraveling cis-regulatory mechanisms at the abdominal-A and Abdominal-B genes in the Drosophila bithorax complex. Dev. Biol. *293*, 294-304.

Allen, E., Xie, Z., Gustafson, A., Sung, G., Spatafora, J., and Carrington, J. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat. Genet. *36*, 1282-90.

Alvarez-Saavedra, E., and Horvitz, H. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Curr. Biol. *20*, 367-73.

Babak, T., Zhang, W., Morris, Q., Blencowe, B., and Hughes, T. (2004). Probing microRNAs with microarrays: tissue specificity and functional inference. RNA *10*, 1813-9.

Bae, E., Calhoun, V., Levine, M., Lewis, E., and Drewell, R. (2002) Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA *99*, 16847-52.

Baek, D., Villen, J., Shin, C., Camargo, F., Gygi, S., and Bartel, D. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Bender, W. (2008). MicroRNAs in the Drosophila bithorax complex. Genes Dev. *22*, 14-9.

Bermingham, J., Martinez-Arias, A., Petitt, M., and Scott, M. (1990). Different patterns of transcription from the two Antennapedia promoters during Drosophila embryogenesis. Development *109*, 553-66.

Brend, T., Gilthorpe, J., Summerbell, D., and Rigby, P. (2003). Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. Development *130*, 2717-28.

Brennecke, J., Stark, A., Russell, R., and Cohen, S. (2005). Principles of microRNA-target recognition. PLoS Biol. *3*, e85.

Chi, S., Zang, J., Mele, A., and Darnell, R. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-8.

Doench, J., and Sharp, P. (2004). Specificity of microRNA target selection in translational repression. Genes Dev. *18*, 504-11.

Drewell, R., Bae, E., Burr, J., and Lewis, E. (2002). Transcription defines the embryonic domains of cis-regulatory activity at the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA. *99*, 16853-8.

Enright, A., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. (2003). MicroRNA targets in Drosophila. Genome Biol. *5*, R1.

Farh, K., Grimson, A., Jan, C., Lewis, B., Johnston, W., Lim, L., Burge, C., and Bartel, D. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-21.

Hornstein, E., Mansfield, J., Yekta, S., Hu, J., Harfe, B., McManus, M., Baskerville, S., Bartel, D., and Tabin, C. (2005). The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. Nature *438*, 671-4.

Kiriakidou, M., Nelson, P., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. Genes Dev. *18*, 1165-78.

Kloosterman, W., Wienholds, E., Ketting, R., and Plasterk, R. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res. *32*, 6284-91.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. Science *305*, 846.

Lee, R., Feinbaum, R., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-54.

Lewis, B., Shih, I., Jones-Rhoades, M., Bartel, D., and Burge, C. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-98.

Li, X., Cassidy, J., Reinke, C., Fischboeck, S., and Carthew, R. (2009). A microRNA imparts robustness against environmental fluctuation during development. Cell *137*, 273-82.

Lim, L., Lau, N., Garrett-Engele, P., Grimson, A., Schelter, J., Castle, J., Bartel, D., Linsley, P., and Johnson, J. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-73.

Mansfield, J., Harfe, B., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A., Ruvkun, G., Sharp, P., Tabin, C., and McManus, M. (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat. Genet. *36*, 1079-83.

McGlinn, E., Yekta, S., Mansfield, J., Soutschek, J., Bartel, D., and Tabin, C. (2009). In ovo application of antagomiRs indicates a role for miR-196 in patterning the chick axial skeleton through Hox gene regulation. Proc. Natl. Acad. Sci. USA *106*, 18610-5.

Miska, E., Alvarez-Saavedra, E., Abbott, A., Lau, N., Hellman, A., McGonagle, S., Bartel, D., Ambros, V., and Horvitz, H. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genet. *3*, e215.

Nagaso, H., Murata, T., Day, N., and Yokoyama, K. (2001). Simultaneous detection of RNA and protein by in situ hybridization and immunological staining. J. Histochem. Cytochem. *49*, 1177-1182.

Nelson, C., Morgan, B., Burke, A., Laufer, E., DiMambro, E., Murtaugh, L., Gonzales, E., Tessarollo, L., Parada, L., and Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. Development *122*, 1449-66.

Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009). Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr. Biol. *19*, 2037-42.

Rajewsky, N. (2006). microRNA target predictions in animals. Nat Genet *38 Suppl.*, S8-13.

Reinhart, B., Slack, F., Basson, M., Pasquinelli, A., Bettinger, J., Rougvie, A., Horvitz, H., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-6.

Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E. C. (2005). The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. Genes Dev. *19*, 2947-52.

Sandberg, R., Neilson, J., Sarma, A., Sharp, P., and Burge, C. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science *320*, 1543-7.

Schneuwly, S., Kuroiwa, A., Bumgartner, P., and Gehring, J. (1986). Structural organization and sequence of the homeotic gene Antennapedia of Drosophila melanogaster. EMBO J. *5*, 733-9.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proc. Natl. Acad. Sci. USA *103*, 2746-51.

Stark, A., Brennecke, J., Bushati, N., Russell, R., and Cohen, S. (2005). Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-46.

Stark, A., Bushati, N., Jan, C., Kheradpour, P., Hodges, E., Brennecke, J., Bartel, D., Cohen, S., and Kellis, M. (2008). A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands. Genes Dev. *22*, 8-13.

Thomsen, S., Azzam, G., Kaschula, R., Williams, L., and Alonso, C. (2010). Developmental RNA processing of 3'UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to miRNAs. Development *137*, 2951-60.

Tyler, D., Okamura, K., Chung, W., Hagen, J., Berezikov, E., Hannon, G., and Lai, E. (2008). Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. Genes Dev. *22*, 26-36.

Wang, X., and Wang, X. (2006). Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic Acids Res. *34*, 1646-52.

Wienholds, E., Kloosterman, W., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H., Kauppinen, S., and Plasterk, R. (2005). MicroRNA expression in zebrafish embryonic development. Science *309*, 310-1.

Wightman, B., and Ruvkin, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-62.

Winter, J., Jung, S., Keller, S., Gregory, R., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat. Cell Biol. *11*, 228-34.

Woltering, J., and Durston, A. (2008). miR-10 represses HoxB1a and HoxB3a in zebrafish. PLoS One *3*, e1396.

Yekta, S., Shih, I., and Bartel, D. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-6.

Zhao, Z., Boyle, T., Liu, Z., Murray, J., Wood, W., and Waterston, R. (2010). A negative regulatory loop between microRNA and Hox gene controls posterior identities in Caenorhabditis elegans. PLoS Genet. *6*, e1001089.

Zisoulis, D., Lovci, M., Wilbert, M., Hutt, K., Liang, T., Pasquinelli, A., and Yeo, G. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat. Struct. Mol. Biol. *17*, 173-9.

Chapter III

The function of Grainy head transcription factors in Animals and Fungi

ABSTRACT

Members of the Grainy head (GRH) family of transcription factors are crucial for epidermal barrier development and repair in all animals in which they have been studied. This appears to be a very high-level conservation of function, as the specific structural and enzymatic genes regulated by GRHs vary widely between species, depending on the specific type of epidermal barrier being formed. Interestingly, GRH-like proteins are also found in many species of fungi, organisms that lack an epidermis. To shed light on the role that GRH proteins play in Fungi, I characterized at the phenotypic and genomic level a null mutant for the GRH homolog (GHH) in the fungus *Neurospora crassa*. As a comparison dataset, I also carried out microarray analysis on late-stage *Drosophila grh* null embryos. Transcriptome analyses of *Neurospora ghh* strains indicate that while this gene does play a role in the development of the cell wall (the closest fungal analog to the metazoan epidermis) it appears to play a larger role in organismal defense and virulence. This points to an intriguing connection between physical barrier formation and detoxification/defense in the last common ancestor of Animals and Fungi.

INTRODUCTION

Grainy head (GRH) transcription factors carry out a number of important biological functions. In *Drosophila,* GRH (a.k.a. Elf-1 or NTF-1) is crucial for proper development of the epidermis and head-skeleton (Bray and Kafatos, 1991), wound-healing (Mace et al., 2005; Pearson et al., 2009; Wang et al., 2009), neuroblast proliferation (Almeida and Bray, 2005; Cenci and Gould, 2005), early-embryo patterning (Huang et al., 1995; Liaw et al., 1995), and tracheal tube morphology (Hemphala et al., 2003). However, the role of GRH in epidermal barrier formation and wound healing has received the most attention, as GRH proteins appear to carry out similar functions in all animal model systems in which they have been studied.

Drosophila grh null mutations are lethal - mutant embryos fail to develop past the embryonic/larval transition point due to their extremely fragile epidermal/cuticular barriers. They also have "grainy" and discontinuous head-skeletons, from which the mutation derives its name (Nusslein-Volhard et al., 1984; Bray and Kafatos, 1991; Ostrowski et al., 2002). These defects in cuticle and head-skeleton development are at least partly due to a requirement of GRH for the proper expression of dopa decarboxylase (encoded by *Ddc*) and tyrosine hydroxylase (encoded by *ple*), two enzymes involved in the production of reactive quinone molecules that are used in the melanization pathway during cuticle protein cross-linking (Bray and Kafatos, 1991; Mace et al., 2005). Some of the genes necessary for cuticle generation during development are also activated during the regenerative process following epidermal wounding, several of which are dependent on GRH for their expression (Mace et al., 2005; Pearson et al., 2009). In addition to these defects in cuticle formation and healing, *grh* embryos are also permeable to exogenously

applied dyes, indicating deficiencies in the integrity of the underlying epithelium as well (Kim and McGinnis, 2011).

In *C. elegans*, RNAi targeted against *Ce-Grh-1* results in embryos with a fragile and puckered hypodermis. Ce-Grh-1 was shown to bind the same consensus sequence as *Drosophila* GRH, and several of the homologous genes regulated by GRH in flies also contain nearby GRH binding sites in the nematode (Venkatessen et al., 2003). In *Xenopus*, expression of a dominant negative form of XGRHL1 leads to a malformed epidermis, partly due to lowered keratin expression levels (Tao et al., 2005). Mice have three versions of GRH, encoded by the genes *Grainyhead-like-1, -2, and -3* (*Grhl1, -2, and -3*), which are all expressed in the surface ectoderm during development (Auden et al., 2006). Knockouts of the mouse *Grhl1* and *Grhl3* genes both result in a malformed epidermis. *Grhl3* mutants display more severe phenotypes, including neural tube closure defects (Ting et al., 2003), increased permeability to exogenous dyes, severe water loss, impaired wound healing (Ting et al., 2005a), and irregular skin morphology at the cellular level (Ting et al., 2005b); *Grhl1* deficient mice display delayed coat growth, thickened paw skin, and hair loss due to poor anchoring of the hair shafts within the follicle (Wilanowski et al., 2008). Strikingly, mouse Grhl3 was also shown to bind to the same consensus site as *Drosophila* and *C. elegans* GRHs, several of which were found upstream of *Transglutaminase 1* (*TGase1*), which plays an important role in mammalian skin development (Ting et al., 2005a).

Taken together these results demonstrate a high-level (one might say "abstract"), functional conservation of GRH proteins as regulators of epidermal integrity and wound-healing in both protostome and deuterostome animals. While the consensus DNA binding

sequence of GRH proteins has been conserved between these diverse taxa (quite striking

in itself), the downstream effectors of GRH are often not homologous, but instead carry

out analogous functions suited to the specific barrier needed to be formed/healed. For

instance, the epidermal defects in *Grhl3* deficient mice are partly due to reduced levels of

TGase1, which is necessary for keratinocyte cross-linking. This enzyme plays an

analogous role to *Ddc* and *ple,* which create the reactive quinone molecules used to cross-

link chitin fibers and proteins in the *Drosophila* cuticle. This situation is reminiscent of

other cases in which high-level transcription factor function has been conserved over

great evolutionary time (e.g., *Pax6/eyeless, Nkx2.5/tinman,* and *Dlx/Dll* in eye, heart, and

limb specification, respectively), despite the drift of specific downstream effectors. In

other words, there appears to be a large amount of evolutionary "inertia" when it comes

to the use of GRH proteins for controlling physical barrier formation in animals.

GRH and the closely related LSF-type proteins (e.g., LSF/CP2 in mice) comprise

the CP2 superfamily of transcription factors, which bears little resemblance to other

known transcription factor families. While they share extensive homology throughout

their DNA binding domains, GRH and LSF proteins have diverged both in their

biological roles as well as in their modes of DNA binding. For instance, GRH proteins

bind the consensus sequence (A/T)AACCGGTTT (Venkatesen et al., 2003; Ting et al.,

2005) as dimers (Uv et al., 1994), while mammalian LSF binds the consensus site

CTGG-N_6-CTGG as a tetramer (Shirra & Hansen, 1998). A comprehensive review of

LSF family proteins is beyond the scope of this introduction, but there appears to be little

overlap between the biological functions of GRH and LSF-type proteins. For example,

mammalian LSF is a ubiquitously expressed cofactor that has been associated with transcriptional complexes mediating immunoglobulin class-switching, hemoglobin synthesis, HIV and SV40 viral expression, BMP4 expression in osteoblasts, and T-cell proliferation, to name just a few (for a review see Kokoszynska et al., 2008).

The CP2 family is specific to the opisthokont lineage (i.e., Metazoa, Fungi, and their sister-groups). With a few exceptions, all sequenced metazoan genomes possess one or more copies of both GRH and LSF-type proteins. These can be distinguished by the presence of a SAM domain (found only in LSF proteins) as well as by other characteristic residues in the DNA binding domain (see below). On the other hand, the choanoflagellate *M. brevicolis* (believed to resemble the single-celled ancestor of animals) contains only a single GRH/LSF-type protein with a SAM domain and a highly derived DNA binding domain. The nucleariid *C. owczarzaki* (a member of an independent opisthokont lineage) possesses a GRH/LSF-type protein which shares features of its DNA binding domain with both GRH and LSF proteins. Amongst the Fungi, ascomycete and zygomycete genomes possess a single GRH/LSF-type protein of uncertain characterization (certain species contain several duplicates of this gene), while basidiomycetes appear to have lost the CP2 family completely. No GRH/LSF-type proteins have yet been identified in plants or other eukaryotes. Thus, it appears that the last common ancestor of Metazoa and Fungi possessed a single GRH/LSF-type protein, which then split into distinct GRH and LSF families in the metazoan lineage (Traylor-Knowles et al., 2010).

Since the function of GRH proteins in epidermal barrier formation and wound-healing appears well conserved in triploblastic animals, it is intriguing to hypothesize

what role CP2 proteins might have been playing in the last common ancestor of Metazoa and Fungi - presumably an organism which lacked a true epidermis. To shed light on this question, I have characterized the function of the Grainyhead homolog (GHH) in the ascomycete fungus *Neurospora crassa* using microarray and phenotypic analyses. I chose *Neurospora* because it is a fairly typical representative of an ascomycete, it has a fair number of molecular tools (including simple gene-knockout procedures), and most importantly it has a fully sequenced genome. I show that loss of GHH in *Neurospora* leads to a developmental defect in cell wall degradation, as well as the down-regulation of numerous genes involved in defense and virulence. Our results point to an ancestral role for GRH-type proteins as mediators of defense, virulence, and barrier formation in the opisthokont ancestor.

RESULTS

Fungal CP2 proteins are more similar to GRH than LSF at the sequence level

Phylogenic analysis of the DNA binding domain sequences from numerous metazoan and fungal CP2-class proteins shows that metazoan GRH, metazoan LSF, and fungal CP2 proteins form three well-structured groupings (Figure 10), similar to previous reports which used gap-less alignments of the full protein sequences (Traylor-Knowles et al., 2010). However, at this resolution it does not indicate whether fungal CP2 proteins are more similar to GRH-class or LSF-class proteins.

The *Neurospora* genome contains a single member of the CP2 family (NCU06095), which we have renamed "Grainy head Homolog" (GHH). Using RT-PCR and primers specific to the predicted start and stop sites, I cloned and sequenced the full length GHH coding region and found the sequence and exon structure to be identical to that predicted by the Broad Institute *Neurospora* database (http://www.broadinstitute.org/ annotation/genome/neurospora/MultiHome.html). No splice variants were detected, although I cannot rule out the possibility of *ghh* transcripts that include additional upstream exons or alternate 3'UTRs. The predicted protein is similar in structure to both *Drosophila* and mammalian GRH-proteins, and contains a well conserved DNA binding domain as well as a putative oligomerization domain in the C-terminus, which has limited homology to the GRH dimerization domain (Figure 11A).

An alignment of the DNA binding domains from two GRH-class proteins (*D. melanogaster* GRH and *H. sapiens* Grhl1), two LSF-class proteins (*D. melanogaster* dCP2 and *H. sapiens* LSF), and *Neurospora* GHH highlights the extensive sequence

conservation throughout this domain (Figure 11B). It has been predicted that the CP2 DNA binding domain adopts a similar tertiary structure to the binding domain of p53, a protein with a very well characterized three-dimensional structure (Kokoszynska et al., 2008). Strikingly, the residues at and around positions predicted to be crucial for DNA binding, based on mapping to the p53 structure (i.e., major and minor groove contacts, zinc-binding residues, and residues involved in dimerization), are extremely well conserved (Figures 11B and 12). Furthermore, it is often these residues that are most informative when distinguishing GRH from LSF proteins at the sequence level, and in all cases these residues point to fungal CP2 proteins being more similar to GRH than to LSF proteins. For example, at positions 194-198 of the GHH protein (a region predicted to be involved in major groove interaction) the same GAERK residues are found in nearly all available GRH and fungal CP2 sequences, while the residues GADRK are found in all available LSF sequences. Similarly, at two other sites predicted to be important for DNA-binding (142-147, and 150-153) we find that the *Neurospora* residues are identical to residues found in nearly all metazoan GRH proteins, and dissimilar to those found in LSF proteins (Figures 11B and 12; residue numbers are relative to the *Neurospora* GHH DNA binding domain sequence). It should also be noted that no fungal CP2 proteins were found to contain a SAM domain, which is characteristic of LSF proteins (data not shown). These data strongly suggest that fungal CP2 proteins are likely to bind DNA in a similar manner to GRH-class (as opposed to LSF-class) proteins.

The *Neurospora* Grainy head homolog can bind to GRH sites in vitro, but does not bind LSF sites

To test this hypothesis, I synthesized full-length *Neurospora* GHH in order to characterize its DNA binding properties using gel-shift analyses. It has been shown that *Drosophila* and *C. elegans* GRH proteins can both bind strongly to a site from the *Ddc* promoter with the sequence AACCGGT (Venkatesen et al., 2003), and that the optimal binding site for murine Grhl3 is (A/T)AAACCGGTT(T/A)(T/A)(T/A) (Ting et al., 2005). Mammalian LSF has been shown to bind oligos containing the consensus sequence CTGG-N6-CTGG, but it will not interact with oligos containing a GRH-site from the *Ubx* promoter; *Drosophila* GRH, however, can weakly bind to an LSF full-site, or a CTGG half-site (Shirra and Hansen, 1998). Therefore, I tested the ability of *Neurospora* GHH to bind oligos containing sequences from the *Drosophila Ddc* promoter, including one of the following sites: the endogenous GRH site (GRH-Ddc), a mutated GRH site (GRH-mut), or a consensus GRH binding site (GRH-con) (Figure 13A). I also tested the ability of GHH to bind oligos containing sequences from the late SV40 promoter, including one of the following sites: the endogenous LSF consensus site (LSF-con), an LSF half-site (LSF-1/2), or a mutated LSF site (LSF-mut) (Figure 13A). Full-length *Drosophila* GRH was used as a comparison.

Drosophila GRH behaved as expected, binding strongly to the GRH-Ddc and GRH-con oligos, but not to the GRH-mut oligo. It also bound very weakly to both the LSF-con and LSF-1/2 oligos, but not to the mutated LSF-mut oligo (Figure 13B, right panels, black arrowheads). *Neurospora* GHH behaved in a very similar manner, albeit with a lower affinity than *Drosophila* GRH. GHH bound strongly to the GRH-con oligo and weakly to the GRH-Ddc oligo, but did not bind appreciably to any of the LSF oligos (a very weak band was perceptible in the LSF-con lane after a long exposure) (Figure

13B, left panels, black arrowheads). Considering these results, along with the DNA

binding domain sequence analyses, I would predict that fungal CP2 proteins bind DNA in

a similar manner to metazoan GRH proteins (as opposed to LSF proteins) in vivo. Thus,

it appears likely that the last common ancestor of metazoans and fungi contained a true

GRH-like protein, which gave rise to the LSF family in metazoans, and not vice versa.

Neurospora ghh **deletion strains do not produce** *ghh* **transcripts**

Neurospora strains containing precise deletions of the *ghh* locus (NCU06095)

were obtained from the Fungal Genetics Stock Center (FGSC) for both mating types:

FGSC13563 (mat A) and FGSC13564 (mat a). In addition, I reisolated multiple

independent *ghh* deletion strains using targeted homologous recombination to replace the

ghh locus with a hygromycin cassette. These all appeared indistinguishable from those

stocks obtained from the FGSC, indicating the phenotypes described below are indeed

due to a precise deletion of the NCU06095 locus. PCR amplification of a region within

the *ghh* locus verified that all strains were missing the *ghh* gene (Figure 14A).

Furthermore, RT-PCR amplification of a region of the *ghh* mRNA yielded no product

using *ghh* RNA as a template, compared to robust detection of the *ghh* transcripts in

wildtype strains (Figure 14B).

The *Neurospora crassa* **life-cycle**

Neurospora crassa is a fairly typical representation of an ascomycete fungus, and

it has a fairly simple cellular organization and life-cycle compared to most animals and

plants (for an in depth treatment on the subject, see Davis, 2000). The most visually

obvious phase of the *Neurospora* life-cycle is the asexual one: single spores (referred to as conidia) germinate on a food source and form a dense, interwoven mat of thread-like mycelia, which spread quickly over the food source to form a colony. *Neurospora* colonies exist as syncitial collections of "cells" which share a common cell wall. While there are regular divisions along length of the mycelial and hyphal axes, these divisions are not complete, and the "cells" use vigorous cytoplasmic streaming to move nutrients and other molecules throughout the colony. After about a day (and every day after that, according to a circadian rhythm) aerial hyphae grow up and away from the food source and bud off chains of new conidia. These conidial chains are quite delicate, and single spores can easily break off and disperse on air currents to found new colonies. Sexual reproduction, which occurs only during phases of nutrient and nitrogen starvation, is quite complex and interesting, but it is beyond the scope of this discussion.

Phenotypes of the *Neurospora ghh* deletion strains

Mutant *ghh* strains are viable and can be propagated asexually as homokaryonic (i.e., all nuclei in the colony are clonal) lines on minimal media. Both mating type strains can serve as males or females in sexual crosses to wildtype or *ghh* deficiency strains of the opposite mating type, indicating no defects in sexual reproduction in these mutants (data not shown). In general, the strains appear quite healthy, and in most ways are indistinguishable from wildtype (Figures 14D-F).

The *ghh* strains display a slightly altered circadian rhythm (Brody et al., 2010), develop orange pigmentation more quickly than wildtype (Figure 14C), and sometimes have paler mycelia than wildtype (data not shown). However, the most striking

phenotype is a pronounced conidial separation defect . As described earlier, during the normal course of asexual reproduction *Neurospora* generate aerial hyphae which bud off chains of spores called conidia. These chains are normally quite delicate, allowing individual conidia to detach and disperse. In the *ghh* lines, however, conidia completely fail to detach, even upon physical stress or immersion in liquid (Figures 14G-I).

This phenotype is identical to that observed in the *conidial separation* strains *csp-1* and *csp-2*, which have been investigated in some depth. Normally during development, the thick, chitinous cross-walls between conidia are digested away, until only thin connectives remain between adjacent conidia. It was shown that the *csp* strains fail to initiate this process, and the cross-walls remain intact, precluding conidial separation. This was correlated with a deficiency in cell wall autocatalytic activity, presumably due to the loss of expression of a secreted enzyme(s) such as chitinase (Selitrennikoff et al., 1974; Springer and Yanofsky, 1989).

While *csp-1* and *csp-2* have long been popular background strains for *Neurospora* researchers (they help prevent cross-contamination of stocks), the actual mutations responsible for these phenotypes were unknown for many years. Recently, it had been shown that *csp-1* maps to a zinc-finger transcription factor on chromosome 1 (Lambreghts et al., 2009). However, the identity of the lesion responsible for the *csp-2* phenotype remained unclear, except that it mapped to chromosome 7 between the genes *thi-3* and *ace-8*. As it turned out, this is precisely the region where *ghh* is located; therefore, I thought a lesion in the *ghh* gene must almost certainly be responsible for the *csp-2* phenotype.

I undertook genetic complementation tests to show that *csp-2* and *ghh* are allelic.

Different *Neurospora* strains (given they have compatible genetic backgrounds) can fuse

to form heterokaryonic colonies containing nuclei from both parental strains. This allows

doubly auxotrophic colonies to survive on minimal media, as each nucleus-type will

complement the nutritional requirements of the other, which also allows one to test for

genetic complementation at another locus. Using standard sexual crosses, both *csp-2* and

ghh strains were placed in different auxotrophic backgrounds (*inos* and *his-3*,

respectively), and conidia from each strain were combined on minimal media. As

expected, all viable heterokaryonic fusings resulted in colonies which still displayed the

csp phenotype, demonstrating that *csp-2* and *ghh* do not complement (see Materials and

Methods for details). Therefore, I conclude that *ghh* and *csp-2* are allelic, and that the

conidial separation phenotype observed in *ghh* strains is due to a loss of cell wall

autocatalytic activity, which in turn leads to a defect in cross-wall digestion.

Microarray profiling of *Neurospora ghh* mutants

To determine the genes directly and indirectly under the control of GHH in

Neurospora, I carried out microarray-based transcriptome profiling of three different

sample types: 1) MYC - actively growing pure mycelial samples; 2) AHC - aerial hyphae

and conidia from 48 hr old colonies; and 3) ALL - all cell types from 48 hr old colonies

(see Materials and Methods for details). I will only describe the results from the AHC

samples, as these are the cell-types that displayed the *csp* phenotype, and because they

are by far the most interesting (for lists of the misexpressed genes in the MYC samples

see the Appendix; the ALL samples did yield as many significant hits as the other two

cell-types and the results are not shown here; see Materials and Methods for details on the microarray and its construction).

Out of the 10,526 genes that were probed on the array, 167 were seen to be misregulated in the AHC samples of *ghh* mutants. This was at a False Discovery Rate (FDR) threshold of < 0.01 (meaning 1%, or about 2 of these would be expected to be false positives), which roughly corresponds to a greater than twofold change in expression up or down. Nearly equal numbers of genes were seen to be up- and down-regulated (84 and 83 genes, respectively).

In order to parse microarray results, researchers often use the Gene Ontology (GO) functional annotation system (http://www.geneontology.org/) to look for highly enriched classes of genes. As there does not exist a high-quality GO annotation of the *Neurospora crassa* genome, I used an alternative classification system - The Functional Catalogue (FunCat) - for which there does exist a good annotation (Ruepp et al., 2004). For the 83 genes that were seen to be significantly down-regulated, there were highly significant enrichments in only five FunCat categories, compared to the genome as a whole (Table 4, top). Three of these categories were involved in amino acid metabolism (specifically that of cysteine, phenylalanine, and tryptophan). A fourth category, "C-compound and carbohydrate transport", was comprised mainly of membrane transport proteins. The fifth highly significant category found was "Disease, Virulence, and Defense", which was comprised of genes that are involved in fungal pathogenicity, defense against other organisms, and certain stress responses. As I had carried out these microarrays in the hopes of finding some connection between the functions of GRH-like transcription factors in fungi and metazoans, I interpreted these results with that in mind.

As the cell wall is the closest analog to the animal epidermis in fungi, the appearance of the class "C-compound and carbohydrate transport" was initially most intriguing to me, due to the fact that the cell wall is comprised largely of carbohydrate polymers (i.e., beta-glucans and chitin). However, the specific down-regulated proteins from this class were either not directly connected to cell wall biogenesis, or else were of unclear function (Figure 15).

I could find no direct connections in the literature between cysteine metabolism and barrier formation in animals. However, most amino-acid metabolism networks are interlinked, and three of these genes (NCU05499, NCU09183, and NCU01402) are also part of the tryptophan and phenylalanine metabolism classes, for which there are some intriguing connections to barrier formation. Melanization reactions in *Drosophila* are used to harden and cross-link cuticular structures as well as to fight pathogenic infection, and are well-known (at least in the epidermis) to rely on GRH for activation (Mace et al., 2005). The reactive quinone molecules which chemically carry out these processes are derivatives of dopamine, which is itself a derivative of the amino acids tyrosine and phenylalanine (for a review see Tang, 2009). I believe it is possible that an ancestral role in phenylalanine regulation by GRH-like transcription factors could have been co-opted in cuticle forming animals to feed into the melanization process. As for the third amino acid-related FunCat category, "degradation of tryptophan", there is some evidence that it is a general mechanism of all cells to degrade tryptophan in response to infection, as a means to slow pathogenic growth (Zelante et al., 2009). If so, this would link it to the last category, "Disease, Virulence, and Defense".

Indeed, it was this fifth category, "disease, virulence, and defense" that I found most intriguing, due to the numerous connections between physical epidermal barriers and chemical defense against pathogens (for a review of the *Drosophila* literature, see Lemaitre and Hoffmann, 2007). All of the genes in this category have some connection to either virulence in other fungal species (NCU05376, NCU06328, and NCU07787), detoxification of foreign molecules (NCU03415 and NCU10051), or have known anti-bacterial activities (NCU04850) (see below).

To investigate the misregulated genes from the *Neurospora ghh* AHC samples in more detail, I undertook a manual classification of the genes based on database and literature searches for studies carried out either directly on the *Neurospora crassa* genes or on their close homologs in other fungal species (Figure 15).

Strikingly, the most strongly down-regulated gene on the entire microarray (other than *ghh* itself) was seen to be *chitinase 1*, the lack of which is almost certainly responsible for the conidial separation phenotype observed in the *ghh* and *csp-2* mutants. I was glad to find this gene highest on the list, both as an explanation of the *csp* phenotype, and as a validation that the microarrays were picking up relevant down-regulated genes.

As stated earlier, I had hoped to see a large number of genes involved in cell wall formation down-regulated in the *ghh* mutants. I did find 7 genes involved in "Cell Wall Synthesis/Morphology" (Figure 15), most of which are enzymes involved in beta-glucan synthesis (which is a major constituent of the fungal cell wall); four of these genes have secretion signals and are almost certainly bona fide cell wall proteins. Also, it has been shown that the close homologs of several of these genes are found in the cell wall of *S.*

cerevisiae, and that mutations in these genes often have cell wall defects (Yuan et al., 2005).

Further classification of the down-regulated genes identified several more genes involved in the functions of defense and virulence (a category I termed "Virulence, Defense, Detoxification; Figure 15). Genes potentially involved in defense include the following: *kynureninase* (NCU09183) and *indoleamine 2,3-dioxygenase* (NCU01402) are involved in tryptophan catabolism (Zelante et al., 2009); NCU05495 is a putative anti-viral factor (Percudani et al., 2005); *exo-beta-1,3-glucanase* (NCU04850) is likely to be involved in the lysis of bacterial cell walls; and the homolog of the metalloprotease MEP1 (NCU07200) has been shown in *C. posadasii* to be crucial for the evasion of host-detection (Hung et al., 2005). Genes thought to be involved in fungal virulence include the following: The homolog of the p450 monooxygenase *lovA* (NCU05376) has been shown in a *Fusarium* species to be directly involved in mycotoxin synthesis (Meek et al., 2003); *cerato-platanin* (NCU07787) could be important for phytotoxin synthesis; the integral membrane protein *pth11* (NCU06328) has been shown in another fungal species to be important for appressorium formation (DeZwaan et al., 1999); and NCU03643 encodes a cutinase transcription factor that has been implicated in the control of plant cuticle digestion during fungal infection (Li et al., 2002). Finally, several genes potentially involved in the detoxification of harmful chemicals, as well as in the stress response, were also identified: the p450 gene *pisatin demethylase* (NCU06327) has been shown in fungal pea-pathogens to be important for detoxifying host defensive chemicals (Delserone et al., 1999); *aldehyde dehydrogenase* (NCU03415) is a broadly acting detoxification and metabolic enzyme; *catalase-3* (NCU00355) and *NAD(P)*

transhydrogenase (NCU01140) are both enzymes known to be important for oxygen-radical detoxification; and the *YBH1 flavohemoglobin* (NCU10051) may be involved in the stress response (Buisson and Labbe-Bois, 1998). Verification of microarray fold-change directionality for 10 genes using quantitative RT-PCR is shown in Figure 16.

Taken together these results indicate that while GHH appears to be playing some role in cell wall formation, its larger role appears to be in controlling the transcription of multiple genes involved in virulence and defense, many of which have secretion signals and are likely components of the cell wall.

Microarray profiling of late-stage *Drosophila grh* embryos

I also carried out microarray-based transcriptome profiling of late-stage *Drosophila grh^IM* and wildtype embryos, as a comparison dataset, and as a resource for the epidermal development and wound healing communities. I used flies homozygous for the *grh^IM* allele because it is the strongest *grh* allele available (with respect to its cuticle and head-skeleton phenotypes) and because homozygous embryos do not produce any detectable GRH protein (assayed using an antibody against the C-terminal half of the protein, Kim and McGinnis, 2011). However the mutation responsible for this allele was unknown.

While preparing mRNA samples for microarray analysis I found that *grh^IM* homozygous embryos were still producing *grh* transcripts at roughly the same level as wildtype (Figure 17A). Therefore, I guessed that a stop-codon introduction or frame-shift mutation must be responsible for the lack of GRH protein production. Using RT-PCR, I sequenced nearly all of the coding region of the *grh^IM* transcript, and found numerous

synonymous substitutions (data not shown), but only one mutation that would be predicted to affect protein production: a TAT to TAA stop-codon introduction in exon 7, shortly into the DNA binding domain and about half-way through the protein (Figure 11B, amino acid Y29 in the *D.mel GRH* DNA binding domain; Figures 17B, E, and F). This was also verified by sequencing heterozygous genomic *grh^{IM}* DNA, which showed an obvious double-peak at this position (Figures 17C and D). This mutation is consistent with a functional null phenotype for *grh^{IM}*, and the lack of GRH detection with a C-terminal specific antibody.

Gene Ontology classification of misregulated genes from the *grh* microarrays

GRH appears to have a huge impact on the transcriptome as a whole, as over 1,200 genes were seen to misregulated in the mutants compared to wildtype (at an FDR threshold of < 0.01) (see the Appendix for a full gene list). A search for enriched GO "Biological Process" (BP) and "Molecular Function" (MF) categories was performed, and the top 10 and 11 most significant classes, respectively, are shown in Table 4 (see the Appendix for the full category lists). As GRH is known to be very important for cuticle development and healing in *Drosophila*, I expected to see numerous genes involved in chitin metabolism down-regulated in the microarrays. Indeed several of the most significant BP classes (e.g., "chitin metabolic process" and "aminoglycan metabolic process") and MF classes (e.g.,"structural constituent of the cuticle" and "chitin binding") reflected the role of GRH in these processes.

Surprisingly, many of the most significant BP GO categories turned out to be involved in either innate immunity or the stress response (e.g., "defense response",

"immune response", and "humoral immune response", to name just a few). Amongst the MF GO categories I saw a large number of classes involving serine proteases and serine protease inhibitors (a.k.a. "serpins"), which are known to be important for hemolymph melanization-reactions used in response to infection (Lemaitre and Hoffmann, 2007). This was puzzling to me, as many of these genes were seen to be up-regulated in *grh* mutants, and GRH has no known function as an inhibitor of the immune response.

Upon some reflection, what I believe to be happening is that these *grh* embryos are wounding themselves during development. During the stages they were collected (late stage 16 or early stage 17 according to Campos-Ortega and Hartenstein, 2007), the *grh^{IM}* embryos were quite mobile and were probably tearing their fragile cuticles as they moved about. Apparently, this was sufficient to induce the expression of large numbers of genes involved in innate immunity and the stress response.

Many proteins involved in cuticle development are potentially under direct GRH-regulation

Similarly to what I did for the *Neurospora* microarrays, I carried out a manual classification of the genes both and up- and down-regulated from the *Drosophila grh* microarrays. The results are shown in Figure 18. I classified 64 genes in the category "Cuticle Formation / Chitin Metabolism", which includes genes involved in the generation and degradation of chitin molecules, as well as many cuticle proteins, which are deposited into the cuticle and mediate aspects of cuticle-shape and elasticity (Karouzou et al., 2007). Most of these genes were seen to be down-regulated, consistent with potential direct regulation by GRH, although many were also seen to be up-

regulated. It is likely that many of the up-regulated genes are not under the direct control of GRH, and are perhaps being expressed to compensate for the lack of the GRH-responsive genes, or equally likely, are being activated in response to cuticular breaks (see above). It was quite impressive to see so many cuticle genes down-regulated in grh^{IM} mutants, reinforcing the idea that GRH is truly a master regulator of physical barrier formation in *Drosophila*.

It is also interesting that one of the most strongly down-regulated genes on the microarray was *chitinase 3*, which is the closest *Drosophila* homolog to *Neurospora chitinase 1*. There are numerous strong GRH binding sites upstream of the *chitinase 3* gene, and it is extensively coexpressed with GRH throughout the epidermis and tracheal system during embryogenesis (data not shown). Therefore, it is possible that regulation of chitinase genes by GRH-proteins is ancestral, although convergent regulatory evolution is also likely.

I also classified 9 genes as belonging in the "Melanization / Wound Healing" category (Figure 18), 4 of which were down-regulated. The 3 most strongly down-regulated genes in this class (*a methyl dopa-resistant*, *Pcd*, and *prophenol oxidase A1*) are all known or suspected to be directly involved in the melanization pathway, and are potentially direct targets of GRH activation. *Dopa decarboxylase*, a known GRH-responsive gene in the melanization pathway, was also seen to be down-regulated, although it did not quite pass the stringent FDR threshold and is not shown in Figure 18. *Cad96Ca/Stitcher*, another known GRH-responsive "wound response" gene (Wang et al., 2009), was seen to be significantly down-regulated.

Up-regulated melanization genes are not likely to be direct-targets of GRH, and are most likely expressed either in a compensatory developmental role, or in direct response to cuticular breaks. A good example of this is the *pale* gene, which is known to not require GRH for its expression, but which is activated in response to wounding (Mace et al., 2005; Pearson et al., 2009).

grh embryos trigger a massive immune and defensive response, in the absence of infection

While a comprehensive analysis of all the genes involved in innate immunity, stress, and detoxification that were seen to be misregulated in *grh* mutants is beyond the scope and focus of this dissertation (and because the misexpression is due to an indirect effect of the *grh* mutation) I will only quickly review the major classes of genes.

35 genes were placed in the category "Innate Immunity" (Figure 18), and they include genes from nearly every aspect of *Drosophila* innate immunity (Lemaitre and Hoffmann, 2007). These genes were almost all up-regulated, and include known anti-microbial peptides (e.g., IMs, Attacins, Drosomycins, Diptericin B, and Metchnikowin), lysozymes, Pattern-Recognition Receptors (PRRs), and the Toll-signaling activator *Spaetzle-Processing Enzyme* (*SPE*). Only 1 AMP was seen to be down-regulated, IM1, although it was nearly 60-fold down-regulated, hinting at the possibility of direct activation of IM1 by GRH in the epidermis. Consistent with this possibility, there is a near perfect GRH binding site (AACTGGTTT) found less than ~600 bp upstream of the *IM1* gene (data not shown). Other down-regulated genes include 2 lectins (potential PRRs), PGRP-LD, and *Helical Factor* (a putative cytokine).

18 genes were placed in the category "Defense / Stress Response" (Figure 18). The two most strongly down-regulated genes were *methuselah-like 8* and *3*. Mutations in the paralogous gene *methuselah*, have been correlated with longer lifespan and increased resistance to stress in *Drosophila* (Lin et al., 1998), so it seems plausible that down-regulation of these paralogs is a reaction to tissue damage and stress in these embryos. 7 heat shock proteins (Hsps) are also on the list, 5 of which were up-regulated and have been shown elsewhere to be differentially expressed upon infection (*Hsp26*, *70Bc*, *70Bb*, *70Bbb*, and *70Aa*) (Chapter 12, Rolff and Reynolds, 2009). *Turandot A*, *C*, and *Victoria* were seen to be up-regulated, which are genes believed to act as extra-cellular chaperones, binding to denatured proteins in the hemolymph that are released upon tissue damage or stress (Ekengren and Hultmark, 2001).

29 genes were place in the category "Detoxification" (Figure 18). These included many cytochrome p450 genes (Chung et al., 2009), glutathione S-transferases, and UDP-glucuronosyltransferases, which function by chemically modifying toxic compounds in the cytoplasm or hemolymph, in order to render them less active. The presence of these genes (which were both down- and up-regulated in equal proportions) is still not clear to me, although I believe their expression is changing in response to the release of toxic endogenous compounds during tissue damage.

Taken together these results demonstrate that grh^{IM} embryos at this stage are triggering a massive immune/wound response and are undergoing (or at least act like they are undergoing) extreme stress, likely due to global tissue damage in response to cuticular tearing. While this is not directly germane to the topic of this chapter, it should be of interest to many in the immunity and wound-healing communities, as a way to investigate

genes that are activated in response to cuticular breaks *per se*, and not in response to

septic infection (if indeed these embryos are "sterile" within their vitelline envelopes).

DISCUSSION

"Is there an evolutionary connection between the fungal cell wall and animal epidermal barriers?" - it may be an impossible question to answer definitively. However, because of their high-level conservation of function in animals, I believed that by studying the function of GRH proteins in Fungi we might shed some light on the matter. I show in this chapter that the Grainy head homolog in *Neurospora crassa* is a true GRH-class homolog (and not an LSF-class homolog), and that it can bind the same DNA consensus site as animal GRHs, albeit with a lower affinity. Therefore, since its binding properties have been conserved, one could imagine that some vestiges of its function in the last common ancestor (LCA) of opisthokonts might also be conserved.

While the morphology of the opisthokont LCA is still a debated subject (an argument that will never be fully settled), it is generally agreed that the LCA was single-celled. Whether this LCA had a cell wall is unclear, although one could make the argument that, if it did, that cell wall would likely utilize the polymer chitin, which is the only extracellular biopolymer that is utilized in both Animals and Fungi (where it is used to strengthen extracellular physical barriers). If there is a true evolutionary connection between chitin usage in Animals and Fungi (as opposed to an independent "invention" of the molecule in both clades), then there might also be an evolutionary connection between transcriptional regulation of these chitinous extracellular barriers.

Chitin (β-1,4-N-acetylglucosamine) is very similar to another biopolymer, cellulose, and indeed the enzymes that synthesize chitin, cellulose, and hyaluronan are all closely related (making up the Glycosyltransferase Family 2). While it is a very real possibility that chitin production was independently developed in Animals and Fungi,

there is good evidence that the chitin synthase enzymes from Fungi and protostome animals (in this case, *Drosophila* and *C. elegans*) are most closely related to each other, and not to the other enzymes in the Glycosyltransferase Family 2. This indicates that a true homology exists between fungal and animal chitin synthases, and that this enzyme might have been present in the LCA (Merzendorfer, 2006).

While there is no evidence to suggest that GRH proteins regulate chitin synthase expression directly in animals (and I found no evidence to suggest it does so in *Neurospora*), there is abundant evidence that they regulate aspects of the melanization process (Bray and Kafatos, 1991; Venkatessan et al. 2003; Mace et al., 2005; Pearson et al., 2009), which is used to strengthen the cuticular chitin-network. In this chapter, I show that in addition to components of the melanization pathway (for which I identified several new putative GRH targets) *Drosophila* GRH is necessary for the proper expression of a large number of cuticle proteins and chitin-metabolizing enzymes (e.g., chitinase 3) during development, reinforcing the theory that GRH is the master control gene for cuticle development in animals.

I had hoped to see defects in cell wall formation in *Neurospora* strains lacking the *ghh* gene, hoping to draw some parallel between the fungal cell wall and animal epidermal barriers. I did uncover clear evidence that GHH is controlling the expression of a key chitin degrading enzyme (chitinase 1), and I find it very interesting (if a bit perplexing) that the most obvious function/phenotype of GHH in *Neurospora* does involve cell wall morphology, except it is involved in breaking down a physical barrier during conidial development, instead of building one up, as it does in animals. It is also intriguing that chitinase 1 in *Neurospora* and chitinase 3 in *Drosophila* both appear to be

directly regulated directly by GRH proteins (as described earlier), perhaps implying some ancestral regulation in the LCA. Why GRH-proteins would be regulating physical barrier formation in both Fungi and Animals, but in opposite capacities, is still unclear to me.

I also found evidence that GHH plays a role in the expression of enzymes involved in the synthesis of the other key biopolymer of the fungal cell wall: β-1,3/β-1,6-glucan. Therefore, GHH may turn out to have a more general role in cell wall development, but I have found no phenotypic evidence for this so far, despite testing the growth of *ghh* strains under several conditions previously shown to inhibit the growth of *S. cerevisiae* strains with compromised cell walls (high-osmolarity medium, high-temperature incubation, and medium containing the chitin-binding molecule Calcofluor-White; data not shown) (Yuan et al., 2005).

I think it is important to keep in mind that the possibility of a direct evolutionary connection between physical barriers in Animals an Fungi is a tenuous one, especially considering the lack of extracellular physical barriers in many dipoloblastic intermediate species, such as cnidarians. However, it has been shown that the external epithelium of the Hydra (a cnidarian, and one of the simplest multicellular animals known), despite a complete lack of mechanical barriers, is able to fight off pathogens by the induction of potent anti-microbial peptide (AMP) expression (Bosch et al., 2009). The position of hydra near the base of the metazoan lineage indicates this is likely to be a basic function of all animal epitheliums. Therefore, it was by researching the defense mechanisms of just such intermediate species, that I was again pointed towards the other possible function of fungal GRH proteins suggested by the microarrays - defense against pathogens.

While "defense" and "virulence" are separate categorizations (usually for good reason) the distinction in pathogenic fungi is often semantic, as they are really two sides of the same coin - the best way to become more virulent is to better defend yourself against your host, and vice versa. With this in mind, it is very interesting to see many of the down-regulated genes on the *Neurospora ghh* microarrays classified as defense and virulence genes. Furthermore, many of these have secretion signals and are likely to be produced and deposited in the cell wall (in a GHH dependent manner). This is analogous to how many epithelial barriers throughout the animal and plant kingdoms produce AMPs, both pro-actively and in response to infection (e.g., the *Drosophila* trachea and epidermis, mammalian lungs, and plant cuticles) (Tzou et al., 2000; Lemaitre and Hoffmann, 2007). Unfortunately, *Neurospora crassa* does not have any characterized host-pathogen interactions, so I was not able to directly test the function of any of these genes in terms of their effects on virulence or defense. Verification of the functions of these genes will have to wait for studies in other closely related ascomycete species with gene-knockout technologies and well-characterized host-pathogen interactions (e.g., *Magnaporthe grisea*).

While direct regulation of anti-microbial defense does not appear to be a major function *Drosophila* GRH, we did see one AMP gene (*IM1*) strongly down-regulated on the *Drosophila* microarrays. If this gene is truly under the expression of GRH in the epidermis or tracheal system, this would be a link between transcriptional control of physical and anti-microbial defense in *Drosophila*, which might be a vestige of the ancestral role of GRH-proteins in the opisthokont LCA. Consistent with this, research

carried out in our lab by M. Kim, demonstrates that knocking down the function of GRH (using RNAi) in the larval or adult epidermis increases the susceptibility of *Drosophila* to septic (bacterial) wounding (data not shown). This suggests a role for GRH proteins in epidermal-mediated anti-microbial defense in *Drosophila*, which has not been reported before.

Taking all of this into account, I now believe the main function of the Grainy head homolog in *Neurospora* is not to regulate physical barrier formation (although it may do that so some extent), but is instead to regulate the expression of numerous genes involved in defense and virulence. Conversely, transcriptome analysis of GRH function in *Drosophila* indicates that this transcription factor controls a large number of the genes involved in cuticle formation, with only minor indications (e.g., *IM1* down-regulation) that it might be involved in anti-microbial defense. However, considering the many connections between physical barriers, anti-microbial defense, and virulence, I find it an intriguing idea that function of GRH-like proteins in the opisthokont LCA was to regulate "chemical" defense and virulence (a function GHH likely still plays in extant fungal species), and that this somehow "primed" GRH transcription factors to adopt the transcriptional control of physical barrier formation in animals (which is now their main function).

MATERIALS AND METHODS

Neurospora **stocks and conditions**

Wildtype strains [FGSC2489 (74-OR23-1V, mat A), FGSC4200 (ORS-SL6, mat a)], *ghh* deficient strains [FGSC13563 (delta NCU06095, mat A), FGSC13564 (delta NCU06095, mat a)], and the NHEJ deficient strain [FGSC9720 (delta *mus-52::bar+*; *his-3*, mat A)] were obtained from the Fungal Genetics Stock Center. Stocks were maintained on minimal Vogel's agar slants with 1.5% sucrose and appropriate supplements (Davis, R., 2000). Genomic DNA for PCR analysis was obtained according to Guo et al. (2005).

Re-isolation of *ghh* deletion strains was performed by transforming a NCU06095 hygromycin replacement cassette (courtesy of the Dunlap lab, Dartmouth) into FGSC9720, as described previously (Ninomiya et al., 2004). Hygromycin resistant colonies were selected, and homokaryonic *ghh* KO strains were verified using PCR to detect the *ghh* locus. All strains not containing the *ghh* locus displayed the *csp* phenotype. These new strains (delta *mus-52::bar+*; *ghh*; *his-3*, mat A) were also used in the complementation assay fusings described below. The primer sequences for verifying the *ghh* deletions and for detecting *ghh* transcripts (Figures 14A and B) were as follows: grh-For: CACCAGTCAAGCTGGCATC and grh-Rev: GGCTTATGTCGCTGCTTTTC. Positive control primers were as follows: actin-For: ATCCGACACTTTTCGTCACC and actin-Rev: TGCAACAACCACCTCTCAAG.

Genetic complementation assays between *ghh* and *csp-2* were carried out by fusing one of the re-isolated *ghh* deletion strains described above (delta *mus-52::bar+*; *ghh*; *his-3*, mat A) to 10 different isolates of *csp-2*; *bd*; *inos*. The *csp-2*; *bd*; *inos* strains

were created using standard crossing methods (*csp-2; bd* x *inos*). Exactly as expected, half (5) of the fusings were viable on minimal media (due opposite mating-type incompatibility), all of which displayed the *csp* phenotype. Using PCR, the *csp-2/ghh* heterokaryons were verified as positive for the *ghh* locus, which apparently is still largely intact in *csp-2* strains (data not shown).

Protein production and gel-shifts

Full-length *Drosophila* GRH coding sequences were cloned into the plasmid pcDNA 3.1/myc-His(-) A (Invitrogen) as described previously (Pearson et al., 2009). Full length *Neurospora* GHH coding sequence was amplified using Phusion Polymerase (New England Biolabs) from an oligo-dT primed reverse transcription library (RETROscript kit from Ambion). Primers Ghh5'XbaIKozak (GCGTCTAGAGCCACCATGTTCAGT CAACGAACAAG) and Ghh3'HindIII (CGCAAGCTTGTAGAGCAGTCGCAGT TCAT) were used to introduce a Kozak sequence for efficient translation, and restriction endonuclease sites as well. The fragment was cloned into pcDNA™3.1/myc-His(-) A using the *XbaI* and *HindIII* sites in the MCS. The insert was fully sequenced and was found to be identical in sequence and intron structure to that predicted by the Broad *Neurospora* database.

GRH and GHH proteins were translated using the TNT T7 Quick Coupled Transcription/Translation System (Promega) by adding 1 μg of template to each reaction, following the manufacturers instructions. Protein expression levels were assayed using Western Blots against the C-terminal Myc tags, as described previously (Pearson et al., 2009). Translated protein was directly used in gel-shift assays, as freezing was found to

negatively affect protein binding activity. 500 pmol of each oligonucleotide were annealed in a final volume of 100 μl in annealing buffer (10 mM Tris-HCl pH7.5, 20 mM NaCl) by heating to 95°C for 5 min and slowly cooling to 25°C. 5 pmol of double-stranded oligos were labeled with polynucleotide kinase (New England BioLabs) in the presence of ATP-[^{32}P] for 30 min at 37°C. The double-stranded probes were purified by QIAquick Nucleotide Removal Kit (Qiagen). Ten to 20 fmol of radiolabeled DNA probe and 1.5 μl of protein from the in vitro transcription/translation reactions were added to 10 μl binding buffer (25 mM Hepes pH7.9, 100 mM KCl, 1 mM DTT, 1% polyvinylalcohol, 1% Nonidet P-40, 0.1% BSA, 10% glycerol, and 20 g/ml poly(dI-dC)) and incubated with DNA for 30 min at 4°C. The binding reaction was then electrophoresed through a 4% native polyacrylamide gel containing 0.5 TBE at 4°C. Gels were dried and autoradiographed with the use of intensifying screens.

Microarray sample collection

Neurospora samples for microarray analysis were collected according to the following procedures. Seeder slants of wildtype (FGSC2489) and *ghh* (FGSC13563) strains were grown for 3 days at 30°C in a 12 hr light/dark cycle, and conidia were harvested in 1 ml H$_2$0. As the *csp* phenotype makes homogenous resuspension of *ghh* conidia impossible, accurate conidial counts of the suspensions could not be obtained. Therefore, plates and flasks were innoculated with approximately equivalent masses of conidia suspended in water. As Neurospora comes to confluence quite quickly on plates, and growth in liquid culture for short periods of time should not be nutrient limiting, I believed the number of starting conidia not to be crucial to the outcome of these

experiments.

The following collection procedures were carried out in triplicate for both wildtype and *ghh* samples: 1) "ALL" samples were collected by densely plating conidia on minimal Vogel's agar medium + 1.5% sucrose in 10 cm Petri dishes that had been overlain with disks of cellophane (VWR, 100357-652). After 48 hr at 30°C in a light/dark cycle, the plates were densely covered with conidiating colonies. Samples were scraped off the cellophane using cell-scrapers, submerged in 5 ml Trizol (Invitrogen), and quickly frozen in liquid nitrogen. 2) "AHC" samples were collected by densely plating conidia on minimal Vogel's agar medium + 1.5% sucrose in a deep 10 cm Petri dishes. Disks of medium gauge wire mesh were suspended ~0.5 cm above the surface of the agar using a rings of plastic tubing around the periphery of each Petri dish. After 48 hr at 30°C in a light/dark cycle, aerial hyphae and conidia had grown abundantly through the mesh. The mesh disks were carefully peeled off, and the adherent cells were harvested in H_20 using cell-scrapers. Samples were dried by vacuum filtration, removed from the filter paper using cell-scrapers, submerged in 5 ml Trizol, and quickly frozen in liquid nitrogen. 3) "MYC" samples were collected by innoculating 25 ml of liquid Vogel's medium + 1.5% sucrose in 125 ml Erlenmeyer flasks with sponge stoppers. After 28 hr at 28°C, with constant shaking in the dark, log-phase mycelial mats were recovered by vacuum filtration, removed from the filter paper using cell-scrapers, and frozen in 2 ml Trizol with liquid nitrogen.

To obtain total RNA from the *Neurospora* samples for microarray analysis, I followed a protocol similar to that reported by Kasuga et al. (2008). Samples in Trizol were thawed and quickly homogenized by vortexing and passing through a P1000 pipet

tip multiple times. ~100 µl of cells were placed in an eppendorf tube with 1 ml Trizol and 200 mg of 0.5 mm Zirconia/Silica Beads (Biospec). Samples were disrupted twice with a MiniBeadBeater (Biospec) at maximum speed for 30 sec each time. RNA was then extracted using standard Trizol procedures and resuspended in 100 µl H_2O. RNA was quantified in 10 mM Tris pH 7.5, and 50 µg of total RNA was cleaned further using the RNeasy miniprep kit (Qiagen). RNA integrity was assayed by gel electrophoresis, RT-PCR, and Bioanalyzer (Agilent) analysis (data not shown).

The following *Drosophila* embryo collection procedures were carried out in duplicate. To aid in the collection of homozygous *grh* deficient embryos, the *cn, grh^{IM}, bw, sp* chromosome (Nusslein-Volhardt et al., 1984) was placed over the fluorescent balancer TM3, Kruppel-GFP. Embryos collected from these parents (+; *cn, grh^{IM}, bw, sp* / TM3, Kruppel-GFP; +) were allowed to develop at 25°C until ~16-20 hr of age. Embryos were aligned on a thin agar slab on a slide, and Kruppel-GFP negative embryos were chosen under an epifluorescence microscope. Autofluorescence in the GFP channel allowed for the selection of viable and properly aged embryos (late stage 16 and early stage 17) using gut morphology as a guide. Wildtype (*y; cn, bw, sp; +*) embryos were similarly raised and selected using gut autofluorescence as a guide. Approximately 500 mutant and wildtype embryos were collected and stored frozen in Trizol. Embryos were ground in Trizol using a pestle, and RNA was purified and assayed as described above.

Microarray design and analysis

Neurospora microarrays were custom synthesized by Agilent using the sequences from the *Neurospora crassa* arrays available from the FGSC (Kasuga et al., 2006). All

probe sequences were shortened from 70mers to 60mers by removing the first ten 5'

nucleotides. 10,526 unique spots were printed on each chip, corresponding to predicted

genes from several databases. Once a finalized list of significant genes was obtained,

probe sequences were BLASTed against the genome to verify the Broad (or MIPS) gene

IDs annotations.

Primers for the qPCR *Neurospora* microarray validations were as follows:

NCU04883: TACCTCTGCTGACACCAACG and CTTTGAGGTTGGCAAAGGAG;

NCU07787: AAGATCCTCAGCCTTTTCACC and GTCGTAGCCCGTGTCGTAG;

NCU04850: TCTCTACAGCGGTCGTGGTC and CCGACCATGATATCGACGAC;

NCU03415: CTTAGGGCTGGTACCGTCTG and ACCGATACCGGACTCCTTG;

NCU10051: ATCTGCATTTGGCGGATAAG and CCGTAGCAAAAAGCTCCAAG;

NCU07610: GATTTGCAGGTGCGGTTTAG and ATCCAACCGTACGATTACCG;

NCU04533: CTTGAAGGTGGATGCGAGAG and GACCAGCCCATACTCGTCTC;

NCU07819: AAAGCATTGTGGGTGAATCG and TCAGAATCACATCGCTCTCG;

NCU07821: TACCCGGGTCTGTTGTTCTC and GGGAGAAAGGGGTAGGACAC;

NCU07232: AGCGCAGCTATGGAGAGTTC and TATCCTGATCCACCGGAGTC.

Pre-designed *Drosophila melanogaster* arrays were ordered from Agilent (Design

ID# 18972). 43,603 spots were printed on each chip, which were mapped to ~13,000

unique FlyBase genes. Fluorescence values from redundant probes, or unique probes

targeting the same gene, were grouped and only the highest fold-change values were used

in these analyses. Once a finalized list of significant genes was obtained, probe sequences

were BLASTed against the genome to verify FlyBase CG # annotations.

RNA labeling, hybridizations, fluorescent quantification, data normalization, and calculation of False Discovery Rates (Sasik et al., 2004) were carried out by the UCSD Biogem Core facility.

ACKNOWLEDGEMENTS

I would like to thank S. Brody, for large amounts of technical advice on handling *Neurospora*, as well as numerous reagents. I would also like to thank Roman Sasik for carrying out the FDR and GO analyses on the microarray data. Finally, I would like to thank Myungjin Kim, for allowing me to cite her work in this chapter.

Chapter III, in full, is currently being prepared for submission for publication of the material. **Paré, A., Kim, M., and McGinnis, W.** The homolog of the metazoan epidermal integrity regulator Grainy head is involved in cell-wall formation, defense, and virulence in the fungus Neurospora crassa (in preparation). I was the primary investigator and author of this material.
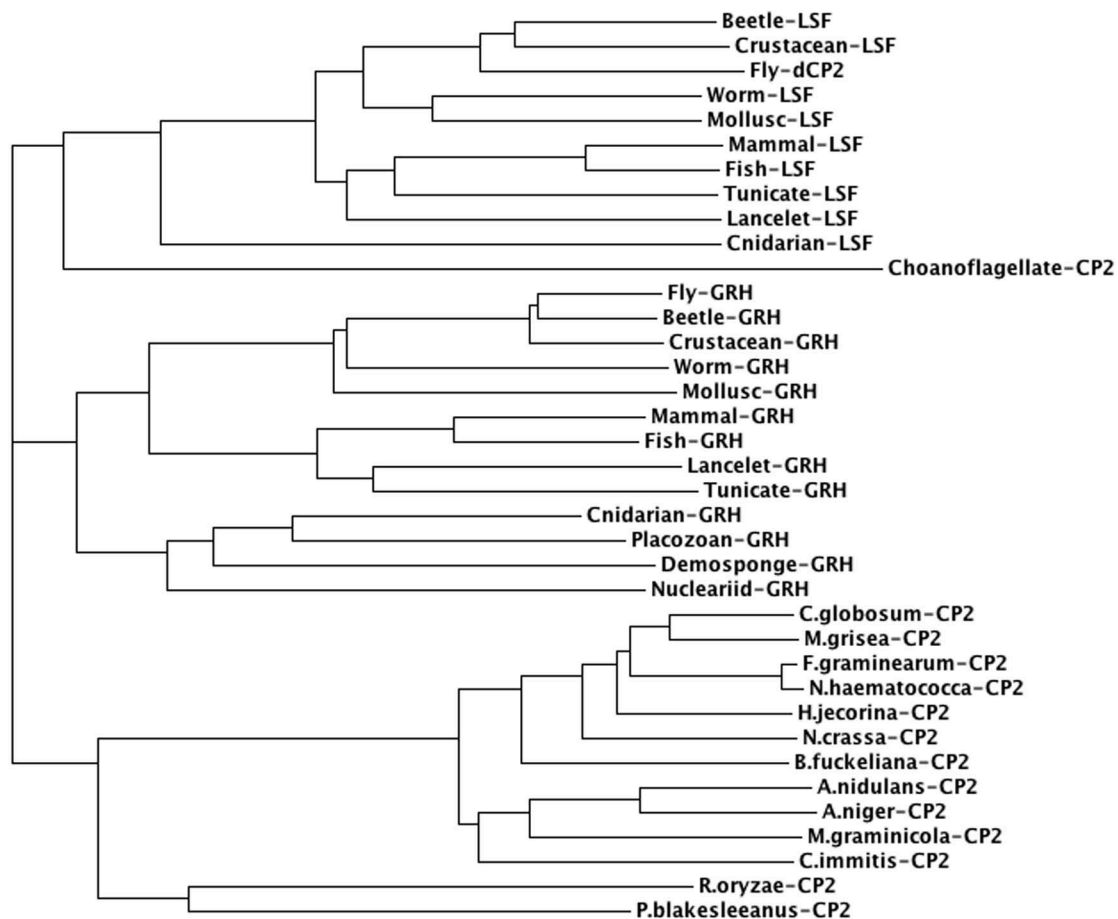
Figure 10. Metazoan and fungal CP2-class proteins form three distinct clades

A simple Neighbor-Joining Tree using the full DNA binding domain sequences from
each CP2 class protein (see the Appendix for alignments). Branch-lengths approximate
relative sequence divergence, and CP2-clade structures closely recapitulate what is
currently known regarding the evolutionary relationships of these species. Metazoan LSF,
metazoan GRH, and fungal CP2 proteins all cluster separately. The degenerate
Choanoflagellate-CP2 protein clusters most closely with metazoan LSF proteins, but the
long branch-length indicates high sequence divergence (consistent with manual
observation). Fly-dCP2 clusters most closely with Beetle-LSF (as expected) upon
removal of a long stretch of approximately 30 amino acids in its DNA binding domain,
which appears only in *Drosophila* (data not shown). Fungal CP2 proteins split into two
main sub-clades, which reflects the evolutionary divergence between the zygomycota
(bottom two species) and the ascomycota.

Figure 11. The *Neurospora* Grainyhead homolog (GHH) is similar in structure to metazoan GRH proteins and has a conserved DNA binding domain

(A) *Neurospora* GHH is similar in structure to both *Drosophila* and mammalian GRHs, containing a highly conserved DNA binding domain, as well a C-terminal domain with limited homology to the dimerization domain of animal GRH. **(B)** The DNA binding domain of GHH (*N.cra* GHH) shares extensive homology with *Drosophila* and human GRH-like proteins (*D.mel* GRH and *H.sap* Grhl1) and LSF-like proteins (*D.mel* dCP2 and *H.sap* LSF). Residues predicted to be important for DNA interaction (Kokoszynska et al., 2008) are marked above the alignment: "D" - dimerization; "Z" - zinc-binding; "m" - minor-groove interaction; "M" - major-groove interaction. Highly conserved residues that are useful in distinguishing GRH from LSF proteins are indicated below the alignment with dots.

A

H. sapiens
Grainyhead-like-1, -2, & -3 — DNA Binding / Dimerization — ~600 aa

D. melanogaster
Grainyhead — DNA Binding / Dimerization — 1063 aa

N. crassa
Grainyhead Homolog — DNA Binding / Oligomerization — 794 aa
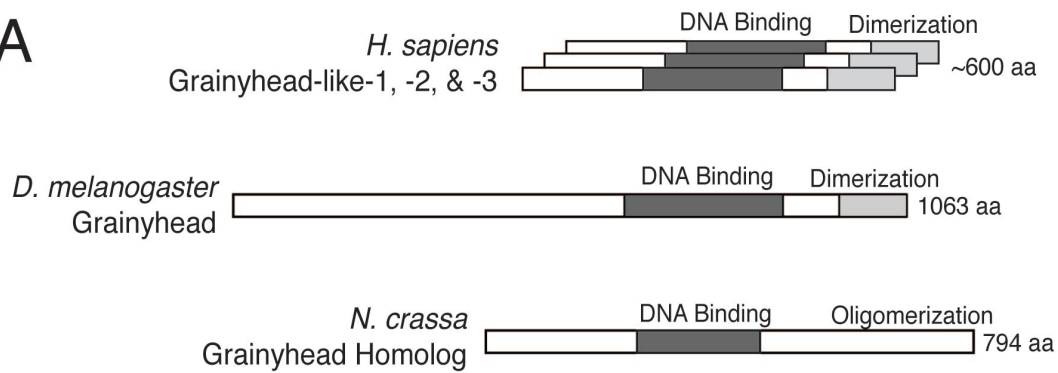
B

```
                         D   DDDD
D.mel GRH    FRYHLESPISSSQRREDDRITYINKGQFYGITLEYVHDAEKPIKN--TTV 48
H.sap Grhl1  FEYTLEASKSLRQKPGDSTMTYLNKGQFYPITLKEVSSSEG-IHHPISKV 49
N.cra GHH    FHTTLNAPTAMIKNTDEIPVTYLNKGQAYSLSVVD--TAPTLPIVPGTRF 48
D.mel dCP2   FQYILAAATSIATKNNEETLTYLNQGQSYEIKLK---KIGDLSLYRDKIL 47
H.sap LSF    FQYVMCAATSPAVKLHDETLTYLNQGQSYEIRMLDNRKMGDMPEINGKLV 50

                                    DD        DDDD
D.mel GRH    KSVIMLMFREEKSPEDEIKAWQFWH-------SRQHSVKQRILD-ADTKN 90
H.sap Grhl1  RSVIMVVFAEDKSREDQLRHWKYWH-------SRQHTAKQRCIDIADYKE 92
N.cra GHH    RTFVRISFEEEKQRHKPGMCWSLWKEGRGTNEAHQRGGKLQAVEFVEATQ 98
D.mel dCP2   KSVIKICFHERRLQFMEREQMQQWQ-------QSRPGERIIEVDVPLSYG 90
H.sap LSF    KSIIRVVFHDRRLQYTEHQQLEGWK-------WNRPGDRLLDLDIPMSVG 93

                                                         Zm mZ
D.mel GRH    SVG-----LVGCIEEVSHNAIAVYWNP---LESSAKINIAVQCLSTDFSS 132
H.sap Grhl1  SFN-----TISNIEEIAYNAISFTWD----INDEAKVFISVNCLSTDFSS 133
N.cra GHH    PAEGDDKRTRIELESASFDGFSVIWTPGINGSVECNIAVRFNFLSTDFSH 148
D.mel dCP2   LCH-----VSQPLSSGSLNTVEIFWDP---LKEVG-VYIKVNCISTEFTP 131
H.sap LSF    IID-----TRTNPSQ--LNAVEFLWDP---AKRTS-AFIQVHCISTEFTP 132
                                                      *   *  *

                D m
D.mel GRH    QK--GVKGLPLHVQIDTFE-DPRDTAV---------------------- 156
H.sap Grhl1  QK--GVKGLPLNIQVDTYSYNNRSNKP---------------------- 158
N.cra GHH    SK--GVKGIPVRLCAKTQPYLPNSPQSP---------------------- 174
D.mel dCP2   KKHGGEKGVPFRLQIETYIENTNSATASGSGGSNNSAIASGSGSSGSAAP 181
H.sap LSF    RKHGGEKGVPFRIQVDTFKQNENGEYT---------------------- 159
                *  * *

                                    MMMMMMMMMMM
D.mel GRH    ----------------FHRGYCQIKVFCDKGAERKTRDEERRAAKRKM 188
H.sap Grhl1  ----------------VHRAYCQIKVFCDKGAERKIRDEERKQSKRKV 190
N.cra GHH    -----------NTSDGAEICYCMVKLFRDHGAERKLSNDVAHVRKNIE 211
D.mel dCP2   ASPERTPSAGSNGKQAVHAAACQIKVFKLKGADRKHKQDREKIQKRPQ 229
H.sap LSF    --------------DHLHSASCQIKVFKPKGADRKQKTDREKMEKRTA 193
                                              *
```

Figure 12. Characteristic residues in the DNA binding domain of fungal CP2 proteins point to them being most homologous to metazoan GRH, and not LSF, proteins

An alignment of three highly conserved regions from various CP2-type proteins predicted to be important for DNA binding (Kokoszynska et al., 2008). Numbers are in reference to the DNA binding domain of *N. crassa* GHH (see Figure 10B). In all cases fungal CP2 residues are clearly more similar to metazoan GRH proteins than they are to metazoan LSF proteins. Species are as follows: Fruit Fly (*Drosophila melanogaster*), Beetle (*Tribolium castaneum*), Crustacean (*Daphnia pulex*), Mollusc (*Capitella capitata*), Mammal (*Homo sapiens*), Teleost Fish (*Danio rerio*), Lancelet (*Branchiostoma floridae*), Tunicate (*Ciona intestinalis*), Sea Anemone (*Nematostella vectensis*), Demosponge (*Amphimedon queenslandica*).

	141-147	150-154	193-198
Fruit Fly	CISTEFT	KHGGEKG	KGADRK
Beetle	CISTEFT	KHGGEKG	KGADRK
Crustacean	CISTEFT	KHGGEKG	KGADRK
Polychaete Worm	CISTEFT	KHGGEKG	KGADRK
Mollusc	CISTEFT	KHGGEKG	KGADRK
Mammal	CISTEFT	KHGGEKG	KGADRK
Teleost Fish	CISTEFT	KHGGEKG	KGADRK
Lancelet	CISTEFT	KHGGEKG	KGADRK
Tunicate	CISTEFT	RHGGEKG	KGADRK
Sea Anemone	CVSSEFT	KSGGESG	KGADRK
Fruit Fly	CLSTDFS	K--GVKG	KGAERK
Beetle	CLSTDFS	K--GVKG	KGAERK
Crustacean	CLSTDFS	K--GVKG	KGAERK
Polychaete Worm	CLSTDFS	K--GVKG	KGAERK
Mollusc	CLSTDFS	K--GVKG	KGAERK
Mammal	CLSTDFS	K--GVKG	KGAERK
Teleost Fish	CLSTDFS	K--GVKG	KGAERK
Lancelet	CLSTDFS	K--GIKG	KGAERK
Tunicate	CLSTDFS	K--GIKG	KGAERK
Sea Anemone	CLSTDFS	K--GVKG	KGAERK
Placozoan	CLSTDFS	K--GVKG	KGAERK
Demosponge	CLSTEFS	K--GVKG	KGAERK
Nucleariid	CLSTEFS	K--GVKG	KGSERK
C. globosum	FLSTDFS	K--GVKG	HGAERK
M. grisea	FLSTDFS	K--GVKG	HGAERK
F. graminearum	FLSTDFS	K--GVKG	HGAERK
N. haematococca	FLSTDFS	K--GVKG	HGAERK
H. jecorina	FLSTDFS	K--GVKG	HGAERK
N. crassa	FLSTDFS	K--GVKG	HGAERK
B. fuckeliana	FLSTDFS	K--GVKG	HGAERK
A. nidulans	FLSTDFS	K--GVKG	HGAERK
A. niger	FLSTDFS	K--GVKG	HGAERK
M. graminicola	FLSTDFS	K--GVKG	HGAERK
C. immitis	FLSTDFS	K--GVKG	HGAERK
R. oryzae	CLSTDFS	K--GVKG	KGAERK
P. blakesleeanus	CLSTDFS	K--GVKG	KGAERK
M. brevicolis	CLSTTFV	R-GGQAG	KGIIRK
Consensus	CLSTDFS	K GVKG	KGAERK

Animal LSF / Animal GRH / Fungal CP2

Figure 13. Gel-shift analyses show that *Neurospora* GHH binds DNA similarly to *Drosophila* GRH

(A) Oligos used in the gel-shifts. Residues in the LSF and GRH consensus binding-sites are marked with dots. **(B)** Gel-shift assays testing *Drosophila* GRH and *Neurosopora* GHH binding to the oligos in (A). Bottom panels were exposed for 16.5 hr, and top panels were exposed for 75 hr. Specific bands are indicated with black arrowheads, and weaker bands are also highlighted with asterisks in the top panel. Nonspecific (NS) bands were also detected in the no-template negative controls (data not shown) and are indicated with white arrowheads in the top panels, and with a bar in the bottom panels.

A

```
                           ****         ****
LSF-con    CTAGCCATATGGCTGGTTATGGCTGGTCAGAGATC
LSF-1/2    CTAGCCATATGGCTGGTTATGTATGTTCAGAGATC
LSF-mut    CTAGCCATATGTATGTTTATGTATGTTCAGAGATC


                           **********
GRH-Ddc    CGGACTGCGATTGAACCGGTCCTGCGGAATTGG
GRH-mut    CGGACTGCGATTCCCAAGGTCCTGCGGAATTGG
GRH-con    CGGACTGCGATTAAACCGGTTTTGCGGAATTGG
```

B

N. crassa GHH   *D. mel* GRH
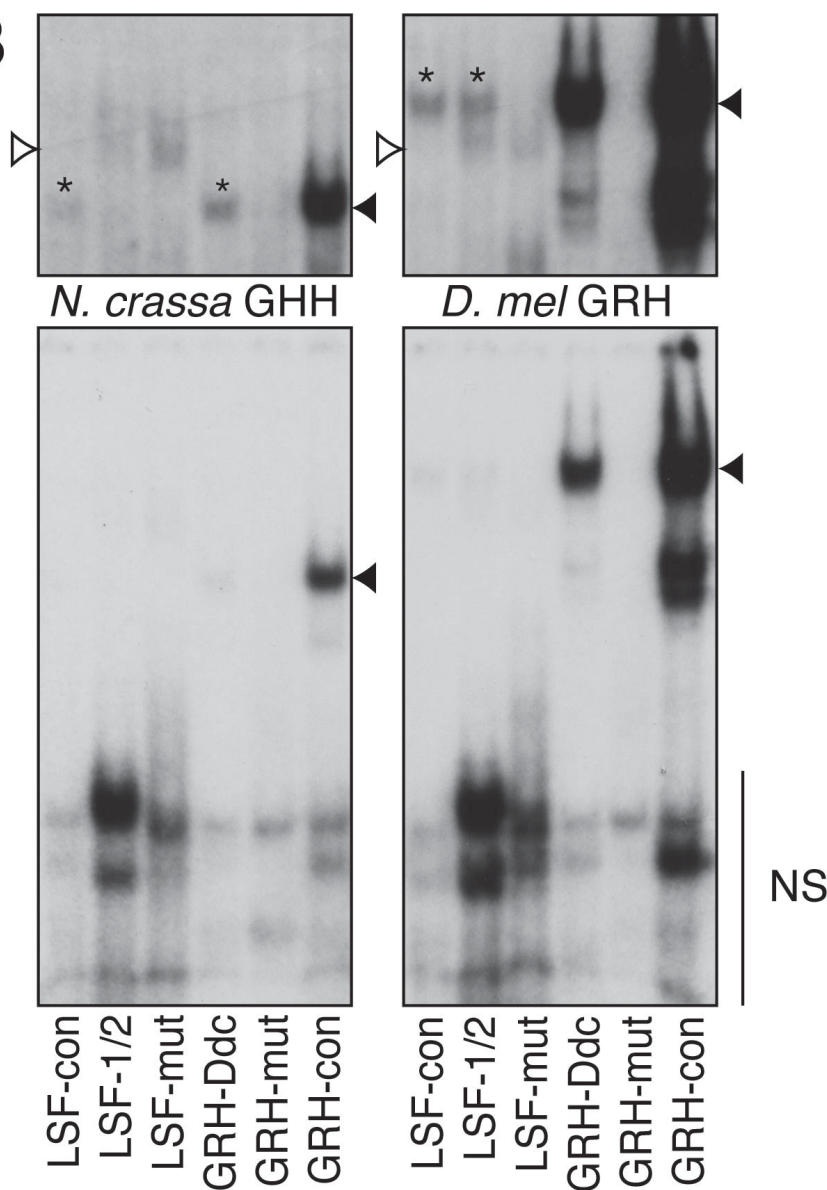
Figure 14. *Neurospora ghh* **mutants display a conidial separation phenotype**

(A) PCR analysis verifies that the *ghh* locus has been deleted in two independently generated *ghh* knockout lines (*ghh-1* and *ghh-2*), as well as in the knockout obtained from the FGSC (*ghh*), compared to wildtype. Actin is shown as a control. **(B)** RT-PCR analysis verifies that no *ghh* transcripts are being generated in the FGSC *ghh* knockout strains (mating type a or A), compared to wildtype. Similar results were obtained using total RNA from conidia or mycelia. Actin is shown as a control. **(C)** *ghh* strains reach full pigmentation more quickly than wildtype. Colonies were grown on Petri dishes for 48 hr at room temperature in a 12 hr light/dark cycle. Approximately 4 cm^2 of each mature conidiating colony is shown here. **(D-F)** *ghh* and wildtype strains have similar growth rates and eventually reach equal pigmentation levels: growth after 5 d at 30°C in constant light (D), a 12 hr light/dark cycle (E), or constant dark (F). **(G-I)** *ghh* strains display a pronounced conidial separation phenotype. (G) Wildtype conidial chains readily disperse in glycerol to yield individual spores. (H) *ghh* conidial chains remain intact in glycerol, even after vigorous agitation. (I) A close-up view of a *ghh* conidial chain showing that conidia remain connected by thick connectives (arrows).
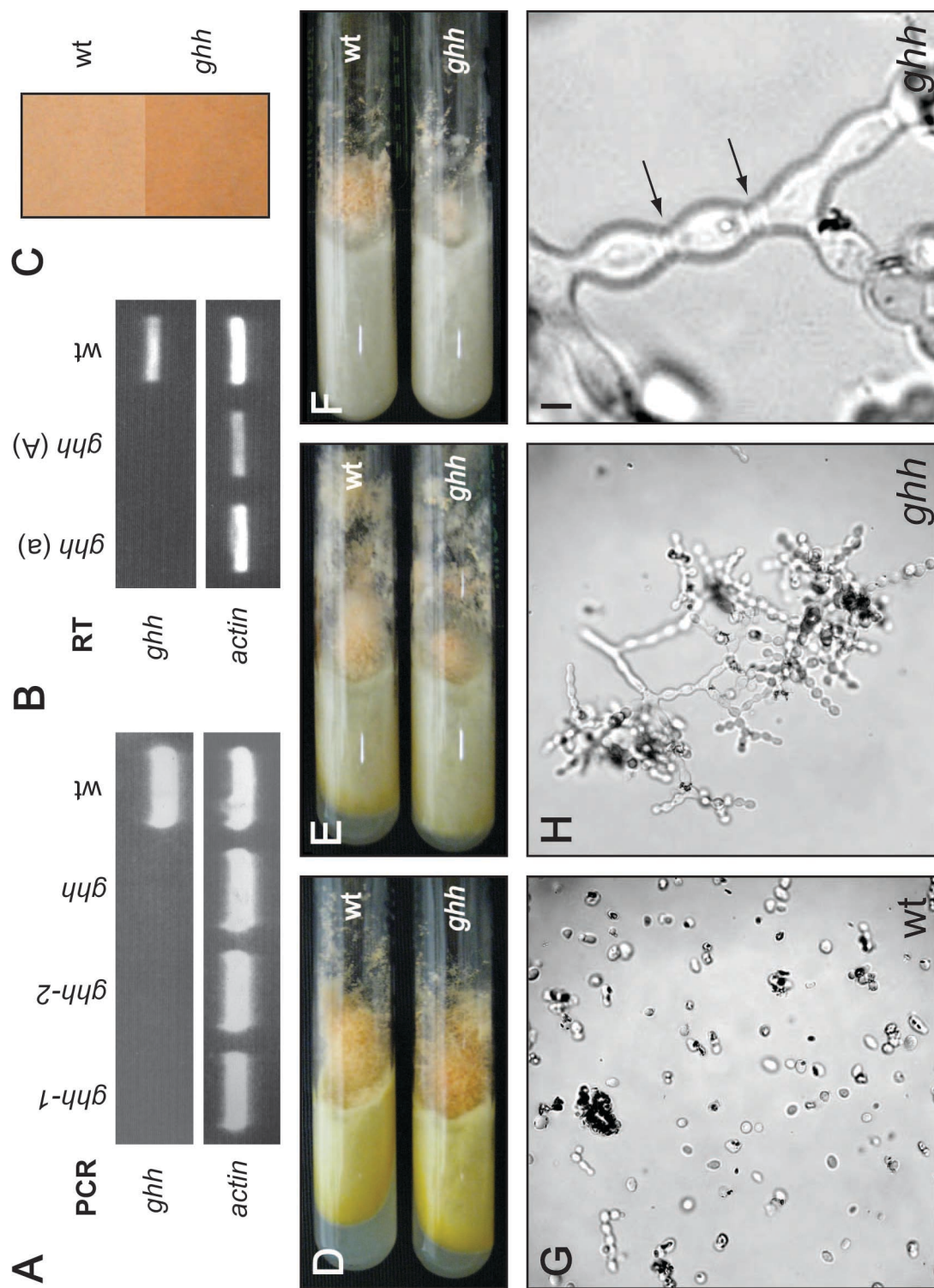
Figure 15. Down-regulated genes from the AHC samples of the *Neurospora ghh* microarrays

A manual annotation of the down-regulated genes from the *Neurospora ghh* AHC samples. "Broad ID" corresponds to the gene IDs in the Broad Institute *Neurospora crassa* database (http://www.broadinstitute.org/annotation/genome/neurospora/ MultiHome.html). The two italicized genes in the first column refer to probes that do not correspond to genes in the Broad database, but which correspond to genes in the MIPS database (http://mips.helmholtz-muenchen.de/genre/proj/ncrassa/). Gene functions were assigned based on literature searches and annotations in the Broad and MIPS databases. Numbers in brackets indicate whether that gene belonged to one of the 5 highly enriched FunCat categories: [1] metabolism of the cysteine - aromatic group, [2] metabolism of phenylalanine, [3] C-compound and carbohydrate transport, [4] degradation of tryptophan, and [5] disease, virulence and defense. "Fold (wt value)" refers to the fold change in *ghh* mutants relative to wildtype, and the absolute wildtype fluorescence-level values are shown in parentheses ("background" is approximately equal to 100 units). "FDR" refers to the False Discovery Rate values calculated for each gene; only genes with FDR values less than 0.01 are shown. The first two columns of the grid show whether the gene was also seen to be down- or up-regulated, respectively, in the MYC microarrays. Columns 2-10 represent a simplification of the FunCat classification system; solid colored blocks indicate that gene was annotated as belonging to that category in the FunCat database; dashes indicate that I found evidence in the literature to suggest the gene might also belong in that category. The final column shows whether the protein is predicted to be secreted (according to the TargetP algorithm); the presence of a number indicates it is likely that the protein will be secreted, with a "1" signifying the highest confidence. Other significant "genes" that could not be assigned a function are not shown.

Virulence, Defense, Detoxification

Broad ID	Gene name or Description	Function (Putative or known)	Fold (wt value)	FDR	Myc Down (> 1.5 Fold)	Myc Up (>1.5 Fold)	Virulence, Defense, Detoxification	Cell Wall Synthesis / Morphology	Membrane Transport / Cellular Import	Metabolism (Carbohydrates)	Metabolism (Amino Acids)	Metabolism (Nitrogen, Sulfur, Selenium)	Metabolism (Lipid, Fatty Acid, Isoprenoid)	Protein Fate / Modification	Environmental Interaction / Homeostasis	Secreted (TargetP prediction)
NCU09183	kynureninase [1, 4]	tryptophan catabolism	-4.54 (2455)	4.53E-05			-				■	■		■		
NCU05376	p450 monooxygenase (lovA) [5]	mycotoxin (trichothecene) synthesis	-3.55 (5445)	1.58E-04			■				■					
NCU01402	indoleamine 2,3-dioxygenase [1, 4]	tryptophan catabolism	-3.94 (633)	5.16E-04			-									
NCU06327	pisatin demethylase / p450 monooxygenase	phytoalexin detoxification	-3.92 (843)	1.90E-03			-									
NCU07787	cerato-platanin [5]	phytotoxic factor	-2.82 (9239)	1.93E-03		■	■									1
NCU05495	Cyanovirin-N domain-containing protein	putative anti-viral factor	-2.41 (24836)	2.24E-03		■	-									
NCU04850	exo-beta-1,3-glucanase [5]	foreign glucan degradation	-3 (12261)	2.24E-03			■			■						2
NCU06328	integral membrane protein (pth11) [5]	appressorium formation	-3.07 (546)	2.53E-03			■	-	■	■			■	■		2
NCU03415	aldehyde dehydrogenase [1, 2, 5]	detoxification / stress response	-2.26 (1121)	3.88E-03		■	■		■		■					
NCU00355	catalase-3	oxidative stress response	-3.08 (384)	4.44E-03			■									2
NCU07200	metalloprotease 1 (MEP1)	host detection evasion	-3.07 (1071)	5.59E-03			-							■		1
NCU10051	YHB1 flavohemoglobin [5]	stress response	-2.33 (17582)	6.34E-03			■					■				
NCU03643	cutinase transcription factor 1 beta	activation of plant cuticle digestion	-2.17 (845)	6.74E-03			■									
NCU01140	NAD(P) transhydrogenase	oxygen/radical detoxification	-2.23 (561)	8.77E-03			-		■							
NCU09402	1,4-butanediol diacrylate esterase	detoxification / stress response	-2.45 (554)	2.19E-03			-									
NCU09859	metalloprotease 1	proteolysis / evasion of host detection	-3.06 (3193)	4.41E-03			-							■		

Cell Wall Synthesis / Morphology

Broad ID	Gene name or Description	Function (Putative or known)	Fold (wt value)	FDR	Myc Down (> 1.5 Fold)	Myc Up (>1.5 Fold)	Virulence, Defense, Detoxification	Cell Wall Synthesis / Morphology	Membrane Transport / Cellular Import	Metabolism (Carbohydrates)	Metabolism (Amino Acids)	Metabolism (Nitrogen, Sulfur, Selenium)	Metabolism (Lipid, Fatty Acid, Isoprenoid)	Protein Fate / Modification	Environmental Interaction / Homeostasis	Secreted (TargetP prediction)
NCU04883	chitinase 1	cell wall morphology / conidial detachment	-6.97 (11576)	1.50E-05												1
NCU05974	cell wall glucanosyltransferase (Mwg1)	glucan synthesis	-2.9 (455)	1.06E-03				-		■						2
NCU00716	non-anchored cell wall protein-5	cell wall protein	-2.76 (3324)	1.60E-03		■		-								1
NCU04431	endo-1,3(4)-beta-glucanase	glucan degradation	-2.05 (542)	6.18E-03				-		■						1
NCU07253	1,3-beta-glucanosyltransferase (gel1)	glucan synthesis	-1.95 (10267)	9.06E-03				■		-						
NCU06505	glucanase B	glucan degradation	-3.15 (1651)	8.57E-04												
NCU04725	N-acetylglucosamine-6-phosphate deacetylase	glycan synthesis	-2.97 (1098)	4.23E-04									■			

Transporters

Broad ID	Gene name or Description	Function (Putative or known)	Fold (wt value)	FDR	Myc Down (>1.5 Fold)	Myc Up (>1.5 Fold)	Virulence, Defense, Detoxification	Cell Wall Synthesis / Morphology	Membrane Transport / Cellular Import	Metabolism (Carbohydrates)	Metabolism (Amino Acids)	Metabolism (Nitrogen, Sulfur, Selenium)	Metabolism (Lipid, Fatty Acid, Isoprenoid)	Protein Fate / Modification	Environmental Interaction / Homeostasis	Secreted (TargetP prediction)
NCU09771	DUF895 domain membrane protein	MFS Transporter	-3.95 (4045)	6.96E-05					-							
NCU02238	pantothenate transporter (liz1) [3]	MFS Transporter	-3.49 (1151)	2.16E-03												
NCU10276	dicarboxylic amino acid permease	amino acid Transporter	-3.01 (1416)	6.84E-03					▉						▉	
NCU01132	MFS monosaccharide transporter [3]	MFS Transporter	-2.68 (5779)	2.91E-03					▉							
NCU04568	pantothenate transporter (liz1)	transporter	-2.35 (1466)	3.95E-03					▉							
NCU03388	stomatin family protein	putative transporter	-2.32 (740)	2.66E-03		▉			-							
NCU09160	hpp family protein	putative transporter	-2.72 (1094)	5.51E-03					-							3
NCU08152	high affinity glucose transporter [3]	carbohydrate transporter	-2.25 (1582)	7.26E-03					-	▉						1
NCU09287	sugar transporter [3]	carbohydrate transporter	-3.8 (8152)	9.47E-03					▉	▉			▉		▉	
NCU01231	carboxylic acid transporter [3]	transporter	-3.16 (510)	1.63E-03						▉						

Miscellaneous Categorization

Broad ID	Gene name or Description	Function (Putative or known)	Fold (wt value)	FDR	Myc Down (>1.5 Fold)	Myc Up (>1.5 Fold)	Virulence, Defense, Detoxification	Cell Wall Synthesis / Morphology	Membrane Transport / Cellular Import	Metabolism (Carbohydrates)	Metabolism (Amino Acids)	Metabolism (Nitrogen, Sulfur, Selenium)	Metabolism (Lipid, Fatty Acid, Isoprenoid)	Protein Fate / Modification	Environmental Interaction / Homeostasis	Secreted (TargetP prediction)
NCU02328	O-methyltransferase	methyltransferase	-3.23 (1200)	7.23E-04		▉										
NCU05136	endo alpha-1,4 polygalactosaminidase precusor	metabolism	-2.5 (57714)	2.19E-03						▉						
NCU04727	glucosamine-6-phosphate deaminase	carbohydrate metabolism	-2.19 (3605)	8.13E-03					-	▉						
NCU09873	phosphoenolpyruvate carboxykinase	gluconeogenesis	-2.02 (1744)	7.42E-03		▉										
NCU08771	acetolactate synthase [1, 2]	amino acid metabolism	-2.74 (2273)	3.52E-03							▉					
NCU05499	homogentisate 1,2-dioxygenase [1, 2]	Tyr/Phe metabolism	-2.46 (2550)	5.48E-03							▉					
NCU06625	cysteine dioxygenase [1]	cysteine metabolism	-2.36 (397)	2.61E-03							-					
NCU11774	alpha-ketoacid dehydrogenase kinase	metabolism	-2.09 (29217)	5.97E-03							▉					
NCU00461	NAD-specific glutamate dehydrogenase	metabolism	-4.31 (53040)	5.18E-05		▉						▉				
NCU09169	NmrA family transcriptional regulator	metabolism	-3.79 (1979)	9.70E-05	▉							▉				
NCU08356	acetamidase	metabolism	-2.8 (1488)	2.03E-03								▉				
NCU08554	SSP-1 parvulin / peptidylprolyl isomerase	protein isomerase	-2.14 (17919)	4.51E-03		▉										1
NCU07063	aspartic-type signal peptidase	protease	-2.31 (700)	5.46E-03		▉								▉		
NCU09772	heterokaryon incompatibility protein	mating-type determination	-2.37 (659)	6.77E-03											▉	
NCU11562	all development altered-1	transcription factor	-2.08 (28601)	6.18E-03					transcription factor							
NCU21757	conserved hypothetical protein	transposon movement	-2.66 (1060)	3.28E-03					transposon movement							
NCU08907	blastomyces yeast phase-specific protein 1	unclear function	-2.29 (180259)	3.06E-03					unclear function							2
NCU05165	pyridoxamine phosphate oxidase	unclear function	-2.33 (625)	3.52E-03					unclear function							1
NCU06884	LipA and NB-ARC domain-containing protein	unclear function	-2.99 (313)*	4.33E-04	▉				unclear function							
NCU01449	GTP cyclohydrolase II	unclear function	-2.51 (6473)	1.89E-03					unclear function							
NCU05609	related to anaphylatoxin/fibulin	unclear function	-2.26 (2669)	5.40E-03					unclear function						▉	

Figure 16. Quantitative RT-PCR verification of fold-change directionality of the *Neurospora ghh* AHC microarrays

Quantitative RT-PCR (qPCR) was carried out on a selection of 10 genes (5 up- and 5 down-regulated) seen to be misregulated in the *Neurospora ghh* AHC microarrays. Genes were chosen to span a wide range of fold-changes. The qPCR results verify the directionality of the fold-changes seen on the microarrays, as well as (in most cases) the approximate fold-change values. Results were analyzed using two different housekeeping genes as controls (*actin* and *cbp*). Labels correspond to the following genes: NCU04883 - *chitinase 1*; NCU07787 - *cerato-platanin*; NCU04850 - *exo-beta-1,3-glucanase*; NCU03415 - *aldehyde dehydrogenase*; NCU10051 - *YHB1 flavohemoglobin*; NCU07610 - *taurine dioxygenase*; NCU04533 - *abundant perithecial protein*; NCU07819 - *alpha-ketoglutarate-dependent taurine dioxygenase*; NCU07821 - *dimethylaniline monooxygenase*; and NCU07232 - *heat shock protein 30*.

Figure 17. The lesion responsible for the *grh^IM* allele is a stop-codon introduction shortly into the DNA binding domain

(A) RT-PCR verifies that *grh^IM* embryos still produce *grh* transcripts at roughly the same levels as wildtype embryos. PCRs were carried out with biological replicates. **(B)** A schematic showing the location of the TAT>TAA stop codon introduction in the *grh^IM* mRNA, shortly after the start of the DNA binding domain (tyrosine Y29, from the *D.mel GRH* protein sequence in Figure 11B). **(C-F)** Sequencing reactions from both genomic DNA and mRNA templates unambiguously verify this mutation. Sequencing results from: heterozygous (+; *cn, grh^IM, bw, sp* / TM3, Kruppel-GFP; +) genomic DNA from adults (C), wildtype (*w^1118*) DNA from adults (D), homozygous deficiency (+; *cn, grh^IM, bw, sp*; +) RNA from embryos (E), and wildtype (*y; cn, bw, sp*; +) RNA from embryos.

A RT-PCR

B *D. melanogaster*
Grainyhead

C *grh/+* DNA

D *+/+* DNA

E *grh/grh* RNA

F *+/+* RNA
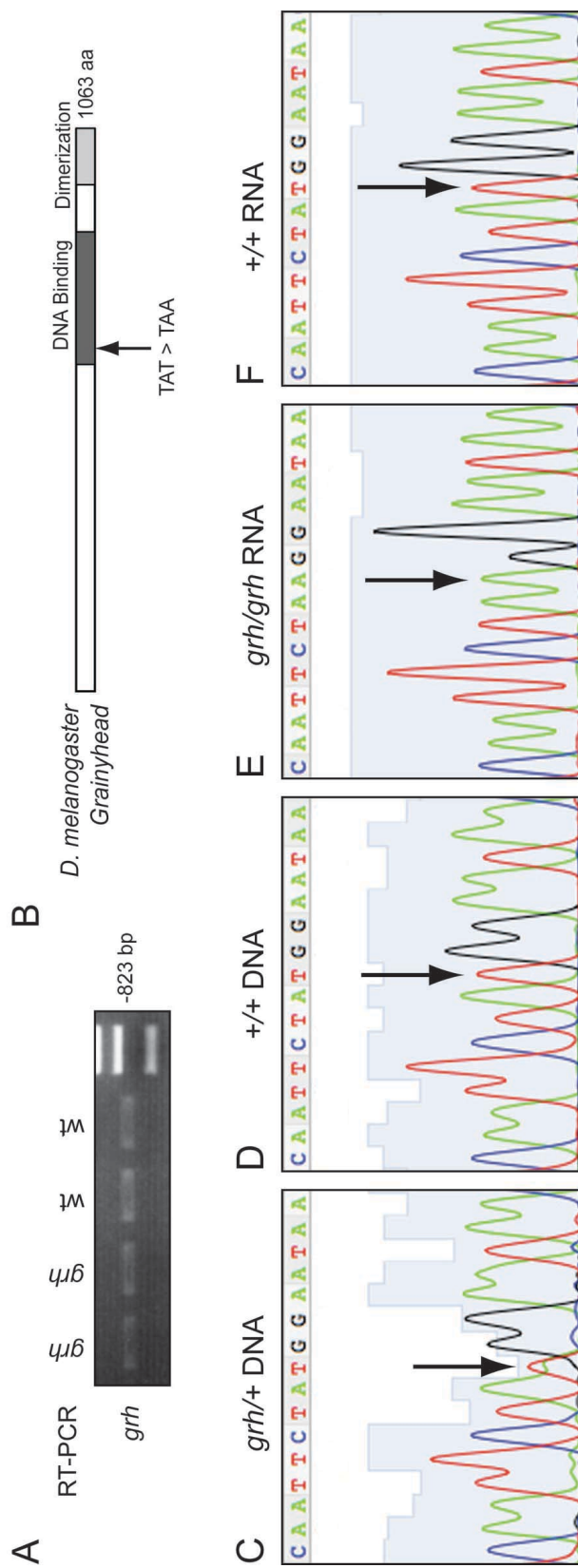
Figure 18. Selected misregulated genes from the late-stage *grh^IM* microarrays

Selected genes were manually classified in into the following categories: Cuticle Formation / Chitin Metabolism; Melanization / Wound Healing; Serine Proteases / Serpins; Innate Immunity; Defense / Stress Response; and Detoxification. "CG #" refers to the accession numbers from FlyBase (http://flybase.org). "Gene Name or Symbol" refers to either the full gene name, or the gene symbol. This column is left blank if no assigned gene name was found in FlyBase. "Function" refers to experimentally verified or putative (usually based on homology) functions assigned to the genes. Numbers in brackets refer to selected cross-referenced studies in which these genes were also seen to be misregulated upon the following treatments: [1] Bacterial infection (DeGregorio et al., 2001; Irving et al., 2001; Boutros et al., 2002; Roxstrom-Lindquist et al., 2004); [2] Fungal Infection (DeGregorio et al., 2001; Irving et al., 2001; Roxstrom-Lindquist et al., 2004); [3] Viral infection (Roxstrom-Lindquist et al., 2004; Dostert et al., 2005); [4] Wolbachia infection (Roxstrom-Lindquist et al. 2004; Xi et al., 2008); [5] Microsporidia infection (Roxstrom-Lindquist et al., 2004); and [6] Parasitoid infection (Wertheim et al., 2005; Schlenke et al., 2007). [1-6] were adapted from Chapter 12 of Rolff and Reynolds, 2009. [7] refers a systematic analysis of the expression patterns of the *Drosophila* p450 genes (Chung et al., 2009). "Fold (wt value)*" refers to the fold changes seen in the expression of these genes on the *ghh* microarrays, relative to wildtype. Absolute wildtype fluorescence values are shown in parentheses. An asterisk next to these values mean the lowest value in the GRH/WT ratio was near baseline (~100 units of fluorescence), which would artificially inflate the fold-change values. "FDR" refers to the False Discovery Rates calculated for each gene. All genes shown have an FDR value of less than 0.01.

Cuticle Formation / Chitin Metabolism (64)

CG #	Gene Name or Symbol	Function	Fold (wt value)*	FDR
CG2044	Lcp4	cuticle protein	-139.41 (13854)*	2.36E-07
CG30163	Cpr60D	cuticle protein	-51.83 (5170)*	1.20E-06
CG18066	Cpr57A	cuticle protein	-32.87 (38041)	4.27E-06
CG15515		cuticle protein	-28.37 (75931)	7.15E-06
CG2043	Lcp3	cuticle protein	-14.34 (1067)*	3.64E-05
CG18140	Chitinase 3	chitin metabolism	-12.9 (2331)	2.48E-05
CG7941	Cpr67Fa1	cuticle protein	-11 (109921)	4.05E-05
CG6955	Lcp65Ad	cuticle protein	-9.9 (1824)	5.80E-05
CG4052	Cpr5C	cuticle protein	-8.72 (4115)	7.31E-05
CG8697	Lcp2	cuticle protein	-6.65 (422)*	1.60E-04
CG32400	Lcp65Ab1	cuticle protein	-6.64 (73278)	1.65E-04
CG17052	obstructor-A	cuticle organization	-5.67 (106803)	2.64E-04
CG8510	Cpr49Af	cuticle protein	-5.66 (5870)	2.75E-04
CG9070	Cpr47Eg	cuticle protein	-5.54 (11869)	2.79E-04
CG6217	knickkopf	cuticle organization	-5.29 (5171)	3.19E-04
CG14250	TweedleQ	cuticle protein; body shape	-5.09 (533)	3.72E-04
CG7287	Lcp65Aa	cuticle protein	-4.94 (1261)	3.91E-04
CG4778	obstructor-B	cuticle organization	-4.87 (16760)	3.88E-04
CG18773	Lcp65Ab2	cuticle protein	-4.45 (15581)	5.24E-04
CG7216	Acp1	cuticle protein	-4.43 (3406)	5.02E-04
CG14643	TweedleG	cuticle protein; body shape	-4.02 (117093)	6.66E-04
CG9369	miniature	cuticle organization	-3.77 (696)	1.33E-03
CG14639	TweedleF	cuticle protein; body shape	-3.32 (30236)	1.79E-03
CG10297	Acp65Aa	cuticle protein	-3.28 (4421)	1.38E-03
CG11650	Lcp1	cuticle protein	-3.23 (235)*	1.18E-03
CG10529	Lcp65Ae	cuticle protein	-3.17 (12315)	1.31E-03
CG5883		chitin metabolism	-3.06 (2218)	1.38E-03
CG7548		cuticle metabolism	-2.75 (8307)	2.01E-03
CG9535	mummy	chitin biosynthesis	-2.63 (43714)	2.37E-03
CG11142	obstructor-E	cuticle organization	-2.56 (6997)	2.77E-03
CG5494	Cpr92F	cuticle protein	-2.5 (39488)	2.92E-03
CG33302	Cpr31A	cuticle protein	-2.5 (142683)	2.95E-03
CG18779	Lcp65Ag3	cuticle protein	-2.46 (233203)	3.08E-03
CG12009		chitin metabolism	-2.4 (2092)	6.86E-03
CG7252		chitin metabolism	-2.24 (3566)	4.48E-03
CG9295	Cpr76Bc	cuticle protein	-2.2 (303)	4.82E-03
CG12755	l(3)mbn	cuticle protein	-2.18 (380)	4.95E-03
CG32499	Cda4	chitin metabolism	-2.13 (24891)	6.23E-03
CG15008	Cpr64Ac	cuticle protein	-2.09 (2006)	5.95E-03
CG18778	Cpr65Au	cuticle protein	-2.07 (534)	6.47E-03
CG32404	Cpr65Aw	cuticle protein	-2.03 (211)	6.89E-03
CG5812	TweedleF	cuticle protein; body shape	-1.91 (27604)	9.30E-03
CG9781	obstructor-G	cuticle organization	1.78 (89)*	9.48E-03
CG9307	Chitinase 5	chitin metabolism	1.78 (15030)	8.48E-03
CG9079	Cpr47Ea	cuticle protein	1.84 (4076)	8.29E-03
CG8515	Cpr49Ah	cuticle protein	1.85 (78)*	6.72E-03
CG2555	Cpr11B	cuticle protein	1.9 (1069)	5.96E-03
CG6773	sec13	cuticle organization	1.95 (1358)	4.89E-03
CG9665	Cpr73D	cuticle protein	1.99 (1041)	6.30E-03
CG7876	Muc18B	chitin metabolism	2.12 (5161)	3.78E-03
CG10725		chitin metabolism	2.22 (4127)	2.45E-03
CG7539	Edg91	cuticle protein	2.26 (20178)	2.19E-03
CG4784	Cpr72Ec	cuticle protein	2.29 (180)	2.77E-03
CG10533	Lcp65Af	cuticle protein	2.34 (61521)	2.70E-03
CG15006	Cpr64Aa	cuticle protein	2.46 (148)	1.48E-03
CG10531	Chitinase 9	chitin metabolism	2.46 (108)	1.43E-03
CG10140		chitin metabolism	2.84 (470)	7.52E-04
CG16963	Crystallin	cuticle protein; eye lens protein	3.42 (399)	3.81E-04
CG31080	TweedleH	cuticle protein; body shape	3.55 (8962)	4.30E-04
CG9076	Cpr47Ed	cuticle protein	3.91 (87)*	2.19E-04
CG13224	Cpr47Eb	chitin metabolism	5.33 (4635)	8.76E-05
CG32284		chitin metabolism	5.57 (221)	7.66E-05
CG34271	Cpr65Ay	cuticle protein	9.56 (211)	1.61E-05
CG8836	Cpr49Ad	cuticle protein	35.79 (68)*	8.68E-07

Melanization / Wound Healing (9)

CG #	Gene Name or Symbol	Function	Fold (wt value)*	FDR
CG10501	a methyl dopa-resistant	dopamine metabolism	-5.46 (15116)	3.05E-04
CG1963	Prd	dopamine metabolism	-2.92 (6569)	1.65E-03
CG42639	prophenol oxidase A1	melanization effector	-2.29 (14798)	4.14E-03
CG10244	Cad96Ca / Stitcher	atypical RTK; wound healing	-1.91 (12245)	9.35E-03
CG1102	MP1	serine protease; melanization activator	1.75 (4362)	9.94E-03
CG12534	Alr	wound healing	1.82 (8107)	8.12E-03
CG15825	fondue	haemolymph coagulation	2.02 (23174)	4.32E-03
CG3066	MP2 / Sp7 / PAE1	serine protease; melanization activator	2.53 (1283)	1.27E-03
CG10118	pale	dopamine metabolism	3.77 (6615)	2.53E-04

Serine Proteases / Serpins (31)

CG #	Gene Name or Symbol	Function	Fold (wt value)*	FDR
CG11912		serine protease [6]	-42.33 (2411)*	2.37E-06
CG7722	Spn47C	serine protease inhibitor	-22.9 (1389)*	8.99E-06
CG16704		serine protease inhibitor [2,6]	-5.39 (973)	2.94E-04
CG1342	Spn100A	serine protease inhibitor	-5.39 (18723)	2.94E-04
CG31200		serine protease [2]	-4.05 (593)	6.74E-04
CG11843		serine protease [2]	-3.11 (212)*	1.63E-03
CG2071	Ser6	serine protease	-2.66 (8292)	2.29E-03
CG12172	Spn43Aa	serine protease inhibitor	-2.4 (3524)	3.63E-03
CG12385	thetaTry	Trypsin	-2.36 (1322)	3.73E-03
CG18477		serine protease [6]	-2.17 (234)	5.14E-03
CG6483	Jonah 65Aiii	serine protease [5,6]	-1.89 (29930)	9.49E-03
CG9649		serine protease [2]	2.01 (553)	4.25E-03
CG12388	kappaTry	Trypsin	2.04 (2297)	4.13E-03
CG9456	Spn1	serine protease inhibitor	2.14 (422)	4.06E-03
CG33329	Sp212	serine protease	2.15 (424)	3.84E-03
CG3344		serine protease [6]	2.19 (4432)	2.52E-03
CG8869	Jonah 25Bii	serine protease [1,3,5,6]	2.4 (569)	1.62E-03
CG8871	Jonah 25Biii	serine protease [1,5,6]	2.51 (4743)	1.43E-03
CG7754	iotaTry	Trypsin	2.57 (1749)	1.26E-03
CG18180		serine protease [1,5]	2.63 (65)*	1.07E-03
CG1859	Spn43Ad	serine protease inhibitor	2.79 (3765)	8.27E-04
CG18681	epsilonTry	Trypsin [6]	2.96 (122)	6.83E-04
CG11911		serine protease [2]	3.12 (20537)	5.85E-04
CG17571		serine protease [5]	3.59 (1111)	3.56E-04
CG2045	Ser7	serine protease [1,2]	3.7 (146)	2.80E-04
CG12351	deltaTry	Trypsin	3.93 (355)	6.21E-04
CG30028	gammaTry	Trypsin	6.98 (181)	4.31E-05
CG8867	Jonah 25Bi	serine protease [3,5]	8.72 (109)	2.17E-05
CG9733		serine protease	9.14 (727)	1.81E-05
CG18211	betaTry	Trypsin	24.12 (106)	2.13E-06
CG4821	Tequila	Trypsin; Neurotrypsin ortholog [6]	46.32 (731)	5.32E-07

CG #	Gene Name or Symbol	Function	Fold (wt value)*	FDR
Innate Immunity (35)				
CG18108	IM1	Fat Body secreted peptide [1,2,6]	-58.95 (3622)*	9.45E-07
CG14823		lysozyme	-9.02 (6023)	6.43E-05
CG10658	Helical Factor	putative cytokine	-5.37 (331)*	3.03E-04
CG7106	lectin-28C	PRR	-2.31 (420)	4.25E-03
CG6124	eater	PRR; phagocytosis	-1.94 (356)	8.79E-03
CG1179	LysB	lysozyme	1.84 (382)	6.80E-03
CG5008	GNBP3	PRR (Fungi); Toll-signaling	1.96 (189)	5.55E-03
CG18279	IM10	Fat Body secreted peptide	1.97 (6532)	6.84E-03
CG6426		lysozyme	2.02 (26238)	4.31E-03
CG10146	Attacin-A	AMP (GN Bacteria) [1,2,3,4,6]	2.02 (80)*	5.19E-03
CG16705	Spatzle-Processing Enzyme	serine protease; Toll-signaling	2.06 (3498)	4.04E-03
CG14704	PGRP-LB	catalytic PGRP [1,6]	2.13 (240)	2.97E-03
CG8696	LvpH	lysozyme	2.23 (4787)	2.86E-03
CG1180	LysE	lysozyme	2.44 (439)	1.50E-03
CG33717	PGRP-LD	PRR	2.63 (1653)	1.08E-03
CG4432	PGRP-LC	PRR (GN Bacteria); Imd-signaling [1]	2.63 (474)	1.06E-03
CG15678	pirk	response to symbiotic bacteria	2.87 (998)	7.21E-04
CG9697	PGRP-SB2	catalytic PGRP	3.12 (67)*	5.07E-04
CG8175	Metchnikowin	AMP (Fungi) [1,2,3,6]	3.3 (296)	4.05E-04
CG15065	IM2-like	Fat Body secreted peptide [1,2]	4.01 (1102)	1.95E-04
CG1165	LysS	lysozyme	4.17 (204)	1.70E-04
CG10794	Diptericin B	AMP (GN Bacteria) [1,3,4]	4.19 (193)	1.70E-04
CG15231	IM4	Fat Body secreted peptide [1,6]	4.35 (13040)	1.88E-04
CG16844	IM3	Fat Body secreted peptide [1,3,6]	5.45 (10195)	7.45E-05
CG32279	drosomycin-2	AMP (Fungi)	5.49 (177)	7.38E-05
CG15066	IM23	Fat Body secreted peptide [1,6]	5.69 (853)	7.10E-05
CG9120	LysX	lysozyme	5.82 (74)*	6.24E-05
CG18372	Attacin-B	AMP (GN Bacteria) [1,2,3,4,6]	6.44 (77)*	5.62E-05
CG10810	Drosomycin	AMP (Fungi) [1,2,3]	6.6 (1095)	4.31E-05
CG4740	Attacin-C	AMP (GN Bacteria) [1,3,4]	6.63 (72)*	4.31E-05
CG13422		PRR [1,2,6]	7.12 (64)*	3.36E-05
CG18106	IM2	Fat Body secreted peptide [1,2,3,6]	7.95 (3140)	3.45E-05
CG2958	lectin-24Db	PRR	9.12 (75)*	1.67E-05
CG9118	LysD	lysozyme	11.69 (319)	7.15E-06
CG10812	drosomycin-5	AMP (Fungi) [1,2]	138 (81)*	1.33E-07
Defense / Stress Response (18)				
CG32475	methuselah-like 8	GPCR	-44.54 (2412)*	1.68E-06
CG6530	methuselah-like 3	GPCR	-5.91 (1075)	2.36E-04
CG16954	Hsp60D	heat shock protein	-5.85 (546)*	2.42E-04
CG33117	Victoria	Turandot-like	-3.99 (500)	7.06E-04
CG2830	Hsp60B	heat shock protein	-3.06 (6325)	1.97E-03
CG4604	Glial Lazarillo	ApoD ortholog	-2.78 (9228)	1.94E-03
CG12002	Peroxidasin	ROS metabolism; ECM peroxidase [1,2,6]	-2.62 (6843)	2.91E-03
CG6646	DJ-1alpha	oxidative stress response [4]	-2.5 (686)	3.29E-03
CG7052	TepII	opsonization; humoral response [1,2,6]	-2.19 (6648)	5.88E-03

CG #	Gene Name or Symbol	Function	Fold (wt value)*	FDR
Defense / Stress Response (18) Continued				
CG6871	Catalase	ROS metabolism; H2O2 breakdown	2.25 (19730)	2.44E-03
CG31509	Turandot A	humoral stress response [6]	2.51 (68)*	1.34E-03
CG6186	Transferrin 1	Iron sequestration [2]	2.92 (105)	6.96E-04
CG4183	Hsp26	heat shock protein [1]	3.76 (563)	2.47E-04
CG6489	Hsp70Bc	heat shock protein [1,3,4]	4.75 (156)	1.11E-04
CG31449	Hsp70Ba	heat shock protein [4]	5.51 (297)	7.73E-05
CG31508	Turandot C	humoral stress response [6]	5.54 (70)*	7.52E-05
CG31366	Hsp70Aa	heat shock protein [4]	7.48 (1513)	3.13E-05
CG5834	Hsp70Bbb	heat shock protein [4]	7.54 (383)	2.69E-05
Detoxification (29)				
CG1944	Cyp4p2	P450 (Fat Body [71])	-87.19 (4690)*	4.20E-07
CG10241	Cyp6a17	P450 (Hindgut [71])	-56.78 (3284)*	8.68E-07
CG33503	Cyp12d1-d	P450 (Fat Body, Midgut, Malphigian Tubes [71])	-22.22 (1314)*	9.99E-06
CG10842	Cyp4p1	P450 (Midgut, Malphigian Tubes [71])	-10.72 (5972)	4.31E-05
CG33546	gfzf	glutathione S-transferase	-10.04 (12617)	4.58E-05
CG1488	Cyp311a1	P450 (Midgut [71])	-4.11 (454)	6.29E-04
CG30489	Cyp12d1-p	P450 (Fat Body, Midgut, Malphigian Tubes [71])	-4.02 (1004)	7.24E-04
CG8652	Ugt37c1	UDP-glucuronosyltransferase	-2.56 (1005)	2.70E-03
CG17527	GstE5	glutathione S-transferase	-2.22 (3038)	6.11E-03
CG12242	GstD5	glutathione S-transferase	-2.21 (223)	4.83E-03
CG13271	Ugt36Bb	UDP-glucuronosyltransferase	-2.19 (240)	4.97E-03
CG17525	GstE4	glutathione S-transferase	-2.16 (1686)	5.26E-03
CG5137	Cyp312a1	P450 (Gonads [7])	-2.11 (322)	5.94E-03
CG8453	Cyp6g1	P450 (Fat Body, Midgut, Malphigian Tubes [71])	-2.03 (463)	7.11E-03
CG1829	Cyp6v1	P450 (Gonads [7])	1.8 (97)*	8.38E-03
CG4772	Ugt86Dh	UDP-glucuronosyltransferase	1.82 (2621)	7.46E-03
CG6633	Ugt86Dd	UDP-glucuronosyltransferase	1.85 (591)	9.71E-03
CG4381	GstD3	glutathione S-transferase	1.89 (537)	6.69E-03
CG10248	Cyp6a8	P450 (Malphigian Tubes [7])	1.89 (2545)	9.56E-03
CG17534	GstE9	glutathione S-transferase	1.95 (4213)	5.17E-03
CG10240	Cyp6a22	P450 (Gonads [7])	1.96 (951)	4.99E-03
CG3943	kraken	detoxification	2.05 (8581)	4.02E-03
CG4485	Cyp9b1	P450 (no in situ expression [71])	2.06 (533)	3.49E-03
CG13270	Ugt36Ba	UDP-glucuronosyltransferase	2.73 (6015)	9.76E-04
CG11012	Ugt37a1	UDP-glucuronosyltransferase	4.47 (93)*	1.37E-04
CG3481	Adh	alcohol dehydrogenase	4.76 (21147)	1.21E-04
CG10245	Cyp6a20	P450 (no in situ expression [71])	5.85 (2512)	6.56E-05
CG8345	Cyp6w1	P450 (Fat Body, Midgut, Malphigian Tubes [71])	9.13 (128)	1.67E-05
CG18559	Cyp309a2	P450 (Gonads [7])	28.2 (94)*	1.58E-06

Table 4. Enriched functional categories for down-regulated *Neurospora ghh* genes and misregulated *Drosophila grh* genes

(Top) Enriched Functional Catalogue (FunCat) categories for the 83 significantly down-regulated (FDR < 0.01) genes from the *Neurospora* AHC microarrays (http://mips.helmholtz-muenchen.de/genre/proj/ncrassa/Search/Catalogs/ searchCatfirstFun.html). (Bottom) The top enriched Gene Ontology (GO) "Biological Process" and "Molecular Function" categories for all misregulated genes from the *Drosophila grh* microarrays (analysis carried out by the Biogem Core, UCSD).

Down-regulated genes from *Neurospora* AHC samples

Enriched FunCat Categories	FunCat ID	# of genes	p-value
metabolism of the cysteine - aromatic group	01.01.09	6	3.07E-04
metabolism of phenylalanine	01.01.09.04	3	1.63E-03
C-compound and carbohydrate transport	20.01.03	5	1.76E-03
degradation of tryptophan	01.01.09.06.02	2	3.44E-03
disease, virulence, and defense	32.05	6	4.13E-03

Misregulated genes from *Drosophila grh* samples

Enriched GO Biological Process Categories	GO term ID	# of genes	p-value
carbohydrate metabolic process	5975	244	1.06E-06
chitin metabolic process	6030	77	2.11E-06
defense response	6952	117	2.67E-06
response to biotic stimulus	9607	109	2.98E-06
aminoglycan metabolic process	6022	95	5.55E-06
response to other organism	51707	104	6.35E-06
immune response	6955	121	1.57E-05
polysaccharide metabolic process	5976	102	2.17E-05
humoral immune response	6959	74	5.40E-05
response to stress	6950	347	8.36E-05

Enriched GO Molecular Function Categories	GO term ID	# of genes	p-value
structural constituent of cuticle	42302	96	1.02E-015
structural constituent of chitin-based cuticle	5214	92	2.32E-014
serine-type endopeptidase activity	4252	165	7.63E-014
serine hydrolase activity	17171	185	1.54E-013
serine-type peptidase activity	8236	183	1.85E-013
peptidase activity, acting on L-amino acid peptides	70011	352	8.71E-009
endopeptidase activity	4175	274	1.96E-008
structural constituent of chitin-based larval cuticle	8010	35	4.81E-008
peptidase activity	8233	361	1.17E-007
chitin binding	8061	66	1.85E-007
polysaccharide binding	30247	87	4.75E-007

REFERENCES

Almeida, M., and Bray, S. (2005), Regulation of post-embryonic neuroblasts by Drosophila Grainyhead. Mech. Develop. *122*, 1282-93.

Auden, A., Caddy, J., Wilanowski, T., Ting, S., Cunningham, J., and Jane, S. (2006). Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. Gene Expr. Patterns *6,* 964-70.

Bray, S., and Kafatos, F. (1991) Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila. Gene. Dev. *5,* 1672-83.

Brody, S., Oelhafen, K., Schneider, K., Perrino, S., Goetz, A., Wang, C., and English, C. (2010). Circadian rhythms in Neurospora crassa: Downstream effectors. Fungal Genet. Biol. *47,* 159-68.

Buisson, N., and Labbe-Bois, R. (1998). Flavohemoglobin expression and function in Saccharomyces cerevisiae. No relationship with respiration and complex response to oxidative stress. J. Biol. Chem. *273,* 9527-33.

Campos-Ortega, J., and Hartenstein, V. (1997). The Embryonic Development of *Drosophila melanogaster*, Second Edition (Berlin, New York: Springer).

Cenci, C., and Gould, A. (2005). Drosophila Grainyhead specifies late programmes of neural proliferation by regulating the mitotic activity and Hox-dependent apoptosis of neuroblasts. Development *132,* 3835-45.

Chung, H., Sztal, T., Pasricha, S., Sridhar, M., Batterham, P., and Daborn, P. (2009). Characterization of Drosophila melanogaster cytochrome P450 genes. Proc. Natl. Acad. Sci. USA *106,* 5731-6.

Davis, R. (2000). Neurospora: contributions of a model organism, First Edition (Oxford, New York: Oxford University Press).

Delserone, L., McCluskey, K., Matthews, D., and Vanetten, H. (1999) Pisatin demethylation by fungal pathogens and nonpathogens of pea: association with pisatin tolerance and virulence. Physiol. Mol. Plant Pathol. *55,* 317-26.

DeZwaan, T., Carroll, A., Valent, B., and Sweigard, J. (1999). Magnaporthe grisea pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. The Plant Cell *11,* 2013-30.

Ekengren, S., and Hultmark, D. (2001). A family of Turandot-related genes in the humoral stress response of Drosophila. Biochem. Bioph. Res. Co. *284,* 998-1003.

Guan, X., Middlebrooks, B., Alexander, S., and Wasserman, S. (2006). Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila. Proc. Natl. Acad. Sci. USA *103*, 16794-9.

Hemphälä, J., Uv, A., Cantera, R., Bray, S., and Samakovlis, C. (2003). Grainy head controls apical membrane growth and tube elongation in response to Branchless/FGF signalling. Development *130*, 249-58.

Huang, J., Dubnicoff, T., Liaw, G., Bai, Y., Valentine, S., Shirokawa, J., Lengyei, J., and Courey, A. (1995). Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of decapentaplegic. Gene. Dev. *9*, 3177-89.

Hung, C., Seshan, K., Yu, J., Schaller, R., Xue, J., Basrur, V., Gardner, M., and Cole, G. (2005). A metalloproteinase of Coccidioides posadasii contributes to evasion of host detection. Infect. Immun. *73*, 6689-703.

Karouzou, M., Spyropoulos, Y., Iconomidou, A., Cornman, R., Hamodrakas, and Willis, J. (2007). Drosophila cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. Insect Biochem. Molec. *37*, 754-60.

Kim, M., and McGinnis, W. (2010). Phosphorylation of Grainy head by ERK is essential for wound-dependent regeneration but not for development of an epidermal barrier. Proc. Natl. Acad. Sci USA *108*, 650-5.

Kokoszynska, K., Ostrowski, J., Rychlewski, L., and Wyrwicz, L. (2008). The fold recognition of CP2 transcription factors gives new insights into the function and evolution of tumor suppressor protein p53. Cell Cycle *7*, 2907-15.

Kudryavtseva, K., Sugihara, T., Wang, N., Lasso, R., Gudnason, J., Lipkin, S., and Andersen, B. (2003). Identification and characterization of Grainyhead-like epithelial transactivator (GET-1), a novel mammalian Grainyhead-like factor. Dev. Dynam. *226*, 604-17.

Lambreghts, R., Shi, M., Belden, W., Decaprio, D., Park, D., Henn, M., Galagan, J., Basturkmen, M., Birren, B., Sachs, M., Dunlap, J., and Loros, J. (2009). A high-density single nucleotide polymorphism map for Neurospora crassa. Genetics *181*, 767-81.

Lemaitre, B., and Hoffmann, J. (2007). The host defense of Drosophila melanogaster. Annu. Rev. Immunol. *25*, 697-743.

Li, D., Sirakova, T., Rogers, L., Ettinger, W., and Kolattukudy, P. (2002). Regulation of constitutively expressed and induced cutinase genes by different zinc finger transcription factors in Fusarium solani f. sp. pisi (Nectria haematococca). J. Biol. Chem. *277*, 7905-12.

Liaw, G., Rudolph, K., Huang, J., Dubnicoff, T., Courey, A., and Lengyel, J. (1995). The torso response element binds GAGA and NTF-1/Elf-1, and regulates tailless by relief of repression. Gene. Dev. *9*, 3163-76.

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

Mace, K., Pearson, J, and McGinnis, W. (2005). An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head. Science *308*, 381-5.

Meek, I., Peplow, A., Ake, C., Phillips, T., and Beremand, M. (2003). Tri1 encodes the cytochrome P450 monooxygenase for C-8 hydroxylation during trichothecene biosynthesis in Fusarium sporotrichioides and resides upstream of another new Tri gene. Appl. Environ. Microb. *69*, 1607-13.

Merzendorfer, H. (2006). Insect chitin synthases: a review. J. Comp. Physiol. B. *176*, 1-15.

Ninomiya, Y., Suzuki, K., Ishii, C., and Inoue, H. (2004). Highly efficient gene replacements in Neurospora strains deficient for nonhomologous end-joining. Proc. Natl. Acad. Sci. USA *101*, 12248-53.

Ostrowski, S., Dierick, H., and Bejsovec, A. (2002) Genetic control of cuticle formation during embryonic development of Drosophila melanogaster. Genetics *161*, 171-82.

Pearson, J., Juarez, M., Kim, M., Drivenes, O., and McGinnis, W. (2009). Multiple transcription factor codes activate epidermal wound-response genes in Drosophila. Proc. Natl. Acad. Sci. USA *106*, 2224-9.

Percudani, R., Montanini, B., and Ottonello, S. (2005). The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. Proteins *60*, 670-8.

Rolff, J., and Reynolds, S. (2009). Insect Infection and Immunity: Evolution, Ecology, and Mechanisms. (Oxford, New York: Oxford University Press).

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., and Mewes, H. (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res. *32*, 5539-45.

Sásik, R., Woelk, C., and Corbeil, J. (2004). Microarray truths and consequences. J. Mol. Endocrinol. *33*, 1-9.

Selitrennikoff, C., Nelson, R., and Siegel, R. (1974). Phase-specific genes for macroconidiation in Neurospora crassa. Genetics *78*, 679-90.

Shirra, M., and Hansen, U. (1998). LSF and NTF-1 share a conserved DNA recognition motif yet require different oligomerization states to form a stable protein-DNA complex. J. Biol. Chem. *273*, 19260-8.

Springer, M., and Yanofsky, C. (1989). A morphological and genetic analysis of conidiophore development in Neurospora crassa. Gene. Dev. *3*, 559-71.

Tang, H. (2009). Regulation and function of the melanization reaction in Drosophila. Fly *3*, 105-11.

Tao, J., Kuliyev, E., Wang, X., Li, X., Wilanowski, T., Jane, S., Mead, P., and Cunningham, J. (2005). BMP4-dependent expression of Xenopus Grainyhead-like 1 is essential for epidermal differentiation. Development *132*, 1021-34.

Ting, S., Wilanowski, T., Cerruti, L., Zhao, L., Cunningham, J., and Jane, S. (2003). The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. Biochem. J. *370*, 953-62.

Ting, S., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W., Elias, P., Cunningham, J., and Jane, S. (2005). A homolog of Drosophila grainy head is essential for epidermal integrity in mice. Science *308*, 411-3.

Ting, S., Caddy, J., Wilanowski, T., Auden, A., Cunningham, J., Elias, P., Holleran, W., and Jane, S. (2005) The epidermis of grhl3-null mice displays altered lipid processing and cellular hyperproliferation. Organogenesis *2*, 33-5.

Traylor-Knowles, N., Hansen, U., Dubuc, T., Martindale, M., Kaufman, L., and Finnerty, J. (2010). The evolutionary diversification of LSF and Grainyhead transcription factors preceded the radiation of basal animal lineages. BMC Evol. Biol. *10*, 101.

Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J., Lemaitre, B., Hoffmann, J., and Imler, J. (2000) Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. Immunity *13*, 737-48.

Uv, A., Thompson, C., and Bray, S. (1994). The Drosophila tissue-specific factor Grainyhead contains novel DNA-binding and dimerization domains which are conserved in the human protein CP2. Mol. Cell. Biol. *14*, 4020-31.

Uv, A., Harrison, E., and Bary, S. (1997). Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol. Cell. Biol. *17*, 6727-35.

Venkatesan, K., McManus, H., Mello, C., Smith, T., and Hansen, U. (2003). Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. Nucleic Acids Res. *31*, 4304-16.

Wang, S., Tsarouhas, V., Xylourgidis, N., Sabri, N., Tiklova, K., Nautiyal, N., Gallio, and Samakovlis, C. (2009). The tyrosine kinase Stitcher activates Grainy head and epidermal wound healing in Drosophila. Nat. Cell Biol. *11*, 890-5.

Yin, Q., de Groot, P., Dekker, H., de Jong, L., Klis, F., and de Koster, C. (2005). Comprehensive proteomic analysis of Saccharomyces cerevisiae cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. J. Biol. Chem. *280*, 20894-901.

Yu, Z., Lin, K., Bhandari, A., Spencer, J., Xu, X., Wang, N., Lu, Z., Gill, G., Roop, D., Wertz, P., and Anderson, B. (2006). The Grainyhead-like epithelial transactivator Get-1/Grhl3 regulates epidermal terminal differentiation and interacts functionally with LMO4. Dev. Biol. *299*, 122-36.

Zelante, T., Fallarino, F., Bistoni, F., Puccetti, P., and Romani, L. (2009). Indoleamine 2,3-dioxygenase in infection: the paradox of an evasive strategy that benefits the host. Microbes Infect. *11*, 133-41.

APPENDIX

Accession numbers for CP2 sequences

GenBank (http://www.ncbi.nlm.nih.gov/genbank/); JGI - Joint Genome Institute (http://www.jgi.doe.gov/genome-projects/); NCBI - National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/); Broad OOM - Broad Institute Origins of Multicellularity Database (http://www.broadinstitute.org/annotation/genome/multicellularity_project/GenomesIndex.html); Broad FGI - Broad Institute Fungal Genome Initiative (http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative); KEGG - Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/).

Species	Common Name	Database	Accession
Metazoan GRH-like Proteins			
Drosophila melanogaster	Fruit Fly	GenBank	AAF57782.3
Tribolium castaneum	Beetle	GenBank	EFA11004.1
Daphnia pulex	Crustacean	JGI	jgi|Dappu1|64332|e_gw1.171.21.1
Capitella capitata	Polychaete Worm	JGI	jgi|Capca1|198092|fgenesh1_pg.C_scaffold_3000030; Capca1|198091|fgenesh1_pg.C_scaffold_3000029
Lottia gigantea	Mollusc	JGI	jgi|Lotgi1|157385|fgenesh2_pg.C_sca_14000008
Homo sapiens	Mammal	GenBank	AAH67520.1
Danio rerio	Teleost Fish	NCBI	XP_001923763
Branchiostoma floridae	Lancelet	NCBI	XP_002586459.1
Ciona intestinalis	Tunicate	NCBI	XP_002131671
Nematostella vectensis	Sea Anemone	JGI	jgi|Nemve1|95157|e_gw.38.94.1
Trichoplax adhaerens	Placozoan	JGI	jgi|Triad1|25702|e_gw1.5.1214.1
Amphimedon queenslandica	Demosponge	N/A	genomic trace archive
Metazoan LSF-like Proteins			
Drosophila melanogaster	Fruit Fly	GenBank	AAM68771.2
Tribolium castaneum	Beetle	NCBI	XP_974530.1
Daphnia pulex	Crustacean	JGI	jgi|Dappu1|192185|estExt_Genewise1Plus.C_50175
Capitella capitata	Polychaete Worm	JGI	jgi|Capca1|222821|estExt_fgenesh1_pg.C_600043
Lottia gigantea	Mollusc	JGI	jgi|Lotgi1|166840|fgenesh2_pg.C_sca_61000088
Homo sapiens	Mammal	GenBank	AAH47235.1
Danio rerio	Teleost Fish	NCBI	XP_001336482.1
Branchiostoma floridae	Lancelet	NCBI	XP_002612877
Ciona intestinalis	Tunicate	JGI	jgi|Cioin2|262956|gw1.03q.546.1
Nematostella vectensis	Sea Anemone	JGI	jgi|Nemve1|189242|estExt_GenewiseH_1.C_1380090
Metazoan outgroup CP2 proteins			
Capsaspora owczarzaki	Nucleariid	Broad OOM	C. owczarzaki: Supercontig 7: 1313062-1317017 -
Monosiga brevicollis	Choanoflagellate	JGI	jgi|Monbr1|29664|fgenesh2_pg.scaffold_37000034
Fungal CP2 Proteins			
Chaetomium globosum	Ascomycete	Broad FGI	CHGG_09727.1
Magnaporthe grisea / oryzae	Ascomycete	Broad FGI	MGG_00856.6
Fusarium graminearum / oxysporum	Ascomycete	Broad FGI	FGSC_06356.3
Nectria haematococca / Fusarium solani	Ascomycete	JGI	jgi|Necha2|99900|estExt_fgenesh1_pg.C_sca_3_chr4_2_00062
Trichoderma reesei / Hypocrea jecorina	Ascomycete	JGI	jgi|Trire2|60761|e_gw1.8.498.1
Neurospora crassa	Ascomycete	Broad FGI	NCU06095.4
Botryotinia fuckeliana	Ascomycete	Kegg	BC1G_14920
Aspergillus nidulans	Ascomycete	Broad FGI	ANID_04878.1
Aspergillus niger	Ascomycete	NCBI	XP_001399828.2
Mycosphaerella graminicola	Ascomycete	JGI	jgi|Mycgr1|80283|estExt_gwp_gw1.C_110736
Coccidioides immitis	Ascomycete	Broad FGI	CIMG_03814.3
Rhizopus oryzae	Zygomycete	Broad FGI	RO3G_07526.3
Phycomyces blakesleeanus	Zygomycete	JGI	jgi|Phybl1|75515|estExt_fgeneshPB_pg.C_10328

CP2 DNA binding domain sequence-alignments

```
>Beetle-LSF
FQYVLAAATSIATKVNEDTLTYLNQGQSYEIKLKKLG---DLSMYRGKLLKSVIRMCFHE
RRLQYMEKEQMAAWQR----ARPGDRILEVDVPLSYGAFDI---------VQPTNA--LN
IIHFNWDPTKE-----VGVYIKVNCISTEFTPKKHGGEKGVPFRIQVETYQNGDNLDSS-
-------------------------------VRLHAAACQIKVFKLKGADRK-HK
QDREKIMKRPL-

>Crustacean-LSF
FQYVLAAATSIATKVNEESLTYLNQGQPYEIKMKKLG---DLSNFRGKLLRSVVRLCFHE
RRLQYMEREQIAAWRM----SRPGDRIVEIDVPLSYGIYEV---------VQDNSN--LN
VVEFAWDPTKE-----VGVYIKVNCISTEFTPKKHGGEKGVPFRIQVETYSHGDGDGTP-
-------------------------------KRLHVAGCQIKVFKLKGADRK-HK
QDREKIYKRPM-

>Fly-dCP2
FQYILAAATSIATKNNEETLTYLNQGQSYEIKLKKIG---DLSLYRDKILKSVIKICFHE
RRLQFMEREQMQQWQQ----SRPGERIIEVDVPLSYGLCHV---------SQPLSSGSLN
TVEIFWDPLKE-----VGVYIKVNCISTEFTPKKHGGEKGVPFRLQIETYIENTNSATAS
GSGGSNNSAIASGSGSSGSAAPASPERTPSAGSNGKQAVHAAACQIKVFKLKGADRK-HK
QDREKIQKRPQ-

>Worm-LSF
FQYVLAAATSPATKMYEETLTYLNQGQSYEIKLKKLG---EMVRPQTLRVRSIVRVVFHE
RRLQYMESEQITTWKH----NRPGDRILDIDIPLSYGLLDV---------NVDPNS--LN
SIDFSWDPSKS-----AGIYIRVNCISTEFTAKKHGGEKGVPFRIQVETYIGDIHPARI-
-------------------------------VHCSSCQVKVFKPKGADRK-HK
TDRERIEKRSE-

>Mollusc-LSF
FQYILGAATSPAVKMNEETLTYLNQGQSYEIKLKKLG---DLSNSHGKLLKSLVRVHFHE
RRLQYMEKQQIETWKQ----NRPGERILDIDIPLSYGLIDV---------NLDPIK--LN
EAEFIWDPTKS-----TGIYIRVHCISTEFTAKKHGGEKGVPFRIQLDTFSHEEEEEKL-
-------------------------------LHSASCQVKVFKPKGADRK-HK
TDREKMDKRSE-

>Mammal-LSF
FQYVMCAATSPAVKLHDETLTYLNQGQSYEIRMLDNRKMGDMPEINGKLVKSIIRVVFHD
RRLQYTEHQQLEGWKW----NRPGDRLLDLDIPMSVGIIDT---------RTNPSQ--LN
AVEFLWDPAKR-----TSAFIQVHCISTEFTPRKHGGEKGVPFRIQVDTFKQNENGEYT-
-------------------------------DHLHSASCQIKVFKPKGADRK-QK
TDREKMEKRTA-

>Fish-LSF
FQYVMCAATSPAVKLHEETLTYLNQGQSYEIRMLDNRKKGEMPELN-KIVKSIVRVVFHD
RRLQYMEHQQLEGWKW----NRPGDRLLDIDIPMSVGITEA---------QAHPSQ--LN
AAEFLWDLSKR-----ASVFVQVHCISTEFTPRKHGGEKGVPFRIQIDTFKQDESGEYI-
-------------------------------EHLHSASCQIKVFKPKGADRK-QK
TDREKMEKRTP-
```

```
>Tunicate-LSF
LQYMLCAPTSPATKVYEETLTYLNQGLPYEIKLKKLRDIPDLGTLK--HVKSQLRVVFHD
RRLQYTEYEQFQNWKF----NRPGDRLLNIDIPMSVGVIQP---------REHPEQ--LN
LVEFIWDVEKE-----ASVFIQVHCISTEFTVRKHGGEKGVPFRIQIDTYALQNNNEYG-
------------------------------------RYIHSASCQIKVFKPKGADRK-QK
TDKDKMERRTA-

>Lancelet-LSF
FQYILCAATSPASKINEETVTYLNQRQSYELRLKRLGDNSSFQGGE--LLKSVVRVVFHD
RRLQYTEYEQLAQWRS----IRPNERLLDIDIPLSVNIYDI---------RKDPTA--VN
KVEFQWDPNKD-----TSVAIQVHCISTEFTAHRHGGEKGIPFRIQVDSYSIEE------
------------------------------------EHLHSASCQIKVFKPKGADRK-IK
TDREKMEKKPD-

>Cnidarian-LSF
FCFILKAPTAPGKKVNEDTLTYLNQGQSYPIDVQYIG---SISLFKGSLLTSVVTLTFYE
RKLQVVEAEKFEEWRN----NHPLERIFEIDVPMSTGLQNI---------RSKGNL--TN
AYEFDWNPEEDQ----IKLYIIINCVSSEFTKGKSGGESGVPLQIQIETWSIDASFDELP
------------------------------------MSCNFCQVKVFKSKGADRK-HK
MELQKLESKSH-

>Fly-GRH
FRYHLESPISSSQRREDDRITYINKGQFYGITL-EYVH-DAEKPIKNTTVKSVIMLMFRE
EKSPEDEIKAWQFWHS--RQHSVKQRILDADTKNSVGLVG------------CIEEVSHN
AIAVYWNPLESS----AKINIAVQCLSTDFSSQK--GVKGLPLHVQIDTFEDPRD-----
-TAVFHR------------------------------GYCQIKVFCDKGAERK-TR
DEERR-AAKRKM

>Beetle-GRH
FKYQLETPISTSQRREDDRITYINKGQFYGITL-DYIP-DPDKQLKSQTVKSIVMLMFRE
EKSPEDEIKAWQFWHG--RQHSVKQRILDADTKNSVGLVG------------CIEEVAHN
AIAIYWNPLESP----AKINIAVQCLSTDFSSQK--GVKGLPLHLQIDTYEDPRD-----
-TNVYHR------------------------------GYCQIKVFCDKGAERK-TR
DEERR-AAKRKM

>Crustacean-GRH
FRYYLESPISTSQRREDDRITYINKGQFYGITL-EYVP-DPERPLKNQTVKTMVMLVFRE
EKSPEDEAKAWQFWHG--RQHSAKQRILDADTKNSSGLIG------------CIEEVAHN
AICIYWNPLESS----AKINVAVQCLSTDFSSQK--GVKGLPLHLQIDTFDDPRDS----
-IPVFHR------------------------------GYCQVKVFCDKGAERK-TR
DEERR-AAKRKM

>Worm-GRH
YRYYLETNISTTQRINEDRITYLNKGQFYGLVL-EYKP--IQRMLPCSTVKSIIMVVFRE
EKPGQDDAKAWDFWYS--RQHSIKMRILDYDTKNSEGVVAQ----------YIEETAHN
AVAIRWNPLDKP----AKINIAINCLSTDFSNQK--GVKGLPLHLQIDTFEDTSS-----
-ATPIHR------------------------------GYCQVKVFCDKGAERK-TR
DEERR-KDKSKP

>Mollusc-GRH
YRYFLESPISTTQKIDEDRITYLNKSQYYGLTL-ENIN--TERIPKSATVKSIIMLVFRD
HKSPEDERKAFEFWHS--RQHSYKQRLLDVDIKNSQGIGPG----------SIEERAFN
AVVIKWNPREGQ----VKVNIAANCLSTDFSNQK--GVKGISLHVQIDTFEENHS-----
--VPIHR------------------------------GYCQIKVFCDKGAERK-TR
DEERRKTAKSKA
```

```
>Mammal-GRH
FEYTLEASKSLRQKPGDSTMTYLNKGQFYPITLKEVSS-SEGIHHPISKVRSVIMVVFAE
DKSREDQLRHWKYWHS--RQHTAKQRCIDIADYKESFNTIS----------NIEEIAYN
AISFTWD-INDE----AKVFISVNCLSTDFSSQK--GVKGLPLNIQVDTYSYNNRS----
-NKPVHR------------------------------AYCQIKVFCDKGAERK-IR
DEERK-QSKRKV

>Fish-GRH
FQYTLEASKSLRQKQGEGPMTYLNKGQFYAITLSETSA-NKRLRHPISKVRSVVMVVFSE
DKNRDEQLKYWKYWHS--RQHTAKQRVLDIADYKESFNTIG----------NIEEIAYN
AVSFTWD-LNEE----AKIFITVNCLSTDFSSQK--GVKGLPLMIQIDTYSYNNRS----
-NKPLHR------------------------------AYCQIKVFCDKGAERK-IR
DEERK-QNRKKT

>Lancelet-GRH
FEYILEAPKSLRQKPGEASMSYVNKGQFYAVTLTEAGS-SP-WRHHSTKVRSVIQVVFGD
GRSEEEQLKHWRYWHA--RQHTARQRIIDMADYKESTNVIE----------NIDEIAHN
AIAFSWD-VRET----GKVFISVNCLSTDFSSQK--GIKGLPLNLQIDTYTDYFKG----
-ATPVHR------------------------------AYVQIKVFCDKGAERK-IR
DEVRK-ISKKKQ

>Tunicate-GRH
FEYAMEAPKSLKQKDGEPTMSYINKGQFYCISLRECAG-RP-WRYKNTRVTSVVQIVFGD
GKPEDEQLRHWKYWHA--RQHTAKQRIIDIADYKESC-MIS----------DIDEFAHN
AISFNWD-VNDV----AKIFVSCNCLSTDFSAQK--GIKGLPLLLQIDTYMDNRRG----
-AAPAHR------------------------------GMCQLKVFCDKGAERK-IR
DEERK-AMRKRQ

>Cnidarian-GRH
YTFILEAPTSIVQRRGDDTLTYLNKGQFYAIDFEGNFD-PPSTEEDIIRVKSVVHLVFRD
EKDPRAELEHWNYWHS--QQPNPQQRAFDIDR-KSCQNIDE----------NITDQAYN
AAGFTWS-PHLN----AKIVIRINCLSTDFSPQK--GVKGIPLHLQIDTYEDVDN----P
DAEPVHR------------------------------AFCQIKVFRDKGAERK-NK
DESKS-AERRMQ

>Placozoan-GRH
YKFSLEAPASIVHKWTGDALTYINKGQFYNINFEA----TPESNPSTTRLKSILHLVFRE
EKEPDNEMSHWQYWYS--QQPNPNQRAFDIDR-KSCQNVEE----------RIEEIAYN
AVAFYWN-PADN----AKIAARINCLSTDFSPQK--GVKGIPLHLQIDTYEDLTS----S
DVEPVHR------------------------------AFCKVKIFRDKGAERK-NK
DESKT-ARKRIQ

>Demosponge-GRH
YLIILGAPTSIAQRQGEDTLTYLNKGQFYSIYFRAN-----NEMVLPSQAKSVIHLSFLD
ESDRTVEQSHWQYWYE--LQANPNQKAFDIDR-KNCEGLIE----------KPLDLGYN
AASFIWD-PRLG----ARVVLRINCLSTEFSGQK--GVKGLPLHVVVDTYEFQDNDRMTE
HDEPSHR------------------------------AYCRVKIFRDKGAERK-NK
DETKS-VERRLQ

>Nucleariid-GRH
YSFILDAPTSIAQKLEEGTLTYVNKGQAYAVTFEGMRGRRGSAGELPHTVKSIIHLVFHD
EHDQKNERGLWEYWRW--QQP-PTLRAMEVDR-KTCSGLTD-----------INELAFN
AFEFKWN-PREG----GKIVVRVNCLSTEFSTQK--GVKGMPLRIQIDTYENVTEP--YE
TSMPVSR------------------------------DYCQIKVFRDKGSERK-SK
DEAKS-AEKKLL
```

```
>C.globosum-CP2
FHTALNAPTAMIKHADEIPVTYLNKGQAYSLSIVDTMP--TLPIVPGTRFRTFVRVSFED
EQQRAKPGVCWSLWKEGRGTNEAHQRGGKLQAVEYVEAGQP-AEGDDKKTRIELETSSFD
GFSVIWTPGVNGTAVECNIAVRFNFLSTDFSHSK--GVKGIPVRLCAKTHLFPADGSSPS
DTANA-------------------------------PEICYCKVKLFRDHGAERK-LS
NDVAHIRKTID-

>M.grisea-CP2
FHSTLNAPTAMIKHADEIPVTYLNKGQAYSLSIVDTTP--TIPIQPGTRFRTFVRVSFED
EQQRQKPGVCWSLWKEGRGTNEAHQRGGKLQAVEYVEAGQP-AEGDDKRTRIELETSSFD
GFSVIWTPGIHG-AAECTIAVRFNFLSTDFSHSK--GVKGIPVRLCAKTSALPLDNSQTS
PDPAA-------------------------------TEICYCKVKLFRDHGAERK-LA
NDVQHVKKTID-

>F.graminearum-CP2
FHSTLNAPTAMIKHADEIPVTYLNKGQAYSLSVADTNA--TMPVAPGTKYRTFVRVSFED
EQQRQRPGVCWGLWKEGRGTNEAHQRGGKLQAVEYVEAGQP-AEGDDKRTRVELESSSFD
GFCVTWTPGING-PPEVNIAVRFNFLSTDFSHSK--GVKGIPVRLCAKTHPIPCDPSQPA
ADAN--------------------------------PEICYCKVKLFRDHGAERK-LS
NDVAHVKKSID-

>N.haematococca-CP2
FHSTLNAPTAMIKHSDEIPVTYLNKGQAYSLSVADTNA--TMPVAPGTKYRTFVRVSFED
DQQRQRPGVCWGLWKEGRGTNEAHQRGGKLQAVEYVEAGQP-AEGDDKRTRVELESSSFD
GFCVTWTPGING-PPEVNIAVRFNFLSTDFSHSK--GVKGIPVRLCAKTHPVPCDPSQPA
ADAN--------------------------------PDICYCKVKLFRDHGAERK-LS
NDVAHVKKSID-

>H.jecorina-CP2
FHATLNAPTAMVKHAAEIPVTYLNKGQAYTLSIMDTGV--TLPVSPGTKYRTYVRISFED
EQQRQKPGVCWSLWKEGRGTNEAHQRGGRLQAVEYVEAGQP-AEGDDKRTRVELESSSFD
GFSVIWTPGANG-AAEANIPVRFNFLSTDFSHSK--GVKGIPVRLCAKTGVFASDQLPSP
PDAA--------------------------------PETCYCKVKLFRDHGAERK-LS
NDVAHVKKSID-

>N.crassa-CP2
FHTTLNAPTAMIKNTDEIPVTYLNKGQAYSLSVVDTAP--TLPIVPGTRFRTFVRISFEE
EKQRHKPGMCWSLWKEGRGTNEAHQRGGKLQAVEFVEATQP-AEGDDKRTRIELESASFD
GFSVIWTPGING-SVECNIAVRFNFLSTDFSHSK--GVKGIPVRLCAKTQPYLPNSPQSP
NTSDG-------------------------------AEICYCMVKLFRDHGAERK-LS
NDVAHVRKNIE-

>B.fuckeliana-CP2
FQAVLNAPTAMVKNSQEIPVTYLNKGQAYSVSILDTEA--GHPLQPGTRYRTFVRISFED
EQQRQRPASCWQLWKEGRGTNEAHQRGGKLQAVEYVESTQI-GESDEKRTRMELDTASFD
GFSVIWTPAANC-VPECNIAVRFNFLSTDFSHSK--GVKGIPVRFCAKTETLS---SGSP
HTKEA-------------------------------SEVAFCKVKLFRDHGAERK-LS
NDIAHIKKTID-

>A.nidulans-CP2
YNVTLRAPTAMINHQNEIPVTYLNKGQAYSLSVVDTAP--PQTTSQPVKYRTFVRVSFQD
DEQRSKPAACWQLWKEGRGTSEAHQRGGKLQAVEFVDPTQG-NVEDQKNRQIQLESSSFD
GFCVTWTANPTTKASDCAISVRFNFLSTDFSHSK--GVKGIPVRLCAKTEMVA-GGSTGE
SS-NE--------------------------------AEVCFCKVKLFRDHGAERK-LS
NDVAHVKKTIE-
```

```
>A.niger-CP2
YHVTLRAPTAMINHQAEIPVTYLNKGQAYSVSVIDSTP--PPMTTQPIKYRTFIRVSFQE
DEQRAKPAACWQLWKEGRGSNEAHQRGGKLQAVEYVDPIQG-GIEDTKNRQIQLESSSFD
GFCVTWTLNPSTGVSECSIPVRFNFLSTDFSHSK--GVKGIPVRLCAKTEMVS-PDDTSS
TSGKE--------------------------------SEVCYCKVKLFRDHGAERK-LS
NDVAHVKKTIE-


>M.graminicola-CP2
YHATLNAPTAMIKHADEIPITYLNKGQAYTLNVVDTQV--QHIMPG-MKYRTFVRVSFED
EQQRQKPAACWQLWKEGRGTNEAHQRGGRLQAVEYVDPGQLCGADDPGKPKLELECASFD
GFCVTWTPAP--GAAECPISVRFNFLSTDFSHSK--GVKGIPVRLCAKTEMIE-DPSGNA
SKVQE--------------------------------SEVCYCKVKLFRDHGAERK-LS
NDVAHVKKTID-

>C.immitis-CP2
FHATLRASTAMVKDPDEIPVTYLNKGQAYTINIIDTAP--MASGSQQLRYRTYIRVSFEE
EEQRSKPASCWQLWKEGRGTNEAHQRDGKLLAVEHVDPNQG-GAGDGRHPRVQLEKANFD
GFSVLWVPSRTTGNPECAISVRFNFLSTDFSHSK--GVKGIPVRLCAKTEVIA-TSSSEP
PLGDS--------------------------------PEVCYCKVKLFRDHGAERK-LS
NDVAHIKKLME-

>R.oryzae-CP2
FQAILHSPTAITKKEDEKPITYLNRGQSYLLDL--TSS--TE---QRGILTSTISIEFHE
PIHRKAAESYWRFWLGEQMTTE-------ARAIGLEE--------SQTTGIFNVRYPSFD
QISFDWYGCFG-----ANIYIRFYCLSTDFSRIK--GVKGIPMRIMVETTARYDQIPENL
SLFTGTFNHKSKDKYE--------------------YVERCFCQIKLFRDKGAERK-NK
DDTKQINKQME-

>P.blakesleeanus-CP2
FYIGLEAPTAAAQKLEESPLTYLNKGQYYSVTLKDTET--SH---ADQIVKSTIIIMFND
ESHRKVAQSYWKFWLSQQKDAQS------ARAIDIGNDSSY--NTSRSSGVHNAEYSSFD
RIAFEWNTKKG-----AVINIRFNCLSTDFSRIK--GVKGIPLRLQMGTQVVGEQH----
------------------------------------IEKAYCRIKLFRDKGAERK-NK
DDAKHIERKN--

>Choanoflagellate-CP2
FHAAQPMLNCHPLCYQLVNVTYMAQTQPYALQLSSSTS---------QQLTSRLTLRIMD
IVKAAQAQTAIEDWHK----SHPDQTMLEYVQHMSQSVDHIQIG--------HEADQHLG
CISFDWEGRQR-----ASFGFQINCLSTTFVSGR-GGQAGVTLALRVDTALKASGVP---
------------------------------------VHSCFCAIKVFAGKGIIRKRLQ
SDLAKLTKAGL-
```

Up-regulated genes from the *Neurospora ghh* AHC microarrays

Broad ID	Gene name or Description	Fold (wt value)	FDR	Myc Down (> 1.5 Fold)	Myc Up (>1.5 Fold)	Virulence, Defense, Detoxification	Cell Wall Synthesis / Morphology	Membrane Transport / Cellular Import	Metabolism (Carbohydrates)	Metabolism (Amino Acids)	Metabolism (Nitrogen, Sulfur, Selenium)	Metabolism (Lipid, Fatty Acid, Isoprenoid)	Protein Fate / Modification	Environmental Interaction / Homeostasis	Secreted (TargetP prediction)
NCU00701	lysozyme	2.62 (856)	5.83E-03	■		■		■				■			1
NCU02476	alcohol dehydrogenase III	2.3 (1507)	1.28E-03			■		■							
NCU00755	*ethionine resistance protein*	2.71 (3671)	1.10E-03	■		■			■						
NCU05885	flavin-binding monooxygenase	3.12 (558)	7.62E-04			■									
NCU01095	MFS drug transporter	3.22 (1567)	3.62E-04			■			■	■					
NCU03428	Cyanovirin-N domain-containing protein	3.99 (374)	1.16E-04			-									
NCU09420	hsp20-like	3.54 (587)	1.08E-04	■		■									
NCU07232	hsp30	7.04 (91)*	1.07E-05	■		■									
NCU07130	endo-beta-1,4-D-xylanase	2.42 (490)	1.45E-03					■							3
NCU09050	FAD binding domain-containing protein	2.21 (136)	1.62E-03												
NCU10015	methanesulfonate monooxygenase	2.61 (2671)	3.56E-03					■			■	■			
NCU08364	choline sulfatase	2.4 (18596)	8.15E-04	■				■			■				
NCU07819	alpha-ketoglutarate-dependent taurine dioxygenase	6.33 (541)	4.01E-06								■				
NCU10016	thermophilic desulfurizing enzyme family protein	4.36 (1578)	1.55E-04								■				
NCU05888	dibenzothiophene desulfurization enzyme A	4.1 (1729)	3.81E-04								■				
NCU07610	taurine dioxygenase	3.87 (205)	8.18E-05								■				
NCU07821	dimethylaniline monooxygenase	2.7 (202)	1.67E-03								■				
NCU05340	alkanesulfonate monooxygenase	2.49 (805)	5.59E-03	■							■				
NCU06041	aryl sulfatase-1	2.11 (139)	3.53E-03								■	■			3
NCU04474	sulfite oxidase	2.01 (4336)	7.72E-03								■				
NCU01195	amination-deficient	2.73 (2952)	1.27E-03				■			■					
NCU03235	sulfate permease II	2.08 (2919)	5.51E-03					■		■	■			■	
NCU04433	cysteine-14	1.94 (7738)	5.89E-03	■				■			■			■	
NCU08325	phosphorus-5	4.18 (133)	1.63E-04					■						■	
NCU07820	pantothenate transporter	5.21 (358)	2.50E-05					■							
NCU09678	MFS transporter	3.85 (2155)	1.11E-04					■							
NCU02195	high affinity methionine permease	3.18 (83)*	2.02E-04					■							
NCU00290	ABC transporter	3.12 (258)	1.85E-04					■							
NCU05884	MFS transporter Seo1	2.58 (152)	5.95E-03					■							
NCU04435	general amino acid permease / AGP3	2.35 (628)	3.55E-03	■				■							
NCU07861	related to maltose permease	2.28 (68)*	1.45E-03					■							
NCU07609	MFS transporter	2.17 (203)	2.56E-03					■							
NCU00803	MFS transporter	2.16 (528)	2.69E-03					■							
NCU05069	FAD dependent oxidoreductase	2.05 (17320)	7.82E-03					■							
NCU02502	S-adenosyl-methionine-sterol-C	2.33 (1635)	1.70E-03												
NCU06468	midasin	2 (1824)	5.54E-03										■		
NCU05762	related to monophenol monooxygenase (tyrosinase)	4.34 (10871)	7.79E-05							■					
NCU02500	clock-controlled pheromone CCG-4	2.86 (105)	3.45E-04	■								■		■	2
NCU04533	abundant perithecial protein	3.76 (56355)	2.16E-04		■										
NCU09345	no message in thiamine-1	2.35 (2367)	7.06E-03				nucleotide metabolism								
NCU04197	CipC protein	6.34 (15978)	9.72E-06		■		unclear function								
NCU05887	oxidoreductase	4.3 (1150)	1.80E-04				unclear function								
NCU05886	DUF895 domain membrane protein	4.24 (2583)	7.64E-04				unclear function								
NCU01057	taurine catabolism dioxygenase TauD	3.88 (3306)	9.78E-05				unclear function								
NCU00305	short chain dehydrogenase	3.26 (3455)	2.73E-04				unclear function								
NCU07899	CAIB/BAIF family protein	2.44 (510)	1.03E-03				unclear function								
NCU03501	von Willebrand factor	2.24 (5735)	1.90E-03				unclear function								
NCU01112	lipid binding protein	2.04 (933)	8.17E-03				unclear function								3
NCU02267	mitochondrial protein Fmp25	1.97 (3106)	9.06E-03				unclear function								
NCU01163	GAF domain nucleotide-binding protein	1.85 (806)	8.80E-03				unclear function								
NCU03918	related to spliceosome-associated protein SAP-49	2.14 (1759)	4.52E-03				unclear function								

Down-regulated genes from the *Neurospora ghh* MYC microarrays

Broad ID	Gene Name or Description	Fold	FDR	Category / Secreted (TargetP prediction)
NCU01873	related to cellobiose dehydrogenase	-3.16	6.33E-05	Metabolism (Carbohydrates); Secreted 1
NCU00838	3-dehydroshikimate dehydratase	-1.75	8.23E-03	Metabolism (Carbohydrates)
NCU04419	2-oxoglutarate-dependent ethylene/succinate-forming enzyme	-1.79	5.57E-03	Metabolism (Carbohydrates)
NCU05855	O-methyltransferase	-2.27	5.40E-04	Metabolism (Carbohydrates)
NCU07338	alpha-1,6-mannosyltransferase Och1	-2.05	1.61E-03	Metabolism (Carbohydrates); Protein Fate / Modification
NCU04823	NADP-dependent alcohol dehydrogenase C	-2.89	1.06E-04	AHC Down; Cell Wall Synthesis / Morphology; Metabolism (Carbohydrates)
NCU09904	glucan 1,3-beta-glucosidase	-1.92	3.72E-03	Cell Wall Synthesis / Morphology; Metabolism (Carbohydrates)
NCU11983	zinc-binding alcohol dehydrogenase	-1.93	2.32E-03	Cell Wall Synthesis / Morphology
NCU04528	laccase precursor	-5.53	3.24E-06	AHC Down; Virulence, Defense, Detoxification; Metabolism (Carbohydrates); Secreted 1
NCU04510	aldose reductase	-2.21	6.09E-04	Metabolism (Carbohydrates); Metabolism (Nitrogen, Sulfur, Selenium); Metabolism (Lipid, Fatty Acid, Isoprenoid); Environmental Interaction / Homeostasis
NCU02704	branched-chain alpha-keto acid dehydrogenase E2 component	-1.74	9.17E-03	AHC Down; Metabolism (Carbohydrates); Metabolism (Nitrogen, Sulfur, Selenium); Metabolism (Lipid, Fatty Acid, Isoprenoid)
NCU06009	oxidoreductase	-1.83	5.86E-03	Membrane Transport / Cellular Import; Metabolism (Carbohydrates); Metabolism (Nitrogen, Sulfur, Selenium); Environmental Interaction / Homeostasis
NCU08880	neutral amino acid permease	-2.02	3.59E-03	Membrane Transport / Cellular Import; Environmental Interaction / Homeostasis
NCU06281	P-type ATPase	-1.98	2.23E-03	Membrane Transport / Cellular Import
NCU09564	phosphorus-4	-3.29	3.51E-05	AHC Up; Membrane Transport / Cellular Import; Environmental Interaction / Homeostasis; Secreted 1
NCU07621	zinc-regulated transporter 1	-2.63	3.26E-04	Membrane Transport / Cellular Import; Environmental Interaction / Homeostasis
NCU04433	cysteine-14	-1.83	4.20E-03	AHC Up; Membrane Transport / Cellular Import; Metabolism (Nitrogen, Sulfur, Selenium)
NCU04452	cytoplasmic ribosomal protein-5	-2.19	7.31E-04	Membrane Transport / Cellular Import
NCU09443	related to ABC transporter	-1.90	3.64E-03	Membrane Transport / Cellular Import
NCU00790	high affinity potassium transporter-1	-1.73	8.16E-03	Membrane Transport / Cellular Import
NCU04076	related to Cu-transporting P1-type ATPase	-3.76	3.21E-05	Membrane Transport / Cellular Import
NCU05650	related to protein carrier KAP123	-1.78	9.29E-03	Membrane Transport / Cellular Import; Protein Fate / Modification
NCU02680	nuclear protein localization protein 4	-1.84	4.14E-03	Membrane Transport / Cellular Import; Protein Fate / Modification
NCU00104	heat shock protein 98	-2.01	2.04E-03	AHC Up; Virulence, Defense, Detoxification; Membrane Transport / Cellular Import; Protein Fate / Modification
NCU02630	heat shock protein 78	-1.97	6.52E-03	Virulence, Defense, Detoxification; Membrane Transport / Cellular Import; Protein Fate / Modification
NCU00755	related to ethionine resistance proteins	-2.25	2.34E-03	Virulence, Defense, Detoxification; Membrane Transport / Cellular Import
NCU05770	peroxidase/catalase 2	-2.23	8.91E-04	Virulence, Defense, Detoxification; Membrane Transport / Cellular Import
NCU08429	related to integral membrane protein / pth 11	-2.79	1.66E-04	Virulence, Defense, Detoxification; Secreted 1
NCU09364	hsp30-like protein	-2.11	3.54E-03	Virulence, Defense, Detoxification
NCU07287	related to DNA damage-responsive protein 48	-1.80	6.95E-03	Virulence, Defense, Detoxification
NCU00281	UDP-glucose,sterol transferase	-2.17	1.13E-03	Virulence, Defense, Detoxification; Metabolism (Lipid, Fatty Acid, Isoprenoid)
NCU08372	triacylglycerol lipase	-2.40	2.32E-03	Metabolism (Lipid, Fatty Acid, Isoprenoid)
NCU09692	phosphatidic acid phosphatase beta	-3.21	4.36E-05	Metabolism (Lipid, Fatty Acid, Isoprenoid)
NCU00461	NAD-specific glutamate dehydrogenase	-3.49	3.03E-05	Metabolism (Amino Acids); Metabolism (Nitrogen, Sulfur, Selenium)
NCU02084	arginase	-3.41	3.37E-05	Metabolism (Amino Acids)
NCU08127	related to DFG5 protein	-2.59	8.82E-04	Cell Wall Synthesis / Morphology; Secreted 2
NCU02500	clock-controlled pheromone CCG-4	-2.20	8.75E-04	AHC Up; Environmental Interaction / Homeostasis; Secreted 2
NCU07082	aspartyl-tRNA synthetase	-3.36	3.45E-05	translation
NCU04635	mismatched base pair and cruciform dna recognition protein	-3.19	5.21E-05	DNA repair
NCU07953	alternative oxidase-1	-2.28	8.93E-04	aerobic respiration
NCU03552	related to zinc finger protein odd-paired-like (opl)	-2.36	1.85E-03	transcription factor
NCU08507	zinc finger protein zpr1	-2.23	9.87E-04	transcription factor
NCU09758	Ras guanine-nucleotide exchange protein	-2.24	6.16E-04	signaling
NCU03369	related to pirin	-1.71	8.21E-03	transcription; Secreted 1
NCU06586	AN1 zinc finger protein	-2.51	4.70E-04	unclear function
NCU06883	FHA domain-containing protein SNIP1	-1.96	2.45E-03	unclear function

Up-regulated genes from the *Neurospora ghh* MYC microarrays

Broad ID	Gene Name or Description	Fold	FDR	Secreted (TargetP prediction)
NCU04118	probable aspartate kinase	1.7429	9.41E-03	
NCU08011	conserved hypothetical protein	1.6711	6.64E-03	
NCU03796	related to branched-chain alpha-ketoacid dehydrogenase kinase	1.7582	4.19E-03	
NCU07732	carbamoyl-phosphate synthase small chain, arginine-specific	1.78	4.34E-03	
NCU08771	related to BENZOYLFORMATE DECARBOXYLASE	2.043	1.77E-03	
NCU10031	related to monophenol monooxygenase	2.4521	1.83E-04	
NCU01348	conserved hypothetical protein	2.5014	1.09E-04	3
NCU09798	aryl-alcohol dehydrogenase	1.9945	2.63E-03	
NCU09209	probable galactose oxidase precursor	4.0978	2.24E-06	
NCU09775	probable alpha-L-arabinofuranosidase	3.7056	5.99E-06	1
NCU08199	related to reticuline oxidase (berberine bridge enzyme)	2.2751	3.08E-04	
NCU09267	related to glyoxal oxidase precursor	2.2239	3.16E-04	
NCU04395	beta-1,6-glucanase Neg1	2.1264	6.26E-04	2
NCU06153	related to ACB 4-hydroxyacetophenone monooxygenase	2.0508	4.86E-03	
NCU02252	probable phosphoglyceromutase	1.6809	6.69E-03	
NCU00187	related to carboxyphosphonoenolpyruvate phosphonomutase	1.7187	6.84E-03	
NCU02328	related to O-methyltransferase	1.7093	7.44E-03	
NCU09170	probable alpha-N-arabinofuranosidase	1.6632	9.04E-03	1
NCU01788	related to C4-dicarboxylate transport protein mae1	1.7508	4.16E-03	
NCU07607	related to carboxylic acid transporter protein	1.8011	4.54E-03	
NCU06042	related to soluble fumarate reductase (NADH)	1.7603	8.72E-03	1
NCU00195	related to nicotinamide mononucleotide permease	2.4078	2.27E-04	
NCU10368	conserved hypothetical protein	2.6922	2.27E-04	
NCU01568	related to PET8 protein	2.0616	8.80E-04	1
NCU07884	probable oxaloacetate/sulfate carrier	2.4401	1.06E-03	
NCU04446	related to transporter protein	1.7343	6.49E-03	
NCU07894	related to sexual differentiation process protein	3.0468	5.78E-05	
NCU09909	probable urea active transporter	1.8869	2.07E-03	
NCU05046	Na+-transporting ATPase ENA-1 (sodium P-type ATPase ENA-1)	2.0315	1.29E-03	
NCU04533	related to calcium-dependent cell adhesion molecule-1	3.935	2.49E-06	
NCU07742	related to NAD kinase	2.5545	9.90E-05	
NCU00829	related to ferric reductase FRE2 precursor	2.0363	1.23E-03	
NCU09772	related to HETEROKARYON incompatibility protein	1.8236	5.61E-03	
NCU11307	related to CCC1 protein (involved in calcium homeostasis)	1.6917	7.71E-03	
NCU09933	related to large-conductance mechanosensitive channel	2.5685	1.01E-04	
NCU07787	probable SnodProt1 PRECURSOR	3.0644	1.32E-05	1
NCU09210	related to peroxidase DyP	3.0307	1.55E-05	
NCU07232	related to heat shock protein 30	2.5413	2.28E-04	
NCU07569	related to heatshock protein Hsp150	2.1707	3.74E-04	1
NCU07769	related to integral membrane protein PTH11	2.1984	4.49E-04	1
NCU00355	catalase-3	2.368	2.24E-04	2
NCU03646	related to beta-1,3 exoglucanase precursor	1.8354	5.02E-03	3
NCU09185	related to pisatin demethylase cytochrome P450	1.84	4.09E-03	2
NCU00701	probable lysozyme	2.4287	1.70E-04	1
NCU08852	related to NAD+ ADP-ribosyltransferase	2.3737	2.27E-04	
NCU05598	rhamnogalacturonase B precursor	1.87	2.26E-03	1
NCU09473	related to 3-oxoacyl-[acyl-carrier-protein] reductase	1.9773	1.32E-03	
NCU04621	related to lipoxygenase 1	1.7098	5.94E-03	
NCU06460	related to lysophosphatidic acid phosphatase	1.6842	9.04E-03	
NCU05858	related to linoleate diol synthase	2.4048	1.74E-04	
NCU09169	conserved hypothetical protein	1.8677	2.22E-03	
NCU06726	probable nitrilase (NIT3)	1.7897	3.56E-03	
NCU09102	related to glu/asp-tRNA amidotransferase subunit A	1.6165	9.13E-03	1
NCU01605	probable MOB1 protein	1.7172	5.75E-03	
NCU09155	related to secreted aspartic protease (barrierpepsin precursor)	2.1576	1.19E-03	1
NCU10048	3-methyl-2-oxobutanoate hydroxymethyltransferase	1.9596	1.94E-03	
NCU08907	related to blastomyces yeast phase-specific protein 1	4.9772	1.14E-07	2
NCU00751	related to 7alpha-cephem-methoxylase P8 chain	1.7016	6.66E-03	
NCU03576	probable hymA gene	1.6703	8.74E-03	
NCU06416	related to gibberellin 20-oxidase	1.741	8.76E-03	
NCU03968	related to DNA repair protein MMS21	1.9134	2.95E-03	
NCU01097	related to nitrate assimilation regulatory protein nirA	1.9014	1.78E-03	
NCU05518	probable peroxisomal amine oxidase (copper-containing)	1.8928	1.87E-03	
NCU03255	related to phytase	2.9949	1.94E-05	
NCU08423	related to bromodomain protein BDF1	2.6435	9.21E-05	

Column categories (matrix headers): AHC Down (> 1.5 Fold); AHC Up (>1.5 Fold); Virulence, Defense, Detoxification; Cell Wall Synthesis / Morphology; Membrane Transport / Cellular Import; Metabolism (Carbohydrates); Metabolism (Amino Acids); Metabolism (Nitrogen, Sulfur, Selenium); Metabolism (Lipid, Fatty Acid, Isoprenoid); Protein Fate / Modification; Environmental Interaction / Homeostasis; Secreted (TargetP prediction).

Gene Ontology classes for the *Drosophila grh* microarrays

GO ID	Molecular Function	p	Bonferroni p
42302	structural constituent of cuticle	9.66E-19	1.02E-15
5214	structural constituent of chitin-based cuticle	2.20E-17	2.32E-14
4252	serine-type endopeptidase activity	7.24E-17	7.63E-14
17171	serine hydrolase activity	1.46E-16	1.54E-13
8236	serine-type peptidase activity	1.75E-16	1.85E-13
70011	peptidase activity, acting on L-amino acid peptides	8.26E-12	8.71E-09
4175	endopeptidase activity	1.86E-11	1.96E-08
8010	structural constituent of chitin-based larval cuticle	4.56E-11	4.81E-08
8233	peptidase activity	1.11E-10	1.17E-07
8061	chitin binding	1.76E-10	1.85E-07
1871	pattern binding	4.51E-10	4.75E-07
30247	polysaccharide binding	4.51E-10	4.75E-07
16614	oxidoreductase activity, acting on CH-OH group of donors	2.64E-09	2.79E-06
30246	carbohydrate binding	7.23E-09	7.62E-06
3796	lysozyme activity	3.58E-08	3.77E-05
16616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	3.11E-07	3.28E-04
16491	oxidoreductase activity	4.20E-07	4.43E-04
4497	monooxygenase activity	4.22E-07	4.45E-04
5506	iron ion binding	4.30E-07	4.53E-04
15020	glucuronosyltransferase activity	5.34E-07	5.63E-04
4553	hydrolase activity, hydrolyzing O-glycosyl compounds	1.50E-06	1.58E-03
61135	endopeptidase regulator activity	2.66E-06	2.80E-03
5198	structural molecule activity	4.07E-06	4.29E-03
30414	peptidase inhibitor activity	4.82E-06	5.08E-03
16798	hydrolase activity, acting on glycosyl bonds	5.26E-06	5.54E-03
46906	tetrapyrrole binding	5.27E-06	5.56E-03
61134	peptidase regulator activity	7.00E-06	7.38E-03
20037	heme binding	8.01E-06	8.45E-03
4866	endopeptidase inhibitor activity	8.79E-06	9.27E-03
9055	electron carrier activity	1.21E-05	1.28E-02
4364	glutathione transferase activity	2.36E-05	2.49E-02
15291	secondary active transmembrane transporter activity	2.42E-05	2.55E-02
15101	organic cation transmembrane transporter activity	2.59E-05	2.73E-02
5550	pheromone binding	2.89E-05	3.05E-02
4857	enzyme inhibitor activity	2.98E-05	3.14E-02
46873	metal ion transmembrane transporter activity	4.48E-05	4.72E-02
4568	chitinase activity	5.18E-05	5.46E-02
22804	active transmembrane transporter activity	8.21E-05	8.65E-02
4867	serine-type endopeptidase inhibitor activity	8.63E-05	9.10E-02
8194	UDP-glycosyltransferase activity	1.50E-04	1.58E-01
16765	transferase activity, transferring alkyl or aryl (other than methyl) groups	1.95E-04	2.06E-01
16490	structural constituent of peritrophic membrane	1.97E-04	2.08E-01
19842	vitamin binding	2.41E-04	2.54E-01
16758	transferase activity, transferring hexosyl groups	2.53E-04	2.67E-01
5549	odorant binding	4.34E-04	4.57E-01

GO ID	Biological Process	p	Bonferroni p
5975	carbohydrate metabolic process	3.69E-10	1.06E-06
6030	chitin metabolic process	7.32E-10	2.11E-06
6952	defense response	9.29E-10	2.67E-06
9607	response to biotic stimulus	1.04E-09	2.98E-06
6022	aminoglycan metabolic process	1.93E-09	5.55E-06
51707	response to other organism	2.21E-09	6.35E-06
6955	immune response	5.44E-09	1.57E-05
5976	polysaccharide metabolic process	7.55E-09	2.17E-05
6959	humoral immune response	1.87E-08	5.40E-05
6950	response to stress	2.90E-08	8.36E-05
9308	amine metabolic process	4.03E-08	1.16E-04
19730	antimicrobial humoral response	4.87E-08	1.40E-04
35080	heat shock-mediated polytene chromosome puffing	1.55E-07	4.48E-04
35079	polytene chromosome puffing	1.55E-07	4.48E-04
51704	multi-organism process	8.07E-07	2.32E-03
34605	cellular response to heat	1.07E-06	3.07E-03
6508	proteolysis	1.30E-06	3.75E-03
2376	immune system process	2.75E-06	7.91E-03
45087	innate immune response	3.32E-06	9.55E-03
19731	antibacterial humoral response	4.09E-06	1.18E-02
55114	oxidation reduction	4.19E-06	1.21E-02
16052	carbohydrate catabolic process	7.83E-06	2.25E-02
9617	response to bacterium	8.65E-06	2.49E-02
6635	fatty acid beta-oxidation	1.21E-05	3.47E-02
42742	defense response to bacterium	1.51E-05	4.36E-02
50802	circadian sleep/wake cycle, sleep	1.55E-05	4.46E-02

GO ID	Cellular Component	p	Bonferroni p
5576	extracellular region	3.64E-10	1.96E-07
5792	microsome	1.60E-06	8.63E-04
42598	vesicular fraction	1.60E-06	8.63E-04
5626	insoluble fraction	3.71E-06	2.00E-03
5624	membrane fraction	4.93E-06	2.65E-03
267	cell fraction	6.27E-06	3.37E-03
5615	extracellular space	1.39E-05	7.45E-03
42579	microbody	1.20E-04	6.45E-02
5777	peroxisome	1.20E-04	6.45E-02
43190	ATP-binding cassette (ABC) transporter complex	1.75E-04	9.40E-02
5956	protein kinase CK2 complex	4.13E-04	2.22E-01

Misregulated genes from the *Drosophila grh* microarrays

symbol	f(Grh/Cbs)	FDR
5-HT1A	-2.4007	3.47E-03
a10	8.8011	1.74E-05
AcCoAS	2.5146	1.81E-03
achi	-2.1918	5.21E-03
Acox57D-d	7.7106	2.86E-05
Acox57D-p	2.0263	5.19E-03
Acp1	-4.428	5.02E-04
Acp65Aa	-3.2849	1.38E-03
Act88F	2.4295	1.53E-03
Adgf-D	1.8127	7.62E-03
Adh	4.7558	1.21E-04
agt	-2.2972	4.12E-03
alpha-Man-IIb	-2.4083	3.60E-03
Alr	1.8216	8.12E-03
Ama	-2.0803	6.12E-03
amd	-5.4582	3.05E-04
Amy-d	-5.8569	2.39E-04
Amy-p	-11.5433	3.31E-05
ana1	-2.198	4.99E-03
Ank2	-1.9784	7.70E-03
antdh	-3.7649	7.79E-04
Appl	1.9961	4.28E-03
Arc1	2.8512	2.39E-03
Arc2	3.5005	3.29E-04
Arc42	1.8588	6.50E-03
Asn-synthetase	-3.6624	1.32E-03
ast	-10.2244	4.91E-05
Ate1	1.9566	4.97E-03
AttA	2.0225	5.19E-03
AttB	6.4445	5.62E-05
AttC	6.6309	4.31E-05
baz	-2.3141	3.96E-03
be	29.5529	1.38E-06
beat-Vb	-1.9656	7.93E-03
Best2	-3.0405	1.41E-03
beta4GalNAcTB	1.9958	5.06E-03
betaNACtes6	2.0344	5.34E-03
betaTry	24.1207	2.13E-06
BG642312	-2.5348	2.79E-03
Bmcp	10.8862	1.02E-05
Bsg	-2.2813	4.23E-03
bw	1.919	5.29E-03
by	-6.284	1.91E-04
Cad74A	-2.3562	4.05E-03
Cad96Ca	-1.911	9.35E-03
Cad99C	-4.0075	6.87E-04
cag	-2.1055	6.09E-03
CalpA	-2.1178	5.70E-03
CalpC	-2.4082	3.35E-03

symbol	f(Grh/Cbs)	FDR
Cap-G	-1.9993	7.21E-03
capt	2.1266	3.60E-03
capu	-2.469	3.05E-03
Cat	2.2457	2.44E-03
Cct2	-2.0787	6.07E-03
Cda4	-2.1253	6.23E-03
Cda9	3.7948	2.43E-04
Cdk5alpha	2.1075	3.23E-03
cenG1A	1.8555	6.68E-03
cer	1.7996	8.09E-03
CHKov1	-2.9363	1.61E-03
Cht3	-12.9022	2.48E-05
Cht5	1.7832	8.48E-03
Cht9	2.4606	1.43E-03
Ciao1	1.8422	7.01E-03
cngl	-3.416	1.13E-03
Cog7	1.8599	6.49E-03
Cp1	-2.3146	3.96E-03
cpo	2.118	3.69E-03
Cpr11B	1.8981	5.96E-03
Cpr31A	-2.5022	2.95E-03
Cpr47Ea	1.8393	8.29E-03
Cpr47Eb	5.3321	8.76E-05
Cpr47Ed	3.9123	2.19E-04
Cpr47Eg	-5.5376	2.79E-04
Cpr49Ad	35.786	8.68E-07
Cpr49Af	-5.6622	2.75E-04
Cpr49Ah	1.8521	6.72E-03
Cpr57A	-32.8685	4.27E-06
Cpr5C	-8.7172	7.31E-05
Cpr60D	-51.8265	1.20E-06
Cpr64Aa	2.46	1.48E-03
Cpr64Ac	-2.0909	5.95E-03
Cpr65Au	-2.066	6.47E-03
Cpr65Aw	-2.0349	6.89E-03
Cpr65Ay	9.555	1.61E-05
Cpr67Fa1	-11.0038	4.05E-05
Cpr72Ec	2.2858	2.77E-03
Cpr73D	1.9888	6.30E-03
Cpr76Bc	-2.2002	4.82E-03
Cpr92F	-2.5025	2.92E-03
CR14033	-4.9161	3.78E-04
CR31044	2.7557	8.87E-04
CR31292	-2.6571	2.50E-03
Cry	3.417	3.81E-04
CycH	1.7373	9.88E-03
Cyp1	-2.5611	2.70E-03
Cyp12d1-d	-22.2229	9.99E-06
Cyp12d1-p	-4.0193	7.24E-04

symbol	f(Grh/Cbs)	FDR
Cyp301a1	1.817	7.51E-03
Cyp309a2	28.1979	1.58E-06
Cyp311a1	-4.1071	6.29E-04
Cyp312a1	-2.1103	5.94E-03
Cyp4p1	-10.7188	4.31E-05
Cyp4p2	-87.1892	4.20E-07
Cyp6a17	-56.7846	8.68E-07
Cyp6a20	5.852	6.56E-05
Cyp6a22	1.9627	4.99E-03
Cyp6a8	1.893	9.56E-03
Cyp6g1	-2.0262	7.11E-03
Cyp6v1	1.8043	8.38E-03
Cyp6w1	9.1282	1.67E-05
Cyp9b1	2.0636	3.49E-03
cype	-4.0058	6.85E-04
d	-4.413	5.45E-04
dally	-2.0405	7.01E-03
DAT	-2.669	2.36E-03
Dbp45A	-3.045	1.41E-03
Dcp1	1.9197	5.96E-03
deltaTry	3.931	6.21E-04
Dg	-4.812	4.49E-04
Dgp-1	-3.0988	1.35E-03
dgt3	1.7746	9.05E-03
Dip-C	-2.4725	3.02E-03
DJ-1alpha	-2.5044	3.29E-03
dnc	1.7327	9.98E-03
Dot	3.2179	5.06E-04
DptB	4.1949	1.70E-04
dro2	5.489	7.38E-05
dro5	138.0037	1.33E-07
Drs	6.5984	4.31E-05
Dscam3	-6.2128	2.03E-04
Dsp1	1.918	5.38E-03
dyn-p25	-6.4004	1.73E-04
ea	-6.59	1.63E-04
eater	-1.9367	8.79E-03
Ect3	-2.454	3.93E-03
Edg91	2.2585	2.19E-03
EfTuM	-2.8229	3.32E-03
eIF4E-3	-1.8892	9.56E-03
eIF4E-5	1.7404	9.81E-03
Elo68beta	-2.6615	2.26E-03
Elp2	-2.2245	4.74E-03
epsilonTry	2.9576	6.83E-04
Esp	3.1782	5.09E-04
Est-Q	2.5005	1.97E-03
exu	-2.3927	3.47E-03
fal	1.8627	6.30E-03

symbol	f(Grh/Cbs)	FDR
fas	-2.2893	4.87E-03
fbp	1.7964	8.00E-03
Fcp3C	1.7549	9.30E-03
fh	-1.929	9.46E-03
Fhos	-3.673	9.04E-04
FK506-bp1	-2.0243	8.43E-03
Fmo-1	-9.3469	7.17E-05
Fmr1	1.9374	5.75E-03
fon	2.0192	4.32E-03
fra	2.233	2.38E-03
fred	-2.9331	1.65E-03
frtz	3.503	3.43E-04
fru	1.8319	7.29E-03
fry	1.9291	5.59E-03
fs(1)h	2.0716	3.73E-03
FucTC	3.3819	3.72E-03
GalNAc-T1	-3.2339	1.16E-03
gammaTry	6.9784	4.31E-05
Gas8	-2.7081	2.18E-03
Gen	2.3533	1.80E-03
Gfat1	-11.419	3.50E-05
gfzf	-10.0385	4.58E-05
gkt	-10.1977	4.36E-05
GLaz	-2.7785	1.94E-03
GlcAT-P	-2.6436	2.42E-03
Gli	-1.9922	8.61E-03
Glycogenin	-5.5003	2.85E-04
GNBP3	1.9585	5.55E-03
Gr59a	-2.1381	5.86E-03
GRHRII	2.5552	1.24E-03
GstD3	1.8907	6.69E-03
GstD5	-2.2103	4.83E-03
GstE4	-2.1621	5.26E-03
GstE5	-2.2227	6.11E-03
GstE9	1.9494	5.17E-03
Gyc76C	1.7445	9.67E-03
Herp	1.8229	7.40E-03
Hf / CG18806	-5.3696	3.03E-04
hgo	-18.7434	1.13E-05
His2A:CG31618	6.6368	4.10E-05
His2B:CG17949	3.6484	2.80E-04
His2B:CG33868	2.2994	1.98E-03
HLHmgamma	2.551	1.22E-03
Hmu	1.8144	7.59E-03
hoe2	1.7398	9.74E-03
HP6	2.3805	1.68E-03
Hrb87F	-4.2083	5.83E-04
Hsp26	3.7634	2.47E-04
Hsp60B	-3.0579	1.97E-03

symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR
Hsp60D	-5.8457	2.42E-04	Lcp65Aa	-4.9398	3.91E-04	neb	2.4199	1.92E-03	PGRP-LC	2.6347	1.06E-03
Hsp70Aa	7.4784	3.13E-05	Lcp65Ab1	-6.6432	1.65E-04	nemy	-6.1475	2.02E-04	PGRP-LD	2.6314	1.08E-03
Hsp70Ba	5.5068	7.73E-05	Lcp65Ab2	-4.4488	5.24E-04	ninaA	-3.718	8.87E-04	PGRP-SB2	3.1247	5.07E-04
Hsp70Bb	7.5429	2.69E-05	Lcp65Ad	-9.9028	5.80E-05	ninaD	-2.8904	1.68E-03	PH4alphaSG1	-2.7403	2.30E-03
Hsp70Bc	4.7508	1.11E-04	Lcp65Ae	-3.1701	1.31E-03	nompC	-2.0106	7.47E-03	phr6-4	-11.9984	3.13E-05
Idgf3	1.9267	6.76E-03	Lcp65Af	2.3428	2.70E-03	Nox	-1.9777	9.39E-03	pigs	1.9103	5.47E-03
Ilp6	6.0282	6.11E-05	Lcp65Ag3	-2.458	3.08E-03	Npc2d	1.8048	7.93E-03	pirk	2.8668	7.21E-04
IM1	-58.9451	9.45E-07	lectin-24Db	9.1249	1.67E-05	nrm	1.9569	4.75E-03	ple	3.7707	2.53E-04
IM10	1.9746	6.84E-03	lectin-28C	-2.3099	4.25E-03	NT1	-9.5533	5.44E-05	png	2.3002	2.17E-03
IM2	7.9458	3.45E-05	Lk	-2.6046	2.50E-03	Ntl	-3.1509	1.28E-03	Ppat-Dpck	1.8004	8.02E-03
IM23	5.6878	7.10E-05	Lmpt	-2.1718	8.02E-03	numb	-2.0621	7.67E-03	proPO-A1	-2.29	4.14E-03
IM3	5.447	7.45E-05	Lnk	1.8606	6.60E-03	Obp46a	-2.6378	2.53E-03	ps	3.9467	2.07E-04
IM4	4.3485	1.88E-04	loco	-3.3059	1.16E-03	Obp47a	1.9368	5.33E-03	Psf3	-2.3088	4.05E-03
ImpL1	-6.454	2.26E-04	Lpin	1.9764	4.40E-03	Obp56a	-2.3249	4.07E-03	pst	-2.2316	4.57E-03
ImpL2	1.9573	9.80E-03	Lsd-1	2.6299	1.06E-03	Obp56d	-2.3811	3.54E-03	Ptp99A	-2.3837	3.59E-03
InR	-2.3841	3.69E-03	LSm7	-87.0595	4.20E-07	Obp57d	-6.5647	1.67E-04	pum	1.8096	7.88E-03
iotaTry	2.568	1.26E-03	Lsp1gamma	2.5027	1.38E-03	Obp57e	-2.1235	5.59E-03	Pvr	1.8762	6.43E-03
IP3K1	1.7987	8.64E-03	LvpH	2.2305	2.86E-03	Obp59a	2.3464	2.07E-03	Pxn	-2.6226	2.91E-03
Ir40a	-2.7613	7.01E-03	LysB	1.8448	6.80E-03	Obp83g	-7.1452	1.22E-04	qm	1.84	6.88E-03
Ir56a	1.9355	5.01E-03	LysD	11.6928	7.15E-06	Obp8a	-2.7539	2.10E-03	Qtzl	2.147	2.88E-03
Ir76a	2.239	2.30E-03	LysE	2.4384	1.50E-03	Obp99b	-3.7299	1.12E-03	r-cup	-2.0121	7.11E-03
Ir87a	4.9691	9.93E-05	LysS	4.1716	1.70E-04	Obp99c	-2.2526	4.75E-03	Rab1	-2.5653	2.56E-03
Irk2	2.2398	2.54E-03	LysX	5.8216	6.24E-05	Obp99d	-1.9116	9.92E-03	Rala	13.0023	5.15E-06
Iswi	2.1433	3.03E-03	lz	3.3217	3.91E-04	obst-A	-5.6724	2.64E-04	raw	1.747	9.96E-03
Jarid2	-2.6705	3.54E-03	m	-3.7689	1.33E-03	obst-B	-4.8657	3.88E-04	Rbp9	1.9393	4.98E-03
Jhe	4.0498	2.06E-04	Madm	-2.8624	1.81E-03	obst-E	-2.5614	2.77E-03	rec	-2.1267	6.72E-03
Jheh2	1.9939	4.67E-04	Mec2	2.2401	2.66E-03	obst-G	1.7785	9.48E-03	Rep	-28.5954	6.93E-06
jim	1.7493	9.41E-03	MED11	-4.9287	3.80E-04	Or19a	10.3114	1.08E-05	Rh50	2.0096	5.07E-03
jing	2.9328	6.66E-04	MED16	-3.1562	1.26E-03	Orct2	1.826	7.49E-03	Rh6	4.7549	1.16E-04
Jon25Bi	8.7151	2.17E-05	MED8	-1.9982	7.36E-03	Os-C	1.7874	8.58E-03	RhoGAP100F	2.2671	2.18E-03
Jon25Bii	2.4003	1.62E-03	mei-S332	-2.2684	4.26E-03	Osi15	-3.1504	1.38E-03	RhoGAP18B	2.7774	9.04E-04
Jon25Biii	2.5057	1.43E-03	mey	-6.1844	1.98E-04	Osi18	-3.6007	1.65E-03	Rnp4F	-2.0524	9.21E-03
Jon65Aiii	-1.8917	9.49E-03	mfrn	-2.2517	4.91E-03	Osi19	-3.9223	1.32E-03	rost	1.8469	6.87E-03
kappaTry	2.0372	4.13E-03	mid	-2.2224	5.29E-03	Osi2	3.621	2.99E-04	RPA2	-2.5474	2.72E-03
kek4	-4.3565	5.26E-04	mmy	-2.6289	2.37E-03	Osi20	-3.2621	1.89E-03	Rpb4	-2.4421	3.14E-03
klu	2.1877	9.07E-03	MP1	1.7466	9.94E-03	Osi9	-3.7253	3.46E-03	RpL3	-5.3138	3.13E-04
knk	-5.2881	3.19E-04	mre11	-7.7369	9.77E-05	par-1	2.4931	1.40E-03	RpL7	-3.153	1.34E-03
kraken	2.0498	4.02E-03	mRpL48	-35.7219	3.47E-06	path	-2.0638	8.03E-03	RpS19b	1.7956	8.83E-03
Ku80	-2.1057	5.78E-03	mthl3	-5.9138	2.36E-04	Pbgs	1.9072	5.70E-03	RpS21	2.9888	6.05E-04
l(1)G0469	1.983	6.30E-03	mthl8	-44.5384	1.68E-06	Pbprp2	2.3821	1.93E-03	RpS23	-2.892	1.68E-03
l(2)34Fc	-3.8053	7.59E-04	Mtk	3.2993	4.05E-04	Pcd	-2.9244	1.65E-03	rt	-2.733	2.10E-03
l(3)07882	-12.2626	2.73E-05	MtnB	-2.5166	4.29E-03	Pde11	2.3198	2.83E-03	rtp	-2.0921	6.20E-03
l(3)mbn	-2.1831	4.95E-03	MtnD	-3.8236	9.28E-04	Pde1c	-2.1127	7.20E-03	ru	-1.9692	8.53E-03
lace	-6.1655	2.07E-04	mtTFB1	-2.1532	5.54E-03	Pdh	-2.6552	2.32E-03	Rx	-2.7611	1.99E-03
LBR	3.5342	3.24E-04	Muc18B	2.12	3.78E-03	pen-2	2.0718	3.63E-03	sa	1.7564	9.25E-03
Lcp1	-3.2349	1.18E-03	Muc91C	-2.6279	2.48E-03	pex10	2.1998	2.50E-03	sala	-7.6503	1.09E-04
Lcp2	-6.6495	1.60E-04	Myo28B1	1.9034	5.60E-03	pex12	1.8947	5.90E-03	Sb	2.07	4.89E-03
Lcp3	-14.339	3.64E-05	Ncc69	-9.6438	5.20E-05	pgant5	-4.512	4.87E-04	Scgdelta	-3.1975	1.32E-03
Lcp4	-139.4136	2.36E-07	Ndae1	-2.3302	4.00E-03	PGRP-LB	2.127	2.97E-03	ScpX	1.9416	5.10E-03

symbol	f(Grh/Cbs)	FDR
scrib	1.8311	7.13E-03
scu	2.0191	4.03E-03
scyl	1.9957	4.57E-03
sec13	1.9526	4.89E-03
SelG	-8.9454	6.43E-05
Ser6	-2.6632	2.29E-03
Ser7	3.7001	2.80E-04
sfl	-2.3344	3.82E-03
Sgs5	10.2811	1.16E-05
Sh	-2.0457	7.89E-03
shep	3.0056	6.89E-03
sick	1.996	4.63E-03
Sirt6	3.7951	2.43E-04
Skeletor	1.7901	8.15E-03
skpE	1.808	7.84E-03
Slob	-2.3807	3.58E-03
sls	4.5178	1.33E-04
SMSr	1.8699	6.29E-03
Snoo	-3.3213	1.10E-03
Sop2	2.2634	2.34E-03
Sp212	2.1546	3.84E-03
Sp7	2.5284	1.27E-03
SPE	2.0644	4.04E-03
spel1	-6.5397	1.73E-04
Spn	2.5999	1.45E-03
Spn1	2.1381	4.06E-03
Spn100A	-5.3854	2.94E-04
Spn43Aa	-2.4049	3.63E-03
Spn43Ad	2.7862	8.27E-04
Spn47C	-22.9042	8.99E-06
Spt6	-1.9581	8.49E-03
sr	-2.2892	4.36E-03
Sry-alpha	1.8738	6.86E-03
Ssl2	3.4858	3.35E-04
ssp	-2.0612	6.47E-03
ssp3	-2.1613	6.21E-03
st	-3.8417	7.57E-04
stau	-3.2053	1.20E-03
Ste:CG33238	2.1302	2.93E-03
Ste:CG33240	1.772	8.82E-03
Ste:CG33245	1.8494	6.64E-03
Ste:CG33246	3.7598	2.57E-04
Su(fu)	-2.8379	1.80E-03
Su(Ste):CR4082	1.8358	7.66E-03
SuUR	-1.9926	7.51E-03
T3dh	1.8773	6.06E-03
Takr99D	-1.9824	7.57E-03
Tango2	2.0702	3.52E-03
TepII	-2.1932	5.88E-03

symbol	f(Grh/Cbs)	FDR
Tequila	46.3219	5.32E-07
thetaTry	-2.3616	3.73E-03
tim	-2.2518	5.10E-03
Tim17a2	-2.2413	4.53E-03
Tim17b2	3.2253	5.35E-04
toc	2.135	3.29E-03
Top1	-5.7064	2.53E-04
Top3alpha	2.1712	2.89E-03
TotA	2.5053	1.34E-03
TotC	5.5376	7.52E-05
toy	2.7418	8.89E-04
TpnC25D	-2.0466	6.63E-03
Tsf1	2.9177	6.96E-04
Tsp42Ec	-1.906	9.20E-03
tutl	-3.3814	1.02E-03
TwdlF	-3.3223	1.79E-03
TwdlG	-4.0183	6.66E-04
TwdlH	3.5528	4.30E-04
TwdlQ	-5.0851	3.72E-04
TwdlT	-1.9101	9.30E-03
U26	-2.8851	1.73E-03
Ugt36Ba	2.7348	9.76E-03
Ugt36Bb	-2.1872	4.97E-03
Ugt37a1	4.474	1.37E-04
Ugt37c1	-2.5583	2.70E-03
Ugt86Dd	1.847	9.71E-03
Ugt86Dh	1.8183	7.46E-03
Use1	1.8099	7.74E-03
Vdup1	-6.0853	2.16E-04
Vha100-4	-6.7154	1.56E-04
Victoria	-3.9856	7.06E-04
vlc	1.9557	5.88E-03
w	-3.4478	9.76E-03
wal	1.789	8.35E-03
Wnt2	1.8725	6.39E-03
Wnt6	1.918	5.29E-03
yellow-d2	-6.3166	1.90E-04
yellow-e2	-13.9196	2.13E-05
yin	-5.7767	2.41E-04
yip2	2.4196	1.54E-03
Zasp52	1.7935	8.67E-03
zfh1	6.659	4.31E-05
zpg	2.2071	2.63E-03
CG10005	-2.0873	5.92E-03
CG10096	-3.7854	7.75E-04
CG10097	-2.6384	3.79E-03
CG10133	-2.2918	4.14E-03
CG10140	2.8361	7.52E-04

symbol	f(Grh/Cbs)	FDR
CG10182	9.0209	1.88E-05
CG10195	3.8925	2.25E-04
CG10200	-2.1595	5.19E-03
CG10202	3.2082	4.75E-04
CG10514	-10.0324	5.32E-05
CG10570	-121.9309	2.36E-07
CG10585	-2.039	7.11E-03
CG10591	1.7673	8.94E-03
CG10638	1.8816	5.92E-03
CG10664	-2.5261	2.97E-03
CG10672	2.7778	8.34E-04
CG10725	2.2218	2.45E-03
CG1077	7.6715	3.13E-05
CG10799	2.3844	1.66E-03
CG10814	1.9961	5.67E-03
CG10877	-3.8855	7.30E-04
CG10887	-3.2621	1.15E-03
CG10898	-9.1653	5.99E-05
CG10907	-2.7232	2.07E-03
CG10914	-2.2443	5.14E-03
CG10924	4.7247	1.31E-04
CG10953	-2.7807	2.87E-03
CG11000	1.8588	7.00E-03
CG11034	13.8662	4.61E-06
CG11055	-2.036	6.66E-03
CG11147	-2.3055	4.17E-03
CG11159	2.2436	2.52E-03
CG11170	-2.6432	2.34E-03
CG11200	-2.8419	1.93E-03
CG11208	-12.582	2.36E-05
CG11211	2.5827	1.14E-03
CG11289	-2.0533	6.70E-03
CG11309	2.3267	1.90E-03
CG11349	-2.0703	6.88E-03
CG11360	1.8704	6.27E-03
CG11370	4.1813	1.73E-04
CG11378	1.7503	9.44E-03
CG11381	1.9504	4.78E-03
CG11395	-5.0547	3.56E-04
CG11400	1.8954	5.87E-03
CG11413	6.641	4.25E-05
CG11425	3.0103	5.87E-04
CG1143	1.8781	6.09E-03
CG11436	-3.4695	9.76E-04
CG11437	-2.5787	2.79E-03
CG11470	1.801	9.49E-03
CG11550	-2.8872	1.68E-03
CG11570	6.1836	5.62E-05
CG11585	3.2691	4.96E-04

symbol	f(Grh/Cbs)	FDR
CG11592	2.0249	4.58E-03
CG11600	2.3257	1.97E-03
CG11668	2.9843	6.19E-04
CG11672	1.7841	8.49E-03
CG11786	-3.3882	1.51E-03
CG11791	-3.215	1.38E-03
CG11796	-47.8162	1.38E-06
CG11811	2.1325	3.01E-03
CG11843	-3.1125	1.63E-03
CG11854	2.5658	1.18E-03
CG11878	-2.5109	4.09E-03
CG11880	-2.8049	1.87E-03
CG11883	1.9862	4.30E-03
CG11885	1.9054	6.62E-03
CG11889	-2.5724	3.31E-03
CG11892	-3.7421	7.97E-04
CG11893	17.9359	3.78E-06
CG11911	3.1233	5.85E-04
CG11912	-42.3251	2.37E-06
CG11951	9.0232	1.67E-05
CG12009	-2.4049	6.86E-03
CG12011	-5.4945	2.81E-04
CG12018	2.9718	6.17E-04
CG12057	-5.983	2.64E-04
CG12129	-7.4214	1.12E-04
CG12177	3.1738	4.95E-04
CG12229	1.8293	7.64E-03
CG12256	3.3153	3.94E-04
CG12262	1.8393	6.98E-03
CG1236	-2.0352	6.68E-03
CG12370	2.406	1.62E-03
CG12375	1.8427	6.94E-03
CG12428	3.228	4.80E-04
CG12484	2.9285	6.60E-04
CG12512	4.2389	1.77E-04
CG12523	-2.4977	3.32E-03
CG12581	-2.7266	2.11E-03
CG12643	-3.9105	7.15E-04
CG12645	-2.0283	6.96E-03
CG1265	-4.8618	4.07E-04
CG12655	1.8745	7.00E-03
CG12766	-2.3432	3.84E-03
CG12826	3.4684	3.41E-04
CG12971	2.0604	4.36E-03
CG12998	-29.8885	5.91E-06
CG13031	2.5239	1.27E-03
CG1304	-440.8433	4.10E-08
CG13041	-4.8114	4.02E-04
CG13048	-2.4181	3.33E-03

symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR
CG13049	-2.0338	7.51E-03	CG13827	-2.1508	5.56E-03	CG14820	-2.147	5.69E-03	CG16743	-4.1884	5.91E-04
CG13059	-2.9801	1.53E-03	CG13830	1.9723	4.49E-03	CG14823	-9.0246	6.43E-05	CG16772	4.4431	1.38E-04
CG13060	-4.6811	4.42E-04	CG13893	1.8444	6.77E-03	CG14851	3.3959	4.02E-04	CG16782	-2.7829	1.95E-03
CG13064	-2.67	2.31E-03	CG13896	2.0873	3.36E-03	CG14866	3.4645	3.46E-04	CG16836	4.9795	1.01E-04
CG13065	-2.623	2.38E-03	CG13905	2.0615	7.07E-03	CG14935	2.2566	2.42E-03	CG16898	7.7307	2.65E-05
CG13077	2.1824	3.55E-03	CG13912	1.901	5.96E-03	CG14962	1.9435	5.18E-03	CG16947	-2.728	2.18E-03
CG13078	1.922	5.21E-03	CG13936	-2.3899	3.49E-03	CG14963	-1.9382	8.99E-03	CG1695	2.4444	1.52E-03
CG13086	2.0366	3.93E-03	CG13946	18.7504	3.03E-06	CG14984	-4.2707	5.99E-04	CG16959	1.882	6.24E-03
CG1309	-2.3124	3.96E-03	CG13957	-1.9933	8.51E-03	CG14997	-2.3969	3.45E-03	CG1698	-2.7006	2.17E-03
CG13102	3.83	2.48E-04	CG13982	1.8198	7.44E-03	CG15065	4.0075	1.95E-04	CG16984	-2.1286	5.51E-03
CG13157	3.085	5.50E-04	CG14050	1.7463	9.88E-03	CG15067	8.2501	2.25E-05	CG16997	-6.9819	1.38E-04
CG13159	-1.9436	9.56E-03	CG14054	-2.5209	3.50E-03	CG15068	3.082	8.77E-04	CG17018	-2.5774	2.59E-03
CG13175	3.0188	6.21E-04	CG14059	4.6997	1.61E-04	CG15073	-2.6863	2.29E-03	CG1702	2.486	1.49E-03
CG13177	1.888	5.85E-03	CG1407	1.7549	9.58E-03	CG15155	1.8849	6.03E-03	CG17036	1.9325	5.84E-03
CG1319	1.746	9.95E-03	CG14096	2.062	4.83E-03	CG15211	1.773	9.06E-03	CG17047	1.9009	5.92E-03
CG13196	-2.2909	4.11E-03	CG14100	-3.2777	1.16E-03	CG15226	2.4263	1.91E-03	CG17104	-2.5609	2.88E-03
CG13203	3.6019	3.06E-04	CG14105	-17.4412	1.67E-05	CG15253	-7.1849	1.35E-04	CG17107	-5.3389	3.09E-04
CG13215	1.8816	6.20E-03	CG14110	-1.8991	9.92E-03	CG15254	-2.3503	4.10E-03	CG17118	1.9392	5.04E-03
CG13227	2.2405	3.86E-03	CG14132	1.9431	5.26E-03	CG15279	-2.1559	5.85E-03	CG17121	-2.6256	2.36E-03
CG13228	8.0274	2.36E-05	CG14135	-3.3259	1.21E-03	CG15282	-124.7027	3.22E-07	CG17127	-2.1538	5.36E-03
CG13258	-2.6287	2.50E-03	CG14147	-2.8707	1.77E-03	CG15293	2.5231	1.28E-03	CG17129	-2.2503	5.10E-03
CG13288	-2.0347	7.14E-03	CG14196	1.8381	7.07E-03	CG15308	9.0359	1.84E-05	CG1718	2.0191	3.89E-03
CG13305	-2.3791	3.92E-03	CG14205	2.3259	1.91E-03	CG15317	-2.4108	3.41E-03	CG17244	2.0587	5.36E-03
CG13306	1.7731	9.03E-03	CG14222	-1.9091	9.11E-03	CG15394	12.092	6.31E-06	CG17264	1.8578	6.60E-03
CG13311	-3.3208	1.19E-03	CG14223	-2.1718	5.29E-03	CG15414	2.0291	3.80E-03	CG17298	-2.5874	2.60E-03
CG13313	1.9195	5.76E-03	CG14253	1.8503	6.78E-03	CG15431	-2.5723	2.88E-03	CG17325	-477.1214	1.48E-08
CG13314	2.8713	7.24E-04	CG14265	10.5541	1.13E-05	CG15515	-28.3676	7.15E-06	CG17328	2.401	1.62E-03
CG13315	2.7943	1.62E-03	CG14273	2.4105	1.59E-03	CG15530	-1.9714	8.07E-03	CG17329	4.0106	2.20E-04
CG13330	2.2482	2.25E-03	CG14285	-1.9835	8.40E-03	CG15597	-1.9855	7.95E-03	CG17341	-2.1367	8.46E-03
CG13337	-5.2386	3.21E-04	CG14324	-4.2567	9.58E-04	CG15653	-2.4399	3.18E-03	CG17386	2.0442	3.82E-03
CG13377	-1.9222	8.86E-03	CG14326	-3.635	1.38E-03	CG15661	2.014	6.75E-03	CG17470	4.0506	1.88E-04
CG13386	-2.1261	5.53E-03	CG1434	-2.6988	2.25E-03	CG15695	-2.3766	3.60E-03	CG17549	1.9951	6.41E-03
CG13403	2.0647	3.65E-03	CG14369	1.8076	7.79E-03	CG15705	3.2233	4.40E-04	CG17562	1.8189	7.43E-03
CG13422	7.1191	3.36E-05	CG14418	2.7934	8.07E-04	CG15712	-3.8464	7.48E-04	CG17571	3.59	3.56E-04
CG13427	-2.4136	3.42E-03	CG14419	-3.8677	7.19E-04	CG1572	2.4018	1.66E-03	CG17618	-3.412	9.93E-04
CG13449	2.8637	7.61E-04	CG14439	2.1529	2.82E-03	CG15773	-2.3383	3.81E-03	CG17633	-1.9023	9.28E-03
CG13488	-13.1057	2.90E-05	CG14478	1.768	9.24E-03	CG15818	-2.1895	5.13E-03	CG17636	1.9156	6.60E-03
CG13562	-4.0248	6.49E-04	CG14495	-2.0094	8.43E-03	CG15822	-3.4321	9.71E-04	CG17669	-2.1181	5.88E-03
CG13565	-4.4363	5.18E-04	CG14515	-3.5733	9.48E-04	CG15887	-3.2896	1.72E-03	CG17707	-3.3636	1.03E-03
CG13606	2.2057	3.96E-03	CG14527	-3.2321	1.18E-03	CG15888	2.3434	1.93E-03	CG1773	-4.5416	5.22E-04
CG13627	-3.8702	1.45E-03	CG14535	-2.5881	2.64E-03	CG15917	2.1383	3.04E-03	CG17739	-3.1097	1.31E-03
CG13658	-2.0495	7.09E-03	CG14564	8.2307	2.56E-05	CG1602	-2.0997	5.83E-03	CG1774	3.3685	4.05E-04
CG13663	-2.3908	3.53E-03	CG14572	1.8242	7.39E-03	CG1621	-2.2537	4.55E-03	CG17742	-13.8092	2.02E-05
CG13699	2.6487	1.06E-03	CG1461	-4.0297	6.60E-04	CG16704	-5.3937	2.94E-04	CG17843	-37.5609	2.76E-06
CG13728	-4.4398	5.02E-03	CG14621	2.0669	4.06E-03	CG16723	-1.9627	7.94E-03	CG17855	2.0886	3.35E-03
CG13742	2.6109	1.08E-03	CG14630	4.1603	1.71E-04	CG16727	1.8637	6.70E-03	CG17904	-2.1202	5.67E-03
CG13784	-2.0381	6.57E-03	CG14642	2.3505	1.82E-03	CG1673	-2.1537	5.84E-03	CG17919	-4.3104	5.41E-04
CG13795	-2.8769	1.76E-03	CG14694	2.4089	1.60E-03	CG16733	-3.7948	7.81E-04	CG17999	1.8725	6.99E-03
CG13813	-2.3407	6.20E-03	CG14741	2.0473	4.52E-03	CG16739	1.9385	6.53E-03	CG18067	4.7133	1.19E-04

symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR
CG18094	1.9894	4.36E-03	CG30121	-3.4752	9.78E-04	CG31778	-2.3766	3.73E-03	CG33255	2.2125	2.66E-03
CG18107	2.7606	1.10E-03	CG30151	2.7935	8.65E-04	CG31792	-1.9406	9.14E-03	CG33267	-3.2014	1.20E-03
CG18131	2.234	2.36E-03	CG30194	-2.9039	1.78E-03	CG31826	-5.6248	2.63E-04	CG33268	-3.571	9.21E-04
CG18170	-2.2763	4.48E-03	CG30197	-2.0765	6.28E-03	CG31862	1.9539	4.79E-03	CG33301	2.0981	3.27E-03
CG18180	2.6298	1.07E-03	CG30280	-2.1788	5.38E-03	CG31869	2.5155	1.37E-03	CG3335	-1.9654	8.13E-03
CG18249	-4.1503	6.41E-04	CG30285	1.9643	4.58E-03	CG31875	-3.0246	1.45E-03	CG3344	2.1949	2.52E-03
CG18278	-2.6675	2.37E-03	CG30334	2.3761	2.37E-03	CG31898	2.9405	6.56E-04	CG33459	6.6815	4.15E-05
CG18301	2.6357	1.05E-03	CG30345	2.0046	4.18E-03	CG31909	3.526	3.69E-04	CG33460	1.8932	5.72E-03
CG1835	3.4388	3.55E-04	CG30360	3.2854	6.68E-04	CG31918	-2.1458	5.45E-03	CG33468	-15.1777	1.74E-05
CG18367	-2.0435	6.79E-03	CG3038	1.8234	7.48E-03	CG31928	2.108	3.52E-03	CG33474	5.136	8.76E-05
CG18369	-2.1727	5.38E-03	CG30380	-3.6405	8.48E-04	CG31953	-4.6877	4.42E-04	CG3348	-1.9778	8.92E-03
CG18428	-2.0074	7.36E-03	CG30440	-6.1218	2.09E-04	CG31954	3.2517	4.25E-04	CG33489	2.7362	9.04E-04
CG18477	-2.1652	5.14E-03	CG30457	-8.9911	8.24E-05	CG31997	-3.022	1.49E-03	CG33493	2.0464	3.71E-03
CG18585	-1.9041	9.27E-03	CG30458	-2.2367	5.38E-03	CG32054	-2.01	7.03E-03	CG33510	2.5742	1.29E-03
CG18600	-2.2786	4.28E-03	CG30466	-7.4714	1.09E-04	CG32055	-2.1578	5.29E-03	CG33511	3.8426	2.33E-04
CG18607	3.5182	3.28E-04	CG30480	-2.1187	7.05E-03	CG32066	1.9062	6.16E-03	CG33723	3.7607	2.56E-04
CG18622	1.7487	9.45E-03	CG30488	8.3717	2.02E-05	CG32073	-9.7221	5.38E-05	CG33774	-2.1453	6.47E-03
CG18641	2.0334	3.88E-03	CG30492	-6.8306	1.53E-04	CG32082	1.9178	8.59E-03	CG3397	-2.3344	3.88E-03
CG18661	4.3249	1.60E-04	CG3071	-2.2281	4.75E-03	CG32091	-2.3941	3.53E-03	CG33977	-1.9063	9.51E-03
CG18749	-2.1372	6.14E-03	CG31002	-2.2651	4.29E-03	CG32100	-1.9682	7.93E-03	CG33993	-2.8536	1.77E-03
CG18754	-10.1854	4.31E-05	CG31029	2.5157	1.52E-03	CG32111	-2.784	1.99E-03	CG34007	-2.7499	2.21E-03
CG18787	1.7775	9.62E-03	CG31030	-2.7941	1.89E-03	CG32115	-2.4652	3.02E-03	CG34010	-2.76	2.15E-03
CG18788	-7.7698	9.85E-05	CG31054	-2.7722	1.99E-03	CG32155	-2.6744	2.26E-03	CG34032	-1.9836	7.89E-03
CG1942	2.0774	4.40E-03	CG3106	2.0944	3.26E-03	CG32214	1.8308	9.62E-03	CG34038	-2.4338	4.04E-03
CG1969	-4.3001	5.48E-04	CG31075	-2.6416	2.97E-03	CG32264	-2.0028	7.39E-03	CG34045	2.0234	4.00E-03
CG2004	1.7648	9.34E-03	CG31082	-2.5346	2.70E-03	CG32278	-2.1473	5.59E-03	CG34054	2.161	2.93E-03
CG2023	1.9429	5.14E-03	CG31097	-3.1314	1.43E-03	CG32284	5.568	7.66E-05	CG34056	-7.6251	1.05E-04
CG2070	2.5018	1.42E-03	CG31103	-32.3506	4.61E-06	CG32299	2.1667	2.69E-03	CG3408	-13.0417	2.32E-05
CG2233	1.8839	6.33E-03	CG31200	-4.0516	6.74E-04	CG32368	-2.9652	1.81E-03	CG34104	-1.9032	9.48E-03
CG2249	2.5325	1.35E-03	CG31219	3.2562	4.35E-04	CG32379	2.7406	8.67E-04	CG34120	-1.992	7.41E-03
CG2316	1.958	5.90E-03	CG31268	9.0064	1.74E-05	CG3238	-34.4813	2.20E-03	CG34166	4.7027	1.18E-04
CG2444	10.9626	1.08E-05	CG31321	-14.9128	2.02E-05	CG3239	2.3075	2.09E-03	CG34180	-2.5919	2.61E-03
CG2543	-2.0402	6.82E-03	CG31337	-2.5852	2.82E-03	CG32407	1.8838	5.89E-03	CG3419	-18.3009	1.21E-05
CG2604	1.9616	4.99E-03	CG31386	-2.075	6.19E-03	CG32444	-8.1808	8.53E-05	CG34195	-1.9047	9.37E-03
CG2616	-2.1671	5.30E-03	CG31436	6.496	4.69E-05	CG32557	1.8507	7.15E-03	CG34207	1.7977	8.13E-03
CG2650	-3.1656	1.31E-03	CG31437	-1.9424	9.74E-03	CG32808	1.8737	6.13E-03	CG34223	2.1725	3.04E-03
CG2767	-2.15	5.52E-03	CG31445	1.9766	4.39E-03	CG32816	1.9889	4.44E-03	CG34224	1.8612	6.83E-03
CG2862	-2.034	6.94E-03	CG31475	-2.638	2.32E-03	CG32821	2.0241	4.23E-03	CG34228	-2.0072	8.38E-03
CG2962	2.9193	1.05E-03	CG31477	-2.4226	3.29E-03	CG3285	-3.0423	1.41E-03	CG34242	-6.6328	1.59E-04
CG2974	1.805	8.51E-03	CG31523	1.8234	7.40E-03	CG32850	-2.7213	2.10E-03	CG34251	2.1747	2.63E-03
CG2983	-1.955	8.11E-03	CG31538	1.7438	9.60E-03	CG32984	-2.0252	6.98E-03	CG34253	-2.0543	6.66E-03
CG30000	2.3856	1.79E-03	CG31668	-1.932	9.05E-03	CG3301	-3.8267	7.50E-04	CG34276	-2.3249	4.25E-03
CG30005	2.2166	2.41E-03	CG31674	6.1547	5.56E-05	CG3303	-1.9779	8.21E-03	CG34281	4.6938	1.15E-04
CG30026	2.1004	3.36E-03	CG31689	2.2392	2.93E-03	CG33082	4.228	1.66E-04	CG34282	1.8171	9.88E-03
CG30049	2.0451	3.72E-03	CG31698	2.2789	2.20E-03	CG33127	1.8056	7.85E-03	CG34305	-31.6237	5.15E-06
CG30050	2.0445	3.77E-03	CG31739	-2.1195	6.15E-03	CG33128	1.8793	6.10E-03	CG34353	-3.0397	1.47E-03
CG30060	-2.124	5.54E-03	CG31759	4.1676	1.72E-04	CG33160	-1.9909	7.50E-03	CG34355	-1.9145	8.93E-03
CG30062	-2.0497	6.60E-03	CG31769	2.4413	1.79E-03	CG33178	-10.1107	4.63E-05	CG34370	-4.2832	5.54E-04
CG30080	7.1668	3.69E-05	CG31775	-3.3973	1.56E-03	CG33226	-2.2589	4.44E-03	CG34375	2.8507	7.93E-04

symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR
CG34376	1.8225	7.44E-03	CG42749	2.007	4.57E-03	CG5402	-3.655	8.51E-04	CG6733	-2.5435	2.84E-03
CG34383	-2.0051	8.32E-03	CG42784	-1.9198	8.83E-03	CG5421	-2.0739	6.12E-03	CG6738	-2.5143	2.80E-03
CG34409	1.8186	7.59E-03	CG42807	2.1339	4.37E-03	CG5493	4.7298	1.78E-04	CG6865	2.7429	9.31E-04
CG34417	1.8636	9.81E-03	CG42808	-1.8789	9.94E-03	CG5577	1.9567	4.95E-03	CG6983	-2.2232	6.20E-03
CG34420	-1.9225	9.41E-03	CG42822	-11.6307	3.08E-05	CG5612	2.4908	1.37E-03	CG7025	-1.9981	7.65E-03
CG34426	-1.8972	9.53E-03	CG42825	-9.0293	6.83E-05	CG5618	-2.2462	4.69E-03	CG7031	-5.3505	3.89E-04
CG34434	1.8386	7.64E-03	CG42827	2.0965	3.19E-03	CG5639	2.364	1.79E-03	CG7046	5.2816	8.53E-05
CG34436	-3.0517	1.42E-03	CG42856	2.2395	2.31E-03	CG5644	2.3159	1.95E-03	CG7051	-2.5622	2.75E-03
CG34437	-2.0142	7.91E-03	CG4288	4.832	1.10E-04	CG5707	2.752	8.75E-04	CG7054	-4.0874	6.41E-04
CG3502	-4.5471	4.72E-04	CG4302	6.5498	4.20E-05	CG5724	9.1254	1.74E-05	CG7059	3.3003	4.02E-04
CG3513	2.0816	3.44E-03	CG43066	-2.8038	1.93E-03	CG5731	-7.6948	1.01E-04	CG7080	-38.7306	3.03E-06
CG3523	2.2854	2.36E-03	CG43163	-4.0101	6.79E-04	CG5770	-2.3603	3.68E-03	CG7102	-3.8628	7.35E-04
CG3588	-4.9509	3.99E-04	CG4324	-4.0537	6.91E-04	CG5778	7.3151	3.17E-05	CG7110	2.1205	4.44E-03
CG3635	2.3586	1.76E-03	CG4367	2.5929	1.39E-03	CG5791	6.1737	5.56E-05	CG7214	-5.5671	2.80E-04
CG3699	-2.2172	4.92E-03	CG4386	-4.3524	7.19E-04	CG5802	1.9885	4.40E-03	CG7224	-2.9872	1.51E-03
CG3726	1.9401	4.91E-03	CG4389	2.4991	1.33E-03	CG5804	-1.967	9.06E-03	CG7252	-2.2416	4.48E-03
CG3734	-1.9512	8.58E-03	CG4393	3.1954	4.73E-04	CG5810	-3.5432	1.09E-03	CG7264	-1.8928	9.56E-03
CG3822	-2.5235	2.77E-03	CG4415	-2.9543	1.57E-03	CG5849	1.7653	8.97E-03	CG7294	-2.1709	5.39E-03
CG3841	1.9047	6.08E-03	CG4440	-2.823	1.91E-03	CG5866	-2.0159	6.99E-03	CG7296	2.2189	2.64E-03
CG3857	2.5388	1.25E-03	CG4542	-1.9231	9.04E-03	CG5883	-3.0588	1.38E-03	CG7330	-2.0456	6.82E-03
CG3868	-1.9643	8.59E-03	CG4563	1.8667	6.24E-03	CG5888	-2.3964	3.51E-03	CG7341	-3.2216	1.69E-03
CG3880	-10.1473	5.03E-05	CG4576	-3.1763	1.24E-03	CG5892	3.2473	4.61E-04	CG7365	2.5334	1.34E-03
CG3918	1.998	4.25E-03	CG4598	2.6198	1.09E-03	CG5953	1.9552	5.48E-03	CG7367	1.8962	5.83E-03
CG3940	1.926	5.67E-03	CG4623	1.8695	6.19E-03	CG5999	2.2844	2.05E-03	CG7377	-3.457	1.01E-03
CG4000	-6.3138	1.91E-04	CG4627	-2.1364	5.44E-03	CG6000	-4.2412	5.99E-04	CG7402	-2.6838	2.27E-03
CG40002	1.9413	7.84E-03	CG4645	-2.5854	2.57E-03	CG6023	1.8178	7.54E-03	CG7432	3.1531	4.96E-04
CG40006	-4.8099	4.22E-04	CG4688	1.7794	8.52E-03	CG6043	-1.9901	7.43E-03	CG7433	-2.5297	2.94E-03
CG4017	3.9717	2.08E-04	CG4734	-4.4506	4.98E-04	CG6055	2.1685	2.75E-03	CG7544	-5.2365	3.22E-04
CG40198	-33.1895	4.27E-06	CG4766	-1.9455	8.28E-03	CG6067	2.2046	3.16E-03	CG7548	-2.7546	2.01E-03
CG4038	-2.0649	6.30E-03	CG4815	2.0596	3.66E-03	CG6071	-3.4228	9.73E-04	CG7564	-2.1213	6.48E-03
CG40485	2.6373	1.08E-03	CG4830	2.4923	1.43E-03	CG6083	1.7911	8.13E-03	CG7567	3.4525	3.44E-04
CG40486	2.3972	1.63E-03	CG4839	2.4125	1.71E-03	CG6106	2.7237	8.99E-04	CG7582	3.0444	5.77E-04
CG4066	-1.9703	7.86E-03	CG4842	-4.1208	6.09E-04	CG6178	2.0509	4.05E-03	CG7601	2.7179	9.36E-04
CG41087	1.7904	8.21E-03	CG4860	-2.7656	2.01E-03	CG6188	4.4245	1.43E-04	CG7632	1.8682	7.61E-03
CG41443	-2.8932	4.00E-03	CG4914	-2.047	7.12E-03	CG6279	2.1372	3.10E-03	CG7724	-2.4769	2.99E-03
CG41520	-6.1352	2.06E-04	CG4927	3.6823	2.96E-04	CG6283	2.4809	1.52E-03	CG7741	-4.386	5.26E-04
CG4194	-2.1893	4.90E-03	CG4962	3.8706	2.22E-04	CG6296	2.123	2.99E-03	CG7742	-2.1138	5.71E-03
CG42249	3.0046	6.11E-04	CG4972	-3.0066	1.46E-03	CG6347	-2.7682	1.98E-03	CG7763	-1.9647	8.56E-03
CG42254	-1.9736	7.79E-03	CG4998	2.9712	6.21E-04	CG6405	-6.8199	1.48E-04	CG7778	2.0671	6.68E-03
CG42255	1.8289	7.39E-03	CG5126	3.2142	4.52E-04	CG6409	-2.3112	4.21E-03	CG7816	-2.614	2.63E-03
CG42272	-2.3528	3.81E-03	CG5130	-35.5187	3.63E-06	CG6426	2.0224	4.31E-03	CG7860	-3.5709	1.12E-03
CG42329	-4.1781	6.29E-04	CG5144	-2.4868	3.31E-03	CG6439	2.0506	4.45E-03	CG7888	-2.1631	5.15E-03
CG42335	24.4998	2.13E-06	CG5167	2.4387	1.54E-03	CG6472	2.2207	2.48E-03	CG7900	3.1982	4.66E-04
CG42345	-3.4077	1.00E-03	CG5171	-6.3978	2.59E-04	CG6486	-2.5204	2.80E-03	CG7912	1.9238	6.74E-03
CG42365	-8.5671	1.47E-04	CG5172	2.0411	3.76E-03	CG6512	-2.6462	2.30E-03	CG7950	-2.1142	6.03E-03
CG42389	-5.5201	2.87E-03	CG5246	1.9557	6.13E-03	CG6520	2.1826	2.60E-03	CG8012	1.8613	6.35E-03
CG42541	-1.9563	8.52E-03	CG5288	1.927	6.29E-03	CG6543	1.843	9.61E-03	CG8028	-4.517	5.24E-04
CG42562	-2.3262	4.02E-03	CG5316	2.3103	2.21E-03	CG6602	-2.1225	5.88E-03	CG8034	2.3605	2.06E-03
CG42669	1.7557	9.68E-03	CG5379	2.8088	8.37E-04	CG6707	2.2458	2.29E-03	CG8066	-3.2277	1.29E-03
CG42678	2.4626	1.53E-03	CG5380	-2.0598	6.40E-03	CG6719	-1.8776	9.96E-03	CG8086	-4.2508	5.71E-04

symbol	f(Grh/Cbs)	FDR	symbol	f(Grh/Cbs)	FDR
CG8105	4.0942	1.82E-04	CG9444	-3.1247	1.43E-03
CG8160	-1.8837	9.99E-03	CG9449	-3.4636	1.07E-03
CG8204	-2.8508	1.75E-03	CG9468	-2.5365	3.22E-03
CG8249	2.6232	1.07E-03	CG9509	2.2345	2.30E-03
CG8306	-2.0865	8.26E-03	CG9510	-5.7533	2.56E-04
CG8311	2.1774	3.14E-03	CG9512	1.8836	6.08E-03
CG8317	-2.0571	6.78E-03	CG9518	1.776	8.64E-03
CG8419	6.7993	4.10E-05	CG9527	-7.855	9.30E-05
CG8486	2.0822	3.78E-03	CG9531	-2.7854	2.03E-03
CG8550	1.7852	8.38E-03	CG9542	-2.8681	1.71E-03
CG8646	-2.2933	4.11E-03	CG9555	2.2073	2.80E-03
CG8690	2.2042	4.48E-03	CG9573	3.1973	5.15E-04
CG8778	-2.2189	4.89E-03	CG9577	2.1083	3.10E-03
CG8783	2.1208	3.02E-03	CG9616	2.901	7.17E-04
CG8785	-2.9722	1.52E-03	CG9641	-32.869	4.44E-06
CG8795	-7.7333	1.00E-04	CG9649	2.0058	4.25E-03
CG8801	-5.7434	2.50E-04	CG9664	-7.2987	1.20E-04
CG8833	-2.0704	6.60E-03	CG9672	2.5371	1.40E-03
CG8837	2.2061	3.23E-03	CG9682	5.7581	6.83E-05
CG8888	-2.2429	4.53E-03	CG9689	-13.586	2.10E-05
CG8889	1.7762	8.93E-03	CG9717	3.15	4.87E-04
CG8918	-12.1318	2.95E-05	CG9733	9.1433	1.81E-05
CG8960	-3.2132	1.46E-03	CG9757	-19.188	1.02E-05
CG9009	2.4898	1.39E-03	CG9821	1.8384	7.29E-03
CG9021	2.4626	1.59E-03	CG9837	1.7844	9.88E-03
CG9027	-3.8181	7.59E-04	CG9877	-517.2344	6.56E-09
CG9083	-4.4027	1.24E-03	CG9928	2.8572	7.66E-04
CG9104	-1.914	9.48E-03	CG9961	-2.0512	7.18E-03
CG9129	2.3631	1.90E-03	CG9967	-1.9754	9.10E-03
CG9153	2.2803	2.21E-03			
CG9171	3.4494	3.53E-04			
CG9173	-2.6266	2.38E-03			
CG9184	-5.2961	3.41E-04			
CG9192	2.0663	3.52E-03			
CG9231	1.9288	5.15E-03			
CG9259	3.7217	2.60E-04			
CG9270	3.4771	5.35E-04			
CG9312	-2.329	3.87E-03			
CG9313	-2.0348	8.69E-03			
CG9319	2.5835	1.15E-03			
CG9331	-1.9282	9.65E-03			
CG9339	1.8937	6.04E-03			
CG9360	-2.6281	2.81E-03			
CG9362	-2.3327	3.84E-03			
CG9363	-5.4922	2.91E-04			
CG9394	-2.3471	4.02E-03			
CG9396	2.2203	2.42E-03			
CG9400	-1.9209	8.83E-03			
CG9416	-1.9688	8.18E-03			
CG9427	-2.6506	2.40E-03			