2009-09-11

A Multiparameter Network Reveals Extensive Divergence Between *C. elegans* bHLH Transcription Factors: A Dissertation

Christian A. Grove
University of Massachusetts Medical School Worcester

A Multiparameter Network Reveals Extensive Divergence between *C. elegans* bHLH Transcription Factors

A Dissertation Presented

By

CHRISTIAN A. GROVE

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 11, 2009

INTERDISCIPLINARY GRADUATE PROGRAM

A MULTIPARAMETER NETWORK REVEALS EXTENSIVE DIVERGENCE BETWEEN *C. ELEGANS* BHLH TRANSCRIPTION FACTORS

A Dissertation Presented By

Christian A. Grove

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

A. J. Marian Walhout. Ph.D.. Thesis Advisor

Michael Green, Ph.D., Member of Committee

Michelle Kelliher, Ph.D., Member of Committee

Mark Alkema, Ph.D., Member of Committee

T. Keith Blackwell, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

Scot Wolfe, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the School.

Anthony Carruthers, Ph.D.
Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

September 11, 2009

COPYRIGHT INFORMATION

ACKNOWLEDGEMENTS

I would first like to thank my parents, Margareta and Jeffrey Grove, for their unwavering support of my choices: career, personal, and otherwise. They have always been 110% behind what I've wanted to do, and invested an immense amount of their own personal energy to be sure that I've had what I needed. For as long as I've been a student, my parents have appreciated my studies as my top priority and have given me the space and opportunity to embrace and nourish my efforts.

I would also like to thank my brother, Eric Grove, whose thirst and passion for knowledge inspired me as a young scientist. There have been many late nights where my brother and I have discussed a whole range of topics, none of which were less than profound.

I thank my mentor, Marian Walhout, for providing me the opportunity to dive into a field of research that is as cutting edge as her thoughts on the topic. It was my fortune that her arrival at UMass Medical School coincided with my growing curiosity about systems and network biology as well as the field of transcription factors and gene regulatory networks. Marian represents the epitome of hard work, perseverance, and pushing of limits, scientific and personal. My graduate studies would not have been nearly as successful without her guidance and support.

I thank Heidi Tissenbaum, who single-handedly introduced me to the nematode *Caenorhabditis elegans* as a model organism, when she hired me as a

technician right out of college (eight years ago). Heidi is a very dedicated scientist and an inspiration to those she works with.

I thank Jenna Balestrini, whose companionship throughout the six years of my graduate studies has been both wonderful and insightful. Jenna has always provided me with support through difficult times, and has been a peer scientist with whom I could bounce ideas off of and learn things about biology that I had never even thought of.

I thank Federico (Fred) De Masi, whose hard work, insight, and endless sense of humor provided a fruitful collaboration of which I am very grateful. I also thank Martha Bulyk for inviting me to her laboratory to learn the techniques of her lab and perform experiments relevant to my thesis. Martha's knowledge and expertise has been crucial to the success of our work.

I thank members of the Walhout and Dekker labs, both past and present, for their help and friendship over the years of graduate school. I would particularly like to thank Bart Deplancke, Natalia Martinez, Efsun Arda, Vanessa Vermeirssen, John Reece-Hoyes, and Inmaculada Barrasa for their years of camaraderie and scientific discussion in the Walhout lab. I would like to express my gratitude to the faculty and staff of the Program in Gene Function and Expression, particularly Nina Bhabhalia, Sharon Briggs, Judy Mondor, Sara Evans, Elizabeth Nourse, and Shanna Spencer for their persistent and friendly administrative and technical support.

I thank the members of my thesis committee, Scot Wolfe, Michael Green, Michelle Kelliher, and Mark Alkema for their years of guidance, support, and encouragement. I also thank Keith Blackwell for agreeing to take time from his busy schedule and travel to Worcester to be an external examiner for my thesis defense.

And last, but not least, I'd like to thank my community of friends in and around Worcester and beyond, whose friendship, music, knowledge, and interests have enriched my life outside of school.

ABSTRACT

It has become increasingly clear that transcription factors (TFs) play crucial roles in the development and day-to-day homeostasis that all biological systems experience. TFs target particular genes in a genome, at the appropriate place and time, to regulate their expression so as to elicit the most appropriate biological response from a cell or multicellular organism. TFs can often be grouped into families based on the presence of similar DNA binding domains, and these families are believed to have expanded and diverged throughout evolution by several rounds of gene duplication and mutation. The extent to which TFs within a family have functionally diverged, however, has remained unclear. We propose that systematic analysis of multiple aspects, or parameters, of TF functionality for entire families of TFs could provide clues as to how divergent paralogous TFs really are.

We present here a multiparameter integrated network of the activity of the basic helix-loop-helix (bHLH) TFs from the nematode *Caenorhabditis elegans*. Our data, and the resulting network, indicate that several parameters of bHLH function contribute to their divergence and that many bHLH TFs and their associated parameters exhibit a wide range of connectivity in the network, some being uniquely associated to one another, whereas others are highly connected to multiple parameter associations.

We find that 34 bHLH proteins dimerize to form 30 bHLH dimers, which are expressed in a wide range of tissues and cell types, particularly during the

development of the nematode. These dimers bind to E-Box DNA sequences and E-Box-like sequences with specificity for nucleotides central to and flanking those E-Boxes and related sequences.

Our integrated network is the first such network for a multicellular organism, describing the dimerization specificity, spatiotemporal expression patterns, and DNA binding specificities of an entire family of TFs. The network elucidates the state of bHLH TF divergence in *C. elegans* with respect to multiple functional parameters and suggests that each bHLH TF, despite many molecular similarities, is distinct from its family members. This functional distinction may indeed explain how TFs from a single family can acquire different biological functions despite descending from common genetic ancestry.

TABLE OF CONTENTS

TITLE PAGE i

SIGNATURE PAGE ii

COPYRIGHT PAGE iii

ACKNOWLEDGEMENTS iv

ABSTRACT vii

TABLE OF CONTENTS ix

LIST OF FIGURES xiv

LIST OF TABLES xvii

LIST OF ABBREVIATIONS xviii

PREFACE TO CHAPTER I 1

CHAPTER I: Transcription Regulatory Networks, Transcription Factor Parameters, and the Divergence of Transcription Factor Families 2

Introduction 3

Transcription Factors and Transcription Regulatory Networks 4

Transcription Factor Predictions 4

DNA Binding Domains 5

Emerging Properties of Transcription Regulatory Networks 5

Parameters of TF Function 8

Alternative Splicing 8

Dimerization 10

DNA-Binding Specificity 11

Spatiotemporal Expression 12

Post-Translational Modifications 13

Ligands 14

Co-Factors 15

Transcription Factor Variants and Disease 16

Integrating Multiple Parameters to Visualize and Understand TRNs 17

Divergence of TF Families 17
The *C. elegans* bHLH TFs 20
Classification of C. elegans *bHLH Proteins* 20
An Experimental Approach to Integrate Multiple Functional Parameters of the C.elegans *bHLH TFs* 21
Synopsis 24
PREFACE TO CHAPTER II 33
CHAPTER II: The *C. elegans* bHLH Dimerization Network 34
Abstract 35
Introduction 36
Results 38
Updates to Various C. elegans *bHLH Gene Models* 38
Auto-Activation of C. elegans *bHLH Proteins in Yeast* 39
The C. elegans *bHLH Dimerization Network* 40
Dimerization Interactions are Class Specific 40
Two Dimerization Modules: The HLH-2 and AHA-1 Modules 41
HLH-3, HLH-4, and HLH-10 vs. the TF-Array 42
Discussion 44
Yeast Two-Hybrid: Data Quality 44
Dimerization Hub Proteins May Confer Transcriptional Activation Activity 45
bHLH Structural Components that Potentially Contribute to Dimerization Specificity 46
bHLH Proteins with no Yeast Two-Hybrid Data 49
Known Human bHLH Network 49
HLH-33: A Novel bHLH-PAS Protein? 51
Summary 52
Materials and Methods 54
Acknowledgements 56

PREFACE TO CHAPTER III 73

CHAPTER III: The Spatiotemporal Expression Patterns of the *C. elegans* bHLH Transcription Factors 74

Abstract 75

Introduction 76

Results 78

Worm PCR Verification 78

Temporal Expression Summary 78

Spatial Expression Summary 79

Class III Expression Patterns 80

Class IV Expression Patterns 80

REF-1 Family Expression Patterns 81

Class VII (bHLH-PAS) Expression Patterns 82

Co-Expression of Heterodimerizing bHLH Proteins 83

Design of Co-expression Experiments 83

Drawbacks of Transcriptional mCherry Fusions 84

Drawbacks of Analyzing Co-expression under Hypoxic Conditions 85

Drawbacks of Analyzing Co-expression where GFP Expression is Intense 86

ClassI/II (HLH-2 Module) Co-Expression 87

Expression Annotation at Cellular Resolution 88

Neuronal Expression of ngn-1 and hlh-13 88

Distal Tip Cells: hnd-1, hlh-12, and lin-32 89

Discussion 91

Spatiotemporal Expression Analysis: Data Quality 91

Non-overlapping Expression Patterns 91

GFP Negative Transgenic Lines 92

Partial Intestine, Coelomocytes, & Head Muscle GFP Expression:

Possible Artifacts? 93
Summary 94
Materials and methods 96
Acknowledgements 100
PREFACE TO CHAPTER IV 120
CHAPTER IV: The DNA Binding Specificities of the *C. elegans* bHLH Transcription Factors 121
Abstract 122
Introduction 123
Results 125
The Experimental Approach – Protein Binding Microarrays 125
DNA Binding Specificity Analysis of Homo- and Heterodimeric bHLH TFs 125
PBM Results Support Y2H Findings 127
Two Clusters of DNA-Binding Specificity 128
The REF-1 Family: the Missing C. elegans *Class VI bHLH Proteins?* 129
C. elegans *bHLH Dimers Bind E-box and E-box-related Sequences, but Avoid CAA-Containing E-boxes* 131
DNA Binding Preferences of Individual bHLH Dimers 132
Half-Site Preferences of Individual bHLH Proteins 132
E-Box Flanking Nucleotides Contribute to bHLH DNA Binding Specificity 135
E-Box-dependent Flanking Nucleotide Preferences 136
Dimer-dependent Flanking Nucleotide Preferences 137
Negative Control Experiments 138
Identifying bHLH Dimer Candidate Target Genes 139
Functional Annotation of Candidate bHLH Target Genes 141
Discussion 143
Protein Binding Microarrays: Data Quality 143

Two Clusters of bHLH DNA Binding Profiles 143
E-Boxes and E-Box-like Sequences 144
Flanking Nucleotide Contribution to bHLH Dimer Binding 145
Candidate bHLH Target Genes and Enriched GO Terms 146
Materials and methods 147
Acknowledgements 152
PREFACE TO CHAPTER V 211
CHAPTER V: A Multiparameter Integrated Network of *C. elegans* bHLH Transcription Factors 212
Abstract 213
Introduction 214
Results 217
An Integrated bHLH Dimerization, DNA Binding and Expression Network 217
HLH-30: in vivo *Validation of the Network* 218
Multiparameter Analysis of bHLH TFs - Similarity Scores 220
HLH-4, HLH-10, HLH-15, and HLH-19 – A Closer Look 220
Target Gene Overlap of CACGTG-Binding bHLH Dimers 222
Parameter Similarity Score Heatmaps 223
Discussion 224
Materials and Methods 228
Acknowledgements 231
REFERENCES 262

LIST OF FIGURES

Figure I-1. Emerging features of transcription regulatory networks (TRNs) 25
Figure I-2. Examples of parameters of TF functionality 26
Figure I-3. The integration of multiple TF parameters into novel TRNs 28
Figure I-4. Functional and molecular divergence in paralogous TF families 29
Figure I-5. Cladogram of aligned *C. elegans* bHLH domains 30
Figure I-6. Schematic overview of experiments 32
Figure II-1. Auto-activation of DB-bHLH Y2H baits 60
Figure II-2A. Example of Y2H matrix assay using DB-HLH-15 as bait 61
Figure II-2B. Raw yeast two-hybrid data for the *C. elegans* bHLH TFs 62
Figure II-3. The bHLH dimerization network 63
Figure II-4. Individual bHLH proteins tested against AD-TF-Array 65
Figure II-5. Alignment of bHLH domains of HLH-2, orthologs, and partners 69
Figure II-6. Structural insight from the E47/NeuroD crystal structure 70
Figure II-7. The human bHLH dimerization network 71
Figure III-1. Diagnostic PCR on individual worms carrying *Phlh*::GFP transgenes 106
Figure III-2. Tissue overlap coefficient analysis 112
Figure III-3A. Embryonic co-expression of HLH-2 and its partners 113
Figure III-3B. Cartoon summary of embryonic co-expression of HLH-2 and its partners 114
Figure III-4. Post-embryonic co-expression of HLH-2 and its partners 115
Figure III-5. Cellular resolution expression annotations 118
Figure IV-1. Protein binding microarray (PBM) raw data 154
Figure IV-2. Binding site logos for 19 *C. elegans* bHLH dimers 155
Figure IV-3. Clustergram of all bHLH dimers that yielded DNA binding profiles at a PBM ES ≥ 0.40 156
Figure IV-4. Individual bHLH monomers that heterodimerize do not

bind DNA independently 157
Figure IV-5. Box plots of all bHLH dimer-E-Box and dimer-E-Box-like Combinations 158
Figure IV-6A. Statistics of E-box and E-box-like sequence binding 160
Figure IV-6B. The *C. elegans* bHLH dimer/binding site network 161
Figure IV-7. Half-site logic scheme 162
Figure IV-8. Deducing half-site preference for HLH-3 and HLH-8 164
Figure IV-9. Half-site preference prediction for HLH-2 and partners 166
Figure IV-10. Variation in ES distribution suggests flanking nucleotide affects 167
Figure IV-11A. Flanking nucleotide preference for Cluster I dimers 168
Figure IV-11B. Flanking nucleotide preference for Cluster II dimers 170
Figure IV-12. Influence of flanking nucleotides on bHLH dimer binding 171
Figure IV-13. Negative control PBM experiments with bHLH-bHLH pairs that do not dimerize in Y2H assays 173
Figure IV-14. Pipeline for predicting bHLH dimer target genes 174
Figure V-1. The integrated *C. elegans* bHLH network 232
Figure V-2. GO Subnetworks 234
Figure V-3. Expression pattern of *hlh-30* 237
Figure V-4. Summary of HLH-30 DNA binding specificity 238
Figure V-5. HLH-30 acts predominately as an activator of transcription 240
Figure V-6. Enrichment for HLH-30 binding sites upstream of HLH-30 target genes 244
Figure V-7. Enrichment for HLH-30 binding sites downstream of HLH-30 target genes 246
Figure V-8. HLH-30 targets tend to have multiple HLH-30 binding sites 248
Figure V-9. A measure of divergence: the Similarity Score 250
Figure V-10A. Heatmaps of similarity scores for comparisons of dimerization specificities between bHLH TFs 251

Figure V-10B. Heatmaps of similarity scores for comparisons of expression patterns between individual bHLH gene promoters 252
Figure V-10C. Heatmaps of similarity scores for comparisons between bHLH dimer expression patterns 253
Figure V-10D. Heatmaps of similarity scores for comparisons between bHLH dimer 10-mer binding sites 254
Figure V-10E. Heatmaps of similarity scores for comparisons between bHLH dimer candidate target genes 255
Figure V-10F. Heatmaps of similarity scores for comparisons between bHLH “network paths” 256
Figure V-11. Sub-network for HLH-2/HLH-4, HLH-2/HLH-10, HLH-2/HLH-15, and HLH-2/HLH-19 257
Figure V-12. Similarity score profiles for HLH-2/HLH-4, HLH-2/HLH-10, HLH-2/HLH-15, HLH-2/HLH-19 258
Figure V-13. Cellular resolution expression annotation for *Phlh-4*, *Phlh-10*, and *Phlh-15* 259
Figure V-14. Candidate target gene overlap for CACGTG binders 261

LIST OF TABLES

Table II-1. *C. elegans* bHLH ORF cloning information 57
Table II-2. Interologs 64
Table III-1. *C. elegans* bHLH promoter cloning information 101
Table III-2. Transgenic *C. elegans* lines 104
Table III-3. Promoter-specific diagnostic primers 105
Table III-4. The bHLH expression matrix 108
Table III-5. bHLH heterodimer co-expression matrix 117
Table IV-1. ORF sequencing primers 153
Table IV-2. 8-mer enrichment score data (electronic file only)
Table IV-3. List of predicted half sites specified by each individual bHLH protein 165
Table IV-4. bHLH dimer candidate target gene lists (electronic file only)
Table IV-5. GO categories enriched among bHLH dimer candidate target genes 175
Table V-1. Genes with significant difference in expression in *hlh-30* mutant 241
Table V-2. GO categories enriched among HLH-30 target genes 249

LIST OF ABBREVIATIONS

AD = Gal4p Activation Domain

βGal = Beta-Galactosidase

bHLH = Basic Region Helix-Loop-Helix

bZIP = Basic Region Leucine Zipper

cDNA = complementary DNA

C. elegans = *Caenorhabditis elegans*

DB = Gal4p DNA Binding Domain

DIC = Differential Interference Contrast Microscopy

DNA = Deoxyribonucleic Acid

DTC = Distal Tip Cell

ER = Estrogen Receptor

ES = Enrichment Score

GFP = Green Fluorescent Protein

GO = Gene Ontology

GST = Glutathione-S-Transferase

NHR = Nuclear Hormone Receptor

ORF = Open Reading Frame

PBM = Protein Binding Microarrays

PCR = Polymerase Chain Reaction

Phlh = Promoter of an *hlh* gene

RNA = Ribonucleic Acid

RNAi = RNA Interference

SS = Similarity Score

TF = Transcription Factor

TRN = Transcription Regulatory Network

Y2H = Yeast Two-Hybrid Assay

PREFACE TO CHAPTER I

This chapter introduces the concepts of transcription factors (TFs), transcription regulatory networks (TRNs), and the impact of a variety of TF parameters on the functionality of individual TFs as well as TRNs in which TFs play the major role. Also introduced is the concept of DNA binding domains, TF families, and TF family expansion, and questions are raised about the divergence of TF parameters among the *C. elegans* bHLH family of TFs.

Much of this chapter has been published separately in:

Grove CA, Walhout AJM. Transcription factor functionality and transcription regulatory networks. Mol Biosyst. 2008 Apr ;4(4):309-314.

and

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJM. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell. 2009 Jul 23;138(2):314-327.

CHAPTER I

Transcription Regulatory Networks, Transcription Factor Parameters, and the Divergence of Transcription Factor Families

Introduction

Now that numerous high-quality complete genome sequences are available, many efforts are focusing on the “second genomic code”, namely the code that determines how the precise temporal and spatial expression of each gene in the genome is achieved. In this regard, the elucidation of transcription regulatory networks that describe combined transcriptional circuits for an organism of interest has become valuable to our understanding of gene expression at a systems level. Such networks describe physical and regulatory interactions between transcription factors (TFs) and the target genes they regulate under different developmental, physiological, or pathological conditions. The mapping of high-quality transcription regulatory networks depends not only on the accuracy of the experimental or computational method chosen, but also relies on the quality of TF predictions and experimental identification/verification. Moreover, the total repertoire of TFs is not only determined by the protein-coding capacity of the genome, but also by different protein properties, including dimerization, co-factor interactions and post-translational modifications. Here, we discuss the factors that influence TF functionality and, hence, the functionality of the networks in which they operate.

Transcription Factors and Transcription Regulatory Networks

Transcription factor predictions

There are two classes of TFs: basal TFs that are involved in transcription of most, if not all, genes, and regulatory TFs that control only subsets of genes (1). For the understanding of differential gene expression at a systems level, we only consider regulatory TFs (hereafter referred to as TFs). TFs interact with their target genes by binding specific *cis*-regulatory gene elements through a sequence-specific DNA binding domain. Different DNA binding domains are used to group TFs into TF families. Examples of DNA binding domains include basic region helix–loop–helix domains (bHLH), homeodomains and various types of zinc fingers. Computational tools have been developed both to define consensus DNA binding domains and to predict additional TFs of that family encoded by a genome of interest (2; 3). We found that, although such computational tools are powerful, they do incorporate false predictions and miss many known TFs (4). For instance, by using a combination of computational tools and extensive manual curation we predicted a high-quality compendium of 934 TFs in the nematode *Caenorhabditis elegans*, which extended purely computational predictions by ~50% (4). However, even though this compendium is more complete than previous collections, it is not yet comprehensive as algorithms and experimental assays continue to improve and, therefore, additional TFs continue to be discovered (5).

DNA binding domains

Over the past decades, many different sequence-specific DNA binding domains have been uncovered. However, we propose that it is unlikely that all DNA binding domains are known. This is because, by applying yeast one-hybrid assays to only 112 *C. elegans* gene promoters, we have already discovered 11 *C. elegans* proteins that robustly interact with their target promoters in yeast, but that do not possess a known DNA binding domain. By using chromatin immunoprecipitation assays in yeast, we confirmed that these interactions are direct for nine of these proteins (6; 7). We do not know yet whether these proteins directly bind to DNA or if they are recruited to their target promoters by interacting with other DNA binding proteins. Future structure–function analysis will provide insight into the mechanism of action of these novel putative TFs. Importantly, the cataloging of the DNA binding domain(s) of these proteins may enable the identification of additional proteins with similar domains in *C. elegans*, and perhaps in other organisms as well.

Emerging Properties of Transcription Regulatory Networks

Transcription regulatory networks (TRNs) can be represented as graph models that combine physical and regulatory interactions between TFs and their target genes (reviewed in (8)). Several methods that can be used to identify physical interactions between TFs and their targets have been developed and applied to the study of both yeast and metazoan transcription regulatory networks. These include TF-centered methods such as chromatin

immunoprecipitations (9-11), protein binding microarrays (12), DamID (13; 14) and bacterial one-hybrid assays (15), as well as gene centered methods such as high-throughput yeast one-hybrid assays (16). The TF-DNA interaction data obtained by these methods are often visualized into network models. Figure I-1 depicts a hypothetical network that contains several of the architectural and topological network features described to date, including TF hubs (TFs that bind a large number of target genes), target gene hubs (genes bound by a large number of TFs), and TF modules (sets of TFs that share numerous target genes). Metazoan TF hubs were uniquely revealed in the nematode *Caenorhabditis elegans* by using gene-centered yeast one-hybrid assays (6; 7). Specifically, we found that, whereas the majority of TFs bind only few promoters, a small subset of TFs bind up to 40% of all promoters tested. This suggests that these proteins play a more essential role in regulating transcription than most other TFs. By performing chromatin immunoprecipitation with eight TFs in the budding yeast *Saccharomyces cerevisiae*, Snyder and colleagues (17) showed that the genes that code for two of these TFs, *MGA1* and *PHD1*, were themselves targeted by all eight TFs, suggesting that these two TFs may be target gene hubs. Interestingly, *MGA1* and *PHD1* are master regulators of pseudohyphal growth. The authors suggest that genes regulated by large number of TFs may function as master regulators of biological processes. In agreement with this, we found that the promoter of the master regulator of D-type

GABA-ergic motor neurons *unc-30* in *C. elegans* can interact with 36 different TFs (7).

TF modules are another feature of transcription regulatory networks that have recently emerged (7). TF modules are distinct from more extensively studied general network modules that are defined as highly interconnected groups of nodes without regard for directionality or node type (*i.e.*, TFs *vs.* target genes) (18; 19). We have defined TF modules as sets of TFs that share many of their target genes (Figure I-1) (7). As such, these are uniquely found in bipartite transcription regulatory networks. Both general network modules and TF modules may reflect functionality, for instance within a cell or tissue type, or regarding a particular biological process. Finally, transcription regulatory networks are composed of recurring circuits, referred to as network motifs. For instance, feed forward loops are found frequently in networks from both prokaryotic and eukaryotic organisms. Such motifs represent widely used regulatory mechanisms that can, for instance, be used to stabilize gene expression (20; 21).

The transcription regulatory networks that have been described to date represent compilations of multiple events that take place during the lifetime of an organism collapsed into a single model. However, in reality, only a subset of the network is active in particular cell types, under different developmental or physiological conditions, or at any given time (22-24). In addition, each TF in regulatory networks is represented as a single node, whereas it is known that

TFs exist in many different functional forms that are determined by a variety of factors including post-translational modifications and dimerization. These factors themselves may depend on specific developmental or physiological conditions. Each TF form may interact with distinct target genes, or with the same target gene, but under different conditions. Here, we discuss the factors that need to be taken into account to determine how many functional TFs occur in an organism of interest, and how this information can be incorporated into transcription regulatory network models to study differential gene expression at a systems level.

Parameters of TF Function

We refer to the various aspects of TF functionality as parameters of TF function. As additional experimental evidence is acquired for individual TFs, we can include these parameters as important aspects of TF functionality and incorporate them into more holistic models of TF function within the grand scheme of transcription regulatory networks. The following list of TF parameters describes parameters that are known to be important for a variety of TFs, yet we do not yet know the extent to which these parameters play important roles in the function of TFs in general.

Alternative splicing

In metazoans, many gene transcripts, including those encoding TFs, are alternatively spliced, which often leads to multiple variants of a protein. Interestingly, it has been found that TF-encoding genes in mice undergo

alternative splicing more frequently than other genes (25). Alternative splicing may lead to TFs with different functions. For instance, DNA binding domains or transcription regulatory domains may be included or excluded from the TF variant. At least 144 *C. elegans* TFs undergo alternative splicing, resulting in 379 different proteins, 30 of which lack a DNA-binding domain (4). The latter may function as regulators of TF function, for instance by titrating interaction partners of the corresponding TFs that do possess a DNA binding domain. Several *C. elegans* TFs contain more than one DNA binding domain and alternative splicing can affect which domains are present in the different protein products. For instance, several DAF-16 variants are generated as a result of alternative splicing, and each variant carries a unique combination of domains (26) (Figure I-2A). DAF-16 is a critical regulator of various physiological processes including fat storage, aging, and the formation of dauers (an alternative larval stage of development, resistant to many forms of stress). It contains two potential forkhead DNA binding domains and is known to bind or regulate numerous target genes (27). It is tempting to speculate that each forkhead domain is responsible for the interaction with a distinct set of target genes, each of which may be involved in a particular biological process. A human example involves the homeodomain proteins hepatocyte nuclear factor HNF1 and vHNF1. There are three HNF1 splice variants, each of which encodes a protein with varying transcription activation properties. vHNF1 is also differentially spliced and one of the resulting protein products, vHNF1-C, functions as a *trans*-dominant repressor

of all three HNF1 variants (28). TF variants may be expressed in different cell types or under particular conditions, leading to variable outputs of the transcriptional circuits in which they function. For instance, the HNF isoforms are differentially expressed in the human digestive tract, liver and kidney, where they may either regulate distinct target genes or, alternatively, the same target genes but at different levels. Genome-scale analyses of alternative splicing, for instance using whole genome or exon junction tiling arrays (29), will greatly facilitate the accurate identification of all TF variants that are produced in each cell and tissue type, and in various model organisms.

Dimerization

Several TFs bind DNA as obligatory dimers, including members from the basic region leucine zipper (bZIP), bHLH and nuclear hormone receptor (NHR) families (30-32). Dimerization should be taken into account when considering the total complement of functional TFs because, if a particular TF only functions when it dimerizes with another TF, the dimer should be considered a single functional unit. Dimerization can affect the total number of functional TFs in different ways (Figure I-2B). In one model, each TF from a family can dimerize with itself and any other member of that family. If this would be the case, the number of functional TFs would be dramatically greater than the number of individually predicted TFs. For instance, dimerization between all 274 *C. elegans* NHRs (4) would result in a total of 37,675 NHR dimers. In another model, each TF dimerizes exclusively with one other TF of the same family. If this were the

case for the *C. elegans* NHRs, this would result in 137 functional TF complexes, which would reduce the total number of functional TFs. In a third, intermediate model, one or more TFs from a family could serve as central dimerization partners that can interact with multiple members of the relevant TF family. Although no comprehensive data regarding TF dimerization is as yet available, our data indicate that this third model is likely most relevant (33). Systematic protein–protein interaction mapping efforts will be required to identify all functional TF dimers. High-throughput yeast two-hybrid assays (34) are particularly well suited for this task as they identify binary interactions. For instance, we found by high-throughput yeast two-hybrid assays that NHR-49 serves as a dimerization hub that can interact with at least 15 other NHRs (5). In the future, it will be important not only to identify all TF dimers, but to also determine the DNA-binding specificities of each dimer and where and when dimerization partners are co-expressed.

DNA-Binding Specificity

Arguably one of the most important features of a TF's function is its ability to bind DNA and to do so in a sequence-specific manner. This functionality enables a TF to specifically target specific subsets of genes for regulation in response to developmental or environmental cues. Despite its importance, complete DNA-binding specificity profiles have only been described for a minority of TFs. This has been in large part due to the lack of appropriate techniques for the elucidation of comprehensive DNA binding profiles. Some methods that have

been used to determine the complete DNA binding specificity of TFs are SELEX (systematic evolution of ligands by exponential enrichment) (35; 36), SAAB (selected and amplified binding-sequence) (37; 38), bacterial one-hybrid assays (15; 39), and protein binding microarrays (PBM) (12; 40). These techniques are powerful, yet their application to entire families of TFs has only been recently implemented (41-43) and is likely to provide great insight into the nature of TF-DNA binding. Future experimentation on the comprehensive determination of DNA binding specificities will allow a more thorough understanding of how TFs specifically recognize their cognate target genes, despite large genomic sequence-search spaces and similarities with other TF family members.

Spatiotemporal Expression

One important question about the function of TFs that remains to be determined is the extent to which any given TF determines the identity of a particular cell type and, conversely, how much the intracellular molecular environment of a given cell type affects the functionality of TFs expressed within that cell. For instance, in *Drosophila melanogaster*, the TF *Antennapedia* normally serves as an important determinant for the thorax in developing *Drosophila* larvae (44). Experiments have shown that ectopic expression of *Antennapedia* at the site of developing antennae will cause legs to form in the place of antennae, suggesting that this TF can override the developmental program normally intended to produce antennae (45). When ectopically expressed elsewhere in the developing larva, however, *Antennapedia* appears to

have no effect, suggesting that the tissue or cell type in which this TF is expressed must be somehow receptive to its activity in order to undergo the transition in developmental program (45). In addition, there is a specific window of time during development when *Antennapedia* expression must be induced in order to see the antenna-to-leg phenotype (45). Therefore, the time at which this TF is expressed must also play a crucial role in its function. Taken together, it is evident that the spatiotemporal expression patterns of TFs are important parameters of TF functionality.

Post-Translational Modifications

Many proteins, including TFs, are post-translationally modified under different conditions and by different modifiers. Several post-translational modifications of TFs have been reported, including phosphorylation, hydroxylation, acetylation, ubiquitination and sumoylation (Figure I-2C) (46). Such modifications often result from the activation of signal transduction pathways in response to environmental stimuli or developmental cues. Post-translational modifications can affect the regulatory activity of a TF, as well as its localization or stability. For instance, estrogen receptor β (ERβ), normally activated by ligand, can be phosphorylated by MAP kinase (MAPK) which leads to recruitment of the steroid receptor coactivator-1 (SRC-1) and, subsequently, ligand-independent activation of target genes (47). Erα can be acetylated by p300 at conserved lysine residues resulting in enhanced DNA-binding activity and, perhaps, ligand-dependent transcriptional activation (48). For most TFs it is

not clear which modified forms exist and how these forms function to regulate gene expression. For a thorough understanding of gene regulatory networks, it will be important to determine which modification each TF is subjected to, under which circumstances, and how these modifications affect TF functionality.

Ligands

Many TFs become activated or inactivated as a result of ligand binding (Figure I-2D). One of the most prominent classes of ligand-dependent TFs is the NHR family, which includes ER, androgen receptor (AR), peroxisome proliferator-activated receptors (PPARs), retinoic acid receptor (RAR) and others. NHR ligands are hydrophobic molecules that can freely diffuse into the nucleus where they specifically interact with their target receptors. Most NHRs become potent transcriptional activators upon ligand binding. Human ER has been studied extensively because of its association with the development of breast cancer (49). A number of ER ligands (endogenous and exogenous) have been identified, some of which are non-steroidal compounds that are referred to as selective ER modulators (SERMs). Whereas steroidal compounds such as estrogen function to naturally modulate ER activity, SERMs such as tamoxifen are used in cancer treatment. Different ligands bind to ER with varying affinities and have different effects. Depending on tissue and cell-type context, ligands induce conformational changes in ER that promote transcriptional activation, whereas others promote transcriptional repression (50). The ligand(s) for most NHRs remain to be identified, and, therefore, such NHRs are referred to as “orphan” receptors. For

instance, the nematode *C. elegans* has 274 predicted NHRs (4), but a ligand (dafachronic acid) has only been identified for a single NHR, DAF-12 (51).

Another class of ligand-binding TFs is the bHLH-PAS sub-family that includes the aryl hydrocarbon receptor (AHR). AHR can interact with a variety of exogenous compounds or toxins such as dioxin, and mediate a biological response (for a review see ref. (52)). The range of compounds that can activate AHR is still under investigation, and although most appear to be exogenous in origin, it has been proposed that endogenous AHR ligands may play a role in organism development or homeostasis (53). Indeed, the *C. elegans* ortholog of AHR, *ahr-1*, is required for the proper development and specification of touch-receptor neurons, interneurons, and motor neurons (54; 55).

Co-Factors

Regulatory TFs often activate or repress transcription, either by recruiting the RNA polymerase II machinery, or by preventing its access to the transcription start site. While many TFs interact directly with general TFs or components of RNA polymerase II, others function by interacting with intermediate proteins called co-factors (Figure I-2E) (56-58). Depending on environmental or developmental circumstances, the same TF can interact with different coactivators or corepressors, illustrating the versatility of TFs in carrying out opposing regulatory effects in different contexts. RAR, for instance, can recruit the NcoR/SMRT corepressor complexes thereby repressing transcriptional activity of its target genes. Upon binding its ligand retinoic acid, however, RAR

changes conformation and adopts a form capable of recruiting coactivator complexes and subsequently activates transcription (57). Interestingly, cofactor interactions can also affect DNA binding specificity (59), implying that different TF-cofactor pairs may interact with different sets of target genes. Future large-scale genomic and proteomic experiments are needed to identify the full spectrum of ligands and co-factors each TF in an organism can interact with and to unravel how these interactions affect the biochemical and biological function of each TF.

Transcription Factor Variants and Disease

TFs play a crucial role in numerous diseases, including congenital disorders and cancer. Mutations in TF-encoding genes can result in loss-of-function, gain-of-function or neomorph TFs that attain a function not shared by the original TF. One of the best-studied TFs mutated in cancer is p53, a tumor suppressor gene that is inactivated by mutation in most human cancers. Interestingly, it appears that some mutations can also convert p53 into an oncogene (60). P53 regulates the expression of various cell cycle inhibitors and proteins involved in apoptosis. It will be interesting to see how the different forms of mutant p53 are affected in their biochemical and biological functions.

Several mutated TFs have been found to result in a variety of human congenital disorders. For instance, altered dimerization between the bHLH TFs Twist1 and Hand2 was found in patients with Saethre–Chotzen syndrome (61). Common neomorph TF variants that are found in instances of leukemia are

fusion proteins resulting from chromosomal translocation/inversion (Figure I-2F). For example, an inversion on murine chromosome 16 leads to an aberrant Cbfb-MYH11 fusion protein, resulting in the development of acute myeloid leukemia (62). It will be important to understand the variety of mutant forms of TFs that exist in different diseases and how they perturb the regulatory networks that contribute to a disease state.

Integrating Multiple Parameters to Visualize and Understand TRNs

Although complete genome sequences have provided a great first step toward the comprehensive identification of the compendium of TFs that function in an organism of interest, we are far from having a complete picture of all the protein variants that may exist for each predicted TF. As we've discussed here, there are numerous factors that affect the functional states of TFs throughout development, homeostasis, and in disease. Since the gene count is strikingly similar between organisms of widely different complexity, a larger number of TF permutations may contribute to more intricate regulatory networks in higher eukaryotes such as humans. In the future, different TF forms need to be incorporated as individual nodes in transcription regulatory networks to facilitate network modeling and hypothesis derivation (Figure I-3).

Divergence of TF Families

Paralogous TFs are grouped into families based on the type of DNA binding domain they possess. Such families grow by gene duplications upon which identical and therefore fully redundant TFs emerge. After acquiring

mutations, duplicate TFs diverge and may become partially redundant. Upon further mutation completely non-redundant, yet paralogous TFs may emerge (Figure I-4A).

TF families often expand with organismal complexity. For instance, whereas the nematode *Caenorhabditis elegans* has 42 basic helix-loop-helix (bHLH) proteins (4), the human genome encodes more than 100 (63). The expansion and divergence of TFs has been proposed to lead to increased regulatory complexity, biological specificity and organismal complexity.

Paralogous TFs often have different biological functions. For example, loss of *C. elegans* bHLH TFs results in phenotypes ranging from neuronal defects to embryonic lethality (see *e.g.* (64-66)). In humans, mutations in paralogous TFs can result in different diseases. Mutations in the human bHLH TFs *TWIST* and *HAND1* can result in Saethre-Chotzen syndrome and heart hypoplasia, respectively (67; 68).

As described above, differences in TF parameters are thought to be important determinants of regulatory and biological specificity. However, both the extent of TF functional divergence and the relative contribution of individual TF parameters remain undetermined. A main challenge in regulatory and genome biology is to understand the mechanisms of TF divergence and to disentangle the contribution of each of the parameters to this process. Specific questions are to what extent members of a TF family differ in each of these parameters, and if

differences in any one parameter are more prevalent than differences in another (Figure I-4B).

Assessment of metazoan TF divergence requires the comprehensive and standardized measurement of multiple TF parameters and the incorporation of these parameters into a single, integrated network. Initial studies in the yeast revealed a large degree of redundancy for the eight Yap TFs, as well as functional divergence through DNA binding specificities and interactions with chromatin proteins (69; 70). However, the mechanisms of divergence in large metazoan TF families remain unexplored (Figure I-4B). Numerous metazoan TFs have been studied individually, but the resulting data are sparse due to assay incompleteness and heterogeneity. Therefore, such data could not be used to determine the extent and mechanisms of divergence of complete TF families.

In the following chapters, we comprehensively identify dimerization partners, spatiotemporal expression patterns and DNA binding specificities for the *C. elegans* bHLH family of TFs, and model these data into an integrated network. This network displays both specificity and promiscuity, as some bHLH proteins, DNA sequences, and tissues are highly connected, whereas others are not. By comparing all bHLH TFs, we find extensive divergence, and that all three parameters contribute equally to bHLH divergence. Our approach provides a framework for examining divergence for other protein families in *C. elegans* and in other complex multicellular organisms, including humans. Cross-species

comparisons of integrated networks may provide further insights into molecular features underlying protein family evolution.

The *C. elegans* bHLH TFs

Classification of C. elegans *bHLH Proteins*

Various bHLH classifications based on the presence of different domains, amino acid conservation, and known DNA-binding and dimerization specificities have been reported. We grouped the *C. elegans* bHLH proteins according to the classes outlined by Murre and colleagues (71) and supplemented by our own data described below (Figure I-5). Class I contains HLH-2, the *C. elegans* ortholog of human E12/E47 and *Drosophila melanogaster* daughterless; class II contains known and newly identified HLH-2 partners (see below); class III contains HLH-30 and SBP-1 (ortholog of the human Sterol Response Element Binding Protein); class IV contains bHLH-ZIP proteins (Max-like); class VI contains LIN-22 and the REF-1 family (based on DNA binding specificity data we generated; see Chapter IV), and class VII contains the *C. elegans* bHLH-PAS proteins. *C. elegans* does not have any known class V bHLH proteins (e.g. Id, emc) that lack the basic region. These classes provide a framework for our network and facilitate extrapolations of our findings to other organisms.

An Experimental Approach to Integrate Multiple Functional Parameters of the* C. elegans *bHLH TFs

Although there are numerous studies describing the characteristics of individual bHLH TFs in many different organisms, there are no systematically derived datasets that comprehensively determine the contributions of several parameters to an entire family of bHLH TFs in a single multicellular organism. We decided that using several high-throughput methods to systematically analyze all members of the *C. elegans* bHLH TF family could provide a reasonable approach to study the extent to which multiple parameters play a role in bHLH TF functionality.

The first dataset, described in Chapter II, is that of dimerization between all members of the *C. elegans* bHLH family. By performing pair-wise bHLH-bHLH interaction assays using the yeast two-hybrid (Y2H) system, we successfully identified all known dimers (for which clones were available) and identified several novel dimers. The dimerization network recapitulates what one might expect since many of the interactions appear to have interologs (i.e. interactions between orthologous proteins in a different organism). This yeast two-hybrid dimerization network more than doubles the number of dimerization interactions that were previously known for *C. elegans* bHLH proteins.

The second dataset, described in Chapter III, is that of bHLH TF spatiotemporal expression patterns. Generating transgenic worms carrying transcriptional fusion constructs of bHLH gene promoters fused to the gene

encoding green fluorescent protein (GFP), we annotated the transcriptional activity of each bHLH gene promoter as an indicator of bHLH TF expression. Our expression analyses, where applicable, agree closely with previously reported expression analyses using a variety of other methods, including immuno-staining of bHLH proteins and in situ RNA-hybridization. We describe numerous novel expression patterns which suggest when and where the various bHLH dimers identified in the network may be functioning *in vivo*.

The third dataset, presented in Chapter IV, describes the DNA binding specificity for a majority of the bHLH dimers in our dimerization network. Using protein binding microarrays (PBMs) we were able to determine the entire spectrum of DNA binding specificities for 19 *C. elegans* bHLH dimers in an unbiased manner. The data reveal two distinct clusters of DNA binding specificity, suggesting an early evolutionary divergence of the bHLH TFs with regards to DNA binding properties. The data also reveal the binding of several bHLH dimers to a variety of non-canonical bHLH binding sites, the specificity for nucleotides central to the bHLH-binding E-Boxes, and the contribution of E-Box-flanking nucleotides to DNA binding. We then used the PBM-derived binding site data to make predictions of bHLH dimer target genes and functional annotation.

In Chapter V, we integrate all of the TF parameters from the previous three chapters into a multiparameter integrated network for the *C. elegans* bHLH TFs. This network, the first of its kind, displays the bHLH dimers and the aforementioned TF parameters as nodes in a network, connected by edges,

which are indicative of their relevant associations. As with the individual parameters, we see that nodes in the network display a wide array of connectivity, as some nodes are highly connected to other nodes, whereas others are more uniquely associated. The overall extent to which each bHLH TF in *C. elegans* has functionally diverged from one another is assessed by the use of a similarity score, a measure of overlap in parameter associations for a given bHLH-bHLH pair. Ultimately, we find that each bHLH TF is remarkably distinct from all others when all parameters are taken into account.

The systematic nature with which the data above was collected provides us with unique datasets that can now be used to address questions about bHLH TF functionality that were previously unresolved. Potential biases introduced by each method used are described in the following chapters, and, although biases may exist, we know that they are applied consistently within each dataset, allowing us to make adequate comparisons from each bHLH TF to another. For a complete overview of the experimental approach used, see Figure I-6.

Synopsis

The understanding of the underlying principles of gene expression regulation may be revealed by the study of transcription regulatory networks (TRNs). Current models of TRNs are constructed from individual TF/target gene interactions identified by a variety of techniques, and typically ignore aspects of TF functionality such as protein-protein interactions, post-translational modifications, and TF spatiotemporal expression patterns. We refer to such aspects of TF functionality as TF parameters and suggest their inclusion in future models of TRNs to generate a more accurate and holistic understanding of the principles of regulation of gene expression. We have chosen to begin to create a more integrated TRN for the *C. elegans* bHLH family of TFs for a number of technical reasons, including the experimental tractability of the nematode *C. elegans* for the study of gene expression as well as the foundation of knowledge already present for the bHLH TFs in a variety of other organisms. In the following chapters, we will describe the experimental approaches for mapping such an integrated network and the implications of the results of such experiments.

Figure I-1

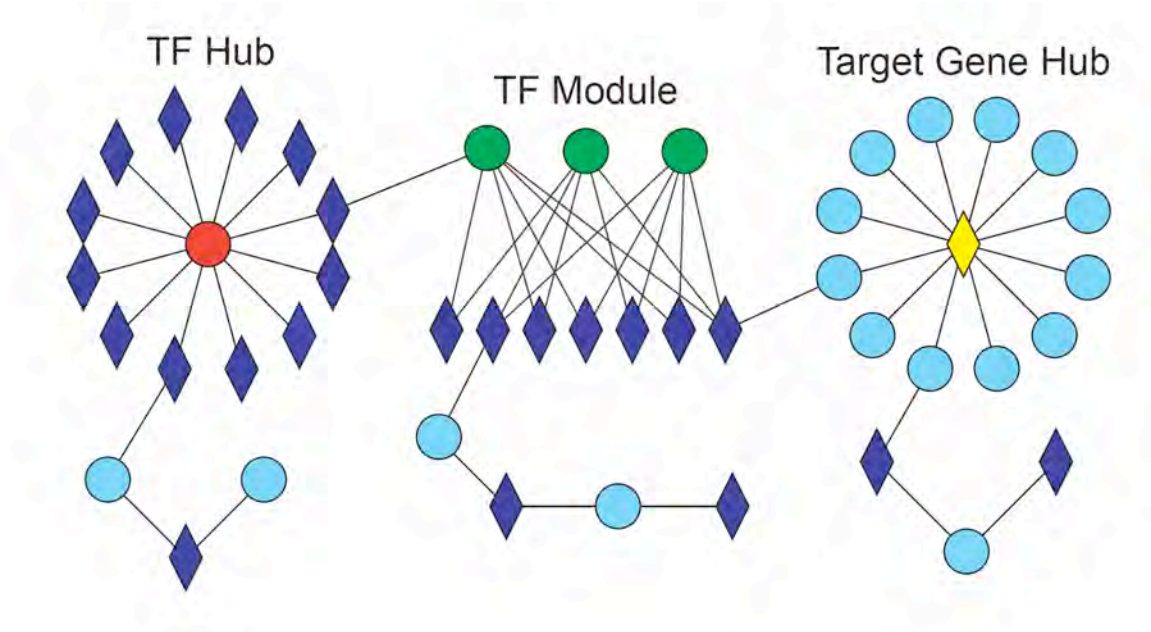


Figure I-1. Emerging features of transcription regulatory networks (TRNs)
The mapping of physical interactions between transcription factors (TFs) and their target genes has resulted in the discovery of several interesting network features, some of which are shown here. Some TFs (circles) target a disproportionately large number of genes (diamonds) and are referred to as TF hubs. Sets of TFs that share many target genes are referred to as TF modules. Target gene hubs interact with a disproportionately large number of TFs.

Figure I-2

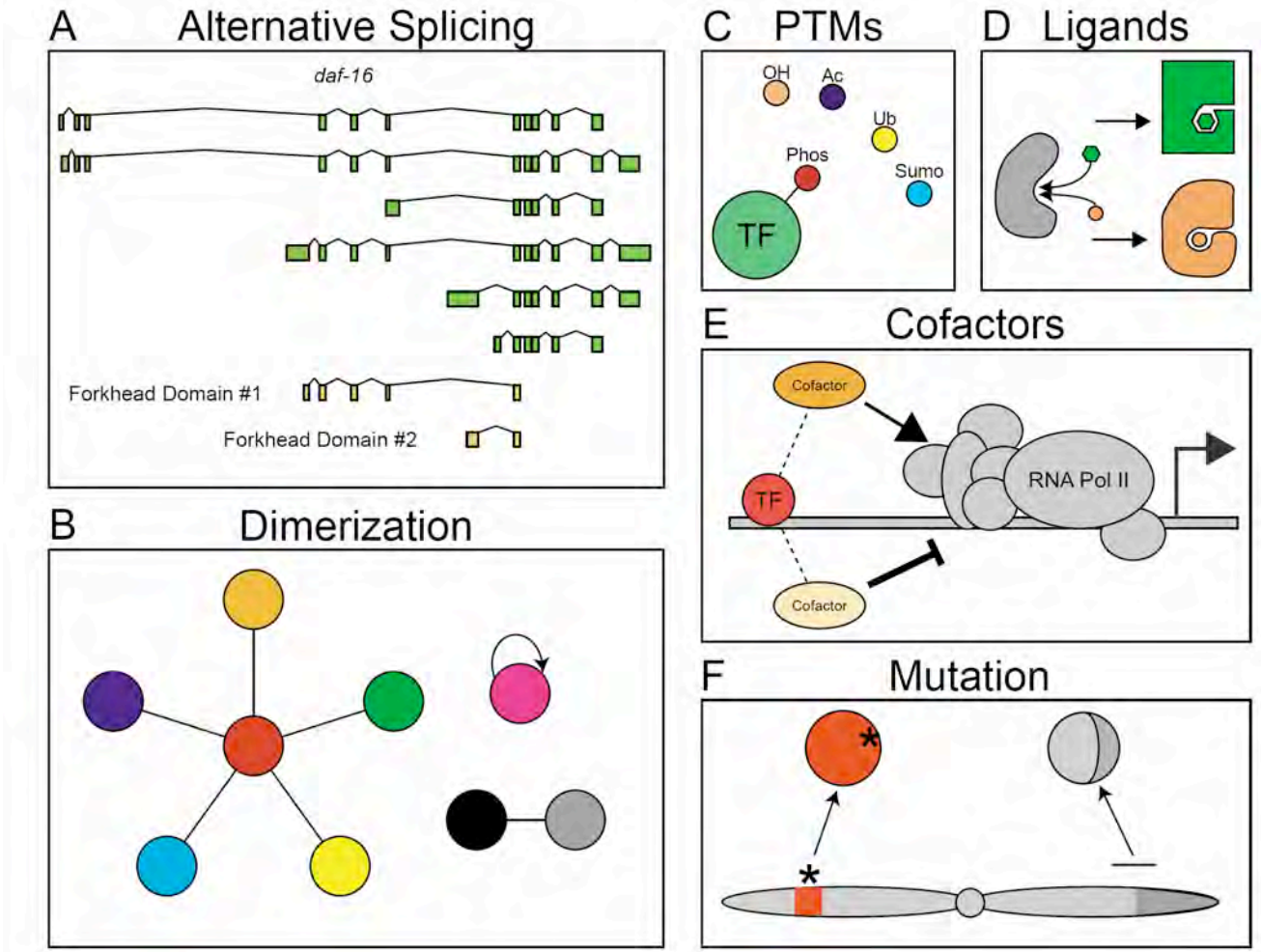

A Alternative Splicing
daf-16
Forkhead Domain #1
Forkhead Domain #2
B Dimerization
C PTMs
OH
Ac
Ub
Phos
Sumo
TF
D Ligands
E Cofactors
Cofactor
TF
RNA Pol II
Cofactor
F Mutation

Figure I-2. Examples of parameters of TF functionality
Various factors influence the functionality of TFs and impact transcription regulatory network modeling and analysis. Whether a TF binds to a promoter, and activates or represses transcription, depends on: (A) alternative splicing that may produce TF variants that carry unique combinations of functional domains involved in the regulation of gene expression. The example shows the *C. elegans daf-16* gene that produces many splice variants, resulting in TFs that contain either of two possible forkhead DNA-binding domains, or neither. (B) Dimerizing TFs potentially combine in different ways to generate a large array of different hetero- or homodimers, each with its own function. (C, D) Some TFs may have an altered function after ligand binding or post-translational modifications (PTMs), such as phosphorylation (Phos), hydroxylation (OH), acetylation (Ac), ubiquitination (Ub), or sumoylation (Sumo). Such modifications can induce different conformational changes, thereby affecting TF functionality. (E) Co-factors can mediate varying affects of TF activity. (F) Mutation or translocation of TF-encoding genes can result in TFs with reduced, enhanced, or novel activity. The asterisk indicates a point mutation and the horizontal line depicts the breaking point of translocation within the chromosome that carries the TF-encoding gene.

Figure I-3

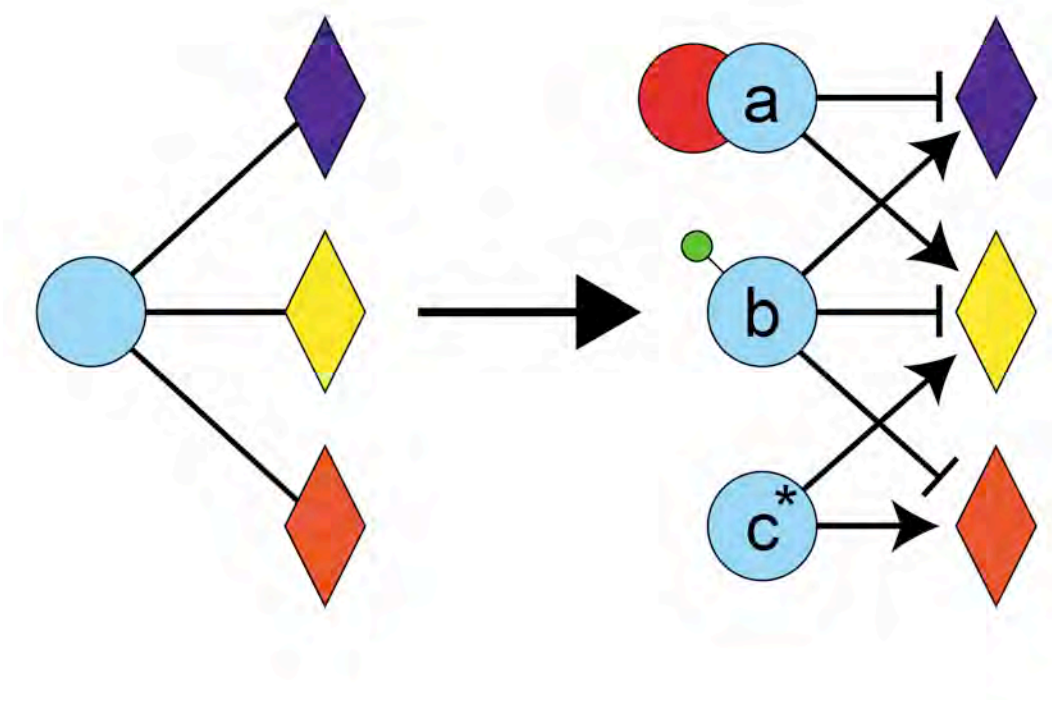


Figure I-3. The integration of multiple TF parameters into novel TRNs
Whereas traditional transcription regulatory networks have been visualized using a single node for a TF protein, the use of individual nodes for each functional TF state may help to depict the regulatory capacity of each TF. TF (blue circle) variant “a” forms heterodimers, variant “b” is post-translationally modified and variant “c” is a result of a mutation in the gene encoding the TF. Each variant has different target genes (diamonds) and/or different effects on those targets (arrows = transcriptional activation, flat arrows = repression).

Figure I-4

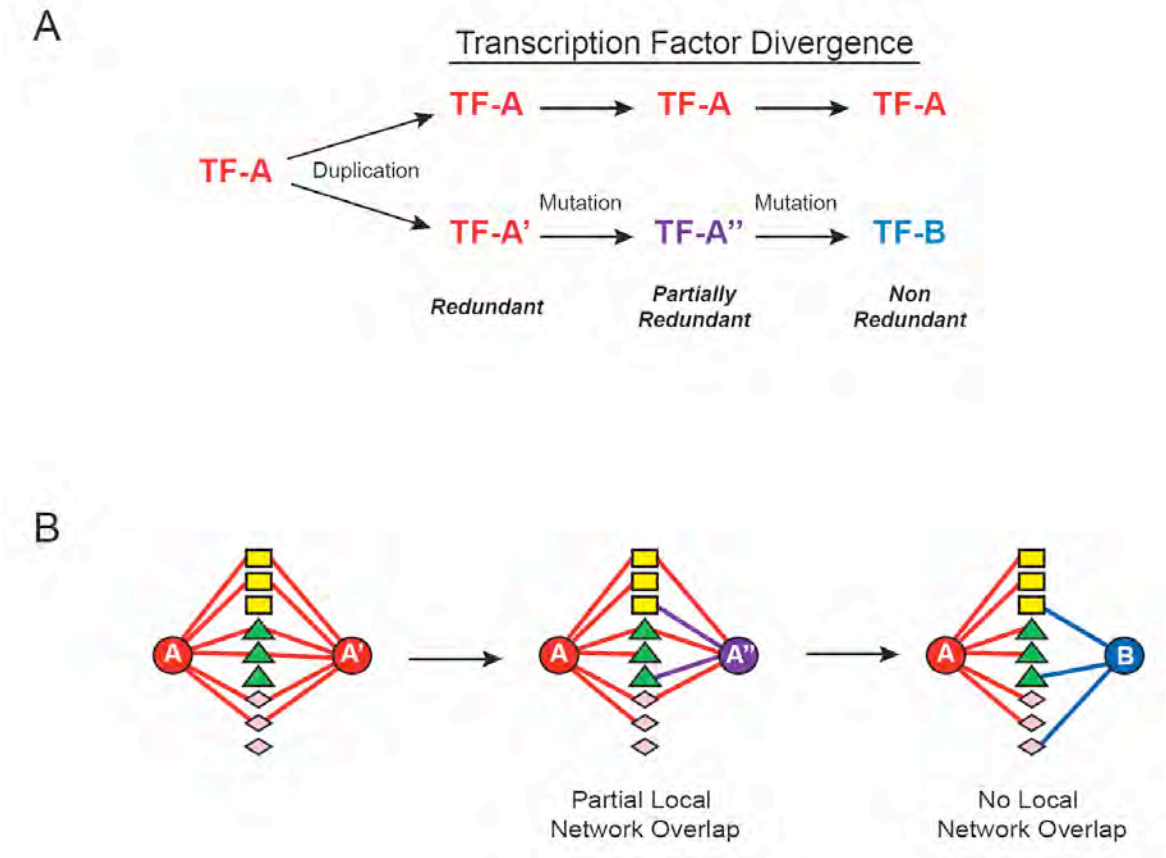
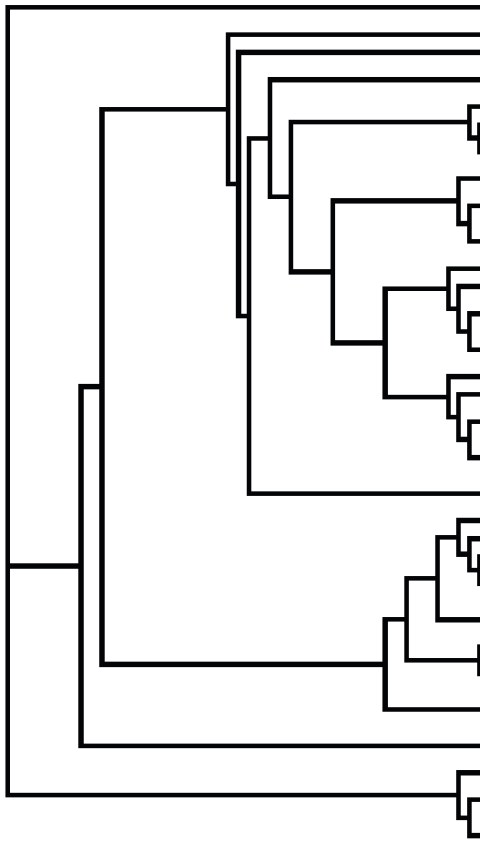


Figure I-4. Functional and molecular divergence in paralogous TF families
(A) Paralogous TFs arise by gene duplication and mutation.
(B) TF divergence can be achieved by the accumulation of molecular and functional differences. Differently shaped nodes (rectangles, triangles and diamonds) between TFs (circles) represent different TF parameters (e.g. dimerization partners, spatiotemporal expression and DNA binding specificities).

Figure I-5



		Closest *D. melanogaster* Homolog	BLASTP E-Value	Closest Human Homolog	BLASTP E-Value
HLH-1	Class Unknown	nautilus CG10250-PA	2.00E-19	myogenic factor 6	8.00E-18
HLH-2	Class I	daughterless CG5102-PA	5.00E-14	transcription factor 3 (E47/E12)	1.00E-15
HLH-11	Class Unknown	cropped CG7664-PA	1.00E-15	transcription factor AP-4 (activating enhancer binding protein 4)	2.00E-14
UNC-3	Class Unknown	knot CG10197-PC, isoform C	1.00E-15	early B-cell factor	1.00E-17
HLH-30	Class III	Mitf CG17469-PB, isoform B	4.00E-15	TFE3	3.00E-18
AHA-1	Class VII	tango CG11987-PA	1.00E-19	aryl hydrocarbon receptor nuclear translocator isoform 3	8.00E-21
HLH-34	Class VII	single-minded CG7771-PB, isoform B	5.00E-08	single-minded homolog 1	2.00E-07
HLH-33	Class VII	germ cell-expressed bHLH-PAS CG6211-PA	0.004	aryl hydrocarbon receptor nuclear translocator-like 2	0.004
AHR-1	Class VII	spineless CG6993-PA	2.00E-14	arylhydrocarbon receptor repressor	6.00E-13
CKY-1	Class VII	dysfusion CG32474-PC, isoform C	3.00E-12	neuronal PAS domain protein 4	3.00E-11
MDL-1	Class IV	Mnt CG13316-PB, isoform B	0.002	MAX interactor 1 isoform b	1.00E-11
MXL-3	Class IV	Max CG9648-PA	1.00E-13	MAX protein isoform e	3.00E-13
MXL-1	Class IV	Max CG9648-PA	4.00E-10	MAX protein isoform e	9.00E-09
MXL-2	Class IV	bigmax CG3350-PA	3.00E-04	transcription factor-like protein 4 isoform alpha	0.006
MML-1	Class IV	Mlx interactor CG18362-PC, isoform C	7.00E-07	transcription factor-like protein 4 isoform alpha	3.00E-05
REF-1A	Class VI	Hairy/E(spl)-related with YRPW motif CG11194-PA	0.12	hairy and enhancer of split 5	0.001
HLH-26A	Class VI	Hairy/E(spl)-related with YRPW motif CG11194-PA	0.014	hairy/enhancer-of-split related with YRPW motif 2	0.1
HLH-25A	Class VI	CG10971 CG10971-PA, isoform A	1.0	golgi autoantigen, golgin subfamily a, 4	0.3
HLH-27A	Class VI	CG10971 CG10971-PA, isoform A	1.7	golgi autoantigen, golgin subfamily a, 4	0.3
HLH-28A	Class VI	Clock CG7391-PE, isoform E	0.092	hairy and enhancer of split 6	0.17
HLH-29A	Class VI	Clock CG7391-PE, isoform E	0.092	hairy and enhancer of split 6	0.17
REF-1B	Class VI	cullin-5 CG1401-PA	1.7	forkhead-associated (FHA) phosphopeptide binding domain 1	0.18
HLH-26B	Class VI	CG4030 CG4030-PA	2.9	inositol hexaphosphate kinase 2 isoform a	0.24
HLH-25B	Class VI	E(spl) region transcript mdelta CG8328-PA	0.053	hairy and enhancer of split 5	0.016
HLH-27B	Class VI	E(spl) region transcript mdelta CG8328-PA	0.053	hairy and enhancer of split 5	0.016
HLH-28B	Class VI	E(spl) region transcript mdelta CG8328-PA	1.7	ecotropic viral integration site 1 isoform a	5.8
HLH-29B	Class VI	E(spl) region transcript mdelta CG8328-PA	1.7	ecotropic viral integration site 1 isoform a	5.8
SBP-1	Class III	Helix loop helix protein 106 CG8522-PA, isoform A	6.00E-12	sterol regulatory element binding transcription factor 1 isoform a	4.00E-13
LIN-22	Class VI	hairy CG6494-PA, isoform A	3.00E-12	hairy and enhancer of split 1	1.00E-13
HLH-16	Class II*	Olig family CG5545-PA	1.00E-10	basic helix-loop-helix domain containing, class B, 4	1.00E-10
CND-1	Class II*	target of Poxn CG7659-PA	2.00E-11	neurogenic differentiation 4	2.00E-15
HLH-31	Class II*	Olig family CG5545-PA	4.00E-15	basic helix-loop-helix domain containing, class B, 5	3.00E-15
HLH-17	Class II*	Olig family CG5545-PA	2.00E-19	basic helix-loop-helix domain containing, class B, 5	5.00E-20
HLH-32	Class II*	Olig family CG5545-PA	2.00E-19	basic helix-loop-helix domain containing, class B, 5	5.00E-20
LIN-32	Class II*	cousin of atonal CG7760-PA	5.00E-13	atonal homolog 1	4.00E-14
NGN-1	Class II*	target of Poxn CG7659-PA	6.00E-12	neurogenin 3	1.00E-12
HLH-8	Class II*	twist CG2956-PA, isoform A	1.00E-13	twist	3.00E-13
HLH-13	Class II*	48 related 2 CG5952-PA	9.00E-15	pancreas specific transcription factor, 1a	1.00E-14
HLH-15	Class II*	Helix loop helix protein 4C CG3052-PA	2.00E-16	nescient helix loop helix 2	1.00E-17
HLH-10	Class II*	atonal CG7508-PA	7.00E-08	transcription factor 21	2.00E-08
HLH-19	Class II*	Hand CG18144-PA	3.00E-07	basic helix-loop-helix transcription factor 15	7.00E-09
HLH-12	Class II*	CG33557 CG33557-PA	6.00E-06	basic helix-loop-helix transcription factor 15	2.00E-05
HIF-1	Class VII	similar CG7951-PA	4.00E-05	hypoxia-inducible factor 1, alpha subunit isoform 2	7.00E-06
HND-1	Class II	Hand CG18144-PA	2.00E-06	pancreas specific transcription factor, 1a	6.00E-07
HLH-3	Class II	asense CG3258-PA	2.00E-09	achaete-scute complex homolog 1	1.00E-10
HLH-14	Class II	asense CG3258-PA	9.00E-07	achaete-scute complex homolog-like 2	3.00E-08
HLH-6	Class II	scute CG3827-PA	2.00E-09	achaete-scute complex-like 4	3.00E-14
HLH-4	Class II	net CG11450-PA	9.00E-07	ASCL3	2.00E-06

*Although dimerization data does not definitively classify HLH-16, HLH-17, HLH-31, and HLH-32 as Class II, their similarity in sequence to the other Class II proteins prompts us to tentatively classify them as Class II bHLH proteins

Figure I-5. Cladogram of aligned *C. elegans* bHLH domains
(Left) bHLH domains were aligned using the Clustal X algorithm (72) and visualized as a cladogram with Tree View (73) (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). The bHLH class designation for each *C. elegans* bHLH protein is indicated immediately to the right of each bHLH protein name. *(Right)* Each *C. elegans* bHLH domain was BLASTed (BLASTP) against the RefSeq protein database for *Drosophila melanogaster* and humans. The closest homolog for each *C. elegans* bHLH domain is indicated with their respective BLASTP E-values. Note that there are 48 bHLH domains aligned for this cladogram because bHLH proteins HLH-25, HLH-26, HLH-27, HLH-28, HLH-29, and REF-1 (the "REF-1 Family") each have two bHLH domains.

Figure I-6

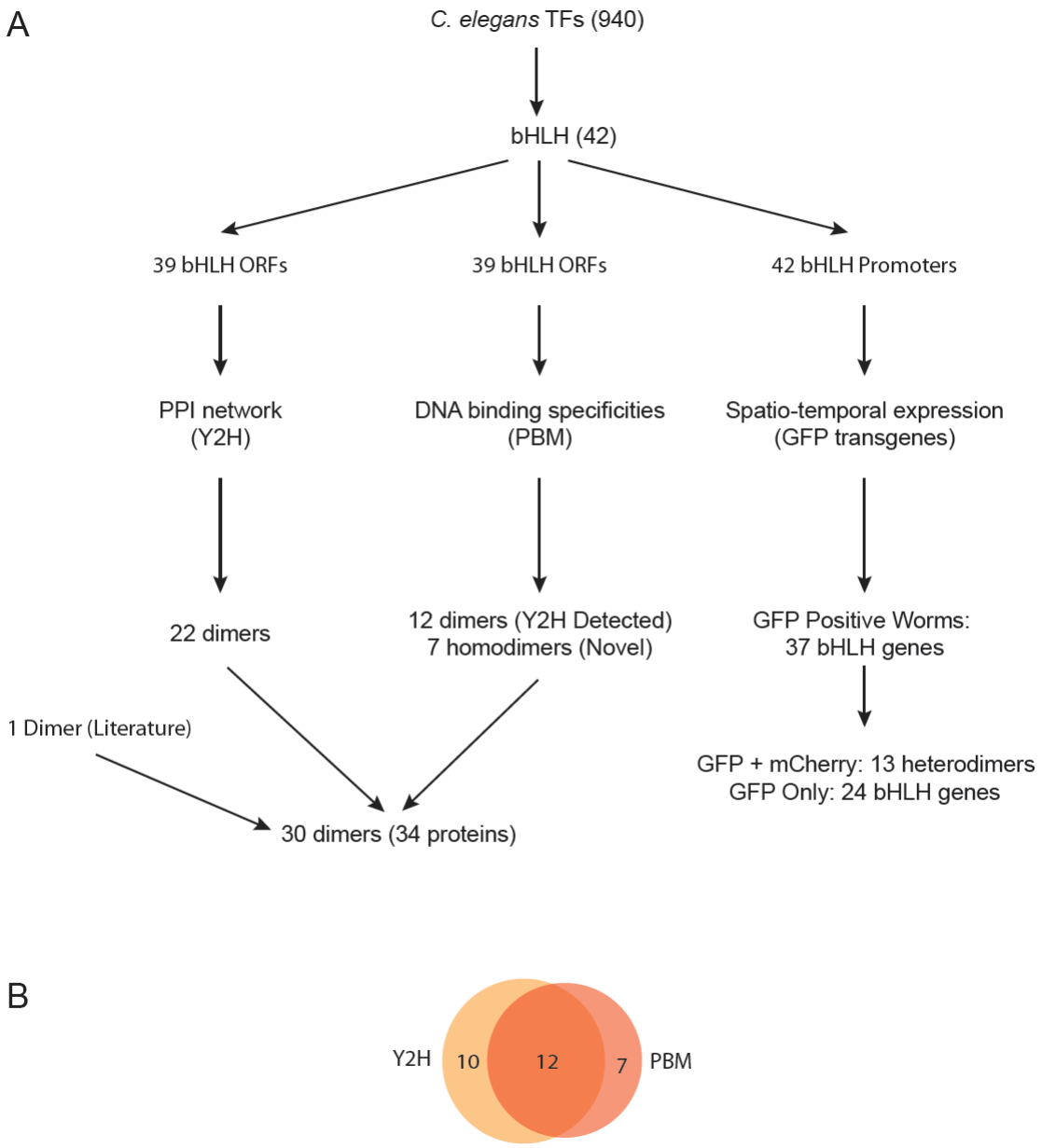


Figure I-6. Schematic overview of experiments
(A) Experimental pipeline, outlining the numbers of bHLH ORF and promoter clones used in each experiment and the data resulting from each experiment. ORFs = open reading frames, PPI = protein-protein interactions, Y2H = yeast two-hybrid, PBM = protein binding microarrays, GFP = green fluorescent protein. *(B)* Venn diagram delineating the overlap in the number of bHLH dimers present in the Y2H and PBM datasets.

PREFACE TO CHAPTER II

This chapter discusses the importance of TF dimerization for many TF families and the implications of the capacity to dimerize. We describe an experimental approach to determine the dimerization network for the *C. elegans* bHLH TFs and discuss the findings that emerged as a result of the analysis.

Much of this chapter has been published separately in:

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJM. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell. 2009 Jul 23;138(2):314-327.

CHAPTER II

The *C. elegans* bHLH Dimerization Network

Abstract

Often integral to the functionality of TFs is the array of protein interaction partners that a particular TF has. TFs may interact with signaling proteins, cofactors, chromatin modifying enzymes as well as other TFs. The most frequent type of TF/TF interaction observed is in the form of homotypic dimerization, in which two TFs from the same family form dimers (homodimers or heterodimers), usually by virtue of the structural properties of the DNA binding domain itself. The basic helix-loop-helix (bHLH) TFs are believed to bind DNA as obligatory dimers, and, therefore, understanding which bHLH TFs dimerize with each other within an organism is critical to understanding the function of this family of TFs as a whole. In this chapter, we describe the technical approach for elucidating the dimerization interaction specificity of the *C. elegans* bHLH TFs and the dimerization interaction network as a whole.

Introduction

Several TF families exhibit the property of homotypic dimerization, whereby two TFs from the same family interact stably to form dimers. These families include the basic leucine zipper (bZIP), the nuclear hormone receptor (NHR), and the basic helix-loop-helix (bHLH) TFs (reviewed in (74)). This dimerization is the result of the structural features of the DNA binding domain, which provides physical complementarity between certain members of each family of TFs, and, hence, enables a stable physical association between the two proteins. Whereas many TFs bind to DNA and carry out their regulatory activity as monomers, many families of TFs (like the bHLH, bZIP and NHR TFs) can only bind DNA as dimers, underlining the importance of dimerization to the functionality of these TFs.

Dimerization creates the possibility for a large number of different functional TFs (dimers) to form from a relatively small number of individual proteins. For example, although there are only 42 predicted *C. elegans* bHLH TFs, the possibility exists that each TF dimerizes with all 42 bHLH TFs, generating as many as 903 possible functional homo- and heterodimers. Alternatively, each TF could heterodimerize exclusively with just one other protein, reducing the number of functional dimers to as low as 21. To elucidate the dimerization specificity, and hence to identify the functional dimers, of the *C. elegans* bHLH TFs, we have systematically analyzed dimerization interaction

specificity by the yeast two-hybrid (Y2H) method (75; 76). In this chapter, we describe the results of the yeast two-hybrid dimerization analysis and discuss the implications of our findings.

Results

Updates to Various C. elegans *bHLH Gene Models*

While attempting to clone all 42 *C. elegans* bHLH ORFs it became apparent that the current gene models for these bHLH genes were not all accurate (clone sources and cloning primers provided in Table II-1). Because ORF cloning begins at the step of PCR amplification of the ORF from a highly representative cDNA library, whether or not we can successfully amplify an ORF can indicate whether or not we have correctly predicted the 5' or 3' limits of the transcribed unit, or whether or not the predicted gene is even expressed. We were unable to amplify ORFs for three of the predicted *C. elegans* bHLH genes (*hlh-12*, *hlh-31*, and *lin-22*), and therefore tested only 39 bHLH ORFS. Improvements in gene finding algorithms and prediction methods, as well as experimental identifications of gene limits using such methods as 5'-RACE and 3'-RACE, can dramatically improve definitions of genetic limits.

For example, the Gene-Finder (77) model for *hlh-19* was originally missing what we now know to be the first exon, which, importantly, encodes the N-terminal portion of the bHLH domain for the HLH-19 protein. As a result, the original ORF clone for *hlh-19* that was used in yeast two-hybrid (Y2H) experiments failed to yield any dimerization interactions, presumably because of the lack of a complete bHLH domain. With the release of the Twinscan *C. elegans* gene predictions (78; 79), an additional exon was predicted 5' to what was thought to be the original first exon. When included into the *hlh-19* gene

model, it became clear that the new exon encoded the missing amino acids of the bHLH domain. When this new ORF was then cloned and tested in the Y2H matrix experiments, we found a novel dimerization interaction between HLH-19 and HLH-2 (see below).

Auto-Activation of C. elegans *bHLH Proteins in Yeast*

As a prerequisite to carrying out well-controlled Y2H experiments, we perform a test for the ability of individual Y2H baits (the Gal4p DNA-binding domain hybrid protein) to activate reporter gene expression in the absence of any interacting partner. We refer to the activation of reporter genes by a single bait protein as “auto-activation”. If left unchecked, auto-activation can lead to the identification of a large number of spurious false positive interactions. Therefore, we systematically test each bait hybrid protein for auto-activation and remove from our experimental matrix any Y2H baits that activate both reporter genes above some threshold. Five of the *C. elegans* bHLH proteins (HLH-30, HLH-2, MXL-3, SBP-1, and HIF-1) exhibited strong auto-activation as Y2H baits and two (AHA-1 and HLH-1) exhibited weak to moderate auto-activation (Figure II-1). Because of strong auto-activation, the five auto-active bHLH baits were removed from further experiments. These same bHLH proteins could, however, be tested as Y2H preys. This allows us to detect heterodimerization interactions between any of these proteins (as prey) and a different non-auto-active bHLH bait. This does, however, exclude us from detecting homodimerization interactions for these five strongly auto-active bHLH proteins.

The C. elegans *bHLH Dimerization Network*

Previous studies in *C. elegans* have identified ten bHLH homo- and heterodimers involving 14 TFs (80-84; 66; 85-87). However, dimerization partners for the majority of *C. elegans* bHLH TFs remained unidentified. Thus, we performed pair-wise Y2H assays to identify all bHLH-bHLH dimers (34). In total, we examined 765 bHLH-bHLH combinations involving 39 bHLH proteins (Figure I-6). In total, we detected 22 dimers (2 homodimers and 20 heterodimers) involving 26 bHLH proteins (Figure II-2A and II-2B). The complete dimerization network is shown in Figure II-3. We supplemented this network with homodimeric interactions for HLH-25, HLH-27, HLH-29, REF-1, HLH-11, MXL-3, and HLH-30, because we detected their specific DNA binding in protein binding microarray (PBM) and/or yeast one-hybrid assays (6) (see Chapter IV). Together, the resulting bHLH network contains 9 homodimers and 21 heterodimers involving 34 proteins.

Dimerization Interactions are Class Specific

One benefit of systematic dimerization analysis is the confidence to make statements about what kind of dimers can or cannot happen. Previously, data regarding dimerization interactions between bHLH proteins were generally acquired as the result of hypothesis-driven experiments that directly test the capability of two bHLH proteins to interact with one another. What these experiments do not reveal, however, is the extent to which any one bHLH protein will dimerize with any other bHLH protein. By performing systematic Y2H matrix

experiments, we can make a call as to how specific any dimerization interactions are for individual bHLH proteins. One interesting finding of the data presented here is that, after testing 765 pairwise bHLH-bHLH dimerization interactions, we find that all but two of the bHLH proteins in our dimerization network dimerize specifically with just one other bHLH protein. The two exceptions, HLH-2 and AHA-1, which we call dimerization hubs, are discussed in more detail below. This finding demonstrates that bHLH dimerization interactions are indeed highly specific. A subsequent finding is that all of the heterodimerization interactions appear to be restricted to the designated bHLH subclasses defined earlier. To this end, HLH-2 (Class I) interacts only with bHLH proteins predicted by sequence to belong to the Class II bHLH proteins. Likewise, Class IV bHLH proteins only dimerize with other Class IV proteins, and Class VII bHLH-PAS proteins only dimerize with AHA-1, the Class VII bHLH-PAS dimerization hub. As described later in this thesis, we believe we have provided further evidence that the bHLH proteins from the “REF-1 Family” are likely members of the Hairy/Enhancer-of-Split class (Class VI) of bHLH proteins.

Two Dimerization Modules: The HLH-2 and AHA-1 Modules

The majority of bHLH proteins exhibit highly specific dimerization as they interact with only a single other bHLH protein (Figure II-3). However, there are two bHLH proteins that interact with multiple other bHLH proteins. The first is AHA-1, the *C. elegans* ortholog of Arnt, that dimerizes with all known class VII members. Members of this class contain a PAS domain that mediates protein-

protein interactions and ligand binding (88). The second is HLH-2, which binds to 14 other bHLH proteins, many orthologs of which are known to interact with HLH-2 orthologs in other organisms (interologs, Table II-2). We refer to the group of interactions between HLH-2 and AHA-1 and their partners as the HLH-2 and AHA-1 modules, respectively. Taken together, the dimerization network displays both specificity and promiscuity as most bHLH proteins interact with one, but some interact with many other bHLH proteins.

HLH-3, HLH-4, and HLH-10 vs. the TF-Array

The fact that so many *C. elegans* bHLH proteins dimerize with HLH-2 raises the question as to how these different dimers distinguish themselves functionally. One possibility is that each HLH-2 partner interacts specifically with other TFs to provide combinatorial control over the expression of target genes. To test the possibility of whether or not HLH-2 partners may provide functional specificity by interacting with other *C. elegans* TFs, we tested three HLH-2 partners for protein-protein interactions with any of the other *C. elegans* TFs for which we have clones available. This collection of 785 *C. elegans* TFs is present in the form of AD-TFs in the appropriate yeast strain for mating to bait strains (either Y2H baits or Y1H baits) (5). We tested HLH-3, HLH-4, and HLH-10 for the ability to interact with any of the TFs in the array, the results of which are displayed in Figure II-4A – II-4C. To our surprise, we found that, although we could reproduce the interactions of each of these TFs with HLH-2, there were no

other TFs with which they could interact, demonstrating the high degree of specificity these proteins exhibit with regards to dimerization.

Discussion

Yeast Two-Hybrid Assays: Data Quality

Several observations indicate that the dimerization network is of high quality. (1) We recovered all nine previously reported dimers for which bHLH ORF clones were available in addition to identifying 20 novel interactions (Figure II-3). (2) We found evolutionarily conserved “interologs” (evolutionarily conserved interactions) (34) (Table II-2). (3) All dimers fall within the bHLH classes, or between class I and class II as expected, which indicates dimerization specificity within the context of the yeast two-hybrid system (Figure II-3). (4) We tested several bHLH proteins for interactions with our entire collection of 785 full-length TFs (5) and found that bHLH-bHLH interactions are indeed highly specific (Figure II-4A, II-4B, II-4C). (5) We used extensively validated Y2H methodology with low expression levels and multiple reporters. This has been shown to result in a low degree of “technical false positives” (89). (6) The Y2H data is validated by the protein binding microarray (PBM) data and *vice versa* since five combinations of bHLH-bHLH proteins that do not dimerize in yeast failed to confer sequence-specific DNA binding in PBM assays (see Chapter IV). (7) The tissue overlap coefficient (TsOC) analysis demonstrates that bHLH proteins that bind to each other in Y2H assays are more co-expressed than bHLH proteins that do not bind to each other in Y2H assays (see Chapter III). (8) All of the proteins that dimerize with each other in yeast “meet” in the worm as indicated by our analysis of spatiotemporal bHLH promoter activity and demonstrated by our co-expression analysis of HLH-

2 and its partners (see Chapter III). We did not detect dimerization partners by Y2H for 13 of the 39 available *C. elegans* bHLH proteins, even though it is likely that all bHLH proteins function as dimers. For some we cannot identify homodimers because they are highly auto-active. This is likely the case for HLH-30 and SBP-1, orthologs of which homodimerize (90; 91). For three, no clone was available (see above).

Dimerization Hub Proteins May Confer Transcriptional Activation Activity

Interestingly, both AHA-1 and HLH-2, the dimerization hubs, are auto-activators in Y2H assays (Figure II-1) whereas their dimerization partners are not (except for HIF-1, Figure II-1, Figure II-3). Previously, we had observed that known transcriptional activators confer strong auto-activation in the context of the Y2H system when fused to the Gal4 DNA binding domain (34). Thus, the bHLH dimerization hubs may confer the transcriptional activation activity to the different dimers of which they are a part, whereas their dimerization partners may contribute specificity in DNA binding. Indeed, HLH-2 orthologs are known transcriptional activators in a variety of organisms (71). Together, these results suggest that HLH-2 and AHA-1 may function as transcriptional activators in *C. elegans* as well. Except for HIF-1, all the other auto-activators form homodimers that may activate transcription *in vivo*. Finally, none of the four proteins that constitute the two class IV heterodimers (Figure II-1, Figure II-3) activate transcription in yeast, which suggests that these dimers may function as

transcriptional repressors, as do some of their mammalian counterparts (Max, Mad) (92).

bHLH Structural Components that Potentially Contribute to Dimerization Specificity

The α-helices of bHLH TFs are the physical components that enable dimerization between partner bHLH proteins. When dimerized, two helices from each bHLH monomer face the two helices from the other bHLH protein to form a left-handed four-helix bundle, the core of which is almost entirely comprised of hydrophobic residues (71). It is thought that this hydrophobic interface is most important for the formation of stable dimers of bHLH TFs. Therefore, it is possible that specificity of dimerization may be achieved when two individual bHLH monomers have compatible hydrophobic surfaces that can fit snugly into one another. Determining this compatibility *a priori* has proven a difficult task, as there may be no simple relationship between amino acid sequence and dimerization specificity. Our systematically obtained dimerization network, however, provides a unique opportunity to examine sequence and/or structural determinants that may dictate partner choice within bHLH classes.

An interesting question that arises from the dimerization network is how a protein like HLH-2 can specifically interact with so many bHLH partners, but that none of the partners can interact with each other. To investigate this we performed a Clustal W alignment (93) of the individual bHLH domain amino acid sequences from HLH-2 and all of its partners, as well as HLH-2 with several

HLH-2 orthologs (i.e. E proteins) from other species (Figure II-5). As can be seen from the alignment of the bHLH domain amino acid sequences, there are several similarities between HLH-2 partners, as well as HLH-2 itself. There are two residues, however, that are consistently present among HLH-2 partner sequences that are lacking from HLH-2 and HLH-2 orthologs.

The first distinguishing residue is a proline at either position 28 or 29 (according to the numbering scheme of Atchley and Fitch (94)) at the end of helix #1 of HLH-2 partners (P28/29, Figure II-5 bottom panel, left red box). HLH-2 has a threonine at the equivalent position, and its orthologs have either a methionine or a glutamine (Figure II-5 top panel, left red box). Previously, Ellenberger et al. observed that helix #1 of mammalian Max (class IV) is one helical turn shorter than helix #1 of the HLH-2 ortholog E47, and that this is likely due to P28, which could act to break the turn of the helix (95). Thus, P28 of the HLH-2 partners may also shorten helix #1, whereas helix #1 of HLH-2 may be extended due to a lack of a proline at this position.

The second distinguishing residue is a bulky aromatic tyrosine at position 60 in helix #2 of HLH-2 partners (Y60, Figure II-5 bottom panel, right red box). There is a smaller, hydrophobic valine at the analogous position in HLH-2 and its orthologs (V60, Figure II-5 top panel, right red box). Y60 in Max has been reported to “cap” the C-terminal end of helix #1 (96). The aromatic ring of tyrosine 60 (in helix #2), which may be resting at the top of helix #1 in HLH-2

partners, may sterically hinder further turns of helix #1, thereby contributing to the shorter helix #1 for those bHLH proteins.

The fact that all HLH-2 partners contain both P28/29 and Y60 suggests that their helix #1 is shorter than that of HLH-2 (Figure II-5). This is supported by the recently published co-crystal structure of E47/NeuroD heterodimers bound to DNA that shows that helix #1 of E47 is one turn longer than helix #1 of NeuroD (Figure II-6) (97). Taken together, it is likely that an extra helical turn in helix #1 of HLH-2 enables additional surface contacts between helix #1 of HLH-2 and helix #2 of HLH-2 partners, and between helix #1 of HLH-2 and its own helix #2. Based on these observations, we propose that HLH-2 partners fail to interact with each other because helix #1 is not long enough to stabilize dimerization.

Most of the members of the other bHLH classes do contain P28, but lack Y60. This suggests that they may have a shorter helix #1, which may help to explain why they do not interact with HLH-2 partners, but does not explain why they do not dimerize with HLH-2. We could not find any other obvious differences between the AHA-1 module or class IV proteins and the HLH-2 partners in primary amino acid sequence. However, in contrast to the HLH-2 partners, class IV members exhibit less similarity to each other in their two dimerizing helices, and the same is true for the AHA-1 module. This may suggest that additional protein domains such as the leucine zipper (class IV) and the PAS domain (class VII) may contribute to specific dimerization of these proteins.

bHLH Proteins with no Yeast Two-Hybrid Data

Although most evidence on bHLH proteins suggests that bHLH proteins must dimerize and bind to DNA to carry out their function, a limited number of cases suggest that some bHLH proteins may perform a major part of their biological roles independent of DNA binding even though relevant "target" genes are affected by their presence (98; 99). This raises the possibility that at least some bHLH proteins can perform their biological function independent of what is thought to be their primary molecular action: dimerization and DNA-binding. That said, it is possible that this is the case for many of the *C. elegans* bHLH proteins for which no dimerization data (e.g. HLH-16, HLH-17, HLH-31, etc.) or DNA binding data (e.g. HLH-6, HLH-13, HND-1, etc., see below) were obtained.

Known Human bHLH Network

Although a detailed and thorough dimerization network analysis of human bHLH TFs would be tremendously insightful, performing such an analysis on the human bHLH TFs would certainly prove to be difficult for a number of reasons: (1) The number of bHLH TFs in humans is likely to be at least 110 members, making pairwise dimerization analysis much more intensive (~6100 experiments) than for the 42 members of *C. elegans* (~900 experiments); (2) Performing any *in vivo* analysis of bHLH gene expression or target gene expression in humans is practically infeasible, or at least limited, since data collection may be limited to voluntarily contributed tissue samples, biopsies, blood samples, etc. This type of data collection certainly could not be applied to all tissue types and all life stages.

(3) Likewise, the creation of a comprehensive/representative human cDNA library for the cloning of human bHLH ORFs would also be difficult or impossible to acquire. Beginning such an integrated network analysis is more practical in *C. elegans* both at the level of data acquisition and data analysis.

To get an idea of what is currently known about the human bHLH dimerization network, we can make use of publicly available datasets provided through online databases. The online database DBD (100) provides lists of predicted transcription factors from a variety of organisms. Another online database, BioGRID (101), provides information about experimentally defined protein-protein interactions for many different organisms. By defining the list of human bHLH TFs using the DBD database and querying the BioGRID database, we can reconstruct the currently known human bHLH dimerization network as shown in Figure II-7. An interesting observation is that, like the *C. elegans* bHLH network, the human bHLH dimerization interactions, with some exceptions, appear to be primarily restricted to class definitions: Class I bHLH TFs (TCF3, TCF4, TCF12, etc.) dimerize with Class II and Class V (ID1-3) bHLH TFs, Class III bHLH TFs dimerize with Class III and Class IV, and Class VII bHLH TFs (the bHLH-PAS TFs ARNT, NPAS, SIM, AHR) dimerize exclusively within their own class.

Another interesting observation is that, as a result of the expansion of the bHLH family, members such as TCF3, TCF4, and TCF12, which likely share a common ancestor with HLH-2, also exhibit characteristics of dimerization hubs.

This also appears to be the case for ARNT and ARNTL, both of which likely share a common ancestor with AHA-1. Perhaps this expansion of bHLH TF hubs is a contributor to the biological complexity that is conferred onto higher eukaryotic systems such as humans.

It is important to note, however, that this human bHLH network is likely incomplete, since dimerization interactions between human bHLH proteins have not been systematically determined. As can be seen from the human network (Figure II-7), almost half of the human bHLH TFs lack any dimerization interaction data at all. Future systematic studies of human bHLH proteins will be required to asses the full degree of connectivity within the human bHLH network, and whether or not the aspects of this connectivity contribute to the complexity of human biology.

HLH-33: A Novel bHLH-PAS protein?

As mentioned earlier, each of the bHLH proteins appear to dimerize specifically with members of its own class. Such is the case for the bHLH-PAS proteins, as we can see that every known bHLH-PAS protein in *C. elegans* is grouped into a single module of the dimerization network. These bHLH-PAS TFs include AHA-1, HIF-1, AHR-1, CKY-1, and HLH-34. The other TF that dimerizes with AHA-1 in this module that doesn't immediately appear to belong to the bHLH-PAS class of bHLH TFs (Class VII) is the HLH-33 protein. At first glance, this protein does not contain a readily apparent PAS domain. If we look at some of HLH-33's closest homologs in other species, however, we find that HLH-33 is

most closely related (albeit weakly) outside the *Caenorhabditis* clade to the zebrafish Clock3 protein, a known bHLH-PAS protein. It is interesting to speculate that HLH-33 may indeed be a member of the bHLH-PAS family that is undergoing divergence to the point of being barely recognizable as a bHLH-PAS protein, and may lose it's capacity to dimerize with this class in the near evolutionary future.

Summary

Critical to understanding the function of dimerizing TFs is the knowledge of the identity of the dimerization partners for each of these TFs. Because these TFs often function as obligate dimers, knowing if a TF forms a homodimer or a heterodimer with a different TF is essential to the identification of functional TF dimer-complexes which carry out the function of gene regulation. We have systematically analyzed the dimerization preferences for the *C. elegans* bHLH TFs using the yeast two-hybrid assay. Our results indicate the presence of two dimerization hubs, HLH-2 and AHA-1, show that most bHLH TFs dimerize specifically with just one other bHLH TF, and show that heterodimerization interactions take place between bHLH TFs within the same class or group of classes. We also observe that some *C. elegans* bHLH TFs, including the two dimerization hubs, can activate gene expression in yeast, suggesting that they may be transcriptional activators in *C. elegans* as well. Incorporating data from subsequent chapters reveals a dimerization network of 30 bHLH dimers involving 34 bHLH proteins. This network represents the first systematically derived

dimerization network for all bHLH TFs in any organism. The following chapters will describe the analysis of spatiotemporal bHLH TF expression and the DNA-binding specificities and their incorporation, along with the dimerization network, into a multiparameter integrated bHLH network.

Materials and Methods

Open Reading Frame Cloning

We obtained 28 *C. elegans* bHLH-encoding open reading frames (ORFs) as Gateway Entry clones from the *C. elegans* ORFeome (102), and one from the wTF2.1 clone array (5). Ten additional ORFs were PCR-amplified *ab initio* from a *C. elegans* cDNA library and subsequently Gateway-cloned into the Donor vector pDONR-221 as described (103). Finally, we failed to PCR-amplify three ORFs (*hlh-12*, *hlh-31* and *lin-22*) from the cDNA library. All available ORFs were cloned by Gateway LR reactions into pAD-DEST and pDB-DEST for yeast two-hybrid (Y2H) experiments as described (104). All Y2H constructs were sequence-verified. Clone sources and primer sequences are provided in Table II-1.

Y2H Assays

Y2H assays were performed as described (104). Briefly, all DB-bHLH and AD-bHLH destination clones were transformed into yeast strains MaV103 and MaV203, respectively (104). All DB-bHLH clones were tested for auto-activation, *i.e.,* the ability to activate reporter gene expression in the absence of any AD clone (105). Highly auto-activating DB-bHLH strains were omitted from further experiments. The remaining DB-bHLH strains were grown as a lawn on 15-cm plates of synthetic complete media lacking leucine (Sc-Leu). AD-bHLH strains were spotted in rows of 12 onto 15-cm plates of Sc medium lacking tryptophan (Sc-Trp). The DB-bHLH lawn and AD-bHLH spots were replica-plated on top of each other onto YEPD media and incubated overnight at 30°C. Diploids were

selected on Sc-Leu-Trp media by overnight incubation at 30°C. The next day, diploids were replica-plated onto Sc-Leu-Trp media lacking histidine and containing 20 mM 3-aminotriazole (3AT) (Sc-Leu-Trp-His+20 mM 3AT) and onto YEPD media with nitrocellulose filters for β-galactosidase assays. Interaction phenotypes were assessed 3-5 days after plating. See (104) for details.

Acknowledgements

We would like to thank Kristen Alexa for help with the original *C. elegans* bHLH ORF cloning into Y2H vectors and pilot Y2H experiments. We would also like to thank Bart Deplancke for cloning the hlh-32 bHLH ORF into a Gateway compatible Donor Vector. We thank Job Dekker, and members (current and past) of the Walhout and Dekker labs for critical reading of relevant manuscripts.

Table II-1. *C. elegans* bHLH ORF cloning information

ORF	*gene*	Clone Source	Forward cloning primers	Reverse cloning primers	Full Length? (WS190)	Mutation(s) in GST-Clone for PBMs
C25A1.11	*aha-1*	ORFeome 1.1			yes	
C41G7.5	*ahr-1*	ORFeome 1.1			yes	
C15C8.2	*cky-1*	ORFeome 1.1			yes	I476V
C34E10.7	*cnd-1*	ORFeome 1.1			yes	
F38A6.3	*hif-1*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGAAGACAATCGGAAAAGAA	GGGGACCACTTTGTACAAGAAAGCTGGGCAAGAGAGCATTGGAAATGGG	no	
B0304.1	*hlh-1*	ORFeome 1.1			yes	
ZK682.4	*hlh-10*	ORFeome 1.1			yes	
F58A4.7	*hlh-11*	ORFeome 1.1			no	
C28C12.8	*hlh-12*	unable to be cloned			not cloned	
F48D6.3	*hlh-13*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGATTCATCGTATGATTCATATTACTG	GGGGACCACTTTGTACAAGAAAGCTGGGCGCCACTTGATCCAATTGAGC	no	
C18A3.8	*hlh-14*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGCCAAGAAGAATCAAGTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTAATGGTGTGGATAATTGGAATATGA	no	
C43H6.8	*hlh-15*	ORFeome 1.1			yes	
DY3.3	*hlh-16*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAGGCTTGAATGAGCAAGAACA	GGGGACCACTTTGTACAAGAAAGCTGGGTTTGTTGACACTTTGAGCATTTTG	no	
F38C2.2	*hlh-17*	ORFeome 3.1			yes	

F57C12.3	*hlh-19*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTCACGTGAACGTGCTAAC	GGGGACCACTTTGTACAAGAAAGCTGGGCAATCATAGTTCACAACAAAATGACA	yes
M05B5.5	*hlh-2*	ORFeome 1.1			yes
C17C3.7	*hlh-25*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCCAAAAGTTATTCAGTCTTCAA	GGGGACCACTTTGTACAAGAAAGCTGGGAGTGATGGAAGAATGAATCGGAG	no
C17C3.8	*hlh-26*	ORFeome 1.1			yes
C17C3.10	*hlh-27*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCCAAAAGTTATCCCATCTTC	GGGGACCACTTTGTACAAGAAAGCTGGGTAGTTACTAATATCGACGGTTTCTTCATT	no
F31A3.2	*hlh-28*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCCAAAAGTACATCAAGCAAC	GGGGACCACTTTGTACAAGAAAGCTGGGCAGCCAATAATATCGATATCTTCCTC	no
F31A3.4	*hlh-29*	ORFeome 1.1			yes
T24B8.6	*hlh-3*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTACATCCACCAAAATTCCGTCGTCA	GGGGACCACTTTGTACAAGAAAGCTGGGTATACGGGAGACTGTTCTGGAGTT	no
W02C12.3	*hlh-30*	ORFeome 1.1			yes
F38C2.8	*hlh-31*	unable to be cloned			not cloned
Y105C5B.29	*hlh-32*	wTF array			yes
Y39A3CR.6	*hlh-33*	ORFeome 1.1			no
T01D3.2	*hlh-34*	ORFeome 1.1			yes
T05G5.2	*hlh-4*	ORFeome 1.1			yes
T15H9.3	*hlh-6*	ORFeome 1.1			yes
C02B8.4	*hlh-8*	ORFeome 1.1			yes

C44C10.8	*hnd-1*	ORFeome 1.1			yes	
Y54G2A.1	*lin-22*	unable to be cloned			not cloned	
T14F9.5	*lin-32*	ORFeome 1.1			no	
R03E9.1	*mdl-1*	ORFeome 1.1			yes	D48E
T20B12.6	*mml-1*	ORFeome 1.1			yes	
T19B10.11	*mxl-1*	ORFeome 1.1			yes	
F40G9.11	*mxl-2*	ORFeome 1.1			yes	
F46G10.6	*mxl-3*	ORFeome 1.1			yes	
Y69A2AR.29	*ngn-1*	ORFeome 3.1			yes	
T01E8.2	*ref-1*	cloned *ab initio*	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGTCCTCATCAGTACCCCAC	GGGGACCACTTTGTACAAGAAAGCTGGGTATTCCCATGGTCTGAACAGCTT	yes	S255G
Y47D3B.7	*sbp-1*	ORFeome 1.1			no	E406G, T896A
Y16B4A.1	*unc-3*	ORFeome 1.1			yes	F73L, S125G

Table II-1. *C. elegans* bHLH ORF cloning information
Listed are the ORF names and gene names for each *C. elegans* bHLH TF in this study. Also listed are the clone source (e.g. ORFeome), cloning primers (where applicable), whether or not the clone is full length (based on Wormbase version WS190 annotations), and mutations that remained in the ORFs used in GST vectors for PBM experiments. Red text in primer sequences indicates ORF-specific sequence; black text indicates Gateway tails.

Figure II-1

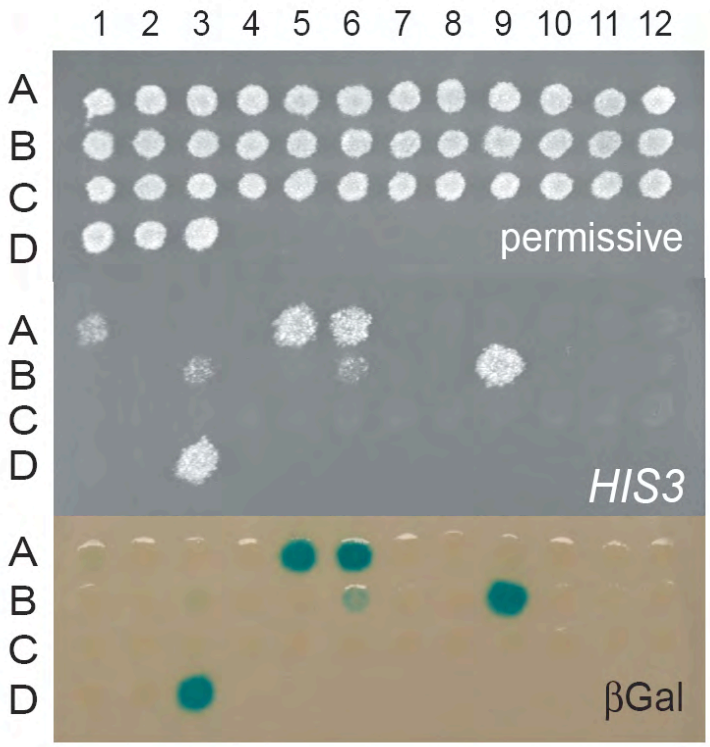


Figure II-1. Auto-activation of DB-bHLH Y2H baits.
Top - DB-bHLH yeast strains were plated in spots on permissive media; middle - activation of the *HIS3* reporter gene ; bottom – activation of the LacZ reporter gene (βGal). Auto-activators are: A1 - DB-AHA-1; A5 - DB-HLH-30; A6 - DB-HLH-2; B3 - DB-HLH-1; B6 - DB-MXL-3; B9 - DB-SBP-1; D3 - DB-HIF-1.

Figure II-2A

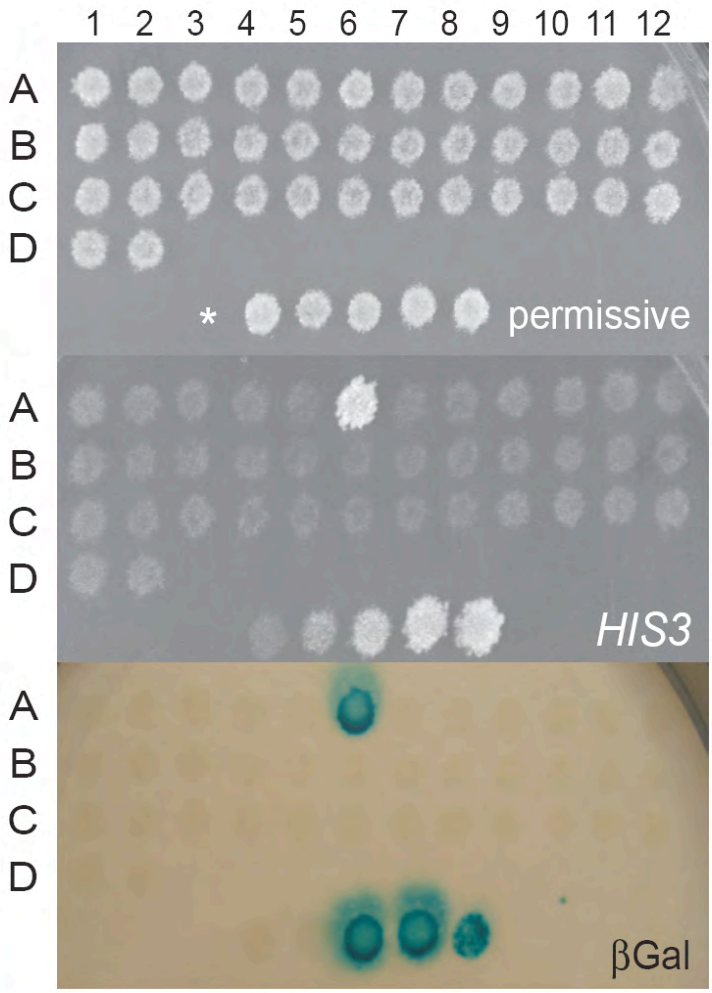


Figure II-2A. Example of Y2H matrix assay using DB-HLH-15 as bait
Top – permissive media; middle - activation of the *HIS3* reporter gene; bottom - activation of the *LacZ* reporter gene (βGal). * Bottom spots in each panel - Y2H controls (104). The Y2H positive in spot A6 is AD-HLH-2.

Figure II-2B

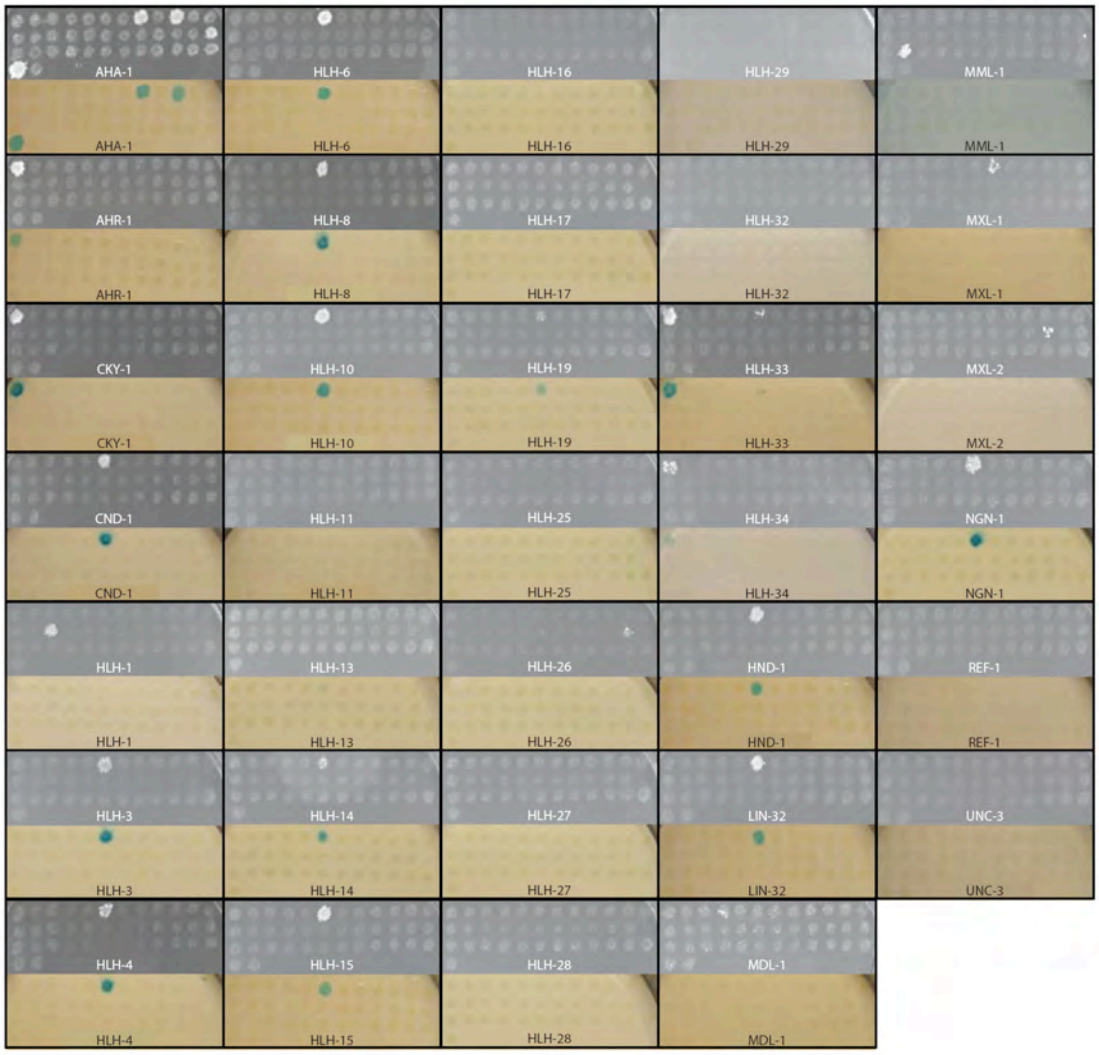


Figure II-2B. Raw yeast two-hybrid data for the *C. elegans* bHLH TFs
As in Figure II-2A, shown are the *HIS3* and LacZ reporter gene readouts for each DB-bHLH bait hybrid protein (permissive plates not shown). Growth of yeast in the top panel for each protein indicates *HIS3* expression; blue colored spots of yeast in the bottom panels indicate expression of the LacZ reporter gene.

Figure II-3

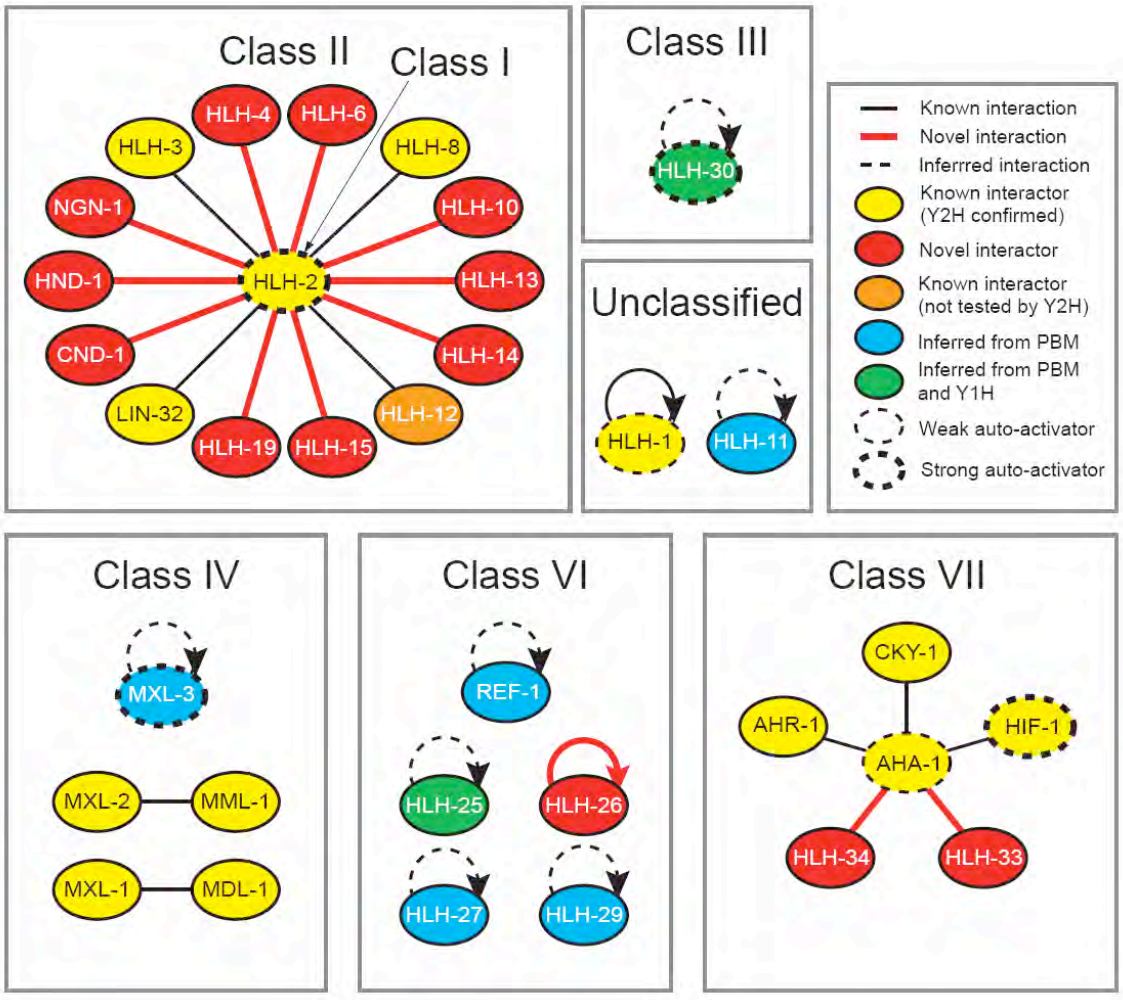


Figure II-3. The bHLH dimerization network.
Each oval represents a single *C. elegans* bHLH protein. Lines connecting ovals represent dimerization interactions between the indicated bHLH proteins. Refer to the legend for what each color represents. Note that curved arrows indicate homodimerization interactions, and dashed curved arrows indicate inferred homodimerization interactions. PBM – protein binding microarray experiments, Y1H – yeast one-hybrid experiments.

Table II-2

bHLH #1	bHLH #2	Interolog	Homolog #1	Homolog #2	Species
AHA-1	AHR-1	yes	ARNT	AHR	Human
AHA-1	HIF-1	yes	ARNT	HIF1A	Human
AHA-1	CKY-1	yes	ARNT2	NPAS4 (NXF)	Human
AHA-1	HLH-33	no			
AHA-1	HLH-34	yes	ARNT2	SIM2	Human
HLH-2	CND-1	yes	E47 (TCF3)	NeuroD1	Mouse
HLH-2	HLH-3	yes	TCF4	ASCL1	Human
HLH-2	HLH-4	yes	TCF4	ASCL1	Human
HLH-2	HLH-6	yes	TCF3	TAL1	Human
HLH-2	HLH-8	yes	TCF3	TWIST1	Human
HLH-2	HLH-10	yes	E2A (TCF3)	TCF21	Human
HLH-2	HLH-13	yes	TCF3	PTF1A	Human
HLH-2	HLH-14	no			
HLH-2	HLH-15	yes	TCF3	LYL1	Human
HLH-2	HLH-19	yes	TCF12	ASCL4	Human
HLH-2	HND-1	yes	TCF3	PTF1A	Human
HLH-2	LIN-32	no, but found before in worms			
HLH-2	NGN-1	yes	E47 (TCF3)	NeuroD1	Mouse
MXL-1	MDL-1	yes	MAX	MXI1	Human
MML-1	MXL-2	yes	MONDOA	MLX	Human
HLH-26	HLH-26	worm-specific			
HLH-1	HLH-1	no			

Table II-2. Interologs
Several bHLH dimerization interactions identified are evolutionarily conserved interologs.

Figure II-4A. DB-HLH-3 vs. AD-TF-Array

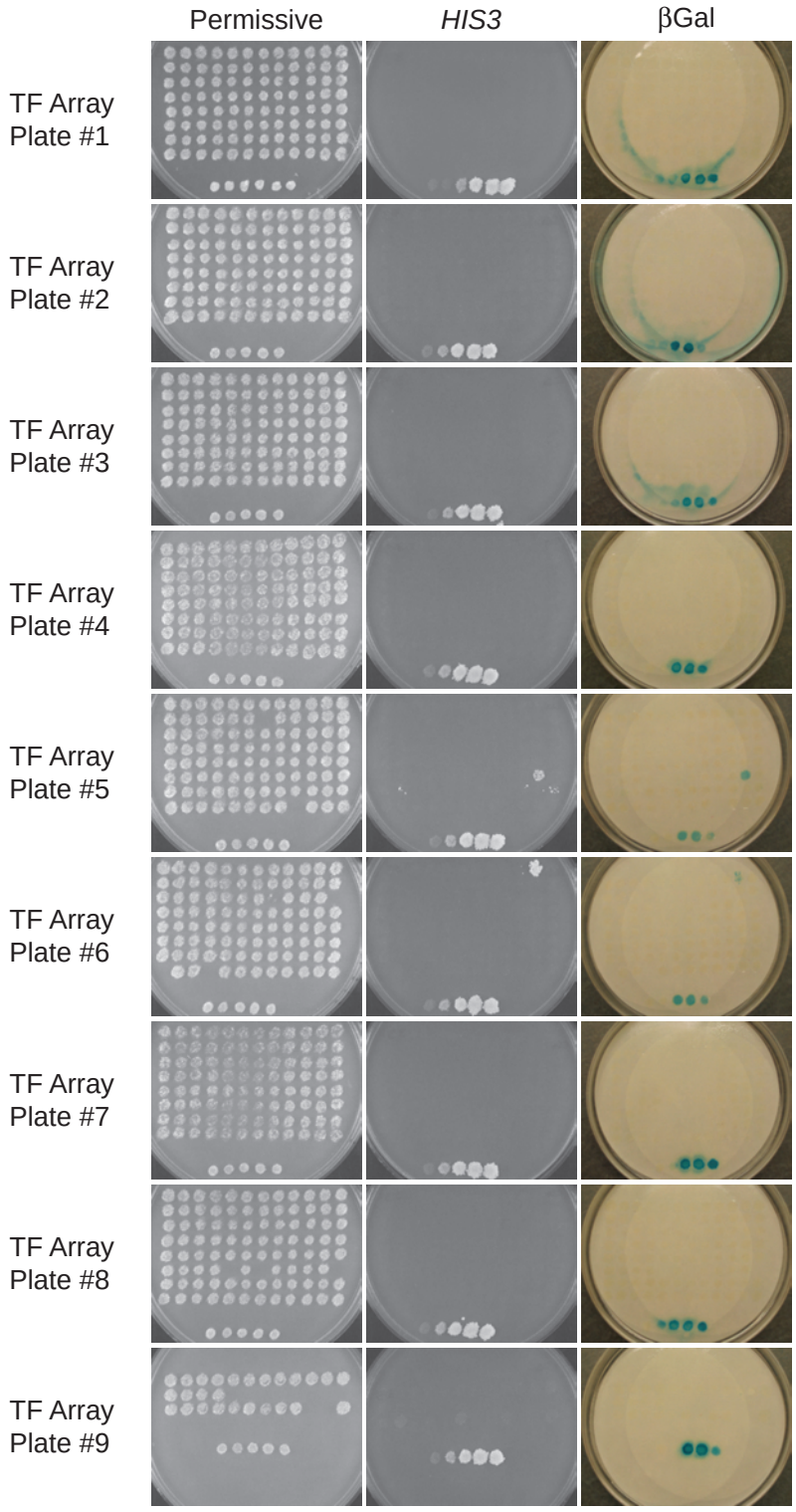


Figure II-4B. DB-HLH-4 vs. AD-TF-Array

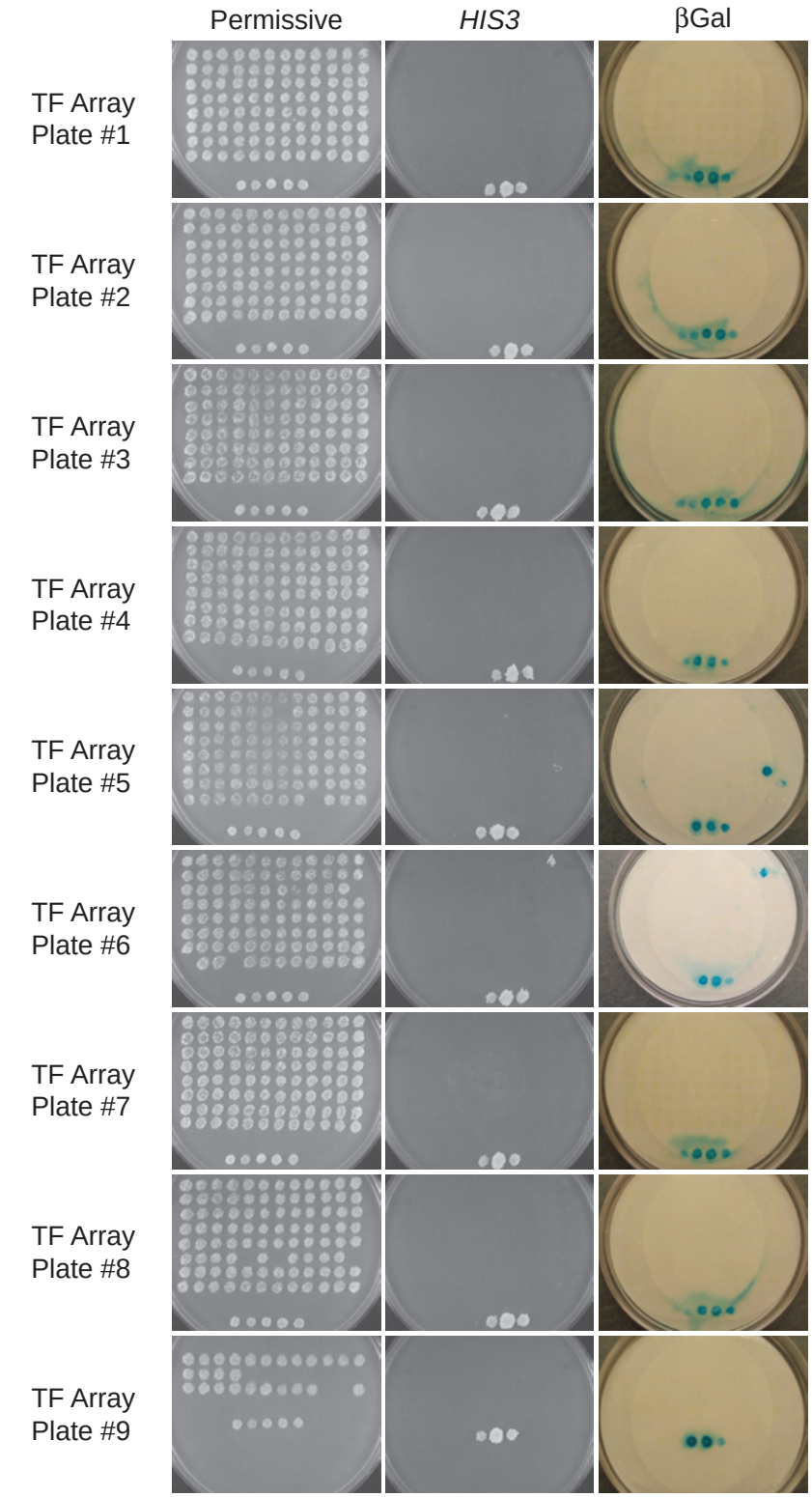


Figure II-4C. DB-HLH-10 vs. AD-TF-Array

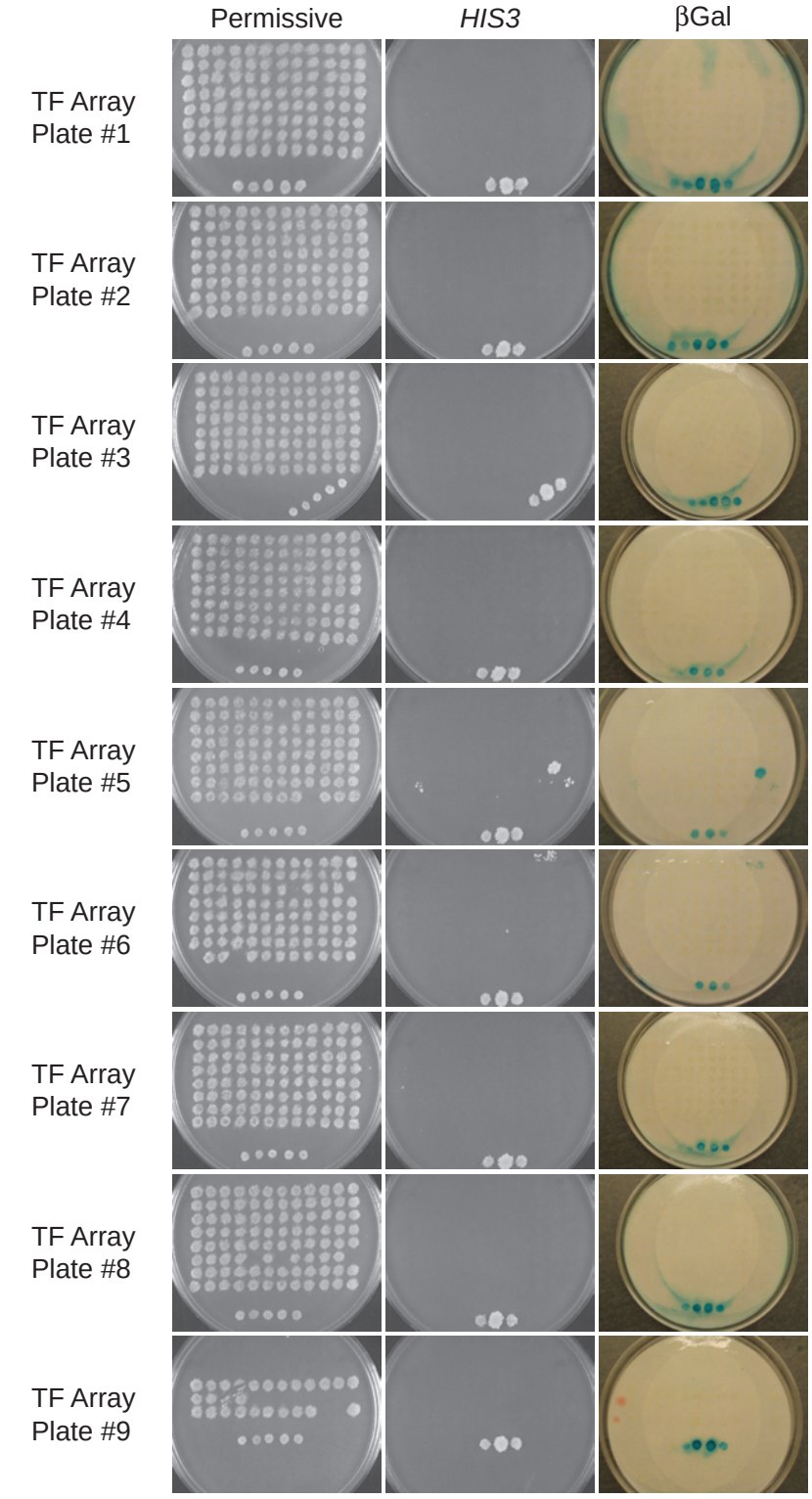


Figure II-4. Individual bHLH proteins tested against AD-TF-Array
Three different bHLH proteins were tested for protein-protein interactions against the entire collection of AD-TFs available in our TF clone resource (5). In this particular version of the TF-array, all AD-TFs are plated over nine 15cm plates, which can then be used in a yeast two-hybrid or yeast one-hybrid mating assay. The results of a mating experiment between the AD-TF array and DB-HLH-3 (A), DB-HLH-4 (B), and DB-HLH-10 (C) are shown. The left column of pictures in each panel represents permissive growth of diploid yeast resulting from the mating of each DB-bHLH TF strain to the TF-array. The second and third columns, respectively, represent the *HIS3* and LacZ ("βGal") reporter gene expression readout. Growth of yeast spots in the second column indicates *HIS3* expression and blue colored spots in the third column indicate β–Galactosidase expression. The experiment reproduces the interaction of these three TFs with HLH-2 (TF Array plate #5, row F, column 11). The other apparent interaction (TF Array plate #6, row A, column 11) is a contamination, as the interaction cannot be reproduced.

Figure II-5

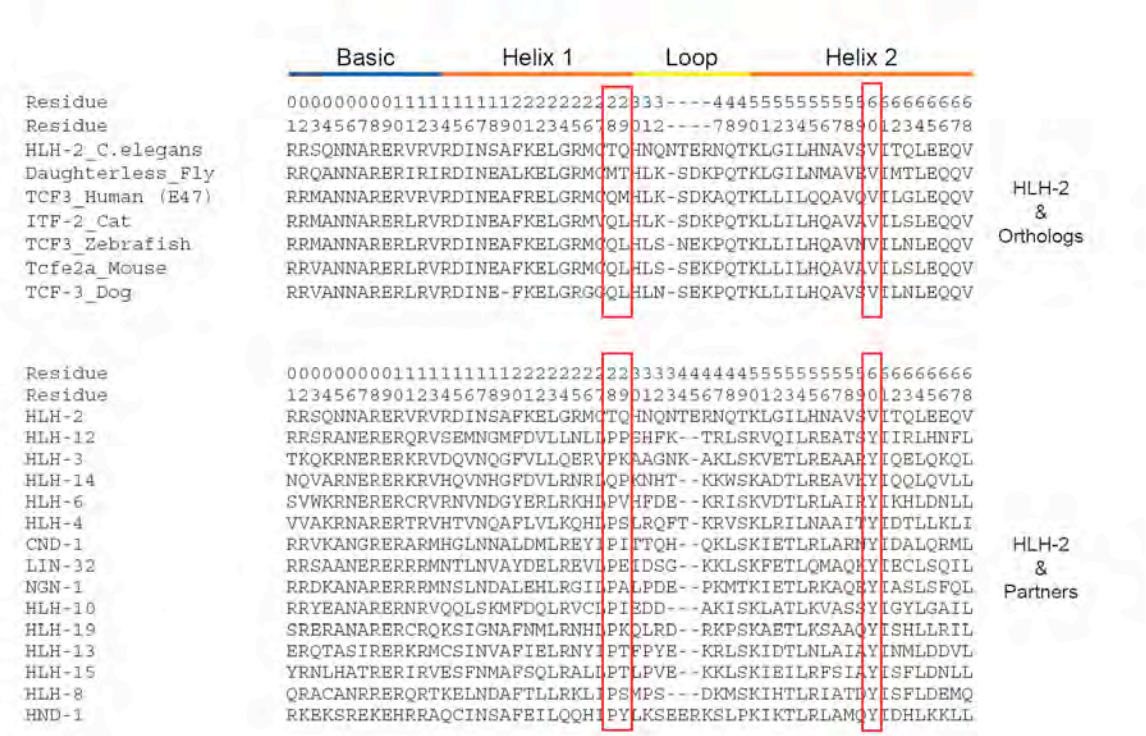


Figure II-5. Alignment of bHLH domains of HLH-2, orthologs, and partners
Residue numbering and secondary structure demarcations are based on bHLH domain consensus annotations by Atchley and Fitch (94). Red boxes indicate the 28th (or 29th) and 60th residues of each bHLH domain (see text for details). (Top panel) Clustal W (93) alignment of HLH-2 bHLH domain and the bHLH domains of several of its orthologs. (Bottom panel) Clustal W alignment of HLH-2 and its dimerization partners.

Figure II-6

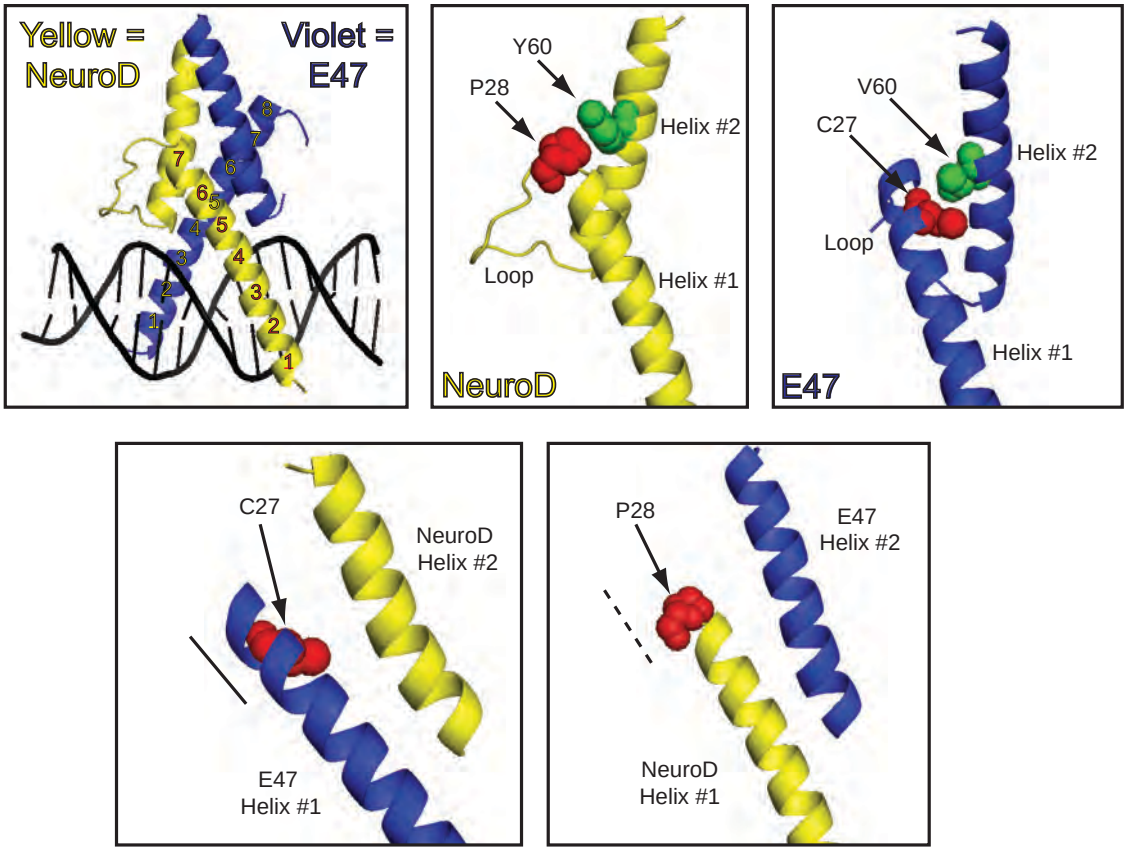


Figure II-6. Structural insight from the E47/NeuroD crystal structure
Utilizing amino acid sequence alignments and known crystal structures of bHLH protein dimers (like this one from Longo *et al* (2008) *Biochemistry* (97)) allows the generation of specific hypotheses as to which amino acid residues are likely important determinants of bHLH dimerization specificity. The solid and dashed lines in the bottom two panels indicate the extra helical turn helix #1 for E47 (solid line, left) and the proline-28-truncated helix #1 of NeuroD (dashed line, right), respectively.

Figure II-7

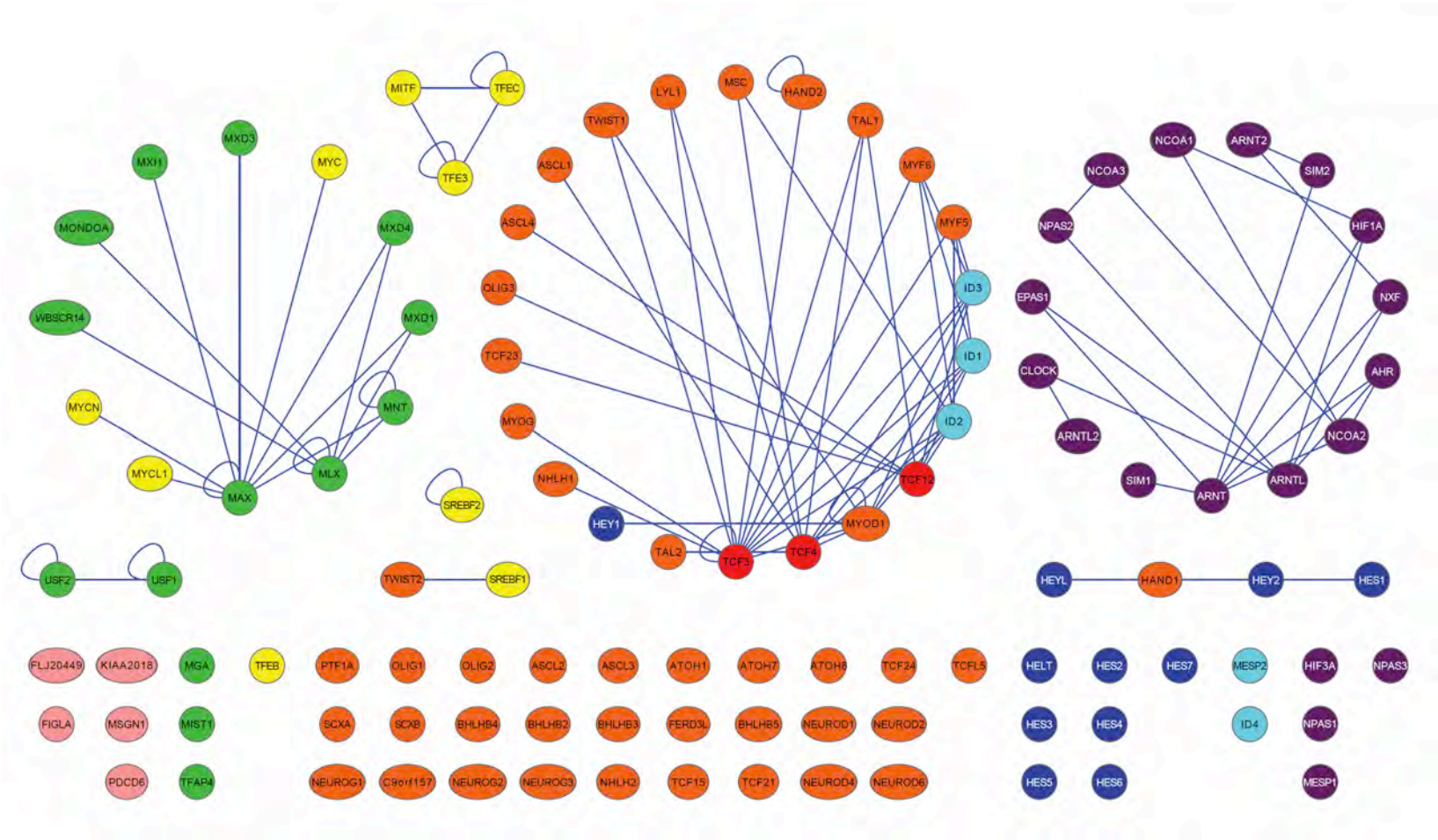

MITF
TFEC
TFE3
MXD3
MXI1
MYC
MONDOA
MXD4
WBSCR14
MXD1
MYCN
MNT
MYCL1
MLX
MAX
USF2
USF1
LYL1
MSC
HAND2
TWIST1
TAL1
ASCL1
MYF6
ASCL4
MYF5
OLIG3
ID3
TCF23
ID1
MYOG
ID2
NHLH1
TCF12
SREBF2
HEY1
MYOD1
TAL2
TCF3
TCF4
TWIST2
SREBF1
NCOA1
ARNT2
NCOA3
SIM2
NPAS2
HIF1A
EPAS1
NXF
CLOCK
AHR
ARNTL2
NCOA2
SIM1
ARNT
ARNTL
HEYL
HAND1
HEY2
HES1
FLJ20449
KIAA2018
MGA
TFEB
PTF1A
OLIG1
OLIG2
ASCL2
ASCL3
ATOH1
ATOH7
ATOH8
TCF24
TCFL5
HELT
HES2
HES7
MESP2
HIF3A
NPAS3
FIGLA
MSGN1
MIST1
SCXA
SCXB
BHLHB4
BHLHB2
BHLHB3
FERD3L
BHLHB5
NEUROD1
NEUROD2
HES3
HES4
ID4
NPAS1
PDCD6
TFAP4
NEUROG1
C9orf157
NEUROG2
NEUROG3
NHLH2
TCF15
TCF21
NEUROD4
NEUROD6
HES5
HES6
MESP1

Figure II-7. The human bHLH dimerization network

Human bHLH TFs (identified by the DBD online database (100; 2)) and interactions between them (identified by the BioGRID interaction database (101)) have been visualized using the Cytoscape software (106-108). Colors of nodes indicate which class each bHLH TF likely belongs to: Red = Class I, Orange = Class II, Yellow = Class III, Green = Class IV, Light Blue = Class V, Dark Blue = Class VI, Purple = Class VII, and Pink = Unknown Class. Although this network is incomplete due to the lack of systematic experimental determination of dimerization specificities for the human bHLH TFs, we can see global similiarities with the *C. elegans* bHLH dimerization network. For example, dimerization interactions appear to be class specific, although perhaps less so than the worm. We also see multiple dimerization hubs such as TCF3, TCF4, TCF12 (homologs of HLH-2), MAX (homolog of MXL-1), ARNT, and ARNTL (homologs of AHA-1). Perhaps the expansion of hubs and interaction specificity are contributing factors to biological complexity.

PREFACE TO CHAPTER III

This chapter describes the spatiotemporal expression patterns of the *C. elegans* bHLH TFs, based on the analysis of transgenic strains of *C. elegans* carrying bHLH-promoter::GFP fusion constructs. In addition, we describe a two-fluorescent-protein reporter approach to more distinctly define the co-expression patterns of pairs of bHLH genes respectively encoding heterodimerizing bHLH TF proteins.

Much of this chapter has been published separately in:

Grove C. A., De Masi F., Barrasa M. I., Newburger D. E., Alkema M. J., Bulyk M. L., Walhout A. J. M. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell. 2009 Jul 23; 138(2): 314-327.

CHAPTER III

The Spatiotemporal Expression Patterns of the *C. elegans* bHLH Transcription Factors

Abstract

As described in chapter I of this thesis, when and where a TF is expressed in the body of a multicellular organism can play a critical role in the ultimate functionality of that TF. It appears that certain tissues and cell types can actually potentiate the function of a TF by providing a cellular/molecular environment that is suitable for its activity. The degree to which a TF determines the identity of a cell type and, conversely, the degree to which a cell type determines the functionality of a TF remain mysteries of TF functionality. The functionality of dimerizing TFs, such as the bHLH TFs, is additionally influenced by spatiotemporal expression since heterodimerizing TFs require co-expression at the same time and place in order to physically interact and carry out their cooperative function. In this chapter, we describe the spatiotemporal expression patterns for the *C. elegans* bHLH TFs by way of transgenic *C. elegans* expressing green fluorescent protein (GFP) under the control of bHLH gene promoters. The resulting expression patterns for individual bHLH genes are described, as well as the analysis of co-expression patterns for HLH-2 and its partners.

Introduction

With the rise of multicellularity came a dramatic expansion of the bHLH family of TFs (109) which alone seems to suggest that these TFs play an important role in specifying and coordinating the functions of multiple cell types in higher eukaryotes. It has been long known that bHLH TFs are expressed in a variety of tissues and cell types in multicellular organisms and are often critical to the proper development and/or functioning of those tissues and cell types. These tissues include, but are not limited to, the nervous system (e.g. the achaete-scute complex, Atonal, Neurogenin, BETA2/Neurod) immune system (e.g. E2A, ABF-1), blood (Lyl-1, Tal-1), muscle (e.g. MyoD, myogenin), and pancreas (BETA2/NeuroD) (reviewed in (71)).

Because most studies of bHLH proteins suggest that they must form dimers in order to function, when and where heterodimeric bHLH protein partners are expressed are important factors in the functionality of these proteins. Regulation of bHLH protein expression could in fact be a mode of regulating the function of these dimers, since the lack of the appropriate partner essentially renders that bHLH protein non-functional. Co-expression of bHLH dimer partners is an obvious prerequisite to dimer formation and is, therefore, an important determinant of bHLH TF function. Homodimeric bHLH TFs, of course, escape this limitation since, as long as more than one copy of a homodimeric bHLH protein is present in any cell, co-expression of partners is possible.

Although tissue-specific expression patterns for various bHLH genes have been reported, there has been no systematic or comprehensive analysis of spatiotemporal bHLH TF expression in any whole multicellular organism. The work presented in this chapter describes the analysis of bHLH expression using a consistent methodology (i.e. the observation of GFP expression in transgenic worms carrying bHLH-promoter::GFP fusion constructs) and a controlled vocabulary to enable systematic annotations and comparisons of bHLH expression patterns.

Results

Worm PCR Verification

To analyze the spatiotemporal expression pattern of bHLH genes, we generated transgenic animals that express the green fluorescent protein (GFP) under the control of bHLH gene promoters (see Table III-1 and III-2 for cloning and transgenic line details; see Materials and Methods). To confirm that each of the transgenic strains carrying bHLH-gene promoter::GFP fusions is indeed carrying the promoter in question, we carried out PCR on individual worms using primers specific to each promoter. The resulting PCR reactions were separated and visualized by gel electrophoresis (Figure III-1; see Table III-3 for promoter-specific primer sequences; see Materials and Methods).

Temporal Expression Summary

GFP expression was observed throughout the lifespan of the worm, from early embryogenesis (about 24 cell stage) to aged adults, with different bHLH gene promoters driving expression for various windows of time (Table III-4). There was no detectable GFP expression in the germline (aside from the spermatheca), which may be a result of transgene silencing of transgenic arrays. Therefore, our expression analysis was limited to post-zygotic expression patterns. For a couple of bHLH genes (*cnd-1* and *ngn-1*) we observed expression as early as the 24-48 cell stage, but the majority of bHLH promoters began driving GFP expression in the mid-stage embryo (about the comma stage of

embryogenesis). This observation is in agreement with the fact that bHLH genes are known to be important regulators of development.

Spatial Expression Summary

bHLH gene promoters drove GFP expression in a wide variety of cell types, but were clearly more predominant in some tissues (Table III-4). We observed GFP expression in the hypodermis, body muscles, head muscles, neurons, intestine, excretory cell, pharynx, coelomocytes, spermatheca, vulva, distal tip cells, pharyngeal glands, pharyngeal-intestinal valve, P cells, seam cells, and rectum. As mentioned above, we did not observe GFP expression in the developing oocytes (germline), which may or may not reflect the actual expression of bHLH proteins. There appear to be some tissues that only express specific bHLH proteins or classes of bHLH proteins. For example, only REF-1 is expressed in the pharyngeal intestinal valve. HLH-6 is the only bHLH protein that appears to be expressed in the pharyngeal glands (HLH-2 appears to be expressed embryonically in cells that eventually become the pharyngeal glands). Other class-specific patterns are described below.

Although many bHLH genes appear to be expressed in *C. elegans* neurons, there may only be a minority of the 302 neurons in the worm that actually express bHLH genes. This is because of the difficulty of determining the exact neurons that express the fluorescent protein in question. A description of some higher resolution analyses of bHLH expression is described below.

Class III Expression Patterns

The Class III bHLH TFs in *C. elegans* are represented by the genes *hlh-30* (the *C. elegans* ortholog of MITF/TFE3) and *sbp-1* (the *C. elegans* ortholog of the Sterol Response Element Binding Protein, SREBP). Interestingly, both of these genes are expressed strongly throughout the intestine from the early larval stages through adulthood. This observation is in concordance with the fact that the intestine is the major site of fat storage and metabolism in *C. elegans*, as SREBP is a known regulator of cholesterol and lipid metabolism (reviewed in (110)) and TFE3 is a regulator of metabolic genes in mice (111). Whereas the *sbp-1* gene is expressed exclusively in the intestine, *hlh-30* is also expressed in a number of other tissues, including the spermatheca, head muscle, body muscle, the pharynx, the head and tail hypodermis, head neurons and the vulva.

Class IV Expression Patterns

The Class IV bHLH genes are represented by *mxl-1*, *mxl-2*, *mxl-3*, *mdl-1*, and *mml-1*. The promoter of the *mxl-1* gene did not drive detectable GFP expression in the worm (see discussion). The remaining four genes are all expressed in the hypodermis, in the head, body, and/or tail. Based on the fact that few other bHLH TFs in *C. elegans* appear to be expressed in the hypodermis, it seems likely that these Class IV bHLH TFs play an important role in the biology of the hypodermis. Although two-color fluorescence co-expression analysis (as described below) was not performed for the *mml-1* and *mxl-2* genes, we can deduce the co-expression of these two genes based on their individual

expression patterns. The MML-1/MXL-2 heterodimer, therefore, appears to be expressed exclusively in the head, body, and tail hypodermis (Table III-5). Aside from the tail hypodermis, MXL-3 homodimers are expressed in multiple head neurons. Although we cannot confirm the expression pattern of the MDL-1/MXL-1 dimer (because of lacking information regarding the *mxl-1* expression pattern), the promoter of *mdl-1* drives expression in body and tail neurons, the pharynx, coelomocytes, the vulva, rectum, and head, body, and tail hypodermis. Once the correct expression pattern of the *mxl-1* gene is deduced, we can make conclusions as to where the respective heterodimer is expressed.

REF-1 Family Expression Patterns

The REF-1 family bHLH TFs, as described earlier, are a unique family of bHLH proteins in that they each contain two recognizable bHLH domains (112). The implications of this are not currently understood, and this phenomenon (two bHLH domains per protein) is unique to *C. elegans* and rice (109; 112). The promoters of the *hlh-25*, *hlh-26*, and *hlh-28* genes did not drive detectable GFP expression in the worm (see discussion). Therefore, expression could only be analyzed for *hlh-27*, *hlh-29*, and *ref-1*. The *hlh-28* and *hlh-29* genes are in opposite orientations and separated by a bidirectional promoter. Interestingly, the same DNA sequence that drives GFP expression in head and tail neurons (the *hlh-29* promoter) does not drive any detectable GFP expression when cloned in the opposite orientation upstream of GFP (the *hlh-28* promoter). This unexpected result demonstrates that either the promoter sequence orientation is important for

proper gene expression, the relevant *cis*-regulatory element placement is important, or both. The *hlh-27* promoter drives expression in a variety of cells including head, body, and tail neurons, the pharynx, coelomocytes, and the rectum. The ref-1 promoter drives expression in tail neurons, the pharynx, the vulva, and, unique to all other bHLH promoters, drives expression in the pharyngeal-intestinal valve.

Class VII (bHLH-PAS) Expression Patterns

As described above for the MML-1/MXL-2 co-expression, two-color fluorescence analysis was not performed on the Class VII bHLH-PAS gene promoters, yet we could reasonably assess co-expression patterns based on the individual promoter::GFP expression patterns for *aha-1* and AHA-1 dimerization partners (Table III-5). The AHA-1/CKY-1 dimer is distinctly and exclusively expressed in the pharynx from early larval stages into adulthood. The AHA-1/HLH-34 dimer also appears to have a restricted pattern of expression as the *hlh-34* promoter drives expression only in a few head neurons. The other three heterodimers exhibit much broader expression: AHA-1/AHR-1 is expressed in coelomocytes and the vulva, AHA-1/HIF-1 is expressed in the pharynx, vulva, and tail hypodermis, and AHA-1/HLH-33 is expressed in body muscle, the pharynx, coelomocytes, body and tail hypodermis, and the vulva. Except for AHA-1/CKY-1, it appears that all dimers of this class are expressed in a variety of neurons in the head and tail of the worm, although definitive identification of co-expression in these cells will require more detailed co-expression analysis. It has

been reported that the AHA-1/AHR-1 dimer is indeed expressed in the RMEL and RMER neurons and performs important developmental functions (55). Although determining co-expression of dimer partners is relatively easy in some tissues and cells (e.g. intestine, hypodermis, pharynx), the co-expression analyses based on GFP expression alone should be interpreted with caution.

Co-Expression of Heterodimerizing bHLH Proteins

If spatiotemporal expression plays an important role in functional TF divergence, one could expect that proteins that dimerize exhibit greater co-expression than proteins that do not dimerize. To test this, we annotated the spatiotemporal expression of the bHLH gene promoters using a controlled vocabulary and calculated the tissue overlap coefficient (TsOC) (113) between all bHLH-bHLH pairs. As expected, dimerization partners are more likely to be co-expressed than bHLH proteins that do not dimerize with each other (Figure III-2, Fisher's exact test $p < 0.001$).

Design of Co-expression Experiments

To more accurately annotate which cells express both members of a particular bHLH dimer (in the case of heterodimers), we took advantage of a two-color fluorescent protein approach, whereby expression of one bHLH gene can be indicated by expression of a red fluorescent protein, another can be indicated by expression of green fluorescent protein, and a merged fluorescent image can indicate when and where both proteins are co-expressed. We decided to apply this experimental approach to the HLH-2 module of bHLH dimers, as most of the

heterodimers in the network belong to this module. The fact that HLH-2 is a dimerization hub makes it a suitable candidate for tracking expression using a red fluorescent protein reporter, as all of the HLH-2 dimerization partners may be represented by the green fluorescent protein transgenic lines we have already generated. A number of technical considerations that came to our attention during these experiments are described here.

Drawbacks of Transcriptional mCherry Fusions

The red fluorescent protein reporter mCherry (short for “monomeric Cherry”) was originally designed to create a vibrant red fluorescent protein with optimal spectral properties for fluorescent microscopy as well as a protein that did not form aggregates in cells, as was known to happen with many fluorescent proteins, including the protein dsRed, from which mCherry was derived (114-116). Fluorescent protein aggregates can make it difficult to assess expression patterns as fluorescence images can have a very speckled appearance, making cellular/sub-cellular identification challenging. It is also possible that these aggregates could potentially lead to deleterious effects on the health of the cells or organism being studied, thereby artificially altering results. Unfortunately, the original mCherry construct did not work well in worms, as expression of mCherry appeared to be stochastic and unreliable. Therefore, a *C. elegans*-specific version of mCherry was generated, in which *C. elegans*-preferred codons were incorporated into the reading frame and *C. elegans*-specific introns were introduced (117). This modified version of mCherry has given much more reliable

and reproducible results for expression in worms. When we generated transgenic *C. elegans* expressing mCherry under the control of the *hlh-2* gene promoter, we observed a very speckled pattern of red fluorescence in transgenic animals, reminiscent of fluorescent protein aggregates. As an attempt to alleviate the speckled fluorescence problem, we constructed an mCherry::histone translational fusion construct which would express a translational, in-frame fusion protein of mCherry fused to the HIS-11 protein (the *C. elegans* H2B histone). This translational fusion construct indeed improved the appearance of the red fluorescence and created a non-speckled, nuclear localized red fluorescence in cells in which the *hlh-2* promoter was active. Thus, all further experiments made use of mCherry::HIS-11 translational fusions for *hlh-2* co-expression analysis.

Drawbacks of Analyzing Co-expression under Hypoxic Conditions

Another technical consideration that came to our attention was the photoactivation of GFP to a red fluorescent protein under hypoxic conditions (i.e. under a cover slip sealed to the microscope slide). We began to observe what appeared to be red fluorescence in *exactly* the same location as the green fluorescence, even in the absence of any red fluorescent protein expression, but only after visualizing the green fluorescence for at least a few seconds. Before observing green fluorescence, no red fluorescence could be observed in the absence of mCherry expression. We found that this had been observed previously by a number of research groups (118; 119) and it is caused by a photo-activatable red shift of green fluorescent protein to a more red fluorescent

protein under hypoxic (<2% oxygen) conditions. This red shift was also apparently more frequent in circumstances in which there was intense GFP fluorescence in a particular tissue or group of cells. We found that this shift could be alleviated (if not entirely avoided) when we mounted anesthetized worms on a thicker (1-2mm) agarose pad with an unsealed cover slip. Presumably, these conditions allowed for more oxygen penetration into the sample (as opposed to the previous methodology using a very thin pad and a sealed cover slip) thereby eliminating the red shift of GFP.

Drawbacks of Analyzing Co-expression where GFP Expression is Intense

One other consideration that came to our attention was that of fluorescence “bleed through”, a common concern for fluorescence microscopy. When the GFP fluorescence is particularly intense/bright, those tissues will appear to be simultaneously red when viewed through a Rhodamine filter, as the intense green fluorescence signal can “spill over” into the red spectrum. This phenomenon was independent of oxygenation, presence of red fluorescent protein, or whether or not the green fluorescence had been observed first. One solution to this problem was to reduce the intensity of GFP excitation so as to reduce the bleed through into the red portion of the visible spectrum. This was sometimes insufficient, however, and made it impossible to confidently determine co-expression of green and red fluorescent protein expression at the same time and place.

Class I/II (HLH-2 Module) Co-Expression

To assess the spatiotemporal co-expression of *hlh-2* and its dimerization partners, worms carrying an integrated *Phlh-2*::mCherry::*his-11* fusion construct (as described above and in the materials and methods) were genetically crossed with worms carrying *Phlh*::GFP reporters representing each of the HLH-2 dimerization partners. Thus, worms carrying both transgenes could be used to determine the cells in which HLH-2 and its partners are co-expressed, and therefore where the proteins likely "meet" *in vivo*.

The *hlh-2* promoter exhibits broad activity in the embryo, and its activity becomes more restricted in larvae and adults, consistent with previous HLH-2 immunofluorescence data (Figure III-3, Figure III-4, Table III-4) (82). HLH-2 and most of its partners are first expressed at the comma stage of embryogenesis (Figure III-3), which is associated with the onset of cellular differentiation. This is in agreement with observations that orthologs of HLH-2 partners are important regulators of cell lineage commitment and differentiation (71). However, there is some temporal specificity as some HLH-2 dimers are expressed only during embryogenesis and in the first larval stage (*e.g.* HLH-2/HLH-3) whereas others are expressed throughout the lifetime of the animal (*e.g.* HLH-2/HLH-8). As has been observed for other organisms, we found that the HLH-2 partners exhibit a more tissue-restricted expression pattern as compared to HLH-2 (71) (Table III-4). Post-hatching, most HLH-2 heterodimers are expressed only in a subset of

tissues, including neurons, the vulva, some hypodermal cells and distal tip cells (Figure III-4, Table III-5).

Expression Annotation at Cellular Resolution

To improve the overall spatiotemporal expression annotation resolution, we worked with Mark Alkema, a specialist in *C. elegans* neurobiology, to more precisely identify which neurons were expressing various bHLH genes. The results of this higher resolution analysis are shown in Figure III-5.

Neuronal Expression of ngn-1 and hlh-13

We were able to identify the likely neurons (or pairs of neurons) that the genes *hlh-4*, *hlh-10*, *hlh-15*, *ngn-1*, and *hlh-13* are expressed in. Detailed expression annotation for *hlh-4*, *hlh-10*, and *hlh-15* are provided in Chapter V as part of the discussion on the integrated network and so will not be discussed further here. The *ngn-1* gene appears to be expressed in a single neuron, which we believe to be the RID dorsal motor neuron (Figure III-5). This may suggest involvement of *ngn-1* in *C. elegans* locomotion and/or behavior. Although the expression pattern of *hlh-13* cannot be definitively narrowed down to individual cells, we can make observations about which ganglia *hlh-13*-positive neuronal cell bodies are likely part of and, in some instances, what specific cells may be expressing *hlh-13*. This gene appears to be expressed in 2 neurons of the retrovesicular ganglion, a bilaterally symmetric pair of interneurons of the lateral ganglion, a bilaterally symmetric pair of interneurons posterior to the posterior

pharyngeal bulb (likely the ADA or RMG neurons), and a single tail neuron (likely ALN) (Figure III-5).

Distal Tip Cells: hnd-1, hlh-12, and lin-32

One cell type in *C. elegans* that is very distinctive and therefore relatively easy to identify is the distal tip cell (DTC). These cells lie at the tip of each gonad arm (the germline of hermaphroditic *C. elegans* is a symmetric organ with two arms bending to form the shape of a narrow letter C). Part of the somatic gonad, these cells have distinct morphologies and can be seen leading the migration of the gonad arms throughout development of the *C. elegans* germline (120). One interesting finding of bHLH expression patterns was that three dimerization partners of HLH-2 are expressed in the DTCs at different times during development. The *hnd-1* gene, the *C. elegans* homolog of the mammalian Hand-1 gene, is expressed in the earliest precursors of the distal tip cells at about the first larval stage of development. By the second larval stage of development, *hnd-1* expression dissipates, but expression of *hlh-12* begins in the DTCs and continues through to adulthood. By the third larval stage, the DTCs are also expressing *lin-32*. The *hlh-2* gene is present in the DTCs at all of these stages and can clearly be seen co-expressed with these three genes.

These results indicate that there may be a sequential role of the HLH-2/HND-1, HLH-2/HLH-12, and HLH-2/LIN-32 heterodimers in the DTCs throughout the development of the worm. Consistent with this hypothesis, mutations in and/or RNAi against *hlh-2*, *hlh-12*, or *hnd-1* all lead to DTC

migration defects and somatic gonad developmental defects (121; 85; 122). It will be interesting to see if HLH-2/HND-1 heterodimers “pave the way” for the HLH-2/HLH-12 heterodimers which subsequently “pave the way” for HLH-2/LIN-32 heterodimers by establishing a gene expression profile that is ultimately receptive to these later-forming dimers.

Discussion

Spatiotemporal Expression Analysis: Data Quality

A transcriptional reporter approach such as the one we used often recapitulates endogenous expression patterns (113; 123; 124). Indeed, our findings recapitulate broad and early expression for HLH-2 and more restricted expression for its partners (85; 122; 124; 125; 65; 126; 80; 127). However, parts of our expression analysis are relatively crude as some tissues could be annotated with greater resolution than others. For instance, we could confidently ascribe expression in several individual cells such as distal tip cells, and coelomocytes (Figure III-4, III-5). In contrast, we grouped expression in head neurons as a single category, even though there are approximately 200 functionally distinct head neurons. We did annotate several neuronal patterns in detail for higher resolution network analysis (Figure III-5, Figure V-11). In the future it will be important to further annotate co-expression of each bHLH-bHLH dimer at single-cell resolution, such as by crossing our GFP lines into specific neuronal marker lines. In this regard, it is important to note that all strains are available through the *Caenorhabditis Genetics Center (CGC).*

Non-overlapping Expression Patterns

An interesting question that remains after our dimerization and spatiotemporal analyses is whether or not individual bHLH proteins which appear to only be active as heterodimers have any function when they are expressed in cells that lack their dimerization partners. If these proteins are truly only

functional when dimerized to a partner bHLH protein, then it would be assumed that these proteins are non-functional when expressed alone. The observation of some bHLH proteins being expressed (apparently) alone raises some interesting questions: (1) Are these proteins non-functional in these particular cells, and, if so, does this represent an evolutionary artifact and perhaps an example of spurious expression? (2) Does this represent a false positive expression annotation for the bHLH protein of interest, or, perhaps, a false negative expression annotation for the partner bHLH protein(s)? (3) Are we missing genuine dimerization partners from the yeast two-hybrid data (i.e. false negatives in the yeast two-hybrid system)? Future analyses of cell-specific bHLH gene knockdown will be required to address the question of whether or not these bHLH TFs have a functional role in cells in which they appear to be expressed in the absence of a dimerization partner.

GFP Negative Transgenic Lines

Five bHLH gene promoters that were tested for GFP expression activity appeared to be inactive, according to analysis of GFP expression in transgenic worms. The promoters for genes *hlh-14*, *hlh-25*, *hlh-26*, *hlh-28*, and *mxl-1* did not drive any detectable levels of GFP expression at any life stages observed (Table III-4). There may be several reasons for this: (1) These promoters may not contain all cis-regulatory elements required for proper expression of these genes (e.g. other elements may exist in more distal upstream portions of the chromosome or in introns), (2) these genes may only be active under specific

environmental conditions (i.e. temperature, osmolarity, dauers, males, etc.) which were not tested, (3) expression of GFP may have been too weak to observe under the microscope, or (4) the genes may have been incorrectly annotated and therefore we have not cloned the correct promoter region into our GFP constructs. The fact that we can successfully clone the open reading frames for all five of these bHLH genes from a cDNA library suggests that these sequences are indeed transcribed (all ORF clones were verified by sequencing) and so lack of GFP expression is likely not the result of these genes being pseudogenes and/or simply unexpressed.

Partial Intestine, Coelomocytes, & Head Muscle GFP Expression: Possible Artifacts?

An observation that we have made while performing promoter::GFP transgenic analyses in *C. elegans* is that some tissues seem to sporadically and inconsistently express GFP in a surprisingly large number of transgenic lines representing many different promoter fragments. For instance, the intestine of *C. elegans* appears to be a common site of such potentially spurious GFP expression. Gradients of GFP can sometimes appear in the anterior and/or posterior of the intestine, with the highest levels at the most extreme anterior or posterior cells. This “partial intestine” GFP expression does not present itself in all GFP transgenic lines, and it is inconsistent in those lines that do exhibit this pattern. Another cell type that appears to express GFP somewhat inconsistently in some transgenic lines is the coelomocyte. There are 6 distinct coelomocytes

that generally occupy certain regions of the worm, acting as macrophage-like cells, endocytosing small particles (128; 129). Given the nature of this cell type it is not difficult to imagine that GFP observed inside coelomocytes may be the result of GFP-uptake during the scavenging process, as opposed to actual endogenous GFP expression inside of the coelomocyte cell. This is indeed known to be the case in some instances, although coelomocytes do exhibit their own endogenous gene expression profiles (130-132). The third cell type in *C. elegans* that appears frequently in these GFP transgenic lines is head muscle. There are 16 head muscles in *C. elegans* that are responsible for moving the head of the worm during locomotion, olfactory perception, and foraging (133). These cells appear to express GFP in a surprisingly large number of transgenic lines, and often do so very weakly, such that GFP can only be observed at very low fluorescence intensities.

Summary

We found that some *hlh* promoters are active broadly, whereas others drive GFP expression in a more restricted fashion (Table III-4). The promoters corresponding to both bHLH proteins that dimerize with multiple partners, AHA-1 and HLH-2, confer broad GFP expression, whereas their partners are generally expressed in a more restricted manner. Conversely, some tissues express few bHLH TFs, whereas other tissues express many. For instance, numerous *hlh* promoters drive expression in the vulva, but only the *ref-1* promoter is active in the pharyngeal-intestinal valve. Together, our observations identify specificity

and promiscuity in the spatiotemporal expression network, both from the bHLH and from the tissue standpoint (visualized in the integrated network in Figure V-1, Chapter V).

In summary, we observed broader, or “tissue-promiscuous”, activity for several bHLH promoters, including those that correspond to the bHLH proteins that interact with multiple partners, and we observed more restricted, or “tissue-restricted”, activity for others. Conversely, we observed that some tissues express many, whereas others express few, bHLH genes.

Materials and Methods

C. elegans *Transgenesis*

Transgenic *C. elegans* were generated as described (13). Briefly, uncoordinated *unc-119(ed3)* mutant worms were bombarded with 1µm gold beads coated with precipitated, linearized promoter::GFP (or mCherry) construct DNA, each molecule of which also carries a wild type copy of the *unc-119* locus. Worms that have taken up the DNA into the developing germline will produce offspring that have wild type movement, which allows for easy scoring and picking of transgenic worms from bombarded plates. Transgenic worms are singled out, allowed to self fertilize, and scored for transmission of transgene to subsequent generations. Worms that reliably express GFP (or mCherry) and transmit the transgene at relatively high frequency are kept for experiments. Double transgenic animals were generated by crossing males that carry *Phlh-2*::mCherry::*his-11* constructs into *Phlh*::GFP carrying hermaphrodites. Each transgenic line carrying a *Phlh*::GFP fusion was independently verified by PCR using promoter-specific primers (primer sequences are provided in Table III-3).

Gateway Cloning of bHLH Promoters

Fourteen bHLH gene promoters were obtained as Gateway Entry clones from the *C. elegans* Promoterome (15) (Table III-1). Twenty eight promoters were PCR-amplified from *C. elegans* genomic DNA *ab initio* as described (16), and subsequently cloned by Gateway BP reactions into the Donor vector,

pDONR-P4-P1R as described (15) (see Table III-1 for primer sequences). All 42 bHLH gene promoters were cloned by Gateway LR reactions into pDEST-DD04, generating promoter::GFP transcriptional fusions. *Phlh-2* was cloned upstream of mCherry::*his-11* (see below) by a multisite LR reaction into pDEST-DD03 (15; 17). All constructs were sequence verified.

***Generation of pDEST-mCherry::*his-11**

The mCherry ORF was PCR-amplified from pAA64 plasmid DNA (generously provided by A. Audhya, Oegema Lab, University of California, San Diego) using the following att-B1 and att-B2 Gateway tailed primers: (GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGTCTCAAAGGGTGAAGAAG and GGGGACCACTTTGTACAAGAAAGCTGGGTTATACAATTCATCCATGCCAC) and the resulting amplicon was cloned by a Gateway BP reaction into pDONR-221 to generate an mCherry Entry clone. A PCR fusion strategy was implemented to create an mCherry::*his-11* fusion ORF. The *his-11* gene encodes for *C. elegans* histone H2B, generating a nuclear localized mCherry when translationally fused. The *his-11* ORF was amplified from pJH4.52 (generously provided by K. Hagstrom, University of Massachusetts Medical School, Worcester) using a *his11*-specific forward primer (CCACCAAAGCCATCTGCCAA) and an att-B2 Gateway-tailed (103) reverse primer specific to the 3' end of the *his11* ORF (GGGGACCACTTTGTACAAGAAAGCTGGGTACTTGCTGGAAGTGTACTTG).

PCR amplification was carried out for 15 cycles to minimize the introduction of mutations. A similar PCR reaction was used to amplify the mCherry ORF using the same att-B1 tailed forward primer mentioned above (to generate the mCherry Entry clone) and an mCherry-specific reverse primer carrying a *his-11*-specific tail at the 5' end of the primer (TTGGCAGATGGCTTTGGTGGCTTATACAATTCATCCATGCCAC). Both PCR products were simultaneously cloned by Gateway BP reactions into pDONR221. The resulting plasmid contained the mCherry ORF fused in frame to the *his-11* ORF, generating an mCherry::*his-11* fusion ORF. This fragment was then cloned by a Multisite Gateway LR reaction into pDEST-DD03 (134) along with *Phlh-2*. The resulting *Phlh-2*::mCherry::*his-11* Destination clone was used directly in microparticle bombardment to create transgenic *C. elegans* (see below), and in a subsequent Gateway BP reaction with pDONR-P4-P1R to replace *Phlh-2* with the *ccdB* Gateway cassette resulting in the novel Destination vector pDEST-mCherry::*his-11*.

Microscopy

Nomarski and fluorescence images were obtained using a Zeiss Axioscope 2+ microscope. Image capture was performed with OpenLab 3.1.7 (Improvision) and Axiovision (Zeiss) software. GFP fluorescent images were obtained using a FITC filter (excitation 460-500 nm, emission 510-565 nm). mCherry fluorescence images were taken using a rhodamine filter (excitation 525-555 nm, emission 575-630 nm). Animals were placed into a drop of 0.1%

sodium azide in S-Basal on a fresh 2% agarose pad for observation. We examined GFP expression in mixed-stage populations of hermaphrodites. None of the promoters drove detectable GFP expression in the germline, which may be the result of transgene silencing (113; 123; 124). Therefore, our expression data are limited to post-zygotic expression. We annotated several head neuronal patterns in more detail. These are provided in Figure III-5 and Figure V-11.

C. elegans *PCR*

Worm lysis solution was prepared by mixing 2 ml proteinase K (15 mg/ml) in 1 ml of worm lysis buffer (50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, 0.01% gelatin). Individual worms were placed in 3 ml of worm lysis solution in PCR tubes and heated in an MJ Research thermocycler at 60^{o}C for 60 minutes and 95^{o}C for 15 minutes. The lysate was then vortexed and briefly (30 seconds) centrifuged at 15,000 *g*. 35 cycles of PCR were performed using our “GFP-FW” primer (TTCTACTTCTTTTACTGAAGC) and promoter-specific reverse primers (see Table III-3).

Online Data

Additional expression pattern data available online at http://edgedb.umassmed.edu (135).

Acknowledgements

Several of the earliest transgenic worms carrying bHLH gene promoter::GFP fusion constructs were generated by John Reece-Hoyes and Jane Shingles at the University of Leeds, in Leeds, UK in the laboratory of Ian Hope. In addition, John Reece-Hoyes and Jane Shingles provided instruction on how to carry out microparticle bombardment so that we could generate the majority of the GFP lines used in these experiments.

Plasmids used for the generation of the *Phlh-2*::mCherry::*his-11* construct were provided by Denis Dupuy (pDest-DD03), Kirsten Hagstrom (pJH4.52, carrying the *C. elegans his-11* gene), and Anjon Audhya (pAA64, carrying the mCherry gene, modified to carry *C. elegans*-preferred codons and *C. elegans* introns, which enhance expression of the gene in worms).

We thank Mark Alkema for his assistance in annotating neuronal expression patterns for several bHLH TFs.

We thank Job Dekker, and members (current and past) of the Walhout and Dekker labs for critical reading of relevant manuscripts.

Table III-1. *C. elegans* bHLH promoter cloning information

ORF	*gene*	Clone Source	Cloning Primers FW	Cloning Primer RV
C25A1.11	*aha-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGATCCCCCATCATCCTAAAG	GGGGACTGCTTTTTTGTACAAACTTGTCATTCTGAAAACTATAAAGAAAAGTTTTTTTC
C41G7.5	*ahr-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGAAGTGTTTTTGAAGTTTTACCGTTTTTTT	GGGGACTGCTTTTTTGTACAAACTTGTCATTTCACAAATTCGAACGTCTGTTTAACCT
C15C8.2	*cky-1*	Promoterome		
C34E10.7	*cnd-1*	Promoterome		
F38A6.3	*hif-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTCAGAAGGGTATCCGTCA	GGGGACTGCTTTTTTGTACAAACTTGTCATATTGAATAGTGTGCGATTTGGAGA
B0304.1	*hlh-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGATTTCAGGAAAATTTTTTCAAAACTGTAAAAC	GGGGACTGCTTTTTTGTACAAACTTGTCATTTCTGGAAAATTATTGGAAAATTTGG
ZK682.4	*hlh-10*	Promoterome		
F58A4.7	*hlh-11*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCAGAAAATGTTTCTTGGATCGGTT	GGGGACTGCTTTTTTGTACAAACTTGTCATTTTCTACTATTGATCTACCTGA
C28C12.8	*hlh-12*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCTGAAAAATGGGAGTGTATTGCTTCTA	GGGGACTGCTTTTTTGTACAAACTTGTCATTTTAATAAAATTGTGTAAGATGACGCTA
F48D6.3	*hlh-13*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGGTGGAGATGCGACCCGGC	GGGGACTGCTTTTTTGTACAAACTTGTCATTTACATAAGTGGTTCTGCTC
C18A3.8	*hlh-14*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGAATAATCAATAGAAAAGTTTGTCAAT	GGGGACTGCTTTTTTGTACAAACTTGTCATCCCCCCATTTTTAAGATTTCCA
C43H6.8	*hlh-15*	Promoterome		
DY3.3	*hlh-16*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCTGGAACATCAGAAATTTGAGAC	GGGGACTGCTTTTTTGTACAAACTTGTCATAATCGGAGATGAAGCTGT
F38C2.2	*hlh-17*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGGTTTATACATCAGTATAGCAAGATGAA	GGGGACTGCTTTTTTGTACAAACTTGTCATGACTGGGGTGTAAGTGAA
F57C12.3	*hlh-19*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTGGGTAGAGAAAGGCGCAGA	GGGGACTGCTTTTTTGTACAAACTTGTCATGTTGAATGCGTTGCCGA

M05B5.5	*hlh-2*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGAATCAGTTAGCTTTTTGAGAATAATTTTTTGT	GGGGACTGCTTTTTTGTACAAACTTGTCATTGAGGTTCTGAAAACTTTATTTCTGCGA
C17C3.7	*hlh-25*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTATGAGTCAAGGATCTGAAATATAATAATAA	GGGGACTGCTTTTTTGTACAAACTTGTCATTAACCATTTTTATCACTCATTTAATC
C17C3.8	*hlh-26*	Promoterome		
C17C3.10	*hlh-27*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGGACTAACACTGCGAGACG	GGGGACTGCTTTTTTGTACAAACTTGTCATTTTGTAGCAGATGTGAAGGTAAGA
F31A3.2	*hlh-28*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCTGAAAAATTTTAAAGTAATCAAAACATACGA	GGGGACTGCTTTTTTGTACAAACTTGTCATGGAGAATGTGGAGGT
F31A3.4	*hlh-29*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGGGAGAATGTGGAGGTGTA	GGGGACTGCTTTTTTGTACAAACTTGTCATCTGAAAAATTTTAAAGTAATCAA
T24B8.6	*hlh-3*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCAAGTTAAGTTTATTTGTCTGCTGT	GGGGACTGCTTTTTTGTACAAACTTGTCATGTTTCTATAACTTTCCTTGGAT
W02C12.3	*hlh-30*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGGTGTCTAAACTTTCTGATCGGGACCT	GGGGACTGCTTTTTTGTACAAACTTGTCATGAATGCTCTTATTCGCTG
F38C2.8	*hlh-31*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGACATTGAAAAGTTTTCAATTTTTTCAATC	GGGGACTGCTTTTTTGTACAAACTTGTCATCTTTATTTATTTAATTTGTAGATAAT
Y105C5B.29	*hlh-32*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTTTTTTGGCGAAAGTTAAAATACATATTTC	GGGGACTGCTTTTTTGTACAAACTTGTCATTTTCTCTCCAAATCTAACGACA
Y39A3CR.6	*hlh-33*	Promoterome		
T01D3.2	*hlh-34*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTTTACCTTCTTCTCCATGCCAA	GGGGACTGCTTTTTTGTACAAACTTGTCATTTCTCAAGTGGTTATAAGTCAA
T05G5.2	*hlh-4*	Promoterome		
T15H9.3	*hlh-6*	Promoterome		
C02B8.4	*hlh-8*	Promoterome		
C44C10.8	*hnd-1*	Promoterome		
Y54G2A.1	*lin-22*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTTTTATTAGTCCTAAGAACTTTTT	GGGGACTGCTTTTTTGTACAAACTTGTCATTTCTGTAATAATTGTAATAATATTAGTA
T14F9.5	*lin-32*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCCTAATCGGAACGGTGTCT	GGGGACTGCTTTTTTGTACAAACTTGTCATGGTTGGTCTGACTGAA

R03E9.1	*mdl-1*	Promoterome		
T20B12.6	*mml-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTCATTCATAAACGAGAATAATGTCAC	GGGGACTGCTTTTTTGTACAAACTTGTCATTTTTGTCTGAAAATATTGAAATCTTTTG
T19B10.11	*mxl-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGTCTTTGTGGCTGCGGAAATTATGT	GGGGACTGCTTTTTTGTACAAACTTGTCATTCTGAAATAGCTTCAGTGAGAT
F40G9.11	*mxl-2*	Promoterome		
F46G10.6	*mxl-3*	Promoterome		
Y69A2AR.29	*ngn-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGAGAAAAAGAAGTGGTGCATATTTTGT	GGGGACTGCTTTTTTGTACAAACTTGTCATTGTGCAAAACAAAAACACGTGG
T01E8.2	*ref-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCCACAGTGTCAAAATATAATACCGAAAA	GGGGACTGCTTTTTTGTACAAACTTGTCATTTCTGGAAAAAAAATTAAGTT
Y47D3B.7	*sbp-1*	cloned *ab initio*	GGGGACAACTTTGTATAGAAAAGTTGCCAGGAGTTTTTGAAAAAATTCAAAATTCAAT	GGGGACTGCTTTTTTGTACAAACTTGTCATTCTGAAAAAAAAAAGTCAAATTTTGAG
Y16B4A.1	*unc-3*	Promoterome		

Table III-1. *C. elegans* bHLH promoter cloning information
Listed are the ORF names and gene names for each *C. elegans* bHLH TF in this study. Also listed are the clone source (e.g. Promoterome) and cloning primers (where applicable). Red text in primer sequences indicates promoter-specific sequence; black text indicates Gateway tails.

Table III-2. Transgenic *C. elegans* lines

bHLH promoter	# of Lines	# of Lines with an Integrated Transgene
aha-1	9	0
ahr-1	8	0
cky-1	4	0
hif-1	8	1
hlh-33	3	0
hlh-34	9	0
hlh-2	8	0
cnd-1	3	0
hlh-3	7	0
hlh-4	8	0
hlh-6	3	1
hlh-8	8	0
hlh-10	5	0
hlh-12	8	0
hlh-13	3	0
hlh-14	3	0
hlh-15	3	0
hlh-19	3	0
hnd-1	4	0
lin-32	8	0
ngn-1	3	0
hlh-1	1	0
hlh-11	3	0
hlh-30	3	0
mdl-1	3	1
mxl-1	4	0
mml-1	9	0
mxl-2	8	0
mxl-3	4	0
hlh-25	6	0
hlh-26	8	0
hlh-27	2	0
hlh-28	7	0
hlh-29	7	0
ref-1	5	0
hlh-16	3	0
hlh-17	7	0
hlh-31	3	0
hlh-32	3	0
lin-22	3	0
sbp-1	2	0
unc-3	3	0

Table III-3. Promoter-specific diagnostic primers

ORF	*gene*	Promoter Specific Reverse Primer
C25A1.11	*aha-1*	GTTGTGCAATGGATCGAGAGTA
C41G7.5	*ahr-1*	TTGTATCGTTCGGCTTATGCTA
C15C8.2	*cky-1*	CATTGTTTGCTCGATCACTCTC
C34E10.7	*cnd-1*	AGATAAGTGCACGGGATGTTCT
F38A6.3	*hif-1*	ACCTTGCACTCACAAGTTCTCA
B0304.1	*hlh-1*	AAAAGAGTTGAGCCGAGAGTTG
ZK682.4	*hlh-10*	ACAACGATGTCACGTCAATAGG
F58A4.7	*hlh-11*	TTTCTCAATTTTTGGAGGCAAT
C28C12.8	*hlh-12*	ATGACGAGTTGTGAGCCAAGTA
F48D6.3	*hlh-13*	TAGCTTTCTGACAACACCCAAC
C18A3.8	*hlh-14*	GATGACTTACCCGAAAAATGGA
C43H6.8	*hlh-15*	AAATTGCCTGTAAAAGGTTTGG
DY3.3	*hlh-16*	AAGTGACACGTTTGGTTCACAC
F38C2.2	*hlh-17*	ATTGGAAAAATTTTGAGCAAGC
F57C12.3	*hlh-19*	AAACGAATGGAAAGCATATTGG
M05B5.5	*hlh-2*	ATACTTCCAATGCCCGTCTCTA
C17C3.7	*hlh-25*	CGAGCTCCGATTTTACCAAATA
C17C3.8	*hlh-26*	TTTACGAGGTTTCCATCTCAGC
C17C3.10	*hlh-27*	TCTCGAGTTTGACGTCAGGTAA
F31A3.2	*hlh-28*	TGTTTGGAGTAGAATGCCAAGA
F31A3.4	*hlh-29*	AAACTTATGGAAACAGTGGCAAA
T24B8.6	*hlh-3*	CTTAGACAGTTCCGATGGGTTC
W02C12.3	*hlh-30*	CAGAGAATTGTTACTCGGCAAA
F38C2.8	*hlh-31*	TAGCAAAATTTTCCGCTCAAAC
Y105C5B.29	*hlh-32*	ATTGAGGAATGAAGGGAAATGA
Y39A3CR.6	*hlh-33*	TACGACAAAATGCACAGCTTCT
T01D3.2	*hlh-34*	CAATTGAATTCTCCCCTCTTTG
T05G5.2	*hlh-4*	CTGACGTTAGTGGGAGATCAGA
T15H9.3	*hlh-6*	TTCCTCGTGATTTGACAAAATG
C02B8.4	*hlh-8*	TGAAAGCATTTCACACGTTTTT
C44C10.8	*hnd-1*	AGCACAACTTTTTGCCTCAAAT
Y54G2A.1	*lin-22*	TTAACGGCTTTATTTTGTGGTG
T14F9.5	*lin-32*	GCGATAATTTTCATGTCAATGC
R03E9.1	*mdl-1*	ATTGAGCAGGGTTCAAGAAGAG
T20B12.6	*mml-1*	GTGTGTGTGTGTGTGAGCGTAT
T19B10.11	*mxl-1*	CAGTTAGGTCAACGGAAAACAGT
F40G9.11	*mxl-2*	ATTTTCTGAACGGGTTTCTCAA
F46G10.6	*mxl-3*	TGCTCTATCACGTAGGTCATGC
Y69A2AR.29	*ngn-1*	TAATTTATCTCGCGGAAATTCG
T01E8.2	*ref-1*	TTACCGCCAATTTATTTCTCGT
Y47D3B.7	*sbp-1*	TTTTGCATGAAACATTTGAACG
Y16B4A.1	*unc-3*	TGACGTCAGCACATTTTTAACC

Figure III-1

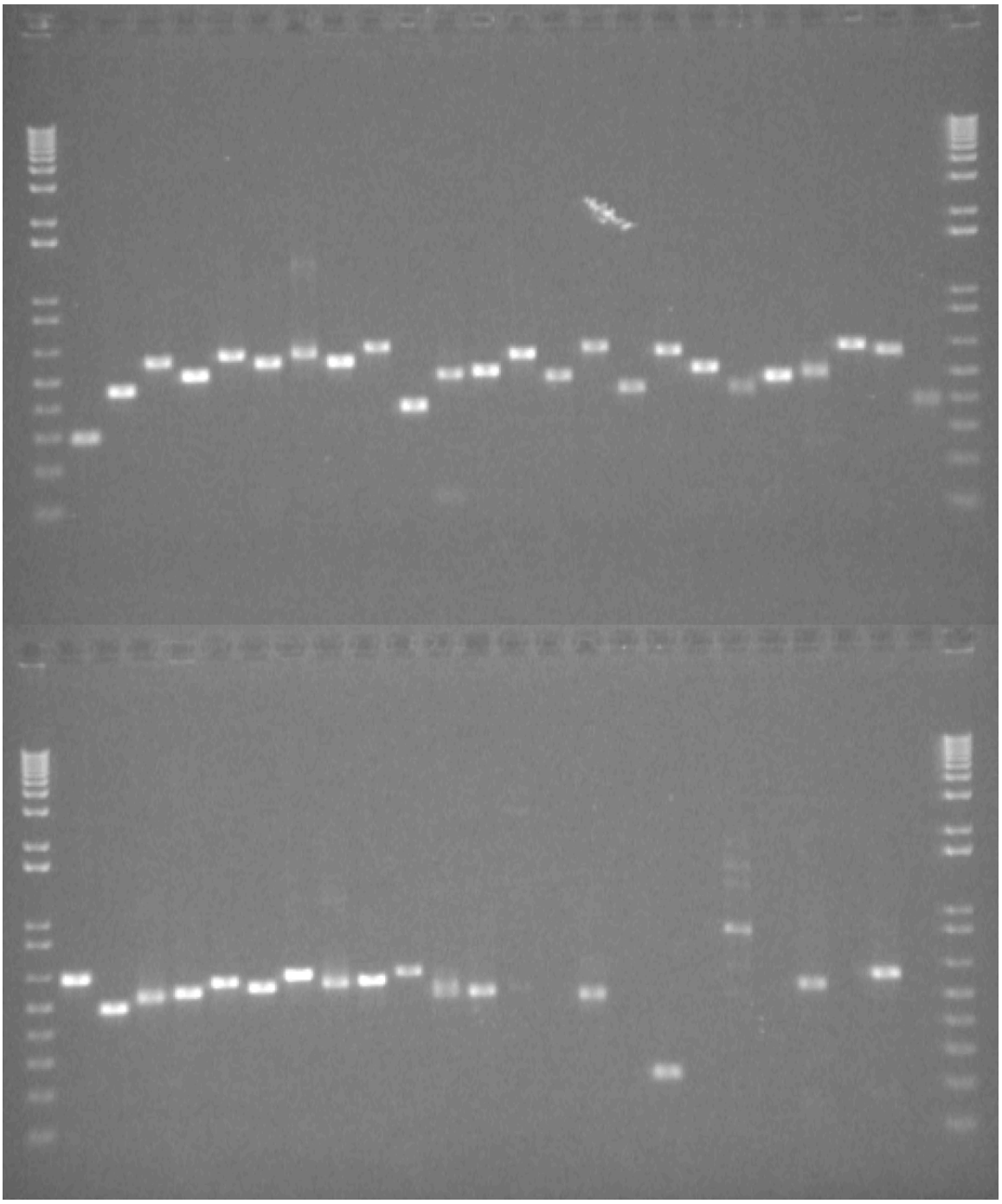


Figure III-1. Diagnostic PCR on individual worms carrying *Phlh*::GFP transgenes

Each *Phlh*::GFP transgenic line of *C. elegans* was tested to ensure that the appropriate promoter was upstream of GFP. This was done by performing single worm PCR on individual worms using a generic forward primer and a promoter-specific reverse primer. The identity of each promoter represented is listed to the right of the figure. (Top panel) PCR bands representing (from left to right) *Paha-1*, *Phif-1*, *Pahr-1*, *Phlh-1*, *Pcky-1*, *Phlh-3*, *Pcnd-1*, *Phlh-8*, *Phlh-31*, *Phlh-12*, *Phlh-2*, *Phlh-17*, *Phlh-4*, *Phlh-25*, *Phlh-6*, *Phlh-26*, *Phlh-10*, *Phlh-28*, *Phlh-11*, *Phlh-29*, *Phlh-13*, *Phnd-1*, *Phlh-14*, *Plin-22*. (Bottom panel) *Plin-32*, *Phlh-15*, *Pmml-1*, *Phlh-16*, *Pmxl-1*, *Phlh-19*, *Pmxl-2*, *Phlh-27*, *Pmxl-3*, *Pmdl-1*, *Pngn-1*, *Phlh-33*, *Pref-1*, *Psbp-1*, *Phlh-34*, *Punc-3*, *Phlh-30*, and *Phlh-32*. See Table III-3 for diagnostic PCR primers for each promoter. “1 Kb Plus DNA Ladder” from Invitrogen was used to identify PCR product sizes.

Table III-4. The bHLH expression matrix

		Temporal Expression								Spatial Expression																					
ORF Name	*Gene*	0	1	2	3	4	5	6	7	I	PI	DTC	GS	HM	BM	HN	BN	TN	P	PG	IV	HH	BH	TH	PC	SC	X	C	S	V	R
C25A1.11	*aha-1*	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	1	0	1	0	1	0
C41G7.5	*ahr-1*	0	0	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1
T01D3.2	*hlh-34*	0	1	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
C15C8.2	*cky-1*	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Y39A3CR.6	*hlh-33*	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	1	1	1	0	0	0	1	0	1	1
F38A6.3	*hif-1*	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	0	1	1
M05B5.5	*hlh-2*	0	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	0	1	0
T24B8.6	*hlh-3*	0	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1
T05G5.2	*hlh-4*	0	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ZK682.4	*hlh-10*	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0
C43H6.8	*hlh-15*	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
F57C12.3	*hlh-19*	0	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
T14F9.5	*lin-32*	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
C34E10.7	*cnd-1*	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0

		Temporal Expression								Spatial Expression																					
ORF Name	*Gene*	0	1	2	3	4	5	6	7	I	PI	DTC	GS	HM	BM	HN	BN	TN	P	PG	IV	HH	BH	TH	PC	SC	X	C	S	V	R
C02B8.4	*hlh-8*	0	0	1	1	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
T15H9.3	*hlh-6*	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
F48D6.3	*hlh-13*	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
C44C10.8	*hnd-1*	0	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Y69A2AR.29	*ngn-1*	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
C28C12.8	*hlh-12*	0	0	0	0	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
F58A4.7	*hlh-11*	0	0	0	1	1	1	1	1	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B0304.1	*hlh-1*	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R03E9.1	*mdl-1*	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	1	0	1	1
T20B12.6	*mml-1*	0	0	0	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
F40G9.11	*mxl-2*	0	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0	1	1	1	0	0	1	0	0	1	1
F46G10.6	*mxl-3*	0	0	0	1	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
W02C12.3	*hlh-30*	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	1	0
T01E8.2	*ref-1*	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0
C17C3.10	*hlh-27*	0	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	1
F31A3.4	*hlh-29*	0	0	0	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
DY3.3	*hlh-16*	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

		Temporal Expression								Spatial Expression																						
ORF Name	*Gene*	0	1	2	3	4	5	6	7	I	PI	DTC	GS	HM	BM	HN	BN	TN	P	PG	PIV	HH	BH	TH	PC	SC	X	C	S	V	R
F38C2.2	*hlh-17*	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	1
F38C2.8	*hlh-31*	0	0	0	1	1	1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Y105C5B.29	*hlh-32*	0	0	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Y54G2A.1	*lin-22*	0	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0
Y47D3B.7	*sbp-1*	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y16B4A.1	*unc-3*	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
C18A3.8	*hlh-14*	No GFP Expression																														
C17C3.7	*hlh-25*	No GFP Expression																														
C17C3.8	*hlh-26*	No GFP Expression																														
F31A3.2	*hlh-28*	No GFP Expression																														
T19B10.11	*mxl-1*	No GFP Expression																														

Table III-4. The bHLH expression matrix

The expression matrix makes use of a controlled vocabulary to systematically apply a binary code to each bHLH promoter to indicate observed expression (1) or lack of observed expression (0) for the indicated developmental time points (Temporal Expression) or tissues (Spatial Expression). The temporal expression code for the matrix is as follows: **0** = early embryogenesis, **1** = mid-embryogenesis, **2** = late embryogenesis, **3** = first larval stage (L1), **4** = second larval stage (L2), **5** = third larval stage (L3), **6** = fourth larval stage (L4), **7** = adult. The spatial expression code for the matrix is as follows: **I** = intestine, **PI** = partial intestine, **DTC** = distal tip cells, **GS** = gonadal sheath, **HM** = head muscle, **BM** = body wall muscle, **HN** = head neurons, **BN** = body neurons, **TN** = tail neurons, **P** = pharynx, **PG** = pharyngeal glands, **PIV** = pharyngeal-intestinal valve, **HH** = head hypodermis, **BH** = body hypodermis, **TH** = tail hypodermis, **PC** = P-cells, **SC** = seam cells**, X** = excretory cells, **C** = coelomocytes**, S** = spermatheca, **V** = vulva, **R** = rectum.

Figure III-2

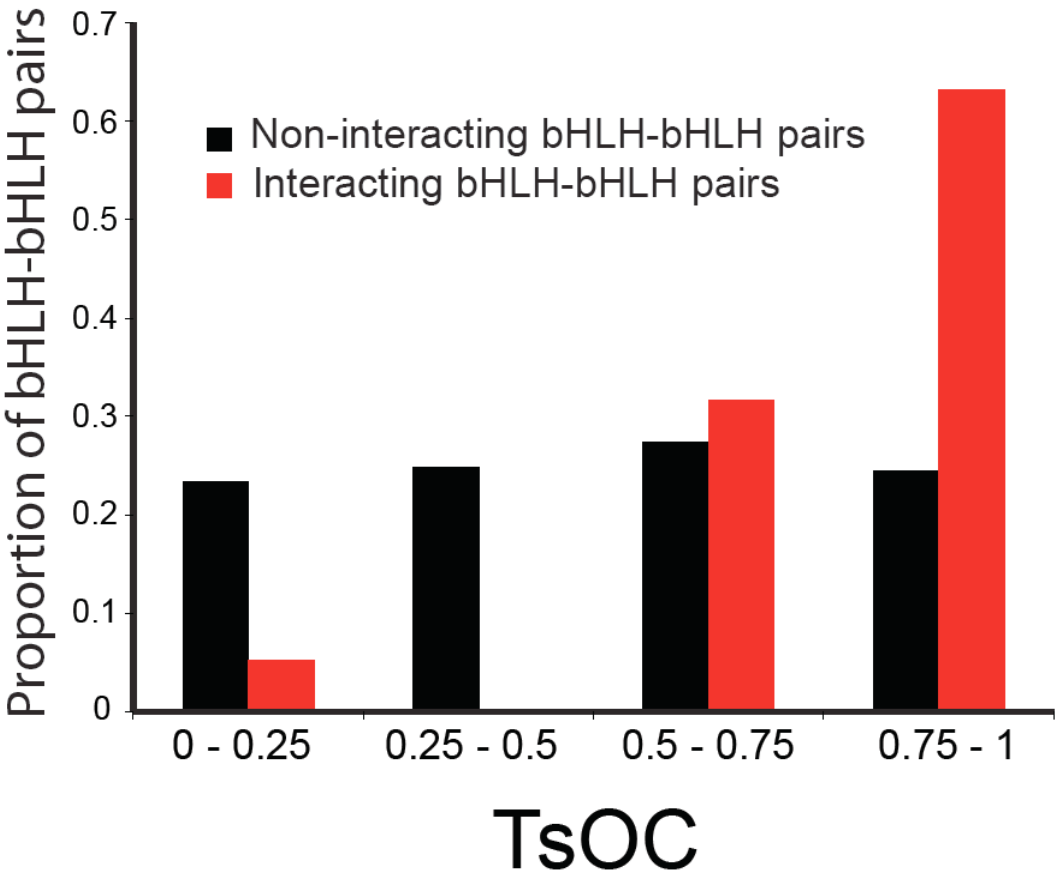


Figure III-2. Tissue overlap coefficient analysis
Tissue overlap coefficient (TsOC) analysis was done as described (113)

$$TsOC = \frac{(HLH - X) \bigcap (HLH - Y)}{(HLH - N)}$$

where HLH-X is the number of tissues where HLH-X is expressed, and HLH-Y is the number of tissues where HLH-Y is expressed. HLH-N is the smallest total number of tissues for either HLH-X or HLH-Y.

Figure III-3A

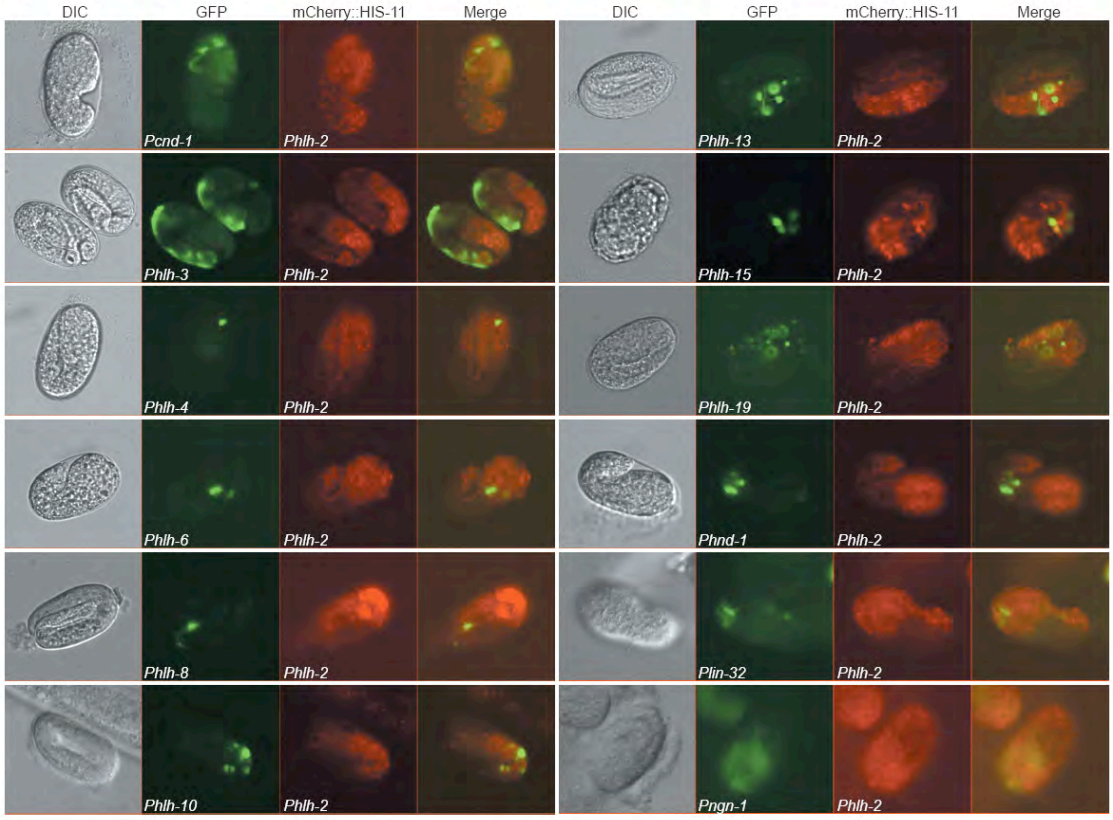


Figure III-3A. Embryonic co-expression of HLH-2 and its partners
Phlh-2::mCherry::*his-11* transgenic animals were crossed with each of the *Phlh-x*::GFP animals to determine co-expression The panels show DIC, GFP, mCherry and merged images depicting co-expression of HLH-2 and its indicated partners in the developing *C. elegans* embryo.

Figure III-3B

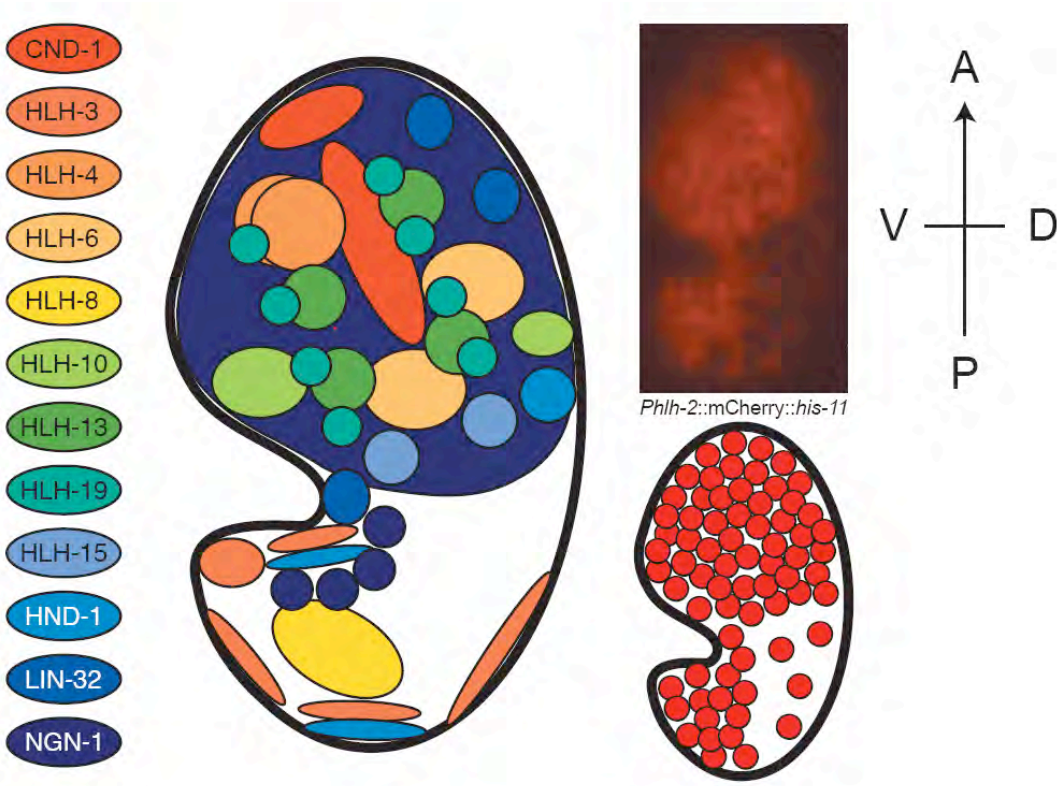


Figure III-3B. Cartoon summary of embryonic co-expression of HLH-2 and its partners

Cartoon summarizing the general distribution of mCherry::HIS-11 expression from the *hlh-2* promoter and GFP expression from the HLH-2 partner gene promoters. Note that colored regions do not depict specific cells, only general regions of the embryo where promoter activity was observed.

Figure III-4

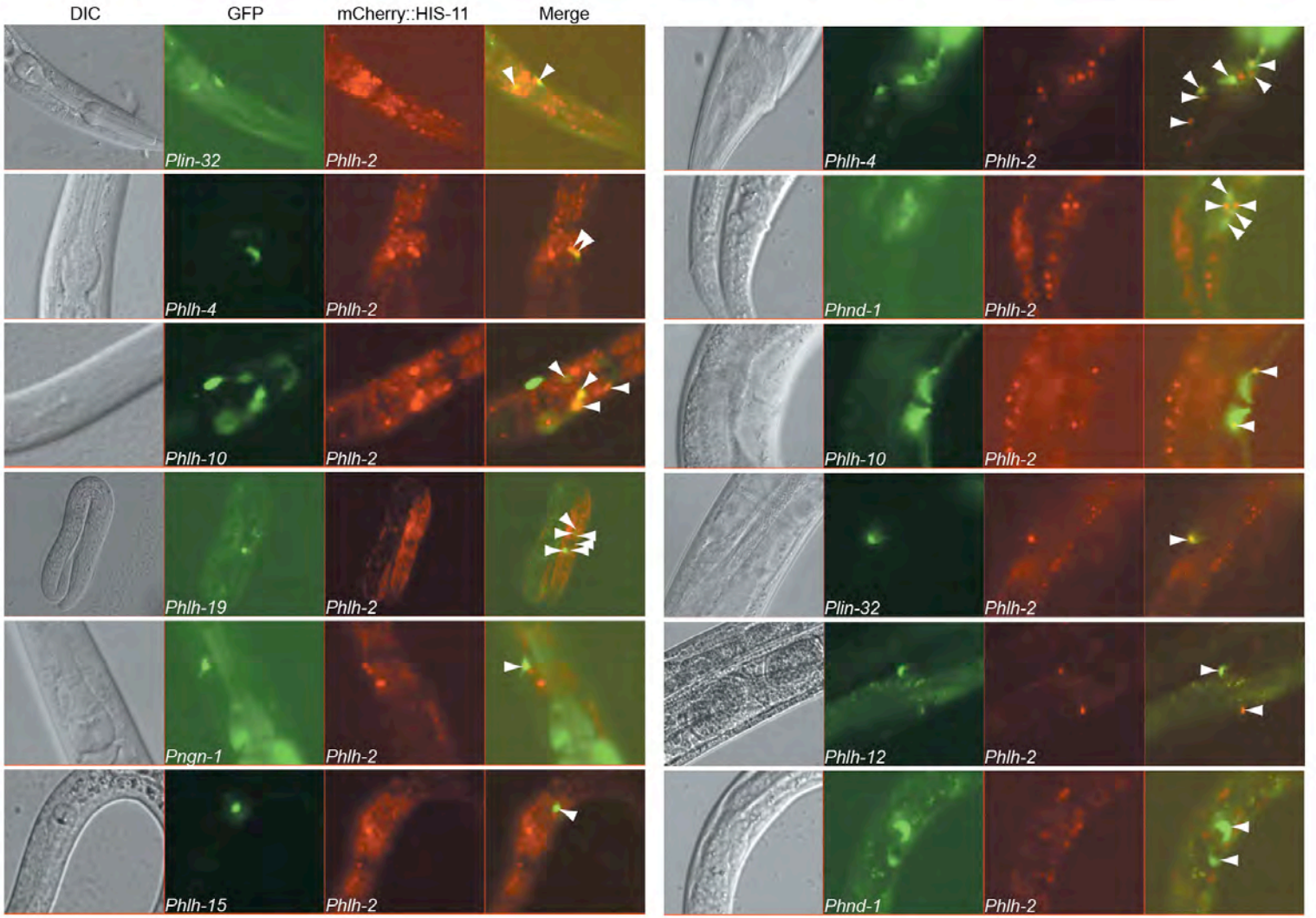

DIC
GFP
mCherry::HIS-11
Merge
Plin-32
Phlh-2
Phlh-4
Phlh-2
Phlh-10
Phlh-2
Phlh-19
Phlh-2
Pngn-1
Phlh-2
Phlh-15
Phlh-2
Phlh-4
Phlh-2
Phnd-1
Phlh-2
Phlh-10
Phlh-2
Plin-32
Phlh-2
Phlh-12
Phlh-2
Phnd-1
Phlh-2

Figure III-4. Post-embryonic co-expression of HLH-2 and its partners
Phlh-2::mCherry::*his-11* transgenic animals were crossed with each of the *Phlh-x*::GFP animals to determine co-expression (indicated by white arrowheads). We observed co-expression of HLH-2 and its partners in head neurons (left), tail neurons (upper right), the vulva (middle right), and distal tip cells (bottom right).

Table III-5. bHLH heterodimer co-expression matrix

	Temporal Expression								Spatial Expression																							
Dimer	0	1	2	3	4	5	6	7	I	PI	DTC	GS	HM	BM	HN	BN	TN	P	PG	PIV	HH	BH	TH	PC	SC	X	C	S	V	R
HLH-2/HLH-3	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
HLH-2/HLH-4	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/HLH-10	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
HLH-2/HLH-15	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/HLH-19	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/LIN-32	0	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0
HLH-2/CND-1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
HLH-2/HLH-8	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
HLH-2/HLH-6	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
HLH-2/HLH-13	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/HND-1	0	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/NGN-1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HLH-2/HLH-12	0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
MML-1/MXL-2	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
AHA-1/AHR-1	0	0	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0
AHA-1/CKY-1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
AHA-1/HIF-1	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	1	0
AHA-1/HLH-33	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0	1	0
AHA-1/HLH-34	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure III-5.

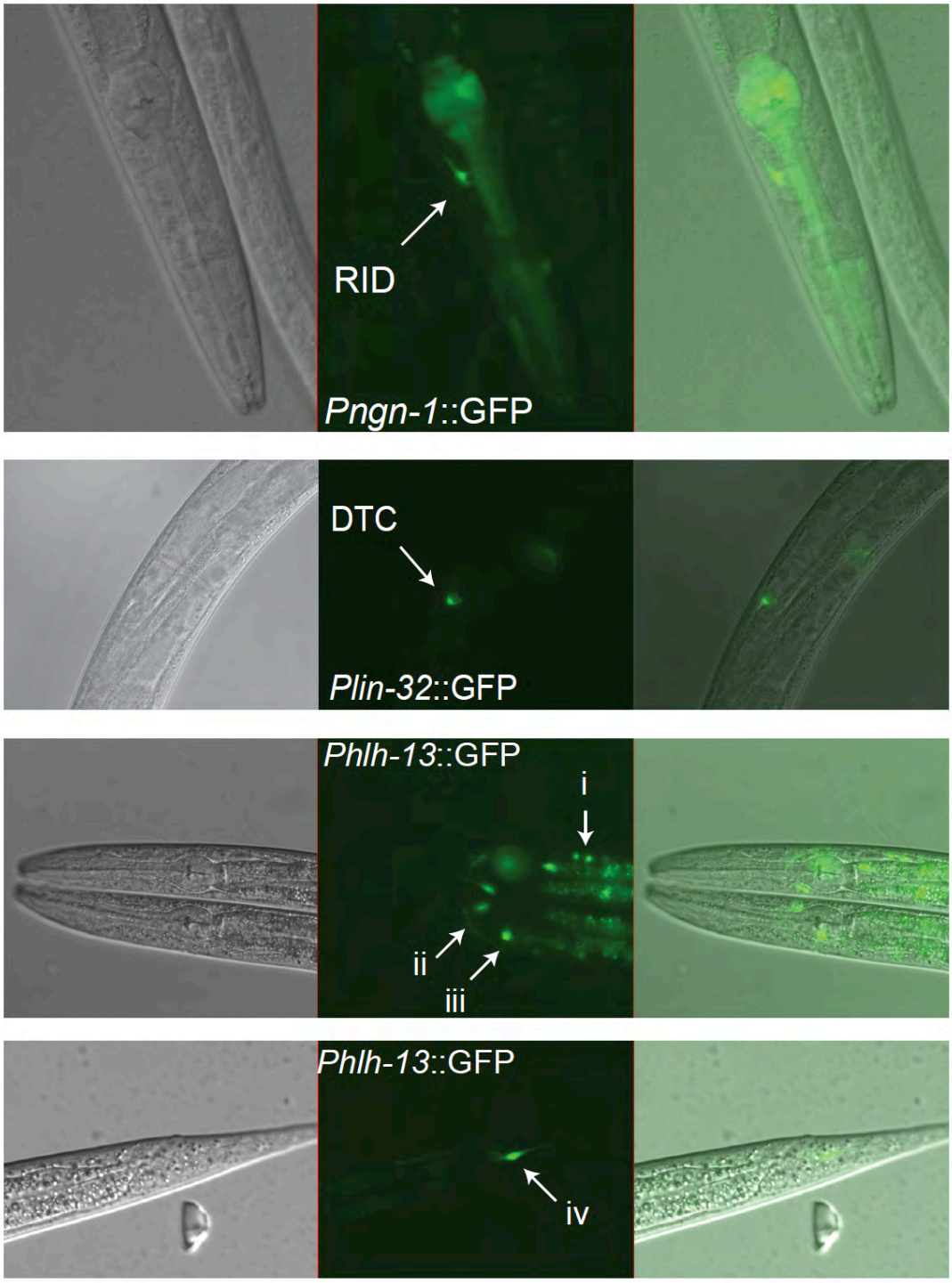


Figure III-5. Cellular resolution expression annotations
Pngn-1 appears to drive GFP expression in the RID neuron. *Plin-32* drives GFP expression in both distal tip cells. *Phlh-13* drives GFP expression in: i) 2 neurons of the retrovesicular ganglion, ii) 2 interneurons (1 bilaterally symmetric pair) of the lateral ganglion, iii) 2 interneurons (1 bilaterally symmetric pair) posterior to the posterior pharyngeal bulb; likely ADA or RMG, iv) 1 neuron of the tail; likely ALN.

PREFACE TO CHAPTER IV

This chapter describes the comprehensive and unbiased determination of the DNA binding specificity for 19 *C. elegans* bHLH dimers. We then describe the use of the DNA binding profiles of the bHLH dimers to identify candidate target genes and, subsequently, enriched Gene Ontology terms associated with those candidate target genes to generate predictions of each dimer's function.

Much of this chapter has been published separately in:

Grove C. A., De Masi F., Barrasa M. I., Newburger D. E., Alkema M. J., Bulyk M. L., Walhout A. J. M. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell. 2009 Jul 23; 138(2): 314-327.

CHAPTER IV

The DNA Binding Specificities of the *C. elegans* bHLH Transcription Factors

Abstract

The DNA-binding specificity of a TF is arguably the most important determinant of TF function, at least with regards to target gene recognition and regulation, thought to be the major functional role of a TF. The determination of the identity of *C. elegans* bHLH dimers in Chapter II of this thesis provides a unique opportunity to assess the complete DNA binding specificity of these dimers. Without information about which dimers form, systematic and comprehensive analysis of bHLH TF DNA binding profiles cannot be performed. Here we describe the use of protein binding microarrays (PBMs) to determine the *in vitro* DNA binding specificity for 19 *C. elegans* bHLH dimers, 9 homodimers and 10 heterodimers. Such an analysis provides novel information regarding the degree to which different bHLH TFs in an organism recognize and bind to different or similar DNA sequences, and what potential implications this has for target gene identification *in vivo*.

Introduction

The bHLH TFs, like all regulatory TFs, bind DNA in a sequence-specific manner, thereby binding to cis-regulatory elements within and around target genes. These proteins are generally thought to bind to E-Box sequences, defined by the sequence CANNTG, where N is any nucleotide. The abundance of E-boxes in any typical genome, however, makes it difficult to search for target genes based on this information alone. Hence, it is likely that the central and flanking nucleotides of E-boxes must be important for bHLH dimer binding. However, the degree to which these nucleotides affect binding is, for the most part, undetermined. There has been no systematic analysis of the DNA binding preferences for an entire family of bHLH TFs in any organism. In this chapter, we describe the use of protein binding microarrays (PBMs) to determine the entire spectrum of sequences that 19 *C. elegans* bHLH dimers bind to, and the order of preference that each dimer has for these sequences.

We find that we can retrieve all seven known DNA binding sites for the *C. elegans* bHLH dimers, as well as 52 novel binding sites, including E-Boxes and E-Box-like sequence variants, similar to sites previously described for the Hairy/Enhancer of Split class of bHLH TFs. The bHLH dimers can be clustered into two distinct clusters based on sequence preference, suggesting an early evolutionary divergence of two classes of bHLH proteins with respect to DNA binding specificity. One cluster binds exclusively to E-Box sequences of various types, but typically do not bind to CACGTG sites. The second cluster binds

predominately to CACGTG sites as well as E-Box variants, such as CACGCG and CATGCG. Interestingly, neither cluster binds to CAA half-site-containing E-Boxes, with the exception of HLH-1 homodimers, which bind to CAACTG sites, albeit relatively weakly. Analysis of the impact of E-Box flanking nucleotides suggests that a majority of bHLH dimers avoid binding to E-Boxes that are flanked at the 5' end of the site by the nucleotide thymine (i.e. TCANNTG), with the exception of HLH-11 and HLH-30, which seem to either be ambivalent to the presence of a 5' thymine (HLH-11) or bind preferentially to E-Boxes flanked by a 5' thymine (HLH-30).

We made full use of the PBM-ranked and scored list of DNA sequences to predict candidate target genes for each bHLH dimer. The lists of putative target genes were then analyzed for enriched Gene Ontology (GO) terms, allowing the prediction of functions for each TF dimer. As may be expected, bHLH dimers expressed in neurons often associated with neuro-related GO terms, bHLH dimers expressed in the intestine often associated with metabolism GO terms, and many bHLH dimers expressed during embryogenesis are associated with developmental GO terms.

Results

The Experimental Approach – Protein Binding Microarrays

Protein binding microarrays (PBMs) are a relatively recent addition to the array of techniques for determining a protein's DNA binding specificity. PBMs are double-stranded DNA (dsDNA) microarrays onto which transcription factors (or other DNA binding proteins) are incubated (and their binding subsequently detected) to determine their DNA binding specificity. Once binding equilibrium has been reached, a fluorescently labeled antibody (primary or secondary) is used to tag the spots of dsDNA that are bound by the protein of interest. Using various statistical analyses, one can acquire a relatively complete *in vitro* DNA binding profile for any soluble DNA-binding protein in an unbiased manner (40).

Previous to this study, PBMs were applied almost exclusively to monomeric transcription factors or homo-oligomeric DNA binding complexes (e.g. homodimers). This standard approach was suitable for testing homodimeric bHLH complexes for *in vitro* DNA binding specificity, but we needed a slightly modified approach to identify DNA binding specificity for heterodimeric bHLH complexes, which we describe below.

DNA Binding Specificity Analysis of Homo- and Heterodimeric bHLH TFs

bHLH TFs bind DNA as obligatory homo- or heterodimers and are classically described as recognizing E-box sequences (CANNTG) (71). Previously, a handful of DNA sequences that can be bound by seven of the known *C. elegans* bHLH dimers had been identified (80; 82; 83; 66; 87; 86).

However, in those studies only one or a few of all possible E-boxes were considered, and no experiments were done to determine, comprehensively, the DNA binding preferences of all *C. elegans* bHLH dimers (all bHLH ORFs used in PBMs were clonal and, unless otherwise noted (see Table II-1) fully wild type; see Table IV-1 for a list of ORF sequencing primers).

We used PBM assays (40; 43; 136) to comprehensively identify the sequence preferences of 19 bHLH dimers. We first tested each available bHLH TF individually in PBM assays, as a GST fusion protein and obtained DNA binding profiles for MXL-3, HLH-1, HLH-11 HLH-25, HLH-26, HLH-27, HLH-29, HLH-30 and REF-1, demonstrating that these proteins can bind DNA without protein partners, presumably as homodimers. We speculate that proteins that yielded sequence-specific DNA binding profiles in PBM assays but that were not detected as interacting with any bHLH protein by Y2H assays (*e.g.* HLH-25) may dimerize in a DNA-dependent manner, as has been reported previously for some TFs (137).

Importantly, none of the bHLH proteins that participate in heterodimeric interactions exhibited significant sequence-specific DNA binding on their own (Figure IV-1, IV-4, and IV-5). This presented a convenient strategy to determine the DNA binding profiles of heterodimeric TFs, by incubating the DNA microarrays simultaneously with a GST-fusion bHLH protein that did not bind to DNA on its own, and a FLAG-tagged partner protein with subsequent detection

using a fluorophore-conjugated anti-GST antibody. We examined each of the bHLH heterodimers identified by our Y2H screen in this manner.

We obtained DNA binding profiles for 9 homodimers and 10 heterodimers, including most heterodimers involving HLH-2, two class IV dimers, and five out of six REF-1 family proteins (Class VI) (Figure IV-2, IV-3, and IV-5). We did not detect any sequence-specific DNA binding by the bHLH-PAS class of dimers, even though these readily form heterodimers in the Y2H system. It is possible that sequence-specific DNA binding by members of this class requires ligands or post-translational modifications (88).

PBM Results Support Y2H Findings

There are a number of results from the findings of the PBM experiments that corroborate the findings of the Y2H dimerization results, and *vice versa*. First, the only proteins for which dimerization interactions were found in the Y2H assay and were capable of binding DNA as single proteins (or presumably homodimers in solution) were those that were found to homodimerize in the Y2H experiments. All other proteins from the Y2H dimerization network (involved exclusively in heterodimers) were incapable of binding DNA individually. Second, when most of the proteins implicated in heterodimerization via Y2H experiments were specifically incubated on PBMs with their Y2H-determined dimerization partners, a specific DNA binding profile was detected. Third, negative control experiments (described below) in which bHLH proteins within the same class or from different classes were incubated pairwise on PBMs, no DNA binding was

detected, indicating that the DNA binding activity of these bHLH heterodimers was specific to those heterodimers found by the Y2H assay.

Two Clusters of DNA-Binding Specificity

The PBM-derived 8-mer data span the full affinity range of DNA binding preferences (1). We calculated enrichment scores (ESs) from the PBM signal intensities for all possible 8-mers, and for each bHLH dimer that yielded sequence-specific DNA binding, and derived position weight matrices (PWMs) for each dimer (Table IV-2, Figure IV-2). We imposed a conservative threshold (ES ≥ 0.40) to identify significantly bound 8-mers. We then hierarchically clustered both the dimers and the 8-mers and found that the bHLH proteins can be grouped into two clusters corresponding to different bHLH classes: Cluster I contains HLH-2 and its partners, HLH-1 and HLH-11, and cluster II contains class III, IV, and VI bHLH proteins (Figure IV-3).

As expected, HLH-2-containing dimers (cluster I) exhibit a strong preference for E-box sequences (CANNTG) (2). Surprisingly, however, cluster II dimers, in addition to binding a few E-boxes, also bind multiple non-E-box sequences. These resemble E-boxes, but contain a C or A in the fifth position and a G or T in the sixth position of the binding site (CAYRMK). These “E-box-like sequences” include the reported CACGCG binding site of *Drosophila* Hairy, and N-boxes (CACNAG), which are bound by *Drosophila* Enhancer of Split (138).

We determined the statistical significance of the preference of each bHLH dimer for E-box and E-box-like sequences as compared to all other 8-mers

(Figure IV-4, IV-5, and IV-6A). As shown in Figure IV-4, neither HLH-2 nor HLH-10 alone can bind significantly to any E-box or E-box-like sequence. However, when combined, they can bind five different sequences. Figure IV-6B shows that the bHLH DNA binding network also displays degrees of specificity and promiscuity. For instance, only HLH-1 homodimers can bind CAA-containing E-boxes (Figure IV-6A). Some E-boxes and E-box-like sequences are preferred by relatively few dimers, whereas others are bound by many dimers. For example, CACATG is bound by only four dimers, but CACCTG is bound by ten distinct dimers. Conversely, some bHLH dimers bind few E-boxes or E-box-like sequences whereas others bind many: HLH-30 only binds CACGTG, but HLH-2/HLH-10 binds five different E-boxes (Figure IV-6B). This demonstrates that there is specificity and promiscuity in the bHLH DNA binding network, both from the perspective of the dimers and from the perspective of their DNA binding sequences.

The REF-1 Family: The Missing* C. elegans *Class VI bHLH Proteins?

The group of *C. elegans* bHLH genes referred to as the "REF-1 family" of bHLH proteins are a unique class of bHLH genes, each predicted to encode proteins with two bHLH domains. These genes include *hlh-25*, *hlh-26*, *hlh-27*, *hlh-28*, *hlh-29*, and *ref-1*. These genes are highly similar to one another with regards to amino acid sequence, and are presumably the result of evolutionarily recent gene duplications (Figure I-5). Although there have been no clear representatives of the Class VI bHLH genes (e.g. Hairy, Enhancer of Split) in *C. elegans*, a

number of observations have indicated that the REF-1 family may represent an evolutionarily divergent form of the Class VI bHLH genes (112). Our DNA binding specificity analysis also support this hypothesis, as some of the non-canonical E-Box binding sites identified for this group (particularly HLH-25, HLH-27, and HLH-29) are similar (or identical) to those previously identified for Drosophila enhancer of split (E(spl)) and mammalian HES1 and HES5 Class VI bHLH proteins (the E-box-related N-box, CACNAG) (139-141) and for Drosophila hairy homodimers (the N-box related CACGCG) (142; 143).

Although the HLH-25 and HLH-27 proteins are, with the exception of four amino acids, identical (likely the result of a gene duplication in recent evolutionary history) their DNA binding profiles are reproducibly distinct. In fact, there is only one amino acid difference between the two proteins within the bHLH domain, presumed to be the sole determinant of DNA binding specificity. Their DNA binding specificity profiles are indeed very similar in that virtually all DNA binding sites bound by HLH-27 are also bound by HLH-25, but HLH-25 also binds to a large number of other related sites, suggesting that HLH-25 binds to DNA much more promiscuously than its bHLH TF “twin”. This observation could be the result of a relatively increased affinity of HLH-25 for DNA, whereby a reduced nuclear concentration of HLH-25 compared to HLH-27 could effectively occupy the same (or similar) binding sites.

C. elegans ***bHLH Dimers Bind E-box and E-box-related Sequences, but Avoid CAA-Containing E-boxes***

Next, we examined E-box and E-Box-like sequences bound by the different dimers in a more detailed and unbiased manner (i.e., without choosing an ES threshold). First, we determined the statistical significance of the preference of each bHLH dimer for E-box and E-box-like sequences as compared to all non-E-Box/E-box-like sequences using one-way ANOVA applied to the distributions of enrichment scores of a dimer for each site (Figure IV-4, IV-5) (see Materials and Methods). The statistical results from the overall analysis are shown in Figure IV-6A and 6B.

We successfully retrieved all known *C. elegans* bHLH binding sites for those bHLH dimers that yielded DNA binding profiles at our PBM threshold (Figure IV-6A, red boxes). Moreover, we obtained 52 novel bHLH dimer binding sites, including novel sites for several bHLH dimers for which interacting sequences had previously been published (Figure IV-6A). Gel shift studies had previously detected DNA binding for HLH-2 and HLH-8 homodimers (3; 4), suggesting that these two proteins can form homodimers. In contrast, we did not observe HLH-8 homodimers in Y2H assays, nor did we observe DNA binding of HLH-2 and HLH-8 individually, even at high (up to 1 μM) protein concentrations (Figure IV-5 and data not shown).

Surprisingly, these results revealed that, except for the HLH-1 homodimer, none of the bHLH dimers bound CAA-containing E-boxes (Figure IV-6A). In

addition, these results show that some binding sites are preferred by relatively few dimers, whereas others are bound by many dimers; for example, the CACATG E-box is bound by only four different dimers, whereas the CACCTG E-box is bound by ten distinct dimers.

DNA Binding Preferences of Individual bHLH Dimers

The DNA binding profiles revealed that the bHLH dimers can be further sub-divided into five groups: four that most highly prefer different E-box sequences, and one that prefers an E-box-like sequence (Figure IV-2). We found numerous subtle differences among bHLH dimers that belong to the same class or module, with regard to both the preferred E-Box/E-Box-like sequences and the preferred 5' flanking nucleotide position (see below). These differences may help in functional specification. Even though the different dimers belonging to the HLH-2 module exhibit overall similar DNA binding profiles with specificity for E-boxes (Figure IV-6A,6B), some of these dimers bind more distinct sets of sequences. Indeed, HLH-2/HLH-3 and HLH-2/HLH-8 heterodimers share only one E-Box sequence (CATCTG). Taken together, we observed different degrees of specificity and promiscuity in DNA binding both for different bHLH dimers, and for the different E-Box and E-Box-like sequences that were bound.

Half-Site Preferences of Individual bHLH Proteins

In order to gain further insight into the functional specification of each individual bHLH protein, we examined their individual DNA binding preferences. Since each bHLH protein contacts a half-site within the CANNTG sequence (96;

95; 97), we reasoned that the DNA binding profiles of individual dimers may allow the inference of bHLH monomer half-site preferences. Specifically, we assumed that when a dimer binds a palindromic DNA sequence, each of the monomers must recognize the same half-site, and used this reasoning to infer half-site preferences of each bHLH protein.

There are no palindromic E-Box-like sequences and thus all of the palindromic sites retrieved from the PBM data are E-boxes. There are 10 possible E-boxes when both the forward and reverse complement orientations are considered (e.g., we consider CACCTG and CAGGTG to indicate the same E-box), including four palindromes, one of which (CAATTG) is not bound by any dimer (Figure IV-6A). We first focused on homodimer-palindrome interactions. Class III, IV and VI homodimers (HLH-30, MXL-3, and some REF-1 family members) exclusively interact with the CACGTG palindrome, and so each protein must be able to specify the CAC half-site. Similarly, HLH-1 binds only the CAGCTG palindrome, and so each HLH-1 monomer likely contacts CAG. The observation that most of the homodimers bind only a single palindrome suggests that these proteins most highly prefer the corresponding half-site. Indeed, in most cases these palindromes exhibited the highest PBM enrichment scores (Table IV-2, Figure IV-5). Many homodimers can also bind non-palindromic sequences, most of which contain the half-site present in the preferred palindromic sequences. Finally, we found that palindromes of less preferred half-sites are typically not bound at ES ≥ 0.40. For example, whereas the MXL-3 homodimer

most highly prefers the CACGTG E-box, it can also interact with CACATG sites (Figure IV-5, IV-6A, IV-6B). However, it does not interact with CATATG sites, indicating that it prefers CAC over CAT. These data suggest that when one of the half-sites provides a high-affinity platform for one of the members of a bHLH dimer, the other member can tolerate a lower affinity half-site.

Next, we investigated the half-site preferences of heterodimeric bHLH proteins. The MDL-1/MXL-1 dimer binds the CACGTG palindrome and a number of non-palindromic E-Box-like sequences, most of which contain a CAC half-site (Figure IV-5, Figure IV-6A,6B). However, since neither MDL-1 nor MXL-1 dimerizes with any other bHLH proteins, it is not possible to determine which of these proteins is contacting which half-site in the non-palindromic sequences.

HLH-2-containing dimers bind all palindromic E-boxes except CAATTG. Thus, HLH-2 must be able to bind CAT, CAC and CAG half-sites (Figure IV-7). We cannot be sure that the HLH-2 monomer has preference for these half sites or if it merely tolerates their presence, but the conclusion still stands. This is in agreement with crystal structures of E47 dimers that show that E proteins are capable of binding to each of these half-sites (21; 22). HLH-2 binds 14 other bHLH proteins, and these heterodimers have partially overlapping DNA binding specificities (Figure II-3, Figure IV-3, IV-5, IV-6). The fact that these specificities are non-identical suggests that each HLH-2 partner must, at least in part, be responsible for the specification of particular half-sites. For example, upon close examination of the binding specificity of the HLH-2/HLH-3 dimer, we noticed that

in addition to binding CAGCTG palindromes, this dimer recognizes CACCTG and CATCTG sequences (Figure IV-8). Each of these sequences contains a CAG half-site that HLH-3 could bind and another half-site that could be bound by HLH-2 (Figure IV-7, Table IV-3). Thus, we propose that HLH-3 specifies the CAG half-site, and that HLH-2 contacts the other half-site, although we cannot exclude the possibility that HLH-3 half-site recognition may be E-box-dependent. By similar reasoning, we propose that HLH-8, which together with its partner HLH-2 recognizes the CATATG palindromic E-box and the CATCTG and CACATG non-palindromic E-boxes (Figure IV-8), contacts the CAT half-site within these sequences. A general scheme of the half-site logic is provided in Figure IV-7 and IV-9. All deduced half-site preferences for HLH-2 and its partners are provided in Table IV-3.

E-Box Flanking Nucleotides Contribute to bHLH DNA Binding Specificity

The PBM ES of a particular DNA sequence bound by a dimer is a reflection of relative DNA binding affinities (1). We noticed that the ES distribution for 8-mers corresponding to a particular dimer/sequence combination varied greatly. For instance, both HLH-26 and MDL-1/MXL-1 bind CACGTG E-boxes, but HLH-26 does so with a broad ES range and MDL-1/MXL-1 with a very narrow ES range (Figure IV-10). This suggests that, in contrast to MDL-1/MXL-1, not all CACGTG E-boxes are bound equally well by HLH-26. We considered the possibility that differences may be due to effects of nucleotides flanking the core CACGTG E-box. Indeed, flanking nucleotides have been reported previously to

contribute to bHLH dimer DNA binding (144-146). However, the effects of nucleotides flanking the E-box and E-box-like sequences had not been analyzed systematically for most bHLH TFs. Since each bHLH monomer may directly contact the flanking nucleotide immediately 5' of the E-box (95; 144), we examined the influence of this position on relative DNA binding preferences. We found that for the MDL-1/MXL-1 dimer each of the four possible nucleotides flanking the CACGTG core sequence is recognized approximately equally well; the ES for each relevant 8-mer is between 0.49 and 0.50 (Figure IV-10). However, HLH-26 exhibits a strong preference for a 5' A or G (median 8-mer ES > 0.40), and disfavors a 5' T (median 8-mer ES < 0.10) and, to a lesser extent, a 5' C ($0 \leq ES \leq 0.40$) (Figure IV-10).

We found that most bHLH proteins exhibit preferences at the 5' flanking nucleotide position (Figure IV-11). We found that most dimers disfavor a 5' T; this observation is similar to what has been reported for the yeast bHLH homodimer Pho4p (144). However, there are exceptions: HLH-11 and MDL-1/MXL-1 heterodimer both tolerate a 5' T, and HLH-30 actually favors a 5' T (Figure IV-11).

E-Box-dependent Flanking Nucleotide Preferences

Another common question in the field of TF-DNA interaction is whether or not there are nucleotide inter-dependencies with respect to TF DNA binding affinity. In other words, are certain nucleotides in a binding site favorable or disfavored only in the presence of other specific nucleotides? The implications of

the answer to this question affect the reliability of position weight matrices and associated logos as a common format for representing DNA binding sites (147; 148). Our data indicate that there are indeed nucleotide inter-dependencies as indicated by E-Box-dependent flanking nucleotide preferences exhibited by dimers in our study. For example, the HLH-2/HLH-10 heterodimer binds to both CACGTG and CAGCTG palindromic E-Boxes. However, in the context of CACGTG, HLH-2/HLH-10 favors a flanking adenine or a guanine and disfavors both a cytosine and a thymine, whereas the same heterodimer disfavors only a flanking thymine when bound to the CAGCTG binding site (Figure IV-12A). This context-specific affect of a flanking cytosine suggests that nucleotides central to these E-Boxes can affect whether or not flanking nucleotides influence the DNA binding affinity for the same dimer. The important consideration of this result is that, when observing position weight matrices and their corresponding logos, the individual nucleotide positions cannot always be considered as independent entities. Therefore, it will be important in future TF/DNA interaction studies to pay close attention to the actual individual sequences bound by a TF and the order of preference that a TF has for each sequence.

Dimer-dependent Flanking Nucleotide Preferences

The converse to E-Box dependent flanking nucleotide preference is dimer-dependent flank preference for a given E-Box. For example, we find that HLH-2/HLH-3 heterodimers and HLH-2/HLH-4 heterodimers differentially bind the CACCTG/CAGGTG E-Box with respect to flanking nucleotide preference. We

found that whereas the HLH-2/HLH-3 heterodimer binds to CACCTG/CAGGTG with distinct “avoidance” of a 5’ flanking thymine, the HLH-2/HLH-4 heterodimer binds to the same core E-Box with less preference for flanking nucleotides, particularly at the 5’ end of the CACCTG E-Box, perhaps as a result of one monomer of the dimer being influenced more than the other (Figure IV-12B). Our half-site analysis (described above) would suggest that it is HLH-2 that is differentially affected by the flanking nucleotides in a partner-dependent manner.

In summary, we identified both prominent and subtle differences in E-box or E-box-like sequence recognition and flanking site preferences between different bHLH dimers, which likely contribute to target site selection and gene regulation *in vivo*.

Negative Control Experiments

Although we found that none of the bHLH proteins implicated in heterodimers (by the Y2H system) were capable of binding to DNA on PBMs in the absence of a bHLH partner, we wanted to be sure that combinations of bHLH TFs not found to heterodimerize in Y2H would also not bind to DNA on PBMs. To address this possibility, we randomly chose bHLH TFs from different or similar bHLH classes and incubated them in equimolar concentrations (at concentrations equivalent to previous experiments (i.e. ~200nM)) as a 1:1 protein mixture on PBMs. We found that, as expected, none of the randomly assigned pairs of bHLH proteins were found to bind to DNA (Figure IV-13). This observation

corroborates the Y2H data, suggesting that the dimers tested here are, in fact, truly negative interactions.

Identifying bHLH Dimer Candidate Target Genes

Aside from achieving an understanding about how transcription factors recognize specific sequences of DNA, it is hoped that determining DNA binding specificities for various TFs will allow researchers to make predictions of target genes. To this end, we tried to make the best use of the PBM data to predict which target gene promoters the various bHLH dimers may bind to. The PBM results provided us with a rank-ordered list of 8-mers, the vast majority of which contained E-box or E-box related sequences.

Because each 8-mer could contain an NN-E-box, N-E-box-N, or an E-Box-NN sequence, and because it is possible that flanking nucleotides could positively or negatively influence DNA binding by a particular dimer, we wanted to incorporate all of the 8-mer data into a more comprehensive predictor of binding. We reasoned that by averaging enrichment scores from three 8-mers that constitute a single 10-mer, we could create a new rank-ordered list of 10-mer DNA sequences which we could use to predict target genes. For example, the three 8-mers TGCACGTG, GCACGTGA, and CACGTGAT can be combined to generate the 10-mer sequence TGCACGTGAT, and the enrichment scores from these three 8-mers could be averaged to approximate the relative binding preference for the 10-mer sequence. The additional benefit of using 10-mers is

that we can be more specific with our target gene predictions, as 10-mers will predict 16-fold fewer sites in a genome than 8-mers.

To assess the validity of this approach, we generated the rank ordered list of 10-mers for each bHLH dimer and looked at the raw PBM intensity scores for each 60 nucleotide deBruijn sequence to see if there is a genuine correlation between PBM spot intensity and the presence of a high-scoring 10-mer sequence. We noticed a strong correlation between the presence of the highest scoring 10-mers and the deBruijn sequences of the brightest spots on the PBM. Although we did not perform a systematic analysis of all dimers and 10-mers, this approach seemed like the best use of PBM 8-mer data for making target gene predictions.

The highest level of sequence conservation of gene regulatory regions within related nematode species lies in the 500 bp upstream of gene starts (149). Therefore, we searched this genomic region for all predicted *C. elegans* genes for the different bHLH binding sequences (10-mers with average ES above 0.3) to identify candidate bHLH target genes (Table IV-4, provided only electronically due to size; see Figure IV-14 for the candidate target gene identification pipeline). We calculated a cumulative ES for each gene (sum of all 10-mer average ESs), with respect to each of the bHLH dimers, and kept only genes that score 0.4 or greater to identify genes with either single “high-affinity” binding sites, or with multiple “lower affinity” binding sites, or a combination of both.

Functional Annotation of Candidate bHLH Target Genes

We used lists of candidate target genes mentioned above to initiate functional annotation of the dimers by searching for over-represented Gene Ontology (GO) categories (150) within each list. We then identified over-represented GO annotation terms associated with these putative target genes, and, hence, with the relevant bHLH dimer (Table IV-5; see Materials and Methods).

We identified multiple enriched GO terms, including Molecular Function terms associated with transcription and signaling, and Biological Process terms associated with development and metabolism. Some of the annotations we obtained are in agreement with what was previously known, either in *C. elegans* or for orthologs in other organisms. For instance, the connection of MDL-1/MXL-1 to “cell division” is evolutionarily conserved with the orthologous human dimer MAD/MAX (8). However, the majority of functional annotations are novel.

As described in the next chapter, we find that HLH-30 is associated through its DNA binding specificity to several metabolic and reproductive terms. The fact that we also see *Phlh-30* driving GFP expression in the intestine, vulva, and spermatheca suggests that HLH-30 may indeed regulate genes involved in metabolism and reproduction. To test this hypothesis, we performed a microarray experiment to assess the gene expression profile of worms harboring a deletion of the bHLH domain-encoding region of the *hlh-30* gene locus. The results, as

will be explained, corroborate our findings of predicted HLH-30 target genes and functionality.

Discussion

Protein Binding Microarrays: Data Quality

Several observations indicate that the PBM data are of high quality: (1) As mentioned above, for bHLH-bHLH pairs that do not form heterodimers in Y2H assays, no specifically bound E-boxes or E-box-like sequences were obtained (Figure IV-13). (2) The PBM experiments yielded E-boxes and similar sequences, several of which constitute known bHLH binding sequences. (3) We identified all previously known *C. elegans* bHLH binding sequences for those bHLH dimers that yielded sequence-specific DNA binding profiles (Figure IV-6A,6B). (4) We have previously demonstrated that PBM assays retrieve DNA binding sites that are highly relevant *in vivo* (43; 136; 12; 151). Taken together, we are confident that the DNA binding specificities determined by PBMs are of high quality and are relevant to the biology of *C. elegans*.

Two Clusters of bHLH DNA Binding Profiles

The PBM data collected here allowed us the opportunity to compare and contrast the bHLH dimers with respect to their DNA binding preferences. By clustering each of the 19 bHLH dimers for which PBM-derived DNA binding specificities were acquired, we found that two groups, or clusters, of bHLH dimers emerged, the members of which share general similarities in sequences bound (Figure IV-3). This observation suggests that the bHLH TFs underwent a relatively early evolutionary divergence into two classes with regards to DNA binding specificity. Further systematic analysis of DNA binding for bHLH TFs in

other organisms will be needed to see if these two clusters of DNA binding profiles exist in other species as well.

E-Boxes and E-Box-like Sequences

The 19 bHLH dimers for which we acquired PBM data bound DNA sequences that almost always contained E-Box sequences (CANNTG) or E-Box-like sequences (e.g. CACGCG, CATGCG, etc.). The question of how nucleotides central to an E-Box binding site affect bHLH dimer binding seemed to be answered by the observation that all possible E-Box sequences were bound except for E-Boxes containing a CAA half-site (i.e. CAATTG, CAACTG, CAAGTG, or CAAATG). HLH-1 homodimers can bind to CAACTG sites, although they are not the most preferred of HLH-1 binding sites. Further experiments will be necessary to understand the biochemical reason behind this observation.

In addition to E-Boxes, several E-Box-like sequences were bound by bHLH dimers: CACGCG, CATGCG, CACACG, CATACG, CACGAG, and CACGCT. These sequences, particularly CACGCG, resemble the binding site that has been reported for homodimers of the *Drosophila* Hairy bHLH protein. The sequence CACGCG is predominately bound by HLH-25, HLH-27 and HLH-29, members of the REF-1 family of bHLH proteins. This family of proteins has been identified as a target of Notch signaling in the developing *C. elegans* embryo (14), which, taken together with the fact that these proteins have weak homology to Hairy and/or Enhancer of Split proteins (Figure I-5), suggests that

the REF-1 family of proteins may in fact represent the *C. elegans* Class VI bHLH proteins.

Flanking Nucleotide Contribution to bHLH Dimer Binding

There has been little conclusive evidence to suggest the extent to which nucleotides flanking E-Box binding sites contribute to bHLH dimer binding affinity. Studies as early as 1992 suggested that the Pho4p yeast bHLH TF bound to E-Boxes less effectively when a 5' flanking thymine was present in the binding site (144). Maerkl and Quake used their MITOMI technique to demonstrate that the bHLH TFs Pho4p, Cbf1p, and MAX each had varying degrees of sensitivity to nucleotides flanking the E-Box (146). Our data also suggest that bHLH TF dimers are sensitive to E-Box flanking nucleotides, although this sensitivity can be E-Box-dependent for a given dimer, and dimer-dependent for a given E-Box. Our systematically derived DNA binding specificity data provide a novel opportunity to address the questions of overall flanking nucleotide affects for most bHLH dimers in a multicellular organism.

One interesting observation was that a large number of bHLH dimers exhibited an avoidance of E-Box sites bearing a thymine at the 5' flanking position. Fisher and Goding had made the observation that the yeast Pho4p bHLH TF also avoided binding to E-Boxes bearing a 5' thymine, and showed that it was most likely the presence of the methyl group of thymine in the major groove of DNA that sterically hindered binding of Pho4p to the E-Box site (144). It is not entirely clear from the bHLH amino acid sequences which residues or

general architectural features of the bHLH domain may be responsible for this avoidance of sites with a 5' thymine, although bHLH crystal structures suggest that it may be the 8th and/or 12th residues of the bHLH domain that make direct contact with the 5' flanking nucleotide (152; 153; 96; 97; 154)

Candidate bHLH Target Genes and Enriched GO Terms

We used the PBM-derived DNA binding specificity data to make predictions about which genes are targeted for regulation by the different bHLH dimers. Based on sequence conservation between nematode species, we chose a search space of 500bp upstream of gene starts (start codons, ATG, in this case). We scanned these genomic regions for 10-mers with an average ES above 0.3, summed these scores for each promoter region, and kept only gene promoters that score 0.4 or greater. This enables the favoring of genes that have multiple bHLH binding sites in their promoter regions.

We then took the candidate target gene list for each bHLH dimer to search for Gene Ontology (GO) terms that are statistically enriched for specific Biological Process and/or Molecular Function terms. Virtually all functional associations made by this approach are novel annotations for the designated bHLH dimers. Assigning molecular functions and biological process terms to TFs simultaneously may begin to bridge the gap in our understanding of exactly how the molecular activities of TFs and target genes manifest as more macroscopic biological processes.

Materials and Methods

Protein Binding Microarray Experiments

Microarray design, preparation and PBM experiments were performed as described previously (1). All experiments were performed using previously described custom-designed “all 10-mer” arrays synthesized in the “4x44K” array format (Agilent Technologies, Inc.) (43; 136). TFs were diluted to a final concentration, in the protein binding reactions, of 200 nM except for HLH-1, HLH-2/HLH15, HLH-2/HLH-19, HLH-2/HLH-14, and HLH-29, which were diluted to 400 nM, and HLH-11, which was diluted to 700 nM. In order to minimize potential bias resulting from the array design, each sample was assayed in duplicate, on arrays designed using different de Bruijn sequences (40; 155).

Binding Site Annotation, Mapping and Prediction of bHLH Target Genes

Target genes were predicted by initially calculating for each dimer the average 8-mer enrichment score (AvgES) within all 10-mers that contained an E-box (NN-E-box, N-E-box-N, E-box-NN)(similar for E-box-like sequence). For each bHLH dimer, genomic sequences 500 bp upstream of each WBGene (referred to as transcriptional start) were scanned with the corresponding set of 10-mers with AvgES ≥ 0.3. Each gene was scored by summing the AvgES of all 10-mers found in the 500 bp upstream sequence. All genes having a Sum of AvgESs ≥ 0.4 were considered for analysis of functional category enrichment using the GOMiner algorithm (http://discover.nci.nih.gov/gominer/) (156).

Gateway Cloning of bHLH ORFs into Expression Vectors

All *C. elegans* bHLH ORFs were cloned into the pDEST-15 Destination vector (Invitrogen Technologies) by Gateway cloning from bHLH ORF Entry clones. Select *C. elegans* bHLH ORFs were similarly cloned into the pCAL-n-FLAG Destination vector (157). All bHLH constructs were sequence verified to obtain wild-type clones (see exceptions in Table II-1).

Protein Expression and Quantification

All bHLH proteins were individually expressed using the Active Pro coupled T7 *in vitro* transcription and translation kit (Ambion) following the manufacturer's specifications. Briefly, proteins were expressed using a final concentration of 20 ng/µl of plasmid DNA, shaking at 1350 rpm in an Eppendorf Thermomixer for 90 minutes at 37°C. Concentrations of expressed proteins were estimated by Western blotting using a dilution series of recombinant GST (Sigma) or of the amino terminal BAP-FLAG control protein (Sigma), as described previously (Berger et al., 2006). 1:3000 dilution (vol/vol) of rabbit anti-GST polyclonal primary antibody (Sigma), followed by 1:5000 dilution (vol/vol) of HRP-conjugated goat anti-rabbit IgG secondary antibody (Pierce), was used in Western blotting. For FLAG-tagged HLH-2 we used HRP-conjugated murine monoclonal anti-FLAG antibody (M2 clone, Sigma) at 1:1000 dilution (vol/vol). Proteins were aliquoted and stored at –80°C.

Protein Binding Microarray Data Normalization and Motif Analysis

Microarray scanning, spot quantification, and data filtering and normalization were performed as described previously (9). Briefly, we calculated a PBM enrichment score (ES) for each contiguous and gapped 8-mer. Enrichment scores for each 8-mer from each of the two array designs were averaged. We constructed position weight matrices (PWMs) using the previously described "Seed-and-Wobble" algorithm (40; 158). Sequence logos were then generated from the Seed-and-Wobble PWMs essentially as described previously (1) using enoLOGOS (159).

Protein Binding Microarray Data Clustering

All data clustering and associated statistics were performed using Matlab with the "statistics" and "bioinformatics" toolboxes (MathWorks, MA). We performed two-dimensional hierarchical clustering of all the TFs' 8-mer ES data using only those 8-mers that had ES ≥ 0.40 in at least one bHLH dataset, using average linkage and the Pearson correlation distance metric.

Statistical Significance of bHLH Dimer Binding to E-box and E-box Variants

We grouped 8-mers according to the presence or absence of each considered 6 nucleotide E-box or E-box-like sequence. We compared each resulting 8-mer group (*i.e.*, foreground set) to all 8-mers lacking E-box or E-box-like sequences (*i.e.*, background set) by calculating the area under a receiver operating characteristic curve (AUC). AUC is a non-parametric statistic that represents the

probability that a random 8-mer from the foreground set will rank higher than a random 8-mer from the background set. In order to assess the statistical significance of the AUC for each foreground set, we utilized a permutation-based approach, as described and implemented previously (160). This approach provides an estimate of the statistical significance of each AUC value by calculating false discovery rate (FDR) Q-values from 1,000 permutations of the foreground/background assignments of the 8-mers. By controlling for both AUC and Q-value, we can identify 8-mer groups that are significantly different from the background set of 8-mers in a manner that does not rely upon assumptions about the underlying distribution of 8-mer enrichment scores and that addresses potential sample size variability of the groups (*i.e.*, the number of spots containing a given 8-mer).

Gene Ontology Annotation Term Analysis

Predicted target genes for each bHLH dimer were analyzed for enriched GO terms using the GOMiner high-throughput analysis tool (http://discover.nci.nih.gov/gominer) (156). The following parameters were chosen for the GOMiner search: Data source: Wormbase (WB), Organism: *C. elegans*, Root Category: All/Gene Ontology. Default settings were used for all other parameters. A significance threshold of $p < 0.05$ ($Log10(p) < -1.3$) for GOMiner results was used to determine which GO terms were considered to be significantly enriched. For incorporation of predicted target gene GO terms into the integrated network (see Chapter V), any term represented by fewer than 10

genes in the predicted target gene list were not included. However, these were included in the overall parameter analysis.

Online Data

Additonal PBM data are available at http://thebrain.bwh.harvard.edu/pbms/webworksW/ and in the UniPROBE database (161).

Acknowledgements

Many of the PBM experiments and DNA binding specificity analyses were carried out by Federico De Masi in the laboratory of Martha Bulyk at Harvard Medical School. All statistical analyses of PBM data were carried out by Federico De Masi and Martha Bulyk. Candidate target gene predictions were performed using PBM-derived data by Inmaculada Barrasa.

We thank Job Dekker, and members (current and past) of the Walhout and Dekker labs for critical reading of relevant manuscripts.

Table IV-1

ORF	*gene*	ORF Sequencing Primer	FW or RV
C25A1.11	*aha-1*	ATTTGATTCTAGAAGCAGCG	FW
C25A1.11	*aha-1*	AAAAATGCTCCACCACAGGG	FW
C41G7.5	*ahr-1*	GGAACACCAATGTTGGACCC	FW
C41G7.5	*ahr-1*	ATGTAATGGGAAGAACTGCC	FW
C41G7.5	*ahr-1*	CGAAGTGGAAGGTACAATGC	FW
C15C8.2	*cky-1*	CAAATGTCATCTCCAATGCC	FW
C15C8.2	*cky-1*	TTCTCGGCTTATTGTCAACC	FW
C15C8.2	*cky-1*	AAGCTGCGACTCTTCAAGCG	FW
F38A6.3	*hif-1*	TCATCCATCCGACTATGACG	FW
F38A6.3	*hif-1*	CGAATTGGTACATCCAGCCG	FW
F38A6.3	*hif-1*	CGTTGGTGGAGAAGAACCTG	FW
F38A6.3	*hif-1*	GACGCTCGATCTATGGGACG	FW
B0304.1	*hlh-1*	CCAGATCTCAAGATTTCGCATC	FW
F58A4.7	*hlh-11*	CAAAGGAGAAGATATACCTGACGG	FW
W02C12.3	*hlh-30*	TGACGATTGGTGGAGAAAAAAC	FW
W02C12.3	*hlh-30*	ACAAACGATCCGGTCGG	RV
Y39A3CR.6	*hlh-33*	ATGATTGGAAAAGATATCCG	FW
T20B12.6	*mml-1*	AGCCGAAAAAACCATTCTGC	FW
T20B12.6	*mml-1*	CTCTGAGTGAACCTTACATG	FW
T20B12.6	*mml-1*	TGGACTATCGATTAATGCCG	FW
T20B12.6	*mml-1*	CTCCATCGAGATCATGGTGG	FW
T20B12.6	*mml-1*	CATATAGTAGAAGGATCGCC	FW
T01E8.2	*ref-1*	CTTTGAAGAACTTTATCATTGAGAACA	FW
Y47D3B.7	*sbp-1*	AGAAGGTCCTGCAAGTATGCTT	FW
Y47D3B.7	*sbp-1*	TCGATTGAAGATGCTCCAGAG	FW
Y47D3B.7	*sbp-1*	GGGATGAAGCTAAGCTTTCAAA	FW
Y47D3B.7	*sbp-1*	TGGAAGAGTGATTGATGACCC	FW
Y47D3B.7	*sbp-1*	GTCTCAGTCGTTTGTAGGAGTCAT	FW
Y47D3B.7	*sbp-1*	TGTTGAGCTCGAAGCAGAAG	FW
Y47D3B.7	*sbp-1*	CTGATCAGGAACTGTCCGC	FW
Y16B4A.1	*unc-3*	AATGTTTCAGAATGGAATCCG	FW
Y16B4A.1	*unc-3*	ACATGTTCGTCCACAACAACTC	FW
General Forward Primer		TTCTACTTCTTTTACTGAAGC	FW
General Reverse Primer		CTCCACTGACAGAAAATTTG	RV

Table IV-1. ORF sequencing primers

All *C. elegans* bHLH ORFs were sequenced to verify full-length wild type sequences in all GST clones to be used in PBM assays (see Table II-1). All ORFs were sequenced at 5' and 3' ends with the general forward and reverse primers. In cases where the ORF was too long to get reliable sequence reads for the entire ORF, additional ORF-specific primers (listed in this table) were generated. Rows for every other bHLH ORF are shaded gray for clarity.

Figure IV-1

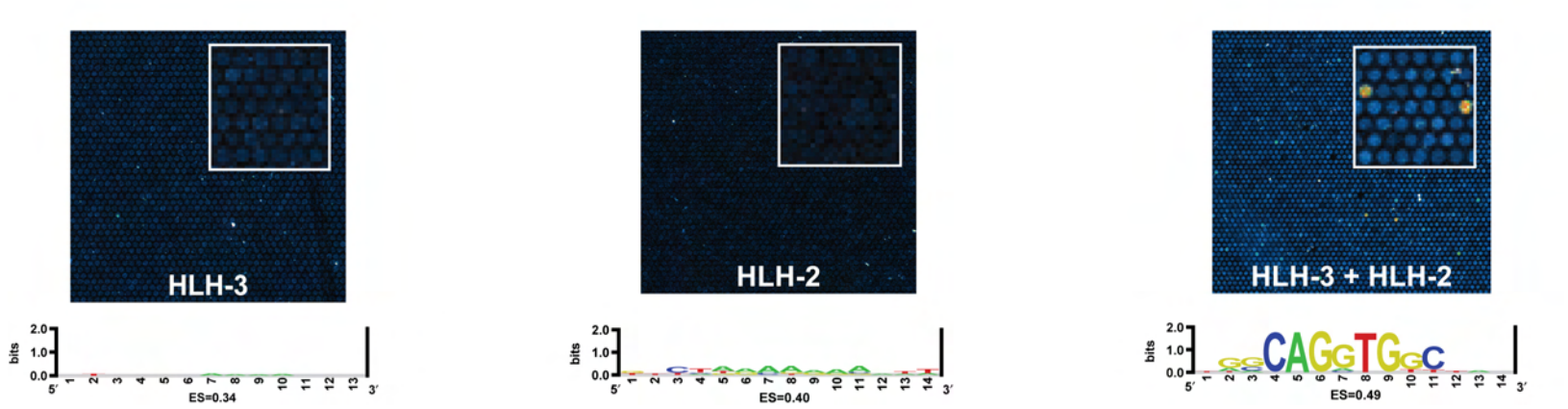


Figure IV-1. Protein binding microarray (PBM) raw data
Three panels depicting the results of typical PBM experiments. The top image in each panel depicts an image from a portion of the scanned microarray, in which dark blue and black spots indicate little or no TF binding, whereas light blue and yellow/orange/red spots indicate strong binding. The bottom image in each panel represents the Seed-and-Wobble output sequence logo for the indicated bHLH protein or dimer (see Materials and Methods). As can be seen from the PBM experiments for HLH-3 and HLH-2 alone, there is no significant DNA binding. Upon mixing the two proteins at equimolar concentrations, however, we observe significant binding to sequences containing E-Box bHLH binding sites.

Table IV-2. 8-mer enrichment score data (Due to space limitations, see electronic file for data)

Figure IV-2

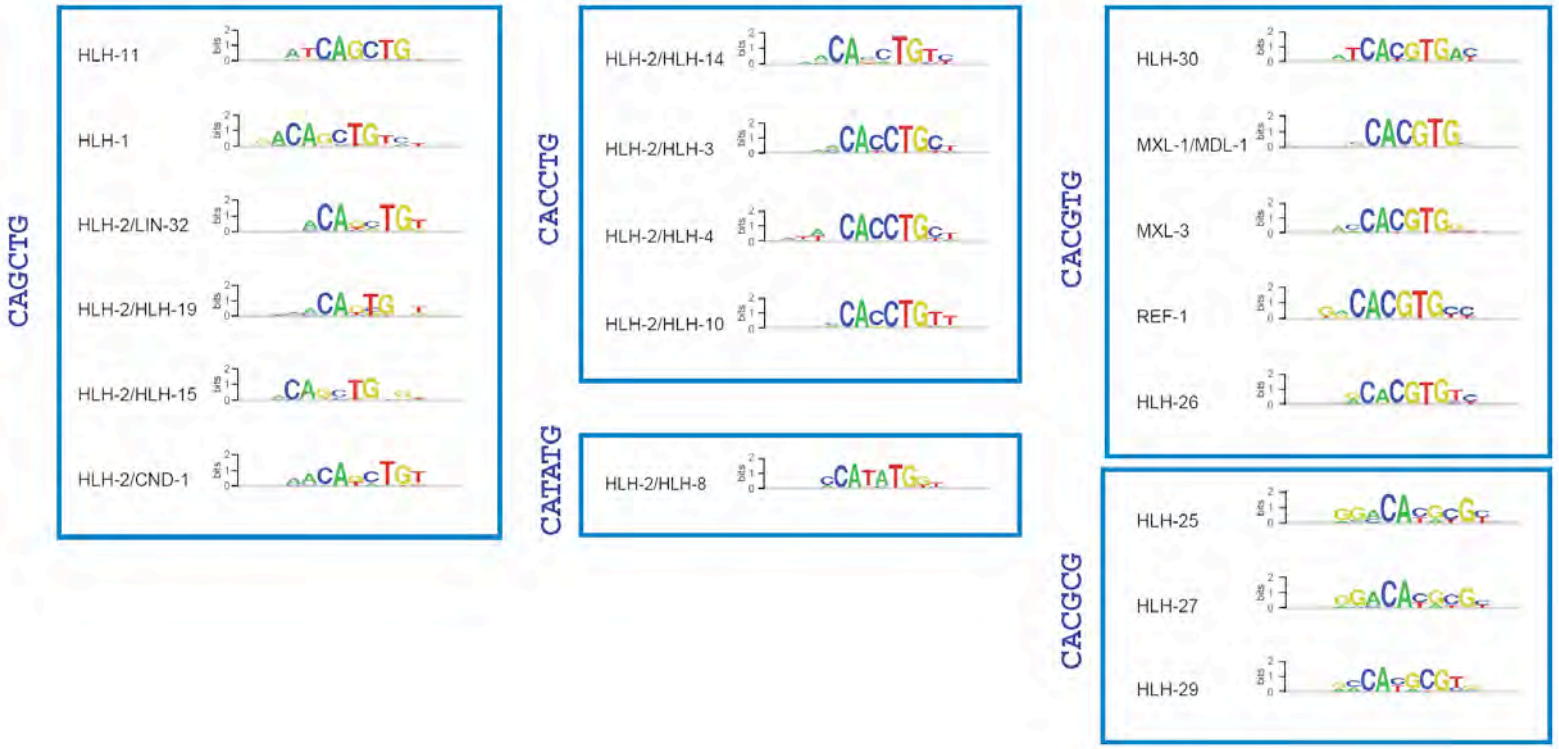


Figure IV-2. Binding site logos for 19 *C. elegans* bHLH dimers
Logo representations of all position weight matrices generated from the PBM data using the Seed-and-Wobble algorithm. We can generally divide the 19 bHLH dimers into five categories of DNA binding specificity, dimers that tend to prefer CAGCTG, CATATG, CACCTG, CACGTG, or CACGCG binding sites.

Figure IV-3

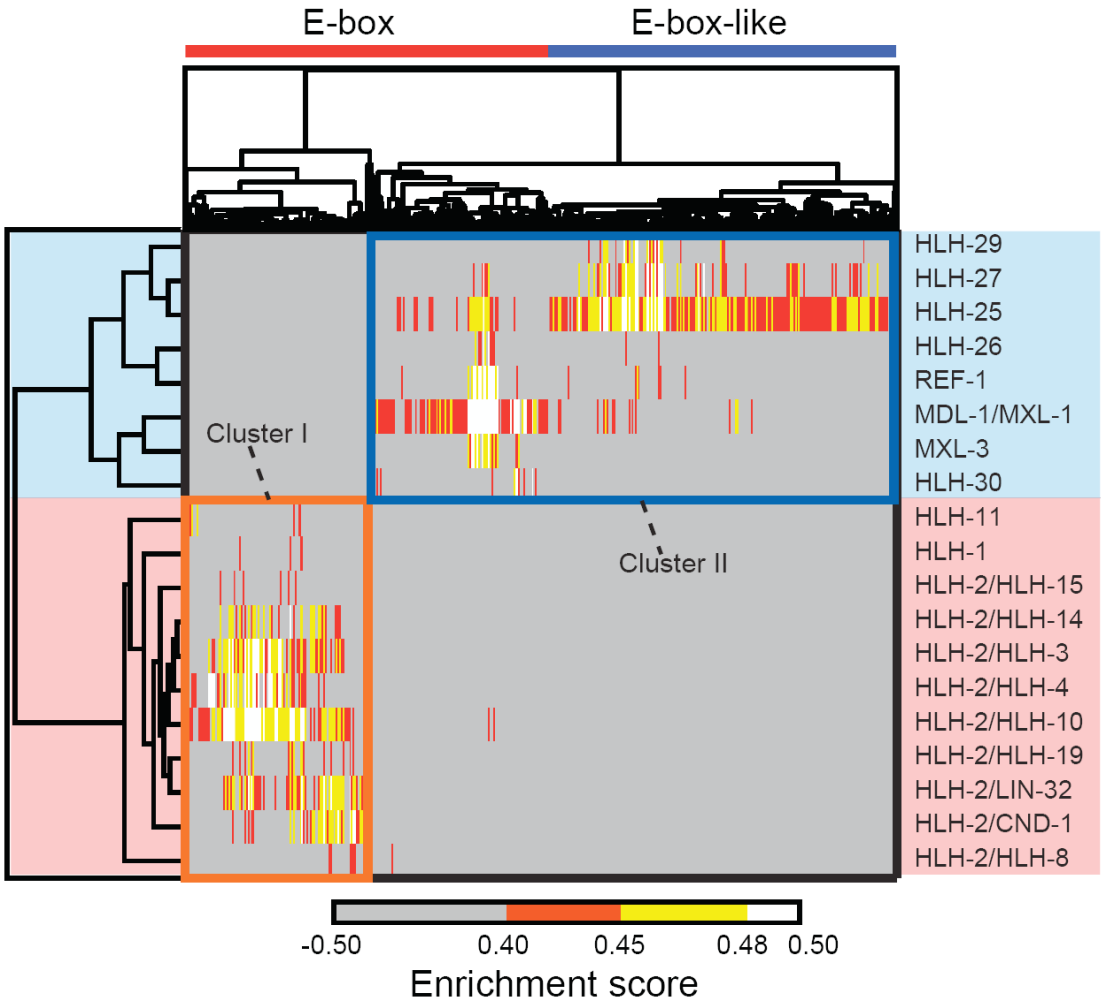


Figure IV-3. Clustergram of all bHLH dimers that yielded DNA binding profiles at a PBM ES ≥ 0.40

8-mer sequences of DNA (horizontal axis) were clustered according to sequence similarity, and bHLH dimers (vertical axis) were clustered according to similarities in DNA binding profiles. Orange box – cluster I; blue box – cluster II.

Figure IV-4

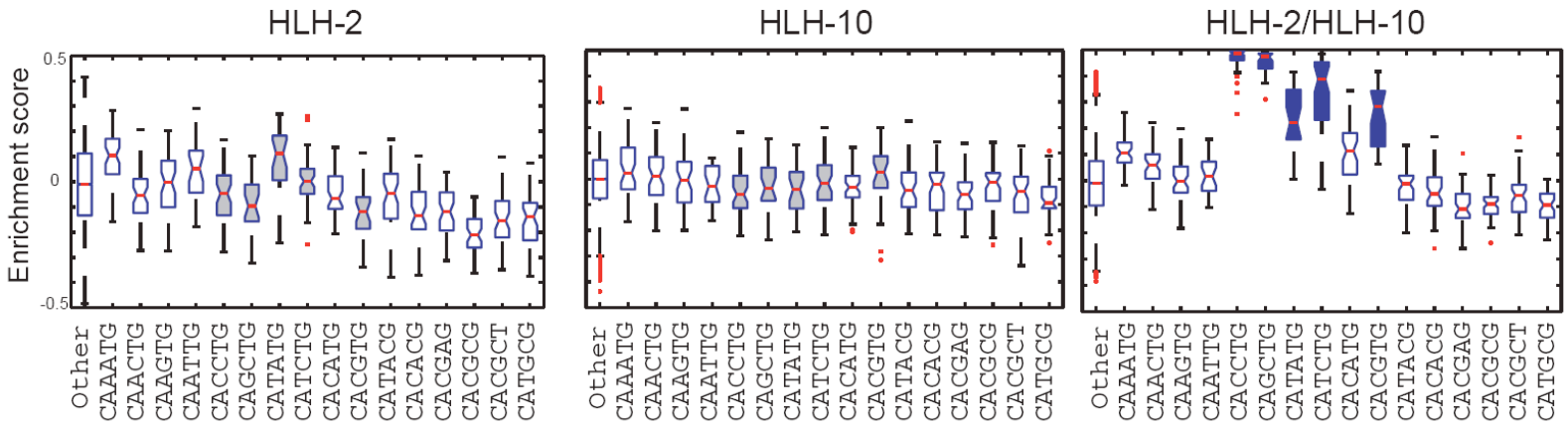


Figure IV-4. Inidividual bHLH monomers that heterodimerize do not bind DNA independently
8-mer enrichment score (ES) distributions of HLH-2, HLH-10 and HLH-2/HLH-10 binding to E-boxes and E-box-related sequences. E-boxes bound preferentially (AUC ≥ 0.85, Q < 0.001) by HLH-2/HLH-10 are indicated in blue (right panel). The corresponding E-boxes are colored gray in the single protein box plots for comparison (left and middle panel). In each box plot, the central bar indicates the median, the edges of the box indicate the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and individual points that are plotted correspond to outliers.

Figure IV-5

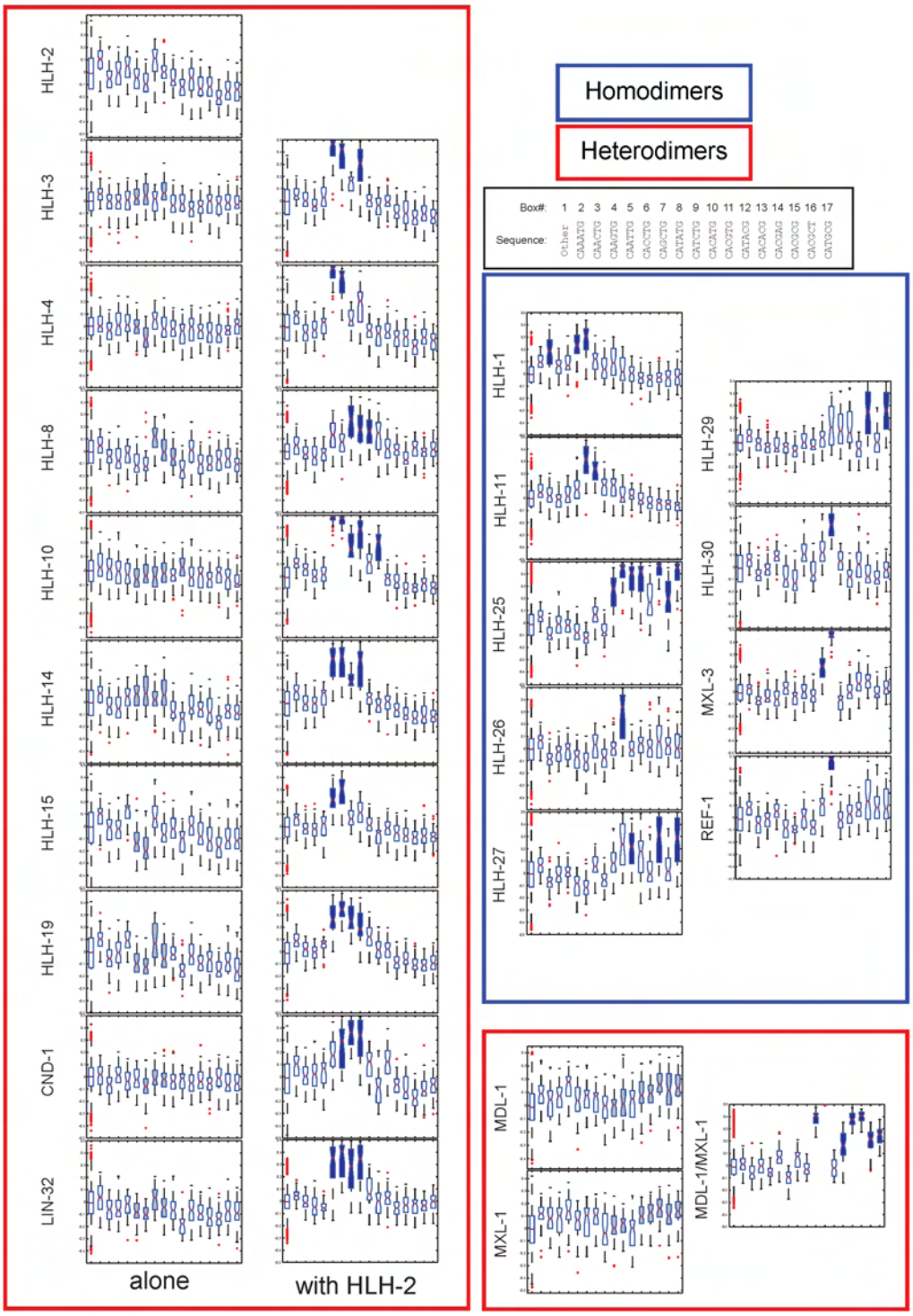

HLH-2
HLH-3
HLH-4
HLH-8
HLH-10
HLH-14
HLH-15
HLH-19
CND-1
LIN-32
alone
with HLH-2
Homodimers
Heterodimers
Box#: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Sequence: Other CAAATG CAACTG CAAGTG CAATTG CACCTG CAGCTG CATATG CATCTG CACATG CACGTG CATACG CACACG CACGAG CACGCG CACGCT CATGCG
HLH-1
HLH-11
HLH-25
HLH-26
HLH-27
HLH-29
HLH-30
MXL-3
REF-1
MDL-1
MXL-1
MDL-1/MXL-1

Figure IV-5: Box plots of all bHLH dimer-E-Box and dimer-E-Box-like combinations

Shown are graphical representations of the ES distributions of each group of 8-mers containing an E-Box or E-Box-like sequence. Dark blue indicates significant binding (AUC ≥ 0.85 and Q-val < 0.001); for heterodimers the corresponding E-Box or E-Box-like sequences are colored gray in the single protein box plots for comparison; white indicates ‘non-binding’. All box plots and error bars are as indicated in the legend of Figure IV-4.

Figure IV-6A

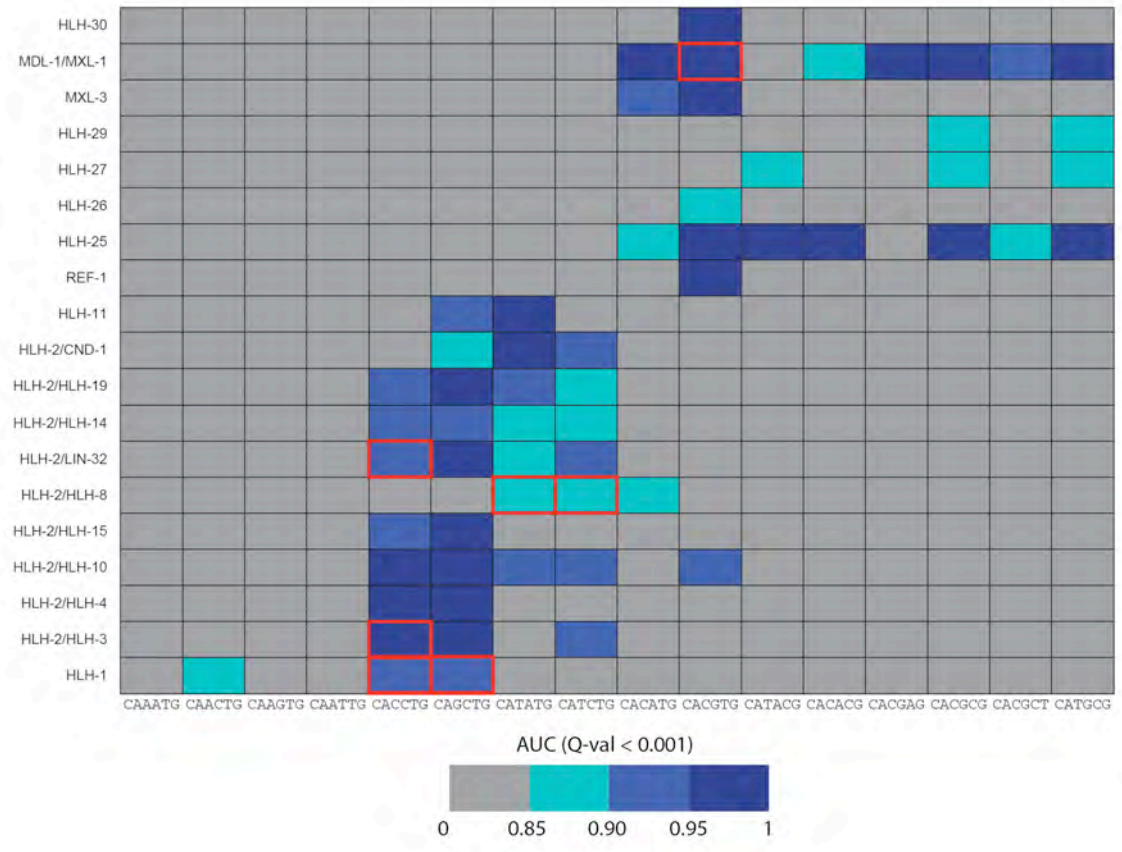


Figure IV-6A. Statistics of E-box and E-box-like sequence binding
Grid representation of the statistical significance of each bHLH dimer binding each E-box and E-box-like sequence according to their AUC and Q-values. We considered highly significant any combination showing a Q-value < 0.001 and AUC ≥ 0.85. The color scheme shows increasing significance according to AUC value. Red boxes indicate previously reported E-Box binding sites for the respective bHLH dimers.

Figure IV-6B

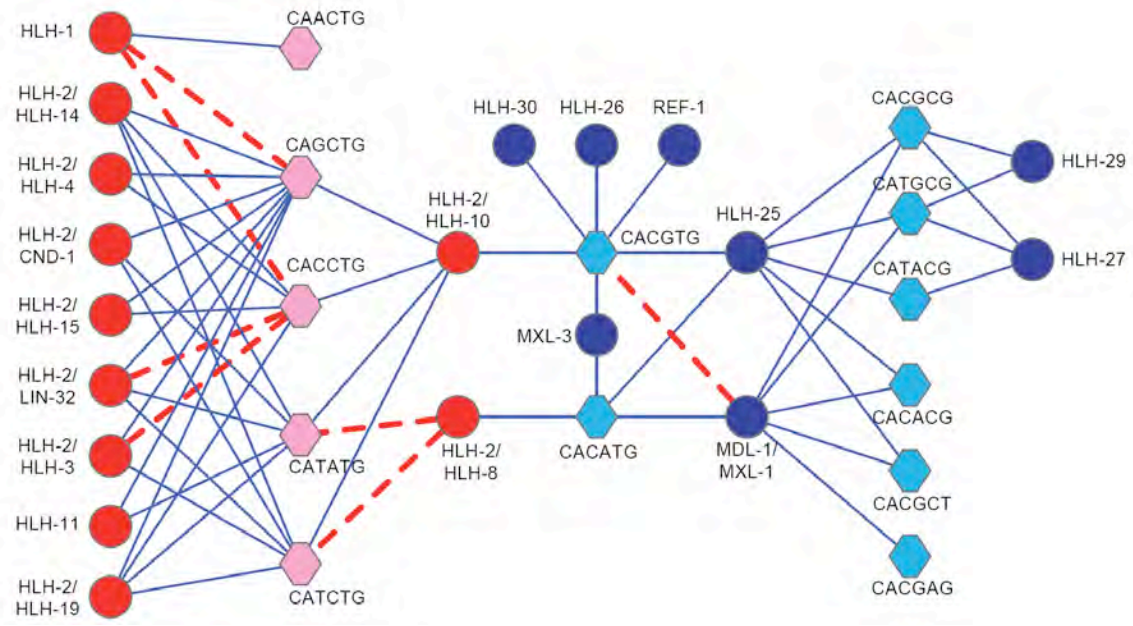
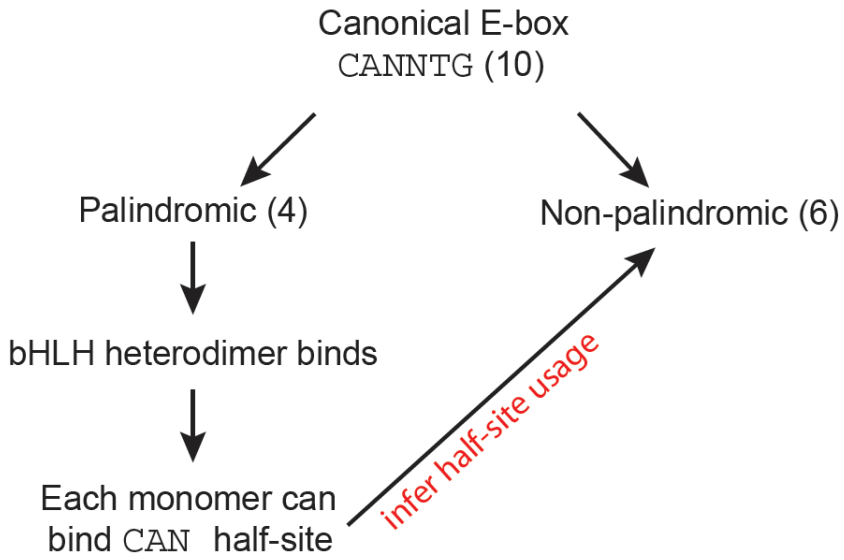


Figure IV-6B. The *C. elegans* bHLH dimer/binding site network
bHLH dimers are indicated in circles, E-Box and E-box-like sequences are indicated in hexagons. Red – cluster I; blue – cluster II. Blue lines – novel interactions; dashed red lines – previously reported interactions.

Figure IV-7


Canonical E-box
CANNTG (10)
Palindromic (4)
Non-palindromic (6)
bHLH heterodimer binds
Each monomer can
bind CAN half-site
infer half-site usage

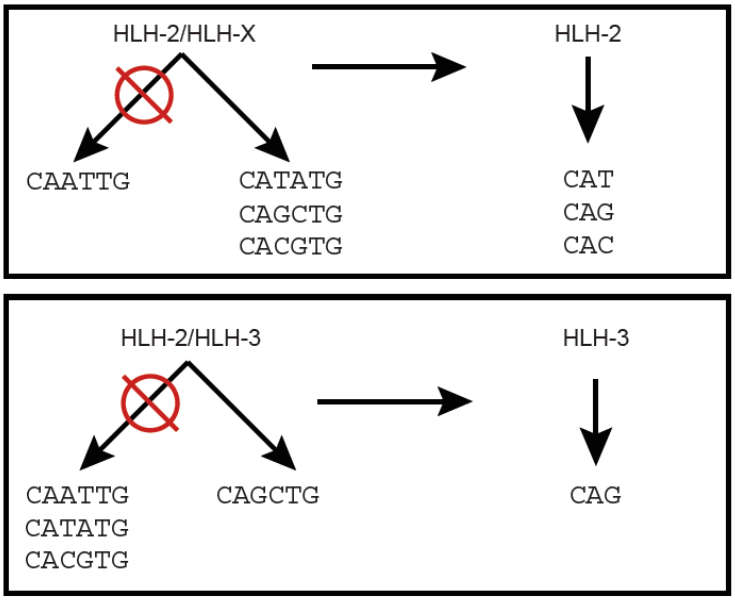

HLH-2/HLH-X
HLH-2
CAATTG
CATATG
CAGCTG
CACGTG
CAT
CAG
CAC
HLH-2/HLH-3
HLH-3
CAATTG
CATATG
CACGTG
CAGCTG
CAG
HLH-3
CAG GTG
GTC CAC
>
CAG ATG
GTC TAC
~
CAG CTG
GTC GAC
HLH-2

Figure IV-7. Half-site logic scheme
The general scheme of deducing half-site preferences for bHLH proteins is shown. There are 10 possible E-boxes (when redundant reverse complements are removed), four of which are palindromes (CAATTG, CACGTG, CAGCTG, and CATATG). By noting which palindromic E-Box each bHLH dimer binds to, we can deduce which half-sites each monomer of a heterodimer must be capable of binding. Once likely half-site preferences have been deduced, we can examine bHLH dimer binding to the six remaining non-palindromic E-Boxes to see if we can infer which monomer of a dimer specifies which half-site. In the example provided, the HLH-2 heterodimers can bind to CACGTG, CAGCTG, or CATATG palindromic E-Boxes, but not to CAATTG palindromic E-Boxes. This suggests that HLH-2 must be able to bind the CAC, CAG, and CAT half-sites, respectively. We then notice that the HLH-2/HLH-3 heterodimer only binds to the CAGCTG palindromic E-Box, indicating that HLH-3 can bind to the CAG half-site, but maybe not CAC or CAT. Indeed, the only E-Boxes bound by the HLH-2/HLH-3 heterodimer are CAGCTG, CAGGTG, and CAGATG, perhaps suggesting that HLH-3 binds to the CAG half-site in all three cases, whereas HLH-2 may bind the CAG, CAC, and CAT half-sites, respectively.

Figure IV-8

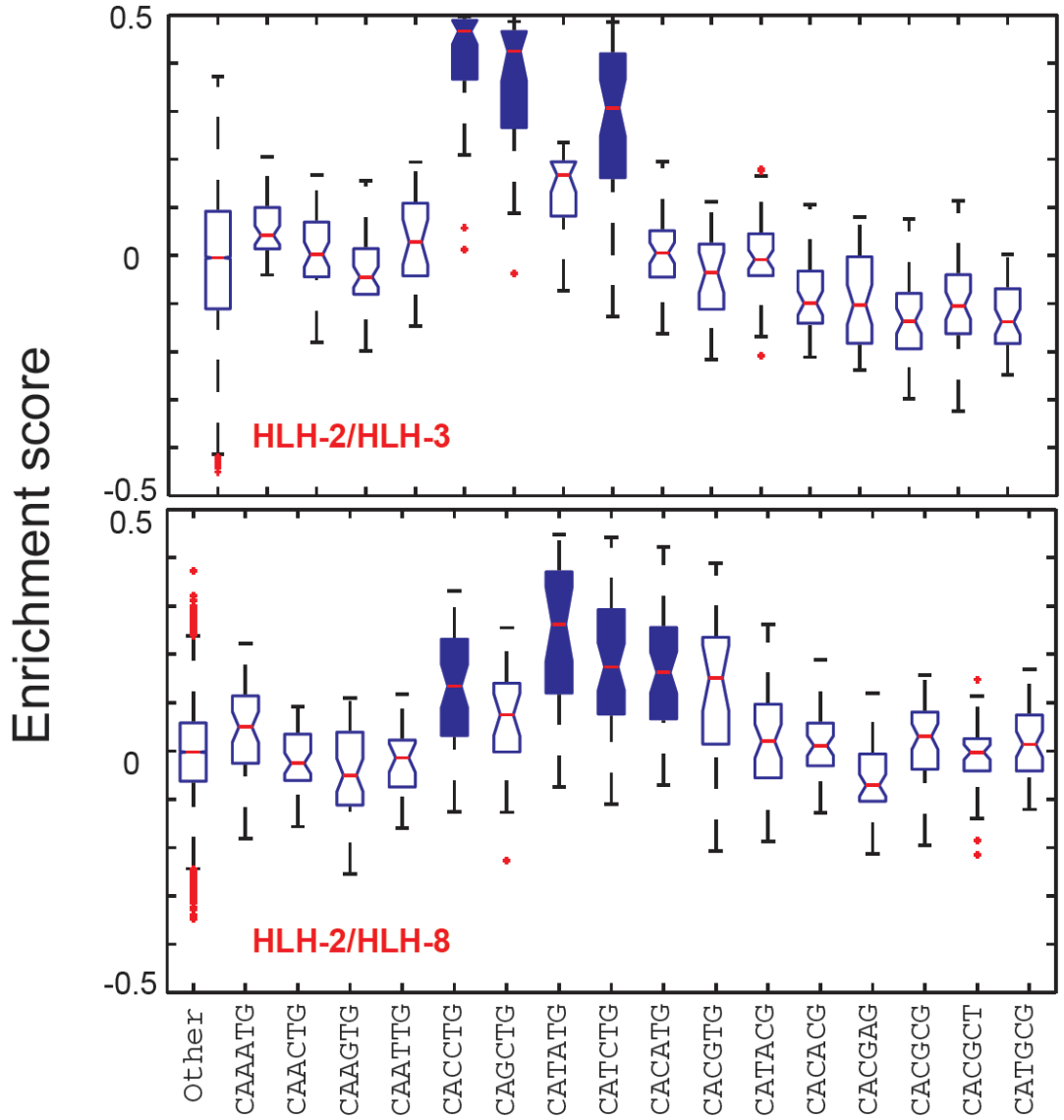


Figure IV-8. Deducing half-site preference for HLH-3 and HLH-8
HLH-2/HLH-3 heterodimers most prefer binding to the CAGCTG palindromic E-Box (i.e. two CAG half-sites), and bind to other E-Boxes containing CAG half sites. HLH-2/HLH-8, on the other hand most prefers binding to the CATATG palindromic E-Box (two CAT half-sites), and binds two other CAT containing E-Boxes. Our model suggests that CAG half-sites are specified by the HLH-3 monomer in HLH-2/HLH-3 heterodimers and the CAT half-site is specified by the HLH-8 monomer in HLH-2/HLH-8 heterodimers.

Table IV-3

bHLH protein	palindrome specified	half-site specified
HLH-30	CACGTG	CAC
MDL-1	CACGTG	CAC
MXL-1	CACGTG	CAC
MXL-3	CACGTG	CAC
HLH-29	none	NA
HLH-27	none	NA
HLH-26	CACGTG	CAC
HLH-25	CACGTG	CAC
REF-1	CACGTG	CAC
HLH-11	CATATG; CAGCTG	CAT; CAG
CND-1	CATATG; CAGCTG	CAT; CAG
HLH-19	CATATG; CAGCTG	CAT; CAG
HLH-14	CATATG; CAGCTG	CAT; CAG
LIN-32	CATATG; CAGCTG	CAT; CAG
HLH-8	CATATG	CAT
HLH-15	CAGCTG	CAG
HLH-10	CACGTG; CATATG; CAGCTG	CAC; CAT; CAG
HLH-4	CAGCTG	CAG
HLH-3	CAGCTG	CAG
HLH-1	CAGCTG	CAG
HLH-2	CACGTG; CATATG; CAGCTG	CAC; CAT; CAG

Table IV-3. List of predicted half sites specified by each individual bHLH protein

Figure IV-9

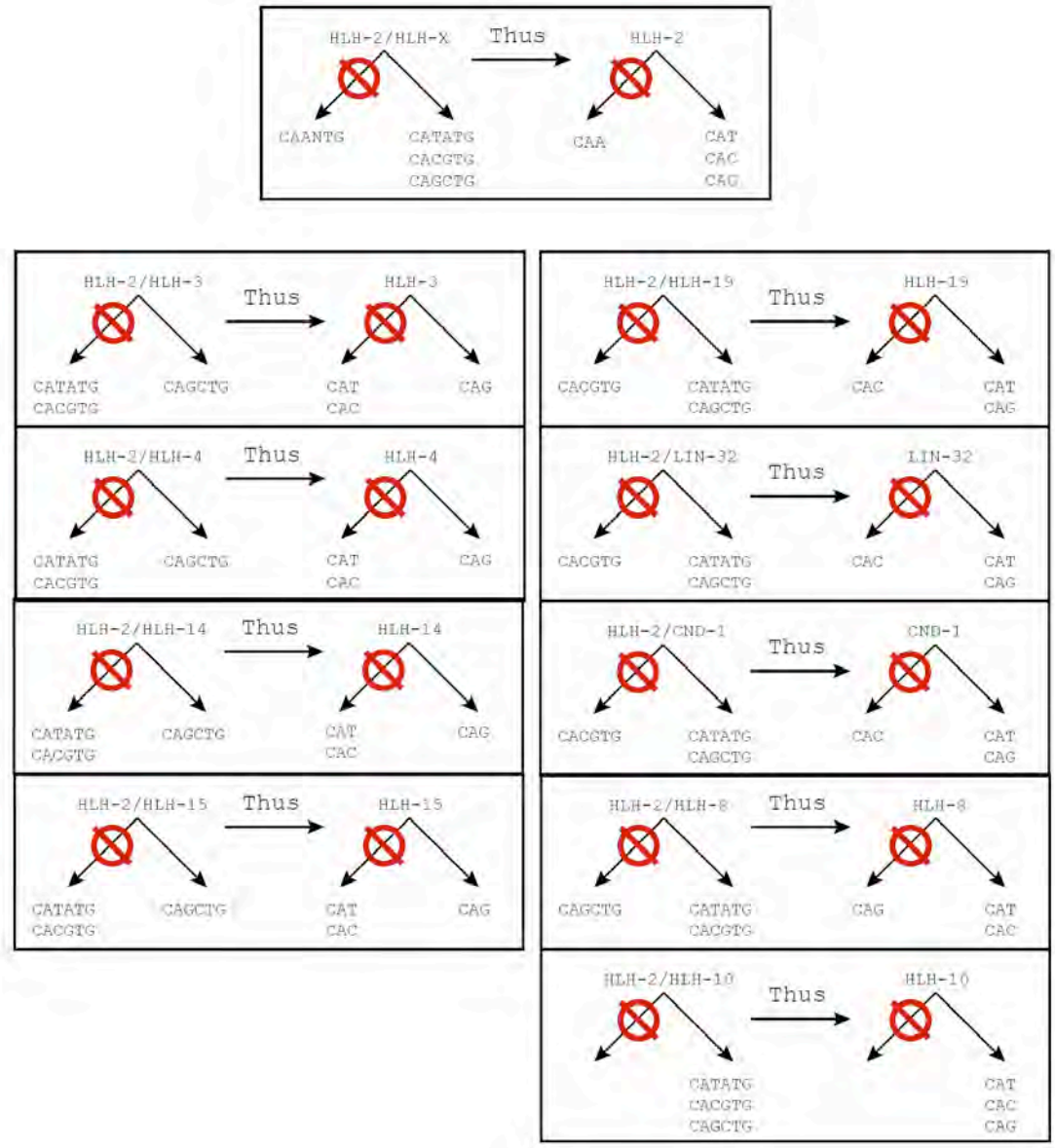


Figure IV-9. Half-site preference prediction for HLH-2 and partners

Figure IV-10

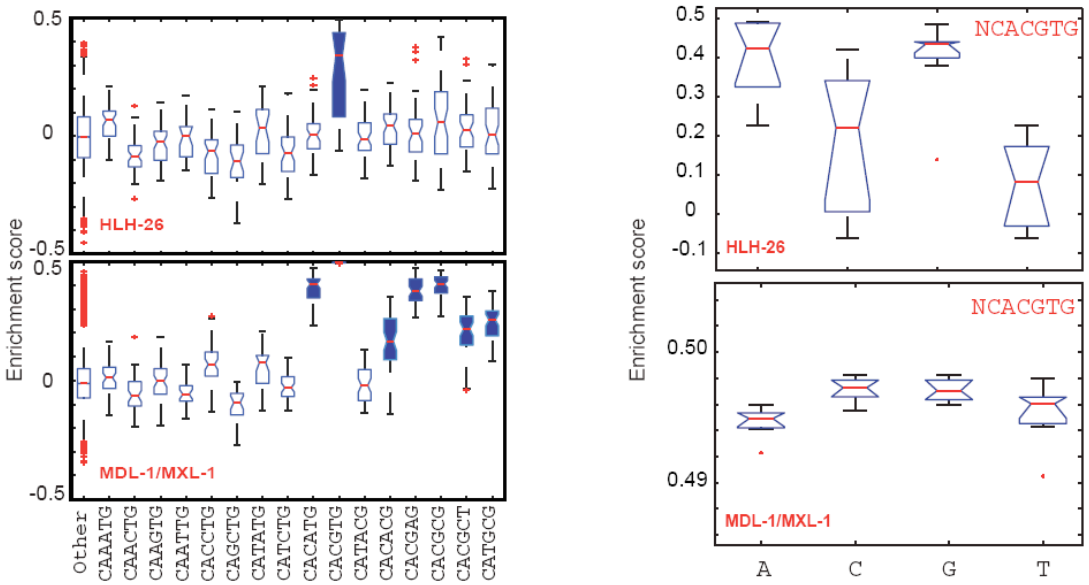


Figure IV-10. Variation in ES distribution suggests flanking nucleotide affects

(Left) Shown are box plots of 8-mer ES distributions for the 6-mers indicated at the bottom of the panel (as in Figure IV-4 and IV-5). Whereas HLH-26 binds to the CACGTG E-Box with a wide variation in relative affinity, MDL-1/MXL-1 heterodimers appear to bind CACGTG with a very narrow range of ES distribution (note that the box plot is barely visible due to such high ES), suggesting the importance of additional nucleotides flanking the E-Box when bound by HLH-26, but not MDL-1/MXL-1. (Right) If we break up the 8-mer ES distributions to represent 7-mers with the indicated fixed 5' flanking nucleotide, we find that a 5' flanking cytosine or thymine can be inhibitory to CACGTG binding by the HLH-26 homodimer. MDL-1/MXL-1, on the other hand, is ambivalent towards flanking nucleotides, as indicated by the nearly equivalent ES distributions for each flank (note the scale of each panel). All box plots and error bars are as indicated in the legend of Figure IV-4.

Figure IV-11A

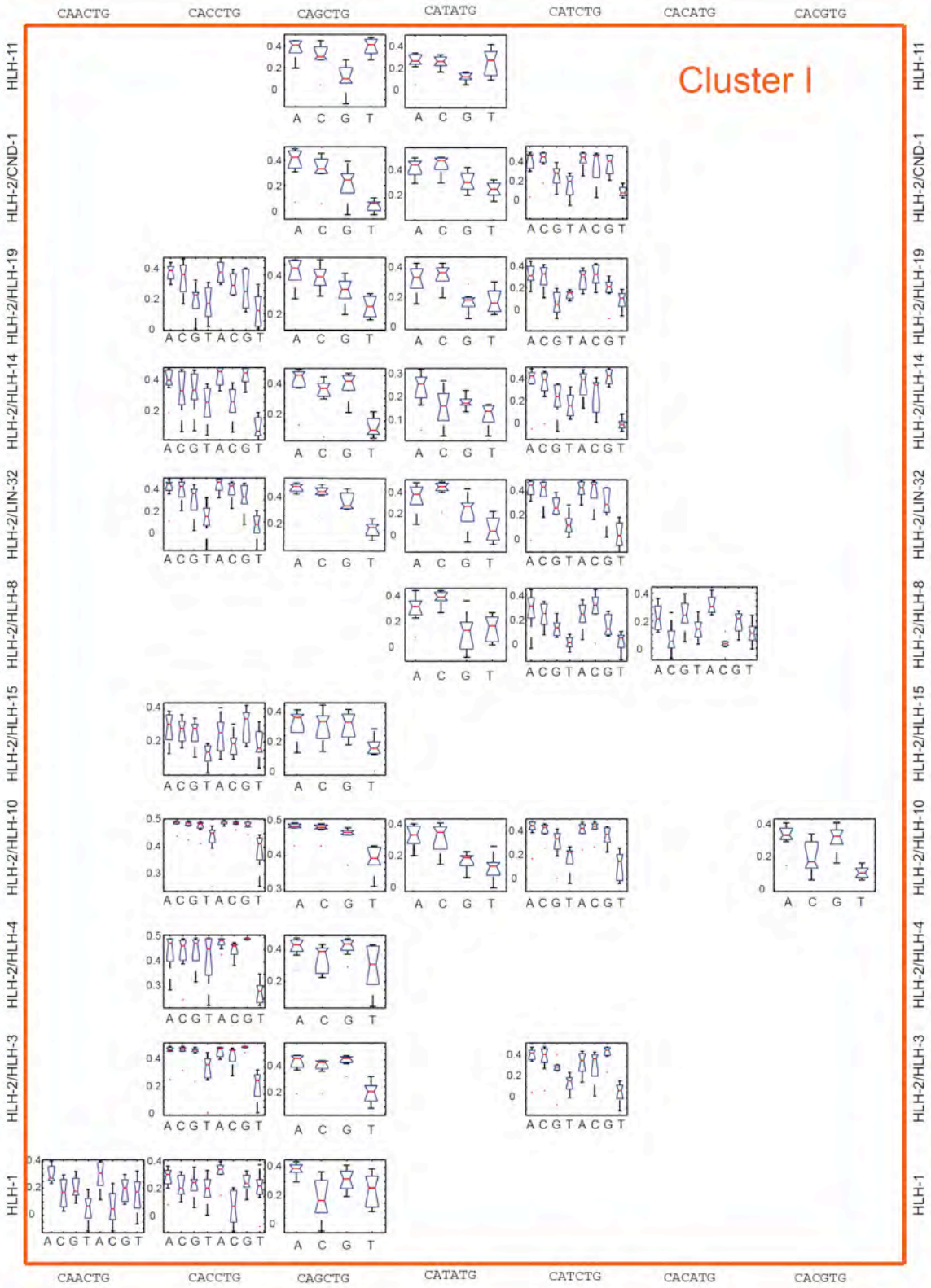


Figure IV-11A. Flanking nucleotide preference for Cluster I dimers
ES box plots for 7-mers having the indicated 5' flanking nucleotides adjacent to indicated E-boxes and E-box-like sequences. Dimers are represented in rows; sequences are represented in columns (as in Figure IV-6A). Flanking nucleotides are indicated below each box plot. Because we cannot assign a 5' flanking nucleotide to a particular side of a palindromic E-box, we merged all data for 5' flanks of palindromic E-boxes into four box plots. All non-palindromic sites are represented by eight box plots: the four left box plots for flanks of the site indicated at the top and bottom of the figure, and the four right box plots for flanks of the reverse complement (as in Figure IV-12). Note that Y-axis scales are different for each series of box plots. All box plots and error bars are as indicated in the legend of Figure IV-4.

Figure IV-11B

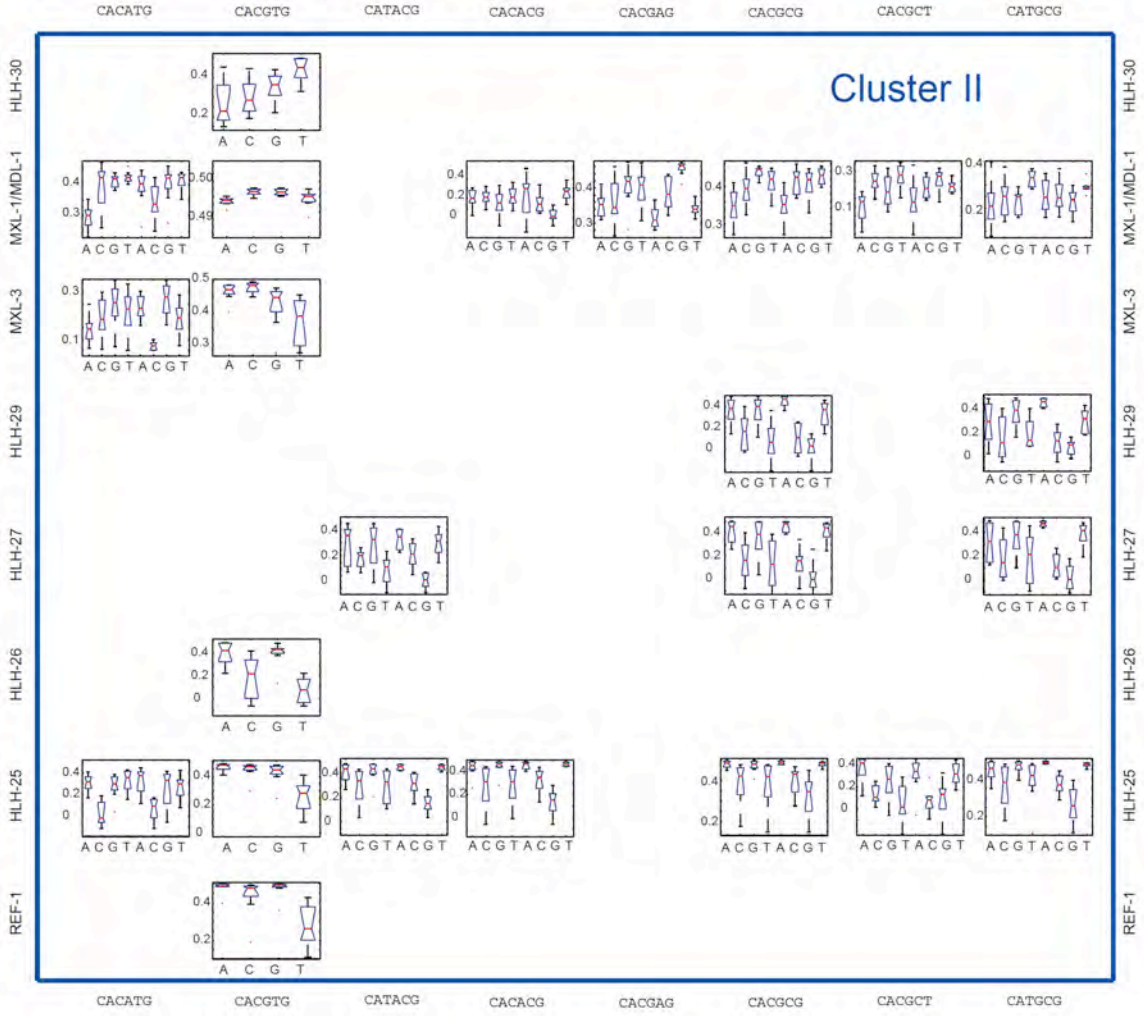


Figure IV-11B. Flanking nucleotide preference for Cluster II dimers
Same as Figure IV-11A

Figure IV-12

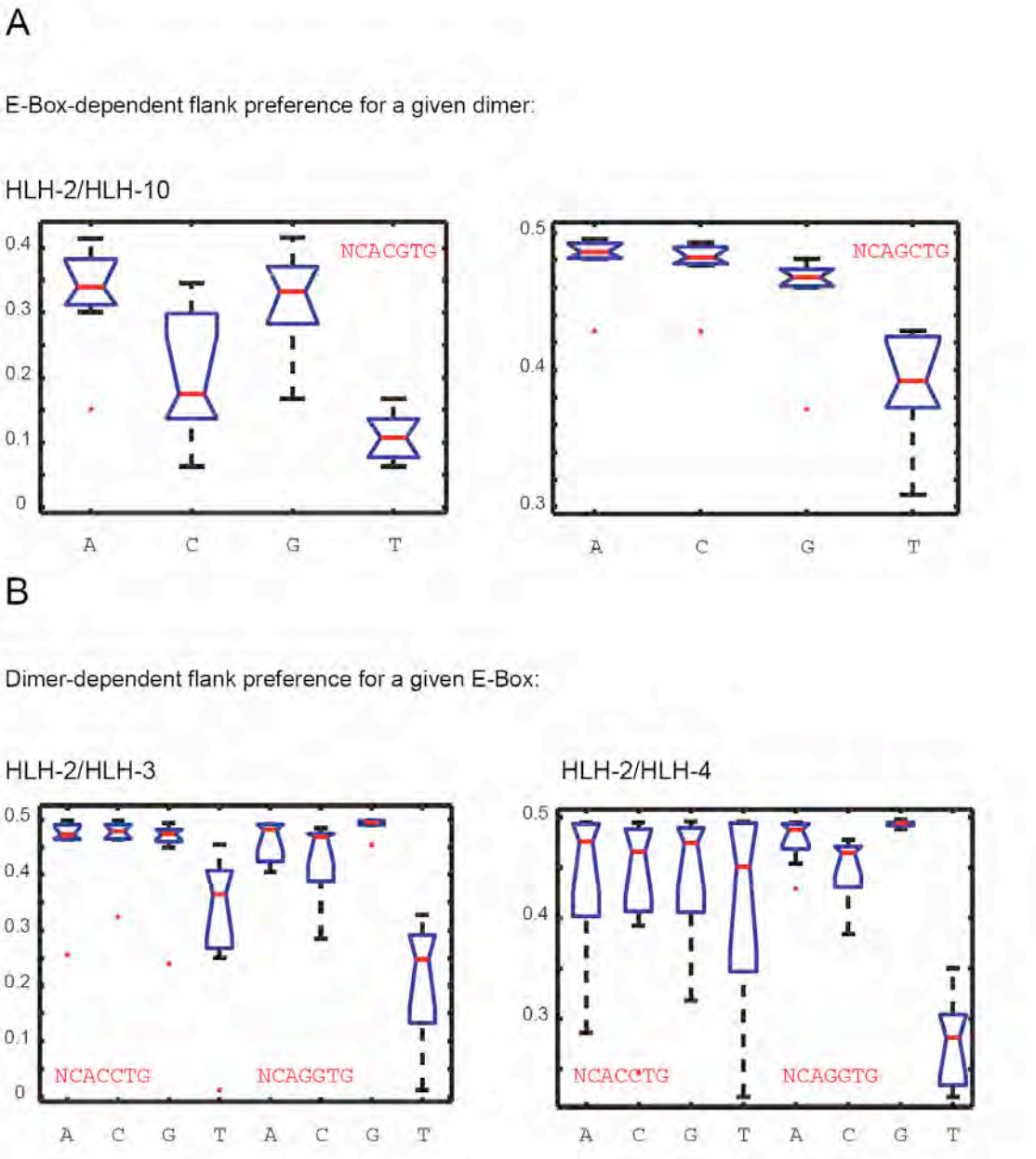

A
E-Box-dependent flank preference for a given dimer:
HLH-2/HLH-10
NCACGTG
NCAGCTG
A C G T
B
Dimer-dependent flank preference for a given E-Box:
HLH-2/HLH-3
HLH-2/HLH-4
NCACCTG
NCAGGTG
A C G T A C G T

Figure IV-12. Influence of flanking nucleotides on bHLH dimer binding
(A) Enrichment score distributions for the HLH-2/HLH-10 heterodimer bound to CACGTG (left) and CAGCTG (right) with respect to 5' flanking nucleotides (indicated at the bottom of each panel). The influence of flanking nucleotides on HLH-2/HLH-10 binding is dependent on the core E-Box bound. (B) Enrichment score distributions for the CACCTG/CAGGTG E-Box bound by HLH-2/HLH-3 heterodimers (left) and HLH-2/HLH-4 heterodimers (right). These two heterodimers bind to the same E-Box binding site with different relative affinities for flanking nucleotides. All box plots and error bars are as indicated in the legend of Figure IV-4.

Figure IV-13

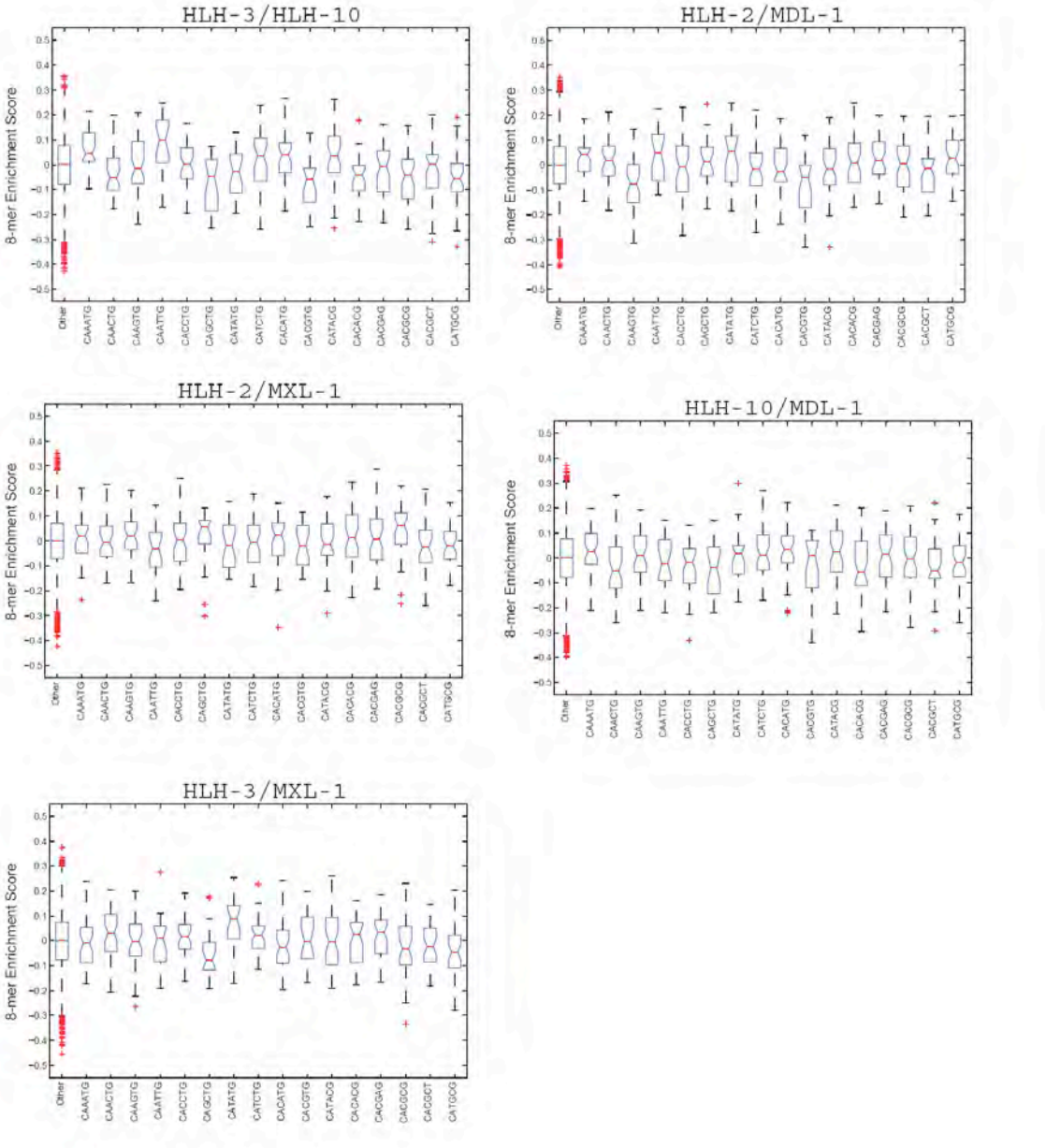


Figure IV-13. Negative control PBM experiments with bHLH-bHLH pairs that do not dimerize in Y2H assays

Box plot representation of the E-box and E-box-like sequence ES distributions for PBM experiments carried out with bHLH pairs that did not dimerize in Y2H assays.

Figure IV-14

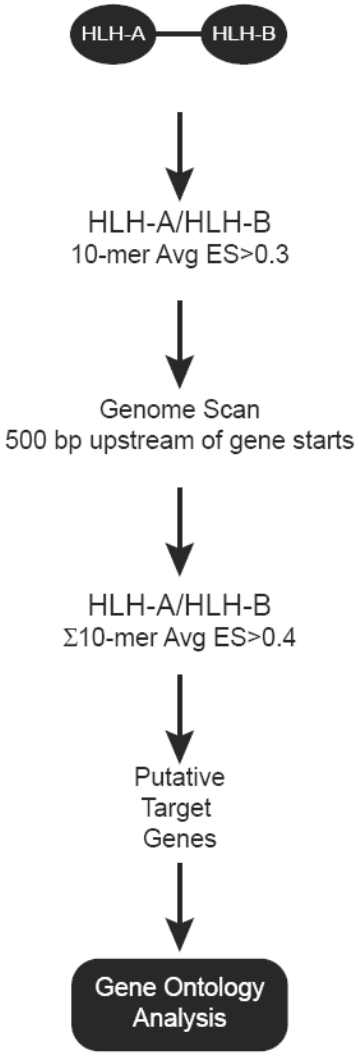


Figure IV-14. Pipeline for predicting bHLH dimer target genes
Flow diagram describing how GO annotations were obtained (see Materials and Methods for details).

Table IV-4. bHLH dimer candidate target gene lists (Due to space limitations, see electronic file for data)

Table IV-5. GO categories enriched among bHLH dimer candidate target genes

HLH-2/CND-1

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005488_binding	5055	457	-2.704494
GO:0005575_cellular_component	4970	444	-2.161453
GO:0009987_cellular_process	4977	438	-1.644553
GO:0005623_cell	4804	434	-2.491572
GO:0044464_cell_part	4788	432	-2.429721
GO:0008152_metabolic_process	4176	372	-1.675568
GO:0032501_multicellular_organismal_process	3705	336	-1.944052
GO:0007275_multicellular_organismal_development	3428	308	-1.573588
GO:0044237_cellular_metabolic_process	3372	304	-1.627429
GO:0044238_primary_metabolic_process	3216	290	-1.562883
GO:0005622_intracellular	2808	277	-3.789383
GO:0043170_macromolecule_metabolic_process	2663	255	-2.673899
GO:0009790_embryonic_development	2693	247	-1.66762
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	240	-1.436608
GO:0044424_intracellular_part	2280	225	-3.05959
GO:0043226_organelle	1749	179	-3.204146
GO:0043229_intracellular_organelle	1740	178	-3.177865
GO:0043231_intracellular_membrane-bounded_organelle	1392	155	-4.571543
GO:0043227_membrane-bounded_organelle	1399	155	-4.455018
GO:0003676_nucleic_acid_binding	1510	151	-2.346986
GO:0005634_nucleus	1121	132	-5.099949
GO:0006139_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	1228	128	-2.665592
GO:0040011_locomotion	1157	113	-1.58267
GO:0010467_gene_expression	1101	109	-1.692772
GO:0050794_regulation_of_cellular_process	1029	103	-1.745097
GO:0003677_DNA_binding	873	98	-3.136263
GO:0009653_anatomical_structure_morphogenesis	879	87	-1.441118
GO:0010468_regulation_of_gene_expression	760	85	-2.736186
GO:0016043_cellular_component_organization_and_biogenesis	804	85	-2.073686
GO:0016070_RNA_metabolic_process	837	83	-1.412752
GO:0032991_macromolecular_complex	721	80	-2.51094
GO:0031323_regulation_of_cellular_metabolic_process	755	80	-2.003723
GO:0019222_regulation_of_metabolic_process	783	80	-1.644839
GO:0006350_transcription	747	79	-1.965138
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	724	78	-2.137655
GO:0045449_regulation_of_transcription	709	77	-2.201273
GO:0030528_transcription_regulator_activity	647	71	-2.172277
GO:0043234_protein_complex	549	70	-3.90039
GO:0044422_organelle_part	553	70	-3.804085
GO:0044446_intracellular_organelle_part	546	69	-3.735784
GO:0043228_non-membrane-bounded_organelle	535	61	-2.286223
GO:0043232_intracellular_non-membrane-bounded_organelle	535	61	-2.286223
GO:0006996_organelle_organization_and_biogenesis	418	50	-2.361946
GO:0007610_behavior	389	43	-1.561443

GO:0005694_chromosome	200	38	-6.111269
GO:0022607_cellular_component_assembly	202	37	-5.569751
GO:0044427_chromosomal_part	156	36	-8.056878
GO:0051276_chromosome_organization_and_biogenesis	161	34	-6.644436
GO:0005216_ion_channel_activity	288	34	-1.698373
GO:0022838_substrate_specific_channel_activity	288	34	-1.698373
GO:0015267_channel_activity	290	34	-1.658314
GO:0022803_passive_transmembrane_transporter_activity	290	34	-1.658314
GO:0000785_chromatin	123	33	-9.213668
GO:0006333_chromatin_assembly_or_disassembly	111	32	-9.830051
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	32	-8.880766
GO:0065003_macromolecular_complex_assembly	178	32	-4.728201
GO:0031497_chromatin_assembly	89	28	-9.653004
GO:0006323_DNA_packaging	96	28	-8.806505
GO:0065004_protein-DNA_complex_assembly	104	28	-7.94889
GO:0006334_nucleosome_assembly	85	27	-9.43661
GO:0000786_nucleosome	86	27	-9.307694
GO:0006629_lipid_metabolic_process	233	27	-1.365751
GO:0046873_metal_ion_transmembrane_transporter_activity	206	25	-1.504615
GO:0030001_metal_ion_transport	215	25	-1.314771
GO:0005261_cation_channel_activity	172	23	-1.86649
GO:0005525_GTP_binding	161	20	-1.389152
GO:0032561_guanyl_ribonucleotide_binding	161	20	-1.389152
GO:0044255_cellular_lipid_metabolic_process	118	18	-2.121962
GO:0006813_potassium_ion_transport	115	16	-1.591345
GO:0005267_potassium_channel_activity	108	15	-1.516194
GO:0016810_hydrolase_activity__acting_on_carbon-nitrogen_(but_not_peptide)_bonds	59	10	-1.679823
GO:0006631_fatty_acid_metabolic_process	36	9	-2.692298
GO:0032787_monocarboxylic_acid_metabolic_process	51	9	-1.662792
GO:0016071_mRNA_metabolic_process	52	9	-1.611933
GO:0001708_cell_fate_specification	57	9	-1.381327
GO:0048598_embryonic_morphogenesis	33	7	-1.797955
GO:0007517_muscle_development	34	7	-1.728819
GO:0008134_transcription_factor_binding	38	7	-1.481863
GO:0045138_tail_tip_morphogenesis	30	6	-1.486068
GO:0005777_peroxisome	17	5	-2.002492
GO:0042579_microbody	17	5	-2.002492
GO:0042692_muscle_cell_differentiation	18	5	-1.890973
GO:0003997_acyl-CoA_oxidase_activity	8	4	-2.615709
GO:0006635_fatty_acid_beta-oxidation	10	4	-2.196601
GO:0009062_fatty_acid_catabolic_process	10	4	-2.196601
GO:0016054_organic_acid_catabolic_process	10	4	-2.196601
GO:0016634_oxidoreductase_activity__acting_on_the_CH-CH_group_of_donors__oxygen_as_acceptor	10	4	-2.196601
GO:0019395_fatty_acid_oxidation	10	4	-2.196601
GO:0046395_carboxylic_acid_catabolic_process	10	4	-2.196601
GO:0007519_skeletal_muscle_development	13	4	-1.750832
GO:0014706_striated_muscle_development	13	4	-1.750832
GO:0048741_skeletal_muscle_fiber_development	13	4	-1.750832
GO:0048747_muscle_fiber_development	13	4	-1.750832
GO:0044242_cellular_lipid_catabolic_process	14	4	-1.633283
GO:0016812_hydrolase_activity__acting_on_carbon-	16	4	-1.430509

nitrogen_(but_not_peptide)_bonds__in_cyclic_amides			
GO:0044463_cell_projection_part	8	3	-1.645371
GO:0040040_thermosensory_behavior	9	3	-1.496191

HLH-1

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0009790_embryonic_development	2693	29	-1.654142
GO:0040007_growth	2387	28	-2.094258
GO:0043231_intracellular_membrane-bounded_organelle	1392	17	-1.517205
GO:0043227_membrane-bounded_organelle	1399	17	-1.498714
GO:0005634_nucleus	1121	15	-1.691271
GO:0044446_intracellular_organelle_part	546	9	-1.654586
GO:0043234_protein_complex	549	9	-1.640948
GO:0044422_organelle_part	553	9	-1.622947
GO:0043228_non-membrane-bounded_organelle	535	8	-1.305504
GO:0043232_intracellular_non-membrane-bounded_organelle	535	8	-1.305504
GO:0000785_chromatin	123	7	-4.407705
GO:0044427_chromosomal_part	156	7	-3.756051
GO:0005694_chromosome	200	7	-3.103464
GO:0006334_nucleosome_assembly	85	5	-3.343131
GO:0000786_nucleosome	86	5	-3.319684
GO:0031497_chromatin_assembly	89	5	-3.251161
GO:0006323_DNA_packaging	96	5	-3.101028
GO:0065004_protein-DNA_complex_assembly	104	5	-2.944127
GO:0006333_chromatin_assembly_or_disassembly	111	5	-2.817888
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	5	-2.6686
GO:0051276_chromosome_organization_and_biogenesis	161	5	-2.125467
GO:0065003_macromolecular_complex_assembly	178	5	-1.948036
GO:0008237_metallopeptidase_activity	192	5	-1.817287
GO:0022607_cellular_component_assembly	202	5	-1.731187
GO:0006629_lipid_metabolic_process	233	5	-1.496402
GO:0004175_endopeptidase_activity	258	5	-1.335978
GO:0004222_metalloendopeptidase_activity	120	4	-1.885649
GO:0006816_calcium_ion_transport	19	2	-2.048368
GO:0008652_amino_acid_biosynthetic_process	22	2	-1.924135
GO:0015674_di-__tri-valent_inorganic_cation_transport	23	2	-1.886752
GO:0009309_amine_biosynthetic_process	31	2	-1.639376
GO:0044271_nitrogen_compound_biosynthetic_process	31	2	-1.639376
GO:0004245_neprilysin_activity	36	2	-1.518046
GO:0009219_pyrimidine_deoxyribonucleotide_metabolic_process	5	1	-1.427393
GO:0030658_transport_vesicle_membrane	5	1	-1.427393
GO:0030660_Golgi-associated_vesicle_membrane	5	1	-1.427393
GO:0005125_cytokine_activity	6	1	-1.349835
GO:0006694_steroid_biosynthetic_process	6	1	-1.349835
GO:0008060_ARF_GTPase_activator_activity	6	1	-1.349835
GO:0032312_regulation_of_ARF_GTPase_activity	6	1	-1.349835
GO:0032934_sterol_binding	6	1	-1.349835
GO:0040032_post-embryonic_body_morphogenesis	6	1	-1.349835

HLH-2/HLH-3

GO CATEGORY	TOTAL	TARGET	LOG10(p)

	GENES	GENES	
GO:0005575_cellular_component	4970	786	-7.176496
GO:0005623_cell	4804	761	-6.948169
GO:0044464_cell_part	4788	758	-6.850771
GO:0032501_multicellular_organismal_process	3705	553	-1.845643
GO:0005622_intracellular	2808	435	-2.564235
GO:0009790_embryonic_development	2693	404	-1.508977
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	397	-1.499263
GO:0016020_membrane	2210	356	-3.293018
GO:0044424_intracellular_part	2280	349	-1.787246
GO:0043229_intracellular_organelle	1740	273	-1.96183
GO:0043226_organelle	1749	273	-1.853084
GO:0003676_nucleic_acid_binding	1510	254	-3.515542
GO:0043231_intracellular_membrane-bounded_organelle	1392	242	-4.294566
GO:0043227_membrane-bounded_organelle	1399	242	-4.145268
GO:0044425_membrane_part	1452	232	-2.071235
GO:0005634_nucleus	1121	207	-5.403964
GO:0031224_intrinsic_to_membrane	1270	203	-1.871259
GO:0016021_integral_to_membrane	1261	199	-1.631515
GO:0004871_signal_transducer_activity	1141	193	-2.873047
GO:0060089_molecular_transducer_activity	1141	193	-2.873047
GO:0004872_receptor_activity	1061	188	-3.858788
GO:0003677_DNA_binding	873	164	-4.771814
GO:0004888_transmembrane_receptor_activity	740	145	-5.24884
GO:0016043_cellular_component_organization_and_biogenesis	804	138	-2.414859
GO:0032991_macromolecular_complex	721	134	-3.765392
GO:0016070_RNA_metabolic_process	837	134	-1.410078
GO:0010468_regulation_of_gene_expression	760	126	-1.772008
GO:0031323_regulation_of_cellular_metabolic_process	755	121	-1.330397
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	724	117	-1.389919
GO:0045449_regulation_of_transcription	709	116	-1.517665
GO:0043234_protein_complex	549	114	-5.409836
GO:0032774_RNA_biosynthetic_process	646	108	-1.684094
GO:0006351_transcription__DNA-dependent	643	107	-1.620811
GO:0006355_regulation_of_transcription__DNA-dependent	628	106	-1.771823
GO:0051252_regulation_of_RNA_metabolic_process	634	106	-1.662851
GO:0044446_intracellular_organelle_part	546	105	-3.615467
GO:0044422_organelle_part	553	105	-3.404
GO:0007166_cell_surface_receptor_linked_signal_transduction	590	99	-1.62447
GO:0043228_non-membrane-bounded_organelle	535	91	-1.671474
GO:0043232_intracellular_non-membrane-bounded_organelle	535	91	-1.671474
GO:0007186_G-protein_coupled_receptor_protein_signaling_pathway	506	85	-1.475904
GO:0006996_organelle_organization_and_biogenesis	418	84	-3.627829
GO:0004930_G-protein_coupled_receptor_activity	464	78	-1.401326
GO:0001584_rhodopsin-like_receptor_activity	446	76	-1.495359
GO:0005694_chromosome	200	57	-7.369008
GO:0051276_chromosome_organization_and_biogenesis	161	55	-10.330653
GO:0022607_cellular_component_assembly	202	55	-6.414193
GO:0044427_chromosomal_part	156	54	-10.404097

GO:0003008_system_process	222	54	-4.719048
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	52	-14.568796
GO:0000785_chromatin	123	52	-14.030152
GO:0065003_macromolecular_complex_assembly	178	52	-7.160061
GO:0006333_chromatin_assembly_or_disassembly	111	51	-15.594476
GO:0006323_DNA_packaging	96	49	-17.389627
GO:0031497_chromatin_assembly	89	48	-18.350666
GO:0065004_protein-DNA_complex_assembly	104	48	-14.808971
GO:0050877_neurological_system_process	189	48	-4.782442
GO:0006334_nucleosome_assembly	85	47	-18.577576
GO:0000786_nucleosome	86	47	-18.295871
GO:0005886_plasma_membrane	245	47	-1.906753
GO:0007600_sensory_perception	128	42	-7.461293
GO:0007606_sensory_perception_of_chemical_stimulus	109	40	-8.717423
GO:0044459_plasma_membrane_part	198	38	-1.645373
GO:0006813_potassium_ion_transport	115	25	-1.854533
GO:0005267_potassium_channel_activity	108	23	-1.647311
GO:0031226_intrinsic_to_plasma_membrane	100	22	-1.750675
GO:0005887_integral_to_plasma_membrane	99	21	-1.529933
GO:0005244_voltage-gated_ion_channel_activity	74	18	-1.952278
GO:0022832_voltage-gated_channel_activity	74	18	-1.952278
GO:0022843_voltage-gated_cation_channel_activity	66	15	-1.4598
GO:0005249_voltage-gated_potassium_channel_activity	61	14	-1.426046
GO:0046983_protein_dimerization_activity	56	13	-1.39283
GO:0008076_voltage-gated_potassium_channel_complex	57	13	-1.337103
GO:0005179_hormone_activity	43	11	-1.526099
GO:0004527_exonuclease_activity	33	10	-1.93529
GO:0008134_transcription_factor_binding	38	10	-1.507279
GO:0035121_tail_morphogenesis	35	9	-1.341139
GO:0045138_tail_tip_morphogenesis	30	8	-1.321873
GO:0003712_transcription_cofactor_activity	24	7	-1.398925
GO:0004198_calcium-dependent_cysteine-type_endopeptidase_activity	15	6	-1.934236
GO:0005089_Rho_guanyl-nucleotide_exchange_factor_activity	20	6	-1.309746
GO:0007178_transmembrane_receptor_protein_serine_threonine_kinase_signaling_pathway	20	6	-1.309746
GO:0007635_chemosensory_behavior	20	6	-1.309746
GO:0035023_regulation_of_Rho_protein_signal_transduction	20	6	-1.309746
GO:0043068_positive_regulation_of_programmed_cell_death	20	6	-1.309746
GO:0009109_coenzyme_catabolic_process	13	5	-1.598197
GO:0051187_cofactor_catabolic_process	13	5	-1.598197
GO:0022411_cellular_component_disassembly	15	5	-1.332949
GO:0030512_negative_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	5	4	-2.781575
GO:0008060_ARF_GTPase_activator_activity	6	4	-2.355246
GO:0032312_regulation_of_ARF_GTPase_activity	6	4	-2.355246
GO:0006563_L-serine_metabolic_process	7	4	-2.037777
GO:0009070_serine_family_amino_acid_biosynthetic_process	7	4	-2.037777
GO:0012502_induction_of_programmed_cell_death	7	4	-2.037777
GO:0017015_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	8	4	-1.786963

GO:0007218_neuropeptide_signaling_pathway	9	4	-1.581607
GO:0007509_mesoderm_migration	9	4	-1.581607
GO:0008078_mesodermal_cell_migration	9	4	-1.581607
GO:0003746_translation_elongation_factor_activity	10	4	-1.409369
GO:0042074_cell_migration_involved_in_gastrulation	10	4	-1.409369
GO:0000097_sulfur_amino_acid_biosynthetic_process	5	3	-1.666796
GO:0006534_cysteine_metabolic_process	5	3	-1.666796
GO:0006917_induction_of_apoptosis	6	3	-1.412708
GO:0008629_induction_of_apoptosis_by_intracellular_signals	6	3	-1.412708
GO:0008630_DNA_damage_response__signal_transduction_resulting_in_induction_of_apoptosis	6	3	-1.412708

HLH-2/HLH-4

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005575_cellular_component	4970	673	-6.732784
GO:0005623_cell	4804	646	-5.784778
GO:0044464_cell_part	4788	643	-5.648466
GO:0016020_membrane	2210	323	-5.335643
GO:0044425_membrane_part	1452	224	-5.258831
GO:0003676_nucleic_acid_binding	1510	203	-1.688151
GO:0031224_intrinsic_to_membrane	1270	201	-5.475017
GO:0043231_intracellular_membrane-bounded_organelle	1392	199	-2.816067
GO:0043227_membrane-bounded_organelle	1399	199	-2.709831
GO:0016021_integral_to_membrane	1261	197	-4.993389
GO:0004871_signal_transducer_activity	1141	191	-6.923088
GO:0060089_molecular_transducer_activity	1141	191	-6.923088
GO:0004872_receptor_activity	1061	186	-8.279762
GO:0007154_cell_communication	1099	169	-3.949147
GO:0005634_nucleus	1121	166	-3.079829
GO:0007165_signal_transduction	1016	160	-4.293935
GO:0004888_transmembrane_receptor_activity	740	146	-10.069319
GO:0003677_DNA_binding	873	131	-2.738187
GO:0032991_macromolecular_complex	721	103	-1.670375
GO:0007166_cell_surface_receptor_linked_signal_transduction	590	99	-3.811668
GO:0045449_regulation_of_transcription	709	99	-1.392539
GO:0032774_RNA_biosynthetic_process	646	93	-1.634084
GO:0006351_transcription__DNA-dependent	643	92	-1.55996
GO:0006355_regulation_of_transcription__DNA-dependent	628	91	-1.675912
GO:0051252_regulation_of_RNA_metabolic_process	634	91	-1.579709
GO:0043234_protein_complex	549	86	-2.467896
GO:0007186_G-protein_coupled_receptor_protein_signaling_pathway	506	85	-3.364551
GO:0004930_G-protein_coupled_receptor_activity	464	81	-3.77396
GO:0001584_rhodopsin-like_receptor_activity	446	79	-3.905268
GO:0044446_intracellular_organelle_part	546	79	-1.507306
GO:0044422_organelle_part	553	79	-1.395518
GO:0006996_organelle_organization_and_biogenesis	418	64	-1.769134
GO:0003008_system_process	222	49	-5.035934
GO:0050877_neurological_system_process	189	45	-5.586121
GO:0005694_chromosome	200	41	-3.571147
GO:0007600_sensory_perception	128	40	-8.473058

GO:0005886_plasma_membrane	245	39	-1.498678
GO:0007606_sensory_perception_of_chemical_stimulus	109	38	-9.611467
GO:0044427_chromosomal_part	156	38	-5.059084
GO:0051276_chromosome_organization_and_biogenesis	161	38	-4.72284
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	36	-7.179089
GO:0000785_chromatin	123	35	-6.377593
GO:0006333_chromatin_assembly_or_disassembly	111	34	-7.059986
GO:0022607_cellular_component_assembly	202	34	-1.68464
GO:0030234_enzyme_regulator_activity	200	33	-1.537309
GO:0006323_DNA_packaging	96	32	-7.654577
GO:0065003_macromolecular_complex_assembly	178	32	-2.007881
GO:0044459_plasma_membrane_part	198	32	-1.395954
GO:0031497_chromatin_assembly	89	31	-7.948749
GO:0065004_protein-DNA_complex_assembly	104	31	-6.202928
GO:0006334_nucleosome_assembly	85	30	-7.865185
GO:0000786_nucleosome	86	30	-7.730227
GO:0009966_regulation_of_signal_transduction	114	21	-1.599593
GO:0004857_enzyme_inhibitor_activity	89	18	-1.820808
GO:0031226_intrinsic_to_plasma_membrane	100	18	-1.353661
GO:0030414_protease_inhibitor_activity	82	17	-1.850147
GO:0005102_receptor_binding	77	16	-1.779935
GO:0004866_endopeptidase_inhibitor_activity	78	16	-1.728971
GO:0004867_serine-type_endopeptidase_inhibitor_activity	69	15	-1.872832
GO:0051056_regulation_of_small_GTPase_mediated_signal_transduction	76	15	-1.512215
GO:0046578_regulation_of_Ras_protein_signal_transduction	61	14	-1.986134
GO:0007265_Ras_protein_signal_transduction	69	14	-1.5359
GO:0005083_small_GTPase_regulator_activity	66	13	-1.366395
GO:0005179_hormone_activity	43	11	-2.015887
GO:0005089_Rho_guanyl-nucleotide_exchange_factor_activity	20	7	-2.215012
GO:0035023_regulation_of_Rho_protein_signal_transduction	20	7	-2.215012
GO:0005088_Ras_guanyl-nucleotide_exchange_factor_activity	21	7	-2.084807
GO:0007266_Rho_protein_signal_transduction	21	7	-2.084807
GO:0007178_transmembrane_receptor_protein_serine_threonine_kinase_signaling_pathway	20	6	-1.616573
GO:0043068_positive_regulation_of_programmed_cell_death	20	6	-1.616573
GO:0008652_amino_acid_biosynthetic_process	22	6	-1.420819
GO:0009069_serine_family_amino_acid_metabolic_process	16	5	-1.485148
GO:0030512_negative_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	5	4	-3.055733
GO:0006563_L-serine_metabolic_process	7	4	-2.296064
GO:0009070_serine_family_amino_acid_biosynthetic_process	7	4	-2.296064
GO:0017015_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	8	4	-2.037449
GO:0001764_neuron_migration	11	4	-1.490029
GO:0000097_sulfur_amino_acid_biosynthetic_process	5	3	-1.86349
GO:0006534_cysteine_metabolic_process	5	3	-1.86349
GO:0008060_ARF_GTPase_activator_activity	6	3	-1.602134

GO:0032312_regulation_of_ARF_GTPase_activity	6	3	-1.602134
GO:0000096_sulfur_amino_acid_metabolic_process	7	3	-1.398489
GO:0012502_induction_of_programmed_cell_death	7	3	-1.398489
GO:0044272_sulfur_compound_biosynthetic_process	7	3	-1.398489

HLH-2/HLH-8

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0009987_cellular_process	4977	60	-1.532413
GO:0044238_primary_metabolic_process	3216	41	-1.390553
GO:0043170_macromolecule_metabolic_process	2663	36	-1.58965
GO:0043167_ion_binding	1592	24	-1.630094
GO:0043169_cation_binding	1534	23	-1.551212
GO:0046872_metal_ion_binding	1525	22	-1.337147
GO:0016772_transferase_activity__transferring_phosphorus-containing_groups	622	13	-2.047723
GO:0003700_transcription_factor_activity	539	10	-1.367622
GO:0043234_protein_complex	549	10	-1.322679
GO:0005216_ion_channel_activity	288	7	-1.598116
GO:0022838_substrate_specific_channel_activity	288	7	-1.598116
GO:0015267_channel_activity	290	7	-1.583817
GO:0022803_passive_transmembrane_transporter_activity	290	7	-1.583817
GO:0005509_calcium_ion_binding	221	6	-1.632579
GO:0016779_nucleotidyltransferase_activity	76	5	-3.019638
GO:0005261_cation_channel_activity	172	5	-1.538768
GO:0022836_gated_channel_activity	194	5	-1.348094
GO:0015276_ligand-gated_ion_channel_activity	115	4	-1.552589
GO:0022834_ligand-gated_channel_activity	115	4	-1.552589
GO:0003899_DNA-directed_RNA_polymerase_activity	27	3	-2.622522
GO:0034062_RNA_polymerase_activity	29	3	-2.532167
GO:0006260_DNA_replication	57	3	-1.716095
GO:0008236_serine-type_peptidase_activity	74	3	-1.423152
GO:0017171_serine_hydrolase_activity	74	3	-1.423152
GO:0015464_acetylcholine_receptor_activity	11	2	-2.289691
GO:0030258_lipid_modification	11	2	-2.289691
GO:0042166_acetylcholine_binding	11	2	-2.289691
GO:0034061_DNA_polymerase_activity	17	2	-1.913465
GO:0043176_amine_binding	18	2	-1.865129
GO:0004889_nicotinic_acetylcholine-activated_cation-selective_channel_activity	20	2	-1.776695
GO:0043087_regulation_of_GTPase_activity	30	2	-1.444965
GO:0044445_cytosolic_part	30	2	-1.444965
GO:0045138_tail_tip_morphogenesis	30	2	-1.444965
GO:0051336_regulation_of_hydrolase_activity	32	2	-1.393546
GO:0005231_excitatory_extracellular_ligand-gated_ion_channel_activity	33	2	-1.369178
GO:0035121_tail_morphogenesis	35	2	-1.322855
GO:0005942_phosphoinositide_3-kinase_complex	5	1	-1.309873
GO:0006188_IMP_biosynthetic_process	5	1	-1.309873
GO:0009374_biotin_binding	5	1	-1.309873
GO:0030155_regulation_of_cell_adhesion	5	1	-1.309873
GO:0031344_regulation_of_cell_projection_organization_and_biogenesis	5	1	-1.309873
GO:0045178_basal_part_of_cell	5	1	-1.309873

GO:0046040_IMP_metabolic_process	5	1	-1.309873

HLH-2/HLH-10

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005575_cellular_component	4970	1125	-6.34995
GO:0005623_cell	4804	1086	-5.868558
GO:0044464_cell_part	4788	1082	-5.799364
GO:0005488_binding	5055	1082	-1.672207
GO:0005622_intracellular	2808	620	-1.957025
GO:0016020_membrane	2210	508	-3.037127
GO:0044424_intracellular_part	2280	498	-1.332135
GO:0043229_intracellular_organelle	1740	387	-1.502163
GO:0043226_organelle	1749	387	-1.386178
GO:0003676_nucleic_acid_binding	1510	357	-3.054953
GO:0043231_intracellular_membrane-bounded_organelle	1392	337	-3.674603
GO:0043227_membrane-bounded_organelle	1399	337	-3.503225
GO:0044425_membrane_part	1452	327	-1.587401
GO:0004871_signal_transducer_activity	1141	295	-5.545571
GO:0060089_molecular_transducer_activity	1141	295	-5.545571
GO:0031224_intrinsic_to_membrane	1270	290	-1.733837
GO:0016021_integral_to_membrane	1261	286	-1.578752
GO:0004872_receptor_activity	1061	283	-6.573909
GO:0005634_nucleus	1121	281	-4.198579
GO:0007154_cell_communication	1099	258	-2.158765
GO:0007165_signal_transduction	1016	243	-2.470132
GO:0050794_regulation_of_cellular_process	1029	237	-1.652531
GO:0003677_DNA_binding	873	227	-4.519128
GO:0004888_transmembrane_receptor_activity	740	214	-7.89794
GO:0010468_regulation_of_gene_expression	760	185	-2.320829
GO:0019222_regulation_of_metabolic_process	783	181	-1.426213
GO:0006350_transcription	747	177	-1.77864
GO:0031323_regulation_of_cellular_metabolic_process	755	177	-1.611655
GO:0032991_macromolecular_complex	721	176	-2.284326
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	724	173	-1.88524
GO:0045449_regulation_of_transcription	709	172	-2.133755
GO:0032774_RNA_biosynthetic_process	646	158	-2.136343
GO:0006351_transcription__DNA-dependent	643	157	-2.098036
GO:0006355_regulation_of_transcription__DNA-dependent	628	155	-2.258037
GO:0051252_regulation_of_RNA_metabolic_process	634	155	-2.098831
GO:0030528_transcription_regulator_activity	647	154	-1.677266
GO:0007166_cell_surface_receptor_linked_signal_transduction	590	151	-2.862353
GO:0043234_protein_complex	549	148	-3.880272
GO:0044446_intracellular_organelle_part	546	136	-2.18655
GO:0044422_organelle_part	553	136	-1.992626
GO:0007186_G-protein_coupled_receptor_protein_signaling_pathway	506	131	-2.751069
GO:0003700_transcription_factor_activity	539	130	-1.662237
GO:0004930_G-protein_coupled_receptor_activity	464	119	-2.403912
GO:0001584_rhodopsin-like_receptor_activity	446	113	-2.137444
GO:0043565_sequence-specific_DNA_binding	464	111	-1.403095
GO:0006996_organelle_organization_and_biogenesis	418	107	-2.191251

GO:0015267_channel_activity	290	76	-1.955539
GO:0022803_passive_transmembrane_transporter_activity	290	76	-1.955539
GO:0005216_ion_channel_activity	288	75	-1.870845
GO:0022838_substrate_specific_channel_activity	288	75	-1.870845
GO:0005694_chromosome	200	70	-5.913745
GO:0003008_system_process	222	67	-3.405486
GO:0051276_chromosome_organization_and_biogenesis	161	63	-7.369018
GO:0022607_cellular_component_assembly	202	63	-3.666815
GO:0044427_chromosomal_part	156	61	-7.142651
GO:0050877_neurological_system_process	189	58	-3.226783
GO:0000785_chromatin	123	57	-9.994762
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	55	-9.441926
GO:0065003_macromolecular_complex_assembly	178	55	-3.169975
GO:0006333_chromatin_assembly_or_disassembly	111	53	-9.941449
GO:0006323_DNA_packaging	96	51	-11.822753
GO:0022836_gated_channel_activity	194	50	-1.347782
GO:0031497_chromatin_assembly	89	49	-12.159596
GO:0065004_protein-DNA_complex_assembly	104	49	-8.990191
GO:0006334_nucleosome_assembly	85	48	-12.482245
GO:0000786_nucleosome	86	48	-12.223149
GO:0007600_sensory_perception	128	47	-4.789661
GO:0005261_cation_channel_activity	172	45	-1.358565
GO:0007606_sensory_perception_of_chemical_stimulus	109	44	-5.769042
GO:0006813_potassium_ion_transport	115	32	-1.426241
GO:0005267_potassium_channel_activity	108	30	-1.3583
GO:0031226_intrinsic_to_plasma_membrane	100	28	-1.339525
GO:0004857_enzyme_inhibitor_activity	89	27	-1.733652
GO:0030414_protease_inhibitor_activity	82	25	-1.669051
GO:0004866_endopeptidase_inhibitor_activity	78	23	-1.414188
GO:0005244_voltage-gated_ion_channel_activity	74	22	-1.410379
GO:0022832_voltage-gated_channel_activity	74	22	-1.410379
GO:0000151_ubiquitin_ligase_complex	18	8	-1.720684
GO:0043068_positive_regulation_of_programmed_cell_death	20	8	-1.429652
GO:0006576_biogenic_amine_metabolic_process	16	7	-1.514812
GO:0030512_negative_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	5	5	-3.437137
GO:0017015_regulation_of_transforming_growth_factor_beta_receptor_signaling_pathway	8	5	-1.931659
GO:0007218_neuropeptide_signaling_pathway	9	5	-1.659285
GO:0031461_cullin-RING_ubiquitin_ligase_complex	9	5	-1.659285
GO:0003746_translation_elongation_factor_activity	10	5	-1.437481
GO:0031625_ubiquitin_protein_ligase_binding	10	5	-1.437481
GO:0005024_transforming_growth_factor_beta_receptor_activity	6	4	-1.728402
GO:0008060_ARF_GTPase_activator_activity	6	4	-1.728402
GO:0032312_regulation_of_ARF_GTPase_activity	6	4	-1.728402
GO:0006563_L-serine_metabolic_process	7	4	-1.43691
GO:0009070_serine_family_amino_acid_biosynthetic_process	7	4	-1.43691
GO:0012502_induction_of_programmed_cell_death	7	4	-1.43691
GO:0031594_neuromuscular_junction	7	4	-1.43691

HLH-11

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0007165_signal_transduction	1016	28	-2.143064
GO:0007154_cell_communication	1099	28	-1.713111
GO:0004871_signal_transducer_activity	1141	28	-1.523651
GO:0060089_molecular_transducer_activity	1141	28	-1.523651
GO:0004872_receptor_activity	1061	27	-1.660433
GO:0004888_transmembrane_receptor_activity	740	23	-2.441307
GO:0007166_cell_surface_receptor_linked_signal_transduction	590	16	-1.355865
GO:0004930_G-protein_coupled_receptor_activity	464	15	-1.890111
GO:0007186_G-protein_coupled_receptor_protein_signaling_pathway	506	15	-1.585825
GO:0001584_rhodopsin-like_receptor_activity	446	14	-1.695705
GO:0008168_methyltransferase_activity	77	4	-1.375304
GO:0004866_endopeptidase_inhibitor_activity	78	4	-1.358032
GO:0016741_transferase_activity__transferring_one-carbon_groups	78	4	-1.358032
GO:0005783_endoplasmic_reticulum	81	4	-1.307945
GO:0043068_positive_regulation_of_programmed_cell_death	20	3	-2.349399
GO:0043067_regulation_of_programmed_cell_death	41	3	-1.491334
GO:0009219_pyrimidine_deoxyribonucleotide_metabolic_process	5	2	-2.554716
GO:0009262_deoxyribonucleotide_metabolic_process	9	2	-2.017912
GO:0008083_growth_factor_activity	10	2	-1.92586
GO:0006665_sphingolipid_metabolic_process	11	2	-1.843561
GO:0006221_pyrimidine_nucleotide_biosynthetic_process	12	2	-1.769224
GO:0006220_pyrimidine_nucleotide_metabolic_process	14	2	-1.639395
GO:0009123_nucleoside_monophosphate_metabolic_process	15	2	-1.58207
GO:0009124_nucleoside_monophosphate_biosynthetic_process	15	2	-1.58207
GO:0032269_negative_regulation_of_cellular_protein_metabolic_process	16	2	-1.528895
GO:0051248_negative_regulation_of_protein_metabolic_process	16	2	-1.528895
GO:0033043_regulation_of_organelle_organization_and_biogenesis	18	2	-1.432998
GO:0051493_regulation_of_cytoskeleton_organization_and_biogenesis	18	2	-1.432998

HLH-2/HLH-14

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005575_cellular_component	4970	487	-4.820709
GO:0005623_cell	4804	472	-4.733827
GO:0044464_cell_part	4788	469	-4.526545
GO:0005488_binding	5055	467	-1.961882
GO:0032501_multicellular_organismal_process	3705	361	-2.998512
GO:0032502_developmental_process	3651	342	-1.716646
GO:0007275_multicellular_organismal_development	3428	321	-1.60695
GO:0005622_intracellular	2808	287	-3.697191
GO:0009790_embryonic_development	2693	266	-2.493713

GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	259	-2.227347
GO:0043170_macromolecule_metabolic_process	2663	253	-1.565201
GO:0044424_intracellular_part	2280	226	-2.210878
GO:0000003_reproduction	1895	188	-1.89959
GO:0043229_intracellular_organelle	1740	177	-2.227614
GO:0043226_organelle	1749	177	-2.137404
GO:0003676_nucleic_acid_binding	1510	164	-3.298471
GO:0043231_intracellular_membrane-bounded_organelle	1392	157	-3.955295
GO:0043227_membrane-bounded_organelle	1399	157	-3.845916
GO:0005634_nucleus	1121	132	-4.203501
GO:0046914_transition_metal_ion_binding	1256	131	-2.087031
GO:0005515_protein_binding	1159	117	-1.521189
GO:0008270_zinc_ion_binding	1083	113	-1.864879
GO:0003677_DNA_binding	873	99	-2.710427
GO:0016043_cellular_component_organization_and_biogenesis	804	98	-3.758583
GO:0009653_anatomical_structure_morphogenesis	879	91	-1.515363
GO:0032991_macromolecular_complex	721	90	-3.848509
GO:0043234_protein_complex	549	78	-5.336245
GO:0044446_intracellular_organelle_part	546	74	-4.374889
GO:0044422_organelle_part	553	74	-4.190904
GO:0043228_non-membrane-bounded_organelle	535	66	-2.826726
GO:0043232_intracellular_non-membrane-bounded_organelle	535	66	-2.826726
GO:0006996_organelle_organization_and_biogenesis	418	59	-4.083772
GO:0005694_chromosome	200	43	-7.944692
GO:0044427_chromosomal_part	156	42	-11.019662
GO:0051276_chromosome_organization_and_biogenesis	161	42	-10.535748
GO:0000785_chromatin	123	40	-13.466246
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	39	-13.134164
GO:0022607_cellular_component_assembly	202	39	-5.980671
GO:0006333_chromatin_assembly_or_disassembly	111	38	-13.634837
GO:0006323_DNA_packaging	96	37	-15.203773
GO:0065003_macromolecular_complex_assembly	178	37	-6.52909
GO:0031497_chromatin_assembly	89	36	-15.602111
GO:0048609_reproductive_process_in_a_multicellular_organism	303	36	-1.55938
GO:0032504_multicellular_organism_reproduction	304	36	-1.540429
GO:0006334_nucleosome_assembly	85	35	-15.474457
GO:0000786_nucleosome	86	35	-15.283967
GO:0065004_protein-DNA_complex_assembly	104	35	-12.354032
GO:0018991_oviposition	295	34	-1.338932
GO:0033057_reproductive_behavior_in_a_multicellular_organism	295	34	-1.338932
GO:0002009_morphogenesis_of_an_epithelium	254	32	-1.761678
GO:0003008_system_process	222	30	-2.084554
GO:0005886_plasma_membrane	245	30	-1.533379
GO:0050877_neurological_system_process	189	26	-1.972386
GO:0044459_plasma_membrane_part	198	25	-1.497042
GO:0005525_GTP_binding	161	23	-1.991043
GO:0032561_guanyl_ribonucleotide_binding	161	23	-1.991043
GO:0019001_guanyl_nucleotide_binding	172	23	-1.671436
GO:0007606_sensory_perception_of_chemical_stimulus	109	21	-3.486622
GO:0007600_sensory_perception	128	21	-2.556589

GO:0006813_potassium_ion_transport	115	18	-2.055666
GO:0005887_integral_to_plasma_membrane	99	16	-2.013716
GO:0031226_intrinsic_to_plasma_membrane	100	16	-1.972776
GO:0005267_potassium_channel_activity	108	16	-1.672505
GO:0004518_nuclease_activity	81	14	-2.075229
GO:0005244_voltage-gated_ion_channel_activity	74	13	-2.021745
GO:0022832_voltage-gated_channel_activity	74	13	-2.021745
GO:0007369_gastrulation	81	13	-1.706825
GO:0003924_GTPase_activity	88	13	-1.440197
GO:0001703_gastrulation_with_mouth_forming_first	60	10	-1.521435
GO:0050660_FAD_binding	60	10	-1.521435
GO:0005249_voltage-gated_potassium_channel_activity	61	10	-1.476963
GO:0004527_exonuclease_activity	33	9	-2.853284
GO:0008134_transcription_factor_binding	38	9	-2.395747
GO:0005179_hormone_activity	43	8	-1.558121
GO:0003712_transcription_cofactor_activity	24	6	-1.863562
GO:0045138_tail_tip_morphogenesis	30	6	-1.408894
GO:0008632_apoptotic_program	23	5	-1.377149
GO:0009109_coenzyme_catabolic_process	13	4	-1.689393
GO:0051187_cofactor_catabolic_process	13	4	-1.689393
GO:0005253_anion_channel_activity	14	4	-1.573074
GO:0044242_cellular_lipid_catabolic_process	14	4	-1.573074
GO:0008138_protein_tyrosine_serine_threonine_phosphatase_activity	15	4	-1.468075
GO:0022411_cellular_component_disassembly	15	4	-1.468075
GO:0006576_biogenic_amine_metabolic_process	16	4	-1.372726
GO:0009069_serine_family_amino_acid_metabolic_process	16	4	-1.372726
GO:0006563_L-serine_metabolic_process	7	3	-1.772616
GO:0009070_serine_family_amino_acid_biosynthetic_process	7	3	-1.772616
GO:0012502_induction_of_programmed_cell_death	7	3	-1.772616
GO:0003997_acyl-CoA_oxidase_activity	8	3	-1.596711
GO:0008308_voltage-gated_anion_channel_activity	8	3	-1.596711
GO:0006821_chloride_transport	9	3	-1.448693
GO:0007509_mesoderm_migration	9	3	-1.448693
GO:0008078_mesodermal_cell_migration	9	3	-1.448693
GO:0003746_translation_elongation_factor_activity	10	3	-1.321721
GO:0006308_DNA_catabolic_process	10	3	-1.321721
GO:0006309_DNA_fragmentation_during_apoptosis	10	3	-1.321721
GO:0006635_fatty_acid_beta-oxidation	10	3	-1.321721
GO:0009062_fatty_acid_catabolic_process	10	3	-1.321721
GO:0016054_organic_acid_catabolic_process	10	3	-1.321721
GO:0016634_oxidoreductase_activity__acting_on_the_CH-CH_group_of_donors__oxygen_as_acceptor	10	3	-1.321721
GO:0019395_fatty_acid_oxidation	10	3	-1.321721
GO:0030262_apoptotic_nuclear_changes	10	3	-1.321721
GO:0042074_cell_migration_involved_in_gastrulation	10	3	-1.321721
GO:0046395_carboxylic_acid_catabolic_process	10	3	-1.321721

HLH-2/HLH-15

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0003674_molecular_function	8423	37	-1.857026
GO:0005488_binding	5055	31	-4.151057

GO:0009987_cellular_process	4977	26	-1.914066
GO:0008152_metabolic_process	4176	24	-2.303481
GO:0005622_intracellular	2808	22	-4.053039
GO:0050789_regulation_of_biological_process	2843	22	-3.966008
GO:0065007_biological_regulation	2922	22	-3.775308
GO:0043170_macromolecule_metabolic_process	2663	21	-3.865139
GO:0032502_developmental_process	3651	21	-1.952353
GO:0003676_nucleic_acid_binding	1510	20	-7.223284
GO:0007275_multicellular_organismal_development	3428	20	-1.918418
GO:0032501_multicellular_organismal_process	3705	20	-1.538642
GO:0043231_intracellular_membrane-bounded_organelle	1392	19	-7.010799
GO:0043227_membrane-bounded_organelle	1399	19	-6.975486
GO:0043229_intracellular_organelle	1740	19	-5.479309
GO:0043226_organelle	1749	19	-5.444945
GO:0048518_positive_regulation_of_biological_process	1955	19	-4.716498
GO:0044424_intracellular_part	2280	19	-3.757226
GO:0040007_growth	2387	19	-3.483217
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	19	-2.893077
GO:0009790_embryonic_development	2693	19	-2.793215
GO:0005634_nucleus	1121	18	-7.682881
GO:0045927_positive_regulation_of_growth	1836	18	-4.484069
GO:0040008_regulation_of_growth	1886	18	-4.322361
GO:0003677_DNA_binding	873	17	-8.466021
GO:0040010_positive_regulation_of_growth_rate	1727	17	-4.22856
GO:0040009_regulation_of_growth_rate	1728	17	-4.225278
GO:0000785_chromatin	123	15	-19.036322
GO:0044427_chromosomal_part	156	15	-17.435083
GO:0005694_chromosome	200	15	-15.790495
GO:0043228_non-membrane-bounded_organelle	535	15	-9.558229
GO:0043232_intracellular_non-membrane-bounded_organelle	535	15	-9.558229
GO:0044446_intracellular_organelle_part	546	15	-9.434727
GO:0044422_organelle_part	553	15	-9.35756
GO:0016043_cellular_component_organization_and_biogenesis	804	15	-7.143331
GO:0006334_nucleosome_assembly	85	14	-19.651493
GO:0000786_nucleosome	86	14	-19.575311
GO:0031497_chromatin_assembly	89	14	-19.352484
GO:0006323_DNA_packaging	96	14	-18.863106
GO:0065004_protein-DNA_complex_assembly	104	14	-18.349463
GO:0006333_chromatin_assembly_or_disassembly	111	14	-17.934117
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	14	-17.43998
GO:0051276_chromosome_organization_and_biogenesis	161	14	-15.603629
GO:0065003_macromolecular_complex_assembly	178	14	-14.98536
GO:0022607_cellular_component_assembly	202	14	-14.212398
GO:0006996_organelle_organization_and_biogenesis	418	14	-9.902316
GO:0043234_protein_complex	549	14	-8.354568
GO:0032991_macromolecular_complex	721	14	-6.856478
GO:0040011_locomotion	1157	13	-3.766113
GO:0002119_nematode_larval_development	1625	13	-2.362779
GO:0002164_larval_development	1627	13	-2.35808
GO:0009791_post-embryonic_development	1643	13	-2.320806
GO:0006694_steroid_biosynthetic_process	6	1	-1.657372
GO:0008060_ARF_GTPase_activator_activity	6	1	-1.657372

GO:0032312_regulation_of_ARF_GTPase_activity	6	1	-1.657372
GO:0032934_sterol_binding	6	1	-1.657372
GO:0046839_phospholipid_dephosphorylation	6	1	-1.657372
GO:0000790_nuclear_chromatin	7	1	-1.591207
GO:0012502_induction_of_programmed_cell_death	7	1	-1.591207
GO:0043071_positive_regulation_of_non-apoptotic_programmed_cell_death	7	1	-1.591207
GO:0005496_steroid_binding	9	1	-1.483626
GO:0040013_negative_regulation_of_locomotion	9	1	-1.483626
GO:0000323_lytic_vacuole	10	1	-1.43865
GO:0005764_lysosome	10	1	-1.43865
GO:0008202_steroid_metabolic_process	11	1	-1.398038
GO:0030258_lipid_modification	11	1	-1.398038
GO:0032011_ARF_protein_signal_transduction	11	1	-1.398038
GO:0032012_regulation_of_ARF_protein_signal_transduction	11	1	-1.398038
GO:0043070_regulation_of_non-apoptotic_programmed_cell_death	12	1	-1.36103
GO:0044454_nuclear_chromosome_part	12	1	-1.36103
GO:0016244_non-apoptotic_programmed_cell_death	13	1	-1.327047

HLH-2/HLH-19

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005575_cellular_component	4970	193	-1.812954
GO:0005623_cell	4804	185	-1.568829
GO:0044464_cell_part	4788	183	-1.420881
GO:0005622_intracellular	2808	112	-1.352614
GO:0043170_macromolecule_metabolic_process	2663	109	-1.610909
GO:0016787_hydrolase_activity	1451	67	-2.124863
GO:0043231_intracellular_membrane-bounded_organelle	1392	64	-2.006368
GO:0043227_membrane-bounded_organelle	1399	64	-1.960577
GO:0005634_nucleus	1121	51	-1.605036
GO:0050794_regulation_of_cellular_process	1029	47	-1.538911
GO:0032991_macromolecular_complex	721	37	-1.992413
GO:0010468_regulation_of_gene_expression	760	37	-1.663633
GO:0031323_regulation_of_cellular_metabolic_process	755	36	-1.512538
GO:0019222_regulation_of_metabolic_process	783	36	-1.315839
GO:0045449_regulation_of_transcription	709	34	-1.484214
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	724	34	-1.374273
GO:0043234_protein_complex	549	32	-2.57648
GO:0044446_intracellular_organelle_part	546	27	-1.408581
GO:0044422_organelle_part	553	27	-1.35126
GO:0006508_proteolysis	465	26	-1.96454
GO:0008233_peptidase_activity	441	24	-1.722676
GO:0044459_plasma_membrane_part	198	16	-2.848582
GO:0005886_plasma_membrane	245	16	-1.952024
GO:0008237_metallopeptidase_activity	192	14	-2.162931
GO:0005694_chromosome	200	14	-2.013425
GO:0044427_chromosomal_part	156	13	-2.528638
GO:0050793_regulation_of_developmental_process	201	12	-1.330815
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	11	-2.533863
GO:0000785_chromatin	123	11	-2.450606

GO:0051276_chromosome_organization_and_biogenesis	161	11	-1.617745
GO:0006333_chromatin_assembly_or_disassembly	111	10	-2.296456
GO:0005887_integral_to_plasma_membrane	99	9	-2.141559
GO:0031226_intrinsic_to_plasma_membrane	100	9	-2.113629
GO:0031497_chromatin_assembly	89	8	-1.929144
GO:0006323_DNA_packaging	96	8	-1.745337
GO:0016477_cell_migration	96	8	-1.745337
GO:0006928_cell_motility	102	8	-1.603527
GO:0051674_localization_of_cell	102	8	-1.603527
GO:0007606_sensory_perception_of_chemical_stimulus	109	8	-1.453985
GO:0006334_nucleosome_assembly	85	7	-1.559217
GO:0000786_nucleosome	86	7	-1.534909
GO:0001708_cell_fate_specification	57	6	-1.868242
GO:0016810_hydrolase_activity__acting_on_carbon-nitrogen_(but_not_peptide)_bonds	59	6	-1.799113
GO:0000902_cell_morphogenesis	69	6	-1.497894
GO:0032989_cellular_structure_morphogenesis	69	6	-1.497894
GO:0005921_gap_junction	25	4	-1.998291
GO:0014704_intercalated_disc	25	4	-1.998291
GO:0009888_tissue_development	29	4	-1.771667
GO:0004180_carboxypeptidase_activity	31	4	-1.672912
GO:0004527_exonuclease_activity	33	4	-1.582131
GO:0008134_transcription_factor_binding	38	4	-1.383955
GO:0007509_mesoderm_migration	9	3	-2.529174
GO:0008078_mesodermal_cell_migration	9	3	-2.529174
GO:0042074_cell_migration_involved_in_gastrulation	10	3	-2.385486
GO:0001707_mesoderm_formation	14	3	-1.948221
GO:0048332_mesoderm_morphogenesis	14	3	-1.948221
GO:0007498_mesoderm_development	16	3	-1.783338
GO:0009069_serine_family_amino_acid_metabolic_process	16	3	-1.783338
GO:0001704_formation_of_primary_germ_layer	18	3	-1.641957
GO:0043068_positive_regulation_of_programmed_cell_death	20	3	-1.518778
GO:0048729_tissue_morphogenesis	20	3	-1.518778
GO:0004181_metallocarboxypeptidase_activity	21	3	-1.462818
GO:0004182_carboxypeptidase_A_activity	21	3	-1.462818
GO:0048646_anatomical_structure_formation	21	3	-1.462818
GO:0022603_regulation_of_anatomical_structure_morphogenesis	23	3	-1.360358
GO:0003712_transcription_cofactor_activity	24	3	-1.313288
GO:0016811_hydrolase_activity__acting_on_carbon-nitrogen_(but_not_peptide)_bonds__in_linear_amides	24	3	-1.313288
GO:0008305_integrin_complex	6	2	-1.786783
GO:0003705_RNA_polymerase_II_transcription_factor_activity__enhancer_binding	7	2	-1.650622
GO:0006563_L-serine_metabolic_process	7	2	-1.650622
GO:0009070_serine_family_amino_acid_biosynthetic_process	7	2	-1.650622
GO:0012502_induction_of_programmed_cell_death	7	2	-1.650622
GO:0017145_stem_cell_division	7	2	-1.650622
GO:0042078_germ-line_stem_cell_division	7	2	-1.650622
GO:0003997_acyl-CoA_oxidase_activity	8	2	-1.535621
GO:0004407_histone_deacetylase_activity	8	2	-1.535621
GO:0006476_protein_amino_acid_deacetylation	8	2	-1.535621
GO:0008037_cell_recognition	8	2	-1.535621

GO:0008038_neuron_recognition	8	2	-1.535621
GO:0033558_protein_deacetylase_activity	8	2	-1.535621
GO:0043235_receptor_complex	8	2	-1.535621
GO:0019213_deacetylase_activity	9	2	-1.436384
GO:0006612_protein_targeting_to_membrane	10	2	-1.349351
GO:0006635_fatty_acid_beta-oxidation	10	2	-1.349351
GO:0009062_fatty_acid_catabolic_process	10	2	-1.349351
GO:0010558_negative_regulation_of_macromolecule_biosynthetic_process	10	2	-1.349351
GO:0016054_organic_acid_catabolic_process	10	2	-1.349351
GO:0016634_oxidoreductase_activity__acting_on_the_CH-CH_group_of_donors__oxygen_as_acceptor	10	2	-1.349351
GO:0017148_negative_regulation_of_translation	10	2	-1.349351
GO:0019395_fatty_acid_oxidation	10	2	-1.349351
GO:0046395_carboxylic_acid_catabolic_process	10	2	-1.349351

HLH-25

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0032501_multicellular_organismal_process	3705	927	-2.823828
GO:0032502_developmental_process	3651	926	-3.684049
GO:0007275_multicellular_organismal_development	3428	877	-4.022807
GO:0065007_biological_regulation	2922	730	-2.132016
GO:0050789_regulation_of_biological_process	2843	713	-2.246649
GO:0005622_intracellular	2808	692	-1.524683
GO:0009790_embryonic_development	2693	689	-3.084198
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	674	-2.823476
GO:0040007_growth	2387	620	-3.481976
GO:0048518_positive_regulation_of_biological_process	1955	500	-2.265009
GO:0000003_reproduction	1895	494	-2.93751
GO:0040008_regulation_of_growth	1886	481	-2.099479
GO:0045927_positive_regulation_of_growth	1836	465	-1.832267
GO:0040010_positive_regulation_of_growth_rate	1727	442	-2.061449
GO:0040009_regulation_of_growth_rate	1728	442	-2.044078
GO:0043226_organelle	1749	440	-1.575104
GO:0043229_intracellular_organelle	1740	439	-1.647489
GO:0009791_post-embryonic_development	1643	437	-3.365444
GO:0002164_larval_development	1627	433	-3.359765
GO:0002119_nematode_larval_development	1625	432	-3.310844
GO:0048856_anatomical_structure_development	1280	328	-1.662375
GO:0040011_locomotion	1157	297	-1.583171
GO:0005737_cytoplasm	1068	275	-1.559051
GO:0022414_reproductive_process	860	224	-1.547576
GO:0048731_system_development	734	193	-1.546293
GO:0048513_organ_development	692	182	-1.489829
GO:0044444_cytoplasmic_part	612	173	-2.690686
GO:0003006_reproductive_developmental_process	594	156	-1.328339
GO:0048806_genitalia_development	515	138	-1.47661
GO:0040035_hermaphrodite_genitalia_development	511	137	-1.476265
GO:0051641_cellular_localization	294	86	-1.989035
GO:0051649_establishment_of_cellular_localization	288	83	-1.773207
GO:0007049_cell_cycle	248	76	-2.346025
GO:0033036_macromolecule_localization	242	74	-2.273743
GO:0008104_protein_localization	240	73	-2.189875

GO:0022402_cell_cycle_process	230	70	-2.129431
GO:0048519_negative_regulation_of_biological_process	235	68	-1.58452
GO:0005856_cytoskeleton	179	58	-2.483006
GO:0007010_cytoskeleton_organization_and_biogenesis	186	56	-1.718332
GO:0018988_molting_cycle__protein-based_cuticle	194	56	-1.376243
GO:0018996_molting_cycle__collagen_and_cuticulin-based_cuticle	194	56	-1.376243
GO:0042303_molting_cycle	194	56	-1.376243
GO:0044430_cytoskeletal_part	141	48	-2.621317
GO:0051301_cell_division	148	48	-2.160875
GO:0031090_organelle_membrane	122	40	-1.975025
GO:0006470_protein_amino_acid_dephosphorylation	132	40	-1.414144
GO:0000910_cytokinesis	121	39	-1.811677
GO:0000278_mitotic_cell_cycle	113	36	-1.634364
GO:0045165_cell_fate_commitment	84	33	-3.119818
GO:0015630_microtubule_cytoskeleton	99	33	-1.830827
GO:0012505_endomembrane_system	87	31	-2.202756
GO:0007369_gastrulation	81	28	-1.848385
GO:0000226_microtubule_cytoskeleton_organization_and_biogenesis	87	28	-1.434997
GO:0005783_endoplasmic_reticulum	81	27	-1.590304
GO:0045926_negative_regulation_of_growth	83	27	-1.457251
GO:0005739_mitochondrion	80	26	-1.416338
GO:0031975_envelope	78	25	-1.312241
GO:0007267_cell-cell_signaling	70	24	-1.612738
GO:0031967_organelle_envelope	73	24	-1.399411
GO:0001708_cell_fate_specification	57	21	-1.828603
GO:0006732_coenzyme_metabolic_process	60	21	-1.569353
GO:0005938_cell_cortex	40	18	-2.690873
GO:0030029_actin_filament-based_process	47	18	-1.818595
GO:0032940_secretion_by_cell	50	18	-1.533155
GO:0044429_mitochondrial_part	49	17	-1.324066
GO:0016790_thiolester_hydrolase_activity	42	16	-1.646552
GO:0030036_actin_cytoskeleton_organization_and_biogenesis	42	15	-1.323797
GO:0005813_centrosome	30	14	-2.377473
GO:0005815_microtubule_organizing_center	31	14	-2.219849
GO:0044432_endoplasmic_reticulum_part	25	12	-2.225118
GO:0000793_condensed_chromosome	28	12	-1.762929
GO:0019866_organelle_inner_membrane	30	12	-1.510927
GO:0005635_nuclear_envelope	31	12	-1.39873
GO:0005789_endoplasmic_reticulum_membrane	24	11	-1.897328
GO:0016407_acetyltransferase_activity	24	11	-1.897328
GO:0042175_nuclear_envelope-endoplasmic_reticulum_network	26	11	-1.60759
GO:0001704_formation_of_primary_germ_layer	18	10	-2.503584
GO:0008483_transaminase_activity	21	10	-1.898833
GO:0048646_anatomical_structure_formation	21	10	-1.898833
GO:0008080_N-acetyltransferase_activity	22	10	-1.735425
GO:0016410_N-acyltransferase_activity	23	10	-1.587098
GO:0008105_asymmetric_protein_localization	25	10	-1.3289
GO:0051728_cell_cycle_switching__mitotic_to_meiotic_cell_cycle	16	8	-1.745168
GO:0051729_germline_cell_cycle_switching__mitotic_to_meiotic_cell_cycle	16	8	-1.745168
GO:0060184_cell_cycle_switching	16	8	-1.745168

GO:0015980_energy_derivation_by_oxidation_of_organic_compounds	18	8	-1.407866
GO:0043176_amine_binding	18	8	-1.407866
GO:0044453_nuclear_membrane_part	18	8	-1.407866
GO:0005882_intermediate_filament	9	7	-3.068523
GO:0045111_intermediate_filament_cytoskeleton	10	7	-2.643664
GO:0009109_coenzyme_catabolic_process	13	7	-1.779429
GO:0051187_cofactor_catabolic_process	13	7	-1.779429
GO:0008287_protein_serine_threonine_phosphatase_complex	14	7	-1.574378
GO:0007219_Notch_signaling_pathway	15	7	-1.396791
GO:0009306_protein_secretion	8	6	-2.540213
GO:0033058_directional_locomotion	9	6	-2.158655
GO:0006099_tricarboxylic_acid_cycle	11	6	-1.607584
GO:0008593_regulation_of_Notch_signaling_pathway	11	6	-1.607584
GO:0009060_aerobic_respiration	11	6	-1.607584
GO:0015464_acetylcholine_receptor_activity	11	6	-1.607584
GO:0016055_Wnt_receptor_signaling_pathway	11	6	-1.607584
GO:0042166_acetylcholine_binding	11	6	-1.607584
GO:0045333_cellular_respiration	11	6	-1.607584
GO:0046356_acetyl-CoA_catabolic_process	11	6	-1.607584
GO:0043070_regulation_of_non-apoptotic_programmed_cell_death	12	6	-1.400231
GO:0005759_mitochondrial_matrix	7	5	-2.026309
GO:0031980_mitochondrial_lumen	7	5	-2.026309
GO:0033205_cytokinesis_during_cell_cycle	7	5	-2.026309
GO:0043057_backward_locomotion	7	5	-2.026309
GO:0030031_cell_projection_biogenesis	8	5	-1.692614
GO:0034330_cell_junction_assembly_and_maintenance	8	5	-1.692614
GO:0045216_cell-cell_junction_assembly_and_maintenance	8	5	-1.692614
GO:0004016_adenylate_cyclase_activity	5	4	-1.919786
GO:0006637_acyl-CoA_metabolic_process	5	4	-1.919786
GO:0030055_cell-substrate_junction	5	4	-1.919786
GO:0045178_basal_part_of_cell	5	4	-1.919786
GO:0045747_positive_regulation_of_Notch_signaling_pathway	6	4	-1.531317

HLH-26

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0003674_molecular_function	8423	162	-1.339745
GO:0009987_cellular_process	4977	102	-1.34402
GO:0008152_metabolic_process	4176	93	-2.232527
GO:0003824_catalytic_activity	3551	85	-2.987214
GO:0044238_primary_metabolic_process	3216	75	-2.271962
GO:0044237_cellular_metabolic_process	3372	75	-1.748455
GO:0043170_macromolecule_metabolic_process	2663	59	-1.375545
GO:0043283_biopolymer_metabolic_process	1859	47	-2.078
GO:0044260_cellular_macromolecule_metabolic_process	1588	46	-3.223559
GO:0044267_cellular_protein_metabolic_process	1536	45	-3.263391
GO:0019538_protein_metabolic_process	1591	45	-2.937299
GO:0017076_purine_nucleotide_binding	1136	32	-2.151675
GO:0000166_nucleotide_binding	1283	32	-1.452077
GO:0032553_ribonucleotide_binding	1066	31	-2.293838

GO:0032555_purine_ribonucleotide_binding	1066	31	-2.293838
GO:0043687_post-translational_protein_modification	663	29	-5.115601
GO:0006464_protein_modification_process	752	29	-4.102843
GO:0043412_biopolymer_modification	771	29	-3.912379
GO:0030554_adenyl_nucleotide_binding	976	29	-2.299923
GO:0005524_ATP_binding	916	28	-2.403068
GO:0032559_adenyl_ribonucleotide_binding	917	28	-2.396483
GO:0006793_phosphorus_metabolic_process	642	26	-4.060057
GO:0006796_phosphate_metabolic_process	642	26	-4.060057
GO:0016772_transferase_activity__transferring_phosphorus-containing_groups	622	23	-3.085129
GO:0016310_phosphorylation	497	20	-3.188608
GO:0016773_phosphotransferase_activity__alcohol_group_as_acceptor	505	20	-3.102204
GO:0016301_kinase_activity	531	20	-2.836682
GO:0006468_protein_amino_acid_phosphorylation	444	19	-3.376583
GO:0004672_protein_kinase_activity	452	19	-3.281161
GO:0004674_protein_serine_threonine_kinase_activity	411	17	-2.90969
GO:0004713_protein-tyrosine_kinase_activity	324	14	-2.654386
GO:0019787_small_conjugating_protein_ligase_activity	48	4	-1.967618
GO:0016881_acid-amino_acid_ligase_activity	58	4	-1.689218
GO:0016853_isomerase_activity	66	4	-1.506925
GO:0008316_structural_constituent_of_vitelline_membrane	20	3	-2.278979
GO:0030704_vitelline_membrane_formation	20	3	-2.278979
GO:0030198_extracellular_matrix_organization_and_biogenesis	23	3	-2.104878
GO:0043062_extracellular_structure_organization_and_biogenesis	35	3	-1.60557
GO:0008533_astacin_activity	41	3	-1.427624
GO:0016638_oxidoreductase_activity__acting_on_the_CH-NH2_group_of_donors	8	2	-2.073401
GO:0007283_spermatogenesis	11	2	-1.795646
GO:0048232_male_gamete_generation	11	2	-1.795646
GO:0030239_myofibril_assembly	12	2	-1.7216
GO:0031032_actomyosin_structure_organization_and_biogenesis	12	2	-1.7216
GO:0045445_myoblast_differentiation	12	2	-1.7216
GO:0048627_myoblast_development	12	2	-1.7216
GO:0048628_myoblast_maturation	12	2	-1.7216
GO:0051146_striated_muscle_cell_differentiation	12	2	-1.7216
GO:0055002_striated_muscle_cell_development	12	2	-1.7216
GO:0007519_skeletal_muscle_development	13	2	-1.654176
GO:0014706_striated_muscle_development	13	2	-1.654176
GO:0048741_skeletal_muscle_fiber_development	13	2	-1.654176
GO:0048747_muscle_fiber_development	13	2	-1.654176
GO:0055001_muscle_cell_development	13	2	-1.654176
GO:0009069_serine_family_amino_acid_metabolic_process	16	2	-1.482424
GO:0016563_transcription_activator_activity	16	2	-1.482424
GO:0042692_muscle_cell_differentiation	18	2	-1.387097

HLH-27

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)

GO:0008150_biological_process	8808	595	-3.84523
GO:0009987_cellular_process	4977	344	-1.790902
GO:0032502_developmental_process	3651	292	-6.260899
GO:0008152_metabolic_process	4176	287	-1.347597
GO:0032501_multicellular_organismal_process	3705	286	-4.678458
GO:0007275_multicellular_organismal_development	3428	274	-5.730074
GO:0044237_cellular_metabolic_process	3372	236	-1.432818
GO:0065007_biological_regulation	2922	222	-3.04874
GO:0005622_intracellular	2808	215	-3.134729
GO:0050789_regulation_of_biological_process	2843	215	-2.837089
GO:0040007_growth	2387	205	-6.219237
GO:0009790_embryonic_development	2693	204	-2.725049
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	200	-2.632067
GO:0044424_intracellular_part	2280	170	-2.012929
GO:0048518_positive_regulation_of_biological_process	1955	165	-4.475021
GO:0000003_reproduction	1895	164	-5.072773
GO:0040008_regulation_of_growth	1886	159	-4.282779
GO:0009791_post-embryonic_development	1643	156	-7.303282
GO:0045927_positive_regulation_of_growth	1836	156	-4.380721
GO:0002119_nematode_larval_development	1625	155	-7.387346
GO:0002164_larval_development	1627	155	-7.351304
GO:0040010_positive_regulation_of_growth_rate	1727	147	-4.16627
GO:0040009_regulation_of_growth_rate	1728	147	-4.153675
GO:0043229_intracellular_organelle	1740	131	-1.745789
GO:0043226_organelle	1749	131	-1.679313
GO:0019538_protein_metabolic_process	1591	124	-2.126253
GO:0044260_cellular_macromolecule_metabolic_process	1588	122	-1.905378
GO:0044267_cellular_protein_metabolic_process	1536	121	-2.231382
GO:0048856_anatomical_structure_development	1280	111	-3.50053
GO:0000166_nucleotide_binding	1283	101	-1.918427
GO:0040011_locomotion	1157	96	-2.473693
GO:0017076_purine_nucleotide_binding	1136	91	-1.958153
GO:0005737_cytoplasm	1068	90	-2.535812
GO:0032553_ribonucleotide_binding	1066	86	-1.949427
GO:0032555_purine_ribonucleotide_binding	1066	86	-1.949427
GO:0009653_anatomical_structure_morphogenesis	879	74	-2.154856
GO:0022414_reproductive_process	860	71	-1.898912
GO:0043412_biopolymer_modification	771	65	-1.963447
GO:0006464_protein_modification_process	752	63	-1.862976
GO:0044444_cytoplasmic_part	612	60	-3.329858
GO:0032991_macromolecular_complex	721	59	-1.589923
GO:0043687_post-translational_protein_modification	663	58	-2.116703
GO:0048731_system_development	734	58	-1.319089
GO:0006793_phosphorus_metabolic_process	642	57	-2.2187
GO:0006796_phosphate_metabolic_process	642	57	-2.2187
GO:0048513_organ_development	692	56	-1.453249
GO:0003006_reproductive_developmental_process	594	51	-1.778961
GO:0007548_sex_differentiation	574	50	-1.861278
GO:0040035_hermaphrodite_genitalia_development	511	48	-2.396776
GO:0048806_genitalia_development	515	48	-2.331967
GO:0044422_organelle_part	553	46	-1.44585
GO:0043228_non-membrane-bounded_organelle	535	45	-1.494467
GO:0043232_intracellular_non-membrane-bounded_organelle	535	45	-1.494467
GO:0044446_intracellular_organelle_part	546	45	-1.368279

GO:0044249_cellular_biosynthetic_process	459	40	-1.591247
GO:0007610_behavior	389	34	-1.441578
GO:0048609_reproductive_process_in_a_multicellular_organism	303	30	-1.985031
GO:0032504_multicellular_organism_reproduction	304	30	-1.966002
GO:0007242_intracellular_signaling_cascade	330	29	-1.325089
GO:0009059_macromolecule_biosynthetic_process	283	28	-1.883608
GO:0018991_oviposition	295	28	-1.665163
GO:0033057_reproductive_behavior_in_a_multicellular_organism	295	28	-1.665163
GO:0019098_reproductive_behavior	300	28	-1.580555
GO:0035264_multicellular_organism_growth	309	28	-1.437145
GO:0040014_regulation_of_multicellular_organism_growth	309	28	-1.437145
GO:0051641_cellular_localization	294	27	-1.465144
GO:0008104_protein_localization	240	26	-2.272768
GO:0033036_macromolecule_localization	242	26	-2.2259
GO:0002009_morphogenesis_of_an_epithelium	254	25	-1.711476
GO:0040018_positive_regulation_of_multicellular_organism_growth	229	23	-1.706659
GO:0006412_translation	232	23	-1.647266
GO:0019001_guanyl_nucleotide_binding	172	22	-2.883252
GO:0005525_GTP_binding	161	21	-2.885566
GO:0032561_guanyl_ribonucleotide_binding	161	21	-2.885566
GO:0030529_ribonucleoprotein_complex	183	20	-1.908971
GO:0045184_establishment_of_protein_localization	189	20	-1.767266
GO:0042578_phosphoric_ester_hydrolase_activity	211	20	-1.323599
GO:0015031_protein_transport	184	19	-1.604922
GO:0005856_cytoskeleton	179	18	-1.447153
GO:0006470_protein_amino_acid_dephosphorylation	132	16	-2.024315
GO:0016311_dephosphorylation	135	16	-1.934891
GO:0004721_phosphoprotein_phosphatase_activity	136	16	-1.905929
GO:0003735_structural_constituent_of_ribosome	137	16	-1.877377
GO:0005840_ribosome	138	16	-1.84923
GO:0040012_regulation_of_locomotion	143	16	-1.71434
GO:0040017_positive_regulation_of_locomotion	119	13	-1.411057
GO:0031090_organelle_membrane	122	13	-1.337293
GO:0045165_cell_fate_commitment	84	12	-2.181534
GO:0005739_mitochondrion	80	10	-1.527578
GO:0001708_cell_fate_specification	57	9	-2.032311
GO:0006457_protein_folding	75	9	-1.323364
GO:0006732_coenzyme_metabolic_process	60	8	-1.45417
GO:0044429_mitochondrial_part	49	7	-1.467722
GO:0043284_biopolymer_biosynthetic_process	19	6	-3.06713
GO:0005921_gap_junction	25	6	-2.394387
GO:0014704_intercalated_disc	25	6	-2.394387
GO:0008135_translation_factor_activity__nucleic_acid_binding	35	6	-1.666156
GO:0045182_translation_regulator_activity	38	6	-1.504958
GO:0016917_GABA_receptor_activity	41	6	-1.362216
GO:0043067_regulation_of_programmed_cell_death	41	6	-1.362216
GO:0005911_intercellular_junction	42	6	-1.318225
GO:0006414_translational_elongation	14	5	-2.89794
GO:0000793_condensed_chromosome	28	5	-1.526935
GO:0008287_protein_serine_threonine_phosphatase_complex	14	4	-2.013085

GO:0015980_energy_derivation_by_oxidation_of_organic_compounds	18	4	-1.615541
GO:0033205_cytokinesis_during_cell_cycle	7	3	-2.131925
GO:0004722_protein_serine_threonine_phosphatase_activity	8	3	-1.948648
GO:0003746_translation_elongation_factor_activity	10	3	-1.659119
GO:0032153_cell_division_site	10	3	-1.659119
GO:0032154_cleavage_furrow	10	3	-1.659119
GO:0032155_cell_division_site_part	10	3	-1.659119
GO:0006099_tricarboxylic_acid_cycle	11	3	-1.541436
GO:0009060_aerobic_respiration	11	3	-1.541436
GO:0045333_cellular_respiration	11	3	-1.541436
GO:0046356_acetyl-CoA_catabolic_process	11	3	-1.541436
GO:0043070_regulation_of_non-apoptotic_programmed_cell_death	12	3	-1.437041
GO:0003887_DNA-directed_DNA_polymerase_activity	13	3	-1.343565
GO:0006084_acetyl-CoA_metabolic_process	13	3	-1.343565
GO:0009109_coenzyme_catabolic_process	13	3	-1.343565
GO:0016244_non-apoptotic_programmed_cell_death	13	3	-1.343565
GO:0051187_cofactor_catabolic_process	13	3	-1.343565
GO:0004016_adenylate_cyclase_activity	5	2	-1.449963
GO:0007143_female_meiosis	5	2	-1.449963
GO:0007187_G-protein_signaling__coupled_to_cyclic_nucleotide_second_messenger	5	2	-1.449963
GO:0007188_G-protein_signaling__coupled_to_cAMP_nucleotide_second_messenger	5	2	-1.449963
GO:0019933_cAMP-mediated_signaling	5	2	-1.449963
GO:0019935_cyclic-nucleotide-mediated_signaling	5	2	-1.449963
GO:0030104_water_homeostasis	5	2	-1.449963
GO:0043405_regulation_of_MAP_kinase_activity	5	2	-1.449963

HLH-29

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0008150_biological_process	8808	310	-4.207307
GO:0032501_multicellular_organismal_process	3705	166	-6.972812
GO:0032502_developmental_process	3651	163	-6.681953
GO:0007275_multicellular_organismal_development	3428	158	-7.394551
GO:0065007_biological_regulation	2922	124	-3.661856
GO:0050789_regulation_of_biological_process	2843	118	-3.069353
GO:0009790_embryonic_development	2693	117	-3.869317
GO:0040007_growth	2387	116	-6.205295
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	115	-3.809886
GO:0005622_intracellular	2808	107	-1.613919
GO:0048518_positive_regulation_of_biological_process	1955	93	-4.483244
GO:0040008_regulation_of_growth	1886	92	-4.870525
GO:0000003_reproduction	1895	91	-4.54299
GO:0045927_positive_regulation_of_growth	1836	89	-4.597618
GO:0002119_nematode_larval_development	1625	85	-5.681008
GO:0002164_larval_development	1627	85	-5.65894
GO:0009791_post-embryonic_development	1643	85	-5.484656
GO:0040010_positive_regulation_of_growth_rate	1727	85	-4.633011

GO:0040009_regulation_of_growth_rate	1728	85	-4.623489
GO:0048856_anatomical_structure_development	1280	64	-3.705399
GO:0040011_locomotion	1157	50	-1.736268
GO:0005737_cytoplasm	1068	46	-1.608726
GO:0022414_reproductive_process	860	44	-2.853894
GO:0009653_anatomical_structure_morphogenesis	879	39	-1.604739
GO:0048731_system_development	734	35	-1.9052
GO:0048513_organ_development	692	34	-2.047711
GO:0032991_macromolecular_complex	721	33	-1.591945
GO:0003006_reproductive_developmental_process	594	32	-2.534751
GO:0007548_sex_differentiation	574	31	-2.486938
GO:0044444_cytoplasmic_part	612	31	-2.087321
GO:0040035_hermaphrodite_genitalia_development	511	30	-2.971658
GO:0048806_genitalia_development	515	30	-2.918269
GO:0044422_organelle_part	553	26	-1.471022
GO:0044446_intracellular_organelle_part	546	25	-1.318597
GO:0048609_reproductive_process_in_a_multicellular_organism	303	18	-2.035098
GO:0032504_multicellular_organism_reproduction	304	18	-2.021177
GO:0002009_morphogenesis_of_an_epithelium	254	17	-2.452937
GO:0009059_macromolecule_biosynthetic_process	283	17	-1.994667
GO:0035264_multicellular_organism_growth	309	17	-1.653169
GO:0040014_regulation_of_multicellular_organism_growth	309	17	-1.653169
GO:0018991_oviposition	295	16	-1.53436
GO:0033057_reproductive_behavior_in_a_multicellular_organism	295	16	-1.53436
GO:0019098_reproductive_behavior	300	16	-1.476665
GO:0005525_GTP_binding	161	14	-3.159096
GO:0032561_guanyl_ribonucleotide_binding	161	14	-3.159096
GO:0019001_guanyl_nucleotide_binding	172	14	-2.87868
GO:0040018_positive_regulation_of_multicellular_organism_growth	229	13	-1.467001
GO:0006412_translation	232	13	-1.428321
GO:0008104_protein_localization	240	13	-1.329808
GO:0033036_macromolecule_localization	242	13	-1.306192
GO:0000279_M_phase	137	9	-1.483309
GO:0022403_cell_cycle_phase	142	9	-1.398607
GO:0065008_regulation_of_biological_quality	142	9	-1.398607
GO:0042592_homeostatic_process	85	8	-2.225755
GO:0016192_vesicle-mediated_transport	104	8	-1.723981
GO:0045165_cell_fate_commitment	84	7	-1.735365
GO:0043284_biopolymer_biosynthetic_process	19	6	-4.692126
GO:0008415_acyltransferase_activity	60	6	-1.906842
GO:0019725_cellular_homeostasis	64	6	-1.777652
GO:0006457_protein_folding	75	6	-1.474917
GO:0006414_translational_elongation	14	5	-4.275626
GO:0008135_translation_factor_activity__nucleic_acid_binding	35	5	-2.306798
GO:0048878_chemical_homeostasis	36	5	-2.253234
GO:0045182_translation_regulator_activity	38	5	-2.151636
GO:0022613_ribonucleoprotein_complex_biogenesis_and_assembly	52	5	-1.595399
GO:0006732_coenzyme_metabolic_process	60	5	-1.362169
GO:0016407_acetyltransferase_activity	24	4	-2.171132
GO:0005921_gap_junction	25	4	-2.106375

GO:0014704_intercalated_disc	25	4	-2.106375
GO:0008533_astacin_activity	41	4	-1.376582
GO:0005911_intercellular_junction	42	4	-1.343848
GO:0033205_cytokinesis_during_cell_cycle	7	3	-2.977836
GO:0003746_translation_elongation_factor_activity	10	3	-2.474056
GO:0008080_N-acetyltransferase_activity	22	3	-1.48947
GO:0016410_N-acyltransferase_activity	23	3	-1.438967
GO:0051258_protein_polymerization	24	3	-1.391152
GO:0055080_cation_homeostasis	25	3	-1.345793
GO:0007143_female_meiosis	5	2	-2.013423
GO:0030104_water_homeostasis	5	2	-2.013423
GO:0043405_regulation_of_MAP_kinase_activity	5	2	-2.013423
GO:0003705_RNA_polymerase_II_transcription_factor_activity__enhancer_binding	7	2	-1.709772
GO:0030728_ovulation	7	2	-1.709772
GO:0004129_cytochrome-c_oxidase_activity	8	2	-1.594079
GO:0006743_ubiquinone_metabolic_process	8	2	-1.594079
GO:0006744_ubiquinone_biosynthetic_process	8	2	-1.594079
GO:0015002_heme-copper_terminal_oxidase_activity	8	2	-1.594079
GO:0016675_oxidoreductase_activity__acting_on_heme_group_of_donors	8	2	-1.594079
GO:0016676_oxidoreductase_activity__acting_on_heme_group_of_donors__oxygen_as_acceptor	8	2	-1.594079
GO:0042375_quinone_cofactor_metabolic_process	8	2	-1.594079
GO:0045426_quinone_cofactor_biosynthetic_process	8	2	-1.594079
GO:0030005_cellular_di-__tri-valent_inorganic_cation_homeostasis	10	2	-1.406438
GO:0032153_cell_division_site	10	2	-1.406438
GO:0032154_cleavage_furrow	10	2	-1.406438
GO:0032155_cell_division_site_part	10	2	-1.406438
GO:0015464_acetylcholine_receptor_activity	11	2	-1.328456
GO:0030003_cellular_cation_homeostasis	11	2	-1.328456
GO:0042166_acetylcholine_binding	11	2	-1.328456
GO:0048546_digestive_tract_morphogenesis	11	2	-1.328456
GO:0055066_di-__tri-valent_inorganic_cation_homeostasis	11	2	-1.328456

HLH-30

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0003674_molecular_function	8423	697	-2.017732
GO:0008152_metabolic_process	4176	370	-2.496802
GO:0003824_catalytic_activity	3551	346	-5.924826
GO:0044237_cellular_metabolic_process	3372	309	-2.974913
GO:0044238_primary_metabolic_process	3216	294	-2.747489
GO:0043283_biopolymer_metabolic_process	1859	175	-2.177126
GO:0044260_cellular_macromolecule_metabolic_process	1588	150	-1.965251
GO:0044267_cellular_protein_metabolic_process	1536	148	-2.248618
GO:0019538_protein_metabolic_process	1591	148	-1.725343
GO:0016787_hydrolase_activity	1451	135	-1.610795
GO:0016740_transferase_activity	1098	111	-2.404356
GO:0030554_adenyl_nucleotide_binding	976	94	-1.578953
GO:0043412_biopolymer_modification	771	93	-4.633085
GO:0006464_protein_modification_process	752	91	-4.596523
GO:0005524_ATP_binding	916	90	-1.729857

GO:0032559_adenyl_ribonucleotide_binding	917	90	-1.718669
GO:0043687_post-translational_protein_modification	663	83	-4.767324
GO:0006793_phosphorus_metabolic_process	642	79	-4.298272
GO:0006796_phosphate_metabolic_process	642	79	-4.298272
GO:0016772_transferase_activity__transferring_phosphorus-containing_groups	622	73	-3.379342
GO:0016301_kinase_activity	531	63	-3.108459
GO:0016773_phosphotransferase_activity__alcohol_group_as_acceptor	505	58	-2.577493
GO:0016310_phosphorylation	497	55	-2.139457
GO:0004672_protein_kinase_activity	452	52	-2.380791
GO:0006468_protein_amino_acid_phosphorylation	444	51	-2.331825
GO:0016788_hydrolase_activity__acting_on_ester_bonds	426	49	-2.272783
GO:0004674_protein_serine_threonine_kinase_activity	411	46	-1.957772
GO:0042578_phosphoric_ester_hydrolase_activity	211	33	-3.897682
GO:0016791_phosphoric_monoester_hydrolase_activity	185	31	-4.266473
GO:0006082_organic_acid_metabolic_process	196	29	-3.093127
GO:0019752_carboxylic_acid_metabolic_process	196	29	-3.093127
GO:0006629_lipid_metabolic_process	233	27	-1.519513
GO:0004721_phosphoprotein_phosphatase_activity	136	24	-3.787757
GO:0006470_protein_amino_acid_dephosphorylation	132	23	-3.570789
GO:0016311_dephosphorylation	135	23	-3.424656
GO:0006807_nitrogen_compound_metabolic_process	159	23	-2.444194
GO:0009308_amine_metabolic_process	155	22	-2.26225
GO:0004725_protein_tyrosine_phosphatase_activity	121	21	-3.287292
GO:0006519_amino_acid_and_derivative_metabolic_process	148	21	-2.182392
GO:0016874_ligase_activity	146	19	-1.652261
GO:0006520_amino_acid_metabolic_process	129	18	-1.874562
GO:0007276_gamete_generation	116	17	-2.000067
GO:0031975_envelope	78	13	-2.0928
GO:0005739_mitochondrion	80	12	-1.628589
GO:0016879_ligase_activity__forming_carbon-nitrogen_bonds	87	12	-1.378191
GO:0031967_organelle_envelope	73	11	-1.545207
GO:0007292_female_gamete_generation	76	11	-1.431533
GO:0006913_nucleocytoplasmic_transport	52	10	-2.1524
GO:0051169_nuclear_transport	52	10	-2.1524
GO:0048477_oogenesis	70	10	-1.309391
GO:0032787_monocarboxylic_acid_metabolic_process	51	9	-1.747664
GO:0008238_exopeptidase_activity	53	9	-1.645581
GO:0007268_synaptic_transmission	57	9	-1.460143
GO:0016881_acid-amino_acid_ligase_activity	58	9	-1.417327
GO:0019787_small_conjugating_protein_ligase_activity	48	8	-1.469265
GO:0008134_transcription_factor_binding	38	7	-1.550675
GO:0022603_regulation_of_anatomical_structure_morphogenesis	23	6	-2.111871
GO:0006605_protein_targeting	32	6	-1.422853
GO:0031966_mitochondrial_membrane	33	6	-1.36474
GO:0006606_protein_import_into_nucleus	12	5	-2.81023
GO:0051170_nuclear_import	12	5	-2.81023
GO:0009064_glutamine_family_amino_acid_metabolic_process	17	5	-2.062577
GO:0048489_synaptic_vesicle_transport	18	5	-1.950078
GO:0017038_protein_import	19	5	-1.846211
GO:0008652_amino_acid_biosynthetic_process	22	5	-1.577046

GO:0001505_regulation_of_neurotransmitter_levels	23	5	-1.499059
GO:0004842_ubiquitin-protein_ligase_activity	23	5	-1.499059
GO:0007283_spermatogenesis	11	4	-2.080105
GO:0048232_male_gamete_generation	11	4	-2.080105
GO:0016860_intramolecular_oxidoreductase_activity	12	4	-1.931835
GO:0022604_regulation_of_cell_morphogenesis	12	4	-1.931835
GO:0009072_aromatic_amino_acid_family_metabolic_process	13	4	-1.799866
GO:0005643_nuclear_pore	14	4	-1.68138
GO:0046930_pore_complex	14	4	-1.68138
GO:0004198_calcium-dependent_cysteine-type_endopeptidase_activity	15	4	-1.574232
GO:0016903_oxidoreductase_activity__acting_on_the_aldehyde_or_oxo_group_of_donors	15	4	-1.574232
GO:0043623_cellular_protein_complex_assembly	15	4	-1.574232
GO:0010564_regulation_of_cell_cycle_process	16	4	-1.476751
GO:0006887_exocytosis	17	4	-1.387609
GO:0044453_nuclear_membrane_part	18	4	-1.305734
GO:0030151_molybdenum_ion_binding	5	3	-2.353515
GO:0004869_cysteine_protease_inhibitor_activity	6	3	-2.078865
GO:0008360_regulation_of_cell_shape	6	3	-2.078865
GO:0016079_synaptic_vesicle_exocytosis	6	3	-2.078865
GO:0016641_oxidoreductase_activity__acting_on_the_CH-NH2_group_of_donors__oxygen_as_acceptor	6	3	-2.078865
GO:0016861_intramolecular_oxidoreductase_activity__interconverting_aldoses_and_ketoses	7	3	-1.862089
GO:0046546_development_of_primary_male_sexual_characteristics	7	3	-1.862089
GO:0046661_male_sex_differentiation	7	3	-1.862089
GO:0051932_synaptic_transmission__GABAergic	7	3	-1.862089
GO:0016638_oxidoreductase_activity__acting_on_the_CH-NH2_group_of_donors	8	3	-1.684113
GO:0007218_neuropeptide_signaling_pathway	9	3	-1.534046
GO:0007270_nerve-nerve_synaptic_transmission	9	3	-1.534046
GO:0040022_feminization_of_hermaphroditic_germ-line	10	3	-1.405047

HLH-2/LIN-32

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0005575_cellular_component	4970	602	-3.512121
GO:0005488_binding	5055	590	-1.757538
GO:0005623_cell	4804	586	-3.748586
GO:0044464_cell_part	4788	584	-3.725366
GO:0032501_multicellular_organismal_process	3705	456	-3.048328
GO:0032502_developmental_process	3651	438	-2.038032
GO:0007275_multicellular_organismal_development	3428	415	-2.198174
GO:0005622_intracellular	2808	349	-2.567713
GO:0009790_embryonic_development	2693	337	-2.676414
GO:0009792_embryonic_development_ending_in_birth_or_egg_hatching	2645	329	-2.443957
GO:0043170_macromolecule_metabolic_process	2663	324	-1.850566
GO:0044424_intracellular_part	2280	284	-2.16352
GO:0043226_organelle	1749	219	-1.831338
GO:0043229_intracellular_organelle	1740	218	-1.835096
GO:0003676_nucleic_acid_binding	1510	198	-2.531601

GO:0043231_intracellular_membrane-bounded_organelle	1392	191	-3.417999
GO:0043227_membrane-bounded_organelle	1399	191	-3.30343
GO:0005634_nucleus	1121	168	-5.079146
GO:0006139_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	1228	155	-1.519222
GO:0040011_locomotion	1157	148	-1.644167
GO:0010467_gene_expression	1101	139	-1.410777
GO:0004872_receptor_activity	1061	134	-1.380474
GO:0050794_regulation_of_cellular_process	1029	131	-1.451061
GO:0003677_DNA_binding	873	128	-3.539742
GO:0009653_anatomical_structure_morphogenesis	879	115	-1.627649
GO:0016070_RNA_metabolic_process	837	111	-1.742537
GO:0016043_cellular_component_organization_and_biogenesis	804	110	-2.113892
GO:0010468_regulation_of_gene_expression	760	109	-2.749291
GO:0032991_macromolecular_complex	721	106	-3.054823
GO:0019222_regulation_of_metabolic_process	783	106	-1.925343
GO:0006350_transcription	747	105	-2.388503
GO:0031323_regulation_of_cellular_metabolic_process	755	105	-2.247729
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	724	103	-2.513991
GO:0045449_regulation_of_transcription	709	102	-2.647821
GO:0004888_transmembrane_receptor_activity	740	98	-1.580774
GO:0043234_protein_complex	549	93	-4.976706
GO:0030528_transcription_regulator_activity	647	92	-2.294054
GO:0032774_RNA_biosynthetic_process	646	90	-2.018667
GO:0006351_transcription__DNA-dependent	643	89	-1.930068
GO:0006355_regulation_of_transcription__DNA-dependent	628	87	-1.90748
GO:0051252_regulation_of_RNA_metabolic_process	634	87	-1.807126
GO:0044422_organelle_part	553	83	-2.759586
GO:0044446_intracellular_organelle_part	546	82	-2.740507
GO:0010171_body_morphogenesis	551	76	-1.689753
GO:0003700_transcription_factor_activity	539	72	-1.357591
GO:0006996_organelle_organization_and_biogenesis	418	63	-2.265408
GO:0005694_chromosome	200	44	-5.354823
GO:0044427_chromosomal_part	156	42	-7.686028
GO:0051276_chromosome_organization_and_biogenesis	161	42	-7.258139
GO:0022607_cellular_component_assembly	202	42	-4.511352
GO:0005216_ion_channel_activity	288	41	-1.31026
GO:0022838_substrate_specific_channel_activity	288	41	-1.31026
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	120	39	-9.763166
GO:0000785_chromatin	123	39	-9.404255
GO:0006333_chromatin_assembly_or_disassembly	111	38	-10.28504
GO:0065003_macromolecular_complex_assembly	178	38	-4.402151
GO:0006323_DNA_packaging	96	34	-9.724069
GO:0065004_protein-DNA_complex_assembly	104	34	-8.668833
GO:0031497_chromatin_assembly	89	33	-10.06869
GO:0003008_system_process	222	33	-1.361734
GO:0006334_nucleosome_assembly	85	32	-9.983647
GO:0000786_nucleosome	86	32	-9.829107
GO:0030001_metal_ion_transport	215	32	-1.341625
GO:0044459_plasma_membrane_part	198	30	-1.374677
GO:0005261_cation_channel_activity	172	27	-1.450142
GO:0007600_sensory_perception	128	22	-1.659363

GO:0007606_sensory_perception_of_chemical_stimulus	109	21	-2.146624
GO:0006813_potassium_ion_transport	115	19	-1.346755
GO:0005887_integral_to_plasma_membrane	99	17	-1.389811
GO:0031226_intrinsic_to_plasma_membrane	100	17	-1.353741
GO:0005244_voltage-gated_ion_channel_activity	74	14	-1.53722
GO:0022832_voltage-gated_channel_activity	74	14	-1.53722
GO:0001708_cell_fate_specification	57	11	-1.360515
GO:0008076_voltage-gated_potassium_channel_complex	57	11	-1.360515
GO:0008134_transcription_factor_binding	38	10	-2.183906
GO:0045138_tail_tip_morphogenesis	30	8	-1.875526
GO:0004180_carboxypeptidase_activity	31	8	-1.788655
GO:0003702_RNA_polymerase_II_transcription_factor_activity	33	8	-1.628404
GO:0004527_exonuclease_activity	33	8	-1.628404
GO:0035121_tail_morphogenesis	35	8	-1.484095
GO:0004245_neprilysin_activity	36	8	-1.417261
GO:0008158_hedgehog_receptor_activity	30	7	-1.395202
GO:0003712_transcription_cofactor_activity	24	6	-1.384662
GO:0000151_ubiquitin_ligase_complex	18	5	-1.396129
GO:0005319_lipid_transporter_activity	18	5	-1.396129
GO:0006869_lipid_transport	18	5	-1.396129
GO:0042692_muscle_cell_differentiation	18	5	-1.396129
GO:0031461_cullin-RING_ubiquitin_ligase_complex	9	4	-1.935878
GO:0030258_lipid_modification	11	4	-1.595493
GO:0005254_chloride_channel_activity	12	4	-1.458006
GO:0008060_ARF_GTPase_activator_activity	6	3	-1.688653
GO:0032312_regulation_of_ARF_GTPase_activity	6	3	-1.688653
GO:0004365_glyceraldehyde-3-phosphate_dehydrogenase_(phosphorylating)_activity	7	3	-1.482166
GO:0008943_glyceraldehyde-3-phosphate_dehydrogenase_activity	7	3	-1.482166
GO:0012502_induction_of_programmed_cell_death	7	3	-1.482166
GO:0017145_stem_cell_division	7	3	-1.482166
GO:0042078_germ-line_stem_cell_division	7	3	-1.482166
GO:0003997_acyl-CoA_oxidase_activity	8	3	-1.314354
GO:0008037_cell_recognition	8	3	-1.314354
GO:0008038_neuron_recognition	8	3	-1.314354

MDL-1/MXL-1

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0044237_cellular_metabolic_process	3372	782	-1.354231
GO:0044238_primary_metabolic_process	3216	751	-1.54768
GO:0043283_biopolymer_metabolic_process	1859	448	-1.817169
GO:0044260_cellular_macromolecule_metabolic_process	1588	382	-1.563098
GO:0043412_biopolymer_modification	771	196	-1.844939
GO:0006464_protein_modification_process	752	189	-1.609212
GO:0048513_organ_development	692	175	-1.619905
GO:0043687_post-translational_protein_modification	663	173	-2.121919
GO:0006793_phosphorus_metabolic_process	642	166	-1.905225
GO:0006796_phosphate_metabolic_process	642	166	-1.905225
GO:0003006_reproductive_developmental_process	594	154	-1.845748
GO:0007548_sex_differentiation	574	149	-1.821264
GO:0048806_genitalia_development	515	135	-1.8356
GO:0040035_hermaphrodite_genitalia_development	511	134	-1.831034

GO:0009056_catabolic_process	306	83	-1.658205
GO:0044248_cellular_catabolic_process	253	69	-1.513988
GO:0007049_cell_cycle	248	68	-1.547044
GO:0022402_cell_cycle_process	230	64	-1.613903
GO:0042578_phosphoric_ester_hydrolase_activity	211	63	-2.279446
GO:0016791_phosphoric_monoester_hydrolase_activity	185	59	-2.882965
GO:0009057_macromolecule_catabolic_process	190	53	-1.442962
GO:0004721_phosphoprotein_phosphatase_activity	136	49	-3.831556
GO:0006470_protein_amino_acid_dephosphorylation	132	48	-3.877282
GO:0016311_dephosphorylation	135	48	-3.609383
GO:0004725_protein_tyrosine_phosphatase_activity	121	41	-2.707323
GO:0044262_cellular_carbohydrate_metabolic_process	117	36	-1.727661
GO:0031090_organelle_membrane	122	36	-1.457538
GO:0000278_mitotic_cell_cycle	113	33	-1.31864
GO:0007369_gastrulation	81	29	-2.460716
GO:0016052_carbohydrate_catabolic_process	70	24	-1.874941
GO:0045137_development_of_primary_sexual_characteristics	77	24	-1.376469
GO:0001703_gastrulation_with_mouth_forming_first	60	23	-2.481713
GO:0044275_cellular_carbohydrate_catabolic_process	67	23	-1.82474
GO:0006732_coenzyme_metabolic_process	60	21	-1.808086
GO:0007051_spindle_organization_and_biogenesis	64	21	-1.490998
GO:0007052_mitotic_spindle_organization_and_biogenesis	58	20	-1.672808
GO:0016810_hydrolase_activity__acting_on_carbon-nitrogen_(but_not_peptide)_bonds	59	20	-1.591394
GO:0005976_polysaccharide_metabolic_process	53	19	-1.797553
GO:0044264_cellular_polysaccharide_metabolic_process	53	19	-1.797553
GO:0006030_chitin_metabolic_process	45	17	-1.898321
GO:0006040_amino_sugar_metabolic_process	47	17	-1.698556
GO:0006041_glucosamine_metabolic_process	47	17	-1.698556
GO:0006044_N-acetylglucosamine_metabolic_process	47	17	-1.698556
GO:0032940_secretion_by_cell	50	17	-1.435095
GO:0045045_secretory_pathway	35	14	-1.880192
GO:0005819_spindle	33	13	-1.71929
GO:0000272_polysaccharide_catabolic_process	36	13	-1.40043
GO:0004568_chitinase_activity	36	13	-1.40043
GO:0006032_chitin_catabolic_process	36	13	-1.40043
GO:0006043_glucosamine_catabolic_process	36	13	-1.40043
GO:0006046_N-acetylglucosamine_catabolic_process	36	13	-1.40043
GO:0044247_cellular_polysaccharide_catabolic_process	36	13	-1.40043
GO:0046348_amino_sugar_catabolic_process	36	13	-1.40043
GO:0031966_mitochondrial_membrane	33	12	-1.348288
GO:0005743_mitochondrial_inner_membrane	27	11	-1.635954
GO:0006887_exocytosis	17	9	-2.271112
GO:0048489_synaptic_vesicle_transport	18	9	-2.064008
GO:0008287_protein_serine_threonine_phosphatase_complex	14	8	-2.326366
GO:0009064_glutamine_family_amino_acid_metabolic_process	17	8	-1.695716
GO:0007283_spermatogenesis	11	7	-2.436139
GO:0048232_male_gamete_generation	11	7	-2.436139
GO:0009072_aromatic_amino_acid_family_metabolic_process	13	7	-1.903167
GO:0009451_RNA_modification	13	7	-1.903167
GO:0016903_oxidoreductase_activity__acting_on_the_ald	15	7	-1.510283

ehyde_or_oxo_group_of_donors			
GO:0006904_vesicle_docking_during_exocytosis	8	6	-2.663235
GO:0007218_neuropeptide_signaling_pathway	9	6	-2.276422
GO:0022406_membrane_docking	9	6	-2.276422
GO:0048278_vesicle_docking	9	6	-2.276422
GO:0006084_acetyl-CoA_metabolic_process	13	6	-1.321925
GO:0009109_coenzyme_catabolic_process	13	6	-1.321925
GO:0012506_vesicle_membrane	13	6	-1.321925
GO:0051187_cofactor_catabolic_process	13	6	-1.321925
GO:0046546_development_of_primary_male_sexual_characteristics	7	5	-2.127406
GO:0046661_male_sex_differentiation	7	5	-2.127406
GO:0004722_protein_serine_threonine_phosphatase_activity	8	5	-1.788679
GO:0016638_oxidoreductase_activity__acting_on_the_CH-NH2_group_of_donors	8	5	-1.788679
GO:0043073_germ_cell_nucleus	8	5	-1.788679
GO:0007200_G-protein_signaling__coupled_to_IP3_second_messenger_(phospholipase_C_activating)	9	5	-1.52306
GO:0030120_vesicle_coat	9	5	-1.52306
GO:0030659_cytoplasmic_vesicle_membrane	9	5	-1.52306
GO:0030662_coated_vesicle_membrane	9	5	-1.52306
GO:0042770_DNA_damage_response__signal_transduction	9	5	-1.52306
GO:0006402_mRNA_catabolic_process	10	5	-1.307886
GO:0009110_vitamin_biosynthetic_process	10	5	-1.307886
GO:0042364_water-soluble_vitamin_biosynthetic_process	10	5	-1.307886
GO:0030976_thiamin_pyrophosphate_binding	5	4	-2.003957
GO:0004448_isocitrate_dehydrogenase_activity	6	4	-1.610647
GO:0006917_induction_of_apoptosis	6	4	-1.610647
GO:0008060_ARF_GTPase_activator_activity	6	4	-1.610647
GO:0008629_induction_of_apoptosis_by_intracellular_signals	6	4	-1.610647
GO:0008630_DNA_damage_response__signal_transduction_resulting_in_induction_of_apoptosis	6	4	-1.610647
GO:0016211_ammonia_ligase_activity	6	4	-1.610647
GO:0016641_oxidoreductase_activity__acting_on_the_CH-NH2_group_of_donors__oxygen_as_acceptor	6	4	-1.610647
GO:0016880_acid-ammonia_(or_amide)_ligase_activity	6	4	-1.610647
GO:0032312_regulation_of_ARF_GTPase_activity	6	4	-1.610647
GO:0012502_induction_of_programmed_cell_death	7	4	-1.325617
GO:0030728_ovulation	7	4	-1.325617
GO:0051087_chaperone_binding	7	4	-1.325617
GO:0051932_synaptic_transmission__GABAergic	7	4	-1.325617

MXL-3

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0003674_molecular_function	8423	586	-3.089456
GO:0009987_cellular_process	4977	357	-2.00616
GO:0008152_metabolic_process	4176	315	-3.17905
GO:0003824_catalytic_activity	3551	303	-8.077067
GO:0044237_cellular_metabolic_process	3372	258	-2.878978
GO:0044238_primary_metabolic_process	3216	247	-2.84082

GO:0043283_biopolymer_metabolic_process	1859	140	-1.429268
GO:0044260_cellular_macromolecule_metabolic_process	1588	124	-1.729896
GO:0019538_protein_metabolic_process	1591	122	-1.493006
GO:0044267_cellular_protein_metabolic_process	1536	121	-1.806387
GO:0016740_transferase_activity	1098	92	-2.125625
GO:0017076_purine_nucleotide_binding	1136	90	-1.504063
GO:0032553_ribonucleotide_binding	1066	85	-1.50373
GO:0032555_purine_ribonucleotide_binding	1066	85	-1.50373
GO:0043412_biopolymer_modification	771	80	-4.669794
GO:0030554_adenyl_nucleotide_binding	976	80	-1.692051
GO:0006464_protein_modification_process	752	78	-4.557738
GO:0043687_post-translational_protein_modification	663	75	-5.747817
GO:0005524_ATP_binding	916	75	-1.605281
GO:0032559_adenyl_ribonucleotide_binding	917	75	-1.595624
GO:0006793_phosphorus_metabolic_process	642	72	-5.398527
GO:0006796_phosphate_metabolic_process	642	72	-5.398527
GO:0016772_transferase_activity__transferring_phosphorus-containing_groups	622	66	-4.222939
GO:0016301_kinase_activity	531	56	-3.590667
GO:0016773_phosphotransferase_activity__alcohol_group_as_acceptor	505	51	-2.891475
GO:0016491_oxidoreductase_activity	566	51	-1.906936
GO:0016310_phosphorylation	497	50	-2.809528
GO:0006468_protein_amino_acid_phosphorylation	444	46	-2.883955
GO:0004672_protein_kinase_activity	452	46	-2.726625
GO:0016788_hydrolase_activity__acting_on_ester_bonds	426	44	-2.755887
GO:0004674_protein_serine_threonine_kinase_activity	411	42	-2.56907
GO:0009056_catabolic_process	306	33	-2.488363
GO:0004713_protein-tyrosine_kinase_activity	324	33	-2.112419
GO:0044248_cellular_catabolic_process	253	30	-2.93866
GO:0042578_phosphoric_ester_hydrolase_activity	211	29	-3.932518
GO:0016791_phosphoric_monoester_hydrolase_activity	185	27	-4.147818
GO:0006082_organic_acid_metabolic_process	196	24	-2.645986
GO:0019752_carboxylic_acid_metabolic_process	196	24	-2.645986
GO:0005975_carbohydrate_metabolic_process	222	24	-1.96814
GO:0006629_lipid_metabolic_process	233	23	-1.490568
GO:0006807_nitrogen_compound_metabolic_process	159	22	-3.163177
GO:0048037_cofactor_binding	205	22	-1.8133
GO:0004721_phosphoprotein_phosphatase_activity	136	21	-3.698753
GO:0006519_amino_acid_and_derivative_metabolic_process	148	21	-3.190354
GO:0009308_amine_metabolic_process	155	21	-2.92632
GO:0006470_protein_amino_acid_dephosphorylation	132	20	-3.439131
GO:0016311_dephosphorylation	135	20	-3.309118
GO:0006520_amino_acid_metabolic_process	129	19	-3.140481
GO:0044265_cellular_macromolecule_catabolic_process	164	18	-1.662259
GO:0004725_protein_tyrosine_phosphatase_activity	121	16	-2.256424
GO:0050662_coenzyme_binding	138	16	-1.731733
GO:0016874_ligase_activity	146	16	-1.526701
GO:0019842_vitamin_binding	93	13	-2.12724
GO:0006066_alcohol_metabolic_process	102	13	-1.803936
GO:0044262_cellular_carbohydrate_metabolic_process	117	13	-1.369208
GO:0016853_isomerase_activity	66	12	-2.96938
GO:0016879_ligase_activity__forming_carbon-nitrogen_bonds	87	11	-1.577763
GO:0005739_mitochondrion	80	10	-1.446673

GO:0019318_hexose_metabolic_process	48	9	-2.44917
GO:0005996_monosaccharide_metabolic_process	49	9	-2.386439
GO:0007267_cell-cell_signaling	70	9	-1.413975
GO:0019787_small_conjugating_protein_ligase_activity	48	8	-1.919635
GO:0016881_acid-amino_acid_ligase_activity	58	8	-1.461532
GO:0008652_amino_acid_biosynthetic_process	22	6	-2.629088
GO:0009310_amine_catabolic_process	28	6	-2.071815
GO:0044270_nitrogen_compound_catabolic_process	28	6	-2.071815
GO:0009309_amine_biosynthetic_process	31	6	-1.853287
GO:0015669_gas_transport	31	6	-1.853287
GO:0015671_oxygen_transport	31	6	-1.853287
GO:0044271_nitrogen_compound_biosynthetic_process	31	6	-1.853287
GO:0019825_oxygen_binding	32	6	-1.787191
GO:0016616_oxidoreductase_activity__acting_on_the_CH-OH_group_of_donors__NAD_or_NADP_as_acceptor	40	6	-1.351449
GO:0009072_aromatic_amino_acid_family_metabolic_process	13	5	-3.004165
GO:0009064_glutamine_family_amino_acid_metabolic_process	17	5	-2.417648
GO:0004842_ubiquitin-protein_ligase_activity	23	5	-1.824064
GO:0009063_amino_acid_catabolic_process	24	5	-1.746082
GO:0000287_magnesium_ion_binding	28	5	-1.475464
GO:0006512_ubiquitin_cycle	28	5	-1.475464
GO:0016860_intramolecular_oxidoreductase_activity	12	4	-2.225699
GO:0008287_protein_serine_threonine_phosphatase_complex	14	4	-1.965503
GO:0042692_muscle_cell_differentiation	18	4	-1.570768
GO:0048489_synaptic_vesicle_transport	18	4	-1.570768
GO:0030151_molybdenum_ion_binding	5	3	-2.594352
GO:0030976_thiamin_pyrophosphate_binding	5	3	-2.594352
GO:0004448_isocitrate_dehydrogenase_activity	6	3	-2.314997
GO:0016861_intramolecular_oxidoreductase_activity__interconverting_aldoses_and_ketoses	7	3	-2.093556
GO:0004722_protein_serine_threonine_phosphatase_activity	8	3	-1.910956
GO:0019319_hexose_biosynthetic_process	10	3	-1.622765
GO:0031625_ubiquitin_protein_ligase_binding	10	3	-1.622765
GO:0046165_alcohol_biosynthetic_process	10	3	-1.622765
GO:0046364_monosaccharide_biosynthetic_process	10	3	-1.622765
GO:0006099_tricarboxylic_acid_cycle	11	3	-1.505743
GO:0007283_spermatogenesis	11	3	-1.505743
GO:0009060_aerobic_respiration	11	3	-1.505743
GO:0009084_glutamine_family_amino_acid_biosynthetic_process	11	3	-1.505743
GO:0045333_cellular_respiration	11	3	-1.505743
GO:0046356_acetyl-CoA_catabolic_process	11	3	-1.505743
GO:0048232_male_gamete_generation	11	3	-1.505743
GO:0006071_glycerol_metabolic_process	13	3	-1.309177
GO:0006084_acetyl-CoA_metabolic_process	13	3	-1.309177
GO:0009109_coenzyme_catabolic_process	13	3	-1.309177
GO:0019751_polyol_metabolic_process	13	3	-1.309177
GO:0051187_cofactor_catabolic_process	13	3	-1.309177
GO:0006072_glycerol-3-phosphate_metabolic_process	5	2	-1.424391
GO:0006094_gluconeogenesis	5	2	-1.424391
GO:0006560_proline_metabolic_process	5	2	-1.424391
GO:0007026_negative_regulation_of_microtubule_depoly	5	2	-1.424391

merization			
GO:0009374_biotin_binding	5	2	-1.424391
GO:0016854_racemase_and_epimerase_activity	5	2	-1.424391
GO:0030055_cell-substrate_junction	5	2	-1.424391
GO:0031110_regulation_of_microtubule_polymerization_or_depolymerization	5	2	-1.424391
GO:0031111_negative_regulation_of_microtubule_polymerization_or_depolymerization	5	2	-1.424391
GO:0031114_regulation_of_microtubule_depolymerization	5	2	-1.424391
GO:0032886_regulation_of_microtubule-based_process	5	2	-1.424391

REF-1

GO CATEGORY	TOTAL GENES	TARGET GENES	LOG10(p)
GO:0003674_molecular_function	8423	470	-1.804929
GO:0009987_cellular_process	4977	295	-2.213976
GO:0008152_metabolic_process	4176	259	-3.06592
GO:0003824_catalytic_activity	3551	249	-7.22154
GO:0044237_cellular_metabolic_process	3372	207	-2.145293
GO:0044238_primary_metabolic_process	3216	199	-2.216952
GO:0044260_cellular_macromolecule_metabolic_process	1588	102	-1.645129
GO:0019538_protein_metabolic_process	1591	100	-1.398088
GO:0044267_cellular_protein_metabolic_process	1536	99	-1.644548
GO:0016787_hydrolase_activity	1451	94	-1.633452
GO:0000166_nucleotide_binding	1283	86	-1.867948
GO:0017076_purine_nucleotide_binding	1136	80	-2.280431
GO:0032553_ribonucleotide_binding	1066	75	-2.154812
GO:0032555_purine_ribonucleotide_binding	1066	75	-2.154812
GO:0016740_transferase_activity	1098	75	-1.859755
GO:0030554_adenyl_nucleotide_binding	976	71	-2.403511
GO:0005524_ATP_binding	916	66	-2.172803
GO:0032559_adenyl_ribonucleotide_binding	917	66	-2.16232
GO:0006464_protein_modification_process	752	61	-3.230948
GO:0043412_biopolymer_modification	771	61	-2.959117
GO:0043687_post-translational_protein_modification	663	60	-4.485026
GO:0006793_phosphorus_metabolic_process	642	58	-4.32935
GO:0006796_phosphate_metabolic_process	642	58	-4.32935
GO:0016772_transferase_activity__transferring_phosphorus-containing_groups	622	56	-4.155174
GO:0016301_kinase_activity	531	48	-3.67278
GO:0016773_phosphotransferase_activity__alcohol_group_as_acceptor	505	43	-2.819066
GO:0016310_phosphorylation	497	41	-2.463102
GO:0006468_protein_amino_acid_phosphorylation	444	37	-2.344385
GO:0004672_protein_kinase_activity	452	37	-2.219881
GO:0004674_protein_serine_threonine_kinase_activity	411	34	-2.149186
GO:0016788_hydrolase_activity__acting_on_ester_bonds	426	33	-1.709584
GO:0004713_protein-tyrosine_kinase_activity	324	29	-2.369058
GO:0009056_catabolic_process	306	24	-1.42908
GO:0044248_cellular_catabolic_process	253	22	-1.789953
GO:0006082_organic_acid_metabolic_process	196	21	-2.752497
GO:0019752_carboxylic_acid_metabolic_process	196	21	-2.752497
GO:0042578_phosphoric_ester_hydrolase_activity	211	20	-2.050702
GO:0016791_phosphoric_monoester_hydrolase_activity	185	19	-2.338912
GO:0005975_carbohydrate_metabolic_process	222	19	-1.551421

GO:0006807_nitrogen_compound_metabolic_process	159	17	-2.322239
GO:0048037_cofactor_binding	205	17	-1.329747
GO:0006520_amino_acid_metabolic_process	129	16	-2.869336
GO:0006519_amino_acid_and_derivative_metabolic_process	148	16	-2.262837
GO:0009308_amine_metabolic_process	155	16	-2.073021
GO:0006470_protein_amino_acid_dephosphorylation	132	15	-2.354375
GO:0016311_dephosphorylation	135	15	-2.263178
GO:0004721_phosphoprotein_phosphatase_activity	136	15	-2.233571
GO:0050662_coenzyme_binding	138	13	-1.487841
GO:0016874_ligase_activity	146	13	-1.320522
GO:0004725_protein_tyrosine_phosphatase_activity	121	12	-1.55863
GO:0016853_isomerase_activity	66	9	-2.098082
GO:0007267_cell-cell_signaling	70	8	-1.494461
GO:0008238_exopeptidase_activity	53	7	-1.665565
GO:0019787_small_conjugating_protein_ligase_activity	48	6	-1.38735
GO:0000287_magnesium_ion_binding	28	5	-1.824216
GO:0004180_carboxypeptidase_activity	31	5	-1.643159
GO:0008235_metalloexopeptidase_activity	32	5	-1.588209
GO:0016860_intramolecular_oxidoreductase_activity	12	4	-2.552078
GO:0009072_aromatic_amino_acid_family_metabolic_process	13	4	-2.410895
GO:0042692_muscle_cell_differentiation	18	4	-1.87151
GO:0004181_metallocarboxypeptidase_activity	21	4	-1.634926
GO:0004182_carboxypeptidase_A_activity	21	4	-1.634926
GO:0008652_amino_acid_biosynthetic_process	22	4	-1.56595
GO:0022603_regulation_of_anatomical_structure_morphogenesis	23	4	-1.50111
GO:0009063_amino_acid_catabolic_process	24	4	-1.440023
GO:0016861_intramolecular_oxidoreductase_activity__interconverting_aldoses_and_ketoses	7	3	-2.34803
GO:0019319_hexose_biosynthetic_process	10	3	-1.865066
GO:0046165_alcohol_biosynthetic_process	10	3	-1.865066
GO:0046364_monosaccharide_biosynthetic_process	10	3	-1.865066
GO:0016564_transcription_repressor_activity	12	3	-1.636335
GO:0016323_basolateral_plasma_membrane	14	3	-1.451959
GO:0005912_adherens_junction	15	3	-1.372121
GO:0004555_alpha_alpha-trehalase_activity	5	2	-1.593966
GO:0006094_gluconeogenesis	5	2	-1.593966
GO:0007143_female_meiosis	5	2	-1.593966
GO:0009374_biotin_binding	5	2	-1.593966
GO:0015927_trehalase_activity	5	2	-1.593966
GO:0030055_cell-substrate_junction	5	2	-1.593966
GO:0030151_molybdenum_ion_binding	5	2	-1.593966
GO:0030976_thiamin_pyrophosphate_binding	5	2	-1.593966
GO:0006090_pyruvate_metabolic_process	6	2	-1.433319
GO:0007050_cell_cycle_arrest	6	2	-1.433319
GO:0008360_regulation_of_cell_shape	6	2	-1.433319
GO:0045747_positive_regulation_of_Notch_signaling_pathway	6	2	-1.433319
GO:0048500_signal_recognition_particle	6	2	-1.433319
GO:0005759_mitochondrial_matrix	7	2	-1.30256
GO:0005984_disaccharide_metabolic_process	7	2	-1.30256
GO:0005991_trehalose_metabolic_process	7	2	-1.30256
GO:0006563_L-serine_metabolic_process	7	2	-1.30256
GO:0009070_serine_family_amino_acid_biosynthetic_pro	7	2	-1.30256

cess			
GO:0031980_mitochondrial_lumen	7	2	-1.30256
GO:0033205_cytokinesis_during_cell_cycle	7	2	-1.30256
GO:0043057_backward_locomotion	7	2	-1.30256
GO:0046546_development_of_primary_male_sexual_characteristics	7	2	-1.30256
GO:0046661_male_sex_differentiation	7	2	-1.30256
GO:0051932_synaptic_transmission__GABAergic	7	2	-1.30256

PREFACE TO CHAPTER V

This chapter describes an integrated network of *C. elegans* bHLH dimers with their associated TF parameters: dimerization, spatiotemporal expression, DNA binding specificity, and functional annotation. Analysis of the network reveals that, despite similarities in individual parameters, individual bHLH proteins and dimers are quite distinct from one another when many parameters are considered.

Much of this chapter has been published separately in:

Grove C. A., De Masi F., Barrasa M. I., Newburger D. E., Alkema M. J., Bulyk M. L., Walhout A. J. M. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell. 2009 Jul 23; 138(2): 314-327.

CHAPTER V

A Multiparameter Integrated Network of *C. elegans* bHLH Transcription Factors

Abstract

We have combined the data presented in the previous chapters for dimerization, spatiotemporal expression, and DNA binding specificity of the *C. elegans* bHLH TFs combined with the lists of candidate target genes and predicted functional annotations into a single, integrated network. As described in Chapter I, such integrated networks will likely provide important insight into the functionality of TFs, transcription regulatory networks, and perhaps other protein families and their relevant networks as well. In this chapter, we describe some features of the integrated network, and provide *in vivo* support for our network by identifying target genes of HLH-30 by assessing the gene expression profile of an *hlh-30* deletion mutant. Additionally we calculate similarity scores for each bHLH-bHLH pair and TF parameter to measure the extent to which the bHLH TFs in *C. elegans* have diverged from each other.

Introduction

The incorporation of multiple TF parameters into transcription regulatory networks (TRNs) can provide clarity and understanding of such networks, not only because of the cognitive simplicity of viewing data in this manner, but also because different data types are often so reliant on each other. For example, determination of the DNA binding specificities of the bHLH TFs from *C. elegans* required knowledge of which bHLH proteins dimerize with which other bHLH proteins. Without such knowledge, biochemistry of individual bHLH dimer-complexes would be difficult, if not impossible. Thus, the incorporation of dimerization specificity is crucial to approaching the question of systematic bHLH DNA binding specificity determination. Likewise, understanding which bHLH dimers form requires knowledge of when and where different bHLH proteins are co-expressed, since two bHLH proteins could form strong dimers *in vitro* (or, in our case, in the yeast two hybrid system), but if they never encounter each other *in vivo*, the dimer will be non-existent and, hence, irrelevant.

Although our datasets end with spatiotemporal analysis, it is reasonable to suggest that complete understanding of spatiotemporal expression requires the knowledge of post-transcriptional gene regulation, among, perhaps, other aspects of gene function and regulation. Our data do, however, provide a substantial step towards the systematic determination of how various TF parameters contribute to the functionality of TFs in a family-wide manner.

To test the integrity of the information provided by the network, we have performed microarray analysis of the gene expression profile of *C. elegans* mutants in which the bHLH-domain-encoding portion of the *hlh-30* locus has been deleted, presumably rendering the HLH-30 protein non-functional, at least with respect to the DNA binding properties of the protein. We chose *hlh-30* because it is expressed strongly in the intestine of the worm for most of its life, and since the intestine is the largest organ volumetrically, we considered it a good candidate for studying the effects of its loss. Overall gene expression changes assessed by total RNA from worms in which the mutation affects only a small portion of the total RNA may be more difficult to detect. Our dimerization, spatiotemporal expression, and DNA binding data suggest that HLH-30 functions as a homodimer predominately in the intestine, spermatheca, vulva, and to a lesser extent in muscle, hypodermis, pharynx, and small number of neurons. Our data also indicate that HLH-30 likely targets genes predominately involved in metabolism (lipid metabolism and amino acid metabolism), signaling (kinase and phosphatase activity), and reproduction (oogenesis, spermatogenesis). The results of the microarray, as explained below, support our predictions of target genes for HLH-30, both in terms of the actual target genes as well as the enriched Gene Ontology (GO) terms, providing *in vivo* support for our network.

The ultimate goal of this work has been to determine the extent to which the *C. elegans* bHLH TFs have functionally diverged from each other with respect to various TF parameters. To assess the degree to which any bHLH TF

is distinct from all other bHLH TFs, we formulated a simple calculation, called a Similarity Score (SS), to measure overlap in parameter associations. This score is simply the overlap in the number of parameter associations divided by the union of all parameter associations exhibited by both TFs. We find that, although many of the *C. elegans* bHLH TFs are apparently similar in one or more parameters, they are actually quite distinct. This is most obvious when we compare lists of predicted target genes for bHLH TFs that initially appear to have very similar DNA binding specificities. The target genes of several bHLH TFs that bind to CACGTG E-Boxes were compared and found predominately to be distinct from one another (SS < ~0.30). This distinction in target genes as a result of apparently subtle DNA binding specificities could help explain how various TFs from the same family can achieve distinct biological functions despite their molecular similarities.

Results

An Integrated bHLH Dimerization, DNA Binding and Expression Network

We assembled all separately measured functional bHLH parameters into the first integrated network for any TF family, combining dimerization, spatiotemporal expression patterns, DNA binding specificities, and enriched GO annotations of candidate target genes (Figure V-1).

As discussed in previous chapters, all the nodes, *i.e.* dimers, tissues and DNA binding sequences, exhibit specificity and promiscuity in this network. In addition, we observed specificity and promiscuity for the different GO categories: some are associated with few bHLH dimers, whereas others are associated with many. For instance, “cell division” is associated only with MDL-1/MXL-1 and HLH-25, whereas “development” is associated with 11 different dimers (Figure V-1). Conversely, some bHLH dimers are associated with few categories, whereas others are associated with many; HLH-1 is connected solely to “development”, but HLH-25 is connected to nine different GO terms. However, it is important to note that “development” can be divided into “embryonic development”, “larval development” and several other terms that exhibit only partial overlap between different bHLH dimers. Similarly, “signaling”, “metabolism” and “reproduction” can be divided into more specific terms that enable the further differentiation between distinct bHLH dimers (Figure V-2 and Figure V-10).

HLH-30: in vivo *Validation of the Network*

To assess the validity of our integrated network, we focused on HLH-30, for which we had a viable deletion mutant available [*hlh-30(tm1978)*]. HLH-30 is strongly expressed in the intestine and weakly in other tissues (Figure V-3). This enables the identification of downstream target genes by expression profiling *in vivo* (*i.e.* this would be more difficult for bHLH TFs that exhibit more restricted expression patterns). RNAi knockdown of *hlh-30* leads to a reduced fat phenotype (162). Our integrated bHLH network contains a unique path that connects HLH-30 to the intestine, the main organ of fat storage, and to the GO categories “metabolism”, “reproduction” and “signaling” (Figure V-1). HLH-30 specifically binds CACGTG E-boxes, and favors a flanking 5’ T (Figure V-4). This leads to the prediction that HLH-30 regulates (fat) metabolism in the intestine by binding target genes that contain HLH-30-bound CACGTG E-boxes in their promoter.

To test this prediction, we performed gene expression profiling of wild type and *hlh-30(tm1978)* mutant animals and compared the resulting expression data. We identified 134 genes that were significantly differentially expressed: 122 exhibited decreased, and 12 exhibited increased expression in the mutant (Figure V-5, Table V-1). This suggests that HLH-30 is primarily a transcriptional activator, which is in agreement with our observation that it is a strong auto-activator in Y2H assays (Figure II-1). We refer to all genes that change in expression in the *hlh-30(tm1978)* mutant as “HLH-30 target genes”, although

some may change in expression due to indirect effects rather than direct regulation by HLH-30.

HLH-30 target genes more frequently possess an HLH-30 binding site in 500 bp promoter sequences than non-target genes (Figure V-6; Fisher's exact test $p = 1.9 \times 10^{-9}$). The consistency between the PBM-derived and experimentally identified HLH-30 target genes supports our overall approach for identifying candidate bHLH target genes using PBM data. When we searched genomic sequences downstream of the transcriptional start, we also observed an increase in HLH-30 binding sites in targets versus non-targets, albeit less significantly (Figure V-7; Fisher's exact test $p = 0.007$). Finally, we found that HLH-30 targets significantly more frequently possess multiple HLH-30 binding sites than non-targets (Figure V-8, chi-square test $p = 2.2 \times 10^{-16}$).

Next, we examined the experimentally determined HLH-30 target genes for over-represented GO terms, and found enrichment for various metabolic, as well as aging terms (Table V-2). Interestingly, the human ortholog of HLH-30, TFE3, has been reported to activate metabolic genes through E-boxes as well (111). This suggests that both the molecular and biological functions of HLH-30 are evolutionarily conserved.

We have likely underestimated the number of *in vivo* HLH-30 target genes because only changes in genes that are broadly or highly expressed can be detected in whole animal gene expression analysis. Thus, it is more difficult to evaluate the association of HLH-30 with the GO term "reproduction" even though

Phlh-30 drives expression in the spermatheca and the vulva (Figure V-3). Nevertheless, the whole animal gene expression analysis does provide support for our overall method and approach.

Multiparameter Analysis of bHLH TFs – Similarity Scores

To examine the overall extent to which bHLH TFs differ from each other we compared all possible 861 bHLH-bHLH pairs. We derived a Similarity Score (SS) for each pair and for each parameter (Figure V-9), clustered the bHLH TFs and dimers according to these scores, and visualized these as heat maps, resulting in one heat map per parameter (Figure V-10). We observed that for each parameter the majority of the pairs have a low SS. For instance, more than 80% of the bHLH-bHLH pairs share fewer than 25% of their target genes (Figure V-9). We observed the lowest degree of divergence in spatial expression; however, this is likely because not all expression could be resolved to the level of individual cells.

HLH-4, HLH-10, HLH-15, and HLH-19 – A Closer Look

Several bHLH-bHLH pairs are more similar in one or more parameter than most other pairs. A sub-network of the most similar bHLH TFs is shown in Figure V-11. These all share HLH-2 as their dimerization partner and, for clarity, heterodimers are depicted as single nodes. The parameter comparisons among these dimers are provided in Figure V-12. Several observations can be made from this analysis. First, several tissues and GO categories can be connected by paths that go through these different dimers. For instance, head neurons can be

connected to sensory perception via both HLH-2/HLH-4 and HLH-2/HLH10. We refer to such similar connections as “network paths”. In fact, we found that HLH-4 and HLH-10 share ~40% of their network paths in the integrated network (SS = 0.43, Figure V-1). This suggests that they may be highly similar in various TF parameters. Indeed, they share more than 50% of each of the parameters measured (SS = 0.52 – 0.67, Figure V-12). HLH-15 and HLH-19 also share ~40% of their network paths in the integrated network (SS = 0.4, Figure V-1). These two dimers connect head and tail neurons to chromatin. Surprisingly, in this case they are quite divergent in each of the individual parameters. In fact, they share fewer than 10% of their predicted target genes (SS = 0.06, Figure V-12). This means that HLH-4 and HLH-10 may regulate an overlapping set of target genes in head neurons to control sensory perception, whereas HLH-15 and HLH-19 may regulate different sets of chromatin genes in (developing) head neurons. The annotation “head neurons” is very broad as there are ~200 different neurons comprising this category. Therefore, we further refined the expression annotations of HLH-4, HLH-10 and HLH-15 (the expression of HLH-19 diminishes after the animals hatch and could not be annotated in more detail). We found that HLH-4 and HLH-10 may be expressed in a similar set of neurons, whereas the expression of HLH-15 is clearly distinct (Figure V-13). This supports the hypothesis that HLH-4 and HLH-10 may share target genes in the same cell(s).

HLH-4 and HLH-15 confer different loss-of-function phenotypes: RNAi of *hlh-15* results in high fat content (162), but no other detectable phenotype, and RNAi of *hlh-4* results in slow growth and protruding vulva (163). These two TFs share almost 25% of their DNA binding sites (SS = 0.24) but less than 5% of their candidate target genes (SS = 0.01), most likely because HLH-2/HLH-4 has a broader DNA binding specificity than HLH-2/HLH15. In addition, HLH-4 and HLH-15 are expressed in distinct neurons (Figure V-13). This indicates that the functional divergence of these two bHLH TFs is likely accomplished by relatively small changes in spatiotemporal expression and DNA binding specificities.

Even though the bHLH TFs shown in Figure V-11 exhibit a relatively high degree of similarity, there are also important differences that can contribute to TF divergence. For instance, of the four bHLH dimers shown, only one is expressed in the vulva (HLH-2/HLH-10). Similarly, only two of the dimers are expressed in later stages of development (HLH-2/HLH-4 and HLH-2/HLH-10), whereas the other two are exclusively expressed during embryogenesis and in the first larval stage (Figure III-3, III-4, Table III-5).

Target Gene Overlap of CACGTG-Binding bHLH Dimers

Finally, we analyzed molecular and functional divergence among a set of bHLH dimers that can all bind the CACGTG E-box. Three of these dimers exclusively bind this E-box (HLH-30, HLH-26 and REF-1) whereas the others (HLH-2/HLH-10, MXL-3, HLH-25 and MDL-1/MXL-1) also bind other E-box and/or E-box-like sequences (Figure IV-6A, IV-6B). Interestingly, we find little

overlap between these different dimers in their candidate target genes (Figure V-14). This indicates that several of these dimers may utilize multiple different E-box and E-box-like sequences in their target genes and that target genes may discriminate bHLH dimers by harboring different combinations of E-box and E-box-like sequences. Even for dimers that exclusively bind the CACGTG E-box, we find little overlap in their candidate target genes. Indeed, HLH-30 favors a flanking T, HLH-26 favors an A or G and REF-1 disfavors a T, indicating that flanking nucleotides may play an important role in functional TF divergence. Finally, the pair that shares the largest proportion of predicted target genes, REF-1 and MXL-3, exhibits non-overlapping spatiotemporal expression patterns, which likely contributes to their functional divergence (Table III-4).

Parameter Similarity Score Heatmaps

To visualize the pairwise bHLH similarity score analysis, we generated similarity score heatmaps for each parameter (Figure V-10). Yellow boxes indicate a high degree of similarity (SS ~ 1.0) for each parameter/bHLH-bHLH pair combination, black indicates intermediate similarity (SS ~0.5) and light blue indicates no similarity (SS ~ 0). As you can see from the heatmaps, most bHLH-bHLH pairs are quite distinct from each other. The large yellow blocks in the dimerization heatmap are a reflection of several bHLH proteins sharing HLH-2 or AHA-1 as a dimerization partner.

DISCUSSION

We present the first integrated network for any TF family that provides connections between proteins, the tissues in which they are expressed, the DNA sequences they preferentially bind, their candidate target genes and enriched GO categories associated with these target genes. Several observations indicate that our individual TF parameter datasets are of high quality, and most importantly, each of the different datasets validate each other. For instance, PBM assays with five combinations of bHLH proteins that did not heterodimerize in Y2H assays did not yield any specific DNA binding motifs (Figure IV-13). This indicates that PBMs validate Y2H, and *vice versa*. Similarly, the observation that bHLH proteins that dimerize are more likely co-expressed than those that do not dimerize validates the Y2H data.

The integrated bHLH network is likely not yet complete. For instance, we used only bHLH promoter activity, and did not include other potential regulatory sequences. In addition, we did not annotate bHLH expression in males or dauers, or under different conditions. Finally, for future models of gene regulation it will be important to incorporate expression levels of different bHLHs in different cell types, because protein levels will determine the binding to high or low affinity binding sites, and, hence the selection of tissue-specific target genes.

Previously, two other integrated networks were reported for *C. elegans* genes. The first connects genes involved in early embryogenesis by protein-protein interactions, phenotypes and expression profiles (164). The second is a

probabilistic network that used various data types and that can be used to predict genetic interactions (165). Although powerful, neither network focused on TFs or provided interactions between proteins, DNA sequences, and tissues or cell types, and therefore could not address the question of divergence in paralogous TF families.

A priori, we reasoned that paralogous TFs could attain functional specificity by individualizing a single molecular parameter. However, we found a spectrum of differences among the TFs in all parameters; some bHLH TFs are relatively similar in one or more parameters, whereas others are highly divergent. This is reflected by the observation of both specificity and promiscuity in the integrated network; some nodes (*e.g.* DNA sequences, tissues) are highly connected to many bHLH TFs, and others are not. Considering all the parameters measured, most bHLH TFs differ substantially from each other. There are several relatively similar bHLH TFs that exhibit only limited divergence in one or more TF parameters. However, we found that a minor difference in DNA binding specificity, either in the core E-box or E-box-like sequence, or in the flanking nucleotides, can result in little overlap in candidate target genes.

Even though many paralogous TFs have distinct biological functions, there are also examples of redundant TF paralogs. For example, members of mammalian ETS family of TFs can function partially redundantly by binding to overlapping sets of target genes (166). Similarly, FLH TFs in *C. elegans* can redundantly regulate microRNA expression (167). Finally, in *C. elegans*,

paralogous TFs such as paired homeodomains can function in modules in the context of neuronal regulatory networks (7). Future systematic studies of genetic interactions will reveal the extent of genetic redundancy within TF families.

In addition to enabling studies of TF divergence, this integrated network is also useful for generating specific hypotheses, as demonstrated by our gene expression profiling analysis of *hlh-30* mutant animals. Moreover, each of the individual data types provides a first comprehensive catalog of dimers, expression patterns and binding sites for a metazoan TF family. These data will be useful for gaining insight into the molecular determinants of the interactions in which the various bHLH proteins participate.

The integrated bHLH network confirms previously reported features for the bHLH family, including a promiscuous role in dimerization, DNA binding specificity and expression for the E/Daughterless homolog, HLH-2, and more specific roles for its dimerization partners (71). AHA-1 and HLH-2, both of which dimerize with multiple bHLH proteins, are auto-activators in Y2H assays whereas most of their dimerization partners are not. Based on these observations, we propose that the bHLH dimerization hubs may confer the transcriptional activation activity to the different dimers, whereas their dimerization partners may contribute specificity in DNA binding.

Our data and methods provide a framework for similar studies of other *C. elegans* TF families and of TF families in other organisms, including humans. Similar studies will likely be useful for other protein families, such as kinases, in

the context of other types of regulatory networks. Such studies of paralogous genes, including comparisons of integrated networks across species, may provide further insights into the molecular features underlying the evolution of gene families.

Materials and Methods

RNA Isolation

N2 and *hlh-30(tm1978)* worms were grown to high density at 20°C on 10 cm egg plates and bleached. Embryos were washed in M9 buffer, collected, and quantified. Approximately 100,000 embryos were isolated immediately for each of three biological replicates per genotype. Embryos were pelleted to 100 microliters, resuspended in 1.0 ml Trizol reagent (Invitrogen), and flash frozen in liquid nitrogen and stored at -80° C for at least 16 hours. The remaining semi-synchronized embryos were plated on 10 cm NGM plates (seeded with OP50) such that three biological replicates of approximately 80,000 L1s and 40,000 L2s could be harvested for each genotype. When harvesting, worms were washed several times with ~10 ml sterile water until no OP50 bacteria could be seen in solution. Worms were then incubated in 10 ml sterile M9 at 4°C for 30 minutes to allow digestion of any bacteria remaining in the digestive tracts of the worms. Immediately following this incubation, worms were pelleted to ~100 microliters, resuspended in 1.0 ml Trizol reagent (as above), flash frozen in liquid nitrogen, and stored for at least 16 hours at -80°C. RNA was then isolated from the samples as described previously (168).

Gene Expression Microarray Experiments

Three biological replicates were done. For each biological replicate, RNA from embryos, L1s, and L2s were pooled for processing and hybridization to each microarray. Affymetrix GeneChip® *C. elegans* Genome arrays were used to

analyze gene expression in wild type and *hlh-30(tm1978)* mutant worms. One cycle cDNA synthesis was performed as suggested by the Affymetrix GeneChip® Expression Analysis Technical Manual. For cRNA *in vitro* transcription/labeling, the Enzo BioArray™ HighYield™ RNA Transcript Labeling Kit (Enzo Life Sciences) was used with overnight incubation. Remaining steps including fragmentation, hybridization, and wash were performed as recommended by the Affymetrix GeneChip® Expression Analysis Technical Manual.

Gene Expression Microarray Analysis

For the normalization and analysis of the microarray data we used R and the affy (169), gcrma (170) and limma (171) packages from Bioconductor (www.bioconductor.org). All microarrays for each experiment were normalized using the gcrma package. To avoid artifacts in fold-change due to very low probe intensities in both conditions, we trimmed the data keeping only probe sets that had a mean $\log_2$(signal intensity) of the three replicas of 8 or higher, in at least one condition. This threshold was chosen based on the signal intensities of the external polyA RNA and cRNA controls. The trimmed data were analyzed for differential expression using the limma package. We considered as differentially expressed those genes whose probe sets had a $P < 0.01$ and $\log_2$(fold-change) ≥ 0.7. Probe sets were converted to WBGene IDs using the Affymetrix annotation file “Celegans.na26.annottrimmed. csv” and WormMart (www.wormbase.org). To avoid redundancy we kept only one gene per probe set and one probe set per

gene. We also removed genes encoded by the mitochondrial genome. The total number of genes detected was 6,591.

Parameter Overlap Analysis

For each pair-wise bHLH-bHLH parameter comparison Similarity Scores (SS) were calculated as follows:

$$SS = \frac{HLH - X \cap HLH - Y}{HLH - X \cup HLH - Y}$$

For instance, when bHLH-X binds 10 target genes and bHLH-Y binds 20 target genes, and they have 5 target genes in common, the SS would be 5/25 = 0.2. Heat maps were created by clustering the bHLH TFs based on their SSs. The clustering of heat maps depicting parameter comparisons were performed using MultiExperiment Viewer version 4.0 (172).

Microarray Accession Number

Gene expression microarray data have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession ID GSE15762.

Acknowledgements

We thank Inmaculada Barrasa for performing the analysis and statistics of microarray data, statistics of HLH-30 binding sites in HLH-30 target genes, and similarity score statistics (including the generation of SS heatmaps). We also thank Phyllis Spatrick at the Genomics Core Facility at UMass Medical School for performing the RNA processing and microarray hybridization and scanning. We thank Mark Alkema for his assistance in annotating neuronal expression patterns for several bHLH TFs.

We thank Job Dekker, and members (current and past) of the Walhout and Dekker labs for critical reading of relevant manuscripts.

Figure V-1

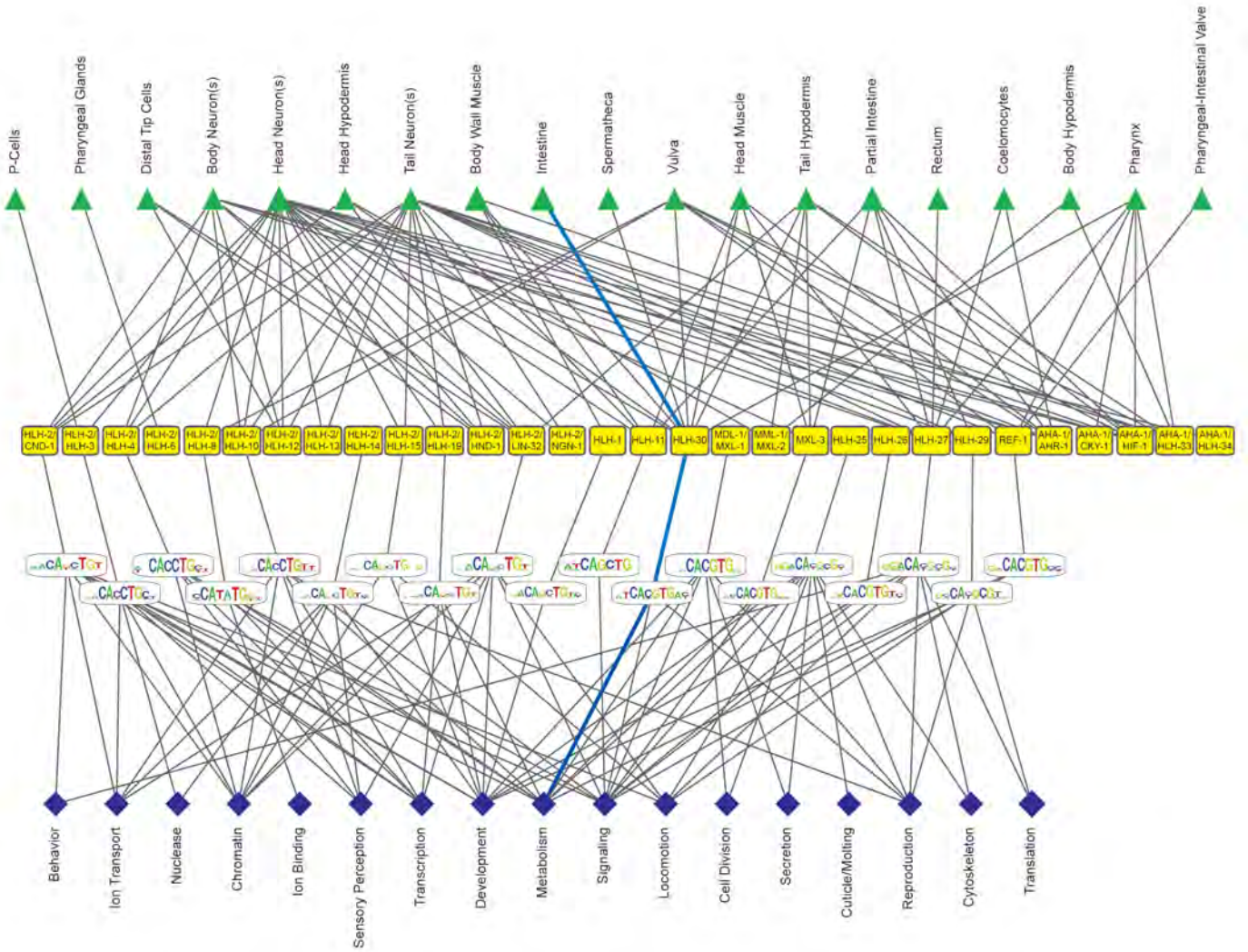

P-Cells
Pharyngeal Glands
Distal Tip Cells
Body Neuron(s)
Head Neuron(s)
Head Hypodermis
Tail Neuron(s)
Body Wall Muscle
Intestine
Spermatheca
Vulva
Head Muscle
Tail Hypodermis
Partial Intestine
Rectum
Coelomocytes
Body Hypodermis
Pharynx
Pharyngeal-Intestinal Valve
HLH-2/ CND-1
HLH-2/ HLH-3
HLH-2/ HLH-4
HLH-2/ HLH-6
HLH-2/ HLH-8
HLH-2/ HLH-10
HLH-2/ HLH-12
HLH-2/ HLH-13
HLH-2/ HLH-14
HLH-2/ HLH-15
HLH-2/ HLH-19
HLH-2/ HND-1
HLH-2/ LIN-32
HLH-2/ NGN-1
HLH-1
HLH-11
HLH-30
MDL-1/ MXL-1
MML-1/ MXL-2
MXL-3
HLH-25
HLH-26
HLH-27
HLH-29
REF-1
AHA-1/ AHR-1
AHA-1/ CRY-1
AHA-1/ HIF-1
AHA-1/ HLH-33
AHA-1/ HLH-34
Behavior
Ion Transport
Nuclease
Chromatin
Ion Binding
Sensory Perception
Transcription
Development
Metabolism
Signaling
Locomotion
Cell Division
Secretion
Cuticle/Molting
Reproduction
Cytoskeleton
Translation

Figure V-1. The integrated *C. elegans* bHLH network
Integrated bHLH network that combines dimerization, spatiotemporal expression, DNA binding specificities and GO categories. The blue lines depict a “network path” connecting the intestine to the “metabolism” GO category through HLH-30.

Figure V-2A

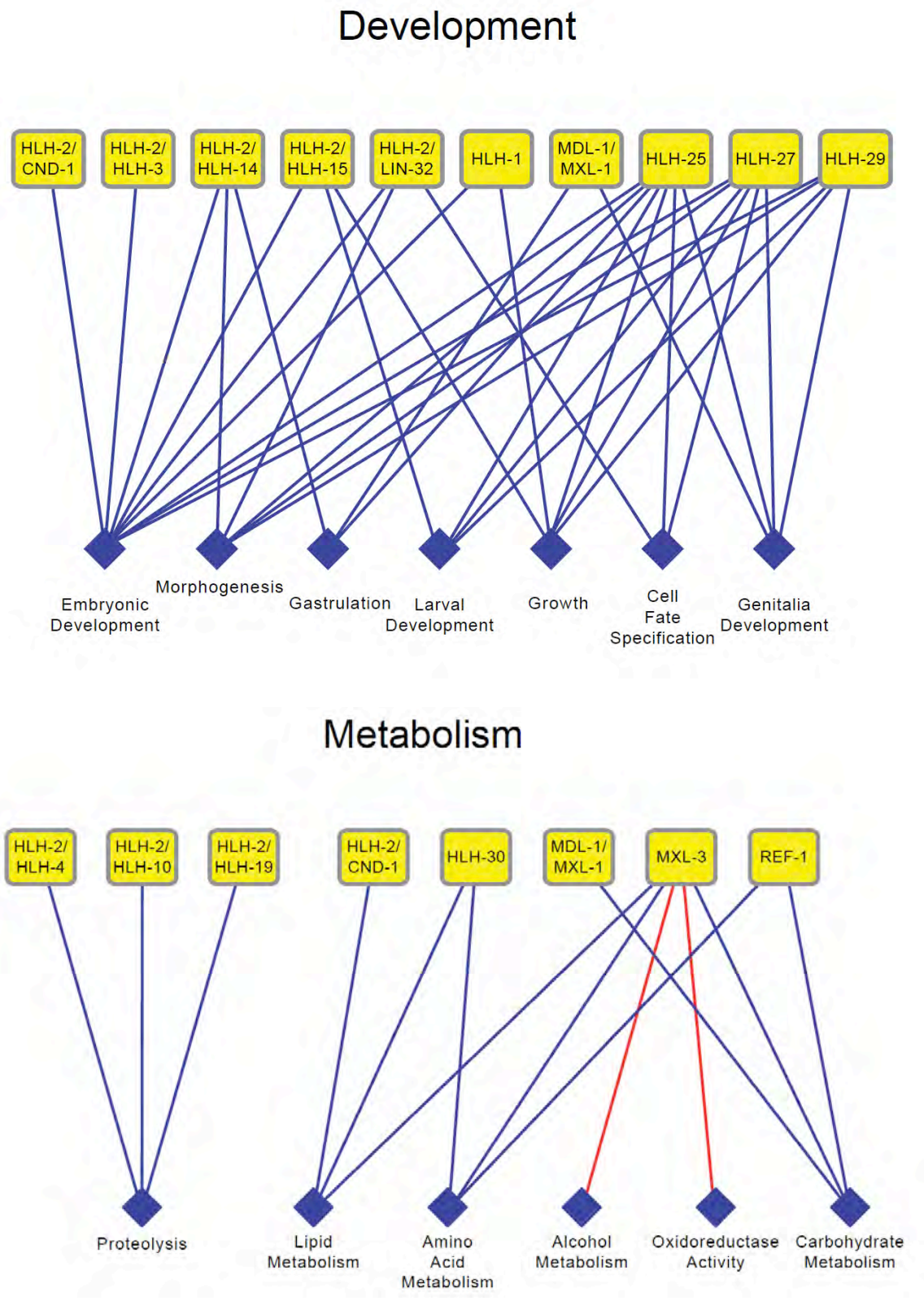

Development
HLH-2/ CND-1
HLH-2/ HLH-3
HLH-2/ HLH-14
HLH-2/ HLH-15
HLH-2/ LIN-32
HLH-1
MDL-1/ MXL-1
HLH-25
HLH-27
HLH-29
Embryonic Development
Morphogenesis
Gastrulation
Larval Development
Growth
Cell Fate Specification
Genitalia Development
Metabolism
HLH-2/ HLH-4
HLH-2/ HLH-10
HLH-2/ HLH-19
HLH-2/ CND-1
HLH-30
MDL-1/ MXL-1
MXL-3
REF-1
Proteolysis
Lipid Metabolism
Amino Acid Metabolism
Alcohol Metabolism
Oxidoreductase Activity
Carbohydrate Metabolism

Figure V-2B

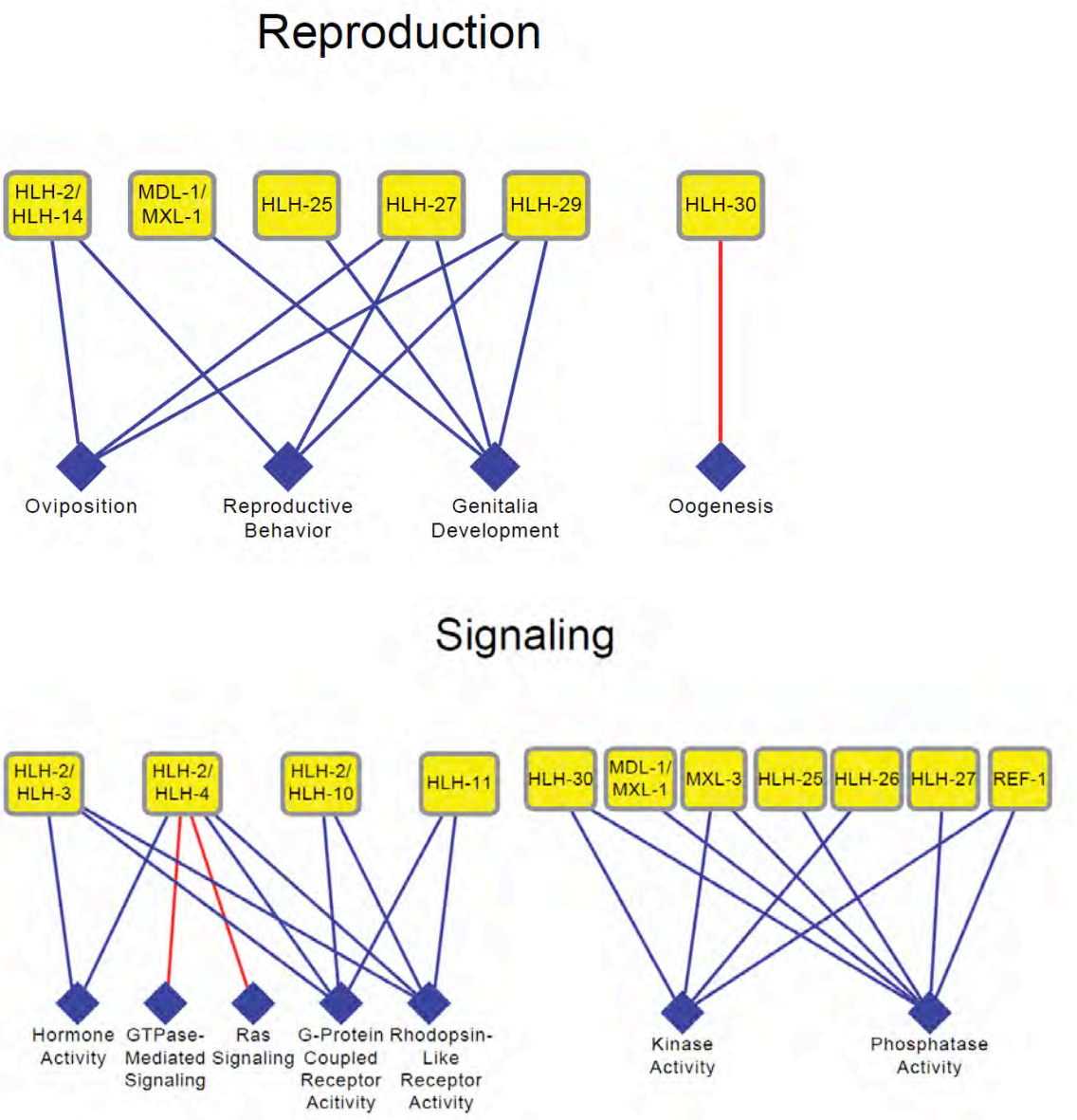
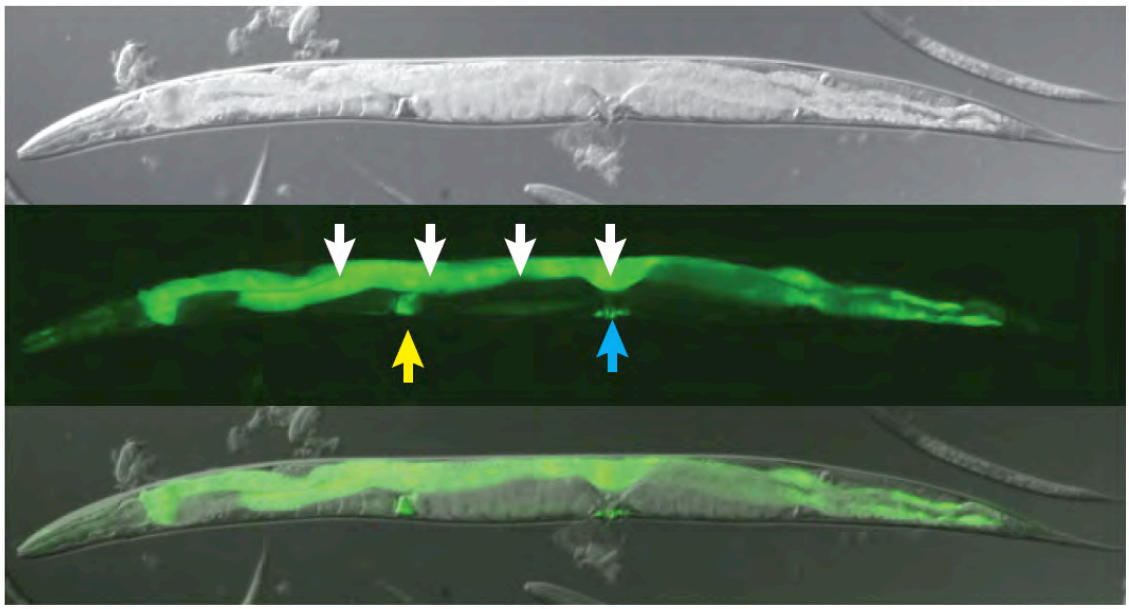

Reproduction
HLH-2/ HLH-14
MDL-1/ MXL-1
HLH-25
HLH-27
HLH-29
HLH-30
Oviposition
Reproductive Behavior
Genitalia Development
Oogenesis
Signaling
HLH-2/ HLH-3
HLH-2/ HLH-4
HLH-2/ HLH-10
HLH-11
HLH-30
MDL-1/ MXL-1
MXL-3
HLH-25
HLH-26
HLH-27
REF-1
Hormone Activity
GTPase-Mediated Signaling
Ras Signaling
G-Protein Coupled Receptor Acitivity
Rhodopsin-Like Receptor Activity
Kinase Activity
Phosphatase Activity

Figure V-2A,B. GO Subnetworks
Many GO terms depicted in the integrated network in Figure V-1 can be further subdivided into more specific terms. Shown here are subcategories of development, metabolism (A), reproduction, and signaling (B) and the bHLH dimers which are associated with them via their candidate target genes. Red lines connecting dimers to GO terms indicate unique associations.

Figure V-3



Figue V-3. Expression pattern of *hlh-30*
Phlh-30 drives GFP expression in different tissues, including the intestine (white arrows), spermatheca (yellow arrow) and vulva (blue arrow). Top – DIC image; middle – GFP image; bottom – merged images.

Figure V-4

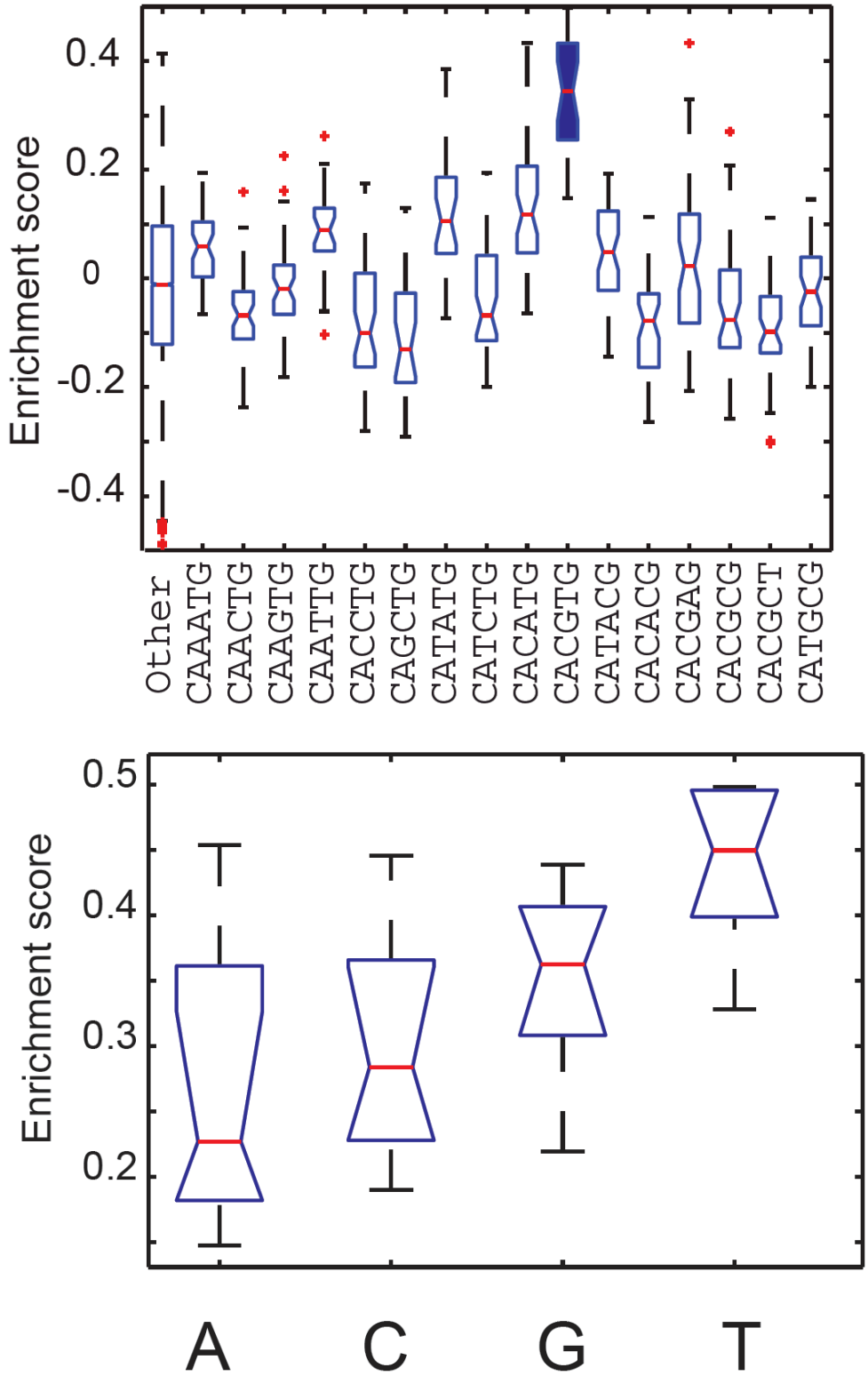

Enrichment score
0.4
0.2
0
-0.2
-0.4
Other
CAAATG
CAACTG
CAAGTG
CAATTG
CACCTG
CAGCTG
CATATG
CATCTG
CACATG
CACGTG
CATACG
CACACG
CACGAG
CACGCG
CACGCT
CATGCG
Enrichment score
0.5
0.4
0.3
0.2
A
C
G
T

Figure V-4. Summary of HLH-30 DNA binding specificity
(Top panel) HLH-30 strongly prefers the CACGTG E-box. (Bottom panel) HLH-30 strongly favors a 5' T flanking the CACGTG E-box.

Figure V-5

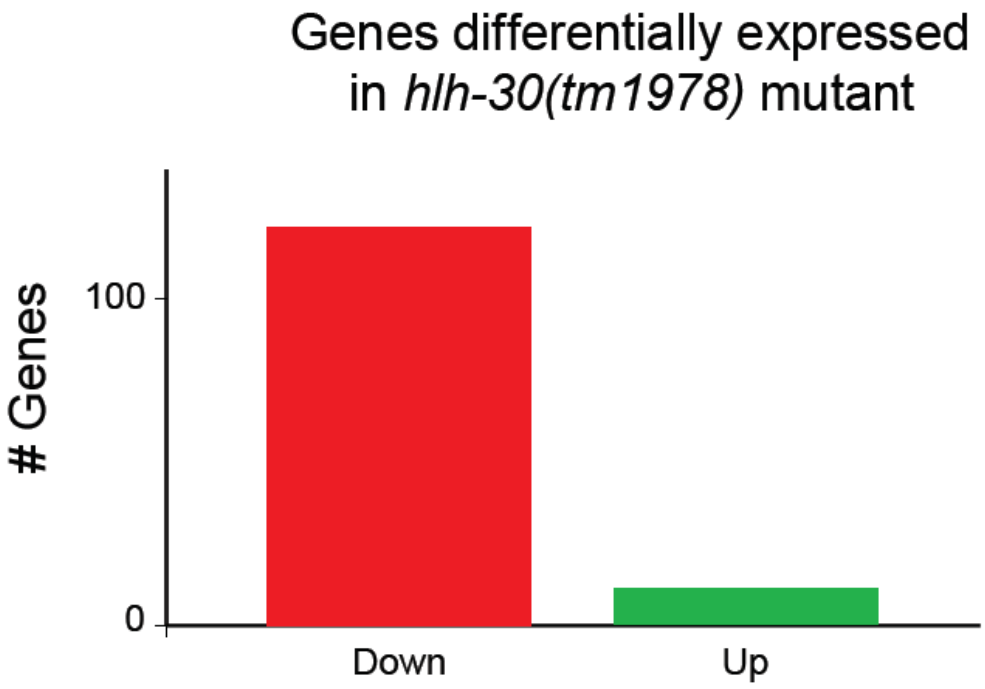


Figure V-5. HLH-30 acts predominately as an activator of transcription
HLH-30 activates gene expression. The majority of genes that change significantly in *hlh-30(tm1978)* mutant animals exhibit reduced expression (red), while the expression of a minority is increased (green).

Table V-1. Genes with significant difference in expression in *hlh-30* mutant

Genes with increased expression in *hlh-30(tm1978)* mutant

Gene WB ID	Gene Public Name	Sequence Name (Gene)	logFC	P.Value	adj.P.Val
WBGene00021224	Y19D10A.9	Y19D10A.9	6.824813	4.87E-08	0.00041051
WBGene00009874	F49C12.4	F49C12.4	3.104021	0.000296071	0.05551002
WBGene00021219	Y19D10A.4	Y19D10A.4	3.076091	0.0000486	0.03205797
WBGene00044734	Y19D10A.16	Y19D10A.16	1.815645	0.008236328	0.26222604
WBGene00021731	Y49G5A.1	Y49G5A.1	1.712395	0.0000239	0.02885601
WBGene00019311	K02E7.6	K02E7.6	1.534168	0.001469218	0.11268904
WBGene00009429	F35E12.5	F35E12.5	1.205518	0.000185591	0.05340381
WBGene00018911	F56A4.3	F56A4.3	1.045228	0.000295695	0.05551002
WBGene00020869	T28A11.2	T28A11.2	1.003069	0.001212684	0.10204113
WBGene00019538	K08D12.4	K08D12.4	0.92112	0.005186948	0.21586081
WBGene00017560	F18C5.5	F18C5.5	0.80831	0.007041085	0.24530028
WBGene00013007	Y48E1B.8	Y48E1B.8	0.732119	0.005593652	0.22313167

Genes with decreased expression in *hlh-30(tm1978)* mutant

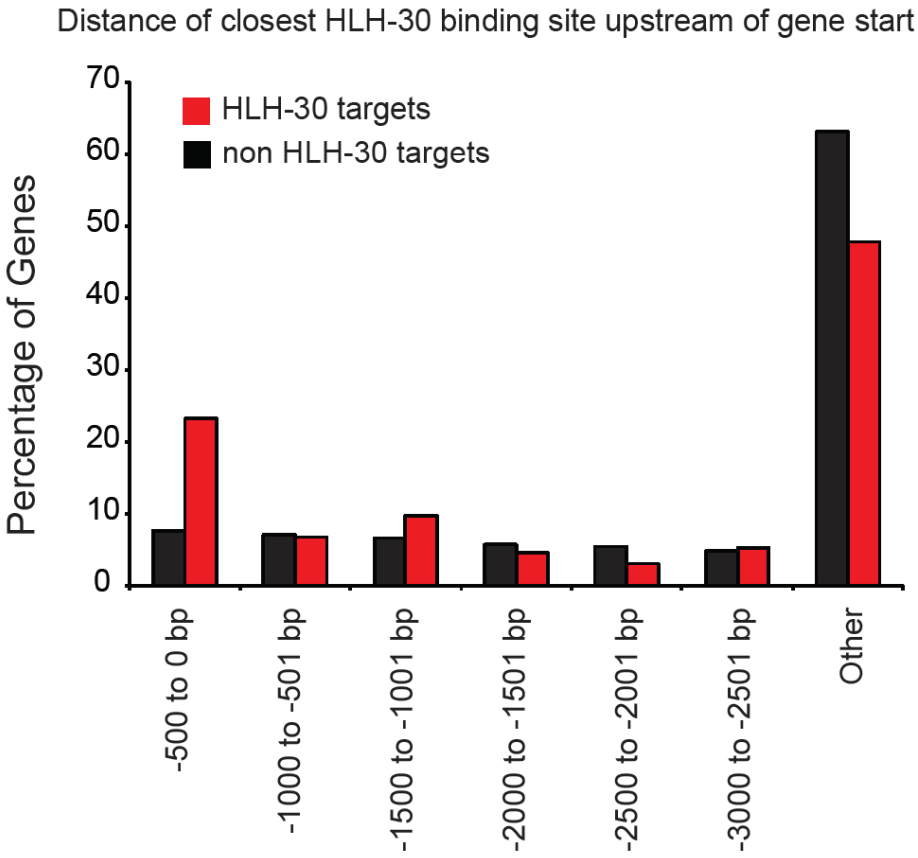
Gene WB ID	Gene Public Name	Sequence Name (Gene)	logFC	P.Value	adj.P.Val
WBGene00006650	tts-1	F09E10.11	-7.22606	0.0000041	0.00865787
WBGene00005832	srw-85	C25F9.1	-6.6146	0.00000101	0.00424488
WBGene00016119	C25H3.10	C25H3.10	-3.18186	0.000106522	0.04085125
WBGene00017964	F31F7.1	F31F7.1	-3.12091	0.00000664	0.01120184
WBGene00010790	sodh-1	K12G11.3	-2.63264	0.000153385	0.04977336
WBGene00011733	T12D8.5	T12D8.5	-2.59226	0.000038	0.03205797
WBGene00022597	ZC395.5	ZC395.5	-2.54336	0.0000113	0.01593798
WBGene00008547	F07A11.4	F07A11.4	-2.5423	0.00000309	0.00865787
WBGene00009221	acs-2	F28F8.2	-2.47643	0.000290157	0.05551002
WBGene00018564	F47D12.9	F47D12.9	-2.44567	0.000348539	0.0588125
WBGene00007365	C06B3.6	C06B3.6	-2.2987	0.0000822	0.03649601
WBGene00020930	hlh-30	W02C12.3	-2.27688	0.00362126	0.17730184
WBGene00011089	R07B7.5	R07B7.5	-1.87825	0.0000765	0.03596451
WBGene00011185	R10D12.1	R10D12.1	-1.79713	0.007323373	0.25116787
WBGene00010480	K01G5.9	K01G5.9	-1.76149	0.002069184	0.13326494
WBGene00016628	C44B7.6	C44B7.6	-1.6754	0.001307127	0.10407137
WBGene00016052	dod-3	C24B9.9	-1.67455	0.002526118	0.14745864
WBGene00009895	F49E11.10	F49E11.10	-1.67031	0.003760656	0.17925791
WBGene00021086	W08E12.5	W08E12.5	-1.52355	0.000899225	0.09031858
WBGene00009142	F26A3.4	F26A3.4	-1.50007	0.0000352	0.03205797
WBGene00003616	nhr-17	C02B4.2	-1.46088	0.000057	0.03205797
WBGene00004196	prx-11	C47B2.8	-1.43753	0.000459175	0.06692982
WBGene00008915	F17C11.4	F17C11.4	-1.41426	0.001702288	0.12381209
WBGene00018294	atgr-18	F41E6.13	-1.40806	0.0000521	0.03205797
WBGene00044728	Y53F4B.45	Y53F4B.45	-1.39508	0.001179598	0.10204113
WBGene00015350	C02F5.7	C02F5.7	-1.38723	0.000248714	0.05551002
WBGene00021876	Y54G2A.11	Y54G2A.11	-1.38536	0.000887067	0.0901709
WBGene00004145	pqn-62	T03G11.1	-1.38241	0.000701532	0.08141863
WBGene00007958	C35C5.9	C35C5.9	-1.35352	0.000398745	0.0618289
WBGene00006591	toh-1	T24A11.3	-1.33453	0.006237335	0.2332574
WBGene00017484	F15E6.3	F15E6.3	-1.32847	0.006434345	0.23463087

WBGene00020031	R12E2.2	R12E2.2	-1.32523	0.000403057	0.0618289
WBGene00016845	C50F7.5	C50F7.5	-1.28913	0.0000767	0.03596451
WBGene00017565	F18E3.7	F18E3.7	-1.28859	0.000292377	0.05551002
WBGene00009935	F52E10.4	F52E10.4	-1.2759	0.002442995	0.14722533
WBGene00011662	T09F3.2	T09F3.2	-1.26448	0.000748934	0.08314155
WBGene00017968	F32A5.2	F32A5.2	-1.24987	0.000927059	0.09094884
WBGene00008925	F17H10.1	F17H10.1	-1.22162	0.00212637	0.13488862
WBGene00019298	K02D7.1	K02D7.1	-1.21205	0.0000719	0.03596451
WBGene00007811	C29F7.2	C29F7.2	-1.17477	0.002441965	0.14722533
WBGene00013516	Y73F4A.3	Y73F4A.3	-1.17191	0.007483796	0.25459994
WBGene00015159	psd-1	B0361.5	-1.1526	0.0000437	0.03205797
WBGene00017789	F25E5.8	F25E5.8	-1.1405	0.002814173	0.15723958
WBGene00001391	far-7	K01A2.2	-1.13633	0.000202608	0.05341898
WBGene00018464	F45E1.3	F45E1.3	-1.12281	0.001290801	0.10407137
WBGene00000981	dhs-18	C45B11.3	-1.11871	0.000524462	0.07136906
WBGene00000784	cpr-4	F44C4.3	-1.11724	0.000644686	0.07882924
WBGene00015403	clec-10	C03H5.1	-1.11379	0.009366678	0.28525464
WBGene00016630	C44B7.10	C44B7.10	-1.10939	0.0002403	0.05551002
WBGene00023382	Y57E12AL.3	Y57E12AL.3	-1.1048	0.008836526	0.27715156
WBGene00021554	Y45G5AL.1	Y45G5AL.1	-1.09607	0.000273238	0.05551002
WBGene00015484	C05D11.7	C05D11.7	-1.07189	0.001764893	0.12458812
WBGene00013898	ZC443.3	ZC443.3	-1.06266	0.000958699	0.09171647
WBGene00009042	F22B5.4	F22B5.4	-1.05929	0.000272573	0.05551002
WBGene00009797	F46G10.2	F46G10.2	-1.03934	0.000938063	0.09097054
WBGene00018731	F53A9.8	F53A9.8	-1.01692	0.002793373	0.15723958
WBGene00014058	ZK673.2	ZK673.2	-1.01687	0.00078161	0.08564217
WBGene00006587	tnt-2	F53A9.10	-1.00861	0.0000465	0.03205797
WBGene00009724	F45D3.4	F45D3.4	-1.0065	0.002880857	0.15934456
WBGene00022078	Y69A2AR.7	Y69A2AR.7	-1.0025	0.000055	0.03205797
WBGene00007574	C14B1.3	C14B1.3	-1.00159	0.00069204	0.08141863
WBGene00018268	F41C3.2	F41C3.2	-1.00042	0.002688612	0.15326903
WBGene00010408	H19N07.4	H19N07.4	-0.99385	0.000267029	0.05551002
WBGene00002078	imb-4	ZK742.1	-0.98687	0.001052524	0.09758397
WBGene00003378	mml-1	T20B12.6	-0.98479	0.009142447	0.28151397
WBGene00001523	gbh-2	M05D6.7	-0.98291	0.000815531	0.08667344
WBGene00009050	F22D6.2	F22D6.2	-0.97914	0.00011333	0.04157255
WBGene00008317	C54G4.7	C54G4.7	-0.97354	0.003502307	0.17663761
WBGene00010681	K08F8.1	K08F8.1	-0.96258	0.000319101	0.05728201
WBGene00001843	hgo-1	W06D4.1	-0.95973	0.000328279	0.0577019
WBGene00016201	C28H8.11	C28H8.11	-0.95736	0.000987496	0.09257229
WBGene00016594	C42D4.1	C42D4.1	-0.9485	0.000345269	0.0588125
WBGene00008412	D2030.2	D2030.2	-0.9391	0.000839682	0.08667344
WBGene00006603	tps-2	F19H8.1	-0.9352	0.000189891	0.05340381
WBGene00008211	C49F5.7	C49F5.7	-0.93004	0.00433794	0.19571765
WBGene00004997	spp-12	T22G5.7	-0.92809	0.00410788	0.18982535
WBGene00000781	cpr-1	C52E4.1	-0.92087	0.006291293	0.2332574
WBGene00016894	C53B7.3	C53B7.3	-0.91322	0.003640376	0.17730184
WBGene00019294	K02A6.3	K02A6.3	-0.89876	0.000967496	0.09171647
WBGene00004222	ptr-8	F44F4.4	-0.89309	0.006032982	0.23031795
WBGene00004258	pyc-1	D2023.2	-0.88915	0.002056346	0.13326494
WBGene00018953	F56C9.10	F56C9.10	-0.88647	0.009808621	0.29139203
WBGene00000245	bca-1	T13C5.5	-0.88503	0.006656657	0.23807267
WBGene00009595	F40F12.7	F40F12.7	-0.88178	0.003029141	0.16488297

WBGene00016061	C24G6.6	C24G6.6	-0.87862	0.00319743	0.17182621
WBGene00003970	pek-1	F46C3.1	-0.87838	0.00036257	0.05937464
WBGene00003096	lys-7	C02A12.4	-0.87361	0.00841689	0.26596743
WBGene00019719	M01H9.3	M01H9.3	-0.85811	0.000252018	0.05551002
WBGene00022233	Y73B6BL.4	Y73B6BL.4	-0.85534	0.000187371	0.05340381
WBGene00006996	zyg-11	C08B11.1	-0.85296	0.001555447	0.11717239
WBGene00014258	ZK1320.9	ZK1320.9	-0.84855	0.005880382	0.22833143
WBGene00001993	hpd-1	T21C12.2	-0.84378	0.008953113	0.27873584
WBGene00012140	T28F4.5	T28F4.5	-0.83931	0.000242625	0.05551002
WBGene00015776	C14F5.1	C14F5.1	-0.83922	0.000714114	0.08141863
WBGene00015251	B0546.4	B0546.4	-0.83305	0.001307522	0.10407137
WBGene00007666	C18B12.4	C18B12.4	-0.82327	0.000302757	0.05552959
WBGene00008233	C50F4.8	C50F4.8	-0.80952	0.000100703	0.0404586
WBGene00017045	D2007.5	D2007.5	-0.8079	0.000599243	0.07778176
WBGene00000110	alh-4	T05H4.13	-0.80153	0.004393162	0.19643203
WBGene00012850	Y44A6C.1	Y44A6C.1	-0.79983	0.00055112	0.07380633
WBGene00020142	aak-2	T01C8.1	-0.79976	0.001605363	0.11833233
WBGene00008436	DH11.2	DH11.2	-0.79689	0.005606722	0.22313167
WBGene00012988	Y48C3A.4	Y48C3A.4	-0.78479	0.00346959	0.17663761
WBGene00011953	T23F11.1	T23F11.1	-0.76733	0.007953828	0.26010251
WBGene00010628	K07C5.5	K07C5.5	-0.7646	0.00158302	0.11819414
WBGene00006819	unc-87	F08B6.4	-0.7634	0.005899757	0.22833143
WBGene00016610	C43G2.1	C43G2.1	-0.76051	0.006279515	0.2332574
WBGene00009850	F48F7.5	F48F7.5	-0.75952	0.003725521	0.17859217
WBGene00011939	T23B5.1	T23B5.1	-0.75939	0.009100501	0.2812488
WBGene00003476	mtm-3	T24A11.1	-0.75634	0.007531771	0.25520304
WBGene00007392	fbxa-156	C06H5.1	-0.75218	0.000521192	0.07136906
WBGene00017262	F08F3.4	F08F3.4	-0.74575	0.003656575	0.17730184
WBGene00010661	tyr-2	K08E3.1	-0.74191	0.005993894	0.23031795
WBGene00020200	T04A6.1	T04A6.1	-0.74112	0.008935641	0.27873584
WBGene00006754	unc-15	F07A5.7	-0.73998	0.003349763	0.17527954
WBGene00002981	lgg-2	ZK593.6	-0.73302	0.001203552	0.10204113
WBGene00019456	K06H7.2	K06H7.2	-0.7271	0.002516235	0.14745864
WBGene00020658	T21F4.1	T21F4.1	-0.72159	0.005521606	0.2228321
WBGene00020917	W01B11.6	W01B11.6	-0.72114	0.000433286	0.06527923
WBGene00009660	F43G6.8	F43G6.8	-0.70498	0.001914129	0.12817069
WBGene00002583	let-363	B0261.2	-0.70418	0.007710454	0.25814723
WBGene00001561	gei-4	W07B3.2	-0.70299	0.004191101	0.19113686
WBGene00022401	Y97E10AR.6	Y97E10AR.6	-0.66727	0.001612921	0.11833233

Table V-1. Genes with significant difference in expression in *hlh-30* mutant
WB Gene ID = Wormbase gene identification number, FC = Fold change, adj.P.Val = Adjusted p-value

Figure V-6


Distance of closest HLH-30 binding site upstream of gene start
HLH-30 targets
non HLH-30 targets
Percentage of Genes
70
60
50
40
30
20
10
0
-500 to 0 bp
-1000 to -501 bp
-1500 to -1001 bp
-2000 to -1501 bp
-2500 to -2001 bp
-3000 to -2501 bp
Other

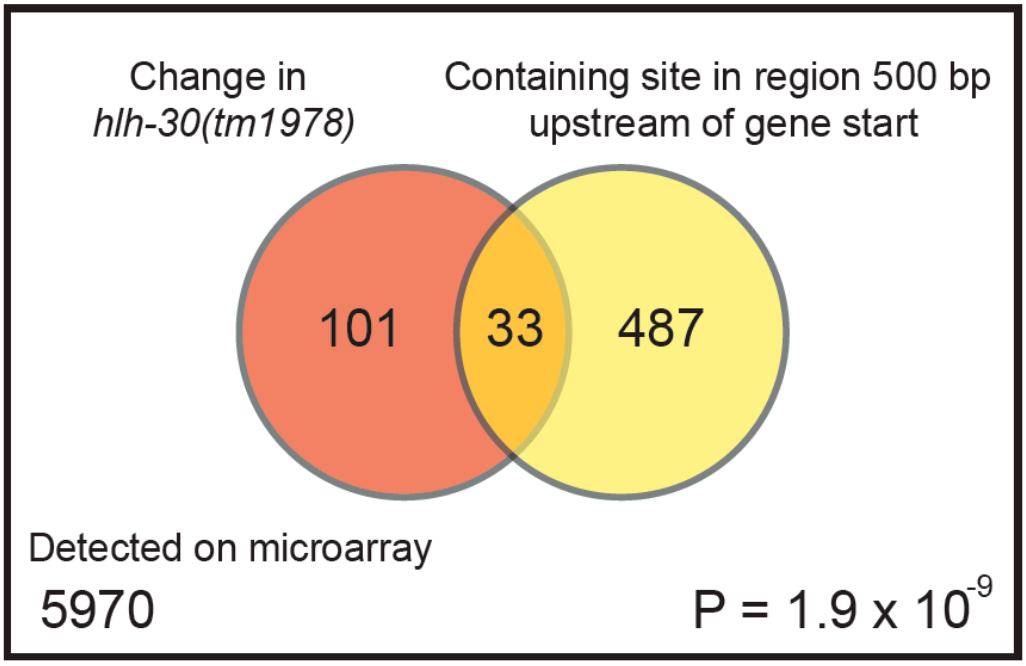
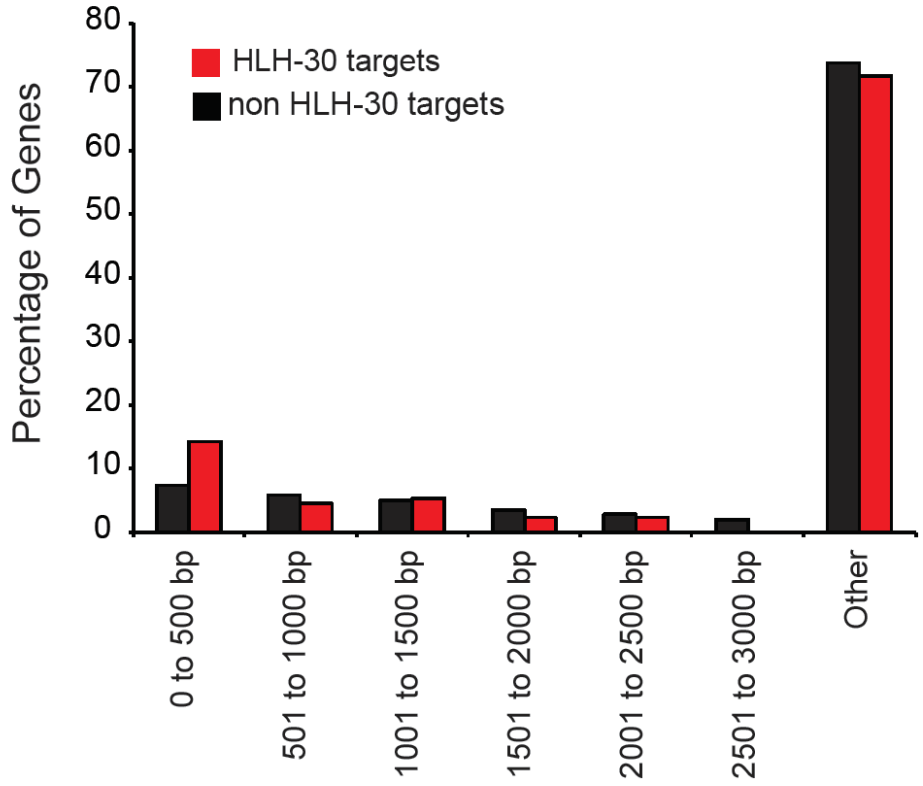

Change in
hlh-30(tm1978)
Containing site in region 500 bp
upstream of gene start
101
33
487
Detected on microarray
5970
P = 1.9 x 10-9

Figure V-6 Enrichment for HLH-30 binding sites upstream of HLH-30 target genes

(Top panel) Distribution of genes for which the location of the closest HLH-30 binding site upstream of the transcriptional start is in the indicated window of distance (in increments of 500 bp). (Bottom panel) Venn diagram demonstrating association of gene expression change in *hlh-30(tm1978)* mutant animals with the region 500 bp upstream of the gene start harboring an HLH-30 binding site.

Figure V-7

Distance of closest HLH-30 binding site downstream of gene start



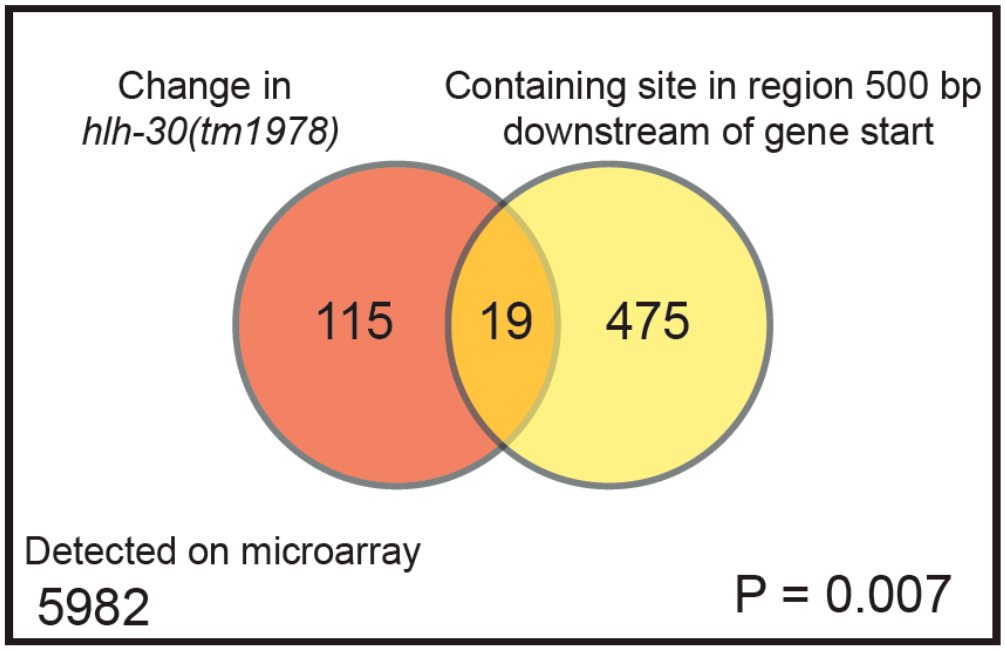


Figure V-7. Enrichment for HLH-30 binding sites downstream of HLH-30 target genes

(Top panel) Distribution of genes for which the location of the closest HLH-30 binding site downstream of the gene start is in the indicated genomic regions (in increments of 500 bp). (Bottom panel) Venn diagram demonstrating association of gene expression change in *hlh-30(tm1978)* mutant animals with the region 500 bp downstream of the gene start harboring an HLH-30 binding site.

Figure V-8

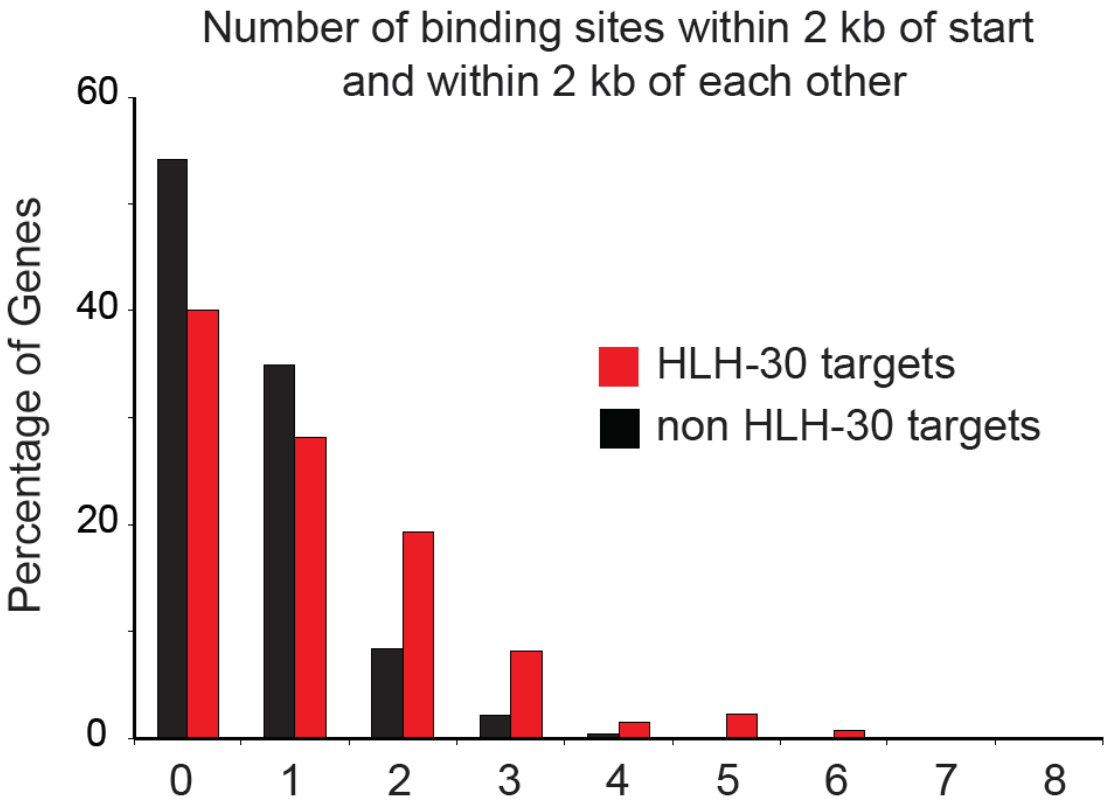


Figure V-8. HLH-30 targets tend to have multiple HLH-30 binding sites
HLH-30 targets have two or more HLH-30 binding sites within 2 kb of each other in the region up or downstream of the gene start more often than do non-HLH-30 targets.

Table V-2. GO categories enriched among HLH-30 target genes

GO CATEGORY	TOTAL GENES	CHANGED GENES	LOG10(p)
GO:0003674_molecular_function	3503	69	-1.383628
GO:0008152_metabolic_process	2043	53	-3.897661
GO:0003824_catalytic_activity	1648	45	-3.612949
GO:0044237_cellular_metabolic_process	1635	39	-1.852611
GO:0044238_primary_metabolic_process	1538	37	-1.809855
GO:0016491_oxidoreductase_activity	313	14	-2.993111
GO:0006082_organic_acid_metabolic_process	137	7	-1.992595
GO:0019752_carboxylic_acid_metabolic_process	137	7	-1.992595
GO:0008340_determination_of_adult_life_span	143	7	-1.896173
GO:0010259_multicellular_organismal_aging	143	7	-1.896173
GO:0007568_aging	145	7	-1.86529
GO:0009056_catabolic_process	181	7	-1.397965
GO:0006519_amino_acid_and_derivative_metabolic_process	98	6	-2.132734
GO:0009308_amine_metabolic_process	103	6	-2.030035
GO:0006807_nitrogen_compound_metabolic_process	107	6	-1.952535
GO:0006629_lipid_metabolic_process	121	6	-1.709467
GO:0022900_electron_transport_chain	136	6	-1.48916
GO:0055114_oxidation_reduction	136	6	-1.48916
GO:0016788_hydrolase_activity__acting_on_ester_bonds	140	6	-1.436243
GO:0006520_amino_acid_metabolic_process	89	5	-1.698956
GO:0005975_carbohydrate_metabolic_process	103	5	-1.457529
GO:0016701_oxidoreductase_activity__acting_on_single_donors_with_incorporation_of_molecular_oxygen	6	4	-5.875566
GO:0016702_oxidoreductase_activity__acting_on_single_donors_with_incorporation_of_molecular_oxygen__incorporation_of_two_atoms_of_oxygen	6	4	-5.875566
GO:0051213_dioxygenase_activity	6	4	-5.875566
GO:0006725_aromatic_compound_metabolic_process	44	4	-2.140964
GO:0044255_cellular_lipid_metabolic_process	72	4	-1.422421
GO:0009072_aromatic_amino_acid_family_metabolic_process	10	3	-3.231808
GO:0009063_amino_acid_catabolic_process	19	3	-2.374612
GO:0009310_amine_catabolic_process	19	3	-2.374612
GO:0044270_nitrogen_compound_catabolic_process	19	3	-2.374612
GO:0006644_phospholipid_metabolic_process	25	3	-2.032319
GO:0006470_protein_amino_acid_dephosphorylation	28	3	-1.895201
GO:0006643_membrane_lipid_metabolic_process	28	3	-1.895201
GO:0004721_phosphoprotein_phosphatase_activity	29	3	-1.853267
GO:0016311_dephosphorylation	29	3	-1.853267
GO:0016791_phosphoric_monoester_hydrolase_activity	43	3	-1.401091
GO:0030258_lipid_modification	7	2	-2.2122
GO:0006576_biogenic_amine_metabolic_process	8	2	-2.092259
GO:0006575_amino_acid_derivative_metabolic_process	10	2	-1.89618
GO:0006084_acetyl-CoA_metabolic_process	12	2	-1.739797
GO:0016051_carbohydrate_biosynthetic_process	16	2	-1.499975

Figure V-9

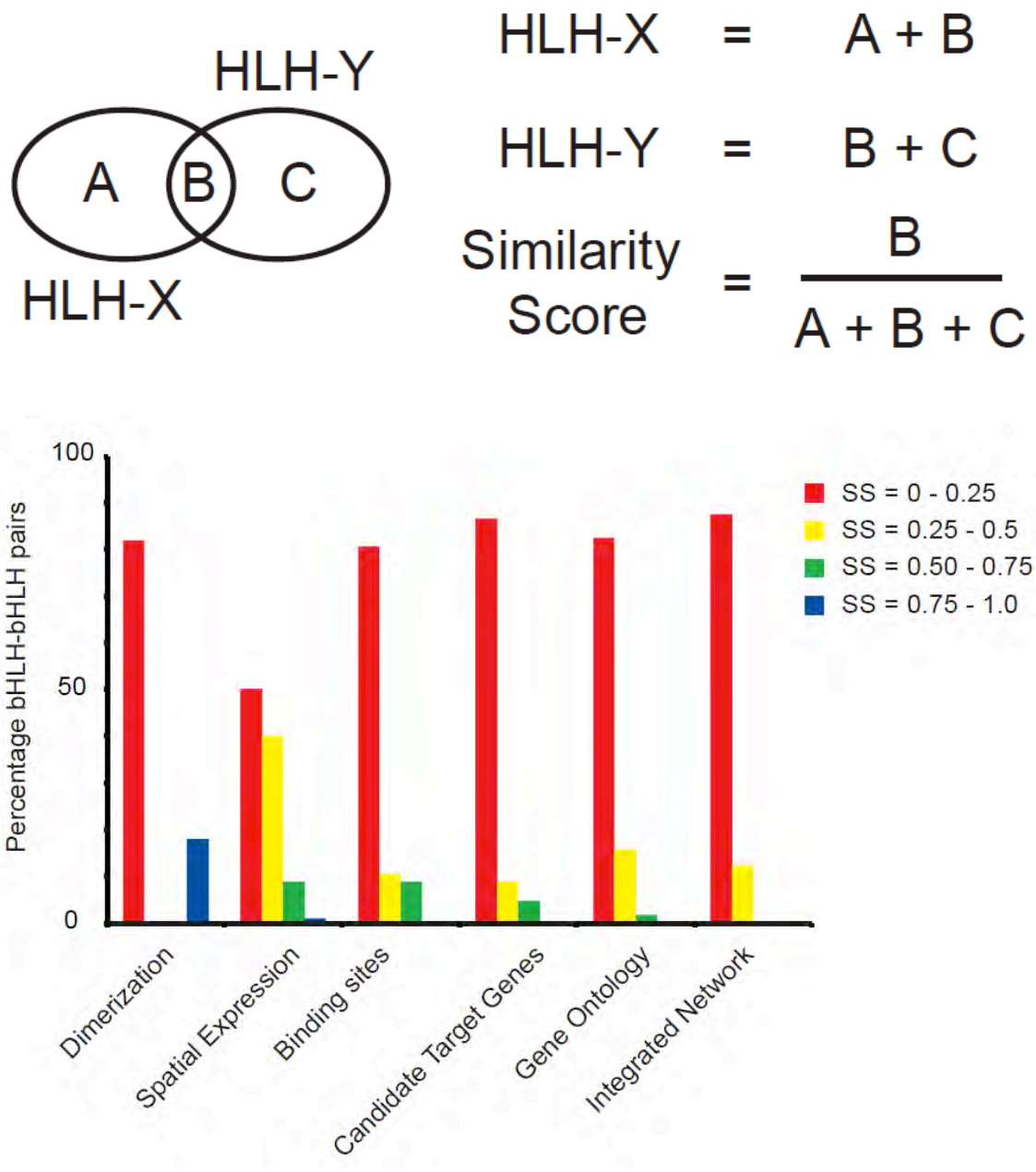


Figure V-9. A measure of divergence: the Similarity Score
(Top panel) For each bHLH-bHLH pair we calculated a Similarity Score (SS) for each functional TF parameter as indicated. (Bottom panel) Integrated parameter overlap analysis of all bHLH-bHLH pairs and dimer pairs (see Figure V-10 for individual parameter analysis). SSs were binned into four groups as indicated.

Figure V-10A

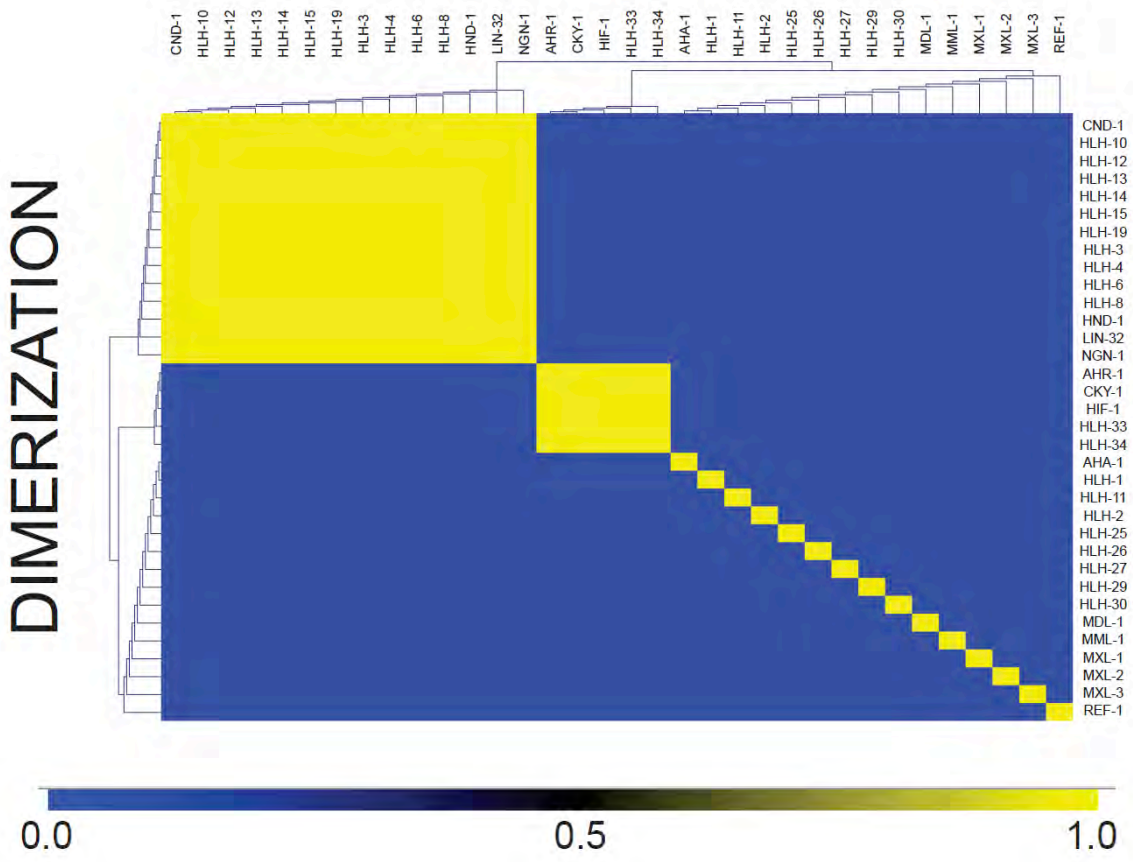


Figure V-10A. Heatmaps of similarity scores for comparisons of dimerization specificities between bHLH TFs

Figure V-10B

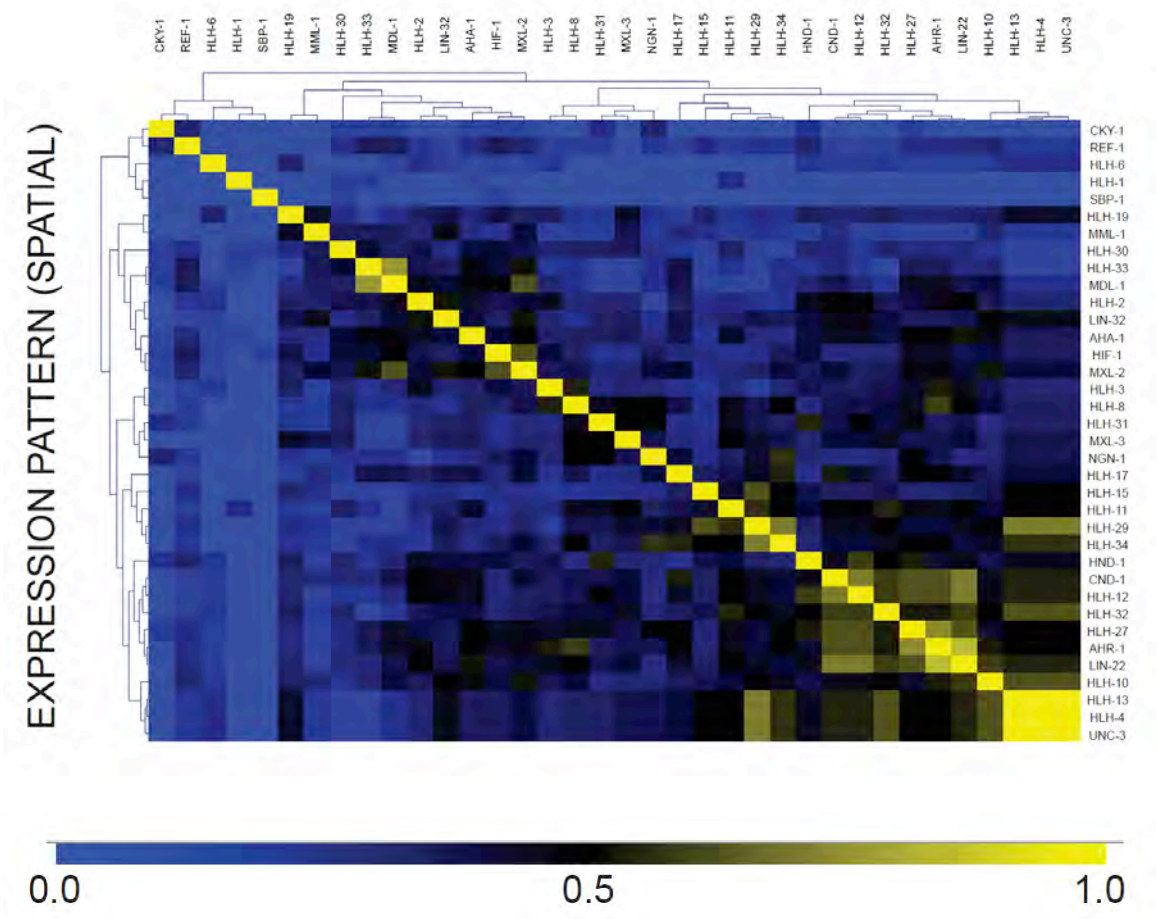


Figure V-10B. Heatmaps of similarity scores for comparisons of expression patterns between individual bHLH gene promoters

Figure V-10C

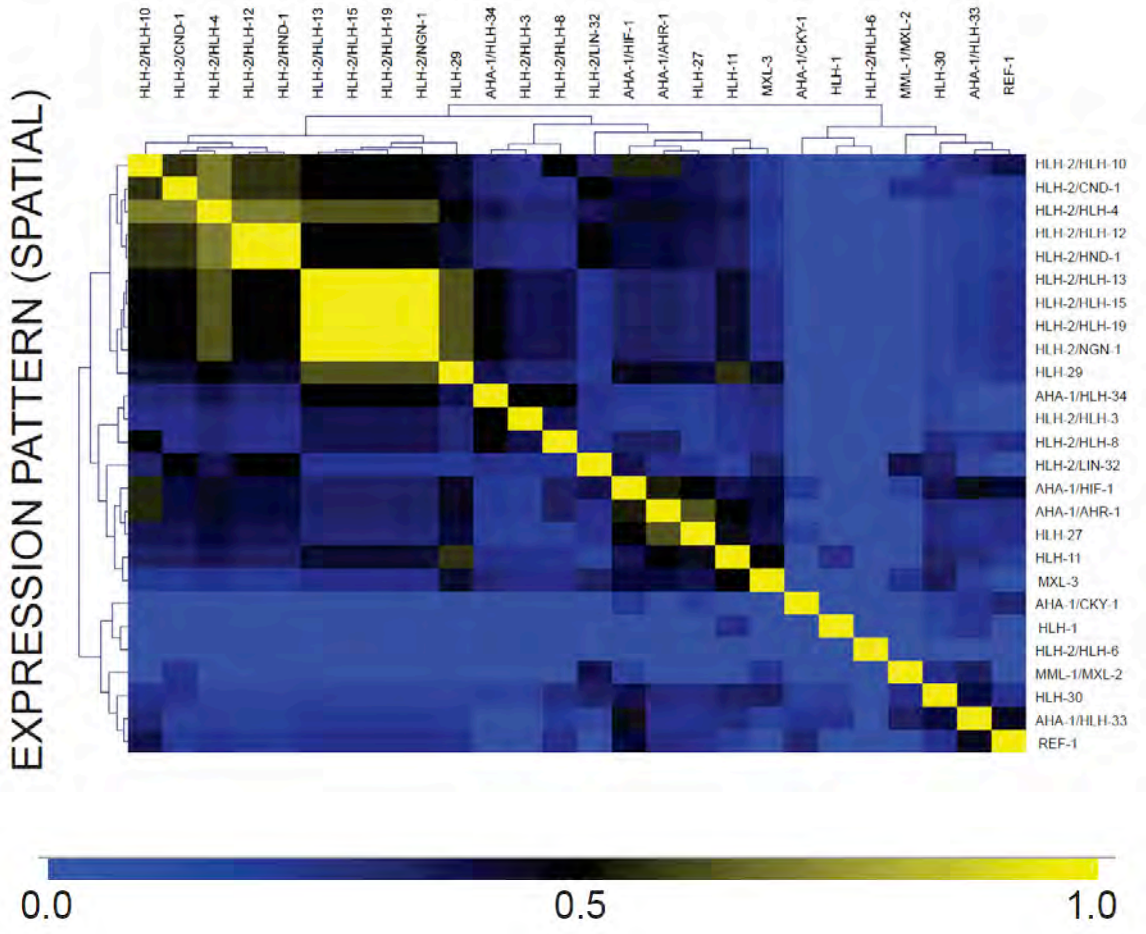


Figure V-10C. Heatmaps of similarity scores for comparisons between bHLH dimer expression patterns

Figure V-10D

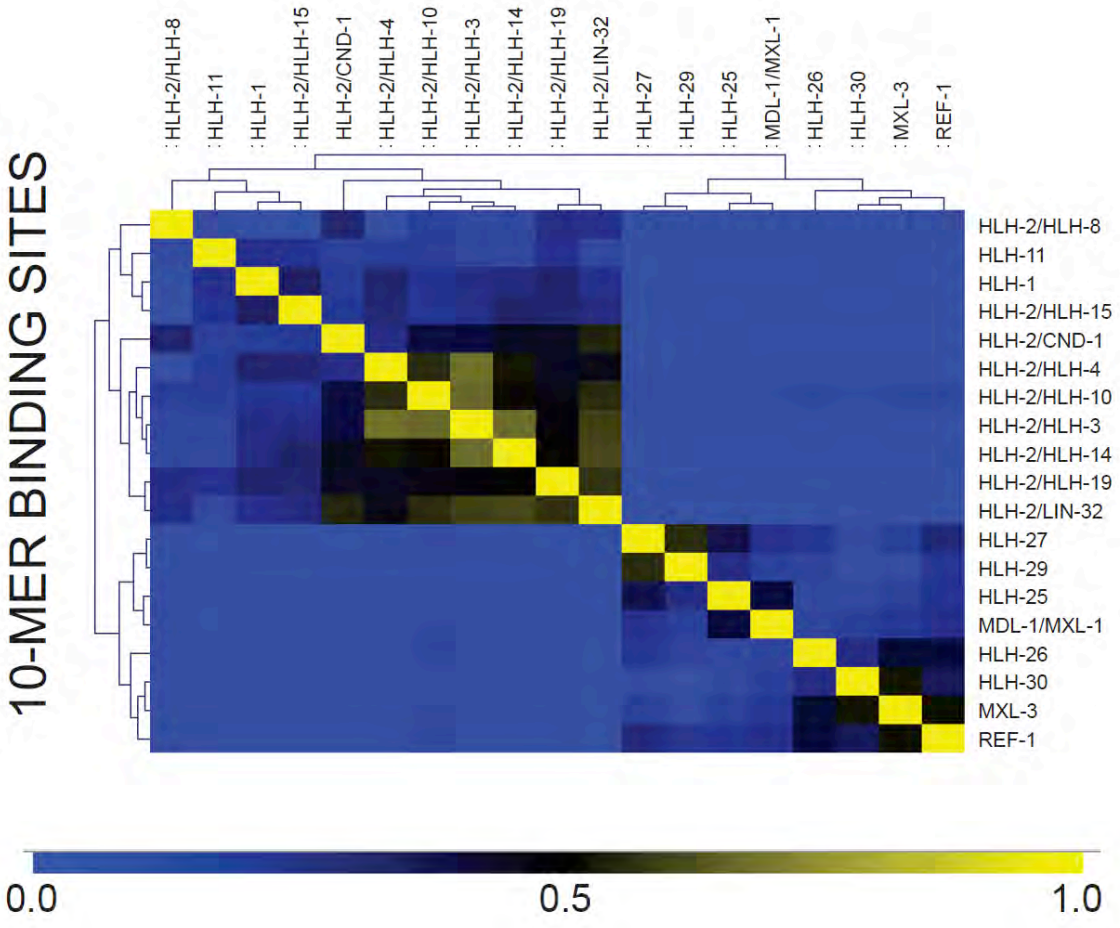


Figure V-10D. Heatmaps of similarity scores for comparisons between bHLH dimer 10-mer binding sites

Figure V-10E

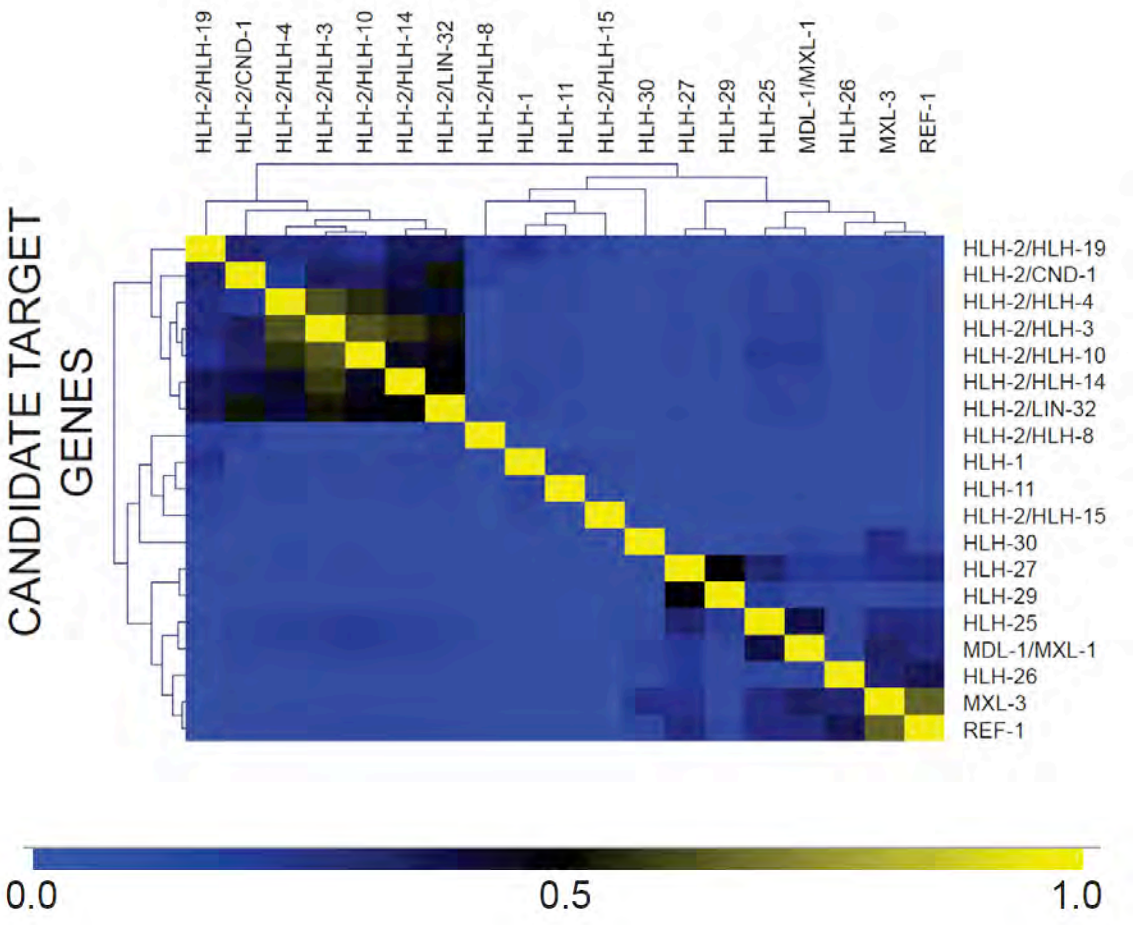


Figure V-10E. Heatmaps of similarity scores for comparisons between bHLH dimer candidate target genes

Figure V-10F

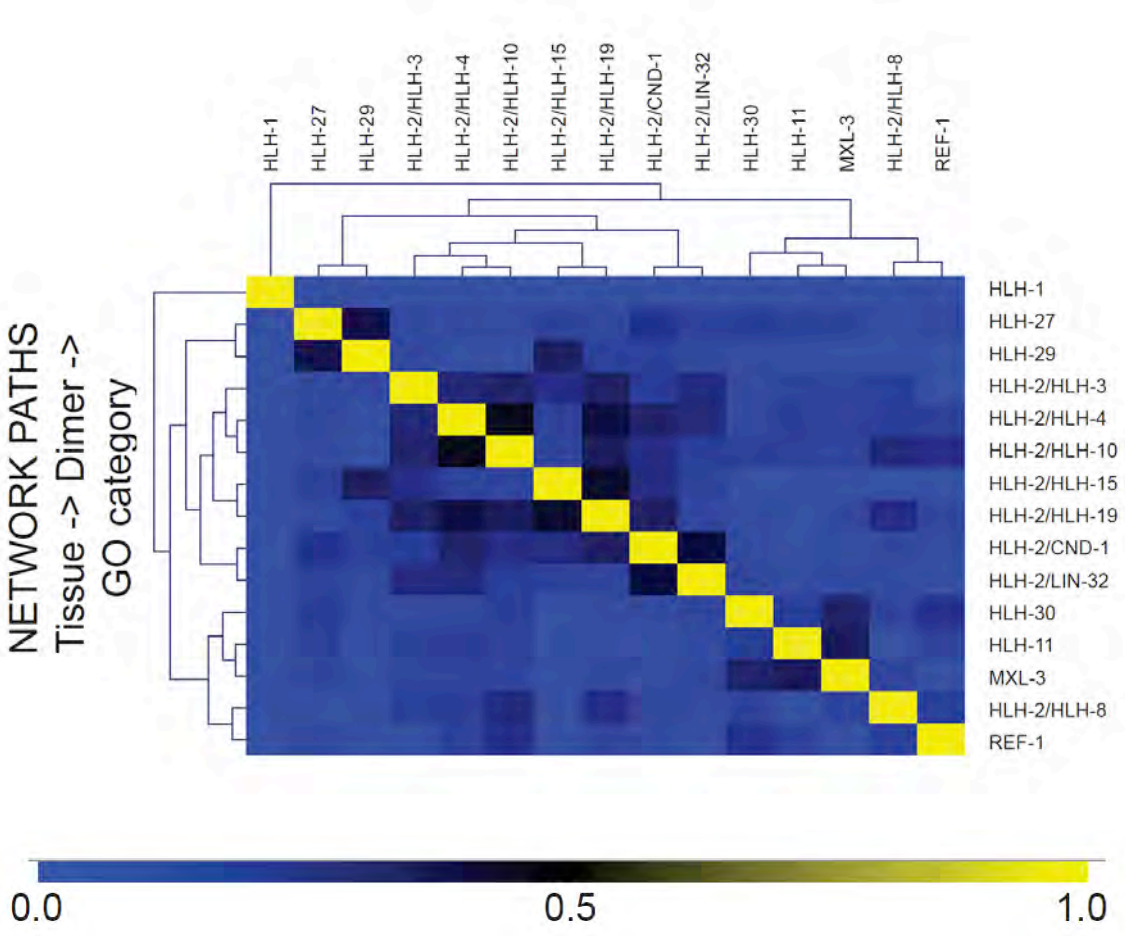


Figure V-10F. Heatmaps of similarity scores for comparisons between bHLH "network paths"

Figure V-11

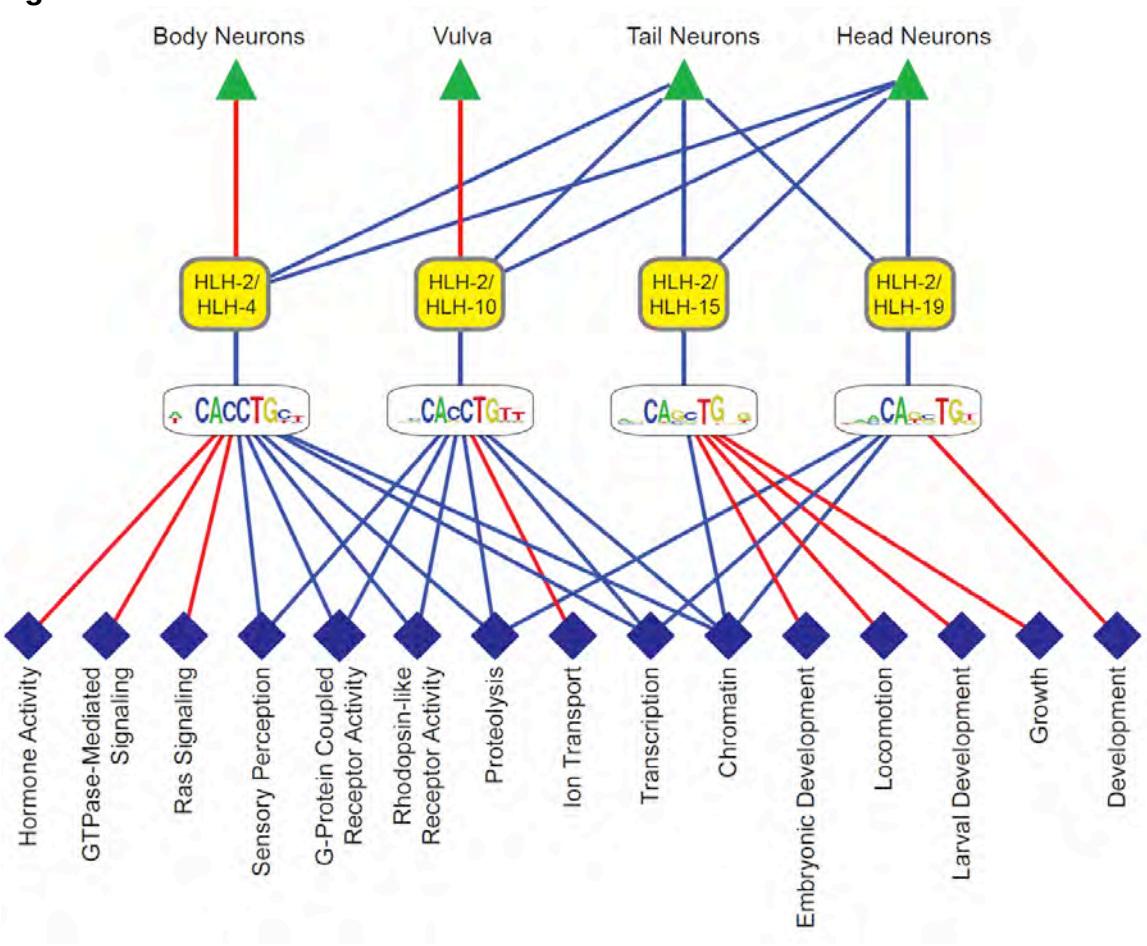


Figure V-11. Sub-network for HLH-2/HLH-4, HLH-2/HLH-10, HLH-2/HLH-15, and HLH-2/HLH-19

Sub-network of bHLH proteins with the highest degree of similarity. Red lines – unique functional parameters; blue lines – shared functional parameters. Blue Diamonds – Gene ontologies

Figure V-12

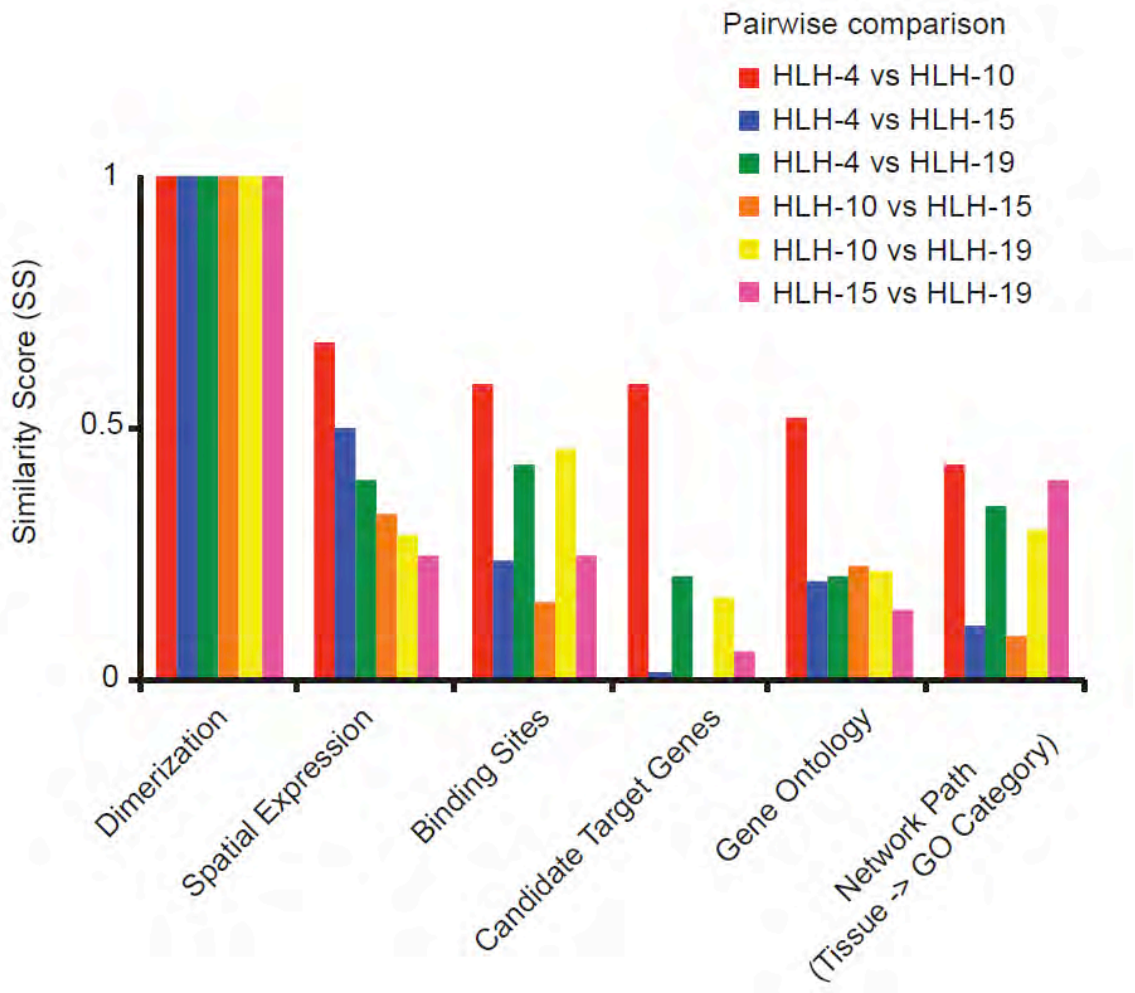


Figure V-12. Similarity score profiles for HLH-2/HLH-4, HLH-2/HLH-10, HLH-2/HLH-15, HLH-2/HLH-19
Individual similarity scores for all bHLH-bHLH pairs shown in Figure V-11. Bar graphs presented as in Figure V-9.

Figure V-13

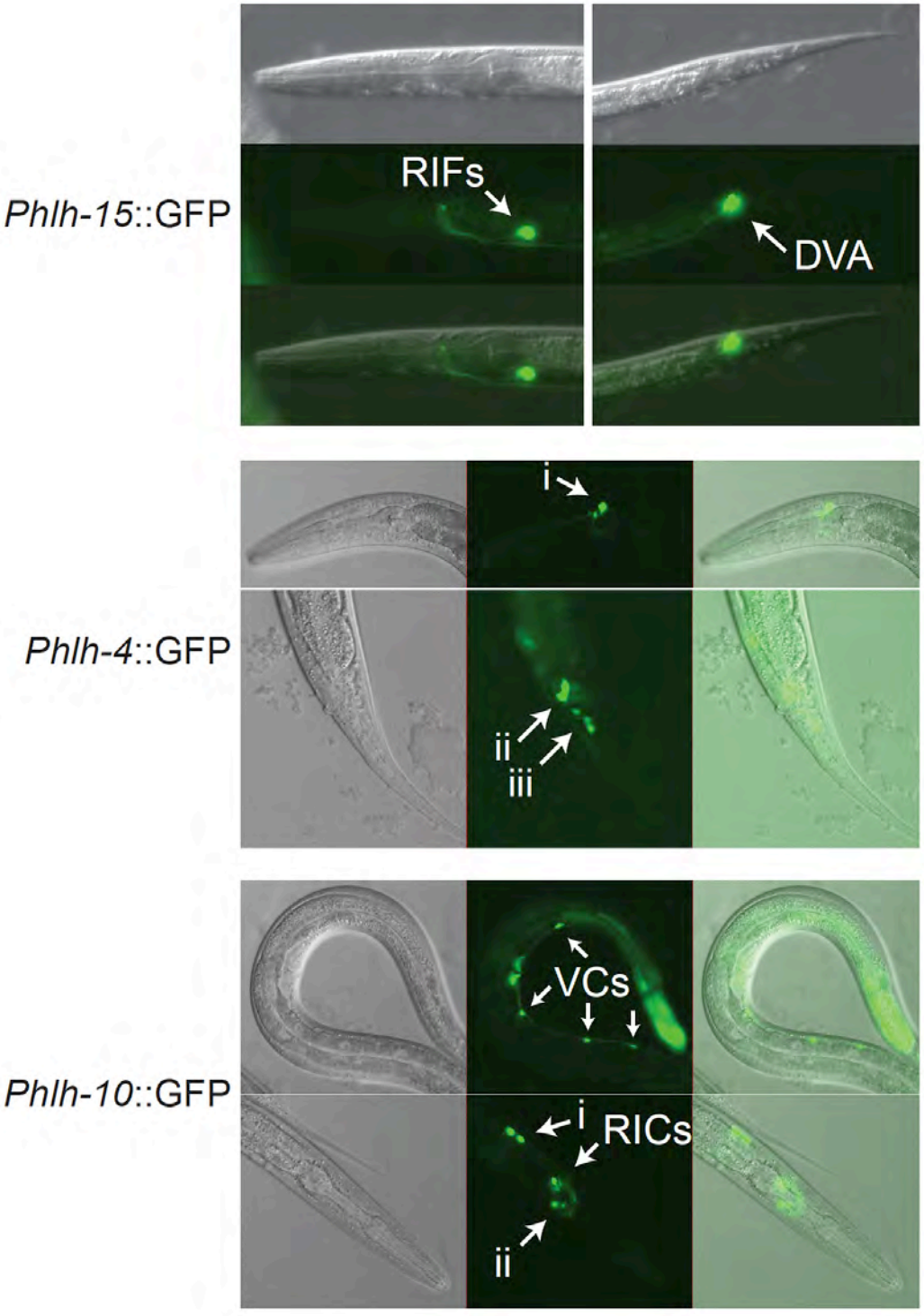


Figure V-13. Cellular resolution expression annotation for *Phlh-4*, *Phlh-10*, and *Phlh-15*

Detailed analysis of neuronal expression conferred by *Phlh-15, Phlh-4* and *Phlh-10*. *Phlh-15* appears to drive GFP expression in the pair of RIF neurons of the retrovesicular ganglion and the single DVA tail neuron. *Phlh-4* drives GFP expression in: i) two sensory head neurons (one bilaterally symmetric pair) of the lateral ganglion, likely AWA or AWB; ii) three pairs of tail neurons of the lumbar ganglion, likely PVQ, PVC, PVW, and/or LUA; iii) two tail neurons (likely a bilaterally symmetric pair) of the lumbar ganglion with processes to the tail. *Phlh-10* drives GFP expression in: i) two interneurons (one bilaterally symmetric pair) of the retrovesicular ganglion, likely RIF or RIG; ii) two sensory head neurons (one bilaterally symmetric pair) of the lateral ganglion, likely AWA or AWB.

Figure V-14

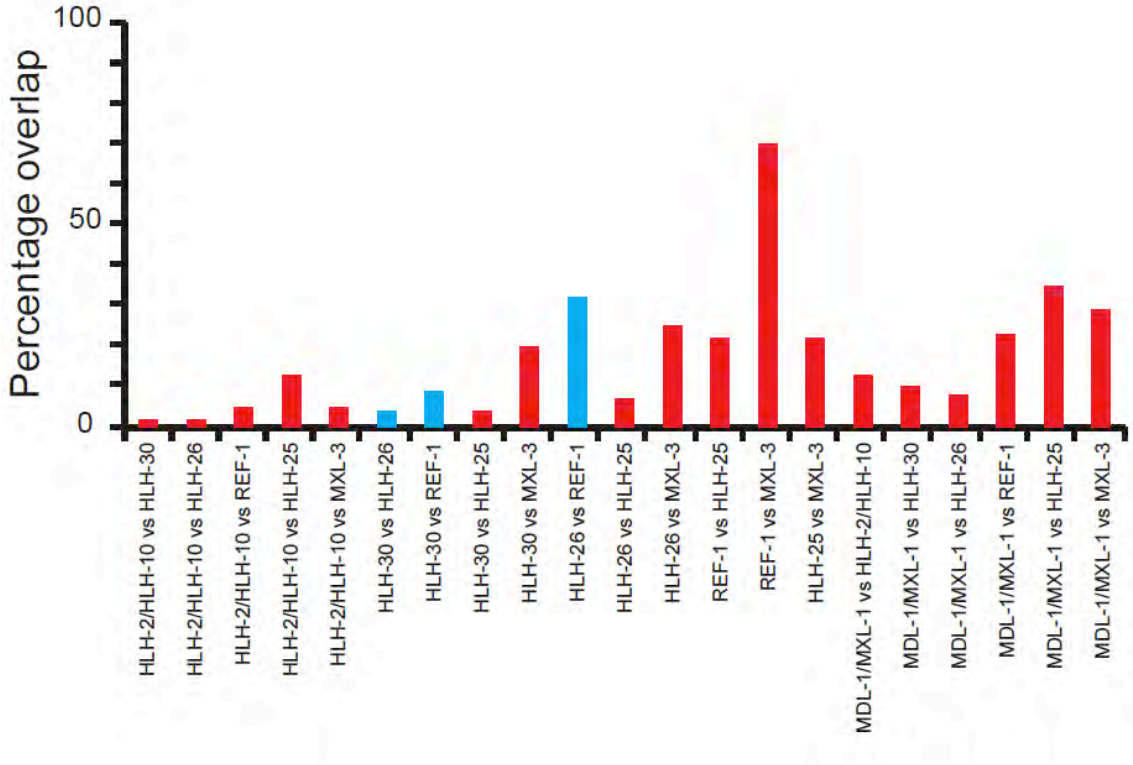


Figure V-14. Candidate target gene overlap for CACGTG binders
Percentage overlap of candidate target genes comparing bHLH dimers that can bind CACGTG E-boxes. Blue bars indicate comparisons in which both dimers exclusively bind CACGTG, red indicates comparisons in which one or both dimers can also bind other E-boxes or E-box-like sequences.

REFERENCES

1. Lemon B, Tjian R. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. 2000 Oct 15;14(20):2551-2569.

2. Kummerfeld SK, Teichmann SA. DBD: a transcription factor prediction database. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D74-81.

3. Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. Genome Res. 2004 Oct ;14(10B):2041-2047.

4. Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJM. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 2005 ;6(13):R110.

5. Vermeirssen V, Deplancke B, Barrasa MI, Reece-Hoyes JS, Arda HE, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Brent MR, Walhout AJM. Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. Nat. Methods. 2007 Aug ;4(8):659-664.

6. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, Tissenbaum HA, Mango SE, Walhout AJM. A gene-centered C. elegans protein-DNA interaction network. Cell. 2006 Jun 16;125(6):1193-1205.

7. Vermeirssen V, Barrasa MI, Hidalgo CA, Babon JAB, Sequerra R, Doucette-Stamm L, Barabási A, Walhout AJM. Transcription factor modularity in a gene-centered C. elegans core neuronal protein-DNA interaction network. Genome Res. 2007 Jul ;17(7):1061-1071.

8. Walhout AJM. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res. 2006 Dec ;16(12):1445-1454.

9. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004 Sep 2;431(7004):99-104.

10. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. Core transcriptional regulatory circuitry in human embryonic

stem cells. Cell. 2005 Sep 23;122(6):947-956.

11. Sandmann T, Jensen LJ, Jakobsen JS, Karzynski MM, Eichenlaub MP, Bork P, Furlong EEM. A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev. Cell. 2006 Jun ;10(6):797-807.

12. Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat. Genet. 2004 Dec ;36(12):1331-1339.

13. van Steensel B, Henikoff S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat. Biotechnol. 2000 Apr ;18(4):424-428.

14. Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu X, White KP, Bussemaker HJ, van Steensel B. Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 2006 Aug 8;103(32):12027-12032.

15. Meng X, Brodsky MH, Wolfe SA. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nat. Biotechnol. 2005 Aug ;23(8):988-994.

16. Deplancke B, Dupuy D, Vidal M, Walhout AJM. A gateway-compatible yeast one-hybrid system. Genome Res. 2004 Oct ;14(10B):2093-2101.

17. Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, Snyder M. Target hub proteins serve as master regulators of development in yeast. Genes Dev. 2006 Feb 15;20(4):435-448.

18. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. 2004 Jun ;14(3):283-291.

19. Guelzim N, Bottani S, Bourgine P, Képès F. Topological and causal structure of the yeast transcriptional regulatory network. Nat. Genet. 2002 May ;31(1):60-63.

20. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat. Genet. 2002 May ;31(1):64-68.

21. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002 Oct

25;298(5594):824-827.

22. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, Pan ZJ, Schilstra MJ, Clarke PJC, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H. A genomic regulatory network for development. Science. 2002 Mar 1;295(5560):1669-1678.

23. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature. 2004 Sep 16;431(7006):308-312.

24. Smith J, Theodoris C, Davidson EH. A gene regulatory network subcircuit drives a dynamic pattern of gene expression. Science. 2007 Nov 2;318(5851):794-797.

25. Taneri B, Snyder B, Novoradovsky A, Gaasterland T. Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. Genome Biol. 2004 ;5(10):R75.

26. Ogg S, Paradis S, Gottlieb S, Patterson GI, Lee L, Tissenbaum HA, Ruvkun G. The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. Nature. 1997 Oct 30;389(6654):994-999.

27. Oh SW, Mukhopadhyay A, Dixit BL, Raha T, Green MR, Tissenbaum HA. Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. Nat. Genet. 2006 Feb ;38(2):251-257.

28. Bach I, Yaniv M. More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. EMBO J. 1993 Nov ;12(11):4229-4242.

29. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003 Dec 19;302(5653):2141-2144.

30. Wolberger C. Multiprotein-DNA complexes in transcriptional regulation. Annu Rev Biophys Biomol Struct. 1999 ;2829-56.

31. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 Jun 27;300(5628):2097-2101.

32. Lamb P, McKnight SL. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. Trends Biochem. Sci. 1991 Nov ;16(11):417-422.

33. Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJM. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell. 2009 Jul 23;138(2):314-327.

34. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M. Protein interaction mapping in C. elegans using proteins involved in vulval development. Science. 2000 Jan 7;287(5450):116-122.

35. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. 1990 Aug 3;249(4968):505-510.

36. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990 Aug 30;346(6287):818-822.

37. Blackwell TK, Weintraub H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science. 1990 Nov 23;250(4984):1104-1110.

38. Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN, Weintraub H. Sequence-specific DNA binding by the c-Myc protein. Science. 1990 Nov 23;250(4984):1149-1151.

39. Meng X, Wolfe SA. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. Nat Protoc. 2006 ;1(1):30-45.

40. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. 2006 Nov ;24(11):1429-1435.

41. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell. 2008 Jun 27;133(7):1277-1289.

42. Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH, Wolfe SA. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic Acids Res. 2008 May ;36(8):2547-2560.

43. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell. 2008 Jun 27;133(7):1266-1276.

44. Schneuwly S, Gehring WJ. Homeotic transformation of thorax into head: Developmental analysis of a new Antennapedia allele in Drosophila melanogaster. Developmental Biology. 1985 Apr ;108(2):377-386.

45. Schneuwly S, Klemenz R, Gehring WJ. Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature. 1987 Mar 26;325(6107):816-818.

46. Berk AJ. Regulation of eukaryotic transcription factors by post-translational modification. Biochim. Biophys. Acta. 1989 Nov 2;1009(2):103-109.

47. Tremblay A, Tremblay GB, Labrie F, Giguère V. Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1. Mol. Cell. 1999 Apr ;3(4):513-519.

48. Kim MY, Woo EM, Chong YTE, Homenko DR, Kraus WL. Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. Mol. Endocrinol. 2006 Jul ;20(7):1479-1493.

49. Dahlman-Wright K, Cavailles V, Fuqua SA, Jordan VC, Katzenellenbogen JA, Korach KS, Maggi A, Muramatsu M, Parker MG, Gustafsson J. International Union of Pharmacology. LXIV. Estrogen receptors. Pharmacol. Rev. 2006 Dec ;58(4):773-781.

50. Smith CL, O'Malley BW. Coregulator function: a key to understanding tissue specificity of selective receptor modulators. Endocr. Rev. 2004 Feb ;25(1):45-71.

51. Motola DL, Cummins CL, Rottiers V, Sharma KK, Li T, Li Y, Suino-Powell K, Xu HE, Auchus RJ, Antebi A, Mangelsdorf DJ. Identification of ligands for DAF-12 that govern dauer formation and reproduction in C. elegans. Cell. 2006 Mar 24;124(6):1209-1223.

52. Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. Annu. Rev. Pharmacol. Toxicol. 2003 ;43309-334.

53. Denison MS, Heath-Pagliuso S. The Ah receptor: a regulator of the

biochemical and toxicological actions of structurally diverse chemicals. Bull Environ Contam Toxicol. 1998 Nov ;61(5):557-568.

54. Qin H, Powell-Coffman JA. The Caenorhabditis elegans aryl hydrocarbon receptor, AHR-1, regulates neuronal development. Dev. Biol. 2004 Jun 1;270(1):64-75.

55. Huang X, Powell-Coffman JA, Jin Y. The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in C. elegans. Development. 2004 Feb ;131(4):819-828.

56. Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. 2005 Feb 7;579(4):909-915.

57. Rosenfeld MG, Lunyak VV, Glass CK. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes Dev. 2006 Jun 1;20(11):1405-1428.

58. Perissi V, Rosenfeld MG. Controlling nuclear receptors: the circular logic of cofactor cycles. Nat. Rev. Mol. Cell Biol. 2005 Jul ;6(7):542-554.

59. Yu Y, Li W, Su K, Yussa M, Han W, Perrimon N, Pick L. The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. Nature. 1997 Feb 6;385(6616):552-555.

60. Strano S, Dell'Orso S, Di Agostino S, Fontemaggi G, Sacchi A, Blandino G. Mutant p53: an oncogenic transcription factor. Oncogene. 2007 Apr 2;26(15):2212-2219.

61. Firulli BA, Krawchuk D, Centonze VE, Vargesson N, Virshup DM, Conway SJ, Cserjesi P, Laufer E, Firulli AB. Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. Nat. Genet. 2005 Apr ;37(4):373-381.

62. Castilla LH, Garrett L, Adya N, Orlic D, Dutra A, Anderson S, Owens J, Eckhaus M, Bodine D, Liu PP. The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia. Nat. Genet. 1999 Oct ;23(2):144-146.

63. Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. 2007 ;733.

64. Chen L, Krause M, Sepanski M, Fire A. The Caenorhabditis elegans MYOD homologue HLH-1 is essential for proper muscle function and complete morphogenesis. Development. 1994 Jun ;120(6):1631-1641.

65. Hallam S, Singer E, Waring D, Jin Y. The C. elegans NeuroD homolog cnd-1 functions in multiple aspects of motor neuron fate specification. Development. 2000 Oct ;127(19):4239-4252.

66. Portman DS, Emmons SW. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. Development. 2000 Dec ;127(24):5415-5426.

67. Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, Garcia Delgado C, Gonzalez-Ramos M, Kline AD, Jabs EW. Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. Nat. Genet. 1997 Jan ;15(1):36-41.

68. Reamon-Buettner SM, Ciribilli Y, Inga A, Borlak J. A loss-of-function mutation in the binding domain of HAND1 predicts hypoplasia of the human hearts. Hum. Mol. Genet. 2008 May 15;17(10):1397-1405.

69. Fernandes L, Rodrigues-Pousada C, Struhl K. Yap, a novel family of eight bZIP proteins in Saccharomyces cerevisiae with distinct biological functions. Mol. Cell. Biol. 1997 Dec ;17(12):6982-6993.

70. Tan K, Feizi H, Luo C, Fan SH, Ravasi T, Ideker TG. A systems approach to delineate functions of paralogous transcription factors: role of the Yap family in the DNA damage response. Proc. Natl. Acad. Sci. U.S.A. 2008 Feb 26;105(8):2934-2939.

71. Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell. Biol. 2000 Jan ;20(2):429-440.

72. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997 Dec 15;25(24):4876-4882.

73. Page RD. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 1996 Aug ;12(4):357-358.

74. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. Trends Biochem. Sci. 2008 May ;33(5):220-229.

75. Fields S, Song O. A novel genetic system to detect protein-protein

interactions. Nature. 1989 Jul 20;340(6230):245-246.

76. Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. U.S.A. 1991 Nov 1;88(21):9578-9582.

77. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. 1998 Dec 11;282(5396):2012-2018.

78. Korf I, Flicek P, Duan D, Brent MR. Integrating genomic homology into gene structure prediction. Bioinformatics. 2001 ;17 Suppl 1S140-148.

79. Wei C, Lamesch P, Arumugam M, Rosenberg J, Hu P, Vidal M, Brent MR. Closing in on the C. elegans ORFeome by cloning TWINSCAN predictions. Genome Res. 2005 Apr ;15(4):577-582.

80. Harfe BD, Vaz Gomes A, Kenyon C, Liu J, Krause M, Fire A. Analysis of a Caenorhabditis elegans Twist homolog identifies conserved and divergent aspects of mesodermal patterning. Genes Dev. 1998 Aug 15;12(16):2623-2635.

81. Jiang H, Guo R, Powell-Coffman JA. The Caenorhabditis elegans hif-1 gene encodes a bHLH-PAS protein that is required for adaptation to hypoxia. Proc. Natl. Acad. Sci. U.S.A. 2001 Jul 3;98(14):7916-7921.

82. Krause M, Park M, Zhang JM, Yuan J, Harfe B, Xu SQ, Greenwald I, Cole M, Paterson B, Fire A. A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. Development. 1997 Jun ;124(11):2179-2189.

83. Ooe N, Saito K, Oeda K, Nakatuka I, Kaneko H. Characterization of Drosophila and Caenorhabditis elegans NXF-like-factors, putative homologs of mammalian NXF. Gene. 2007 Oct 1;400(1-2):122-130.

84. Pickett CL, Breen KT, Ayer DE. A C. elegans Myc-like network cooperates with semaphorin and Wnt signaling pathways to control cell migration. Dev. Biol. 2007 Oct 15;310(2):226-239.

85. Tamai KK, Nishiwaki K. bHLH transcription factors regulate organ morphogenesis via activation of an ADAMTS protease in C. elegans. Dev. Biol. 2007 Aug 15;308(2):562-571.

86. Yuan J, Tirabassi RS, Bush AB, Cole MD. The C. elegans MDL-1 and MXL-1 proteins can functionally substitute for vertebrate MAD and MAX. Oncogene. 1998 Sep 3;17(9):1109-1118.

87. Powell-Coffman JA, Bradfield CA, Wood WB. Caenorhabditis elegans orthologs of the aryl hydrocarbon receptor and its heterodimerization partner the aryl hydrocarbon receptor nuclear translocator. Proc. Natl. Acad. Sci. U.S.A. 1998 Mar 17;95(6):2844-2849.

88. Crews ST. Control of cell lineage-specific development and transcription by bHLH-PAS proteins. Genes Dev. 1998 Mar 1;12(5):607-620.

89. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual J, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet A, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabási A, Tavernier J, Hill DE, Vidal M. High-quality binary protein interaction map of the yeast interactome network. Science. 2008 Oct 3;322(5898):104-110.

90. Hemesath TJ, Steingrímsson E, McGill G, Hansen MJ, Vaught J, Hodgkinson CA, Arnheiter H, Copeland NG, Jenkins NA, Fisher DE. microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. Genes Dev. 1994 Nov 15;8(22):2770-2780.

91. Nagoshi E, Yoneda Y. Dimerization of sterol regulatory element-binding protein 2 via the helix-loop-helix-leucine zipper domain is a prerequisite for its nuclear localization mediated by importin beta. Mol. Cell. Biol. 2001 Apr ;21(8):2779-2789.

92. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu. Rev. Cell Dev. Biol. 2000 ;16653-699.

93. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994 Nov 11;22(22):4673-4680.

94. Atchley WR, Fitch WM. A natural classification of the basic helix-loop-helix class of transcription factors. Proc. Natl. Acad. Sci. U.S.A. 1997 May 13;94(10):5172-5176.

95. Ellenberger T, Fass D, Arnaud M, Harrison SC. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. 1994 Apr 15;8(8):970-980.

96. Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature. 1993

May 6;363(6424):38-45.

97. Longo A, Guanga GP, Rose RB. Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry. 2008 Jan 8;47(1):218-229.

98. Thattaliyath BD, Firulli BA, Firulli AB. The basic-helix-loop-helix transcription factor HAND2 directly regulates transcription of the atrial naturetic peptide gene. J. Mol. Cell. Cardiol. 2002 Oct ;34(10):1335-1344.

99. Porcher C, Liao EC, Fujiwara Y, Zon LI, Orkin SH. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. Development. 1999 Oct ;126(20):4603-4615.

100. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. DBD--taxonomically broad transcription factor predictions: new content and functionality. Nucl. Acids Res. 2008 Jan 18;36(suppl_1):D88-92.

101. Stark C, Breitkreutz B, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D535-539.

102. Reboul J, Vaglio P, Rual J, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, Moore T, Hudson JR, Hartley JL, Brasch MA, Vandenhaute J, Boulton S, Endress GA, Jenna S, Chevet E, Papasotiropoulos V, Tolias PP, Ptacek J, Snyder M, Huang R, Chance MR, Lee H, Doucette-Stamm L, Hill DE, Vidal M. C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat. Genet. 2003 May ;34(1):35-41.

103. Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Meth. Enzymol. 2000 ;328575-592.

104. Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods. 2001 Jul ;24(3):297-306.

105. Walhout AJ, Vidal M. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. Genome Res. 1999 Nov ;9(11):1128-1134.

106. Christmas R, Avila-Campillo I, Bolouri H, Schwikowski B, Anderson M, Kelley R, Landys N, Workman C, Ideker T, Cerami E, Sheridan R, Bader GD, Sander C. Cytoscape: A Software Environment for Integrated Models of

Biomolecular Interaction Networks [Internet]. 2005. p. 12-16.[cited 2009 Sep 2] Available from: http://educationbook.aacrjournals.org/cgi/content/full/2005/1/12

107. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003 Nov ;13(11):2498-2504.

108. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang P, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007 ;2(10):2366-2382.

109. Ledent V, Paquet O, Vervoort M. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 2002 ;3(6):RESEARCH0030.

110. Horton JD, Shimomura I. Sterol regulatory element-binding proteins: activators of cholesterol and fatty acid biosynthesis. Curr. Opin. Lipidol. 1999 Apr ;10(2):143-150.

111. Nakagawa Y, Shimano H, Yoshikawa T, Ide T, Tamura M, Furusawa M, Yamamoto T, Inoue N, Matsuzaka T, Takahashi A, Hasty AH, Suzuki H, Sone H, Toyoshima H, Yahagi N, Yamada N. TFE3 transcriptionally activates hepatic IRS-2, participates in insulin signaling and ameliorates diabetes. Nat. Med. 2006 Jan ;12(1):107-113.

112. Neves A, Priess JR. The REF-1 family of bHLH transcription factors pattern C. elegans embryos through Notch-dependent and Notch-independent pathways. Dev. Cell. 2005 Jun ;8(6):867-879.

113. Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJM. Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res. 2008 Dec ;18(12):2005-2015.

114. Baird GS, Zacharias DA, Tsien RY. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. Proc. Natl. Acad. Sci. U.S.A. 2000 Oct 24;97(22):11984-11989.

115. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. 2004 Dec ;22(12):1567-1572.

116. Wall MA, Socolich M, Ranganathan R. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. Nat. Struct. Biol. 2000 Dec ;7(12):1133-1138.

117. Green RA, Audhya A, Pozniakovsky A, Dammermann A, Pemble H, Monen J, Portier N, Hyman A, Desai A, Oegema K. Expression and imaging of fluorescent proteins in the C. elegans gonad and early embryo. Methods Cell Biol. 2008 ;85179-218.

118. Takahashi E, Takano T, Numata A, Hayashi N, Okano S, Nakajima O, Nomura Y, Sato M. Genetic oxygen sensor: GFP as an indicator of intracellular oxygenation. Adv. Exp. Med. Biol. 2005 ;56639-44.

119. Takahashi E, Takano T, Nomura Y, Okano S, Nakajima O, Sato M. In vivo oxygen imaging using green fluorescent protein. Am. J. Physiol., Cell Physiol. 2006 Oct ;291(4):C781-787.

120. Hirsh D, Oppenheim D, Klass M. Development of the reproductive system of Caenorhabditis elegans. Dev. Biol. 1976 Mar ;49(1):200-219.

121. Cram EJ, Shang H, Schwarzbauer JE. A systematic RNA interference screen reveals a cell migration gene network in C. elegans. J. Cell. Sci. 2006 Dec 1;119(Pt 23):4811-4818.

122. Mathies LD, Henderson ST, Kimble J. The C. elegans Hand gene controls embryogenesis and early gonadogenesis. Development. 2003 Jul ;130(13):2881-2892.

123. Reece-Hoyes JS, Shingles J, Dupuy D, Grove CA, Walhout AJM, Vidal M, Hope IA. Insight into transcription factor gene duplication from Caenorhabditis elegans Promoterome-driven expression patterns. BMC Genomics. 2007 ;827.

124. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, McKay S, Okada HM, Pan J, Schulz AK, Tu D, Wong K, Zhao Z, Alexeyenko A, Burglin T, Sonnhammer E, Schnabel R, Jones SJ, Marra MA, Baillie DL, Moerman DG. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. PLoS Biol. 2007 Sep ;5(9):e237.

125. McKay SJ, Johnsen R, Khattra J, Asano J, Baillie DL, Chan S, Dube N, Fang L, Goszczynski B, Ha E, Halfnight E, Hollebakken R, Huang P, Hung K, Jensen V, Jones SJM, Kai H, Li D, Mah A, Marra M, McGhee J, Newbury R, Pouzyrev A, Riddle DL, Sonnhammer E, Tian H, Tu D, Tyson JR, Vatcher G, Warner A, Wong K, Zhao Z, Moerman DG. Gene expression profiling of

cells, tissues, and developmental stages of the nematode C. elegans. Cold Spring Harb. Symp. Quant. Biol. 2003 ;68159-169.

126. Doonan R, Hatzold J, Raut S, Conradt B, Alfonso A. HLH-3 is a C. elegans Achaete/Scute protein required for differentiation of the hermaphrodite-specific motor neurons. Mech. Dev. 2008 Oct ;125(9-10):883-893.

127. Etchberger JF, Lorch A, Sleumer MC, Zapf R, Jones SJ, Marra MA, Holt RA, Moerman DG, Hobert O. The molecular signature and cis-regulatory architecture of a C. elegans gustatory neuron. Genes Dev. 2007 Jul 1;21(13):1653-1674.

128. Yanowitz J, Fire A. Cyclin D involvement demarcates a late transition in C. elegans embryogenesis. Dev. Biol. 2005 Mar 1;279(1):244-251.

129. Fares H, Grant B. Deciphering endocytosis in Caenorhabditis elegans. Traffic. 2002 Jan ;3(1):11-19.

130. Loria PM, Hodgkin J, Hobert O. A conserved postsynaptic transmembrane protein affecting neuromuscular signaling in Caenorhabditis elegans. J. Neurosci. 2004 Mar 3;24(9):2191-2201.

131. Fares H, Greenwald I. Genetic analysis of endocytosis in Caenorhabditis elegans: coelomocyte uptake defective mutants. Genetics. 2001 Sep ;159(1):133-145.

132. Zhang Y, Grant B, Hirsh D. RME-8, a conserved J-domain protein, is required for endocytosis in Caenorhabditis elegans. Mol. Biol. Cell. 2001 Jul ;12(7):2011-2021.

133. White J, Southgate E, Thomson J, Brenner S. The structure of the nervous system of the nematode C. elegans. Phil. Trans. Royal Soc. London. Series B, Biol Scien. 1986 Nov 12;314(1165):1-340.

134. Dupuy D, Li Q, Deplancke B, Boxem M, Hao T, Lamesch P, Sequerra R, Bosak S, Doucette-Stamm L, Hope IA, Hill DE, Walhout AJM, Vidal M. A first version of the Caenorhabditis elegans Promoterome. Genome Res. 2004 Oct ;14(10B):2169-2175.

135. Barrasa MI, Vaglio P, Cavasino F, Jacotot L, Walhout AJM. EDGEdb: a transcription factor-DNA interaction database for the analysis of C. elegans differential gene expression. BMC Genomics. 2007 ;821.

136. Zhu C, Byers KJRP, McCord RP, Shi Z, Berger MF, Newburger DE, Saulrieta K, Smith Z, Shah MV, Radhakrishnan M, Philippakis AA, Hu Y, De Masi F, Pacek M, Rolfs A, Murthy T, Labaer J, Bulyk ML. High-resolution

DNA-binding specificity analysis of yeast transcription factors. Genome Res. 2009 Apr ;19(4):556-566.

137. Peirano RI, Wegner M. The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic Acids Res. 2000 Aug 15;28(16):3047-3055.

138. Davis RL, Turner DL. Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. Oncogene. 2001 Dec 20;20(58):8342-8357.

139. Akazawa C, Sasai Y, Nakanishi S, Kageyama R. Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. J. Biol. Chem. 1992 Oct 25;267(30):21879-21885.

140. Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. Genes Dev. 1992 Dec ;6(12B):2620-2634.

141. Tietze K, Oellers N, Knust E. Enhancer of splitD, a dominant mutation of Drosophila, and its use in the study of functional domains of a helix-loop-helix protein. Proc. Natl. Acad. Sci. U.S.A. 1992 Jul 1;89(13):6152-6156.

142. Ohsako S, Hyer J, Panganiban G, Oliver I, Caudy M. Hairy function as a DNA-binding helix-loop-helix repressor of Drosophila sensory organ formation. Genes Dev. 1994 Nov 15;8(22):2743-2755.

143. Van Doren M, Bailey AM, Esnayra J, Ede K, Posakony JW. Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of achaete. Genes Dev. 1994 Nov 15;8(22):2729-2742.

144. Fisher F, Goding CR. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. 1992 Nov ;11(11):4103-4109.

145. Walhout AJ, van der Vliet PC, Timmers HT. Sequences flanking the E-box contribute to cooperative binding by c-Myc/Max heterodimers to adjacent binding sites. Biochim. Biophys. Acta. 1998 Apr 29;1397(2):189-201.

146. Maerkl SJ, Quake SR. A systems approach to measuring the binding energy landscapes of transcription factors. Science. 2007 Jan 12;315(5809):233-237.

147. van Nimwegen E. Finding regulatory elements and regulatory motifs: a general probabilistic framework. BMC Bioinformatics. 2007 ;8 Suppl 6S4.

148. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990 Oct 25;18(20):6097-6100.

149. Castillo-Davis CI, Hartl DL, Achaz G. cis-Regulatory and protein evolution in orthologous and duplicate genes. Genome Res. 2004 Aug ;14(8):1530-1536.

150. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 2000 May ;25(1):25-29.

151. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang C, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML. Diversity and Complexity in DNA Recognition by Transcription Factors [Internet]. Science. 2009 May 14;[cited 2009 Jun 2] Available from: http://www.ncbi.nlm.nih.gov/pubmed/19443739

152. Párraga A, Bellsolell L, Ferré-D'Amaré AR, Burley SK. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 A resolution. Structure. 1998 May 15;6(5):661-672.

153. Ferré-D'Amaré AR, Pognonec P, Roeder RG, Burley SK. Structure and function of the b/HLH/Z domain of USF. EMBO J. 1994 Jan 1;13(1):180-189.

154. Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell. 1994 May 6;77(3):451-459.

155. Philippakis AA, Qureshi AM, Berger MF, Bulyk ML. Design of compact, universal DNA microarrays for protein binding microarray experiments. J. Comput. Biol. 2008 Sep ;15(7):655-665.

156. Zeeberg B, Feng W, Wang G, Wang M, Fojo A, Sunshine M, Narasimhan S, Kane D, Reinhold W, Lababidi S, Bussey K, Riss J, Barrett J, Weinstein J. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biology. 2003 ;4(4):R28.

157. Braun P, Hu Y, Shen B, Halleck A, Koundinya M, Harlow E, LaBaer J. Proteome-scale purification of human proteins from bacteria. Proc. Natl. Acad. Sci. U.S.A. 2002 Mar 5;99(5):2654-2659.

158. Berger MF, Bulyk ML. Universal protein-binding microarrays for the

comprehensive characterization of the DNA-binding specificities of transcription factors. Nat Protoc. 2009 ;4(3):393-411.

159. Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV. enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W389-392.

160. Warner JB, Philippakis AA, Jaeger SA, He FS, Lin J, Bulyk ML. Systematic identification of mammalian regulatory motifs' target genes and functions. Nat. Methods. 2008 Apr ;5(4):347-353.

161. Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. 2009 Jan ;37(Database issue):D77-82.

162. Ashrafi K, Chang FY, Watts JL, Fraser AG, Kamath RS, Ahringer J, Ruvkun G. Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes. Nature. 2003 Jan 16;421(6920):268-272.

163. Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PVE, Kamath RS, Fraser AG, Ahringer J, Plasterk RHA. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol. 2003 Oct ;1(1):E12.

164. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JJ, Hao T, Berriz GF, Bertin N, Huang J, Chuang L, Li N, Mani R, Hyman AA, Sönnichsen B, Echeverri CJ, Roth FP, Vidal M, Piano F. Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. Nature. 2005 Aug 11;436(7052):861-865.

165. Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. Nat. Genet. 2008 Feb ;40(2):181-188.

166. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev. 2007 Aug 1;21(15):1882-1894.

167. Ow MC, Martinez NJ, Olsen PH, Silverman HS, Barrasa MI, Conradt B, Walhout AJM, Ambros V. The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in C. elegans. Genes Dev. 2008 Sep 15;22(18):2520-2534.

168. Van Gilst MR, Hadjivassiliou H, Jolly A, Yamamoto KR. Nuclear hormone receptor NHR-49 controls fat consumption and fatty acid composition in C. elegans. PLoS Biol. 2005 Feb ;3(2):e53.

169. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004 Feb 12;20(3):307-315.

170. Irizarry RA, Wu Z, Jaffee HA. Comparison of Affymetrix GeneChip expression measures. Bioinformatics. 2006 Apr 1;22(7):789-794.

171. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004 ;3Article3.

172. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. BioTechniques. 2003 Feb ;34(2):374-378.