ABSTRACT

LARGE, EDWARD ELLIS. *hunchback* and Ikaros-like Zinc Finger Genes Control Reproductive System Development in *Caenorhabditis elegans*. (Under the direction of Dr. Laura D. Mathies and Dr. Stephanie E. Curtis).

The entire *C. elegans* somatic reproductive system is generated from two somatic gonadal precursor (SGP) cells. The SGPs are specified during embryogenesis from the developing mesoderm and migrate to meet two primordial germ cells (PGCs). The SGPs and the PGCs form a four-cell gonad primordium that remains undivided until the L1 larvae hatches from its egg. Each L1 SGP is a multipotent blast cell that has the potential to generate all five tissues of the somatic reproductive system: one distal tip cell (DTC), one anchor cell (AC), dorsal uterine (DU) cells, ventral uterine (VU) cells, and sheath/spermathecae (SS) cells. When combined with a PGC one SGP can generate a self-fertilizing gonad arm complete with sperm, eggs, and embryos. Investigating *C. elegans* SGP specification and differentiation, therefore, provides an excellent system to understand the molecular mechanisms underlying lineage progression.

Two conserved transcription factors have been previously characterized in early SGP development. The C2H2 zinc finger GLI ortholog *tra-1* was found to control late SGP division and polarity whereas the bHLH dHand ortholog *hnd-1* was found to control early SGP survival. These investigations lead to the identification of the novel C2H2 zinc finger gene *ehn-3*, which interacts genetically with either *tra-1* or *hnd-1* to cause a nearly complete loss of gonad arms.

This dissertation focuses on the characterization of *ehn-3* using a variety of techniques. A combination of bioinformatics, genetic interactions, phenotypic

characterization, and chimeric rescue constructs indicates *ehn-3* is part of a previously unrecognized gene family similar to mammalian Ikaros and *Drosophila hunchback*; both of which are involved in lineage progression. An RNAi screen was used to search for genes upstream, downstream, and in parallel to *ehn-3*. Several chromatin factors known to physically interact with either *Drosophila hunchback* or mammalian Ikaros were discovered to genetically interact with *ehn-3*. Additional analyses provide support for a model where two chromatin remodeling complexes, SWI/SNF and NuRD, are acting in parallel with or antagonistically to *ehn-3* to control SGP lineage progression. This work, in turn, provides the starting point for an understanding of how *hunchback* and Ikaros-like genes control SGP lineage progression utilizing chromatin factors.

hunchback and Ikaros-like Zinc Finger Genes Control Reproductive System Development in *Caenorhabditis elegans*

by
Edward Ellis Large

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy.

Genetics

Raleigh, NC

2010

APPROVED BY:

Dr. Laura D. Mathies, Chair of Advisory Committee, Genetics	Dr. Stephanie E. Curtis Co-Chair of Advisory Committee, Genetics
Dr. David M. Bird Committee Member Functional Genomics	Dr. Patricia A. Estes Committee Member Genetics

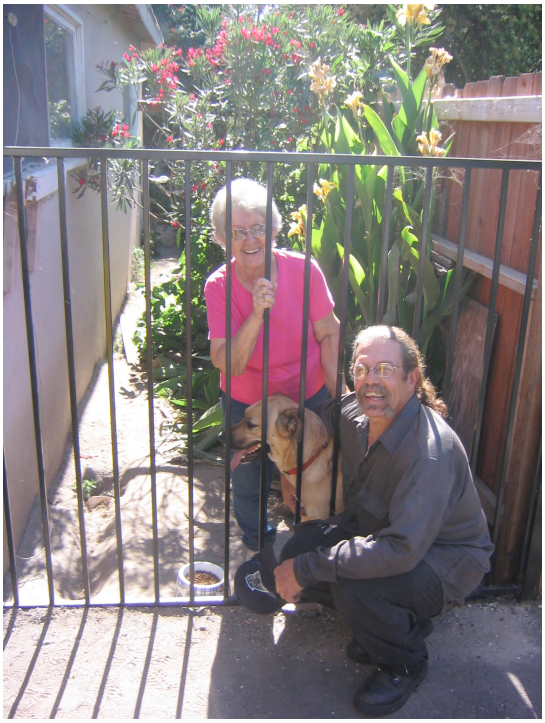
DEDICATION

To my Uncle Glen “Bo” Gilman, a rare blend of family and friend.

1954-2006

To my grandmother, Laraine E. Simoni, a strong independent woman who loved her family and her pets. You enjoyed a unique blend of cooking, bowling, riding dirt bikes, and target practice with pistols. Your quick wit and sense of humor are missed.

1928-2008



Grandma Laraine and Uncle Bo. Anaheim, CA (July 2005)

BIOGRAPHY

Edward Large is a native of Wyoming. He was born in Riverton, WY and raised in Fremont County, WY both within and outside the current borders of the Wind River Indian Reservation. Edward's paternal grandfather is an enrolled member of the Northern Arapaho tribe and his paternal grandmother is an enrolled member of the Eastern Shoshone tribe*. His maternal grandmother is a first-generation Italian American whose parents migrated from the Abruzzo and Lazio Regions of Italy. Edward's maternal grandfather is a World War II veteran and third-generation Californian whose own great-grandfather fought and died in the U.S. Civil War.

Edward excelled at academics, football, track, and swimming in high school (although he was one of two people not to make the basketball team) and served honorably in the Wyoming National Guard following graduation. He worked several full-time and part-time jobs while attending college full-time and was the first person in his family to both attend and graduate from college. Edward graduated from the University of Wyoming with a B.S. in Molecular Biology and worked for several years as a laboratory technician following graduation. While a technician at North Carolina State University, Edward was impressed with the caliber of the NCSU Genetics Department after taking Genetic Data Analysis (GN 721) with Dr. Bruce Weir. He subsequently began his doctoral studies at NCSU in 2004 and graduated with a PhD in Genetics and a minor in Biotechnology in 2010. Edward is looking forward to continuing a career in research, empowered with the skills and knowledge gleaned from those at the NCSU Genetics Department.

*Genotype as determined by the Illumina HumanHap550+ Genotyping BeadChip (~550,000 SNPs) is 0.1% African, 9.9% Asian, and 90.0% European.

ACKNOWLEDGMENTS

To begin, I would like to thank the NCSU Department of Genetics. In addition to a well-rounded curriculum, the department provided me with several opportunities for professional development and generously supported my travel to several conferences. The department also actively sought out funding opportunities that provided both financial support and excellent health insurance. I also had the pleasure of interacting with several graduate students, postdocs, technicians, and faculty members outside the lab and departmental environment. A stimulating atmosphere was often provided by the gracious Dr. Trudy Mackay and Dr. Robert Anholt over better than average wine. In addition, several grad students, postdocs, and technicians were occasionally willing to grab a few beers at Mitch's or East Village. Dr. Steve Spiker was thoughtful enough to open his house up during Oktoberfest and continues to host a friendly annual March Madness NCAA bracket competition. The Department of Genetics is also fortunate to have an excellent cadre of support staff and faculty members. Jennifer McCall, Lanakila Alexander, Suzanne Quick, Franda Sagcal, and Jeff Hawley have been incredibly helpful the past several years. While serving as Department Head, Dr. Stephanie Curtis managed to steer the department through a complete renovation. On a few occasions, a wall that was present when I entered the lab would be missing when I left later in the day. Floods and dust storms occasionally provided interesting experimental hurdles. It was truly an adventure and through it all, Stephanie managed to ensure that every student, including myself, felt safe and secure. A special thanks goes out to Dr. Bob Franks. Bob was kind enough to allow me to rotate through his

lab and was instrumental in my graduate education with courses such as Genetics of Reproduction (GN 810D) and Mechanisms of Plant Development (GN 810R). Bob also spearheaded the NCSU Genetics Outreach program; a program that provides digestible scientific information to the underserved general public. I firmly believe in the transformational powers of science and education on individuals and society and the Genetics Outreach program exemplifies these virtues. I would also like to thank Dr. Bill Atchley for his willingness to provide many contributions to my education both inside and outside the class. His careful instruction in both Population and Quantitative Genetics (GN 703) and Computational Molecular Evolution (GN 756) helped me appreciate the historic excellence of NCSU Genetics. Dr. Atchley was also instrumental in emphasizing the value of planning experiments and Statistics in scientific research. If it were not for Bill, I would not have been exposed to several excellent works on evolutionary biology. These texts, Bill's careful instruction, and his warranted admonition regarding the importance of timeliness have made me appreciate what it will take to evaluate the mechanisms underlying molecular evolution in the current century.

I am especially fortunate to have been part of the Mathies lab. I am grateful that Dr. Laura Mathies took me on as a new faculty member. Laura played a strong role as committee chair, advisor, and mentor. She is a logical, rigorous, and smart scientist and has provided a wealth of technical and theoretical insights. We had many lively discussions both inside and outside of classes ranging from Evolution of Development (GN 810T) to Chromatin Structure and Gene Regulation (GN 810C), co-taught with Dr. Steve Spiker. Dr. Mathies also provided guidance for me during two tough years as a TA for Honors Genetics (GN

411H) co-taught by Dr. Jeff Thorne. Laura's patience, guidance, and support have been greatly appreciated over the past several years and she has given me the tools necessary to develop into a mature scientist. Much of the work could not have been carried out as efficiently were it not for the important contributions of Zigrida Smith, a superb Research Technician. Furthermore, it was always a pleasure to chat with Zigrida about the pleasures of travel and she provided an excellent sounding board for graduate school dilemmas. I would be remiss not to thank Diana Klompstra for providing a spirited addition to the Mathies lab. Diana, now an NYU graduate student, entered the Mathies lab as an undergraduate at NCSU. In addition to an attention to detail and work ethic well beyond her years, Diana made it a point to put a smile on every face. Without this solid support group my laboratory life would have been substantially less fulfilling.

I would also like to thank my family and friends for their incredible patience. It has been a long six years and friends and family alike have suffered both personal and economic setbacks. I look forward to spending more time with family and friends now that the commitment to finishing my PhD program is finally over. I would especially like to thank LAJ for her constant companionship during the latter half of my PhD program. Liz, you are the one person I could count on to talk to and I deeply appreciate everything you have done and continue to do for me.

Finally, I would like to thank my committee members for their support. Dr. David Bird provided valuable advice during the Genomics Journal Club and during committee meetings. His background in *C. elegans* and his current investigations into plant parasitic nematodes

often provided relevant and thoughtful insights into my own project. Dr. Pat Estes provided one of the first courses, Molecular Genetics (GN 701), when I entered graduate school. I still have fond memories of all-night study sessions and my own research has benefited from the class. Pat's lab was next door to the Mathies lab and it was a pleasure to interact with her and her lab members over the years both inside and outside the lab. I was even more fortunate that Pat was willing to contribute her vast knowledge of developmental and molecular biology to my own research. I would also like to thank Dr. Stephanie Curtis for her willingness to be my committee co-chair. Stephanie was Department Head and co-taught Professionalism and Ethics (GN 850) with Dr. Julie Pederson during my first semester of graduate school. This provided a variety of opportunities to hone my writing skills in addition to opening my eyes to a wide array of personal, professional, and ethical academic dilemmas. Stephanie had and continues to have an open door policy as Department Head and now Director of Academic Programs. I am privileged to have had the opportunity to benefit from her wise council on more than a few occasions.

Last but not least I would like to thank the relevant funding agencies that provided graduate student support and laboratory support. I was fortunate to receive both a GAANN Biotechnology fellowship and a predoctoral fellowship from the National Institutes of Health. The Mathies lab research was further supported by a grant from the National Science Foundation.

TABLE OF CONTENTS

	Page
LIST OF TABLES…….……..	xii
LIST OF FIGURES…….…….	xiii
CHAPTER ONE: General Introduction ……...……....	1
Gene duplication and divergence…….……..	2
Versatility and expansion of C2H2 zinc finger genes……...	4
The deuterostome Ikaros family of C2H2 zinc finger transcription factors…..	7
The protostome *hunchback* family of C2H2 zinc finger transcription factors..	18
The search for the ancestral role of *hunchback* in animal development……	21
The *C. elegans* reproductive system as a model for lineage progression……..	25
The protostome HIL family of C2H2 zinc finger transcription factors…….	28
References……..	30
CHAPTER TWO: *hunchback* and Ikaros-like zinc finger genes control reproductive system development in *C. elegans*…….	49
Abstract……..	52
Introduction…….....	53
Materials and methods……..…….	57
Results…….....	64
A *hunchback*/Ikaros-like gene family in *C. elegans*……..……	64
Expression and function of the HIL genes in the developing gonad..……	65
ehn-3 and *R08E3.4* function in SGPs to control the development of differentiated gonadal cell types…….	68
The EHN-3 and R08E3.4 isoforms have different activities……..	74
EHN-3 and R08E3.4 have functionally distinct zinc finger domains…….	75
A new family of *hunchback* and Ikaros-like genes in protostomes……	77
Discussion…….....	79
Pleiotropic regulators of somatic gonad development……	79
Duplication and divergence of the HIL family in nematodes…….	81
A common origin for *hunchback*, Ikaros, and HIL transcription factors?.....	83
Acknowledgements…….……	86
References…….……..	87
CHAPTER THREE: *hunchback* and Ikaros-like zinc finger genes interact with chromatin factors to control reproductive system development in *Caenorhabditis elegans*……...	114
Summary……..	115
Introduction……...……..	116
Materials and methods…….	121
Results……..	127
SWI/SNF functions in somatic gonad development……..…….	128

SWI/SNF subunits act in parallel to *ehn-3*…….…….…. 131
ehn-3 has distinct interactions with NuRD components…….....…….…. 133
ehn-1 is *hda-2*……..…….…. 139
egl-27 has unique interactions with *ztf-16*……..…….…. 140
Discussion……...…….. 141
Distinct functions for PBAP and BAP in *C. elegans* somatic gonadogenesis…… 141
Antagonistic roles for the *C. elegans* NuRD complex in somatic gonad development……..... 143
Are EGL-27 and HDA-2 NuRD components?.....……. 145
How might SWI/SNF, NuRD, and EHN-3 regulate genes in *C. elegans*?....…… 146
Acknowledgements…… 149
References……. 150

CHAPTER FOUR: Summary and Future Work……..……..... 178
Summary……...…….. 180
Conclusion and Future Directions……. 181
References…….. 189

LIST OF TABLES

Page

CHAPTER TWO

Table 1: *R08E3.4* and *ehn-3* have overlapping functions in somatic gonad development…….. 94
Table 2: Correlation of L1 and adult phenotype in *ehn-3* and *ehn-3*; *R08E3.4* mutants…….. 95
Table 3: *ehn-3* and *R08E3.4* have overlapping functions in the generation of DTCs……. 96
Table 4: EHN-3 and R08E3.4 are not functionally equivalent……. 97
Table S1: Primer sequences…….. 98
Table S2: Additional genetic interaction tests…….. 100

CHAPTER THREE

Table 1: Chromatin factors interact genetically with *ehn-3*……. 158
Table 2: SWI/SNF acts alone and in parallel to *ehn-3* during somatic gonad development…..... 159
Table 3: Different roles for *C. elegans* PBAP and BAP during gonadogenesis 160
Table 4: Candidate NuRD subunits are modifiers of the *ehn-3* phenotype….. 161
Table 5: *egl-27* and *let-418* have opposing roles in parallel with *ehn-3*……... 162
Table 6: *egl-27* has unique interactions with *ztf-16*…….. 163
Table S1: *psa-4(os13)* and *psa-1(os22)* have temperature-sensitive defects... 164
Table S2: Maternal effect lethality of SWI/SNF subunits…… 165
Table S3: Qualitative differences between SWI/SNF and NuRD interactions. 166
Table S4: *ehn-1/hda-2* has strong synergistic interactions with *ehn-3*…….. 167

LIST OF FIGURES

Page

CHAPTER ONE

Figure 1: Versatility of C2H2 zinc finger proteins……..……..…. 43
Figure 2: Duplication and divergence of the Ikaros family of transcription factors...……..…….. 45
Figure 3: Representative sample of known *hunchback* family members……..……. 46
Figure 4: The *C. elegans* somatic reproductive system as a model for organogenesis…… 48

CHAPTER TWO

Figure 1: Three HIL genes are expressed in the early somatic gonad lineage….. 101
Figure 2: *ehn-3* and *R08E3.4* are required for somatic gonad development….. 102
Figure 3: *ehn-3; R08E3.4* double mutants have abnormal SGPs…….. 103
Figure 4: *ehn-3; R08E3.4* double mutants have ectopic *arg-1* expression in the somatic gonad…… 104
Figure 5: EHN-3 and R08E3.4 have functionally distinct zinc finger domains… 105
Figure 6: Comparison of Ikaros, *hunchback*, and HIL genes……..... 106
Figure S1: Structure of the *C. elegans* HIL genes……. 108
Figure S2: Chimeric rescue constructs…….. 109
Figure S3: Alignment of six conserved zinc fingers from Ikaros, *hunchback*, and HIL genes…….. 110
Figure S4: Neighbor-Joining tree…….. 111
Figure S5: Organization of zinc fingers from representative protostome *hunchback* and HIL genes and deuterostome IFLs…….. 112
Figure S6: Additional alignments……..... 113

CHAPTER THREE

Figure 1: SWI/SNF and *ehn-3* have a synergistic phenotype…… 168
Figure 2: *egl-27* is expressed at multiple stages of SGP and somatic gonad development…… 170
Figure 3: Maternal *egl-27* and *ehn-3* have antagonistic functions in somatic gonad Development……. 171
Figure 4: *let-418* and *ehn-3* have antagonistic functions in somatic gonad development……..…… 173
Figure 5: *ehn-1* is allelic to *hda-2* and is expressed in the somatic gonad…….. 175
Figure 6: SWI/SNF and NuRD act at multiple times in somatic gonad development…… 177

CHAPTER ONE

General Introduction

Gene duplication and divergence

The duplication and divergence of genes is an important mechanism for generating new genes with novel or redundant functions. Early geneticists recognized the evolutionary importance of gene duplication (BRIDGES 1936; HALDANE 1932; MULLER 1935; MULLER 1936) but the formal treatment of duplication and divergence as a model for the generation of evolutionary change was not developed until the 1970s (OHNO 1970). The geneticist and evolutionary biologist Susumu Ohno proposed that a gene can duplicate to generate two genes; one gene that retains the ancestral function and another gene that accumulates gene duplicate mutations, which lead to either new functions and preservation (i.e. neofunctionalization) or the acquisition of degenerate mutations leading to a pseudogene (i.e. nonfunctionalization). The theoretical work of Ohno helped revitalize the duplication and divergence field and subsequent work has found that gene preservation also occurs via a third process dubbed subfunctionalization. Subfunctionalization occurs when gene duplicates complement one another after accumulating degenerative mutations (FORCE *et al.* 1999; LYNCH and CONERY 2000; LYNCH and KATJU 2004; PRINCE and PICKETT 2002). Several additional models have also been proposed although the processes of neofunctionalization, nonfunctionalization, and subfunctionalization remain the best-recognized (INNAN and KONDRASHOV 2010).

The availability of whole genome sequences now provides the means to accurately evaluate the entire set of paralogs encoded in an organisms DNA. An estimated 30-60% of eukaryotic genes are the result of gene duplications indicating that much of our genome is recycled (BALL and CHERRY 2001). One explanation for the large number of paralogs is the

high rate of eukaryotic gene duplications. Eukaryotic gene duplications are estimated to occur at a frequency of 0.01 duplications per gene per million years (LYNCH and CONERY 2000), which is the same as the rate of mutation per nucleotide site (LI 1999). At this rate, even in the absence of whole genome duplication, an average of half of the genes in a genome would be duplicated every 35 to 350 million years (LYNCH and CONERY 2000). Gene duplication and divergence is, therefore, a potential source of significant contributions to evolutionary change. Characterizing gene duplicates at the molecular, functional, and phenotypic levels will provide insight into the extent gene duplications contribute to morphological and physiological change.

Mutations leading to subfunctionalization or neofunctionalization can occur in either coding or *cis*-regulatory regions. The end product of coding regions can be a protein while the final product of a non-coding RNA gene can be a variety of mature RNA molecules such as tRNA (transfer RNA), rRNA (ribosomal RNA), or miRNA (micro RNA). Finding and characterizing mutations within coding regions is, therefore, a relatively straightforward process given that molecular lesions affect both the DNA and the end product. *Cis*-regulatory regions are more complex, however, and can occur in both coding and non-coding regions. Furthermore, natural genetic variation or mutations affecting *cis*- regulatory regions typically affect the gene transcription level through a variety of molecular mechanisms (DAVIDSON 2006) making the study of *cis*-regulatory regions particularly challenging. Numerous examples of both coding and *cis*-regulatory region mutations have been characterized but the relative importance of either mutation class to gene duplicate subfunctionalization or neofunctionalization continues to be a subject of debate (BENJAMIN

PRUD'HOMME 2007; COYNE 2007; WRAY 2007). The systematic evaluation of gene duplicate *cis*-regulatory regions and coding regions in different species should allow for the testing and validation of the relative contribution of each mutant class to adaptive change.

Versatility and expansion of C2H2 zinc finger genes

Zinc finger proteins are so named because they both bind to zinc and "grasp" DNA. The discovery of the first zinc finger protein is intimately intertwined with the characterization of the first eukaryotic transcription factor. Researchers were intent on finding the factor responsible for the transcription of 5SRNA by RNA polymerase III in *Xenopus* oocytes. Transcription factor IIIA (TFIIIA) was identified as the responsible protein and found to bind a DNA sequence called the internal control region within 5SRNA genes (BROWN 1984). Subsequent work revealed that TFIIIA also binds to 5S RNA to form 7S ribonucleoprotein particles (RNP), which may, in turn, mediate autoregulation of 5S RNA transcription (PELHAM and BROWN 1980). A biochemical analysis of the 7S RNPs revealed high amounts of zinc that corresponded proportionally with nine 30-amino acid repeats derived from proteolytic digestion (MILLER *et al.* 1985). Further scrutiny of the amino acid sequence revealed a repeating motif consisting of two cysteines followed by a pair of histidines. This amino acid pattern came to be known as the C2H2 zinc finger motif (GINSBERG *et al.* 1984; MILLER *et al.* 1985).

The C2H2 domain structure was predicted to form an antiparallel β-sheet with the two cysteines and an α-helix with the two histidines (MILLER *et al.* 1985). In this model, the cysteines and histidines would bind to a centrally located zinc cation with DNA recognition postulated to reside in the helical region (Fig. 1). NMR studies validated the folding patterns

of zinc finger α-helix and anti-parallel β-sheet domains (LEE *et al.* 1989; NAKASEKO *et al.* 1992; NEUHAUS *et al.* 1992) while X-ray crystallography determined that three α-helix amino acids at positions -1, 3, and 6 are responsible for binding three DNA bases in the major groove (PAVLETICH and PABO 1991). A second crystal structure would reveal an additional contribution from position 2 of the α-helix to the opposite strand of DNA (FAIRALL *et al.* 1993). A nearly one-to-one correlation of specific amino acids to base pairs led to speculation that zinc fingers could be designed to target specific DNA sequences.

This observation has been borne out in a variety of studies utilizing engineered C2H2 zinc fingers. A pioneering paper showed that an engineered C2H2 zinc finger protein could bind to a 9-bp sequence *in vitro* or *in vivo* to suppress an oncogene (CHOO *et al.* 1994). When fused to an activation domain, the same protein could activate a reporter containing the same 9-bp sequence in cell culture (CHOO *et al.* 1994). Similar approaches have been used successfully in applications ranging from HIV (human immunodeficiency virus) and herpes simplex virus repression to VEGF (vascular endothelial growth factor) activation in cell lines and a mouse model (JAMIESON *et al.* 2003). By attaching a nuclease to targeted zinc fingers, investigators have also been able to facilitate gene conversion events and targeted gene knockouts in both human cell lines and animal models (BIBIKOVA *et al.* 2003; DOYON *et al.* 2008; MENG *et al.* 2008; PORTEUS and BALTIMORE 2003; SANTIAGO *et al.* 2008). Furthermore, the increasing knowledge of C2H2 zinc finger structures and defined binding sites continues to refine predictive C2H2 binding site algorithms (BENOS *et al.* 2002; KAPLAN *et al.* 2005; LIU *et al.* 2005; MAEDER *et al.* 2008; MANDEL-GUTFREUND and MARGALIT 1998; PERSIKOV *et al.* 2009; SUZUKI *et al.* 1995).

C2H2 zinc fingers have expanded throughout the eukaryotic lineage with increasing organismal complexity. The number of zinc finger genes increases from *Sachromyces cerevisiae* (34-48), to *C. elegans* (68-151) and *D. melanogaster* (234-357), to humans (564-706) (LANDER *et al.* 2001; VENTER *et al.* 2001). Furthermore, the number of zinc fingers per gene has similarly risen from a few zinc finger motifs per gene in yeast to multiple zinc finger motifs per gene in humans (yeast = 1.5, worms = 2.5, flies = 3.5, humans = 8) (EMERSON and THOMAS 2009; LOOMAN *et al.* 2002). The C2H2 zinc finger motif has therefore expanded in terms of the number of zinc fingers per gene and the number of genes containing zinc fingers to become the most abundant protein motif encoded in mammalian genomes.

The evolutionary expansion of C2H2 zinc finger genes may be linked to the inherent versatility of this unique motif. As mentioned previously, C2H2 zinc fingers have tremendous adaptive potential as they can bind DNA, RNA, or other proteins (MACKAY and CROSSLEY 1998). The modular nature of C2H2 zinc fingers has lead to the rise of poly-zinc finger proteins that bind both DNA and other proteins. Recent evidence suggests that the DNA-binding domains of poly-zinc finger genes may undergo higher rates of positive selection than non-DNA-binding zinc finger domains (EMERSON and THOMAS 2009). The ability to switch gene or genomic targets while retaining protein-protein interactions with other transcription factors or chromatin complexes provides the flexibility necessary to respond to environmental changes. Understanding how similar poly-zinc finger families regulate genes in different species may, in turn, help us quantify the extent to which poly-zinc fingers contribute to adaptive evolutionary change.

The deuterostome Ikaros family of C2H2 zinc finger transcription factors

Two independent research groups originally identified the Ikaros DNA-binding C2H2 transcription factor based on its ability to bind to immune-system specific enhancers. One team used an enhancer for the TdT (terminal deoxynucleotidyl transferase) gene (HAHM *et al.* 1994; LO *et al.* 1991), which is expressed in immature B and T cells (BOLLUM 1979), while the second team used the CD3δ (cluster of differentiation) chain enhancer, which is an early T-cell differentiation marker (FURLEY *et al.* 1986; HAYNES *et al.* 1989). Ikaros was then identified as a factor responsible for CD3δ enhancer activation (GEORGOPOULOS *et al.* 1992a; GEORGOPOULOS *et al.* 1992b). Subsequent studies have found that Ikaros regulates multiple immune-system loci; although not necessarily directly as only a few direct targets are known (NG *et al.* 2009).

A hallmark of Ikaros family members is the distinct C2H2 zinc finger domain spacing. An N-terminal DNA-binding domain, consisting of four zinc fingers clustered together, is easily distinguishable from the separate C-terminal protein-protein interaction domain consisting of two zinc fingers (Fig. 1C) (MCCARTY *et al.* 2003). The four N-terminal zinc fingers bind DNA while the two C-terminal zinc fingers serve as vehicles for homodimeration, heterodimerization, and multimerization with other members of the Ikaros family (HAHM *et al.* 1994; MOLNAR and GEORGOPOULOS 1994). The two middle fingers of the DNA binding domain recognize the sequence 5'-TGGGAA-3' and binding affinity to this motif is facilitated by the inclusion of one or both of the flanking zinc fingers (COBB *et al.* 2000; MOLNAR and GEORGOPOULOS 1994). Ikaros family member isoforms utilize variable numbers of N-terminal zinc fingers thereby providing less or more specificity and affinity for

the canonical motif which may, in turn, allow for a potentially wider range of DNA targets (MOLNAR and GEORGOPOULOS 1994). Furthermore, the N-terminal DNA binding domain contains a nuclear localization sequence (COBB *et al.* 2000; MOLNAR and GEORGOPOULOS 1994). All natural Ikaros isoforms contain the C-terminal multimerization domain, which is unable to bind DNA (MOLNAR and GEORGOPOULOS 1994). Mutant mice lacking only the DNA binding N-terminal zinc fingers generate dominant negative proteins that can form multimers without binding DNA (GEORGOPOULOS *et al.* 1994). Some naturally occurring Ikaros isoforms also lack a functional N-terminal DNA binding domain, which points to a built-in mechanism to generate dominant negative isoforms (SUN *et al.* 1996). Therefore, the presence or absence of DNA binding zinc fingers serves as a mechanism to increase DNA binding specificity and affinity or to inhibit DNA binding altogether.

In addition to regulating DNA binding, both zinc finger regions and non-zinc finger regions physically interact with chromatin remodeling complexes (CRCs), such as NuRD (nucleosome remodeling and deacetylating) and SWI/SNF (KIM *et al.* 1999; KOIPALLY *et al.* 1999a; SRIDHARAN and SMALE 2007). SWI/SNF and NuRD represent two well-characterized classes of CRCs, with the latter containing histone deacetylase (HDAC) activity. SWI/SNF was the first CRC to be discovered and characterized in yeast (CAIRNS *et al.* 1994; KRUGER *et al.* 1995; NEIGEBORN and CARLSON 1984; STERN *et al.* 1984) while NuRD was subsequently found in vertebrates (TONG *et al.* 1998; WADE *et al.* 1998; XUE *et al.* 1998; ZHANG *et al.* 1998). While SWI/SNF has been identified in yeast, plants, and animals, NuRD has not been clearly identified in yeast. Both CRCs utilize ATP-dependent mechanisms to catalyze nucleosome dynamics, such as the replacement of histone variants or

the shifting of nucleosomes into looser or more compact chromatin states (SAHA *et al.* 2006). In this way, CRCs are able to affect gene transcription by making cis-regulatory regions more or less accessible. NuRD has the additional capacity to deacetylate histones, which is predominantly associated with chromatin compaction and subsequent transcriptional repression (STRUHL 1998). A biochemical extraction of Ikaros complexes from the thymus of Ikaros transgenic mouse lines has revealed an association of Ikaros with large amounts of NuRD and smaller amounts of SWI/SNF (KIM *et al.* 1999). Both NuRD and SWI/SNF were also found in conjunction with Ikaros in erythroleukemia cell lines implying a large complex containing both CRCs or multiple complexes (O'NEILL *et al.* 2000). Independent studies have also found the simultaneous presence of the respective NuRD and SWI/SNF ATPases, BRG1 and Mi-2, in large protein complexes indicating common regulatory functions or a unique complex containing both CRCs(NAKAMURA *et al.* 2002; SHIMONO *et al.* 2003). In addition to NuRD and SWI/SNF, Ikaros has also been shown to physically associate with Sin3, which is a core component of another HDAC containing complex distinct from NuRD (KOIPALLY *et al.* 1999b). Therefore, Ikaros may utilize multiple mechanisms to attract HDACs and, in turn affect the acetylation state of histones and subsequent active or repressed chromatin states. Ikaros has also been found to physically associate with both CtBP (C-terminal binding protein) and CtIP (CtBP-interacting protein) (KOIPALLY and GEORGOPOULOS 2000). CtBP and CtIP are both repressors that are thought to inhibit transcription by interacting with and inhibiting basal transcription factors. Thus, although Ikaros is known to predominantly associate with NuRD and SWI/SNF, it has the potential to

affect gene expression by chromatin remodeling, alteration of histone acetylation states, or by directly inhibiting basal transcription factors.

Ikaros family members are expressed in tissues within and outside the developing immune system. Of the five known Ikaros family members Ikaros, Helios, and Aiolos are expressed primarily in immune system lineages while Eos and Pegasus are broadly expressed both within and outside the immune system (PERDOMO *et al.* 2000). During embryogenesis Ikaros is expressed in the aorta gonad mesonephros (AGM), which gives rise to the hematopoietic stem cells (HSCs) and the gonad in addition to several other tissues (GEORGOPOULOS *et al.* 1992a). Ikaros is expressed in other sites important for immune system development such as the fetal liver and thymus, blood islands of the yolk sac, and both proliferating and mature populations of T and B cell precursors in the fetus and adult. The adult HSCs, which are both pluripotent and self-renewing, and myelo-erythroid precursors also express Ikaros (KELLEY *et al.* 1998; KLUG *et al.* 1998; MORGAN *et al.* 1997). Downregulation of Ikaros expression in these lineages coincides with their differentiation into macrophages, monocytes, and erythrocytes with the exception of granulocytes where Ikaros expression is retained and T cells where Ikaros expression peaks. Aiolos is expressed in the common lymphoid progenitors (CLPs), originally thought to give rise to all B and T cells, with expression subsequently increasing in mature B and T cells. The highest expression of Aiolos is observed in B cells, which are predominantly affected by Aiolos defects (AKASHI *et al.* 2000; MORGAN *et al.* 1997). Helios expression in the immune system is similar to Ikaros but at reduced levels with additional expression observed in ectoderm derivatives (HAHM *et al.* 1994; KELLEY *et al.* 1998). Eos and Pegasus expression is,

however, found predominantly outside the immune system (HONMA *et al.* 1999; PERDOMO *et al.* 2000). The expression of Ikaros also extends outside the immune system into neuronal, neuroendocrine, and pituitary cell populations (AGOSTON *et al.* 2007; EZZAT *et al.* 2006; MOLNAR and GEORGOPOULOS 1994). Ikaros family member expression points to both unique and redundant functions in a variety of tissues.

Subsequent work has found major roles for Ikaros in lymphocyte specification, homeostastis, and lineage progression (SMALE and DORSHKIND 2006). Dominant negative mutants of Ikaros have also been generated in which all but the first DNA binding zinc finger are deleted resulting in a multimerization domain and an intact transcriptional activation domain (GEORGOPOULOS *et al.* 1994). Mice homozygous for dominant negative mutations result in the complete elimination of HSC activity including the postnatal production of αβ-T cells. αβ-T cells are the most common T cells in the blood and express a heterodimer of two proteins (α and β) derived from somatic recombination. The αβ heterodimer, in turn, serves as a TCR (T Cell Receptor) that recognizes self versus non-self antigens presented on MHC (major histocompatibility complex) molecules displayed on all cells. Null mutants of Ikaros cause a substantial decrease in HSC activity in fetal and adult mice (NICHOGIANNOPOULOU *et al.* 1999). HSCs that are produced do not make natural killer cells, B cells, fetal T cells or CLP precursors (WANG *et al.* 1996). Ikaros null mutants do, however, result in the postnatal production of αβ-T cells pointing to an alternate pathway that is 1) independent of CLPs and 2) less affected by Ikaros null mutations. The Ikaros null mutant phenotype helped overturn the notion that CLPs were the only cell that could give rise to B and T cells. The more

severe dominant dominant negative Ikaros phenotype also implied that Ikaros was physically interacting with an unidentified factor controlling the non-canonical production of αβ-T cells; perhaps even another Ikaros family member (WANG *et al.* 1996). In contrast to lymphoid derivatives, null Ikaros mutant HSCs generate a normal amount of myeloid precursors and subsequent myelocytes despite a decrease in both number and activity of HSCs (NICHOGIANNOPOULOU *et al.* 1999). This points to a negative role for Ikaros in the HSC myeloid pathway and a positive role for Ikaros in the HSC lymphoid pathway. Mice heterozygous for dominant negative mutations and null mutations develop T and B cells, but they have a tendency to develop aggressive T cell malignancies. Therefore, Ikaros plays roles in the differentiation and proliferation of HSCs in addition to opposing effects on HSC derived myeloid and lymphoid populations.

Ikaros has also been found to regulate cell cycle progression in T cells. *In vitro* TCR stimulation of mature CD4 and CD8 αβ-T cells from null Ikaros mutants results in hyperproliferation compared to wild-type controls (AVITAHL *et al.* 1999). αβ-T cells heterozygous for dominant negative mutants are more responsive to TCR stimulation than null Ikaros heterozygote αβ-T cells. This indicates the phenotype is dependent upon decreasing Ikaros levels because dominant negative mutant heterozygotes result in less Ikaros activity than heterozygote null Ikaros mutants. The root cause of αβ-T cell hyperproliferation can be traced to the early initiation of DNA replication corresponding with early entry into S phase. Nuclear staining reveals Ikaros and the NuRD ATPase (Mi-2β) dispersed widely in resting T cells but colocalized into distinct clusters upon activation

(AVITAHL *et al.* 1999; BROWN *et al.* 1997; KIM *et al.* 1999). The clusters correspond with heterochromatin foci associated with DNA replication during the G1/S phase of the cell cycle. The subsequent removal of Ikaros results in an inability of Mi-2β to form foci. This suggests that Ikaros is responsible for the targeting of NuRD to heterochromatin foci for proper DNA replication during T cell proliferation.

Perhaps not surprisingly Ikaros mutant mice develop aggressive T cell leukemias (WINANDY *et al.* 1995). This is likely due to the accumulation of chromosomal abnormalities in rapidly dividing T cells, which is observed in primary cultures (AVITAHL *et al.* 1999). Acute leukemia and resultant tumors are found in homozygous mutants or heterozygous dominant negative mutants but less so in null heterozygotes (WANG *et al.* 1996; WINANDY *et al.* 1995). Recombinase activating gene (RAG) homozygous mutants cannot form TCRs and double mutants of RAG(-/-) and Ikaros dominant negative heterozygotes eliminates aberrant T cell proliferation (WINANDY *et al.* 1999). Dominant negative heterozygote Ikaros mice develop leukemenia and lymphoma with 100% penetrance by three months of age, which was absent in a RAG(-/-) background by up to eleven months of age. Transgenic expression of the TCR, in turn, restored the malignancies. This implies that TCR signaling is necessary for oncogenic transformation of Ikaros deficient T cells. In contrast, the first study to examine the loss of Mi-2β in mammals focused on T cells and observed defects in T cell proliferation. A significant number of T cells were both unable to respond to TCR stimulation *in vitro* and activated T cells had cell cycle defects (WILLIAMS *et al.* 2004). The results imply that Ikaros has negative roles in TCR-mediated proliferation while Mi-2β has positive roles in TCR-mediated proliferation.

Subsequent studies have verified an antagonistic relationship between Mi-2β and Ikaros specifically at the CD4 locus during T cell lineage progression (NAITO *et al.* 2007). Early and late T cell development can be characterized by the expression of CD8 and CD4 coreceptors. In early T cell development, cells do not express either CD4 or CD8 coreceptors and are called double negative (DN). T cells progress to a later stage of development in which they express both CD4 and CD8 coreceptors and are called double positive (DP). Ikaros was found associated with a well-defined silencer of CD4 during both the DN and DP stages of T cell lineage progression. The activity of Mi-2β was found to be responsible for overriding the silencing of the CD4 locus during the DN-DP transition (NAITO *et al.* 2007). SWI/SNF also represses CD4 and binds the CD4 silencer at both the DN and DP stages of development (CHI *et al.* 2002). One attractive model is that Ikaros, Helios, and/or Aiolos bind the CD4 silencer and serve as intermediates for both positive and negative regulators of chromatin dynamics. The DN stage of T cell development may be characterized by the presence of SWI/SNF and HDACs at the CD4 silencer. The compact chromatin state of the CD4 locus, in turn, may be overridden by the recruitment of NuRD and HATs to the CD4 silencer during the DN-DP transition. This provides a useful model for the opposing effects of SWI/SNF and NuRD mediated by Ikaros family members at other loci.

Recent research suggests additional roles for mammalian Ikaros in lineage progression outside the immune system (ELLIOTT *et al.* 2008). Mouse embryonic retinas generate multiple cell types from multipotent retinal progenitor cells (RPCs) during early (i.e. cone photoreceptors, retinal ganglion cells, horizontal cells), middle (i.e. amacrine cells), and

late (i.e. Muller glial cells, rod photoreceptors, bipolar cells) embryonic retinal development (LIVESEY and CEPKO 2001). The order and fate of cell birth is reproducible making this an excellent mammalian system for studying temporal cell lineage progression.

The embryonic retinas of mice were examined for Ikaros family member expression. Ikaros is expressed in early multipotent retinal progenitor cells (RPCs) while Eos and Pegasus are expressed throughout retinal development, Helios was expressed later in embryonic development, and Aiolos was not detected (ELLIOTT *et al.* 2008). In addition, both homozygous null mutants of Ikaros and overexpression of dominant negative Ikaros in RPCs were found to specifically decrease the number of early born retinal cells without affecting later born retinal cells (ELLIOTT *et al.* 2008). Surprisingly, when wild-type Ikaros was expressed in later cell types, early-born cells were produced indicating a resetting of the temporal clock. Therefore, it is now thought that Ikaros acts as a temporal competence factor in mouse RPCs similar to previous research with the related transcription factor *hunchback* in *Drosophila* neuroblasts (PEARSON and DOE 2003).

In addition to roles outside the immune system, the mechanisms of Ikaros gene regulation have been refined over time. As previously mentioned, Ikaros physically associates with NuRD, which has activation potential via chromatin remodeling activity and repressive potential via HDACs. It has subsequently been proposed and observed that these antagonistic activities create a "bivalent" state of activation and repression at Ikaros-associated immune system loci (NG *et al.* 2009).

The bivalent state was originally discovered at embryonic stem cell promoters and is thought to be a mechanism that leaves cells "poised" for activation or repression (BERNSTEIN

et al. 2006). Methylation of Histone H3 lysine 4 (H3K4) is associated with trithorax group proteins, a relaxed chromatin state, and transcriptional activation whereas lysine 27 (H3K27) is connected with polycomb group proteins, a compact chromatin state, and gene repression. ~200 highly conserved noncoding elements (HCNEs) are found in mammalian genomes and are known to harbor a high frequency of transcription factors important to embryogenesis. A survey of HCNEs in embryonic stem (ES) cells revealed large regions of H3K27 and smaller regions of H3K4 at developmentally important genes pointing to a "bivalent" state of activation and inactivation. These bivalent states were subsequently resolved upon cellular differentiation (BERNSTEIN *et al.* 2006). This provides a mechanism to potentially poise genes in HSCs for activation or repression and subsequent immune system lineage progression.

The striking phenotypes of Ikaros mutants in the mammalian adaptive immune system have led to a search for Ikaros-like genes in the adaptive immune systems of other species. Ikaros family members have subsequently been discovered in the adaptive immune systems of a variety of animals (JOHN *et al.* 2009). Additional Ikaros orthologs have been identified in mammalian lymphocytes, zebrafish lymphocytes, and even in the most primitive vertebrate immune systems such as the lymphoid-like cells of hagfish and lampreys (ROTHENBERG and PANT 2004). Ikaros family members are, therefore, consistently found in the lymphoid tissues of animals possessing an adaptive immune system. It has been further proposed that the duplication and divergence of Ikaros family members coincides with adaptive immune system evolution.

Additional searches have found Ikaros family-like (IFL) proteins in non-vertebrate chordates, such as the asicidian *Oikopleura dioica* and *Ciona intestinali*s, both of which lack an adaptive immune system. The expression of basal deuterostome IFL genes points to additional roles outside the immune system. For example, the lancelet IFL1 gene was found to be expressed in the ovary and gills while the urochordate equivalent was found in the oocyte and filter house (CUPIT *et al.* 2003; HUANG *et al.* 2007). Lamprey IFL genes were highly expressed in the ovary and intestinal epithelium (MAYER *et al.* 2002). The possibility arises that at least one of the functions of the ancestral gene leading to the Ikaros family of transcription factors is related to oocyte expression. These results indicate a possible additional role for IFLs in the reproductive system and reinforce the notion that IFLs have roles outside the immune system.

A model for the evolutionary expansion of IFL genes has recently been proposed for deuterostomes (Fig. 2) (JOHN *et al.* 2009). Within the vertebrate Ikaros gene family, the gene pairs Ikaros and Aiolos and the gene pairs Helios and Eos are thought to be paralogs derived from a whole genome duplication event in fish following the divergence of the agnanthans. The pre-agnanthan pair of ancestral genes (IKFL1 and IKFL2) are, in turn, thought to be paralogs derived from a whole genome duplication event following the divergence of chordates from cephalochordates and urochordates. This leaves an ancestral gene leading to Pegasus (IFL2) and an ancestral gene leading to the other four family members (IFL1) in cephalochordates and urochordates, which lack adaptive immune systems. Despite the ancient conservation of this important group of proteins, a universal protostome and deuterostome ancestor to IFL1/2 awaits to be discovered. Ikaros family members have not

been found in either *D. melanogaster* or *C. elegans* but the unique arrangement of Ikaros zinc finger domains is similar to that seen with *Drosophila hunchback (hb)* and has led to the proposal that *hunchback* and IFLs are derived from a common protostome/deuterostome ancestor (GEORGOPOULOS *et al.* 1992a; JOHN *et al.* 2009).

The protostome *hunchback* family of C2H2 zinc finger transcription factors

hunchback was first characterized as a gap gene in early *Drosophila* embryogenesis (RIVERA-POMAR and JACKLE 1996; ST JOHNSTON and NUSSLEIN-VOLHARD 1992). The role of *hunchback* in anterior-posterior patterning built upon earlier discoveries of *nanos* and *bicoid* as maternal effect genes. During *Drosophila* oogenesis all three mRNAs, *hunchback*, *nanos*, and *bicoid*, are supplied to the oocyte (TAUTZ 1988). At the time of fertilization *hunchback* mRNA is diffused equally while *bicoid* and *nanos* mRNAs are concentrated at the anterior and posterior ends of the egg, respectively. Bicoid is a homeodomain transcription factor that activates zygotic *hunchback* expression in the anterior of the early embryo while Nanos is a conserved RNA-binding protein that inhibits translation of maternal *hunchback* and *bicoid* mRNAs in the posterior of the embryo. The translational repression of *hunchback* by Nanos is achieved in concert with Pumilio, which is also a conserved RNA-binding protein. Pumilio binds directly to *nanos* response elements (NRE) in the *hunchback* 3'UTR and recruits Nanos (MURATA and WHARTON 1995; SONODA and WHARTON 1999; WHARTON and STRUHL 1991). *hunchback* translation is restricted to the anterior of the embryo. Hunchback is, in turn, responsible for the repression of posterior patterning genes such as *giant* and *knirps* and for the activation of genes involved in head development. Hunchback and other gap genes also directly bind to and repress HOX genes (MULLER and BIENZ 1992;

QIAN *et al.* 1991; SHIMELL *et al.* 1994; ZHANG and BIENZ 1992). Hunchback is thought to mediate HOX gene repression and subsequent Polycomb-group (PcG) recruitment via physical interactions with the Mi-2 ATPase component of the NuRD CRC (KEHLE *et al.* 1998). Whether or not the ancestral role of *hunchback* lies in anterior-posterior patterning, HOX gene repression, or some other process continues to be a subject of inquiry.

New insights into *hunchback* have been provided by studies focused on *Drosophila* embryonic neuroblast lineage progression. The asymmetric divisions of *Drosophila* neuroblast 7-1 (NB7-1) are well characterized and have led to the identification of *hunchback* as an early regulator of temporal identity. NB7-1 progenitors give rise to ganglion mother cells (GMC-1, GMC-2, GMC-3, etc.), which subsequently divide to create two post-mitotic neurons named according to birth order (e.g. U1, U2, U3, etc.). Each embryonic cellular division of NB7-1 is characterized by the sequential expression of a specific transcription factor in the following order at approximately forty minute intervals: Hunchback (Hb), Kruppel (Kr), Pdm1/Pdm2 (Pdm), and Castor (Cas) (GROSSKORTENHAUS *et al.* 2005; ISSHIKI *et al.* 2001; PEARSON and DOE 2003). This expression pattern, in turn, dictates whether an early-born or late-born post-mitotic neuron will be generated in the following sequence: high Hb= U1, low Hb= U2, Kr=U3, Pdm=U4, Pdm/Cas=U5 (CLEARY and DOE 2006). If Hb or Kr are aberrantly expressed during the competence window of the first five divisions U1-U3 neurons can be produced (CLEARY and DOE 2006; PEARSON and DOE 2003). The NB7-1 competence window is limited to five divisions (PEARSON and DOE 2003) but can be maintained if Hb, and only Hb, expression is maintained (GROSSKORTENHAUS *et al.* 2005). If Hb is mutated the early born Hb-dependent cells die or undergo cell fate transformation

without affecting late-born neurons (ISSHIKI *et al.* 2001). This suggests similar roles for *Drosophila hunchback* and mammalian Ikaros as temporal competence factors in neuroblast lineage progression.

In contrast to *Drosophila hunchback*, the *C. elegans hunchback* orthologue *hbl-1* is not involved in anterior-posterior patterning but has been found to play a role in epithelial cell differentiation. *C. elegans hbl-1* is expressed in early embryonic ephithelial cells, the embryonic pharynx and in neurons from embryogenesis to young adults (FAY *et al.* 1999). In addition, *hbl-1*(RNAi) causes post morphogenesis embryonic lethality, and other defects that are not suggestive of a maternal role for *hbl-1* (FAY *et al.* 1999). The recruitment of *nanos* and *pumilio* family proteins to the NRE in the *hunchback* 3'UTR is conserved in *C. elegans* but takes place in the hypodermis and ventral nerve cord (NOLDE *et al.* 2007). Therefore *C. elegans hunchb*ack is neither maternally expressed nor controls axial patterning but is, instead, primarily required for early morphogenesis of *C. elegans* embryos and epithelial cell differentiation.

Recent work has found that *C. elegans hunchback* controls the temporal lineage progression of hypodermal cells known as seam cells. Genes affecting the timing of the seam cell divisions are known as "heterochronic" genes and mutations in these genes can cause "precocious" phenotypes in which the cells prematurely differentiate or "retarded" phenotypes in which the terminal differentiation is delayed. *hbl-1* has also been identified as a heterochronic gene that, when mutated, causes a precocious phenotype leading to the premature differentiation of seam cells in the L4 stages (ABRAHANTE *et al.* 2003; LIN *et al.* 2003). Overexpression of HBL-1 causes a retarded phenotype in seam cells at the L4 stage

but not at the L1 or adult stage, possibly pointing to a competence window for *C. elegans hbl-1* in seam cell divisions. This work reinforces the importance of *hunchback* as a factor affecting temporal competence and emphasizes the possible shared ties between temporal regulation in the *C. elegans* seam cells, *Drosophila* neuroblasts, and mammalian retina.

A search for the ancestral role of *hunchback* in animal development

The role of *Drosophila hunchback* in lineage progression and a similar role for mammalian Ikaros in lineage progression provide potential clues to the ancestral functions of *hunchback*- and Ikaros-like genes. This makes the presence or absence of *hunchback* or *hunchback*-related genes in animal genomes important. For example, if lineage progression were the ancestral role it would be interesting to find a *hunchback*-like gene in an animal lacking tissues or organs.

Animal organogenesis is unique to bilaterally symmetrical animals (i.e. bilaterians) and dependent upon three germ layers. The sophisticated adult body plan of bilaterians has its origins in embryogenesis. Early anterior-posterior (A-P) and dorsal-ventral (D-V) patterning results in an organism with two axes: 1) a front and back and 2) a top and bottom. A-P and D-V patterning provide the regional specification necessary to define three embryonic germ layers during early development: ectoderm, endoderm, and mesoderm. A process known as gastrulation leads to the physical separation of the three germ layers. An opening in the early developing embryo, known as the blastopore, leads to the migration of groups of cells into the interior of the developing embryo. The cells on the exterior become the ectoderm and give rise to the several tissues such as the nervous system and epidermis. Some interior cells become the endoderm and give rise to a variety of structures such as the

gut and the linings of many internal organs. The mesoderm is also derived from the inner cells and is crucial to organogenesis as it provides a barrier between the endoderm and the ectoderm. This allows for organs to develop independently from the entire organism and other organs. In addition to a buffer zone, the mesoderm generates various connective tissues, muscle, bone, and blood. The spatial organization and patterning of the three germ layers, in particular the mesoderm, is therefore essential to organogenesis.

In addition to organization of the developing embryo, gastrulation also provides a useful feature for bilaterian classification. There are three major bilaterally symmetrical superclades: ecdysozoans, lophotrochozoans, and deuterostomes. Deuterostomes include chordates, hemichordates, and echinoderms and are distinct from lophotrochozoans and ecdysozoans, collectively known as protostomes. The ecdysozoan clade contains molting animals with a hard cuticle such as arthropods and nematodes. Lophotrochozoans, on the other hand, are animals with a lophophore (i.e. ciliated adult feeding organ) or a distinct larval stage (i.e. trochophore), including annelids and mollusks. Molecular evidence provides further support for a distinction between ecdysozoans and lophotrochozoans while placing both within the protostome clade (AGUINALDO *et al.* 1997; HALANYCH *et al.* 1995). One of the primary differences between deuterostomes and protostomes is that the initial opening during gastrulation becomes the anus in the deuterostomes and the mouth in protostomes. The classification of ecdysozoans, lophotrochozoans, and deuterostomes has been further substantiated and refined using 18S rDNA sequences, nuclear and mitochondrial markers, and morphological-molecular classification (ADOUTTE *et al.* 2000; HALANYCH 2004; PETERSON and EERNISSE 2001). Classification of the three bilaterian superclades

continues to be a rich source of debate with the broad biological importance of molecules discovered in *C. elegans*, *Drosophila*, and other model organisms at stake.

Less complex animals lack bilateral symmetry and gastrulation. In contrast to bilaterians, animals with radial symmetry (i.e. radiata) do not have anterior-posterior patterning but do establish a dorsal-ventral axis. Likewise, while bilaterians possess three germ layers (i.e. triploblastic) radiata lack a definitive mesoderm but retain both an ectodermal and endodermal germ layer (i.e. diploblastic). Radiata are, therefore, able to generate tissues due to the presence of an endodermal germ layer but cannot form true organs due to the lack of a mesodermal germ layer. For example, cnidarians are diploblastic, possess radial symmetry and have true tissues but are not bilaterally symmetrical and do not have organs. Even less complex animals have no symmetry and also lack both tissues and organs. This group includes sponges, which contain only one germ layer but do possess ten to twenty cell types. The sole representative of the Placozoan phylum, *Trichoplax adhaerens*, is one of the simplest animals and contains only four or five cell types with recent phylogenetic analyses placing it close to sponges (SRIVASTAVA *et al.* 2008). The unicellular and colony-forming choanoflagellate *Monosiga brevicollis* serves as a useful outgroup for animals at both the molecular and structural level with only one cell type. The transition from unicellular marine eukaryote to animals with one or more germ layers and radial or bilateral symmetry likely involved many evolutionary innovations. Understanding the molecular underpinnings of early A-P/D-V patterning and mesodermal formation may provide some insight into the morphological leap from tissues to organs.

Finding and characterizing genes unique to bilaterians can help identify the molecular mechanisms underlying organogenesis. The recently sequenced genomes of several animals provide insight into the breadth and depth of molecular toolkits available to diverse phyla. It was surprising to discover that cnidarians possess the molecular toolkit to produce both A-P and D-V patterning in addition to all three germ layers. This suggests that the common ancestor of currently recognized bilaterians and cnidarians was triploblastic and bilaterian with extant cnidarians possibly having many derived features (BAGUNA *et al.* 2008). It has also been proposed that cnidarians are using a toolkit that was found in the common ancestor to generate structures that are not necessarily homologous to those found in true bilaterians (BALL 2007). The discovery of genes typically thought to control axial patterning and mesoderm formation in much simpler animals indicates these genes may have originally been used to control cell specification. One possibility is that the evolution of organismal complexity and organogenesis may have more to do with how genes are used (i.e. gene regulation) than the presence or absence of genes in a genome. Finding and characterizing genes and/or novel mechanisms of gene regulation unique to bilaterians is consequently a valuable and rewarding area of research.

The ancestral role of *hunchback* in animal development and ancestral gene structure may be inferred by the first presence of *hunchback* or *hunchback*-like genes in the animal kindgdom. Current evidence suggests that *hunchback* is unique to protostomes (KERNER *et al.* 2006). Developmental and molecular models have been established for *hunchback* in *D. melanogaster* and *C. elegans*. The lophotrochozoan clade does not, however, contain a genetically tractable model organism equivalent to that seen in ecdysozoans and there are

fewer sequenced lophotrochozoan genomes. This has hampered the functional characterization of *hunchback* orthologs in diverse species but allowed for a broader search in more genomes. The evaluation of *hunchback* orthologs in several protostomes has found the inconsistent presence of different C2H2 zinc finger domains in different species. Two C-terminal domain zinc fingers and four DNA-binding zinc fingers are consistently found in *hunchback* orthologs but one or two extra N-terminal zinc fingers are occasionally observed (Fig. 3). An extra zinc finger is also sometimes located between the DNA-binding domain and the C-terminal domain. Therefore, the variable C2H2 zinc finger domains of *hunchback* orthologs point to an ancestral protein structure consisting of two N-terminal zinc fingers, four middle zinc fingers, an extra finger, and two C-terminal fingers (PATEL *et al.* 2001). The expression patterns of various protostome *hunchback* orthologs and gene knockdown studies have provided a wide variety of potential ancestral *hunchback* functions. *hunchback* has been implicated in processes as diverse as mesoderm and neuroblast development with possible ancillary roles in oocytes, embryonic epithelia, and the nervous system (JAMIE PINNELL 2006; KERNER *et al.* 2006; PATEL *et al.* 2001). The continued identification and characterization of protostome *hunchback* and *hunchback*-like genes should provide additional clues to the evolutionary roles of *hunchback* and, in turn, may reveal unknown functions of Ikaros in mammals.

The *C. elegans* reproductive system as a model for lineage progression

A fully developed organ relies upon the specification and differentiation of multiple cell types. Each differentiated cell type in the adult organ has a precursor cell which, in turn, has its own precursor cell. This sequence of precursor cells extends back in time to the

fertilized egg and is known as the cell lineage. Understanding the progression of cell fates in time and space leading to the adult differentiated cell type can shed light on processes as seemingly diverse as organogensis, cancer, and tissue engineering.

C. elegans is a useful model organism for understanding lineage progression due to both its transparency and a complete map of the somatic cell lineage (SULSTON *et al.* 1983). This provides the opportunity to determine developmental defects in time and space with high precision. A variety of molecular and genetic tools, a sequenced genome, and a rapid three-day life cycle further facilitate a continued understanding of organogenesis in real time.

The somatic reproductive system develops from only two blast cells known as Z1 and Z4. Together, the two cells are referred to as somatic gonadal precursors (SGPs). The SGPs are specified ~260 minutes post-fertilization from mesodermal precursors and migrate to meet the primordial germ cells (PGCs) (Fig. 4) (SULSTON *et al.* 1983). The four cells come together to form the gonad primordium with two PGCs found in the center flanked by an SGP on either side. This results in a two-fold rotational symmetry whose function is not clear but is consistently found in all nematode species (CHITWOOD and CHITWOOD 1950). Each SGP of a *C. elegans* hermaphrodite is able to generate 72 cells leading to the production of the entire 143 cells of the somatic reproductive system from two multipotent SGPs (KIMBLE and HIRSH 1979). Only 143 cells are formed because, in one of the few examples of *C. elegans* intercellular signaling, the decision between an anchor cell (AC) and a ventral uterine (VU) precursor cell is determined by lateral inhibition using the Delta-Notch pathway (GREENWALD *et al.* 1983; KIMBLE 1981; SEYDOUX and GREENWALD 1989). The AC then fuses with several other cells to form the uterine seam cell leaving 143 cells instead

of 144 cells (NEWMAN and STERNBERG 1996). Two hermaphrodite distal tip cells (DTCs) are also generated by the SGPs. The DTCs promote germline proliferation and also serve as leader cells that guide the gonad arm through the worm body (KIMBLE and WHITE 1981). This creates two symmetrical U-shaped gonad arms emanating from the center of the rod-shaped worm (KIMBLE and HIRSH 1979). Sheath/spermathecae (SS) cells are also generated from the SGPs and promote male germline development, germline proliferation, germline maturation, and several important reproductive system structural roles (MCCARTER *et al.* 1997; ROSE *et al.* 1997). The remaining dorsal uterine (DU) and VU cells form the uterus, which serve as a holding area for embryos prior to egg-laying. The two SGPs can be thought of as mesodermally-derived multipotetent precursors with the capacity to produce 72 cells consisting of five cell types: DTC, SS, DU, VU, and AC (Fig. 4).

Research on the SGP blast cells has primarily focused on the first division although recent progress has been made in finding genes controlling early SGP development. Several genes have been characterized for the onset and timing of the first *C. elegans* SGP division (FRIEDMAN *et al.* 2000; SUN and LAMBIE 1997) and the first asymmetric SGP division (BRYAN T. PHILLIPS 2007; CHANG *et al.* 2004; KIDD *et al.* 2005; MASAKO ASAHINA 2006; MATHIES *et al.* 2004; MISKOWSKI *et al.* 2001; SAWA *et al.* 1996; SIEGFRIED *et al.* 2004; SIEGFRIED and KIMBLE 2002; STERNBERG and HORVITZ 1988; TILMANN and KIMBLE 2005). On the other hand, one gene has been characterized for the earliest stage of SGP development and a few additional genes have been elucidated for later SGP development. The conserved *hnd-1*/dHand transcription factor has been found to control early SGP survival while the *tra-1*/GLI ortholog, also a conserved transcription factor, controls late SGP polarity and division

(MATHIES *et al.* 2003; MATHIES *et al.* 2004). A C2H2 zinc finger gene with no known ortholog, *ehn-3*, was found to be expressed solely in the SGPs. Both weak and strong alleles of *ehn-3* interact genetically with both *hnd-1* during early SGP development and with *tra-1* during late SGP development to cause an almost complete loss of gonad arms (MATHIES *et al.* 2003; MATHIES *et al.* 2004). The discovery of a C2H2 gene only expressed in the SGPs that interacts with two highly conserved transcription factors controlling early SGP development served as the starting point for my graduate work.

The protostome HIL family of C2H2 zinc finger transcription factors

My research objective was to functionally characterize a distinct family of *C. elegans* genes similar in structure to both *hunchback* and Ikaros (LARGE and MATHIES 2010). Chapter II focuses on the characterization of a *C. elegans* mult-gene family, the *hunchback* and Ikaros-like (HIL) gene family, of which *ehn-3* is the founding member. We provide evidence that the family has structural and functional similarities to *hunchback* and Ikaros. The overlapping expression and phenotypic patterns of *C. elegans* HIL family members was characterized. Chimeric rescue experiments were used to support the idea that the HIL family has similar functions to Ikaros. HIL genes have not been previously described but a subsequent search of sequenced genomes reveals HIL genes in most protostome genomes.

In Chapter III, I use an RNAi screen to identify genetic enhancers of *ehn-3*. The enhancers include components of chromatin remodeling complexes such as SWI/SNF and NuRD, with which mammalian Ikaros physically interacts. Several alleles of SWI/SNF components are examined in addition to putative NuRD subunits. The results point to opposing actions of SWI/SNF and NuRD in somatic reproductive system development.

In Chapter IV, I provide a summary of the two previous chapters and provide ideas for future work. Possible common roles are posited for HIL, *hunchback*, and Ikaros family members both within and outside lineage progression. The potential for physical interactions between HIL family members and chromatin factors is also explored.

REFERENCES

ABRAHANTE, J. E., A. L. DAUL, M. LI, M. L. VOLK, J. M. TENNESSEN *et al.*, 2003 The *Caenorhabditis elegans hunchback*-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. Dev Cell **4:** 625-637.

ADOUTTE, A., G. BALAVOINE, N. LARTILLOT, O. LESPINET, B. PRUD'HOMME *et al.*, 2000 The new animal phylogeny: reliability and implications. Proc Natl Acad Sci U S A **97:** 4453-4456.

AGOSTON, D. V., M. SZEMES, A. DOBI, M. PALKOVITS, K. GEORGOPOULOS *et al.*, 2007 Ikaros is expressed in developing striatal neurons and involved in enkephalinergic differentiation. J Neurochem **102:** 1805-1816.

AGUINALDO, A. M., J. M. TURBEVILLE, L. S. LINFORD, M. C. RIVERA, J. R. GAREY *et al.*, 1997 Evidence for a clade of nematodes, arthropods and other moulting animals. Nature **387:** 489-493.

AKASHI, K., L. I. RICHIE, T. MIYAMOTO, W. H. CARR and I. L. WEISSMAN, 2000 B lymphopoiesis in the thymus. J Immunol **164:** 5221-5226.

AVITAHL, N., S. WINANDY, C. FRIEDRICH, B. JONES, Y. GE *et al.*, 1999 Ikaros sets thresholds for T cell activation and regulates chromosome propagation. Immunity **10:** 333-343.

BAGUNA, J., P. MARTINEZ, J. PAPS and M. RIUTORT, 2008 Back in time: a new systematic proposal for the Bilateria. Philos Trans R Soc Lond B Biol Sci **363:** 1481-1491.

BALL, C. A., and J. M. CHERRY, 2001 Genome comparisons highlight similarity and diversity within the eukaryotic kingdoms. Curr Opin Chem Biol **5:** 86-89.

BALL, E. E., DE JONG, D. M., SCHIERWATER, B., SHINZATO, C.,HAYWARD, D. C. & MILLER, D. J., 2007 Implications of cnidarian gene expression patterns for the origins of bilateralty—is the glass half full or half empty? Int. Comp. Biol **47:** 701-711.

BENJAMIN PRUD'HOMME, N. G., AND SEAN B. CARROLL, 2007 Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A **104**.

BENOS, P. V., A. S. LAPEDES and G. D. STORMO, 2002 Probabilistic code for DNA recognition by proteins of the EGR family. J Mol Biol **323:** 701-727.

BERNSTEIN, B. E., T. S. MIKKELSEN, X. XIE, M. KAMAL, D. J. HUEBERT *et al.*, 2006 A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell **125:** 315-326.

BIBIKOVA, M., K. BEUMER, J. K. TRAUTMAN and D. CARROLL, 2003 Enhancing gene targeting with designed zinc finger nucleases. Science **300:** 764.

BOLLUM, F. J., 1979 Terminal deoxynucleotidyl transferase as a hematopoietic cell marker. Blood **54:** 1203-1215.

BRIDGES, C. B., 1936 The Bar "Gene" a Duplication. Science **83:** 210-211.

BROWN, D. D., 1984 The role of stable complexes that repress and activate eucaryotic genes. Cell **37:** 359-365.

BROWN, K. E., S. S. GUEST, S. T. SMALE, K. HAHM, M. MERKENSCHLAGER *et al.*, 1997 Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell **91:** 845-854.

BRYAN T. PHILLIPS, A. R. K. I., RYAN KING, JEFF HARDIN, AND JUDITH KIMBLE, 2007 Reciprocal asymmetry of SYS-1/-catenin and POP-1/TCF controls asymmetric divisions in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A.

CAIRNS, B. R., Y. J. KIM, M. H. SAYRE, B. C. LAURENT and R. D. KORNBERG, 1994 A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. Proc Natl Acad Sci U S A **91:** 1950-1954.

CHANG, W., C. TILMANN, K. THOEMKE, F. H. MARKUSSEN, L. D. MATHIES *et al.*, 2004 A forkhead protein controls sexual identity of the *C. elegans* male somatic gonad. Development **131:** 1425-1436.

CHI, T. H., M. WAN, K. ZHAO, I. TANIUCHI, L. CHEN *et al.*, 2002 Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. Nature **418:** 195-199.

CHITWOOD, B. B., and M. B. CHITWOOD, 1950 *Introduction to Nematology*. University Park Press, Baltimore.

CHOO, Y., I. SANCHEZ-GARCIA and A. KLUG, 1994 In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature **372:** 642-645.

CLEARY, M. D., and C. Q. DOE, 2006 Regulation of neuroblast competence: multiple temporal identity factors specify distinct neuronal fates within a single early competence window. Genes Dev **20:** 429-434.

COBB, B. S., S. MORALES-ALCELAY, G. KLEIGER, K. E. BROWN, A. G. FISHER *et al.*, 2000 Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. Genes Dev **14:** 2146-2160.

COYNE, H. E. H. A. J. A., 2007 The Locus of Evolution: Evo Devo and the Genetics of Adaptation. Evolution **61:** 995-1016.

CUPIT, P. M., J. D. HANSEN, A. S. MCCARTY, G. WHITE, M. CHIODA *et al.*, 2003 Ikaros family members from the agnathan *Myxine glutinosa* and the urochordate *Oikopleura dioica*: emergence of an essential transcription factor for adaptive immunity. J Immunol **171:** 6006-6013.

DAVIDSON, E. H., 2006 *Genomic Regulatory Systems: Development and Evolution*. Academic Press, San Diego.

DOYON, Y., J. M. MCCAMMON, J. C. MILLER, F. FARAJI, C. NGO *et al.*, 2008 Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol **26:** 702-708.

ELLIOTT, J., C. JOLICOEUR, V. RAMAMURTHY and M. CAYOUETTE, 2008 Ikaros confers early temporal competence to mouse retinal progenitor cells. Neuron **60:** 26-39.

EMERSON, R. O., and J. H. THOMAS, 2009 Adaptive evolution in zinc finger transcription factors. PLoS Genet **5:** e1000325.

EZZAT, S., R. MADER, S. FISCHER, S. YU, C. ACKERLEY *et al.*, 2006 An essential role for the hematopoietic transcription factor Ikaros in hypothalamic-pituitary-mediated somatic growth. Proc Natl Acad Sci U S A **103:** 2214-2219.

FAIRALL, L., J. W. SCHWABE, L. CHAPMAN, J. T. FINCH and D. RHODES, 1993 The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature **366:** 483-487.

FAY, D. S., H. M. STANLEY, M. HAN and W. B. WOOD, 1999 A *Caenorhabditis elegans* homologue of *hunchback* is required for late stages of development but not early embryonic patterning. Dev Biol **205:** 240-253.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y.-L. YAN *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545.

FRIEDMAN, L., S. SANTA ANNA-ARRIOLA, J. HODGKIN and J. KIMBLE, 2000 *gon-4*, a cell lineage regulator required for gonadogenesis in *Caenorhabditis elegans*. Developmental Biology **228:** 350-362.

FURLEY, A. J., S. MIZUTANI, K. WEILBAECHER, H. S. DHALIWAL, A. M. FORD *et al.*, 1986 Developmentally regulated rearrangement and expression of genes encoding the T cell receptor-T3 complex. Cell **46:** 75-87.

GEORGOPOULOS, K., M. BIGBY, J. H. WANG, A. MOLNAR, P. WU *et al.*, 1994 The Ikaros gene is required for the development of all lymphoid lineages. Cell **79:** 143-156.

GEORGOPOULOS, K., D. D. MOORE and B. DERFLER, 1992a Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science **258:** 808-812.

GEORGOPOULOS, K., B. A. MORGAN and D. D. MOORE, 1992b Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activity of a T-cell-specific enhancer. Mol Cell Biol **12:** 747-757.

GINSBERG, A. M., B. O. KING and R. G. ROEDER, 1984 *Xenopus* 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell **39:** 479-489.

GREENWALD, I. S., P. W. STERNBERG and H. R. HORVITZ, 1983 The *lin-12* locus specifies cell fates in *Caenorhabditis elegans*. Cell **34:** 435-444.

GROSSKORTENHAUS, R., B. J. PEARSON, A. MARUSICH and C. Q. DOE, 2005 Regulation of temporal identity transitions in *Drosophila* neuroblasts. Dev Cell **8:** 193-202.

HAHM, K., P. ERNST, K. LO, G. S. KIM, C. TURCK *et al.*, 1994 The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. Mol Cell Biol **14:** 7111-7123.

HALANYCH, K. M., 2004 The new view of animal phylogeny. Annu Rev Ecol Syst **35:** 229-256.

HALANYCH, K. M., J. D. BACHELLER, A. M. AGUINALDO, S. M. LIVA, D. M. HILLIS *et al.*, 1995 Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science **267:** 1641-1643.

HALDANE, J. B. S., 1932 *The Causes of Evolution*. Longmans and Green, London.

HAYNES, B. F., M. J. TELEN, L. P. HALE and S. M. DENNING, 1989 CD44-a molecule involved in leukocyte adherence and T-cell activation. Immunol Today **10:** 423-428.

HONMA, Y., H. KIYOSAWA, T. MORI, A. OGURI, T. NIKAIDO *et al.*, 1999 Eos: a novel member of the Ikaros gene family expressed predominantly in the developing nervous system. FEBS Lett **447:** 76-80.

HUANG, G., X. XIE, Y. HAN, L. FAN, J. CHEN *et al.*, 2007 The identification of lymphocyte-like cells and lymphoid-related genes in amphioxus indicates the twilight for the emergency of adaptive immune system. PLoS One **2:** e206.

INNAN, H., and F. KONDRASHOV, 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet **11:** 97-108.

ISSHIKI, T., B. PEARSON, S. HOLBROOK and C. Q. DOE, 2001 *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. Cell **106:** 511-521.

JAMIE PINNELL, P. S. L., SIERRA COLAVITO, CHRIS LOWE, AND ROBERT M. SAVAGE, 2006 The divergent roles of the segmentation gene *hunchback*. Integrative and Comparitive Biology **46:** 519-532.

JAMIESON, A. C., J. C. MILLER and C. O. PABO, 2003 Drug discovery with engineered zinc-finger proteins. Nat Rev Drug Discov **2:** 361-368.

JOHN, L. B., S. YOONG and A. C. WARD, 2009 Evolution of the Ikaros gene family: implications for the origins of adaptive immunity. J Immunol **182:** 4792-4799.

KAPLAN, T., N. FRIEDMAN and H. MARGALIT, 2005 *Ab initio* prediction of transcription factor targets using structural knowledge. PLoS Comput Biol **1:** e1.

KEHLE, J., D. BEUCHLE, S. TREUHEIT, B. CHRISTEN, J. A. KENNISON *et al.*, 1998 dMi-2, a *hunchback*-interacting protein that functions in polycomb repression. Science **282:** 1897-1900.

KELLEY, C. M., T. IKEDA, J. KOIPALLY, N. AVITAHL, L. WU *et al.*, 1998 Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. Curr Biol **8:** 508-515.

KERNER, P., F. ZELADA GONZALEZ, M. LE GOUAR, V. LEDENT, D. ARENDT *et al.*, 2006 The expression of a hunchback ortholog in the polychaete annelid *Platynereis dumerilii* suggests an ancestral role in mesoderm development and neurogenesis. Dev Genes Evol **216:** 821-828.

KIDD, A. R., 3RD, J. A. MISKOWSKI, K. R. SIEGFRIED, H. SAWA and J. KIMBLE, 2005 A beta-catenin identified by functional rather than sequence criteria and its role in Wnt/MAPK signaling. Cell **121:** 761-772.

KIM, J., S. SIF, B. JONES, A. JACKSON, J. KOIPALLY *et al.*, 1999 Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity **10:** 345-355.

KIMBLE, J., 1981 Alterations in cell lineage following laser ablation of cells in the somatic gonad of *Caenorhabditis elegans*. Dev. Biol. **87:** 286-300.

KIMBLE, J., and D. HIRSH, 1979 The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Developmental Biology **70:** 396-417.

KIMBLE, J. E., and J. G. WHITE, 1981 On the control of germ cell development in *Caenorhabditis elegans*. Developmental Biology **81:** 208-219.

KLUG, C. A., S. J. MORRISON, M. MASEK, K. HAHM, S. T. SMALE *et al.*, 1998 Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. Proc Natl Acad Sci U S A **95:** 657-662.

KOIPALLY, J., and K. GEORGOPOULOS, 2000 Ikaros interactions with CtBP reveal a repression mechanism that is independent of histone deacetylase activity. J Biol Chem **275:** 19594-19602.

KOIPALLY, J., J. KIM, B. JONES, A. JACKSON, N. AVITAHL *et al.*, 1999a Ikaros chromatin remodeling complexes in the control of differentiation of the hemo-lymphoid system. Cold Spring Harb Symp Quant Biol **64:** 79-86.

KOIPALLY, J., A. RENOLD, J. KIM and K. GEORGOPOULOS, 1999b Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. EMBO J **18:** 3090-3100.

KRUGER, W., C. L. PETERSON, A. SIL, C. COBURN, G. ARENTS *et al.*, 1995 Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. Genes Dev **9:** 2770-2779.

LANDER, E. S., L. M. LINTON, B. BIRREN, C. NUSBAUM, M. C. ZODY *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860-921.

LARGE, E. E., and L. D. MATHIES, 2010 *hunchback* and Ikaros-like zinc finger genes control reproductive system development in *Caenorhabditis elegans*. Dev Biol **339:** 51-64.

LEE, M. S., G. P. GIPPERT, K. V. SOMAN, D. A. CASE and P. E. WRIGHT, 1989 Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science **245:** 635-637.

LI, W.-H., 1999 *Molecular Evolution*. Sinauer, Sunderland, MA.

LIN, S. Y., S. M. JOHNSON, M. ABRAHAM, M. C. VELLA, A. PASQUINELLI *et al.*, 2003 The *C elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. Dev Cell **4:** 639-650.

LIU, Z., F. MAO, J. T. GUO, B. YAN, P. WANG *et al.*, 2005 Quantitative evaluation of protein-DNA interactions using an optimized knowledge-based potential. Nucleic Acids Res **33:** 546-558.

LIVESEY, F. J., and C. L. CEPKO, 2001 Vertebrate neural cell-fate determination: lessons from the retina. Nat Rev Neurosci **2:** 109-118.

LO, K., N. R. LANDAU and S. T. SMALE, 1991 LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. Mol Cell Biol **11:** 5229-5243.

LOOMAN, C., M. ABRINK, C. MARK and L. HELLMAN, 2002 KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol Biol Evol **19:** 2118-2130.

LYNCH, M., and J. S. CONERY, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

LYNCH, M., and V. KATJU, 2004 The altered evolutionary trajectories of gene duplicates. Trends Genet **20:** 544-549.

MACKAY, J. P., and M. CROSSLEY, 1998 Zinc fingers are sticking together. Trends Biochem Sci **23:** 1-4.

MAEDER, M. L., S. THIBODEAU-BEGANNY, A. OSIAK, D. A. WRIGHT, R. M. ANTHONY *et al.*, 2008 Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell **31:** 294-301.

MANDEL-GUTFREUND, Y., and H. MARGALIT, 1998 Quantitative parameters for amino acid-base interaction: implications for prediction of protein-DNA binding sites. Nucleic Acids Res **26:** 2306-2312.

MASAKO ASAHINA, T. V., MARIE SILHANKOVA, VLADIMIR KORINEK, AND MAREK JINDRA, 2006 Crosstalk between a Nuclear Receptor and b-Catenin Signaling Decides Cell Fates in the *C. elegans* Somatic Gonad. Developmental Cell **11:** 203-211.

MATHIES, L. D., S. T. HENDERSON and J. KIMBLE, 2003 The *C. elegans* Hand gene controls embryogenesis and early gonadogenesis. Development **130:** 2881-2892.

MATHIES, L. D., M. SCHVARZSTEIN, K. M. MORPHY, R. BLELLOCH, A. M. SPENCE *et al.*, 2004 TRA-1/GLI controls development of somatic gonadal precursors in *C. elegans*. Development **131:** 4333-4343.

MAYER, W. E., C. O'HUIGIN, H. TICHY, J. TERZIC and M. SARAGA-BABIC, 2002 Identification of two Ikaros-like transcription factors in lamprey. Scand J Immunol **55:** 162-170.

MCCARTER, J., B. BARTLETT, T. DANG and T. SCHEDL, 1997 Soma-germ cell interactions in *Caenorhabditis elegans*: multiple events of hermaphrodite germline development require the somatic sheath and spermathecal lineages. Developmental Biology **181:** 121-143.

MCCARTY, A. S., G. KLEIGER, D. EISENBERG and S. T. SMALE, 2003 Selective dimerization of a C2H2 zinc finger subfamily. Mol Cell **11:** 459-470.

MENG, X., M. B. NOYES, L. J. ZHU, N. D. LAWSON and S. A. WOLFE, 2008 Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol **26:** 695-701.

MILLER, J., A. D. MCLACHLAN and A. KLUG, 1985 Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J **4:** 1609-1614.

MISKOWSKI, J., Y. LI and J. KIMBLE, 2001 The *sys-1* gene and sexual dimorphism during gonadogenesis in *Caenorhabditis elegans*. Developmental Biology **230:** 61-73.

MOLNAR, A., and K. GEORGOPOULOS, 1994 The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Mol Cell Biol **14:** 8292-8303.

MORGAN, B., L. SUN, N. AVITAHL, K. ANDRIKOPOULOS, T. IKEDA *et al.*, 1997 Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. EMBO J. **16:** 2004-2013.

MULLER, H. J., 1935 The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. Genetica **17:** 237-252.

MULLER, H. J., 1936 Bar Duplication. Science **83:** 528-530.

MULLER, J., and M. BIENZ, 1992 Sharp anterior boundary of homeotic gene expression conferred by the fushi tarazu protein. EMBO J **11:** 3653-3661.

MURATA, Y., and R. P. WHARTON, 1995 Binding of pumilio to maternal *hunchback* mRNA is required for posterior patterning in *Drosophila* embryos. Cell **80:** 747-756.

NAITO, T., P. GOMEZ-DEL ARCO, C. J. WILLIAMS and K. GEORGOPOULOS, 2007 Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression. Immunity **27:** 723-734.

NAKAMURA, T., T. MORI, S. TADA, W. KRAJEWSKI, T. ROZOVSKAIA *et al.*, 2002 ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell **10:** 1119-1128.

NAKASEKO, Y., D. NEUHAUS, A. KLUG and D. RHODES, 1992 Adjacent zinc-finger motifs in multiple zinc-finger peptides from SWI5 form structurally independent, flexibly linked domains. J Mol Biol **228:** 619-636.

NEIGEBORN, L., and M. CARLSON, 1984 Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. Genetics **108:** 845-858.

NEUHAUS, D., Y. NAKASEKO, J. W. SCHWABE and A. KLUG, 1992 Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional 1H nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet. J Mol Biol **228:** 637-651.

NEWMAN, A. P., and P. W. STERNBERG, 1996 Coordinated morphogenesis of epithelia during development of the *Caenorhabditis elegans* uterine-vulval connection. Proc. Natl. Acad. Sci. USA **93:** 9329-9333.

NG, S. Y., T. YOSHIDA, J. ZHANG and K. GEORGOPOULOS, 2009 Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity **30:** 493-507.

NICHOGIANNOPOULOU, A., M. TREVISAN, S. NEBEN, C. FRIEDRICH and K. GEORGOPOULOS, 1999 Defects in hemopoietic stem cell activity in Ikaros mutant mice. J Exp Med **190:** 1201-1214.

NOLDE, M. J., N. SAKA, K. L. REINERT and F. J. SLACK, 2007 The *Caenorhabditis elegans* pumilio homolog, *puf-9*, is required for the 3'UTR-mediated repression of the *let-7* microRNA target gene, hbl-1. Dev Biol **305:** 551-563.

O'NEILL, D. W., S. S. SCHOETZ, R. A. LOPEZ, M. CASTLE, L. RABINOWITZ *et al.*, 2000 An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. Mol Cell Biol **20:** 7572-7582.

OHNO, S., 1970 *Evolution by Gene Duplication*. Springer, Berlin, Heidelberg, New York.

PATEL, N. H., D. C. HAYWARD, S. LALL, N. R. PIRKL, D. DIPIETRO *et al.*, 2001 Grasshopper *hunchback* expression reveals conserved and novel aspects of axis formation and segmentation. Development **128:** 3459-3472.

PAVLETICH, N. P., and C. O. PABO, 1991 Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science **252:** 809-817.

PEARSON, B. J., and C. Q. DOE, 2003 Regulation of neuroblast competence in *Drosophila.* Nature **425:** 624-628.

PELHAM, H. R., and D. D. BROWN, 1980 A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc Natl Acad Sci U S A **77:** 4170-4174.

PERDOMO, J., M. HOLMES, B. CHONG and M. CROSSLEY, 2000 Eos and Pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. J. Biol. Chem. **275:** 38347-38354.

PERSIKOV, A. V., R. OSADA and M. SINGH, 2009 Predicting DNA recognition by Cys2His2 zinc finger proteins. Bioinformatics **25:** 22-29.

PETERSON, K. J., and D. J. EERNISSE, 2001 Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol Dev **3:** 170-205.

PORTEUS, M. H., and D. BALTIMORE, 2003 Chimeric nucleases stimulate gene targeting in human cells. Science **300:** 763.

PRINCE, V. E., and F. B. PICKETT, 2002 Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet **3:** 827-837.

QIAN, S., M. CAPOVILLA and V. PIRROTTA, 1991 The bx region enhancer, a distant cis-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. EMBO J **10:** 1415-1425.

RIVERA-POMAR, R., and H. JACKLE, 1996 From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. Trends Genet **12:** 478-483.

ROSE, K. L., V. P. WINFREY, L. H. HOFFMAN, D. H. HALL, T. FURUTA *et al.*, 1997 The POU gene *ceh-18* promotes gonadal sheath cell differentiation and function required for meiotic maturation and ovulation in *Caenorhabditis elegans*. Developmental Biology **192:** 59-77.

ROTHENBERG, E. V., and R. PANT, 2004 Origins of lymphocyte developmental programs: transcription factor evidence. Semin Immunol **16:** 227-238.

SAHA, A., J. WITTMEYER and B. R. CAIRNS, 2006 Mechanisms for nucleosome movement by ATP-dependent chromatin remodeling complexes. Results Probl Cell Differ **41:** 127-148.

SANTIAGO, Y., E. CHAN, P. Q. LIU, S. ORLANDO, L. ZHANG *et al.*, 2008 Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A **105:** 5809-5814.

SAWA, H., L. LOBEL and H. R. HORVITZ, 1996 The *Caenorhabditis elegans* gene *lin-17*, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. Genes & Development **10:** 2189-2197.

SEYDOUX, G., and I. GREENWALD, 1989 Cell autonomy of *lin-12* function in a cell fate decision in C. elegans. Cell **57:** 1237-1245.

SHIMELL, M. J., J. SIMON, W. BENDER and M. B. O'CONNOR, 1994 Enhancer point mutation results in a homeotic transformation in *Drosophila*. Science **264:** 968-971.

SHIMONO, Y., H. MURAKAMI, K. KAWAI, P. A. WADE, K. SHIMOKATA *et al.*, 2003 Mi-2 beta associates with BRG1 and RET finger protein at the distinct regions with transcriptional activating and repressing abilities. J Biol Chem **278:** 51638-51645.

SIEGFRIED, K., A. R. I. KIDD, M. A. CHESNEY and J. KIMBLE, 2004 The *sys-1* and *sys-3* genes cooperate with Wnt signaling to establish the proximal-distal axis of the *Caenorhabditis elegans* gonad. Genetics **166:** 171-186.

SIEGFRIED, K., and J. KIMBLE, 2002 POP-1 controls axis formation during early gonadogenesis in *C. elegans*. Development **129:** 443-453.

SMALE, S. T., and K. DORSHKIND, 2006 Hematopoiesis flies high with Ikaros. Nat Immunol **7:** 367-369.

SONODA, J., and R. P. WHARTON, 1999 Recruitment of Nanos to *hunchback* mRNA by Pumilio. Genes Dev **13:** 2704-2712.

SRIDHARAN, R., and S. T. SMALE, 2007 Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. J Biol Chem **282:** 30227-30238.

SRIVASTAVA, M., E. BEGOVIC, J. CHAPMAN, N. H. PUTNAM, U. HELLSTEN *et al.*, 2008 The *Trichoplax* genome and the nature of placozoans. Nature **454:** 955-960.

ST JOHNSTON, D., and C. NUSSLEIN-VOLHARD, 1992 The origin of pattern and polarity in the *Drosophila* embryo. Cell **68:** 201-219.

STERN, M., R. JENSEN and I. HERSKOWITZ, 1984 Five SWI genes are required for expression of the HO gene in yeast. J Mol Biol **178:** 853-868.

STERNBERG, P. W., and H. R. HORVITZ, 1988 *lin-17* mutations of *Caenorhabditis elegans* disrupt certain asymmetric cell divisions. Developmental Biology **130:** 67-73.

STRUHL, K., 1998 Histone acetylation and transcriptional regulatory mechanisms. Genes Dev **12:** 599-606.

SULSTON, J. E., E. SCHIERENBERG, J. G. WHITE and J. N. THOMSON, 1983 The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Developmental Biology **100:** 64-119.

SUN, A. Y., and E. J. LAMBIE, 1997 *gon-2*, a gene required for gonadogenesis in *Caenorhabditis elegans*. Genetics **147:** 1077-1089.

SUN, L., A. LIU and K. GEORGOPOULOS, 1996 Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. EMB0 J. **15:** 5358-5369.

SUZUKI, M., S. E. BRENNER, M. GERSTEIN and N. YAGI, 1995 DNA recognition code of transcription factors. Protein Eng **8:** 319-328.

TAUTZ, D., 1988 Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. Nature **332:** 281-284.

TILMANN, C., and J. KIMBLE, 2005 Cyclin D regulation of a sexually dimorphic asymmetric cell division. Dev Cell **9:** 489-499.

TONG, J. K., C. A. HASSIG, G. R. SCHNITZLER, R. E. KINGSTON and S. L. SCHREIBER, 1998 Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature **395:** 917-921.

VENTER, J. C., M. D. ADAMS, E. W. MYERS, P. W. LI, R. J. MURAL *et al.*, 2001 The sequence of the human genome. Science **291:** 1304-1351.

WADE, P. A., P. L. JONES, D. VERMAAK and A. P. WOLFFE, 1998 A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. Curr Biol **8:** 843-846.

WANG, J. H., A. NICHOGIANNOPOULOU, L. WU, L. SUN, A. H. SHARPE *et al.*, 1996 Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. Immunity **5:** 537-549.

WHARTON, R. P., and G. STRUHL, 1991 RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. Cell **67:** 955-967.

WILLIAMS, C. J., T. NAITO, P. G. ARCO, J. R. SEAVITT, S. M. CASHMAN *et al.*, 2004 The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. Immunity **20:** 719-733.

WINANDY, S., L. WU, J. H. WANG and K. GEORGOPOULOS, 1999 Pre-T cell receptor (TCR) and TCR-controlled checkpoints in T cell differentiation are set by Ikaros. J Exp Med **190:** 1039-1048.

WINANDY, S., P. WU and K. GEORGOPOULOS, 1995 A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. Cell **83:** 289-299.

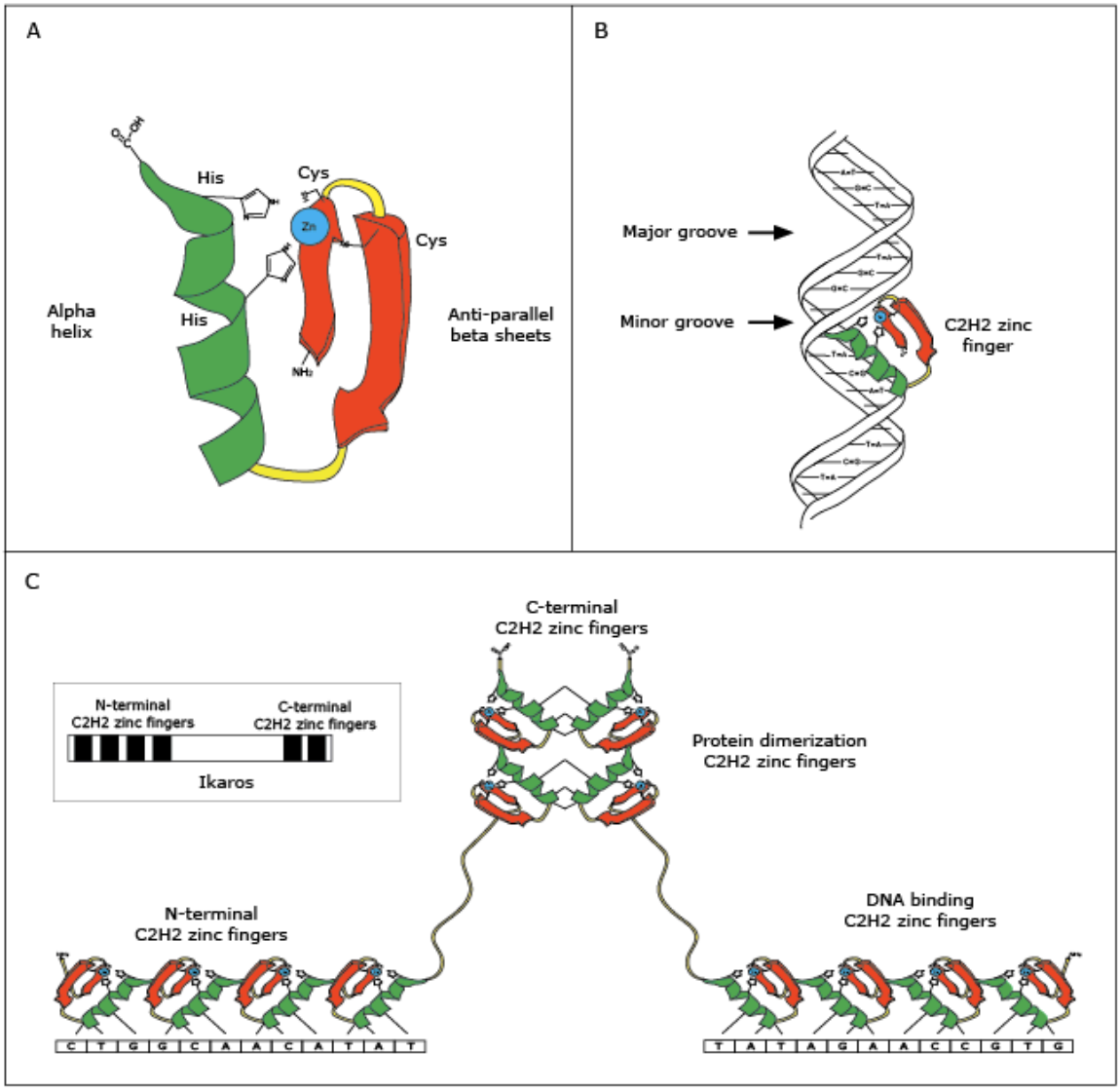
WRAY, G. A., 2007 The evolutionary significance of *cis*-regulatory mutations. Nature **8**.

XUE, Y., J. WONG, G. T. MORENO, M. K. YOUNG, J. COTE *et al.*, 1998 NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell **2:** 851-861.

ZHANG, C. C., and M. BIENZ, 1992 Segmental determination in *Drosophila* conferred by *hunchback (hb)*, a repressor of the homeotic gene *Ultrabithorax (Ubx)*. Proc Natl Acad Sci U S A **89:** 7511-7515.

ZHANG, Y., G. LEROY, H. P. SEELIG, W. S. LANE and D. REINBERG, 1998 The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell **95:** 279-289.

Figure 1. Versatility of C2H2 zinc finger proteins
(A) C2H2 zinc fingers are formed from a N-terminal anti-parallel beta sheet (red) containing two cysteines (Cys) and an C-terminal alpha helix (green) containing two histidines (His). (B) C2H2 zinc fingers bind the major groove of DNA primarily utilizing amino acids found in the alpha helix. (C) The modular nature of C2H2 zinc finger genes and nearly one-to-one correspondence of amino acids to nucleotide recognition is indicated with one module contacting three basepairs. The same amino acids known to bind DNA are also essential for heterodimerization and homodimerization of Ikaros (inset) family members; implying a similar one-to-one correspondence between amino acids in respective dimerization domains (MCCARTY *et al.* 2003). The Ikaros protein (inset) is pictured showing the unique spacing of the four N-terminal DNA binding zinc fingers and the two C-terminal protein-protein dimerization zinc fingers.


A
His
Cys
Zn
Cys
Alpha helix
His
Anti-parallel beta sheets
B
Major groove
Minor groove
C2H2 zinc finger
C
C-terminal C2H2 zinc fingers
N-terminal C2H2 zinc fingers
C-terminal C2H2 zinc fingers
Ikaros
Protein dimerization C2H2 zinc fingers
N-terminal C2H2 zinc fingers
DNA binding C2H2 zinc fingers
C T G G C A A C A T A T
T A T A G A A C C G T G

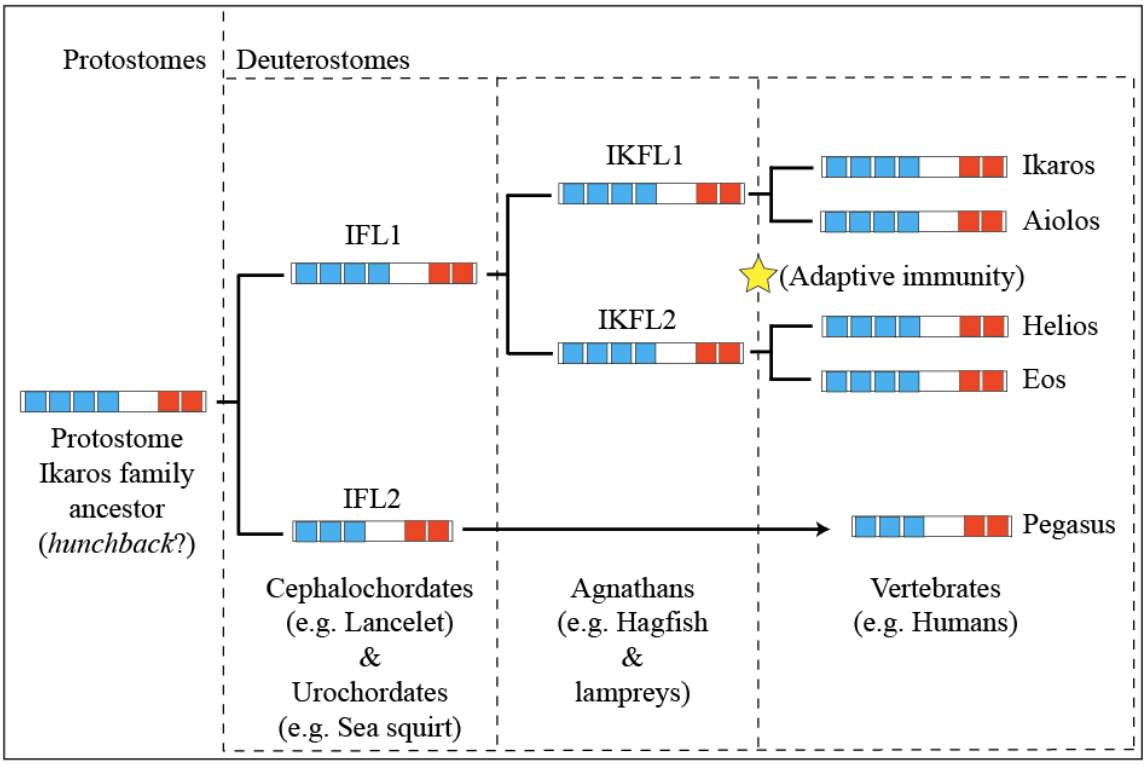
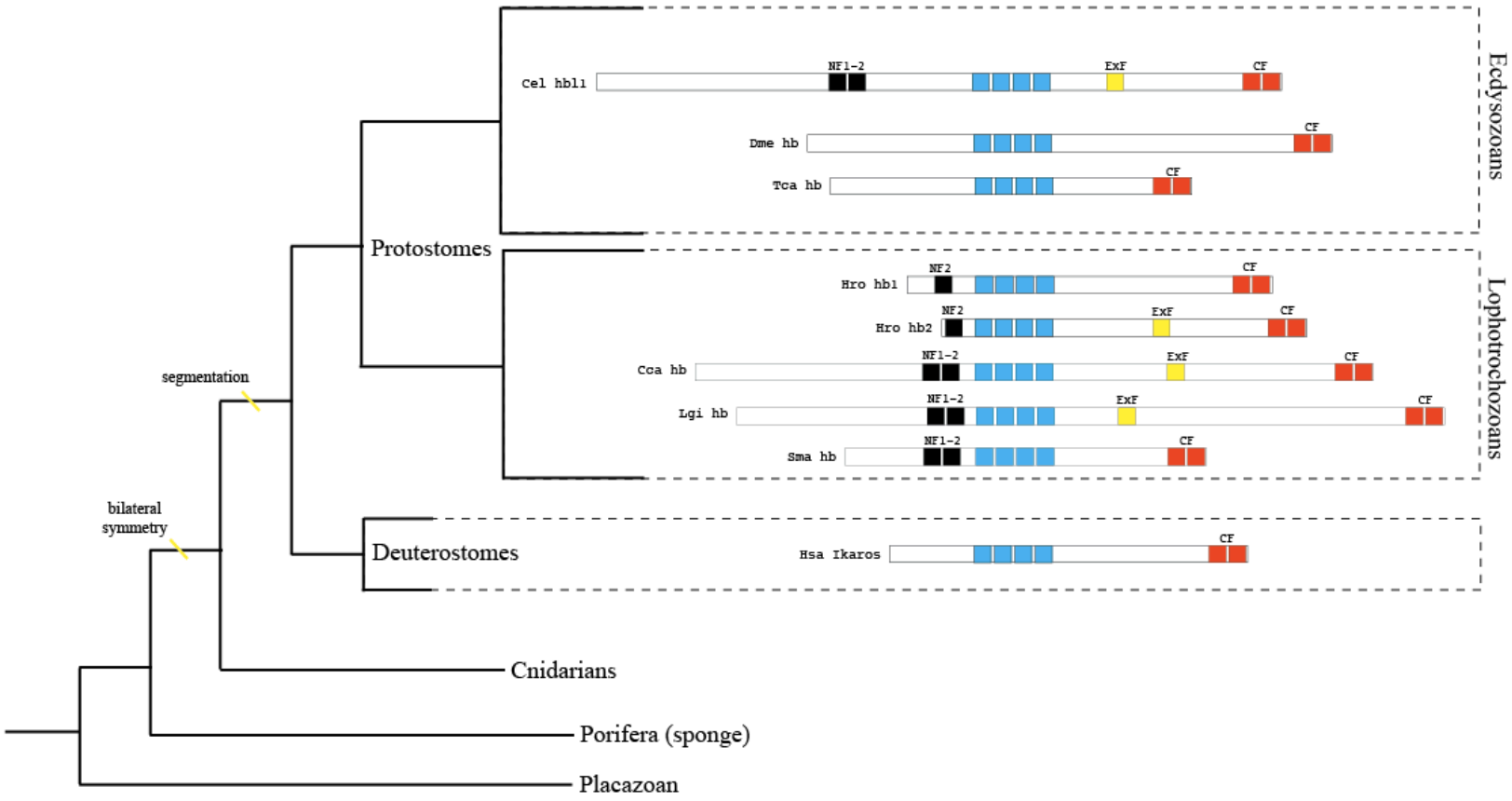


Figure 2. Duplication and divergence of the Ikaros family of transcription factors Adapted from John *et al.*, 2009. Large rectangles containing red and blue boxes represent Ikaros family proteins with the N-terminus located on the left and the C-terminus on the right. Ikaros family proteins have distinct zinc finger domains with characterized or putative DNA binding activity (3-4 blue boxes representing individual zinc fingers) and a zinc finger domain with known or potential protein-protein dimerization activity (2 red boxes representing individual zinc fingers). An ancestral protostome or protostome-deuterostome protein similar to *hunchback* has previously been hypothesized to have given rise to the Ikaros family (far left of panel). An ancient gene duplication event gave rise to Pegasus (lower half of panel), which contains three DNA-binding zinc fingers, and all other Ikaros family members (upper half of panel), which have four DNA-binding zinc fingers. The duplication and divergence of IKFL1 and IKFL2 coincides with the rise of adaptive immunity in vertebrates (yellow star).

Figure 3. Representative sample of known *hunchback* family members
A simplified phylogenetic tree indicating the hypothesized points at which bilaterial symmetry and segmentation evolved (yellow line). A single deuterostome and a representative sample of protostome proteins are shown with the N-terminus located on the left and the C-terminus on the right. Each protein has distinct zinc finger domains consisting of characterized or putative DNA binding activity (4 blue boxes representing individual zinc fingers) and a zinc finger domain with known or potential protein-protein dimerization activity (2 red boxes representing individual zinc fingers). Additional uncharacterized zinc fingers have been found in found in protostome proteins near the N-terminus (NF1-2: N-terminal fingers denoted by black boxes representing individual zinc fingers) and in the interfinger region between the DNA-binding domain and the protein-protein dimerization domain (ExF: Extra finger denoted by a yellow box representing a single zinc finger). Ecdysozoans: Cel (*C. elegans*), Dme (*Drosophila melanogaster*), Tca (*Tribolium castaneum*); Lophotrochozoans: Hro (*Helobdella robusta*), Cca (*Capitella capitata*), Lgi (*Lottia gigantea*), Sma (*Schistosoma mansoni*); Deuterostomes: Hsa (*Homo sapiens*)


Ecdysozoans
Lophotrochozoans
Protostomes
Deuterostomes
segmentation
bilateral symmetry
Cnidarians
Porifera (sponge)
Placazoan
Cel hbl1
Dme hb
Tca hb
Hro hb1
Hro hb2
Cca hb
Lgi hb
Sma hb
Hsa Ikaros
NF1-2
NF2
ExF
CF

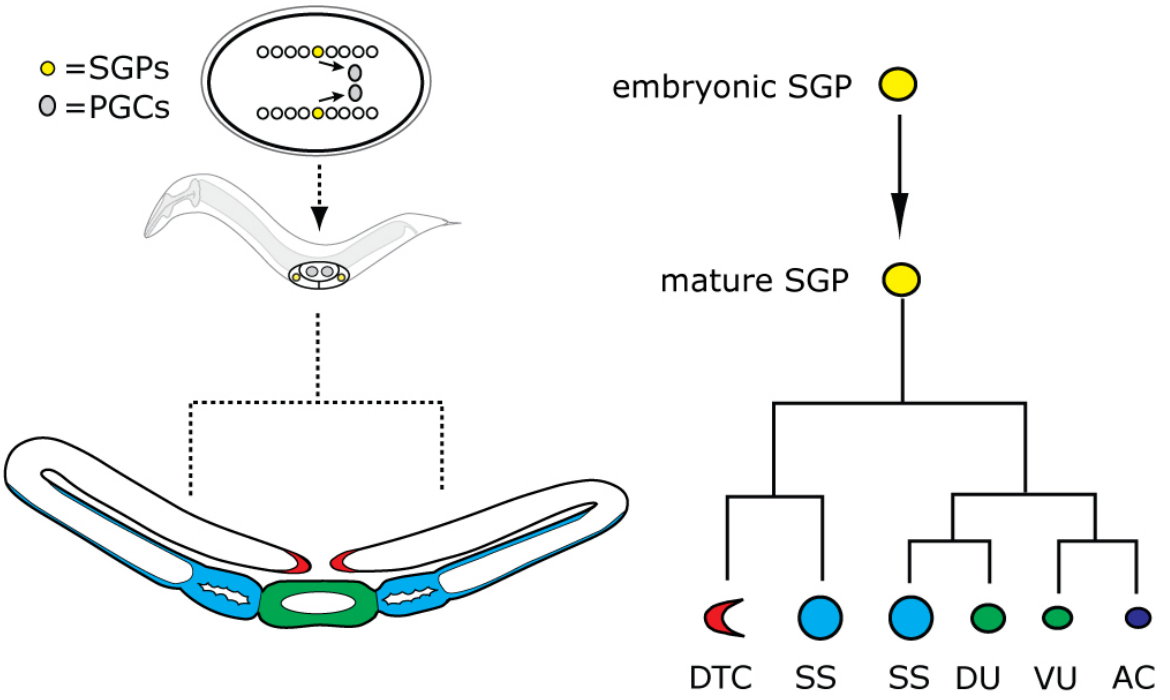


Figure 4. The *C. elegans* reproductive system as a model for lineage progression
Left panel: Somatic gonadadal precursors (SGPs: yellow) are specified within the mesodermal layer (small white circles) and migrate to meet the primordial germ cells (PGCs: light grey). These four cells form the gonad primordium, which is present in the L1 larvae of all nematode species (CHITWOOD and CHITWOOD 1950). Each SGP can generate a somatic gonad arm and the two SGPs together generate two symmetrical U-shaped gonad arms. Right panel: Lineage diagram indicating the embryonic SGP (yellow; top), which develops into the mature SGP (yellow; middle) found in L1 larvae. Each SGP can generate all five tissues of the somatic gonad: distal tip cells (DTC; red), sheath/spermathecae (SS; blue), dorsal uterine (DU; green), ventral uterine (VU; green), and an anchor cell (AC; purple). The colors of individual cells on the right correspond with the location and colors of the two gonadal arms pictured on the left.

CHAPTER TWO

hunchback and Ikaros-like zinc finger genes control reproductive system development in *C. elegans*

The following chapter was published in *Developmental Biology*. 2010. 339(1):51-64.

At the time of publication *ztf-16* was referred to as R08E3.4, which is reflected in Chapter Two.

hunchback and Ikaros-like zinc finger genes control reproductive system development in *C. elegans*

Edward E. Large and Laura D. Mathies*

Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614 USA

*Corresponding author:

Laura Mathies

3510 Thomas Hall

Department of Genetics

North Carolina State University

Raleigh, NC 27695-7614

Phone: 919-515-7079

Fax: 919-515-3355

Email: ldmathies@ncsu.edu

ABSTRACT

Here we provide evidence for a C2H2 zinc finger gene family with similarity to Ikaros and *hunchback*. The founding member of this family is *C. elegans ehn-3*, which has important and poorly understood functions in somatic gonad development. We examined the expression and function of four additional *hunchback/Ikaros*-like (HIL) genes in *C. elegans* reproductive system development. Two genes, *ehn-3* and *R08E3.4*, are expressed in somatic gonadal precursors (SGPs) and have overlapping functions in their development. In *ehn-3; R08E3.4* double mutants, we find defects in the generation of distal tip cells, anchor cells, and spermatheca; three of the five tissues derived from the SGPs. We provide *in vivo* evidence that *C. elegans* HIL proteins have functionally distinct zinc finger domains, with specificity residing in the N-terminal set of four zinc fingers and a likely protein-protein interaction domain provided by the C-terminal pair of zinc fingers. In addition, we find that a chimeric human Ikaros protein containing the N-terminal zinc fingers of EHN-3 functions in *C. elegans*. Together, these results lend support to the idea that the *C. elegans* HIL genes and Ikaros have similar functional domains. We propose that *hunchback,* Ikaros*,* and the HIL genes arose from a common ancestor that was present prior to the divergence of protostomes and deuterostomes

Key words: Ikaros, *hunchback*, EHN-3, R08E3.4, gonadogenesis, *C. elegans*, gene duplication, zinc finger

INTRODUCTION

The *C. elegans* reproductive system develops from a four-celled primordium containing two somatic gonadal precursors (SGPs) and two primordial germ cells (PGCs) (KIMBLE and HIRSH 1979). Through a stereotyped cell lineage, the SGPs generate all somatic tissues of the reproductive organs (KIMBLE and HIRSH 1979). Each SGP generates one of the two hermaphrodite gonadal arms, with somatic tissue derived from the SGP and germ line formed from the PGC. In recent years, we have learned a great deal about the genes controlling the first asymmetric division of the SGPs (ASAHINA *et al.* 2006; CHANG *et al.* 2004; KIDD *et al.* 2005; MATHIES *et al.* 2004; MISKOWSKI *et al.* 2001; PHILLIPS *et al.* 2007; SAWA *et al.* 1996; SIEGFRIED *et al.* 2004; SIEGFRIED and KIMBLE 2002; STERNBERG and HORVITZ 1988; TILMANN and KIMBLE 2005), but much remains to be learned about the early controls of SGP development. Three transcription factors are known to function prior to the first SGP division: 1) the *C. elegans* Hand bHLH gene *hnd-1* controls the survival of the SGPs (MATHIES *et al.* 2003), 2) the *C. elegans* Gli gene *tra-1* controls the polarity and division of the SGPs (CHANG *et al.* 2004; MATHIES *et al.* 2004), and 3) the C2H2 zinc finger gene *ehn-3* interacts genetically with *hnd-1* and *tra-1* to control early SGP development (MATHIES *et al.* 2003; MATHIES *et al.* 2004). *ehn-3* is expressed almost exclusively in the SGPs, but its role(s) in SGP development remain poorly defined. The *ehn-3* gene has no clear ortholog in other sequenced genomes, but it does share sequence similarity and a similar C2H2 zinc finger spacing with mammalian Ikaros and *Drosophila hunchback*.

The Ikaros family of C2H2 zinc finger transcription factors are important regulators of hematopoiesis in vertebrates (SMALE and DORSHKIND 2006). Three of the five family

members, Ikaros, Aiolos, and Helios, are expressed predominantly in the hematopoietic system and control the development of lymphoid precursors from multipotent hematopoietic stem cells and the subsequent development of T and B cells (HAHM *et al.* 1998; MOLNAR and GEORGOPOULOS 1994; MORGAN *et al.* 1997). Ikaros family members share a unique arrangement of zinc fingers, with a set of four N-terminal C2H2 zinc fingers responsible for binding DNA and a set of two C-terminal C2H2 zinc fingers that mediate protein-protein interactions (MCCARTY *et al.* 2003; MOLNAR and GEORGOPOULOS 1994; SUN *et al.* 1996). Ikaros family members are expressed in even in the most primitive vertebrate immune systems, such as the lymphoid-like cells of hagfish and lampreys (ROTHENBERG and PANT 2004). They are also found in more basal deuterostomes that lack an adaptive immune system, such as the ascidian *Ciona intestinalis* (CUPIT *et al.* 2003) and the sea urchin *Strongylocentrotus purpuratus* (HIBINO *et al.* 2006; RAST *et al.* 2006). Ikaros orthologs may therefore have important and ancient roles outside of immune system development. Despite the conservation of this important group of proteins in deuterostomes, Ikaros family members have not been found in representative protostome genomes, such as *Drosophila melanogaster* or *Caenorhabditis elegans*.

hunchback was originally found as a segmentation gene affecting early *Drosophila* embryo patterning (ST JOHNSTON and NÜSSLEIN-VOLHARD 1992). *Drosophila hunchback* encodes four zinc fingers near the middle of the protein and two zinc fingers at the extreme C-terminus of the protein. Like Ikaros, the four tandem zinc fingers are important for DNA binding (SOMMER *et al.* 1992; STANOJEVIC *et al.* 1989; TREISMAN and DESPLAN 1989) and the two C-terminal zinc fingers can form homodimers (MCCARTY *et al.* 2003). Based on

comparisons of *hunchback* genes in annelids, nematodes, and arthropods, it has been proposed that the ancestral Hunchback protein contained nine zinc fingers consisting of two N-terminal fingers (NF), four middle fingers (MF), an extra finger (ExF), and two C-terminal fingers (CF) (PATEL *et al.* 2001; PINNELL *et al.* 2006). Searches for *hunchback*-like genes in deuterostomes have not revealed any unequivocal orthologs (KERNER *et al.* 2006). However, it has been suggested that *hunchback* may share a common origin with Ikaros based on their similar arrangement of zinc fingers and spacing between histidine residues (GEORGOPOULOS *et al.* 1992; JOHN *et al.* 2009; KNIGHT and SHIMELD 2001; PATEL *et al.* 2001).

Here we describe a family of C2H2 zinc finger genes that shares a similar organization of zinc fingers with *hunchback* and Ikaros. We refer to this family as *hunchback/Ikaros*-like (HIL) to distinguish it from previously described Ikaros family-like (IFL) genes in deuterostomes and *hunchback* genes in protostomes. The founding member of the HIL family is the *C. elegans* gene *ehn-3*, which has important roles in early somatic gonad development (MATHIES *et al.* 2003; MATHIES *et al.* 2004). We examined the function and expression of five HIL genes in *C. elegans* and found that three are expressed in the somatic gonad and two have overlapping functions in early SGP development. Using a combination of domain swaps and rescue experiments, we show that HIL genes have functionally distinct zinc finger domains, with specificity determined by their N-terminal zinc fingers. We further show that we can replace the Ikaros DNA binding zinc fingers with the N-terminal zinc fingers of *ehn-3* and this chimeric protein functions in *C. elegans* gonadogenesis. Based on this evidence we suggest that Ikaros and HIL proteins may utilize a

common molecular mechanism for controlling development. We further suggest that Ikaros, *hunchback*, and the HIL genes might share a common evolutionary origin prior to the divergence of protostomes and deuterostomes.

MATERIALS AND METHODS

Strains, genetics, and RNAi

Strains: *C. elegans* strains were cultured as described previously (BRENNER 1974; WOOD 1988). All strains were grown at 20°C unless otherwise specified. *C. elegans* mutants were derived from the N2 wild isolate (HODGKIN 1997). The following mutations were used in this study and are described in *C. elegans II* (HODGKIN 1997), cited references, or this work: *LGII*: *C46E10.8 (tm442), C46E10.9 (tm1692); LGIV*: *ehn-3(q689)* (MATHIES *et al.* 2003), *ehn-3(q766)* (MATHIES *et al.* 2004)*, ehn-3 (rd2); LGX*: *R08E3.4(tm2127, ok1916, za16)*. GFP-marked balancer chromosomes: *hT2[qIs48]* for *LGI* and *LGIII*, and *nT1[qIs50]* for *LGIV* and *LGV*. Molecular markers: *qIs56* [*lag-2::GFP*] (BLELLOCH *et al.* 1999), *qIs70* [*lag-2::YFP*] (KIDD *et al.* 2005), *rdIs2* [*ehn-3A::GFP*] (WELCHMAN *et al.* 2007), *ccIs4444* [*arg-1::GFP*] (KOSTAS and FIRE 2002), *ezIs1* [*K09C8.2::gfp*]*, ezIs2* [*fkh-6::GFP*] (CHANG *et al*. 2004) and *syIs57* [*cdh-3::CFP*] (INOUE *et al*. 2002).

RT-PCR: Using semi-nested RT-PCR with the SL1 trans-spliced leader, we determined the gene structures of *R08E3.4, F12E12.5, C46E10.8* and *C46E10.9*. Briefly, RNA was prepared from mixed-stage populations using Tri-reagent (Invitrogen) and the RNA was reverse transcribed with MMLV reverse transcriptase (Roche) using a dT_{16} primer (Operon). Semi-nested RT-PCR was performed with a primer to the SL1 trans-spliced leader and two nested gene-specific primers (Table S1 and Fig. S1). RT-PCR products were gel purified and sequenced.

ehn-3 alleles: Two *ehn-3* alleles have been previously described (MATHIES *et al.* 2003; MATHIES *et al.* 2004) and a third allele was identified in a PCR-based screen

(KRAEMER *et al.* 1999). The deletion allele was backcrossed ten times to remove linked mutations. All of the *ehn-3* alleles cause incompletely penetrant defects in gonadal morphology, with the most common defect being the absence of one of the two gonadal arms. Using this measure, the *ehn-3* alleles can be placed into an allelic series with *rd2* > *q689* > *q766* > wild-type. Molecularly, the alleles are predicted to affect the *ehn-3* locus differently: *ehn-3(q689)* results in a premature stop in the EHN-3A isoform, *ehn-3(q766)* removes the *ehn-3B* promoter and the first two zinc fingers, and *ehn-3(rd2)* deletes the last three exons common to both isoforms. e*hn-3(rd2)* is therefore likely to be a strong loss-of-function allele.

R08E3.4 alleles*:* Three *R08E3.4* alleles were used in this study. All alleles caused a slight developmental delay. For example, only 76% of a synchronized population of *R08E3.4(tm2127)* worms (*n*=303) reached the L4 stage ~48 hours after hatching, compared to 93% of wild-type worms (*n*=349). Based on their enhancement of the *ehn-3* mutant phenotype, the alleles can be placed into an allelic series with *tm2127 ~ ok1916 > za16*. Molecularly, *R08E3.4(tm2127)* removes the *R08E3.4B* promoter and disrupts the second to fourth zinc fingers and *R08E3.4(ok1916)* disrupts the third and fourth zinc fingers. *R08E3.4(tm2127)* is therefore predicted to be the stronger of the two alleles. The *R08E3.4(za16)* allele has a Mos1 transposon insertion near the end of the 9th exon of *R08E3.4*. Its effect on the locus is less obvious, but it consistently had a less severe enhancement of the *ehn-3* gonadal defects, suggesting that it is a weak loss-of-function allele.

C46E10.8 and *C46E10.9* alleles: We obtained deletion alleles for *C46E10.8* and *C46E10.*9 from the National BioResource Project. *C46E10.8(tm442)* and *C46E10.9(tm1692)*

are homozygous viable and they exhibit no obvious morphological or developmental defects. Molecularly, *C46E10.8(tm442)* removes exons two through four and *C46E10.9(tm1692)* removes the last three exons of the gene. Based on our RT-PCR analysis, these are the only zinc fingers encoded by *C46E10.8* and *C46E10.9*. Therefore, *tm442* and *tm1692* are likely to be null alleles.

RNAi: Double-stranded RNA was synthesized using the Megascript T7 kit (Ambion) and injected into the gonad or intestine at ~1 mg/ml. Template for RNA synthesis was derived from RNAi clones (where available) or from genomic DNA and contained at least 500bp of coding sequence. At least 10 worms were injected and 24-hour collections were taken following an overnight recovery.

Reporter Constructs

ehn-3: Additional upstream sequence was added to an existing translational reporter (MATHIES *et al.* 2004) by cloning a 4985bp XbaI fragment from cosmid ZK616 into pJK939. The resulting plasmid (pRA227) includes 2752bp upstream of the start of transcription for *ehn-3A*. The reporter was injected with pRF4 (MELLO *et al.* 1991) as a co-injection marker to make *rdEx4* and the array was crossed into *ehn-3(rd2); R08E3.4(tm2127)* to assess rescue.

R08E3.4: BC12352 was acquired from the BC Gene Expression Consortium (HUNT-NEWBURY *et al.* 2007). This transcriptional reporter contains 2958bp upstream of the start of transcription for *R08E3.4A*. Expression was seen in several tissues, but not in the somatic gonad at any stage. PCR fusion (HOBERT 2002) was used to generate an *R08E3.4::GFP* transcriptional reporter containing ~5kb upstream of *R08E3.4A*. Briefly, primers RA266 and RA246 were used to amplify a 5kb upstream fragment and primers RA245 and RA247 were

used to generate a 1.8kb fragment containing GFP and the *unc-54* 3'UTR. These fragments were fused using a semi-nested PCR with primers RA244 and RA247 and purified by gel/PCR purification (Qiagen). The A full-length *R08E3.4* reporter was generated using bacterial recombineering, essentially as described in (TURSUN *et al.* 2009). Briefly, GFP was amplified by nested PCR from pBALU1, transformed into SW105 cells containing fosmid WRM0633cF04 (Geneservice), and selected on plates containing galactose. The insertion was verified by PCR and the galK cassette was removed by inducing FLP recombinase. The resulting fosmid (pRA410) was amplified in EPI300 cells (Epicentre) and sequenced to verify the insertion site and GFP coding sequence. The reporters were injected into *unc-119(ed3)* mutants with a rescuing *unc-119* construct as a co-injection marker (MADURO and PILGRIM 1995) and integrated using gamma irradiation to create *rdIs26* and *rdIs27* (transcriptional reporter) and *rdIs25* (translational reporter).

C46E10.8, C46E10.9, and *F12E12.5*: The *C46E10.8* (pUL#JRH6H4) and *F12E12.5* (pUL#JRH6G6) reporters were acquired from the Hope lab (REECE-HOYES *et al.* 2007). A *C46E10.9* reporter (pRA294) was generated using primers RA427 and RA428, cloned into pPD95.77 (Addgene), and injected into *unc-119(ed3*) mutants to create *rdEx45*. The *F12E12.5* reporter was spontaneously integrated to create *rdIs24* and the *C46E10.8* reporter was integrated by gamma irradiation to create *rdIs29*.

Rescue Experiments

The *ehn-3* and *R08E3.4* cDNAs were amplified by RT-PCR with primers containing flanking AgeI sites and cloned downstream of the *ehn-3A* promoter and upstream of GFP to make pRA235 (*ehn-3A*) and pRA260 (*ehn-3B*) or VENUS to make pRA418 (*R08E3.4A*),

pRA309 (*R08E3.4B*) and pRA310 (*R08E3.4C*). Chimeric rescue constructs were generated by amplifying each half of the chimera using the iProof high fidelity Taq polymerase mix (Bio-Rad). Ends were phosphorylated using end conversion mix (Novagen). *ehn-3* sequences were amplified from pRA235, *R08E3.4* sequences were amplified from pRA309, and human Ikaros sequences were amplified from pCMV-SPORT6-hIkaros (Open Biosystems). Primer sequences and positions are indicated in Table S1 and Fig. S2. All constructs were sequenced to verify the integrity of the coding region and expression in SGPs was validated by GFP or VENUS fluorescence. All constructs were injected into *ehn-3(rd2); R08E3.4(tm2127)* or *ehn-3(rd2)* with pRF4 (MELLO *et al.* 1991) as a co-injection marker and at least two independent lines were scored for rescue.

BLAST searches and phylogenetic tree construction

BLAST searches were performed with the concatenated C2H2 zinc fingers of EHN-3 or R08E3.4 and reciprocal BLAST searches were performed against the *C. elegans* genome with the best hit (E-value threshold of $<10^{-16}$). Subsequent searches using the C-terminus of R08E3.4 revealed a conserved C2H2 zinc finger protein with similar zinc finger spacing to that of *hunchback* and Ikaros family members in ecdysozoan genomes. Top BLAST hits were near the C-terminus of the predicted protein and contained at least four additional C2H2 zinc fingers upstream of the C-terminal fingers. Similar searches with the C-terminal zinc fingers of Ikaros and reciprocal BLAST searches with the best hit revealed additional IFL family members in various sequenced genomes. Only predicted genes containing two C-terminal zinc fingers (CFs) with identical C2H2 spacing, and separately, four tandem zinc fingers (MFs) were used for alignments and phylogenetic reconstruction. Sequences were

aligned using Clustal X (LARKIN *et al.* 2007) with some subsequent manual adjustment (Fig. S3). The phylogenetic tree (Fig. S4) was created using the Neighbor-Joining algorithm (SAITOU and NEI 1987) and viewed using FigTree v.1.2.3 (tree.bio.ed.ac.uk/software/figtree). Chordate Ikaros family members were used to root the tree. Node values indicate the robustness over 10000 bootstrap replicates while the scale bar indicates an evolutionary distance of 0.05 amino acid substitutions per site.

The following genomes and sites were used for alignments and phylogenetic tree construction: *Drosophila melanogaster* (Dme), *Aedes aegypti* (Aae), *Acyrthosiphon pisum* (Api), *Tribolium castaneum* (Tca), and *Ixodes scapularis* (Isc); (www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=insects), *Caenorhabditis elegans* (Cel) and *Caenorhabditis briggsae* (Cbr); (www.wormbase.org), *Pristionchus pacificus* (Ppa); (www.pristionchus.org), *Helobdella robusta* (Hro), *Capitella capitata* (Cca), *Lottia gigantea* (Lgi); (genome.jgi-psf.org), *Schistosoma mansoni* (Sma); (www.sanger.ac.uk/Projects/Helminths), *Schmidtea mediterranea* (Sme); (smedgd.neuro.utah.edu/index.html) (ROBB *et al.* 2008), *Homo sapiens* (Hsa), *Danio rerio* (Dre); (www.uniprot.org/uniprot), *Strongylocentrotus purpuratus* (Spu); (www.spbase.org/SpBase), *Ciona intestinalis* (Cin); (genome.jgi-psf.org). Preliminary sequence data for the hemichordate *Saccoglossus kowalevskii* (Sko) was obtained from the Baylor College of Medicine Human Genome Sequencing Center website at http://www.hgsc.bcm.tmc.edu. In some cases an *ab initio* gene prediction program, such as FGENESH (SALAMOV and SOLOVYEV 2000), was used to find additional zinc fingers with a subsequent six-frame translation for verification. The following proteins were used in this

study (genbank identifier or respective scaffold/contig): Hsa Ikaras (17390815), Hsa Eos (122891864), Dre Ikaros (220678854), Dre Eos (189536069), Cel hbl1 (115533442), Cel HIL1(71995006), Cel ehn3(71999776), Cel sdz12(17533143), Cbr hbl1 (157769927), Cbr HIL1 (157769619), Dme hb (24645127), Dme HIL1(28573961), Aae hb (157111582), Aae HIL1 (157108862), Api hb (242117900), Api HIL1 (193606023), Isc hb (241701224), Isc HIL1(241053661), Tca hb (113206058), Tca HIL1 (91081013), Cin IFL1 (118344106), Sma hb (256075149), Cca hb [scaffold 123 (protein ID: 199858)], Cca HIL1 [scaffold 772 (protein ID: 209625)], Ppa hbl1 (Contig8), Ppa HIL1 (Contig101), Sko IFL1 (Contig65674), Spu IFL1 (SPU_011260), Sme hb (mk4.008217.00.01), Lgi hb [Scaffold 67 (protein ID: 167538)], Hro hb1 [scaffold 12 (protein ID: 167650)], Hro hb2 [scaffold 78 (protein ID: 116286)], Hro hb3 [scaffold 37 (protein ID: 176361)], Hro HIL1 [scaffold 8 (similar to protein ID: 131184)]. The following sites were also examined for Ikaros, *hunchback*, and HIL family members: *Trichoplax adhaerens* (genome.jgi-psf.org/Triad1/Triad1.home.html); Cnidarians: *Nematostella vectensis* (genome.jgi-psf.org/Nemve1/Nemve1.home.html), all cnidarians (cnidbase.bu.edu/blast/blast_cs.html); Hydra: *Hydra.magnipapillata* (hydrazome.metazome.net) and (blast.genome.jp/).

RESULTS

A *hunchback*/Ikaros-like gene family in *C. elegans*

Our previous work demonstrated that *ehn-3* controls important aspects of early SGP development, including cell division and survival (MATHIES *et al.* 2003; MATHIES *et al.* 2004). The *ehn-3* gene encodes proteins with distinct C2H2 zinc finger domains: an N-terminal domain with up to four zinc fingers and a C-terminal domain with two zinc fingers. We searched for additional genes similar to *ehn-3* and identified the *C. elegans* gene *R08E3.4/ztf-16* as the best BLAST hit. The *R08E3.4* gene encodes proteins with up to eight zinc fingers: one near the N-terminus, five in the middle, and two at the very C-terminus of the protein. BLAST searches with the conserved zinc fingers of EHN-3 or R08E3.4 identified one additional region in the *C. elegans* genome with significant sequence similarity to *ehn-3* and *R08E3.4* (E-value $<10^{-8}$). This genomic region has a tandem array of three predicted genes, *F12E12.5/sdz-12*, *C46E10.8*, and *C46E10.9,* all of which potentially encode proteins with two zinc finger domains. This arrangement of zinc fingers is similar to mammalian Ikaros and *Drosophila hunchback*. We therefore, refer to this gene family as *hunchback/Ikaros*-like (HIL).

All five of the HIL genes are expressed, as evidenced by RT-PCR products containing the SL1 trans-spliced leader (BLUMENTHAL and STEWARD 1997) and spliced introns (see Materials and Methods). *R08E3.4* has the most compelling gene structure similarity with *ehn-3*, including introns at identical positions within zinc fingers and alternative promoters in homologous positions (Fig. 1A). *F12E12.5* shares one homologous

intron with *ehn-3* and *R08E3.4*. By contrast, *C46E10.8* and *C46E10.9* appear to have undergone significant changes, including the loss of zinc finger encoding exons through the accumulation of stop codons or chromosomal rearrangements (Fig. S1). Of the HIL genes, only *R08E3.4* has an obvious ortholog in other ecdysozoan genomes as assayed by reciprocal best BLAST hit. In *Drosophila*, this is the uncharacterized gene CG12769. Together, these observations suggest that *R08E3.4* is the ancestral gene in the family and the other genes arose via gene duplication.

Expression and function of the HIL genes in the developing gonad

Given the previously defined function of *ehn-3* in gonadogenesis (MATHIES *et al.* 2003; MATHIES *et al.* 2004), we asked if any of the other HIL genes are expressed in the reproductive system using gene reporters. For this initial survey, we generated transcriptional reporters that included 3-5kb upstream of the start of translation or to the next upstream gene and we created multiple reporters for genes with more than one promoter. We previously described an *ehn-3* reporter containing 3kb upstream of the *ehn-3A* promoter (WELCHMAN *et al.* 2007). This reporter was expressed in SGPs from shortly after their birth (Fig. 1B) through their divisions in the first larval stage (Fig. 1C, D). We generated an *R08E3.4* reporter containing 5kb upstream of the start of transcription for *R08E3.4A*. This reporter was expressed in the somatic gonad, first in the SGPs during embryogenesis (Fig. 1E) and then during the L2 stage in the descendants of Z1.a and Z4.p, which are the distal tip cells and their sisters (Fig. 1F). In the L4 stage, expression of the reporter was observed in the spermatheca (not shown). The *R08E3.4A* reporter was also expressed in a variety of non-

gonadal tissues, including the hypodermis and unidentified cells in the head and tail. We obtained an *F12E12.5* reporter containing 479bp upstream of the start of transcription (REECE-HOYES *et al.* 2007) and integrated it into the genome. This reporter was expressed at low levels in the early somatic gonadal lineage. Prominent expression was first seen in Z1.pa and Z4.ap (Fig. 1G), and their descendants (Fig. 1H). In the L3 stage, the *F12E12.5* reporter was expressed in ventral and dorsal uterine precursors (not shown). Despite the previous identification of F12E12.5/*sdz-12* as a SKN-1 dependent zygotic transcript in early *C. elegans* embryos (ROBERTSON *et al.* 2004), reporter expression was not seen in early embryos, perhaps due to the limited sequences used to generate the reporter. Finally, a *C46E10.8* reporter was expressed in the dorsal and ventral nerve cord and unidentified neurons in the head and tail and a *C46E10.9* reporter was expressed in unidentified cells running parallel to the pharynx (not shown). We did not, however, detect expression of these reporters in the somatic gonad at any stage. Our expression analysis suggests that *F12E12.5* and *R08E3.4,* but not *C46E10.8* and *C46E10.9*, are expressed in the somatic gonad and might have overlapping functions with *ehn-3*.

We asked if any of the other HIL genes function in reproductive system development using a combination of RNAi and genetic mutants. For this analysis, we isolated a new deletion allele of *ehn-3* (Materials and Methods) and obtained deletion alleles for *R08E3.4, C46E10.9,* and *C46E10.8* (Fig. 1A; Fig. S1). Because no mutants are currently available for *F12E12.5* we used RNAi to examine this gene. We scored gonadal morphology using a dissecting microscope and report the overall penetrance of defects (Table 1; Table S2). All of the *ehn-3* alleles caused incompletely penetrant defects in gonadal morphology, with the

most common defect being the absence of one of the two gonadal arms. None of the other HIL genes, when inactivated by either genetic mutants or RNAi, resulted in missing gonadal arms or showed any obvious morphological defects. Next, we examined all double mutant combinations with *ehn-3* to ask if any of the HIL genes has overlapping functions in the somatic gonad. Only *R08E3.4* mutants and RNAi enhanced the penetrance of the *ehn-3* mutant phenotype. This enhancement was seen with all *ehn-3* and *R08E3.4* alleles to varying degrees (Table 1, Table S2). Finally, we looked for defects in triple, quadruple, and quintuple mutants and found that none of the other HIL family members enhanced the *ehn-3; R08E3.4* phenotype further (Table 1). Therefore, the functional analysis indicates that *ehn-3* and *R08E3.*4 have overlapping functions during gonadogenesis.

To ascertain that our reporters accurately reflect the expression of *ehn-3* and *R08E3.4* and to demonstrate that disruptions in these genes cause their mutant phenotypes, we generated translational reporters and tested them for rescue. The previously described EHN-3::GFP reporter contained only 429 bp upstream of the start of transcription for *ehn-3A* (MATHIES *et al.* 2004). We generated a second EHN-3::GFP reporter containing additional upstream sequence and found that, like the previously described reporter, it was expressed in SGPs during embryogenesis and was almost undetectable in L1 larvae (not shown). Importantly, this translational reporter rescued *ehn-3(rd2); R08E3.4(tm2127)* mutants to fertile two-armed adults (99%, n=107), suggesting that it is expressed in all cells that require *ehn-3* for their development. Because *ehn-3* translational reporters are only expressed during embryogenesis, whereas transcriptional reporters containing the same upstream sequence were expressed through the L1 larval stage (Fig. 1C, D), it is likely that the *ehn-3* locus is

regulated post-transcriptionally. Similarly, we generated translational reporters for *R08E3.4*. Since the *R08E3.4* locus spans at least 10kb and contains three promoters, we used homologous recombination in *E. coli* to generate a full-length translational fusion (DOLPHIN and HOPE 2006; YU *et al.* 2000). The *R08E3.4* translational reporter was expressed in the somatic gonad in an identical spatial and temporal pattern as the transcriptional reporter and it partially rescues *ehn-3(rd2); R08E3.4(tm2127)* double mutants (from 62% to 27% defects, *n*=179). This penetrance is similar to *ehn-3(rd2)* single mutants (Table 1), suggesting that the reporter is expressed in all cells that require *R08E3.4* in an *ehn-3* mutant background. Therefore, *ehn-3* and *R08E3.4* are both expressed in SGPs during embryogenesis, while *R08E3.4* alone is expressed in the distal somatic gonadal lineages (Fig. 1I). Taken together, our functional and expression analyses suggest that *ehn-3* and *R08E3.4* have overlapping functions in SGPs.

ehn-3 and _R08E3.4_ function in SGPs to control the development of differentiated gonadal cell types

To explore the functional overlap between *ehn-3* and *R08E3.4*, we examined single and double mutants using molecular markers and DIC microscopy. Unless otherwise stated, we used the strongest alleles of *ehn-3* and *R08E3.4*, which are *rd2* and *tm2127*, respectively (Materials and Methods). Wild-type hermaphrodites have two gonadal arms, each developing from one of the two SGPs. Similarly, *R08E3.4* mutants had two gonadal arms (*n*=160). By contrast, *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants were often missing one or both of the gonadal arms (Fig. 2) and the penetrance of these defects was

significantly higher in *ehn-3; R08E3.4* double mutants. In addition to these gross morphological defects, we also observed ectopic germ line proliferation in *ehn-3* and *ehn-3; R08E3.4* mutants. Proliferation of the germ line is controlled by the distal tip cells (DTCs) and is normally confined to the distal end of the gonadal arms (KIMBLE and WHITE 1981). By contrast, we sometimes observed germ line proliferation near the center of the gonad in *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants (5.8%, *n*=378 and 32.7%, *n*=211, respectively). This proximal germ line proliferation was only seen in gonads where one (Fig. 2C) or both gonadal arms were missing. Finally, *ehn-3, R08E3.4,* and *ehn-3; R08E3.4* mutants sometimes had abnormal DTC migration defects. Normally, each DTC migrates in a U-shaped pattern to generate one of the two reflexed gonadal arms (KIMBLE and WHITE 1981). In *ehn-3* and *R08E3.4* single mutants, one or both of the DTCs had abnormal migration patterns (11%, *n*=109 and 4.4%, *n*=160, respectively) and the penetrance of these defects increased in *ehn-3; R08E3.4* double mutants (34.9%, *n*=126). We conclude that *ehn-3* and *R08E3.4* have largely redundant functions in the development and migration of the gonadal arms.

The missing gonadal arms in *ehn*-3 and *ehn-3; R08E3.4* mutants suggested a defect early in gonadogenesis. To explore the cause of this defect, we began with an *ehn-3* transcriptional reporter to visualize the SGPs both shortly after their birth and in newly hatched L1 larvae. Most *ehn-3* and *ehn-3; R08E3.4* mutants had two SGPs in embryos (56/56 and 44/44, respectively) and L1 larvae (140/141 and 50/50, respectively; Fig. 3). Therefore, the missing gonadal arms do not result from missing SGPs. However, *ehn-3* and *ehn-3; R08E3.4* mutants often had SGPs with abnormal morphology or position within the

gonadal primordium. Normally, the two SGPs occupy the poles of the four-celled primordium and make contact with one another on its ventral side. By contrast, *ehn-3* and *ehn-3; R08E3.4* mutants had a variety of SGP configurations, which we grouped into four major classes: 1) normal morphology, ventral contact (Fig. 3A), 2) normal morphology, abnormal contact (Fig. 3B, C), 3) abnormal position within the primordium (Fig. 3D), and 4) abnormal morphology (Fig. 3E). The *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants had the same array of phenotypes, but the spectrum was shifted toward SGPs with abnormal morphology or position in the double mutants (Fig. 3F). Therefore, *ehn-3* and *R08E3.4* are not required for SGP specification, but they are necessary for proper gonadal primordium assembly and SGP morphology.

To learn which SGP defects correlate with adult gonadal defects, we examined individuals at the L1 larval stage, allowed them to develop, and examined them again at the L4 stage (Table 2). In both *ehn-3* and *ehn-3; R08E3.4* mutants, SGPs with abnormal position or morphology often developed into adults that were missing the corresponding gonadal arm. Therefore, these abnormal SGPs typically do not execute their developmental program properly. Interestingly, over half of the L1s with normal-looking SGPs developed into adults with gonadogenesis defects in *ehn-3; R08E3.4* double mutants, as compared to 5% for *ehn-3* single mutants. We conclude that *ehn-3* plays a larger role in the early development of SGPs based upon the low percentage of normal-looking SGPs that go on to form abnormal gonads, while *ehn-3* and *R08E3.4* work in parallel to control subsequent SGP development.

The elongation of each hermaphrodite gonadal arm is led by one DTC (HEDGECOCK *et al.* 1987; KIMBLE and HIRSH 1979). Therefore, we reasoned that the missing gonadal arms

might be due to missing or non-functional DTCs. To test this possibility, we examined *ehn*-3 and *ehn-3; R08E3.4* mutants with *lag-2::GFP*, which marks DTCs (BLELLOCH *et al.* 1999) and found that they are often missing one or both DTCs (Table 3). Since *ehn-3; R08E3.4* mutants almost always had SGPs, this suggests a defect in the lineage leading from the SGPs to DTCs. Normally, each SGP divides asymmetrically to yield one cell that generates a DTC and one cell with the potential to generate an anchor cell (AC) (Fig. 1I). Therefore, one possible cause for missing DTCs is a failure in the asymmetric division of the SGPs, which can lead to extra ACs at the expense of DTCs (MISKOWSKI *et al.* 2001; SIEGFRIED *et al.* 2004; SIEGFRIED and KIMBLE 2002; STERNBERG and HORVITZ 1988). To test this, we used the AC marker *cdh-3::CFP* (INOUE *et al.* 2002) in combination with the DTC marker *lag-2::YFP*. Wild-type worms had a single anchor cell at the L3 stage, whereas *ehn-3* mutants and *ehn-3; R08E3.4(RNAi)* double mutants occasionally had two ACs at a similar stage as seen by both morphology and *cdh-3::CFP* expression (11.2%, *n*=143 and 16.3%, *n*=80, respectively). However, we did not see any correlation between extra ACs and missing DTCs in *ehn-3; R08E3.4(RNAi)* (11/13 had two DTCs) suggesting that extra AC production was not caused by a failure of SGPs to divide asymmetrically. Furthermore, the two ACs were frequently located adjacent to one another. Therefore, it is more likely that the extra ACs in *ehn-3; R08E3.4(RNAi)* result from defects in lateral signaling between the AC/VU precursors (Z1.ppp and Z4.aaa) (KIMBLE 1981). This contact-dependent LIN-12/*Notch* signaling event insures that only one AC is generated (GREENWALD *et al.* 1983) and it normally prevents the development of adjacent ACs. We conclude that the missing DTCs are unlikely to result from defects in the asymmetric division of the SGPs. Instead, the

missing gonadal arms might result from defects in specification, differentiation, or survival of the DTCs. Future analysis of *ehn-3* and *R08E3.4* should shed light on these processes.

In the course of our phenotypic analysis, we observed *lag-2*-expressing cells that did not appear to be typical hermaphrodite DTCs (hDTCs). Instead, they had the morphology of male DTCs (mDTCs). The hDTCs are crescent-shaped cells that lead the elongation of the two hermaphrodite gonadal arms and induce germ line proliferation, whereas the mDTCs are small cells that remain situated at the distal end of the single male gonadal arm, where they induce germ line proliferation (KIMBLE and HIRSH 1979; KIMBLE and WHITE 1981). This observation was intriguing because *ehn-3* mutants have a weak sexual transformation of the somatic gonad that is enhanced by removing one copy of the *tra-1* gene (MATHIES *et al.* 2004). Therefore, *R08E3.4* may redundantly control gonadal sex with *ehn-3*. To distinguish between mDTCs and hDTCs, we sought a marker that is differentially expressed between these cell types. Expression of *arg-1::GFP* was previously reported in the hermaphrodite head mesodermal cell, vulval muscles, and the enteric muscles (KOSTAS and FIRE 2002; ZHAO *et al.* 2007). We noticed that this reporter is also expressed in the mDTC, but not the hDTC (Fig. 4A, B). We examined *arg-1::GFP* in the hermaphrodite somatic gonad from the L1 to L4 stage in wild type and *ehn-3; R08E3.4* double mutants. In wild type somatic gonads, no expression was observed in L1 or L2 larvae and only faint expression was seen in hDTCs in L3 larvae (Fig. 4D). By comparison, we saw expression of *arg-1::GFP* in the somatic gonad of *ehn-3; R08E3.4* double mutants beginning in the L2 stage and peaking at the L3 stage with 13.6% expressing *arg-1::GFP* in the hDTCs and 16.5% expressing *arg-1::GFP* in small cells near the center of the gonad (Fig. 4C, D). These small cells were

morphologically similar to mDTCs, they did not lead gonadal arm migration, and they were always associated with proximal germ line proliferation (*n*=33). Together, these observations suggest that the hDTCs might be sexually transformed into mDTCs in *ehn-3; R08E3.4* double mutants.

In order to determine if the ectopic *arg-1::GFP* expression was indicative of a broad sexual transformation of the hermaphrodite somatic gonad, we used the male-specific reporter *K09C8.2::GFP* (CHANG *et al.* 2004). This marker is only expressed in the male seminal vesicle and a population of vas deferens cells without any expression observed in hermaphrodites. We used RNAi of *R08E3.4* in an *ehn-3* background to examine expression of *K09C8.2::GFP* from the L1 to L4 stage in hermaphrodites. No expression of *K09C8.2::GFP* was observed in *ehn-3(rd2); R08E3.4(RNAi)* hermaphrodites (*n*=601), leading us to conclude that this is not a broad sexual transformation of the gonad and, instead, appears to be restricted to the DTCs.

To ask how broadly *ehn-3* and *R08E3.4* function in the somatic gonadal lineage, we used *fkh-6::GFP*, which marks spermathecal and sheath cells at the L4 stage. Wild-type worms express *fkh-6::GFP* brightly in the spermatheca and more dimly in the sheath cells that surround the proximal two-thirds of each gonadal arm (CHANG *et al.* 2004). Using a fluorescent dissecting microscope, we observed *fkh-6::GFP* in two patches flanking the vulva, corresponding to the two spermathecae (*n*=125). By contrast, *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants sometimes expressed *fkh-6::GFP* in only one patch of spermathecal tissue (7.4%, *n*=258 and 18%, *n*=108, respectively). This patch of *fkh-6::GFP* expression was typically the same size or smaller than a single spermathecum in wild-type

worms (not shown). Therefore, *ehn-3* and *R08E3.4* are required for proper development of the spermathecae, suggesting a defect in one or both of the lineages (SS, Fig. 1I) that give rise to this tissue.

In summary, we find that *ehn-3* and *R08E3.4* function broadly in somatic gonad development and they largely have overlapping functions. In L1 larvae, *ehn-3; R08E3.4* mutants exhibited incompletely penetrant defects in the morphology, placement, and development of the SGPs. At the L4 stage, they had defects in the number or morphology of DTCs, ACs, and spermathecal cells. Generally, *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants had the same array of defects, but the penetrance was much higher in *ehn-3; R08E3.4* double mutants. Together, our phenotypic analysis suggests that *ehn-3* and *R08E3.4* are required for the development of several differentiated tissues in the somatic gonad and they show synergistic interactions consistent with genetic redundancy.

The EHN-3 and R08E3.4 isoforms have different activities

To begin to dissect the functions of the different *ehn-3* and *R08E3.4* isoforms, we used the *ehn-3* promoter to drive expression of each *ehn-3* and *R08E3.4* cDNA and tested them for rescue of *ehn-3; R08E3.4* double mutants. All constructs fused GFP to the C-terminus of the protein and at least two lines were examined for each construct. For these experiments we assayed missing gonadal arms and report the overall penetrance of gonadal defects. We found that EHN-3A::GFP rescues the gonadogenesis defects of *ehn-3; R08E3.4* double mutants almost completely (Table 4). By contrast, none of the other EHN-3 or R08E3.4 isoforms was capable of fully rescuing *ehn-3; R08E3.4* double mutants. Therefore,

EHN-3A is sufficient to control SGP development. Of the other isoforms, R08E3.4A and R08E3.4B had the best rescuing activity. We observed better rescue when the *R08E3.4* isoforms were expressed from the *ehn-3* promoter than when they were expressed in the context of the *R08E3.4* translational reporter. This difference might be due to differences in expression level or timing of the *ehn-3* and *R08E3.4* promoters. To ask if any of the R08E3.4 proteins could substitute for EHN-3, we tested the R08E3.4 isoforms for rescue of *ehn-3* single mutants. We found that R08E3.4A and R08E3.4B partially rescued *ehn-3* single mutants when driven by the *ehn-3* promoter, suggesting that both isoforms retain some functional similarity with EHN-3A. Importantly, neither *R08E3.4A* nor *R08E3.4B* can completely replace *ehn-3A* in these rescue experiments, suggesting that some functional divergence has occurred between the homologous EHN-3A and R08E3.4B proteins.

EHN-3 and R08E3.4 have functionally distinct zinc finger domains

Ikaros proteins have an N-terminal set of DNA binding zinc fingers and a C-terminal pair of zinc fingers that mediate protein-protein interactions. To ask if EHN-3 and R08E3.4 also have functionally distinct zinc finger domains, we created chimeric proteins containing different combinations of the EHN-3A and R08E3.4B zinc finger domains and tested them for rescue of *ehn-3; R08E3.4* double mutants (Fig. 5). Since R08E3.4A and R08E3.4B showed equivalent rescuing activity in our previous experiments, we used the R08E3.4B isoform for these experiments. First, we swapped the N-terminal zinc fingers. We found that the N-terminal zinc fingers of EHN-3A were sufficient to give R08E3.4B the ability to rescue *ehn-3; R08E3.4* double mutants (Fig. 5), indicating that this domain is responsible for

the functional differences between EHN-3A and R08E3.4B. Reciprocally, the N-terminal zinc fingers of R08E3.4B reduced the ability of EHN-3A to rescue *ehn-3; R08E3.4* double mutants. This chimeric protein rescued better than R08E3.4B, suggesting that other regions of the EHN-3A protein are also important in conferring specific functions to EHN-3. We deleted the N-terminal zinc fingers from EHN-3A and found that this significantly reduced its ability to rescue *ehn-3; R08E3.4* mutants. Next, we replaced the C-terminal zinc fingers of *ehn-3* with those of *R08E3.4* and human Ikaros and found that these chimeric proteins rescued the *ehn-3; R08E3.4* defects as efficiently as wild-type EHN-3A. This result supports the idea that the EHN-3 and R08E3.4 C-terminal zinc fingers function in a manner similar to the C-terminal zinc fingers of Ikaros and may also mediate dimerization. Taken together, these results are consistent with the model that EHN-3 and R08E3.4 have functionally distinct zinc finger domains and that their N-terminal zinc fingers largely determine the functional specificity of the proteins.

To more directly test the idea that EHN-3 and Ikaros are homologous proteins, we used human Ikaros as the basis for a chimeric rescue experiment. We reasoned that the putative DNA binding domain of EHN-3 might be sufficient to provide functionality to human Ikaros when expressed in *C. elegans*. We generated a construct containing the four N-terminal zinc fingers of EHN-3A and the C-terminus of human Ikaros and expressed this chimeric protein in SGPs using the *ehn-3A* promoter. This heterologous protein was able to rescue *ehn-3* single mutants and *ehn-3; R08E3.4* double mutants as well as EHN-3A (Fig. 5). By comparison, the N-terminal zinc fingers of EHN-3A only partially rescue *ehn-3; R08E3.4* double mutants, suggesting that the C-terminus of human Ikaros is providing functionality to

this chimeric protein. This reinforces our previous results suggesting that the specificity of EHN-3 resides in its N-terminal putative DNA-binding zinc fingers and indicates that *ehn-3*, *R08E3.4*, and Ikaros encode functionally equivalent C-terminal domains.

A new family of *hunchback* and Ikaros-like genes in protostomes

Our chimeric rescue experiments suggest that the *C. elegans* HIL genes could be homologous to Ikaros and it has been previously suggested that Ikaros and *hunchback* might share a common ancestor. Therefore, to explore the relationship between the HIL genes and *hunchback*, we searched for orthologs of both genes in several sequenced protostome genomes. We used the C-terminal zinc fingers of *hunchback* or *R08E3.4* to identify pairs of zinc fingers near the C-terminus of predicted proteins, then we searched for four tandem zinc fingers encoded upstream of these zinc fingers. Using this method, we identified candidate *hunchback* and HIL genes in representative ecdysozoan genomes (Fig. 6). All of the HIL and *hunchback* genes encode four conserved zinc fingers, which we refer to as middle fingers (MF1-4) using the nomenclature that was established for *hunchback,* and two conserved C-terminal fingers (Fig. 6B, C). For comparison, we show the homologous zinc fingers of several chordate Ikaros family genes (Fig. 6D). We also found *hunchback* and HIL genes in the sequenced lophotrochozoan genomes, including previously unreported *hunchback* genes in the leech, mollusk, and flatworm genomes and HIL genes in the annelid and mollusk genomes (Fig. S3). To establish the relationship among these genes, we used the six conserved zinc fingers (MF1-4 and CF1-2) to construct a phylogenetic tree using the Neighbor-Joining algorithm (SAITOU and NEI 1987). As an outgroup, we used chordate

Ikaros family members. The resulting phylogram (Fig. S4) provides strong support for distinct *hunchback* and HIL clades in protostomes. The presence of distinct *hunchback* and HIL gene sequences in a wide array of sequenced protostome genomes suggests that both gene families existed early in protostome evolution.

In the course of our analysis, we noticed that several HIL genes are predicted to encode one or two N-terminal zinc fingers (NF), as has been reported for *hunchback* genes. For example, the nematode *R08E3.4/ztf-16* genes all encode two additional zinc fingers N-terminal to the four MFs (NF1-2) and several arthropod and annelid HILs have one NF (Fig. S5 and S6). A close inspection of basal deuterostome Ikaros family-like (IFL) members revealed an additional NF in the previously described sea urchin Ikaros ortholog (HIBINO *et al.* 2006) and two NFs in the uncharacterized hemichordate acorn worm Ikaros ortholog (Fig. S6). Therefore IFLs in basal deuterostomes may contain additional zinc finger domains, similar to the ancestral *hunchback* (PATEL *et al.* 2001; PINNELL *et al.* 2006) and nematode HIL genes (this work), reinforcing the idea that Ikaros, *hunchback*, and HIL genes are derived from a common ancestor. To explore the roots of these genes outside the bilateria, we searched for IFL, *hunchback*, and HIL genes in the sequenced cnidarian genomes, including *Nematostella vectensis, Hydra magnipapillata*, and *Trichoplax adhaerens,* and were unable to find any sequences encoding a similar zinc finger arrangement and spacing of the C-terminal histidines. Therefore, the ancestor to *hunchback,* Ikaros, and the HIL genes may have arisen in the lineage leading to bilaterians.

DISCUSSION

In this paper, we identify and characterize a family of *C. elegans* zinc finger genes with a similar arrangement of C2H2 zinc fingers as mammalian Ikaros and *Drosophila hunchback*. The *C. elegans* gene family consists of five genes, three of which are expressed in reproductive system lineages and two that function early in the development of the reproductive system. We find that *ehn-3* and *R08E3.4* function in SGPs to control a variety of cell fate decisions during somatic gonad development. Using a combination of domain swaps and rescue experiments, we provide support for the idea that *ehn-3* and *R08E3.4* encode distinct protein-protein interaction and DNA binding zinc finger domains, as has been shown for mammalian Ikaros family members and *Drosophila hunchback* (MCCARTY *et al.* 2003). We also show that human Ikaros can function in the *C. elegans* reproductive system if the putative DNA binding domain of EHN-3 is used in place of its DNA binding domain. Taken together with the protein structure and sequence similarities between mammalian Ikaros, *Drosophila hunchback*, and the *C. elegans* HIL genes, we suggest that these gene families shared a common origin prior to the divergence of protostomes and deuterostomes.

Pleiotropic regulators of somatic gonad development

C. elegans HIL genes are important for the development of at least three of the five differentiated somatic gonadal tissues. Our phenotypic analysis suggests that *ehn-3* and *R08E3.4* control diverse processes ranging from sexual differentiation to lateral signaling and our rescue experiments demonstrate that they are capable of rescuing these defects when expressed only in SGPs. Finally, our reporter analysis indicates that *ehn-3* and *R08E3.4* have

overlapping expression in embryonic SGPs. Therefore, it is likely that *ehn-3* and *R08E3.4* act in SGPs to control their subsequent development into differentiated cell types. How might the HIL genes function in SGPs to control processes that occur much later in development? One idea comes from the sequence, structural, and functional similarities between the HIL genes and Ikaros. Mammalian Ikaros is important for several aspects of hematopoietic system development. A precise developmental role for Ikaros has been difficult to define, owing in part to its diverse functions and action at multiple stages of development (reviewed in SMALE and DORSHKIND 2006). Similarly, we have shown that the *C. elegans* HIL genes control reproductive system development at multiple times and in a broad range of tissues. Ikaros is involved in the activation and repression of genes in conjunction with chromatin remodeling complexes such as NuRD and SWI/SNF, with which it physically interacts (KIM *et al.* 1999; KOIPALLY *et al.* 1999; SRIDHARAN and SMALE 2007). It has been suggested that the interaction between Ikaros and NuRD, might be important for "priming" genes for expression later in hematopoietic system development (NG *et al.* 2007; NG *et al.* 2009). This epigenetic regulation is thought to be important for the potential of hematopoietic stems cells and their progenitors to generate multiple lymphoid fates. Our domain swap experiments demonstrate that the C-terminus of human Ikaros can replace the C-terminus of EHN-3, suggesting that EHN-3 might utilize a similar molecular mechanism to control reproductive system development. These developmental and molecular similarities lead us to propose that HIL genes are acting in SGPs to establish a permissive chromatin state surrounding genes that will be utilized later in somatic gonad development. Tests of this hypothesis must await the identification of EHN-3 and R08E3.4 target genes.

Duplication and divergence of the HIL family in nematodes

The sequenced *Caenorhabditis* genomes each encode at least one HIL family member and they all contain a highly conserved ortholog of *R08E3.4/ztf-16*. In *C. elegans, R08E3.4* appears to have duplicated to give rise to at least four HIL genes. Of these, *ehn-3* and *R08E3.4* are both expressed in SGPs, have strong synergistic genetic interactions, and group together within the protostome HIL clade in our phylogenetic analysis. The other HIL genes, however, may not function in gonadogenesis. The only gonadal phenotype we observed for *R08E3.4* single mutants is an abnormal migration pattern of the DTCs, a phenotype that is shared by *ehn-3* single mutants. In all other aspects of somatic gonadal development, *R08E3.4* strongly enhanced the defects observed in *ehn-3* single mutants. Therefore, *ehn-3* and *R08E3.4* appear to retain largely overlapping functions, with somatic gonad development being more reliant on *ehn-3* than *R08E3.4*.

Despite the strong functional redundancy between *ehn-3* and *R08E3.4*, other observations suggest that *ehn-3* has acquired new functions since its divergence from *R08E3.4*. First, R08E3.4 was only partially capable of replacing EHN-3 in early SGP development. These functional differences appear to lie predominantly in the four middle zinc fingers (MF1-4), since this domain was sufficient to convey full rescuing activity to R08E3.4. Furthermore, *ehn-3* interacts genetically with the conserved transcription factors, *tra-1*/GLI and *hnd-1*/dHand (MATHIES *et al.* 2003; MATHIES *et al.* 2004), neither of which interacts with *R08E3.4* (our unpublished observation). Finally, *ehn-3* and *R08E3.4* have low levels of amino acid sequence identity, even within their conserved MF domains where they share only six of the twelve amino acids that are predicted to interact with DNA (WOLFE *et*

al. 2000). How do we resolve this apparent paradox of low sequence similarity and strong functional redundancy? We noticed that four of the six differences between EHN-3 and R08E3.4 lie in their third and fourth MFs, suggesting that MF1-2 might be under different selective pressure than MF3-4. Interestingly, EHN-3 and R08E3.4 isoforms have variable numbers of MFs and all isoforms that were capable of rescuing the double mutant phenotype contained the first two MFs. Therefore, the EHN-3 and R08E3.4 isoforms might bind distinct DNA sequences using different pairs of MFs. An analysis of metazoan zinc finger genes encoding more than two tandem C2H2 zinc fingers, the poly-zinc-finger (poly-ZF) genes, suggests that positive selection frequently acts on gene duplicates to change their DNA binding specificity (EMERSON and THOMAS 2009). These adaptive changes would presumably come at the expense of redundancy, since they result in the regulation of different target genes. However, if poly-ZF proteins can utilize different subsets of zinc fingers to bind distinct target genes, this might allow for the asymmetric evolution of different zinc finger domains. This, in turn, could explain the simultaneous retention of overlapping functions by *ehn-3* and *R08E3.4* and evolution of new functions by *ehn-3*.

It is likely that more *hunchback/Ikaros*-like genes will be found in both protostomes and deuterostomes. For example, other *Caenorhabditis* genomes have HIL duplications that are not syntenic with the genes we analyzed here and do not appear to be orthologs of any of the *C. elegans* HILs. In addition, a *hunchback* duplication was previously described in leeches (SAVAGE and SHANKLAND 1996) and we uncovered an additional *hunchback* duplication event in leeches (Fig. S3). Therefore, this gene family has likely undergone duplication multiple times in protostome evolution. The Ikaros family has successfully

duplicated three times in the course of vertebrate evolution (JOHN *et al.* 2009). Some of these duplicates are intimately linked with the evolution of the adaptive immune system, which undoubtedly gave vertebrates an advantage in the battle against pathogens. Poly-ZF genes have undergone extensive expansion in metazoan genomes (EMERSON and THOMAS 2009). Perhaps the evolutionary versatility of poly-ZF proteins, as discussed above, has favored their retention in the genome following gene duplication.

A common origin for *hunchback*, Ikaros, and HIL transcription factors?

We identified a second gene family in protostomes with significant gene structure and sequence similarity to *hunchback* and Ikaros. We found at least one HIL and one *hunchback* gene in most protostome genomes with the exception of two flatworm genomes, which appear to contain only *hunchback*. The *hunchback* and HIL genes form distinct clades in our phylogenetic analysis, indicating that both *hunchback* and HIL genes were likely to be present in the ancestor to lophotrochozoans and ecdysozoans. Members of both gene families contain zinc fingers N-terminal to the conserved MFs, whereas a single zinc finger between the conserved MFs and CFs (the ExF) appears to be unique to *hunchback* genes. Furthermore, we find evidence for sea urchin and acorn worm IFLs with zinc fingers N-terminal to the conserved MFs. This gene structure is more typical of the protostome *hunchback* and HIL genes, lending support to the idea that these three gene families may have shared a common origin in the ancestor to protostomes and deuterostomes.

Ikaros family-like (IFL) genes have been well characterized in deuterostomes (JOHN *et al.* 2009 and references therein) and *hunchback* genes have been well characterized in

protostomes (KERNER *et al.* 2006; PINNELL *et al.* 2006 and references therein). Several groups have suggested that Ikaros and *hunchback* might share an evolutionary history, yet an ancestral role for these genes has not been proposed. Vertebrate Ikaros family members, Ikaros, Aiolos, and Helios, are known for their primary expression and function in the adaptive immune system (reviewed in GEORGOPOULOS 2002). The other family members are more broadly expressed and are therefore likely to have major roles outside immune system development (HONMA *et al.* 1999; PERDOMO *et al.* 2000). *hunchback,* on the other hand, was identified for its function in anterior-posterior patterning of the *Drosophila* embryo, where it acts as a "gap" class segmentation gene (LEHMANN and NUSSLEIN-VOLHARD 1987; ST JOHNSTON and NÜSSLEIN-VOLHARD 1992). *hunchback* appears to function broadly in insect segmentation (LIU and KAUFMAN 2004; MARQUES-SOUZA *et al.* 2008; MITO *et al.* 2006; PATEL *et al.* 2001; PEEL *et al.* 2005; WOLFF *et al.* 1998) and it may function more broadly in arthropod segmentation (SCHWAGER *et al.* 2009). However, *hunchback* genes are not expressed in a pattern consistent with a role in segmentation in some basal arthropods (CHIPMAN and STOLLEWERK 2006; KONTARAKIS *et al.* 2006) and annelids (KERNER *et al.* 2006; PINNELL *et al.* 2006). Expression of *hunchback* in the mesoderm of animals ranging from arthropods to annelids has led to the proposal that the ancestral function of *hunchback* was in mesoderm development (KERNER *et al.* 2006; KONTARAKIS *et al.* 2006; PATEL *et al.* 2001). We note that the developing mesodermal lineage also gives rise to both the adaptive immune system in vertebrates and the somatic gonad in *C. elegans*. Therefore, we speculate that the common ancestor of the HIL, *hunchback*, and Ikaros families might have been involved in mesoderm development.

Consistent with this idea, we were unable to find any genes with this unique gene structure in non-bilaterian species, as has been previously reported for *hunchback* (KERNER *et al.* 2006). The characterization of additional HIL, *hunchback*, and IFL genes, especially in basal lineages, should reveal if mesodermal expression is a common feature of all three gene families. Our description of the HIL gene family in *C. elegans* opens the door to future studies on the function of HIL genes in other protostomes.

ACKNOWLEDGEMENTS

We thank Maureen Barr, Frank Slack, Shohei Mitani and the National BioResource Project, and the *Caenorhabditis* Genetics Center for strains. Confocal microscopy was performed at the Cell and Molecular Imaging Facility at North Carolina State University. We are grateful to Pat Estes and Bob Franks for comments on the manuscript and Zigrida Smith for technical assistance. Preliminary sequence data was obtained from the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) website at http://www.hgsc.bcm.tmc.edu. The DNA sequencing of *S. kowalevskii* was supported by grant number 5U54HG003272 from the National Human Genome Research Institute (NHFRI) and National Institutes of Health (NIH) to Richard Gibbs at BCM-HGSC. E.E.L. was supported by a GAANN Biotechnology fellowship and a predoctoral fellowship from the National Institutes of Health (GM-08443-10). This research was supported by a grant from the National Science Foundation (IOB-0617897) to L.D.M.

REFERENCES

ASAHINA, M., T. VALENTA, M. SILHANKOVA, V. KORINEK and M. JINDRA, 2006 Crosstalk between a nuclear receptor and beta-catenin signaling decides cell fates in the *C. elegans* somatic gonad. Dev Cell **11:** 203-211.

BLELLOCH, R., S. SANTA ANNA-ARRIOLA, D. GAO, Y. LI, J. HODGKIN *et al.*, 1999 The *gon-1* gene is required for gonadal morphogenesis in *Caenorhabditis elegans*. Developmental Biology **216:** 382-393.

BLUMENTHAL, T., and K. STEWARD, 1997 RNA processing and gene structure, pp. 117-145 in *C. elegans II*, edited by D. L. RIDDLE, T. BLUMENTHAL, B. J. MEYER and J. R. PRIESS. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

CHANG, W., C. TILMANN, K. THOEMKE, F. H. MARKUSSEN, L. D. MATHIES *et al.*, 2004 A forkhead protein controls sexual identity of the *C. elegans* male somatic gonad. Development **131:** 1425-1436.

CHIPMAN, A. D., and A. STOLLEWERK, 2006 Specification of neural precursor identity in the geophilomorph centipede *Strigamia maritima*. Dev Biol **290:** 337-350.

CUPIT, P. M., J. D. HANSEN, A. S. MCCARTY, G. WHITE, M. CHIODA *et al.*, 2003 Ikaros family members from the agnathan *Myxine glutinosa* and the urochordate *Oikopleura dioica*: emergence of an essential transcription factor for adaptive immunity. J Immunol **171:** 6006-6013.

DOLPHIN, C. T., and I. A. HOPE, 2006 *Caenorhabditis elegans* reporter fusion genes generated by seamless modification of large genomic DNA clones. Nucleic Acids Res **34:** e72.

EMERSON, R. O., and J. H. THOMAS, 2009 Adaptive evolution in zinc finger transcription factors. PLoS Genet **5:** e1000325.

GEORGOPOULOS, K., 2002 Haematopoietic cell-fate decisions, chromatin regulation and ikaros. Nat Rev Immunol **2:** 162-174.

GEORGOPOULOS, K., D. D. MOORE and B. DERFLER, 1992 Ikaros, an early lymphoid specific transcription factor and a putative mediator for T cell commitment. Science **258:** 808-812.

GREENWALD, I. S., P. W. STERNBERG and H. R. HORVITZ, 1983 The *lin-12* locus specifies cell fates in *Caenorhabditis elegans*. Cell **34:** 435-444.

HAHM, K., B. S. COBB, A. S. MCCARTY, K. E. BROWN, C. A. KLUG *et al.*, 1998 Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes & Development **12:** 782-796.

HEDGECOCK, E. M., J. G. CULOTTI, D. H. HALL and B. D. STERN, 1987 Genetics of cell and axon migration in *Caenorabditis elegans*. Development **100:** 365-382.

HIBINO, T., M. LOZA-COLL, C. MESSIER, A. J. MAJESKE, A. H. COHEN *et al.*, 2006 The immune gene repertoire encoded in the purple sea urchin genome. Dev Biol **300:** 349-365.

HOBERT, O., 2002 PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. BioTechniques **32:** 728-730.

HODGKIN, J., 1997 Appendix 1. Genetics, pp. 881-1047 in *C. elegans II*, edited by D. L. RIDDLE, T. BLUMENTHAL, B. J. MEYER and J. R. PRIESS. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

HONMA, Y., H. KIYOSAWA, T. MORI, A. OGURI, T. NIKAIDO *et al.*, 1999 Eos: a novel member of the Ikaros gene family expressed predominantly in the developing nervous system. FEBS Letters **447:** 76-80.

HUNT-NEWBURY, R., R. VIVEIROS, R. JOHNSEN, A. MAH, D. ANASTAS *et al.*, 2007 High-throughput *in vivo* analysis of gene expression in *Caenorhabditis elegans*. PLoS Biol **5:** e237.

INOUE, T., D. R. SHERWOOD, G. ASPOCK, J. A. BUTLER, B. P. GUPTA *et al.*, 2002 Gene expression markers for *Caenorhabditis elegans* vulval cells. Mech Dev **119 Suppl 1:** S203-209.

JOHN, L. B., S. YOONG and A. C. WARD, 2009 Evolution of the Ikaros gene family: implications for the origins of adaptive immunity. J Immunol **182:** 4792-4799.

KERNER, P., F. ZELADA GONZALEZ, M. LE GOUAR, V. LEDENT, D. ARENDT *et al.*, 2006 The expression of a *hunchback* ortholog in the polychaete annelid *Platynereis dumerilii* suggests an ancestral role in mesoderm development and neurogenesis. Dev Genes Evol **216:** 821-828.

KIDD, A. R., 3RD, J. A. MISKOWSKI, K. R. SIEGFRIED, H. SAWA and J. KIMBLE, 2005 A beta-catenin identified by functional rather than sequence criteria and its role in Wnt/MAPK signaling. Cell **121:** 761-772.

KIM, J., S. SIF, B. JONES, A. JACKSON, J. KOIPALLY *et al.*, 1999 Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity **10:** 345-355.

KIMBLE, J., 1981 Alterations in cell lineage following laser ablation of cells in the somatic gonad of *Caenorhabditis elegans*. Dev. Biol. **87:** 286-300.

KIMBLE, J., and D. HIRSH, 1979 The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Developmental Biology **70:** 396-417.

KIMBLE, J. E., and J. G. WHITE, 1981 On the control of germ cell development in *Caenorhabditis elegans*. Developmental Biology **81:** 208-219.

KNIGHT, R. D., and S. M. SHIMELD, 2001 Identification of conserved C2H2 zinc-finger gene families in the Bilateria. Genome Biol **2:** RESEARCH0016.

KOIPALLY, J., J. KIM, B. JONES, A. JACKSON, N. AVITAHL *et al.*, 1999 Ikaros chromatin remodeling complexes in the control of differentiation of the hemo-lymphoid system. Cold Spring Harb Symp Quant Biol **64:** 79-86.

KONTARAKIS, Z., T. COPF and M. AVEROF, 2006 Expression of *hunchback* during trunk segmentation in the branchiopod crustacean *Artemia franciscana*. Dev Genes Evol **216:** 89-93.

KOSTAS, S. A., and A. FIRE, 2002 The T-box factor MLS-1 acts as a molecular switch during specification of nonstriated muscle in *C. elegans*. Genes Dev **16:** 257-269.

KRAEMER, B., S. CRITTENDEN, M. GALLEGOS, G. MOULDER, R. BARSTEAD *et al.*, 1999 NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. Curr Biol **9:** 1009-1018.

LARKIN, M. A., G. BLACKSHIELDS, N. P. BROWN, R. CHENNA, P. A. MCGETTIGAN *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948.

LEHMANN, R., and C. NUSSLEIN-VOLHARD, 1987 *hunchback*, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. Dev Biol **119:** 402-417.

LIU, P. Z., and T. C. KAUFMAN, 2004 *hunchback* is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect *Oncopeltus fasciatus*. Development **131:** 1515-1527.

MADURO, M., and D. PILGRIM, 1995 Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics **141:** 977-988.

MARQUES-SOUZA, H., M. ARANDA and D. TAUTZ, 2008 Delimiting the conserved features of *hunchback* function for the trunk organization of insects. Development **135:** 881-888.

MATHIES, L. D., S. T. HENDERSON and J. KIMBLE, 2003 The *C. elegans* Hand gene controls embryogenesis and early gonadogenesis. Development **130:** 2881-2892.

MATHIES, L. D., M. SCHVARZSTEIN, K. M. MORPHY, R. BLELLOCH, A. M. SPENCE *et al.*, 2004 TRA-1/GLI controls development of somatic gonadal precursors in *C. elegans*. Development **131:** 4333-4343.

MCCARTY, A. S., G. KLEIGER, D. EISENBERG and S. T. SMALE, 2003 Selective Dimerization of a C2H2 Zinc Finger subfamily. Molecular Cell **11:** 459-470.

MELLO, C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J **10:** 3959-3970.

MISKOWSKI, J., Y. LI and J. KIMBLE, 2001 The *sys-1* gene and sexual dimorphism during gonadogenesis in *Caenorhabditis elegans*. Developmental Biology **230:** 61-73.

MITO, T., H. OKAMOTO, W. SHINAHARA, Y. SHINMYO, K. MIYAWAKI *et al.*, 2006 *Kruppel* acts as a gap gene regulating expression of *hunchback* and *even-skipped* in the intermediate germ cricket *Gryllus bimaculatus*. Dev Biol **294:** 471-481.

MOLNAR, A., and K. GEORGOPOULOS, 1994 The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Molecular and Cellular Biology **14:** 8292-8303.

MORGAN, B., L. SUN, N. AVITAHL, K. ANDRIKOPOULOS, T. IKEDA *et al.*, 1997 Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. EMBO J. **16:** 2004-2013.

NG, S. Y., T. YOSHIDA and K. GEORGOPOULOS, 2007 Ikaros and chromatin regulation in early hematopoiesis. Curr Opin Immunol **19:** 116-122.

NG, S. Y., T. YOSHIDA, J. ZHANG and K. GEORGOPOULOS, 2009 Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity **30:** 493-507.

PATEL, N. H., D. C. HAYWARD, S. LALL, N. R. PIRKL, D. DIPIETRO *et al.*, 2001 Grasshopper *hunchback* expression reveals conserved and novel aspects of axis formation and segmentation. Development **128:** 3459-3472.

PEEL, A. D., A. D. CHIPMAN and M. AKAM, 2005 Arthropod segmentation: beyond the *Drosophila* paradigm. Nat Rev Genet **6:** 905-916.

PERDOMO, J., M. HOLMES, B. CHONG and M. CROSSLEY, 2000 Eos and Pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. J. Biol. Chem. **275:** 38347-38354.

PHILLIPS, B. T., A. R. KIDD, 3RD, R. KING, J. HARDIN and J. KIMBLE, 2007 Reciprocal asymmetry of SYS-1/beta-catenin and POP-1/TCF controls asymmetric divisions in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A **104:** 3231-3236.

PINNELL, J., P. S. LINDEMAN, S. COLAVITO, C. LOWE and R. M. SAVAGE, 2006 The divergent roles of the segmentation gene *hunchback*. Integrative and Comparative Biology **46:** 519-532.

RAST, J. P., L. C. SMITH, M. LOZA-COLL, T. HIBINO and G. W. LITMAN, 2006 Genomic insights into the immune system of the sea urchin. Science **314:** 952-956.

REECE-HOYES, J. S., J. SHINGLES, D. DUPUY, C. A. GROVE, A. J. WALHOUT *et al.*, 2007 Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. BMC Genomics **8:** 27.

ROBB, S. M., E. ROSS and A. SANCHEZ ALVARADO, 2008 SmedGD: the *Schmidtea mediterranea* genome database. Nucleic Acids Res **36:** D599-606.

ROBERTSON, S. M., P. SHETTY and R. LIN, 2004 Identification of lineage-specific zygotic transcripts in early *Caenorhabditis elegans* embryos. Dev Biol **276:** 493-507.

ROTHENBERG, E. V., and R. PANT, 2004 Origins of lymphocyte developmental programs: transcription factor evidence. Semin Immunol **16:** 227-238.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406-425.

SALAMOV, A. A., and V. V. SOLOVYEV, 2000 *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res **10:** 516-522.

SAVAGE, R. M., and M. SHANKLAND, 1996 Identification and characterization of a *hunchback* orthologue, Lzf2, and its expression during leech embryogenesis. Dev Biol **175:** 205-217.

SAWA, H., L. LOBEL and H. R. HORVITZ, 1996 The *Caenorhabditis elegans* gene *lin-17*, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. Genes & Development **10:** 2189-2197.

SCHWAGER, E. E., M. PECHMANN, N. M. FEITOSA, A. P. MCGREGOR and W. G. DAMEN, 2009 *hunchback* functions as a segmentation gene in the spider *Achaearanea tepidariorum*. Curr Biol **19:** 1333-1340.

SIEGFRIED, K., A. R. I. KIDD, M. A. CHESNEY and J. KIMBLE, 2004 The *sys-1* and *sys-3* genes cooperate with Wnt signaling to establish the proximal-distal axis of the *Caenorhabditis elegans* gonad. Genetics **166:** 171-186.

SIEGFRIED, K., and J. KIMBLE, 2002 POP-1 controls axis formation during early gonadogenesis in *C. elegans*. Development **129:** 443-453.

SMALE, S. T., and K. DORSHKIND, 2006 Hematopoiesis flies high with Ikaros. Nat Immunol **7:** 367-369.

SOMMER, R. J., M. RETZLAFF, K. GOERLICH, K. SANDER and D. TAUTZ, 1992 Evolutionary conservation pattern of zinc-finger domains of *Drosophila* segmentation genes. Proc Natl Acad Sci U S A **89:** 10782-10786.

SRIDHARAN, R., and S. T. SMALE, 2007 Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. J Biol Chem **282:** 30227-30238.

ST JOHNSTON, D., and C. NÜSSLEIN-VOLHARD, 1992 The origin of pattern and polarity in the *Drosophila* embryo. Cell **68:** 201-219.

STANOJEVIC, D., T. HOEY and M. LEVINE, 1989 Sequence-specific DNA-binding activities of the gap proteins encoded by *hunchback* and *Kruppel* in *Drosophila*. Nature **341:** 331-335.

STERNBERG, P. W., and H. R. HORVITZ, 1988 *lin-17* mutations of *Caenorhabditis elegans* disrupt certain asymmetric cell divisions. Developmental Biology **130:** 67-73.

SUN, L., A. LIU and K. GEORGOPOULOS, 1996 Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. EMB0 J. **15:** 5358-5369.

TILMANN, C., and J. KIMBLE, 2005 Cyclin D regulation of a sexually dimorphic asymmetric cell division. Dev Cell **9:** 489-499.

TREISMAN, J., and C. DESPLAN, 1989 The products of the *Drosophila* gap genes *hunchback* and *Kruppel* bind to the *hunchback* promoters. Nature **341:** 335-337.

TURSUN, B., L. COCHELLA, I. CARRERA and O. HOBERT, 2009 A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. PLoS ONE **4:** e4625.

WELCHMAN, D. P., L. D. MATHIES and J. AHRINGER, 2007 Similar requirements for CDC-42 and the PAR-3/PAR-6/PKC-3 complex in diverse cell types. Dev Biol **305:** 347-357.

WOLFE, S. A., L. NEKLUDOVA and C. O. PABO, 2000 DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct **29:** 183-212.

WOLFF, C., R. SCHRODER, C. SCHULZ, D. TAUTZ and M. KLINGLER, 1998 Regulation of the *Tribolium* homologues of *caudal* and *hunchback* in *Drosophila*: evidence for maternal gradient systems in a short germ embryo. Development **125:** 3645-3654.

WOOD, W. B., 1988 Introduction to *C. elegans* Biology, pp. 1-16 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

YU, D., H. M. ELLIS, E. C. LEE, N. A. JENKINS, N. G. COPELAND *et al.*, 2000 An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A **97:** 5978-5983.

ZHAO, J., P. WANG and A. K. CORSI, 2007 The *C. elegans* Twist target gene, *arg-1*, is regulated by distinct E box promoter elements. Mech Dev **124:** 377-389.

Table 1. *R08E3.4* and *ehn-3* have overlapping functions in somatic gonad development

Genotype	% Gon[a] +/- SD	*n*	p[b]
N2 (wild type)	0 +/- 0	220	-
N2+F12E12.5 RNAi	0.1 +/- 0.3	933	-
N2+C46E10.8 RNAi	0 +/- 0	563	-
N2+C46E10.9 RNAi	0 +/- 0	550	-
N2+R08E3.4 RNAi	0 +/- 0	323	-
N2+ehn-3 RNAi	5.7 +/- 2.7	418	-
ehn-3(q766)	3.1 +/- 2.3	382	-
ehn-3(q689)	21.5 +/- 1.0	156	-
ehn-3(rd2)	24.2 +/- 3.7	434	-
ehn-3(rd2)+R08E3.4 RNAi[c]	45 +/- 12	864	0.001
ehn-3(rd2); R08E3.4(za16)[c]	45.1 +/- 4.2	717	<0.0001
ehn-3(rd2); R08E3.4(ok1916)[c]	56.5 +/- 1.7	983	<0.0001
ehn-3(rd2);R08E3.4(tm2127)[c]	62.3 +/- 5.8	612	<0.0001
ehn-3(rd2);R08E3.4(tm2127); R08E3.4(RNAi)[d]	51.7 +/- 6.1	333	0.011
ehn-3(rd2)+F12E12.5 RNAi[c]	20.7 +/- 5.1	1183	0.621
C46E10.8(tm442); ehn-3(rd2)[c]	27.3 +/- 5.4	1041	0.282
C46E10.9(tm1692); ehn-3(rd2)[c]	19.3 +/- 2	379	0.062
ehn-3(rd2); R08E3.4(tm2127) + F12E12.5 RNAi[d]	42.1 +/- 6.1	247	0.002
C46E10.8(tm442); ehn-3(rd2); R08E3.4(tm2127)[d]	56.3 +/- 5.6	1336	0.282
C46E10.9(tm1692); ehn-3(rd2); R08E3.4(tm2127)[d]	48.9 +/- 5.1	914	0.002
C46E10.8(tm442);ehn-3(rd2);R08E3.4(tm2127) + F12E12.5 RNAi[d]	67.4 +/- 4.3	678	0.237
C46E10.9(tm1692);ehn-3(rd2);R08E3.4(tm2127)+ F12E12.5 RNAi[d]	47.2 +/- 7.8	865	0.015

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.
[b] Unpaired t-tests were used for statistical comparisons and the p-value is reported.
[c] *ehn-3(rd2)* was compared.
[d] *ehn-3(rd2);R08E3.4(tm2127)* was compared.

Table 2. Correlation of L1 and adult phenotype in *ehn-3* and *ehn-3; R08E3.4* mutants

Genotype[†]	Gonadal primordium	Adult Gonad (%)[#]				
	SGP morphology*	Two arms	One arm	Abnormal	None visible	*n*
wild-type	ventral contact	100	0	0	0	54
ehn-3(rd2)	ventral contact	95	5	0	0	56
	abnormal contact	93	3	0	3	59
	abnormal position	57	43	0	0	7
	abnormal morphology	0	33	0	67	3
ehn-3(rd2); R08E3.4(tm2127)	ventral contact	44	50	6	0	18
	abnormal contact	60	30	10	0	10
	abnormal position	43	29	29	0	7
	abnormal morphology	0	57	21	21	14

[†] All contain *rdIs2*, an integrated *ehn-3::GFP*

[*] Presence, position, and morphology of Z1 and Z4 were scored by *ehn-3::GFP* expression

[#] Rows may not add to 100% due to rounding errors

Table 3. *ehn-3* and *R08E3.4* have overlapping functions in the generation of DTCs

	Percentage of animals				
Genotype[b]	Three hDTCs	Two hDTCs	One hDTC	No hDTCs	*n*
wild-type	0	99.1	0.9	0	106
R08E3.4(tm2127)	1.8	90.9	7.3	0	55
ehn-3(rd2)	0	50.3	43.4	6.3	143
ehn-3(rd2); R08E3.4(tm2127)	0	42.2	45.3	12.5	128

[b]all strains contain *qIs56*, an integrated *lag-2::GFP* reporter

Table 4. EHN-3 and R08E3.4 are not functionally equivalent

Genotype	Rescue construct	% Gon[a] +/- SD	*n*
ehn-3(rd2)	none	24.2 +/- 3.7	434
	EHN-3A::GFP	0.8 +/- 0.3	384
	EHN-3B::GFP	28.6 +/- 3.5	742
	R08E3.4A::GFP	13.8 +/- 1.1	1428
	R08E3.4B::GFP	14.8 +/- 1.7	711
ehn-3(rd2); R08E3.4(tm2127)	none	62.3 +/- 5.8	612
	EHN-3A::GFP	3.3 +/- 2.2	394
	EHN-3B::GFP	47.8 +/- 5.0	812
	R08E3.4A::GFP	18.9 +/- 5.6	792
	R08E3.4B::GFP	26 +/- 8.6	1131
	R08E3.4C::GFP	56.1 +/- 5.8	431

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

Table S1. Primer sequences

Primer	Sequence
RNAi primers	
F12E12.5 5'	GCCGTATTGTCGCTTATTG
F12E12.5 3'	GCTCATTGTGGCATCGTGGAA
R08E3.4 5'	ATCAAGGCCAAATGTACCAA
R08E3.4 3'	GCACTGAGTTCCGCACAA
RA276	TGTGACTGGATTGGACAAGATGC
RA277	ATGGACCTGCCTGTAACCC
RA252	GGAGCTTCTTTGTCAGCAGAAAGA
RA253	ATGAGCATTTGAATGGGCTTCC
RA282	TAATACGACTCACTATAGGG
RA284	TAATACGACTCACTATAGGGAGTTTTCCCAGTCACGACG
RT-PCR	
RA211	GGTTTAATTACCCAAGTTTGAG
RA219	GCACTGAGTTCCGCACAA
RA223	TTTGGAATGTTGAGCATTGG
RA255	CCTATGCCACCAAAATCATGACTC
RA288	GCTCATTGTGGCATCGTGGAA
RA253	ATGAGCATTTGAATGGGCTTCC
RA276	TGTGACTGGATTGGACAAGATGC
Transcriptional Reporters	
RA226	TCGGCTAGACAAACTTGTGGGA
RA246	CTTTGGGTCCTTTGGCCAATCGCATTCTGTGCAAGCGTTCAG
RA245	CTGAACGCTTGCACAGAATGCGATTGGCCAAAGGACCCAAAG
RA247	CCAAACCTTCTTCCGATCT
RA244	CGCCTTTAGAGAGTACTGTA
RA247	CCAAACCTTCTTCCGATCT
RA425	TTCTGCAGGACGAATAGGAAGACTTGTAGAGGC
RA426	TTTCTAGACATTTTCACAGAAGAAATCTCTGCG
RA427	TTCTGCAGGTACTCAATGAATCTCGTTCTCGC
RA428	TTTCTAGACATCGCATTCAATAACTCCCCT
RA429	TTCTGCAGGTGTGAAGATAATCTCGCGTTTACA
RA430	TTTCTAGAGGTTACAGGCAGGTCCATTGATA
RA431	TTCTGCAGGTAGGTTCAAGGATTTTTACCCG
RA432	TTTCTAGACCGATGACGATGCATGTTAGC

Table S1. (cont'd)

Rescue Constructs	
RA696	CTGGATCATACGAAATTCTTCTCTTTC
RA697	TTTTTCCATTTTGTAATTTGGAAGCTGG
RA698	TGCGACATCTGTAAAATGCAGT
RA699	ATGTCTTGTCTGGAAATGAAGGAG
RA708	GACGATTCTGAGTCTTCTCCGA
RA709	TTGGTACATTTTTACCGGTACCC
RA710	TGCGATATTTGTCACCAGAAGT
RA711	GTGTGCAGTATGAAAATGAAGCAG
RA522	AACTTCTATCGGATCCCCG
RA523	GTACCAGCTCTCTGGACCATCCA
RA516	TGCGAACACTGCCGGGT
RA517	GTGCTCCCCTCGCGTTATGT
RA518	TGCTGTCATTGCGGAATGAT
RA519	ATGATCGGCAAAAATGACATGAG
RA522	AACTTCTATCGGATCCCCG
RA523	GTACCAGCTCTCTGGACCATCCA

Table S2. Additional genetic interaction tests

Genotype	% Gon[a] +/- SD	n
ehn-3(q766)	3.1 +/- 2.3	382
ehn-3(q766)+F12E12.5 RNAi	1.9 +/- 0.7	417
ehn-3(q766)+C46E10.9 RNAi	1.6 +/- 2.1	383
ehn-3(q766)+R08E3.4 RNAi	7.3 +/- 1.7	441
ehn-3(q766); R08E3.4(za16)	7.1 +/- 3.1	709
ehn-3(q689)	21.5 +/- 1.0	156
ehn-3(q689)+C46E10.9 RNAi	24.1 +/- 1.7	742
ehn-3(q689)+F12E12.5 RNAi	24.3 +/- 2.9	601
ehn-3(q689); R08E3.4(za16)	47.1 +/- 2.9	1538
ehn-3(q689); R08E3.4(za16)+F12E12.5 RNAi	47.4 +/- 3.8	428
ehn-3(q689); R08E3.4(za16)+C46E10.9 RNAi	39.6 +/- 6.2	384
ehn-3(rd2)	24.2 +/- 3.7	434
ehn-3(rd2)+C46E10.9 RNAi	19.8 +/- 6.2	575
ehn-3(rd2)+C46E10.8 RNAi	31.3 +/- 6.5	211
R08E3.4(tm2127)+F12E12.5 RNAi	0.5 +/- 0.7	202
C46E10.9(tm1692)+F12E12.5 RNAi	0 +/- 0	440
C46E10.8(tm442)+F12E12.5 RNAi	0 +/- 0	527
C46E10.9(tm1692); R08E3.4(tm2127)	0 +/- 0	667
C46E10.8(tm442); R08E3.4(tm2127)	0 +/- 0	665

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

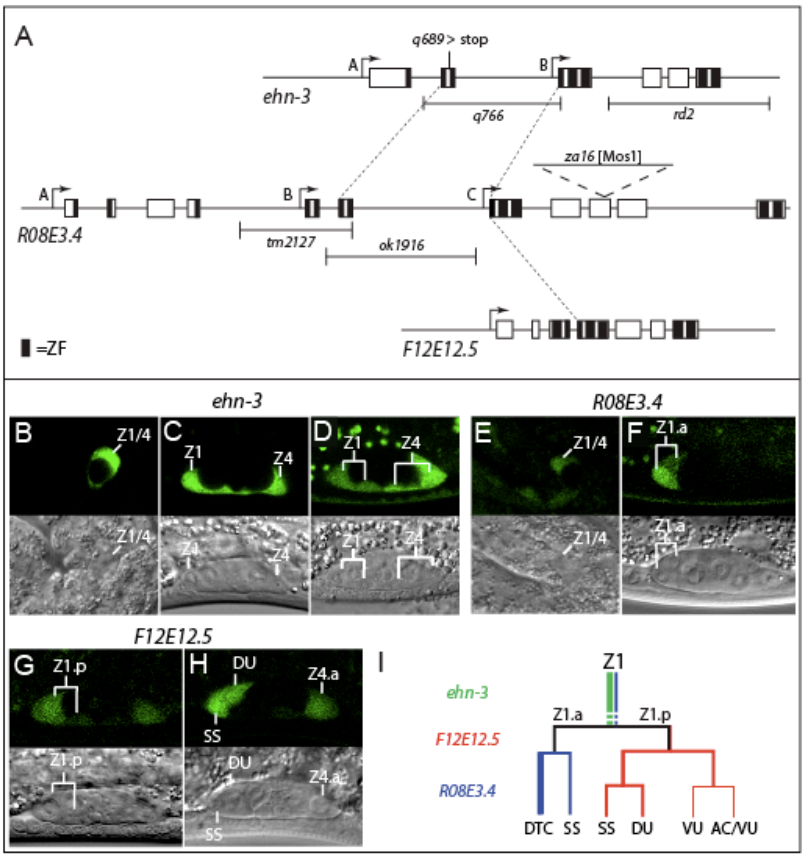
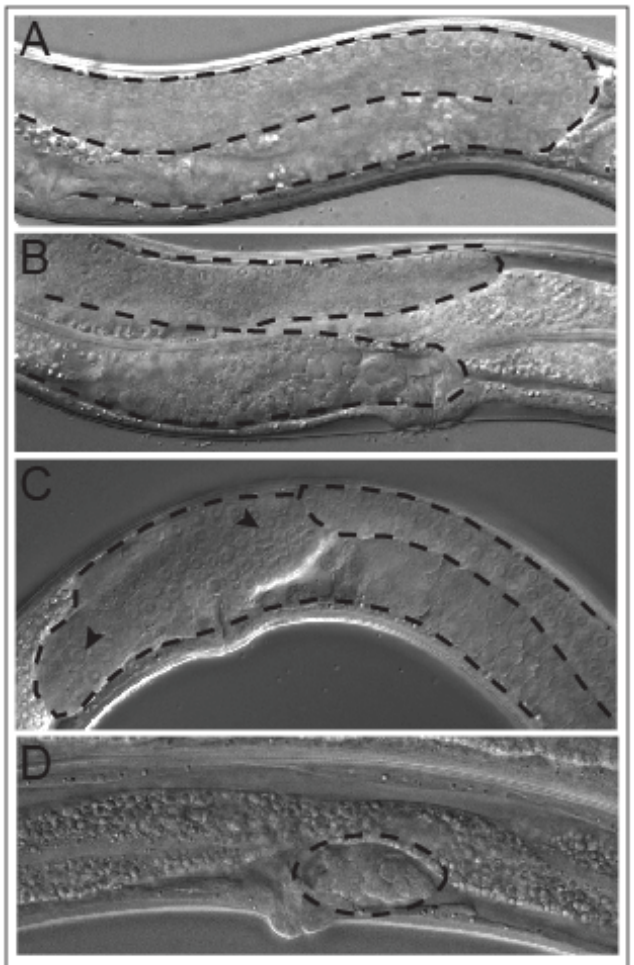


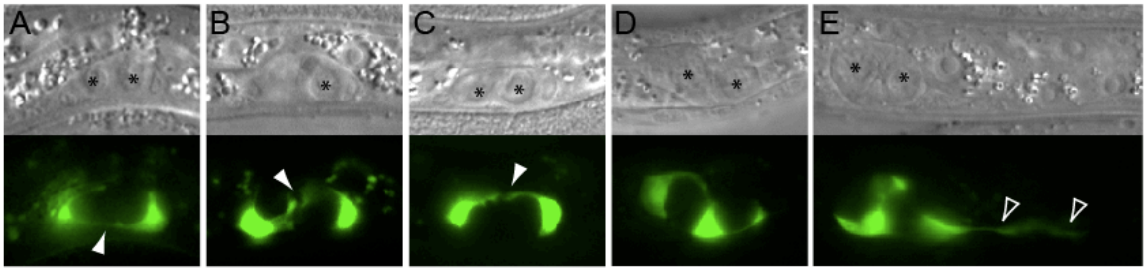
Figure 1. Three HIL genes are expressed in the early somatic gonad lineage
(A) Gene structures of *ehn-3, R08E3.4,* and *F12E12.5*. Three alleles of *ehn-3* and *R08E3.4* are used in this study; the molecular nature of each allele is indicated. Predicted zinc fingers (ZF) are indicated by black boxes. Narrow boxes indicate ZFs that are split by introns; these are the first and second ZFs of *ehn-3,* the first through fourth ZFs of *R08E3.4*, and the second ZF of *F12E12.5*. Promoters are indicated by arrows: *ehn-3* has two promoters and produces isoforms that contain 4-6 zinc fingers, *R08E3.4* has three promoters and produces isoforms that contain 4-8 zinc fingers, and *F12E12.5* has one promoter and produces a protein with 5 zinc finger motifs. Conserved intron-exon boundaries are indicated by dashed lines. (B-H) Confocal micrographs showing GFP (top) and DIC (bottom) images. (B-D) *ehn-3* transcriptional reporters are expressed in the SGPs in embryos (B) and L1 larvae (C). They continue to be expressed in the SGP daughters through the L1 stage (D). *ehn-3* translational reporters are expressed in SGPs only in embryos (not shown). (E-F) *R08E3.4* transcriptional reporters are expressed in SGPs in embryos (E) and in the descendants of Z1.a and Z4.p (F; Z1.a is shown). (G-H) *F12E12.5* reporters are expressed in the daughters of Z1.p and Z4.a (G; Z1.p is shown) and in Z1.pa and Z4.ap and their descendant (H; SS, DU). (I) Summary of the temporal expression of *ehn-3, F12E12.5,* and *R08E3.4* during early somatic gonad development (green, *ehn-3;* red, *F12E12.5*; blue, *R08E3.4*); posttranscriptional regulation is taken into account for *ehn-3*. The width of the bar indicates relative expression level. Only the Z1 lineage is shown; Z4 has 2-fold rotational symmetry as compared to Z1.



E

Genotype	Percentage of animals: Two arms	One arm	No arms	n
wild-type	100	0	0	>200
R08E3.4(tm2127)	100	0	0	160
ehn-3(rd2)	79	19	2	378
ehn-3(rd2); R08E3.4(tm2127)	33	50	17	211

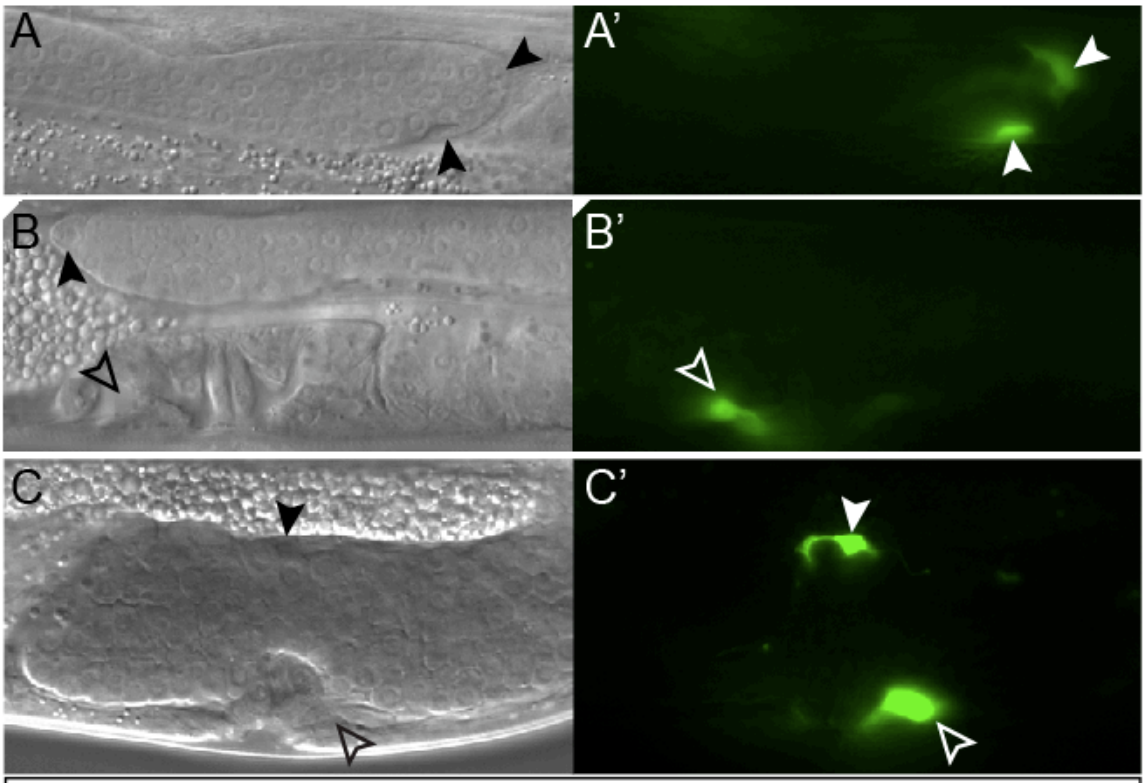
Figure 2. *ehn-3* and *R08E3.4* are required for somatic gonad development.
(A-D) Representative DIC images of young adult *ehn-3(rd2); R08E3.4(tm2127)* double mutants; the gonad is indicated by dashed lines. (A) Two gonadal arms; the posterior arm is outlined. (B) Anterior gonadal arm only. (C) Posterior gonadal arm only; near the vulva is a large patch of germ cells (arrowheads). (D) Severely underdeveloped gonad, with only a small patch of gonadal tissue near the vulva. (E) DIC analysis of gonadal morphology in wild type, *R08E3.4(tm2127)*, *ehn-3(rd2),* and *ehn-3(rd2); R08E3.4(tm2127)* mutants. Percentage of animals with two gonadal arms (Two arms), one gonadal arm (One arm), or no gonadal arms (No arms) is reported.



F

Genotype	Percentage of animals: Wild-type	Contact	Position	Morphology	Absent	n
wild-type	95.9	4.1	0	0	0	73
R08E3.4(tm2127)	83.6	15.8	0.6	0	0	171
ehn-3(rd2)	44.1	46.5	5.5	3.1	0.8	127
ehn-3(rd2); R08E3.4(tm2127)	36.0	22.0	14.0	28.0	0	50

Figure 3. *ehn-3; R08E3.4* double mutants have abnormal SGPs
ehn-3(rd2); R08E3.4(tm2127) mutants typically have two SGPs, but they often have abnormal morphology or position within the primordium. (A-E) DIC images of the gonadal primordium; germ cells are marked by asterisks (top); *ehn-3::GFP* highlights SGP cell bodies (bottom). (A) Wild type morphology: SGPs are at the poles of the primordium and make contact on its ventral surface (arrowhead). (B) Abnormal contact: the anterior SGP body is ventral and the posterior SGP body is dorsal; SGPs contact one another centrally (arrowhead). (C) Abnormal contact: SGPs make contact on the dorsal surface of the primordium (arrowhead). (D) Abnormal position: The posterior SGP is located between the two germ cells (asterisks). (E) Abnormal morphology: the posterior SGP extends a process (open arrowheads) away from the gonadal primordium. The anterior SGP is also located between the two germ cells (asterisks). (F) Analysis of the gonadal primordium of wild-type, *R08E3.4(tm2127)*, *ehn-3(rd2),* and *ehn-3(rd2); R08E3.4(tm2127)* mutants with *ehn-3::GFP*. Percentage of animals with SGPs in a wild-type configuration (A), abnormal contact (B, C), abnormal position (D), abnormal morphology (E), and absent SGPs is reported.



D

		L1	L2	L3	L4
wild-type	hDTC expression	n/a	0	0.9	1.1
	Cells with processes	0	0	0	0
		(*n*=110)	(*n*=156)	(*n*=114)	(*n*=93)
ehn-3(rd2); R08E3.4(tm2127)	hDTC expression	n/a	3.8	13.6	5.2
	Cells with processes	0	4.7	16.5	18.2
		(*n*=245)	(*n*=106)	(*n*=103)	(*n*=77)

Figure 4. *ehn-3; R08E3.4* double mutants have ectopic *arg-1* expression in the somatic gonad.
(A-C) DIC images and (A'-C') *arg-1::*GFP fluorescence. (A) Wild-type males express *arg-1::GFP* in the two mDTCs (arrowhead). (B) Wild-type hermaphrodites express *arg-1::GFP* in the vulval muscles (open arrowhead), but not in the hDTCs (closed arrowhead). (C) *ehn-3; R08E3.4* mutants sometimes express *arg-1::GFP* in the hermaphrodite somatic gonad (closed arrowhead) in cells that resemble mDTCs (compare to A); expression is also seen in vulval muscles (open arrowhead). (D) Analysis of *arg-1::GFP* expression in wild-type and *ehn-3(rd2); R08E3.4(tm2127)* mutant hermaphrodites. The percentage of animals with expression in the hermaphrodite DTC (not shown) and cells with processes (C) is reported at four larval stages (L1 through L4).

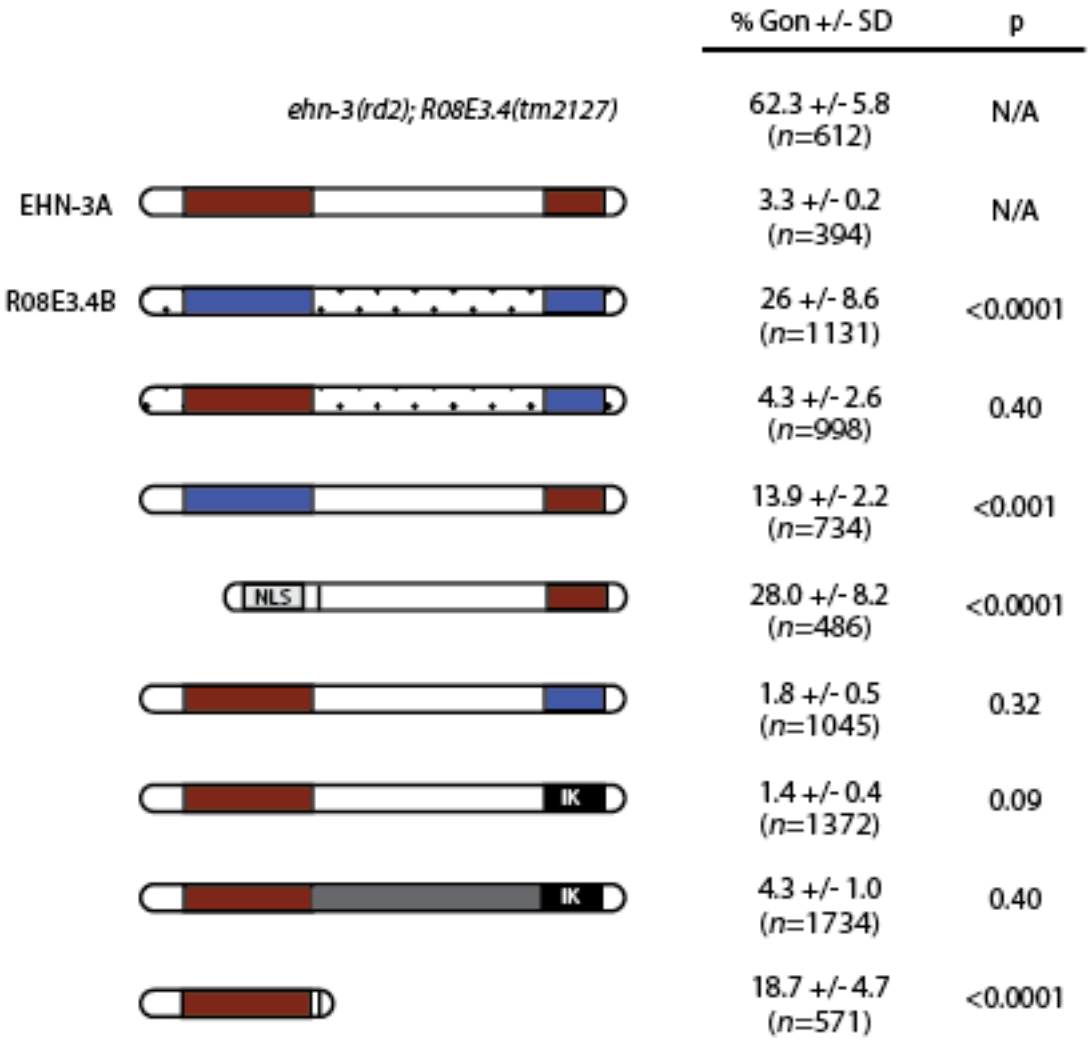
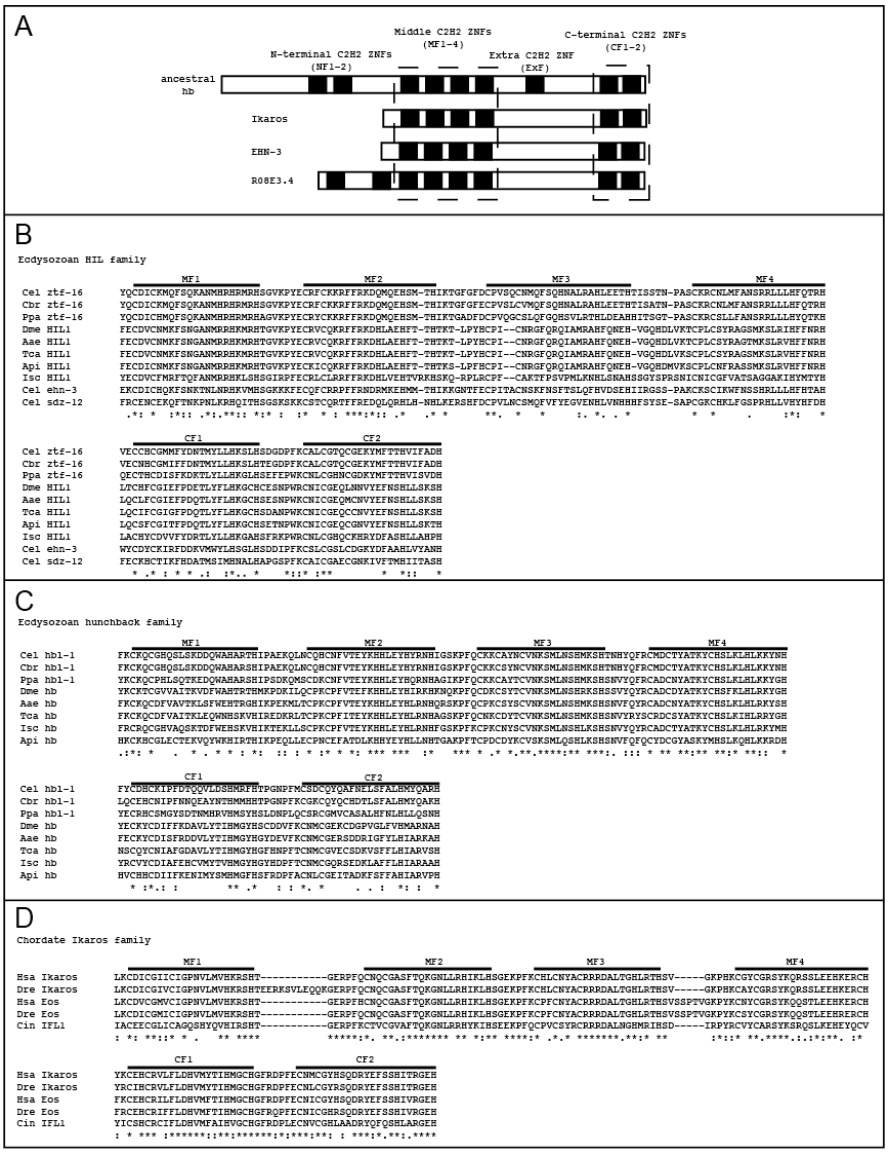


Figure 5. EHN-3 and R08E3.4 have functionally distinct zinc finger domains. Chimeric proteins were expressed in the SGPs under control of the *ehn-3A* promoter and tested for rescue of *ehn-3(rd2); R08E3.4(tm2127)* double mutants. The proteins consisted of different combinations of the EHN-3, R08E3.4, and Ikaros proteins. Zinc fingers are indicated by dark boxes; EHN-3 (red); R08E3.4 (blue); Ikaros (black). Sequences outside the zinc fingers are indicated by light bars; EHN-3 (white); R08E3.4 (stippled); Ikaros (grey). Deletion of the N-terminal zinc fingers of EHN-3A affected nuclear localization of the protein, so an SV40 nuclear localization signal (NLS) was added. The percentage of animals with gonadogenesis defects was scored by dissecting microscope and the average is reported with standard deviation (SD). Unpaired t-tests were used to compare rescue with EHN-3A and the p-value is reported.

Figure 6. Comparison of Ikaros, *hunchback,* and HIL genes.
(A) The ancestral *hunchback,* Ikaros, and HIL genes all have four tandem zinc fingers (MF1-4) and two paired zinc fingers (CF1-2) near their C-terminus; nomenclature as established for *hunchback*. For reference, we include the gene structures of *C. elegans ehn-3* and *R08E3.4*. *hunchback* and HIL genes can have one or two zinc fingers N-terminal to the MFs. A zinc finger between the MF and CF domains, the ExF, is unique to *hunchback* genes. (B-D) Alignment of the MF and CF domains of HIL (B), *hunchback* (C), and Ikaros (D) family genes. MF and CF domains from the following species were aligned using ClustalX: (Cel) *Caenorhabditis elegans*, (Cbr) *Caenorhabditis briggsae*, (Ppa) *Pristionchus pacificus*, (Dme) *Drosophila melanogaster*, (Aae) *Aedes aegypti,* (Tca) *Tribolium castaneum,* (Api) *Acyrthosiphon pisum*, (Isc) *Ixodes scapularis,* (Hsa) *Homo sapiens*, (Dre) *Danio rerio*, (Cin) *Ciona intestinalis*. Conservation of residues is indicated below the columns as identical (asterisk), highly similar (colon), or moderately conserved (period).


A
Middle C2H2 ZNFs
(MF1-4)
C-terminal C2H2 ZNFs
(CF1-2)
N-terminal C2H2 ZNFs
(NF1-2)
Extra C2H2 ZNF
(ExF)
ancestral hb
Ikaros
EHN-3
R08E3.4
B
Ecdysozoan HIL family
MF1
MF2
MF3
MF4
Cel ztf-16 YQCDICKMQFSQKANMHRHRMRHSGVKPYECRFCKKRFFRKDQMQEHSM-THIKTGFGFDCPVSQCNMQFSQHNALRAHLEETHTISSTN-PASCKRCNLMFANSRRLLLHFQTRH
Cbr ztf-16 YQCDICKMQFSQKANMHRHRMRHSGVKPYECRFCKKRFFRKDQMQEHSM-THIKTGFGFECPVSLCVMQFSQHNALRAHLEETHTISATN-PASCKRCNLMFANSRRLLLHFQTRH
Ppa ztf-16 YQCDICKMQFSQKANMHRHRMRHAGVKPYECRYCKKRFFRKDQMQEHSM-THIKTGADFDCPVQGCSLQFGQHSVLRTHLDEAHHITSGT-PASCKRCSLLFANSRRLLLHYQTKH
Dme HIL1 FECDVCNMKFSNGANMRRHKMRHTGVKPYECRVCQKRFFRKDHLAEHFT-THTKT-LPYHCPI--CNRGFQRQIAMRAHFQNEH-VGQHDLVKTCPLCSYRAGSMKSLRIHFFNRH
Aae HIL1 FECDVCNMKFSNGANMRRHKMRHTGVKPYECRVCQKRFFRKDHLAEHFT-THTKT-LPYHCPI--CNRGFQRQIAMRAHFQNEH-VGQHDLVKTCPLCSYRAGSMKSLRVHFFNRH
Tca HIL1 FECDVCNMKFSNGANMRRHKMRHTGVKPYECRVCQKRFFRKDHLAEHFT-THTKT-LPYHCPI--CNRGFQRQIAMRAHFQNEH-VGQHDLVKSCPLCSYRAGSMKSLRVHFFNRH
Api HIL1 FECDICNMKFSNGANMRRHKMRHTGVKPYECRICQKRFFRKDHLAEHFT-THTKS-LPFHCPI--CNRGFQRQIAMRAHFQNEH-VGQHDMVKSCPLCNFRASSMKSLRVHFFNRH
Isc HIL1 YECDVCFMRFTQFANMRRHKLSHSGIRPFECRLCLRRFFRKDHLVEHTVRKHSKQ-RPLRCPF--CAKTFPSVPMLKNHLSNAHSSGYSPRSNICNICGFVATSAGGAKIHYMTYH
Cel ehn-3 EKCDICHQKFSNKTNLNRHKVMHSGKKKFECQFCRRPFFRNDRMKEHMM-THIKKGNTFECPITACNSKFNSFTSLQFHVDSEHIIRGSS-PAKCKSCIKWFNSSHRLLLHFHTAH
Cel sdz-12 FRCENCEKQFTNKPNLKRHQITHSGSKSKKCSTCQRTFFREDQLQRHLH-NHLKERSHFDCPVLWCSMQFVFYEGVENHLVNHHHFSYSE-SAPCGKCHKLFGSPRHLLVHYHFDH
CF1
CF2
Cel ztf-16 VECCHCGMMFYDWTMYLLHKSLHSDGDPFKCALCGTQCGEKYMFTTHVIFADH
Cbr ztf-16 VECNHCGMIFFDWTMYLLEKSLHTEGDPFKCALCGTQCGEKYMFTTHVIFADH
Ppa ztf-16 QECTHCDISFKDKTLYLLSKGLHSEFEPWKCNLCGHNCGDKYMFTTHVISVDH
Dme HIL1 LTCHFCGIEFPDETLYFLHKGCHCESNPWRCNICGEQLNNVYEFNSHLLSKSH
Aae HIL1 LQCLFCGIEFPDQTLYFLHKGCHSESNPWKCNICGEQMCNVYEFNSHLLSKSH
Tca HIL1 LQCIFCGIGFPDQTLYFLHKGCHSDANPWKCNICGEQCCNVYEFNSHLLSKSH
Api HIL1 LQCSFCGITFPDQTLYFLHKGCHSETNPWKCNICGEQCGNVYEFNSHLLSKTH
Isc HIL1 LACHYCDVVFYDRTLYLLHKGAHSFRKPWKCNLCGHQCHERYDFASHLLAHPH
Cel ehn-3 WYCDYCYIRFDDKVMWYLHSGLHSDDIPFKCSLCGSLCDGKYDFAAHLVYAHH
Cel sdz-12 FECKHCTIKFHDATMSIMHNALHAPGSPFKCAICGAECGNKIVFTMHIITASH
C
Ecdysozoan hunchback family
MF1
MF2
MF3
MF4
Cel hb1-1 FKCKQCGHQSLSKDDQWAHARTHIPAEKQLNCQHCNFVTEYKHHLEYHYRNHIGSKPFQCKKCAYNCVNKSMLNSHMKSHTNHYQFRCMDCTYATKYCHSLKLHLKKYNH
Cbr hb1-1 FKCKQCGHQSLSKDDQWAHARSHIPAEKQLNCQHCNFVTEYKHHLEYHYRNHIGSKPFQCKKCSYNCVNKSMLNSHMKSHTNHYQFRCMDCTYATKYCHSLKLHLKKYNH
Ppa hb1-1 YKCKQCPHLSQTKEDQWAHARSHIPSDKQMSCDKCNFVTEYKHHLEYHQRNHAGIKPFQCKKCAYTCVNKSMLNSHMKSHSNVYQFRCADCNYATKYCHSLKLHLKKYGH
Dme hb YKCKTCGVVAITKVDFWAHTRTHMKPDKILQCPKCPFVTEFKHHLEYHIRKHKMQKPFQCDKCSYTCVNKSMLNSHERKSHSSVYQYRCADCDYATKYCHSFKLHLRKYGH
Aae hb FKCKQCDFVAVTKLSFWEHTRGHIKPEKMLTCPKCPFVTEYKHHLEYHLRNHQRSKPFQCPKCSYSCVNKSMLNSHMKSHSNVYQYRCADCNYATKYCHSLKLHLRKYSH
Tca hb FKCKQCDFVAITKLEQWNHSSKVHIREDKRLTCPKCPFITEYKHHLEYHLRNHAGSKPFQCNKCDYTCVNKSMLNSHMKSHSNVYRYSCRDCSYATKYCHSLKIHLRRYGH
Isc hb FRCRQCGHVAQSKTDFWEHSKVHIKTEKLLSCPKCPFVTEYKHHLEYHLRNHFGSKPFKCPKCNYSCVNKSMLNSHLKSHSNVYQYRCADCTYATKYCHSLKLHLRKYMH
Api hb HKCKHCGLECTEKVQYWKHIRTHIKPEQLLECPNCEFATDLKHHYEYHLLNHTGAKPFTCPDCDYKCVSKSMLQSHLKSHSNVFQFQCYDCGYASKYMHSLKQHLKKRDH
CF1
CF2
Cel hb1-1 FYCDECKIPFDSQQVLDSHMRTHTPGNPFMCSDCQYQAFRELSFALHMYQAHH
Cbr hb1-1 LQCEHCNIPFNNQEAYNTHMMHHTPGNPFKCGKCQYQCHDTLSFALHMYQAKH
Ppa hb1-1 YECRHCSMGYSDTNMHRVHMSYHSLDNFLQCSRCGMVCASALHFNLHLLQSNH
Dme hb YECKYCDIFFKDAVLYTIHMGYHSCDDVFKCNMCGEKCDGPVGLFVHMARNAH
Aae hb FCKYCDISFRDDVLYTIHMGYHGYDEVFKCNMCGERSDDRIGFYLHIARKAH
Tca hb NSCQYCNIAFGDAVLYTIHMGYHGFHNPFTCNMCGVECSDKVSFFLHIARVSH
Isc hb YRCVYCDIAFEHCVMYTVHMGYHGYHDPFTCNMCGQRSEDKLAFFLHIARAAH
Api hb HVCHHCDIIFKENIMYSMHMGFHSFRDPFACNLCGEITADKFSFFAHIARVPH
D
Chordate Ikaros family
MF1
MF2
MF3
MF4
Hsa Ikaros LKCDICGIICIGPNVLMVHKRSHT----------GERPFQCNQCGASFTQKGNLLRHIKLHSGEKPFKCHLCNYACRRRDALTGHLRTHSV-----GKPHKCGYCGRSYKQRSSLEEHKERCH
Dre Ikaros LKCDICGIVCIGPNVLMVHKRSHTEERKSVLEQQKGERPFQCNQCGASFTQKGNLLRHIKLHSGEKPFKCHLCNYACRRRDALTGHLRTHSV-----GKPHKCAYCGRSYKQRSSLEEHKERCH
Hsa Eos LKCDVCGMVCIGPNVLMVHKRSHT----------GERPFHCNQCGASFTQKGNLLRHIKLHSGEKPFKCPFCNYACRRRDALTGHLRTHSVSSPTVGKPYKCNYCGRSYKQQSTLEEHKERCH
Dre Eos LKCDICGMICIGPNVLMVHKRSHT----------GERPFQCNQCGASFTQKGNLLRHIKLHSGEKPFKCPFCNYACRRRDALTGHLRTHSVSSPTVGKPYKCSYCGRSYKQQSTLEEHRERCH
Cin IFL1 IACEECGLICAGQSHYQVHIRSHT----------GERPFKCTVCGVAFTQKGNLLRHYKIHSEEKPFQCPVCSYRCRRRDALNGHMRIHSD-----IRPYRCVYCARSYKSRQSLKEHEYQCV
CF1
CF2
Hsa Ikaros YKCEHCRVLFLDHVMYTIHMGCHGFRDPFECNMCGYHSQDRYEFSSHITRGEH
Dre Ikaros YRCIHCRVLFLDHVMYTIHMGCHGFRDPFECNLCGYRSQDRYEFSSHITRGEH
Hsa Eos FKCEHCRILFLDHVMFTIHMGCHGFRDPFECNICGYHSQDRYEFSSHIVRGEH
Dre Eos FRCEHCRIFFLDHVMFTIHMGCHGFRQPFECNICGHRSQDRYEFSSHIVRGEH
Cin IFL1 YICSHCRCIFLDHVMFAIHVGCHGFRDPILCNVCGHLAADRYQFQSHLARGEH

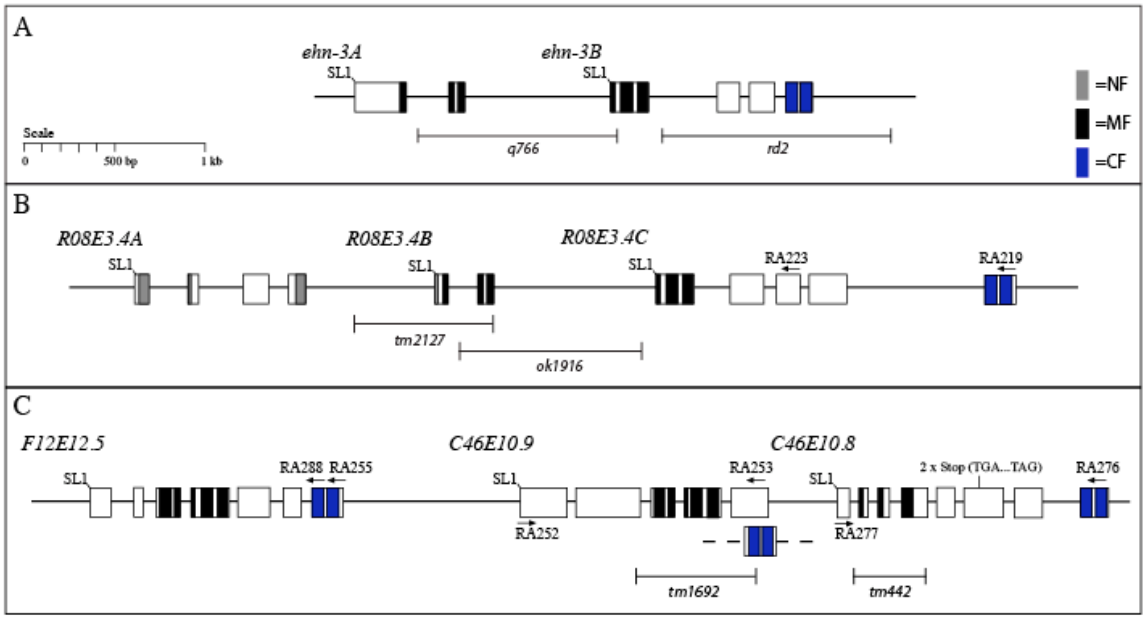


Figure S1. Structure of the *C. elegans* HIL genes.
Grey boxes represent N-terminal fingers (NF); black boxes represent middle zinc fingers (MF); blue boxes represent C-terminal zinc fingers (CF). SL1 trans-splicing events indicate the start of transcription for alternative promoters (SL1). Primers used to determine the gene structures are indicated above and below the gene models with corresponding RA numbers (arrows); deletion alleles are indicated below the gene models. (A) *ehn-3*. (B) *R08E3.4*. (C) The gene cluster on linkage group II containing *F12E12.5, C46E10.9,* and *C46E10.8*. *C46E10.9* produces an mRNA that encodes only the four MFs, which are now located near the C-terminus of the protein; the CFs are encoded in the genome, but are not present in the *C46E10.9* mRNA (shown below the gene model). *C46E10.8* produces an mRNA that encodes only two MFs; its CFs lie downstream of two stop codons.

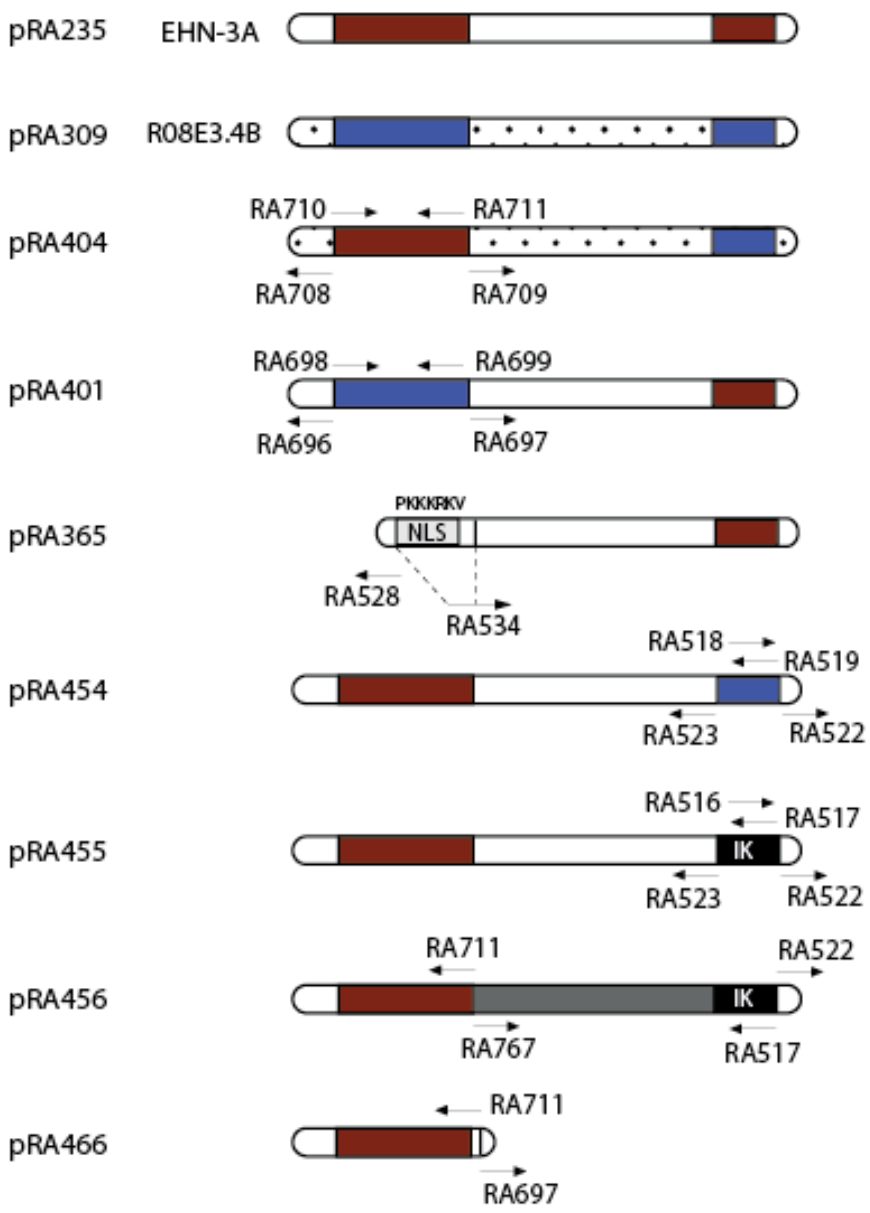


Figure S2. Chimeric rescue constructs.
Zinc fingers are indicated by dark boxes; EHN-3 (red); R08E3.4 (blue); Ikaros (black). Sequences outside the zinc fingers are indicated by light bars; EHN-3 (white); R08E3.4 (stippled); Ikaros (grey). *ehn-3* and *R08E3.4* sequences were amplified from pRA235 or pRA309. Ikaros sequences were amplified from pCMV-SPORT6-hIkaros (Open Biosystems). Primers are shown above and below the chimeric protein with corresponding RA numbers (arrows); resulting plasmids are indicated on the left (pRA#). Constructs were ligated and sequenced as described in methods.

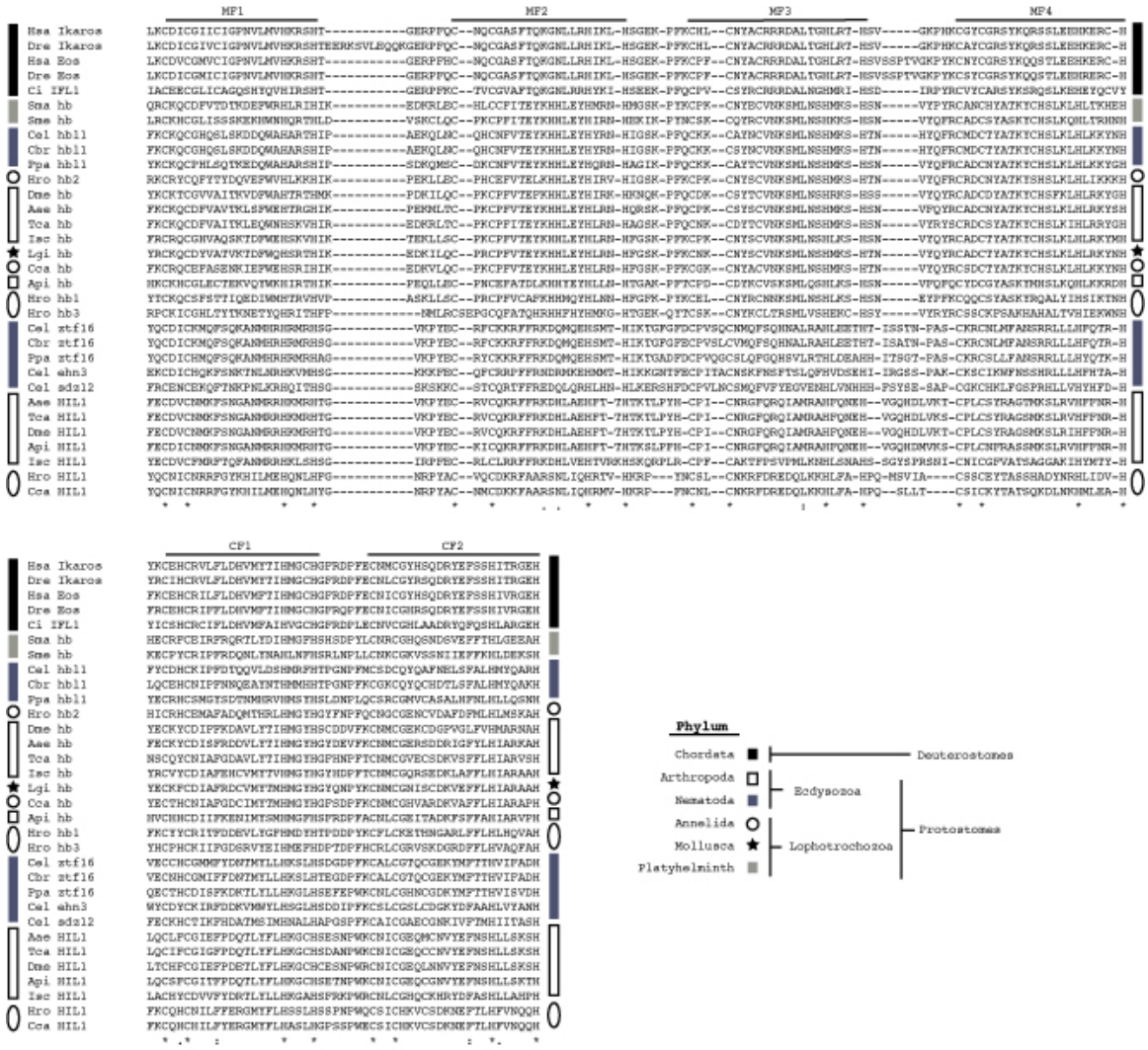


Figure S3. Alignment of six conserved zinc fingers from Ikaros, *hunchback,* and HIL genes.

Middle zinc fingers (MF) and C-terminal zinc fingers (CF) are indicated above the alignment. MF and CF domains from the following species were aligned using ClustalX: (Hsa) *Homo sapiens*, (Dre) *Danio rerio,* (Cin) *Ciona intestinalis,* (Sma) *Schistosoma mansoni,* (Sme) *Schmidtea mediterranea,* (Cel) *Caenorhabditis elegans*, (Cbr) *Caenorhabditis briggsae*, (Ppa) *Pristionchus pacificus*, (Hro) *Helobdella robusta,* (Dme) *Drosophila melanogaster*, (Aae) *Aedes aegypti,* (Tca) *Tribolium castaneum,* (Isc) *Ixodes scapularis*, (Lgi) *Lottia gigantea,* (Cca) *Capitella capitata*, and (Api) *Acyrthosiphon pisum*. Conservation of residues is indicated below the columns as identical (asterisk), highly similar (colon), or moderately conserved (period). The following symbols were used for respective phyla: black boxes (Chordates), white boxes (Arthropods), blue boxes (Nematodes), white circles (Annelids), stars (Molluscs), and grey boxes (Platyhelminths).

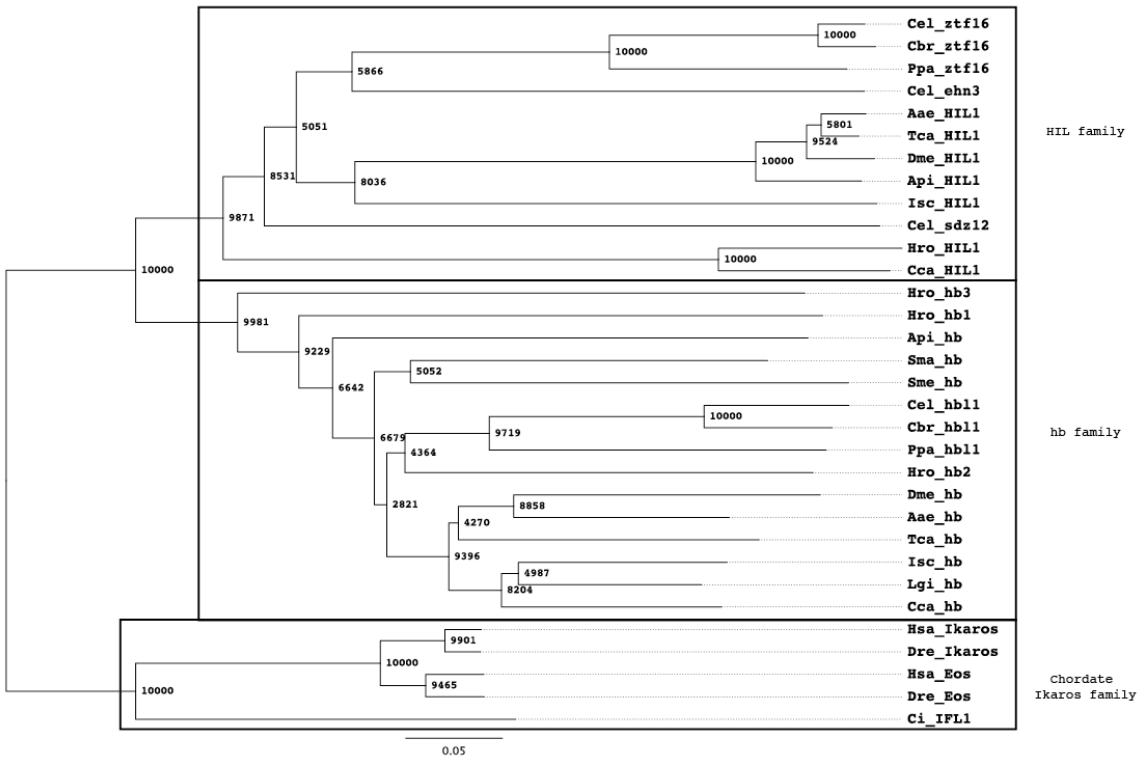


Figure S4. Neighbor-Joining tree.
The sequences from the preceding alignment (Fig. S3) were used to generate a phylogenetic tree using the Neighbor-Joining algorithm. Chordate Ikaros family members were used as an outgroup. Node values indicate the robustness over 10000 bootstrap replicates while the scale bar indicates an evolutionary distance of 0.05 amino acid substitutions per site. HIL and *hunchback* family members group into separate clades as indicated by labeled boxes.

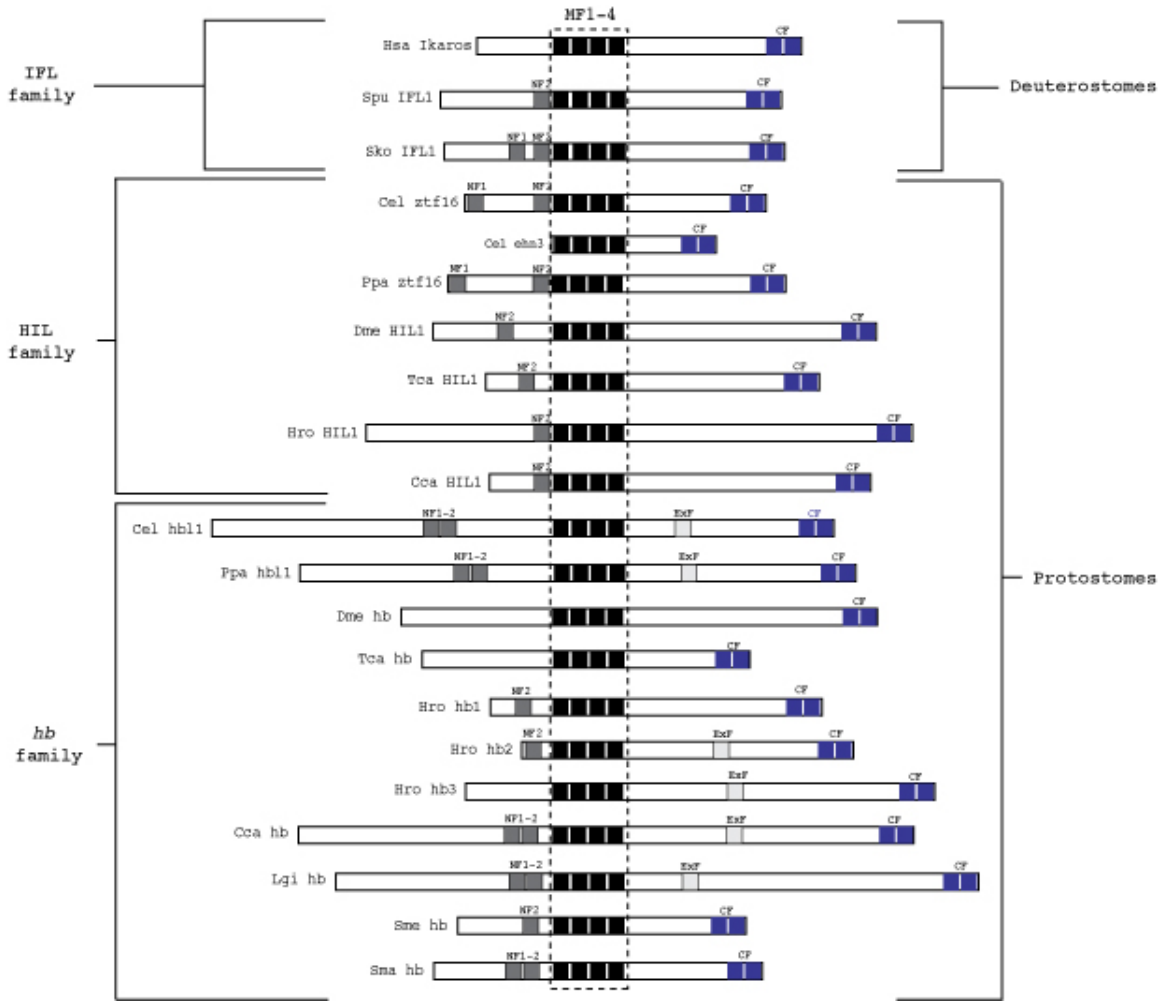


Figure S5. Organization of zinc fingers from representative protostome *hunchback* and HIL genes and deuterostome IFLs.
N-terminal fingers (NFs, grey), middle fingers (MFs, black), extra fingers (ExF, light grey), and C-terminal fingers (CFs, blue) are indicated. Note the absence of ExF motifs from HIL and Ikaros/IFL family members and the inconsistent presence and spacing of NF domains in all three families.

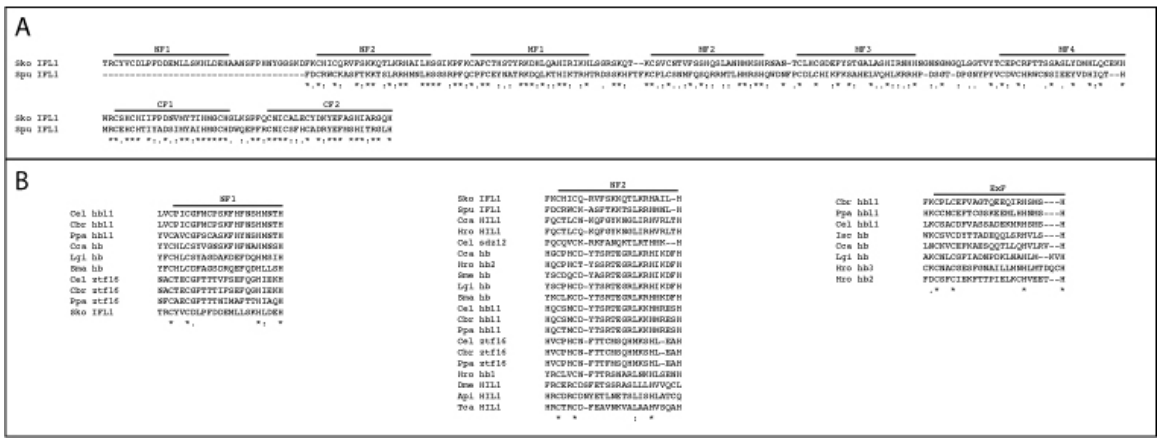


Figure S6. Additional alignments.
(A) NF, MF, and CF domains from the sea urchin *Strongylocentrotus purparatus* (Spu) and acorn worm *Saccoglossus kowalevskii* (Sko) IFL proteins were aligned using ClustalX. Conservation of residues is indicated below the columns as identical (asterisk), highly similar (colon), or moderately conserved (period). (B) N-terminal (NF1-2) and extra fingers (ExF) from HIL, *hunchback,* and IFL genes were aligned using ClustalX. ExF motifs were found only in *hunchback* family members.

CHAPTER THREE

hunchback and Ikaros-like zinc finger genes interact with chromatin factors to control reproductive system development in *Caenorhabditis elegans*

SUMMARY

The *Caenorhabditis elegans* reproductive system develops from two blast cells known as somatic gonadal precursors (SGPs). The SGPs are specified during embryogenesis and later give rise to all somatic tissues of the adult reproductive system. The *C. elegans hunchback* and Ikaros-like (HIL) family members *ztf-16* and *ehn-3* overlap in both expression and function to control SGP development. In order to find novel genes acting in SGP development we used a weak allele of *ehn-3* as the basis for a reverse genetic screen. Feeding RNAi was used to screen ~2,400 clones consisting of transcription factors, signaling components, and chromatin factors. Several candidate genes were found to be similar to components of the chromatin remodeling complexes SWI/SNF and NuRD, which physically interact with mammalian Ikaros family members in immune system development. We find that the SWI/SNF complex and the previously uncharacterized histone deaceylase *hda-2* control SGP development in parallel to *ehn*-3. In contrast, we define at least two phases for NuRD components acting both antagonistically and in parallel to *ehn-3*. Based on these results we provide the first evidence that 1) *hda-2* acts in complexes distinct from NuRD to control SGP development, 2) SWI/SNF and NuRD have opposing roles in development and 3) at least one HIL family member interacts genetically with chromatin factors.

INTRODUCTION

During the first larval stage (L1), all nematodes possess two somatic gonadal precursors (SGPs) and one or two primordial germ cells (PGCs) (CHITWOOD and CHITWOOD 1950). SGPs are multipotent blast cells that generate all somatic tissues of the reproductive system, while PGCs produce gametes. In *C. elegans* hermaphrodites, each SGP generates one of the two gonadal "arms" through identical and rotationally symmetrical cell lineages (KIMBLE and HIRSH 1979). We recently described two *C. elegans hunchback* and Ikaros-like (HIL) genes with overlapping functions in SGP development (LARGE and MATHIES 2010). The conserved HIL family member *ztf-16* and its partial gene duplicate *ehn-3* are expressed in embryonic SGPs and are redundantly required for their development into the differentiated tissues of the somatic gonad (LARGE and MATHIES 2010; MATHIES *et al.* 2003). This is reminiscent of the overlapping expression and similar functions of the Ikaros family of transcription factors, which share structural similarity with HIL family members.

Three Ikaros family members, Ikaros, Aiolos, and Helios are enriched in haematopoietic cells (GEORGOPOULOS *et al.* 1992; HAHM *et al.* 1998; MORGAN *et al.* 1997). Ikaros is the founding member of this gene family and is required for the development of lymphoid (i.e. T cells, B cells) lineage progression (GEORGOPOULOS *et al.* 1994). Mutations in Aiolos cause defects in B cell development (WANG *et al.* 1998) while a Helios mutant causes no obvious T cell defects despite being expressed primarily in T cells (CAI *et al.* 2009). It is thought that Ikaros and Aiolos act in parallel to one another in B cells while other Ikaros family members can compensate for the loss of Helios in T cells. Ikaros is the best studied of the three family members and is likely to control immune system lineage

progression through its association with chromatin remodeling complexes (CRCs) such as NuRD (Nucleosome Remodeling and Deacetylase) and SWI/SNF (SWItching defective and Sucrose Non Fermenting) (KIM *et al.* 1999; O'NEILL *et al.* 2000; SRIDHARAN and SMALE 2007).

SWI/SNF was the first CRC to be thoroughly characterized (CAIRNS *et al.* 1994; KRUGER *et al.* 1995) and was initially identified in genetic screens for mutations in *Saccharomyces cerevisiae* mating type switching and sucrose non-fermentation (NEIGEBORN and CARLSON 1984; STERN *et al.* 1984). SWI/SNF complexes are large (1 to 2MD) and contain eight or more subunits although a core complex of four subunits has chromatin remodeling activity *in vitro* (PHELAN *et al.* 1999). This enzymatic core is conserved between yeast, *Drosophila*, and mammals and consists of a central SNF2-type ATPase (yeast Swi2/Snf2 or Sth1, human hBRM or BRG1, and *Drosophila* Brahma), two SANT/SWIRM/Leu zipper-containing subunits (yeast Swi3 or Rsc8, human BAF155 or BAF170, and *Drosophila* Moira), and one subunit that contains both a coiled-coil domain and a repeat domain (yeast Snf5 or Sfh1, human hSNF5/INI1, and *Drosophila* SNR1) (MOHRMANN and VERRIJZER 2005; PHELAN *et al.* 1999). Two types of SWI/SNF CRCs are found in yeast, animals and plants and can be classified based upon unique accessory proteins. One class contains yeast SWI/SNF, *Drosophila* BAP (Brahma-associated proteins), and mammalian BAF (BRG1- or hBRM-associated factors) while the other includes yeast RSC (Remodel the Structure of Chromatin), *Drosophila* PBAP (Polybromo-associated BAP), and mammalian PBAF (Polybromo-associated BAF) (MOHRMANN *et al.* 2004; MOHRMANN and VERRIJZER 2005). In contrast to mammalian PBAP, which can alternate between one of

two ATPases (hBRM and BRG1), Drosophila BAP and PBAP share one ATPase known as Brahma (BRM) (JOHN W. TAMKUN 1992; SIMON and TAMKUN 2002). The distinguishing subunits in *Drosophila* are OSA in BAP and Polybromo, BAP170, and SAYP (supporter of activation of yellow protein) in PBAP (CHALKLEY *et al.* 2008; KWON 2007; MOHRMANN *et al.* 2004). PBAP and BAP also have different patterns of chromatin targeting and distinct developmental phenotypes (CARRERA *et al.* 2008; MOHRMANN *et al.* 2004). The full catalytic activity of SWI/SNF core components and the distinct chromatin distribution and phenotypes of BAP and PBAP suggests that accessory components allow for differential patterns of chromatin targeting for *Drosophila* BAP and PBAP. Although previous studies have found important roles for SWI/SNF components in *C. elegans* development (CUI *et al.* 2004; SAWA *et al.* 2000) unique functions have not been described for PBAP or BAP in *C. elegans*.

Several groups independently identified the vertebrate NuRD CRC (TONG *et al.* 1998; WADE *et al.* 1998; XUE *et al.* 1998; ZHANG *et al.* 1998). The largest subunit of each complex was one of two related Mi-2/CHD ATPases. Most complexes also contained histone deacetylases (HDAC1 and HDAC2), histone binding proteins (RbAp46 and RbAp48), metastasis-associated proteins (MTA1, MTA2, and MTA3) and methyl CpG-binding domain 3 (MBD3) proteins (TONG *et al.* 1998; WADE *et al.* 1998; XUE *et al.* 1998; ZHANG *et al.* 1998). The Mi-2 ATPases (Mi-2α and Mi-2β or CHD3 and CHD4, respectively) are part of the SNF2-type chromatin remodeling superfamily (EISEN *et al.* 1995) and are found in both plants and animals but have not been clearly identified in yeast. In mammals, NuRD components are known to be involved in lymphocyte development (FUJITA *et al.* 2004;

NAITO *et al.* 2007; WILLIAMS *et al.* 2004). In *C. elegans*, two Mi-2 homologs, *let-418* and *chd-3*, and two MTA1-like genes, *egl-27* and *lin-40,* have unique and overlapping developmental functions (SOLARI *et al.* 1999; VON ZELEWSKY *et al.* 2000) and several NuRD components antagonize RAS-dependent vulval development (SOLARI and AHRINGER 2000). The NuRD complex, unlike other CRCs, also contains HDAC components.

HDACs catalyze the removal of acetyl groups from the lysine tails of histones, which is generally associated with transcriptional repression (STRUHL 1998), and are found in several chromatin complexes such as Sin3, NuRD, and CoREST (GROZINGER and SCHREIBER 2002). HDACs are often divided into two groups, referred to as class I and class II, based upon sequence similarity. Of the class I proteins, HDAC1 and HDAC2 are the most well characterized and have at least one representative in vertebrates (HDAC1 and HDAC2), *Drosophila* (RPD3), and *C. elegans* (*hda-1* and *hda-3*) (GREGORETTI *et al.* 2004; GROZINGER and SCHREIBER 2002). *C. elegans* contains one additional class I protein, *hda-2*, which groups loosely with class I HDAC3 proteins in phylogenetic trees (GREGORETTI *et al.* 2004). While a single allele of *hda-1* has been characterized in *C. elegans* development (DUFOURCQ *et al.* 2002) no molecular lesions of either *hda-2* or *hda-3* have been described.

Here we use a reverse genetic screen to identify genes involved in SGP development. Several of the identified genes are found to have sequence similarity to SWI/SNF and NuRD components. Given the resemblance of *ehn-3* and *ztf-16* to Ikaros (LARGE and MATHIES 2010), we further characterize their interactions with SWI/SNF and NuRD. We find that PBAP acts in parallel to *ehn-3*, but not *ztf-16,* to control somatic gonad development and we identify a unique role for BAP that is independent of *ehn-3*. Our data further suggest that

both BAP and PBAB utilize the *psa-4/*Brahma ATPase. Finally, we find that a NuRD-like complex acts in opposition to *ehn-3* maternally and that the MTA1-like gene *egl-27* has unique roles in parallel with *ehn-3* in addition to roles more similar to NuRD.

MATERIALS AND METHODS

Strains

C. elegans strains were cultured as described previously (BRENNER 1974; WOOD 1988). All strains were grown at 20°C unless otherwise specified. *C. elegans* were derived from the wild isolate N2 (HODGKIN 1997). The following mutations were used in this study and are described in *C. elegans II* (HODGKIN 1997), cited references, or this work:

LGI: *pbrm-1(ok843), let-526(h185*) (JOHNSEN *et al.* 2000)*, C18E3.2(tm3395), hda-3(ok1991)*

LGII: *egl-27(ok1670), egl-27(we3)* (SOLARI *et al.* 1999)*, hda-2(ok1479), rol-6(e187), ehn-1(q690), ehn-1(q638)* (MATHIES *et al.* 2003)

LGIII: *tag-246(tm3309), snfc-5(ok622)*

LGIV: *ehn-3(q766)* (MATHIES *et al.* 2004), *ehn-3(q689)* (MATHIES *et al.* 2003), *ehn-3(rd2)* (LARGE and MATHIES 2010), *psa-4(os13*) (SAWA *et al.* 2000), *psa-4(tm305)*

LGV: *psa-1(os22*) (SAWA *et al.* 2000), *psa-1(ku355)* (CUI *et al.* 2004), *let-418(s1617)* (VON ZELEWSKY *et al.* 2000), *let-418(n3536)* (CEOL *et al.* 2006), *lin-40*(*ku285)* (CHEN and HAN 2001), *hda-1(e1795)* (DUFOURCQ *et al.* 2002)

LGX: *hnd-1(q740)* (MATHIES *et al.* 2003), *ztf-16(tm2127)* (LARGE and MATHIES 2010), *chd-3(eh4)* (VON ZELEWSKY *et al.* 2000)

We used the following GFP-marked balancer chromosomes: *eT1[qIs60]* for LGIII and LGV, *mIn1[dpy-10(e128) mIs14]* for LGII, *hT2[qIs48]* for *LGI* and *LGIII*, *nT1[qIs50]* for *LGIV* and *LGV*, and *eT1[qIs60]* for *LGIII* and *LGV*.

All deletion alleles obtained from the *C. elegans* Gene Knockout Consortium or National BioResource Project were backcrossed alternately to balancer chromosomes and N2 at least five times prior to analysis; the deletion allele was tracked using PCR. The *R07E5.3(ok622)* allele initially displayed gonadogenesis defects, but these defects were not maintained through seven backcrosses. Since the *R07E5.3/snfc-5* phenotype was expected to be embryonic lethal, we used PCR to screen for potential duplication events and were unable to detect any duplications using three primer sets.

Regulome RNAi screen

We compiled a set of ~2,400 genes consisting of transcription factors, cell signaling components, and chromatin factors from both the Vidal and Ahringer RNAi libraries. Specifically, the screen was composed of the published signal transduction, transcriptional regulation, and chromatin remodeling libraries (REECE-HOYES *et al.* 2005), the Ahringer chromatin, phosphatase, and transcription factor sub-libraries (http://www.geneservice.co.uk/products/rnai/datasheets/desp_info.jsp), known chromatin factors affecting *C. elegans* development (CUI and HAN 2007), and candidate cell adhesion proteins. Feeding RNAi was performed essentially as in Kamath (KAMATH *et al.* 2003). Two L4 worms were used for each well containing one RNAi clone. The worms were fed overnight in one well of a 12-well plate and were transferred to new wells in duplicate. The now gravid adult worms were allowed to lay eggs for a 24-hour period and the worms were subsequently removed from the plate. Gonad defects were assessed using a dissecting microscope and at least 50 animals were examined per well. In an initial pass, we scored as positive any clones that resulted in greater than 10% gonadogenesis defects in the

ehn-3(q766) background as compared to wild-type. All positives were rescreened and any that tested positive in the secondary screen were sequenced to verify the clone identity.

Subsequent RNAi was performed by injection of double-stranded RNA (dsRNA) into the gonad or intestine. dsRNA was synthesized using the Megascript T7 kit (Ambion) and injected at 0.2 - 1 mg/ml. Key chromatin regulators that were missing from the commercially available RNAi libraries were amplified and cloned into pPD129.36 (L4440). Template for RNA synthesis was derived from RNAi clones (where available) or from genomic DNA using primers flanking coding regions. In each case, the template contained at least 500 bp of coding sequence.

Phenotypic analysis

Three young adults were placed on each of three plates and 24-hour collections were taken. Gonad defects were assessed in the F1 generation using a dissecting microscope and the number of worms with missing gonadal arms, disorganized gonads ("white patch"), or no visible gonad ("gonadless") was recorded. The average penetrance and standard deviation (SD) of gonadogenesis defects (Gon) was calculated. Unpaired t-tests were used for pair wise statistical comparisons and the raw p-value is reported.

For *let-418(n3536)* temperature-shift experiments thirty L4 worms were allowed to develop to gravid adults at 20°C or 22.5°C for 24 hours. Worm plates were transferred to water maintained at 20°C or 22.5°C for one hour and then transferred to new plates maintained at 20°C or 22.5°C for one hour of egg laying. The thirty gravid adults were removed and the eggs were allowed to develop for 9.5 hours at 20°C or 8 hours at 22.5°C until hatching as L1 larvae. Worms were then either maintained at the same temperature or

shifted to 20°C or 22.5°C until the L4 stage when gonadogenesis defects were assessed via dissecting scope.

For *egl-27(we3)* tempterature-shift experiments thirty L4 worms were maintained at 15°C or 20°C for 24 hours until gravid adults and shifted to water maintained at 15°C for one hour or allowed to develop at a constant temperature. Worms were then transferred to plates maintained at 15°C for one hour of egg laying. The thirty gravid adults were removed and the eggs were allowed to develop until the L4 stage when gonadogenesis defects were evaluated with a dissecting scope.

***ehn-1/hda-2* cloning**

ehn-1(q638) had been previously mapped to the region between *unc-104* and *rol*-6 (MATHIES *et al.* 2003). Deficiency mapping subsequently placed *ehn-1* between *zyg-11* and *ltd-1*. Transformation rescue experiments indicated that the *ehn-1* mutant phenotype could be rescued by the cosmid C08B11, which includes the histone deacetylase *hda-2*. Since our RNAi screen identified *hda-2* as a genetic enhancer of *ehn-3*, we tested for complementation between *ehn-1* alleles and a deletion allele of *hda*-2. *ehn-1 rol-6* homozygotes were crossed with *hda-2(ok1479)/mIn1[mIs14]* males and non-Rol, non-GFP progeny were scored for gonadogenesis defects. From these crosses, 15.9% of *ehn-1(q690)/hda-2(ok1479)* trans-heterozygotes (*n*=170) and 27.6% of *ehn-1(q638)/hda-2(ok1749)* trans-heterozygotes (*n*=199) had gonadogenesis defects. By comparison, *ehn-1(q638) rol-6* crossed to wild-type males resulted in only 1% gonadogenesis defects (MATHIES *et al.* 2003). We amplified and sequenced the *hda-2* coding region from *ehn-1(q638)* and *ehn-1(q690)* mutants and identified molecular lesions in both alleles: *q638* is a C -> T transition that results in a premature stop

codon after amino acid 154 and *q690* is a C ->T transition that changes amino acid 298 from serine to leucine.

SWI/SNF and NuRD reporter constructs

BC12725 (*pbrm-1*::GFP) and BC13898 (*C08B11.3::GFP*) were acquired from the BC Gene Expression Consortium (HUNT-NEWBURY *et al.* 2007). Both reporters were integrated using gamma irradiation to create *rdIs31* (*pbrm-1*) and *rdIs19* and *rdIs20* (*C08B11.3*). The *pbrm-1* transcriptional reporter contains 2690 bp upstream of the start of translation; it is expressed ubiquitously in embryos, but is restricted to SGPs and DTCs and unidentified neurons during larval development. The *C08B11.3* transcriptional reporter contains 1195 bp upstream of the start of translation. It is expressed broadly in embryos, in larval SGPs, and in the intestine and unidentified head and tail neurons in larvae and adults. *C08B11.3* is the upstream gene in an operon with *hda-2*; therefore this reporter likely reflects the expression of both *C08B11.3* and *hda-2*. Additional integrated nuclear-localized transcriptional reporters were acquired from the Kim lab (LIU *et al.* 2009). SD1584 is an *egl-27* reporter that contains 4854 bp upstream of the start of translation and SD1588 is a *C08B11.3* reporter containing 2479 bp upstream of the start of translation.

Tissue-restricted RNAi strains

The *rde-1* genomic sequence including all exons, introns, and UTRs was cloned downstream of the *ehn-3*A promoter from pRA255 to create pRA277 and downstream of the *hnd-1* promoter from pJK848 (MATHIES *et al.* 2003) to create pRA279. pRA277 was injected into *rrf-3(pk1426); rde-1(ne219)* with pRF4 (MELLO *et al.* 1991) as a co-injection marker and integrated with gamma irradiation to create *rdIs6*. pRA279 was injected into

rrf-3(pk1426); unc-119(ed3); rde-1(ne219) with an *unc-119* rescue construct (MADURO and PILGRIM 1995) and integrated with gamma irradiation to create *rdIs7*. Both strains were backcrossed several times and the presence and functionality of the construct was verified with *pop-1(RNAi)*. The *rrf-3* mutation was subsequently removed during outcrossing and strain construction.

HDAC alignment

We aligned the full-length proteins of previously identified Class I HDAC family members from Humans (HDAC 1, 2, 3, and 8), *Drosophila* (Rpd3 and HDAC3), *S. cerevisiae* (Rpd3, Hos1, and Hos2), and *C. elegans* (HDA-1, 2, 3) (GREGORETTI *et al.* 2004) using Clustal X (LARKIN *et al.* 2007). The following proteins were used for the alignment: Hs HDAC1 (gi13128860), Hs HDAC2 (gi21411359), Hs HDAC3 (gi13128862), Hs HDAC8 (gi8923769), Dm Rpd3 (gi7292522), Dm HDAC3 (gi7296744), Sc Rpd3 (gi6323999), Sc Hos1 (gi6325325), Sc Hos2 (gi6321244), Ce HDA-1 (gi17561978), Ce HDA-2 (gi17534739), and Ce HDA-3 (gi17508561).

RESULTS

To identify additional genes acting in parallel to *ehn-3* and to learn more about how *ehn-3* controls SGP development in *C. elegans,* we used a weak allele of *ehn-3* to provide a sensitized background for an RNAi screen. *ehn-3(q766)* mutants are missing one of the two gonadal arms only 3% of the time, but they exhibit strong synergistic interactions with *hnd-1* and *tra-1* causing a nearly complete loss of gonadal arms in double mutants (MATHIES *et al.* 2003; MATHIES *et al.* 2004). Since *ehn-3* is likely to be involved in gene regulation, we limited our screen to transcription factors, signaling components, and chromatin factors (see Materials and Methods). L4 worms were placed on RNAi bacteria for 24 hours then transferred in duplicate to new bacteria for a 24-hour period. Their progeny were examined using a dissecting microscope for gonadogenesis defects (Gon), including missing gonadal arms, disorganized gonads, and no apparent gonad. Any RNAi clones that resulted in >10% gonadogenesis defects were rescreened and the clones sequenced to ascertain their identity. From a screen of ~2400 genes, we identified 33 *ehn-3* enhancers. Several of these were implicated in chromatin regulation, including five members of the SWI/SNF complex and two candidate NuRD subunits (Table 1). We recently reported sequence and functional similarities between EHN-3 and mammalian Ikaros (LARGE and MATHIES 2010), which physically associates with the NuRD and SWI/SNF complexes (KIM *et al.* 1999; O'NEILL *et al.* 2000; SRIDHARAN and SMALE 2007). Therefore, we focus here on characterizing the interactions between *ehn-3* and components of the SWI/SNF and NuRD complexes in *C. elegans*.

SWI/SNF functions in somatic gonad development

Our RNAi screen identified the core SWI/SNF subunit *psa-1/Moira* and four accessory subunits *pbrm-1/*Polybromo*, Y71H2AM.17/*BAP111*, C18E3.2/*BAP60, and *tag-246/*BAP60 as genetic enhancers of *ehn-3(q766)* (Table 1). RNAi of four other subunits, including *snfc-5/SNR1, psa-4/Brahma, let-526/*OSA and *ZK616.4/*BAP55*,* caused embryonic or larval lethality precluding an examination of the somatic gonad. Therefore, several SWI/SNF components act as genetic enhancers of *ehn-3(q766)*.

The *C. elegans* core SWI/SNF complex consists of the ATPase *psa-4/*Brahma, *snfc-5*/SNR1, and *psa-1*/Moira. To explore the function of SWI/SNF in gonadogenesis, we examined available alleles of *psa-1, psa-4,* and *snfc-5*. Temperature-sensitive alleles of *psa-1/*Moira and *psa-4/*Brahma were isolated in a screen for genes affecting asymmetric cell division in *C. elegans* (SAWA *et al.* 2000). We examined *psa-1(os22)* and *psa-4(os13)* using a dissecting microscope and observed temperature-dependent gonadogenesis defects (Table S1). A *psa-1(ku355)* allele isolated in a screen for genes synthetically lethal with *lin-35/Rb* (CUI *et al.* 2004) also had similar gonadogenesis defects (Table 2). We also examined a deletion allele of *psa-4*. This allele, *tm305*, removes 788 bp beginning in the ATPase domain and creates an out-of-frame fusion; it is therefore likely to be a null allele. *psa-4(tm305)* is homozygous lethal, but it exhibits a low penenetrance of gonadogenesis defects as a heterozygote (Table 2). Finally, we examined a deletion allele of *snfc-5/SNR1*. We were surprised to find that the *snfc-5* deletion was homozygous viable. Previous RNAi experiments (this work and SAWA *et al.* 2000) indicated that loss of *snfc-5* function results in embryonic lethality. Molecularly, the *snfc-5(ok622)* deletion removes almost the entire

snfc-5 coding region in addition to the promoter and first two exons of an adjacent gene, *rnp-4*. Possible explanations for the phenotypic difference between the deletion allele and *snfc-5* RNAi include off-target effects of RNAi and a gene duplication event similar to that seen with previous deletion screens (DERRY *et al.* 2001). However, PCR assays were unable to detect an additional copy of the *snfc-5* in the *snfc-5(ok622)* strain (see Materials and Methods) and there are no predicted off targets for this RNAi clone. Therefore, the nature of this allele remains in question. We conclude that at least two subunits of the SWI/SNF core complex, including the central ATPase, function in *C. elegans* somatic gonad development.

SWI/SNF complexes contain accessory subunits that are not necessary for chromatin remodeling activity, but are thought to be involved in target selection (KWON 2007). In *C. elegans,* these are *C18E3.*2/BAP60, *tag-246*/BAP60, *ZK616.4*/BAP55 and *Y71H2AM.17*/BAP111. Alleles of *C18E3.*2/BAP60, *tag-246*/BAP60, and *Y71H2AM.17*/BAP111 but not *ZK616.4*/BAP55 have been identified, but have not been previously described. We examined a deletion allele of *Y71H2AM.17*/BAP111, *tm3647*, which removes 81 bp upstream of the start of translation through the beginning of the third exon, including most of the HMG box. *Y71H2AM.17(tm3647)* could be maintained as a homozygote and it had no gonadogenesis defects (*n*=576). Next, we examined deletion/insertion alleles of *tag-246*/BAP60 and *C18E3.2/*BAP60. The *tag-246* allele, *tm3309*, creates a frame shift early in the protein coding region, while the *C18E3.2* allele, *tm3395*, removes about half of the third exon which encodes most of the protein; both alleles are likely to be strong loss-of-function or null alleles. All of the homozygous progeny of *C18E3.2(tm3395)* heterozygotes were sterile (100%, *n*=41) and neither heterozygotes

(*n*=142) nor homozygotes (*n*=41) had obvious gonadogenesis defects. Similarly, none of the progeny of heterozygous *tag-246(tm3309)* mutants exhibited gonad defects. However, progeny of homozygous *tag-246(tm3309)* mothers were sometimes missing one of the two gonadal arms indicating a maternal requirement for *tag-246* in gonadogenesis (Table 2). In addition, we observed significant embryonic and larval lethality (Table S2), indicating additional roles for *tag-246* during embryogenesis or larval development. Given that *tag-246* and *C18E3.2* are BAP60 paralogs, we considered the possibility that they are functionally redundant and created a strain heterozygous for both alleles. No gonadogenesis defects were observed in *tag-246(tm3309); C18E3.2(tm3395)* double mutants although maternal effects could not be evaluated due to the sterility of *C18E3.2(tm3395)*. Therefore the BAP60 paralogs are not zygotically redundant for gonadogenesis although at least one BAP60 paralog, *tag-246*, has maternal requirements for *C. elegans* gonadogenesis and essential roles in embryonic and larval development. In contrast, alleles of other accessory components *C18E3.*2/BAP60 and *Y71H2AM.17*/BAP111 did not cause gonadogenesis defects alone suggesting they may not be essential for SWI/SNF activity in somatic gonad development.

SWI/SNF complexes can be subdivided based upon the presence of unique accessory subunits. In *Drosophila,* these complexes are BAP and PBAP, in mammals they are BAF and PBAF, and in yeast they are SWI/SNF and RSC (MOHRMANN *et al.* 2004; MOHRMANN and VERRIJZER 2005). The defining subunit for *C. elegans* BAP is *let-526*/OSA (Masahiro Uchida and Hitoshi Sawa, pers. comm.) while the distinguishing subunits of PBAP are *pbrm-1/*Polybromo, *C08B11.3*/BAP170, and *F33E11.6*/SAYP. To ask which complex is involved in somatic gonad development, we examined alleles of *pbrm-1* and *let-526* as

representatives of PBAP and BAP, respectively. The *pbrm-1* deletion allele, *ok843*, removes two bromo domains and creates an out of frame fusion, while the *let-526* allele *h185* results in a premature stop codon that truncates the protein at amino acid 473 (Masahiro Uchida and Hitoshi Sawa, pers. comm.). We found that *pbrm-1(ok843)* mutants had minor gonadogenesis defects as both a heterozygote and a homozygote (Table 2). Most of the progeny of *pbrm-1(ok843)* homozygotes died as embryos or arrested development at the L1 larval stage (Table S2) and the surviving progeny had significant gonadogenesis defects. By contrast, we did not observe any gonadogenesis defects in *let-526(h185)* heterozygotes although we were unable to examine *let-526(h185)* homozygotes due to larval lethality. We conclude that both PBAP and BAP are required for larval development and PBAP may have additional roles in gonadogenesis.

SWI/SNF subunits act in parallel to *ehn-3*

Our RNAi screen utilized a weak allele of *ehn-3* and was therefore able to identify genes acting upstream, downstream, or in parallel with *ehn-3*. To ask if the SWI/SNF complex functions in parallel to *ehn-3*, we generated double mutants with select SWI/SNF components and a null allele of *ehn-3*. We found that double homozygotes of *ehn-3(rd2)* with *psa-1(os22)*, *psa-1(ku355), pbrm-1(ok843)*, *C18E3.2(tm3309),* and *tag-246(tm3309)* displayed non-additive effects (Table 2). Importantly, both the penetrance and the severity of the defects were greater in double mutants than in either single mutant, suggesting that the SWI/SNF complex acts in parallel to *ehn-3* (Fig.1, Table 2). This was particularly notable in *pbrm-1(ok843); ehn-3(rd2)* double mutants where 26.7% (*n*=101) had no visible gonad (Fig.1B, Table S3), whereas *pbrm-1(ok843)* and *ehn-3(rd2)* single mutants were almost never

gonadless. Furthermore, we found that *pbrm-1* dominantly enhanced the *ehn-3(rd2)* phenotype (Table 2). Therefore, *pbrm-1* has dose-sensitive and synergistic interactions with *ehn-3*. The defining subunit of BAP, *let-526/*OSA, did not dominantly enhance *ehn-3(rd2)*. Dominant enhancement of *ehn-3* by *pbrm-1* but not *let-526* suggests that PBAP has a stronger role in parallel to *ehn-3* than BAP.

The developmental roles of chromatin factors are often challenging to characterize due to the phenotypic severity and pleiotropy of loss-of-function mutations (KWON 2007). We used a tissue-specific RNAi strategy to circumvent the embryonic and larval lethality associated with the loss of core SWI/SNF components, to clarify the roles of PBAP and BAP, and to possibly illuminate roles for SWI/SNF at multiple times of development. Our tissue-specific RNAi system uses two lineage-restricted promoters to drive expression of *rde-1* in *rde-1* mutants, thus rescuing RNAi in these tissues (QADOTA *et al.* 2007). We used the *hnd-1* promoter to drive expression of *rde-1* broadly in the MS, C, and D lineages and later in the SGPs and the *ehn-3* promoter to drive expression solely in the SGPs but after the onset of *hnd-1* expression (LARGE and MATHIES 2010; MATHIES *et al.* 2003; MATHIES *et al.* 2004; WELCHMAN *et al.* 2007). We refer to these respective constructs as mesodermal RNAi and SGP RNAi.

A clear synergistic interaction was observed between *ehn-3* and *pbrm-1* using the mesodermal RNAi strain (Table 3). This synergism is reflected in both the penetrance and the severity of the phenotype. For example, 8.0% of *ehn-3(q766); pbrm-1(RNAi)* worms had no visible gonad, a phenotype rarely seen in *ehn-3(q766)* or in *pbrm-1(RNAi)* alone. This result is in agreement with our genetic analysis using null alleles of *ehn-3* and *pbrm-1* and

indicates that *pbrm-1* acts in parallel to *ehn-3*. SGP RNAi is expected to reveal functions later in SGP development and we observed a synergistic interaction between *ehn-3* and *pbrm-1* using the SGP RNAi treatment although this interaction was significantly less penetrant and less severe than the mesodermal RNAi phenotype. These results are consistent with the possibility that PBAP acts with *ehn-3* later in SGP development. Broad expression of *pbrm-1* during embryogenesis and more specific expression in larval SGPs and DTCs indicates *pbrm-1* may have additional roles during somatic gonad development including one in parallel to *ehn-3* (Fig. 1H-K). Using the mesodermal RNAi strains, we also found that *let-526(RNAi)* causes severe gonadogenesis defects typified by a disorganized patch of tissue with clear germline proliferation (Fig. 1G). These defects are distinct from those seen in *pbrm-1* RNAi and they do not appear to be strongly synergistic with *ehn-3*. Therefore, *let-526* plays important roles in SGP development that appear to be distinct from the roles of *pbrm-1*. Finally, we observed strongly synergistic interactions between *psa-4* and *ehn-3* using the mesodermal RNAi strain. Qualitatively, this interaction appears to reflect a stronger role for *psa-4* in parallel to *ehn-3* than either *prbm-1* or *let-526* (Fig 1). Therefore, the relative strength of contributions to the gonadogenesis defects in parallel to *ehn-3* can be represented by *psa-4* > *pbrm-1* > *let-526*.

***ehn-3* has distinct interactions with NuRD components**

Our RNAi screen also identified *egl-27/MTA1* and the class I histone deacetylase C08B11.2/*hda-2* as genetic enhancers of *ehn-3* (Table 1). Both MTA1 and class I HDACs are present in the NuRD complex therefore we explored the interaction between candidate NuRD components and *ehn-3*. We began by examining two alleles of *egl-27*. The *egl-27*

locus is complex and contains multiple transcripts that are derived from different promoters (HERMAN *et al.* 1999; SOLARI *et al.* 1999). These transcripts produce two major protein isoforms: the largest has similarity to MTA1 and the two shorter isoforms contain a glutamine rich domain and several arginine glutamic acid (RE) dipeptides. A transcriptional reporter for the promoter immediately upstream of the longest isoform of *egl-27*/*C04A2.3* (LIU *et al.* 2009) is broadly expressed in embryos in addition to expression in the SGPs and DTCs in later stages of development (Fig. 2B-E). The *ok1670* allele contains a 24 bp insertion and a 1747 bp deletion, which removes the SANT domain and GATA-type zinc finger (Fig. 2A). The deletion is predicted to produce the largest *egl-27* isoform but eliminates one of the smaller *egl-27* isoforms. The *we3* allele is cold-sensitive and contains a premature stop codon toward the end of the coding region. It is predicted to affect all *egl-27* isoforms (SOLARI *et al.* 1999). *egl-27(ok1670)* mutants do not have gonadogenesis defects and do not dominantly enhance *ehn-3(rd2)*. However, *egl-27(ok1670); ehn-3(rd2)* double homozygotes had severe gonadogenesis defects (Table 4). Furthermore, *egl-27(ok1670)* enhanced not only the penetrance, but also the severity of the *ehn-3(rd2)* defects (Table S3). This result indicates that the largest EGL-27 isoform acts in parallel to *ehn-3* to control somatic gonad development. Next, we examined *egl-27(we3)* at 15°C, which is the restrictive temperature. At 15°C, *egl-27(we3)* had a high percentage of embryonic and larval lethality (Table S2), as previously reported, and the few surviving worms had gonadogenesis defects (Table 5). Therefore, we were surprised to find that *egl-27(we3)* suppressed the defects seen in *ehn-3(rd2)* mutants. This result could indicate differences in the activity of the two EGL-27 isoforms or it could reflect differences in maternal versus zygotic activity.

The *egl-27(we3)* allele was maintained as a homozygote, therefore it reflects the loss of both maternal and zygotic product. To explore this possibility, we examined *egl-27(we3); ehn-3(rd2)* mutants derived from *egl-27(we3)* heterozygotes and found that they had a more severe phenotype than *ehn-3(rd2)* single mutants (data not shown). This result is in agreement with our analysis of the *egl-27(ok1670)* allele and suggests that maternal *egl-27* opposes *ehn-3*, while zygotic *egl-27* acts in parallel to *ehn-3*. In order to solidify this result, we shifted the temperature during the maternal phase of development for both *egl-27(we3); ehn-3(rd2)* and *egl-27(we3)* homozygotes. We were surprised to see that a wild-type maternal product was sufficient to rescue the maternal lethality (not shown) and it relieved the suppression of *ehn-3(rd2)* gonad defects (Fig. 3). This reinforces the observation that *egl-27* functions maternally to suppress *ehn-3(rd2)* gonadogenesis defects in addition to zygotic functions acting in parallel to *ehn-3(rd2)*.

There are two MTA-like genes in *C. elegans, egl-27* and *lin-40/egr-1*, and previous studies have found functional redundancy between them (SOLARI *et al.* 1999). Therefore, we also examined an allele of *lin-40*. Null alleles of *lin-40* are lethal or sterile so we used the partial loss-of-function allele *ku285* (CHEN and HAN 2001). *lin-40(ku285)* mutants did not have gonadogenesis defects alone and they did not enhance the *ehn-3(rd2)* phenotype when derived from a heterozygous parent. However, the progeny of *lin-40(ku285)*; *ehn-3(rd2)* worms had less severe defects than *ehn-3(rd2)*, indicating a maternal role for *lin-40* in opposition to *ehn-3* (Table 4). Together with our analysis of *egl-27* alleles, this suggests that MTA-like proteins play both positive and negative roles in parallel to *ehn-3* and it suggests that these functions can be separated based upon maternal versus zygotic requirement. Under

this model, maternal *lin-40* and *egl-27* oppose the activity of *ehn-3*, whereas zygotic *egl-27* but not *lin-40* acts in parallel to *ehn-3*.

The defining subunits of the NuRD complex are Mi-2 and MTA1. In order to better understand the interactions between *ehn-3* and NuRD, we examined alleles of the two *C. elegans* Mi-2 homologs, *let-418* and *chd-3* (VON ZELEWSKY *et al.* 2000). First, we examined strong loss-of-function or null alleles of *let-418* and *chd-3*. *let-418(s1617)* is a substitution that results in an early stop codon and *chd-3(eh4)* is a deletion that results in a frame-shift and subsequent premature stop codon (VON ZELEWSKY *et al.* 2000). *chd-3(eh4)* has no phenotype alone and it showed only moderate enhancement of the *ehn-3(rd2)* phenotype (Table 4). Similarly, *let-418(s1617)* showed modest enhancement of the *ehn-3(rd2)* phenotype as a heterozygote; *let-418(s1617)* homozygotes are larval lethal and were not examined. Next, we examined *let-418(n3536)*, which is a temperature-sensitive mutation that affects the ATPase/helicase domain of LET-418 (CEOL *et al.* 2006). We found that *let-418(n3536)* had temperature-dependent gonadogenesis defects. Similar to *lin-40* and *egl-27*, we found that *ehn-3(rd2); let-418(n3536)* double mutants had less severe defects than *ehn-3(rd2)* single mutants at all temperatures examined (Table 5). This was particularly notable at 22.5°C, where both *let-418(n3536)* and *ehn-3(rd2)* single mutants have gonadogenesis defects and the double mutant phenotype is clearly less than additive. We conclude that loss-of-function mutations in both *let-418/Mi-2* and *lin-40/MTA1* suppress the *ehn-3(rd2)* phenotype, suggesting that a NuRD complex acts antagonistically with *ehn-3*.

We used the temperature-sensitive allele of *let-418* to evaluate the requirement for *let-418* during maternal, embryonic, and post-embryonic development (Fig. 4). These data

reveal several interesting aspects of the interaction between *let-418* and *ehn-3*. First, we observed gonadogenesis defects in *let-418(n3536)* mutants only when they were shifted to 22.5°C during embryonic and post-embryonic development. Second, these defects appear to be suppressed by mutations in *ehn-3*, since this same temperature regimen resulted in approximately equivalent defects in *ehn-3(rd2); let-418(n3536)* double mutants and *ehn-3(rd2)* single mutants. Third, we observed a lower penetrance of defects in *ehn-3(rd2); let-418(n3536)* double mutants than in *ehn-3(rd2)* single mutants when they were shifted to 22.5°C during maternal and embryonic development. Fourth, *ehn-3(rd2); let-418(n3536)* double mutants derived from *let-418(n3536)* heterozygote parents had a more severe phenotype than their heterozygote siblings at 22.5°C (data not shown). Together, these observations suggest that maternal and embryonic *let-418* loss-of-function suppresses the defects caused by *ehn-3* loss-of-function and that *ehn-3(rd2)* suppresses the defects caused by embryonic and post-embryonic *let-418* loss-of-function. We interpret this to mean that *ehn-3(rd2)* and *let-418(n3536)* have mutually antagonistic functions at different times in somatic gonad development.

Finally, we examined alleles of all three *C. elegans* class I HDACs, *hda-1, hda-2,* and *hda-3*. We examined a deletion allele of *hda-2* that results in a frame shift and truncation of the protein at Ser169 (Fig. 5A-B). The *hda-2(ok1479)* deletion removes the most of the catalytic domain and is likely to be a null allele. We found that it had weak gonadogenesis defects on its own and significantly enhanced the *ehn-3(rd2)* defects (Table S4). The sole previously described class I HDAC was also examined alone or in conjunction with *ehn-3(rd2)*. *hda-1(e1795)* is a strong loss-of-function mutation caused by a missense

mutation in the catalytic domain of HDA-1 (DUFOURCQ *et al.* 2002). Nearly all *hda-1(e1795)* single mutants and *ehn-3(rd2)*; *hda-1(e1795)* double mutants had severe gonadogenesis defects that appeared qualitatively similar when derived from heterozygous parents (not shown). We noticed that a small percentage of *hda-1(e1795)* and *ehn-3(rd2); hda-1(e1795)* homozygotes were fertile (1.8%, *n*=109 and 1.5%, *n*=134, respectively) and we were able to establish two independent lines of each strain. High rates of embryonic lethality and no larval lethality was seen in *hda-1(e1795)* homozygotes while double *ehn-3(rd2)*; *hda-1(e1795)* mutants had less embryonic lethality and some larval lethality (Table S2). A similar penetrance of gonadogenesis defects was observed in single and double mutants although qualitatively it appeared that *ehn-3* and *hda-1* are mutually suppressing one another: *ehn-3(rd2)* suppresses the white patch phenotype of *hda-1(e1795)* and *hda-1(e1795)* suppresses the one-arm phenotype of *ehn-3(rd2*) (Table S3). Finally, we examined an uncharacterized allele of *hda-3*. The *hda-3(ok1991)* deletion is predicted to create an in-frame insertion/deletion that eliminates two of the five exons encoding 225 of the 465 amino acids and much of the catalytic domain encoded by *hda-3*. Single mutants of *hda-3(ok1991)* did not exhibit gonad defects alone (*n*=300) while double mutants of *ehn-3(rd2); hda-3(ok1991)* suppressed *ehn-3(rd2)* defects (Table 4). Similar to *let-418(n3536)* and *hda-1(e1795),* double *ehn-3(rd2); hda-3(ok1991)* mutants derived from *hda-3(ok1991)* heterozygotes had a more severe phenotype than *ehn-3(rd2)* single mutants. This indicates maternal *hda-3(ok1991*) is antagonistic to *ehn-3(rd2)* while zygotic *hda-3(ok1991)* acts in parallel to *ehn-3(rd2)*. We conclude that the previously

uncharacterized HDAC *hda-2* acts in parallel to *ehn-3* while *hda-1* and *hda-3* acts in opposition to *ehn-3*.

ehn-1* is *hda-2

Alleles of a gene called *ehn-1* (enhancer of *hnd-1*) cause a low penetrance of gonadogenesis defects, but they strongly enhance the phenotype of *ehn-3* and *hnd-1* mutants (MATHIES *et al.* 2003). We noticed that *ehn-1* maps to the same region as C08B11.2/*hda-2*. It was known that the *ehn-1* mutant phenotype could be rescued by transformation with the cosmid C08B11, but a smaller interval containing the gene had not been determined (L.M. unpublished observation). Complementation tests between the *ehn-1* alleles and the deletion allele *hda-2(ok1479)* suggested that *ehn-1* is *hda-2* (see Materials and Methods). We sequenced the *hda-2* coding region in both *ehn-1* alleles and found a nonsense mutation in *ehn-1(q638)* and a missense mutation in *ehn-1(q690)*. While *q638* results in a premature stop codon, *q690* results in the conversion of serine to leucine at amino acid 298 (Fig. 5A). This amino acid is either serine or a threonine in diverse histone deacetylases, raising the possibility that it is subject to phosphorylation. The identification of two mutations in the *hda-2* coding region combined with evidence from complementation tests and cosmid rescue experiments provides conclusive evidence that *ehn-1* is *hda-2.*

To examine the expression of *hda-2,* we obtained a transcriptional reporter for *C08B11.3* (LIU *et al.* 2009) and examined its expression in the somatic gonad. *hda-2* is the downstream gene in an operon with the SWI/SNF gene *C08B11.3/BAP170*. The *C08B11.3* reporter contains 2479 bp upstream of the start of translation and it is expressed broadly during embryogenesis (Fig. 5C-D). During larval development, it is expressed in SGPs (Fig.

5E) and their descendants, including the uterus and spermatheca in young adults (Fig. 5F). Together with our previous analysis of *ehn-1/hda-2* alleles, we conclude that the Class I HDAC *hda-2* is expressed broadly in the somatic gonad where it acts in parallel to *ehn-3* to control SGP development.

egl-27 has unique interactions with _ztf-16_

ehn-3 and *ztf-16* are both expressed in embryonic SGPs and they have overlapping functions in somatic gonad development. Therefore, we asked if either the SWI/SNF or NuRD complexes also interacted with *ztf-16*. *ztf-16* alleles do not have obvious defects in gonadal morphology, but they strongly enhance the *ehn-3* gonadogenesis defects (LARGE and MATHIES 2010). We chose *hda-2* and *egl-27* for this analysis since alleles of these genes significantly enhanced the *ehn-3* phenotype and *psa-4/Brahma* since it is the core SWI/SNF ATPase. We did not observe synergistic interactions between *psa-4(os13)* or *hda-2(ok1479)* and *ztf-16/R08E3.4(tm2127)* (Table 6). This suggests that SWI/SNF and HDA-2 do not act in parallel to *ztf-16*. Alternatively, *ehn-3* may be sufficient to act in parallel to these chromatin complexes. By contrast, despite the absence of gonad defects in either *egl-27(ok1670)* or *ztf-16(tm2127)* alone, *egl-27(ok1670)*; *ztf-16(tm2127)* double mutants had gonadogenesis defects on par with those observed in *ehn-3*; *ztf-16* double mutants (Table 6, Table S3). We find, therefore, that *egl-27* genetic interactions with *ehn-3* and *ztf-16* are distinct from *hda-2* and SWI/SNF suggesting that *egl-27* may represent a novel mutant class acting in parallel to *ehn-3* and *ztf-16* to control a different aspect of SGP development.

DISCUSSION

In this chapter we identify roles for the *C. elegans* SWI/SNF and NuRD chromatin remodeling complexes in somatic gonad development (Fig. 6). We provide evidence that SWI/SNF and *hda-2* act in parallel to *ehn-3* during embryonic development and we identify distinct phenotypes for defining members of BAP and PBAP. Temperature shift experiments and genetic analyses of candidate NuRD components define at least two phases of NuRD action and point to the possibility of multiple NuRD complexes acting antagonistically in SGP development. Furthermore, the molecular identify of *ehn-1* is revealed to be *hda-2* and we describe a loss-of-function allele that may provide insight into the molecular functions of HDAC proteins. Finally, we provide evidence for genetic interactions between chromatin factors and *hunchback* and Ikaros-like (HIL) genes in protostomes.

Distinct functions for PBAP and BAP in *C. elegans* somatic gonadogenesis

The SWI/SNF core components *psa-1/*Moira, *psa-4/*Brahma, and *snfc-5/*SNR1 are required for *C. elegans* hypodermal T cell development (SAWA *et al.* 2000) and *psa-1* and *snfc-5* are synthetically lethal with *lin-35*/Rb (CUI *et al.* 2004). Using genetic mutants and tissue-restricted RNAi, we show that *psa-1* and *psa-4* also function in SGP development in parallel to *ehn-3*. The gonad defects of temperature sensitive alleles of both *psa-1* and *psa-4* increase with rising temperature, which likely correlates with reduced activity of the PSA-1 and PSA-4 proteins. A deletion allele of *psa-4* was completely penetrant for embryonic lethality as a homozygote and exhibits dominant gonadogenesis defects as a heterozygote. Similar effects have been seen with homozygotes and heterozygotes of the mouse Brahma

ortholog Brg1 (BULTMAN *et al.* 2000). Together, these results support roles for the SWI/SNF complex in early in SGP development alone and in parallel to *ehn-3*.

SWI/SNF complexes can be subdivided based upon unique accessory components. While OSA is unique to BAP, Polybromo/BAP180, BAP170, and SAYP are unique to PBAP (CHALKLEY *et al.* 2008; KWON 2007; MOHRMANN *et al.* 2004). These subunits have not yet been investigated in *C. elegans*. Studies from *Drosophila* provide evidence for overlapping and unique roles for PBAP components, roles for OSA in viability, and distinct roles for the core complex independent of BAP or PBAP (CARRERA *et al.* 2008). Our results point to possible parallels in *C. elegans* SGP development. First, a less severe phenotype is seen with *pbrm-1/*Polybromo compared to core SWI/SNF members similar to non-lethal roles observed for PBAP components in *Drosophila*. Second, RNAi of *let-526/*OSA resulted in phenotypes that are distinct from *pbrm-1/*Polybromo mutants, in addition to larval lethality, suggesting that BAP and PBAP control different aspects of somatic gonad development. Third, *pbrm-1* RNAi caused a synthetic gonadless phenotype with *ehn-3*, while *let-526/*OSA RNAi did not. This provides evidence for a more important role for *pbrm-1* in parallel to *ehn-3*. Taken together our data suggest that PBAP and BAP 1) function during somatic gonad development, 2) have distinct functions with *ehn-3*, and 3) like Polybromo and OSA in *Drosophila* are required for viability.

Vertebrate SWI/SNF complexes contain 11 subunits that can sometimes be used interchangeably to create biological specificity in a process called combinatorial assembly (WU *et al.* 2009). For example, multiple BAP60 paralogs have been implicated in mammalian embryonic stem cell pluripotency and self-renewal, heart development, and the

establishment of left-right asymmetry (HO *et al.* 2009; LICKERT *et al.* 2004; TAKEUCHI and BRUNEAU 2009; TAKEUCHI *et al.* 2007). RNAi has revealed developmental roles for the two *C. elegans* BAP60 paralogs in the asymmetric division of the T cell (SAWA *et al.* 2000) and as synthetic lethal genes with the *C. elegans* Retinoblastoma orthologue, *lin-35*/Rb (CUI *et al.* 2004). Similarly, we show that both *tag-246*/BAP60 and *C18E3.2/BAP60* act in parallel to *ehn-3* during somatic gonad development. We also demonstrate that *tag-246/BAP60* has maternal requirements for gonadogenesis while *C18E3.2*/BAP60 causes sterility. Our work provides the first genetic evidence for distinct functions of BAP60 paralogs in *C. elegans*. The unique functions of *tag-246*/BAP60 and *C18E3.2*/BAP60 point to the potential for combinatorial SWI/SNF assembly outside vertebrate development.

Antagonistic roles for the *C. elegans* NuRD complex in somatic gonad development

We provide genetic evidence for at least four NuRD components acting in opposition to *ehn-3*. Alleles of *let-418*/Mi-2, *lin-40*/MTA1, *hda-1*, and *hda-3* suppress *ehn-3* gonadogenesis defects as homozygotes but not as the homozygous progeny of heterozygotes. This suggests these four NuRD components suppress *ehn-3(rd2)* gonad defects maternally and may, in turn, all be acting in the same complex (Fig. 6). Interestingly, we observed different maternal and zygotic activities for several NuRD components. A null allele of *let-418*/Mi-2 showed slight dominant enhancement of *ehn-3* while a temperature-sensitive allele of *let-418(n3536)* revealed opposing embryonic and post-embryonic roles for *let-418/*Mi-2 in gonadogenesis. Double mutants of *ehn-3(rd2)* and *let-418(n3536)* shifted to the restrictive temperature during embryonic and post-embryonic development have greater defects than double mutants raised at a constant temperature. In contrast, temperature shift

experiments of *let-418(n3536)* during the maternal and embryonic stage caused a suppression of *ehn-3(rd2)* defects. Based on these results, *let-418* appears to be acting antagonistically to *ehn-3* maternally and in parallel with *ehn-3* zygotically. Different alleles of *egl-27* also exhibited opposing functions during gonadogenesis. A deletion allele of *egl-27* acts in parallel to *ehn-3* to cause severe gonad defects while a temperature sensitive allele *egl-27* exhibited maternal suppression and zygotic enhancement of *ehn-3*. Therefore, at least two NuRD components act antagonistically to *ehn-3* maternally and in parallel to *ehn-3* zygotically.

Our genetic data indicates several NuRD components may regulate similar sets of genes with antagonistic activities at different times of development but this is not a clear indication that these proteins are all part of the same complex. NuRD contains multiple subunits, each with uncharacterized or diverse functions that contribute to the activation or repression of genes. Vertebrate NuRD complexes can contain an Mi-2α or Mi-2β ATPase, one of three mutually exclusive MTA proteins, and two different methyl-binding proteins in addition to histone deacetylases (HDAC1 and HDAC2) and histone binding proteins (RbAp46 and RbAp48) (BOWEN *et al.* 2004; DENSLOW and WADE 2007; FENG and ZHANG 2001; LE GUEZENNEC *et al.* 2006). The potential exists, therefore, for multiple NuRD-like complexes defined by a central Mi-2 ATPase acting as either transcriptional activators or repressors at different times in development. Whether or not the general trend of these NuRD-like complexes is to regulate sets of genes with opposing activities at different times of development awaits further investigation.

Are EGL-27 and HDA-2 NuRD components?

Previous studies have suggested that *egl-27* is an MTA-like gene in *C. elegans* and functional overlap with *lin-40*/MTA1 supports this idea. However, we find that *egl-27* has genetic interactions that are distinct from other NuRD components, suggesting that these proteins may not always act in the same complex. Although *lin-40*/MTA, *hda-1*/HDAC, and *hda-3* act in opposition to *ehn-3* maternally, only *egl-27* has strong synergistic interactions with *ehn-3* zygotically. In addition, only *egl-27* had genetic interactions with *ztf-16*. The *egl-27* gene encodes the BAH, SANT, ELM, and GATA domains commonly found in MTA proteins (SOLARI *et al.* 1999), but it also encodes arginine and glutamic acid dipeptide (RERE) repeats and is proline rich, both characteristic of human RERE and *Drosophila* Atrophin (WANG and TSAI 2008). Furthermore, a phylogenetic analysis of MTA family members places EGL-27 closer to Human RERE/Atrophin (BOWEN *et al.* 2004; WANG and TSAI 2008). One idea is that *egl-27* encodes a bifunctional protein(s) with characteristics of both MTA1 and Atrophin. Ultimately, biochemical approaches will be required to provide more insight into the structural nature and physical associations of EGL-27 in *C. elegans.*

HDAC1 and HDAC2 are class I HDACs that are both members of the NuRD complex in mammals. The *C. elegans* genome encodes three class I HDACs, which are more phylogenetically divergent (GREGORETTI *et al.* 2004). Previous research defined a role for *hda-1* in somatic gonad development (DUFOURCQ *et al.* 2002). We define a novel maternal function for *hda-1* (and *hda-3*) in opposition to *ehn-3*. This function is similar to *lin-40*/MTA and *let-418*/Mi-2 NuRD subunits suggesting *hda-1*/HDAC*, hda-3*, *lin-40*/MTA,

and *let-418*/Mi-2 all act in the same complex. We find that *hda-2/ehn-1*, unlike *hda-1* and *hda-3*, has independent functions in somatic gonad development and dramatically enhances *ehn-3* gonad defects. The phenotype of *hda-2* mutants is clearly distinct from *lin-40/MTA1, let-418/Mi-2, hda-1,* and *hda-3* since the latter suppress the *ehn-3* phenotype, while *hda-2* enhances it. Our results strengthen support for HDA-1 and HDA-3 as members of the NuRD complex and suggest that HDA-2 may be part of a chromatin complex that is distinct from NuRD.

How might SWI/SNF, NuRD, and EHN-3 regulate genes in *C. elegans*?

We observed opposing functions of SWI/SNF and NuRD during *C.elegans* SGP development. Both SWI/SNF and NuRD components acted genetically with *ehn-3* to positively or negatively regulate gonadogenesis. The opposing effects of SWI/SNF and NuRD have been similarly observed on mammalian promoters (GAO *et al.* 2009; RAMIREZ-CARROZZI *et al.* 2006). Ikaros can facilitate the recruitment of NuRD or SWI/SNF CRCs with opposing functions to specific loci during T cell lineage progression (NAITO *et al.* 2007). It is thought that Ikaros or Ikaros family members bind the CD4 locus and, in turn, recruit both positive and negative regulators of CD4. SWI/SNF and HDACs may initially inhabit the compact chromatin, which is then overridden by the recruitment of NuRD and HATs. It is possible that variations of Ikaros family member multimers serve as the molecular signposts for SWI/SNF and NuRD recruitment. In *C. elegans*, the HIL family members EHN-3 and ZTF-16 may act similarly during SGP development to facilitate NuRD and SWI/SNF recruitment. Given the shared ancestry with Ikaros, one interpretation of the opposing roles of SWI/SNF and NuRD components in *C. elegans* SGP development is that

individual or multiple HIL family members are recruiting different complexes at different times of development to currently unknown loci.

We have argued that HIL genes share a common origin with *hunchback* and Ikaros and have found that the majority of protostome genomes contain at least one HIL and one *hunchback* gene (LARGE and MATHIES 2010). Therefore, the *hunchback* and HIL genes may have retained different characteristics of the Ikaros gene family in deuterostomes, including physical interactions with specific chromatin factors such as NuRD and SWI/SNF (KIM *et al.* 1999; O'NEILL *et al.* 2000; SRIDHARAN and SMALE 2007). We observed strong synergistic genetic interactions between *ehn-3* and components of the SWI/SNF complex, indicating a role for SWI/SNF in parallel to *ehn-3*. However, SGP-specific RNAi showed a less prominent enhancement of *ehn-3* suggesting that SWI/SNF might also act in association with EHN-3 later in SGP development. Biochemical studies will be required to test this idea. On the other hand, physical interactions between HIL family members and NuRD may have been lost or reduced in *C. elegans* along with DNA methylation. NuRD complexes contain either MBD2 or MBD3 (WADE *et al.* 1999; ZHANG *et al.* 1999). MBD2 recognizes methylated DNA (HENDRICH and BIRD 1998) while MBD3 has lost this ability (HENDRICH and TWEEDIE 2003). The association of Ikaros with Mi-2 in mammalian lymphocytes may therefore serve as a regulatory link between DNA methylation and chromatin modifications. One member of the methyl-CpG-binding domain (MBD) protein family, *mbd-2*, remains in the *C. elegans* genome but it lacks a functional methyl-binding domain and has variable mutant phenotypes in different nematode species (GUTIERREZ and SOMMER 2007). The ancient function of this protein was to presumably link DNA methylation to chromatin

remodeling complexes such as NuRD. Studies in *Drosophila,* where DNA methylation has not been completely lost, may provide more insight into the physical relationship between the protostome HIL family and the NuRD complex.

ACKNOWLEDGEMENTS

We thank Dana Byrd for the *rde-1* genomic sequence, Shohei Mitani and the National BioResource Project for deletion alleles, the *Caenorhabditis* Genetics Center for strains, and the Triangle Worm Group for RNAi clones. We are grateful to Pat Estes for comments on the manuscript and Zigrida Smith for technical assistance. E.E.L. was supported by a GAANN Biotechnology fellowship and a predoctoral fellowship from the National Institutes of Health (GM-08443-10). This research was supported by a grant from the National Science Foundation (IOS-0617897) to L.D.M.

REFERENCES

BOWEN, N. J., N. FUJITA, M. KAJITA and P. A. WADE, 2004 Mi-2/NuRD: multiple complexes for many purposes. Biochim Biophys Acta **1677:** 52-57.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

BULTMAN, S., T. GEBUHR, D. YEE, C. LA MANTIA, J. NICHOLSON *et al.*, 2000 A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol Cell **6:** 1287-1295.

CAI, Q., A. DIERICH, M. OULAD-ABDELGHANI, S. CHAN and P. KASTNER, 2009 Helios deficiency has minimal impact on T cell development and function. J Immunol **183:** 2303-2311.

CAIRNS, B. R., Y. J. KIM, M. H. SAYRE, B. C. LAURENT and R. D. KORNBERG, 1994 A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. Proc Natl Acad Sci U S A **91:** 1950-1954.

CARRERA, I., J. ZAVADIL and J. E. TREISMAN, 2008 Two subunits specific to the PBAP chromatin remodeling complex have distinct and redundant functions during *Drosophila* development. Mol Cell Biol **28:** 5238-5250.

CEOL, C. J., F. STEGMEIER, M. M. HARRISON and H. R. HORVITZ, 2006 Identification and classification of genes that act antagonistically to *let-60* Ras signaling in *Caenorhabditis elegans* vulval development. Genetics **173:** 709-726.

CHALKLEY, G. E., Y. M. MOSHKIN, K. LANGENBERG, K. BEZSTAROSTI, A. BLASTYAK *et al.*, 2008 The transcriptional coactivator SAYP is a trithorax group signature subunit of the PBAP chromatin remodeling complex. Mol Cell Biol **28:** 2920-2929.

CHEN, Z., and M. HAN, 2001 Role of *C. elegans lin-40* MTA in vulval fate specification and morphogenesis. Development **128:** 4911-4921.

CHITWOOD, B. B., and M. B. CHITWOOD, 1950 *Introduction to Nematology*. University Park Press, Baltimore.

CUI, M., D. S. FAY and M. HAN, 2004 *lin-35*/Rb cooperates with the SWI/SNF complex to control *Caenorhabditis elegans* larval development. Genetics **167:** 1177-1185.

CUI, M., and M. HAN, 2007 Roles of chromatin factors in *C. elegans* development. WormBook**:** 1-16.

DENSLOW, S. A., and P. A. WADE, 2007 The human Mi-2/NuRD complex and gene regulation. Oncogene **26:** 5433-5438.

DERRY, W. B., A. P. PUTZKE and J. H. ROTHMAN, 2001 *Caenorhabditis elegans* p53: role in apoptosis, meiosis, and stress resistance. Science **294:** 591-595.

DUFOURCQ, P., M. VICTOR, F. GAY, D. CALVO, J. HODGKIN *et al.*, 2002 Functional requirement for histone deacetylase 1 in *Caenorhabditis elegans* gonadogenesis. Mol Cell Biol **22:** 3024-3034.

EISEN, J. A., K. S. SWEDER and P. C. HANAWALT, 1995 Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res **23:** 2715-2723.

FENG, Q., and Y. ZHANG, 2001 The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev **15:** 827-832.

FUJITA, N., D. L. JAYE, C. GEIGERMAN, A. AKYILDIZ, M. R. MOONEY *et al.*, 2004 MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell **119:** 75-86.

GAO, H., K. LUKIN, J. RAMIREZ, S. FIELDS, D. LOPEZ *et al.*, 2009 Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. Proc Natl Acad Sci U S A **106:** 11258-11263.

GEORGOPOULOS, K., M. BIGBY, J. H. WANG, A. MOLNAR, P. WU *et al.*, 1994 The Ikaros gene is required for the development of all lymphoid lineages. Cell **79:** 143-156.

GEORGOPOULOS, K., D. D. MOORE and B. DERFLER, 1992 Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science **258:** 808-812.

GREGORETTI, I. V., Y. M. LEE and H. V. GOODSON, 2004 Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J Mol Biol **338:** 17-31.

GROZINGER, C. M., and S. L. SCHREIBER, 2002 Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. Chem Biol **9:** 3-16.

GUTIERREZ, A., and R. J. SOMMER, 2007 Functional diversification of the nematode mbd2/3 gene between *Pristionchus pacificus* and *Caenorhabditis elegans*. BMC Genet **8:** 57.

HAHM, K., B. S. COBB, A. S. MCCARTY, K. E. BROWN, C. A. KLUG *et al.*, 1998 Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes Dev **12:** 782-796.

HENDRICH, B., and A. BIRD, 1998 Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol **18:** 6538-6547.

HENDRICH, B., and S. TWEEDIE, 2003 The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet **19:** 269-277.

HERMAN, M. A., Q. CH'NG, S. M. HETTENBACH, T. M. RATLIFF, C. KENYON *et al.*, 1999 EGL-27 is similar to a metastasis-associated factor and controls cell polarity and cell migration in *C. elegans*. Development **126:** 1055-1064.

HO, L., J. L. RONAN, J. WU, B. T. STAAHL, L. CHEN *et al.*, 2009 An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A **106:** 5181-5186.

HODGKIN, J., 1997 Appendix 1. Genetics, pp. 881-1047 in *C. elegans II*, edited by D. L. RIDDLE, T. BLUMENTHAL, B. J. MEYER and J. R. PRIESS. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

HUNT-NEWBURY, R., R. VIVEIROS, R. JOHNSEN, A. MAH, D. ANASTAS *et al.*, 2007 High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. PLoS Biol **5:** e237.

JOHNSEN, R. C., S. J. JONES and A. M. ROSE, 2000 Mutational accessibility of essential genes on chromosome I(left) in *Caenorhabditis elegans*. Mol Gen Genet **263:** 239-252.

KAMATH, R. S., A. G. FRASER, Y. DONG, G. POULIN, R. DURBIN *et al.*, 2003 Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature **421:** 231-237.

KEHLE, J., D. BEUCHLE, S. TREUHEIT, B. CHRISTEN, J. A. KENNISON *et al.*, 1998 dMi-2, a *hunchback*-interacting protein that functions in polycomb repression. Science **282:** 1897-1900.

KIM, J., S. SIF, B. JONES, A. JACKSON, J. KOIPALLY *et al.*, 1999 Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity **10:** 345-355.

KIMBLE, J., and D. HIRSH, 1979 The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Developmental Biology **70:** 396-417.

KOSTAS, S. A., and A. FIRE, 2002 The T-box factor MLS-1 acts as a molecular switch during specification of nonstriated muscle in *C. elegans*. Genes Dev **16:** 257-269.

KRUGER, W., C. L. PETERSON, A. SIL, C. COBURN, G. ARENTS *et al.*, 1995 Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. Genes Dev **9:** 2770-2779.

KWON, C., DORIS WAGNER, 2007 Unwinding chromatin for development and growth: a few genes at a time. Trends Genetics **23:** 403-412.

LARGE, E. E., and L. D. MATHIES, 2010 *hunchback* and Ikaros-like zinc finger genes control reproductive system development in *Caenorhabditis elegans*. Dev Biol **339:** 51-64.

LARKIN, M. A., G. BLACKSHIELDS, N. P. BROWN, R. CHENNA, P. A. MCGETTIGAN *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948.

LE GUEZENNEC, X., M. VERMEULEN, A. B. BRINKMAN, W. A. HOEIJMAKERS, A. COHEN *et al.*, 2006 MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol **26:** 843-851.

LICKERT, H., J. K. TAKEUCHI, I. VON BOTH, J. R. WALLS, F. MCAULIFFE *et al.*, 2004 Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature **432:** 107-112.

LIU, X., F. LONG, H. PENG, S. J. AERNI, M. JIANG *et al.*, 2009 Analysis of cell fate from single-cell gene expression profiles in *C. elegans*. Cell **139:** 623-633.

MADURO, M., and D. PILGRIM, 1995 Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics **141:** 977-988.

MATHIES, L. D., S. T. HENDERSON and J. KIMBLE, 2003 The *C. elegans* Hand gene controls embryogenesis and early gonadogenesis. Development **130:** 2881-2892.

MATHIES, L. D., M. SCHVARZSTEIN, K. M. MORPHY, R. BLELLOCH, A. M. SPENCE *et al.*, 2004 TRA-1/GLI controls development of somatic gonadal precursors in *C. elegans*. Development **131:** 4333-4343.

MELLO, C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J **10:** 3959-3970.

MOHRMANN, L., K. LANGENBERG, J. KRIJGSVELD, A. J. KAL, A. J. HECK *et al.*, 2004 Differential targeting of two distinct SWI/SNF-related *Drosophila* chromatin-remodeling complexes. Mol Cell Biol **24:** 3077-3088.

MOHRMANN, L., and C. P. VERRIJZER, 2005 Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim Biophys Acta **1681:** 59-73.

MOLLER, A., F. W. AVILA, J. W. ERICKSON and H. JACKLE, 2005 Drosophila BAP60 is an essential component of the Brahma complex, required for gene activation and repression. J Mol Biol **352:** 329-337.

MORGAN, B., L. SUN, N. AVITAHL, K. ANDRIKOPOULOS, T. IKEDA *et al.*, 1997 Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. EMBO J. **16:** 2004-2013.

NAITO, T., P. GOMEZ-DEL ARCO, C. J. WILLIAMS and K. GEORGOPOULOS, 2007 Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression. Immunity **27:** 723-734.

NEIGEBORN, L., and M. CARLSON, 1984 Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. Genetics **108:** 845-858.

O'NEILL, D. W., S. S. SCHOETZ, R. A. LOPEZ, M. CASTLE, L. RABINOWITZ *et al.*, 2000 An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. Mol Cell Biol **20:** 7572-7582.

PHELAN, M. L., S. SIF, G. J. NARLIKAR and R. E. KINGSTON, 1999 Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol Cell **3:** 247-253.

QADOTA, H., M. INOUE, T. HIKITA, M. KOPPEN, J. D. HARDIN *et al.*, 2007 Establishment of a tissue-specific RNAi system in *C. elegans*. Gene **400:** 166-173.

RAMIREZ-CARROZZI, V. R., A. A. NAZARIAN, C. C. LI, S. L. GORE, R. SRIDHARAN *et al.*, 2006 Selective and antagonistic functions of SWI/SNF and Mi-2beta nucleosome remodeling complexes during an inflammatory response. Genes Dev **20:** 282-296.

REECE-HOYES, J. S., B. DEPLANCKE, J. SHINGLES, C. A. GROVE, I. A. HOPE *et al.*, 2005 A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol **6:** R110.

SAWA, H., H. KOUIKE and H. OKANO, 2000 Components of the SWI/SNF complex are required for asymmetric cell division in *C. elegans*. Mol Cell **6:** 617-624.

SIMON, J. A., and J. W. TAMKUN, 2002 Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. Curr Opin Genet Dev **12:** 210-218.

SOLARI, F., and J. AHRINGER, 2000 NURD-complex genes antagonise Ras-induced vulval development in *Caenorhabditis elegans*. Curr Biol **10:** 223-226.

SOLARI, F., A. BATEMAN and J. AHRINGER, 1999 The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatic regulatory complex, and are redundantly required for embryonic patterning. Development **126:** 2483-2494.

SRIDHARAN, R., and S. T. SMALE, 2007 Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. J Biol Chem **282:** 30227-30238.

STERN, M., R. JENSEN and I. HERSKOWITZ, 1984 Five SWI genes are required for expression of the HO gene in yeast. J Mol Biol **178:** 853-868.

STRUHL, K., 1998 Histone acetylation and transcriptional regulatory mechanisms. Genes Dev **12:** 599-606.

SULSTON, J. E., and H. R. HORVITZ, 1977 Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Developmental Biology **56:** 110-156.

SULSTON, J. E., E. SCHIERENBERG, J. G. WHITE and J. N. THOMSON, 1983 The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Developmental Biology **100:** 64-119.

TAKEUCHI, J. K., and B. G. BRUNEAU, 2009 Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature **459:** 708-711.

TAKEUCHI, J. K., H. LICKERT, B. W. BISGROVE, X. SUN, M. YAMAMOTO *et al.*, 2007 Baf60c is a nuclear Notch signaling component required for the establishment of left-right asymmetry. Proc Natl Acad Sci U S A **104:** 846-851.

TAMKUN, J.W., R. DEURING, M.P. SCOTT, M. KISSINGER, A.M. PATTATUCCI, T.C. KAUFMAN, AND J.A. KENNISON, 1992 *brahma*: A Regulator of *Drosophila* Homeotic Genes Structurally Related to the Yeast Transcriptional Activator SNF2/SWI2. Cell **68:** 561-572.

TONG, J. K., C. A. HASSIG, G. R. SCHNITZLER, R. E. KINGSTON and S. L. SCHREIBER, 1998 Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature **395:** 917-921.

VON ZELEWSKY, T., F. PALLADINO, K. BRUNSCHWIG, H. TOBLER, A. HAJNAL *et al.*, 2000 The *C. elegans* Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. Development **127:** 5277-5284.

WADE, P. A., A. GEGONNE, P. L. JONES, E. BALLESTAR, F. AUBRY *et al.*, 1999 Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet **23:** 62-66.

WADE, P. A., P. L. JONES, D. VERMAAK and A. P. WOLFFE, 1998 A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. Curr Biol **8:** 843-846.

WANG, J. H., N. AVITAHL, A. CARIAPPA, C. FRIEDRICH, T. IKEDA *et al.*, 1998 Aiolos regulates B cell activation and maturation to effector state. Immunity **9:** 543-553.

WANG, L., and C. C. TSAI, 2008 Atrophin proteins: an overview of a new class of nuclear receptor corepressors. Nucl Recept Signal **6:** e009.

WELCHMAN, D. P., L. D. MATHIES and J. AHRINGER, 2007 Similar requirements for CDC-42 and the PAR-3/PAR-6/PKC-3 complex in diverse cell types. Dev Biol **305:** 347-357.

WILLIAMS, C. J., T. NAITO, P. G. ARCO, J. R. SEAVITT, S. M. CASHMAN *et al.*, 2004 The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. Immunity **20:** 719-733.

WOOD, W. B., 1988 Introduction to *C. elegans* Biology, pp. 1-16 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

WU, J. I., J. LESSARD and G. R. CRABTREE, 2009 Understanding the words of chromatin regulation. Cell **136:** 200-206.

XUE, Y., J. WONG, G. T. MORENO, M. K. YOUNG, J. COTE *et al.*, 1998 NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell **2:** 851-861.

ZHANG, Y., G. LEROY, H. P. SEELIG, W. S. LANE and D. REINBERG, 1998 The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell **95:** 279-289.

ZHANG, Y., H. H. NG, H. ERDJUMENT-BROMAGE, P. TEMPST, A. BIRD *et al.*, 1999 Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev **13:** 1924-1935.

Table 1. Chromatin factors interact genetically with *ehn-3*

S. cerivisiae	*H. sapiens*	*D. melanogaster*	*C. elegans*	*ehn-3* enhancer[a]
SWI/SNF complex				
Swi3, Rsc8	BAF155, BAF170	Moira	*psa-1*	Yes
Swi2/Snf2, Sth1	hBRM, BRG1	Brahma	*psa-4*	Emb[b]
			C52B9.8	No
Snf5, Sfh1	INI1	SNR1	R07E5.3	Emb
None	BAF57	BAP111	Y71H2AM.17	Yes
Arp7, Arp9	BAF53	BAP55	ZK616.4	Emb
Swp73, Rsc6	BAF60a	BAP60	C18E3.2	Yes
			tag-246	Yes
Swi1	hOSA1, hOSA2	OSA	*let-526*	Let
Rsc1, Rsc2, Rsc4	BAF180	Polybromo	*pbrm-1*	Yes
Rsc9	BAF200	BAP170	C08B11.3	Emb
none	BAF45a	SAYP	F33E11.6	Not tested
NuRD complex				
N/A	Mi-2	dMi-2	*let-418*	No
			chd-3	No
N/A	MTA1	MTA1	*lin-40/egr-1*	No
			egl-27[c]	Yes
Rpd3	HDAC1/2	DmHDAC1/2	*hda-1*	Emb
			hda-3	Not tested
	HDAC3	DmHDAC3	*hda-2*[d]	Yes

[a] Feeding RNAi was performed in an *ehn-3(q766)* background and greater than 10% penetrance was scored as positive; Emb=embryonic lethal, Let=Larval lethal

[b] The *psa-4* clone was determined to be incorrect; we generated a new clone and found that *psa-4(RNAi)* caused embryonic lethality similar to previous observations (SAWA *et al.* 2000).

[c] *egl-27* was originally identified as an MTA1 component but more recent phylogenies place it closer to Atrophin, which is not a part of the NuRD complex

[d] Recent *hda-2* phylogenies place it closer to HDAC3, which is present in N-CoR/SMRT (Silencing Mediator of Retinoic acid and Thyroid hormone receptor) complexes but not NuRD complexes.

Table 2. SWI/SNF acts alone and in parallel to *ehn-3* during somatic gonad development

Genotype	% Gon[a] +/- SD	*n*	P[b]
ehn-3(rd2)	22.0 +/- 2.7	441	
psa-4(os13)	2.0 +/- 0.7	1132	
psa-4(tm305)/hT2g	1.1 +/- 0.9	447	
psa-1(os22ts)	5.0 +/- 1.5	636	
psa-1(ku355)	3.7 +/- 1.2	295	
ehn-3(rd2); psa-1(ku355)	54.4 +/- 3.8	406	***
ehn-3(rd2); psa-1(os22ts)	45.8 +/- 14.0	236	*
snfc-5(ok622) [m-, z-]	0.1 +/- 0.1	1125	
snfc-5(ok622);ehn-3(rd2)	21.3 +/- 6.1	441	NS
pbrm-1(ok843)/hT2g	0.5 +/- 1.1	205	
pbrm-1(ok843) [m+, z-]	1.7 +/- 3.3	60	
pbrm-1(ok843) [m-, z-]	26.5 +/- 12.7	34	
pbrm-1(ok843)/hT2g; ehn-3(rd2)	50.0 +/- 4.9	484	***
pbrm-1(ok843); ehn-3(rd2) [m+, z-]	82.2 +/- 4.9	101	***
let-526(h185)/hT2g[c]	0.0 +/- 0.0	332	
let-526(h185)/hT2g ;ehn-3(rd2)[c]	22.4 +/- 7.4	361	NS
tag-246(tm3309) [m- z-]	10.7 +/- 2.1	327	
tag-246(tm3309); ehn-3(rd2) [m+, z-]	68.6 +/- 19.5	70	**
tag-246(tm3309); ehn-3(rd2) [m-, z-]	91.1 +/- 7.2	56	***
Y71H2AM.17(tm3647)	0.0 +/- 0	576	
Y71H2AM.17(tm3647); ehn-3(rd2)	27.0 +/- 5.8	626	NS
C18E3.2(tm3309) [m+, z-]	0.0 +/- 0.0	142	
C18E3.2(tm3309)/hT2g; ehn-3(rd2)	32.5 +/- 5.7	624	*
C18E3.2(tm3309); ehn-3(rd2)	86.1 +/- 7.3	151	***

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.
[b] Unpaired t-tests were used for statistical comparisons; *ehn-3(rd2)* was compared and the significance is indicated (NS=not significant, $p<0.05$*, $p<0.01$**, $p<0.001$***).
[c] *let-526(h185)* is linked to *dpy-5(e61) unc-13(e450)*

Table 3. Different roles for *C. elegans* PBAP and BAP during gonadogenesis

Genotype	RNAi	Tissue-specificity	% Gon[a] +/- SD	*n*
ehn-3(q766); rdIs6[b]	None	SGPs	3.1 +/- 3.0	131
ehn-3(q766); rdIs7[c]		mesoderm	1.4 +/- 1.6	145
N2	*psa-4*	none	Emb	169
rdIs6[b]		SGPs	1.9 +/- 2.2	266
ehn-3(q766); rdIs6[b]			15.2 +/- 1.3	204
rdIs7[f]		mesoderm	99.0 +/- 1.1	99
ehn-3(q766); rdIs7[c]			95.2 +/- 2.5	335
N2	*pbrm-1*	none	Let	105
rdIs6[e]		SGPs	0.5 +/- 0.5	194
ehn-3(q766); rdIs6[b]			13.9 +/- 3.1	462
rdIs7[c, d]		mesoderm	10.2 +/- 5.9	137
ehn-3(q766); rdIs7[c]			56.8 +/- 10.8	88
N2	*let-526*	none	Emb	78
rdIs6[b]		SGPs	51.5 +/- 13.2	439
ehn-3(q766); rdIs6[b]			79.5 +/- 11.7	200
rdIs7[c]		mesoderm	100 +/- 0	379
ehn-3(q766); rdIs7[c]			97.7 +/- 2.9	656

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.
[b] genotype includes *rde-1(ne219); rdIs6* [*ehn-3::rde-1*]
[c] genotype includes *unc-119(ed3); rde-1(ne219); rdIs7* [*hnd-1::rde-1*]
[d] genotype includes *ccIs4443* [*arg-1::GFP*]

Table 4. Candidate NuRD subunits are modifiers of the *ehn-3* phenotype

Genotype	% Gon[a] +/- SD	*n*	p[b]
egl-27(ok1670)	none	>200	
egl-27(ok1670)/mC6g; ehn-3(rd2)	25.1 +/- 3.4	934	NS
egl-27(ok1670); ehn-3(rd2) [m+, z-]	91.3 +/- 5.5	369	***
egl-27(we3) [m-, z-]	0 +/- 0	249	
egl-27(we3); ehn-3(rd2) [m-, z-]	25.4 +/- 1.1	279	NS
lin-40(ku285)	none	>200	
lin-40(ku285); ehn-3(rd2) [m+, z-]	14.6 +/- 9.8	82	NS
lin-40(ku285); ehn-3(rd2) [m-, z-]	5.8 +/- 2.0	589	***
ehn-3(rd2); let-418(s1617)/eT1g[c]	28.0 +/- 2.9	182	*
let-418(n3536)	0.2 +/- 0.2	1576	
let-418(n3536); ehn-3(rd2)	14.6 +/- 3.3	803	*
chd-3(eh4)	none	>200	
ehn-3(rd2); chd-3(eh4)	27.9 +/- 3.9	433	*
hda-1(e1975) [m-, z-]	11.6 +/- 5.4	381	
ehn-3(rd2); hda-1(e1975) [m-, z-]	10.1 +/- 5.1	267	*
hda-2(ok1479)	11.1 +/- 3.8	641	
hda-2(ok1479); ehn-3(rd2)	84.9 +/- 10.3	272	***
hda-3(ok1991)	0 +/- 0	300	
hda-3(ok1991); ehn-3(rd2) [m+, z-]	27.8 +/- 5.7	316	**
hda-3(ok1991); ehn-3(rd2) [m-, z-]	11.6 +/- 0.7	1039	**

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.
[b] Unpaired t-tests were used for statistical comparisons; *ehn-3(rd2)* was compared and the significance is indicated (NS=not significant, $p<0.05$*, $p<0.01$**, $p<0.001$***).
[c] *let-418(s1617)* is linked to *unc-46(e177)*

Table 5. *egl-27* and *let-418* have opposing roles in parallel with *ehn-3*

Genotype	% Gon[a] +/- SD	*n*
ehn-3(rd2) 15°C	21.8 +/- 3.7	193
ehn-3(rd2) 20°C	22.0 +/- 2.7	441
ehn-3(rd2) 22.5°C	25.0 +/- 8.1	328
egl-27(we3) 15°C	33.3 +/- 16.3	21
egl-27(we3) 20°C	0 +/- 0	249
egl-27(we3); ehn-3(rd2) 15°C	12.3 +/- 1.8	995
egl-27(we3); ehn-3(rd2) 20°C	25.4 +/- 1.1	279
let-418(n3536) 15°C	0 +/- 0	604
let-418(n3536) 20°C	0.2 +/- 0.2	1576
let-418(n3536) 22.5°C	11.7 +/- 2.4	1242
ehn-3(rd2); let-418(n3536) 15°C	15.3 +/- 3.3	464
ehn-3(rd2); let-418(n3536) 20°C [m+, z-]	26.5 +/- 18.3	68
ehn-3(rd2); let-418(n3536) 20°C [m-, z-]	14.6 +/- 3.3	803
ehn-3(rd2); let-418(n3536) 22.5°C [m+, z-]	26.7 +/- 9.5	60
ehn-3(rd2); let-418(n3536) 22.5°C [m-, z-]	11.5 +/- 1.1	998

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

Table 6. *egl-27* has unique interactions with *ztf-16*

Genotype	% Gon[a] +/- SD	*n*
egl-27(ok1670)	0.0 +/- 0.0	>200
ztf-16(tm2127)	0.0 +/- 0.0	>200
egl-27(ok1670); ztf-16(tm2127)	44.3 +/- 6.4	176
hda-2(ok1479)	11.1 +/- 3.8	641
hda-2(ok1479); ztf-16(tm2127)	12.2 +/- 5.6	534
psa-4(os13)	3.0 +/- 0.9	435
psa-4(os13); ztf-16(tm2127)	3.0 +/- 0.9	1022

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

Table S1. *psa-4(os13)* and *psa-1(os22)* have temperature-sensitive defects

Temp	Genotype	% Gon[a] +/- SD	*n*
15° C	*psa-4(os13)*	0.8 +/- 0.4	638
	psa-1(os22)	0.6 +/- 0.2	647
20° C	*psa-4(os13)*	2.0 +/- 0.1	1132
	psa-1(os22)	5.0 +/- 1.5	636
22.5° C	*psa-4(os13)*	3.6+/- 3.4	140
	psa-1(os22)	Emb	>100

[a] Gonadogenesis defects were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

Table S2. Maternal effect lethality of SWI/SNF and NuRD subunits

Genotype	% Emb[a] +/- SD	*n*	% Lva[a] +/- SD	*n*
tag-246(tm3309)	46.3 +/- 18.1	105	27.3 +/- 12.8	62
pbrm-1(ok843)	15.4 +/- 1.0	312	44.8 +/- 6.5	907
hda-1(e1795)	74.2 +/- 11.3	676	N.A.	0
ehn-3(rd2); hda-1(e1795)	28.2 +/- 7.7	132	13.5 +/- 4.1	63
egl-27(we3)[b]	86.7% +/- 6.9	467	7.1 +/- 3.1	467

[a] Embryonic (Emb) and larval (Lva) lethality were assessed using a dissecting microscope

[b] *egl-27(we3)* was examined at 15°C

Table S3. Qualitative differences between SWI/SNF and NuRD interactions

Genotype	Two arm[a]	One arm[a]	WP[a]	Gon[a]	*n*
ehn-3(rd2)	75.8	24.2	0	0	434
ehn-3(rd2); ztf-16(tm2127)	37.8	53.3	0	9	612
SWI/SNF					
ehn-3(rd2); psa-1(os22)	54.2	39	0	6.8	236
ehn-3(rd2); psa-1(ku355)	45.6	34	0	20.4	406
pbrm-1(ok843)/hT2g; ehn-3(rd2)	50	47.5	0	2.5	484
pbrm-1(ok843); ehn-3(rd2)	17.8	55.5	0	26.7	101
tag-246(tm3309)/hT2g; ehn-3(rd2)	72	21.5	0.8	5.6	354
tag-246(tm3309); ehn-3(rd2) [m+, z-]	31.4	28.6	2.9	37.1	70
tag-246(tm3309); ehn-3(rd2) [m-, z-]	3.1	10.9	0	86	578
let-526(h185)/hT2g; ehn-3(rd2)	77.6	22.5	0	0	361
NuRD					
egl-27(ok1670); ehn-3(rd2)	8.7	25	4.1	62.3	369
egl-27(ok1670); ztf-16(tm2127)	55.7	31.8	9.1	3.4	176
hda-1(e1795) [m+, z-]	1.8	0	98.2	0	109
hda-1(e1795) [m-, z-]	86.9	0	13.1	0	381
ehn-3(rd2); hda-1(e1795) [m+, z-]	0.7	1.5	97.8	0	135
ehn-3(rd2); hda-1(e1795) [m-, z-]	89.9	7.4	2.2	0.4	267
ehn-3(rd2); lin-40(ku285) [m+, z-]	85.4	13.4	0	1.2	82
ehn-3(rd2); lin-40(ku285) [m-, z-]	94.2	4.4	0.2	1.2	589

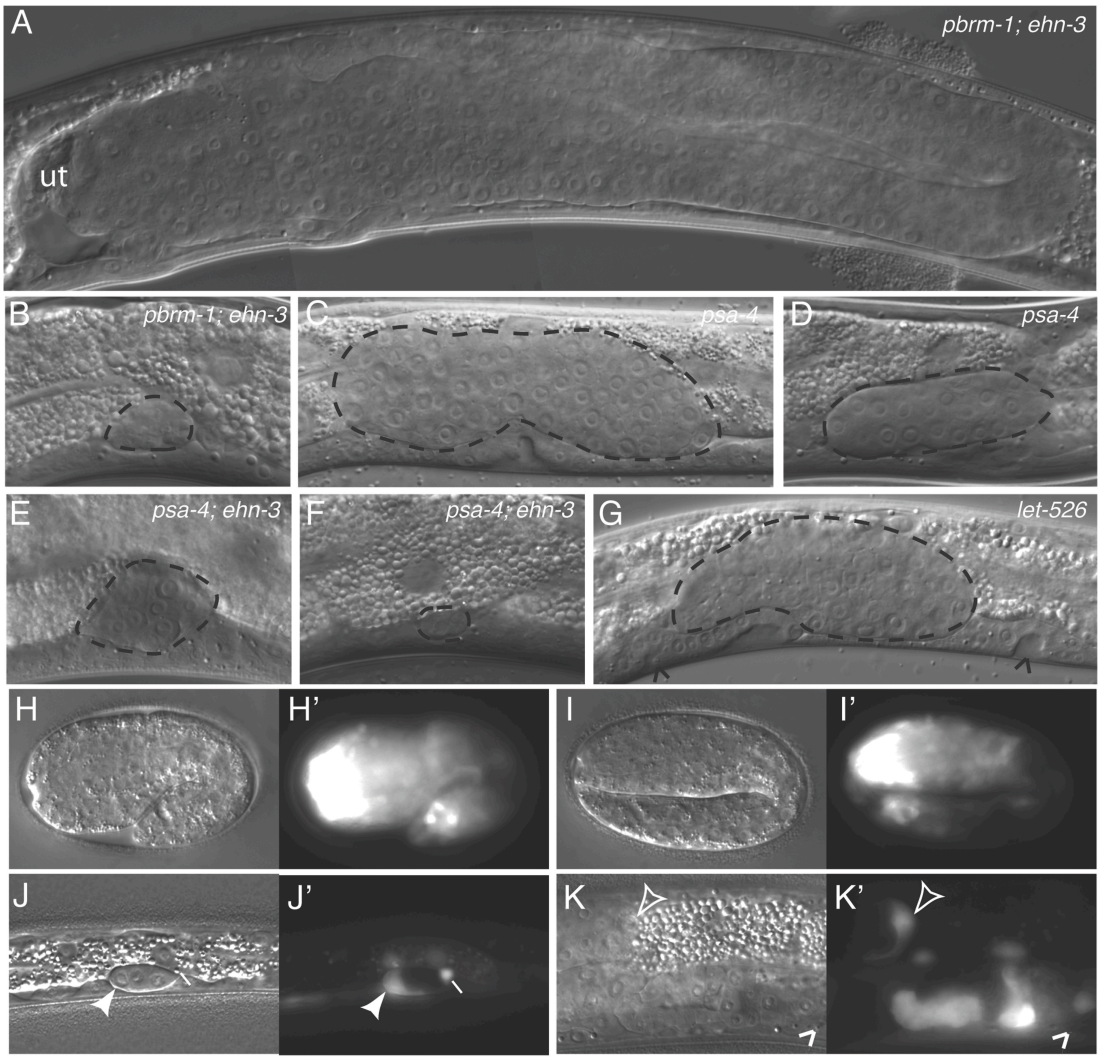
[a] Gonadogenesis defects were assessed using a dissecting microscope. The percentages of two arm, one arm, white patch (WP), and gonadless (Gon) animals are reported.

Table S4. *ehn-1/hda-2* has strong synergistic interactions with *ehn-3*

Genotype	% Gon[a] +/- SD	*n*
hda-2(ok1479)	11.1 +/- 3.8	641
ehn-3(q766)	3.1 +/- 2.3	382
hda-2(ok1479); ehn-3(q766)	70.0 +/- 1.7	310
ehn-3(rd2)	22.0 +/- 2.7	441
hda-2(ok1479); ehn-3(rd2)	84.9 +/- 10.3	272

[a] Gonadogenesis defects (Gon) were assessed using a dissecting microscope. The average penetrance and standard deviation (SD) are reported.

Figure 1. SWI/SNF and *ehn-3* have a synergistic phenotype.
Representative phenotypes of *pbrm-1(ok843); ehn-3(rd2)* (A-B), mesodermal *psa-4(RNAi)* in wild-type (C-D) and *ehn-3(q766)* backgrounds (E-F), and mesodermal *let-526(RNAi)* in a wild-type background (G). *pbrm-1(ok843); ehn-3(rd2)* double mutants can be wild-type but they are often missing one gonadal arm (A) and can be almost completely gonadless (B). Mesodermal *psa-4(RNAi)* causes a high penetrance of disorganized gonads (C, D), which are more severe in the *ehn-3(q766)* background, including disorganized gonads (E) and essentially gonadless worms (F). In (F) there is a small patch of tissue that may be the remnant of the gonad. Mesodermal *let-526(RNAi)* results in large disorganized gonads (G). These worms also had ectopic vulval development at sites that could be distant from the gonad (carats). *pbrm-1* is expressed at multiple stages of SGP and somatic gonad development (H-K); DIC images (H'-K') and *pbrm-1*::GFP fluorescence (H'-I'). Broad expression of *pbrm-1*::GFP is seen in embryos (J'). Expression of *pbrm-1*::GFP is seen in Z1 in L1 hermaphrodites (arrowhead) and in one of two hermaphrodite-specific neurons (HSNs) indicated by the bar (K'). Expression of *pbrm-1*::GFP in the DTC of an L3 hermaphrodite (open arrowhead) and in somatic gonadal tissues near the vulva (carat) (K').


A
pbrm-1; ehn-3
ut
B
pbrm-1; ehn-3
C
psa-4
D
psa-4
E
psa-4; ehn-3
F
psa-4; ehn-3
G
let-526
H
H'
I
I'
J
J'
K
K'

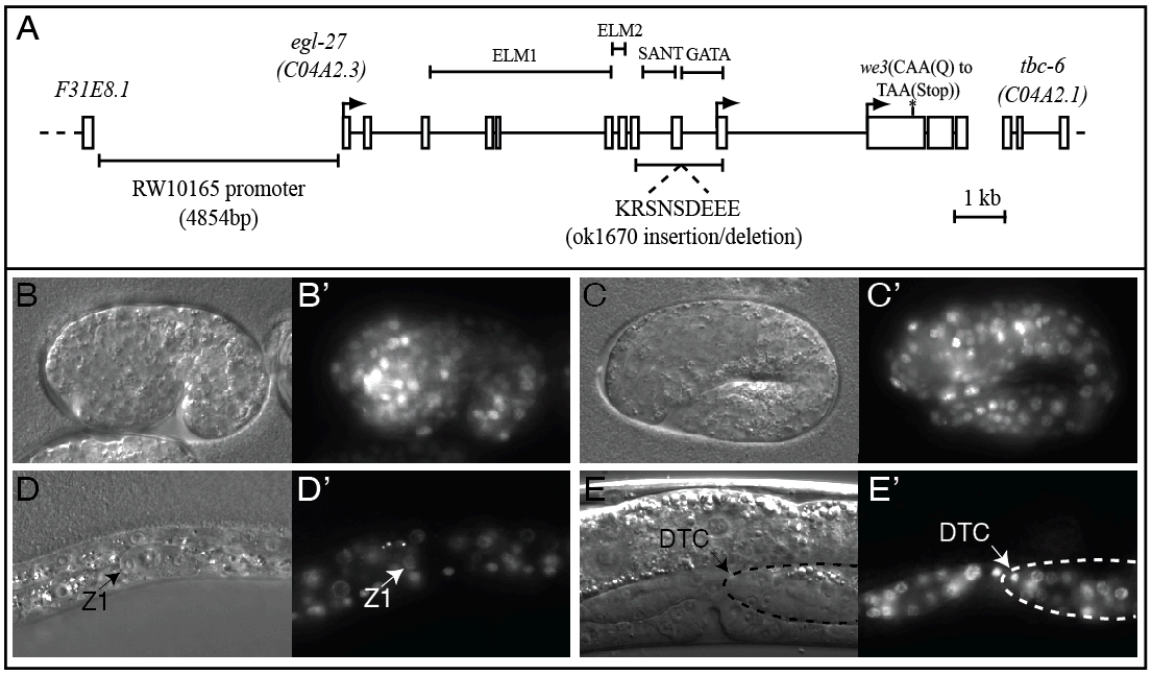


Figure 2. *egl-27* is expressed at multiple stages of SGP and somatic gonad development (A) Gene structure of *egl-27* showing multiple transcripts (arrows) and exons (horizontal boxes). *egl-27(we3)* is an early stop (asterisk) while *egl-27(ok1670)* is a deletion with a nine amino acid insertion indicated with brackets and dashed lines. (B-E) DIC images and (B'-E') *egl-27*::HIS::mCherry fluorescence. Broad expression of *egl-27*::HIS::mCherry is seen in precomma (B) and two-fold (C) embryos. Expression of *egl-27*::HIS::mCherry persists in L1 hermaphrodite SGPs and is indicated by the arrowhead (C'). L4 hermaphrodite expression of *egl-27*::HIS::mCherry is denoted by an arrowhead in the DTC of the posterior arm (E').

Figure 3. Maternal *egl-27* and *ehn-3* have antagonistic functions in somatic gonad development.
(A) *ehn-3(rd2),* (B) *egl-27(we3)* and (C) *egl-27(we3); ehn-3(rd2)* were raised at 15°C or 20°C during maternal (M), embryonic (E), or larval (L) development. Boxes indicate the temperature shift regimen (red, 15°C; grey, 20°C). The percentage of worms with gonadogenesis defects is shown with standard deviation. (A) *ehn-3(rd2)* shows no temperature-sensitivity. (B) *egl-27(we3)* has gonadogenesis defects when reared at 15°C continuously but not when raised at 15°C during embryonic and larval development. (C) *egl-27(we3); ehn-3(rd2)* double mutants have less severe gonadogenesis defects than *egl-27(we3)* or *ehn-3(rd2)* single mutants when reared at 15°C continuously. This effect is not as pronounced if they are reared at 20°C during maternal development.

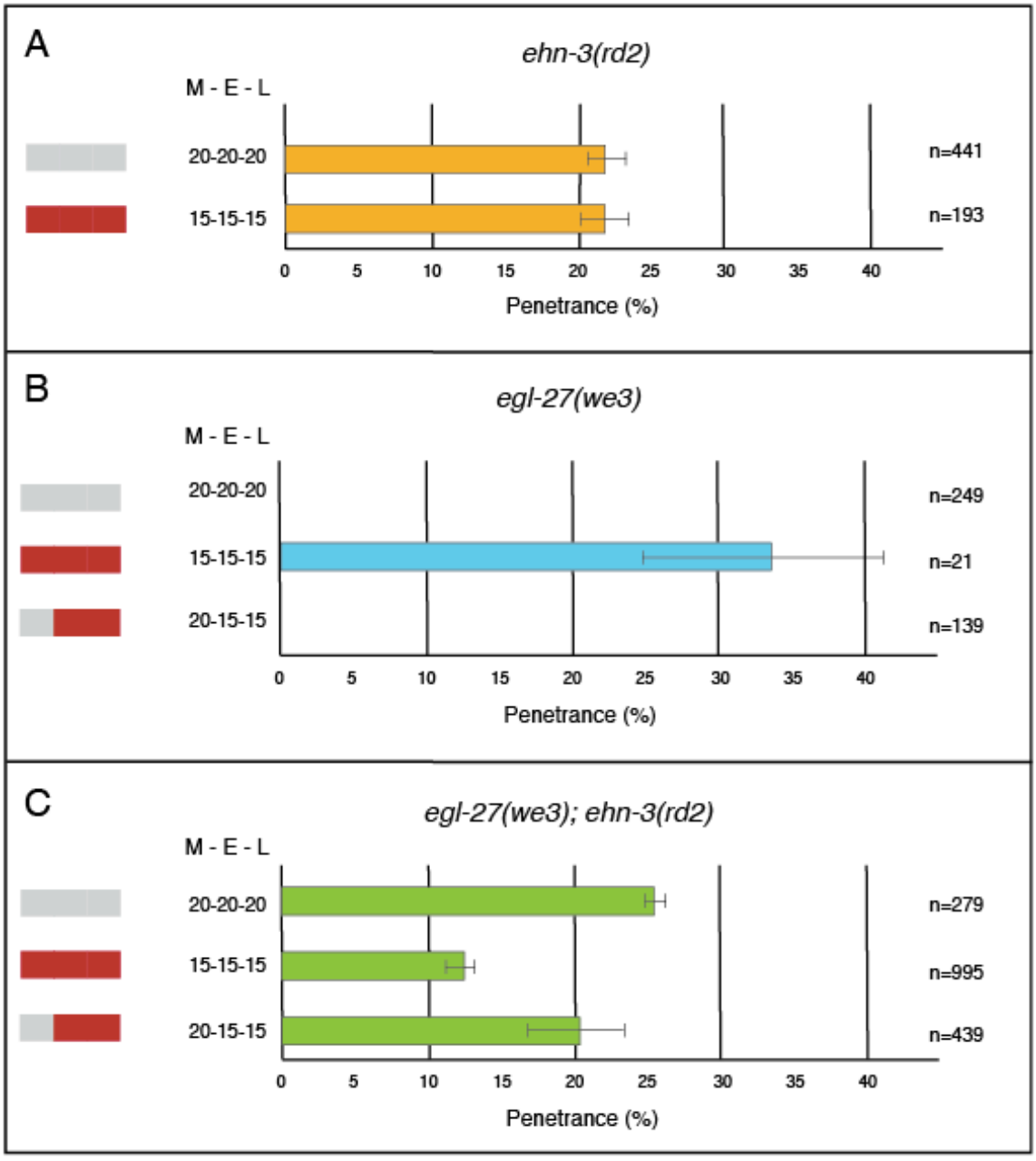

A
ehn-3(rd2)
M - E - L
20-20-20
15-15-15
n=441
n=193
0 5 10 15 20 25 30 35 40
Penetrance (%)
B
egl-27(we3)
M - E - L
20-20-20
15-15-15
20-15-15
n=249
n=21
n=139
0 5 10 15 20 25 30 35 40
Penetrance (%)
C
egl-27(we3); ehn-3(rd2)
M - E - L
20-20-20
15-15-15
20-15-15
n=279
n=995
n=439
0 5 10 15 20 25 30 35 40
Penetrance (%)

Figure 4. *let-418* and *ehn-3* have antagonistic functions in somatic gonad development. (A) *ehn-3(rd2),* (B) *let-418(n3536),* and (C) *ehn-3(rd2); let-418(n3536)* were raised at 20° or 22.5°C during maternal (M), embryonic (E), or larval (L) development. Boxes indicate the temperature shift regimen (red, 22.5°C; grey, 20°C). The percentage of worms with gonadogenesis defects is shown with standard deviation. (A) *ehn-3(rd2)* shows no temperature-sensitivity. (B) *let-418(n3536)* has gonadogenesis defects when reared at 22.5°C continuously or during embryonic and larval development. (C) *ehn-3(rd2); let-418(n3536)* double mutants have a less severe phenotype than *ehn-3(rd2)* mutants and this effect is most significant when reared at 22.5°C during maternal and embryonic development.

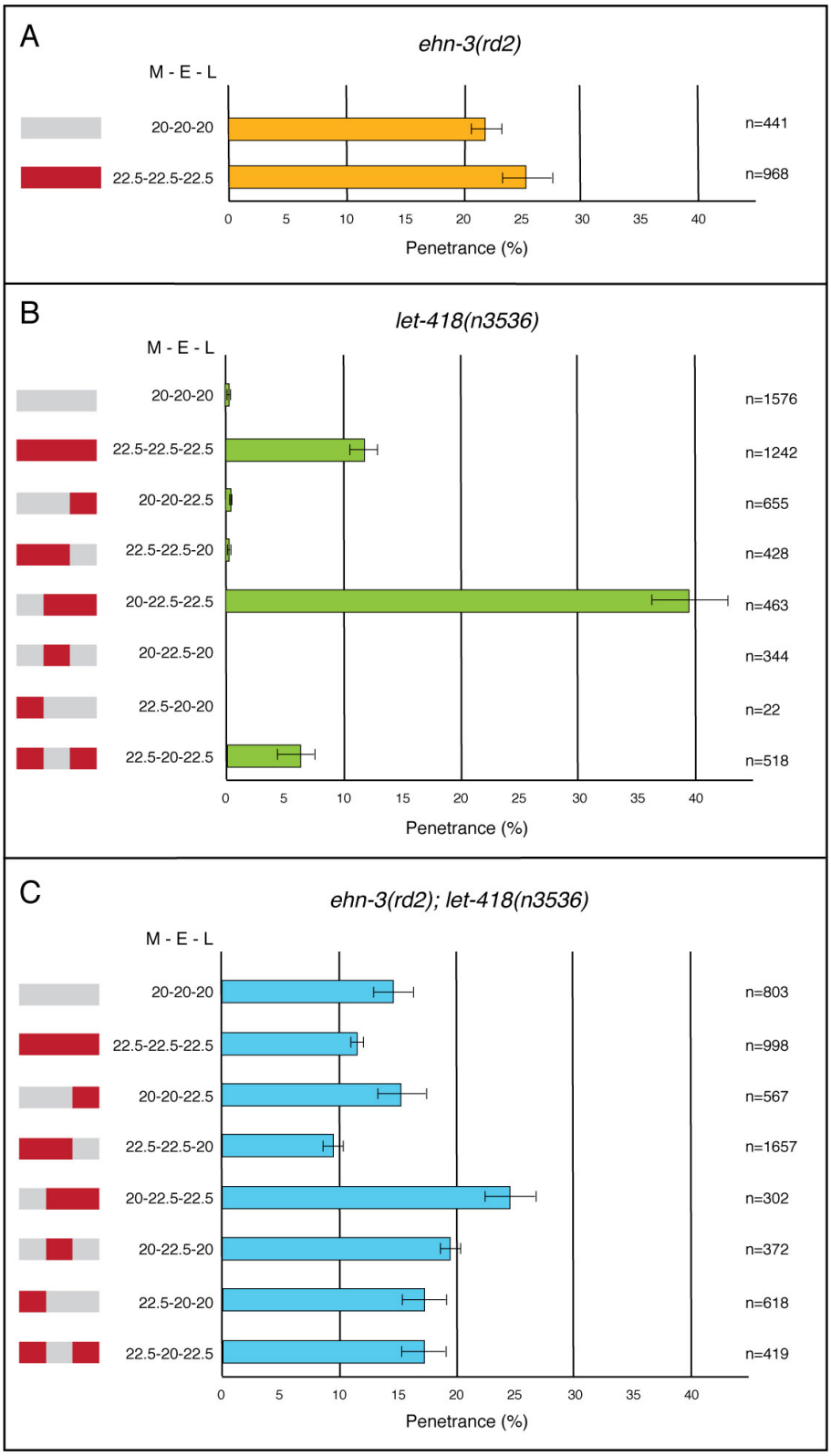
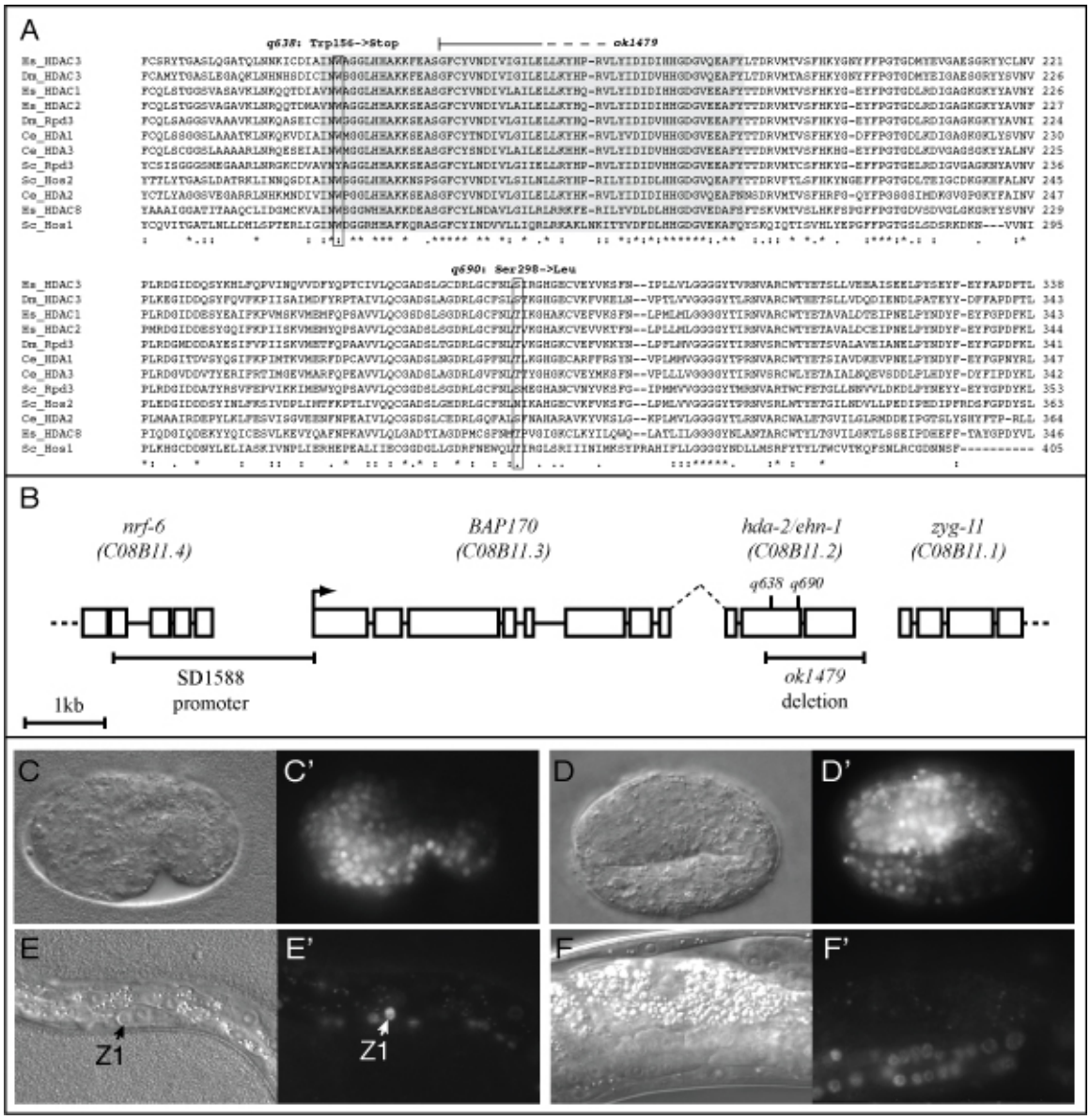

A
ehn-3(rd2)
M - E - L
20-20-20
22.5-22.5-22.5
n=441
n=968
0 5 10 15 20 25 30 35 40
Penetrance (%)
B
let-418(n3536)
M - E - L
20-20-20
22.5-22.5-22.5
20-20-22.5
22.5-22.5-20
20-22.5-22.5
20-22.5-20
22.5-20-20
22.5-20-22.5
n=1576
n=1242
n=655
n=428
n=463
n=344
n=22
n=518
0 5 10 15 20 25 30 35 40
Penetrance (%)
C
ehn-3(rd2); let-418(n3536)
M - E - L
20-20-20
22.5-22.5-22.5
20-20-22.5
22.5-22.5-20
20-22.5-22.5
20-22.5-20
22.5-20-20
22.5-20-22.5
n=803
n=998
n=567
n=1657
n=302
n=372
n=618
n=419
0 5 10 15 20 25 30 35 40
Penetrance (%)

Figure 5. *ehn-1* is allelic to *hda-2* and is expressed in the somatic gonad.
(A) Alignment of Class I and II HDACs from *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Homo sapiens* (Hs), and *Saccharomyces cerevisiae* (Sc). Point mutations in two *ehn-1/hda-2* alleles and the site of the *ok1479* frame shift are indicated. The catalytic site is boxed (light grey). (B) Gene structure of the operon containing *C08B11.3/BAP170* and *C08B11.2/hda-2*. The positions of all alleles used in this study are indicated. The genomic sequence present in the SD1584 reporter (Liu *et al.* 2009) is indicated. Expression of *C08B11.3::HIS::mCherry*; DIC images (C-F); *mCherry* fluorescence (C'-F'). The reporter is expressed broadly during embryogenesis (C, D). In the somatic gonad, the reporter is expressed in the SGPs (E, only Z1 is visible) and later in somatic gonadal tissues including uterus and spermatheca (F).


A
q638: Trp156->Stop
ok1479
Hs_HDAC3
Dm_HDAC3
Hs_HDAC1
Hs_HDAC2
Dm_Rpd3
Ce_HDA1
Ce_HDA3
Sc_Rpd3
Sc_Hos2
Ce_HDA2
Hs_HDAC8
Sc_Hos1
q690: Ser298->Leu
B
nrf-6
(C08B11.4)
BAP170
(C08B11.3)
hda-2/ehn-1
(C08B11.2)
zyg-11
(C08B11.1)
q638 q690
SD1588
promoter
ok1479
deletion
1kb
C
C'
D
D'
E
E'
F
F'
Z1
Z1

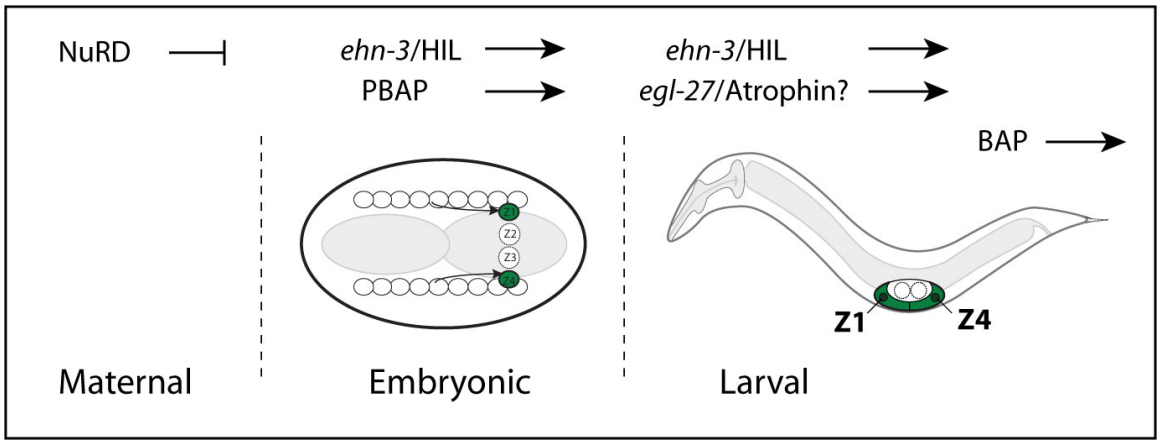


Figure 6. SWI/SNF and NuRD act at multiple times in somatic gonad development. Genetic interactions suggest that maternal NuRD acts in opposition to *ehn-3*, while zygotic NuRD acts in parallel to *ehn-3*. We also find evidence that different SWI/SNF complexes control distinct aspects of somatic gonad development. PBAP acts in parallel to *ehn-3*, while BAP appears to act independently to control different aspects of somatic gonad development.

CHAPTER FOUR

Summary and Future Work

The question of how a fertilized egg can divide and differentiate into multiple cell types leading to the adult tissues and organs continues to challenge biologists. The matter is not purely academic as developmental biology has led to advances in stem cell biology and cancer research, and has fostered the growth of regenerative medicine. Development is a complex process in which each cell must be specified and differentiate in response to internal and/or external inputs. Cells come together over time to form tissues and tissues form organs. Taken as a whole, the entire process seems intractable but, by examining the progression of cell lineages in discrete packets of time, we can begin to understand the process of development as a whole.

Cell lineage progression is defined as the successively limited potential of a cell to generate new fates over the course of development. The temporal activation or repression of different sets of genes often correlates with the subsequent restriction of cell fates. A major focus of research is to understand how different factors orchestrate the commitment of a cell to a particular fate.

Animal models provide well-understood tools to dissect cell lineage progression. For example, in *C. elegans*, the entire pedigree or cell lineage of every adult somatic cell can be traced back to the fertilized egg. These divisions can be observed in real-time facilitated by the transparent body of this useful microscopic nematode. Insights into the mechanisms controlling lineage progression in model organisms can then eventually lead to discoveries that directly impact human health.

Understanding the effect of transcriptional regulation on cell fate and commitment over time is essential to the study of lineage progression. *hunchback* and Ikaros are two

C2H2 zinc finger transcription factors involved in lineage progression. Two decades of research have yielded only a handful of direct gene targets for *hunchback* and Ikaros so there is much to be learned. It is known, however, that both *hunchback* and Ikaros regulate genes by interacting with chromatin factors. For example, the NuRD ATPase Mi-2 physically associates with Ikaros and Hunchback while SWI/SNF also physically interacts with Ikaros. Despite validated gene targets and known biochemical associations with chromatin factors, a molecular model of Ikaros and Hunchback acting in concert with chromatin factors on genes responsible for lineage progression remains elusive.

Summary

This work provides the first example of a *hunchback* and Ikaros-like (HIL) family involved in the lineage progression of a mesodermal tissue in a protostome. We observed overlapping expression patterns of three of the five HIL family members in the somatic reproductive system similar to the overlapping expression patterns observed with Ikaros, Helios, and Aiolos in mammalian lymphoid development. Two HIL family members were subsequently found to be functionally redundant pleiotropic regulators of *C. elegans* somatic reproductive system development, which is derived from the mesoderm. The overlapping expression patterns and functional redundancy of HIL family members in the development of a mesodermally-derived tissue is remininscent of the hypothesized function of Ikaros family members in immune system lineage progression.

Our research also indicates the HIL family is unique to protostomes and distinct from *hunchback* and Ikaros family members. We discovered conserved HIL family members in the genomes of most sequenced protostomes. A phylogenetic analysis of *hunchback* and HIL

family members reveals two distinct clades that likely shared a common ancestor. Similar C2H2 zinc finger domain structures were observed between Ikaros, *hunchback*, and HIL family members. Rescue experiments were executed with heterologous proteins driven by the *ehn-3* promoter. The results point to functional similarities between Ikaros proteins and HIL family proteins and raises the possibility that the HIL family, the *hunchback* family, and the Ikaros family all share a common ancestor.

Finally, we found genetic interactions between HIL family members and chromatin factors (Chapter III). Both SWI/SNF and NuRD chromatin remodeling complex components were found to interact genetically with *ehn-3* to control gonadogenesis. This may mean that HIL family members and Ikaros utilize similar molecular mechanisms to regulate genes involved in lineage progression.

Conclusion and Future Directions

The results of this research reveal unexpected parallels between a new gene family in protostomes and *hunchback* and Ikaros family members. As is often the case, additional questions are raised by the insights from our studies. For example, what are the gene targets of HIL family members in the somatic reproductive system of *C. elegans*? Do HIL family members physically interact with chromatin remodeling complex components? Are there shared roles between HIL family members and *hunchback* both within and outside lineage progression? How can research on *hunchback* and HIL family members yield insights into the role of Ikaros family members in mammalian development?

Gene duplicates can provide both biological robustness and functional redundancy for at least 80 million years (TISCHLER *et al.* 2006). The research presented in Chapter II finds

redundant functions for the HIL gene family members *ehn-3* and *ztf-16* in early and late SGP development (LARGE and MATHIES 2010). Functional and structural similarities between HIL family members and the vertebrate Ikaros family of C2H2 zinc finger transcription factors imply that the two families may share similar molecular mechanisms of gene regulation and the ability of a heterologous Ikaros/EHN-3 protein to rescue *ehn-3;ztf-16* double mutants reinforces this point (LARGE and MATHIES 2010). The conservation of HIL family members in a variety of protostomes, including nematodes, indicates the genetic redundancy of HIL family members may not be unique to *C. elegans*.

Overlapping expression patterns between Ikaros family members also points to potential redundancies. Ikaros family members are on different chromosomes and mutants of Ikaros, Aiolos, and Helios have been generated (CAI *et al.* 2009; GEORGOPOULOS *et al.* 1994; WANG *et al.* 1998), so an analysis of redundant functions is possible. Following the lead of *Drosophila hunchback* neuroblast research, recent evidence also implicates Ikaros in retinal neuroblast progression where four of five family members (Aiolos excluded) are also expressed (ELLIOTT *et al.* 2008). Helios and Aiolos are also strongly expressed in the lymphoid immune system while Eos and Pegasus are broadly expressed. The first reported attempt at making double mutants with Ikaros family members proved to be difficult due to a high mortality rate (CAI *et al.* 2009) pointing to unique and overlapping functions outside immune system development. Our work revealed redundancies between *ehn-3* and *ztf-16* in somatic reproductive system development, which could not have been identified by looking at HIL family members individually. A similar analysis of different mutant combinations of

Ikaros family members may reveal additional redundancies between Ikaros family members in uncharacterized mammalian developmental processes.

Downstream targets of Ikaros and Hunchback have been found but there are no known targets of any HIL family members. As discussed in Chapter I, depletion or overexpression of Ikaros in neuroblasts or *hunchback* in *Drosophila* neuroblasts or *C. elegans* seam cells during crucial phases of development affects the temporal competence of cell lineages (ABRAHANTE *et al.* 2003; ELLIOTT *et al.* 2008; LIN *et al.* 2003; PEARSON and DOE 2003). The strongest effect on somatic reproductive system lineage progression for HIL family members occurs in the transition from an SGP to a DTC. A lineage analysis of *ehn-3* and e*hn-3; ztf-16* double mutants will need to be carried out to determine the precise timing and effect of these mutations on SGP to DTC lineage progression. Recent work has found downstream targets of *C. elegans hunchback* in seam cells (NIWA *et al.* 2009). *C. elegans hunchback* was overexpressed with a heatshock promoter while a subsequent microarray analysis was able to successfully identify both activation and repression of gene targets. A similar approach could be used to identify HIL family targets by overexpressing EHN-3 or ZTF-16 during the critical period leading from SGP to DTC. Heat shock experiments have been used to examine overexpression of EHN-3 in embryos (J. Lonon unpublished results) so at least one of the necessary tools exists to carry out this experiment. Furthermore, given the ability of both Ikaros and *hunchback* overexpression to reset lineage progression clocks (ELLIOTT *et al.* 2008; PEARSON and DOE 2003), it should be interesting to see the effects of HIL family overexpression on different somatic reproductive system lineages. Finding direct

downstream targets of ZTF-16 and/or EHN-3 in the somatic reproductive system may provide a foundation upon which to build a molecular model of HIL family gene regulation.

In Chapter III we found that HIL family members interact genetically with both SWI/SNF and NuRD to control somatic reproductive system development. Further evidence is provided indicating that NuRD and *ehn-3* act at multiple times of development either in parallel or antagonistically. The SWI/SNF complex defined by *pbrm-1* (PBAP) appears to act in parallel to *ehn-3* to control SGP development. It should be interesting to see if the opposing functions of NuRD and SWI/SNF are a common theme with other HIL family members in *C. elegans* and other model organisms. Biochemistry may, in turn, provide additional insight into possible shared physical interactions between HIL family members and chromatin factors similar to that seen with Ikaros and chromatin factors.

Chromatin factors have been found to physically associate with both Ikaros and *Drosophila hunchback*. The strongest support has been provided for physical interactions between mammalian Ikaros and NuRD with SWI/SNF physical interactions coming in a close second. *Drosophila* Hunchback has also been found to physically associate with the NuRD ATPase Mi-2 with additional genetic evidence pointing to a shared mechanism of HOX repression with Polycomb group proteins. *C. elegans* researchers have developed an *in vivo* assay for protein-protein interactions and the subsequent recruitment of chromatin complexes to target promoters (UPDIKE and MANGO 2006). We have not, however, positively identified any direct targets of EHN-3. Furthermore, modENCODE was unable to find direct targets of EHN-3 in the SGPs with a tagged rescuing construct of EHN-3 combined with ChIP-Seq (Laura Mathies, pers. comm.). Additional experiments will be

needed to determine whether EHN-3 or other HIL family members physically associate with chromatin components, which may be greatly facilitated by finding direct gene targets of HIL family members in *C. elegans*. Molecular models of chromatin factors acting in concert with HIL family members may, in turn, provide insight into the mechanisms underlying Ikaros and *hunchback* gene target regulation.

The ability of Ikaros family members to form heterodimers and homodimers is remarkably conserved. Given that Ikaros and Pegasus possibly duplicated in a common protostome-deuterostome ancestor (JOHN *et al.* 2009) it is astonishing that they have retained the same dimerization domain over 550 million years in the same organism. The C-terminal dimerization domain of mammalian Ikaros has even been shown to form heterodimers with hagfish Ikaros indicating this physical interaction has been retained over 450 million years in two independently evolving species (ROTHENBERG and PANT 2004). The physical redundancy appears to be restricted to deuterostomes as mammalian Ikaros cannot form heterodimers with *Drosophila* Hunchback although Hunchback can form homodimers with itself (MCCARTY *et al.* 2003).

We hoped to test the ability of HIL family members to dimerize using a technique known as Bimolecular Fluorescence Complementation (BiFC) in which two constructs are expressed in the same cell but with different halves of yellow fluorescent protein (YFP) (HU *et al.* 2002). In this assay, two complementary nonfluorescent halves of YFP are fused to two separate proteins. The two halves of YFP are brought together if the proteins physically interact and this results in a positive fluorescent signal. The BiFC assay is useful for *in vivo* visualization of physical interactions between proteins and has been successfully used in *C.*

elegans (SHYU *et al.* 2008). A test for the homodimerization of EHN-3 was initiated using two rescuing EHN-3 BiFC constructs expressed in the SGPs. This would have served as the starting point to begin to understand which chromatin factors EHN-3 interacts with in the SGPs and whether EHN-3 forms heterodimers with other HIL family members *in vivo*. We were unable, however, to develop a strong negative control. A deletion of the putative dimerization domain of EHN-3 resulted in YFP expression in SGPs. There are several possible explanations for this result: 1) over expression may have resulted in dimerization even in the absence of the presumptive dimerization domain, 2) other dimerization domains could exist, and 3) N-terminal zinc fingers might bind to adjacent DNA sites and allow the YFP halves to bind even in the absence of a dimerization domain. Future work using single-copy insertions of EHN-3 may provide more insight into the physical interactions of EHN-3 *in vivo*.

Yeast two-hybrid assays could serve as a useful tool to understand HIL family members. The strength of the protein-dimerization domain interaction with Ikaros family members has been used to positively identify previously unknown mammalian Ikaros family members using a yeast two-hybrid screen (PERDOMO *et al.* 2000). Efforts to find physical interactions with EHN-3 have been thwarted by the auto-activation of EHN-3 in yeast two-hybrid assays (L. Mathies, pers. comm.). A dissection of the domain or domains responsible for the autoactivation of EHN-3 has not been carried out but the possibility arises that the dimerization domain of EHN-3 or other HIL family members could be used as a tool to ferret out additional HIL family members. Therefore, the potential to find novel HIL

family members in *C. elegans* and any other protostome cDNA library may be possible if the dimerization potential of this domain is conserved.

Current evidence suggests that the role of *hunchback* in anterior posterior patterning is restricted to arthropods while the role of *hunchback* in HOX gene regulation is found only in insects. I have proposed that *ehn-3* shares a common ancestor with *Drosophila hunchback*, which is expressed maternally and has a well-established role in anterior-posterior patterning (RIVERA-POMAR and JACKLE 1996; ST JOHNSTON and NUSSLEIN-VOLHARD 1992). *C. elegans hbl-1* is not expressed maternally and does not control embryonic patterning (FAY *et al.* 1999). The most conserved HIL family member, *ztf-16* is, however, expressed in the germline (Yuji Kohara, personal communication). The expression of the most conserved HIL family member in the *C. elegans* germline could point to roles for HIL genes in the germline of other species; possibly including embryonic patterning. It will be necessary to use either single-copy transgenes or *in situ* hybridization of known *C. elegans* HIL family members to determine the extent to which other HIL family members are expressed in the germline. Furthermore, we have demonstrated that *ehn-3* interacts genetically in parallel to SWI/SNF and both in parallel and antagonistically to NuRD. *Drosophila hunchback* has also been found to regulate HOX genes by physically interacting with the NuRD ATPase Mi-2 (KEHLE *et al.* 1998) while Ikaros has been found to physically associate with SWI/SNF and NuRD in the immune system (KIM *et al.* 1999; KOIPALLY *et al.* 1999; SRIDHARAN and SMALE 2007). This may indicate a potential ancestral role for *hunchback*-like genes in HOX gene regulation that extends to the protostome-deuterostome ancestor, despite single gene studies indicating otherwise. The ancestral role of

hunchback-like genes, including Ikaros, *hunchback*, and HIL family members, in anterior-posterior patterning and/or HOX gene regulation could therefore be more ancient than previously realized if reexamined in the context of gene duplication and redundancy.

REFERENCES

ABRAHANTE, J. E., A. L. DAUL, M. LI, M. L. VOLK, J. M. TENNESSEN *et al.*, 2003 The *Caenorhabditis elegans hunchback*-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. Dev Cell **4:** 625-637.

CAI, Q., A. DIERICH, M. OULAD-ABDELGHANI, S. CHAN and P. KASTNER, 2009 Helios deficiency has minimal impact on T cell development and function. J Immunol **183:** 2303-2311.

ELLIOTT, J., C. JOLICOEUR, V. RAMAMURTHY and M. CAYOUETTE, 2008 Ikaros confers early temporal competence to mouse retinal progenitor cells. Neuron **60:** 26-39.

FAY, D. S., H. M. STANLEY, M. HAN and W. B. WOOD, 1999 A *Caenorhabditis elegans* homologue of *hunchback* is required for late stages of development but not early embryonic patterning. Dev Biol **205:** 240-253.

GEORGOPOULOS, K., M. BIGBY, J. H. WANG, A. MOLNAR, P. WU *et al.*, 1994 The Ikaros gene is required for the development of all lymphoid lineages. Cell **79:** 143-156.

HU, C. D., Y. CHINENOV and T. K. KERPPOLA, 2002 Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell **9:** 789-798.

JOHN, L. B., S. YOONG and A. C. WARD, 2009 Evolution of the Ikaros gene family: implications for the origins of adaptive immunity. J Immunol **182:** 4792-4799.

KEHLE, J., D. BEUCHLE, S. TREUHEIT, B. CHRISTEN, J. A. KENNISON *et al.*, 1998 dMi-2, a *hunchback*-interacting protein that functions in polycomb repression. Science **282:** 1897-1900.

KIM, J., S. SIF, B. JONES, A. JACKSON, J. KOIPALLY *et al.*, 1999 Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity **10:** 345-355.

KOIPALLY, J., J. KIM, B. JONES, A. JACKSON, N. AVITAHL *et al.*, 1999 Ikaros chromatin remodeling complexes in the control of differentiation of the hemo-lymphoid system. Cold Spring Harb Symp Quant Biol **64:** 79-86.

LARGE, E. E., and L. D. MATHIES, 2010 *hunchback* and Ikaros-like zinc finger genes control reproductive system development in *Caenorhabditis elegans*. Dev Biol **339:** 51-64.

LIN, S. Y., S. M. JOHNSON, M. ABRAHAM, M. C. VELLA, A. PASQUINELLI *et al.*, 2003 The *C elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. Dev Cell **4:** 639-650.

MCCARTY, A. S., G. KLEIGER, D. EISENBERG and S. T. SMALE, 2003 Selective dimerization of a C2H2 zinc finger subfamily. Mol Cell **11:** 459-470.

NIWA, R., K. HADA, K. MOLIYAMA, R. L. OHNIWA, Y. M. TAN *et al.*, 2009 *C. elegans sym-1* is a downstream target of the *hunchback*-like-1 developmental timing transcription factor. Cell Cycle **8:** 4147-4154.

PEARSON, B. J., and C. Q. DOE, 2003 Regulation of neuroblast competence in *Drosophila*. Nature **425:** 624-628.

PERDOMO, J., M. HOLMES, B. CHONG and M. CROSSLEY, 2000 Eos and Pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. J. Biol. Chem. **275:** 38347-38354.

RIVERA-POMAR, R., and H. JACKLE, 1996 From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. Trends Genet **12:** 478-483.

ROTHENBERG, E. V., and R. PANT, 2004 Origins of lymphocyte developmental programs: transcription factor evidence. Semin Immunol **16:** 227-238.

SHYU, Y. J., S. M. HIATT, H. M. DUREN, R. E. ELLIS, T. K. KERPPOLA *et al.*, 2008 Visualization of protein interactions in living *Caenorhabditis elegans* using bimolecular fluorescence complementation analysis. Nat Protoc **3:** 588-596.

SRIDHARAN, R., and S. T. SMALE, 2007 Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. J Biol Chem **282:** 30227-30238.

ST JOHNSTON, D., and C. NUSSLEIN-VOLHARD, 1992 The origin of pattern and polarity in the *Drosophila* embryo. Cell **68:** 201-219.

TISCHLER, J., B. LEHNER, N. CHEN and A. G. FRASER, 2006 Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol **7:** R69.

UPDIKE, D. L., and S. E. MANGO, 2006 Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. PLoS Genet **2:** e161.

WANG, J. H., N. AVITAHL, A. CARIAPPA, C. FRIEDRICH, T. IKEDA *et al.*, 1998 Aiolos regulates B cell activation and maturation to effector state. Immunity **9:** 543-553.