University of Alberta

Forkhead Evolution and the FOXC1 Inhibitory Domain

by

Christina Dawn Fetterman

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Medical Sciences - Medical Genetics

Spring 2011
Edmonton, Alberta
Canada

Examining Committee

Dr. Michael Walter, Department of Medical Genetics

Dr. Warren Gallin, Department of Biological Sciences

Dr. Yves Sauvé, Department of Physiology

Dr. Sarah Hughes, Department of Medical Genetics

Dr. Abbot Clark, Department of Cell Biology and Anatomy, University of North Texas

Abstract

Forkhead (Fox) proteins are transcription factors that function in many processes including development, metabolism and cell cycle regulation. This gene family is divided into subfamilies that appear to originate from a common ancestor. I have identified the evolutionary selection pressures acting on individual amino acid positions in the FoxA, FoxC, FoxD, FoxI, FoxO and FoxP subfamilies. The patterns of selection observed allowed for the prediction of residue function and identification of residues that differentiate orthologs and paralogs. The subfamily structure and negative selection found within the subfamilies indicates that after gene duplication, differentiation of subfamilies through amino acid changes and subsequent negative selection on these changes has occurred. Meanwhile, the observed neutral changes and positive selection allow for further protein differentiation. Within the FoxC subfamily, positive selection was identified at one amino acid site in the inhibitory domain. Mutation of this site in FOXC1 alters transactivation activity and the effects of mutants on transactivation activity are different on different reporters. The mutant effects were consistent with those of known disease causing mutations, supporting the predicted positive selection. The inhibitory domain is known to function in reducing FOXC1 transactivation activity and influences protein stability. Here I additionally show that loss of the inhibitory domain and mutation of the positively selected site can reduce FOXC1 DNA binding. Co-transfection of FOXC1 and TLE4, a repressor protein that can potentially bind to the inhibitory domain, was shown to increase FOXC1 transactivation activity. The effects of a novel disease causing FOXC1 inhibitory

domain mutation on FOXC1 function were also assessed. The mutation reduced FOXC1 transactivation activity and increased protein half-life both of which may lead to disease. Regulation of FOXC1 activity is critical for normal function and this work has furthered our knowledge of how the inhibitory domain influences FOXC1 activity. I have provided biological evidence for the theory that positive selection acts at the amino acid level to optimize protein function. I have also shown that both changes in transcription factor proteins and the cis-regulatory region of target genes have the potential to contribute to evolutionary adaptation.

Acknowledgements

My thanks go out to:

Dr. Mike Walter for his support and guidance throughout the years; for letting me work on a project outside of the lab's main interests.

Dr. Bruce Rannala for initially taking me on as a graduate student.

Past and present members of the Walter and Rannala labs for teaching me techniques, providing opinions and having fun along the way.

My supervisory and examining committee members for their thoughts and time.

My friends and family for their support and patience.

This work was supported by the Alberta Heritage Foundation for Medical Research, the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada.

Table of Contents

List of Tables

List of Figures

List of Symbols

ω	nonsynonymous substitution rate divided by the synonymous substitution rate
$p_{\#}$	proportion of category #
R^2	correlation coefficient
$T_{1/2}$	half-life

List of Species Abbreviations in Sequence Identifiers

amex	*Ambystoma mexicanum*
atep	*Achaearanea tepidariorum*
bflo	*Branchiostoma floridae*
bmor	*Bombyx mori*
clal	*Colisa lalia*
drer	*Danio rerio*
ggal	*Gallus gallus*
ggor	*Gorilla gorilla*
hsap	*Homo sapiens*
mmul	*Macaca mulatta*
mmus	*Mus musculus*
olat	*Oryzias latipes*
omos	*Oreochromis mossambicus*
ppyg	*Pongo pygmaeus*

ptro	*Pan troglodytes*
pvul	*Patella vulgata*
rnor	*Rattus norvegicus*
sscr	*Sus scrofa*
stri	*Spermophilus tridecemlineatus*
tcas	*Tribolium castaneum*
tgut	*Taeniopygia guttata*
tnig	*Tetraodon nigroviridis*
xlae	*Xenopus laevis*
xmac	*Xiphophorus maculatus*
xtro	*Xenopus tropicalis*

List of Gene/Protein Names

AE4	anion exchanger 4
ALR	augmenter of liver regeneration
Antp	antennapedia
BRG1	brahma-related gene 1
CEBPA	CCAAT/enhancer-binding protein alpha
Coch	coagulation factor C homolog
CXCR4	chemokine CXC motif receptor 4
CYP	cytochrome P450
Dll4	distal less 4

Dlx	distal less like
ERalpha	estrogen receptor alpha
FGF19	fibroblast growth factor 19
FOX	forkhead box
GABPA	GA binding protein transcription factor alpha
Hey2	hairy/enhancer of split 2
HMGA1	high mobility group AT-hook1
HMG-IY	high mobility group protein HMG-I/HMG-Y
HNF	hepatocyte nuclear factor
HOX	homeobox
Itgbβ3	integrin beta 3
Jag1	jagged 1
Mef2c	myocyte enhancer factor 2c
MITF	microphthalmia-associated transcription factor
MLL	myeloid lymphoid leukemia
MUC4	mucin 4
myf5	myogenic factor 5
NK2	natural killer cell associated antigen 2
Nkx2-5	NK2 transcription factor related, locus 5
NRF-2	nuclear respiratory factor 2 subunit alpha
otd	orthodenticle
Otx	orthodenticle homolog
PAI-1	plasminogen activator inhibitor 1

PAX	paired box
Pdx1	pancreatic and duodenal homeobox 1
PI3K	phosphoinositide-3 kinase
PKB or Akt	protein kinase B
REL	v-rel reticuloendotheliosis viral oncogene homolog
RREB1	ras responsive element binding protein 1
SGK	serum and glucocorticoid inducible kinase
SPB	lung-specific surfactant protein B
TEAD1	TEA domain family member 1
TEF-1	transcriptional enhancer factor
TFIIB	transcription factor II B
TFIID	transcription factor II D
tin	tinman
TLE	transducin like enhancer of split
TRIM5α	tripartite motif containing protein alpha
Ubx	ultrabithorax

List of Abbreviations

6x BS	six times binding site
ARS	Axenfeld-Rieger syndrome
BEB	Bayes empirical Bayes
BGC	biased gene conversion

BLOSUM	block substitution matrix
BSA	bovine serum albumin
CDART	Conserved Domain Architecture Retrieval Tool
CDD	Conserved Domain Database
ChIP	chromatin immunoprecipitation
CHX	cyclohexamide
DAPI	4’,6-diamidino-2-phenylindole
DBD	DNA binding domain
DTT	dithiothreitol
EH1	engrailed homology 1
EMSA	electrophoretic mobility shift assay
HRP	horse radish peroxidase
HTM	human trabecular meshwork
ID	inhibitory domain
LCA	least common ancestor
LRT	likelihood ratio test
MODY	maturity onset diabetes of the young
NCBI	National Center for Biotechnology Information
NEB	naïve empirical Bayes
NES	nuclear export signal
NLS	nuclear localization signal
PAM	point accepted mutation
PBS	phosphate buffered saline

RT-PCR	reverse transcriptase polymerase chain reaction
SMART	simple modular architecture research tool
SNP	single nucleotide polymorphism
TK	thymidine kinase

Chapter 1: Introduction to molecular evolution and the forkhead gene family.

Genetic Evolution

Adaptation of behavior, morphology and physiology in response to environmental cues has produced the biological spectrum that is present today. Underlying these physically observable adaptations are genetic changes that are much more difficult to discern. There are two main processes regulating genetic change: genetic drift and natural selection. Genetic drift is a random change in allele frequency in a population due to chance (Gillespie 1998; Patthy 2008). These changes are caused by stochastic variations in the number of offspring produced by members of a population (Gillespie 1998). Conversely, natural selection is a change in allele frequency due to the effects that the allele has on fitness (the ability to reproduce) (Gillespie 1998; Yang 2006; Patthy 2008). Alleles that reduce fitness are removed by negative selection whereas alleles that increase fitness are maintained by positive selection. When alleles do not decrease or increase fitness they are considered neutral.. Genetic drift is only observable within a population or species and concerns existing alleles. Natural selection can be observed within a species as well as amongst differing species. Natural selection is a driving force in adaptation because it acts continuously on existing and newly developed alleles.

At the molecular level natural selection can be described in terms of ω, the nonsynonymous (nucleotide changes that result in amino acid changes) substitution rate divided by the synonymous (nucleotide changes that do not result in amino acid changes) substitution rate (Yang 2006). These rates are determined through a comparison of DNA composition at codons among aligned sequences

of interest (Nei and Gojobori 1986). Each nucleotide in the alignment is identified as a nonsynonymous site, a synonymous site or a combination of the two types, which is based on what happens to the amino acid composition if any of the three possible substitutions occur. The site is synonymous if there is no potential for amino acid change or nonsynonymous if there is potential for amino acid change. If both nonsynonymous and synonymous substitutions can occur the site is a combination of the two types. Practically this can be quantified by assigning a value of 1/3 for each of the three potential nucleotide substitutions to either the nonsynonymous or synonymous category depending on what effect the substitution has on amino acid composition. The nonsynonymous and synonymous changes among the sequences are also identified. These values are then used to calculate the substitution rates where the nonsynonymous substitution rate is the number of nonsynonymous substitutions per nonsynonymous sites and the synonymous substitution rate is the number of synonymous substitutions per synonymous site. The value of ω can be calculated over the entire length of a sequence or on individual codons. An $\omega < 1$ indicates negative selection is occurring while $\omega > 1$ suggests positive selection and $\omega = 1$ for neutral changes. Negative or positive selection at an amino acid site implies that the residue is functionally important. Neutral changes at amino acid sites imply that the exact composition of amino acids at these sites is unimportant and that they are not directly involved in protein function.

The calculation of substitution rates described above is the most basic type. More biologically realistic adaptations that consider the

transition/transversion rate ratio, the frequency of codon usage and time separating the sequences are generally used to calculate substitution rates (Li 1993; Pamilo and Bianchi 1993; Goldman and Yang 1994; Yang and Nielsen 2000). Transitions (purine to purine or pyrimidine to pyrimidine nucleotide changes) are known to be more common than transversions (purine to pyridine or pyrimidine to purine changes) (Patthy 2008). This is due to chemical changes of nucleotides that allow for non-standard base pairing, which most often result in transitions, during DNA replication. Codon usage bias is the preferential use of some codons over others even though they code for the same amino acid. tRNAs are not equally abundant in some cells, therefore for proteins that require high levels of production, codons that match the more abundant tRNAs are preferentially used over lower abundant tRNAs (Patthy 2008). The establishment of a novel allele is most often a slow process. Therefore an increase in the time to the most recent common ancestor of a set of sequences allows for more nucleotide changes to occur (Yang 2006). All of these factors affect nucleotide substitution rates.

Natural selection acting on gene duplications is thought to be the major mechanism driving an evolutionary increase in organism complexity (Holland 1999; Taylor and Raes 2004). Duplication events may involve any amount of DNA, from a fraction of a gene to the whole genome. Nondisjunction of chromosomes during meiosis, slipped-strand mispairing during DNA replication, unequal crossing over of chromosomes during recombination and transposition or retrotransposition (RNA is first transcribed into cDNA) of DNA from one locus to

another all contribute to genetic duplication (Holland 1999; Patthy 2008). Transfer of a gene between species, lateral gene transfer, can also introduce a homologous gene family member (Patthy 2008). The fate of duplicates is determined by natural selection (Lynch and Conery 2000; Patthy 2008; Innan and Kondrashov 2010). When duplication of a gene has no effect on fitness or is deleterious, an accumulation of mutations can inactivate one of the copies rendering it a pseudogene (i.e. a gene that will not produce protein). Alternatively, during this period of mutation accumulation the gene may develop advantageous properties or novel functions (neofunctionalization) that prevent it from becoming a pseudogene. Duplicate genes may also subfunctionalize resulting in the requirement of two genes to perform the same function as their single ancestor. If an additional copy of a gene is advantageous, it will be maintained.

This duplication and divergence process results in the formation of gene families (Demuth and Hahn 2009). Gene families are defined by the presence of a particular sequence motif(s). Defining motifs are generally highly conserved and required for normal protein function. These motifs can diverge but still maintain the same function in family members. Outside of the gene family motif, the sequences of family members can diverge to impart novel functions. The number of members in a particular gene family can vary among species due to lineage specific duplications or losses. Orthologous genes are genes in different species that share a common ancestor, while parologous genes are genes within a species that share a common ancestor.

Changes in transcriptional regulation are thought to be a major contributor to phenotypic evolution (King and Wilson 1975; Wray 2003; Wray et al. 2003; Lynch and Wagner 2008; Wagner and Lynch 2008). Transcription factor proteins can change to alter transcription regulation or the cis-regulatory region of targets (i.e. the regulatory elements located on the antisense strand, e.g. transcription factor binding sites) can change resulting in differences in gene expression among individuals and species. The relative contribution of changes in transcription factors vs. changes in the cis-regulatory regions to phenotypic differences is currently under argument (Wray 2007; Lynch and Wagner 2008; Wagner and Lynch 2008). There are studies that show that sequences are different among species however there is a lack of studies showing that these differences can cause an effect. Therefore the arguments regarding the importance of proteins and cis-regulatory elements in evolution are mainly theoretical. Arguments supporting cis-regulatory evolution center around the idea that cis-regulatory elements are modular. These arguments include; 1) cis-regulatory elements can evolve from existing elements and 2) any transcription factor can be used to regulate a gene if the required binding site is present (Wagner and Lynch 2008). The most important argument for evolution through variation in cis-regulation, is that changes should not have highly pleiotropic effects because different cis-elements can be used to control spatial and temporal gene expression (Wagner and Lynch 2008). While these arguments may be true, they can also be applied in support of evolution through variation of transcription factor proteins. Transcription factors are also modular and domain functions can be controlled by the cellular

environment (Locker 2001). In different cell types or in response to signals, alterations in transcription factor expression, co-factor expression and post-translational modifications allow for differential regulation of transcription factor function. Additionally, alternative splicing can create transcription factors with different domain complements. Orthologous transcription factors are rarely (if ever) identical and often only highly conserved in parts of the molecule and not over the entire length. Therefore transcription factors can evolve from existing elements, regulate any gene if the transcription factor and binding site are present and changes in transcription factors are not necessarily highly pleiotropic due to controls in the cellular environment (i.e. differences in cofactor availability among cell types). Biological support of evolution through cis-regulation comes from studies in *Drosophila*. Wing coloration in *Drosophila* was shown to be variable among species due to differences in cis-regulatory elements in regions that are orthologous (Gompel et al. 2005; Prud'homme et al. 2006). In support of evolution through transcription factor proteins, differences in orthologous transcription factor function have been demonstrated. For example, *Drosophila otd* (orthodenticle) cannot completely rescue the nervous system phenotype of mice lacking the orthologs *Otx1* and/or *Otx2* (orthodenticle homolog) (Acampora et al. 1998). Mouse *Otx2* and Drosophila *otd* were also shown to have different target gene profiles (Montalta-He et al. 2002). Overall there is evidence that differences in cis-regulatory regions and differences in transcription factors can lead to evolutionary change. It seems unlikely that these changes are independent of one another.

Transcription Factors

A brief background on the form and function of transcription factors is given here. The data presented is from Locker 2001 (Locker 2001), but can be found in most introductory biochemistry and molecular genetics textbooks. Transcription factors are required for gene expression, the production of mRNA from DNA. In eukaryotes a set of proteins consisting of a number of general transcription factors (e.g. SL1, TFIID or TFIIIB are required) and RNA polymerase make up the general transcription machinery that is used for transcription of all genes. This machinery recognizes a set of nucleotide elements, the promoter of a gene, which directs transcription of that gene. The promoter of a gene must be accessible in order for the transcription machinery to bind. Promoter accessibility is influenced by DNA packaging into chromatin. Chromatin is primarily made up of DNA bound to histone proteins. Chromatin can be tightly packed, which blocks transcription, or open (remodeled), which allows transcription to occur. In addition to the general transcription machinery, transcription factors that function as activators or repressors are required for the control of gene expression. Transcriptional activators are needed for chromatin remodeling, direction of the general transcription machinery to the required promoter and to lift transcriptional repression. Transcriptional repressors do the opposite of activators. Transcription factors that are not part of the general machinery dictate cell and gene specific transcription to mediate cellular differentiation and function. A transcription factor may act as an activator and a repressor on the same or different target genes, in the same or different cell type

(Liu et al. 1999; Locker 2001; Wierstra and Alves 2006; Konopka et al. 2009). Posttranslational modifications or interactions with other proteins are often required for transcription factor function and regulation, and can dictate a switch between activator and repressor. In general, transcription factors that are not part of the general machinery are modular, containing DNA-binding domain(s) and transactivation and/or repression domains. Additional domains that control nuclear import and export, dimerization and transcriptional activity are also often present. These transcription factors are grouped into gene families that are defined through sequence homology of their DNA binding domains. The broadest grouping of transcription factors is by the architecture of their DNA binding domain. Common DNA binding domain architectures include helix-turn-helix, zinc finger and leucine zipper motifs. Different transcription factor gene families exist within these architectures. The similarity in DNA binding motifs among higher and lower organisms allows for the establishment of homology so that the evolution of transcription factors and therefore species differences can be studied.

The Forkhead Gene Family

Forkhead gene family members play diverse roles as transcription activators or repressors during development and in adult tissues. Human diseases known to be caused by mutations and/or chromosomal rearrangements of forkhead genes are: Axenfeld-Rieger syndrome (FOXC1) (Mirzayans et al. 2000),

lymphedema distichiasis (FOXC2) (Fang et al. 2000), vitiligo (FOXD3) (Alkhateeb et al. 2005), thyroid agenesis and cleft palate (FOXE1) (Clifton-Bligh et al. 1998), anterior segment mesenchymal dysgenesis (FOXE3) (Semina et al. 2001), aphakia (FOXE3) (Valleix et al. 2006), alveolar capillary dysplasia (FOXF1) (Stankiewicz et al. 2009), Rett syndrome (FOXG1) (Ariani et al. 2008), Pendred syndrome (FOXI1) (Yang et al. 2007), blepharophimosis ptosis epicanthus inversus syndrome (FOXL2) (Crisponi et al. 2001), t-cell immunodeficiency with alopecia and nail dystrophy (FOXN1) (Frank et al. 1999), developmental verbal dyspraxia (FOXP2) (Lai et al. 2001b), immune dysregulation polyendocrinopathy enteropathy X-lined syndrome (FOXP3) (Bennett et al. 2001) as well as various cancers (FOXA1, FOXC1, FOXC2, FOXO1a, FOXO3a, FOXO4, FOXP1) (Galili et al. 1993; Parry et al. 1994; Hillion et al. 1997; Banham et al. 2001; Lin et al. 2002; Haralambieva et al. 2006; Chanock et al. 2007; Mani et al. 2007).

Forkhead proteins all contain a DNA binding domain, termed forkhead, which defines gene family members. The adopted symbol for forkhead genes is Fox, which comes from Forkhead box. Within the forkhead gene family, subfamilies have been distinguished by their position within a phylogenetic tree that was created using only the forkhead domain sequences (Figure 1-1) (Kaestner et al. 2000). Different subfamilies are identified by letters, with subfamilies A through R noted in humans. Human Fox genes are represented by all capitals (e.g. FOXA), mouse Fox genes have only the F of Fox capitalized (e.g. Foxa), zebrafish Fox genes are written in all lower case (e.g. foxa) and other species, or

when referring to a variety of species, the F of Fox and the subfamily letter are capitalized (e.g. FoxA). For many species, multiple members of a subfamily are known to exist and are further delineated by Arabic numerals then letters.

The forkhead domain was first identified in a comparison of the *Drosophila fork head* and rat *HNF-3A* (subsequently renamed FoxA1) genes (Weigel and Jackle 1990). Mutation of *fork head* resulted in a forked head skeleton in *Drosophila* embryos thus giving rise to the forkhead name (Jurgens et al. 1984). The forkhead domain is a variation of the helix-turn-helix architecture and it folds into at least three α-helices, an antiparallel β-sheet and in most cases at least two loops (Figure 1-2) (Clark et al. 1993; Marsden et al. 1998; van Dongen et al. 2000; Weigelt et al. 2000; Tsai et al. 2006; Brent et al. 2008). Overall the structure resembles a butterfly with a helical thorax and two loop 'wings', leading to the use of 'winged-helix' to describe the forkhead domain. Helix 3 binds the major groove of DNA and plays a role in target selection (Clark et al. 1993; Stroud et al. 2006; Tsai et al. 2006; Brent et al. 2008; Littler et al. 2010). Many forkhead proteins have been shown to have preferred binding site sequences however all of the proteins whose binding sites have been identified have also been shown to bind to non-consensus sequences (Stroud et al. 2006; Tsai et al. 2006; Yan et al. 2006; Brent et al. 2008; Zeng et al. 2008; Koh et al. 2009; Littler et al. 2010). Wings 1 and 2 also contact DNA and may play a role in target affinity (Clark et al. 1993; Shiyanova and Liao 1999; Cirillo and Zaret 2007). There are exceptions to this structure, FOXP2 and FOXM1c do not have typical wings that contact DNA (Stroud et al. 2006; Littler et al. 2010). In these

proteins wing 1 is absent or very short and wing 2 folds back over the helical core. The forkhead domain structure resembles that of the chicken nucleosome linker histone H5 (Clark et al. 1993). The homologous mammalian histone H1 interacts with DNA at the base of nucleosomes to form higher order chromatin structure (Happel and Doenecke 2009). FoxA1, FoxI1 and FoxO1 have all been shown to disrupt this higher order chromatin structure when bound to DNA (Cirillo et al. 2002; Yan et al. 2006; Hatta and Cirillo 2007). The forkhead domain is required for DNA binding and regions outside of the domain that bind histones H3 and H4 facilitate chromatin remodeling (Cirillo et al. 2002; Hatta and Cirillo 2007). FoxA1 preferentially binds sites where H3 lysine 4 (H3K4) is methylated and differences in H3K4 methylation patterns among cell types can account for cell specific FoxA1 activity (Lupien et al. 2008). Forkhead transcriptional activators may function by binding sites formerly occupied by linker histones to disrupt histone-DNA and histone-histone interactions resulting in open chromatin (Hatta and Cirillo 2007). Conversely, the forkhead transcriptional repressor FoxP3 recruits corepressors, such as histone deacetylases and BRG1 (brahma-related gene 1) associate factor complex, to silence chromatin (Li et al. 2006).

The natural selection work presented here focuses on the FoxA, FoxD, FoxI, FoxO, FoxP and FoxC subfamilies while the *in vitro* work focuses on FOXC1. Outside of the forkhead domain, these subfamilies may share similar features (e.g. a transactivation or repression domain, nuclear localization or export signals, dimerization motifs, etc.) however the organization of these features and

lack of sequence conservation makes the subfamilies very different from one another.

The FoxA Subfamily

FoxA genes were the first forkhead genes discovered after *fork head* was found in *Drosophila*. Three FoxA genes were identified in rat liver and subsequently named *hepatocyte nuclear factor 3* (*HNF-3*) *A*, *B* and *C* (Lai et al. 1990). Upon establishment of the forkhead nomenclature committee, these genes were renamed *FoxA1*, *FoxA2* and *FoxA3* respectively. These three FoxA genes are also found in humans. *Foxa1* and *Foxa2* are expressed in many of the same tissues during development and in adulthood. Both are detected in the late primitive streak, axial mesoderm, notochord, neural plate, floor plate of the neural tube and definitive endoderm during development and persist in adult endoderm derived structures (e.g. liver, lung, thyroid, pancreas) and nervous system (reviewed by: Friedman and Kaestner 2006). *Foxa3* expression is dissimilar to that of *Foxa1* and *Foxa2*; it is not expressed in the primitive streak, axial mesoderm or neural structures (reviewed by: Friedman and Kaestner 2006). *Foxa3* is first expressed later than *Foxa1* and *Foxa2* in the endoderm from the hindgut to the midgut/foregut boundary and persists in adult structures derived from this region (e.g. liver, small and large intestine, colon) (reviewed by: Friedman and Kaestner 2006). *Foxa3* is also expressed in the testis, *Foxa1* and *Foxa2* are not, and is required in adults for germ cell maintenance (Behr et al. 2007). Homozygous knockout of Foxa2 is embryonic lethal while Foxa1

knockouts die as neonates and Foxa3 knockouts have a normal lifespan (Weinstein et al. 1994; Kaestner et al. 1999; Shen et al. 2001). Foxa1 and Foxa2 are necessary for the onset of hepatogenesis (Lee et al. 2005a), pancreas development (Gao et al. 2008) and lung development (Wan et al. 2005). Foxa1 is also involved in prostate development and adult kidney function (reviewed by: Friedman and Kaestner 2006). All three Foxa genes are involved in glucose homeostasis and prevention of hypoglycemia (Shih et al. 1999; Shen et al. 2001; Tan et al. 2002; Lee et al. 2005b; Gao et al. 2008). PHA-4, a FoxA ortholog in *C elegans*, increases longevity in response to diet restriction (Panowski et al. 2007). A link between dysfunctional FOXA genes and maturity onset diabetes of the young (MODY) was proposed (Duncan et al. 1998), but subsequent studies have shown that mutation of these genes is not a common cause of MODY (Abderrahmani et al. 2000; Baier et al. 2000; Navas et al. 2000; Yu et al. 2001). However, mutations in *FOXA2* have been associated with late-onset Type II diabetes in one population (Zhu et al. 2000). *FOXA1* is amplified and overexpressed in esophageal and lung adenocarcinomas (Lin et al. 2002) and may act as an oncogene in breast cancer (Williamson et al. 2006).

Important functional regions other than the forkhead domain and an EH1 (engrailed homology 1) motif (Copley 2005) have not been identified in human FOXA proteins but functional residues have been identified in rat FoxA proteins. The rat FoxA proteins share conserved domains I, II and III (Lai et al. 1991). Conserved domain I is the forkhead domain while II and III were shown to be transactivation domains in FoxA2 (Pani et al. 1992). Interestingly, experiments

utilizing FoxA2 demonstrated that conserved domain II contains an EH1 motif that binds TLE1 (transducin like enhancer of split 1) (Wang et al. 2000). This is an interaction normally associated with transcriptional repressors. Conserved domain II was then shown to function as a repression domain in Hela (cervical) cells and an activation domain in HepG2 (liver) cells. This is thought to be due to the higher amount of TLE1 in Hela cells as compared to the amount in HepG2 cells. Therefore the transcriptional activation or repression activity of FoxA2, and potentially other EH1 motif containing Fox proteins, may be cell specific. Additional transactivation domains, conserved domains IV and V, have also been identified in FoxA2 (Pani et al. 1992; Qian and Costa 1995). Numerous genes that are transactivated (directly or potentially directly) by FoxA proteins are known; for example: *ERalpha* (estrogen receptor alpha) (Bernardo et al. 2010), *Hoxb13* (homeobox b 13) (McMullin et al. 2010), *MUC4* (mucin 4) (Jonckheere et al. 2007), *Pdx1* (pancreatic and duodenal homeobox 1) (Gao et al. 2008), ALR (augmenter of liver regeneration) (Dayoub et al. 2010), *SPB* (lung-specific surfactant protein B) (Bohinski et al. 1994) and numerous *CYP* genes (cytochrome P450) (Bort et al. 2004). FoxA1 (Lee et al. 2008) and FoxA2 (Rausa et al. 2003) can also act as indirect transcriptional repressors by binding to other transcription factors and preventing the transactivation of their target genes.

The FoxD Subfamily

There are eight human FOXD proteins: FOXD1, FOXD2, FOXD3, FOXD4, FOXD4L1, FOXD4L2, FOXD4L3 and FOXD4b (also known as

FOXD4L4). *FoxD1*, previously known as brain factor 2, is expressed in the dorsolateral mesoderm, rostral diencephalic neuroepithelium, optic stalk and posterior retina during central nervous system development, stromal cells during kidney development, adult testis, kidney, fetal kidney and kidney cell lines (Hatini et al. 1994; Pierrou et al. 1994; Ernstsson et al. 1996; Hatini et al. 1996; Gomez-Skarmeta et al. 1999). Mice without functioning Foxd1 die within 24 hours of birth (Hatini et al. 1996). FoxD1 functions in forebrain development, posterior retina and optic chiasm establishment, kidney development and modulates T-cell immune response (Hatini et al. 1994; Hatini et al. 1996; Gomez-Skarmeta et al. 1999; Herrera et al. 2004; Levinson et al. 2005; Lin and Peng 2006). *FOXD2* expression has only been observed by northern blot in kidney and kidney cell lines (Ernstsson et al. 1997). During development *FoxD2* is expressed in somites, branchial arches, cranial neural crest cells, and mesoderm including the developing tongue, meninges, nose, whiskers, kidney, limb joints and in the mid- and forebrain (Wu et al. 1998; Pohl and Knochel 2002). *Foxd2*$^{-/-}$ mice are viable however they do have defects in the urinary system (Kume et al. 2000a). Foxd2 functions in kidney development and regulates the T-cell immune response (Kume et al. 2000a; Johansson et al. 2003). *Foxd3* expression was first observed in mouse embryonic stem cells and this expression is down-regulated when the cells are stimulated to differentiate (Sutton et al. 1996). Thus Foxd3 has also been referred to as Genesis. However, *FOXD3* expression was not observed in human embryonic stem cells in one study (Ginis et al. 2004) and expression in human embryonic stem cells has not been demonstrated to date. Northern blots of adult,

fetal and day 18 post conception mouse tissues for *Foxd3* show that *Foxd3* is not expressed in any of these tissues (Sutton et al. 1996). It appears that *FoxD3* is expressed during early developmental stages when cellular identify is determined, and expression is lost after differentiation. For example, *Foxd3* is expressed in neural crest cells and motor neuron progenitors in the developing spinal cord and expression is down regulated once they differentiate (Labosky and Kaestner 1998). Mouse Foxd3 functions in determination of neural crest cell identity and cellular proliferation during neural crest development (Hromas et al. 1999; Dottori et al. 2001; Kos et al. 2001). While zebrafish foxd3 does not appear to function in determination of neural crest identity but does function in determination of cell fate and migration (Lister et al. 2006; Stewart et al. 2006). *Foxd3* is also expressed in epiblast (Hanna et al. 2002) and trophoblast progenitors (Tompers et al. 2005) and functions in maintaining cellular multipotency. Homozygous knockout of Foxd3 is embryonic lethal (Hanna et al. 2002). A mutation in the *FOXD3* promoter is associated with vitiligo (loss of skin pigment) (Alkhateeb et al. 2005). *FOXD4* and *FOXD4b* have broader expression patterns than the other FoxD genes, both are expressed in adult heart, brain, kidney and lung (Freyaldenhoven et al. 2002). FOXD4 is additionally expressed in placenta and skeletal muscle while FOXD4b is also found in pancreas (Freyaldenhoven et al. 2002). A mutation in *FOXD4* has been associated with dilated cardiomyopathy, obsessive-compulsive disorder and suicidality in one family (Minoretti et al. 2007). The FoxD4 genes, including

FOXD4, FOXD4b, FOXD4L1, FOXD4L2, FOXD4L3, are the least studied FoxD proteins and have otherwise not been functionally characterized.

To date, functional domains of the FoxD proteins outside of the forkhead domain have not been extensively characterized. The region N-terminal to the forkhead domain has been described as rich in acidic residues (Ernstsson et al. 1996; Sutton et al. 1996; Ernstsson et al. 1997; Freyaldenhoven et al. 2002). EH1 repression motifs were identified C-terminal to the forkhead domain in FOXD2 and FOXD3 (Copley 2005). *Xenopus* foxd3 also contains an EH1 motif which has been shown to bind Grg4 (groucho related gene 4, a TLE ortholog) resulting in increased repression of transcription (Yaklichkin et al. 2007). The C-terminal region of FOXD proteins is alanine and proline rich, a feature found in transcription repressors, (Cowell 1994; Ernstsson et al. 1996; Sutton et al. 1996; Ernstsson et al. 1997; Freyaldenhoven et al. 2002) and has been shown to be involved in repression in one *Xenopus* FoxD protein (Sullivan et al. 2001). FoxD1 and FoxD2 are known transactivators. FoxD1 transactivates *tyrosine hydroxylase* (Zhang et al. 2010), *RIα subunit of protein kinase A* (Dahle et al. 2002), *PIGF* (placental growth factor) (Zhang et al. 2003) and *foxj1* (Lin and Peng 2006). Foxd2 also activates *RIα subunit of protein kinase A* (Johansson et al. 2003). FoxD3 has transactivation and repression capabilities. FoxD3 transactivates *Foxa1*, *Foxa2*, *osteopontin enhancer* (Guo et al. 2002) and *myf5* (myogenic factor 5) (Lee et al. 2006). *MITF* (microphthalmia-associated transcription factor) is indirectly repressed by FoxD3 in an assay using human FOXD3 on a chicken *MITF* target conducted in a mouse cell line (Thomas and

Erickson 2009) but appears to be directly repressed by foxd3 in zebrafish (Ignatius et al. 2008; Curran et al. 2009). FoxD3 can also repress its own transcription, but it is unknown if the repression is direct or indirect (Pohl and Knochel 2001).

The FoxI Subfamily

There are two FoxI genes in humans: *FOXI1* and *FOXI2*. At the start of this study *FOXI2* was an undiscovered gene, therefore it was not included in the analysis. Very little work has been done on FoxI2 and the human version has not been studied at all. Zebrafish *foxi2* is expressed in the notochord at the three somite stage (3s), in the pharyngeal arch and anterior optic primordium at 18s and continues to be expressed in the developing eye and pharyngeal arch until at least four days post fertilization (Solomon et al. 2003b). Mouse *Foxi2* was examined for expression during craniofacial development and was expressed in the pharyngeal arch and cranial ectoderm excluding the otic placode (Ohyama and Groves 2004). *Foxi2* is also expressed in developing whiskers, teeth, mandibular gland, brain, thymus, kidney, hair follicles and the lining of the inner ear at later stages (Wijchers et al. 2005). *FoxI1* is expressed in non-neural ectoderm prior to somitogenesis, dorsal ectoderm lateral to the neural plate at the beginning of somitogenesis, in the otic placode, otic vesicle, pharyngeal arches and developing kidney (Overdier et al. 1997; Hulander et al. 2003; Solomon et al. 2003a). Mutations in foxi1 result in loss or reduction of the otic placode and vesicle, or a split placode, as well as a smaller malformed jaw (Nissen et al. 2003; Solomon et

al. 2003a). Complete loss of FoxI1 is embryonic lethal in zebrafish but only lethal 50% of the time perinatally in mice (Hulander et al. 1998; Solomon et al. 2003a). Normal ear development is dependant on FoxI1 and *Foxi1*$^{-/-}$ mice are deaf (Hulander et al. 1998; Hulander et al. 2003). Mutations in *FOXI1* cause Pendred syndrome, a disease characterized by childhood or early adulthood onset deafness (Yang et al. 2007). In mice, loss of Foxi1 leads to altered ultrastructure of cells lining the distal nephron of the kidney and distal renal tubular acidosis, however macro and microscopic kidney development appears normal (Blomqvist et al. 2004). Loss of Foxi1 also leads to male sterility due to loss of a proton pump in the epididymal epithelia (Blomqvist et al. 2006). Foxi1 is a transcriptional activator and this activity is dependant on a C-terminal transactivation domain (Overdier et al. 1997). Genes that may be directly or indirectly induced by foxi1 include: *pax2a* (paired box 2a), *pax8*, *dlx3b* (distal less like 3b), *dlx5a*, *dlx4b*, *dlx2a*, *Coch* (coagulation factor C homolog), *Jag1* (jagged 1) (Hulander et al. 2003; Solomon et al. 2003a), all of which function in ear development. *SLC26A4*, also known as pendrin, is directly activated by FOXI1 (Yang et al. 2007). Mutations in *pendrin*, like mutations in *FOXI1*, lead to Pendred syndrome. *AE4* (*anion exchanger 4*) and the B1 and a4 subunits of the vacuolar H^+-ATPase proton pump, which are lost in Foxi1$^{-/-}$ mice, are also directly activated by Foxi1 (Blomqvist et al. 2006; Kurth et al. 2006; Vidarsson et al. 2009).

The FoxO Subfamily

The human FoxO subfamily consists of four genes *FOXO1a*, *FOXO3a*, *FOXO4* and *FOXO6* as well as two known pseudogenes *FOXO1b* and *FOXO3b*. FOXO6, FOXO1b and FOXO3b were not included in analyses conducted here due to predicted or pseudogene status, and therefore will not be discussed. *FoxO1a* and *FoxO3a* are both widely expressed in human and mouse adult tissues including heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary small intestine, colon, and peripheral blood leukocytes (Anderson et al. 1998; Biggs et al. 2001). Conversely, *Foxo4* is only expressed in skeletal muscle (Biggs et al. 2001). Foxo1 is essential for vascular development and *Foxo1*$^{-/-}$ is embryonic lethal (Hosaka et al. 2004). *Foxo3*$^{-/-}$ is not embryonic lethal and the only developmental defect observed in these mice is abnormal ovaries (Hosaka et al. 2004). Female *Foxo3*$^{-/-}$ mice are sterile by 12 weeks of age due to a loss of suppression of follicular activation, resulting in degeneration, while males are unaffected (Castrillon et al. 2003; Hosaka et al. 2004). *Foxo4*$^{-/-}$ is also not embryonic lethal and no developmental defects have been observed in these mice (Hosaka et al. 2004). FoxO proteins regulate cellular functions that include cell cycle arrest, resistance to oxidative stress, differentiation and cell death (reviewed by: Greer and Brunet 2008). Not surprisingly, FoxO genes play a role in longevity and in tumor suppression. The *C elegans* FoxO homolog, daf-16, regulates lifespan and *FOXO3a* variants have been associated with long life in a number of human populations (Kenyon et al. 1993; Willcox et al. 2008; Flachsbart et al. 2009). FOXO1, FOXO3a and FOXO4

have all been shown to suppress or reduce tumorigenesis in human cell lines and simultaneous deletion of all Foxo's in adult tissues results in cancer in mice (Ramaswamy et al. 2002; Hu et al. 2004; Yang et al. 2005; Paik et al. 2007). Mutations in FOXO genes are not known to be disease causing however translocations involving FOXO genes cause cancer. Fusion of *PAX3* to *FOXO1a* results in alveolar rhabdomyosarcoma and *FOXO3a* or *FOXO4* fusions to *MLL* (myeloid lymphoid leukemia) result in leukemia (Galili et al. 1993; Parry et al. 1994; Hillion et al. 1997). FoxO1 also functions in glucose metabolism and FoxO1, FoxO3 and FoxO4 are important for immune cell homeostasis (reviewed by: Gross et al. 2008; Peng 2008).

FoxO proteins are transcriptional activators and repressors and many targets have been identified (reviewed by: Glauser and Schlegel 2007; van der Vos and Coffer 2008). Transactivation domains have been defined in the C-terminus of FOXO1a and FOXO4 (Sublett et al. 1995; So and Cleary 2002). Movement between the nucleus and cytoplasm is a major regulator of FoxO activity. Nuclear localization and nuclear export signals have been delineated in FOXO1a, FOXO3a and FOXO4 (Biggs et al. 1999; Brownawell et al. 2001; Brunet et al. 2002; Zhang et al. 2002b; Zhao et al. 2004). Phosphorylation of FoxOs by PKB (protein kinase B or Akt) or SGK (serum and glucocorticoid inducible kinase) results in nuclear export and/or inactivation of FOXOs thereby promoting PI3K (phosphatidylinositol-3 kinase) mediated cell survival and cell cycle progression (Biggs et al. 1999; Brunet et al. 1999; Kops et al. 1999; Nakae et al. 1999; Rena et al. 1999; Takaishi et al. 1999; Tang et al. 1999; Brunet et al.

2001; Zhang et al. 2002b). Conversely, phosphorylation of FoxOs by oxidative stress induced kinases results in nuclear import and activation of FoxOs (Essers et al. 2004; Lehtinen et al. 2006; Asada et al. 2007). FOXO4 is also monoubiquitinated in response to oxidative stress and subsequently localized to the nucleus (van der Horst et al. 2006).

The FoxP Subfamily

In humans four FOXP genes have been identified: *FOXP1*, *FOXP2*, *FOXP3* and *FOXP4*. *FOXP3* was not included in the analysis here due to its lack of sequence conservation with other FoxP proteins, and therefore will not be discussed. Multiple splice variant isoforms of FOXP1, FOXP2 and FOXP4 exist and the functional differences among the isoforms have not been examined. *FOXP1* is widely expressed in adult and fetal tissues including, but not limited to: brain, heart, lung, kidney, liver, skeletal muscle, pituitary, thalamus, spinal cord, stomach, small and large intestine, colon spleen, pancreas, thymus, blood, bladder, placenta, uterus, ovaries and testes (Banham et al. 2001). *FOXP2* is also widely expressed in adult and fetal tissues including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (Lai et al. 2001a). *Foxp4* is expressed in adult heart, brain, spleen, lung, liver, kidney and testis, but not skeletal muscle (Teufel et al. 2003). During development *Foxp1*, *Foxp2* and *FoxP4* are expressed in the lung, neural, intestinal and cardiovascular tissues (Shu et al. 2001; Lu et al. 2002; Li et al. 2004b; Takahashi et al. 2008). Foxp1 is required for establishment of spinal column motor neuron identity, B cell development, monocyte

differentiation and macrophage function (Shi et al. 2004; Hu et al. 2006; Rousso et al. 2008; Shi et al. 2008). *Foxp1*$^{-/-}$ is embryonic lethal due to heart defects (Hu et al. 2006). *Foxp2*$^{-/-}$ mice die by postnatal day 21, have motor impairment and loss of ultrasonic vocalizations (Shu et al. 2005). *Foxp4*$^{-/-}$ is embryonic lethal and these mice develop two functional hearts (Li et al. 2004b). Altered expression of *FOXP1*, *FOXP2* and *FOXP4* has been observed in various cancers (Banham et al. 2001; Teufel et al. 2003; Campbell et al. 2010). *FOXP2* mutations lead to developmental verbal dyspraxia (Lai et al. 2001a). Recently, *FOXP1* deletions have also been implicated as a causative factor in speech defects (Pariani et al. 2009; Carr et al. 2010).

FoxPs are transcriptional repressors. In order to bind DNA, Foxps must form homo- or heterodimers (Li et al. 2004a). Dimerization is mediated by a region containing a C_2H_2 zinc finger and a leucine zipper (Wang et al. 2003; Li et al. 2004a). All FoxPs contain a glutamine rich region, with the tract lengths varying among the proteins and among species (Banham et al. 2001; Lai et al. 2001a; Shu et al. 2001; Teufel et al. 2003; Shi et al. 2004). The region N-terminal to the forkhead domain, which contains the polyglutamine tract, zinc finger and leucine zipper, is a broadly defined repression domain (Shu et al. 2001; Li et al. 2004a). Foxp1 targets genes involved in B cell development (Wang et al. 2003; Shi et al. 2004; Hu et al. 2006) and Foxp1 and Foxp2 both repress lung epithelial genes (Shu et al. 2001). FOXP2 has also been shown to act as a transcriptional activator and many genes that are repressed or activated by FOXP2 in the brain

have been identified (Spiteri et al. 2007; Vernes et al. 2007). No FOXP4 targets have been identified to date.

Outside of the polyglutamine tract, the *FOXP2* sequence differs at only four amino acids when compared to chimpanzee, gorilla, orangutan, rhesus macaque and mouse *FoxP2*. A test for variation in nucleotide substitution rate along the human lineage in a phylogeny of these six species found evidence for a significantly increased evolutionary rate in humans (Enard et al. 2002; Zhang et al. 2002a). This suggests either relaxation of selective constraints or positive selection in the human lineage. Subsequently, *in vitro* and *in vivo* studies have demonstrated that altering the variable amino acids changes behavior. The function of human FOXP2 was compared to the same construct mutated at the variable sites to form the chimpanzee version in human neuronal cells (Konopka et al. 2009). Comparison of these two proteins demonstrated differences in downstream target gene expression (Konopka et al. 2009). The majority of these differences were replicated in additional cell lines and apparent in gene expression comparisons between human and chimpanzee brains (Konopka et al. 2009). Thus the amino acid changes between human and chimpanzee FoxP2 appears to result in differential downstream target regulation. The mechanism underlying differential target regulation is unknown. These amino acid changes do not disrupt normal homo- or heterodimer formation but can disrupt FoxP2 transactivation/repression activity (Konopka et al. 2009). The variable sites are in a region that may function in transcriptional repression (Li et al. 2004a). Mice with Foxp2 mutated to create a human version have demonstrated altered

ultrasonic vocalizations, decreased exploratory behavior, decreased dopamine concentrations and altered neuron morphology in the striatum when compared to wild type mice (Enard et al. 2009). Therefore, the amino acid changes between human and mouse FoxP2 result in gross differences at the organism level. Taken all together it appears that FOXP2 is under positive selection and not relaxed selective constraint. FOXP2 was the first gene definitively associated with language ability in humans. Therefore work in FOXP2 has garnered much attention and it has been speculated that FOXP2 is key to the development of language (Fisher and Scharff 2009). Recently mutations in FOXP1 were also associated with language impairment (Hamdan et al. 2010). Patients with FOXP1 mutations also have nonsyndromic intellectual disability and autism spectrum disorders where patients with FOXP2 mutations only have neurological defects that involve language. FOXP1 appears to play a more global role than FOXP2 in brain development and potentially contributes to language development.

The FoxC Subfamily

The FoxC subfamily consists of two members in humans, FOXC1 and FOXC2. Mutations in *FOXC1* cause Axenfeld-Rieger syndrome, an ocular disorder that may involve systemic malformations (discussed below). Mutations in *FOXC2* cause lymphedema-distichiasis syndrome, which is characterized by lymphedema in the lower limbs and distichiasis (aberrant eyelashes) (Connell et al. 2008). Patients with lymphedema-distichiasis syndrome may have other defects including ptosis, cleft palate and congenital heart disease (Connell et al.

2008). FOXC2 additionally has a protective effect against Type 2 diabetes (Cederberg et al. 2001). Both FOXC1 and FOXC2 have been associated with breast cancer. Altered *FOXC1* expression (Muggerud et al. 2010; Ray et al. 2010) and mutations in the 5' and 3' extragenic sequence (Chanock et al. 2007) were found in breast tumor samples. In mouse, Foxc2 promotes breast cancer metastasis and in humans high levels of *FOXC2* expression have been found in aggressive and invasive breast cancers (Mani et al. 2007).

Mutations in *FOXC1* cause Axenfeld-Rieger syndrome (ARS), which is characterized by ocular anterior segment defects and may involve other systemic malformations including heart defects, redundant periumbilical skin, dental abnormalities and craniofacial dysmorphism (Tumer and Bach-Holm 2009). ARS is autosomal dominant, highly penetrant and has variable expressivity. The presence of posterior embryotoxon is a strong indicator of ARS useful in distinguishing it from other anterior segment disorders. Posterior embryotoxon is a prominent Schwalbe's line (the interior edge of Descemet's membrane of the cornea) that is displaced anteriorly (Figure 1-3). Iris abnormalities that have been observed are hypoplasia (thinning), corectopia (pupil displacement) and additional holes mimicking multiple pupils (polycoria) (Figure 1-3). Iris strands bridging the iridocorneal angle are also common (Figure 1-3). The iridocorneal angle is the angle formed between the cornea and the iris and marks the boundary between cornea and sclera (Figure 1-4) (Forrester 2007). It contains trabecular meshwork cells and Schlemms canal which allow the aqueous humor fluid from the anterior chamber to flow into the scleral venous system. Disrupted aqueous

humor outflow can increase ocular pressure, which is a risk factor for glaucoma development. The major health consequence of ARS malformations is an increased risk of glaucoma (Strungaru et al. 2007). Other anterior segment disorders can overlap phenotypically with ARS and include iridogoniodysgenesis, Peters anomaly and primary congenital glaucoma (Tumer and Bach-Holm 2009). Iridogoniodysgenesis is malformation of the iridocorneal angle and iris hypoplasia. Peters anomaly is a malformation of the cornea and presents with corneal opacity. Primary congenital glaucoma is early onset and associated with goniodysgenesis and high intraocular pressure. These disorders are distinct form ARS in that individuals do not normally have posterior embryotoxon. Mutations in FOXC1 have been associated with iridogoniodysgenesis (Mears et al. 1998; Lehmann et al. 2000; Mirzayans et al. 2000; Saleem et al. 2003), Peters anomaly (Honkanen et al. 2003; Weisschuh et al. 2008), primary congenital glaucoma (Chakrabarti et al. 2009; Ito et al. 2009) and aniridia (absence of the iris) (Ito et al. 2009). Therefore the spectrum of abnormalities is variable for FOXC1 mutations. Additionally, phenotypic differences have been observed among family members who have the same FOXC1 mutation (Ito et al. 2007; Weisschuh et al. 2008).

Northern blots have shown that *FOXC1* is expressed in numerous adult tissues including, but not limited to, heart, liver, kidney, pancreas, skeletal muscle and brain (Pierrou et al. 1994; Mears et al. 1998). *FOXC2* has been detected in adult adipose tissue by northern blot (Cederberg et al. 2001) and skeletal muscle by RT-PCR (Di Gregorio et al. 2004). Microarray analysis of human *FOXC2* found the highest expression levels in brain, ciliary ganglion, superior cervical

ganglion, skeletal muscle and kidney (Su et al. 2004). FoxC1 and FoxC2 both function in the development of the circulatory, ocular, skeletal and urogenital systems (Winnier et al. 1997; Kume et al. 1998; Hong et al. 1999; Kidson et al. 1999; Winnier et al. 1999; Kume et al. 2000b; Smith et al. 2000; Dagenais et al. 2004; Seo et al. 2006). Additionally, FoxC1 is known to function in brain and gonad development (Mattiske et al. 2006; Zarbalis et al. 2007; Aldinger et al. 2009). Both *Foxc1*$^{-/-}$ and *Foxc*2$^{-/-}$ are embryonic lethal (Winnier et al. 1997; Kume et al. 1998). During development, mouse *Foxc1* and *Foxc2* are expressed throughout the mesoderm, in somites but not the notochord, and in mesenchyme (Sasaki and Hogan 1993; Miura et al. 1997; Hiemisch et al. 1998; Kume et al. 2000b; Kume et al. 2001). *Foxc1* and *Foxc2* have overlapping and distinct expression patterns. For example, *Foxc1* is expressed in the mitral, tricuspid, aortic and pulmonary heart valves while *Foxc2* is expressed only in the aortic and pulmonary valves of newborn mice (Winnier et al. 1999). Similarly, Foxc1 and Foxc2 have overlapping and distinct direct target genes. Both genes can directly activate *Dll4* (distal less 4) and *Hey2* (hairy/enhancer of split 2) promoters (Seo et al. 2006; Hayashi and Kume 2008b) while Foxc2 but not Foxc1 can activate *PAI-1* (plasminogen activator inhibitor 1) (Fujita et al. 2006). Human FOXC1 and FOXC2 can both directly activate *FOXO1a* and *FGF19* (fibroblast growth factor 19) (Tamimi et al. 2006; Berry et al. 2008; Huang 2009). For each shared target gene, chromatin immunoprecipitation demonstrated that FoxC1 and FoxC2 both bind in the same segment of DNA, but it is unknown if they bind the exact same nucleotides. Foxc2 has also been shown to directly activate *Mef2c* (myocyte

enhancer factor 2c), *Itgbβ3* (integrin beta 3) and *CXCR4* (chemokine CXC motif receptor 4) however it is unknown if Foxc1 activates these genes directly (De Val et al. 2008; Hayashi and Kume 2008a; Hayashi et al. 2008). The *in vitro* preferred consensus DNA binding site of FOXC1 is **GTAAA**(**T**/C)**A**(**A**/T/C)(**A**/T/G)(**C**/G/T)(**A**/G/C) as determined by protein binding to random oligonucleotides (Pierrou et al. 1994) (the predominant nucleotides are in bold). For FOXC2 the preferred binding site is (**C**/T/A)(**A**/T)(**A**/G)(**A**/G/T)(**A**/G)(**A**/T)(**A**/C/T)**AA**(**C**/T)**A** as determined through chromatin immunoprecipitation-chip analysis in lymphatic cells (Norrmen et al. 2009). However, both proteins are known to bind segments of DNA that do not contain perfect consensus sites (Hayashi and Kume 2008b; Huang 2009). The forkhead domains of FOXC1 and FOXC2 differ at only two amino acid sites, neither of which are involved in contacting DNA. The similarities in binding sites, shared targets and lack of amino acid differences in the DNA contacting region of the forkhead domain suggest that other factors influence differential target selection by FOXC1 and FOXC2. Deletion studies of human and mouse FoxC1 and FoxC2 have delineated transactivation domains that are responsible for transactivation activity and an inhibitory domain which reduces transactivation activity (Berry et al. 2002; Petrova et al. 2004; Fujita et al. 2006).

Both *FOXC1* and *FOXC2* are found in FOX chromosomal gene clusters (Figure 1-5). *FOXQ1*, *FOXF2* and *FOXC1* are found at 6p25 while *FOXF1, FOXC2* and *FOXL1* are found at 16q24 (Wotton and Shimeld 2006). The orthology of these genes plus others in the two regions has been determined

through phylogenetic analyses in bilaterians and evolutionary models have been proposed (Figure 1-5) (Mazet et al. 2006; Wotton and Shimeld 2006; Wotton et al. 2008). A tree showing the relationship of taxonomic groups is given in Figure 1-6. Basal bilaterians are thought to contain one copy of *FoxQ1*, *FoxF*, *FoxC* and *FoxL1*. Protosomes and non-vertebrate Deuterostomes at most maintain one copy of each of these genes while whole genome duplication along the Vertebrate lineage results in up to two copies of each gene in Vertebrate species. In the majority of lineages gene loss is also predicted. For humans, one copy of *FOXL1* and *FOXQ1* are thought to be lost while *FOXC* and *FOXF* paralogs differentiate into *FOXC1* and *FOXC2* and *FOXF2* and *FOXF1* respectively. Teleosts additionally experience a second round of whole genome duplication followed by gene loss.

Forkhead Gene Family Expansion

The early evolution of forkhead genes family members has been conservatively traced (Figure 1-7) (Larroux et al. 2008). No forkhead homologs have been identified in plants. At the base of the eukaryotes, the Fungi-Metazoa least common ancestor (LCA) likely contained one forkhead gene. From this gene, three forkhead ancestors are thought to develop in the Choanoflagellate-Metazoa LCA. Metazoan specific classes then originate from further expansion of the original forkhead gene and the additional appearance of a second series of forkhead genes that lack introns. Of the gene families of interest here, FoxI, FoxO and FoxP contain introns, while FoxA, FoxC and FoxD are intronless.

Further expansion of both lineages gives rise to the A, B, C, D, F, J, K, L, M, N, O, P, Q subfamilies by the Protostomia-Deuterostomia LCA. FoxD, FoxO and FoxP subfamilies are older and originate in the Metazoa, while FoxA and FoxC subfamilies originate in the Eumetazoa. FoxIs are the most recently developed subfamily of interest here and appear to originate in the Deuterostomia.

Rationale and Hypotheses

The initial goal of this research was to identify the evolutionary selection pressures acting at the codon level in forkhead gene family members. I hypothesized that forkhead genes are experiencing variable selection pressures including: neutral changes, negative and positive selection. To test this hypothesis I used *in silico* methods to estimate ω for each codon in an alignment of Fox sequences. At the start of this analysis, rates of evolutionary change had only been examined in the FoxP2 and FoxL2 subfamilies and various phylogenies of the whole gene family and subfamilies had been created, otherwise the evolution of these genes had not been examined. Analysis of forkhead gene evolution at the molecular level has provided insights into amino acids that are important for gene function. This is important because as discussed above, forkhead genes function in development and adult tissues and have been associated with many diseases. Additionally, changes in transcription factor function are thought to be a contributing factor to speciation. Identification of selection pressures on forkhead genes helps determine changes in these genes that may play a role in speciation.

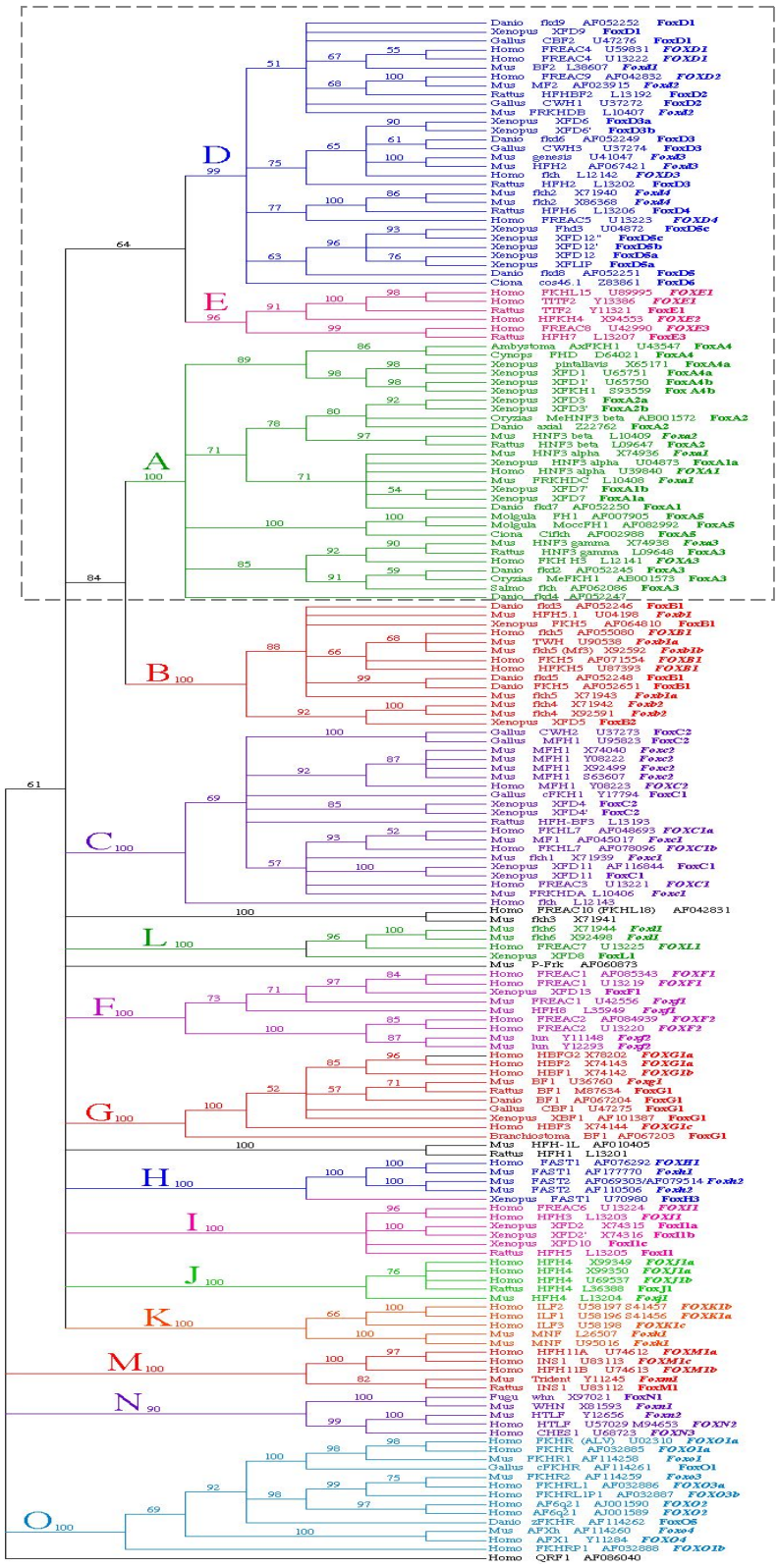
The selection pressure analysis above identified positive selection at one codon in the FoxC subfamily. I hypothesized that altering this amino acid in FOXC1 would alter FOXC1 function. The codon under positive selection was mutated and the effects on FOXC1 function were assessed *in vitro*. Functional domains have been identified in FOXC1 however their mechanisms of action are

still being elucidated. The positive selection observed here is in the inhibitory domain and these analyses have furthered the knowledge of inhibitory domain function. FOXC1 is required for normal development and mutations in *FOXC1* lead to Axenfeld-Rieger Syndrome, therefore elucidating FOXC1 functional mechanisms is important. These analyses are required to confirm the predicted positive selection and provide a basis for further hypotheses regarding FoxC functional evolution.

Finally I examined further aspects of inhibitory domain amino acid function in FOXC1. Inhibitory domains in other transcription factors can inhibit transactivation through impairing DNA binding and/or physically blocking a transactivation domain. I hypothesized that the inhibitory domain of FOXC1 can exert its function through these mechanisms. During the selection analyses, EH1 motifs were identified in the FoxC sequences. In other transcription factors these motifs have been shown to mediate interactions with TLE proteins, which results in transcriptional repression. I hypothesized that the EH1 motif in FOXC1 interacts with TLE4 to mediate repression of FOXC1 activity. These hypotheses were tested using *in vitro* systems. Additionally, a novel polymorphism in the inhibitory domain was identified in two individuals with ocular anterior segment dysgenesis by the Walter lab. I hypothesized that this change would alter FOXC1 function. I created this amino acid change in FOXC1 and compared the function of the resulting protein to wild type FOXC1 using *in vitro* systems. Again, both of these analyses are important for further elucidation of inhibitory domain and therefore FOXC1 function.

Figures

A.



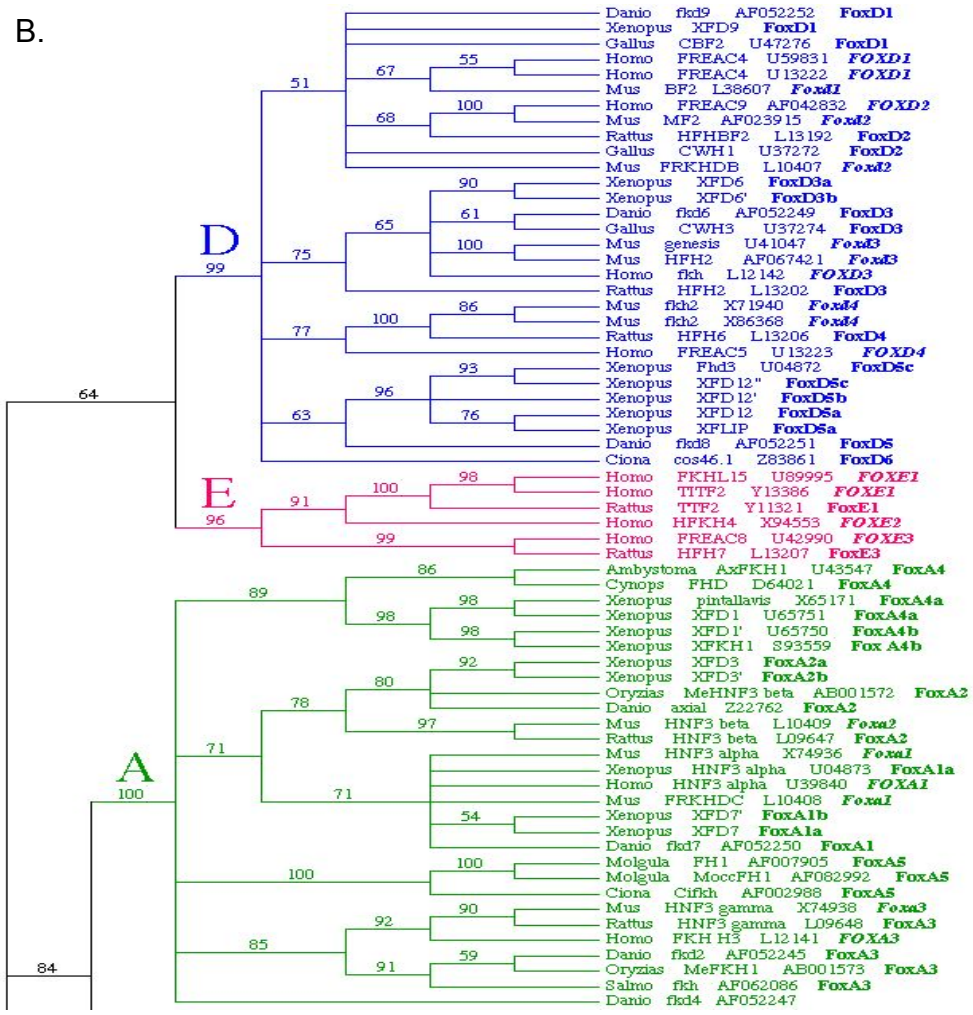
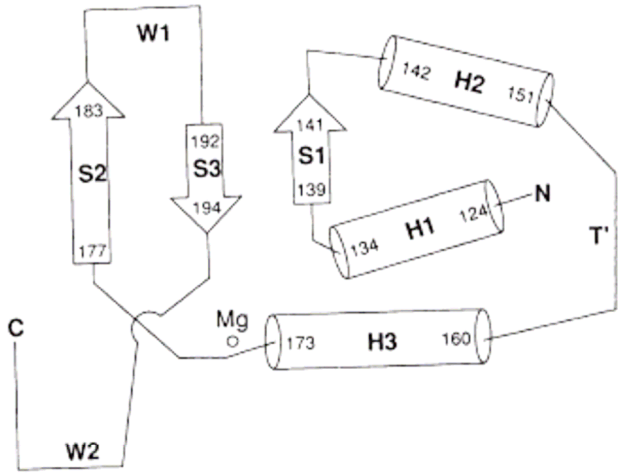


Figure 1-1. Establishment of forkhead nomenclature.

A. The original neighbor-joining tree of chordate forkhead domains that was used to establish family nomenclature (Kaestner et al. 2000). B. Enlargement of the phylogeny section that is boxed in A. Subfamily clades are each assigned a representative letter and color coded here. Within each subfamily, paralogs are assigned representative numbers based on the observed branches and additional letters or numbers if the paralog duplicates further. The tree contains duplicates of sequences, some names have changed since this original version and additional subfamilies (P, Q, R) have been identified.

A.



B.

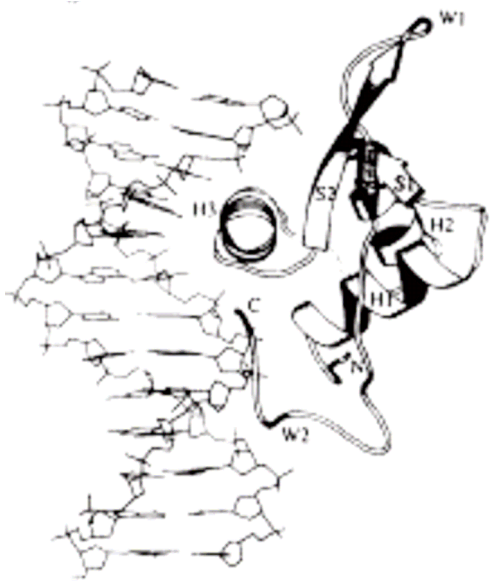


Figure 1-2. Structure of the forkhead domain.

Rat FoxA3 was co-crystallized with a fragment of DNA to determine forkhead domain structure (Clark et al. 1993). A. Schematic showing three α-helices (H1, H2, H3), a three strand β-sheet (S1, S2, S3) and two loop wings (W1, W2). B. The forkhead domain is shown bound to DNA. Helix3 and wing 2 are shown to make DNA contacts. The DNA fragment used here was too short to identify all wing 1 contacts.

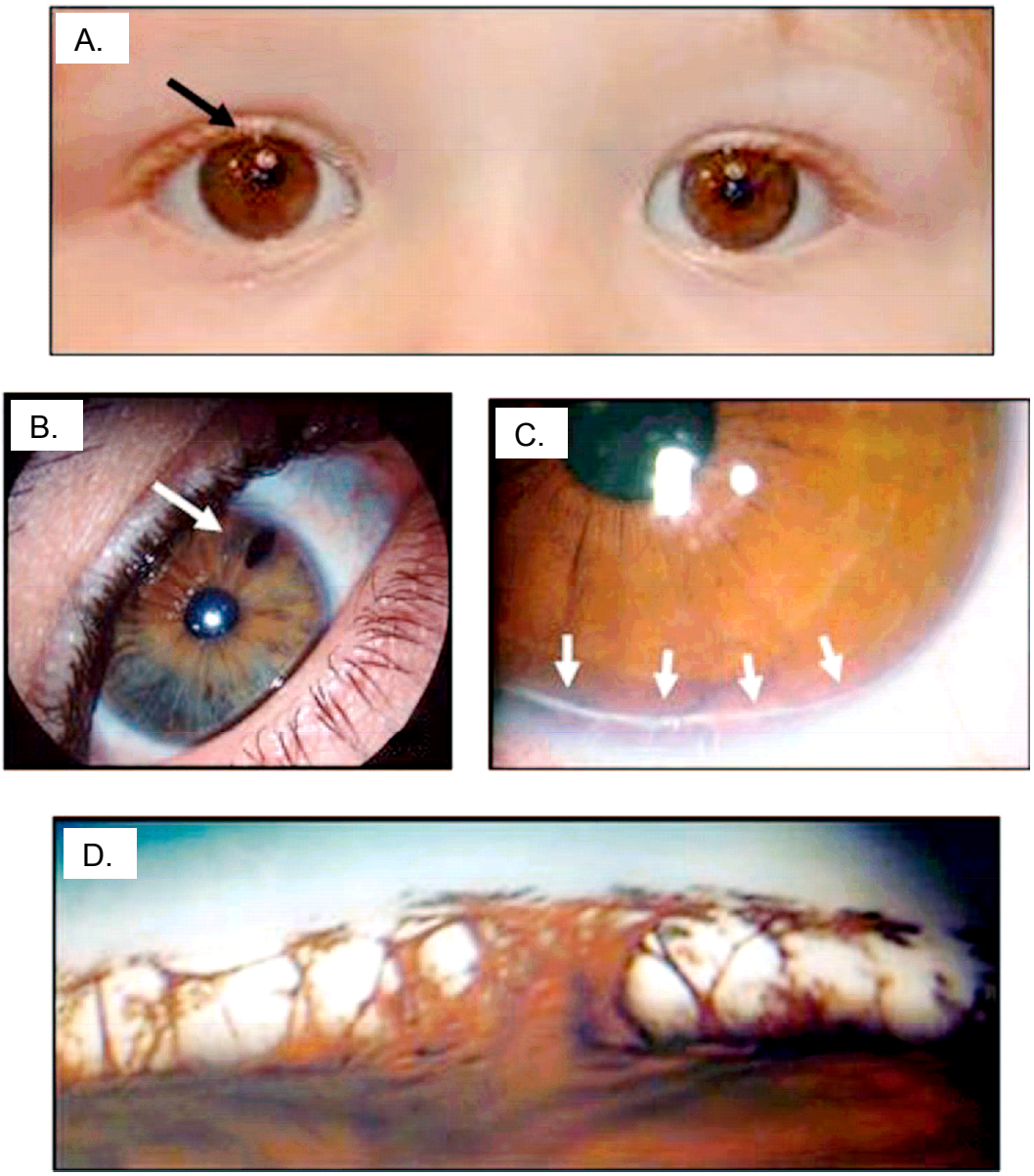


Figure 1-3. Eye phenotypes of ARS (Tumer and Bach-Holm 2009).

A. Corectopia (pupil displacement) is indicated with an arrow. B. Polycoria (extra hole in the iris) is indicated with an arrow. C. Posterior embryotoxon (prominent Schwalbe's line at the edge of the cornea) is indicated with an arrow. D. Iris strands that span the iridocorneal angle.

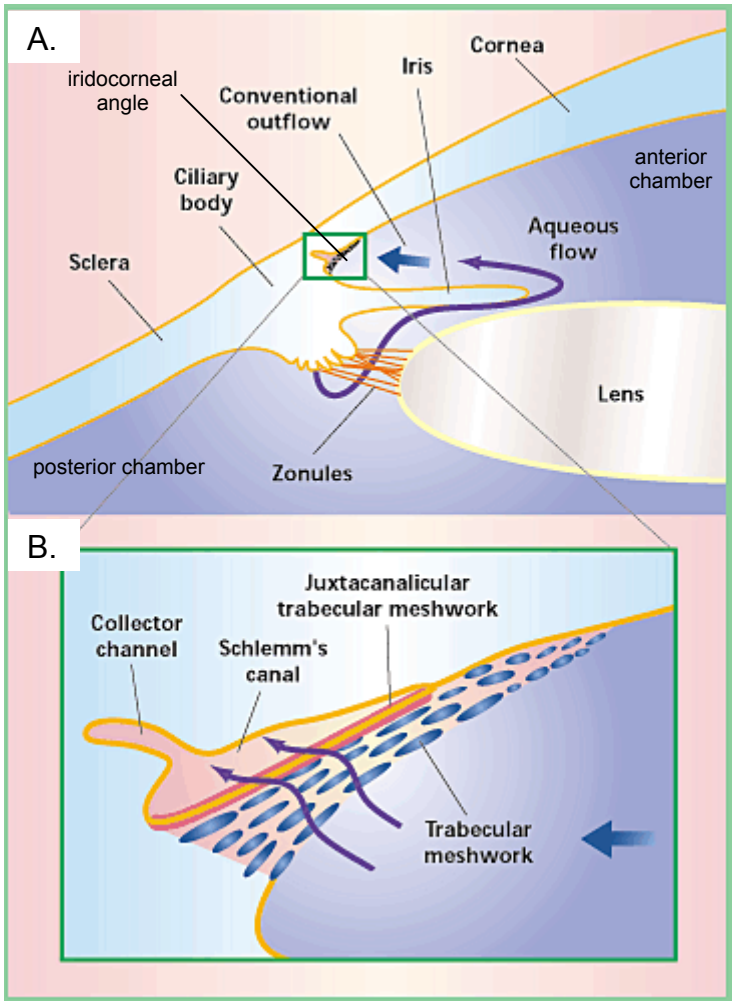
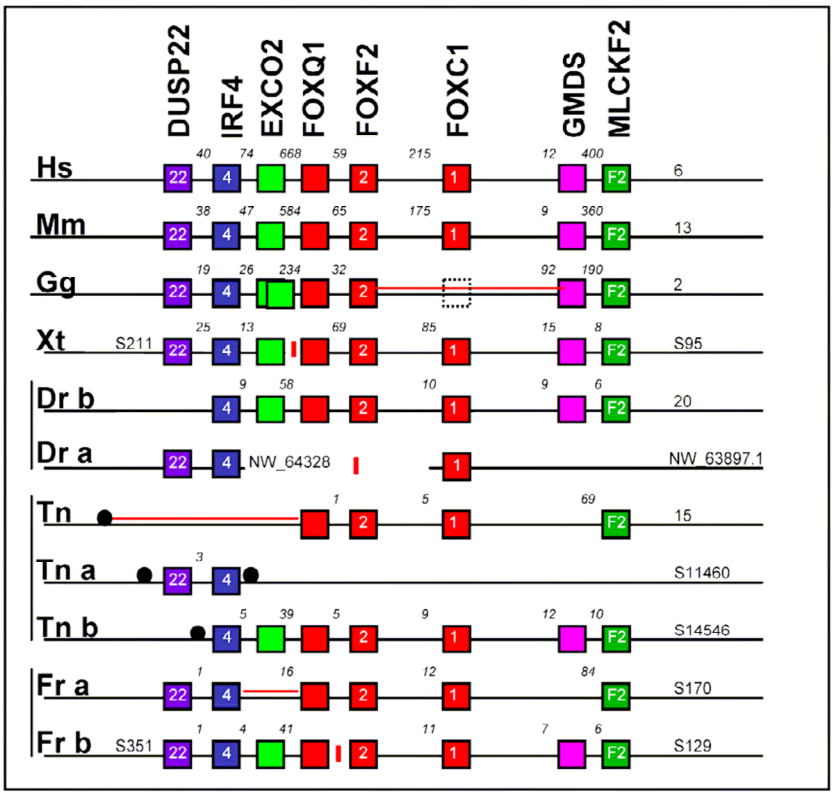
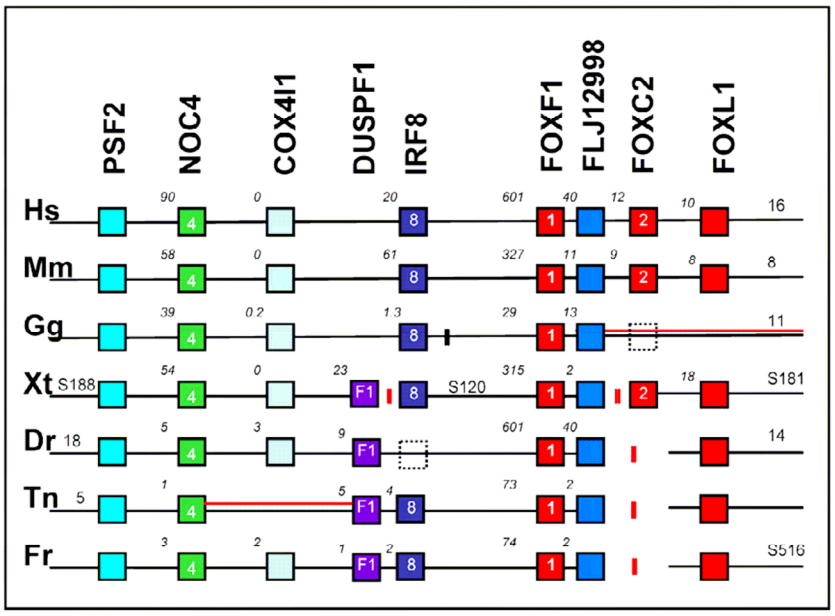
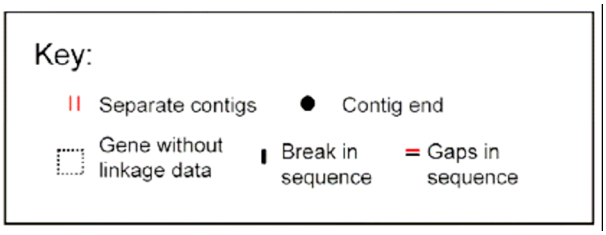


Figure 1-4. Schematic of aqueous humor flow (modified from: Tomarev 2001) A. Aqueous humor is produced by the ciliary body, flows into the anterior chamber and exits the chamber through the iridocorneal angle. B. Expansion of part of the iridocorneal angle. Aqueous humor flows through the trabecular meshwork, into Schlemm's canal and exits through the scleral venous system.

A.


DUSP22
IRF4
EXCO2
FOXQ1
FOXF2
FOXC1
GMDS
MLCKF2
Hs
Mm
Gg
Xt
Dr b
Dr a
Tn
Tn a
Tn b
Fr a
Fr b
NW_64328
NW_63897.1
S211
S95
S11460
S14546
S170
S351
S129


PSF2
NOC4
COX4I1
DUSPF1
IRF8
FOXF1
FLJ12998
FOXC2
FOXL1
Hs
Mm
Gg
Xt
Dr
Tn
Fr
S188
S120
S181
S516


Key:
Separate contigs
Contig end
Gene without linkage data
Break in sequence
Gaps in sequence

B.

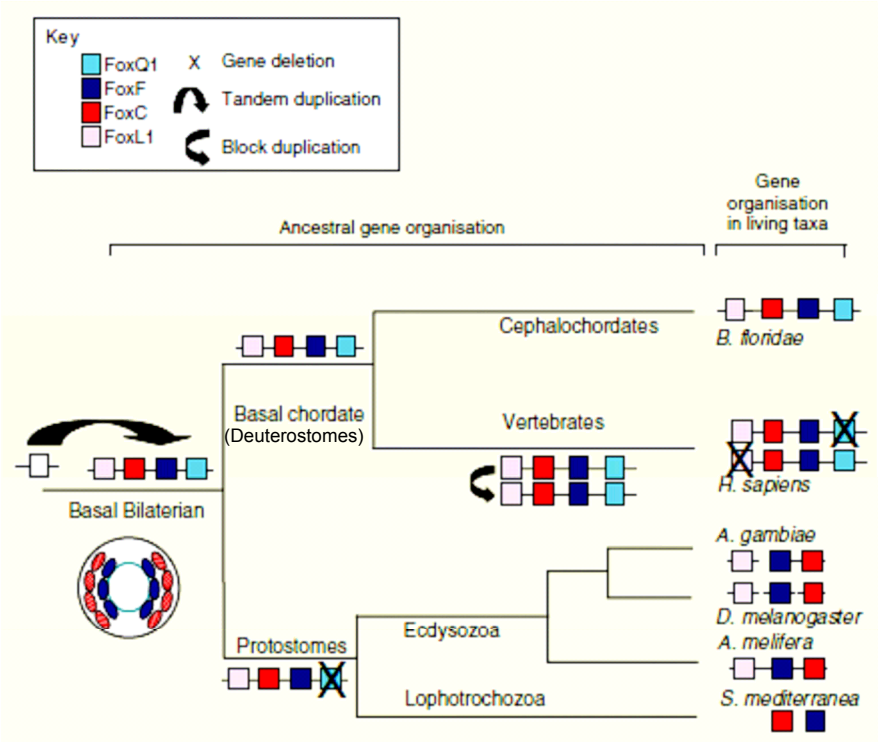


Figure 1-5. The FoxQ, FoxF, FoxC and FoxL gene clusters. Orthologous and parologous relationships were determined through phylogenetic analyses. A. The FoxQ1-FoxF2-FoxC1 and FoxF2-FoxC2-FoxL1 clusters are maintained in many species (modified from: Wotton and Shimeld 2006). Genes are indicated with colored boxes on a horizontal line that represents a chromosome. The chromosome or scaffold location is given at the end of each line. The diagram is not to scale and approximate distances between the genes are given above the chromosome line in kilobases. *Homo sapiens* (Hs) *Mus musculus* (Mm) *Xenopus tropicalis* (Xt) *Gallus gallus* (Gg) *Danio rerio* (Dr) *Tetraodon nigroviridis* (Tn) *Fugu rubripes* (Fr) B. Proposed cluster evolution (modified from: Mazet et al. 2006).

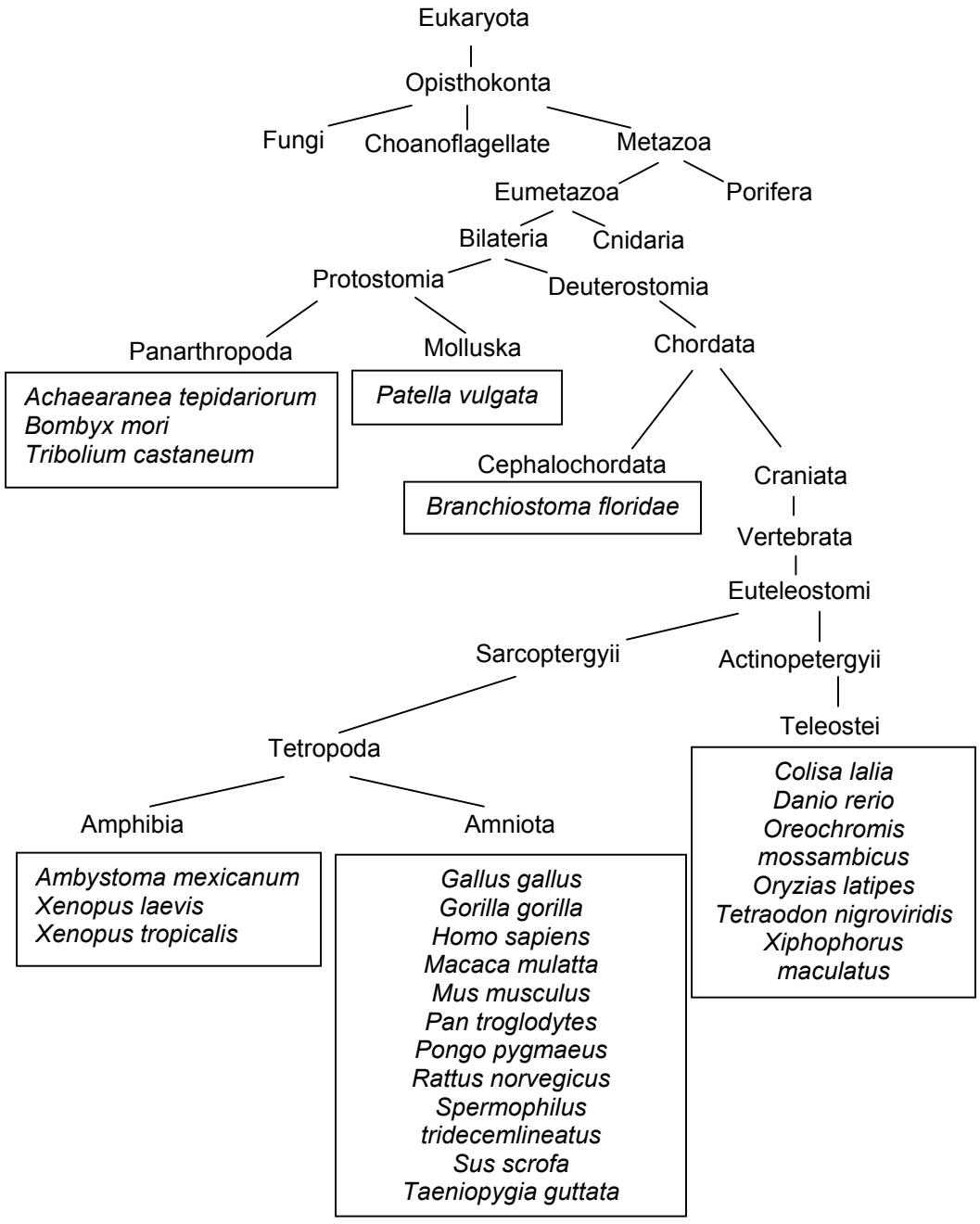


Figure 1-6. Phylogeny of taxonomic groups.

Data is from the NCBI Taxonomy Browser (Wheeler et al. 2006). Species included in Chapter 2 and 3 analyses are shown underneath the relevant group and boxed.

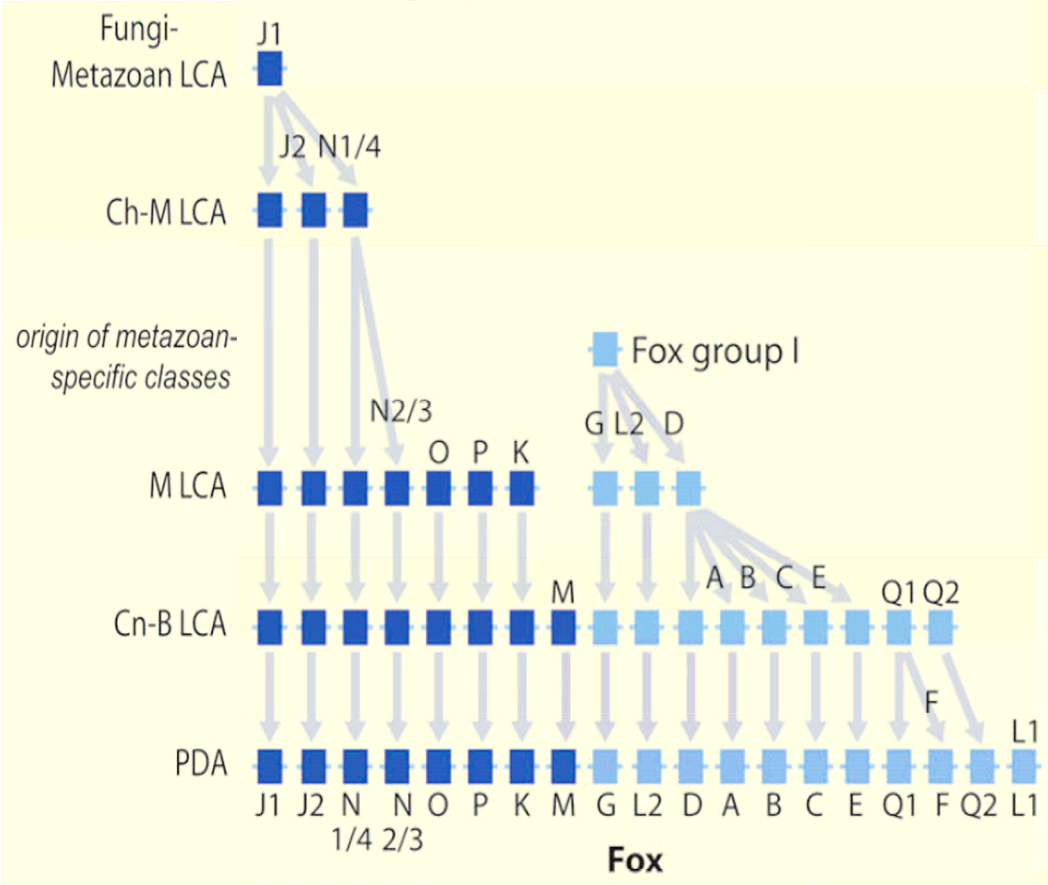


Figure 1-7. Early expansion of the forkhead gene family (modified from: Larroux et al. 2008).

The pathway is based on phylogenetic analyses, intron position and genomic linkage of forkhead genes in various species. Dark blue boxes represent genes with introns; light blue boxes represent genes without introns (Fox group I). Bilaterian (B) Choanoflagellate (Ch) Cnidarian (Cn) Metazoan (M) Last common ancestor (LCA) Protostome-Deuterostome ancestor (PDA)

References

Abderrahmani A, Chevre JC, Otabe S, Chikri M, Hani EH, Vaxillaire M, Hinokio Y, Horikawa Y, Bell GI, Froguel P (2000) Genetic variation in the hepatocyte nuclear factor-3beta gene (HNF3B) does not contribute to maturity-onset diabetes of the young in French Caucasians. Diabetes 49:306-308

Acampora D, Avantaggiato V, Tuorto F, Barone P, Reichert H, Finkelstein R, Simeone A (1998) Murine Otx1 and Drosophila otd genes share conserved genetic functions required in invertebrate and vertebrate brain development. Development 125:1691-1702

Aldinger KA, Lehmann OJ, Hudgins L, Chizhikov VV, Bassuk AG, Ades LC, Krantz ID, Dobyns WB, Millen KJ (2009) FOXC1 is required for normal cerebellar development and is a major contributor to chromosome 6p25.3 Dandy-Walker malformation. Nat Genet 41:1037-1042

Alkhateeb A, Fain PR, Spritz RA (2005) Candidate functional promoter variant in the FOXD3 melanoblast developmental regulator gene in autosomal dominant vitiligo. J Invest Dermatol 125:388-391

Anderson MJ, Viars CS, Czekay S, Cavenee WK, Arden KC (1998) Cloning and characterization of three human forkhead genes that comprise an FKHR-like gene subfamily. Genomics 47:187-199

Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, Pollazzon M, Buoni S, Spiga O, Ricciardi S, Meloni I, Longo I, Mari F, Broccoli V, Zappella M, Renieri A (2008) FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet 83:89-93

Asada S, Daitoku H, Matsuzaki H, Saito T, Sudo T, Mukai H, Iwashita S, Kako K, Kishi T, Kasuya Y, Fukamizu A (2007) Mitogen-activated protein kinases, Erk and p38, phosphorylate and regulate Foxo1. Cell Signal 19:519-527

Baier LJ, Permana PA, Traurig M, Dobberfuhl A, Wiedrich C, Sutherland J, Thuillez P, Luczy-Bachman G, Hara M, Horikawa Y, Hinokio Y, Hanson RL, Bogardus C (2000) Mutations in the genes for hepatocyte nuclear factor (HNF)-1alpha, -4alpha, -1beta, and -3beta; the dimerization cofactor of HNF-1; and insulin promoter factor 1 are not common causes of early-onset type 2 diabetes in Pima Indians. Diabetes Care 23:302-304

Banham AH, Beasley N, Campo E, Fernandez PL, Fidler C, Gatter K, Jones M, Mason DY, Prime JE, Trougouboff P, Wood K, Cordell JL (2001) The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p. Cancer Res 61:8820-8829

Behr R, Sackett SD, Bochkis IM, Le PP, Kaestner KH (2007) Impaired male fertility and atrophy of seminiferous tubules caused by haploinsufficiency for Foxa3. Dev Biol 306:636-645

Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, Kelly TE, Saulsbury FT, Chance PF, Ochs HD (2001) The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet 27:20-21

Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH, Abdul-Karim FW, Montano MM, Keri RA (2010) FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development 137:2045-2054

Berry FB, Saleem RA, Walter MA (2002) FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain. J Biol Chem 277:10292-10297

Berry FB, Skarie JM, Mirzayans F, Fortin Y, Hudson TJ, Raymond V, Link BA, Walter MA (2008) FOXC1 is required for cell viability and resistance to oxidative stress in the eye through the transcriptional regulation of FOXO1A. Hum Mol Genet 17:490-505

Biggs WH, 3rd, Cavenee WK, Arden KC (2001) Identification and characterization of members of the FKHR (FOX O) subclass of winged-helix transcription factors in the mouse. Mamm Genome 12:416-425

Biggs WH, 3rd, Meisenhelder J, Hunter T, Cavenee WK, Arden KC (1999) Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. Proc Natl Acad Sci U S A 96:7421-7426

Blomqvist SR, Vidarsson H, Fitzgerald S, Johansson BR, Ollerstam A, Brown R, Persson AE, Bergstrom GG, Enerback S (2004) Distal renal tubular acidosis in mice that lack the forkhead transcription factor Foxi1. J Clin Invest 113:1560-1570

Blomqvist SR, Vidarsson H, Soder O, Enerback S (2006) Epididymal expression of the forkhead transcription factor Foxi1 is required for male fertility. EMBO J 25:4131-4141

Bohinski RJ, Di Lauro R, Whitsett JA (1994) The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. Mol Cell Biol 14:5671-5681

Bort R, Gomez-Lechon MJ, Castell JV, Jover R (2004) Role of hepatocyte nuclear factor 3 gamma in the expression of human CYP2C genes. Arch Biochem Biophys 426:63-72

Brent MM, Anand R, Marmorstein R (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. Structure 16:1407-1416

Brownawell AM, Kops GJ, Macara IG, Burgering BM (2001) Inhibition of nuclear import by protein kinase B (Akt) regulates the subcellular distribution and activity of the forkhead transcription factor AFX. Mol Cell Biol 21:3534-3546

Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell 96:857-868

Brunet A, Kanai F, Stehn J, Xu J, Sarbassova D, Frangioni JV, Dalal SN, DeCaprio JA, Greenberg ME, Yaffe MB (2002) 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. J Cell Biol 156:817-828

Brunet A, Park J, Tran H, Hu LS, Hemmings BA, Greenberg ME (2001) Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). Mol Cell Biol 21:952-965

Campbell AJ, Lyne L, Brown PJ, Launchbury RJ, Bignone P, Chi J, Roncador G, Lawrie CH, Gatter KC, Kusec R, Banham AH (2010) Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells. Br J Haematol 149:221-230

Carr CW, Moreno-De-Luca D, Parker C, Zimmerman HH, Ledbetter N, Martin CL, Dobyns WB, Abdul-Rahman OA (2010) Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency. Eur J Hum Genet

Castrillon DH, Miao L, Kollipara R, Horner JW, DePinho RA (2003) Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. Science 301:215-218

Cederberg A, Gronning LM, Ahren B, Tasken K, Carlsson P, Enerback S (2001) FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and diet-induced insulin resistance. Cell 106:563-573

Chakrabarti S, Kaur K, Rao KN, Mandal AK, Kaur I, Parikh RS, Thomas R (2009) The transcription factor gene FOXC1 exhibits a limited role in primary congenital glaucoma. Invest Ophthalmol Vis Sci 50:75-83

Chanock SJ, Burdett L, Yeager M, Llaca V, Langerod A, Presswalla S, Kaaresen R, Strausberg RL, Gerhard DS, Kristensen V, Perou CM, Borresen-Dale AL

(2007) Somatic sequence alterations in twenty-one genes selected by expression profile analysis of breast carcinomas. Breast Cancer Res 9:R5

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell 9:279-289

Cirillo LA, Zaret KS (2007) Specific interactions of the wing domains of FOXA1 transcription factor with DNA. J Mol Biol 366:720-724

Clark KL, Halay ED, Lai E, Burley SK (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 364:412-420

Clifton-Bligh RJ, Wentworth JM, Heinz P, Crisp MS, John R, Lazarus JH, Ludgate M, Chatterjee VK (1998) Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nat Genet 19:399-401

Connell F, Brice G, Mortimer P (2008) Phenotypic characterization of primary lymphedema. Ann N Y Acad Sci 1131:140-146

Copley RR (2005) The EH1 motif in metazoan transcription factors. BMC Genomics 6:169

Cowell IG (1994) Repression versus activation in the control of gene transcription. Trends Biochem Sci 19:38-42

Crisponi L, Deiana M, Loi A, Chiappe F, Uda M, Amati P, Bisceglia L, Zelante L, Nagaraja R, Porcu S, Ristaldi MS, Marzella R, Rocchi M, Nicolino M, Lienhardt-Roussie A, Nivelon A, Verloes A, Schlessinger D, Gasparini P, Bonneau D, Cao A, Pilia G (2001) The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. Nat Genet 27:159-166

Curran K, Raible DW, Lister JA (2009) Foxd3 controls melanophore specification in the zebrafish neural crest by regulation of Mitf. Dev Biol 332:408-417

Dagenais SL, Hartsough RL, Erickson RP, Witte MH, Butler MG, Glover TW (2004) Foxc2 is expressed in developing lymphatic vessels and other tissues associated with lymphedema-distichiasis syndrome. Gene Expr Patterns 4:611-619

Dahle MK, Gronning LM, Cederberg A, Blomhoff HK, Miura N, Enerback S, Tasken KA, Tasken K (2002) Mechanisms of FOXC2- and FOXD1-mediated regulation of the RI alpha subunit of cAMP-dependent protein kinase include release of transcriptional repression and activation by protein kinase B alpha and cAMP. J Biol Chem 277:22902-22908

Dayoub R, Groitl P, Dobner T, Bosserhoff AK, Schlitt HJ, Weiss TS (2010) Foxa2 (HNF-3beta) regulates expression of hepatotrophic factor ALR in liver cells. Biochem Biophys Res Commun 395:465-470

De Val S, Chi NC, Meadows SM, Minovitsky S, Anderson JP, Harris IS, Ehlers ML, Agarwal P, Visel A, Xu SM, Pennacchio LA, Dubchak I, Krieg PA, Stainier DY, Black BL (2008) Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. Cell 135:1053-1064

Demuth JP, Hahn MW (2009) The life and death of gene families. BioEssays 31:29-39

Di Gregorio GB, Westergren R, Enerback S, Lu T, Kern PA (2004) Expression of FOXC2 in adipose and muscle and its association with whole body insulin sensitivity. Am J Physiol Endocrinol Metab 287:E799-803

Dottori M, Gross MK, Labosky P, Goulding M (2001) The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Development 128:4127-4138

Duncan SA, Navas MA, Dufort D, Rossant J, Stoffel M (1998) Regulation of a transcription factor network required for differentiation and metabolism. Science 281:692-695

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137:961-971

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872

Ernstsson S, Betz R, Lagercrantz S, Larsson C, Ericksson S, Cederberg A, Carlsson P, Enerback S (1997) Cloning and characterization of freac-9 (FKHL17), a novel kidney-expressed human forkhead gene that maps to chromosome 1p32-p34. Genomics 46:78-85

Ernstsson S, Pierrou S, Hulander M, Cederberg A, Hellqvist M, Carlsson P, Enerback S (1996) Characterization of the human forkhead gene FREAC-4. Evidence for regulation by Wilms' tumor suppressor gene (WT-1) and p53. J Biol Chem 271:21094-21099

Essers MA, Weijzen S, de Vries-Smits AM, Saarloos I, de Ruiter ND, Bos JL, Burgering BM (2004) FOXO transcription factor activation by oxidative stress mediated by the small GTPase Ral and JNK. EMBO J 23:4802-4812

Fang J, Dagenais SL, Erickson RP, Arlt MF, Glynn MW, Gorski JL, Seaver LH, Glover TW (2000) Mutations in *FOXC2 (MFH-1)*, a forkhead family

transcription factor are responsible for the hereditary lymphedema-distichiasis syndrome. Am J Hum Genet 67:1382-1388

Fisher SE, Scharff C (2009) FOXP2 as a molecular window into speech and language. Trends Genet 25:166-177

Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A 106:2700-2705

Forrester JV (2007) The eye : basic sciences in practice. Saunders, Edinburgh ; New York

Frank J, Pignata C, Panteleyev AA, Prowse DM, Baden H, Weiner L, Gaetaniello L, Ahmad W, Pozzi N, Cserhalmi-Friedman PB, Aita VM, Uyttendaele H, Gordon D, Ott J, Brissette JL, Christiano AM (1999) Exposing the human nude phenotype. Nature 398:473-474

Freyaldenhoven BS, Fried C, Wielckens K (2002) FOXD4a and FOXD4b, two new winged helix transcription factors, are expressed in human leukemia cell lines. Gene 294:131-140

Friedman JR, Kaestner KH (2006) The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci 63:2317-2328

Fujita H, Kang M, Eren M, Gleaves LA, Vaughan DE, Kume T (2006) Foxc2 is a common mediator of insulin and transforming growth factor beta signaling to regulate plasminogen activator inhibitor type I gene expression. Circ Res 98:626-634

Galili N, Davis RJ, Fredericks WJ, Mukhopadhyay S, Rausher FJ, 3rd, Emanual BS, Rovera G, Barr FG (1993) Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma. Nature Genetics 5:230-235

Gao N, LeLay J, Vatamaniuk MZ, Rieck S, Friedman JR, Kaestner KH (2008) Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. Genes Dev 22:3435-3448

Gillespie JH (1998) Population genetics : a concise guide. The Johns Hopkins University Press, Baltimore, Md

Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, Gerecht-Nir S, Amit M, Hoke A, Carpenter MK, Itskovitz-Eldor J, Rao MS (2004) Differences between human and mouse embryonic stem cells. Dev Biol 269:360-380

Glauser DA, Schlegel W (2007) The emerging role of FOXO transcription factors in pancreatic beta cells. J Endocrinol 193:195-207

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725-736

Gomez-Skarmeta JL, de la Calle-Mustienes E, Modolell J, Mayor R (1999) Xenopus brain factor-2 controls mesoderm, forebrain and neural crest development. Mech Dev 80:15-27

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433:481-487

Greer EL, Brunet A (2008) FOXO transcription factors in ageing and cancer. Acta Physiol (Oxf) 192:19-28

Gross DN, van den Heuvel AP, Birnbaum MJ (2008) The role of FoxO in the regulation of metabolism. Oncogene 27:2320-2336

Guo Y, Costa R, Ramsey H, Starnes T, Vance G, Robertson K, Kelley M, Reinbold R, Scholer H, Hromas R (2002) The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. Proc Natl Acad Sci U S A 99:3663-3667

Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, Foomani G, Dobrzeniecka S, Krebs MO, Joober R, Lafreniere RG, Lacaille JC, Mottron L, Drapeau P, Beauchamp MH, Phillips MS, Fombonne E, Rouleau GA, Michaud JL (2010) De Novo Mutations in FOXP1 in Cases with Intellectual Disability, Autism, and Language Impairment. Am J Hum Genet

Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA (2002) Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes Dev 16:2650-2661

Happel N, Doenecke D (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431:1-12

Haralambieva E, Adam P, Ventura R, Katzenberger T, Kalla J, Holler S, Hartmann M, Rosenwald A, Greiner A, Muller-Hermelink HK, Banham AH, Ott G (2006) Genetic rearrangement of FOXP1 is predominantly detected in a subset of diffuse large B-cell lymphomas with extranodal presentation. Leukemia 20:1300-1303

Hatini V, Huh SO, Herzlinger D, Soares VC, Lai E (1996) Essential role of stromal mesenchyme in kidney morphogenesis revealed by targeted disruption of Winged Helix transcription factor BF-2. Genes Dev 10:1467-1478

Hatini V, Tao W, Lai E (1994) Expression of winged helix genes, BF-1 and BF-2, define adjacent domains within the developing forebrain and retina. J Neurobiol 25:1293-1309

Hatta M, Cirillo LA (2007) Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. J Biol Chem 282:35583-35593

Hayashi H, Kume T (2008a) Forkhead transcription factors regulate expression of the chemokine receptor CXCR4 in endothelial cells and CXCL12-induced cell migration. Biochem Biophys Res Commun 367:584-589

Hayashi H, Kume T (2008b) Foxc transcription factors directly regulate Dll4 and Hey2 expression by interacting with the VEGF-Notch signaling pathways in endothelial cells. PLoS ONE 3:e2401

Hayashi H, Sano H, Seo S, Kume T (2008) The Foxc2 transcription factor regulates angiogenesis via induction of integrin beta3 expression. J Biol Chem 283:23791-23800

Herrera E, Marcus R, Li S, Williams SE, Erskine L, Lai E, Mason C (2004) Foxd1 is required for proper formation of the optic chiasm. Development 131:5727-5739

Hiemisch H, Monaghan AP, Schutz G, Kaestner KH (1998) Expression of the mouse Fkh1/Mf1 and Mfh1 genes in late gestation embryos is restricted to mesoderm derivatives. Mech Dev 73:129-132

Hillion J, Le Coniat M, Jonveaux P, Berger R, Bernard OA (1997) AF6q21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines a forkhead transcriptional factor subfamily. Blood 90:3714-3719

Holland PW (1999) Gene duplication: past, present and future. Semin Cell Dev Biol 10:541-547

Hong HK, Lass JH, Chakravarti A (1999) Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcriptionfactor gene. Hum Mol Genet 8:625-637

Honkanen RA, Nishimura DY, Swiderski RE, Bennett SR, Hong S, Kwon YH, Stone EM, Sheffield VC, Alward WL (2003) A family with Axenfeld-Rieger syndrome and Peters Anomaly caused by a point mutation (Phe112Ser) in the FOXC1 gene. Am J Ophthalmol 135:368-375

Hosaka T, Biggs WH, 3rd, Tieu D, Boyer AD, Varki NM, Cavenee WK, Arden KC (2004) Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification. Proc Natl Acad Sci U S A 101:2975-2980

Hromas R, Ye H, Spinella M, Dmitrovsky E, Xu D, Costa RH (1999) Genesis, a Winged Helix transcriptional repressor, has embryonic expression limited to the neural crest, and stimulates proliferation in vitro in a neural development model. Cell Tissue Res 297:371-382

Hu H, Wang B, Borde M, Nardone J, Maika S, Allred L, Tucker PW, Rao A (2006) Foxp1 is an essential transcriptional regulator of B cell development. Nat Immunol 7:819-826

Hu MC, Lee DF, Xia W, Golfman LS, Ou-Yang F, Yang JY, Zou Y, Bao S, Hanada N, Saso H, Kobayashi R, Hung MC (2004) IkappaB kinase promotes tumorigenesis through inhibition of forkhead FOXO3a. Cell 117:225-237

Huang L (2009) Function and regulation of the FOXC transcription factors., University of Alberta, Edmonton

Hulander M, Kiernan AE, Blomqvist SR, Carlsson P, Samuelsson EJ, Johansson BR, Steel KP, Enerback S (2003) Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in inner ears of Foxi1 null mutant mice. Development 130:2013-2025

Hulander M, Wurst W, Carlsson P, Enerback S (1998) The winged helix transcription factor Fkh10 is required for normal development of the inner ear. Nat Genet 20:374-376

Ignatius MS, Moose HE, El-Hodiri HM, Henion PD (2008) colgate/hdac1 Repression of foxd3 expression is required to permit mitfa-dependent melanogenesis. Dev Biol 313:568-583

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97-108

Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, Walter MA (2009) Severe molecular defects of a novel FOXC1 W152G mutation result in aniridia. Invest Ophthalmol Vis Sci 50:3573-3579

Ito YA, Footz TK, Murphy TC, Courtens W, Walter MA (2007) Analyses of a novel L130F missense mutation in FOXC1. Arch Ophthalmol 125:128-135

Johansson CC, Dahle MK, Blomqvist SR, Gronning LM, Aandahl EM, Enerback S, Tasken K (2003) A winged helix forkhead (FOXD2) tunes sensitivity to cAMP in T lymphocytes through regulation of cAMP-dependent protein kinase RIalpha. J Biol Chem 278:17573-17579

Jonckheere N, Vincent A, Perrais M, Ducourouble MP, Male AK, Aubert JP, Pigny P, Carraway KL, Freund JN, Renes IB, Van Seuningen I (2007) The human mucin MUC4 is transcriptionally regulated by caudal-related homeobox, hepatocyte nuclear factors, forkhead box A, and GATA endodermal transcription factors in epithelial cancer cells. J Biol Chem 282:22638-22650

Jurgens G, Wieschaus E, Nusslein-Volhard C, Kluding H (1984) Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome. Roux's Arch Dev Biol 193:283-295

Kaestner KH, Katz J, Liu Y, Drucker DJ, Schutz G (1999) Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. Genes Dev 13:495-504

Kaestner KH, Knöchel W, Martínez DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev 14:142-146

Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R (1993) A C. elegans mutant that lives twice as long as wild type. Nature 366:461-464

Kidson SH, Kume T, Deng K, Winfrey V, Hogan BL (1999) The forkhead/winged-helix gene, Mf1, is necessary for the normal development of the cornea and formation of the anterior chamber in the mouse eye. Dev Biol 211:306-322

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107-116

Koh KP, Sundrud MS, Rao A (2009) Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3. PLoS One 4:e8109

Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH (2009) Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 462:213-217

Kops GJ, de Ruiter ND, De Vries-Smits AM, Powell DR, Bos JL, Burgering BM (1999) Direct control of the Forkhead transcription factor AFX by protein kinase B. Nature 398:630-634

Kos R, Reedy MV, Johnson RL, Erickson CA (2001) The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. Development 128:1467-1479

Kume T, Deng K, Hogan BL (2000a) Minimal phenotype of mice homozygous for a null mutation in the forkhead/winged helix gene, Mf2. Mol Cell Biol 20:1419-1425

Kume T, Deng K, Hogan BL (2000b) Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (Mfh1) are required for the early organogenesis of the kidney and urinary tract. Development 127:1387-1395

Kume T, Deng KY, Winfrey V, Gould DB, Walter MA, Hogan BL (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell 93:985-996

Kume T, Jiang H, Topczewska JM, Hogan BL (2001) The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes Dev 15:2470-2482

Kurth I, Hentschke M, Hentschke S, Borgmeyer U, Gal A, Hubner CA (2006) The forkhead transcription factor Foxi1 directly activates the AE4 promoter. Biochem J 393:277-283

Labosky PA, Kaestner KH (1998) The winged helix transcription factor Hfh2 is expressed in neural crest and spinal cord during mouse development. Mech Dev 76:185-190

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001a) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523

Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001b) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523

Lai E, Prezioso VR, Smith E, Litvin O, Costa RH, Darnell JE, Jr. (1990) HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. Genes Dev 4:1427-1436

Lai E, Prezioso VR, Tao WF, Chen WS, Darnell JE, Jr. (1991) Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. Genes Dev 5:416-427

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980-996

Lee CS, Friedman JR, Fulmer JT, Kaestner KH (2005a) The initiation of liver development is dependent on Foxa transcription factors. Nature 435:944-947

Lee CS, Sund NJ, Behr R, Herrera PL, Kaestner KH (2005b) Foxa2 is required for the differentiation of pancreatic alpha-cells. Dev Biol 278:484-495

Lee HC, Huang HY, Lin CY, Chen YH, Tsai HJ (2006) Foxd3 mediates zebrafish myf5 expression during early somitogenesis. Dev Biol 290:359-372

Lee HJ, Hwang M, Chattopadhyay S, Choi HS, Lee K (2008) Hepatocyte nuclear factor-3 alpha (HNF-3alpha) negatively regulates androgen receptor transactivation in prostate cancer cells. Biochem Biophys Res Commun 367:481-486

Lehmann OJ, Ebenezer ND, Jordan T, Fox M, Ocaka L, Payne A, Leroy BP, Clark BJ, Hitchings RA, Povey S, Khaw PT, Bhattacharya SS (2000)

Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma. Am J Hum Genet 67:1129-1135

Lehtinen MK, Yuan Z, Boag PR, Yang Y, Villen J, Becker EB, DiBacco S, de la Iglesia N, Gygi S, Blackwell TK, Bonni A (2006) A conserved MST-FOXO signaling pathway mediates oxidative-stress responses and extends life span. Cell 125:987-1001

Levinson RS, Batourina E, Choi C, Vorontchikhina M, Kitajewski J, Mendelsohn CL (2005) Foxd1-dependent signals control cellularity in the renal capsule, a structure required for normal renal development. Development 132:529-539

Li B, Samanta A, Song X, Furuuchi K, Iacono KT, Kennedy S, Katsumata M, Saouaf SJ, Greene MI (2006) FOXP3 ensembles in T-cell regulation. Immunol Rev 212:99-113

Li S, Weidenfeld J, Morrisey EE (2004a) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 24:809-822

Li S, Zhou D, Lu MM, Morrisey EE (2004b) Advanced cardiac morphogenesis does not require heart tube fusion. Science 305:1619-1622

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96-99

Lin L, Miller CT, I CJ, Prescott MS, Dagenais SL, Wu R, Yee J, Orringer MB, Misek DE, Hanash SM, Glover TW, Beer DG (2002) The hepatocyte nuclear factor 3 alpha gene, HNF3alpha (FOXA1), on chromosome band 14q13 is amplified and overexpressed in esophageal and lung adenocarcinomas. Cancer Res 60:5273-5279

Lin L, Peng SL (2006) Coordination of NF-kappaB and NFAT antagonism by the forkhead transcription factor Foxd1. J Immunol 176:4793-4803

Lister JA, Cooper C, Nguyen K, Modrell M, Grant K, Raible DW (2006) Zebrafish Foxd3 is required for development of a subset of neural crest derivatives. Dev Biol 290:92-104

Littler DR, Alvarez-Fernandez M, Stein A, Hibbert RG, Heidebrecht T, Aloy P, Medema RH, Perrakis A (2010) Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. Nucleic Acids Res

Liu Q, Melnikova IN, Hu M, Gardner PD (1999) Cell type-specific activation of neuronal nicotinic acetylcholine receptor subunit genes by Sox10. J Neurosci 19:9747-9755

Locker J (2001) Transcription factors. BIOS ;

Academic Press, Oxford
San Diego, CA

Lu MM, Li S, Yang H, Morrisey EE (2002) Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. Mech Dev 119 Suppl 1:S197-202

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132:958-970

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155

Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. Evolution 62:2131-2154

Mani SA, Yang J, Brooks M, Schwaninger G, Zhou A, Miura N, Kutok JL, Hartwell K, Richardson AL, Weinberg RA (2007) Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. Proc Natl Acad Sci U S A 104:10069-10074

Marsden I, Jin C, Liao X (1998) Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. J Mol Biol 278:293-299

Mattiske D, Kume T, Hogan BL (2006) The mouse forkhead gene Foxc1 is required for primordial germ cell migration and antral follicle development. Dev Biol 290:447-458

Mazet F, Amemiya CT, Shimeld SM (2006) An ancient Fox gene cluster in bilaterian animals. Curr Biol 16:R314-316

McMullin RP, Dobi A, Mutton LN, Orosz A, Maheshwari S, Shashikant CS, Bieberich CJ (2010) A FOXA1-binding enhancer regulates Hoxb13 expression in the prostate gland. Proc Natl Acad Sci U S A 107:98-103

Mears AJ, Jordan T, Mirzayans F, Dubois S, Kume T, Parlee M, Ritch R, Koop B, Kuo WL, Collins C, Marshall J, Gould DB, Pearce W, Carlsson P, Enerback S, Morissette J, Bhattacharya S, Hogan B, Raymond V, Walter MA (1998) Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. Am J Hum Genet 63:1316-1328

Minoretti P, Arra M, Emanuele E, Olivieri V, Aldeghi A, Politi P, Martinelli V, Pesenti S, Falcone C (2007) A W148R mutation in the human FOXD4 gene segregating with dilated cardiomyopathy, obsessive-compulsive disorder, and suicidality. Int J Mol Med 19:369-372

Mirzayans F, Gould DB, Héon E, Billingsley GD, Cheung JC, Mears AJ, Walter MA (2000) Axenfeld-Rieger syndrome resulting from mutation of the *FKHL7* gene on chromosome 6p25. European Journal of Human Genetics 8:71-74

Miura N, Iida K, Kakinuma H, Yang XL, Sugiyama T (1997) Isolation of the mouse (MFH-1) and human (FKHL 14) mesenchyme fork head-1 genes reveals conservation of their gene and protein structures. Genomics 41:489-492

Montalta-He H, Leemans R, Loop T, Strahm M, Certa U, Primig M, Acampora D, Simeone A, Reichert H (2002) Evolutionary conservation of otd/Otx2 transcription factor action: a genome-wide microarray analysis in Drosophila. Genome Biol 3:RESEARCH0015

Muggerud AA, Ronneberg JA, Warnberg F, Botling J, Busato F, Jovanovic J, Solvang H, Bukholm I, Borresen-Dale AL, Kristensen VN, Sorlie T, Tost J (2010) Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Res 12:R3

Nakae J, Park BC, Accili D (1999) Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a Wortmannin-sensitive pathway. J Biol Chem 274:15982-15985

Navas MA, Vaisse C, Boger S, Heimesaat M, Kollee LA, Stoffel M (2000) The human HNF-3 genes: cloning, partial sequence and mutation screening in patients with impaired glucose homeostasis. Human Heredity 50:370-381

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426

Nissen RM, Yan J, Amsterdam A, Hopkins N, Burgess SM (2003) Zebrafish foxi one modulates cellular responses to Fgf signaling required for the integrity of ear and jaw patterning. Development 130:2543-2554

Norrmen C, Ivanov KI, Cheng J, Zangger N, Delorenzi M, Jaquet M, Miura N, Puolakkainen P, Horsley V, Hu J, Augustin HG, Yla-Herttuala S, Alitalo K, Petrova TV (2009) FOXC2 controls formation and maturation of lymphatic collecting vessels through cooperation with NFATc1. J Cell Biol 185:439-457

Ohyama T, Groves AK (2004) Expression of mouse Foxi class genes in early craniofacial development. Dev Dyn 231:640-646

Overdier DG, Ye H, Peterson RS, Clevidence DE, Costa RH (1997) The winged helix transcriptional activator HFH-3 is expressed in the distal tubules of embryonic and adult mouse kidney. J Biol Chem 272:13725-13730

Paik JH, Kollipara R, Chu G, Ji H, Xiao Y, Ding Z, Miao L, Tothova Z, Horner JW, Carrasco DR, Jiang S, Gilliland DG, Chin L, Wong WH, Castrillon DH, DePinho RA (2007) FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. Cell 128:309-323

Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol 10:271-281

Pani L, Overdier DG, Porcella A, Qian X, Lai E, Costa RH (1992) Hepatocyte nuclear factor 3β contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. Mol Cell Biol 12:3723-3732

Panowski SH, Wolff S, Aguilaniu H, Durieux J, Dillin A (2007) PHA-4/Foxa mediates diet-restriction-induced longevity of C. elegans. Nature 447:550-555

Pariani MJ, Spencer A, Graham JM, Jr., Rimoin DL (2009) A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. Eur J Med Genet 52:123-127

Parry P, Wei Y, Evans G (1994) Cloning and characterization of the t(X;11) breakpoint from a leukemic cell line identify a new member of the forkhead gene family. Genes Chromosomes & Cancer 11:79-84

Patthy L (2008) Protein evolution. Blackwell Science, Malden, MA ; Oxford

Peng SL (2008) Foxo in the immune system. Oncogene 27:2337-2344

Petrova TV, Karpanen T, Norrmen C, Mellor R, Tamakoshi T, Finegold D, Ferrell R, Kerjaschki D, Mortimer P, Yla-Herttuala S, Miura N, Alitalo K (2004) Defective valves and abnormal mural cell recruitment underlie lymphatic vascular failure in lymphedema distichiasis. Nat Med 10:974-981

Pierrou S, Hellqvist M, Samuelsson L, Enerback S, Carlsson P (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J 13:5002-5012

Pohl BS, Knochel W (2001) Overexpression of the transcriptional repressor FoxD3 prevents neural crest formation in Xenopus embryos. Mech Dev 103:93-106

Pohl BS, Knochel W (2002) Temporal and spatial expression patterns of FoxD2 during the early development of Xenopus laevis. Mech Dev 111:181-184

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB (2006) Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature 440:1050-1053

Qian X, Costa RH (1995) Analysis of hepatocyte nuclear factor-3β protein domains required for transcriptional activation and nuclear targeting. Nucleic Acids Res 23:1184-1191

Ramaswamy S, Nakamura N, Sansal I, Bergeron L, Sellers WR (2002) A novel mechanism of gene regulation and tumor suppression by the transcription factor FKHR. Cancer Cell 2:81-91

Rausa FM, Tan Y, Costa RH (2003) Association between hepatocyte nuclear factor 6 (HNF-6) and FoxA2 DNA binding domains stimulates FoxA2 transcriptional activity but inhibits HNF-6 DNA binding. Mol Cell Biol 23:437-449

Ray PS, Wang J, Qu Y, Sim MS, Shamonki J, Bagaria SP, Ye X, Liu B, Elashoff D, Hoon DS, Walter MA, Martens JW, Richardson AL, Giuliano AE, Cui X (2010) FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. Cancer Res 70:3870-3876

Rena G, Guo S, Cichy SC, Unterman TG, Cohen P (1999) Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. J Biol Chem 274:17179-17183

Rousso DL, Gaber ZB, Wellik D, Morrisey EE, Novitch BG (2008) Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. Neuron 59:226-240

Saleem RA, Murphy TC, Liebmann JM, Walter MA (2003) Identification and analysis of a novel mutation in the FOXC1 forkhead domain. Invest Ophthalmol Vis Sci 44:4608-4612

Sasaki H, Hogan BL (1993) Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. Development 118:47-59

Semina EV, Brownell I, Mintz-Hittner HA, Murray JC, Jamrich M (2001) Mutations in the human forkhead transcription factor FOXE3 associated with anterior segment ocular dysgenesis and cataracts. Hum Mol Genet 10:231-236

Seo S, Fujita H, Nakano A, Kang M, Duarte A, Kume T (2006) The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. Dev Biol 294:458-470

Shen W, Scearce LM, Brestelli JE, Sund NJ, Kaestner KH (2001) Foxa3 (hepatocyte nuclear factor 3gamma) is required for the regulation of hepatic GLUT2 expression and the maintenance of glucose homeostasis during a prolonged fast. J Biol Chem 276:42812-42817

Shi C, Sakuma M, Mooroka T, Liscoe A, Gao H, Croce KJ, Sharma A, Kaplan D, Greaves DR, Wang Y, Simon DI (2008) Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte differentiation and macrophage function. Blood 112:4699-4711

Shi C, Zhang X, Chen Z, Sulaiman K, Feinberg MW, Ballantyne CM, Jain MK, Simon DI (2004) Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. J Clin Invest 114:408-418

Shih DQ, Navas MA, Kuwajima S, Duncan SA, Stoffel M (1999) Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3alpha-deficient mice. Proc Natl Acad Sci U S A 96:10152-10157

Shiyanova T, Liao X (1999) The dissociation rate of a winged helix protein-DNA complex is influenced by non-DNA contact residues. Arch Biochem Biophys 362:356-362

Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MA, Rabidou D, Santucci AC, Perl D, Morrisey E, Buxbaum JD (2005) Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A 102:9643-9648

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497

Smith RS, Zabaleta A, Kume T, Savinova OV, Kidson SH, Martin JE, Nishimura DY, Alward WL, Hogan BL, John SW (2000) Haploinsufficiency of the transcription factors FOXC1 and FOXC2 results in aberrant ocular development. Hum Mol Genet 9:1021-1032

So CW, Cleary ML (2002) MLL-AFX requires the transcriptional effector domains of AFX to transform myeloid progenitors and transdominantly interfere with forkhead protein function. Mol Cell Biol 22:6542-6552

Solomon KS, Kudoh T, Dawid IB, Fritz A (2003a) Zebrafish foxi1 mediates otic placode formation and jaw development. Development 130:929-940

Solomon KS, Logsdon JM, Jr., Fritz A (2003b) Expression and phylogenetic analyses of three zebrafish FoxI class genes. Dev Dyn 228:301-307

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81:1144-1157

Stankiewicz P, Sen P, Bhatt SS, Storer M, Xia Z, Bejjani BA, Ou Z, et al. (2009) Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating

mutations of FOXF1 cause alveolar capillary dysplasia and other malformations. Am J Hum Genet 84:780-791

Stewart RA, Arduini BL, Berghmans S, George RE, Kanki JP, Henion PD, Look AT (2006) Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. Dev Biol 292:174-188

Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L (2006) Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14:159-166

Strungaru MH, Dinu I, Walter MA (2007) Genotype-phenotype correlations in Axenfeld-Rieger malformation and glaucoma patients with FOXC1 and PITX2 mutations. Invest Ophthalmol Vis Sci 48:228-237

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101:6062-6067

Sublett JE, Jeon IS, Shapiro DN (1995) The alveolar rhabdomyosarcoma PAX3/FKHR fusion protein is a transcriptional activator. Oncogene 11:545-552

Sullivan SA, Akers L, Moody SA (2001) foxD5a, a Xenopus winged helix gene, maintains an immature neural ectoderm via transcriptional repression that is dependent on the C-terminal domain. Dev Biol 232:439-457

Sutton J, Costa R, Klug M, Field L, Xu D, Largaespada DA, Fletcher CF, Jenkins NA, Copeland NG, Klemsz M, Hromas R (1996) Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. J Biol Chem 271:23126-23133

Takahashi K, Liu FC, Hirokawa K, Takahashi H (2008) Expression of Foxp4 in the developing and adult rat forebrain. J Neurosci Res 86:3106-3116

Takaishi H, Konishi H, Matsuzaki H, Ono Y, Shirai Y, Saito N, Kitamura T, Ogawa W, Kasuga M, Kikkawa U, Nishizuka Y (1999) Regulation of nuclear translocation of forkhead transcription factor AFX by protein kinase B. Proc Natl Acad Sci U S A 96:11836-11841

Tamimi Y, Skarie JM, Footz T, Berry FB, Link BA, Walter MA (2006) FGF19 is a target for FOXC1 regulation in ciliary body-derived cells. Hum Mol Genet 15:3229-3240

Tan Y, Hughes D, Wang X, Costa RH (2002) Adenovirus-mediated increase in HNF-3beta or HNF-3alpha shows differences in levels of liver glycogen and gene expression. Hepatology 35:30-39

Tang ED, Nunez G, Barr FG, Guan KL (1999) Negative regulation of the forkhead transcription factor FKHR by Akt. J Biol Chem 274:16741-16746

Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38:615-643

Teufel A, Wong EA, Mukhopadhyay M, Malik N, Westphal H (2003) FoxP4, a novel forkhead transcription factor. Biochim Biophys Acta 1627:147-152

Thomas AJ, Erickson CA (2009) FOXD3 regulates the lineage switch between neural crest-derived glial cells and pigment cells by repressing MITF through a non-canonical mechanism. Development 136:1849-1858

Tomarev SI (2001) Eyeing a new route along an old pathway. Nat Med 7:294-295

Tompers DM, Foreman RK, Wang Q, Kumanova M, Labosky PA (2005) Foxd3 is required in the trophoblast progenitor cell lineage of the mouse embryo. Dev Biol 285:126-137

Tsai KL, Huang CY, Chang CH, Sun YJ, Chuang WJ, Hsiao CD (2006) Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. J Biol Chem 281:17400-17409

Tumer Z, Bach-Holm D (2009) Axenfeld-Rieger syndrome and spectrum of PITX2 and FOXC1 mutations. Eur J Hum Genet 17:1527-1539

Valleix S, Niel F, Nedelec B, Algros MP, Schwartz C, Delbosc B, Delpech M, Kantelip B (2006) Homozygous nonsense mutation in the FOXE3 gene as a cause of congenital primary aphakia in humans. Am J Hum Genet 79:358-364

van der Horst A, de Vries-Smits AM, Brenkman AB, van Triest MH, van den Broek N, Colland F, Maurice MM, Burgering BM (2006) FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. Nat Cell Biol 8:1064-1073

van der Vos KE, Coffer PJ (2008) FOXO-binding partners: it takes two to tango. Oncogene 27:2289-2299

van Dongen MJ, Cederberg A, Carlsson P, Enerback S, Wikstrom M (2000) Solution structure and dynamics of the DNA-binding domain of the adipocyte-transcription factor FREAC-11. J Mol Biol 296:351-359

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 81:1232-1250

Vidarsson H, Westergren R, Heglind M, Blomqvist SR, Breton S, Enerback S (2009) The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis. PLoS One 4:e4471

Wagner GP, Lynch VJ (2008) The gene regulatory logic of transcription factor evolution. Trends Ecol Evol 23:377-385

Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang SL, Wert S, Stahlman MT, Whitsett JA (2005) Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J Biol Chem 280:13809-13816

Wang B, Lin D, Li C, Tucker P (2003) Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem 278:24259-24268

Wang JC, Waltner-Law M, Yamada K, Osawa H, Stifani S, Granner DK (2000) Transducin-like enhancer of split proteins, the human homologs of Drosophila groucho, interact with hepatic nuclear factor 3beta. J Biol Chem 275:18418-18423

Weigel D, Jackle H (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell 63:455-456

Weigelt J, Climent I, Dahlman-Wright K, Wikstrom M (2000) 1H, 13C and 15N resonance assignments of the DNA binding domain of the human forkhead transcription factor AFX. J Biomol NMR 17:181-182

Weinstein DC, Ruiz i Altaba A, Chen WS, Hoodless P, Prezioso VR, Jessell TM, Darnell JE, Jr. (1994) The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78:575-588

Weisschuh N, Wolf C, Wissinger B, Gramer E (2008) A novel mutation in the FOXC1 gene in a family with Axenfeld-Rieger syndrome and Peters' anomaly. Clin Genet

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, et al. (2006) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 34:D173-180

Wierstra I, Alves J (2006) Transcription factor FOXM1c is repressed by RB and activated by cyclin D1/Cdk4. Biol Chem 387:949-962

Wijchers PJ, Hoekman MF, Burbach JP, Smidt MP (2005) Cloning and analysis of the murine Foxi2 transcription factor. Biochim Biophys Acta 1731:133-138

Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, Masaki KH, Willcox DC, Rodriguez B, Curb JD (2008) FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A 105:13987-13992

Williamson EA, Wolf I, O'Kelly J, Bose S, Tanosaki S, Koeffler HP (2006) BRCA1 and FOXA1 proteins coregulate the expression of the cell cycle-dependent kinase inhibitor p27(Kip1). Oncogene 25:1391-1399

Winnier GE, Hargett L, Hogan BL (1997) The winged helix transcription factor MFH1 is required for proliferation and patterning of paraxial mesoderm in the mouse embryo. Genes Dev 11:926-940

Winnier GE, Kume T, Deng K, Rogers R, Bundy J, Raines C, Walter MA, Hogan BL, Conway SJ (1999) Roles for the winged helix transcription factors MF1 and MFH1 in cardiovascular development revealed by nonallelic noncomplementation of null alleles. Dev Biol 213:418-431

Wotton KR, Mazet F, Shimeld SM (2008) Expression of FoxC, FoxF, FoxL1, and FoxQ1 genes in the dogfish Scyliorhinus canicula defines ancient and derived roles for Fox genes in vertebrate development. Dev Dyn 237:1590-1603

Wotton KR, Shimeld SM (2006) Comparative genomics of vertebrate Fox cluster loci. BMC Genomics 7:271

Wray GA (2003) Transcriptional regulation and the evolution of development. Int J Dev Biol 47:675-684

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206-216

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20:1377-1419

Wu SC, Grindley J, Winnier GE, Hargett L, Hogan BL (1998) Mouse Mesenchyme forkhead 2 (Mf2): expression, DNA binding and induction by sonic hedgehog during somitogenesis. Mech Dev 70:3-13

Yaklichkin S, Steiner AB, Lu Q, Kessler DS (2007) FoxD3 and Grg4 physically interact to repress transcription and induce mesoderm in Xenopus. J Biol Chem 282:2548-2557

Yan J, Xu L, Crawford G, Wang Z, Burgess SM (2006) The forkhead transcription factor FoxI1 remains bound to condensed mitotic chromosomes and stably remodels chromatin structure. Mol Cell Biol 26:155-168

Yang H, Zhao R, Yang HY, Lee MH (2005) Constitutively active FOXO4 inhibits Akt activity, regulates p27 Kip1 stability, and suppresses HER2-mediated tumorigenicity. Oncogene 24:1924-1935

Yang T, Vidarsson H, Rodrigo-Blomqvist S, Rosengren SS, Enerback S, Smith RJ (2007) Transcriptional control of SLC26A4 is involved in Pendred syndrome

and nonsyndromic enlargement of vestibular aqueduct (DFNB4). Am J Hum Genet 80:1055-1063

Yang Z (2006) Computational molecular evolution. Oxford University Press, Oxford

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32-43

Yu L, Wei Q, Jin L, Nishigori H, Nishigori T, Tomura H, Fujita J, Yamada Y, Seino Y, Takeda J (2001) Genetic variation in the hepatocyte nuclear factor (HNF)-3alpha gene does not contribute to maturity-onset diabetes of the young in Japanese. Horm Metab Res 33:163-166

Zarbalis K, Siegenthaler JA, Choe Y, May SR, Peterson AS, Pleasure SJ (2007) Cortical dysplasia and skull defects in mice with a Foxc1 allele reveal the role of meningeal differentiation in regulating cortical development. Proc Natl Acad Sci U S A 104:14002-14007

Zeng J, Yan J, Wang T, Mosbrook-Davis D, Dolan KT, Christensen R, Stormo GD, Haussler D, Lathrop RH, Brachmann RK, Burgess SM (2008) Genome wide screens in yeast to identify potential binding sites and target genes of DNA-binding proteins. Nucleic Acids Res 36:e8

Zhang H, Palmer R, Gao X, Kreidberg J, Gerald W, Hsiao L, Jensen RV, Gullans SR, Haber DA (2003) Transcriptional activation of placental growth factor by the forkhead/winged helix transcription factor FoxD1. Curr Biol 13:1625-1629

Zhang J, Webb DM, Podlaha O (2002a) Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics 162:1825-1835

Zhang K, Zhang L, Rao F, Brar B, Rodriguez-Flores JL, Taupenot L, O'Connor DT (2010) Human tyrosine hydroxylase natural genetic variation: delineation of functional transcriptional control motifs disrupted in the proximal promoter. Circ Cardiovasc Genet 3:187-198

Zhang X, Gan L, Pan H, Guo S, He X, Olson ST, Mesecar A, Adam S, Unterman TG (2002b) Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. Direct and indirect effects on nuclear/cytoplasmic shuttling and DNA binding. J Biol Chem 277:45276-45284

Zhao X, Gan L, Pan H, Kan D, Majeski M, Adam SA, Unterman TG (2004) Multiple elements regulate nuclear/cytoplasmic shuttling of FOXO1: characterization of phosphorylation- and 14-3-3-dependent and -independent mechanisms. Biochem J 378:839-849

Zhu Q, Yamagata K, Yu L, Tomura H, Yamada S, Yang Q, Yoshiuchi I, Sumi S, Miyagawa J, Takeda J, Hanafusa T, Matsuzawa Y (2000) Identification of

missense mutations in the hepatocyte nuclear factor-3beta gene in Japanese subjects with late-onset Type II diabetes mellitus. Diabetologia 43:1197-1200

Chapter 2: Identification and analysis of evolutionary selection pressures acting at the molecular level in five forkhead subfamilies.

This chapter contains work published in: Fetterman CD, Rannala B, Walter MA (2008) Identification and analysis of evolutionary selection pressures acting at the molecular level in five forkhead subfamilies. BMC Evol Biol 8:261.

Authors' Contributions: CDF participated in study design and carried out all experiments. BR and MAW conceived of the study and participated in its design.

Introduction

Members of the forkhead gene family act as transcription regulators in multiple biological processes during development and in adult tissues (Carlsson and Mahlapuu 2002). While some research has examined forkhead gene family evolution (Enard et al. 2002; Mazet et al. 2006; Wotton and Shimeld 2006; Larroux et al. 2008), selection pressures on individual codons have only been measured in the FoxL2 subfamily (Cocquet et al. 2003). Here selection pressures were measured on subfamilies to explore the evolution of forkhead paralogs and orthologs. The FoxA, FoxD, FoxI, FoxO and FoxP subfamilies were each examined individually. *In silico* site analyses were used to estimate ω (the nonsynonymous substitution rate divided by the synonymous substitution rate) for each codon in an alignment of sequences (Yang 1997). All of the subfamilies were experiencing neutral amino acid changes, which do not decrease or increase fitness, and negative selection, which counteracts deleterious amino acid changes. Branch-site models, which allow estimated selection pressures along specified lineages to vary as compared to the remaining phylogeny (Yang 1997), identified positive selection along branches leading to the FoxA3 and Protostomia/Cephalochordata clades in the FoxA cluster and the branch leading to the FoxO3 clade in the FoxO cluster. Consideration of selection pressures observed in conjunction with known functional information allowed prediction of residue function and refinement of domain boundaries. Identification of residues that differentiate orthologs and paralogs provided insight into the development and functional consequences of paralogs and forkhead subfamily composition

differences among species. Overall after gene duplication, differentiation and subsequent fixation of amino acid changes through negative selection has occurred.

Methods

Sequence Data

A list of 672 amino acid sequences containing the forkhead domain was retrieved from the NCBI Entrez Protein Database using the Conserved Domain Architecture Retrieval Tool (CDART) (Geer et al. 2002) in conjunction with the Conserved Domain Database forkhead domain definition, cd00059 (Marchler-Bauer et al. 2005; Wheeler et al. 2006). Sequences described as partial, incomplete, fragment, predicted, putative and hypothetical as well as duplicates and isoforms were excluded resulting in a total of 299 sequences from 51 species analyzed. Initial analysis of all known forkhead genes simultaneously using global or local alignment methods, and parsimony, likelihood or Bayesian phylogenetic methods, produced trees with inconsistent subfamily placement due to low sequence homology outside of the forkhead domain. BLASTCLUST was therefore used to cluster the amino acid sequences in groups of 30% identity over 90% of their length (Altschul et al. 1997). Sequences are generally considered evolutionarily related when they have a percent identity of 25% or greater and this level of relatedness also improves alignment accuracy (Doolittle 1981; Thompson et al. 1999). The combination of percent identity and sequence length over which the identity is measured was initially varied using 25, 30 and 40% identity and 80, 90 and 100% length. For robust selection analysis accuracy and power 10 or more sequences are recommended (Anisimova et al. 2001; Anisimova et al. 2002), therefore only clusters containing 10 or more sequences were included in these analyses. Clusters produced using a 25 or 30% identity level were identical

so the 25% identity cluster results were not considered further. Clusters with 30% or 40% identity over 100% of their length were also the same. Clusters produced when measuring percent identity over 80% of sequence length led to different Fox subclasses clustered together, therefore these clusters were not considered for further use. Clustering at 30 or 40% identity over 90% sequence length each produced six clusters with more than 10 members. At a sequence length of 100%, only two clusters had more than 10 members so clusters at 100% sequence length were not considered. The remaining clusters to be considered, 30 and 40% identity over 90% length, differed only in 2 groups. At 40% identity the FoxD group lost five sequences and the FoxO group lost one sequence as compare to the groups at 30% identity. The clusters produced using a measure of 30% identity over 90% sequence length were maintained and analyzed independently in all further experiments. There were six clusters, named for the subfamily contained within each one, identified: FoxA, FoxC, FoxD, FoxI, FoxO and FoxP (Table 2-1). The FoxC cluster will be addressed in Chapter 3.

Alignment and Phylogenetic Analysis

Each cluster was aligned independently using a combination of CLUSTALX1.83 (Thompson et al. 1997) and CLUSTALW1.81 (Thompson et al. 1994). Amino acid sequences were aligned rather than nucleotide sequences so that gaps would not be introduced into the corresponding codons. The first step of the alignment procedure was optimization of the substitution matrices, BLOSUM, Gonnet or PAM, used during the pairwise and multiple alignment

phases. A full multiple alignment was performed with CLUSTALX using each combination of pairwise and multiple alignment substitution matrices with all other parameters at default values. A neighbor-joining tree was constructed from each alignment using the CLUSTALX program. The quality scores, using default parameters, assigned by CLUSTALX to each column in the alignment were averaged over all sites by the program Tune ClustalX (Hall 2004), giving an overall Q-score for the alignment. Quality scores are a measure of the conservation of amino acids in an aligned column and were calculated using the Gonnet PAM 250 matrix. The neighbor-joining trees were then examined to determine the most common topology. The alignment producing a neighbor-joining tree with the common topology and the highest Q-score was considered to be made with the optimal combination of pairwise and multiple alignment substitution matrices. In all cases, the alignment that produced the highest Q-score also produced a neighbor-joining tree with the most common topology. The next step in aligning the sequences was optimization of the gap open and gap extension penalties used during the multiple alignment phase. The dendrogram produced by the alignment that optimized the substitution matrices was used during this step so that the pairwise gap penalty parameters did not need to be optimized. The optimal substitution matrix for the data set, determined during the first step of the alignment, was also used during this phase. Alignments were created with CLUSTALW varying the gap open penalty from five to fifteen in steps of one and the gap extension penalty from zero to three in steps of 0.5 also including the default value of 0.2. These ranges were chosen based on previous

studies (Vogt et al. 1995; Gotoh 1996; Higgins et al. 1996; Yuan et al. 1999). CLUSTALW was used instead of CLUSTALX so that the alignments could be automated through use of a Perl script. Q-scores were then calculated for each alignment and the alignment with the highest Q-score was considered to be made with the optimal combination of gap penalties. This alignment was used for all further analyses. The amino acid alignments were converted into nucleotide alignments, for phylogeny creation, utilizing the proteins' corresponding nucleotide sequences from GenBank with the program protal2dna2.0 (Letondal and Schuerer). The nucleotide alignment was then converted to nexus format with the ReadSeq2.93 (Gilbert 1999) program for phylogenetic analysis.

MrModeltest2.2 (Nylander 2004) was used in conjunction with PAUP4.0b10 (Swofford 2002) to determine the best nucleotide substitution model for each cluster. The model chosen by the Akaike Information Criterion measure in MrModeltest was implemented in MrBayes3.1.1 (Ronquist and Huelsenbeck 2003) for each cluster. All priors were uninformative and set at default values. Each analysis was run for 1000000 generations, sampling every 100^{th} generation for a total of 10001 samples. A burn-in value, the number of initial samples removed from analysis, of 3000 was chosen based on previous analyses. The generation versus log probability plots were examined to ensure convergence was reached and that a burn-in of 3000 was appropriate. The potential scale reduction factor was also used as a measure of convergence (Gelman and Rubin 1992).

Identification of Selection Pressures

Values of ω were estimated for each non-ambiguous codon in the alignment using the codeml program contained in the PAML3.15 package (Yang 1997). Codon site models M0, M3, M1a, M2a, M7 and M8 that estimate ω, were implemented for each cluster (Yang et al. 1998; Yang et al. 2000a; Wong et al. 2004; Yang et al. 2005). Model M0 allows only one category of ω for all sites. Model M3 allowed three unconstrained ω categories, ω_1, ω_2 and ω_3 with proportions p_1, p_2 and $p_3 = 1-p_1- p_2$. Model M1a contains two categories of ω, $0 < \omega_0 < 1$ and $\omega_1 = 1$ with proportions p_0 and $p_1 = 1-p_0$. Model M2a adds a third category, $\omega_s > 1$ with proportion p_s such that $p_s = 1-p_0-p_1$. Models M7 and M8 both contain 10 equal proportion ω categories approximated from ß(p,q) with $0 < \omega < 1$ while Model M8 adds an additional ω category, $\omega_s > 1$. The proportion of sites with $\omega \sim$ ß(p,q) is represented by p_0 and those with $\omega_s > 1$ are represented by p_s where $p_s = 1-p_0$. Each site is assigned to an ω category using a naïve empirical Bayes (NEB) (models M0, M3, M1a and M7) (Nielsen and Yang 1998) or Bayes empirical Bayes (BEB) (models M2a and M8) (Yang et al. 2005) approach.

Codon frequencies were set as free parameters (CodonFreq = 3) and ambiguous columns in the alignment were removed from the analysis. The transition/transversion ratio and branch lengths were estimated from the data using maximum likelihood methods. Two separate analyses were conducted with initial values of 0.4 and 2.0 for ω to identify and avoid local optima (Anisimova et al. 2002; Wong et al. 2004). Each analysis was repeated once. Comparison of the results for each model using initial $\omega = 0.4$ and $\omega = 2$ and their repeats

revealed that parameter estimates (ln likelihood, p, ω and β(p,q)) for each model were identical when rounded to three decimal places. The accuracy and power of selection analysis are good if different models are tested, initial values of ω are varied and the analysis is consistent when repeated (Wong et al. 2004). Tree topology does not affect the prediction of selection pressures (Nielsen and Yang 1998; Yang 2000; Swanson et al. 2001; Scheffler and Seoighe 2005) therefore different topologies were not tested.

A likelihood ratio test (LRT) comparing M0 and M3 using a χ^2 distribution with four degrees of freedom was used as a test for variation in ω among sites (Yang et al. 2000b; Anisimova et al. 2001). If M3 fits the data better than M0 then there is evidence that three categories of ω better describe the selection pressures than one ω category. Two LRTs were used as a test for positive selection, M1a against M2a and M7 against M8, each using a χ^2 distribution with two degrees of freedom (Nielsen and Yang 1998; Yang et al. 2000a). M2a and M8 differ from their corresponding null models in that they both contain an additional category of $\omega > 1$, a positive selection category. If M2a or M8 fit the data better than M1a or M7 respectively, then there is evidence that a category of $\omega > 1$ is required to describe the data and this is evidence for positive selection. The LRTs were considered significant when the P-value was ≤ 0.05. The critical values are 9.49 and 5.99 for four and two degrees of freedom respectively when $P = 0.05$. A correction for multiple tests was not performed as the two LRTs for positive selection test the fit of different distributions of ω to the data and are therefore performed for robustness (Yang 2006).

If positive selection occurs in only a few lineages in a tree, it may not be identified using site models, therefore branch-site model A, which allows for $\omega > 1$ along a specified lineage, the foreground branch, while ω cannot be greater than one in any of the other lineages, the background branches (Zhang et al. 2005) was applied. This model was implemented for lineages leading to parologous clades in the FoxA, FoxD, FoxO and FoxP clusters as positive selection is a potential evolutionary force driving subfamily paralog functional differentiation. The FoxI cluster was not examined as no lineages of interest were identified. Model A contains four classes of sites; class 0: $0 < \omega_0 < 1$ and class 1: $\omega_1 = 1$, with proportions p_0 and p_1 respectively, for both the foreground and background branches and class 2a or 2b: $\omega_2 \geq 1$ for the foreground branch with corresponding sites in the background lineage falling into class 2a: $0 < \omega_0 < 1$ or class 2b: $\omega_1 = 1$ site classes with proportions $(1-p_0-p_1)p_0/(p_0+p_1)$ and $(1-p_0-p_1)p_1/(p_0+p_1)$ respectively. All other parameters and running conditions were set as described for the site models. Model A is compared to a null model A with $\omega_2 = 1$ fixed using a LRT and χ^2 distribution with one degree of freedom. Statistical significance at $\alpha = 0.05$ was determined after correction for multiple tests using Rom's procedure and the Bonferroni correction when multiple branches were tested in a phylogeny (Anisimova and Yang 2007). If significance was obtained through Rom's procedure but not the more stringent Bonferroni correction, the LRT was referred to as potentially positive. BEB is used to identify sites under positive selection if the LRT is significant and $\omega_2 > 1$.

Identification of EH1 Motifs

The engrailed homology 1 (EH1) motif has previously been identified in many, but not all of the sequences included in this analysis (Copley 2005; Yaklichkin et al. 2007). Visual examination of the sequence alignments in conjunction with known EH1 locations suggested that there were EH1 motifs present in the sequences included here that have not been previously reported. A Perl script was written to search all of the sequences included in this analysis for the EH1 motif of the form XXaXbXXcdXX where X can be any amino acid, a can be Phe, His, Tyr or Trp, b and c can be Ile, Leu or Val and d can be Glu, Phe, His, Ile, Lys, Met, Gln, Arg, Trp or Tyr (Smith and Jaynes 1996; Copley 2005). Sequences with newly identified EH1 motifs are indicated in Table 2-1 and the locations of the motifs can be found in Figures 2-1 to 2-5.

Results

Site Analysis

Codon site models M0, M1a, M2a, M3, M7 and M8 were implemented in codeml for each of the six clusters and compared using likelihood ratio tests (LRTs). For each cluster the M3 vs. M0 LRT was significant ($P < 0.0001$) (Table 2-2), indicating that one category of ω was insufficient to describe the variability in selection pressure across amino acid sites. LRTs testing for positive selection, M2a vs. M1a and M8 vs. M7, were insignificant for each cluster (LRT = 0, P = 1 for both tests for all clusters), therefore the amino acid changes within each cluster are neutral or under negative selection. Table 2-3 reports the parameter estimates for the least parameter rich model, M1a, which best describes the variation in selection pressures across sites. The posterior weighted ω, the mean of ω over the site classes weighted by the posterior probability of each class, as estimated by M1a, for each residue analyzed is shown graphically for each cluster (Figure 2-6).

Branch-Site Analysis

Figure 2-7 shows the branches that were tested for positive selection in each of the gene cluster phylogenies. The phylogenies are gene trees and may have topologies that are different from species trees. This is because the time to the last common ancestor of two DNA sequences is often different (generally longer) than the time to the last common ancestor of two species (reviewed by:

Nichols 2001). Additionally, many of the forkhead paralogs included in the analyses were present in multiple species; which means that they were likely formed in an ancestral species. This results in orthologs grouping together and the formation of clades that are parologous to one another. Therefore the branch-site results represent selection pressures within the forkhead subfamilies and are not inferable to the species as a whole. LRTs (Table 2-4) were significant for branches leading to the FoxA3 and Protostomia/Cephalochordata clades in the FoxA cluster and the FoxD2 lineage in the FoxD cluster and potentially significant for the FoxD1/2/4 lineage in the FoxD cluster and the FoxO3 lineage in the FoxO cluster. These results suggest that positive selection has acted in the diversification of these paralogs from other genes in the cluster. Model A parameter estimates for lineages under positive selection are given in Table 2-5. Positive selection was not identified in any of the other lineages tested.

The most C-terminal site under positive selection in the FoxA Protostomia/Cephalochordata lineage (e.g. S222 of the fkh_pvul sequence in Figure 2-1) is serine for all of the sequences analyzed. Although there are no amino acid differences at this site in the sequences included in the analysis, due to the codons used among different species and the nucleotide substitution rate model of codeml, the ancestral state of this site for the Protostomia/Cephalochordata lineage has a higher probability of not being serine than of being serine. The codons used among all of the sequences are: TCC, TCT, TCG, AGC and AGT. The AGC and AGT codons are used by four of the six Protostomia/Cephalochordata sequences and are not present in the Craniata

sequences. This means that there is a higher probability assigned to TCB (B is not A) codons occurring outside of the Protostomia/Cephalochordata clade and AGY (Y is C or T) codons occurring in the Protostomia/Cephalochordata clade. When changing from the unknown ancestral sequences to TCB and AGY, there are more opportunities for nonsynonymous substitutions than synonymous substitutions, therefore the site is predicted to be under positive selection. This occurs because the model of nucleotide substitution rate used in the codeml program sets the instantaneous substitution rate to zero for codons that differ at more than one nucleotide. Therefore, multiple single nucleotide changes from the ancestral state to TCB or AGY are assigned higher probabilities than two or three nucleotide changes and these multiple steps provide more opportunities for nonsynonymous substitutions than synonymous substitutions. This also occurs because six different codons can code for serine and this allows for codons that differ at two and three sites to still code for serine. No other codons that differ at three sites can code for the same amino acid and only codons for leucine and arginine can differ at two positions and still code for the same amino acid. The majority of codons can only differ at one site and still code for the same amino acid.

In the FoxD2 clade one positively selected site occurs between the forkhead domain and the EH1 motif in a region that has not been functionally characterized while the remaining positively selected sites identified in this lineage and that identified in the FoxD1/2/4 lineage occur within the EH1 motif as identified in the FoxD1, FoxD3 and FoxD5 sequences (Figure 2-2). The LRT for

the FoxD1/2/4 branch was potentially significant (i.e. insignificant with the Bonferroni correction but significant with the Rom's procedure correction), the amino acid residues at the positively selected site identified in the FoxD1/2/4 lineage differ only in the FoxD2 lineage and are otherwise 100 percent conserved in the other sequences analyzed, therefore it is unlikely that positive selection acted along the FoxD1/2/4 lineage. The FoxD2 lineage sequences contain an EH1 motif however it was not aligned with that identified in the FoxD1, FoxD3 or FoxD5 sequences due to additional amino acids, some of which were under positive selection, found in the FoxD2 lineage. It is likely that the positive selection identified in the FoxD2 lineage within this region is due to the high conservation of the EH1 motif in the other sequences analyzed and lack of motif alignment and not due to evolutionary forces.

Discussion

Prediction of Functional and Nonfunctional Residues Using Site Analysis

The site methods described here may be used to predict functionally important residues in gene family members. If a functional domain has been identified in one member of a gene family, but not in a different member and the functional domain is under negative selection, prediction of a similarly functioning domain may be made in the family member where a domain has not been identified. In support of this theory, the forkhead domain, which is most likely functionally active in all of the sequences analyzed, was under negative selection in each cluster. Functional domains were also predicted in the FoxA, FoxO and FoxP cluster sequences.

In the FoxA cluster, conserved domain II has been shown to be involved in transactivation (Pani et al. 1992) and repression (Wang et al. 2000) in rat FoxA2. Since conserved domain II is entirely under negative selection (Figure 2-6A) and contained only one ambiguous column in the alignment (Figure 2-1), it is likely functionally important in all of the sequences analyzed. In the FoxO cluster, a transactivation domain has been identified at the C-terminus of FOXO1a and FOXO4 (Sublett et al. 1995; So and Cleary 2002) while a transactivation domain has yet to be identified in FOXO3a. A portion of the C-terminal transactivation domain in FOXO4 and the entire transactivation domain in FOXO1a was under negative selection (Figure 2-6D), therefore a C-terminal transactivation domain consisting of the negatively selected residues (sites 389-

428 in Figure 2-6D, residues 605-673 in FOXO3a) may be predicted in FOXO3a. A second, weaker, transactivation domain was identified in FOXO4 between the forkhead domain and the C-terminal transactivation domain (So and Cleary 2002). This region is not highly conserved, although small islands of consecutive columns without gaps in the alignment that show negative selection, i.e. sites 315-326 in Figure 2-6D, may be functionally important. C-terminal deletions of PAX3-FOXO1a (a fusion protein consisting of the PAX3 N-terminal region, which includes two DNA binding domains, to the C-terminal region of FOXO1a, that includes part of the forkhead domain and the C-terminal transactivation domain) that include residues within FOXO1a corresponding to the FOXO4 transactivation domain have also shown reduced transactivation (Kempf and Vogt 1999; Lam et al. 1999). The residues under negative selection in this region may be key to the transactivation function seen in FOXO1a and FOXO4, and residues of FOXO3a within this region may also show transactivation function. A N-terminal NES and a NLS at the N-terminus of the forkhead domain have been identified in FOXO1a (Zhao et al. 2004) and were found to be under negative selection (Figure 2-6D). These regions have not been examined for NES or NLS function in FOXO3a and FOXO4. The negative selection of these regions suggests that a NES may be found in the N-terminus of the protein and an NLS at the N-terminus of the forkhead domain in all of the sequences analyzed. Similarly, three phosphorylation sites involved in cellular localization have been identified in FOXO1a, Ser322, Ser325 and Ser329 (Woods et al. 2001; Rena et al. 2002). It is unknown if these sites are phosphorylated or involved in cellular

localization of FOXO3a or FOXO4. The Foxo6_mmus sequence was the only sequence that did not contain serines at these three positions (Figure 2-4) suggesting that these serines may be functionally important in the other sequences analyzed with the exception of Foxo6_mmus. Finally, there are three common phosphorylation sites among the human FOXO proteins (sites 20, 157 and 216 in Figure 2-6D) and two 14-3-3 protein binding sites (sites 17-22 and 153-159 in Figure 2-6D) that are important in regulating cytoplasmic/nuclear localization and therefore transactivation activity (Brunet et al. 1999; Kops et al. 1999; Nakae et al. 1999; Rena et al. 1999; Takaishi et al. 1999; Tang et al. 1999; Brownawell et al. 2001; Brunet et al. 2001; Rena et al. 2001; Brunet et al. 2002; Zhang et al. 2002; Mazumdar and Kumar 2003; Obsil et al. 2003; Zhao et al. 2004). These phosphorylation and 14-3-3 binding sites were are all highly conserved among species and under negative selection suggesting functional importance in all of the sequences analyzed. Within the FoxP cluster the leucine zipper and zinc finger identified in FOXP1 and mouse Foxp1, Foxp2 and Foxp4 (Banham et al. 2001; Teufel et al. 2003; Wang et al. 2003; Li et al. 2004) were under negative selection suggesting that they are present and functional in the other sequences analyzed (Figure 2-6E). The leucine zipper allows FoxP proteins to form homo- and hetero-dimers (Wang et al. 2003; Li et al. 2004) and although the zinc finger function has yet to be determined, it has been suggested that it aids in dimer formation (Wang et al. 2003).

Functional domains may also be predicted in regions under negative selection where a domain is not known to exist. For example, functionally

important residues have not been identified in the N-terminus of FOXD proteins and a series of amino acids under negative selection is found in this region (Figure 2-6B). This series of negatively selected amino acids may be functionally important and forms a starting point to identifying functionally important residues outside of the forkhead domain in the FOXD proteins. Predicting functionally important residues with these methods provides a specific region of amino acids and potential domain boundaries that can be tested when searching for functional domains *in vitro*.

When a functional region has been identified in one gene family member, but the majority of the amino acids making up the functional region are aligned with gaps and/or are experiencing neutral changes, the region is likely not functioning in the same manner in the other sequences analyzed. Examples include conserved domains IV and V in the FoxA cluster and the transactivation domain in the FoxI cluster (Figures 2-1, 2-3, 2-6A, C). This method identifies a region of amino acids that are less likely to be important for a specific function, which may then be examined last for functional significance when using *in vitro* methods.

Refining Domain Boundaries Using Site Analysis

Domain boundaries are often identified by sequence comparison to functionally related proteins or through mutagenesis experiments. When comparing sequences, it is assumed that the domain boundaries are accurately

defined in the protein to which the comparison is made. Often the boundaries of a new domain are loosely defined through mutagenesis experiments, as it is too time consuming to examine every amino acid near the suspected boundary for functional contribution. These loosely defined domains are then used by other researchers in sequence comparisons to identify domains in related proteins. The methods used in this paper provide a new *in silico* procedure for identifying domain boundaries as discussed for the forkhead domain below. Molecular analysis is necessary to confirm the reallocation of domain boundaries.

The assigned boundaries of the forkhead domain vary from source to source. The NCBI Conserved Domain Database (CDD) definition of the forkhead domain, which was taken from the SMART (simple modular architecture research tool) database forkhead definition, was used in this paper. In this definition, the boundaries of the forkhead domain are defined by tertiary structure and sequence comparison of all known forkhead domains (Schultz et al. 2000). Since the C-terminal end of the forkhead domain is unstructured and variable among subfamilies (Clark et al. 1993; Marsden et al. 1998; van Dongen et al. 2000; Weigelt et al. 2000; Stroud et al. 2006), this region is excluded from the CDD forkhead domain definition even though it is involved in DNA binding (Clevidence et al. 1993; Pierrou et al. 1994; Shiyanova and Liao 1999; Cirillo and Zaret 2007; Brent et al. 2008). When a new protein containing a forkhead domain is described in the literature, the forkhead domain is often identified through sequence comparison to the rat FoxA1 forkhead domain, the first forkhead domain containing protein identified in mammals (Lai et al. 1990). The

rat FoxA1 forkhead domain was broadly defined through mutational analysis (Lai et al. 1990) and then succinctly defined through sequence comparison to the rat FoxA2, FoxA3 and *Drosophila* Fork Head proteins (Weigel and Jackle 1990; Lai et al. 1991). When a forkhead domain is defined through sequence comparison to rat FoxA1, the N- and C-terminal domain boundaries vary within the gene family and subfamilies while the CDD definition of the forkhead domain is consistent among gene family members. The N- and C-terminal domain boundaries include additional amino acids when defined through sequence comparison to rat FoxA1 as compared to the CDD definition. In this analysis, a series of residues directly adjacent to the N- and C-termini of the forkhead domain in each of the clusters analyzed were under negative selection (Figure 2-6), suggesting that the forkhead domain definition should include these residues. The forkhead domain definitions supplied in the literature often accounted for some of the negatively selected sites not included in the CDD forkhead definition; however, the literature definitions either included sites that were not conserved among species, included sites with neutral changes, did not include all of the sites under negative selection and all varied in their start and stop points within subfamilies. If the N- and C-terminal boundaries of a domain are defined as the first and last residue respectively of a series of residues under negative selection, the results will be reproducible and consistent among gene family or subfamily members.

Identification of Amino Acids Involved in Paralog or Ortholog Differentiation

The branch-site and site analysis of selection pressures on codons conducted here have identified specific amino acids that may be responsible for differentiation of paralogs in the FoxA and FoxO clusters and orthologs in the FoxA cluster. In the FoxA cluster, the region N-terminal to the forkhead domain appears to contribute to paralog differentiation. One positively selected site identified in the FoxA3 clade occurs within conserved domain IV and one positively selected site identified in the Protostomia/Cephalochordata lineage occurs within conserved domain V as both domains are defined in FoxA2 (Qian and Costa 1995) (Figure 2-1). Overall conserved domains IV and V, which have been shown to play a role in transactivation in FoxA2 proteins (Qian and Costa 1995), are not well conserved in the FoxA3 or Protostomia/Cephalochordata proteins as compared to the FoxA1 and FoxA2 proteins as the majority of the residues making up these domains were not analyzed due to gaps in the alignment and those that were examined by site analysis show variability in selection pressure with most of the sites, 5/7, having experienced neutral changes (Figure 2-6A). Additional sites under positive selection N-terminal to the forkhead domain were also identified through branch-site analysis in the FoxA3 and Protostomia/Cephalochordata lineages (Figure 2-1). Two of these sites in the FoxA3 lineage occur in a nuclear localization signal (NLS) that was broadly defined in rat FoxA2 (Qian and Costa 1995) while the other positively selected sites are found in regions uncharacterized in any FoxA protein. FoxA1 and FoxA2 have more similar expression patterns and functions during development

and metabolism as compared to the FoxA3 proteins (reviewed by: Friedman and Kaestner 2006). This evidence in conjunction with the positive selection identified here suggests that the N-terminal region of sequences not included in the FoxA1 or FoxA2 clades have evolved to differentiate these proteins from the FoxA1 and FoxA2 proteins while the sequences were conserved in the FoxA1 and FoxA2 proteins leading to overlapping expression and function.

Conserved domain III, which has been shown to function in transactivation in rat FoxA2 (Pani et al. 1992) contained many ambiguous sites in the FoxA alignment (Figure 2-1) due to sequences from the Protostomia lineage and variations in selection pressure were observed in the four sites that did contain amino acids from these species (Figure 2-6A). This suggests that conserved domain III is important for FoxA function in the Deuterostomia but not in the Protostomia and that the FoxA genes in the two lineages have evolved to perform species specific functions. Therefore the presence of conserved domain III may differentiate FoxA orthologs between the Protostomia and Deuterostomia lineages.

In the FoxO cluster, the NES(s) located between the forkhead domain and the C-terminus in the FOXO1a, FOXO3a and FOXO4 sequences (Biggs et al. 1999; Brownawell et al. 2001; Brunet et al. 2002; Zhao et al. 2004) are not highly conserved among the FoxO family members as their alignment was not well defined, only three sites, 250-252, in Figure 2-6D contain NES residues from each of the three human FOXO proteins examined and some residues have experienced

neutral changes as demonstrated by site analysis. These NES(s) may be used to differentiate FoxO paralogs.

Only one site was found to be under positive selection in the FoxO3 lineage during branch-site analysis and the LRT was potentially significant. This residue is found in a region important for nuclear localization, C-terminal to the forkhead domain (Figure 2-3). The amino acid located at the positively selected site is serine in the FoxO3 sequences while it is glycine, alanine or aspartic acid in the other sequences analyzed. The presence of serine at this position may be important for regulation of the FoxO3 proteins by phosphorylation and this regulation may be different from the other FoxO sequences analyzed. Molecular testing is required to validate this hypothesis.

In summary, residues that differentiate paralogs were identified in the FoxA and FoxO clusters while residues that differentiate orthologs were also identified in the FoxA cluster. This information provided insights into the evolution of these two subfamilies. Within the FoxD, FoxI, and FoxP clusters, residues that differentiate orthologs or paralogs were unidentifiable due to lack of functional information (FoxD and FoxI clusters only) and overall negative selection in the identified domains.

Subfamily Evolution

Forkhead subfamilies are defined by their homology in the forkhead domain alone. Here I analyzed the entire coding regions of forkhead proteins and

found that the subfamily structures were maintained after sequence analysis with BLASTCLUST. The site analysis also demonstrated distinct regions of homology under negative selection outside the forkhead domain in each of the clusters analyzed. Therefore, sequences outside of the forkhead domain contribute to subfamily evolutionary relationships. These results show that after gene duplication, selective restraints were relaxed enough to allow for subfamily differentiation. Negative selection then acted on these amino acid changes to maintain the differentiated subfamilies. Within subfamilies, after gene duplication selective restraints are relaxed on some amino acids and constrained on others. This maintains the subfamily relationship while allowing for paralog differentiation. While the majority of studies that have used these methods focus only on positive selection, a few involving transcription factor gene families have discussed negative selection as well. My results are similar to those seen in a comparable analysis of *HOX7* where heterogeneous selection pressures but not positive selection were observed during site analysis and positive selection was observed on a single branch separating paralogs during branch-site analysis (Fares et al. 2003). These types of analyses of gene families that were originally defined by a common functional motif may confirm or refute the family relationships and provide insights into their evolutionary development.

Forkhead Domain Evolution

As forkhead subfamilies are defined by and forkhead gene function is reliant on the forkhead domain, identification of selection pressures acting on

codons within the domain provides insights into the functional evolution of subfamilies and their paralogs. In each of the subfamilies, the majority of the residues in the forkhead domain were under negative selection (Figure 2-6) consistent with the general consensus that the domain is highly conserved and important for proper gene function. More interestingly, sites under positive selection and neutral changes were observed in the forkhead domain in some subfamilies and these provide insights into the evolutionary differentiation of forkhead genes.

In the FoxA cluster Protostomia/Cephalochordata lineage a number of residues under positive selection were found in the forkhead domain through branch-site analysis. These residues are located within helix 2, β-sheet 2 and wing 1 as defined by the crystal structure of FoxA3 (Clark et al. 1993) (Figure 2-1, Figure 2-8). The residues corresponding to the positively selected sites in the Protostomia/Cephalochordata lineage are almost 100% conserved among the other FoxA sequences analyzed, only the FoxA2_rnor sequence differs at one site. The residues corresponding to the positively selected sites in disease causing forkhead genes are not known to be mutated in a disease state and are not associated with SNPs (single nucleotide polymorphism) (NCBI Entrez SNP database (Wheeler et al. 2006), Build 126). Therefore, it is possible that these positively selected changes in amino acid composition of the forkhead domain alter the domain to allow for different target binding and/or regulation of FoxA genes in the Protostomia/Cephalochordata as compared to the Craniata. It is interesting to note that to date, in non-Craniata Eumetazoa only one or two FoxA

class genes are normally found in a species while in the Craniata, the presence of three FoxA class genes is common (Shimeld 1997; Odenthal and Nusslein-Volhard 1998; Adell and Muller 2004; Magie et al. 2005; Tu et al. 2006; Hansen et al. 2007; Larroux et al. 2008). If FoxA targets are similar among all lineages, the alterations in the forkhead domain of Protostomia/Cephalochordata FoxA may allow these single proteins to perform the same function that require multiple FoxA proteins in the Craniata. This theory is further supported by the differences observed in the N-terminal region of the Protostomia/Cephalochordata FoxA and in conserved domain III as compared to the Craniata discussed earlier. Alternatively, the alterations in the forkhead domain may allow the Craniata FoxAs to perform functions (i.e. bind different co-regulators) that are not performed by the Protostomia/Cephalochordata.

The FoxA Protostomia/Cephalochordata positive site in helix 2 is tryptophan in the Craniata and phenylalanine in the Protostomia/Cephalochordata. In rat FoxA3 this site is part of a hydrophobic surface patch that is exposed to solvent when the forkhead domain is bound to DNA (Clark et al. 1993). These hydrophobic residues are highly conserved among other forkhead domain members. Since phenylalanine is hydrophobic and the remaining residues that make up the patch are identical, the patch is likely not disrupted in the Protostomia/Cephalochordata. However any specific intermolecular interactions involving the patch have the potential to be altered. To date no interactions or modification of this patch have been identified in any forkhead protein.

The positively selected residues in β-sheet 2 and wing 1 are within the C-terminal nuclear localization signal (NLS) of the forkhead domain. This region does not conform to a specific NLS pattern but does contain a number of basic amino acids that are likely required for localization (Qian and Costa 1995). Only one of the positively selected sites is a basic amino acid, e.g. R217 in the fkh_pvul sequence in Figure 2-1. This site is lysine (K) in all of the Craniata and arginine (R) or lysine in the Protostomia/Cephalochordata. Therefore the charge of this residue is conserved and nuclear localization is likely not disrupted.

The positively selected site within wing 1 of the forkhead domain is serine in all of the sequences analyzed. As noted in the results, this site was identified as being under positive selection due to codon changes even though there is no amino acid change among the lineages. This site is conserved and known to make DNA contact in rat FoxA3, human FOXO1a and FOXK1a (Clark et al. 1993; Tsai et al. 2006; Brent et al. 2008). It appears that selection pressures have acted to keep serine at this site in all of the sequences analyzed. Overall the positively selected sites in the forkhead domain of the Protostomia/Cephalochordata lineage have similar physical properties to those in the Craniata (i.e. comparing phenylalanine to tryptophan, tyrosine to serine, lysine to arginine). This indicates that these properties are important for domain function in both lineages.

One residue within the forkhead domain was experiencing neutral changes in the FoxA, FoxD and FoxP clusters (Figures 2-6A (site 41), B (site 74), E (site 451)). The locations of the residues with neutral changes are shown on the

FoxA3 crystal structure in Figure 2-8. The sites experiencing neutral changes identified in the FoxA and FoxP clusters were found at the C-terminus of alpha helix 1 while the site experiencing neutral changes in the FoxD cluster was located near the C-terminus of alpha helix 2. Neutral changes at a site imply that any amino acid may be present at that site and amino acid changes will not affect protein function. In support of this theory, mutation of the site corresponding to the neutral site identified in the FoxD cluster in rat FoxA3 from aspartate to lysine did not affect DNA binding (Clevidence et al. 1993). The sites with neutral changes identified in the FoxA, FoxD and FoxP clusters and the corresponding sites in other Fox proteins have not been associated with point mutations causing human disease and have not been shown to contact DNA during DNA binding. The NCBI Entrez SNP database (Wheeler et al. 2006), Build 126, was initially used to determine if the sites with neutral changes have naturally occurring single nucleotide polymorphisms in any of the forkhead genes found in humans. Only one forkhead gene, FOXD4, had a known SNP at a location corresponding to one of the sites with neutral changes. The SNP identified in FOXD4 corresponds to the neutrally changed site identified in the FoxD proteins and is either aspartate or glycine. A more recent search of the SNP database Build 131 after publication of these results identified an asparagine/serine SNP at the FoxD neutral site in FOXD3. It would be interesting to determine if amino acid changes at these sites affect forkhead domain function and if the neutrally changed sites are common to the forkhead domain or specific to the subfamilies in which they were identified.

The variations from negative selection in the forkhead domain identified here may account for differences in subfamily and paralog function that are not explained by differences in timing or location of expression or other functional regions in the proteins.

Conclusions

This analysis has provided insights into forkhead gene family and subfamily evolution. Through identification of selection pressures the functional and evolutionary importance of amino acid differences in paralogs and orthologs of subfamilies has been predicted. This work has also supported the forkhead subfamily structure and identified a pattern of evolution in the family. Additionally, these analyses allowed evaluation and extension of domain structural and positional information between gene family members. Future *in vitro* studies may use this information as a starting point or for refinement of protein functional analyses.

Tables

Table 2-1. Composition of the sequence clusters analyzed.

The table shows the composition of the five clusters of sequences with 30% identity over 90% of their length that were analyzed. All sequences, excluding those indicated by *, also grouped in the given clusters at 40% identity over 90% of their length. Sequences in which EH1 motifs were newly identified are indicated by Φ. Protein and nucleotide accession numbers are from the NCBI Entrez Protein and Nucleotide databases respectively.

FoxA Cluster of 31 sequences

Species	Sequence Identifier	Protein Accession #	Nucleotide Accession #
Achaearanea tepidariorum	At.fkh_atep Φ	BAC24088	AB096073.1
Ambystoma mexicanum	FoxA4_amex	AAC60128	U43547.1
Bombyx mori	SGF1_bmor	Q17241	D38514
Branchiostoma floridae	AmHNF31_bflo	CAA65368	X96519.1
Branchiostoma floridae	HNF3_bflo Φ	CAA70438	Y09236.1
Colisa lalia	FoxA2_clal Φ	BAB21570	AB050937.1
Danio rerio	FoxA2_drer	NP_571024	NM_130949.1
Danio rerio	FoxA3_drer	NP_571374	NM_131299.1
Gallus gallus	FoxA2_ggal	NP_990101	NM_204770.1
Homo sapiens	FOXA1_hsap	NP_004487	NM_004496.2
Homo sapiens	FOXA2_hsap	NP_068556	NM_021784.3

Homo sapiens	FOXA3_hsap	NP_004488	NM_004497.2
Mus musculus	Foxa1_mmus	NP_032285	NM_008259.1
Mus musculus	Foxa2_mmus	NP_034576	NM_010446.1
Mus musculus	Foxa3_mmus	NP_032286	NM_008260.1
Oreochromis mossambicus	HNF3B_omos [Φ]	AAL68498	AF251499.1
Oryzias latipes	FoxA2_olat [Φ]	O42097	AB001572
Oryzias latipes	FoxA3_olat [Φ]	BAA23580	AB001573.2
Patella vulgata	fkh_pvul	CAD45552	AJ507424.1
Rattus norvegicus	FoxA1_rnor	NP_036874	NM_012742.1
Rattus norvegicus	FoxA2_rnor	NP_036875	NM_012743.1
Rattus norvegicus	FoxA3_rnor	NP_058773	NM_017077.1
Tetraodon nigroviridis	UN_3_tnig [Φ]	CAF89623	CAAE01007089
Tetraodon nigroviridis	UN_45_tnig [Φ]	CAG09884	CAAE01015009
Tetraodon nigroviridis	UN_51_tnig [Φ]	CAG12727	CAAE01015113
Tribolium castaneum	Tcfkh_tcas	AAF71998	AF217810.1
Xenopus laevis	FoxA1b_xlae	P32315	M93658
Xenopus laevis	FoxA4a_xlae	P33205	X65171
Xenopus laevis	FoxA4b_xlae	P33206	S93559
Xenopus tropicalis	FoxA1_xtro [Φ]	NP_989419	NM_204088.1
Xenopus tropicalis	FoxA2_xtro [Φ]	NP_989423	NM_204092.1

FoxD Cluster of 24 sequences			
Species	**Sequence Identifier**	**Protein Accession #**	**Nucleotide Accession #**
Danio rerio	FoxD3_drer	NP_571365	NM_131290.1
Danio rerio	FoxD5_drer	NP_571345	NM_131270.1
Gallus gallus	FoxD1_ggal	NP_990523	NM_205192.1
Gallus gallus	FoxD2_ggal	NP_990283	NM_204952.1
Gallus gallus	FoxD3_ggal	NP_990282	NM_204951.1
Gorilla gorilla	FoxD4_ggor* [Φ]	AAQ72340	AY345862.1
Homo sapiens	FOXD1_hsap	NP_004463	NM_004472.1
Homo sapiens	FOXD2_hsap	NP_004465	NM_004474.2
Homo sapiens	FOXD3_hsap	NP_036315	NM_012183.1
Homo sapiens	FOXD4b_hsap* [Φ]	NP_954714	NM_199244.1
Homo sapiens	FOXD4L2_hsap* [Φ]	NP_954586	NM_199135.1
Homo sapiens	FOXD4L3_hsap* [Φ]	NP_955390	NM_199358.1
Mus musculus	Foxd1_mmus	NP_032268	NM_008242.1
Mus musculus	Foxd2_mmus	NP_032619	NM_008593.1
Mus musculus	Foxd3_mmus	NP_034555	NM_010425.2
Oreochromis mossambicus	FoxD5_omos [Φ]	AAM75747	AF251498.1
Pan troglodytes	FoxD4_ptro* [Φ]	NP_001009014	NM_001009014
Tetraodon nigroviridis	UN_48_tnig [Φ]	CAG11584	CAAE01015039
Xenopus laevis	FoxD2_xlae	CAC69867	AJ344435.1
Xenopus laevis	FoxD3b_xlae	CAC12895	AJ298866.1

Xenopus laevis	xfd12_xlae $^{\Phi}$	CAB44728	AJ242676.1
Xenopus laevis	xfd12dblprime_xlae $^{\Phi}$	CAB44730	AJ242678.1
Xenopus laevis	xfd12prime_xlae	CAB44729	AJ242677.1
Xenopus laevis	xfd6_xlae $^{\Phi}$	BAA36334	AB014611.1

FoxI Cluster of 10 sequences

Species	Sequence Identifier	Protein Accession #	Nucleotide Accession #
Danio rerio	FoxI1_drer	NP_859424	NM_181735.1
Danio rerio	FoxI2_drer	NP_944598	NM_198916.1
Danio rerio	FoxI3a_drer	NP_944599	NM_198917.1
Danio rerio	FoxI3b_drer	NP_944600	NM_198918.1
Homo sapiens	FOXI1_hsap	NP_036320	NM_012188.3
Mus musculus	Foxi1_mmus	NP_076396	NM_023907.2
Mus musculus	Foxi2_mmus	NP_899016	NM_183193.1
Tetraodon nigroviridis	UN_46_tnig	CAG10122	CAAE01015015
Xenopus laevis	FoxI1_xlae	AAH42303	BC042303.1
Xenopus laevis	FoxI1c_xlae	CAD31849	AJ487620.1

FoxO Cluster of 12 sequences

Species	Sequence Identifier	Protein Accession #	Nucleotide Accession #
Danio rerio	FoxO5_drer	NP_571160	NM_131085.1
Homo sapiens	FOXO1a_hsap	NP_002006	NM_002015.2
Homo sapiens	FOXO3a_hsap	NP_001446	NM_001455.2

Homo sapiens	FOXO4_hsap	NP_005929	NM_005938.1
Mus musculus	Foxo1_mmus	NP_062713	NM_019739.2
Mus musculus	Foxo3_mmus	NP_062714	NM_019740.1
Mus musculus	Foxo4_mmus	NP_061259	NM_018789.1
Mus musculus	Foxo6_mmus	NP_918949	NM_194060.1
Spermophilus tridecemlineatus	FoxO1a_stri	AAO72710	AY255525.1
Sus scrofa	FoxO1a_sscr	NP_999179	NM_214014.1
Tetraodon nigroviridis	UN_53_tnig*	CAG13202	CAAE01015123
Xiphophorus maculatus	FoxO5_xmac Φ	AAK74186	AY040320.1

FoxP Cluster of 10 sequences

Species	Sequence Identifier	Protein Accession #	Nucleotide Accession #
Gorilla gorilla	FoxP2_ggor Φ	AAN03386	AF512948.1
Homo sapiens	FOXP1_hsap Φ	NP_116071	NM_032682.4
Homo sapiens	FOXP2_hsap Φ	NP_055306	NM_014491.1
Homo sapiens	FOXP4_hsap	NP_612466	NM_138457.2
Macaca mulatta	FoxP2_mmul Φ	AAN03388	AF512950.1
Mus musculus	Foxp1_mmus Φ	NP_444432	NM_053202.1
Mus musculus	Foxp2_mmus Φ	NP_444472	NM_053242.3
Pan troglodytes	FoxP2_ptro Φ	Q8MJA0	AY143178
Pongo pygmaeus	FoxP2_ppyg Φ	AAN03387	AF512949.1
Taeniopygia guttata	FoxP2_tgut Φ	AAR28756	AY395709.1

Table 2-2. Site analysis M3 vs. M0 LRT results for each cluster.

Statistically significant results at $\alpha = 0.05$ are in boldface.

	M3 vs. M0 LRT	
Cluster	**$2(\ln_{M3}-\ln_{M0})$**	**P-value**
FoxA	1446.504	**< 0.0001**
FoxD	1252.774	**< 0.0001**
FoxI	649.137	**< 0.0001**
FoxO	687.901	**< 0.0001**
FoxP	135.938	**< 0.0001**

Table 2-3. Parameter estimates of site model M1a for each cluster.

Cluster	**Parameter Estimates**
FoxA	$\omega_0 = 0.024$ $\omega_1 = 1$ $p_0 = 0.682$ $p_1 = 0.318$
FoxD	$\omega_0 = 0.025$ $\omega_1 = 1$ $p_0 = 0.555$ $p_1 = 0.445$
FoxI	$\omega_0 = 0.039$ $\omega_1 = 1$ $p_0 = 0.832$ $p_1 = 0.168$
FoxO	$\omega_0 = 0.042$ $\omega_1 = 1$ $p_0 = 0.839$ $p_1 = 0.161$
FoxP	$\omega_0 = 0.019$ $\omega_1 = 1$ $p_0 = 0.970$ $p_1 = 0.030$

Table 2-4. Statistical significance of the branch-site analysis LRTs after multiple corrections using Rom's procedure and the Bonferroni correction.

Critical values of statistically significant results are in boldface.

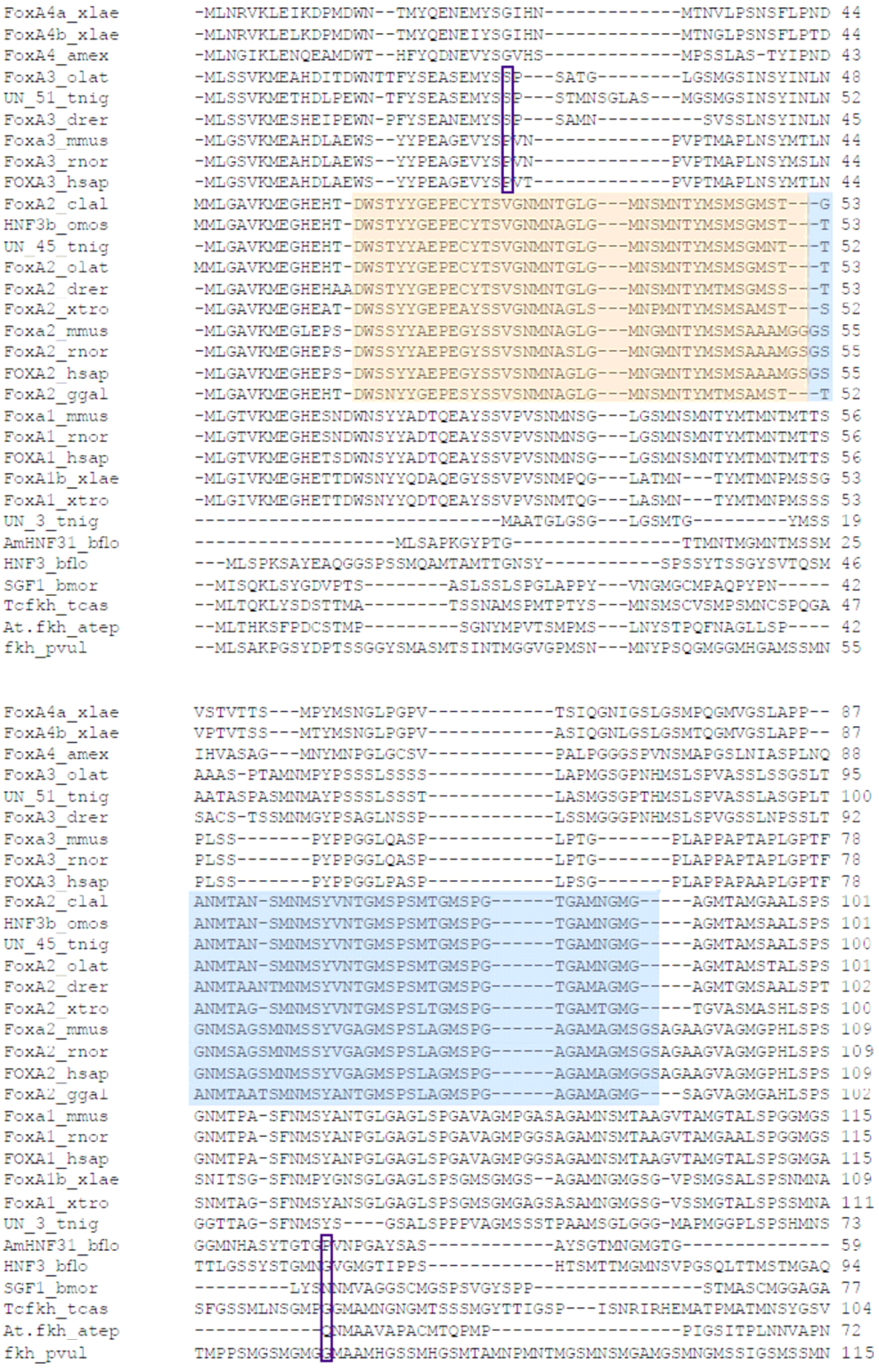
Cluster	Lineage Tested	P-value of LRT	Rom's Procedure critical value	Bonferroni critical value
FoxA	FoxA1	0.62	0.0127	0.0125
	FoxA2	0.0634	0.0169	0.0125
	FoxA3	0.0032	**0.025**	**0.0125**
	Protostomia/ Cephalochordata	< 0.0001	**0.05**	**0.0125**
FoxD	FoxD1	1	0.0102	0.01
	FoxD4	0.0973	0.0127	0.01
	FoxD1/2	0.0347	0.0169	0.01
	FoxD1/2/4	0.0234	**0.025**	0.01
	FoxD2	< 0.0001	**0.05**	**0.01**
FoxO	FoxO4	0.2628	0.0169	0.0167
	FoxO1	0.109	0.025	0.0167
	FoxO3	0.0177	**0.05**	0.0167
FoxP	FoxP2	1	0.05	0.05

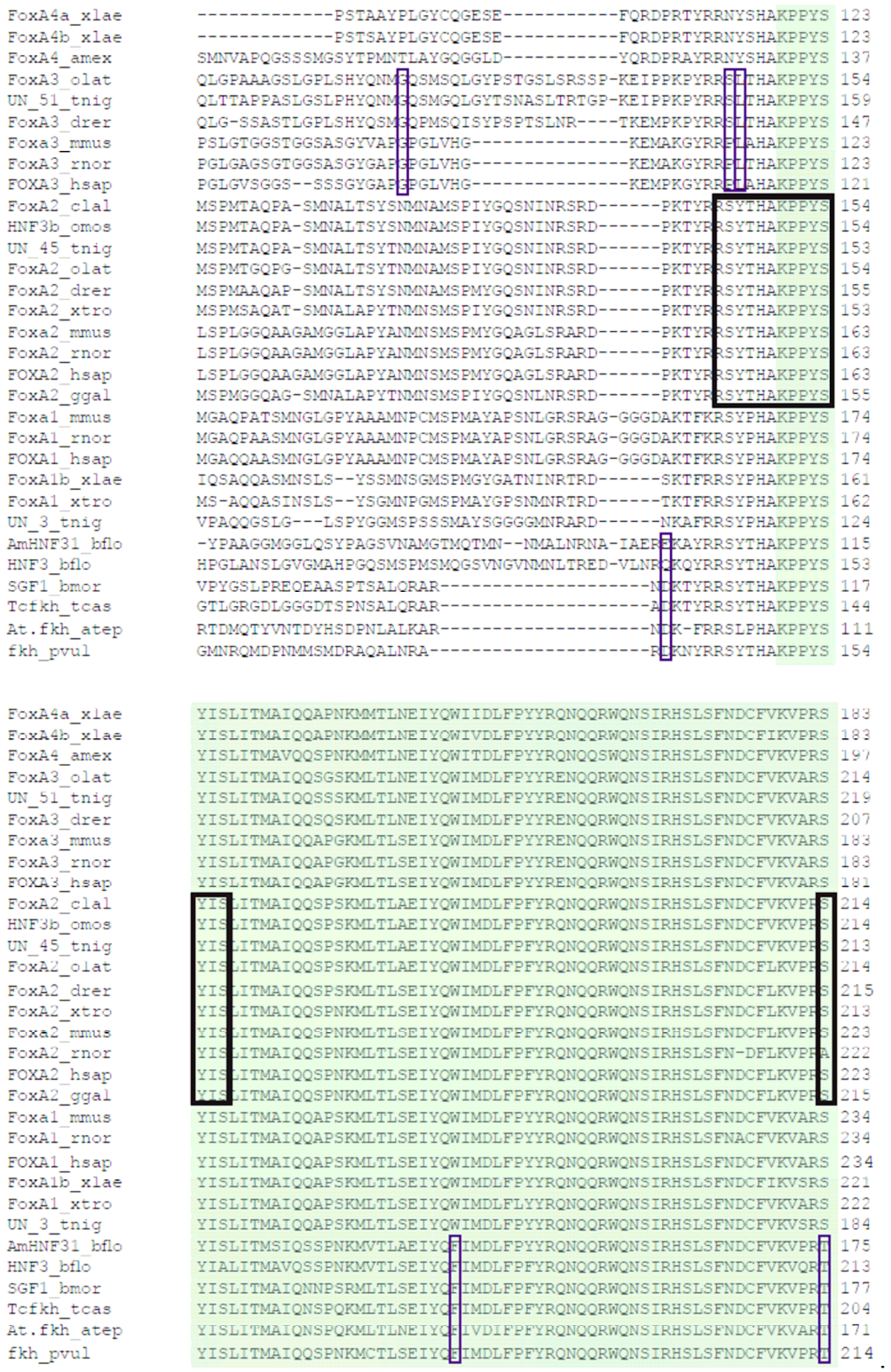
Table 2-5. Model A parameter estimates for significant branch-site LRTs.

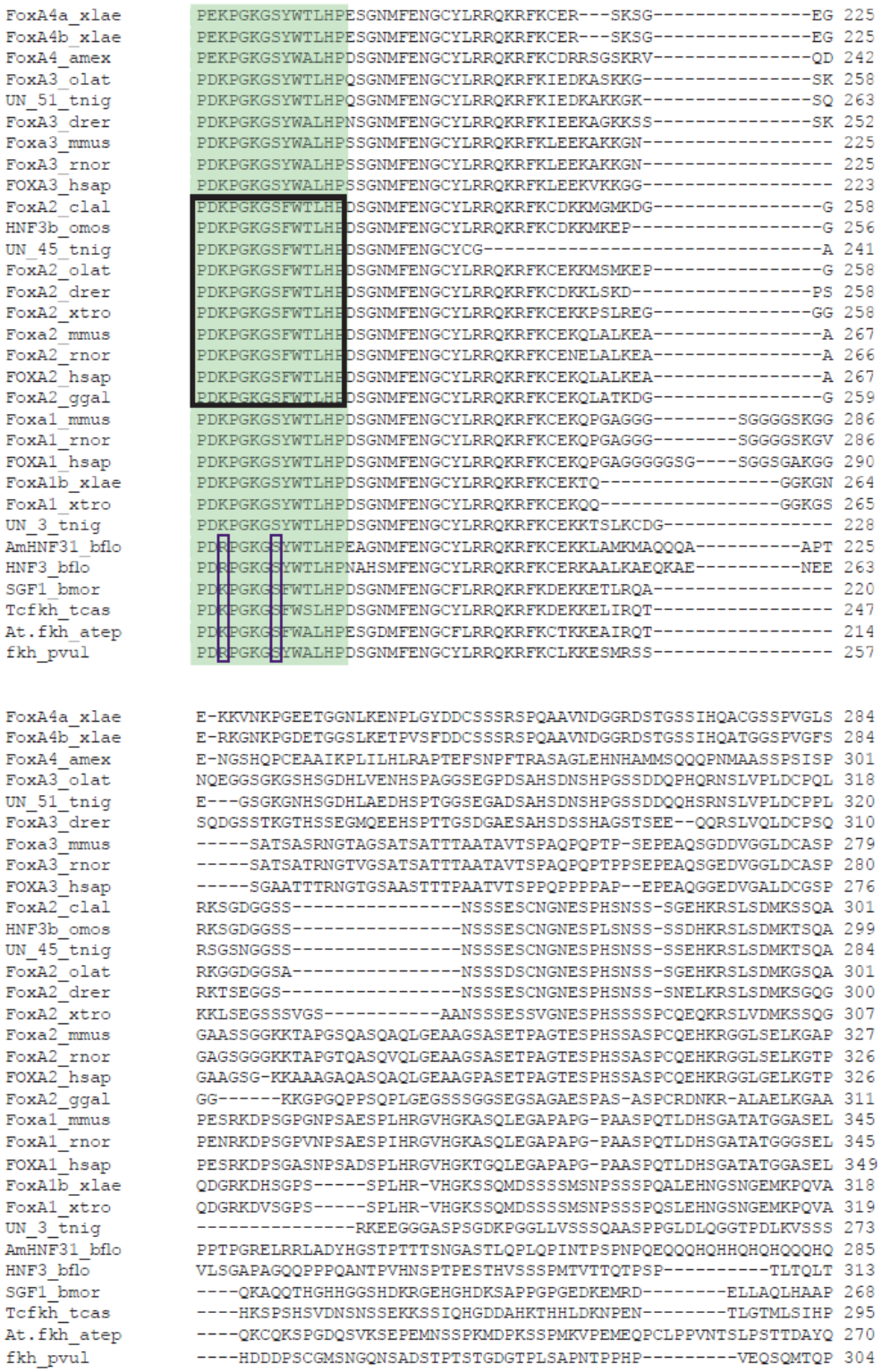
Cluster Lineage	Site Class	Proportion	Backgound ω	Foreground ω	Positively Selected Sites* (P ≥ 0.95)
FoxA					
FoxA3	0	0.640	0.02356	0.02356	FOXA3_hsap 27P, 96G, 112P, 113L
	1	0.300	1	1	
	2a	0.041	0.02356	999	
	2b	0.019	1	999	
Protostomia/ Cephalochordata	0	0.627	0.02362	0.02362	FOXA1_hsap 68Y, 159A, 199W, 234S, 237K, 242S
	1	0.293	1	1	
	2a	0.054	0.02362	999	
	2b	0.026	1	999	
FoxD					
FoxD1/2/4	0	0.537	0.02506	0.02506	FOXD2_hsap R388
	1	0.419	1	1	
	2a	0.025	0.02506	71.85587	
	2b	0.019	1	71.85587	
FoxD2	0	0.517	0.02526	0.02526	FOXD2_hsap E240, T384, L387, R388, Q389, G390, L391, K392, T393
	1	0.368	1	1	
	2a	0.067	0.02526	999	
	2b	0.048	1	999	
FoxO					
FoxO3	0	0.805	0.04277	0.04277	FOXO3_hsap S280
	1	0.144	1	1	
	2a	0.043	0.04277	10.7599	
	2b	0.008	1	10.7599	

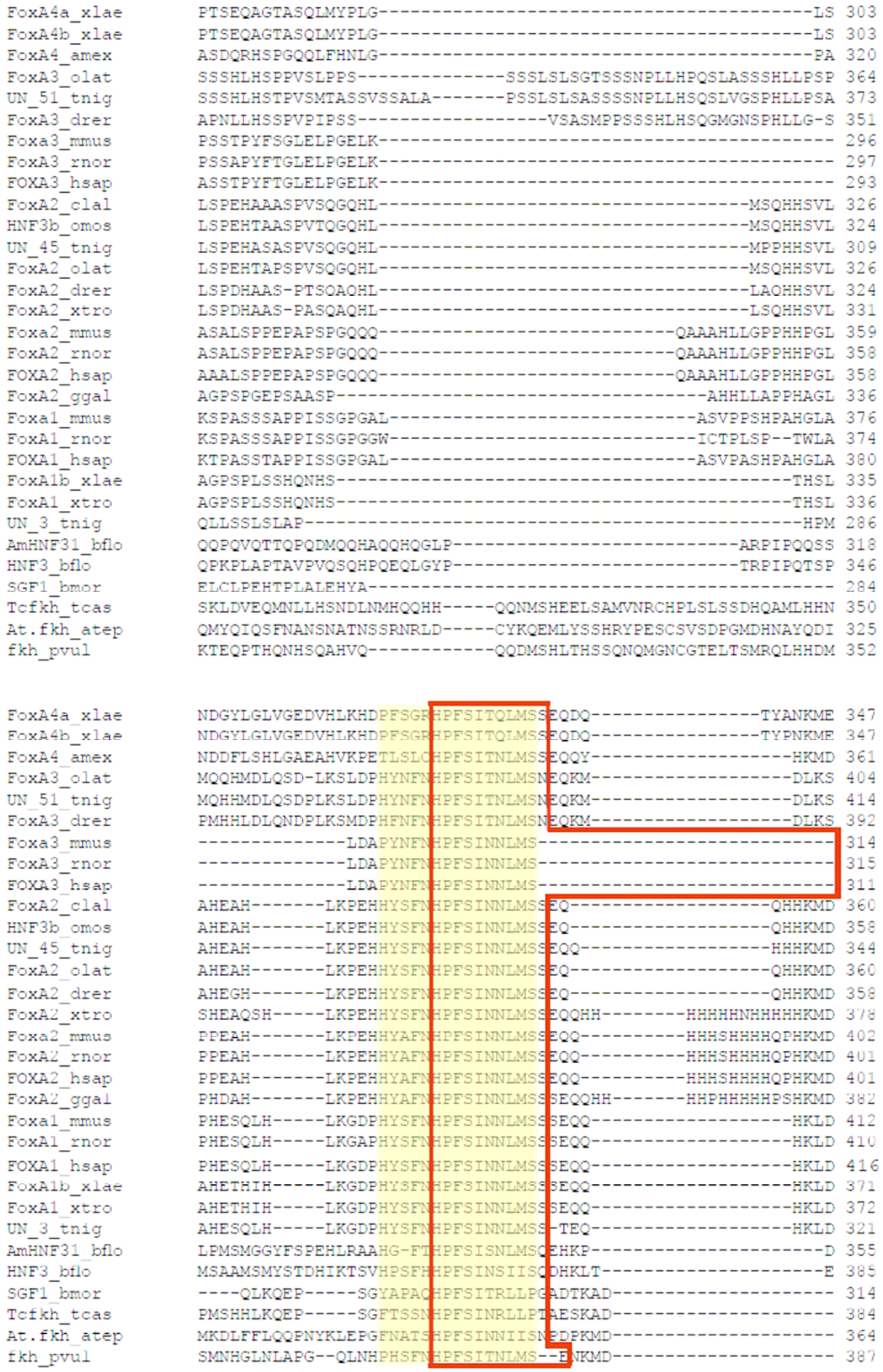
* The sequence to which amino acid residues reported correspond is given for each lineage.

Figures


FoxA4a_xlae -MLNRVKLEIKDPMDWN--TMYQENEMYSGIHN----------MTNVLPSNSFLPND 44
FoxA4b_xlae -MLNRVKLELKDPMDWN--TMYQENEIYSGIHN----------MTNGLPSNSFLPTD 44
FoxA4_amex -MLNGIKLENQEAMDWT--HFYQDNEVYSGVHS----------MPSSLAS-TYIPND 43
FoxA3_olat -MLSSVKMEAHDITDWNTTFYSEASEMYSGP---SATG--------LGSMGSINSYINLN 48
UN_51_tnig -MLSSVKMETHDLPEWN-TFYSEASEMYSGP---STMNSGLAS---MGSMGSINSYINLN 52
FoxA3_drer -MLSSVKMESHEIPEWN-PFYSEANEMYSGP---SAMN----------SVSSLNSYINLN 45
Foxa3_mmus -MLGSVKMEAHDLAEWS--YYPEAGEVYSGVN----------PVPTMAPLNSYMTLN 44
FoxA3_rnor -MLGSVKMEAHDLAEWS--YYPEAGEVYSGVN----------PVPTMAPLNSYMSLN 44
FOXA3_hsap -MLGSVKMEAHDLAEWS--YYPEAGEVYSGVT----------PVPTMAPLNSYMTLN 44
FoxA2_clal MMLGAVKMEGHEHT-DWSTYYGEPECYTSVGNMNTGLG---MNSMNTYMSMSGMST---G 53
HNF3b_omos MMLGAVKMEGHEHT-DWSTYYGEPECYTSVGNMNAGLG---MNSMNTYMSMSGMST---T 53
UN_45_tnig -MLGAVKMEGHEHT-DWSTYYAEPECYTSVGNMNTGLG---MNSMNTYMSMSGMNT---T 52
FoxA2_olat MMLGAVKMEGHEHT-DWSTYYGEPECYTSVGNMNTGLG---MNSMNTYMSMSGMST---T 53
FoxA2_drer -MLGAVKMEGHEHAADWSTYYGEPECYTSVSNMNTGLG---MNSMNTYMTMSGMSS---T 53
FoxA2_xtro -MLGAVKMEGHEAT-DWSSYYGEPEAYSSVGNMNAGLS---MNPMNTYMSMSAMST---S 52
Foxa2_mmus -MLGAVKMEGLEPS-DWSSYYAEPEGYSSVSNMNAGLG---MNGMNTYMSMSAAAMGGGS 55
FoxA2_rnor -MLGAVKMEGHEPS-DWSSYYAEPEGYSSVSNMNASLG---MNGMNTYMSMSAAAMGSGS 55
FOXA2_hsap -MLGAVKMEGHEPS-DWSSYYAEPEGYSSVSNMNAGLG---MNGMNTYMSMSAAAMGSGS 55
FoxA2_ggal -MLGAVKMEGHEHT-DWSNYYGEPESYSSVSNMNAGLG---MNSMNTYMTMSAMST---T 52
Foxa1_mmus -MLGTVKMEGHESNDWNSYYADTQEAYSSVPVSNMNSG---LGSMNSMNTYMTMNTMTTS 56
FoxA1_rnor -MLGTVKMEGHESNDWNSYYADTQEAYSSVPVSNMNSG---LGSMNSMNTYMTMNTMTTS 56
FOXA1_hsap -MLGTVKMEGHETSDWNSYYADTQEAYSSVPVSNMNSG---LGSMNSMNTYMTMNTMTTS 56
FoxA1b_xlae -MLGIVKMEGHETTDWSNYYQDAQEGYSSVPVSNMPQG---LATMN---TYMTMNPMSSG 53
FoxA1_xtro -MLGIVKMEGHETTDWSNYYQDTQEAYSSVPVSNMTQG---LASMN---TYMTMNPMSSS 53
UN_3_tnig ----------MAATGLGSG---LGSMTG---------YMSS 19
AmHNF31_bflo ----------MLSAPKGYPTG----------TTMNTMGMNTMSSM 25
HNF3_bflo ---MLSPKSAYEAQGGSPSSMQAMTAMTTGNSY----------SPSSYTSSGYSVTQSM 46
SGF1_bmor --MISQKLSYGDVPTS--------ASLSSLSPGLAPPY---VNGMGCMPAQPYPN----- 42
Tcfkh_tcas --MLTQKLYSDSTTMA--------TSSNAMSPMTPTYS---MNSMSCVSMPSMNCSPQGA 47
At.fkh_atep --MLTHKSFPDCSTMP---------SGNYMPVTSMPMS---LNYSTPQFNAGLLSP---- 42
fkh_pvul --MLSAKPGSYDPTSSGGYSMASMTSINTMGGVGPMSN---MNYPSQGMGGMHGAMSSMN 55
FoxA4a_xlae VSTVTTS---MPYMSNGLPGPV----------TSIQGNIGSLGSMPQGMVGSLAPP-- 87
FoxA4b_xlae VPTVTSS---MTYMSNGLPGPV----------ASIQGNLGSLGSMTQGMVGSLAPP-- 87
FoxA4_amex IHVASAG---MNYMNPGLGCSV----------PALPGGGSPVNSMAPGSLNIASPLNQ 88
FoxA3_olat AAAS-PTAMNMPYPSSSLSSSS----------LAPMGSGPNHMSLSPVASSLSSGSLT 95
UN_51_tnig AATASPASMNMAYPSSSLSSST----------LASMGSGPTHMSLSPVASSLASGPLT 100
FoxA3_drer SACS-TSSMNMGYPSAGLNSSP----------LSSMGGGPNHMSLSPVGSSLNPSSLT 92
Foxa3_mmus PLSS-------PYPPGGLQASP----------LPTG-------PLAPPAPTAPLGPTF 78
FoxA3_rnor PLSS-------PYPPGGLQASP----------LPTG-------PLAPPAPTAPLGPTF 78
FOXA3_hsap PLSS-------PYPPGGLPASP----------LPSG-------PLAPPAPAAPLGPTF 78
FoxA2_clal ANMTAN-SMNMSYVNTGMSPSMTGMSPG------TGAMNGMG-----AGMTAMGAALSPS 101
HNF3b_omos ANMTAN-SMNMSYVNTGMSPSMTGMSPG------TGAMNGMG-----AGMTAMSAALSPS 101
UN_45_tnig ANMTAN-SMNMSYVNTGMSPSMTGMSPG------TGAMNGMG-----AGMTAMSAALSPS 100
FoxA2_olat ANMTAN-SMNMSYVNTGMSPSMTGMSPG------TGAMNGMG-----AGMTAMSTALSPS 101
FoxA2_drer ANMTAANTMNMSYVNTGMSPSMTGMSPG------TGAMAGMG-----AGMTGMSAALSPT 102
FoxA2_xtro ANMTAG-SMNMSYVNTGMSPSLTGMSPG------TGAMTGMG-----TGVASMASHLSPS 100
Foxa2_mmus GNMSAGSMNMSSYVGAGMSPSLAGMSPG------AGAMAGMSGSAGAAGVAGMGPHLSPS 109
FoxA2_rnor GNMSAGSMNMSSYVGAGMSPSLAGMSPG------AGAMAGMSGSAGAAGVAGMGPHLSPS 109
FOXA2_hsap GNMSAGSMNMSSYVGAGMSPSLAGMSPG------AGAMAGMGGSAGAAGVAGMGPHLSPS 109
FoxA2_ggal ANMTAATSMNMSYANTGMSPSLAGMSPG------AGAMAGMG----SAGVAGMGAHLSPS 102
Foxa1_mmus GNMTPA-SFNMSYANTGLGAGLSPGAVAGMPGASAGAMNSMTAAGVTAMGTALSPGGMGS 115
FoxA1_rnor GNMTPA-SFNMSYANPGLGAGLSPGAVAGMPGGSAGAMNSMTAAGVTAMGAALSPGGMGS 115
FOXA1_hsap GNMTPA-SFNMSYANPGLGAGLSPGAVAGMPGGSAGAMNSMTAAGVTAMGTALSPSGMGA 115
FoxA1b_xlae SNITSG-SFNMPYGNSGLGAGLSPSGMSGMGS--AGAMNGMGSG-VPSMGSALSPSNMNA 109
FoxA1_xtro SNMTAG-SFNMSYANSGLGAGLSPSGMSGMGAGSASAMNGMGSG-VSSMGTALSPSSMNA 111
UN_3_tnig GGTTAG-SFNMSYS----GSALSPPPVAGMSSSTPAAMSGLGGG-MAPMGGPLSPSHMNS 73
AmHNF31_bflo GGMNHASYTGTGPVNPGAYSAS----------AYSGTMNGMGTG---------- 59
HNF3_bflo TTLGSSYSTGMNPVGMGTIPPS----------HTSMTTMGMNSVPGSQLTTMSTMGAQ 94
SGF1_bmor ---------LYSNNMVAGGSCMGSPSVGYSPP----------STMASCMGGAGA 77
Tcfkh_tcas SFGSSMLNSGMPGGMAMNGNGMTSSSMGYTTIGSP---ISNRIRHEMATPMATMNSYGSV 104
At.fkh_atep ----------QNMAAVAPACMTQPMP----------PIGSITPLNNVAPN 72
fkh_pvul TMPPSMGSMGMGGMAAMHGSSMHGSMTAMNPMNTMGSMNSMGAMGSMNGMSSIGSMSSMN 115


FoxA4a_xlae ----------PSTAAYPLGYCQGESE----------FQRDPRTYRRNYSHAKPPYS 123
FoxA4b_xlae ----------PSTSAYPLGYCQGESE----------FQRDPRTYRRNYSHAKPPYS 123
FoxA4_amex SMNVAPQGSSSMGSYTPMNTLAYGQGGLD----------YQRDPRAYRRNYSHAKPPYS 137
FoxA3_olat QLGPAAAGSLGPLSHYQNMSQSMSQLGYPSTGSLSRSSP-KEIPPKPYRRSLTHAKPPYS 154
UN_51_tnig QLTTAPPASLGSLPHYQNMSQSMGQLGYTSNASLTRTGP-KEIPPKPYRRSLTHAKPPYS 159
FoxA3_drer QLG-SSASTLGPLSHYQSMSQPMSQISYPSPTSLNR----TKEMPKPYRRSLTHAKPPYS 147
Foxa3_mmus PSLGTGGSTGGSASGYVAPSPGLVHG----------KEMAKGYRRSLAHAKPPYS 123
FoxA3_rnor PGLGAGSGTGGSASGYGAPSPGLVHG----------KEMAKGYRRSLTHAKPPYS 123
FOXA3_hsap PGLGVSGGS--SSSGYGAPSPGLVHG----------KEMPKGYRRSLAHAKPPYS 121
FoxA2_clal MSPMTAQPA-SMNALTSYSNMNAMSPIYGQSNINRSRD------PKTYRRSYTHAKPPYS 154
HNF3b_omos MSPMTAQPA-SMNALTSYSNMNAMSPIYGQSNINRSRD------PKTYRRSYTHAKPPYS 154
UN_45_tnig MSPMTAQPA-SMNALTSYTNMNAMSPIYGQSNINRSRD------PKTYRRSYTHAKPPYS 153
FoxA2_olat MSPMTGQPG-SMNALTSYTNMNAMSPIYGQSNINRSRD------PKTYRRSYTHAKPPYS 154
FoxA2_drer MSPMAAQAP-SMNALTSYSNMNAMSPMYGQSNINRSRD------PKTYRRSYTHAKPPYS 155
FoxA2_xtro MSPMSAQAT-SMNALAPYTNMNSMSPIYGQSNINRSRD------PKTYRRSYTHAKPPYS 153
Foxa2_mmus LSPLGGQAAGAMGGLAPYANMNSMSPMYGQAGLSRARD------PKTYRRSYTHAKPPYS 163
FoxA2_rnor LSPLGGQAAGAMGGLAPYANMNSMSPMYGQAGLSRARD------PKTYRRSYTHAKPPYS 163
FOXA2_hsap LSPLGGQAAGAMGGLAPYANMNSMSPMYGQAGLSRARD------PKTYRRSYTHAKPPYS 163
FoxA2_ggal MSPMGGQAG-SMNALAPYTNMNSMSPIYGQSNLNRSRD------PKTYRRSYTHAKPPYS 155
Foxa1_mmus MGAQPATSMNGLGPYAAAMNPCMSPMAYAPSNLGRSRAG-GGGDAKTFKRSYPHAKPPYS 174
FoxA1_rnor MGAQPAASMNGLGPYAAAMNPCMSPMAYAPSNLGRSRAG-GGGDAKTFKRSYPHAKPPYS 174
FOXA1_hsap MGAQQAASMNGLGPYAAAMNPCMSPMAYAPSNLGRSRAG-GGGDAKTFKRSYPHAKPPYS 174
FoxA1b_xlae IQSAQQASMNSLS--YSSMNSGMSPMGYGATNINRTRD------SKTFRRSYPHAKPPYS 161
FoxA1_xtro MS-AQQASINSLS--YSGMNPGMSPMAYGPSNMNRTRD------TKTFRRSYPHAKPPYS 162
UN_3_tnig VPAQQGSLG---LSPYGGMSPSSSMAYSGGGGMNRARD------NKAFRRSYPHAKPPYS 124
AmHNF31_bflo -YPAAGGMGGLQSYPAGSVNAMGTMQTMN--NMALNRNA-IAERKAYRRSYTHAKPPYS 115
HNF3_bflo HPGLANSLGVGMAHPGQSMSPMSMQGSVNGVNMNLTRED-VLNRQKQYRRSYTHAKPPYS 153
SGF1_bmor VPYGSLPREQEAASPTSALQRAR----------NLKTYRRSYTHAKPPYS 117
Tcfkh_tcas GTLGRGDLGGGDTSPNSALQRAR----------ALKTYRRSYTHAKPPYS 144
At.fkh_atep RTDMQTYVNTDYHSDPNLALKAR----------NLK-FRRSLPHAKPPYS 111
fkh_pvul GMNRQMDPNMMSMDRAQALNRA----------RLKNYRRSYTHAKPPYS 154
FoxA4a_xlae YISLITMAIQQAPNKMMTLNEIYQWIIDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVPRS 183
FoxA4b_xlae YISLITMAIQQAPNKMMTLNEIYQWIVDLFPYYRQNQQRWQNSIRHSLSFNDCFIKVPRS 183
FoxA4_amex YISLITMAVQQSPNKMMTLNEIYQWITDLFPYYRQNQQSWQNSIRHSLSFNDCFVKVPRS 197
FoxA3_olat YISLITMAIQQSGSKMLTLNEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 214
UN_51_tnig YISLITMAIQQSSSKMLTLNEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 219
FoxA3_drer YISLITMAIQQSQSKMLTLNEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 207
Foxa3_mmus YISLITMAIQQAPGKMLTLSEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 183
FoxA3_rnor YISLITMAIQQAPGKMLTLSEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 183
FOXA3_hsap YISLITMAIQQAPGKMLTLSEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARS 181
FoxA2_clal YISLITMAIQQSPSKMLTLAEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRS 214
HNF3b_omos YISLITMAIQQSPSKMLTLAEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRS 214
UN_45_tnig YISLITMAIQQSPSKMLTLAEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRS 213
FoxA2_olat YISLITMAIQQSPSKMLTLAEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 214
FoxA2_drer YISLITMAIQQSPSKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 215
FoxA2_xtro YISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 213
Foxa2_mmus YISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 223
FoxA2_rnor YISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFN-DFLKVPRA 222
FOXA2_hsap YISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 223
FoxA2_ggal YISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRS 215
Foxa1_mmus YISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVARS 234
FoxA1_rnor YISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNACFVKVARS 234
FOXA1_hsap YISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVARS 234
FoxA1b_xlae YISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFIKVSRS 221
FoxA1_xtro YISLITMAIQQAPSKMLTLSEIYQWIMDLFLYYRQNQQRWQNSIRHSLSFNDCFVKVARS 222
UN_3_tnig YISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVSRS 184
AmHNF31_bflo YISLITMSIQSSPNKMVTLAEIYQFIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVPRT 175
HNF3_bflo YIALITMAVQSSPNKMVTLSEIYQFIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVQRT 213
SGF1_bmor YISLITMAIQNNPSRMLTLSEIYQFIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRT 177
Tcfkh_tcas YISLITMAIQNSPQKMLTLSEIYQFIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRT 204
At.fkh_atep YISLITMAIQNSPQKMLTLNEIYQFIVDIFPFYRQNQQRWQNSIRHSLSFNDCFVKVART 171
fkh_pvul YISLITMAIQQSPNKMCTLSEIYQFIMDLFPFYRQNQQRWQNSIRHSLSFNDCFVKVPRT 214


FoxA4a_xlae PEKPGKGSYWTLHPESGNMFENGCYLRRQKRFKCER---SKSG----------EG 225
FoxA4b_xlae PEKPGKGSYWTLHPESGNMFENGCYLRRQKRFKCER---SKSG----------EG 225
FoxA4_amex PEKPGKGSYWALHPDSGNMFENGCYLRRQKRFKCDRRSGSKRV----------QD 242
FoxA3_olat PDKPGKGSYWTLHPQSGNMFENGCYLRRQKRFKIEDKASKKG----------SK 258
UN_51_tnig PDKPGKGSYWTLHPQSGNMFENGCYLRRQKRFKIEDKAKKGK----------SQ 263
FoxA3_drer PDKPGKGSYWALHPNSGNMFENGCYLRRQKRFKIEEKAGKKSS----------SK 252
Foxa3_mmus PDKPGKGSYWALHPSSGNMFENGCYLRRQKRFKLEEKAKKGN---------- 225
FoxA3_rnor PDKPGKGSYWALHPSSGNMFENGCYLRRQKRFKLEEKAKKGN---------- 225
FOXA3_hsap PDKPGKGSYWALHPSSGNMFENGCYLRRQKRFKLEEKVKKGG---------- 223
FoxA2_clal PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCDKKMGMKDG----------G 258
HNF3b_omos PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCDKKMKEP----------G 256
UN_45_tnig PDKPGKGSFWTLHPDSGNMFENGCYCG----------A 241
FoxA2_olat PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKKMSMKEP----------G 258
FoxA2_drer PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCDKKLSKD----------PS 258
FoxA2_xtro PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKKPSLREG----------GG 258
Foxa2_mmus PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKQLALKEA----------A 267
FoxA2_rnor PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCENELALKEA----------A 266
FOXA2_hsap PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKQLALKEA----------A 267
FoxA2_ggal PDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKQLATKDG----------G 259
Foxa1_mmus PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKQPGAGGG--------SGGGGSKGG 286
FoxA1_rnor PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKQPGAGGG--------SGGGGSKGV 286
FOXA1_hsap PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKQPGAGGGGGSG----SGGSGAKGG 290
FoxA1b_xlae PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKTQ----------GGKGN 264
FoxA1_xtro PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKQQ----------GGKGS 265
UN_3_tnig PDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKKTSLKCDG---------- 228
AmHNF31_bflo PDKPGKGSYWTLHPEAGNMFENGCYLRRQKRFKCEKKLAMKMAQQQA----------APT 225
HNF3_bflo PDKPGKGSYWTLHPNAHSMFENGCYLRRQKRFKCERKAALKAEQKAE----------NEE 263
SGF1_bmor PDKPGKGSFWTLHPDSGNMFENGCFLRRQKRFKDEKKETLRQA---------- 220
Tcfkh_tcas PDKPGKGSFWSLHPDSGNMFENGCYLRRQKRFKDEKKELIRQT---------- 247
At.fkh_atep PDKPGKGSFWALHPESGDMFENGCFLRRQKRFKCTKKEAIRQT---------- 214
fkh_pvul PDKPGKGSYWALHPDSGNMFENGCYLRRQKRFKCLKKESMRSS---------- 257
FoxA4a_xlae E-KKVNKPGEETGGNLKENPLGYDDCSSSRSPQAAVNDGGRDSTGSSIHQACGSSPVGLS 284
FoxA4b_xlae E-RKGNKPGDETGGSLKETPVSFDDCSSSRSPQAAVNDGGRDSTGSSIHQATGGSPVGFS 284
FoxA4_amex E-NGSHQPCEAAIKPLILHLRAPTEFSNPFTRASAGLEHNHAMMSQQQPNMAASSPSISP 301
FoxA3_olat NQEGGSGKGSHSGDHLVENHSPAGGSEGPDSAHSDNSHPGSSDDQPHQRNSLVPLDCPQL 318
UN_51_tnig E---GSGKGNHSGDHLAEDHSPTGGSEGADSAHSDNSHPGSSDDQQHSRNSLVPLDCPPL 320
FoxA3_drer SQDGSSTKGTHSSEGMQEEHSPTTGSDGAESAHSDSSHAGSTSEE--QQRSLVQLDCPSQ 310
Foxa3_mmus -----SATSASRNGTAGSATSATTTAATAVTSPAQPQPTP-SEPEAQSGDDVGGLDCASP 279
FoxA3_rnor -----SATSATRNGTVGSATSATTTAATAVTSPAQPQPTPPSEPEAQSGEDVGGLDCASP 280
FOXA3_hsap -----SGAATTTRNGTGSAASTTTPAATVTSPPQPPPPAP--EPEAQGGEDVGALDCGSP 276
FoxA2_clal RKSGDGGSS----------NSSSESCNGNESPHSNSS-SGEHKRSLSDMKSSQA 301
HNF3b_omos RKSGDGGSS----------NSSSESCNGNESPLSNSS-SSDHKRSLSDMKTSQA 299
UN_45_tnig RSGSNGGSS----------NSSSESCNGNESPHSNSS-SSEHKRSLSDMKTSQA 284
FoxA2_olat RKGGDGGSA----------NSSSDSCNGNESPHSNSS-SGEHKRSLSDMKGSQA 301
FoxA2_drer RKTSEGGS----------NSSSESCNGNESPHSNSS-SNELKRSLSDMKSGQG 300
FoxA2_xtro KKLSEGSSSVGS----------AANSSSESSVGNESPHSSSSPCQEQKRSLVDMKSSQG 307
Foxa2_mmus GAASSGGKKTAPGSQASQAQLGEAAGSASETPAGTESPHSSASPCQEHKRGGLSELKGAP 327
FoxA2_rnor GAGSGGGKKTAPGTQASQVQLGEAAGSASETPAGTESPHSSASPCQEHKRGGLSELKGTP 326
FOXA2_hsap GAAGSG-KKAAAGAQASQAQLGEAAGPASETPAGTESPHSSASPCQEHKRGGLGELKGTP 326
FoxA2_ggal GG------KKGPGQPPSQPLGEGSSSGGSEGSAGAESPAS-ASPCRDNKR-ALAELKGAA 311
Foxa1_mmus PESRKDPSGPGNPSAESPLHRGVHGKASQLEGAPAPG-PAASPQTLDHSGATATGGASEL 345
FoxA1_rnor PENRKDPSGPVNPSAESPIHRGVHGKASQLEGAPAPG-PAASPQTLDHSGATATGGGSEL 345
FOXA1_hsap PESRKDPSGASNPSADSPLHRGVHGKTGQLEGAPAPG-PAASPQTLDHSGATATGGASEL 349
FoxA1b_xlae QDGRKDHSGPS-----SPLHR-VHGKSSQMDSSSSMSNPSSSPQALEHNGSNGEMKPQVA 318
FoxA1_xtro QDGRKDVSGPS-----SPLHR-VHGKSSQMDSSSSMSNPSSSPQSLEHNGSNGEMKPQVA 319
UN_3_tnig ----------RKEEGGGASPSGDKPGGLLVSSSQAASPPGLDLQGGTPDLKVSSS 273
AmHNF31_bflo PPTPGRELRRLADYHGSTPTTTSNGASTLQPLQPINTPSPNPQEQQQHQHHQHQQQHQ 285
HNF3_bflo VLSGAPAGQQPPPQANTPVHNSPTPESTHVSSSPMTVTTQTPSP----------TLTQLT 313
SGF1_bmor ----QKAQQTHGHHGGSHDKRGEHGHDKSAPPGPGEDKEMRD--------ELLAQLHAAP 268
Tcfkh_tcas ----HKSPSHSVDNSNSSEKKSSIQHGDDAHKTHHLDKNPEN--------TLGTMLSIHP 295
At.fkh_atep ----QKCQKSPGDQSVKSEPEMNSSPKMDPKSSPMKVPEMEQPCLPPVNTSLPSTTDAYQ 270
fkh_pvul ----HDDDPSCGMSNGQNSADSTPTSTGDGTPLSAPNTPPHP---------VEQSQMTQP 304


FoxA4a_xlae PTSEQAGTASQLMYPLG---------- LS 303
FoxA4b_xlae PTSEQAGTASQLMYPLG---------- LS 303
FoxA4_amex ASDQRHSPGQQLFHNLG---------- PA 320
FoxA3_olat SSSHLHSPPVSLPPS----------SSSLSLSGTSSSNPLLHPQSLASSSHLLPSP 364
UN_51_tnig SSSHLHSTPVSMTASSVSSALA-------PSSLSLSASSSSNPLLHSQSLVGSPHLLPSA 373
FoxA3_drer APNLLHSSPVPIPSS----------VSASMPPSSSHLHSQGMGNSPHLLG-S 351
Foxa3_mmus PSSTPYFSGLELPGELK---------- 296
FoxA3_rnor PSSAPYFTGLELPGELK---------- 297
FOXA3_hsap ASSTPYFTGLELPGELK---------- 293
FoxA2_clal LSPEHAAASPVSQGQHL----------MSQHHSVL 326
HNF3b_omos LSPEHTAASPVTQGQHL----------MSQHHSVL 324
UN_45_tnig LSPEHASASPVSQGQHL----------MPPHHSVL 309
FoxA2_olat LSPEHTAPSPVSQGQHL----------MSQHHSVL 326
FoxA2_drer LSPDHAAS-PTSQAQHL----------LAQHHSVL 324
FoxA2_xtro LSPDHAAS-PASQAQHL----------LSQHHSVL 331
Foxa2_mmus ASALSPPEPAPSPGQQQ----------QAAAHLLGPPHHPGL 359
FoxA2_rnor ASALSPPEPAPSPGQQQ----------QAAAHLLGPPHHPGL 358
FOXA2_hsap AAALSPPEPAPSPGQQQ----------QAAAHLLGPPHHPGL 358
FoxA2_ggal AGPSPGEPSAASP----------AHHLLAPPHAGL 336
Foxa1_mmus KSPASSSAPPISSGPGAL----------ASVPPSHPAHGLA 376
FoxA1_rnor KSPASSSAPPISSGPGGW----------ICTPLSP--TWLA 374
FOXA1_hsap KTPASSTAPPISSGPGAL----------ASVPASHPAHGLA 380
FoxA1b_xlae AGPSPLSSHQNHS----------THSL 335
FoxA1_xtro AGPSPLSSHQNHS----------THSL 336
UN_3_tnig QLLSSLSLAP----------HPM 286
AmHNF31_bflo QQPQVQTTQPQDMQQHAQQHQGLP----------ARPIPQQSS 318
HNF3_bflo QPKPLAPTAVPVQSQHPQEQLGYP----------TRPIPQTSP 346
SGF1_bmor ELCLPEHTPLALEHYA---------- 284
Tcfkh_tcas SKLDVEQMNLLHSNDLNMHQQHH-----QQNMSHEELSAMVNRCHPLSLSSDHQAMLHHN 350
At.fkh_atep QMYQIQSFNANSNATNSSRNRLD-----CYKQEMLYSSHRYPESCSVSDPGMDHNAYQDI 325
fkh_pvul KTEQPTHQNHSQAHVQ----------QQDMSHLTHSSQNQMGNCGTELTSMRQLHHDM 352
FoxA4a_xlae NDGYLGLVGEDVHLKHDPFSGRHPFSITQLMSSEQDQ----------TYANKME 347
FoxA4b_xlae NDGYLGLVGEDVHLKHDPFSGRHPFSITQLMSSEQDQ----------TYPNKME 347
FoxA4_amex NDDFLSHLGAEAHVKPETLSLCHPFSITNLMSSEQQY----------HKMD 361
FoxA3_olat MQQHMDLQSD-LKSLDPHYNFNHPFSITNLMSNEQKM----------DLKS 404
UN_51_tnig MQHHMDLQSDPLKSLDPHYNFNHPFSITNLMSNEQKM----------DLKS 414
FoxA3_drer PMHHLDLQNDPLKSMDPHFNFNHPFSITNLMSNEQKM----------DLKS 392
Foxa3_mmus ----------LDAPYNFNHPFSINNLMS---------- 314
FoxA3_rnor ----------LDAPYNFNHPFSINNLMS---------- 315
FOXA3_hsap ----------LDAPYNFNHPFSINNLMS---------- 311
FoxA2_clal AHEAH-------LKPEHHYSFNHPFSINNLMSSEQ----------QHHKMD 360
HNF3b_omos AHEAH-------LKPEHHYSFNHPFSINNLMSSEQ----------QHHKMD 358
UN_45_tnig AHEAH-------LKPEHHYSFNHPFSINNLMSSEQQ----------HHHKMD 344
FoxA2_olat AHEAH-------LKPEHHYSFNHPFSINNLMSSEQ----------QHHKMD 360
FoxA2_drer AHEGH-------LKPEHHYSFNHPFSINNLMSSEQ----------QHHKMD 358
FoxA2_xtro SHEAQSH-----LKPEHHYSFNHPFSINNLMSSEQQHH--------HHHHHNHHHHHKMD 378
Foxa2_mmus PPEAH-------LKPEHHYAFNHPFSINNLMSSEQQ----------HHHSHHHHQPHKMD 402
FoxA2_rnor PPEAH-------LKPEHHYAFNHPFSINNLMSSEQQ----------HHHSHHHHQPHKMD 401
FOXA2_hsap PPEAH-------LKPEHHYAFNHPFSINNLMSSEQQ----------HHHSHHHHQPHKMD 401
FoxA2_ggal PHDAH-------LKPEHHYAFNHPFSINNLMSSEQQHH-------HHPHHHHHPSHKMD 382
Foxa1_mmus PHESQLH-----LKGDPHYSFNHPFSINNLMSSEQQ----------HKLD 412
FoxA1_rnor PHESQLH-----LKGAPHYSFNHPFSINNLMSSEQQ----------HKLD 410
FOXA1_hsap PHESQLH-----LKGDPHYSFNHPFSINNLMSSEQQ----------HKLD 416
FoxA1b_xlae AHETHIH-----LKGDPHYSFNHPFSINNLMSSEQQ----------HKLD 371
FoxA1_xtro AHETHIH-----LKGDPHYSFNHPFSINNLMSSEQQ----------HKLD 372
UN_3_tnig AHESQLH-----LKGDPHYSFNHPFSINNLMSS-TEQ----------HKLD 321
AmHNF31_bflo LPMSMGGYFSPEHLRAAHG-FTHPFSISNLMSQEHKP----------D 355
HNF3_bflo MSAAMSMYSTDHIKTSVHPSFHHPFSINSIISQDHKLT----------E 385
SGF1_bmor ----QLKQEP-----SGYAPAQHPFSITRLLPGADTKAD---------- 314
Tcfkh_tcas PMSHHLKQEP-----SGFTSSNHPFSINRLLPTAESKAD---------- 384
At.fkh_atep MKDLFFLQQPNYKLEPGFNATSHPFSINNIISNEDPKMD---------- 364
fkh_pvul SMNHGLNLAPG--QLNHPHSFNHPFSITNLMS--ENKMD---------- 387

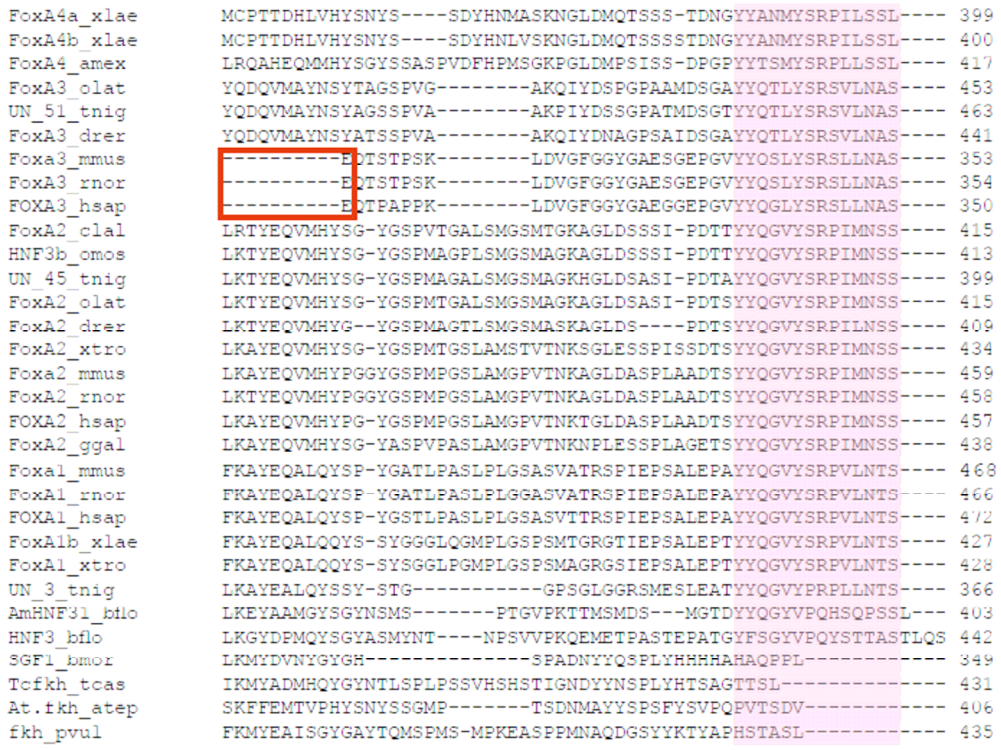
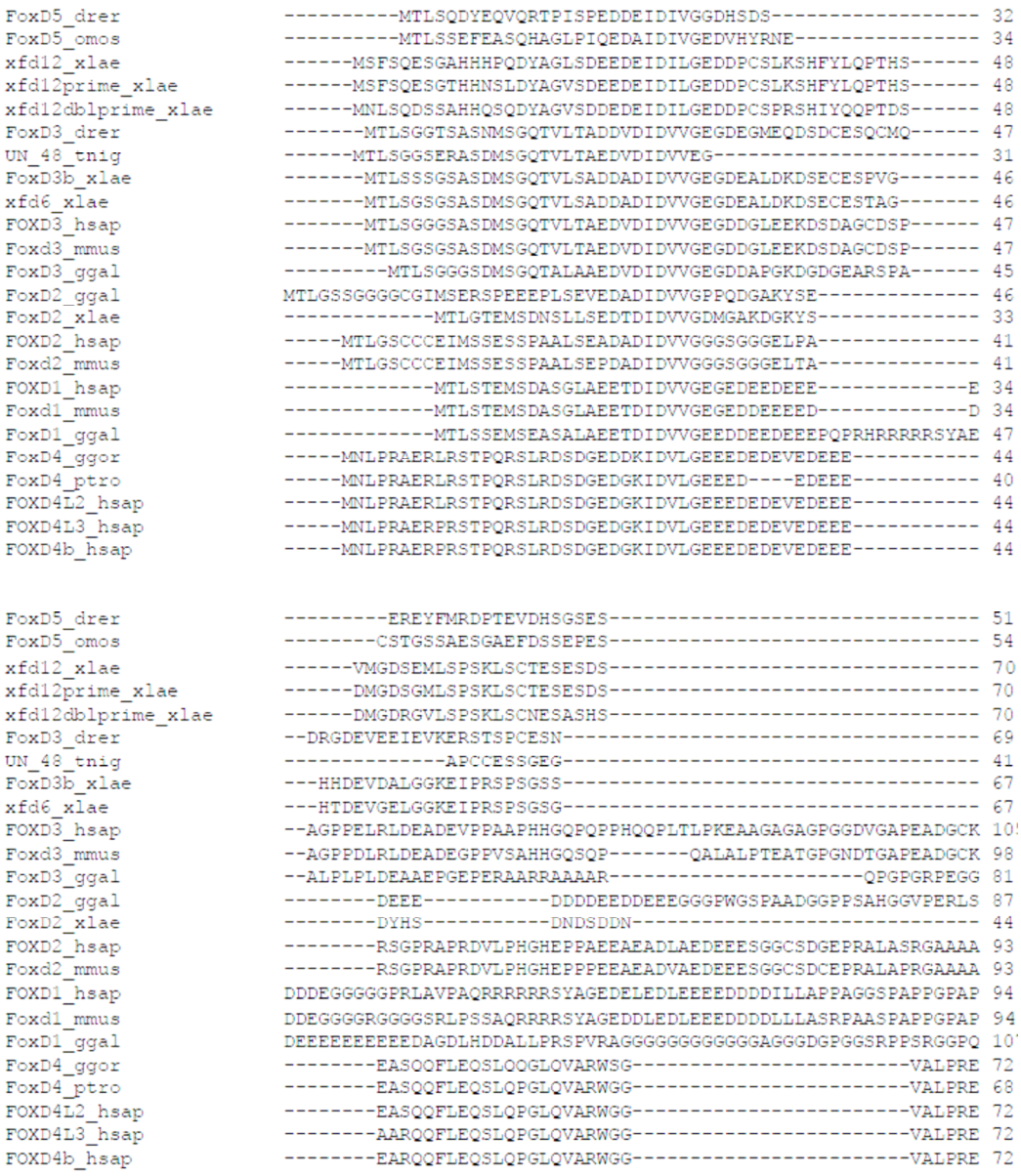
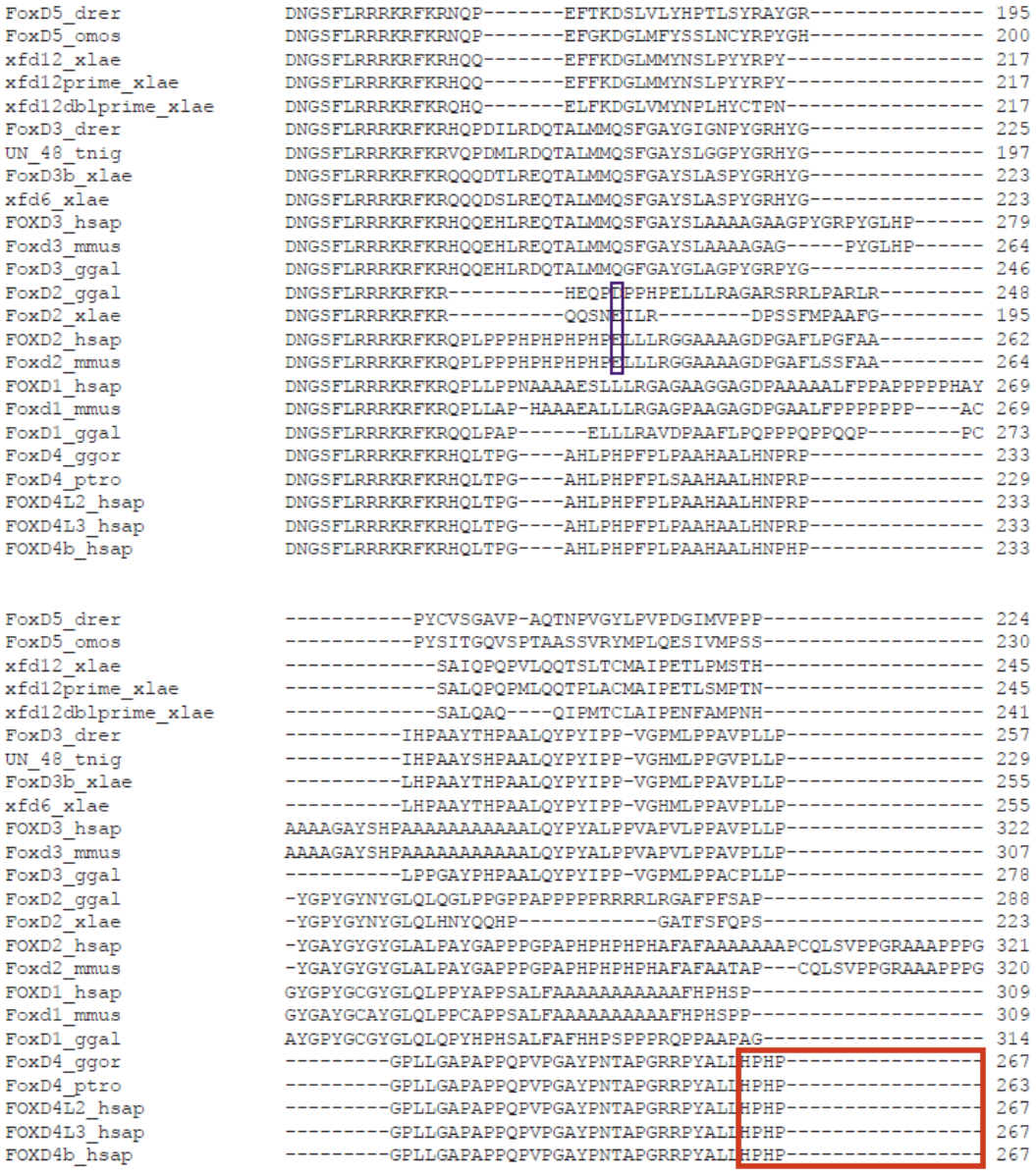


Figure 2-1. Alignment of the FoxA cluster of sequences.

Conserved and functionally important regions noted in the literature are highlighted. Peach: conserved domain IV (Pani et al. 1992; Qian and Costa 1995) Blue: conserved domain V (Pani et al. 1992; Qian and Costa 1995) Black Box: nuclear localization signal (Qian and Costa 1995) Green: forkhead domain (NCBI Protein database, see Table 2-1 for accession numbers) Yellow: conserved domain II (Lai et al. 1991) Red Box: EH1 motif (Copley 2005; Yaklichkin et al. 2007) Pink: conserved domain III (Lai et al. 1991) Purple Box: positively selected sites identified by branch-site analyses


FoxD5_drer ----------MTLSQDYEQVQRTPISPEDDEIDIVGGDHSDS---------- 32
FoxD5_omos ----------MTLSSEFEASQHAGLPIQEDAIDIVGEDVHYRNE---------- 34
xfd12_xlae ------MSFSQESGAHHHPQDYAGLSDEEDEIDILGEDDPCSLKSHFYLQPTHS------ 48
xfd12prime_xlae ------MSFSQESGTHHNSLDYAGVSDEEDEIDILGEDDPCSLKSHFYLQPTHS------ 48
xfd12dblprime_xlae ------MNLSQDSSAHHQSQDYAGVSDDEDEIDILGEDDPCSPRSHIYQQPTDS------ 48
FoxD3_drer -------MTLSGGTSASNMSGQTVLTADDVDIDVVGEGDEGMEQDSDCESQCMQ------ 47
UN_48_tnig ------MTLSGGSERASDMSGQTVLTAEDVDIDVVEG---------- 31
FoxD3b_xlae -------MTLSSSGSASDMSGQTVLSADDADIDVVGEGDEALDKDSECESPVG------- 46
xfd6_xlae -------MTLSGSGSASDMSGQTVLSADDADIDVVGEGDEALDKDSECESTAG------- 46
FOXD3_hsap -------MTLSGGGSASDMSGQTVLTAEDVDIDVVGEGDDGLEEKDSDAGCDSP------ 47
Foxd3_mmus -------MTLSGSGSASDMSGQTVLTAEDVDIDVVGEGDDGLEEKDSDAGCDSP------ 47
FoxD3_ggal ---------MTLSGGGSDMSGQTALAAEDVDIDVVGEGDDAPGKDGDGEARSPA------ 45
FoxD2_ggal MTLGSSGGGGCGIMSERSPEEEPLSEVEDADIDVVGPPQDGAKYSE---------- 46
FoxD2_xlae ------------MTLGTEMSDNSLLSEDTDIDVVGDMGAKDGKYS---------- 33
FOXD2_hsap -----MTLGSCCCEIMSSESSPAALSEADADIDVVGGGSGGGELPA---------- 41
Foxd2_mmus -----MTLGSCCCEIMSSESSPAALSEPDADIDVVGGGSGGGELTA---------- 41
FOXD1_hsap ------------MTLSTEMSDASGLAEETDIDVVGEGEDEEDEEE----------E 34
Foxd1_mmus ------------MTLSTEMSDASGLAEETDIDVVGEGEDDEEEED----------D 34
FoxD1_ggal ------------MTLSSEMSEASALAEETDIDVVGEEDDEEDEEEPQPRHRRRRRSYAE 47
FoxD4_ggor -----MNLPRAERLRSTPQRSLRDSDGEDDKIDVLGEEEDEDEVEDEEE---------- 44
FoxD4_ptro -----MNLPRAERLRSTPQRSLRDSDGEDGKIDVLGEEED----EDEEE---------- 40
FOXD4L2_hsap -----MNLPRAERLRSTPQRSLRDSDGEDGKIDVLGEEEDEDEVEDEEE---------- 44
FOXD4L3_hsap -----MNLPRAERPRSTPQRSLRDSDGEDGKIDVLGEEEDEDEVEDEEE---------- 44
FOXD4b_hsap -----MNLPRAERPRSTPQRSLRDSDGEDGKIDVLGEEEDEDEVEDEEE---------- 44
FoxD5_drer ---------EREYFMRDPTEVDHSGSES---------- 51
FoxD5_omos --------CSTGSSAESGAEFDSSEPES---------- 54
xfd12_xlae ------VMGDSEMLSPSKLSCTESESDS---------- 70
xfd12prime_xlae ------DMGDSGMLSPSKLSCTESESDS---------- 70
xfd12dblprime_xlae ------DMGDRGVLSPSKLSCNESASHS---------- 70
FoxD3_drer --DRGDEVEEIEVKERSTSPCESN---------- 69
UN_48_tnig -------------APCCESSGEG---------- 41
FoxD3b_xlae ---HHDEVDALGGKEIPRSPSGSS---------- 67
xfd6_xlae ---HTDEVGELGGKEIPRSPSGSG---------- 67
FOXD3_hsap --AGPPELRLDEADEVPPAAPHHGQPQPPHQQPLTLPKEAAGAGAGPGGDVGAPEADGCK 10
Foxd3_mmus --AGPPDLRLDEADEGPPVSAHHGQSQP-------QALALPTEATGPGNDTGAPEADGCK 98
FoxD3_ggal --ALPLPLDEAAEPGEPERAARRAAAAR----------QPGPGRPEGG 81
FoxD2_ggal --------DEEE-----------DDDDEEDDEEEGGGPWGSPAADGGPPSAHGGVPERLS 87
FoxD2_xlae --------DYHS-----------DNDSDDN---------- 44
FOXD2_hsap --------RSGPRAPRDVLPHGHEPPAEEAEADLAEDEEESGGCSDGEPRALASRGAAAA 93
Foxd2_mmus --------RSGPRAPRDVLPHGHEPPPEEAEADVAEDEEESGGCSDCEPRALAPRGAAAA 93
FOXD1_hsap DDDEGGGGGPRLAVPAQRRRRRRSYAGEDELEDLEEEEDDDDILLAPPAGGSPAPPGPAP 94
Foxd1_mmus DDEGGGGRGGGGSRLPSSAQRRRRSYAGEDDLEDLEEEDDDDLLLASRPAASPAPPGPAP 94
FoxD1_ggal DEEEEEEEEEEDAGDLHDDALLPRSPVRAGGGGGGGGGGGGAGGGDGPGGSRPPSRGGPQ 10
FoxD4_ggor --------EASQQFLEQSLQQGLQVARWSG----------VALPRE 72
FoxD4_ptro --------EASQQFLEQSLQPGLQVARWGG----------VALPRE 68
FOXD4L2_hsap --------EASQQFLEQSLQPGLQVARWGG----------VALPRE 72
FOXD4L3_hsap --------AARQQFLEQSLQPGLQVARWGG----------VALPRE 72
FOXD4b_hsap --------EARQQFLEQSLQPGLQVARWGG----------VALPRE 72

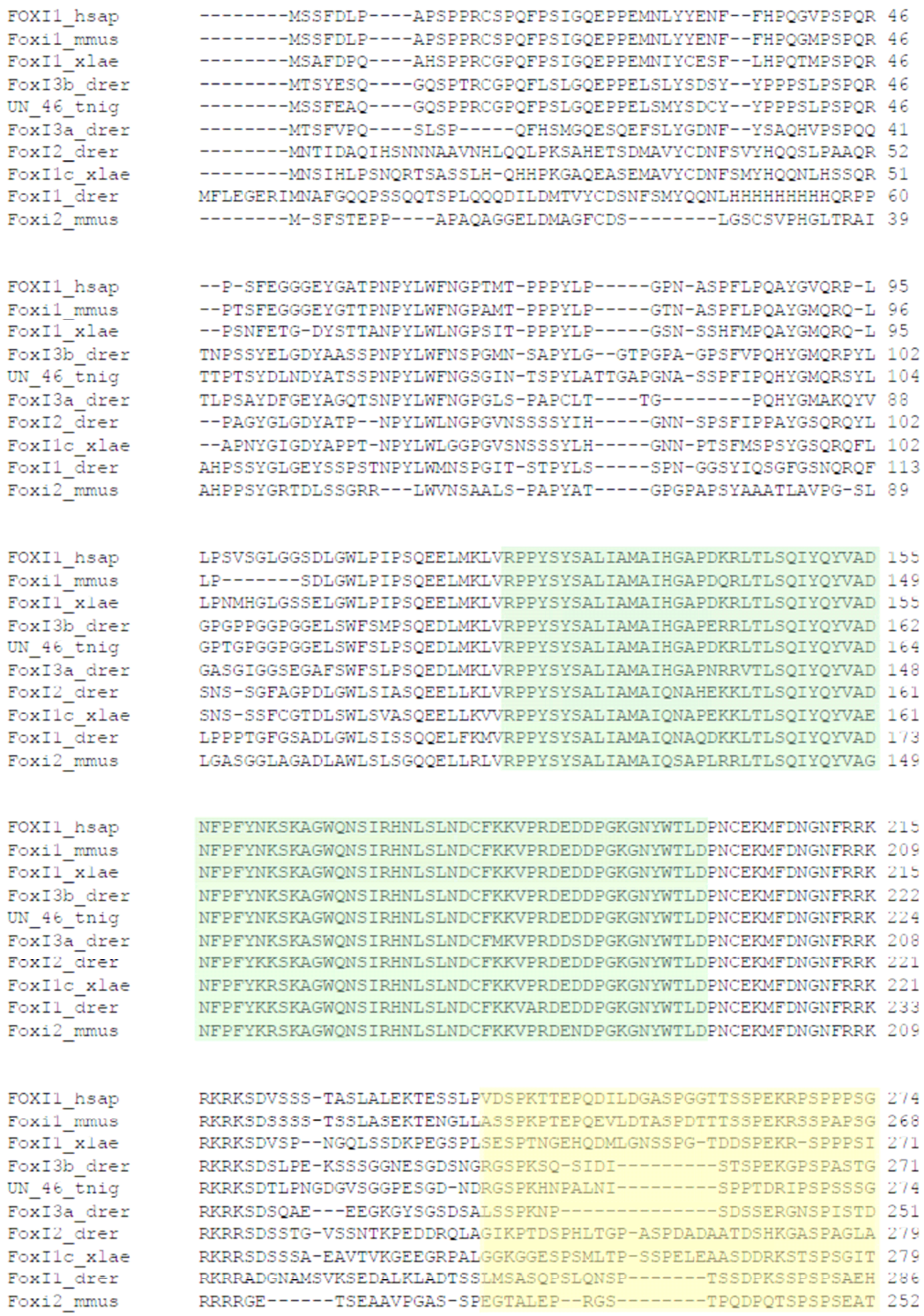
```
FoxD5_drer              ---------SGESE---SDFASSTVAPK--QSSSVKPPYSYIALITMAILQSPMKKLTLS 97
FoxD5_omos              ---------SGESE---NSFCADAPPSRKAQSSSVKPPYSYIALITMAILQSPLKKLTLS 102
xfd12_xlae              ---------SGESEGGTSKDSSTTPTGSKAKRTLVKPPYSYIALITMAILQSPHKKLTLS 121
xfd12prime_xlae         ---------SGESEGGTSKDSPATSPGGKAKRALVKPPYSYIALITMAILQSPHKKLTLS 121
xfd12dblprime_xlae      ---------SGERERGTSKHSLDTTTNGKVKRALVKPPYSYIALITIAIMQSPHKKLTLS 121
FoxD3_drer              ---------ADGETKGDAQESSTGPMQNKPKSSLVKPPYSYIALITMAILQSPQKKLTLS 120
UN_48_tnig              ---------DTGKGEGQEQGARRVGGIQKPKNSLVKPPYSYIALITMAILQSPQKKLTLS 92
FoxD3b_xlae             ---------TDAEGKGESQQQQQEGIQNKPKNSVVKPPYSYIALITMSILQSPQKKLTLS 118
xfd6_xlae               ---------TEAEGKGESQQQQQEGIQNKPKNSLVKPPYSYIALITMSILQSPQKKLTLS 118
FOXD3_hsap              GGVGGEEGGASGGGPGAGSGSAGGLAPSKPKNSLVKPPYSYIALITMAILQSPQKKLTLS 165
Foxd3_mmus              G---GEDAVTGGGGPGAGSGATGGLTPNKPKNSLVKPPYSYIALITMAILQSPQKKLTLS 155
FoxD3_ggal              RGGGGGGGGGEEGASGGGAAAAAAAGQSKPKSSLVKPPYSYIALITMAILQSPQKKLTLS 141
FoxD2_ggal              PAGARSPRAPGPRPGKRAAGGGGGGGGGGGKNPLVKPPYSYIALITMAILQSPKKRLTLS 147
FoxD2_xlae              --VARTPRGDPASPDLSSGSESNQRAEKSPKNALVKPPYSYIALITMSILQSPKKRLTLS 102
FOXD2_hsap              AGSPGPGAAAARGAAGPGPGP--PSGGAATRSPLVKPPYSYIALITMAILQSPKKRLTLS 151
Foxd2_mmus              AGSPGPGVQAARGATGPGPGPGPPSGGAATRSPLVKPPYSYIALITMAILQSPKKRLTLS 153
FOXD1_hsap              AAGAGAG-----GGGGGGGAGGGGSAGSGAKNPLVKPPYSYIALITMAILQSPKKRLTLS 149
Foxd1_mmus              APGTGSGGCSGAGAGGGAGGGTGAGTGGGAKNPLVKPPYSYIALITMAILQSPKKRLTLS 154
FoxD1_ggal              KAAAAGGGGAGGGGGGGGGAGGGGGGGGGKNSLVKPPYSYIALITMAILQSPKKRLTLS 167
FoxD4_ggor              HIEGGGGPSDPSEFGTKFRAPPRSAAASEDARQPAKPPYSYIALITMAILQNPHKRLTLS 132
FoxD4_ptro              HIEGGGGPSDPSEFSTKFRAPPRSAAASEDARQPAKPPYSYIALITMAILQNPHKRLTLS 128
FOXD4L2_hsap            HIEGGGGPSDPSEFGTKFRAPPRSAAASEDARQPAKPPYSYIALITMAILQNPHKRLTLS 132
FOXD4L3_hsap            HIEGGGGPSDPSEFGTKFRAPPRSAAASEDARQPAKPPYSYIALITMAILQNPHKRLTLS 132
FOXD4b_hsap             HIEGGGGPSDPSEFGTKFRAPPRSAAASEDARQPAKPPYSYIALITMAILQNPHKRLTLS 132

FoxD5_drer              GICDFISNKFPYYKEKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWSLDPASEDMF 157
FoxD5_omos              GICDFISNKFPYYRDKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWSLDPASEDMF 162
xfd12_xlae              GICDFISSKFPYYKDKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWTLDPASEDMF 181
xfd12prime_xlae         GICDFISSKFPYYKDKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWTLDPASEDMF 181
xfd12dblprime_xlae      GICDFISSKFPYYKDKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWTLDPASKDMF 181
FoxD3_drer              GICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPQSEDMF 180
UN_48_tnig              GICEFISSRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTMDPASEDMF 152
FoxD3b_xlae             GICEFISSRFPYYREKFPAWQNSIRHNLSLNDCFIKIPREPGNPGKGNYWTLDPQSEDMF 178
xfd6_xlae               GICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPQSEDMF 178
FOXD3_hsap              GICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPQSEDMF 225
Foxd3_mmus              GICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPQSEDMF 215
FoxD3_ggal              GICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPQSEDMF 201
FoxD2_ggal              EICEFISGRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 207
FoxD2_xlae              EICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 162
FOXD2_hsap              EICEFISGRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 211
Foxd2_mmus              EICEFISGRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 213
FOXD1_hsap              EICEFISGRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 209
Foxd1_mmus              EICEFISSRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 214
FoxD1_ggal              EICEFISGRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLDPESADMF 227
FoxD4_ggor              GICAFISGRFPYYRRKFPAWQNSIHHNLSLNDCFFKIPREPGHPGRGNYWSLDPASQDMF 192
FoxD4_ptro              GICAFISGRFPYYRRKFPAWQNSIHHNLSLNDCFVKIPREPGHPGKGNYWSLDPASQDMF 188
FOXD4L2_hsap            GICAFISGRFPYYRRKFPAWQNSIRHNLSLNDCFVKIPREPGHPGKGNYWSLDPASQDMF 192
FOXD4L3_hsap            GICAFISGRFPYYRRKFPAWQNSIRHNLSLNDCFVKIPREPGHPGKGNYWSLDPASQDMF 192
FOXD4b_hsap             GICAFISGRFPYYRRKFPAWQNSIRHNLSLNDCFVKIPREPGHPGKGNYWSLDPASQDMF 192
```


FoxD5_drer DNGSFLRRRKRFKRNQP-------EFTKDSLVLYHPTLSYRAYGR---------- 195
FoxD5_omos DNGSFLRRRKRFKRNQP-------EFGKDGLMFYSSLNCYRPYGH---------- 200
xfd12_xlae DNGSFLRRRKRFKRHQQ-------EFFKDGLMMYNSLPYYRPY---------- 217
xfd12prime_xlae DNGSFLRRRKRFKRHQQ-------EFFKDGLMMYNSLPYYRPY---------- 217
xfd12dblprime_xlae DNGSFLRRRKRFKRQHQ-------ELFKDGLVMYNPLHYCTPN---------- 217
FoxD3_drer DNGSFLRRRKRFKRHQPDILRDQTALMMQSFGAYGIGNPYGRHYG---------- 225
UN_48_tnig DNGSFLRRRKRFKRVQPDMLRDQTALMMQSFGAYSLGGPYGRHYG---------- 197
FoxD3b_xlae DNGSFLRRRKRFKRQQQDTLREQTALMMQSFGAYSLASPYGRHYG---------- 223
xfd6_xlae DNGSFLRRRKRFKRQQQDSLREQTALMMQSFGAYSLASPYGRHYG---------- 223
FOXD3_hsap DNGSFLRRRKRFKRHQQEHLREQTALMMQSFGAYSLAAAAGAAGPYGRPYGLHP------ 279
Foxd3_mmus DNGSFLRRRKRFKRHQQEHLREQTALMMQSFGAYSLAAAAGAG-----PYGLHP------ 264
FoxD3_ggal DNGSFLRRRKRFKRHQQEHLRDQTALMMQGFGAYGLAGPYGRPYG---------- 246
FoxD2_ggal DNGSFLRRRKRFKR----------HEQPPPHPELLLRAGARSRRLPARLR--------- 248
FoxD2_xlae DNGSFLRRRKRFKR----------QQSNPILR--------DPSSFMPAAFG--------- 195
FOXD2_hsap DNGSFLRRRKRFKRQPLPPPHPHPHPHPELLLRGGAAAAGDPGAFLPGFAA--------- 262
Foxd2_mmus DNGSFLRRRKRFKRQPLPPPHPHPHPHPLLLRGGAAAAGDPGAFLSSFAA--------- 264
FOXD1_hsap DNGSFLRRRKRFKRQPLLPPNAAAAESLLLRGAGAAGGAGDPAAAAALFPPAPPPPPHAY 269
Foxd1_mmus DNGSFLRRRKRFKRQPLLAP-HAAAEALLLRGAGPAAGAGDPGAALFPPPPPPP----AC 269
FoxD1_ggal DNGSFLRRRKRFKRQQLPAP------ELLLRAVDPAAFLPQPPPQPPQQP--------PC 273
FoxD4_ggor DNGSFLRRRKRFKRHQLTPG----AHLPHPFPLPAAHAALHNPRP---------- 233
FoxD4_ptro DNGSFLRRRKRFKRHQLTPG----AHLPHPFPLSAAHAALHNPRP---------- 229
FOXD4L2_hsap DNGSFLRRRKRFKRHQLTPG----AHLPHPFPLPAAHAALHNPRP---------- 233
FOXD4L3_hsap DNGSFLRRRKRFKRHQLTPG----AHLPHPFPLPAAHAALHNPRP---------- 233
FOXD4b_hsap DNGSFLRRRKRFKRHQLTPG----AHLPHPFPLPAAHAALHNPHP---------- 233
FoxD5_drer ----------PYCVSGAVP-AQTNPVGYLPVPDGIMVPPP---------- 224
FoxD5_omos ----------PYSITGQVSPTAASSVRYMPLQESIVMPSS---------- 230
xfd12_xlae ----------SAIQPQPVLQQTSLTCMAIPETLPMSTH---------- 245
xfd12prime_xlae ----------SALQPQPMLQQTPLACMAIPETLSMPTN---------- 245
xfd12dblprime_xlae ----------SALQAQ----QIPMTCLAIPENFAMPNH---------- 241
FoxD3_drer ----------IHPAAYTHPAALQYPYIPP-VGPMLPPAVPLLP---------- 257
UN_48_tnig ----------IHPAAYSHPAALQYPYIPP-VGHMLPPGVPLLP---------- 229
FoxD3b_xlae ----------LHPAAYTHPAALQYPYIPP-VGPMLPPAVPLLP---------- 255
xfd6_xlae ----------LHPAAYTHPAALQYPYIPP-VGHMLPPAVPLLP---------- 255
FOXD3_hsap AAAAGAYSHPAAAAAAAAAAALQYPYALPPVAPVLPPAVPLLP---------- 322
Foxd3_mmus AAAAGAYSHPAAAAAAAAAAALQYPYALPPVAPVLPPAVPLLP---------- 307
FoxD3_ggal ----------LPPGAYPHPAALQYPYIPP-VGPMLPPACPLLP---------- 278
FoxD2_ggal -YGPYGYNYGLQLQGLPPGPPAPPPPPRRRRLRGAFPFSAP---------- 288
FoxD2_xlae -YGPYGYNYGLQLHNYQQHP----------GATFSFQPS---------- 223
FOXD2_hsap -YGAYGYGYGLALPAYGAPPPGPAPHPHPHPHAFAFAAAAAAAPCQLSVPPGRAAAPPPG 321
Foxd2_mmus -YGAYGYGYGLALPAYGAPPPGPAPHPHPHPHAFAFAATAP---CQLSVPPGRAAAPPPG 320
FOXD1_hsap GYGPYGCGYGLQLPPYAPPSALFAAAAAAAAAAAFHPHSP---------- 309
Foxd1_mmus GYGAYGCAYGLQLPPCAPPSALFAAAAAAAAAAFHPHSPP---------- 309
FoxD1_ggal AYGPYGCGYGLQLQPYHPHSALFAFHHPSPPPRQPPAAPAG---------- 314
FoxD4_ggor ---------GPLLGAPAPPQPVPGAYPNTAPGRRPYALLHPHP---------- 267
FoxD4_ptro ---------GPLLGAPAPPQPVPGAYPNTAPGRRPYALLHPHP---------- 263
FOXD4L2_hsap ---------GPLLGAPAPPQPVPGAYPNTAPGRRPYALLHPHP---------- 267
FOXD4L3_hsap ---------GPLLGAPAPPQPVPGAYPNTAPGRRPYALLHPHP---------- 267
FOXD4b_hsap ---------GPLLGAPAPPQPVPGAYPNTAPGRRPYALLHPHP---------- 267


FoxD5_drer ----------FFQYQTMNIKIH-------------------------DAPEIQQRPEHK 248
FoxD5_omos ----------SYHLLPQPLNSHGKCVGPKDFRAQ-------------LCAAEPAEPKSGP 267
xfd12_xlae ----------LAPYPDIKRKVSYP---AQGVHRG-------------FKAQDADNHPNNS 279
xfd12prime_xlae ----------LTPYPDIKRKAHYP---DQGAHRG-------------FEGQDANNHPNKS 279
xfd12dblprime_xlae ----------LVPYPDINITVPCP---DQGVHRV-------------LTAQDVDNHPSNS 275
FoxD3_drer ----------SAELNRKAFSS--QLSPSLQLQLN------------SLS--TASIIKSEP 291
UN_48_tnig ----------SAELNRKAFNS--QLSPSLQLQLN------------SLS--TASMIKSEP 263
FoxD3b_xlae ----------SSELTRKAFSS--QLSPSLQLQLS------------SLSSTAASIIKSEP 291
xfd6_xlae ----------SSELTRKAFSS--QLSPSLQLQLS------------SLSSTAASIIKSEP 291
FOXD3_hsap ----------SGELGRKAAAFGSQLGPGLQLQLNSLGAAAAAAGTAGAAGTTASLIKSEP 372
Foxd3_mmus ----------SGELGRKAAAFGSQLGPSLQLQLNTLGAAAAAAGTAGAAG-TTSLIKSEP 356
FoxD3_ggal ----------SGELSRKAFNA--QLGPSLQLQLS------------SLGAAGSIVKSEP 313
FoxD2_ggal ----------HCPLPVGPPSAASVFSAASGLPSFLGGELNCRKSFYHPQLSPTAL----- 333
FoxD2_xlae ----------HCPLPP-----PASVFSSPTLSPFLGNELSRKS-----LYSQLSP----- 258
FOXD2_hsap PPTASVFAGAGSAPAPAPASGSGPGPGPAGLPAFLGAELGCAKAFYPASLSPPAAGTAAG 381
Foxd2_mmus PPTASVFASAASAPAPAPAPGSGPSP--AGLPAFLGAELGCAKAFYPASLSPPAAGTAAS 378
FOXD1_hsap ----------PPPPPP---HGAAAELARTAFGYRPHPLGAALPGPLPASAAKAGGPGASA 356
Foxd1_mmus ----------PPPPPPPPPPGAAAELARTAFGYRSHALAAALPGPLQAAAVKAGGRGAAA 359
FoxD1_ggal ---APAAALPPPPPPPPPPRRRAPLLPAAELARTPFGYPHPLGPALAASLHAAKPGSGAA 371
FoxD4_ggor ---------LRYLLLSAPAYAGAPKKAEGADLATPAPFPCCSPHLVLSLGRRARVWRRHR 318
FoxD4_ptro ---------LRYLLLSAPAYAGAPKKAEGADLATPAPFPCCSPHLVLSLGRRARVWRRHR 314
FOXD4L2_hsap ---------LRYLLLSAPVYAGAPKKAEGADLATPAPFPCCSPHLVLSLGRRARVWRRHR 318
FOXD4L3_hsap ---------LRYLLLSAPVYAGAPKKAEGADLATPAPFPCCSPHLVLSLGRRARVWRRHR 318
FOXD4b_hsap ---------LRYLLLSARVYAAAPKKAEGADLATPAPFPCCSPHLVLSLGRRARVWRRHR 318
FoxD5_drer TQRCSFSIDSIMAKSTESS----------SKSSAHHLTPDYSFVFP 284
FoxD5_omos QAKCSFSIDSIMSPSSIS----------QHNSNPQLGPHGALGYG 303
xfd12_xlae QSKCSFSIENIMRKPKEPE----------PNIQSFNSHWNYN-HVF 314
xfd12prime_xlae QSKCSFSIENIMKKPKEPE----------PSFPSFNSHWNYNNHLL 315
xfd12dblprime_xlae HSKCSFSIENIMGETKEPE----------KHLTSFNQNWNYN-HLL 310
FoxD3_drer SSRPSFSIENIIGVSSSLR----------AIQTFLRPPVTVQSALLS 328
UN_48_tnig SNRPSFSIENIIGVSSASSSS----------PGAAQAFLRPPVTVQSALLS 304
FoxD3b_xlae SSRPSFSIENIIGVSAASS----------VAPQTFLRPPVTVQSALMS 329
xfd6_xlae SSRPSFSIENIIGVSAASS----------AAPHTFLRPPVTVQSALMS 329
FOXD3_hsap SARPSFSIENIIGGPAAPGGSAVGAG---VAGGTGGSGGGSTAQSFLRPPGTVQSAALM 429
Foxd3_mmus SARPSFSIENIIGASAAPGGSAGGGGSGGGAGGGGGSGGGGGAQSFLRPPGTVQSAALM 416
FoxD3_ggal SSRPSFSIENIIGGPAASSAP----------SAQTFLRPPVTVQSGLVA 352
FoxD2_ggal --EAALLQTLKPDPTPAGTGGTAAATTNPSRPSFSIDNIIGGAVPPPPSTNPSAAPAAPY 391
FoxD2_xlae --TLPILHTLKPDAQ----------SRPSFSIDNIIGGSGSTPSPTSPYTAQPGTH 302
FOXD2_hsap LPTALLRQGLKTDAGGGAGGGGAGAGQRPSFSIDHIMGHGGGGAAPPGAGEGSPGPPFAA 441
Foxd2_mmus LSTALLRQGLKTDAGGGAGGGGAGTGQRPSFSIDHIMGHGGGGAAPPGSGDGSPGSPFAA 438
FOXD1_hsap LARSPFSIESIIGGSLGPAAAAAAAAQAAAAAQASPSPSPVAAPPAPGSSGGGCAAQAAV 416
Foxd1_mmus LARSPFSIESLIGPTRGPAAAGAHVSSGAASGTAP----------GPGGGGCAVQAAA 407
FoxD1_ggal VARSPFSIESIIGGPGPGLGAGPAPGAGGSCASQ---------- 406
FoxD4_ggor EADASLSALRVLCKGSGERVQGLRRVCP----------RPRGATATCSSDHQACCIP 365
FoxD4_ptro EADASLSALRVLCKGSGERVQGLRRVCP----------RPRGATATCSSDHQACCIP 361
FOXD4L2_hsap EADASLSALRVLCKGSGERVQGLRRVCP----------RPRGATATCSSDHQACCIP 365
FOXD4L3_hsap EADASLSALRVLCKGSGERVQGLRRVCP----------RPRGATATCSSDHQACCIP 365
FOXD4b_hsap EADASLSALRVLCKGSGERVQGLRRVCP----------RPRGATATCSSDHQACCIP 365

Figure 2-2. Alignment of the FoxD cluster of sequences.

Conserved and functionally important regions noted in the literature are highlighted. Green: forkhead domain (NCBI Protein database, see Table 2-1 for accession numbers) Red Box: EH1 motif (Sullivan et al. 2001; Copley 2005; Yaklichkin et al. 2007) Purple Box: positively selected sites identified by branch-site analyses


FOXI1_hsap --------MSSFDLP----APSPPRCSPQFPSIGQEPPEMNLYYENF--FHPQGVPSPQR 46
Foxi1_mmus --------MSSFDLP----APSPPRCSPQFPSIGQEPPEMNLYYENF--FHPQGMPSPQR 46
FoxI1_xlae --------MSAFDPQ----AHSPPRCGPQFPSIGQEPPEMNIYCESF--LHPQTMPSPQR 46
FoxI3b_drer --------MTSYESQ----GQSPTRCGPQFLSLGQEPPELSLYSDSY--YPPPSLPSPQR 46
UN_46_tnig --------MSSFEAQ----GQSPPRCGPQFPSLGQEPPELSMYSDCY--YPPPSLPSPQR 46
FoxI3a_drer --------MTSFVPQ----SLSP-----QFHSMGQESQEFSLYGDNF--YSAQHVPSPQQ 41
FoxI2_drer --------MNTIDAQIHSNNNAAVNHLQQLPKSAHETSDMAVYCDNFSVYHQQSLPAAQR 52
FoxI1c_xlae --------MNSIHLPSNQRTSASSLH-QHHPKGAQEASEMAVYCDNFSMYHQQNLHSSQR 51
FoxI1_drer MFLEGERIMNAFGQQPSSQQTSPLQQQDILDMTVYCDSNFSMYQQNLHHHHHHHHHQRPP 60
Foxi2_mmus --------M-SFSTEPP----APAQAGGELDMAGFCDS--------LGSCSVPHGLTRAI 39
FOXI1_hsap --P-SFEGGGEYGATPNPYLWFNGPTMT-PPPYLP-----GPN-ASPFLPQAYGVQRP-L 95
Foxi1_mmus --PTSFEGGGEYGTTPNPYLWFNGPAMT-PPPYLP-----GTN-ASPFLPQAYGMQRQ-L 96
FoxI1_xlae --PSNFETG-DYSTTANPYLWLNGPSIT-PPPYLP-----GSN-SSHFMPQAYGMQRQ-L 95
FoxI3b_drer TNPSSYELGDYAASSPNPYLWFNSPGMN-SAPYLG--GTPGPA-GPSFVPQHYGMQRPYL 102
UN_46_tnig TTPTSYDLNDYATSSPNPYLWFNGSGIN-TSPYLATTGAPGNA-SSPFIPQHYGMQRSYL 104
FoxI3a_drer TLPSAYDFGEYAGQTSNPYLWFNGPGLS-PAPCLT----TG--------PQHYGMAKQYV 88
FoxI2_drer --PAGYGLGDYATP--NPYLWLNGPGVNSSSSYIH-----GNN-SPSFIPPAYGSQRQYL 102
FoxI1c_xlae --APNYGIGDYAPPT-NPYLWLGGPGVSNSSSYLH-----GNN-PTSFMSPSYGSQRQFL 102
FoxI1_drer AHPSSYGLGEYSSPSTNPYLWMNSPGIT-STPYLS-----SPN-GGSYIQSGFGSNQRQF 113
Foxi2_mmus AHPPSYGRTDLSSGRR---LWVNSAALS-PAPYAT-----GPGPAPSYAAATLAVPG-SL 89
FOXI1_hsap LPSVSGLGGSDLGWLPIPSQEELMKLVRPPYSYSALIAMAIHGAPDKRLTLSQIYQYVAD 155
Foxi1_mmus LP-------SDLGWLPIPSQEELMKLVRPPYSYSALIAMAIHGAPDQRLTLSQIYQYVAD 149
FoxI1_xlae LPNMHGLGSSELGWLPIPSQEELMKLVRPPYSYSALIAMAIHGAPDKRLTLSQIYQYVAD 155
FoxI3b_drer GPGPPGGPGGELSWFSMPSQEDLMKLVRPPYSYSALIAMAIHGAPERRLTLSQIYQYVAD 162
UN_46_tnig GPTGPGGPGGELSWFSLPSQEDLMKLVRPPYSYSALIAMAIHGAPDKRLTLSQIYQYVAD 164
FoxI3a_drer GASGIGGSEGAFSWFSLPSQEDLMKLVRPPYSYSALIAMAIHGAPNRRVTLSQIYQYVAD 148
FoxI2_drer SNS-SGFAGPDLGWLSIASQEELLKLVRPPYSYSALIAMAIQNAHEKKLTLSQIYQYVAD 161
FoxI1c_xlae SNS-SSFCGTDLSWLSVASQEELLKVVRPPYSYSALIAMAIQNAPEKKLTLSQIYQYVAE 161
FoxI1_drer LPPPTGFGSADLGWLSISSQQELFKMVRPPYSYSALIAMAIQNAQDKKLTLSQIYQYVAD 173
Foxi2_mmus LGASGGLAGADLAWLSLSGQQELLRLVRPPYSYSALIAMAIQSAPLRRLTLSQIYQYVAG 149
FOXI1_hsap NFPFYNKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 215
Foxi1_mmus NFPFYNKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 209
FoxI1_xlae NFPFYNKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 215
FoxI3b_drer NFPFYNKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 222
UN_46_tnig NFPFYNKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 224
FoxI3a_drer NFPFYNKSKASWQNSIRHNLSLNDCFMKVPRDDSDPGKGNYWTLDPNCEKMFDNGNFRRK 208
FoxI2_drer NFPFYKKSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 221
FoxI1c_xlae NFPFYKRSKAGWQNSIRHNLSLNDCFKKVPRDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 221
FoxI1_drer NFPFYKKSKAGWQNSIRHNLSLNDCFKKVARDEDDPGKGNYWTLDPNCEKMFDNGNFRRK 233
Foxi2_mmus NFPFYKRSKAGWQNSIRHNLSLNDCFKKVPRDENDPGKGNYWTLDPNCEKMFDNGNFRRK 209
FOXI1_hsap RKRKSDVSSS-TASLALEKTESSLPVDSPKTTEPQDILDGASPGGTTSSPEKRPSPPPSG 274
Foxi1_mmus RKRKSDSSSS-TSSLASEKTENGLLASSPKPTEPQEVLDTASPDTTTSSPEKRSSPAPSG 268
FoxI1_xlae RKRKSDVSP--NGQLSSDKPEGSPLSESPTNGEHQDMLGNSSPG-TDDSPEKR-SPPPSI 271
FoxI3b_drer RKRKSDSLPE-KSSSGGNESGDSNGRGSPKSQ-SIDI---------STSPEKGPSPASTG 271
UN_46_tnig RKRKSDTLPNGDGVSGGPESGD-NDRGSPKHNPALNI---------SPPTDRIPSPSSSG 274
FoxI3a_drer RKRKSDSQAE---EEGKGYSGSDSALSSPKNP----------SDSSERGNSPISTD 251
FoxI2_drer RKRRSDSSTG-VSSNTKPEDDRQLAGIKPTDSPHLTGP-ASPDADAATDSHKGASPAGLA 279
FoxI1c_xlae RKRRSDSSSA-EAVTVKGEEGRPALGGKGGESPSMLTP-SSPELEAASDDRKSTSPSGIT 279
FoxI1_drer RKRRADGNAMSVKSEDALKLADTSSLMSASQPSLQNSP-------TSSDPKSSPSPSAEH 286
Foxi2_mmus RRRRGE------TSEAAVPGAS-SPEGTALEP--RGS--------TPQDPQTSPSPSEAT 252

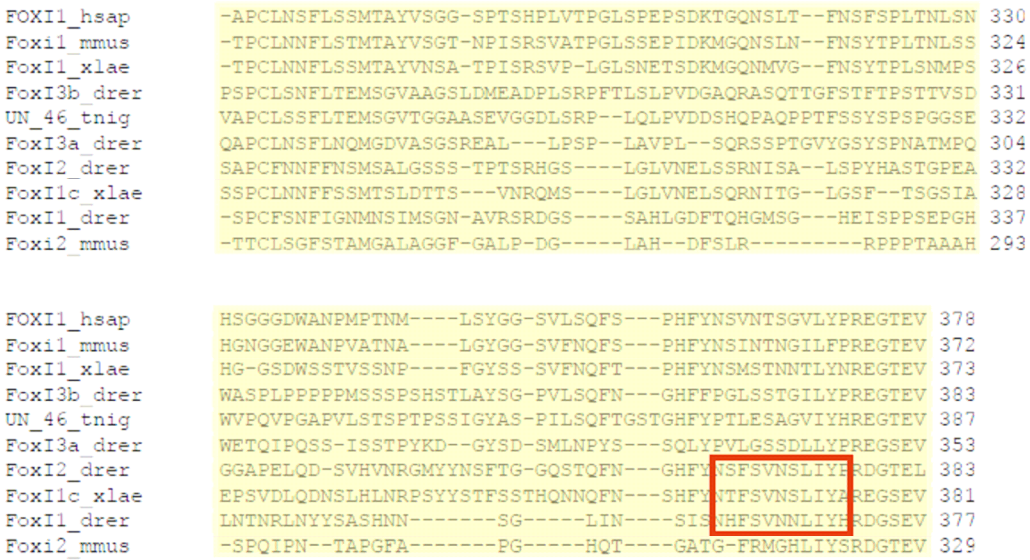
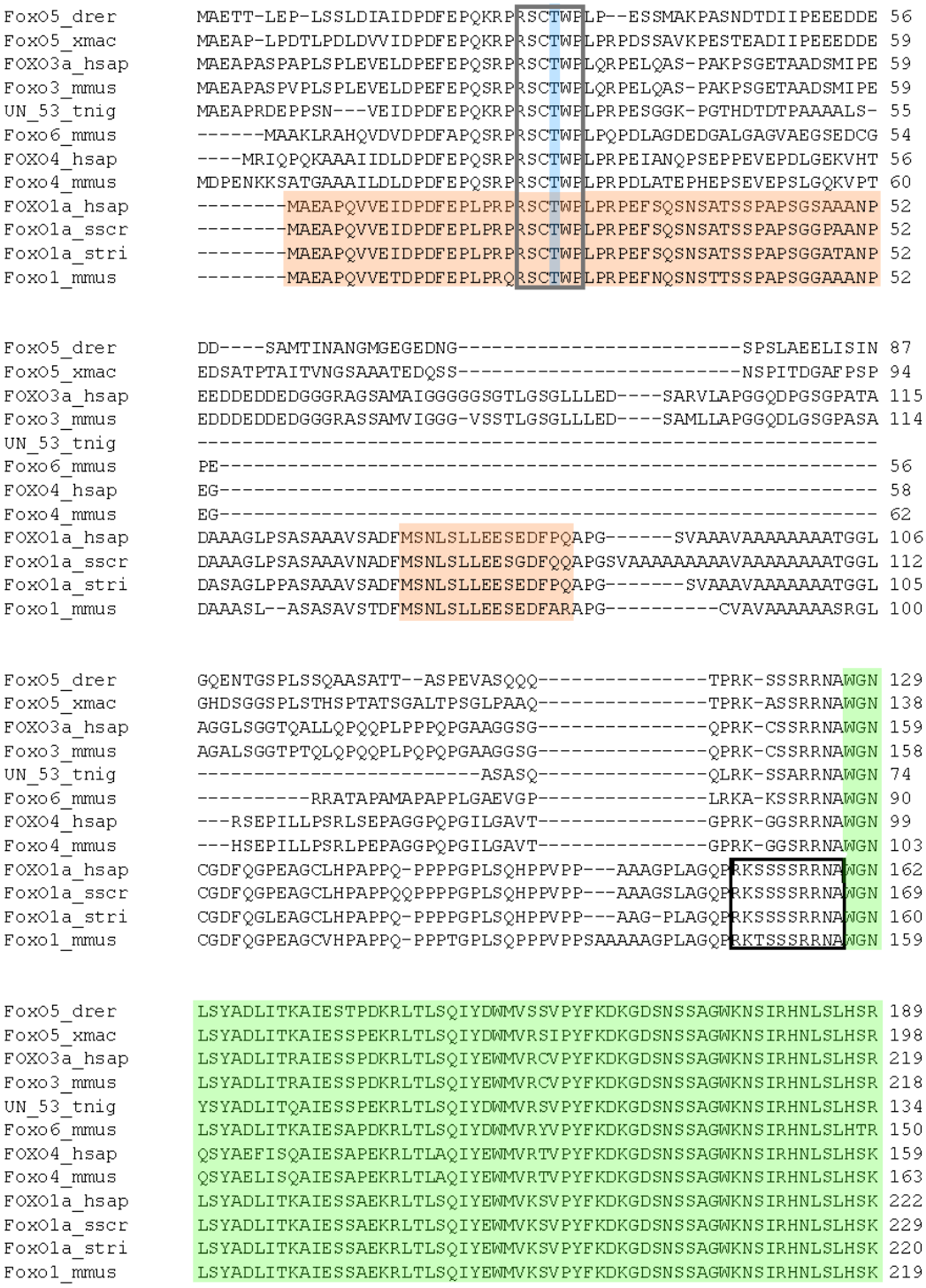
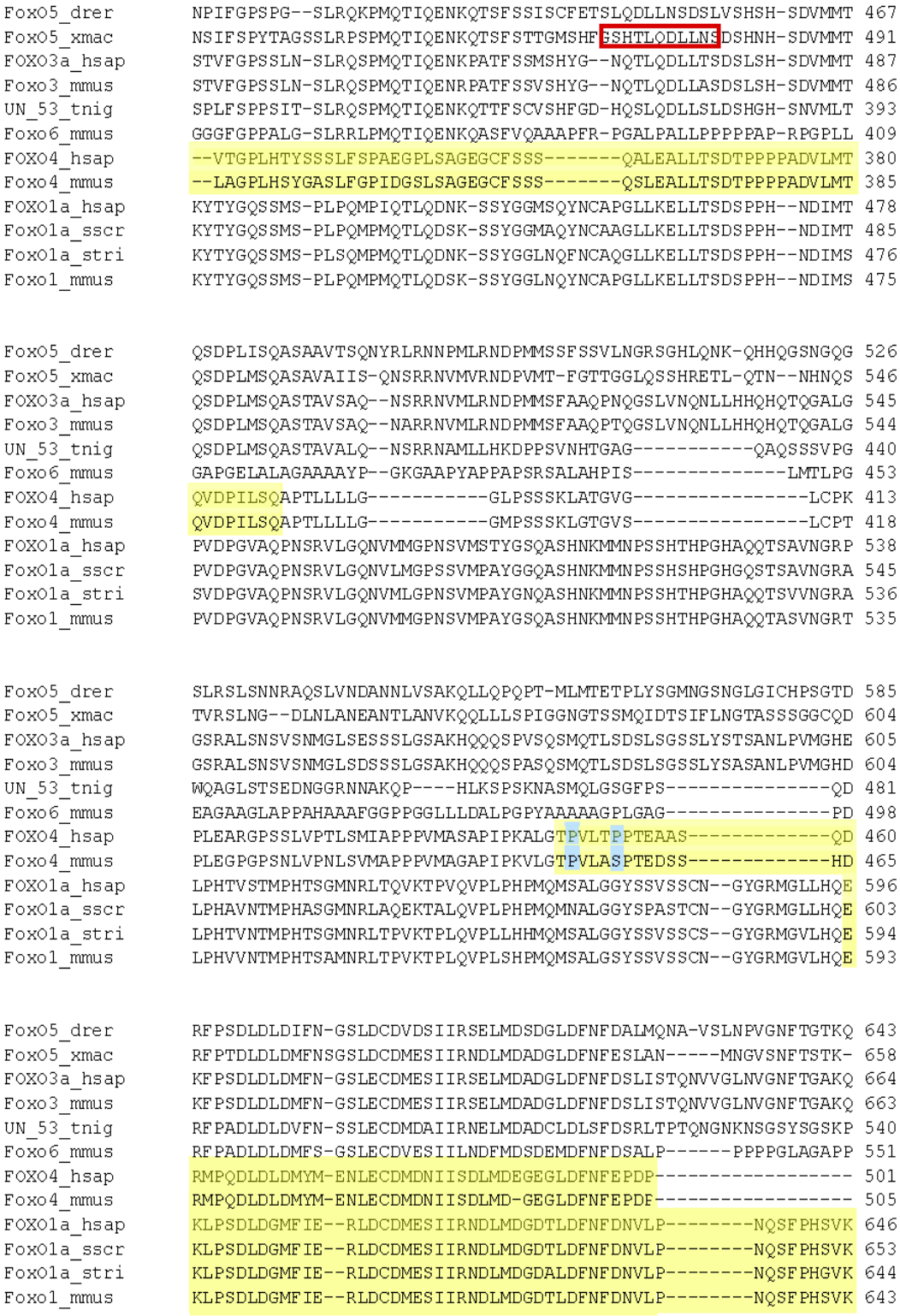


Figure 2-3. Alignment of the FoxI cluster of sequences.

Conserved and functionally important regions noted in the literature are highlighted. Green: forkhead domain (NCBI Protein database, see Table 2-1 for accession numbers) Yellow: transactivation domain (Overdier et al. 1997) Red Box: EH1 motif (Yaklichkin et al. 2007)


Fox05_drer MAETT-LEP-LSSLDIAIDPDFEPQKRPRSCTWPLP--ESSMAKPASNDTDIIPEEEDDE 56
Fox05_xmac MAEAP-LPDTLPDLDVVIDPDFEPQKRPRSCTWPLPRPDSSAVKPESTEADIIPEEEDDE 59
FOXO3a_hsap MAEAPASPAPLSPLEVELDPEFEPQSRPRSCTWPLQRPELQAS-PAKPSGETAADSMIPE 59
Foxo3_mmus MAEAPASPVPLSPLEVELDPEFEPQSRPRSCTWPLQRPELQAS-PAKPSGETAADSMIPE 59
UN_53_tnig MAEAPRDEPPSN---VEIDPDFEPQKRPRSCTWPLPRPESGGK-PGTHDTDTPAAAALS- 55
Foxo6_mmus ------MAAKLRAHQVDVDPDFAPQSRPRSCTWPLPQPDLAGDEDGALGAGVAEGSEDCG 54
FOXO4_hsap ----MRIQPQKAAAIIDLDPDFEPQSRPRSCTWPLPRPEIANQPSEPPEVEPDLGEKVHT 56
Foxo4_mmus MDPENKKSATGAAAILDLDPDFEPQSRPRSCTWPLPRPDLATEPHEPSEVEPSLGQKVPT 60
FOXO1a_hsap --------MAEAPQVVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAPSGSAAANP 52
Fox01a_sscr --------MAEAPQVVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAPSGGPAANP 52
Fox01a_stri --------MAEAPQVVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAPSGGATANP 52
Foxo1_mmus --------MAEAPQVVETDPDFEPLPRQRSCTWPLPRPEFNQSNSTTSSPAPSGGAAANP 52
Fox05_drer DD----SAMTINANGMGEGEDNG-----------------------SPSLAEELISIN 87
Fox05_xmac EDSATPTAITVNGSAAATEDQSS-----------------------NSPITDGAFPSP 94
FOXO3a_hsap EEDDEDDEDGGGRAGSAMAIGGGGGSGTLGSGLLLED----SARVLAPGGQDPGSGPATA 115
Foxo3_mmus EDDDEDDEDGGGRASSAMVIGGG-VSSTLGSGLLLED----SAMLLAPGGQDLGSGPASA 114
UN_53_tnig --
Foxo6_mmus PE-- 56
FOXO4_hsap EG-- 58
Foxo4_mmus EG-- 62
FOXO1a_hsap DAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPG------SVAAAVAAAAAAAATGGL 106
Fox01a_sscr DAAAGLPSASAAAVNADFMSNLSLLEESGDFQQAPGSVAAAAAAAAAVAAAAAAAATGGL 112
Fox01a_stri DASAGLPPASAAAVSADFMSNLSLLEESEDFPQAPG-------SVAAAVAAAAAAATGGL 105
Foxo1_mmus DAAASL--ASASAVSTDFMSNLSLLEESEDFARAPG----------CVAVAAAAAASRGL 100
Fox05_drer GQENTGSPLSSQAASATT--ASPEVASQQQ---------------TPRK-SSSRRNAWGN 129
Fox05_xmac GHDSGGSPLSTHSPTATSGALTPSGLPAAQ---------------TPRK-ASSRRNAWGN 138
FOXO3a_hsap AGGLSGGTQALLQPQQPLPPPQPGAAGGSG---------------QPRK-CSSRRNAWGN 159
Foxo3_mmus AGALSGGTPTQLQPQQPLPQPQPGAAGGSG---------------QPRK-CSSRRNAWGN 158
UN_53_tnig -------------------------ASASQ---------------QLRK-SSARRNAWGN 74
Foxo6_mmus ----------RRATAPAMAPAPPLGAEVGP---------------LRKA-KSSRRNAWGN 90
FOXO4_hsap ---RSEPILLPSRLSEPAGGPQPGILGAVT---------------GPRK-GGSRRNAWGN 99
Foxo4_mmus ---HSEPILLPSRLPEPAGGPQPGILGAVT---------------GPRK-GGSRRNAWGN 103
FOXO1a_hsap CGDFQGPEAGCLHPAPPQ-PPPPGPLSQHPPVPP---AAAGPLAGQPRKSSSSRRNAWGN 162
Fox01a_sscr CGDFQGPEAGCLHPAPPQQPPPPGPLSQHPPVPP---AAAGSLAGQPRKSSSSRRNAWGN 169
Fox01a_stri CGDFQGLEAGCLHPAPPQ-PPPPGPLSQHPPVPP---AAG-PLAGQPRKSSSSRRNAWGN 160
Foxo1_mmus CGDFQGPEAGCVHPAPPQ-PPPTGPLSQPPPVPPSAAAAAGPLAGQPRKTSSSRRNAWGN 159
Fox05_drer LSYADLITKAIESTPDKRLTLSQIYDWMVSSVPYFKDKGDSNSSAGWKNSIRHNLSLHSR 189
Fox05_xmac LSYADLITKAIESSPEKRLTLSQIYDWMVRSIPYFKDKGDSNSSAGWKNSIRHNLSLHSR 198
FOXO3a_hsap LSYADLITRAIESSPDKRLTLSQIYEWMVRCVPYFKDKGDSNSSAGWKNSIRHNLSLHSR 219
Foxo3_mmus LSYADLITRAIESSPDKRLTLSQIYEWMVRCVPYFKDKGDSNSSAGWKNSIRHNLSLHSR 218
UN_53_tnig YSYADLITQAIESSPEKRLTLSQIYDWMVRSVPYFKDKGDSNSSAGWKNSIRHNLSLHSR 134
Foxo6_mmus LSYADLITKAIESAPDKRLTLSQIYDWMVRYVPYFKDKGDSNSSAGWKNSIRHNLSLHTR 150
FOXO4_hsap QSYAEFISQAIESAPEKRLTLAQIYEWMVRTVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 159
Foxo4_mmus QSYAELISQAIESAPEKRLTLAQIYEWMVRTVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 163
FOXO1a_hsap LSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 222
Fox01a_sscr LSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 229
Fox01a_stri LSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 220
Foxo1_mmus LSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLSLHSK 219


FoxO5_drer FIRVQNEGTGKSSWWMINPEGGRGGKAPRRRAVSMDNGNKYTKTARGR-AAKKKAALQAA 248
FoxO5_xmac FIRVQNEGTGKSSWWMINPEGGKGGKAPRRRAVSMDNSNKYTKSARGRAAAKKKAALQAV 258
FOXO3a_hsap FMRVQNEGTGKSSWWIINPDGGKSGKAPRRRAVSMDNSNKYTK-SRGR-AAKKKAALQTA 277
Foxo3_mmus FMRVQNEGTGKSSWWIINPDGGKSGKAPRRRAVSMDNSNKYTK-SRGR-AAKKKAALQAA 276
UN_53_tnig FVKVQNEGTGKSSWWMVNPEGGKGGKAPRRRAVSMDNSKYIKG-ARGR-ATKKKALLQAA 192
Foxo6_mmus FIRVQNEGTGKSSWWMLNPEGGKTGKTPRRRAVSMDNGAKFLR-IKGKASKKKQLHLPER 209
FOXO4_hsap FIKVHNEATGKSSWWMLNPEGGKSGKAPRRRAASMDSSSKLLR-GRSKAPKKKPSVLPAP 218
Foxo4_mmus FIKVHNEATGKSSWWMLNPDGGKGGKAPRRRAASMDSSSKLLR-GRSKGPKKKPSVLPAP 222
FOXO1a_hsap FIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAK-SRSR-AAKKKASLQSG 280
FoxO1a_sscr FIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAK-SRGR-AAKKKASLQSG 287
FoxO1a_stri FIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAK-SRGQ-AAKKKASLQSG 278
Foxo1_mmus FIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAK-SRGR-AAKKKASLQSG 277
QEGSLEN--ISAGGLS---KWPGSPTSRSSDEQESCWTDFRSRTNSNASTVSGRLSPILA 303
AAAAGEGGGDSPTGPS---KWPGSPTSRSSEE-LDAWTDFRSRTNSNASTLSGRLSPILA 314
PESADDS----PSQLS---KWPGSPTSRSSDE-LDAWTDFRSRTNSNASTVSGRLSPIMA 329
PESADDS----PSQLS---KWPGSPTSRSSDE-LDAWTDFRSRTNSNASTVSGRLSPILA 328
QDGSSES----SSSLS---KWTGSPTSRSSDE-LDAWTDFRSRTNSNASTLSGRLSPILA 244
SPDDSPPGAPVPGPLSASAKWAASPASHASDD-YEAWADFRGSRRP--------LLGEAA 260
PEGATPT-----SPVGHFAKWSGSPCSRNREE-ADMWTTFRPRSSSNASSVSTRLSPLRP 272
PEGATPR-----SPLGHFAKWSSSPCPRNREE-ADVWTTFRPRSSSNASTVSTRLSPMRP 276
QEGAGDS------PGSQFSKWPASPGSHSNDD-FDNWSTFRPRTSSNASTISGRLSPIMT 333
QEGAGDS------PGSQFSKWPASPGSHSNDD-FDNWSTFRPRTSSNASTISGRLSPIMT 340
QEGAGDS------PGSQFSKWPASPGSHSNDD-FDNWSTFRPRTSSNASTISGRFSPIMT 331
QEGPGDS------PGSQFSKWPASPGSHSNDD-FDNWSTFRPRTSSNASTISGRLSPIMT 330
NPELDDVSDG-PPLSPILYSSPSSLSPPSSTSSNSKTCPGELPVLADLTGTMNLNE--AD 360
NPELDEVADDEPPLSPMIYSSPGRALSPGNT--NGKAVPTELPRLADLANTMNLNDGITQ 372
STELDEVQDDDAPLSPMLYSSSASLSP-----SVSKPCTVELPRLTDMAGTMNLNDGLTE 384
STELDDVQDDDGPLSPMLYSSSASLSP-----SVSKPCTVELPRLTDMAGTMNLNDGLAE 383
NLELDEVPDDDSPLSPMLYSSPSSMSP-----STGP------TVLSDLAGTMNLNDGLSD 293
ELEDDEALEALAPSSPLMYPSPASALSP----ALGARCPGELPRLAELGGPLGLHGGGVA 316
ESEVLAEEIP-ASVSSYAGGVP----------PTLNEGLELLDGLNL 308
ESEVLAEEEMPASASSYAGGVP----------PTLSEDLELLDGLNL 313
EQDDLGEGDVHSMVYPPSAAKMASTLP----------SLSEISNPENMEN--LL 375
EQDDLGNGDVHSMVYPPSAAKMASTLP----------SLSEISNPENMEN--LL 382
EQDDLGDGDVHSLVYPPSASKMASTLP----------SLSEISNPENMEN--LL 373
EQDDLGDGDVHSLVYPPSAAKMASTLP----------SLSEISNPENMEN--LL 372
NLMVDLLDNI----ATQVQVVQE-SSPFSFGSK--STASPPSTSTSQTVSNNTGG---FS 410
DLMDDFLDNIKLVPTTCSQAMQNGSSGFSFGSKPNGIGSPSSTSSPSSNNSSNGGGNSYS 432
NLMDDLLDNITLPPSQPSPTGGLMQRSSSFPYTTKGSGLGSPTSS----------FN 431
NLMDDLLDNIALPPSQPSPPGGLMQRGSSFPYTAKSSGLGSPTGS----------FN 430
NLMDDLLDNISLTATQQLPPG----EEDSGGQATSVFTFSCSGSS----------LG 336
GLPDALLDGAQDAYGPRARAG----TPSYFGSCKASAYG----------G 352
TSSHSLLSRSGLSGFSLQHPG---------- 329
ASPHSLLSRSGLSGFSLQHPG---------- 334
DNLNLLSSPTSLTVSTQSSPGTMMQQTPCYSFAPPNTSLNSPSPN----------YQ 422
DNLNLLSSPTSLTVSTQSSPGTIMQQTPCYSFAPPNTSLNSPSPN----------YQ 429
DNLNLLSSPTSLTVSTQSSPGSMMQQTPCYSFAPPSTSLNSPSPN----------YQ 420
DNLNLLSSPTSLTVSTQSSPGSMMQQTPCYSFAPPNTSLNSPSPN----------YS 419


FoxO5_drer NPIFGPSPG--SLRQKPMQTIQENKQTSFSSISCFETSLQDLLNSDSLVSHSH-SDVMMT 467
FoxO5_xmac NSIFSPYTAGSSLRPSPMQTIQENKQTSFSTTGMSHFGSHTLQDLLNSDSHNH-SDVMMT 491
FOXO3a_hsap STVFGPSSLN-SLRQSPMQTIQENKPATFSSMSHYG--NQTLQDLLTSDSLSH-SDVMMT 487
Foxo3_mmus STVFGPSSLN-SLRQSPMQTIQENRPATFSSVSHYG--NQTLQDLLASDSLSH-SDVMMT 486
UN_53_tnig SPLFSPPSIT-SLRQSPMQTIQENKQTTFSCVSHFGD-HQSLQDLLSLDSHGH-SNVMLT 393
Foxo6_mmus GGGFGPPALG-SLRRLPMQTIQENKQASFVQAAAPFR-PGALPALLPPPPPAP-RPGPLL 409
FOXO4_hsap --VTGPLHTYSSSLFSPAEGPLSAGEGCFSSS-------QALEALLTSDTPPPPADVLMT 380
Foxo4_mmus --LAGPLHSYGASLFGPIDGSLSAGEGCFSSS-------QSLEALLTSDTPPPPADVLMT 385
FOXO1a_hsap KYTYGQSSMS-PLPQMPIQTLQDNK-SSYGGMSQYNCAPGLLKELLTSDSPPH--NDIMT 478
Foxo1a_sscr KYTYGQSSMS-PLPQMPMQTLQDSK-SSYGGMAQYNCAAGLLKELLTSDSPPH--NDIMT 485
Foxo1a_stri KYTYGQSSMS-PLSQMPMQTLQDNK-SSYGGLNQFNCAQGLLKELLTSDSPPH--NDIMS 476
Foxo1_mmus KYTYGQSSMS-PLPQMPMQTLQDSK-SSYGGLNQYNCAPGLLKELLTSDSPPH--NDIMS 475
FoxO5_drer QSDPLISQASAAVTSQNYRLRNNPMLRNDPMMSSFSSVLNGRSGHLQNK-QHHQGSNGQG 526
FoxO5_xmac QSDPLMSQASAVAIIS-QNSRRNVMVRNDPVMT-FGTTGGLQSSHRETL-QTN--NHNQS 546
FOXO3a_hsap QSDPLMSQASTAVSAQ--NSRRNVMLRNDPMMSFAAQPNQGSLVNQNLLHHQHQTQGALG 545
Foxo3_mmus QSDPLMSQASTAVSAQ--NARRNVMLRNDPMMSFAAQPTQGSLVNQNLLHHQHQTQGALG 544
UN_53_tnig QSDPLMSQASTAVALQ--NSRRNAMLLHKDPPSVNHTGAG-----------QAQSSSVPG 440
Foxo6_mmus GAPGELALAGAAAAYP--GKGAAPYAPPAPSRSALAHPIS--------------LMTLPG 453
FOXO4_hsap QVDPILSQAPTLLLLG-----------GLPSSSKLATGVG----------------LCPK 413
Foxo4_mmus QVDPILSQAPTLLLLG-----------GMPSSSKLGTGVS----------------LCPT 418
FOXO1a_hsap PVDPGVAQPNSRVLGQNVMMGPNSVMSTYGSQASHNKMMNPSSHTHPGHAQQTSAVNGRP 538
Foxo1a_sscr PVDPGVAQPNSRVLGQNVLMGPSSVMPAYGGQASHNKMMNPSSHSHPGHGQSTSAVNGRA 545
Foxo1a_stri SVDPGVAQPNSRVLGQNVMLGPNSVMPAYGNQASHNKMMNPSSHTHPGHAQQTSVVNGRA 536
Foxo1_mmus PVDPGVAQPNSRVLGQNVMMGPNSVMPAYGSQASHNKMMNPSSHTHPGHAQQTASVNGRT 535
FoxO5_drer SLRSLSNNRAQSLVNDANNLVSAKQLLQPQPT-MLMTETPLYSGMNGSNGLGICHPSGTD 585
FoxO5_xmac TVRSLNG--DLNLANEANTLANVKQQLLLSPIGGNGTSSMQIDTSIFLNGTASSSGGCQD 604
FOXO3a_hsap GSRALSNSVSNMGLSESSSLGSAKHQQQSPVSQSMQTLSDSLSGSSLYSTSANLPVMGHE 605
Foxo3_mmus GSRALSNSVSNMGLSDSSSLGSAKHQQQSPASQSMQTLSDSLSGSSLYSASANLPVMGHD 604
UN_53_tnig WQAGLSTSEDNGGRNNAKQP----HLKSPSKNASMQLGSGFPS---------------QD 481
Foxo6_mmus EAGAAGLAPPAHAAAFGGPPGGLLLDALPGPYAAAAAGPLGAG---------------PD 498
FOXO4_hsap PLEARGPSSLVPTLSMIAPPPVMASAPIPKALGTPVLTPPTEAAS-------------QD 460
Foxo4_mmus PLEGPGPSNLVPNLSVMAPPPVMAGAPIPKVLGTPVLASPTEDSS-------------HD 465
FOXO1a_hsap LPHTVSTMPHTSGMNRLTQVKTPVQVPLPHPMQMSALGGYSSVSSCN--GYGRMGLLHQE 596
Foxo1a_sscr LPHAVNTMPHASGMNRLAQEKTALQVPLPHPMQMNALGGYSPASTCN--GYGRMGLLHQE 603
Foxo1a_stri LPHTVNTMPHTSGMNRLTPVKTPLQVPLLHHMQMSALGGYSSVSSCS--GYGRMGVLHQE 594
Foxo1_mmus LPHVVNTMPHTSAMNRLTPVKTPLQVPLSHPMQMSALGSYSSVSSCN--GYGRMGVLHQE 593
FoxO5_drer RFPSDLDLDIFN-GSLDCDVDSIIRSELMDSDGLDFNFDALMQNA-VSLNPVGNFTGTKQ 643
FoxO5_xmac RFPTDLDLDMFNSGSLDCDMESIIRNDLMDADGLDFNFESLAN-----MNGVSNFTSTK- 658
FOXO3a_hsap KFPSDLDLDMFN-GSLECDMESIIRSELMDADGLDFNFDSLISTQNVVGLNVGNFTGAKQ 664
Foxo3_mmus KFPSDLDLDMFN-GSLECDMESIIRSELMDADGLDFNFDSLISTQNVVGLNVGNFTGAKQ 663
UN_53_tnig RFPADLDLDVFN-SSLECDMDAIIRNELMDADCLDLSFDSRLTPTQNGNKNSGSYSGSKP 540
Foxo6_mmus RFPADLDLDMFS-GSLECDVESIILNDFMDSDEMDFNFDSALP------PPPPGLAGAPP 551
FOXO4_hsap RMPQDLDLDMYM-ENLECDMDNIISDLMDEGEGLDFNFEPDP------------------ 501
Foxo4_mmus RMPQDLDLDMYM-ENLECDMDNIISDLMD-GEGLDFNFEPDF------------------ 505
FOXO1a_hsap KLPSDLDGMFIE--RLDCDMESIIRNDLMDGDTLDFNFDNVLP--------NQSFPHSVK 646
Foxo1a_sscr KLPSDLDGMFIE--RLDCDMESIIRNDLMDGDTLDFNFDNVLP--------NQSFPHSVK 653
Foxo1a_stri KLPSDLDGMFIE--RLDCDMESIIRNDLMDGDALDFNFDNVLP--------NQSFPHGVK 644
Foxo1_mmus KLPSDLDGMFIE--RLDCDMESIIRNDLMDGDTLDFNFDNVLP--------NQSFPHSVK 643

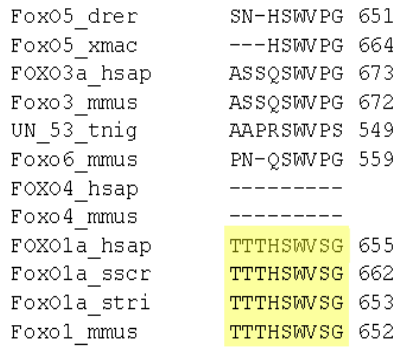
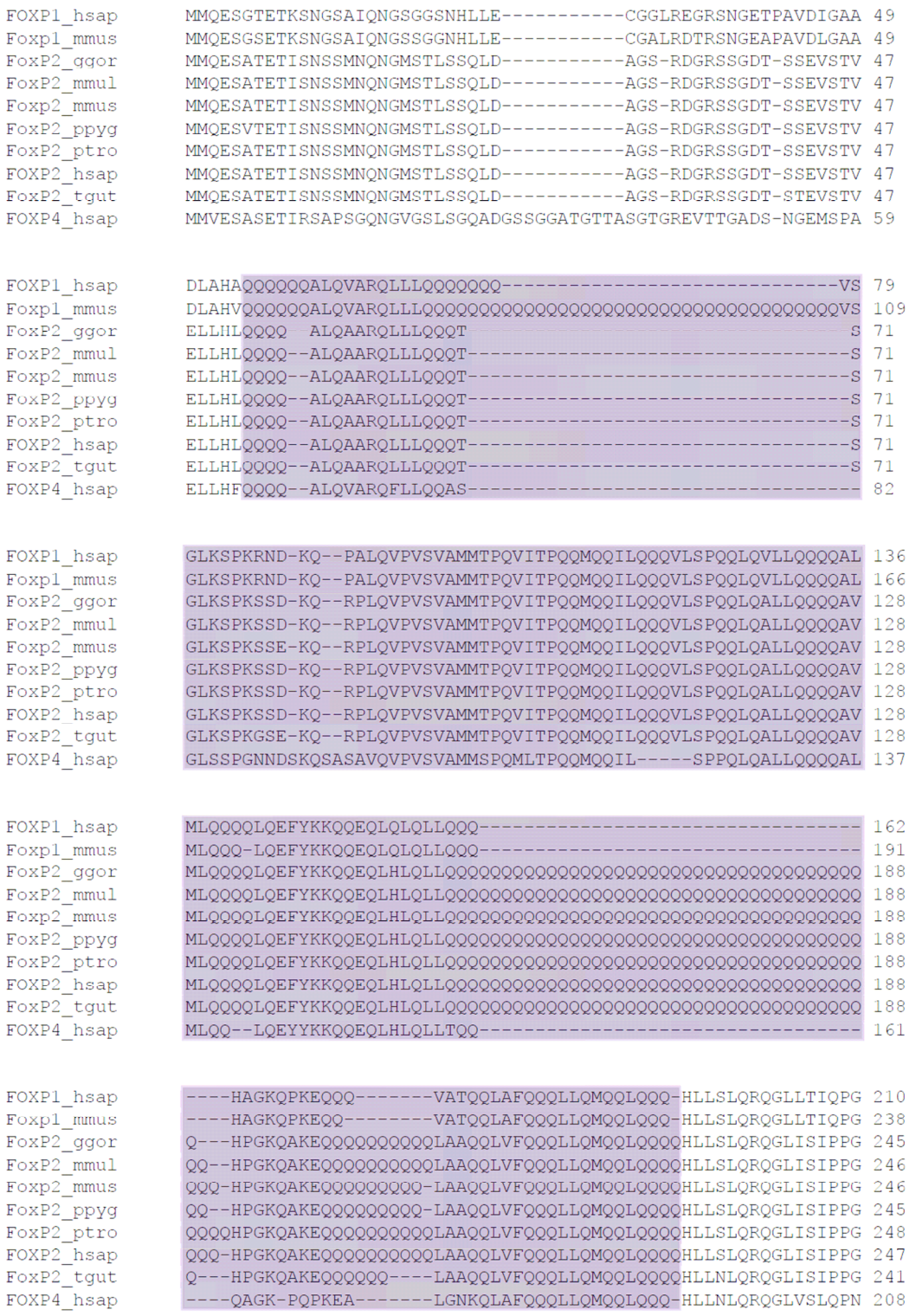


Figure 2-4. Alignment of the FoxO cluster of sequences.

Conserved and functionally important regions noted in the literature are highlighted. Green: forkhead domain (NCBI Protein database, see Table 2-1 for accession numbers) Orange: nuclear export signal (Biggs et al. 1999; Brownawell et al. 2001; Brunet et al. 2002; Zhao et al. 2004) Black Box: nuclear localization signal (Brownawell et al. 2001; Zhang et al. 2002; Zhao et al. 2004) Blue: phosphorylation site (Brunet et al. 1999; Kops et al. 1999; Nakae et al. 1999; Rena et al. 1999; Takaishi et al. 1999; Tang et al. 1999; Brunet et al. 2001; Woods et al. 2001) Grey Box: 14-3-3 binding site (Brunet et al. 1999; Brunet et al. 2001; Rena et al. 2001; Obsil et al. 2003; Zhao et al. 2004) Yellow: transactivation domain (Sublett et al. 1995; So and Cleary 2002) Red Box: EH1 motif Purple Box: positively selected site identified by branch-site analyses

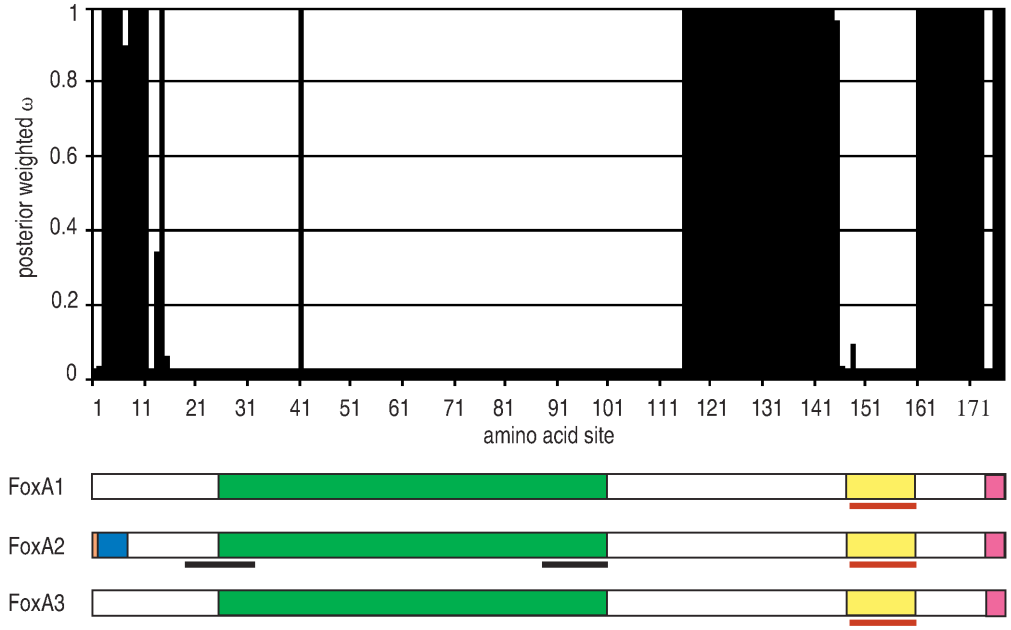

FOXP1_hsap MMQESGTETKSNGSAIQNGSGGSNHLLE----------CGGLREGRSNGETPAVDIGAA 49
Foxp1_mmus MMQESGSETKSNGSAIQNGSSGGNHLLE----------CGALRDTRSNGEAPAVDLGAA 49
FoxP2_ggor MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
FoxP2_mmul MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
Foxp2_mmus MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
FoxP2_ppyg MMQESVTETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
FoxP2_ptro MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
FOXP2_hsap MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-SSEVSTV 47
FoxP2_tgut MMQESATETISNSSMNQNGMSTLSSQLD----------AGS-RDGRSSGDT-STEVSTV 47
FOXP4_hsap MMVESASETIRSAPSGQNGVGSLSGQADGSSGGATGTTASGTGREVTTGADS-NGEMSPA 59
FOXP1_hsap DLAHAQQQQQQALQVARQLLLQQQQQQQ----------VS 79
Foxp1_mmus DLAHVQQQQQQALQVARQLLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQVS 109
FoxP2_ggor ELLHLQQQQ ALQAARQLLLQQQT S 71
FoxP2_mmul ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
Foxp2_mmus ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
FoxP2_ppyg ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
FoxP2_ptro ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
FOXP2_hsap ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
FoxP2_tgut ELLHLQQQQ--ALQAARQLLLQQQT----------S 71
FOXP4_hsap ELLHFQQQQ--ALQVARQFLLQQAS---------- 82
FOXP1_hsap GLKSPKRND-KQ--PALQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQVLLQQQQAL 136
Foxp1_mmus GLKSPKRND-KQ--PALQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQVLLQQQQAL 166
FoxP2_ggor GLKSPKSSD-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FoxP2_mmul GLKSPKSSD-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
Foxp2_mmus GLKSPKSSE-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FoxP2_ppyg GLKSPKSSD-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FoxP2_ptro GLKSPKSSD-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FOXP2_hsap GLKSPKSSD-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FoxP2_tgut GLKSPKGSE-KQ--RPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAV 128
FOXP4_hsap GLSSPGNNDSKQSASAVQVPVSVAMMSPQMLTPQQMQQIL-----SPPQLQALLQQQQAL 137
FOXP1_hsap MLQQQQLQEFYKKQQEQLQLQLLQQQ---------- 162
Foxp1_mmus MLQQQ-LQEFYKKQQEQLQLQLLQQQ---------- 191
FoxP2_ggor MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FoxP2_mmul MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
Foxp2_mmus MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FoxP2_ppyg MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FoxP2_ptro MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FOXP2_hsap MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FoxP2_tgut MLQQQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 188
FOXP4_hsap MLQQ--LQEYYKKQQEQLHLQLLTQQ---------- 161
FOXP1_hsap ----HAGKQPKEQQQ-------VATQQLAFQQQLLQMQQLQQQ-HLLSLQRQGLLTIQPG 210
Foxp1_mmus ----HAGKQPKEQQ--------VATQQLAFQQQLLQMQQLQQQ-HLLSLQRQGLLTIQPG 238
FoxP2_ggor Q---HPGKQAKEQQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 245
FoxP2_mmul QQ--HPGKQAKEQQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 246
Foxp2_mmus QQQ-HPGKQAKEQQQQQQQQQ-LAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 246
FoxP2_ppyg QQ--HPGKQAKEQQQQQQQQQ-LAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 245
FoxP2_ptro QQQQHPGKQAKEQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 248
FOXP2_hsap QQQ-HPGKQAKEQQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPG 247
FoxP2_tgut Q---HPGKQAKEQQQQQQ----LAAQQLVFQQQLLQMQQLQQQQHLLNLQRQGLISIPPG 241
FOXP4_hsap ----QAGK-PQPKEA-------LGNKQLAFQQQLLQMQQLQQQ-HLLNLQRQGLVSLQPN 208


FOXP1_hsap QPALPLQPLAQ-GMIPTELQQLWK-EVTSAHTAEETTGNNHSSLDLTT---TCVSSSAPS 265
Foxp1_mmus QPALPLQPLAQ-GMIPTELQQLWK-EVTSAHTAEETTSSNHSSLDLTS---TCVSSSAPS 293
FoxP2_ggor QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 303
FoxP2_mmul QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 304
Foxp2_mmus QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 304
FoxP2_ppyg QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 303
FoxP2_ptro QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 306
FOXP2_hsap QAALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSNTS 305
FoxP2_tgut QSALPVQSLPQAGLSPAEIQQLWK-EVTGVHSMEDN-GIKHGGLDLTTNNSSSTTSSTTS 299
FOXP4_hsap QASGPLQTLPQ-AVCPTDLPQLWKGEGAPGQPAEDS--VKQEGLDLTGTAATATSFAAPP 265
FOXP1_hsap KTSLIMNPHASTNGQLSVHTPKRESLSHEEHPHSHPLYGHGVCKWPGCEAVCEDFQSFLK 325
Foxp1_mmus KSSLIMNPHASTNGQLSVHTPKRESLSHEEHPHSHPLYGHGVCKWPGCEAVCDDFPAFLK 353
FoxP2_ggor KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 363
FoxP2_mmul KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 364
Foxp2_mmus KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 364
FoxP2_ppyg KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 363
FoxP2_ptro KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 366
FOXP2_hsap KASPPITHHSIVNGQSSVLSARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK 365
FoxP2_tgut KASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESVCEDFGQFLK 359
FOXP4_hsap KVSPPLSHHTLPNGQPTVLTSRRDSSSHEETPGSHPLYGHGECKWPGCETLCEDLGQFIK 325
FOXP1_hsap HLNSEHALDDRSTAQCRVQMQVVQQLELQLAKDKERLQAMMTHLHVKSTEPKAAPQPLNL 385
Foxp1_mmus HLNSEHALDDRSTAQCRVQMQVVQQLELQLAKDKERLQAMMTHLHVKSTEPKAAPQPLNL 413
FoxP2_ggor HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 423
FoxP2_mmul HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 424
Foxp2_mmus HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 424
FoxP2_ppyg HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 423
FoxP2_ptro HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 426
FOXP2_hsap HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 425
FoxP2_tgut HLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNL 419
FOXP4_hsap HLNTEHALDDRSTAQCRVQMQVVQQLEIQLAKESERLQAMMAHLHMRPSEPKPFSQP--- 382
FOXP1_hsap VSSVTLSKSASEASPQSLPHTPTTPTAPLTPVTQGPSVITTTSMHTVGPIRRRYSDKYNV 445
Foxp1_mmus VSSVTLSKSASEASPQSLPHTPTTPTAPLTPVTQGPSVITTTSMHTVGPIRRRYSDKYNV 473
FoxP2_ggor VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 483
FoxP2_mmul VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 484
Foxp2_mmus VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 484
FoxP2_ppyg VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 483
FoxP2_ptro VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 486
FOXP2_hsap VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 485
FoxP2_tgut VSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNI 479
FOXP4_hsap ---VTVS--AADSFPDGLVHPPTSAAAPVTPLR--PPGLGSASLHGGGPARRRSSDKFCS 435
FOXP1_hsap PISSADIAQNQEFYKNAEVRPPFTYASLIRQAILESPEKQLTLNEIYNWFTRMFAYFRRN 505
Foxp1_mmus PISSADIAQNQEFYKNAEVRPPFTYASLIRQAILESPEKQLTLNEIYNWFTRMFAYFRRN 533
FoxP2_ggor PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 542
FoxP2_mmul PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 543
Foxp2_mmus PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 543
FoxP2_ppyg PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 542
FoxP2_ptro PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 545
FOXP2_hsap PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 544
FoxP2_tgut PMSS-EIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRN 538
FOXP4_hsap PISS-ELAQNHEFYKNADVRPPFTYASLIRQAILETPDRQLTLNEIYNWFTRMFAYFRRN 494

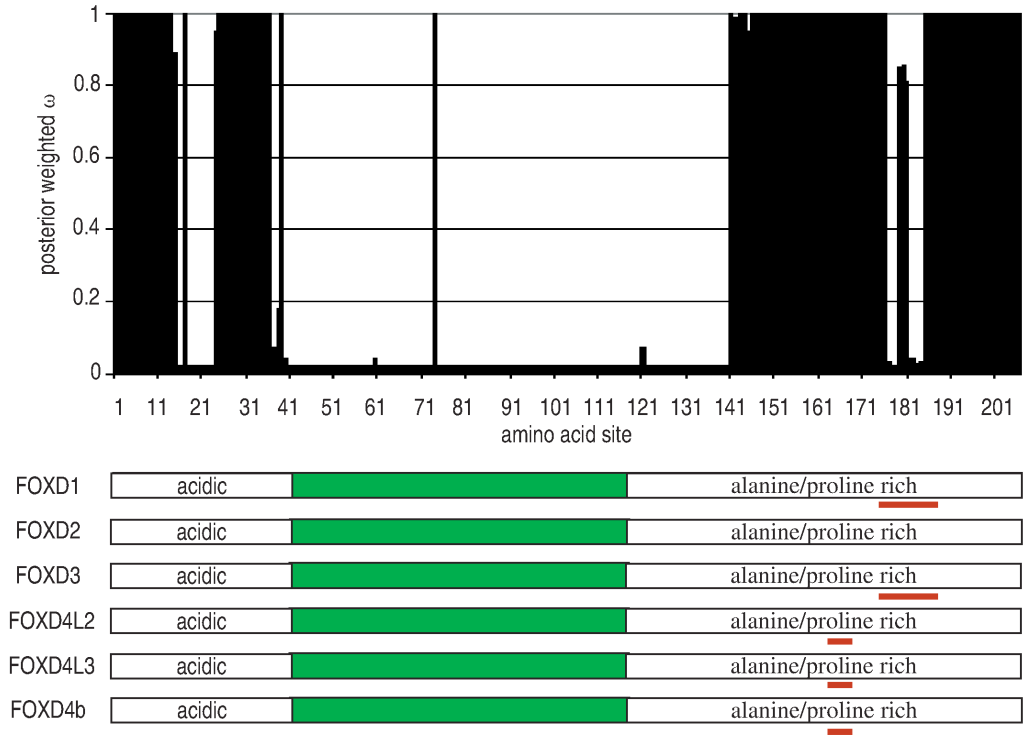
Figure 2-5. Alignment of the FoxP cluster of sequences.

Conserved and functionally important regions noted in the literature are highlighted. Green: forkhead domain (NCBI Protein database, see Table 2-1 for accession numbers) Purple: glutamine (Q) rich region (Banham et al. 2001; Lai et al. 2001; Shu et al. 2001; Teufel et al. 2003; Shi et al. 2004) Blue: zinc finger (Banham et al. 2001; Shu et al. 2001; Teufel et al. 2003; Wang et al. 2003) Yellow: leucine zipper (Banham et al. 2001; Teufel et al. 2003; Wang et al. 2003; Li et al. 2004) Red Box: EH1 motif

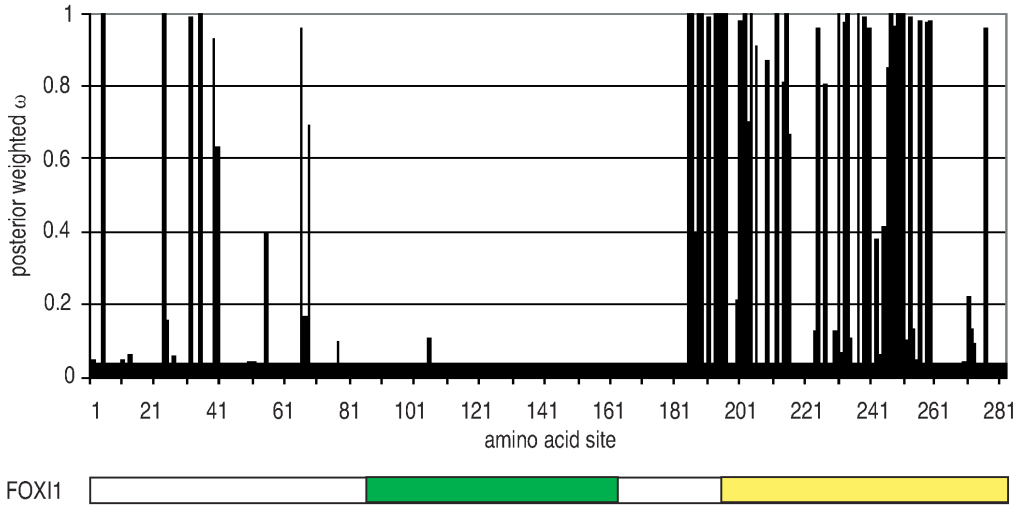
A.


posterior weighted ω
1
0.8
0.6
0.4
0.2
0
1 11 21 31 41 51 61 71 81 91 101 111 121 131 141 151 161 171
amino acid site
FoxA1
FoxA2
FoxA3

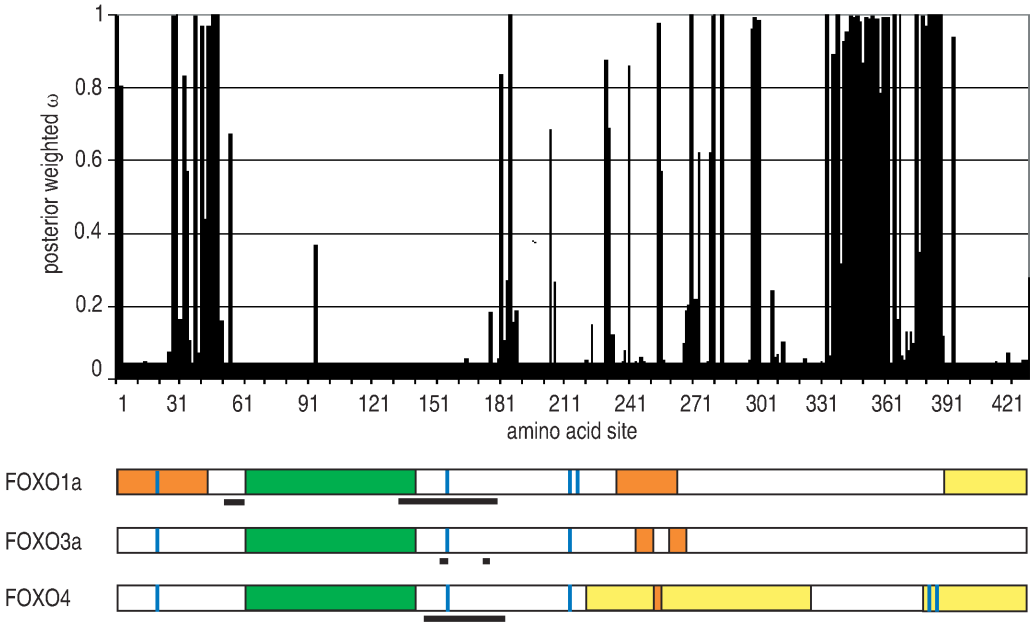
B.


posterior weighted ω
1
0.8
0.6
0.4
0.2
0
1 11 21 31 41 51 61 71 81 91 101 111 121 131 141 151 161 171 181 191 201
amino acid site
FOXD1
FOXD2
FOXD3
FOXD4L2
FOXD4L3
FOXD4b
acidic
alanine/proline rich

C.


posterior weighted ω
1
0.8
0.6
0.4
0.2
0
1 21 41 61 81 101 121 141 161 181 201 221 241 261 281
amino acid site
FOXI1

D.


posterior weighted ω
1
0.8
0.6
0.4
0.2
0
1 31 61 91 121 151 181 211 241 271 301 331 361 391 421
amino acid site
FOXO1a
FOXO3a
FOXO4

E.

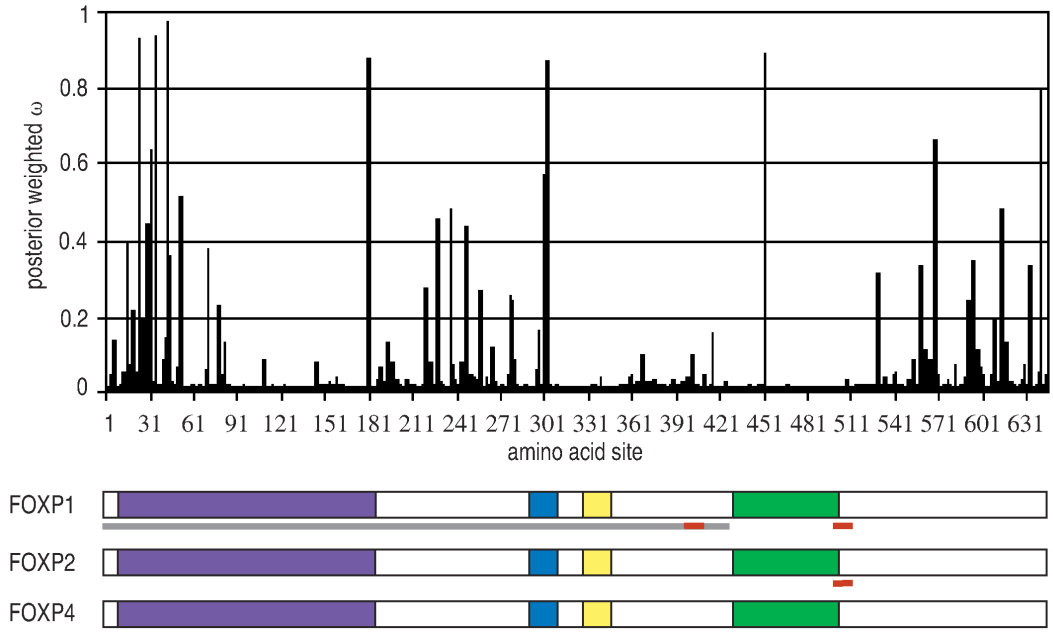


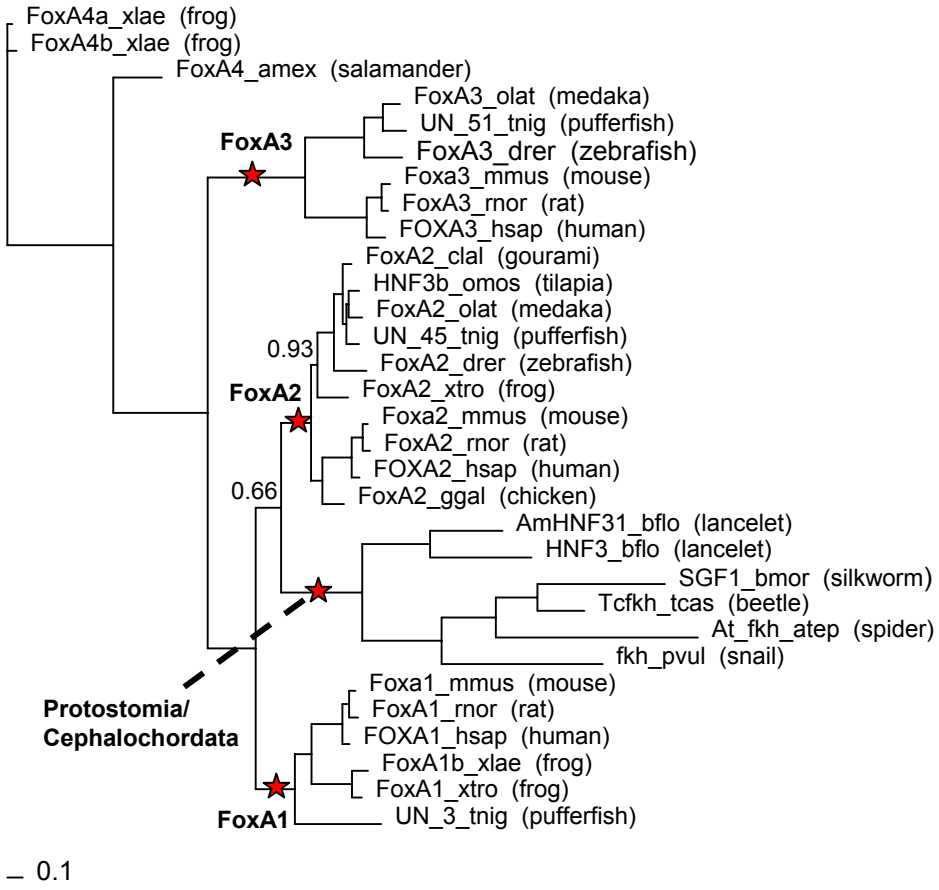
Figure 2-6. Selection pressures on amino acids in each of the five clusters analyzed.

Along the x-axis are the amino acid sites that were analyzed. Since ambiguous sites were removed, the residue numbers along the bottom of the graphs do not correspond to residue numbers of the analyzed sequences. Examination of the alignment and graph of selection pressures simultaneously allows for correlation of selection pressures and actual residue numbers. Underneath each graph is a cartoon of the important regions contained in human forkhead gene(s) within that cluster. Few functional regions have been examined in human FoxA and FoxP proteins therefore functional information identified in rat and mouse protein studies has been included in the FoxA and FoxP figures respectively. The location of the forkhead domain for each human sequence was taken from the NCBI Entrez Protein (Wheeler et al. 2006) database record for that sequence.

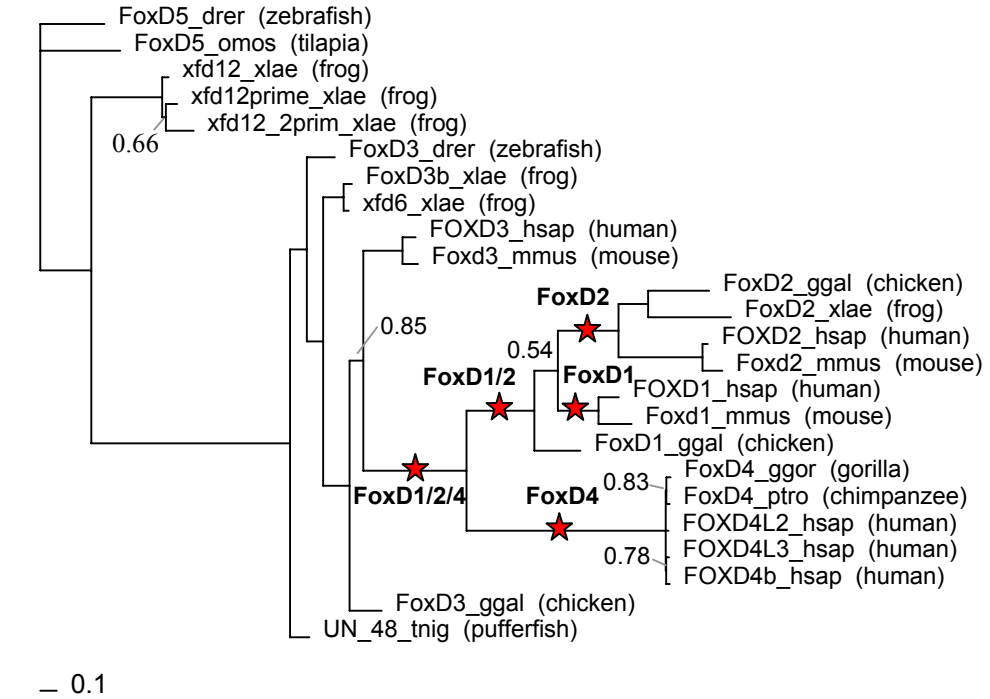
The posterior weighted ω is given on the y-axis. ω < 1 indicates negative selection and ω = 1 indicates neutral change **A.** FoxA cluster of 31 sequences. Peach: conserved domain IV, site 1 (Pani et al. 1992; Qian and Costa 1995) Blue: conserved domain V, sites 2-7 (Pani et al. 1992; Qian and Costa 1995) Green: forkhead domain, sites 25-101 Yellow: conserved domain II, sites 147-160 (Lai et al. 1991) Pink: conserved domain III, sites 174-177 (Lai et al. 1991) Black Bar: nuclear localization signal, sites 19-32 and 88-101 (Qian and Costa 1995) Red Bar: EH1 motif, sites 151-160 (Copley 2005; Yaklichkin et al. 2007) **B.** FoxD cluster of 24 sequences. Green: forkhead domain, sites 42-119 Acidic and alanine/proline rich regions described by Ernstsson et al. 1996; Sutton et al. 1996; Ernstsson et al. 1997; Freyaldenhoven, Fried, and Wielckens 2002. Red Bar: EH1 motif, sites 163-169 for all of the FOXD4s, sites 176-186 for FOXD1 and FOXD3 (Copley 2005; Yaklichkin et al. 2007) **C.** FoxI cluster of 10 sequences. Green: forkhead domain, sites 123-200 Yellow: transactivation domain, sites 196-282 (Overdier et al. 1997) **D.** FoxO cluster of 12 sequences. Green: forkhead domain, sites 61-141 Orange: nuclear export signal, sites 1-43 and 235-264 for FOXO1a, sites 244-252 and 258-266 for FOXO3a, sites 250-256 for FOXO4 (Biggs et al. 1999; Brownawell et al. 2001; Brunet et al. 2002; Zhao et al. 2004) Black Bar: nuclear localization signal, sites 52-60 and 134-180 for FOXO1a, sites 152-154 and 173,174 for FOXO3a, sites 144-183 for FOXO4 (Brownawell et al. 2001; Zhang et al. 2002; Zhao et al. 2004) Yellow: transactivation domain, sites 389-428 for FOXO1a, sites 221-326 and 378-428 for FOXO4 (Sublett et al. 1995; So and Cleary 2002) Blue Bar: phosphorylation

site, sites 20, 157 and 216 for FOXO1a, FOXO3a and FOXO4, additionally site 218 for FOXO1a, sites 379 and 383 for FOXO4 (Brunet et al. 1999; Kops et al. 1999; Nakae et al. 1999; Rena et al. 1999; Takaishi et al. 1999; Tang et al. 1999; Brunet et al. 2001; Woods et al. 2001) **E.** FoxP cluster of 10 sequences. Purple: glutamine rich region, sites 6-182 (Banham et al. 2001; Lai et al. 2001; Shu et al. 2001; Teufel et al. 2003; Shi et al. 2004) Blue: zinc finger, sites 288-311 (Banham et al. 2001; Shu et al. 2001; Teufel et al. 2003; Wang et al. 2003) Yellow: leucine zipper, sites 324-349 (Banham et al. 2001; Teufel et al. 2003; Wang et al. 2003; Li et al. 2004) Green: forkhead domain, sites 434-506 Grey Bar: region involved in repression, sites 1-505 (Shi et al. 2004) Red Bar: EH1 motif , sites 398-408 and 501-511 for FOXP1, sites 501-511 for FOXP2

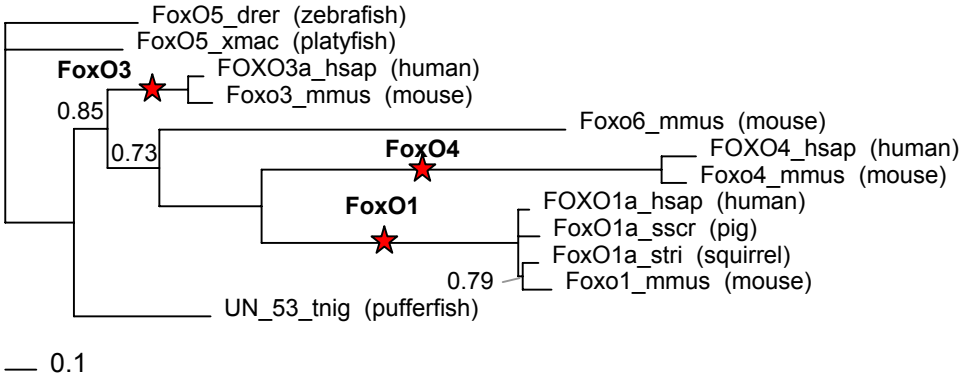
A.


FoxA4a_xlae (frog)
FoxA4b_xlae (frog)
FoxA4_amex (salamander)
FoxA3_olat (medaka)
UN_51_tnig (pufferfish)
FoxA3_drer (zebrafish)
FoxA3
Foxa3_mmus (mouse)
FoxA3_rnor (rat)
FOXA3_hsap (human)
FoxA2_clal (gourami)
HNF3b_omos (tilapia)
FoxA2_olat (medaka)
UN_45_tnig (pufferfish)
0.93
FoxA2_drer (zebrafish)
FoxA2
FoxA2_xtro (frog)
Foxa2_mmus (mouse)
FoxA2_rnor (rat)
FOXA2_hsap (human)
0.66
FoxA2_ggal (chicken)
AmHNF31_bflo (lancelet)
HNF3_bflo (lancelet)
SGF1_bmor (silkworm)
Tcfkh_tcas (beetle)
At_fkh_atep (spider)
fkh_pvul (snail)
Protostomia/
Cephalochordata
Foxa1_mmus (mouse)
FoxA1_rnor (rat)
FOXA1_hsap (human)
FoxA1b_xlae (frog)
FoxA1_xtro (frog)
FoxA1
UN_3_tnig (pufferfish)
0.1

B.


FoxD5_drer (zebrafish)
FoxD5_omos (tilapia)
xfd12_xlae (frog)
xfd12prime_xlae (frog)
xfd12_2prim_xlae (frog)
0.66
FoxD3_drer (zebrafish)
FoxD3b_xlae (frog)
xfd6_xlae (frog)
FOXD3_hsap (human)
Foxd3_mmus (mouse)
FoxD2
FoxD2_ggal (chicken)
FoxD2_xlae (frog)
0.85
FOXD2_hsap (human)
0.54
Foxd2_mmus (mouse)
FoxD1/2
FoxD1
FOXD1_hsap (human)
Foxd1_mmus (mouse)
FoxD1_ggal (chicken)
FoxD4_ggor (gorilla)
0.83
FoxD1/2/4
FoxD4
FoxD4_ptro (chimpanzee)
FOXD4L2_hsap (human)
0.78
FOXD4L3_hsap (human)
FOXD4b_hsap (human)
FoxD3_ggal (chicken)
UN_48_tnig (pufferfish)
0.1

C.


FoxO5_drer (zebrafish)
FoxO5_xmac (platyfish)
FoxO3
FOXO3a_hsap (human)
Foxo3_mmus (mouse)
0.85
Foxo6_mmus (mouse)
0.73
FoxO4
FOXO4_hsap (human)
Foxo4_mmus (mouse)
FoxO1
FOXO1a_hsap (human)
FoxO1a_sscr (pig)
FoxO1a_stri (squirrel)
0.79
Foxo1_mmus (mouse)
UN_53_tnig (pufferfish)
0.1

D.

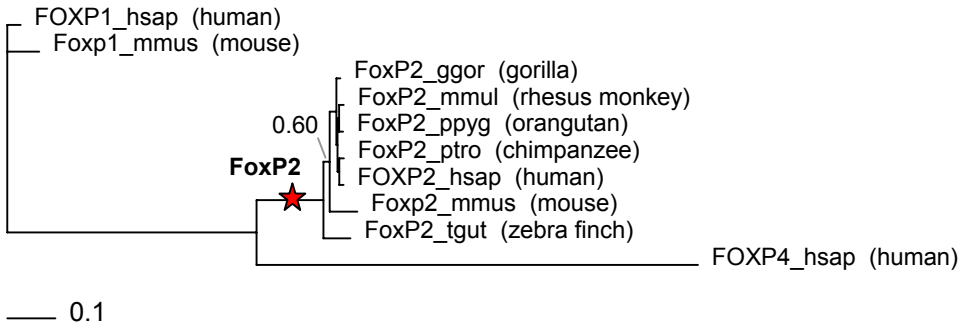


Figure 2-7. Branches tested for positive selection with branch-site models. In each phylogeny the branches tested are indicated with red stars and labels representing the clade of interest. Clade credibility values less than 0.95 are given above the branches. A. FoxA B. FoxD C. FoxO D. FoxP

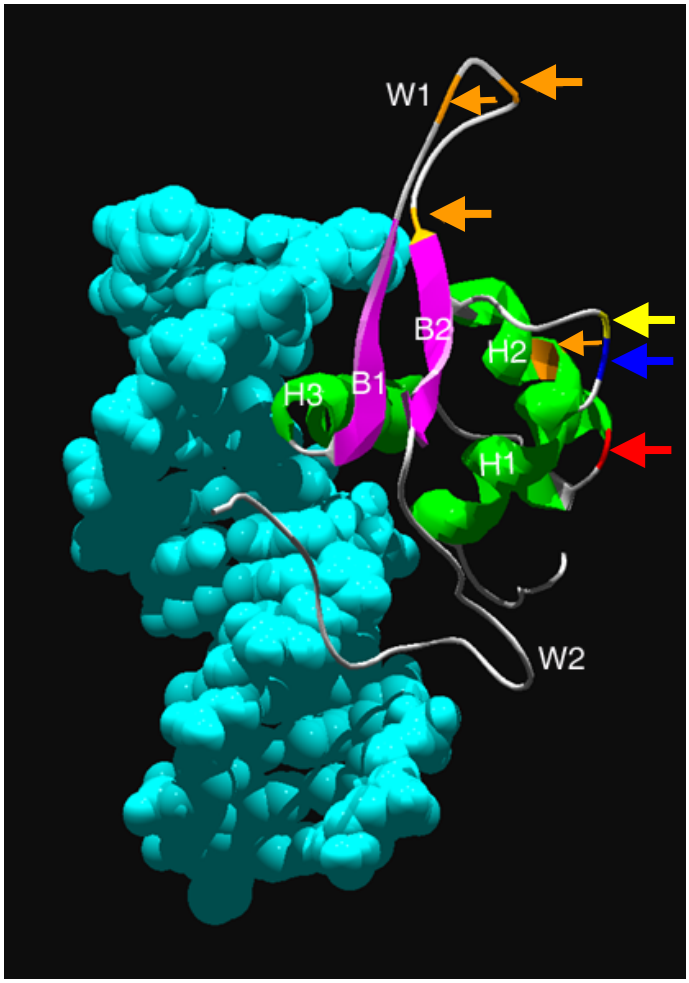


Figure 2-8. Residues in the forkhead domain experiencing neutral changes and positive selection.

.The forkhead domain of FOXA3 is shown as a ribbon model bound to DNA (light blue space filled model) (Clark et al. 1993). Residues with neutral changes identified in the FoxA (dark blue), FoxD (red) and FoxP (yellow) clusters and residues under positive selection identified in the Protostomia/Cephalochordata lineage of the FoxA cluster (orange) are highlighted and indicated with arrows. Alpha helices 1, 2 and 3 (green), beta sheets 1 and 2 (pink), and wings 1 and 2 are denoted H1, H2, H3, B1, B2, W1 and W2 respectively.

References

Adell T, Muller WE (2004) Isolation and characterization of five Fox (Forkhead) genes from the sponge Suberites domuncula. Gene 334:35-46

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585-1592

Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19:950-958

Anisimova M, Yang Z (2007) Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol 24:1219-1228

Banham AH, Beasley N, Campo E, Fernandez PL, Fidler C, Gatter K, Jones M, Mason DY, Prime JE, Trougouboff P, Wood K, Cordell JL (2001) The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p. Cancer Res 61:8820-8829

Biggs WH, 3rd, Meisenhelder J, Hunter T, Cavenee WK, Arden KC (1999) Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. Proc Natl Acad Sci U S A 96:7421-7426

Brent MM, Anand R, Marmorstein R (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. Structure 16:1407-1416

Brownawell AM, Kops GJ, Macara IG, Burgering BM (2001) Inhibition of nuclear import by protein kinase B (Akt) regulates the subcellular distribution and activity of the forkhead transcription factor AFX. Mol Cell Biol 21:3534-3546

Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell 96:857-868

Brunet A, Kanai F, Stehn J, Xu J, Sarbassova D, Frangioni JV, Dalal SN, DeCaprio JA, Greenberg ME, Yaffe MB (2002) 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. J Cell Biol 156:817-828

Brunet A, Park J, Tran H, Hu LS, Hemmings BA, Greenberg ME (2001) Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). Mol Cell Biol 21:952-965

Carlsson P, Mahlapuu M (2002) Forkhead transcription factors: key players in development and metabolism. Dev Biol 250:1-23

Cirillo LA, Zaret KS (2007) Specific interactions of the wing domains of FOXA1 transcription factor with DNA. J Mol Biol 366:720-724

Clark KL, Halay ED, Lai E, Burley SK (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 364:412-420

Clevidence DE, Overdier DG, Tao W, Qian X, Pani L, Lai E, Costa RH (1993) Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family. Proc Natl Acad Sci U S A 90:3948-3952

Cocquet J, De Baere E, Gareil M, Pannetier M, Xia X, Fellous M, Veitia RA (2003) Structure, evolution and expression of the FOXL2 transcription unit. Cytogenet Genome Res 101:206-211

Copley RR (2005) The EH1 motif in metazoan transcription factors. BMC Genomics 6:169

Doolittle RF (1981) Similar amino acid sequences: chance or common ancestry? Science 214:149-159

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872

Fares MA, Bezemer D, Moya A, Marin I (2003) Selection on coding regions determined Hox7 genes evolution. Mol Biol Evol 20:2104-2112

Friedman JR, Kaestner KH (2006) The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci 63:2317-2328

Geer LY, Domrachev M, Lipman DJ, Bryant SH (2002) CDART: Protein homology by domain architecture. Genome Research 12:1619-1623

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7:457-511

Gilbert DG (1999) Readseq version 2, an improved biosequence conversion tool, written in the java language. release 2.93, Bloomington, Indiana

Gotoh O (1996) Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. Journal of Molecular Biology 264:823-838

Hall BG (2004) Improving ClustalX alignments with TuneClustalX. Bellingham Research Institute

Hansen IA, Sieglaff DH, Munro JB, Shiao SH, Cruz J, Lee IW, Heraty JM, Raikhel AS (2007) Forkhead transcription factors regulate mosquito reproduction. Insect Biochem Mol Biol 37:985-997

Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383-402

Kempf BE, Vogt PK (1999) A genetic analysis of PAX3-FKHR, the oncogene of alveolar rhabdomyosarcoma. Cell Growth Differ 10:813-818

Kops GJ, de Ruiter ND, De Vries-Smits AM, Powell DR, Bos JL, Burgering BM (1999) Direct control of the Forkhead transcription factor AFX by protein kinase B. Nature 398:630-634

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523

Lai E, Prezioso VR, Smith E, Litvin O, Costa RH, Darnell JE, Jr. (1990) HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. Genes Dev 4:1427-1436

Lai E, Prezioso VR, Tao WF, Chen WS, Darnell JE, Jr. (1991) Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. Genes Dev 5:416-427

Lam PY, Sublett JE, Hollenbach AD, Roussel MF (1999) The oncogenic potential of the Pax3-FKHR fusion protein requires the Pax3 homeodomain recognition helix but not the Pax3 paired-box DNA binding domain. Mol Cell Biol 19:594-601

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980-996

Letondal C, Schuerer K protal2dna release 2.0, Paris, France

Li S, Weidenfeld J, Morrisey EE (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 24:809-822

Magie CR, Pang K, Martindale MQ (2005) Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. Dev Genes Evol 215:618-630

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz D, I, Jackson JD, Ke Z, Lanczycki CJ, A LC, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2005) CDD: a Conserved Domain Database for protein classification. Nucleic Acids Research 33:D192-D196

Marsden I, Jin C, Liao X (1998) Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. J Mol Biol 278:293-299

Mazet F, Amemiya CT, Shimeld SM (2006) An ancient Fox gene cluster in bilaterian animals. Curr Biol 16:R314-316

Mazumdar A, Kumar R (2003) Estrogen regulation of Pak1 and FKHR pathways in breast cancer cells. FEBS Lett 535:6-10

Nakae J, Park BC, Accili D (1999) Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a Wortmannin-sensitive pathway. J Biol Chem 274:15982-15985

Nichols R (2001) Gene trees and species trees are not the same. Trends Ecol Evol 16:358-364

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936

Nylander JAA (2004) MrModeltest 2.0 release 2.2, Uppsala, Sweden

Obsil T, Ghirlando R, Anderson DE, Hickman AB, Dyda F (2003) Two 14-3-3 binding motifs are required for stable association of Forkhead transcription factor FOXO4 with 14-3-3 proteins and inhibition of DNA binding. Biochemistry 42:15264-15272

Odenthal J, Nusslein-Volhard C (1998) fork head domain genes in zebrafish. Dev Genes Evol 208:245-258

Overdier DG, Ye H, Peterson RS, Clevidence DE, Costa RH (1997) The winged helix transcriptional activator HFH-3 is expressed in the distal tubules of embryonic and adult mouse kidney. J Biol Chem 272:13725-13730

Pani L, Overdier DG, Porcella A, Qian X, Lai E, Costa RH (1992) Hepatocyte nuclear factor 3β contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. Mol Cell Biol 12:3723-3732

Pierrou S, Hellqvist M, Samuelsson L, Enerback S, Carlsson P (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J 13:5002-5012

Qian X, Costa RH (1995) Analysis of hepatocyte nuclear factor-3β protein domains required for transcriptional activation and nuclear targeting. Nucleic Acids Res 23:1184-1191

Rena G, Guo S, Cichy SC, Unterman TG, Cohen P (1999) Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. J Biol Chem 274:17179-17183

Rena G, Prescott AR, Guo S, Cohen P, Unterman TG (2001) Roles of the forkhead in rhabdomyosarcoma (FKHR) phosphorylation sites in regulating 14-3-3 binding, transactivation and nuclear targetting. Biochem J 354:605-612

Rena G, Woods YL, Prescott AR, Peggie M, Unterman TG, Williams MR, Cohen P (2002) Two novel phosphorylation sites on FKHR that are critical for its nuclear exclusion. EMBO J 21:2263-2271

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574

Scheffler K, Seoighe C (2005) A Bayesian model comparison approach to inferring positive selection. Mol Biol Evol 22:2531-2540

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28:231-234

Shi C, Zhang X, Chen Z, Sulaiman K, Feinberg MW, Ballantyne CM, Jain MK, Simon DI (2004) Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. J Clin Invest 114:408-418

Shimeld SM (1997) Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. Dev Biol 183:74-85

Shiyanova T, Liao X (1999) The dissociation rate of a winged helix protein-DNA complex is influenced by non-DNA contact residues. Arch Biochem Biophys 362:356-362

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497

Smith ST, Jaynes JB (1996) A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. Development 122:3141-3150

So CW, Cleary ML (2002) MLL-AFX requires the transcriptional effector domains of AFX to transform myeloid progenitors and transdominantly interfere with forkhead protein function. Mol Cell Biol 22:6542-6552

Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L (2006) Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14:159-166

Sublett JE, Jeon IS, Shapiro DN (1995) The alveolar rhabdomyosarcoma PAX3/FKHR fusion protein is a transcriptional activator. Oncogene 11:545-552

Sullivan SA, Akers L, Moody SA (2001) foxD5a, a Xenopus winged helix gene, maintains an immature neural ectoderm via transcriptional repression that is dependent on the C-terminal domain. Dev Biol 232:439-457

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci U S A 98:2509-2514

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods). release 4.0b10, Sunderland, Massachusetts, USA

Takaishi H, Konishi H, Matsuzaki H, Ono Y, Shirai Y, Saito N, Kitamura T, Ogawa W, Kasuga M, Kikkawa U, Nishizuka Y (1999) Regulation of nuclear translocation of forkhead transcription factor AFX by protein kinase B. Proc Natl Acad Sci U S A 96:11836-11841

Tang ED, Nunez G, Barr FG, Guan KL (1999) Negative regulation of the forkhead transcription factor FKHR by Akt. J Biol Chem 274:16741-16746

Teufel A, Wong EA, Mukhopadhyay M, Malik N, Westphal H (2003) FoxP4, a novel forkhead transcription factor. Biochim Biophys Acta 1627:147-152

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

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

Thompson JD, Plewniak F, Poch O (1999) A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res 27:2682-2690

Tsai KL, Huang CY, Chang CH, Sun YJ, Chuang WJ, Hsiao CD (2006) Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. J Biol Chem 281:17400-17409

Tu Q, Brown CT, Davidson EH, Oliveri P (2006) Sea urchin Forkhead gene family: phylogeny and embryonic expression. Dev Biol 300:49-62

van Dongen MJ, Cederberg A, Carlsson P, Enerback S, Wikstrom M (2000) Solution structure and dynamics of the DNA-binding domain of the adipocyte-transcription factor FREAC-11. J Mol Biol 296:351-359

Vogt G, Etzold T, Argos P (1995) An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. J Mol Biol 249:816-831

Wang B, Lin D, Li C, Tucker P (2003) Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem 278:24259-24268

Wang JC, Waltner-Law M, Yamada K, Osawa H, Stifani S, Granner DK (2000) Transducin-like enhancer of split proteins, the human homologs of Drosophila groucho, interact with hepatic nuclear factor 3beta. J Biol Chem 275:18418-18423

Weigel D, Jackle H (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell 63:455-456

Weigelt J, Climent I, Dahlman-Wright K, Wikstrom M (2000) 1H, 13C and 15N resonance assignments of the DNA binding domain of the human forkhead transcription factor AFX. J Biomol NMR 17:181-182

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, et al. (2006) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 34:D173-180

Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041-1051

Woods YL, Rena G, Morrice N, Barthel A, Becker W, Guo S, Unterman TG, Cohen P (2001) The kinase DYRK1A phosphorylates the transcription factor FKHR at Ser329 in vitro, a novel in vivo phosphorylation site. Biochem J 355:597-607

Wotton KR, Shimeld SM (2006) Comparative genomics of vertebrate Fox cluster loci. BMC Genomics 7:271

Yaklichkin S, Vekker A, Stayrook S, Lewis M, Kessler DS (2007) Prevalence of the EH1 Groucho interaction motif in the metazoan Fox family of transcriptional regulators. BMC Genomics 8:201

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556

Yang Z (2000) Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J Mol Evol 51:423-432

Yang Z (2006) Bonferroni Correction The Genetic Software Forum PAML. Vol. 2007, http://gsf.gc.ucdavis.edu/viewtopic.php?f=1&t=1484&p=2996&hilit=bonferroni+correction&sid=fec34f5a576643cc276fd52c7add4116#p2996

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000a) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449

Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol 15:1600-1611

Yang Z, Swanson WJ, Vacquier VD (2000b) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol 17:1446-1455

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107-1118

Yuan J, Amend A, Borkowski J, DeMarco R, Bailey W, Liu Y, Xie G, Blevins R (1999) MULTICLUSTAL: a systematic method for surveying Clustal W alignment parameters. Bioinformatics 15:862-863

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472-2479

Zhang X, Gan L, Pan H, Guo S, He X, Olson ST, Mesecar A, Adam S, Unterman TG (2002) Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. Direct and indirect effects on nuclear/cytoplasmic shuttling and DNA binding. J Biol Chem 277:45276-45284

Zhao X, Gan L, Pan H, Kan D, Majeski M, Adam SA, Unterman TG (2004) Multiple elements regulate nuclear/cytoplasmic shuttling of FOXO1: characterization of phosphorylation- and 14-3-3-dependent and -independent mechanisms. Biochem J 378:839-849

Chapter 3: *In silico* analysis of selection pressures acting at the molecular level in the FoxC subfamily and biological verification of positive selection.

Introduction

The FoxC forkhead subfamily has originated in the Eumetazoa (Larroux et al. 2008). Paralogs FoxC1 and FoxC2 have only been identified in vertebrates and are found in separate chromosomal Fox gene clusters (Wotton and Shimeld 2006). The gross evolution of these gene clusters has been defined (Mazet et al. 2006; Wotton and Shimeld 2006; Wotton et al. 2008) however the molecular evolution of FoxC1 and FoxC2 has not been examined. Both genes are required for normal development, adult function and when mutated cause disease (Winnier et al. 1997; Kume et al. 1998; Cederberg et al. 2001; Connell et al. 2008; Tumer and Bach-Holm 2009). As discussed in Chapter 1, there are unknown factors that influence differential target selection by FOXC1 and FOXC2. A domain important for regulation of transactivation activity, the inhibitory domain has been identified in human FOXC1 and mouse Foxc1 and Foxc2 (Figure 3-1) (Berry et al. 2002; Fujita et al. 2006). However, the functional mechanisms of this domain have not been fully determined. Analyses of disease causing mutations in FOXC1 and FOXC2 have predicted that too much and demonstrated that too little transactivation activity can lead to disease therefore tight regulation of their function is essential for normalcy (Berry et al. 2005; Saleem et al. 2008; Vreeburg et al. 2008).

Here the selection pressures acting on the FoxC sequences in vertebrates were identified to provide insights into the amino acids important for gene function and to characterize the molecular evolution of this forkhead subfamily.

Selection pressures were measured as ω, the nonsynonymous substitution rate divided by the synonymous substitution rate, for each codon in an alignment of FoxC sequences. The majority of sites were under negative selection with positive selection identified at one site. The positively selected amino acid, within the inhibitory domain, was mutated in *FOXC1* and the effects on FOXC1 function were assessed. Altering the positively selected site affects normal FOXC1 function providing biological support for the predicted positive selection. The effects of mutants on transactivation activity were dependant on the target promoter in question. This is evidence supporting the theory that combination of amino acid change in a transcription factor and differences in the target promoters of that transcription factor can drive optimization of gene expression and evolution. Further insights into inhibitory domain function were also obtained as deletion of the inhibitory domain was shown to reduce FOXC1 DNA binding ability.

Materials and Methods

Sequence Data, Alignment and Phylogenetic Analysis

FoxC sequences were obtained from the NCBI Entrez Protein Database and selected for analysis as described in Fetterman *et al.* 2008. A total of 13 sequences from six species were included: *Homo sapiens* (human) FOXC1 (NP_001444) and FOXC2 (NP_005242); *Mus musculus* (mouse) Foxc1 (NP_032618) and Foxc2 (NP_038547); *Gallus gallus* (chicken) FoxC1 (NP_990337) and FoxC2b (NP_990469); *Danio rerio* (zebrafish) foxc1.1 (NP_571803) and foxc1.2 (NP_571804); *Xenopus laevis* (African clawed frog) FoxC1 (AAC99469), FoxC1a (CAB44727), FoxC2a (CAB54143) and FoxC2b (CAB54144); and *Tetraodon nigroviridis* (green spotted pufferfish) UN (unnamed protein product) (CAF98493). The protein sequences were aligned with a combination of CLUSTALX1.83 (Thompson et al. 1997) and CLUSTALW1.81 (Thompson et al. 1994) then converted into nucleotide alignments, utilizing the proteins' corresponding nucleotide sequences from GenBank, with the protal2dna2.0 program (Letondal and Schuerer). MrBayes3.1.1 (Ronquist and Huelsenbeck 2003) was used to create a phylogeny with models chosen by MrModeltest2.2 (Nylander 2004) in conjunction with PAUP4.0b10 (Swofford 2002). Alignment and phylogenetic analysis details can be found in Fetterman *et al.* 2008.

Identification of Selection Pressures

Site models from the codeml program, within the PAML3.15 package (Yang 1997), were used to estimate the selection pressure on every non-ambiguous codon (site) in the alignment. Model M3, that contains three unconstrained categories of ω, is compared to model M0, which allows only one category of ω, to determine if there is evidence for variation in ω among sites. As a test for positive selection, Model M2a, that has three categories of ω: $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 > 1$ and allows for positive selection ($\omega_2 > 1$), is compared to model M1a, with categories $0 < \omega_0 < 1$, $\omega_1 = 1$ which does not allow for positive selection. Models M8 and M7, which both contain 10 categories of ω approximated from β(p,q) while M8 contains an additional category that allows for positive selection, $\omega_s > 1$, are also compared as a test for positive selection. As a third test for positive selection, model M8 is compared to model M8a which differs from model M8 by setting $\omega_s = 1$. Likelihood ratio tests (LRTs) were used to determine statistical significance of the above model comparisons.

Select branches of the phylogeny were tested for positive selection by comparing branch-site model A to a null model A with the codeml program. Branch-site model A allows for positive selection along pre-specified foreground branches, while the background branches are constrained to negative selection. The null model A does not allow for positive selection along any lineage. A LRT, with corrections for multiple testing by Rom's procedure and the Bonferroni correction, was used to determine if there is a significant difference in the fit of the models to the data.

Site and branch-site analysis parameter details can be found in Fetterman *et al.* 2008.

Plasmids and Cell Culture

Site-directed mutagenesis to alter the amino acid at the positively selected site in FOXC1 was performed with the QuickChange Site Directed Mutagenesis kit (Stratagene) with the addition of 5% dimethylsulfoxide to the reaction. The mutations were chosen to vary the charge and size of the amino acid at the positively selected site. A fragment of FOXC1, amino acids 183-397, in pGEMT (Promega) was mutagenized. Successful mutations were confirmed by sequencing and the mutants were subcloned into full length FOXC1 in pcDNA4-Xpress His/Max B (Invitrogen) which has been previously described (Saleem et al. 2001) creating wild type and mutant FOXC1-Xpress. GAL4 DNA binding domain (GAL4-DBD) fusions to a fragment of FOXC1 consisting of amino acids 215-553 (FOXC1(215-553)) were created in pM (Clonetech) using FOXC1 restriction digest fragments from wild type and mutant FOXC1-Xpress. FOXC1-Xpress lacking the inhibitory domain (ID, amino acids 215-366), FOXC1ΔID, has been previously described (Berry et al. 2002). An Xpress vector containing the inhibitory domain of FOXC1, ID-Xpress, was provided by Dr Fred Berry.

Hela (cervical epithelial cells), HTM (human trabecular meshwork) and Cos-7 (African green monkey kidney cells) cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Transfections were performed with Fugene 6 (Roche) or TransIT-LT1 (Mirus) in

a ratio of 3:1, [Fugene or TransIT-LT1]:[DNA]. For immunoblotting and electrophoretic mobility shift assays (EMSA), cells were harvested by scraping, suspended in nuclear lysis buffer (20mM Hepes (pH 7.6), 500mM NaCl, 1.5mM $MgCl_2$, 0.1% Triton-X 100, 20% glycerol, 1mM dithiothreitol (DTT), 1mM phenylmethylsulfonylfluoride, 0.5% protease inhibitor cocktail (Sigma P8340) and sonicated.

Immunofluorescence

Hela and HTM cells were grown on coverslips in six well plates and transfected with 500ng of wild type FOXC1-Xpress or mutant FOXC1-Xpress or empty Xpress vector and grown for 24 hours. Cells were fixed in 2% paraformaldehyde, permeabilized with 0.05% TritonX-100 in phosphate buffered saline (PBS-X) and blocked with 5% bovine serum albumin (BSA) in PBS-X. Primary anti-Xpress mouse antibody (Invitrogen) and secondary anti-mouse Cy3 conjugated antibody (Jackson ImmunoResearch Laboratories) as well as 4',6-diamidino-2-phenylindole (DAPI) were applied and allowed for visualization of FOXC1 localization with respect to the nucleus through fluorescence microscopy. Cy3 signal was not detected in the cells transfected with empty vector.

Immunoblotting

Xpress tagged proteins from whole cell lysates were resolved on a 10% sodium dodecyl sulfate gel and detected by immunoblotting with primary mouse

anti-Xpress antibody (Invitrogen), secondary anti-mouse horseradish peroxidase (HRP) (Jackson ImmunoResearch Laboratories) and visualized through chemiluminescence. GAL4-DBD FOXC1(215-553) fusion proteins were resolved and detected in the same manner with primary mouse anti-GAL4 antibody (Zymed). Detection of endogenous TFIID with primary rabbit anti-TFIID (Santa Cruz) and secondary anti-rabbit HRP (Jackson ImmunoResearch Laboratories) served as a control for overall protein production levels. Expression levels were quantified through densitometry using ImageJ 1.37, and FOXC1-Xpress or GAL4-DBD FOXC1(215-553) expression was normalized by TFIID expression.

Electrophoretic Mobility Shift Assay (EMSA)

60mm plates of Cos-7 cells were transfected with 2µg of wild type FOXC1-Xpress, mutant FOXC1-Xpress or empty Xpress vector and the proteins were harvested 48 hours post transfection. FOXC1-Xpress expression levels were measured by immunoblotting and the amount of FOXC1-Xpress included in further analyses was equalized among extracts. Protein extracts were incubated at room temperature for 1 hour with 1.25mM DTT, 0.3µg sheared salmon sperm DNA, 0.125µg polydIdC (Sigma) and ^{32}P-dCTP labeled double stranded DNA containing a synthetic ideal FOXC1 binding site (forward; 5'-GATCCAAA**GTAAATAAACA**ACAGA, reverse; 5'-GATCTCTGT**TGTTTATTTAC**TTTG). Reactions including only nuclear lysis buffer or empty vector and not FOXC1-Xpress served as negative controls. The

DNA:protein complexes were resolved on a 6% polyacrylamide gel and visualized through autoradiography.

60mm plates of Cos-7 cells were also transfected with 2μg of wild type FOXC1-Xpress, FOXC1ΔID-Xpress, ID-Xpress or empty Xpress vector and the proteins were harvested 48 hours post transfection. Protein expression was detected by immunoblotting and quantified by Bradford assays. To examine the effects of the inhibitory domain on DNA binding in trans, 10μg of FOXC1-Xpress was combined with 10, 20 or 40μg of either ID-Xpress or empty Xpress vector and subjected to EMSA as described above. To measure the inhibitory domain effects on DNA binding in cis, increasing amounts of FOXC1-Xpress or FOXC1ΔID-Xpress were assayed for binding as described above.

Transactivation Assays

Transactivation assays were performed using the Dual Luciferase Reporter Assay System (Promega). 24 well plates of Hela or Cos-7 cells were cotransfected with 100ng of test plasmid (i.e. wild type, mutant or empty vector), 20ng of the pGL3-TK reporter construct and 2ng of pRL-TK control plasmid. Two different reporter constructs that FOXC1 is known to activate were tested with FOXC1-Xpress plasmids in Hela cells; a synthetic construct containing six ideal FOXC1 binding sites (BS) (Pierrou et al. 1994): 6x BS (Saleem et al. 2001) and a biological reporter containing 830 nucleotides of the FOXO1a promoter: FOXO1a (Berry et al. 2008). GAL4-DBD FOXC1(215-553) fusions were tested for activation of a GAL4-DBD responsive reporter, pG5luc (Promega) in Cos-7

cells. To account for potential differences in wild type FOXC1 and mutant FOXC1 production, transfection of 60mm plates of Hela or Cos-7 cells, protein harvest, immunoblotting and densitometry analysis were performed to measure protein production levels concurrent with the transactivation assays. The expression of transfected test protein was normalized to the expression of transfected wild type FOXC1 and this value was used to weight the test protein transactivation results. Each complete transactivation assay measured three separate reactions, experiments were repeated a minimum of two times and the results were pooled to calculate means and standard errors. Statistical significance was measured using two-tailed t-tests with unequal variances at a significance level of $\alpha = 0.05$.

In silico Prediction of FOXC1 Disorder

The amino acid sequence of FOXC1 was inputted into the DISOPRED2 program (Ward et al. 2004) to predict regions of disorder. A false positive threshold of 2% was implemented.

In silico Analyses of Nucleotides Upstream of *FOXO1a*.

Human *FOXO1a* (ENST00000379561) orthologs were identified from the ortholog list in Ensembl, release 57 (Flicek et al. 2010) and are as follows: *Mus musculus* Foxo1 ENSMUST00000053764, *Gallus gallus* FoxO1A ENSGALG00000017034, *Danio rerio* foxo1a ENSDARG00000063540 and

foxo1b ENSDARG00000061549, *Tetraodon nigroviridis* FoxO1.1 ENSTNIT00000022462 and FoxO1.2 ENSTNIT00000014910. Data for *Xenopus laevis* was unavailable. The 8000nt upstream of the translational start site for each FoxO1a gene were assessed for FOXC1 binding sites using Possum (Fu et al. 2004) and a FOXC1 DNA binding site matrix (M00291) from the TRANSFAC database (Matys et al. 2006). Possum parameters were set as follows: score threshold = 5, residue abundance range = 100, assume residue abundances = 1/4 and pseudocount = 0.375. The mmeta program (Blanco et al. 2007) was used to identify and align transcription factor binding motifs in the first 600nt upstream of the translation start site. The position weight matrices from JASPAR 1.0 TOP-50, which is a collection of the 50 most informative matrices in JASPAR 1.0 (Bryne et al. 2008), were used to identify transcription factor binding sites by mmeta. Other mmeta parameters were set as: mapping threshold quality = 0.25, $\alpha = 0.5$, $\lambda = 0.1$, $\mu = 1.0$, $\gamma = -10$, $c = 100$. *Gallus gallus* data was excluded as much of its sequence is unknown in this region.

Results

Selection Pressures Acting on FoxC Codons

The selection pressures acting on individual codons in the FoxC sequences were assessed by LRTs (likelihood ratio tests) of the site models, M0, M1a, M2a, M3, M7, M8 and M8a implemented in codeml. The LRT comparing models M3 and M0 was significant indicating that selection pressures vary across amino acid sites ($P < 0.0001$) (Table 3-1). Model comparisons testing for positive selection, M8 vs. M7 and M8 vs. M8a, were also significant ($P = 0.03$ for both) providing evidence for positive selection (Table 3-1). The M8 vs. M7 LRT determines if there is class of $\omega > 1$, but doesn't test if the class is significantly greater than one. M8 vs. M8a is implemented to determine if the class of $\omega > 1$ is significantly greater than one and is considered less conservative than M8 vs. M7 (Swanson et al. 2003). Model M8 identified one amino acid under positive selection, estimating $\omega = 6.52$ with a probability of 0.95 (Figure 3-2). In contrast, the M2a vs. M1a LRT for positive selection was not significant ($P = 1$) (Table 3-1). The M2a model is more conservative than the M8 model and may not fit the data significantly better than the M1a model if only a few sites are under positive selection (Wong et al. 2004). It is likely that M2a did not detect positive selection here while M8 did as only one site in the FoxC cluster is under positive selection. The remainder of the sites analyzed are all under negative selection (Figure 3-2).

The amino acid located at the positively selected site for each sequence analyzed is shown in the FoxC gene tree context in Figure 3-3. Branch-site

models were applied to determine if positive selection is acting only along the branch separating the Actionopterygii from the Sarcopterygii or the branch separating FoxC1 and FoxC2 sequences as indicated in Figure 3-3. LRTs comparing Model A, which allows for positive selection along the branch specified, to null Model A, which does not allow for positive selection, were insignificant (P = 0.79 and 0.5 respectively) without correcting for multiple tests (Table 3-1). This indicates that the positive selection observed in the site analysis is not acting solely along either of these lineages, rather positive selection is likely acting on multiple lineages.

The *in vitro* Effects of Amino Acid Substitutions at the Positively Selected Site in FOXC1

The positively selected amino acid identified by the *in silico* site analysis is alanine 337 (A337) in FOXC1. A search of the NCBI Entrez SNP database, Build 126 and 131 (Wheeler et al. 2006), revealed that this site is not noted as a SNP in any FoxC sequence. Site-directed mutagenesis was used to mutate A337 to arginine (R), glutamate (E), glycine (G), proline (P), phenylalanine (F) and also to delete (Δ) it from the protein. Western blots of proteins harvested from human trabecular meshwork (HTM), Hela and Cos-7 cells transfected with wild type or mutated *FOXC1* demonstrated that the mutations do not abolish protein expression (Figure 3-4). Immunofluorescence of Hela and HTM cells transfected with wild type and mutant *FOXC1* showed that alterations at the positive site do not affect normal nuclear localization of the protein (Figure 3-4). Electrophoretic

mobility shift assays (EMSAs) utilizing whole cell lysates from mutant and wild type transfected Cos-7 cells and a radioactively labeled probe containing a consensus FOXC1 DNA binding site showed that the mutations do not abolish DNA binding, however DNA binding by the A337G mutant was reduced (Figure 3-5). Transactivation assays comparing mutant and wild type FOXC1 activity on a luciferase reporter gene containing six ideal FOXC1 binding sites (6x BS) demonstrated significantly reduced FOXC1 transactivation function for the proline ($P < 0.0001$), phenylalanine ($P < 0.0001$) and deletion ($P < 0.0001$) mutants and significantly increased activity for the glutamate mutant ($P = 0.03$) (Figure 3-6). Transactivation assays comparing mutant and wild type activity on a luciferase reporter gene containing a fragment of upstream sequence from a known biological target, *FOXO1a* (Berry et al. 2008), showed a significant reduction in transactivation activity for the arginine ($P = 0.004$) and deletion mutants ($P = 0.0009$) (Figure 3-6). A fragment of FOXC1 3' to the forkhead domain consisting of amino acids 215-553, was ligated to the GAL4 DNA binding domain and transactivation assays with A337 wild type and mutant constructs on a GAL4 responsive promoter showed increased transactivation activity of the arginine ($P = 0.0004$), glycine ($P = 0.03$), proline ($P = 0.05$) and deletion ($P = 0.0006$) mutants with respect to wild type (Figure 3-7).

The Predicted Disorder of FOXC1

Intrinsic disorder of a protein or region of a protein is the lack of constant secondary/tertiary structure. Transcription factors are more likely to contain

regions of disorder than proteins involved in catalytic cellular functions (Iakoucheva et al. 2002; Ward et al. 2004; Garza et al. 2009). These regions of disorder are thought to allow transcription factors to interact with different targets and/or other regulatory proteins. Regions of disorder were predicted in FOXC1 with DISOPRED2. The forkhead domain was correctly predicted to be ordered, while the remainder of the protein is predicted to be disordered.

(un)Conservation of the FOXO1a Upstream Region

Examination of 8000 nucleotides upstream of the translation start site of FoxO1a in human, mouse, chicken, frog and pufferfish using Possum and a FOXC1 DNA binding site matrix, identified a number of potential FOXC1 binding sites in all of the sequences analyzed (Figure 3-8). At least one perfect binding site was found in mouse, frog and pufferfish (FoxO1.2) and at least one almost perfect site, which differed by only one nucleotide at a known variable amino acid, was found in all sequences.

Transcription factor binding motifs, identified with JASPAR 1.0 TOP-50, in the first 600 nucleotides 5' to the translation start site of *FoxO1a*, in the sequences for which data was available, were aligned with mmeta. There were no common alignable elements among the human, mouse, zebrafish, pufferfish *FoxO1a* orthologs. Alignment of only the human and mouse sequences with mmeta identified 15 common transcription factor binding motifs (Table 3-2).

Inhibitory Domain Effects on FOXC1 DNA Binding

EMSAs showed that the DNA binding ability of FOXC1 is not affected by the presence of the inhibitory domain as fragment in trans (Figure 3-9). However, EMSAs comparing wild type FOXC1 binding ability to that of FOXC1 lacking the inhibitory domain (FOXC1ΔID), revealed that at least 10 times more FOXC1ΔID was required to have a band shift intensity that approaches that of wild type FOXC1 (Figure 3-9).

Discussion

Positive Selection is Acting on the FoxC Sequences

Application of site models to predict selection pressures acting on individual codons in an alignment of 13 FoxC sequences resulted in the identification of positive selection at one amino acid site. Positive selection is identified here due to a higher nonsynonymous substitution rate as compared to the synonymous substitution rate at this codon in the alignment. This suggests that the site is evolving to optimize FoxC function. If this site was mistakenly identified as positive selection rather than relaxed selective constraint (neutral changes) altering the amino acid at the positive site is not expected to alter FoxC function. To biologically test this amino acid for positive selection, mutations of the positive site were created in *FOXC1* and the effects on FOXC1 function were determined. The amino acid under positive selection, A337 in FOXC1, is located in the inhibitory domain, a region known to function as a negative regulator of FOXC1 transactivation activity (Berry et al. 2002). Functional analysis of A337 mutations demonstrated that the amino acid composition at the positively selected site affects FOXC1 function. The A337G FOXC1 construct had a reduced DNA binding ability compared to wild type. Transactivation assays using a synthetic (6x BS) and biological (FOXO1a) FOXC1 target revealed that mutation of A337 can alter FOXC1 transactivation activity. Transactivation assays with GAL4-DBD FOXC1(215-553) constructs also revealed differences between wild type and mutant constructs providing evidence that the inhibitory domain is

intrinsically affected. These results provide biological evidence that support the *in silico* identification of positive selection in the FoxC sequences.

It has recently been shown that biased gene conversion (BGC) can result in accelerated evolutionary rates and false detection of positive selection (Berglund et al. 2009; Galtier et al. 2009). BGC is the preferential use of nucleotides G and C over A and T during DNA mismatch repair. This can result in amino acid changes that are not due to natural selection but have accelerated evolutionary rates. If the positive selection observed here is due to BGC, there would be more AT to GC conversions as compared to GC to AT conversions. Figure 3-10 shows the ancestral reconstruction of codons at the positively selected site in the FoxC gene tree context. There are three AT to GC conversions in the FoxC gene tree; node 7 to *FOXC1 Homo sapiens* (T to C), node 10 to *Foxc2 Mus musculus* (A to G) and node 11 to *FoxC2b Xenopus laevis* (A to C). While there are nine GC to AT conversions on the tree; node 4 to *foxc1.2 Danio rerio* (C to T), node 5 to *foxc1.1 Danio rerio* (C to T), node 5 to *UN Tetraodon nigroviridis* (G to T), node 3 to node 6 (C to A and C to T), node 1 to node 2 (C to T), node 9 to node 10 (C to A) and node 9 to node 11 (C to A and C to A). The remaining nucleotide changes at the positive site are G to C or C to G transversions. These observations suggest that positive selection has not been falsely identified due to BGC.

I have identified potential positive selection in FOXC1 and shown that mutations at the positively selected site change protein function, but how would

these changes affect fitness? Characterizations of disease causing missense mutations in FOXC1 have identified mutants that affect nuclear localization, DNA binding, transactivation activity, protein expression, phosphorylation, folding and aggresome formation (Saleem et al. 2001; Saleem et al. 2003a; Saleem et al. 2003b; Murphy et al. 2004; Ito et al. 2007; Fetterman et al. 2009; Ito et al. 2009). Ultimately all missense mutations reduce transactivation activity. Duplications and deletions of *FOXC1* are also disease causing and are thought to respectively increase or reduce FOXC1 transactivation activity (Lehmann et al. 2000; Nishimura et al. 2001; Lehmann et al. 2002). From these mutation analyses a transactivation level in the range of >78% to <150% of wild type activity on the 6x BS reporter is predicted to be normal, while values outside of this range would be disease causing. Here, the A337Δ, A337P and A337F mutants all had transactivation levels that were less than 40% of wild type and A337E transactivation levels were 190% of wild type on the 6x BS reporter (Figure 3-6). Therefore, these amino acid changes at the positively selected site would also likely cause disease. Mutations in FOXC1 primarily cause anterior segment defects of the eye resulting in reduced or lost vision. This would ultimately decrease reproductive fitness through indirect effects on health (e.g. reduced foraging ability leading to inadequate caloric intake) and direct effects on social reproductive value (e.g. being viewed as unhealthy by potential mates). The effects of mutations of the positive site mirror those of disease causing mutation in FOXC1 thus suggesting the positive site has evolved to improve fitness and

providing additional biological evidence to support the identification of positive selection.

The use of *in silico* methods to predict natural selection pressures on individual codons is common (e.g.(Kapralov and Filatov 2007; Premzl and Gamulin 2009; Schwalie and Schultz 2009; Burri et al. 2010; Gomez et al. 2010; Wu et al. 2010; Xu et al. 2010)), however only a few studies have validated these predictions using *in vitro* or *in vivo* systems. Identification of positive selection followed by *in vivo* confirmation of reduced fitness due to mutation of the positively selected sites has only been performed in two studies, both of which compared strains of infectious *E coli* (Chen et al. 2009; Novais et al. 2010). Chen et al. found positive selection in FimH, which is involved in bacterial adhesion to the host. Novais et al. found positive selection in β-lactamase, which is involved in antibiotic resistance. Both studies demonstrated that the amino acids at the sites under positive selection conferred increased pathogenicity to strains experiencing positive selection. Similarly, only two studies of mammalian genes have identified positively selected sites and provided confirmation of altered protein function by mutational analysis of the positively selected sites with *in vitro* systems (Ivarsson et al. 2003; Sawyer et al. 2005). Ivarsson et al. identified sites under positive selection in glutathione transferases and showed that these sites are involved in substrate specificity. Sawyer et al. found positive selection in TRIM5α (tripartite motif containing protein 5 alpha), a gene involved in immune response to retroviruses. The region containing positive selection in TRIM5α was shown to be involved in viral recognition but the role of the

positively selected sites in viral recognition was not examined. For both mammalian studies the effects of the positively selected sites on fitness are unclear. The work presented here is the first complete *in silico* and *in vitro* characterization of an amino acid under positive selection in a transcription factor. I was able to use *in silico* methods to identify potential positive selection and mutational analyses of the positively selected site to show that altering the positive site affects protein function. The observed changes in protein function are the same as those seen in analyses of disease causing mutations of FOXC1, therefore the positively selected site has likely evolved to increase fitness. This work has provided support for the use of *in silico* methods to predict functionally important residues and positive selection that can then be tested in biological systems.

Conservation of the FoxCs Among Orthologs and Paralogs

Negative selection has also played a role in FoxC evolution. Outside of the positively selected site, all other amino acids analyzed are under negative selection (Figure 3-2). This implies that these amino acids are functionally important in all of the sequences analyzed. There were no alignment gaps in the forkhead domain or in a series of residues contiguous to the N- and C-terminal forkhead boundaries. The maintenance of this contiguous residue pattern among orthologs and paralogs suggests that these residues are important for forkhead domain function and should be considered as part of the domain. This is consistent with my previous work in other forkhead subfamilies and supports the

use of these methods for defining functional domain boundaries (Chapter 2 (Fetterman et al. 2008)). Only one fifth of the N-terminal transactivation domain, as defined in FOXC1, was included in the analysis due to gaps in the alignment. However, these gaps are almost exclusively due to a lack of amino acids in the chicken FoxC1 sequence in this region. Visual examination of the alignment shows that this region is conserved among the other sequences analyzed with the exception of a variable length polyalanine tract found only in the FoxC1 sequences (Figure 3-11). This indicates that the domain is functional in all of the sequences analyzed with the exception of chicken FoxC1. The C-terminal transactivation domain, as defined in FOXC1, also contained numerous gaps in the alignment and as a whole is not highly conserved among the sequences analyzed. The residues in this region that are under negative selection may be key to transactivation domain function, while those which are variable may be less important for function or differentiating among the sequences analyzed. These results are also consistent with my previous observations in other forkhead subfamilies (Chapter 2 (Fetterman et al. 2008)). The lack of neutral changes indicates that random change at the sites analyzed is not tolerated and the particular amino acid composition of all of the sites is important for function. Sites that would experience neutral change may not have been included in the analysis due to gaps in the alignment.

The Inhibitory Domain is Disordered and Under Variable Selection Pressures

The inhibitory domain of FOXC1 is a series of 152 amino acids between the forkhead and C-terminal transactivation domains. The inhibitory domain is so named as removal of this region results in increased FOXC1 transactivation activity (Berry et al. 2002). The presence of the inhibitory domain in FOXC1 does not abolish FOXC1 transactivation activity, but it does reduce this activity. An inhibitory domain has also been biologically predicted in mouse Foxc1 and Foxc2 in a region that aligns to that of the inhibitory domain of FOXC1 (Figure 3-1) (Petrova et al. 2004; Fujita et al. 2006). There is no known tertiary structure for the inhibitory domain and analysis of the FOXC1 sequence with DISOPRED2 indicated that the entire region is intrinsically disordered. Disordered regions in transcription factors families have been shown to have lower sequence conservation and higher insertion and deletion rates when compared to structurally conserved domains (Minezaki et al. 2006). Consistent with this observation, the FoxC multiple sequence alignment contained a number of gaps in the alignment in the disordered inhibitory domain but none in the structured forkhead domain (Figure 3-11). The alignment gaps in the inhibitory domain were due to FoxC1 and FoxC2 differences or sequence differences in only one or two of those included in the analysis. Overall, human, mouse and chicken FoxC1 contained approximately 150 amino acids in this region, foxc1.2 in zebrafish contained 102 amino acids and the remaining sequences contained 121-129 amino acids. This indicates that as a whole the domain is not strictly conserved, however the codons included in the site analysis (54% of the FOXC1 inhibitory

domain) were all under negative selection with the exception of the positively selected site. Disordered regions have also been associated with increased evolutionary rates which would aid in facilitating differentiation among gene family members (Brown et al. 2002). However, mutation of disordered regions usually results in loss of disorder suggesting that the maintenance of disordered regions requires directional selection and that these regions are not robust to neutral changes (Schaefer et al. 2010). All together this suggests that here, select amino acids are important for FoxC function in all of the sequences analyzed (sites under negative selection) while others are less functionally constrained and may be differentiating among orthologs and/or paralogs (sites aligned with gaps and positive selection). This is consistent with the *in vitro* analysis results of the positively selected site in FOXC1.

Promoter Differences in FoxC Targets May Drive Positive Selection

Positive selection may be acting to optimize transactivation activity of FoxC's due to target differences between species or paralogs. The transactivation activity of the A337 mutants were not always consistently upregulated, downregulated or equivalent to wild type on the 6x BS and FOXO1a reporters. For example, the A337R mutant transactivation activity was not significantly different from wild type on the 6x BS reporter, but was significantly less than wild type on the FOXO1a reporter. Conversely, the A337P and A337F mutants had significantly less activity than wild type on the 6x BS reporter, but were not significantly different from wild type on the FOXO1a reporter. These results

show that the regulation of transactivation function by the positive site is influenced by the available binding site(s) and/or surrounding sequence. Here the 6x BS reporter contains six ideal (GTAAATAAACA) FOXC1 binding sites while the FOXO1a reporter contains two imperfect FOXC1 sites (GTAAACAAAGT and GCAAACCAGCG). The upstream region of *FoxO1a* in mouse, chicken, frog and pufferfish also contain FOXC1 binding sites. However, the number of perfect and imperfect FOXC1 binding sites and their location vary among species (Figure 3-8). Data for frog was unavailable. Additionally, an attempt to align transcription factor binding motifs in the first 600 nucleotides 5' to the translation start site revealed no common elements among *FoxO1a* orthologs in human, mouse, zebrafish and pufferfish. Chicken FoxO1A data was excluded from analysis as much of its sequence is unknown in this region. In comparison, alignment of only the human and mouse sequences identified 15 common transcription factor binding motifs. The promoter region of a particular gene is not necessarily identical or highly conserved among species (Carninci et al. 2005; Chiba et al. 2008) and these results confirm this for *FoxO1a*. In conjunction with the *in vitro* results, this suggests that for the FoxC sequences positive selection is acting to optimize transactivation activity in response to promoter differences. Comparisons of DNA binding by orthologous transcription factors have shown that the majority of binding events are species specific however, orthologs tend to share binding site specificity and conservation of binding is not correlated to binding site presence therefore other factors are involved in determining DNA occupancy, even for orthologous targets (Odom et al. 2007; Schmidt et al. 2010).

Additionally, actual targets or required transactivation levels of targets may be different among species or paralogs resulting in these promoter region differences and optimization of FoxC transactivation activity in response. Differences in FoxC targets or required transactivation levels of targets among species have not been fully demonstrated to date. However, there is evidence that *ING2*, *NFYB* and *HDGF* are direct FOXC1 targets in human cell lines while foxc1.1/foxc1.2 zebrafish morpholinos did not show differences in expression of these genes in ocular tissue (Berry et al. 2008). Additionally, FoxC2 expression patterns are different in the developing heart of mouse and *Xenopus laevis* supporting the theory that there are species specific requirements for FoxC transactivation activity (Gessert and Kuhl 2009). *Xenopus laevis* also contains two FoxC2 genes that are expressed simultaneously indicating that the level of FoxC2 activity required is different from species expressing only one FoxC2 gene (Koster et al. 2000). With regards to paralogs, FoxC1 and FoxC2 have overlapping as well as distinct expression patterns and targets. For example, mouse Foxc1 and Foxc2 both transactivate *CXCR4* and *Dll4* (Seo et al. 2006; Hayashi and Kume 2008a). However, *CXCR4* is activated to similar levels by both paralogs while *Dll4* levels of activation and control through a FOX binding element differ between the paralogs (Seo et al. 2006; Hayashi and Kume 2008b). The mechanism(s) by which FoxC1 and FoxC2 differentially regulate targets is unknown and the positive site may play a role in directing this regulation.

FoxC Paralog Differences May Drive Positive Selection

The positively selected site is alanine in human FOXC1 and glycine in human FOXC2. Here mutation from alanine to glycine in FOXC1 reduced DNA binding of a FOXC1 target. Interestingly, FOXC1 A337G did not have altered transactivation activity on the 6x BS or FOXO1a reporters suggesting that an increased transactivation activity makes up for a reduced DNA binding ability. The A337G mutant in the GAL4-DBD FOXC1(215-553) construct demonstrated significantly greater transactivation activity than wild type supporting the theory that A337G increases transactivation activity. Since this increase is only visible in the context of the GAL4 DNA binding domain, it also suggests that the A337G effects on DNA binding are specific to the forkhead domain. FOXC2 has also been shown to have a greater transactivation activity on the FOXO1a reporter as compared to FOXC1 (Huang 2009). All together this indicates that the positive site influences FOXC1/FOXC2 functional differentiation. However, in the mouse and chicken sequences, the amino acid at the positive site does not differ between paralogs suggesting that other sequence differences or regulatory mechanisms contribute to paralog differentiation in these species.

Species Differences in FoxC Genome Composition May Drive Positive Selection

Finally, differences in the number of FoxC1 and/or FoxC2 genes among species may cause actual FoxC targets, required FoxC transactivation activities and/or DNA binding abilities to vary among species. Thus, the positively selected site may be adapting inhibitory domain activity in response to these conditions.

FoxC genes originated in the Eumetazoa as they are present in the Bilateria and Cnidaria lineages but not in the Porifera or outside of the Metazoa (Larroux et al. 2008). The FoxC1 and FoxC2 paralogs have only been identified in the vertebrate lineage and may have been created through whole genome or block duplication (Mazet et al. 2006; Wotton and Shimeld 2006). In the species considered here, human, mouse and chicken each have one FoxC1 gene and one FoxC2 gene, while the tetraploidy of the African clawed frog results in two copies of each paralog. Teleosts contain additional FoxC1 paralogs, consistent with an extra round of duplication in this lineage, but have lost FoxC2 sequences (Wotton and Shimeld 2006; Wotton et al. 2008). Here, two foxc1 zebrafish sequences were included but only one FoxC1 sequence from the green spotted pufferfish was initially identified. A 174 amino acid sequence that is 88% identical to the N-terminus of FOXC1 is present in pufferfish but was not included in this analysis due to its short length. There is evidence that human FOXC1 and FOXC2 and zebrafish foxc1.1 and foxc1.2 regulate the transcription of FGF19/fgf19 and FOXO1a/foxO1a.1 and foxO1a.2 ((Tamimi et al. 2006; Berry et al. 2008; Huang 2009). This suggests that the zebrafish foxc1 genes compensate for their loss of Foxc2 genes. Conversely, the African clawed frog contains two FoxC2 genes, both of which are simultaneously expressed (Koster et al. 2000), suggesting that FoxC2 transactivation activity is modified in comparison to species with expression of only one FoxC2 gene. This history in conjunction with the *in vitro* results suggests that positive selection is acting on FoxC genes to

optimize their transactivation function in response to the presence of different FoxC complements in different species.

Positive Selection and the Inhibitory Domain Itself Influence DNA Binding

The positively selected site influences DNA binding. EMSAs demonstrated that the A337G mutation has reduced DNA binding compared to wild type. This is the first time that the inhibitory domain has been shown to play a role in FOXC1 DNA binding. To further investigate if the inhibitory domain affects DNA binding, EMSAs comparing the effects of the inhibitory domain in cis and in trans were performed. To examine the effects in trans, FOXC1 wild type was combined with the inhibitory domain as a fragment or empty vector (as a control) and DNA binding abilities were compared. The addition of the inhibitory domain in trans did not affect DNA binding by FOXC1. To examine the effects of the inhibitory domain in cis, the binding ability of FOXC1 with the inhibitory domain deleted (FOXC1ΔID) was compared to that of wild type FOXC1. FOXC1ΔID demonstrated a decreased binding ability compared to wild type as at least ten times more FOXC1ΔID was required to create band shifts comparable in intensity to wild type FOXC1. The lack of effect on DNA binding when the inhibitory domain is present in trans in conjunction with the effects on DNA binding observed when the inhibitory domain was removed from FOXC1, shows that the overall composition of FOXC1 affects DNA binding and the inhibitory domain plays a role in DNA binding in addition to the forkhead domain. The inhibitory domain is intrinsically disordered and similar to the results here,

disordered regions outside of the DNA binding homeodomain in the HOX proteins Antp, NK2 and Ubx have been shown to affect DNA binding abilities (Liu et al. 2008; Toth-Petroczy et al. 2009). For Antp and NK2, protein conformation, charge and additional DNA contacts mediated by the disordered regions outside of the homeodomain and play a role in DNA binding. The mechanism(s) of disordered region effects on DNA binding of Ubx are unknown but may involved protein conformation, intramolecular interactions or pKa changes. Loss of the inhibitory domain may affect DNA binding due to a number of potential changes to FOXC1 including: loss of overall protein conformation, loss of interaction with other protein(s), loss of additional DNA contacts, altered post-translational modifications, altered charge, etc. Any of these changes may directly or indirectly involve the inhibitory domain. My results show that the inhibitory domain plays a role in FOXC1 DNA binding ability and the positively selected site is specifically involved. Both loss of the inhibitory domain and the A337G mutation reduce FOXC1 DNA binding ability.

Conclusions

The *in silico* and *in vitro* analyses here have provided evidence for positive selection in the FoxC forkhead subfamily. This positive selection may be acting to differentiate or regulate FoxC function between species or paralogs. It is important to recognize that transcription factors have the potential to play powerful roles in speciation through gene regulation. Identification of sites under positive selection in transcription factors provides evidence for differentiation of gene regulation among orthologs and identifies the amino acids that can cause this. This is one of the first reports providing biological evidence for positive selection in a transcription factor.

Advances in genome sequencing and functional analyses of proteins in different species have shown that for many gene families gene function is not grossly different among species and the expansion of families in higher organisms often results in redundancy or complementation rather than the development of unique gene function. These properties are also found in the forkhead gene family. As a result of these observations, differences in cis-regulatory elements of promoters among species is now considered a major factor driving evolutionary change. Of course these observations are quite general and ignore many known differences in gene function among species therefore arguments against the consideration of cis-regulatory elements changes as more important than protein changes during evolution have been made. Here I have provided evidence that a combination of differences in protein and promoter sequence in FoxC proteins

and their targets can drive evolutionary change. To my knowledge, this is the first demonstration that transcription factor amino acid change due to positive selection results in altered transactivation ability and that these alterations are influenced by differences in target promoters.

Mutational analysis of the positive site in FOXC1 has also furthered our knowledge of FOXC1 functional regulation. Analysis of disease causing mutations in FOXC1 has shown that too much or too little transactivation activity can lead to disease therefore regulation of FOXC1 function is essential for normalcy. The inhibitory domain of FOXC1 is known to be critical for regulation of FOXC1 transactivation activity, however the mechanisms of this regulation are unknown. Identification of the positive site as a regulator of FOXC1 transactivation activity is a step in elucidating these mechanisms. Additionally, the inhibitory domain has been shown to play a role in DNA binding for the first time. Concise regulation of FOXC1 activity is imperative for normal development and adult function and this work has increased our knowledge of FOXC1 regulation by the inhibitory domain. It is also the first time that residues outside of the forkhead domain have been shown to influence the DNA binding ability of a forkhead transcription factor.

Tables

Table 3-1. Site and branch-site LRT results.

If the LRT result is greater than the critical value at $\alpha = 0.05$ then the test is statistically significant. The P-value for each test is also given. A P-value < 0.05 is statistically significant and shown in bold.

Models Compared	**LRT Result**	**$\alpha = 0.05$ critical value**	**P-value**
Site Models			
M3 vs. M0	494.906	9.49 χ^2_4	**< 0.0001**
M2a vs. M1a	0	5.99 χ^2_2	1
M8 vs. M7	6.604	5.99 χ^2_2	**0.03**
M8 vs. M8a	3.523	2.71 50:50 χ^2_0 :χ^2_1	**0.03**
Branch-Site Models			
Model A vs. Null Sarcopterygii branch	0.070	3.84 χ^2_1	0.79
Model A vs. Null FoxC2 branch	0.445	3.84 χ^2_1	0.5

Table 3-2. Transcription factors whose binding motifs are found upstream of FoxO1a and whose sites are shared between mouse and human.

Transcription Factor	**Number of Binding Sites**	**General Function***
NRF-2 or GABPA	10	cellular respiration
RREB-1	2	cellular differentiation
HMG-IY or HMGA1	1	multiple, cancer metastasis
REL	1	multiple, oncogene
TEF-1 or TEAD1	1	multiple, muscle twitch

* Data is from the NCBI Gene Database (Wheeler et al. 2006).

Figures

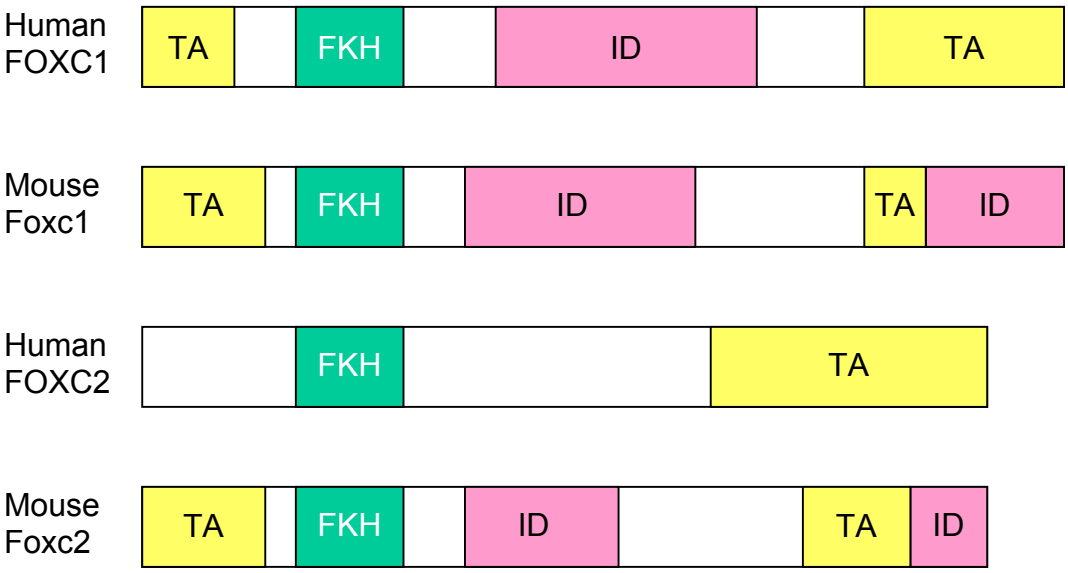


Figure 3-1. Cartoon representation of the experimentally determined functional domains in human and mouse FoxC1 and FoxC2 (Berry et al. 2002; Petrova et al. 2004; Fujita et al. 2006).

Attempts to identify a N-terminal transactivation domain and an inhibitory domain in human FOXC2 have not been made. Yellow: transactivation domain (TA) Green: forkhead domain (FKH) Pink: inhibitory domain (ID)

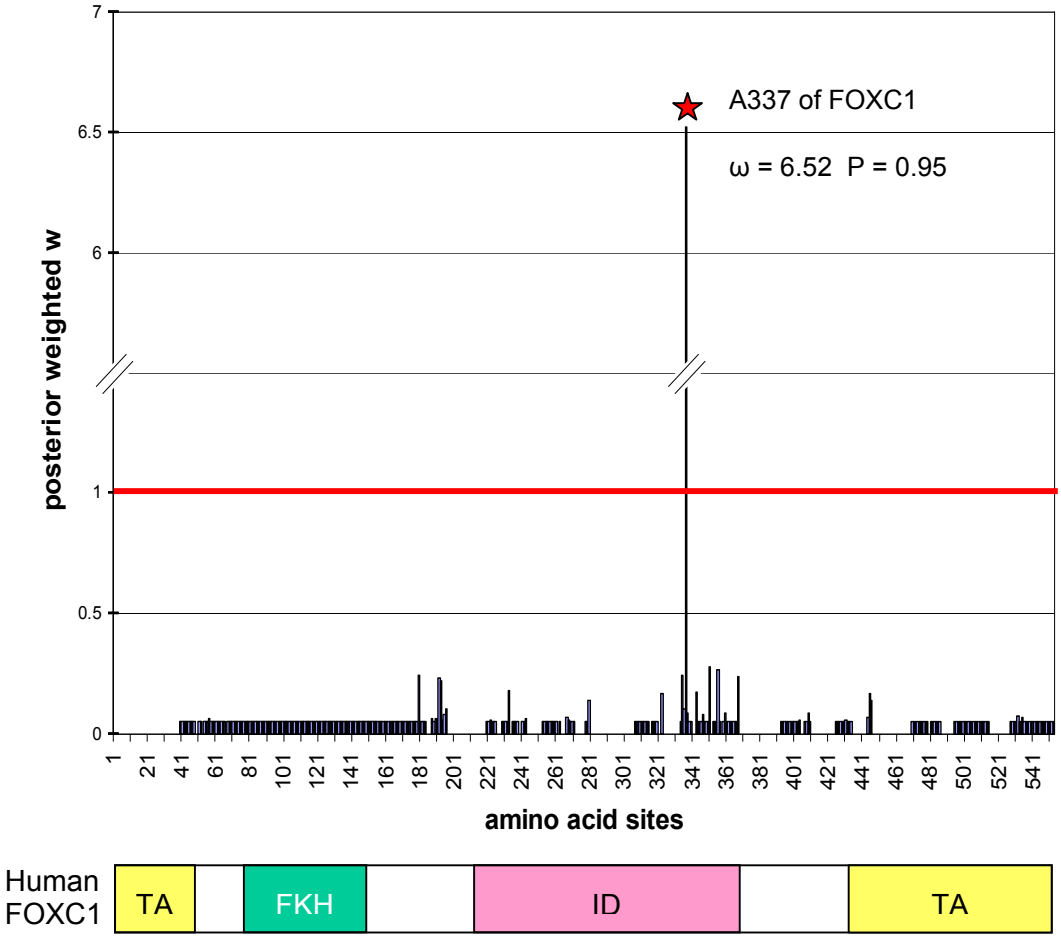


Figure 3-2. Graph of the posterior weighted ω, the mean of ω over the site classes weighted by the probability of each class, for each site as estimated by model M8. A cartoon of the functional domains in human FOXC1 corresponding to the sites analyzed is shown underneath. Gaps in the alignment were not included in the analysis therefore the ω for these amino acids is set to 0. A337 of FOXC1 has ω = 6.5 with a probability of 0.95. Yellow: transactivation domain (TA) Green: forkhead domain (FKH) Pink: inhibitory domain (ID)

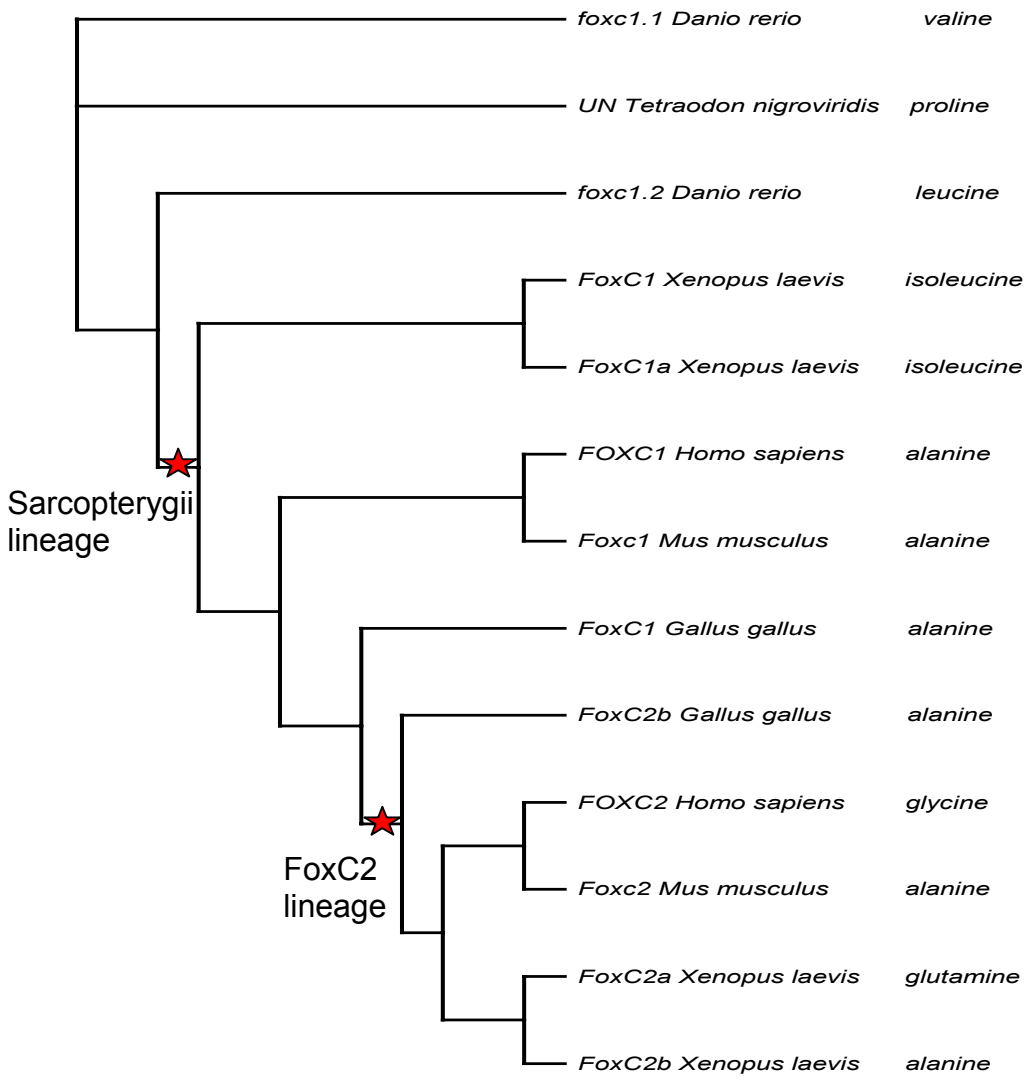
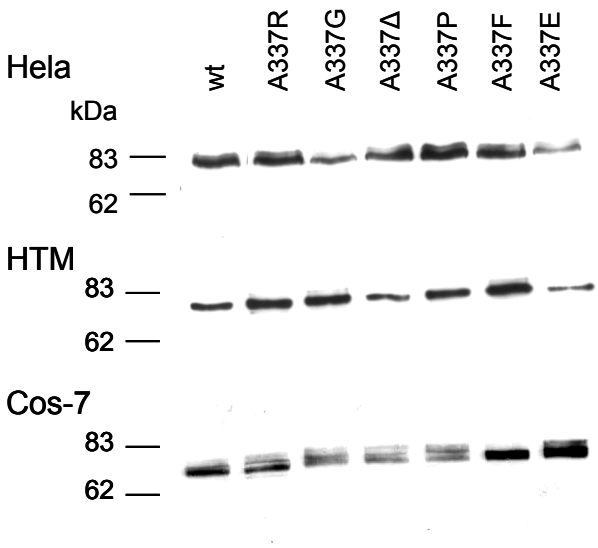


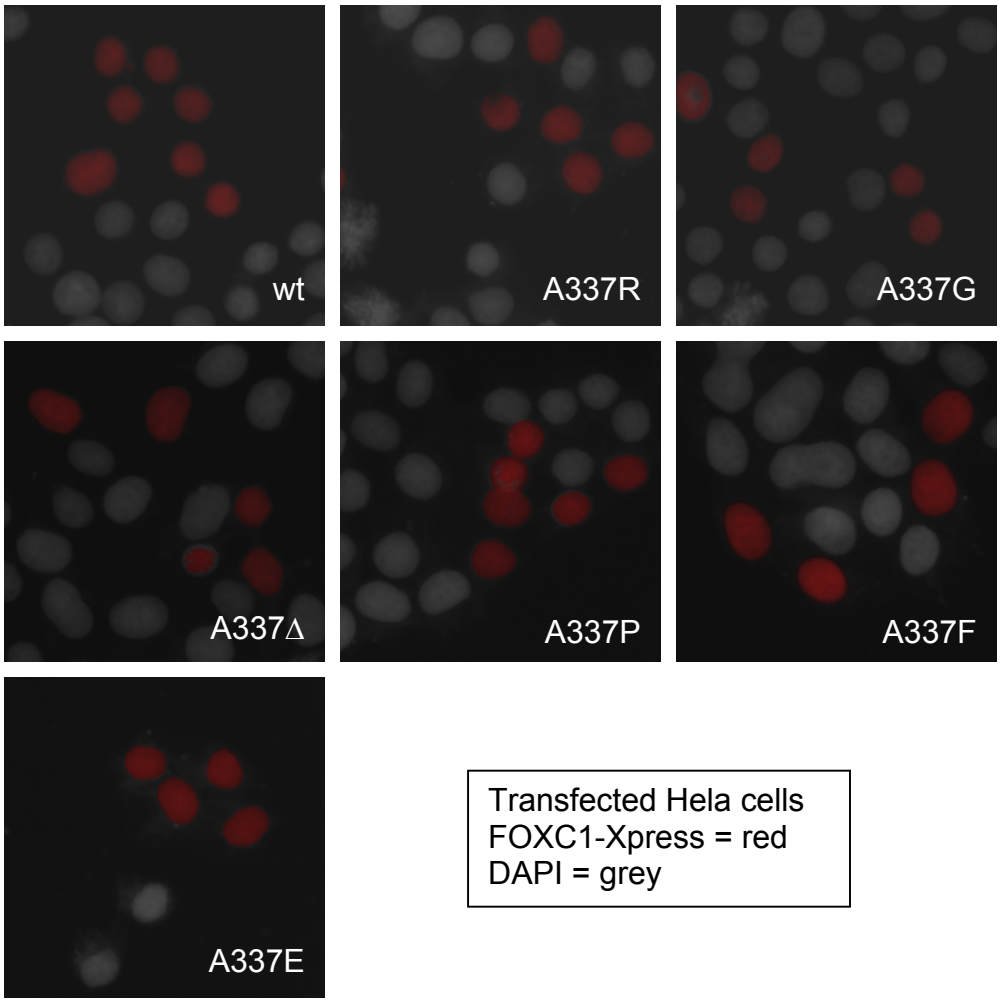
Figure 3-3. *FoxC* phylogeny.

The amino acid composition at the positively selected site is given to the right of each gene. The branch separating the Sarcopterygii (lobed finned fish and tetrapods) lineage from the Actionopterygii (ray finned fish) lineage and that separating the FoxC1 genes from the FoxC2 genes which were tested for positive selection are indicated with stars.

A.



B.



Transfected Hela cells
FOXC1-Xpress = red
DAPI = grey

C.

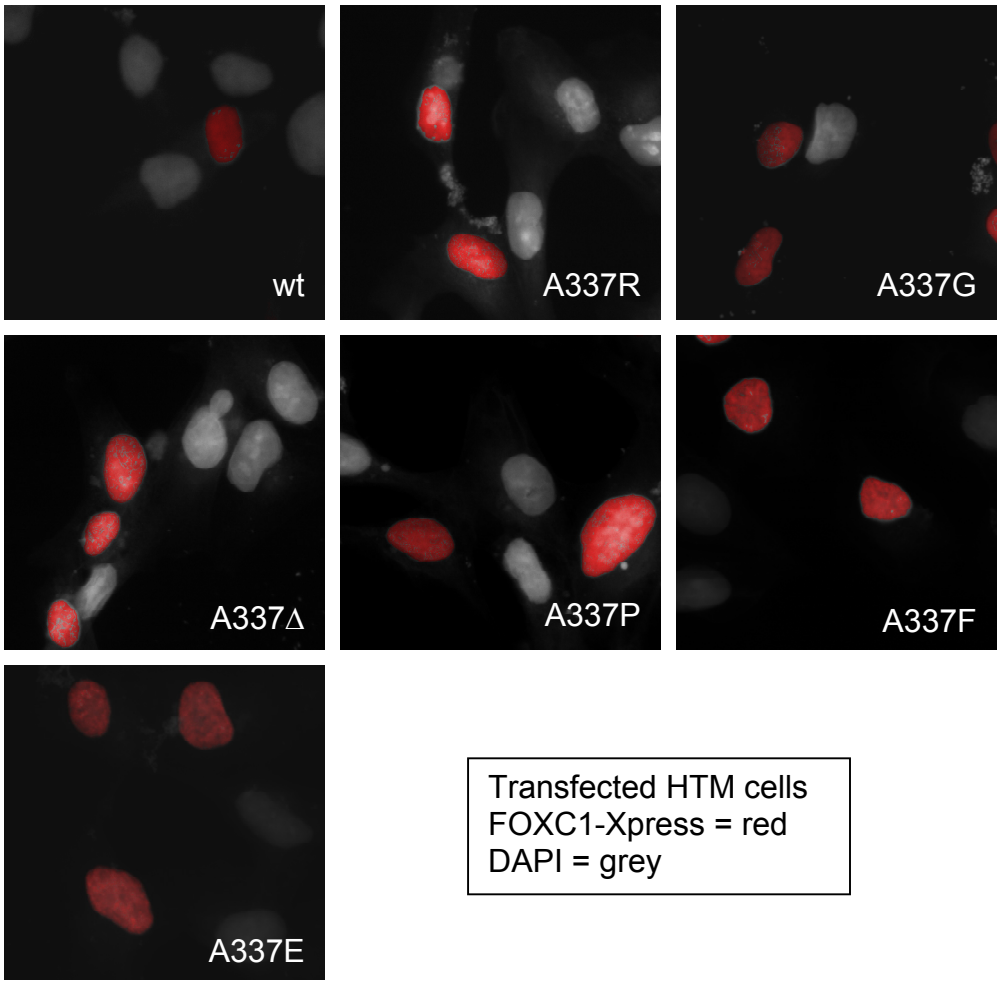
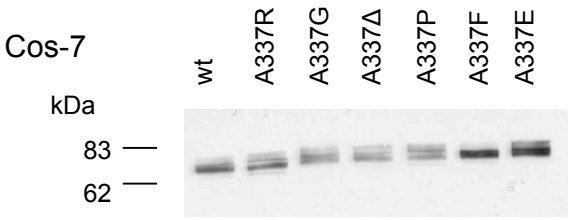


Figure 3-4. Mutation of the positive site does not abolish protein expression or alter normal nuclear localization of FOXC1.

A. Western blots demonstrate that the A337 mutant *FOXC1* constructs express protein. Whole cell lysates from cells transfected with wild type or mutant *FOXC1-Xpress* constructs were resolved by polyacrylamide gel electrophoresis then subjected to western blotting with an anti-Xpress antibody. B/C. Immunofluorescence shows that A337 mutant FOXC1 localize to the nucleus. Cells were transfected with mutant or wild type *FOXC1-Xpress* and the proteins produced were detected with an anti-Xpress Cy3 conjugated antibody (red). DAPI is shown in grey. B. Hela cells C. HTM cells

A.


Cos-7
kDa
83
62
wt
A337R
A337G
A337Δ
A337P
A337F
A337E

B.

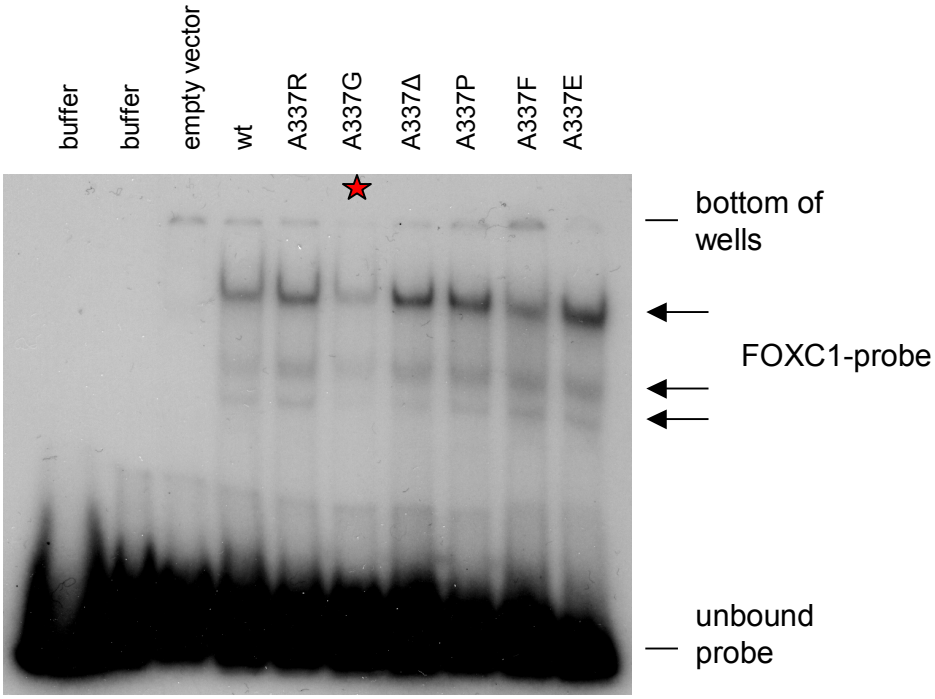

buffer
buffer
empty vector
wt
A337R
A337G
A337Δ
A337P
A337F
A337E
bottom of wells
FOXC1-probe
unbound probe

Figure 3-5. The FOXC1 A337G mutant shows a reduction in DNA binding. Cos-7 cells were transfected with wild type or mutant FOXC1-Xpress constructs and whole cell lysates were collected. A. The amount of Xpress tagged proteins used in the EMSAs were roughly equalized through western blotting as shown here. B. EMSAs using these lysates and a radioactively labeled FOXC1 DNA binding site probe demonstrated band shift patterns equivalent to wild type for all of the constructs however, the intensity of the A337G band shifts was reduced compared to wild type (starred lane). Negative controls (buffer) that do not include cell lysates, but only the buffer in which the cell lysates were collected, show no band shift. A second negative control (empty vector) that includes whole cell lysates from cells transfected with empty Xpress vector shows one band that is not present in the reactions that contain FOXC1-Xpress proteins. This band likely occurs due to the presence of endogenous forkhead domain containing proteins in Cos-7 cells.

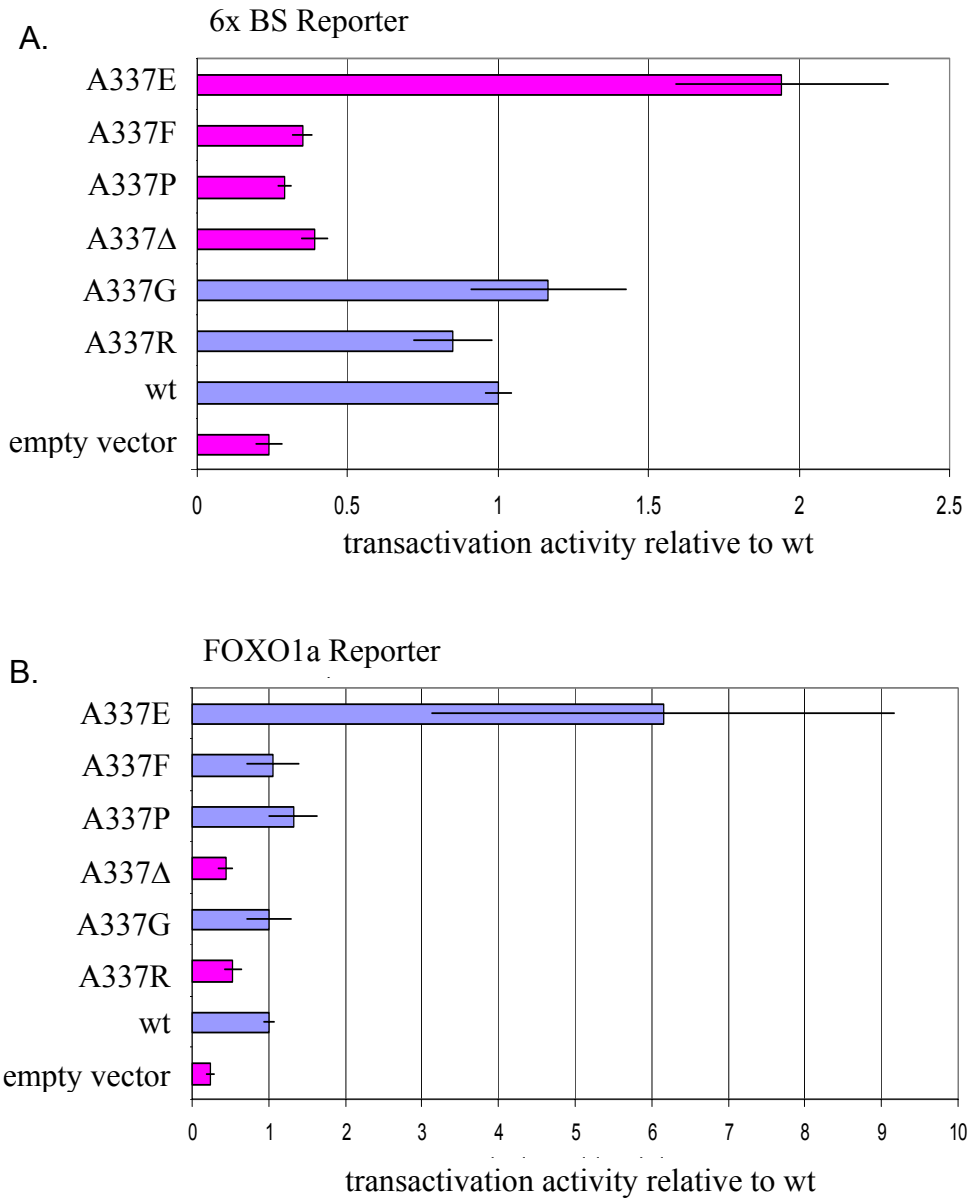


Figure 3-6. The positively selected site influences FOXC1 transactivation activity.

Transactivation activities of the A337 mutants relative to wild type (wt) FOXC1 on the 6x BS reporter (A) and the FOXO1a reporter (B). Statistically significant differences in transactivation activity relative to wild type are indicated with pink bars and standard errors are shown.

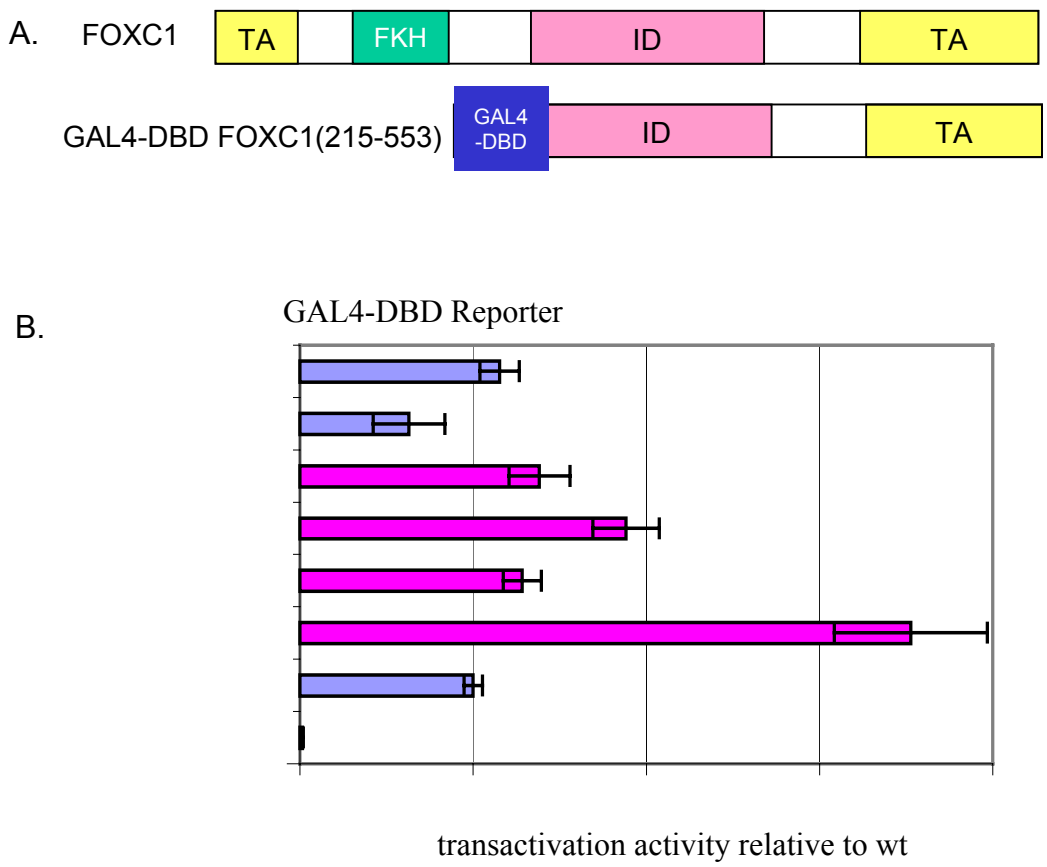


Figure 3-7. Mutating the positively selected site affects the inhibitory domain intrinsically.

A. Schematic of GAL4-DBD FOXC1(215-553) fusion proteins. Yellow: transactivation domain (TA) Green: forkhead domain (FKH) Pink: inhibitory domain (ID) Blue: GAL4 DNA binding domain (GAL4-DBD) B. Transactivation activities of the GAL4-DBD FOXC1(215-553) A337 mutant fusions relative to wild type FOXC1(215-553) fusions on a GAL4-DBD responsive reporter gene. Statistically significant differences in transactivation activity relative to wild type are indicated with pink bars and standard errors are shown.

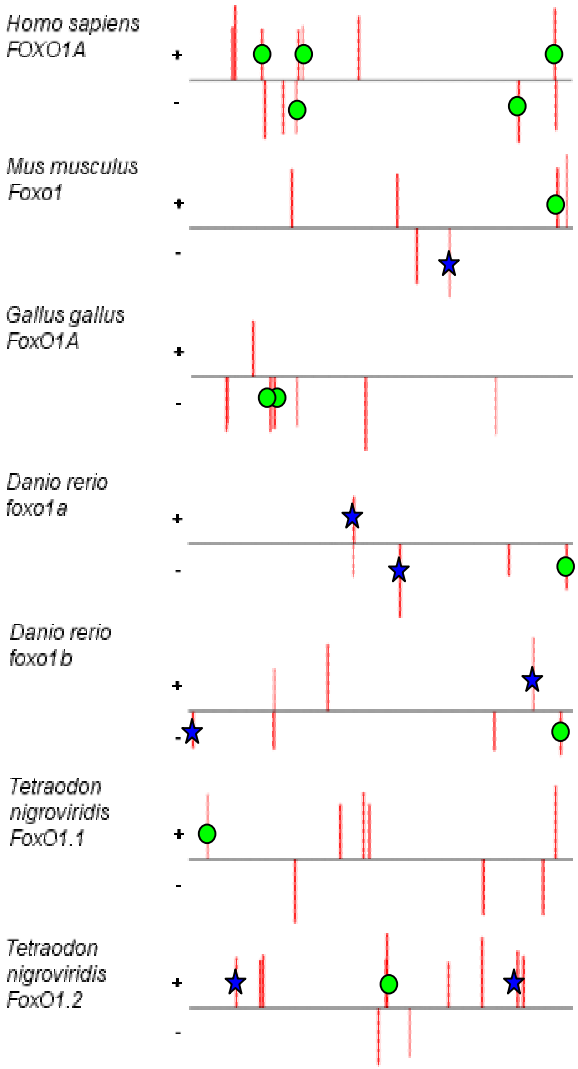
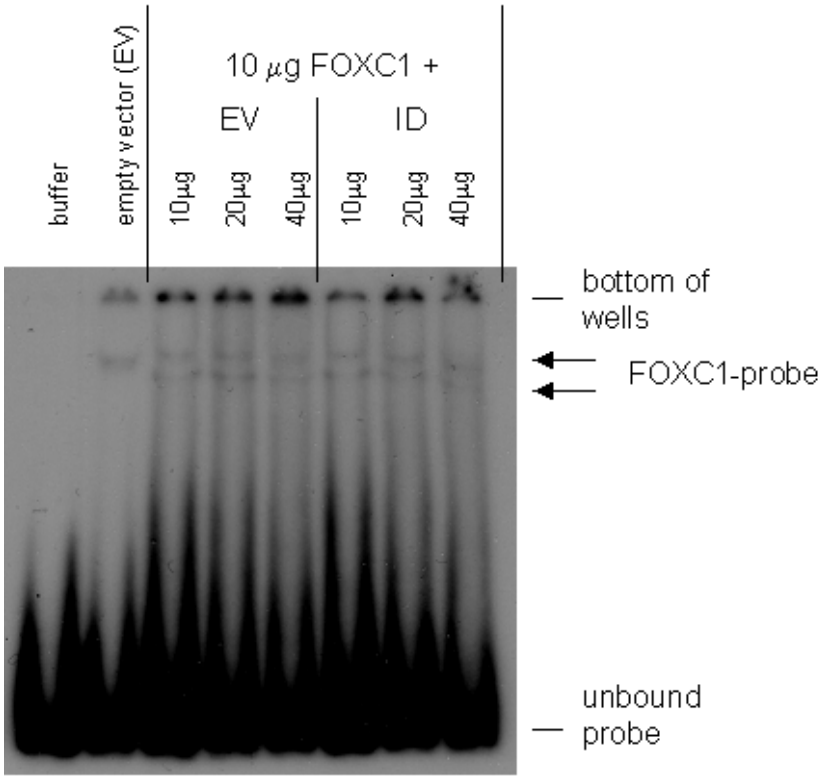


Figure 3-8. FOXC1 binding sites in the 8000 nucleotides upstream of the translation start site of *FOXO1a* orthologs.

The 8000 nucleotides in the 5' to 3' direction are represented by a horizontal grey line. Vertical red lines represent FOXC1 binding sites. Binding sites on the + side of the black line are in trans while those on the – side are in cis with the antisense strand. Perfect FOXC1 binding sites are indicated with blue stars while sites that differ by one nucleotide at a known variable nucleotide are labeled with green circles.

A.


buffer
empty vector (EV)
10 μg FOXC1 +
EV
ID
10μg
20μg
40μg
10μg
20μg
40μg
bottom of wells
FOXC1-probe
unbound probe

B.

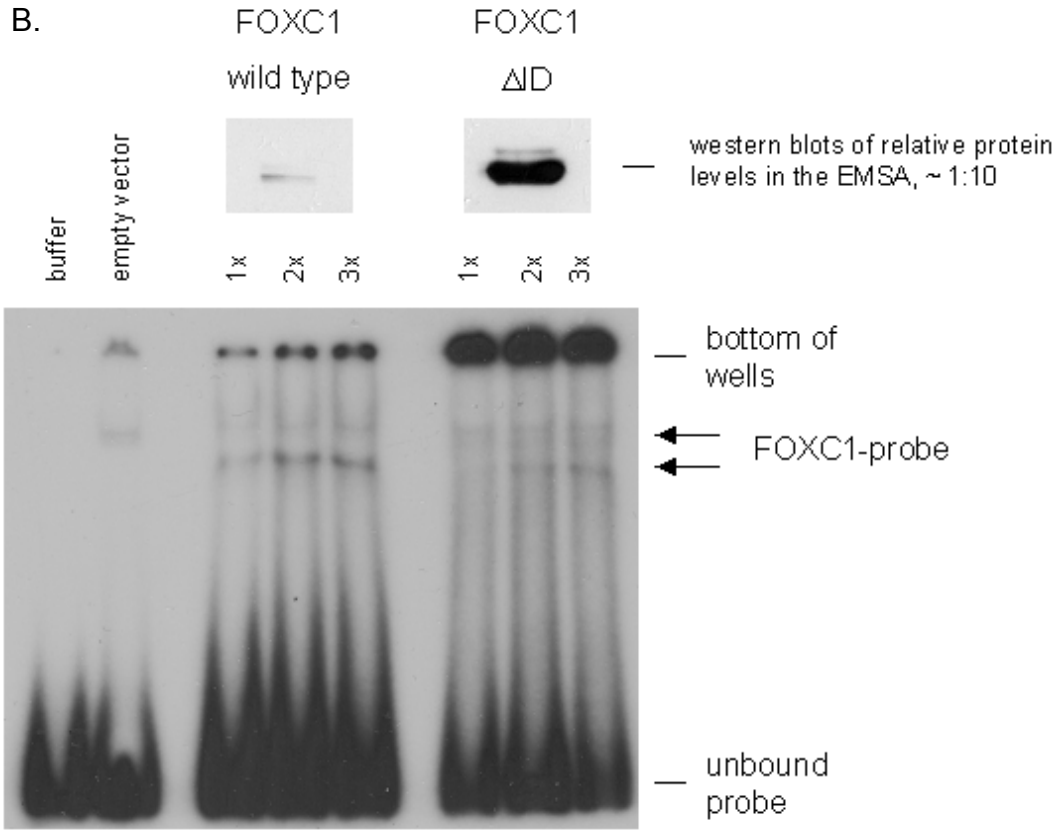

FOXC1
wild type
FOXC1
ΔID
western blots of relative protein levels in the EMSA, ~ 1:10
buffer
empty vector
1x
2x
3x
1x
2x
3x
bottom of wells
FOXC1-probe
unbound probe

Figure 3-9. Loss of the inhibitory domain (ID) reduces FOXC1 binding. Cos-7 cells were transfected with wild type FOXC1, FOXC1ΔID, ID or empty vector (EV) and whole cell lysates were collected. A. EMSAs of FOXC1 plus increasing amounts of EV or increasing amounts of ID. Adding ID does not alter FOXC1 DNA binding. B. EMSAs of increasing amounts of FOXC1 or FOXC1ΔID. The expression of FOXC1 and FOXC1ΔID were not equalized prior to EMSA. The expression of the two constructs at the 3x concentration used in the EMSA is shown above the EMSA radiograph. At least 10 times more FOXC1ΔID is required to generate a band shift that approaches the intensity of that seen for wild type FOXC1. Negative controls (buffer) that do not include cell lysates, but only the buffer in which the cell lysates were collected, show no band shift. A second negative control (empty vector) that includes whole cell lysates from cells transfected with empty Xpress vector shows one band that is not present in the reactions that contain FOXC1-Xpress proteins. This band likely occurs due to the presence of endogenous forkhead domain containing proteins in Cos-7 cells.

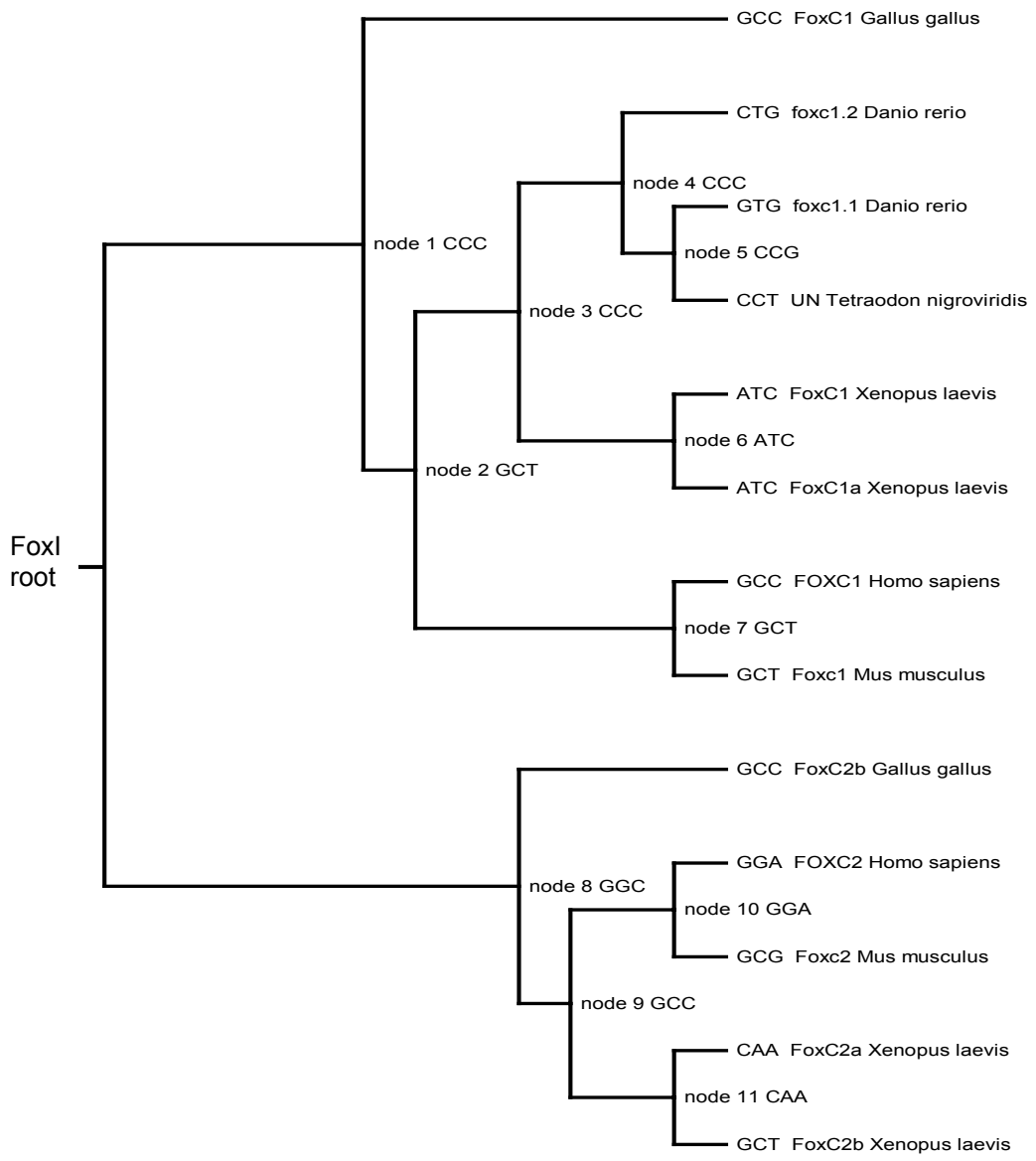
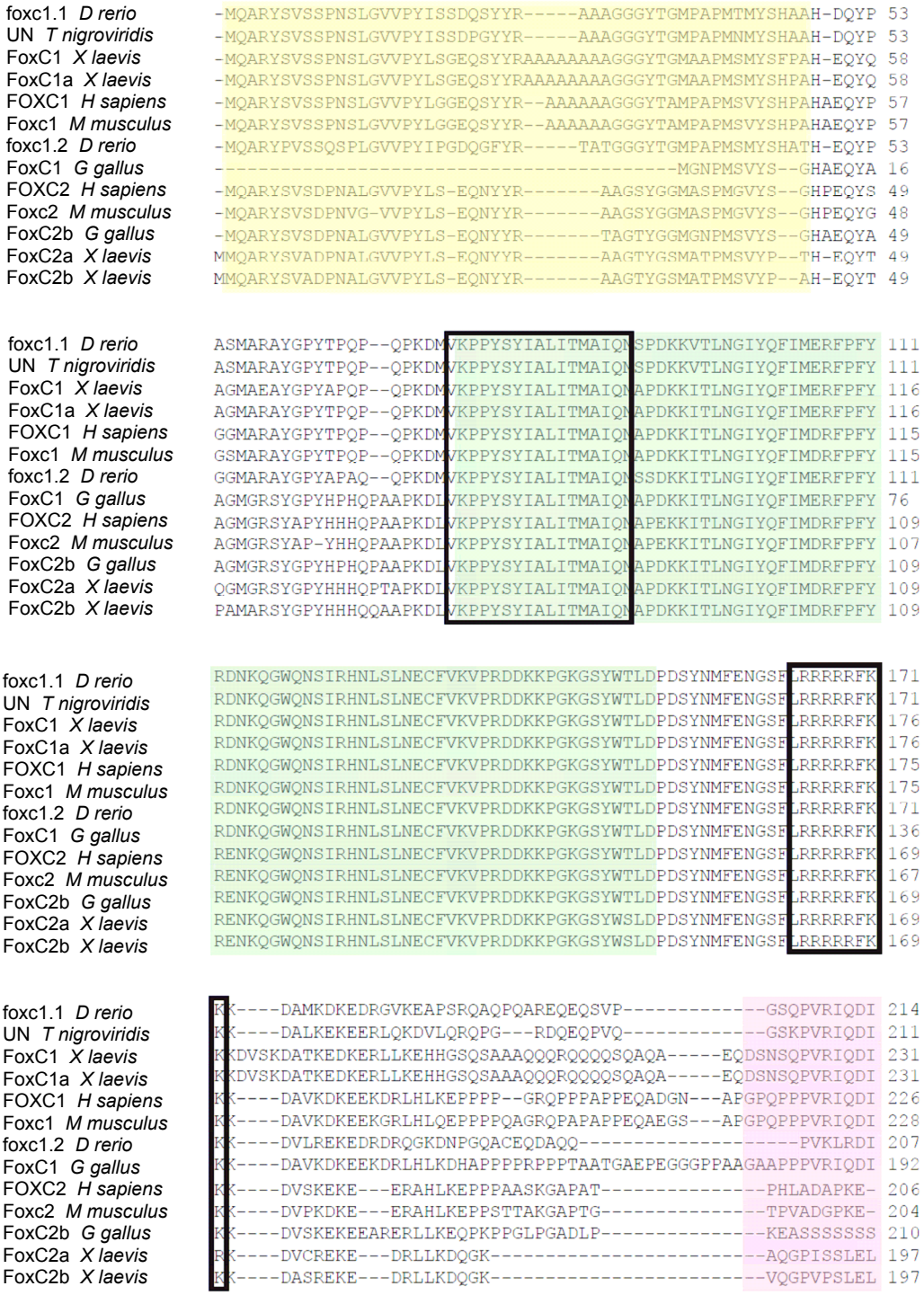
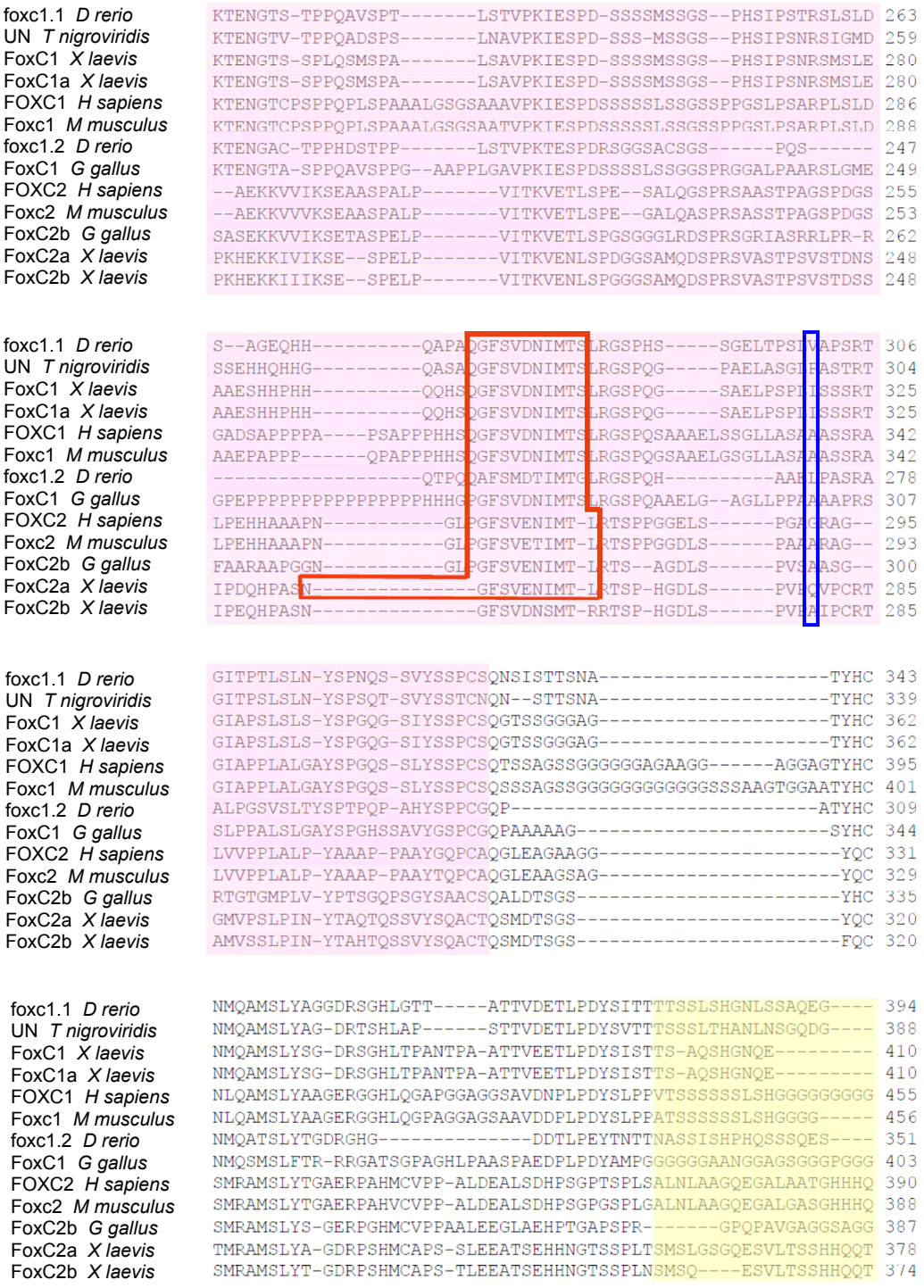


Figure 3-10. The codons located at the positively selected site in a gene tree context.

For each node in the tree the codon located at the positively selected site is given. Codons at nodes labeled 'node #' were reconstructed during the codeml analysis. The FoxC gene tree was rooted using a set of 10 FoxI sequences as an outgroup. No codon is given at the root as the codeml analysis was done using an unrooted tree.


foxc1.1 D rerio -MQARYSVSSPNSLGVVPYISSDQSYYR-----AAAGGGYTGMPAPMTMYSHAAH-DQYP 53
UN T nigroviridis -MQARYSVSSPNSLGVVPYISSDPGYYR-----AAAGGGYTGMPAPMNMYSHAAH-DQYP 53
FoxC1 X laevis -MQARYSVSSPNSLGVVPYLSGEQSYYRAAAAAAAAGGGYTGMAAPMSMYSFPAH-EQYQ 58
FoxC1a X laevis -MQARYSVSSPNSLGVVPYLSGEQSYYRAAAAAAAAGGGYTGMAAPMSMYSHPAH-EQYQ 58
FOXC1 H sapiens -MQARYSVSSPNSLGVVPYLGGEQSYYR--AAAAAAGGGYTAMPAPMSVYSHPAHAEQYP 57
Foxc1 M musculus -MQARYSVSSPNSLGVVPYLGGEQSYYR--AAAAAAGGGYTAMPAPMSVYSHPAHAEQYP 57
foxc1.2 D rerio -MQARYPVSSQSPLGVVPYIPGDQGFYR-----TATGGGYTGMPAPMSMYSHATH-EQYP 53
FoxC1 G gallus ----------MGNPMSVYS--GHAEQYA 16
FOXC2 H sapiens -MQARYSVSDPNALGVVPYLS-EQNYYR-------AAGSYGGMASPMGVYS--GHPEQYS 49
Foxc2 M musculus -MQARYSVSDPNVG-VVPYLS-EQNYYR-------AAGSYGGMASPMGVYS--GHPEQYG 48
FoxC2b G gallus -MQARYSVSDPNALGVVPYLS-EQNYYR-------TAGTYGGMGNPMSVYS--GHAEQYA 49
FoxC2a X laevis MMQARYSVADPNALGVVPYLS-EQNYYR-------AAGTYGSMATPMSVYP--TH-EQYT 49
FoxC2b X laevis MMQARYSVADPNALGVVPYLS-EQNYYR-------AAGTYGSMATPMSVYP--AH-EQYT 49
foxc1.1 D rerio ASMARAYGPYTPQP--QPKDMVKPPYSYIALITMAIQNSPDKKVTLNGIYQFIMERFPFY 111
UN T nigroviridis ASMARAYGPYTPQP--QPKDMVKPPYSYIALITMAIQNSPDKKVTLNGIYQFIMERFPFY 111
FoxC1 X laevis AGMAEAYGPYAPQP--QPKDMVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMERFPFY 116
FoxC1a X laevis AGMARAYGPYTPQP--QPKDMVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMERFPFY 116
FOXC1 H sapiens GGMARAYGPYTPQP--QPKDMVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMDRFPFY 115
Foxc1 M musculus GSMARAYGPYTPQP--QPKDMVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMDRFPFY 115
foxc1.2 D rerio GGMARAYGPYAPAQ--QPKDMVKPPYSYIALITMAIQNSSDKKITLNGIYQFIMERFPFY 111
FoxC1 G gallus AGMGRSYGPYHPHQPAAPKDLVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMERFPFY 76
FOXC2 H sapiens AGMGRSYAPYHHHQPAAPKDLVKPPYSYIALITMAIQNAPEKKITLNGIYQFIMDRFPFY 109
Foxc2 M musculus AGMGRSYAP-YHHQPAAPKDLVKPPYSYIALITMAIQNAPEKKITLNGIYQFIMDRFPFY 107
FoxC2b G gallus AGMGRSYGPYHPHQPAAPKDLVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMDRFPFY 109
FoxC2a X laevis QGMGRSYGPYHHHQPTAPKDLVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMDRFPFY 109
FoxC2b X laevis PAMARSYGPYHHHQQAAPKDLVKPPYSYIALITMAIQNAPDKKITLNGIYQFIMDRFPFY 109
foxc1.1 D rerio RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 171
UN T nigroviridis RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 171
FoxC1 X laevis RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 176
FoxC1a X laevis RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 176
FOXC1 H sapiens RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 175
Foxc1 M musculus RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 175
foxc1.2 D rerio RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 171
FoxC1 G gallus RDNKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 136
FOXC2 H sapiens RENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 169
Foxc2 M musculus RENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 167
FoxC2b G gallus RENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSFLRRRRRFK 169
FoxC2a X laevis RENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWSLDPDSYNMFENGSFLRRRRRFK 169
FoxC2b X laevis RENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWSLDPDSYNMFENGSFLRRRRRFK 169
foxc1.1 D rerio KK----DAMKDKEDRGVKEAPSRQAQPQAREQEQSVP----------GSQPVRIQDI 214
UN T nigroviridis KK----DALKEKEERLQKDVLQRQPG---RDQEQPVQ----------GSKPVRIQDI 211
FoxC1 X laevis KKDVSKDATKEDKERLLKEHHGSQSAAAQQQRQQQQSQAQA-----EQDSNSQPVRIQDI 231
FoxC1a X laevis KKDVSKDATKEDKERLLKEHHGSQSAAAQQQRQQQQSQAQA-----EQDSNSQPVRIQDI 231
FOXC1 H sapiens KK----DAVKDKEEKDRLHLKEPPPP--GRQPPPAPPEQADGN---APGPQPPPVRIQDI 226
Foxc1 M musculus KK----DAVKDKEEKGRLHLQEPPPPQAGRQPAPAPPEQAEGS---APGPQPPPVRIQDI 228
foxc1.2 D rerio KK----DVLREKEDRDRQGKDNPGQACEQDAQQ----------PVKLRDI 207
FoxC1 G gallus KK----DAVKDKEEKDRLHLKDHAPPPPRPPPTAATGAEPEGGGPPAAGAAPPPVRIQDI 192
FOXC2 H sapiens KK----DVSKEKE---ERAHLKEPPPAASKGAPAT----------PHLADAPKE- 206
Foxc2 M musculus KK----DVPKDKE---ERAHLKEPPSTTAKGAPTG----------TPVADGPKE- 204
FoxC2b G gallus KK----DVSKEKEEARERLLKEQPKPPGLPGADLP----------KEASSSSSSS 210
FoxC2a X laevis RK----DVCREKE---DRLLKDQGK----------AQGPISSLEL 197
FoxC2b X laevis KK----DASREKE---DRLLKDQGK----------VQGPVPSLEL 197


foxc1.1 D rerio
UN T nigroviridis
FoxC1 X laevis
FoxC1a X laevis
FOXC1 H sapiens
Foxc1 M musculus
foxc1.2 D rerio
FoxC1 G gallus
FOXC2 H sapiens
Foxc2 M musculus
FoxC2b G gallus
FoxC2a X laevis
FoxC2b X laevis

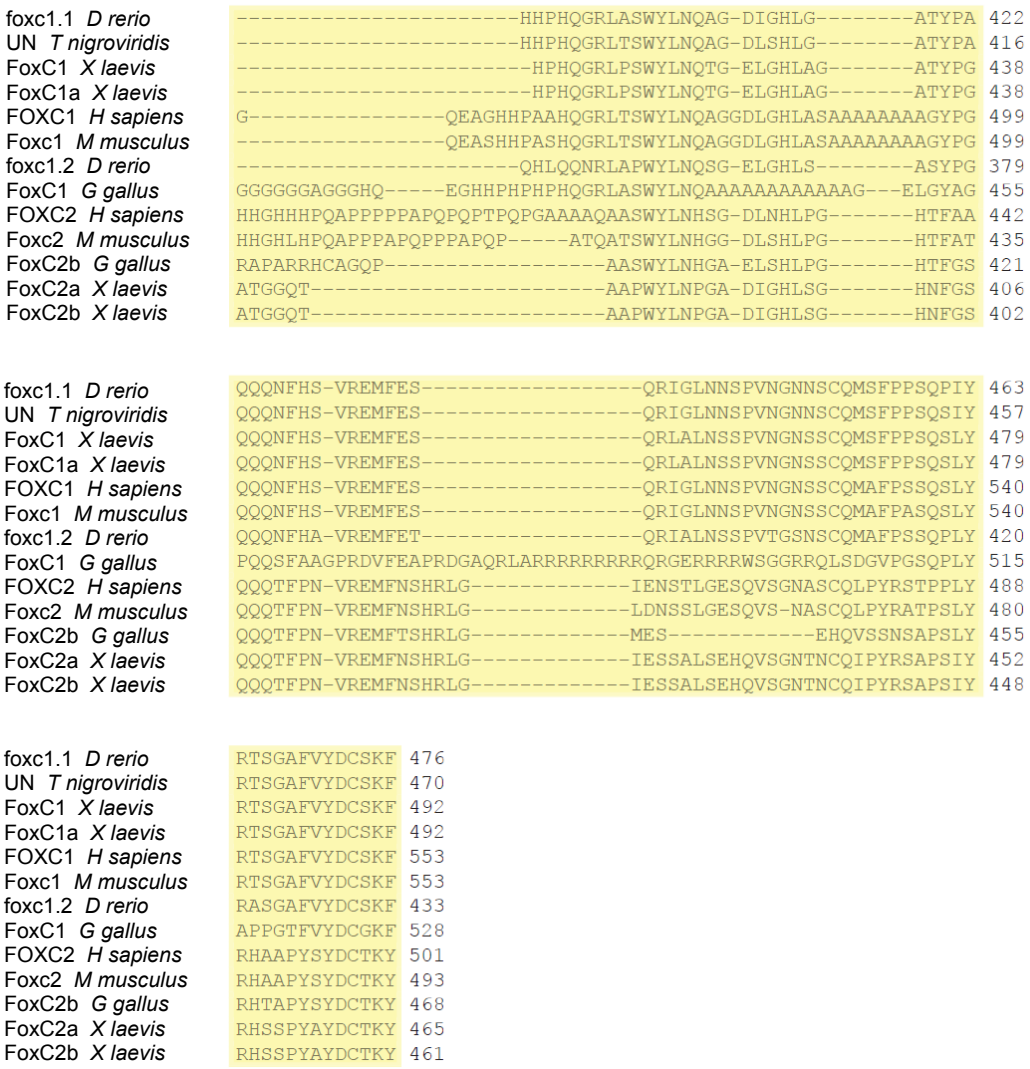


Figure 3- 11. Alignment of the FoxC sequences with functionally important regions highlighted.

Yellow: transactivation domain Green: forkhead domain Black Box: nuclear localization signal (Berry et al. 2002) Pink: inhibitory domain Red Box: EH1 motif (Copley 2005) Blue Box: positively selected site

References

Berglund J, Pollard KS, Webster MT (2009) Hotspots of biased nucleotide substitutions in human genes. PLoS Biol 7:e26

Berry FB, Saleem RA, Walter MA (2002) FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain. J Biol Chem 277:10292-10297

Berry FB, Skarie JM, Mirzayans F, Fortin Y, Hudson TJ, Raymond V, Link BA, Walter MA (2008) FOXC1 is required for cell viability and resistance to oxidative stress in the eye through the transcriptional regulation of FOXO1A. Hum Mol Genet 17:490-505

Berry FB, Tamimi Y, Carle MV, Lehmann OJ, Walter MA (2005) The establishment of a predictive mutational model of the forkhead domain through the analyses of FOXC2 missense mutations identified in patients with hereditary lymphedema with distichiasis. Hum Mol Genet 14:2619-2627

Blanco E, Guigo R, Messeguer X (2007) Multiple non-collinear TF-map alignments of promoter regions. BMC Bioinformatics 8:138

Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 55:104-110

Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36:D102-106

Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L (2010) Adaptive divergence of ancient gene duplicates in the avian MHC class II B. Mol Biol Evol

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, et al. (2005) The transcriptional landscape of the mammalian genome. Science 309:1559-1563

Cederberg A, Gronning LM, Ahren B, Tasken K, Carlsson P, Enerback S (2001) FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and diet-induced insulin resistance. Cell 106:563-573

Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li Z, Bouckaert J, Gordon JI, Hultgren SJ (2009) Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. Proc Natl Acad Sci U S A 106:22439-22444

Chiba H, Yamashita R, Kinoshita K, Nakai K (2008) Weak correlation between sequence conservation in promoter regions and in protein-coding regions of human-mouse orthologous gene pairs. BMC Genomics 9:152

Connell F, Brice G, Mortimer P (2008) Phenotypic characterization of primary lymphedema. Ann N Y Acad Sci 1131:140-146

Copley RR (2005) The EH1 motif in metazoan transcription factors. BMC Genomics 6:169

Fetterman CD, Mirzayans F, Walter MA (2009) Characterization of a novel FOXC1 mutation, P297S, identified in two individuals with anterior segment dysgenesis. Clin Genet 76:296-299

Fetterman CD, Rannala B, Walter MA (2008) Identification and analysis of evolutionary selection pressures acting at the molecular level in five forkhead subfamilies. BMC Evol Biol 8:261

Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, Chen Y, et al. (2010) Ensembl's 10th year. Nucleic Acids Res 38:D557-562

Fu Y, Frith MC, Haverty PM, Weng Z (2004) MotifViz: an analysis and visualization tool for motif discovery. Nucleic Acids Res 32:W420-423

Fujita H, Kang M, Eren M, Gleaves LA, Vaughan DE, Kume T (2006) Foxc2 is a common mediator of insulin and transforming growth factor beta signaling to regulate plasminogen activator inhibitor type I gene expression. Circ Res 98:626-634

Galtier N, Duret L, Glemin S, Ranwez V (2009) GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. Trends Genet 25:1-5

Garza AS, Ahmad N, Kumar R (2009) Role of intrinsically disordered protein regions/domains in transcriptional regulation. Life Sci 84:189-193

Gessert S, Kuhl M (2009) Comparative gene expression analysis and fate mapping studies suggest an early segregation of cardiogenic lineages in Xenopus laevis. Dev Biol 334:395-408

Gomez D, Conejeros P, Marshall SH, Consuegra S (2010) MHC evolution in three salmonid species: a comparison between class II alpha and beta genes. Immunogenetics

Hayashi H, Kume T (2008a) Forkhead transcription factors regulate expression of the chemokine receptor CXCR4 in endothelial cells and CXCL12-induced cell migration. Biochem Biophys Res Commun 367:584-589

Hayashi H, Kume T (2008b) Foxc transcription factors directly regulate Dll4 and Hey2 expression by interacting with the VEGF-Notch signaling pathways in endothelial cells. PLoS ONE 3:e2401

Huang L (2009) Function and regulation of the FOXC transcription factors., University of Alberta, Edmonton

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573-584

Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, Walter MA (2009) Severe molecular defects of a novel FOXC1 W152G mutation result in aniridia. Invest Ophthalmol Vis Sci 50:3573-3579

Ito YA, Footz TK, Murphy TC, Courtens W, Walter MA (2007) Analyses of a novel L130F missense mutation in FOXC1. Arch Ophthalmol 125:128-135

Ivarsson Y, Mackey AJ, Edalat M, Pearson WR, Mannervik B (2003) Identification of residues in glutathione transferase capable of driving functional diversification in evolution. A novel approach to protein redesign. J Biol Chem 278:8733-8738

Kapralov MV, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evol Biol 7:73

Koster M, Dillinger K, Knochel W (2000) Activin A signaling directly activates Xenopus winged helix factors XFD-4/4', the orthologues to mammalian MFH-1. Dev Genes Evol 210:320-324

Kume T, Deng KY, Winfrey V, Gould DB, Walter MA, Hogan BL (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell 93:985-996

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980-996

Lehmann OJ, Ebenezer ND, Ekong R, Ocaka L, Mungall AJ, Fraser S, McGill JI, Hitchings RA, Khaw PT, Sowden JC, Povey S, Walter MA, Bhattacharya SS, Jordan T (2002) Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions. Invest Ophthalmol Vis Sci 43:1843-1849

Lehmann OJ, Ebenezer ND, Jordan T, Fox M, Ocaka L, Payne A, Leroy BP, Clark BJ, Hitchings RA, Povey S, Khaw PT, Bhattacharya SS (2000) Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma. Am J Hum Genet 67:1129-1135

Letondal C, Schuerer K protal2dna release 2.0, Paris, France

Liu Y, Matthews KS, Bondos SE (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. J Biol Chem 283:20874-20887

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34:D108-110

Mazet F, Amemiya CT, Shimeld SM (2006) An ancient Fox gene cluster in bilaterian animals. Curr Biol 16:R314-316

Minezaki Y, Homma K, Kinjo AR, Nishikawa K (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol 359:1137-1149

Murphy TC, Saleem RA, Footz T, Ritch R, McGillivray B, Walter MA (2004) The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions. Invest Ophthalmol Vis Sci 45:2531-2538

Nishimura DY, Searby CC, Alward WL, Walton D, Craig JE, Mackey DA, Kawase K, Kanis AB, Patil SR, Stone EM, Sheffield VC (2001) A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. Am J Hum Genet 68:364-372

Novais A, Comas I, Baquero F, Canton R, Coque TM, Moya A, Gonzalez-Candelas F, Galan JC (2010) Evolutionary trajectories of beta-lactamase CTX-M-1 cluster enzymes: predicting antibiotic resistance. PLoS Pathog 6:e1000735

Nylander JAA (2004) MrModeltest 2.0 release 2.2, Uppsala, Sweden

Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet 39:730-732

Petrova TV, Karpanen T, Norrmen C, Mellor R, Tamakoshi T, Finegold D, Ferrell R, Kerjaschki D, Mortimer P, Yla-Herttuala S, Miura N, Alitalo K (2004) Defective valves and abnormal mural cell recruitment underlie lymphatic vascular failure in lymphedema distichiasis. Nat Med 10:974-981

Pierrou S, Hellqvist M, Samuelsson L, Enerback S, Carlsson P (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J 13:5002-5012

Premzl M, Gamulin V (2009) Positive selection in prion protein. J Mol Evol 68:205-207

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574

Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA (2001) Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1. Am J Hum Genet 68:627-641

Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA (2003a) Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1. Hum Mol Genet 12:2993-3005

Saleem RA, Berry FB, Huang L, Walter MA (2008) FOXC1 and FOXL2 and the Axenfeld-Rieger Malformations and the Blepharophimosis, Ptosis, and Epicanthus Inversus Syndrome. In: Epstein CJ, Erickson RP, Wynshaw-Boris AJ (eds) Inborn errors of development : the molecular basis of clinical disorders of morphogenesis. Oxford University Press, Oxford ; New York, pp 834-840

Saleem RA, Murphy TC, Liebmann JM, Walter MA (2003b) Identification and analysis of a novel mutation in the FOXC1 forkhead domain. Invest Ophthalmol Vis Sci 44:4608-4612

Sawyer SL, Wu LI, Emerman M, Malik HS (2005) Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A 102:2832-2837

Schaefer C, Schlessinger A, Rost B (2010) Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. Bioinformatics 26:625-631

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT (2010) Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. Science

Schwalie PC, Schultz J (2009) Positive selection in tick saliva proteins of the Salp15 family. J Mol Evol 68:186-191

Seo S, Fujita H, Nakano A, Kang M, Duarte A, Kume T (2006) The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. Dev Biol 294:458-470

Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18-20

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods). release 4.0b10, Sunderland, Massachusetts, USA

Tamimi Y, Skarie JM, Footz T, Berry FB, Link BA, Walter MA (2006) FGF19 is a target for FOXC1 regulation in ciliary body-derived cells. Hum Mol Genet 15:3229-3240

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

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

Toth-Petroczy A, Simon I, Fuxreiter M, Levy Y (2009) Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. J Am Chem Soc 131:15084-15085

Tumer Z, Bach-Holm D (2009) Axenfeld-Rieger syndrome and spectrum of PITX2 and FOXC1 mutations. Eur J Hum Genet 17:1527-1539

Vreeburg M, Heitink MV, Damstra RJ, Moog U, van Geel M, van Steensel MA (2008) Lymphedema-distichiasis syndrome: a distinct type of primary lymphedema caused by mutations in the FOXC2 gene. Int J Dermatol 47 Suppl 1:52-55

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337:635-645

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, et al. (2006) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 34:D173-180

Winnier GE, Hargett L, Hogan BL (1997) The winged helix transcription factor MFH1 is required for proliferation and patterning of paraxial mesoderm in the mouse embryo. Genes Dev 11:926-940

Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041-1051

Wotton KR, Mazet F, Shimeld SM (2008) Expression of FoxC, FoxF, FoxL1, and FoxQ1 genes in the dogfish Scyliorhinus canicula defines ancient and derived roles for Fox genes in vertebrate development. Dev Dyn 237:1590-1603

Wotton KR, Shimeld SM (2006) Comparative genomics of vertebrate Fox cluster loci. BMC Genomics 7:271

Wu Q, Li L, Zhang G (2010) Crassostrea angulata Bindin Gene and the Divergence of Fucose-Binding Lectin Repeats Among Three Species of Crassostrea. Mar Biotechnol (NY)

Xu Z, Zhou R, Jin M, Chen H (2010) Selection pressure on the hemagglutinin gene of Influenza A (H1N1) virus: adaptation to human and swine hosts in Asia. Acta Virol 54:113-118

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556

Chapter 4: Characterization of a novel FOXC1 inhibitory domain mutation, P297S, and exploration into the inhibitory domain functional mechanism.

This chapter contains work published in: Fetterman CD, Mirzayans F, Walter MA (2009) Characterization of a novel FOXC1 mutation, P297S, identified in two individuals with anterior segment dysgenesis. Clin Genet 76:296-299.

Authors' Contributions: CDF participated in study design and carried out almost all experiments. FM first identified the mutation and performed some DNA sequencing. MAW conceived of the study and participated in its design.

Introduction

As discussed in Chapter 1, mutations in FOXC1 have been associated with ocular anterior segment abnormalities. To date, 23 different missense or nonsense mutations of FOXC1 have been identified (Figure 4-1) (Mears et al. 1998; Nishimura et al. 1998; Mirzayans et al. 2000; Kawase et al. 2001; Nishimura et al. 2001; Suzuki et al. 2001; Panicker et al. 2002; Komatireddy et al. 2003; Saleem et al. 2003b; Mortemousque et al. 2004; Murphy et al. 2004; Cella et al. 2006; Weisschuh et al. 2006; Fuse et al. 2007; Ito et al. 2007; Weisschuh et al. 2008; Ito et al. 2009). Only three of these are located outside of the forkhead domain and they are nonsense mutations that occur within the first 50 amino acids of the protein. Here I report and functionally examine the first *FOXC1* missense mutation, P297S, that occurs outside of the forkhead domain. P297S is within the inhibitory domain, which is known to regulate FOXC1 transactivation activity and stability (Berry et al. 2002; Berry et al. 2006). The mechanism by which the inhibitory domain regulates transactivation activity remains unknown. In Chapter 3 I determined that the inhibitory domain likely does not exert its effect through a reduction in DNA binding by FOXC1. Here I also attempt to determine if the inhibitory domain directly represses a transactivation domain or interacts with a repressor protein.

Materials and Methods

Patients and Mutation Detection

The 889C>T transition, resulting in P297S, was identified in two unrelated individuals with anterior segment dysgenesis through sequencing of *FOXC1*. Patient 1 presented at 25 years of age while patient 2 presented at 48 years. Other ocular findings for patient 1 include iridogoniodysgenesis, myopia and abnormal tissue in the angle while for patient 2 iridogoniodysgenesis, goniodysgenesis, map dot fingerprint dystrophy, myopia and plateau iris have been noted. It is unknown whether these patients have other systemic defects. Both patients were successfully treated medically and surgically. Sequencing of *FOXC1* in 50 control subjects did not reveal 889C>T. All sequencing was performed with an Applied Biosystems Genetic Analyzer, 3100-*Avant* or 3130xl with the Applied Biosystems Big Dye Terminator kit. This research adhered to the tenets of the Declaration of Helsinki, patient samples and information were gathered following the University of Alberta Ethics Board policies.

P297S Functional Analyses

The FOXC1 pcDNA4-Xpress His/Max B (Invitrogen) plasmid was used in all experiments. Site-directed mutagenesis to create 889C>T was completed with the QuickChange Site Directed Mutagenesis kit (Stratagene). Transient transfections were performed with Fugene 6 (Roche) in a ratio of 3:1, [Fugene]:[DNA] in Hela, human trabecular meshwork (HTM) and Cos-7 cells

grown in DMEM supplemented with 10% fetal bovine serum. For immunoblotting, half-life assays and electrophoretic mobility shift assays (EMSAs), cells were harvested by scraping, suspended in nuclear lysis buffer and sonicated. Cellular localization of FOXC1-Xpress and protein expression levels were determined through immunofluorescence and immunoblotting respectively. EMSAs utilizing a probe containing an ideal FOXC1 binding site (Saleem et al. 2001) were performed to characterize DNA binding ability. Transactivation assays with the Dual Luciferase Reporter Assay System (Promega) tested the activity of wild type FOXC1 and FOXC1 P297S on two different reporter constructs: 6x BS and FOXO1a. Statistical significance between wild type and FOXC1 P297S results was measured using a two-tailed t-test with unequal variances at $\alpha = 0.05$. The above experiments were all performed as those described in Chapter 3.

Treatment with cyclohexamide (CHX), a protein synthesis inhibitor, and measurement of FOXC1-Xpress protein levels was used to determined relative wild type and FOXC1 P297S protein half-life. Hela and HTM cells in 60mm plates were transfected with 2µg of wild type FOXC1 or FOXC1 P297S. At 24 hours post transfection, the cells were treated with 100µg/ml CHX for 0, 15, 30, 60, 180 and 360 minutes then harvested. Immunoblotting and ImageJ densitometry analysis of the proteins allowed for FOXC1-Xpress quantification. TFIID protein levels were also measured and used as a loading control to normalize FOXC1 levels. The levels of FOXC1 were then normalized to the time 0 (t_0) FOXC1 level and then linearized and plotted. The correlation coefficient

(R^2) of the line was used as a measure of the reliability of the fit of the line. Only experiments with $R^2 > 0.7$ were considered for further analysis. The slope of the line was used to determine half-life where $T_{1/2} = \ln(0.5)/\text{slope}$. To determine statistical significance, $T_{1/2}$ was weighted by wild type FOXC1 $T_{1/2}$ for each experiment independently then two experiments were compared with a two-tailed t-test with unequal variances at a significance level of $\alpha = 0.05$.

TLE4 Transactivation Assays

TLE (transducin like enhancer of split) proteins are known as transcriptional repressors (Chen and Courey 2000; Cinnamon and Paroush 2008). HA tagged TLE4 isoforms (TLE4a, TLE4b and TLE4c) in the pCI (Promega) vector were a gift from Dr. Gareth Correy. Each TLE4 isoform was independently cotransfected with FOXC1-Xpress in Hela and HTM cells for dual luciferase reporter assays on the FOXC1 6x BS reporter. These results were compared to cells cotransfected with FOXC1-Xpress and empty HA vector. The amount of FOXC1 DNA and total DNA transfected was the same for each assay. As a vector control, empty Xpress and empty HA vectors were cotransfected and assayed for activity on the reporter. The experiments were performed two times in Hela cells and three times in HTM cells. To determine if TLE4 has an effect on the 6x BS reporter, cells were cotransfected with each of the isoforms independently with empty Xpress vector and assayed for activity on the reporter. These results were compared to the activity of cotransfected empty Xpress and empty HA on the reporter. As a positive control, FOXC1 cotransfected with

empty HA activity was also measured. Statistical significance was determined with a two-tailed t-test with unequal variances at a significance level of $\alpha = 0.05$. For each experiment, FOXC1 and TLE4 expression levels in simultaneously transfected 60mm plates of cells were measured by western blotting with anti-Xpress and anti-HA respectively. Immunoblotting of TFIID was used as a loading control. The level of FOXC1 expression was used to weight the transactivation activity observed. These experiments were performed following the same procedures as those described in Chapter 3.

Results

Identification and Functional Analysis of FOXC1 P297S

Routine sequencing of *FOXC1* in patients with anterior segment dysgenesis revealed a heterozygous 889C>T in two individuals (Figure 4-1) (screening was conducted by F Mirzayans). This change was not observed through sequencing in 50 control subjects therefore it is likely not a polymorphism. The transition, resulting in a proline to serine change at amino acid 297 (P297S), was created in a FOXC1-Xpress vector and its function was compared to wild type FOXC1. Immunoblotting of wild type FOXC1 and FOXC1 P297S samples revealed roughly equal expression, equal molecular weight and number of bands in proteins harvested from transfected Hela, HTM and Cos cells (Figure 4-2). Wild type FOXC1 is known to be located to the nucleus (Saleem et al. 2001) and here immunofluorescence of transfected Hela and HTM cells showed nuclear localization for both wild type and FOXC1 P297S (Figure 4-2). The DNA binding ability of FOXC1 P297S was equivalent to wild type as demonstrated by EMSA (Figure 4-2).

The half-life of FOXC1 P297S was 0.96x and 1.25x the half-life of wild type FOXC1 in two experiments utilizing HTM cells (Figure 4-3). These values are not significantly different from wild type FOXC1 ($P = 0.6$). However in Hela cells, the half-life of FOXC1 P297S was 1.45x the half-life of wild type FOXC1 in two experiments and this difference is statistically significant ($P = 0.0007$) (Figure 4-3).

Transactivation assays with the 6x BS reporter, which contains six ideal FOXC1 binding sites, were performed in Hela cells. FOXC1 P297S transactivation activity was 75% of wild type activity (P = 0.02) (Figure 4-4). On the FOXO1a reporter, which contains 830 nucleotides of the *FOXO1a* cis-regulatory region with two FOXC1 binding sites, FOXC1 P297S transactivation activity was 60% of wild type (P = 0.0006) (Figure 4-4).

The Effects of TLE4 on FOXC1 Transactivation Activity

The FoxC sequences analyzed in Chapter 3 contain an EH1 motif in the inhibitory domain, with the exception of frog FoxC2b and zebrafish foxc1.2 (Figure 4-5) (Copley 2005; Yaklichkin et al. 2007b). This motif has the potential to interact with TLE (transducin like enhancer of split) proteins to mediate repression of transactivation activity (Zhu et al. 2002; Jennings et al. 2006; Heimbucher et al. 2007; Yaklichkin et al. 2007a; Hoffman et al. 2008). There are five TLE paralogs in humans, TLE1, TLE2, TLE3, TLE4 and AES (amino terminal enhancer of split). AES is a dominant negative form that does not repress transcription. All TLE paralogs are broadly expressed in adult tissues (Stifani et al. 1992; Miyasaka et al. 1993) and their expression during mammalian development is still being elucidated. TLE4 has been shown to interact with and reduce Six3 activity and has potential to function in eye development (Zhu et al. 2002). Therefore TLE4 was examined as a potential corepressor to mediate FOXC1 inhibitory domain function. There are three TLE4 isoforms in humans:

TLE4a, TLE4b and TLE4c (Figure 4-6). Hela and HTM cells were cotransfected with FOXC1-Xpress and each of the TLE4-HA isoforms independently and transactivation of the FOXC1 6x BS reporter was measured (Figure 4-6). The addition of TLE4 resulted in a significant increase in transactivation activity of the reporter in HTM ($P = 0.002$ for TLE4a, $P = 0.001$ for TLE4b, $P = 0.006$ for TLE4c) but not in Hela cells. Assessment of TLE4 activity on the 6x BS reporter in the absence of FOXC1 in HTM cells demonstrated reduced activity when compared to the activity of the empty vectors (Figure 4-6). This decrease was statistically significant for TLE4b ($P = 0.03$) however the experiment was only performed once. Nevertheless, TLE4 does not appear to activate the 6x BS reporter.

Discussion

P297S is a Novel FOXC1 Mutation.

A newly identified FOXC1 alteration, P297S, was found in two unrelated individuals with anterior segment dysgenesis. Contiguously with publication of this work, a different study reported identification of P297S in two other unrelated individuals, both with microphthalmia, one with additional sclerocornea and one with additional contralateral myopia (Kaur et al. 2009). However Kaur et al. ruled out P297S as a causative mutation due to identification of the change in 19 out of 100 control individuals. P297S introduces a novel HgaI restriction site in FOXC1 and this was used by Kaur et al. to check controls for the change. Prior to sequencing I also attempted to use HgaI to check controls for P297S, however the enzyme was unreliable: not cutting when it was expected to cut on experimental positive controls or cutting at too many nonspecific sites creating a ladder effect. Therefore I opted to sequence 50 control individuals and did not find P297S. The difference in methods used to test controls for P297S likely accounts for the difference in identification in control samples. It is unlikely that the controls used in Kaur et al. contain 19 unrelated individuals with FOXC1 mutations however we don't know if the controls were examined for disease and it is possible that a family with FOXC1 mutations was included among the controls. The controls used in the work presented here were 50 unrelated individuals (spouses of affected individuals) with unknown phenotypes. P297S has not been reported as a SNP in the NCBI Entrez SNP database, Build 126, by any other studies (Wheeler et al. 2006). Most importantly, since the effects of P297S on FOXC1

function are consistent with those of other disease causing mutations, as discussed below, I can conclude that P297S is a disease causing mutation and not a polymorphism.

P297S Alters FOXC1 Transactivation Activity and Stability.

P297 is within the inhibitory domain of FOXC1, a region known to regulate transactivation activity and protein stability (Berry et al. 2002; Berry et al. 2006). This is the first identification of a missense mutation outside of the forkhead domain in FOXC1. Previous characterizations of disease causing missense mutations in FOXC1 have shown that mutations can affect nuclear localization, DNA binding, protein expression, phosphorylation, transactivation activity and can cause aggregate formation (Saleem et al. 2001; Saleem et al. 2003a; Saleem et al. 2003b; Murphy et al. 2004; Ito et al. 2007; Ito et al. 2009). The P297S mutation does not affect protein localization or DNA binding, two known forkhead domain functions. Additionally, FOXC1 P297S expression levels were not grossly altered as compared to wild type FOXC1 and aggregates were not observed by immunofluorescence. The P297S mutation creates a potential additional phosphorylation site in FOXC1, however differences in phosphorylation status or other post-translational modifications between P297S and wild type are unlikely, as the number and molecular weight of protein bands on immunoblots were not different between the two constructs. The transactivation activity of FOXC1 is significantly reduced by P297S on both

reporters tested. Analyses of *FOXC1* mutations and the presence of disease when *FOXC1* is duplicated have provided evidence that 78-150% of wild type transactivation activity is normal (Lehmann et al. 2000; Saleem et al. 2001). Here, the transactivation ability of FOXC1 P297S is ≤ 75% of wild type on two different reporters. This is consistent with other FOXC1 mutations as all other FOXC1 missense mutations that have been functionally characterized also reduce transactivation activity (Saleem et al. 2001; Saleem et al. 2003a; Saleem et al. 2003b; Murphy et al. 2004; Ito et al. 2007; Ito et al. 2009).

A previous study in Hela cells has demonstrated that an amino acid change, S272A, in the inhibitory domain of FOXC1 reduces protein stability (Berry et al. 2006) therefore the stability of FOXC1 P297S was assessed. Here, FOXC1 P297S has a half-life that is 45% longer than wild type in Hela cells, while in HTM cells the differences in half-life were not statistically significant. The differences in half-lives between wild type and P297S as well as between Hela and HTM cells, which are a cancer cell line and an ocular cell line respectively, may be due to differences in protein degradation pathways or signals utilized. The majority of proteins in eukaryotes are degraded by proteasomes (ubiquitin dependant or independent) or lysosomes (reviewed by: Knecht et al. 2009). Different pathways can be used for the same protein and the pathway choice is dependant on cell signaling (reviewed by: Attaix et al. 2001). FOXC1 is known to be polyubiquitinated and degraded by the 26S proteasome pathway in Hela cells (Berry et al. 2006). Phosphorylation of S272 prevents FOXC1 degradation while a C-terminal degron is required for ubiquitin dependant

degradation (Berry et al. 2006). P297S may potentially alter or interfere with these or other yet unidentified degradation or stabilization signal(s).

Missense Mutations Outside of the Forkhead Domain in Other FOX Genes Also Alter Transactivation Activity.

Missense mutations outside of the forkhead domain have also been identified in the forkhead genes *FOXA2*, *FOXC2*, *FOXE1*, *FOXE3*, *FOXI1*, *FOXL2* and *FOXP2*. The functional consequences of missense mutations outside of the forkhead domain have been briefly examined in FOXA2 (Zhu et al. 2000), FOXC2 (van Steensel et al. 2009), FOXI1 (Yang et al. 2007) and FOXL2 (Beysen et al. 2008). The nuclear localization of mutant FOXC2, FOXI1 and FOXL2 was shown to be normal. The FOXC2, FOXI1 and one FOXA2 mutation are located in transactivation domains and resulted in increased transactivation activity. A second FOXA2 mutation was located in a region with undefined function and did not alter transactivation. A FOXL2 mutation was also located in an uncharacterized region and displayed increased transactivation activity. The effects of disease causing mutations on DNA binding ability and protein stability have only been examined in FOXC1. A potential missense mutation (T368N) near the C-terminus of the FOXC1 inhibitory domain was identified by Kaur et al. however no controls were screened for the change. Although there is little data available, so far the effects of missense mutations outside of the forkhead domain on protein function are similar among all forkhead genes. Namely these mutations do not affect nuclear localization but do enhance the activity of the

domain in which they are found (i.e. if in a transactivation domain they increase activity, if in an inhibitory domain they increase inhibition).

P297S and FOXC1 Duplications May Cause Disease Through a Mechanism That is Different From That of Forkhead Domain Missense Mutations.

Interestingly, both P297S patients examined here have iridogoniodysgenesis, a malformation that is more commonly associated with FOXC1 duplications rather than mutations (Strungaru et al. 2007). There are three other FOXC1 mutations associated with iridogoniodysgenesis; Q23X (N-terminal to the forkhead domain) (Mirzayans et al. 2000), L86F (within the forkhead domain) (Saleem et al. 2003b) and a deletion of 10 base pairs resulting in a frameshift and stop codon after 10 amino acids (N-terminal to the forkhead domain) (Mears et al. 1998). FOXC1 L86F has reduced DNA binding ability and a 50% reduction in transactivation activity. The Q23X and 10 base pair deletion mutations would both produce a protein lacking the forkhead domain most likely rendering FOXC1 nonfunctional. In addition to iridogoniodysgenesis, all of the individuals with these mutations have Axenfeld Rieger Syndrome with posterior embryotoxon. Patients with P297S or FOXC1 duplications do not have posterior embryotoxon (Nishimura et al. 2001; Lehmann et al. 2002; Strungaru et al. 2007). Considering this phenotypic data in conjunction with the increase in inhibitory function observed in FOXC1 P297S, it suggests that FOXC1 duplications and mutations that disrupt the inhibitory domain may lead to disease through similar mechanisms and thus have more similar phenotypes as compared to disease

caused by missense mutations with reduced protein function. One potential mechanistic explanation is that FOXC1 P297S and FOXC1 duplications exert their effects through target gene(s) that are different from the target gene(s) affected in patients without iridogoniodysgenesis. In Chapter 3 I demonstrated that the effects of an amino acid substitution at A337 in FOXC1 on transactivation activity are dependant on FOXC1 binding site context, thus supporting the theory that different FOXC1 mutations will impact different target genes.

The Inhibitory Domain May Function By Blocking a Transactivation Domain and/or Binding Corepressors.

As discussed in Chapter 3, the inhibitory domain is not highly conserved among species and is unstructured. P297S occurs within a region that is proline rich in the human, mouse and chicken FoxC1 sequences (Figure 4-5). The fish and frog FoxC1 sequences and all of the FoxC2 sequences analyzed in Chapter 3 are not proline rich in this region, resulting in a span of gaps in the alignment that includes FOXC1 P297S. Therefore the role of P297 in FOXC1 function is specific to FoxC1 in higher organisms. Proline rich regions are known to mediate protein-protein interactions with signaling molecules and influence local protein conformation (reviewed by: Kay et al. 2000). Elucidation of the function of this proline rich region will help determine how the inhibitory domain functions and potentially identify a mechanism whereby FOXC1 activity is differentially regulated from FOXC2.

The mechanism by which the inhibitory domain of FOXC1 reduces transactivation activity has not been determined. In other transcription factors, inhibitory domains have been shown to lower transactivation activity by reducing DNA binding (Hagman and Grosschedl 1992; Giovane et al. 1994; Chan et al. 1996; Chen et al. 2009), binding to corepressors (Barr 2001; Wierstra and Alves 2006b; Seo et al. 2009) or binding to a transactivation domain (Serber et al. 2002; Wierstra and Alves 2006a). Multiple mechanisms may be used by an inhibitory domain to reduce activity, for example, the inhibitory region in FOXM1c binds retinoblastoma protein as a corepressor (Wierstra and Alves 2006b) as well as to its own transactivation domain in cis (Wierstra and Alves 2006a). FOXM1c also uses an additional mechanism to inhibit transactivation activity, binding of the inhibitory domain to itself in trans (Wierstra and Alves 2007). In Chapter 3 I used EMSAs to determine if the inhibitory domain of FOXC1 reduces DNA binding. Deletion of the inhibitory domain resulted in reduced FOXC1 binding suggesting that the inhibitory domain does not function by reducing DNA binding. Addition of the inhibitory domain fragment in trans did not alter FOXC1 DNA binding ability suggesting that the inhibitory domain does not physically interact with FOXC1 to alter DNA binding. To determine if the inhibitory domain physically interacts with a transactivation domain or other domains in FOXC1, I began using nickel pull down assays with full length FOXC1 or the inhibitory domain fragment attached to nickel agarose beads (data not shown). I applied a cell lysate containing full length V5 tagged FOXC1 (V5-FOXC1) to the beads and determined if V5-FOXC1 bound to FOXC1 or the inhibitory domain attached to

the beads. However, V5-FOXC1 bound to nickel agarose beads that did not have FOXC1 or the inhibitory domain attached rendering the experiments uninformative.

Finally, the inhibitory domain may mediate its effects through interaction(s) with corepressors. To date there are no known proteins that interact with the inhibitory domain. However, there is an EH1 motif adjacent to the proline rich region of the inhibitory domain in the FoxC sequences. EH1 motifs in transcription factors are known mediate interactions with transducin like enhancer of split (TLE) proteins resulting in a reduction of transcription factor activity (Zhu et al. 2002; Jennings et al. 2006; Heimbucher et al. 2007; Yaklichkin et al. 2007a; Hoffman et al. 2008). The mechanism of TLE repression has not been widely examined however TLEs have been shown to compete with coactivators and to recruit histone deacetylases to silence chromatin (reviewed by: Chen and Courey 2000; Cinnamon and Paroush 2008). Here the addition of TLE4 to FOXC1 transactivation assays resulted in an increase in FOXC1 transactivation activity. The effect of increased transactivation activity of a transcription factor upon addition of a TLE protein has been observed in two different reports involving TLE1 (Hentschke and Borgmeyer 2003; Riz et al. 2009). In both cases, TLE1 was shown to directly interact with the transcription factor of interest. It is unknown whether TLE1 is mediating activation or out competing an even stronger repressor. Additionally, an EH1 motif in *Drosophila* Dorsal has been shown to have transactivation activity (Flores-Saaib et al. 2001) and the yeast TLE homolog Tup1 interacts with an additional cofactor to activate

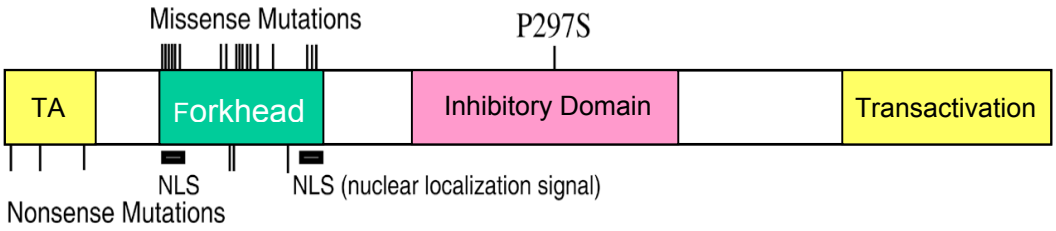
transcription in response to metabolic signals (Conlan et al. 1999). The observed increase in transactivation activity of FOXC1 upon the addition of TLE4 may be due to TLE4 mediation of activation or blocking of a different repressor. The increase in transactivation activity was observed in the HTM but not the Hela cell lines used here suggesting that additional cell specific cofactors are involved in this reaction. A similar phenomenon has been observed for FoxA2. A domain containing an EH1 motif in FoxA2 was shown to function as an activation domain in HepG2 cells (hepatocyte carcinoma line) and a repression domain in Hela cells (Wang et al. 2000). This EH1 motif was also shown to bind TLE1 resulting in reduced transactivation activity. It was speculated that the greater endogenous expression of TLE1 in Hela cells as compared to HepG2 cells is responsible for the differences in domain activity among the cell types. I attempted to determine if FOXC1 and TLE4 physically interact using nickel pull downs with FOXC1 attached to nickel agarose bead and cell lysates containing TLE4, however TLE4 bound to beads that did not contain FOXC1 (data not shown). Overall there is evidence that cell type specific cofactors are involved in the regulation of FOXC1 activity and the role of TLEs in this regulation is still unclear.

Conclusions

P297S is the first missense mutation identified outside of the forkhead domain in FOXC1. It is a recurrent mutation as it was present in two unrelated individuals with anterior segment dysgenesis. FOXC1 is under tight regulatory control, thus any disturbances to this regulation are likely to be pathogenic. Here I have demonstrated a reduction in transactivation ability and potentially extended half-life for FOXC1 P297S. It is likely that P297S alters FOXC1 interaction with other yet unidentified factors involved in transactivation and degradation. Both of these alterations may be disease causing and their effects may be tissue and timing of expression dependent. These analyses extend the possible pathological effects on protein function of mutations of FOXC1, namely stability. It also appears that mutations outside of the forkhead domain and FOXC1 duplications share a disease mechanism that is different from that of missense mutations within the forkhead domain. In general, mutations outside of the forkhead domain in any forkhead protein increase the activity of the domain in which they are found. Attempts to determine how the inhibitory domain decreases FOXC1 activity have revealed that the domain does not reduce binding. However, there is evidence that FOXC1 interacts with an unknown repressor as the addition of TLE4 increased transactivation activity.

Figures

A.



B.

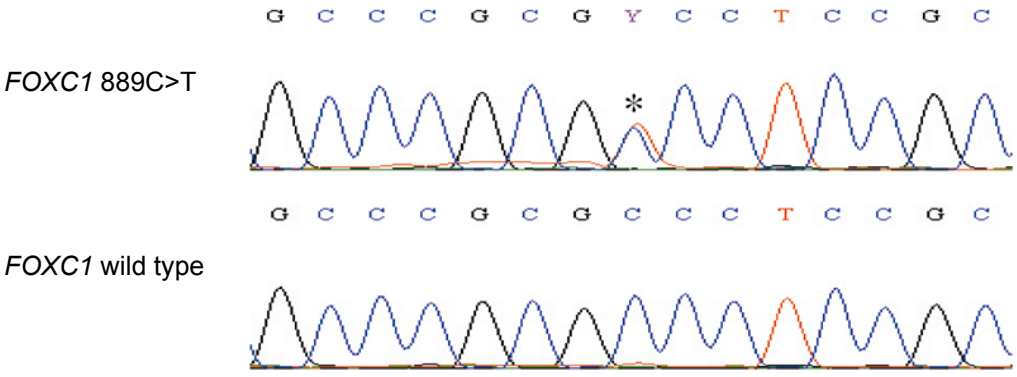
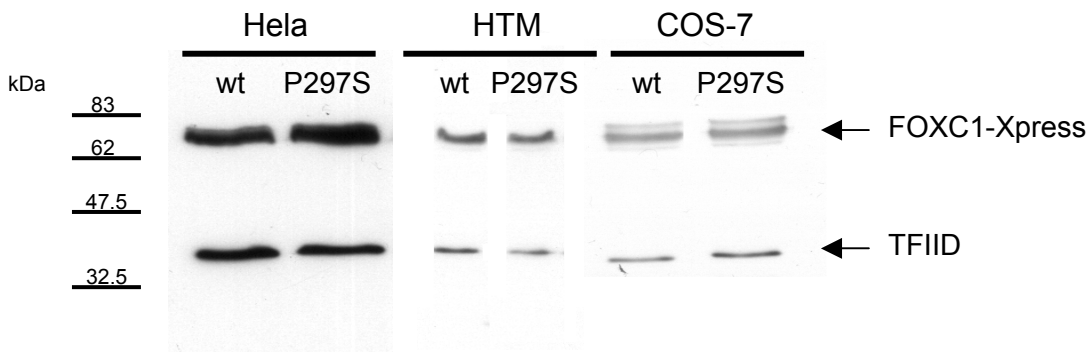


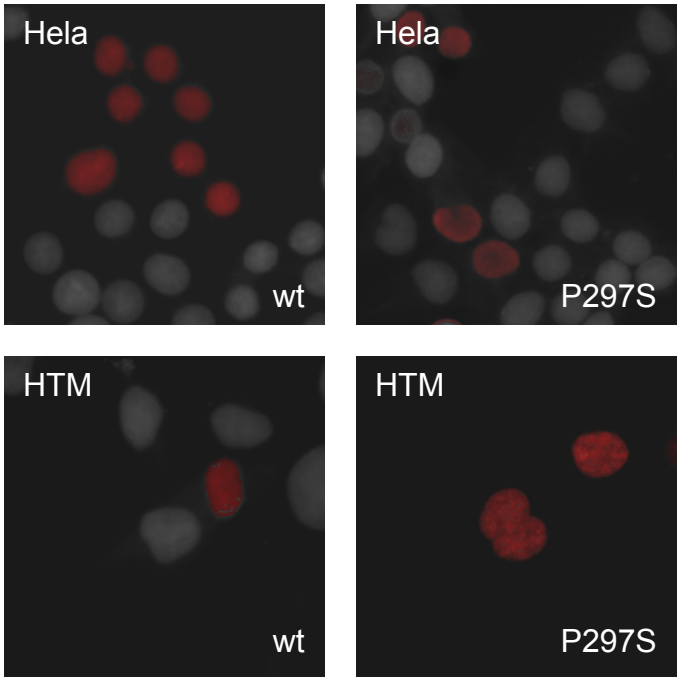
Figure 4-1. FOXC1 mutations and functional domains.

A. Schematic of FOXC1 functional domains (Saleem et al. 2001; Berry et al. 2002) with the location of known missense mutations shown above and nonsense mutations shown below the protein. Yellow: transactivation domain (TA) Green: forkhead domain Pink: inhibitory domain Black horizontal bar: nuclear localization signal (NLS) B. Chromatograms showing *FOXC1* 889C>T (indicated with a star) and wild type sequences.

A.


Hela
HTM
COS-7
wt
P297S
wt
P297S
wt
P297S
kDa
83
62
47.5
32.5
FOXC1-Xpress
TFIID

B.


Hela
wt
Hela
P297S
HTM
wt
HTM
P297S

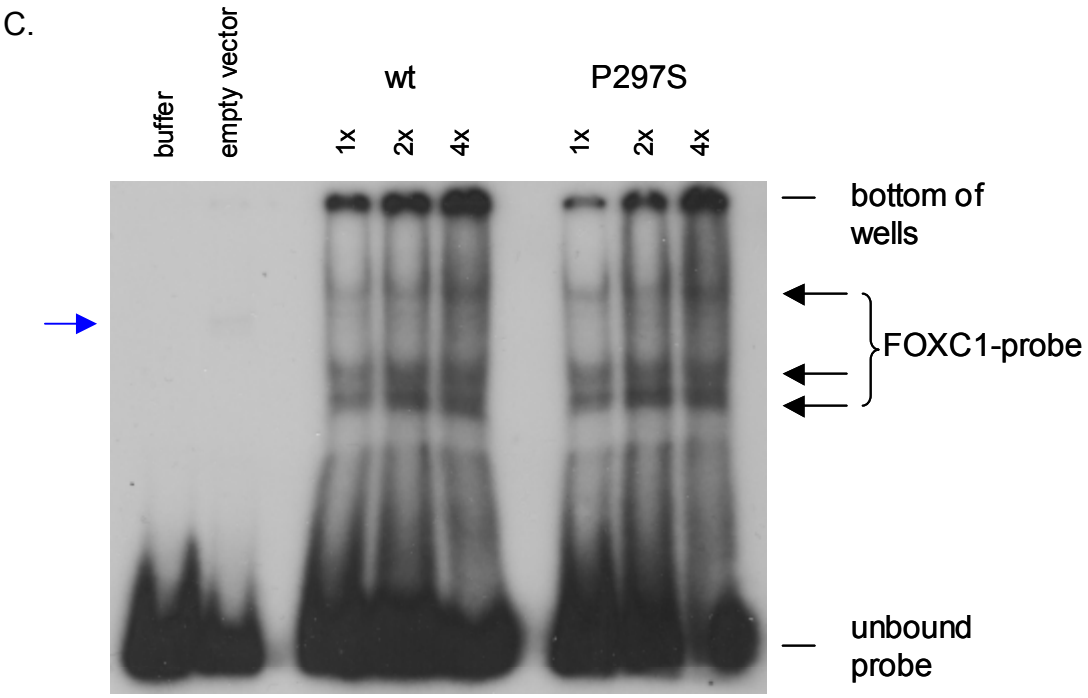
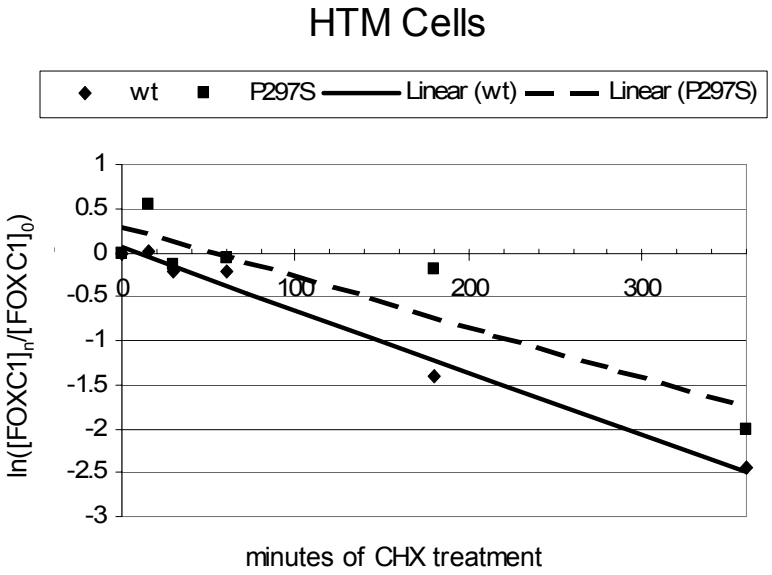


Figure 4-2. FOXC1 P297S does alter protein expression, localization or DNA binding.

A. Western blotting of cell lysates from cells transfected with wild type (wt) FOXC1 or P297S. Probing for TFIID serves as a gel loading control. B. Immunofluorescence of cells transfected with wt FOXC1 or P297S. Red = FOXC1 Grey = DAPI. C. EMSA with increasing amounts of wt FOXC1 or P297S and a radioactively labeled FOXC1 DNA binding site probe. FOXC1 was transfected into COS-7 cells and whole cell lysates were utilized. Panel A demonstrates relative protein expression levels. The buffer lane (cell lysis buffer is added but no cell lysate is included) shows no band shifts. The empty vector lane (cell lysates from cells transfected with empty vector) shows a faint band shift at a position that does not correspond with FOXC1-probe bands (blue arrow). This shift is probably due to endogenous proteins binding to the probe.

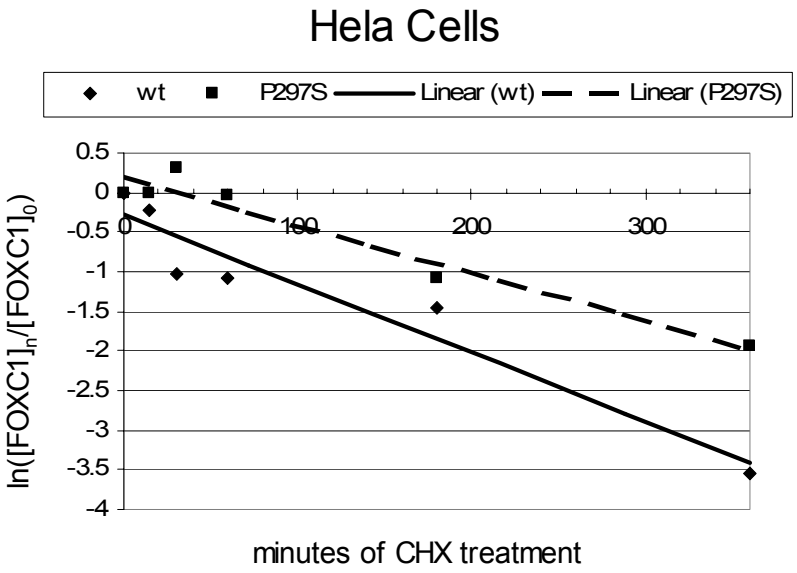
A.



Experiment	Sample	R^2	$T_{1/2}$ (min)	$T_{1/2}$ weighted by wt
1	wt	0.7089	108.28	1
	P297S	0.9864	103.43	0.96
2	wt	0.9848	97.61	1
	P297S	0.8401	121.58	1.25

ttest (wt vs P297S) P-value = 0.615

B.



Experiment	Sample	R^2	$T_{1/2}$(min)	$T_{1/2}$ weighted by wt
1	wt	0.9251	79.67	1
	P297S	0.9485	115.52	1.45
2	wt	0.8550	53.72	1
	P297S	0.8505	77.87	1.45

ttest (wt vs P297S) P-value = 0.0007

Figure 4-3. P297S increases FOXC1 half-life ($T_{1/2}$) in Hela cells but not in HTM cells.

Cells transfected with FOXC1 wild type (wt) or P297S were treated with cyclohexamide (CHX) for 0, 15, 30, 60, 180 or 360 minutes. The amount of FOXC1-Xpress remaining after each time point was measured by immunoblotting and densitometry analysis, weighted by a TFIID loading control, $\ln([FOXC1]_n/[FOXC1]_0)$ was calculated and plotted (n = time point). Linear regression was used to fit a line (Linear (wt or P297S)) to the data and the slope of the line was used to calculate $T_{1/2} = \ln(0.5)/\text{slope}$. R^2 = correlation coefficient Representative graphs are shown for HTM cells (A) and Hela cells (B).

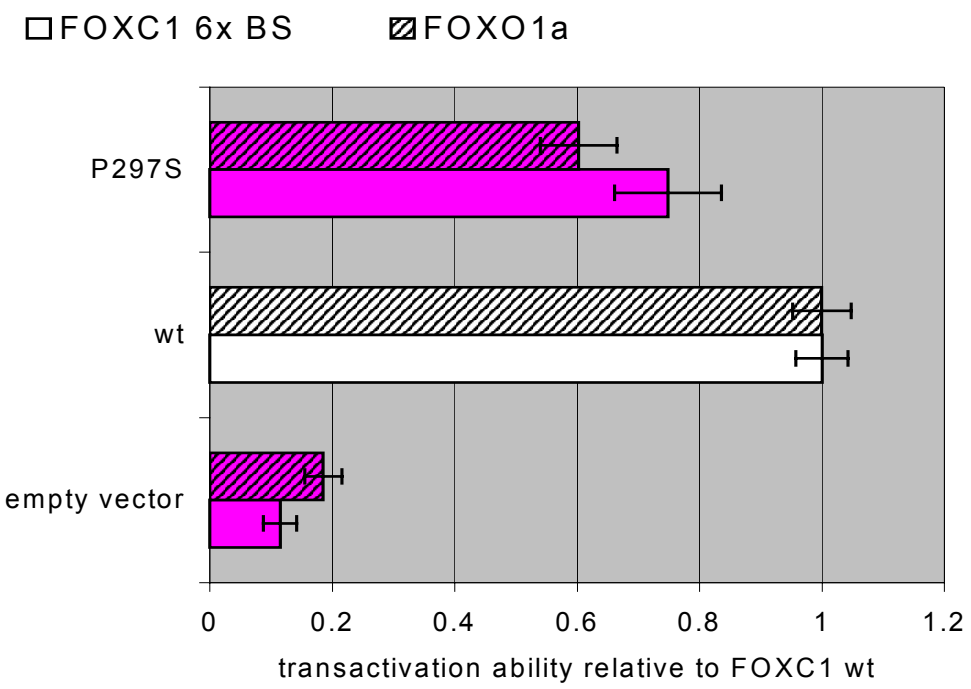


Figure 4-4. FOXC1 P297S transactivation activity is significantly less than wild type (wt) FOXC1.

The activity of wt FOXC1 is set to 1. Pink = statistically significant difference from wt Open bars = 6x BS reporter Hatched bars = FOXO1a reporter Standard errors are shown.

A.

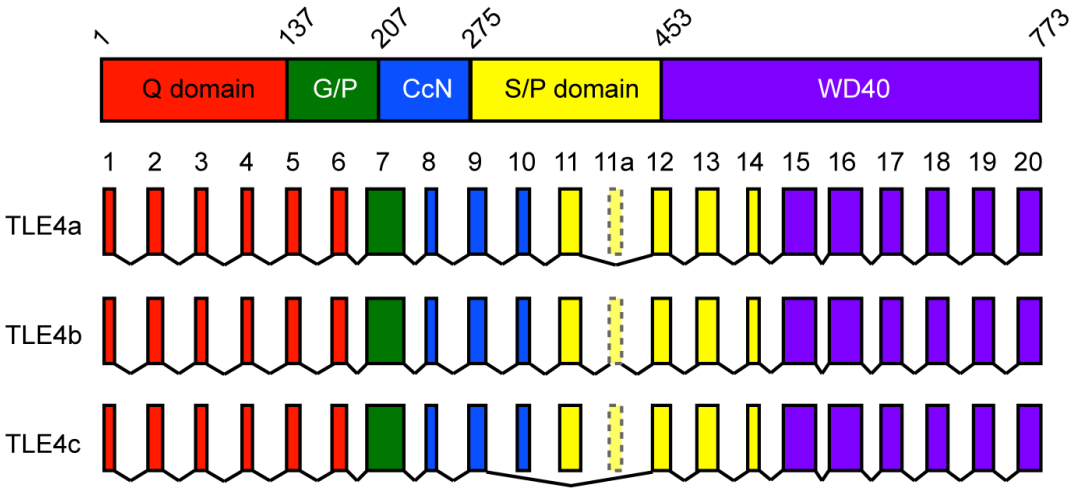
```
FOXC1  H sapiens     GPQPPPVRIQDIKTENGTCPSPPQPLSPAAALGSGSAAAVPKIESPDSSSSSLSSGSSPPGSLPSARPLSLDGADSAP 292
Foxc1  M musculus    GPQPPPVRIQDIKTENGTCPSPPQPLSPAAALGSGSAATVPKIESPDSSSSSLSSGSSPPGSLPSARPLSLDAAEPAP 294
FoxC1  G gallus      GAAPPPVRIQDIKTENGTA-SPPQAVSPPG--AAPPLGAVPKIESPDSSSSLSSGGSPRGGALPAARSLGMEGPEPPP 255
foxc1.1  D rerio     --GSQPVRIQDIKTENGTS-TPPQAVSPT-------LSTVPKIESPD-SSSSMSSGS--PHSIPSTRSLSLDS--AGE 267
foxc1.2  D rerio     -----PVKLRDIKTENGAC-TPPHDSTPP-------LSTVPKTESPDRSGGSACSGS------PQS------------ 247
UN  T nigroviridis   --GSKPVRIQDIKTENGTV-TPPQADSPS-------LNAVPKIESPD-SSS-MSSGS--PHSIPSNRSIGMDSSEHHQ 265
FoxC1  X laevis      DSNSQPVRIQDIKTENGTS-SPLQSMSPA-------LSAVPKIESPD-SSSSMSSGS--PHSIPSNRSMSLEAAESHH 286
FoxC1a  X laevis     DSNSQPVRIQDIKTENGTS-SPPQSMSPA-------LSAVPKIESPD-SSSSMSSGS--PHSIPSNRSMSLEAAESHH 286
FOXC2  H sapiens     --PHLADAPKE---AEKKVVIKSEAASPALP-------VITKVETLSPE--SALQGSPRSAASTPAGSPDGSLPEHHA 261
Foxc2  M musculus    --TPVADGPKE---AEKKVVVKSEAASPALP-------VITKVETLSPE--GALQASPRSASSTPAGSPDGSLPEHHA 259
FoxC2b  G gallus     --KEASSSSSSSASEKKVVIKSETASPELP-------VITKVETLSPGSGGGLRDSPRSGRIASRRLPR-RFAARAA 268
FoxC2a  X laevis     --AQGPISSLELPKHEKKIVIKSE--SPELP-------VITKVENLSPDGGSAMQDSPRSVASTPSVSTDNSIPDQHP 254
FoxC2b  X laevis     --VQGPVPSLELPKHEKKIIIKSE--SPELP-------VITKVENLSPGGGSAMQDSPRSVASTPSVSTDSSIPEQHP 254

FOXC1  H sapiens     PPPA----PSAPPPHHSQGFSVDNIMTSLRGSPQSAAAELSSGLLASAAASSRAGIAPPLALGAYSPGQS-SLYSSPCS 370
Foxc1  M musculus    PP------QPAPPPHHSQGFSVDNIMTSLRGSPQGSAAELGSGLLASAAASSRAGIAPPLALGAYSPGQS-SLYSSPCS 368
FoxC1  G gallus      PPPPPPPPPPPPPHHHGPGFSVDNIMTSLRGSPQAAELG--AGLLPPAAAAPRSSLPPALSLGAYSPGHSSAVYGSPCG 339
foxc1.1  D rerio     QHH----------QAPAQGFSVDNIMTSLRGSPHS-----SGELTPSLVAPSRTGITPTLSLN-YSPNQS-SVYSSPCS 332
foxc1.2  D rerio     ------------QTPQQAFSMDTIMTGLRGSPQH----------AAELPASRAALPGSVSLTYSPTPQP-AHYSPPCG 302
UN  T nigroviridis   HHG----------QASAQGFSVDNIMTSLRGSPQG-----PAELASGLPASTRTGITPSLSLN-YSPSQT-SVYSSTCN 330
FoxC1  X laevis      PHH----------QQHSQGFSVDNIMTSLRGSPQG-----SAELPSPLISSSRTGIAPSLSLS-YSPGQG-SIYSSPCS 351
FoxC1a  X laevis     PHH----------QQHSQGFSVDNIMTSLRGSPQG-----SAELPSPLISSSRTGIAPSLSLS-YSPGQG-SIYSSPCS 351
FOXC2  H sapiens     AAPN-----------GLPGFSVENIMT-LRTSPPGGELS------PGAGRAG--LVVPPLALP-YAAAP-PAAYGQPCA 322
Foxc2  M musculus    AAPN-----------GLPGFSVETIMT-LRTSPPGGDLS------PAAARAG--LVVPPLALP-YAAAP-PAAYTQPCA 320
FoxC2b  G gallus     PGGN-----------GLPGFSVENIMT-LRTS--AGDLS------PVSAASG--RTGTGMPLV-YPTSGQPSGYSAACS 328
FoxC2a  X laevis     ASN---------------GFSVENIMT-LRTSP-HGDLS------PVPQVPCRTGMVPSLPIN-YTAQTQSSVYSQACT 312
FoxC2b  X laevis     ASN---------------GFSVDNSMT-RRTSP-HGDLS------PVPAIPCRTAMVSSLPIN-YTAHTQSSVYSQACT 312
```

EH1 motif: XXaXbXXcdXX where X can be any amino acid
a can be Phe (F), His (H), Tyr (Y) or Trp (W)
b and c can be Ile (I), Leu (L) or Val (V)
d can be Glu (E), Phe (F), His (H), Ile (I), Lys (K), Met (M), Gln (Q), Arg (R), Trp (W), Tyr (Y)

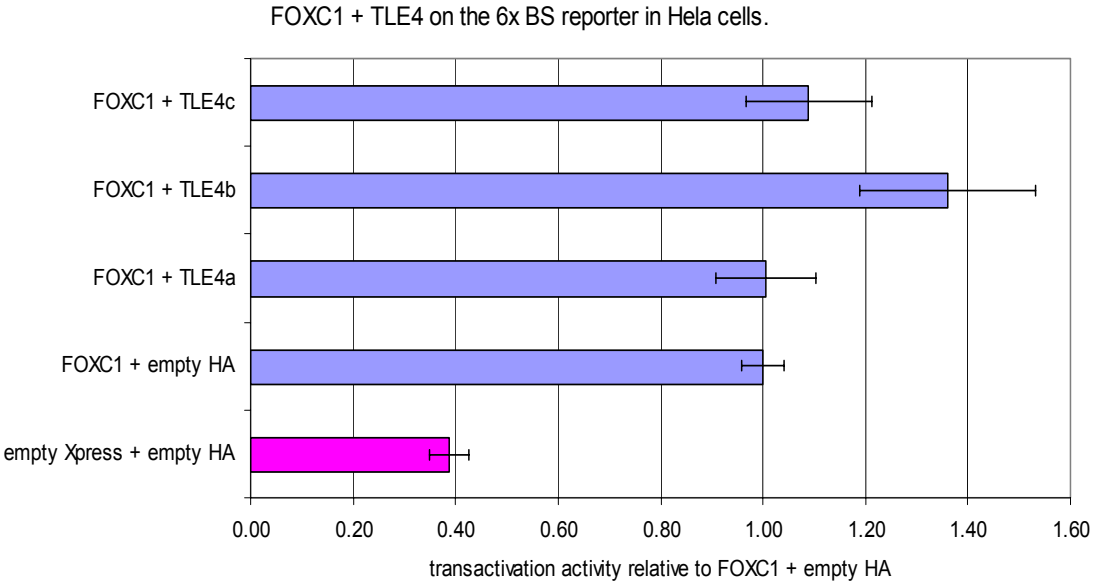
Figure 4-5. Alignment of the FOXC1 inhibitory domain with other FoxC1 and FoxC2 sequences.

This alignment is the same one created in Chapter 3, the sequences have been rearranged to place FOXC1 first. Only the inhibitory domain, as defined in FOXC1, is shown. P297 is marked in red in FOXC1. The EH1 motif is highlighted in blue and the EH1 motif definition is shown below the alignment (Smith and Jaynes 1996; Copley 2005).

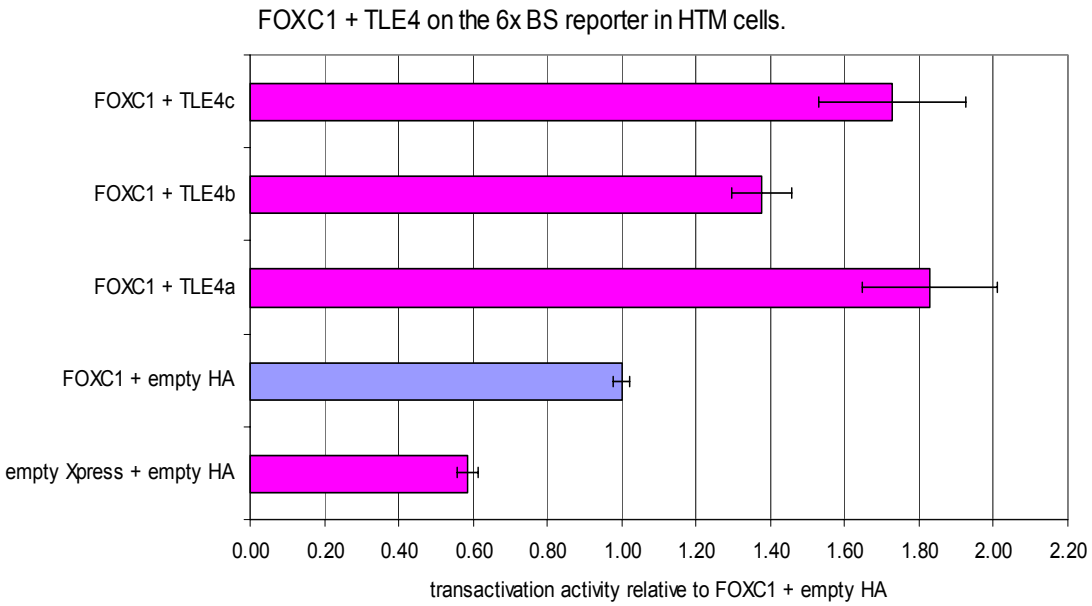
A.


1
137
207
275
453
773
Q domain
G/P
CcN
S/P domain
WD40
1 2 3 4 5 6 7 8 9 10 11 11a 12 13 14 15 16 17 18 19 20
TLE4a
TLE4b
TLE4c

B.


FOXC1 + TLE4 on the 6x BS reporter in Hela cells.
FOXC1 + TLE4c
FOXC1 + TLE4b
FOXC1 + TLE4a
FOXC1 + empty HA
empty Xpress + empty HA
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
transactivation activity relative to FOXC1 + empty HA

C.


FOXC1 + TLE4 on the 6x BS reporter in HTM cells.
FOXC1 + TLE4c
FOXC1 + TLE4b
FOXC1 + TLE4a
FOXC1 + empty HA
empty Xpress + empty HA
0.00 0.20 0.40 0.60 0.80 1.00 1.20 1.40 1.60 1.80 2.00 2.20
transactivation activity relative to FOXC1 + empty HA

D.

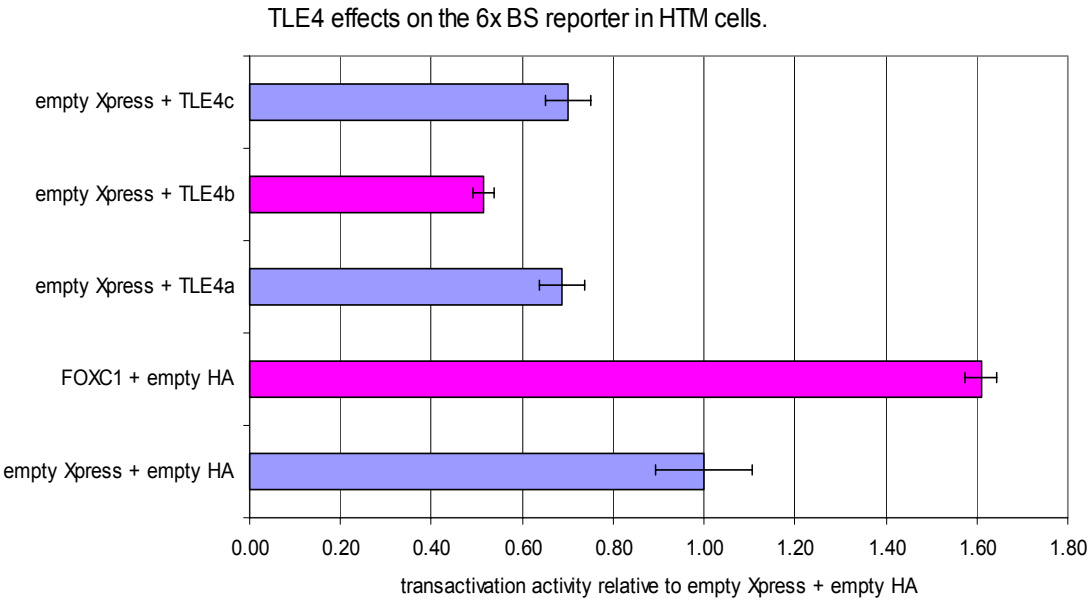

TLE4 effects on the 6x BS reporter in HTM cells.
empty Xpress + TLE4c
empty Xpress + TLE4b
empty Xpress + TLE4a
FOXC1 + empty HA
empty Xpress + empty HA
0.00 0.20 0.40 0.60 0.80 1.00 1.20 1.40 1.60 1.80
transactivation activity relative to empty Xpress + empty HA

Figure 4-6. The presence of TLE4 proteins increases FOXC1 transactivation activity in HTM cells but not Hela cells.

A. Schematic of the three TLE4 isoforms naturally created through alternate splicing (courtesy of Dr Gareth Cory). The exon composition of each isoform is shown. General functions of the protein domains are as follows: the Q domain (red) is involved in tetramerization and other protein interactions, the GP domain (green) can interact with the histone deacetylase HDAC1, the CcN domain (blue) may be phosphorylated and function in nuclear localization, the SP domain (yellow) may function in repression however it is not well studied and the WD40 domain (purple) can bind to other proteins (Stifani et al. 1992; Brantjes et al. 2001; Rave-Harel et al. 2005; Jennings et al. 2006; Heimbucher et al. 2007). The addition of TLE4 to transactivation assays of FOXC1 on the 6x BS reporter has no effect on reporter activation in Hela cells (B) and increases reporter expression in HTM cells (C). For both assays empty Xpress + empty HA had significantly less activity than FOXC1 + empty HA ($P < 0.0001$ for B and C). D. TLE4 alone did not activate the 6x BS reporter in HTM cells. As a positive control, FOXC1 + empty HA demonstrated significant activation of the reporter, $P = 0.02$.

References

Attaix D, Combaret L, Pouch M-N, Taillandier D (2001) Cellular control of protein degradation. International Animal Agriculture and Food Science Conference, Indianapolis, IN, USA

Barr FG (2001) Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. Oncogene 20:5736-5746

Berry FB, Mirzayans F, Walter MA (2006) Regulation of FOXC1 stability and transcriptional activity by an epidermal growth factor-activated mitogen-activated protein kinase signaling cascade. J Biol Chem 281:10098-10104

Berry FB, Saleem RA, Walter MA (2002) FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain. J Biol Chem 277:10292-10297

Beysen D, Moumne L, Veitia R, Peters H, Leroy BP, De Paepe A, De Baere E (2008) Missense mutations in the forkhead domain of FOXL2 lead to subcellular mislocalization, protein aggregation and impaired transactivation. Hum Mol Genet 17:2030-2038

Brantjes H, Roose J, van De Wetering M, Clevers H (2001) All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res 29:1410-1419

Cella W, de Vasconcellos JP, de Melo MB, Kneipp B, Costa FF, Longui CA, Costa VP (2006) Structural assessment of PITX2, FOXC1, CYP1B1, and GJA1 genes in patients with Axenfeld-Rieger syndrome with developmental glaucoma. Invest Ophthalmol Vis Sci 47:1803-1809

Chan SK, Popperl H, Krumlauf R, Mann RS (1996) An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. EMBO J 15:2476-2487

Chen G, Courey AJ (2000) Groucho/TLE family proteins and transcriptional repression. Gene 249:1-16

Chen LW, Raghavan V, Chang PJ, Shedd D, Heston L, Delecluse HJ, Miller G (2009) Two phenylalanines in the C-terminus of Epstein-Barr virus Rta protein reciprocally modulate its DNA binding and transactivation function. Virology 386:448-461

Cinnamon E, Paroush Z (2008) Context-dependent regulation of Groucho/TLE-mediated repression. Curr Opin Genet Dev 18:435-440

Conlan RS, Gounalaki N, Hatzis P, Tzamarias D (1999) The Tup1-Cyc8 protein complex can shift from a transcriptional co-repressor to a transcriptional co-activator. J Biol Chem 274:205-210

Copley RR (2005) The EH1 motif in metazoan transcription factors. BMC Genomics 6:169

Flores-Saaib RD, Jia S, Courey AJ (2001) Activation and repression by the C-terminal domain of Dorsal. Development 128:1869-1879

Fuse N, Takahashi K, Yokokura S, Nishida K (2007) Novel mutations in the FOXC1 gene in Japanese patients with Axenfeld-Rieger syndrome. Mol Vis 13:1005-1009

Giovane A, Pintzas A, Maira SM, Sobieszczuk P, Wasylyk B (1994) Net, a new ets transcription factor that is activated by Ras. Genes Dev 8:1502-1513

Hagman J, Grosschedl R (1992) An inhibitory carboxyl-terminal domain in Ets-1 and Ets-2 mediates differential binding of ETS family factors to promoter sequences of the mb-1 gene. Proc Natl Acad Sci U S A 89:8889-8893

Heimbucher T, Murko C, Bajoghli B, Aghaallaei N, Huber A, Stebegg R, Eberhard D, Fink M, Simeone A, Czerny T (2007) Gbx2 and Otx2 interact with the WD40 domain of Groucho/Tle corepressors. Mol Cell Biol 27:340-351

Hentschke M, Borgmeyer U (2003) Identification of PNRC2 and TLE1 as activation function-1 cofactors of the orphan nuclear receptor ERRgamma. Biochem Biophys Res Commun 312:975-982

Hoffman BG, Zavaglia B, Beach M, Helgason CD (2008) Expression of Groucho/TLE proteins during pancreas development. BMC Dev Biol 8:81

Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, Walter MA (2009) Severe molecular defects of a novel FOXC1 W152G mutation result in aniridia. Invest Ophthalmol Vis Sci 50:3573-3579

Ito YA, Footz TK, Murphy TC, Courtens W, Walter MA (2007) Analyses of a novel L130F missense mutation in FOXC1. Arch Ophthalmol 125:128-135

Jennings BH, Pickles LM, Wainwright SM, Roe SM, Pearl LH, Ish-Horowicz D (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. Mol Cell 22:645-655

Kaur K, Ragge NK, Ragoussis J (2009) Molecular analysis of FOXC1 in subjects presenting with severe developmental eye anomalies. Mol Vis 15:1366-1373

Kawase C, Kawase K, Taniguchi T, Sugiyama K, Yamamoto T, Kitazawa Y, Alward WL, Stone EM, Nishimura DY, Sheffield VC (2001) Screening for

mutations of Axenfeld-Rieger syndrome caused by FOXC1 gene in Japanese patients. J Glaucoma 10:477-482

Kay BK, Williamson MP, Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J 14:231-241

Knecht E, Aguado C, Carcel J, Esteban I, Esteve JM, Ghislat G, Moruno JF, Vidal JM, Saez R (2009) Intracellular protein degradation in mammalian cells: recent developments. Cell Mol Life Sci 66:2427-2443

Komatireddy S, Chakrabarti S, Mandal AK, Reddy AB, Sampath S, Panicker SG, Balasubramanian D (2003) Mutation spectrum of FOXC1 and clinical genetic heterogeneity of Axenfeld-Rieger anomaly in India. Mol Vis 9:43-48

Lehmann OJ, Ebenezer ND, Ekong R, Ocaka L, Mungall AJ, Fraser S, McGill JI, Hitchings RA, Khaw PT, Sowden JC, Povey S, Walter MA, Bhattacharya SS, Jordan T (2002) Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions. Invest Ophthalmol Vis Sci 43:1843-1849

Lehmann OJ, Ebenezer ND, Jordan T, Fox M, Ocaka L, Payne A, Leroy BP, Clark BJ, Hitchings RA, Povey S, Khaw PT, Bhattacharya SS (2000) Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma. Am J Hum Genet 67:1129-1135

Mears AJ, Jordan T, Mirzayans F, Dubois S, Kume T, Parlee M, Ritch R, Koop B, Kuo WL, Collins C, Marshall J, Gould DB, Pearce W, Carlsson P, Enerback S, Morissette J, Bhattacharya S, Hogan B, Raymond V, Walter MA (1998) Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. Am J Hum Genet 63:1316-1328

Mirzayans F, Gould DB, Héon E, Billingsley GD, Cheung JC, Mears AJ, Walter MA (2000) Axenfeld-Rieger syndrome resulting from mutation of the *FKHL7* gene on chromosome 6p25. European Journal of Human Genetics 8:71-74

Miyasaka H, Choudhury BK, Hou EW, Li SS (1993) Molecular cloning and expression of mouse and human cDNA encoding AES and ESG proteins with strong similarity to Drosophila enhancer of split groucho protein. Eur J Biochem 216:343-352

Mortemousque B, Amati-Bonneau P, Couture F, Graffan R, Dubois S, Colin J, Bonneau D, Morissette J, Lacombe D, Raymond V (2004) Axenfeld-Rieger anomaly: a novel mutation in the forkhead box C1 (FOXC1) gene in a 4-generation family. Arch Ophthalmol 122:1527-1533

Murphy TC, Saleem RA, Footz T, Ritch R, McGillivray B, Walter MA (2004) The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions. Invest Ophthalmol Vis Sci 45:2531-2538

Nishimura DY, Searby CC, Alward WL, Walton D, Craig JE, Mackey DA, Kawase K, Kanis AB, Patil SR, Stone EM, Sheffield VC (2001) A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. Am J Hum Genet 68:364-372

Nishimura DY, Swiderski RE, Alward WL, Searby CC, Patil SR, Bennet SR, Kanis AB, Gastier JM, Stone EM, Sheffield VC (1998) The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. Nat Genet 19:140-147

Panicker SG, Sampath S, Mandal AK, Reddy AB, Ahmed N, Hasnain SE (2002) Novel mutation in FOXC1 wing region causing Axenfeld-Rieger anomaly. Invest Ophthalmol Vis Sci 43:3613-3616

Rave-Harel N, Miller NL, Givens ML, Mellon PL (2005) The Groucho-related gene family regulates the gonadotropin-releasing hormone gene through interaction with the homeodomain proteins MSX1 and OCT1. J Biol Chem 280:30975-30983

Riz I, Lee HJ, Baxter KK, Behnam R, Hawley TS, Hawley RG (2009) Transcriptional activation by TLX1/HOX11 involves Gro/TLE corepressors. Biochem Biophys Res Commun 380:361-365

Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA (2001) Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1. Am J Hum Genet 68:627-641

Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA (2003a) Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1. Hum Mol Genet 12:2993-3005

Saleem RA, Murphy TC, Liebmann JM, Walter MA (2003b) Identification and analysis of a novel mutation in the FOXC1 forkhead domain. Invest Ophthalmol Vis Sci 44:4608-4612

Seo HW, Kim EJ, Na H, Lee MO (2009) Transcriptional activation of hypoxia-inducible factor-1alpha by HDAC4 and HDAC5 involves differential recruitment of p300 and FIH-1. FEBS Lett 583:55-60

Serber Z, Lai HC, Yang A, Ou HD, Sigal MS, Kelly AE, Darimont BD, Duijf PH, Van Bokhoven H, McKeon F, Dotsch V (2002) A C-terminal inhibitory domain controls the activity of p63 by an intramolecular mechanism. Mol Cell Biol 22:8601-8611

Smith ST, Jaynes JB (1996) A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. Development 122:3141-3150

Stifani S, Blaumueller CM, Redhead NJ, Hill RE, Artavanis-Tsakonas S (1992) Human homologs of a Drosophila Enhancer of split gene product define a novel family of nuclear proteins. Nat Genet 2:119-127

Strungaru MH, Dinu I, Walter MA (2007) Genotype-phenotype correlations in Axenfeld-Rieger malformation and glaucoma patients with FOXC1 and PITX2 mutations. Invest Ophthalmol Vis Sci 48:228-237

Suzuki T, Takahashi K, Kuwahara S, Wada Y, Abe T, Tamai M (2001) A novel (Pro79Thr) mutation in the FKHL7 gene in a Japanese family with Axenfeld-Rieger syndrome. Am J Ophthalmol 132:572-575

van Steensel MA, Damstra RJ, Heitink M, Bladergroen RS, Veraart J, Steijlen PM, van Geel M (2009) Novel missense mutations in the FOXC2 gene alter transcriptional activity. Hum Mutat 30:E1002-1009

Wang JC, Waltner-Law M, Yamada K, Osawa H, Stifani S, Granner DK (2000) Transducin-like enhancer of split proteins, the human homologs of Drosophila groucho, interact with hepatic nuclear factor 3beta. J Biol Chem 275:18418-18423

Weisschuh N, Dressler P, Schuettauf F, Wolf C, Wissinger B, Gramer E (2006) Novel mutations of FOXC1 and PITX2 in patients with Axenfeld-Rieger malformations. Invest Ophthalmol Vis Sci 47:3846-3852

Weisschuh N, Wolf C, Wissinger B, Gramer E (2008) A novel mutation in the FOXC1 gene in a family with Axenfeld-Rieger syndrome and Peters' anomaly. Clin Genet

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, et al. (2006) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 34:D173-180

Wierstra I, Alves J (2006a) Despite its strong transactivation domain, transcription factor FOXM1c is kept almost inactive by two different inhibitory domains. Biol Chem 387:963-976

Wierstra I, Alves J (2006b) Transcription factor FOXM1c is repressed by RB and activated by cyclin D1/Cdk4. Biol Chem 387:949-962

Wierstra I, Alves J (2007) The central domain of transcription factor FOXM1c directly interacts with itself in vivo and switches from an essential to an inhibitory domain depending on the FOXM1c binding site. Biol Chem 388:805-818

Yaklichkin S, Steiner AB, Lu Q, Kessler DS (2007a) FoxD3 and Grg4 physically interact to repress transcription and induce mesoderm in Xenopus. J Biol Chem 282:2548-2557

Yaklichkin S, Vekker A, Stayrook S, Lewis M, Kessler DS (2007b) Prevalence of the EH1 Groucho interaction motif in the metazoan Fox family of transcriptional regulators. BMC Genomics 8:201

Yang T, Vidarsson H, Rodrigo-Blomqvist S, Rosengren SS, Enerback S, Smith RJ (2007) Transcriptional control of SLC26A4 is involved in Pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (DFNB4). Am J Hum Genet 80:1055-1063

Zhu CC, Dyer MA, Uchikawa M, Kondoh H, Lagutin OV, Oliver G (2002) Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. Development 129:2835-2849

Zhu Q, Yamagata K, Yu L, Tomura H, Yamada S, Yang Q, Yoshiuchi I, Sumi S, Miyagawa J, Takeda J, Hanafusa T, Matsuzawa Y (2000) Identification of missense mutations in the hepatocyte nuclear factor-3beta gene in Japanese subjects with late-onset Type II diabetes mellitus. Diabetologia 43:1197-1200

Chapter 5: Discussion and Future Directions

Discussion

In Chapters 2 and 3 I assessed the selection pressures on the FoxA, FoxC, FoxD, FoxI, FoxO and FoxP subfamilies. The results allowed for the prediction of functional domains, domain boundaries and residues that differentiate orthologs and paralogs. It appears that after forkhead gene duplication, selective restraints were relaxed enough to allow for differentiation into subfamilies. Neutral changes, positive selection and gaps in the alignments allow for paralog diversification.

In Chapter 3 I identified positive selection at one amino acid site in the FoxC subfamily. Mutation of the site resulted in altered FOXC1 transactivation activity and DNA binding. The mutant effects are the same as those seen in disease causing FOXC1 mutations therefore providing biological evidence for positive selection. The mutant effects on transactivation activity were dependant on the cis-regulatory region of the reporter used. Examination of the upstream region of the FoxC1 target, *FoxO1a*, for FoxC1 binding sites showed that the location and sequence composition of binding sites varies among species. These results provide evidence that changes in the cis-regulatory region of FoxC target genes as well as the FoxC proteins can alter regulation in a species specific manner.

In Chapters 3 and 4 I assessed the effects of mutations within the FOXC1 inhibitory domain on FOXC1 function. For the first time, mutation and deletion of the inhibitory domain was shown to alter FOXC1 DNA binding. The first FOXC1 missense mutation outside of the forkhead domain, P297S, was identified

and characterized. This mutation altered FOXC1 transactivation activity and stability.

Evidence That Alterations in Both Proteins and the Cis-Regulatory Regions of Genes Contribute to Evolutionary Change is Provided by the Forkhead Gene Family

Currently there is some question as to whether changes in proteins or changes in the cis-regulatory regions of genes contribute more to adaptive evolution. Analysis of the forkhead gene family has provided evidence that both types of change are present and not mutually exclusive within a transcription factor gene family and its targets.

The genome composition of forkhead genes varies among species. Phylogenetic analyses have demonstrated expansion of the family that is concurrent with increasing organism complexity (Figure 1-6) (Kaestner et al. 2000; Mazet et al. 2003; Adell and Muller 2004; Lee and Frasch 2004; Magie et al. 2005; Szilagyi et al. 2005; Tu et al. 2006; Larroux et al. 2008). Differences in forkhead domain sequence among family members have allowed for the establishment of subfamilies whose origins can be traced through the family expansion (Kaestner et al. 2000; Larroux et al. 2008). In Chapter 2 I used BLASTCLUST to cluster full length forkhead sequences into groups with 30% identity over 90% of their length. The resulting clusters concurred with previously identified subfamily divisions that were created using the forkhead domain alone. Identification of selection pressures in the FoxA, FoxC, FoxD,

FoxI, FoxO and FoxP subfamilies in Chapters 2 and 3 identified regions of conservation (negative selection) as well as regions of change (alignment gaps, neutral changes, positive selection) outside of the forkhead domain in each of the subfamilies. Sequence differences among family members allow for differential target selection and regulation (Pani et al. 1992; Li et al. 2004; Stroud et al. 2006). All together this data shows that forkhead proteins have differentiated from one another during family expansion while maintaining enough sequence similarities to establish family and subfamily relationships.

The differentiation of forkhead genes into subfamilies combined with differences in family member complement among species suggests that forkhead protein changes contribute to the evolution of species. Evidence supporting this theory comes from each thesis chapter. In Chapter 2 a domain important for transactivation activity (conserved domain III) was shown to be conserved in the Deuterostomia but not the Protostomia lineage of the FoxA subfamily. Additionally, positive selection was observed in the forkhead domain of the lineage separating the Protostomia/Cephalochordata from the Craniata. These species specific changes both have the potential to create differential activity of the FoxA proteins among species. In Chapter 3, the identification of positive selection in FoxC transcription factors and biological support obtained in FOXC1 mutational analyses indicates that the amino acid composition of the positively selected site is due to species specific requirements for FoxC activity or regulation. Altering the positively selected site can change FOXC1 transactivation activity and DNA binding ability. Therefore in other species,

which have an amino acid that is different from human at the positively selected site, FoxC transactivation activity and DNA binding ability may be altered. This would allow for differential regulation of FoxC target genes among species. Work in Chapters 3 and 4 shows that P297 is not conserved among FoxC subfamily members however the P297S mutation in FOXC1 reduces FOXC1 transactivation activity and causes disease. If mutation of a non-conserved amino acid can cause disease, then any sequence differences among species have the potential to alter protein function in a species specific manner.

Additional evidence that differentiation of forkhead proteins contributes to speciation comes from studies of FOXP2. Analysis of FoxP2 among mammals has identified two amino acids that may be under positive selection (Enard et al. 2002). In one study the two amino acids were mutated in human FOXP2 to the amino acids found at these sites in chimpanzee FoxP2 (Konopka et al. 2009). Microarray analyses of total RNA from a human neuronal cell line expressing either FOXP2 or mutated FOXP2 revealed that changing these two sites resulted in differential expression of over 100 different target genes. A comparison of gene expression between human and chimp brain tissue identified expression differences between the species that overlapped with those seen in the cell line experiment. Therefore amino acid changes in FoxP2 allow for the differential regulation of genes between humans and chimps. A second study mutated the two potentially positively selected amino acids in endogenous mouse Foxp2 to those found in human FOXP2 (Enard et al. 2009). The mutant mice were shown to have altered ultrasonic vocalizations, decreased exploratory behavior and

altered neuronal growth as compared to wild type mice. This is evidence that protein changes among species can lead to phenotypic differences among species.

There is evidence that other transcription factors have species specific activity as well. First, transcription factors have been shown to have species specific targets through ChIP (chromatin immunoprecipitation)-chip analyses. Human and mouse transcription factors HNF1A (hepatocyte nuclear factor 1A), HNF4A and HNF6 bind species specific as well as shared target genes (Odom et al. 2007). The most commonly occupied DNA sequence for each transcription factor varied slightly between human and mouse. The transcription factor CEBPA (CCAAT/enhancer-binding protein alpha) has also been shown bind species specific targets in human, mouse, dog, opossum and chicken (Schmidt et al. 2010). Second, comparison of gene expression profiles with microarrays or RT-PCR among species have demonstrated species specific gene expression. For example, comparison of the genes expressed in the lacrimal gland between human and mouse demonstrated species specific gene expression as well as differences in the abundance of common transcripts (Ozyildirim et al. 2005). Differences in gene expression between human and mouse kidney (Si et al. 2009) and embryonic stem cells (Ginis et al. 2004) have also been observed. Finally, attempts to rescue mutant phenotypes caused by loss of a transcription factor in one species with a homologous gene from another species have shown that complete rescue is not always possible. As an example, a null mutation of *tin* (tinman) in *Drosophila* results in failure of heart and visceral mesoderm formation (Ranganayakulu et al. 1998). The mouse homolog of *tin*, *Nkx2-5* (NK2 transcription factor related,

locus 5), is important for normal cardiac morphology and myogenesis but is not involved in visceral mesoderm formation (Lyons et al. 1995). Introduction of mouse *Nkx2-5*, into *Drosophila tin* mutant germlines resulted in rescue of the visceral mesoderm but not heart, demonstrating divergence of transcription factor function (Ranganayakulu et al. 1998). These examples provide evidence that there are species differences in orthologous transcription factor function, however the relative contribution of change in proteins compared to change in cis-regulatory regions of target genes to these differences is unknown.

The contribution of changes in the cis-regulatory regions of genes to evolution has not been as widely examined as the contribution of protein changes. In one study, the DNA occupancy of human HNF1A, HNF4A and HNF6 on human chromosome 21 and the orthologous mouse genes on mouse chromosome 16 was shown to be different (Wilson et al. 2008). Interestingly, within a mouse carrying human chromosome 21 the DNA occupancy of mouse HNF1A, HNF4A and HNF6 on human chromosome 21 was equivalent to the occupancy of the human genes on chromosome 21 within a human cell. This demonstrates that species differences in the cis-regulatory regions of genes exist and can change transcription factor occupancy of DNA in the absence of protein changes.

In Chapter 3, analysis of the effects of mutations at the positively selected site in FOXC1 on transactivation activity of different reporters demonstrated that changes in the cis-regulatory region of a gene can change FOXC1 activity. For example, when compared to wild type FOXC1, FOXC1 A337P has decreased transactivation activity on one reporter but no change in activity on a different

reporter. The cis-regulatory region of a known FoxC1 target, FoxO1a was also shown to be different among species. These data show that species specific differences in the regulatory region of forkhead target genes exist and can potentially lead to species specific gene regulation.

Overall there is evidence that differences in forkhead gene family member sequence as well differences in the cis-regulatory region of forkhead target genes can both contribute to evolutionary change.

Non-Conservation of a Sequence Does Not Mean That it is Unimportant for Function

It is common to perform sequence comparisons among orthologs and paralogs when looking for functionally important regions in proteins or cis-regulatory regions of genes. Regions that are conserved are thought to be functional while regions of non-conservation are considered non-functional or less important for function (Ponting 2001). In Chapter 2 I used these theories to predict functional and non-functional residues in the forkhead proteins based on the observed selection pressures. The idea that conserved residues are functional does have biological support. For example, the forkhead domain has been shown to be conserved and functional in many different forkhead proteins (Li et al. 2004; Saleem et al. 2004; Wierstra and Alves 2006; Cirillo and Zaret 2007). The idea that non-conserved regions are non-functional does not have experimental support because these regions are generally ignored by researchers. My analyses of mutations of the non-conserved FOXC1 positively selected site, the patient

mutation P297S and the variable disordered inhibitory domain in Chapters 3 and 4 all show that these non-conserved amino acids are important for normal protein function. Therefore it is not correct to simply say that if a sequence is not conserved among species it is not functionally important. Logically, differences must exist among species to actually create species. The concept that non-conserved does not mean non-functional is important when searching for disease causing mutations. Programs that predict if a single nucleotide polymorphism is deleterious consider changes at non-conserved sites to be non-deleterious (Ng and Henikoff 2006). There is evidence that disease causing mutations are more likely to occur at an evolutionarily conserved site, however this does not mean that they cannot occur at non-conserved sites (Miller and Kumar 2001). The FOXC1 P297S mutation discussed in Chapter 4 is of a non-conserved site, demonstrating that non-conserved sites can be as important as conserved sites for normal protein function. Similarity, the non-conserved FoxC positively selected site discussed in Chapter 3 is also important for normal FOXC1 function. The work in this thesis demonstrates that evolutionarily non-conserved residues can be functionally important.

The Inhibitory Domain is Important for Regulation of FOXC1 Target Genes.

Previous work has shown that deletion of the inhibitory domain of FOXC1 results in an increase in FOXC1 transactivation activity on the 6x BS reporter. In Chapters 3 and 4 I found that mutation of a single amino acid in the inhibitory domain would also lead to alterations in FOXC1 transactivation activity.

Interestingly, the effects of the mutations at A337 on transactivation activity varied with the reporters tested. This shows that inhibitory domain activity is influenced by the cis-regulatory region of a FOXC1 target. Therefore the inhibitory domain is not simply reducing FOXC1 activity by a constant percentage on every target gene; instead the inhibitory domain is fine tuning FOXC1 activity in response to other signals. One of these signals may be the binding site itself. Analysis of the structure of glucocorticoid receptor (GR) protein bound to DNA targets with variable sequences has shown that the protein adopts a different conformation when bound to different targets (Meijsing et al. 2009). This results in transactivation activities that are target specific. The change in GR transactivation activities was not correlated with changes in target site affinity but appears to be related to the differential use of cofactors. If FOXC1 adopts different conformations in response to binding site differences the mechanism of inhibitory domain function may be affected. A second signal influencing inhibitory domain activity may be interactions or lack thereof with other proteins due to cis-regulatory region changes that affect the availability of interacting partners. i.e. the loss or gain of binding sites for cofactors that are involved in inhibitory domain function. Finally, the inhibitory domain has the potential to influence FOXC1 target affinity. In Chapter 3 both deletion of the inhibitory domain and the A337G mutant decreased FOXC1 DNA binding. This shows that the inhibitory domain can influence target binding. These three mechanisms may all be involved in inhibitory domain regulation of FOXC1 target activation. Determining how the inhibitory domain actually reduces FOXC1

activity will aid in determining which mechanisms are more likely to be used by the domain to fine tune FOXC1 activity.

The P297S patient mutation resulted in a less severe defect of FOXC1 transactivation activity as compared to mutations within the forkhead domain. Patients with P297S were phenotypically more similar to patients with FOXC1 duplications rather than forkhead domain mutations. Taken together with the idea that the level of FOXC1 activity is influenced by cis-regulatory target sequences, this suggests that mutations in different functional domains of FOXC1 lead to various phenotypes through differential regulation of target genes. Mutations in different domains of FOXC1 may simply alter the activation level to different extents resulting in different levels of target gene activation. Mutations in different domains may also result in abnormal activity on FOXC1 targets that are specific to the mutation. A combination of domain mutation specific effects on activity level and particular targets affected could also occur. Genotype-phenotype correlations cannot be established at this time as only two individuals the mutations outside of the forkhead domain have been identified. However, examination of the relative effects of different FOXC1 mutations on different target genes would establish if mutations in different domains can exert their effects through differential target regulation.

Future Directions

While there has been much research focusing on the forkhead domain family, the mechanisms behind how these proteins actually function and why they are different from one another remain poorly understood. Forkhead family member expression patterns overlap greatly during development and in adults. The forkhead domain itself is highly conserved among orthologs and paralogs, but different enough to allow for preferential DNA targets to exist among family members. However, forkhead genes are promiscuous and do not exclusively bind to their preferred target sequences. The mechanisms behind target choice are still being determined. Helix 3 of the forkhead domain is known to bind DNA and aid in target selection, as discussed in Chapter 1. However helix 3 is usually 100% conserved among paralogs and would not allow for differential target selection by paralogs if it were the only factor used for target selection. How forkhead genes actually activate or repress transcription also remains to be elucidated. As discussed in Chapter 1, FoxA proteins appear to activate transactivation by opening compacted chromatin. While the FoxP transcriptional repressors interact with corepressor proteins to form compact chromatin. Confirmation of the use of these mechanisms by other family members or identification of novel mechanisms is still pending. The experiments discussed below will contribute to answering these questions. Determination of what forkhead genes are actually doing and how they are doing it will allow for a better understanding of speciation, development, body function and how to treat forkhead related disease.

The *in silico* analysis in Chapter 2 has shown which amino acids are conserved and which are not conserved among subfamily members. Conserved residues may perform similar functions in paralogs while non-conserved residues may be differentiating paralogs. Positive and potentially positive selected sites in the FoxA3 and Protostomia lineages of the FoxA cluster and the FoxO3 lineage of the FoxO cluster can all be biologically verified with the methods that were used in Chapter 3. Predicted domains or refinement of domain boundaries can be tested in the same manner through the utilization of deletion constructs. These types of analyses will aid in a fuller understanding of how forkhead proteins perform precise functions and verify or refute the theories proposed in Chapter 1.

Determining why a set of proteins is under positive selection is difficult in the forkhead gene family because protein functions and the mechanisms of function are still being elucidated. For the FoxA subfamily, I predicted that the positive selection in the lineage separating the Protostomia/Cephalochordata from the Craniata has acted to allow one or two copies of FoxA genes in the Protostomia/Cephalochordata to regulate the same targets of FoxA1, FoxA2 and FoxA3 in the Craniata. Alternatively, the subfamily expansion in the Craniata may have allowed for the establishment of novel targets in this lineage. To test these theories the regulation of FoxA targets by the different FoxA paralogs and orthologs can be compared using chromatin immunoprecipitation and transactivation analyses. These analyses would confirm or refute that the amino acids under positive selection are important for target selection. Additionally, if

target genes are different among species, FoxA subfamily members can be said to play a role in speciation.

The effects of mutations at the neutrally changed sites within the forkhead domain identified in the FoxA, FoxD and FoxP clusters can also be examined with the methods used in Chapter 3. These experiments would test the theory that any amino acid may be present at a neutral site and protein function will not be impaired. If the mutated protein activities are the same as the wild type protein activities, this would support the above theory of neutral change. If mutations at the neutral site are beneficial or detrimental, the above theory of neutral change is not supported. The experiments would also determine the importance of and potentially the role of the neutral sites in forkhead domain function.

Altering the positively selected site in FOXC1 demonstrated that the amino acid composition at this site was important for regulation of transactivation activity. Testing the corresponding site for functional effects in FOXC1 orthologs and paralogs using the methods in Chapter 3 would determine if positive selection is having the same effects on the different proteins. If a difference is not observed, the positive selection is not acting on that particular protein to modify its function, but is acting to differentiate those proteins in which changes in activity are observed.

The effects of mutations at the positively selected site in FOXC1 were different on different promoters. The reason for these differences is unknown. One possibility is that different mutations result in the use of different FOXC1 binding sites. In support of this, FOXC1 with the forkhead domain mutation

I126M is known to have higher affinity for mutated FOXC1 DNA consensus sequences than a perfect FOXC1 consensus sequence (Saleem et al. 2001). The actual use of different DNA binding sites can be tested through the use of deletion construction of the FOXC1 binding sites within the FOXO1a reporter in conjunction with transactivation assays. If deletion of a binding site results in reduced transactivation by wild type FOXC1 but not mutant FOXC1 (or vice versa), then different binding sites are being used by the different constructs. If binding site targets do not seem to differ, then other yet unknown inhibitory domain mechanisms are being altered by the mutation.

Identifying the FOXC1 inhibitory domain functional mechanism requires a number of theories to be pursued. As discussed in Chapter 4, the inhibitory domain may physically interact with FOXC1 to reduce activity. Nickel pull down or co-immunoprecipitation assays can be used to determine if this occurs. My attempts to identify FOXC1-inhibitory domain interactions with nickel pull downs have resulted in negative controls with positive results (i.e. FOXC1 binds to empty nickel agarose beads). The inhibitory domain may also interact with other protein(s) to reduce activity. Protein-protein interactions can be identified with yeast two hybrid assays. Additionally, since the presence of TLE4 increased the transactivation activity of FOXC1 in HTM cells, but not in Hela cells, a FOXC1 interacting partner (which is out competed by TLE4) may be present in HTM cells but not in Hela cells. A comparison of the proteome of the two cell lines could identify potential interacting partners. However, it is unknown if TLE4 is acting to reduce FOXC1 repression or activate transactivation. The

effect of deleting the inhibitory domain on FOXC1 transactivation activity has not been tested in HTM cells. Comparison of the magnitude of transactivation activity change when the inhibitory domain is deleted between different cell types will establish if there are FOXC1 cofactor differences among cell types. This is one mechanism by which FOXC1 target activation can be controlled. Establishment of how the inhibitory domain of FOXC1 exerts its functional effects may provide a mechanism that can be exploited in the treatment of individuals with disease caused by duplications of or increased dosage of FOXC1.

References

Adell T, Muller WE (2004) Isolation and characterization of five Fox (Forkhead) genes from the sponge Suberites domuncula. Gene 334:35-46

Cirillo LA, Zaret KS (2007) Specific interactions of the wing domains of FOXA1 transcription factor with DNA. J Mol Biol 366:720-724

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137:961-971

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872

Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, Gerecht-Nir S, Amit M, Hoke A, Carpenter MK, Itskovitz-Eldor J, Rao MS (2004) Differences between human and mouse embryonic stem cells. Dev Biol 269:360-380

Kaestner KH, Knöchel W, Martínez DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev 14:142-146

Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH (2009) Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 462:213-217

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980-996

Lee HH, Frasch M (2004) Survey of forkhead domain encoding genes in the Drosophila genome: Classification and embryonic expression patterns. Dev Dyn 229:357-366

Li S, Weidenfeld J, Morrisey EE (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 24:809-822

Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev 9:1654-1666

Magie CR, Pang K, Martindale MQ (2005) Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. Dev Genes Evol 215:618-630

Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM (2003) Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. Gene 316:79-89

Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, Yamamoto KR (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. Science 324:407-410

Miller MP, Kumar S (2001) Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 10:2319-2328

Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61-80

Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet 39:730-732

Ozyildirim AM, Wistow GJ, Gao J, Wang J, Dickinson DP, Frierson HF, Jr., Laurie GW (2005) The lacrimal gland transcriptome is an unusually rich source of rare and poorly characterized gene transcripts. Invest Ophthalmol Vis Sci 46:1572-1580

Pani L, Overdier DG, Porcella A, Qian X, Lai E, Costa RH (1992) Hepatocyte nuclear factor 3β contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. Mol Cell Biol 12:3723-3732

Ponting CP (2001) Issues in predicting protein function from sequence. Brief Bioinform 2:19-29

Ranganayakulu G, Elliott DA, Harvey RP, Olson EN (1998) Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. Development 125:3037-3048

Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA (2001) Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1. Am J Hum Genet 68:627-641

Saleem RA, Banerjee-Basu S, Murphy TC, Baxevanis A, Walter MA (2004) Essential structural and functional determinants within the forkhead domain of FOXC1. Nucleic Acids Res 32:4182-4193

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT (2010) Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. Science

Si H, Banga RS, Kapitsinou P, Ramaiah M, Lawrence J, Kambhampati G, Gruenwald A, Bottinger E, Glicklich D, Tellis V, Greenstein S, Thomas DB, Pullman J, Fazzari M, Susztak K (2009) Human and murine kidneys show gender- and species-specific gene expression differences in response to injury. PLoS One 4:e4802

Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L (2006) Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14:159-166

Szilagyi Z, Batta G, Enczi K, Sipiczki M (2005) Characterisation of two novel fork-head gene homologues of Schizosaccharomyces pombe: their involvement in cell cycle and sexual differentiation. Gene 348:101-109

Tu Q, Brown CT, Davidson EH, Oliveri P (2006) Sea urchin Forkhead gene family: phylogeny and embryonic expression. Dev Biol 300:49-62

Wierstra I, Alves J (2006) Despite its strong transactivation domain, transcription factor FOXM1c is kept almost inactive by two different inhibitory domains. Biol Chem 387:963-976

Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT (2008) Species-specific transcription in mice carrying human chromosome 21. Science 322:434-438