The Sea Urchin Regulome in Development

Thesis by

Meredith Ashby

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended October 5, 2006)

Acknowledgements

So many people have contributed to the success of this effort. I would foremost like to thank my advisor, Eric Davidson, for creating an intellectual atmosphere where I could indulge my passion for science. Thank you for all the discussions about both the big picture and the fine details, probing questions, encouragement, and especially all the delicious cookies.

I would also like to acknowledge Stefan Materna, Titus Brown, Andy Cameron, Rachel Gray, and Lili Chen for their work on the research presented here, and the whole Davidson Group, for making the lab such a fun place to work. Many of you I owe particular thanks to: Cathy Yuh and Titus Brown, who sacrificed hours of potentially productive time helping me out when I first joined the lab; Gabriele Amore, for being the best labmate ever, silly walk and opera singing included; the whole crazy crew who made it fun to get up at 4 am to filter gallons of urchin embryos; and Deanna Thomas and Jane Rigg, for making everything happen smoothly, from the magical appearance of supplies, to travel arrangements without bankruptcy, to publications with beautiful figures and accurate references.

Finally, I would like to thank all my friends in Pasadena, who have become family to me ─ Jen Ma, Lara, Raffi, Natasha, Maro, Sevan, and Nancy ─ I never would have made it without all of you; Stefan again, for being the best roommate ever; Mom and Dad, for undying patience and encouragement; and Jen, to whom I still owe three weeks of vacation. And most of all, Ashby, thank you for bringing new magic into my life, and putting up with all the crazy travel while I pursued my passion.

Abstract

During development an organism undergoes many rounds of pattern formation, generating ever greater complexity with each ensuing round of cell division and specification. The instructions for executing this process are encoded in the DNA, in *cis*-regulatory modules that direct the expression of developmental transcription factors and signaling molecules. Each transcription factor binding site within a *cis*-regulatory module contributes information about when, where or how much a gene is turned on, and by dissecting the modules driving a given gene, all the inputs governing expression of the gene can be accurately identified. Furthermore, by mapping the output of each gene to the inputs of other genes, it is possible to reverse engineer developmental circuits and even whole networks, revealing common bilaterian strategies for specifying progenitor fields, locking down regulatory states, and driving development forward. The *S. purpuratus* endomesodermal gene network is one of the best-characterized developmental networks, with interactions between over 40 regulatory genes mapped by perturbation experiments. With the sequencing of the sea urchin genome, it is possible to move towards the definitive completion of this network. By identifying all the transcription factors in the genome and determining their expression patterns, any previously unrecognized players can be incorporated into the network. In addition, such a comprehensive examination of transcription factor usage in maximally indirect development has not been done and will itself yield interesting conclusions.

Keywords: *cis*-regulatory module; gene regulatory network; repression; feedback loop

Table of Contents

Introduction

Understanding Development through Gene Regulatory Networks

Molecular biology has illuminated how DNA encodes amino acid sequences, and how the cell is able to translate those blueprints into proteins. Understanding in similar detail how DNA also encodes where, when and how much each protein will be expressed has yet to be achieved. Written into genomic DNA is a self-executing set of instructions which precisely directs developmental pattern formation and cell division, ultimately producing the complex body plan of the adult organism. Decrypting that information is one of the most interesting problems in biology. Only recently, with the availability of large amounts of genomic DNA sequence and the advent of high throughput *cis*-regulatory analysis, has it become possible to peer into the black box, and begin to understand at the molecular level how *cis*-regulatory information is processed to generate complexity during development, both at the individual gene level and at the gene network level.

In essence, *cis*-regulatory elements are information processing devices hardwired into the genomic DNA sequence, the function of which is to regulate gene expression (Davidson, 2006b). Most commonly, *cis*-regulatory elements or modules are several hundred base pairs long and are located within a few kilobases of the exons or within the introns of the gene they control, though there are many examples of modules which exert their influence over distances as great as 100 kb. A *cis*-regulatory module is comprised of multiple binding sites for transcription factors, plus some inter-site sequence, with each

specific binding interaction having a functional meaning. A *cis*-regulatory module typically includes many sites for ubiquitous DNA binding proteins, some of which are involved in DNA looping or required for interaction with the basal transcription apparatus. On average, a module will have binding sites for four to eight different transcription factors (Arnone and Davidson, 1997), and several sites may be present for some factors. To a rough approximation, more sites for a given factor afford the module greater sensitivity to a given regulator. Frequently two or more different transcription factors must be bound to a module in order for a gene to be activated (AND logic). Alternately, any one of several different transcription factors may be sufficient to generate an output, and the strength of the module's regulatory activity depends additively on the number of relevant interactions (OR logic). Repressor binding sites (NOT logic) are often used to delineate expression boundaries.

A gene receives information about when and where it is in the course of development by way of these transcription factor binding interactions. For example, when a signal is received from a neighboring cell at a receptor, it typically causes a cascade of protein-protein interactions, and the information conveyed by the signal ultimately arrives at the nucleus in the form of a DNA binding transcription factor. If the transcription factor is present at sufficient concentration, it will occupy target sites in an array of target *cis*-regulatory modules, and thus communicate important spatial data to the regulatory apparatus of the cell. Information about the current developmental state of the cell itself is expressed via other transcription factors, which may be turned on or off as a result of previous regulatory events. In this way, *cis*-regulatory elements read cellular conditions. They function by resolving the multiple developmental inputs they receive

into a single directive to the basal transcription apparatus, thereby specifying the appropriate outputs.

The recent wealth of genomic data has confirmed that bilaterians as simple as nematodes and as complex as humans use the same basic tool kit of transcription factors and signaling molecules to process spatial and temporal information during development (Erwin and Davidson, 2002). The qualitative complexity of the developmental regulatory tool kit is thus not correlated with genome or proteome size. Rather than relying upon a vastly larger tool kit, complexity is increased with remarkable economy by reusing transcription factors in additional unique ways in the course of later rounds of pattern formation. Every regulatory gene has not just one but many *cis*-regulatory modules which control the expression of the transcription factor it encodes in different spatial domains at different times in development. One module may activate a gene in one embryonic domain, while other modules assure that the same gene is repressed simultaneously in neighboring domains. Yet another module may direct the gene's later involvement in patterning specialized structures or organs, while a late-acting module is involved in cell differentiation. Hardwired into these individual modules is the correct response of the gene to every diverse circumstance the cells of the organism will encounter, throughout development and the lifetime of the organism. While to date only a few *cis*-regulatory modules have been mapped in fine detail, it is clear that the same strategies are used across the bilaterians to encode when and where in development genes are expressed (Davidson, 2006b; Levine and Davidson, 2005).

even-skipped

One of the first *cis*-regulatory modules to be characterized at the target site level is the *Drosophila* gene *even-skipped*, or *eve*. The early *Drosophila* embryo is syncytial: the nuclei exist within a common cytoplasm. Maternal mRNA localized at the anterior of the embryo generates a diffusion gradient of Bicoid (Bcd) protein, a maternally encoded transcription factor. Likewise, the transcription factor Caudal diffuses from the posterior of the embryo, uninhibited by cellular membranes. These opposing gradients are used to generate gradients of additional regulatory proteins, including Hunchback (Hb), Krüppel (Kr), and Giant (Gt). By the early blastoderm stage, the syncytial nuclei have migrated to the periphery of the embryo in preparation for the specification of territories corresponding to future segments. The formation of individual cell membranes occurs late in cleavage, but *eve* is activated in seven thin circumferential stripes only a few nuclei wide prior to this (fig. I.1A), in response to the earlier established transcription factor gradients.

The five *cis*-regulatory modules responsible for the expression of these stripes have been identified (Andrioli et al., 2002; Frasch and Levine, 1987; Fujioka et al., 1999; Harding et al., 1989; Macdonald et al., 1986; Small et al., 1996). Three modules drive the expression of one stripe each, while two other modules control two stripes each (fig. I.1B). The *eve* stripe 2 module is understood in the most detail, and is an excellent example of how both positive and negative inputs can be combined to delineate very precise spatial expression patterns. Two activators, Bicoid and Hunchback, are required for stripe 2 expression, and four functional binding sites for Bcd plus one for Hb are located in the minimal *eve* 2 module (Stanojevic et al., 1991). However, as both

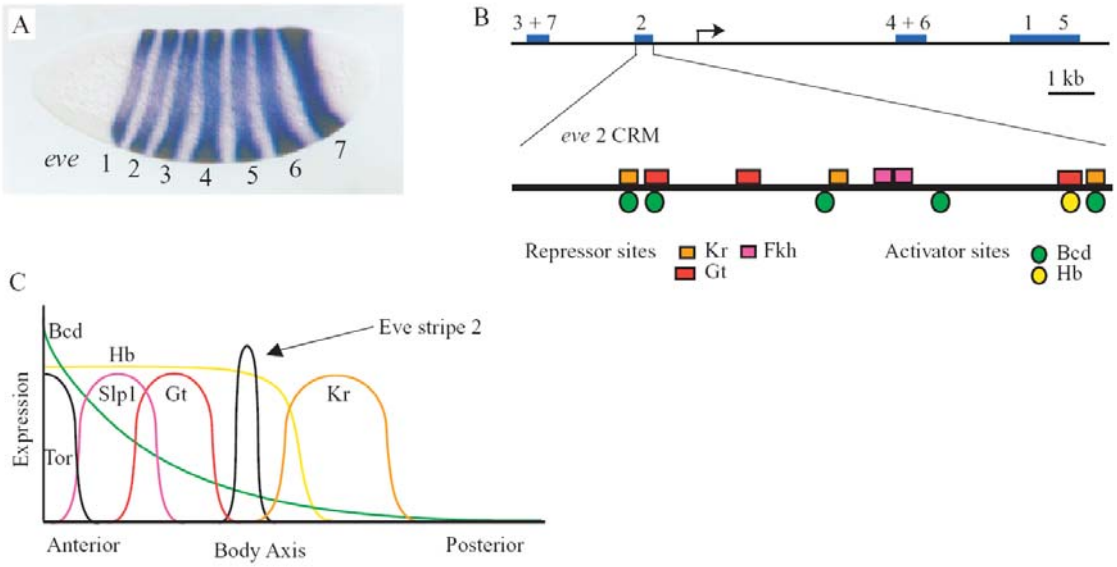


Figure I.1. The cis-regulatory module of *eve* stripe 2. A. RNA in situ hybridization with a digoxigenin-labelled antisense probe reveals the seven stripes of expression in the *Drosophila* embryo (Small et al., 1996). B. The five *cis*-regulatory modules that direct expression are located both upstream and downstream of the transcription start site. Below is an expanded view of the *eve* stripe 2 module. The squares above the line show the location of repressor binding site; circles below the line mark activator sites. Adapted from Andrioli et al., 2002. C. A qualitative graph summarizing the expression domains of transcription factors that are inputs to the *eve* stripe 2 regulatory module. Adapted from Stanojevic et al., 1991.

transcription factors are present in a broad section of the anterior embryo, they alone are insufficient to produce the narrow band seen in stripe 2. The gap gene repressors Giant, acting with an unknown corepressor, and Krüppel, are required to constrain the anterior and posterior boundaries of the stripe, respectively (Arnosti et al., 1996; Gray and Levine, 1996; Small et al., 1992). Accordingly, three binding sites for each repressor can also be found in the module. Mutation of these repressor sites results in the ectopic expansion of stripe 2. Anterior of the Giant domain, yet another repressor acts on the *eve* 2 module. Repression is actuated through two adjacent TTTGTTT motifs, most likely by the forkhead factor Sp1 and an unknown corepressor (Andrioli et al., 2002). Finally, repression of *eve* in the anterior tip of the embryo is controlled by a tyrosine receptor kinase phosphorylation cascade, which may act by interfering with Bcd-dependent activation (Andrioli et al., 2002). The precisely bounded *eve* 2 stripe arises from a

combination of AND and NOT logic hardwired into the DNA upstream of the gene. The *eve* 2 module draws on an array of spatial inputs established in the previous round of specification, and it integrates this information into a single new transcriptional output, activating the gene in the spatial domain that it uniquely specifies.

dorsal and pha-4

As part of a network, regulatory genes not only take in multiple inputs, but also make numerous downstream connections. It is via these downstream outputs that information from the previous round of specification is both locked in and combined with neighboring specification events. One mechanism by which a single gene can directly activate a whole range of regulatory genes within an embryonic territory while adding new information is by making use of a gradient. The importance of gradients in development also illustrates how *cis*-regulatory elements can be tuned to respond to precise cellular conditions.

Dorsal directs a well-known gradient network in *Drosophila* development. Maternal Dorsal is most concentrated in the presumptive mesoderm in the ventral region of the embryo, with the concentration falling off steeply in the lateral regions of the embryo at the future neurogenic ectoderm boundary (Stathopoulos and Levine, 2002). Genes downstream of *dorsal* have target sites with an array of sensitivities such that individual genes are activated only in specific spatial territories of the embryo (Levine and Davidson, 2005). For example, *dorsal* activates *twist* in only the most ventral region of the embryo via a pair of low-affinity sites; in other regions the concentration of Dorsal is too low to saturate the sites and turn on the gene (Jiang and Levine, 1993). In all, the

Dorsal gradient directly specifies between four and seven different thresholds of activation.

C. elegans pharynx development is another example of how a regulatory gene gradient can be used to orchestrate a gene network. In this case, the organ identity gene *pha-4* presides over temporal, rather than spatial, specification patterns. It has been shown that the regulatory regions of most genes expressed during pharyngeal development carry copies of the Pha-4 consensus binding sequence TRTTKRY (Gaudet and Mango, 2002). Furthermore, higher and lower affinity Pha-4 sites are correlated with earlier and later pharyngeal activation, respectively, and the onset of expression can be advanced or delayed by altering the affinity of these sites. A picture emerges in which *pha-4* is at the nexus of pharynx development. Cued by rising Pha-4 levels, sets of genes with differing Pha-4 sensitivity are sequentially activated, perhaps helping to coordinate organogenesis among the five different cell types present in the mature pharynx. Because *pha-4* is so central to pharynx development, disabling it results in ablation of the whole organ. Both *dorsal* and *pha-4* exemplify how a concentration gradient can enable one gene to send different signals to an array of targets over either developmental time or space. They also demonstrate that *cis*-regulatory modules can be sensitive to not just the presence or absence of key regulators, but can be set to respond to only very precise cellular conditions.

endo16

The upstream regulatory region of *endo16* in the sea urchin, *Strongylocentrotus purpuratus,* has been mapped in detail and is an excellent example of how *cis*-regulatory

modules function as hardwired information processing devices during development. The *endo16* gene is first expressed in the vegetal plate of blastula-stage embryos in a ring of cells that will give rise to endodermal and mesodermal cell types. Expression is specifically repressed in the skeletogenic progenitors at the center of the vegetal plate, and in the surrounding ectoderm. During gastrulation, *endo16* is expressed throughout the archenteron. Subsequently, expression is turned off in the foregut, then in the hindgut, while intensifying in the midgut. A 2300 bp region upstream of the coding region

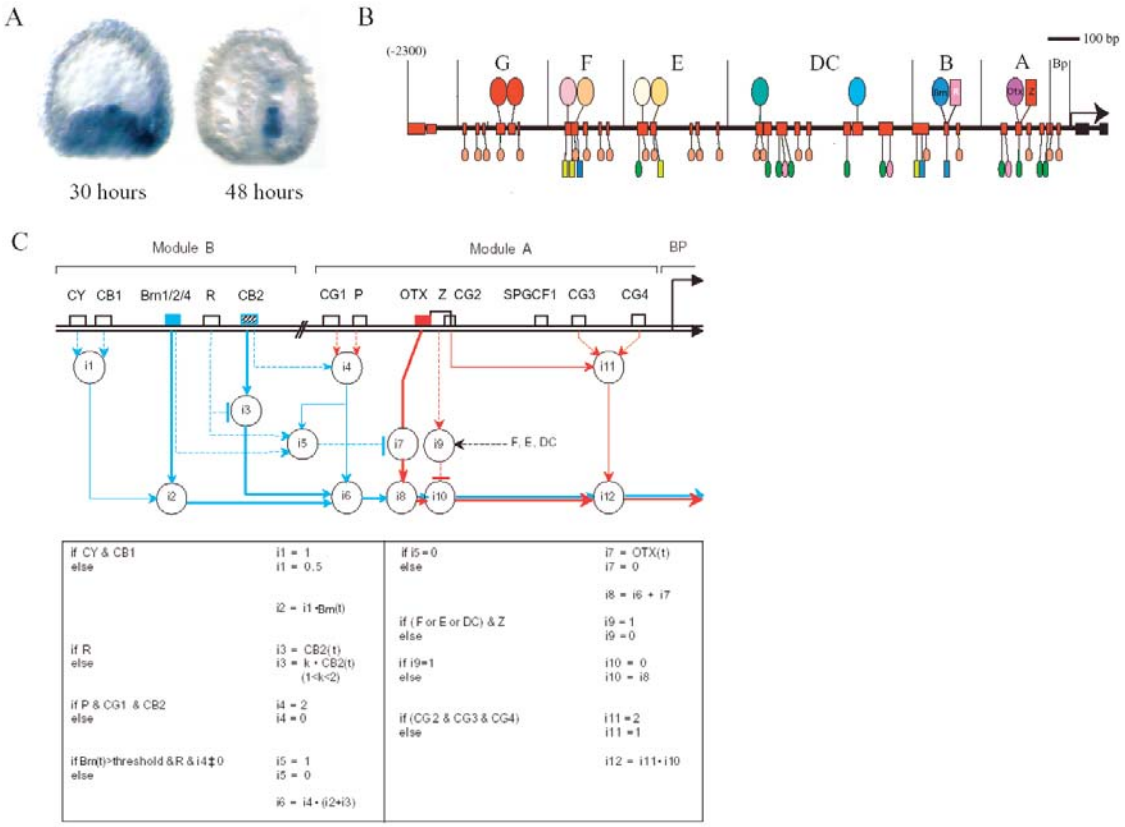


Figure I.2. The *cis*-regulatory logic of the endo16 promoter. Whole mount *in situ* hybridization shows *endo16* expression at 30 h and 48 h. B. A detailed map of transcription factor binding sites within the 2300 bp region which correctly recapitulates *endo16* gene expression. C. The behavior of modules A and B can be modeled as a logic map. Boolean functions are shown with dashed lines, and scalar inputs are shown as solid lines. Each individual step in processing the inputs to this system is enumerated in the boxed logic statements.

recapitulates this expression pattern when fused to a CAT reporter gene (Yuh et al., 1994). The protein binding sites within the region were mapped in detail and it was determined that nine different proteins bind at unique sites, and five additional proteins bind at multiple sites within the regulatory domain.

The 2300 bp regulatory region can be understood as a collection of discrete *cis*-regulatory modules A–G (fig. I.2B), each having a distinct role in defining correct spatial and temporal expression of *endo16*. Likewise, within each module, every target site has a specific function. As seen in figure I.2B, the overall domain naturally divides into clusters of binding sites, with one or two uniquely occurring sites within each cluster. Target sites for the architectural protein SpGCF1 are scattered throughout the regulatory region and probably act to facilitate communication between non-adjacent modules; SpGCF1 enables DNA looping by forming multimers once bound to DNA (Zeller et al., 1995).

The most proximal subregion, module A, has a dual role as both gatekeeper and activator. At all stages, it relays the output of the other modules to the basal promoter (Yuh and Davidson, 1996). Also, boosted synergistically by module G, it drives the initial appearance of *endo16* in the vegetal plate. The primary activating input to module A, SpOtx, is widely expressed in the early embryo (Yuh et al., 2001). Hence module DC is required to repress ectopic *endo16* expression in the skeletogenic precursors, while modules E and F repress expression in the ectoderm. Mutagenesis studies indicate that these repression signals are transmitted via site Z, directly adjacent to the SpOtx target in module A (Yuh et al., 1998).

Module B is the second main activator of *endo16* and controls the late surge in expression during gastrulation and differentiation of the midgut. The gut-specific transcription factor Brn1/2/4 is the primary driver of the module, though interactions at the CY, CB1, and CB2 target sites provide additional boosts in expression (Yuh et al., 2001; Yuh et al., 2004). Module G continues to exert its synergistic influence. Once again, the output of this module passes through module A, in this instance from the CB2 site in module B to sites P and CG1 in module A. If any of these sites is abolished, the strong late rise in expression driven by module B is entirely absent.

Perhaps the most interesting aspect of the *endo16* regulatory system is the manner in which control of expression is handed off from module A to module B in the late blastula stage. The switching function is encoded at target site R in module B, such that when the output from Brn1/2/4 exceeds some threshold, a protein bound at R blocks further input from the SpOtx site in module A (Yuh et al., 2001; Yuh et al., 2004). The role of module A in this condition is then only to amplify the output of module B linearly, by a factor close to four, and it is this enhanced regulatory impetus that is passed on to the basal transcription apparatus. Throwing the "R"-mediated Otx vs. Brn1/2/4 switch relieves the ongoing dependence of the system on the repressors binding in modules E, F, and DC. These are needed for correct specification when the gene is driven by the ubiquitously present SpOtx factor. Once under control of the gut-specific Brn1/2/4 in module B, *endo16* is enrolled in the process of gut differentiation.

In summary, the most important general aspects of the *endo16 cis*-regulatory system revealed by the experimental and computational analyses of Yuh et al. are threefold. First, the functional significance of each and every target site in the most

important regions of the system was tested, and each was demonstrated to play a specific regulatory role. Second, these roles are distinct, and are qualitatively unique with respect to one another. Third, as a whole, the system functions in a conditional manner, depending on the inputs, and its operation can be summarized and predicted accurately by a set of conditional logic statements. The *endo16 cis*-regulatory system is thus an example, indeed the best known such example, of a developmental logic processor that equips the gene it controls to respond appropriately to all regulatory conditions it will encounter in any cell of the embryo, over all developmental time.

Regulatory circuits and networks

Of course, no one gene can convey sufficient spatial information to generate complex morphologies. Rather, morphology is generated by successive rounds of pattern formation directed by networks of regulatory genes encoding transcription factors and signaling proteins. A single tissue or structure in a bilaterian organism is the result of the expression of hundreds or thousands of genes. Each node of a developmental network is a *cis*-regulatory element, which translates upstream regulatory gene outputs into the more refined expression of the next transcription factor in the cascade. Networks allow complexity to be built up, with each round of specification adding information about the structure of the developing body part.

A useful way to visually depict information flow in developmental gene regulatory networks is by use of "logic maps," treating various *cis*-regulatory interactions as Boolean AND/OR logic gates or switches (Bolouri and Davidson, 2002a, b; Istrail and

Davidson, 2005). The object of a gene regulatory network model that portrays the logic map for a given aspect of development is to connect the inputs and outputs of network. Network logic can only be appreciated in this larger context. Depicting gene networks this way highlights some common strategies that organisms use to achieve the remarkable level of precision and control seen in developmental gene expression.

Positive feedback loops are one such mechanism, and are commonly used to enforce the strictly forward progress of development (Bolouri and Davidson, 2002b; Davidson et al., 2002b). In the generalized scheme depicted in figure I.3A, a signal from an initial activator turns on gene 1. Gene 1 in turn activates gene 2, which passes the activation signal to a group of downstream genes. However, gene 2 also generates a feedback loop with gene 1, such that when the initial activator subsides, genes 1 and 2

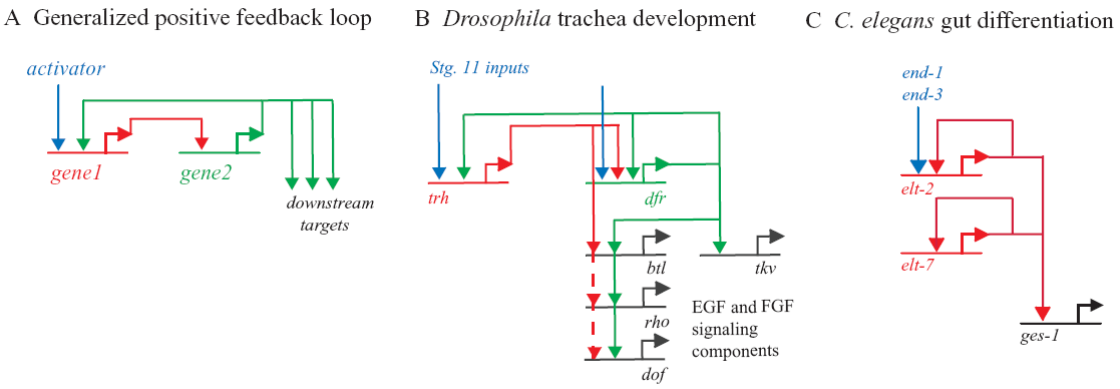


Figure I.3 Circuit diagrams of positive feedback loops. A. Once turned on by an activator, gene 1 (shown in red) activates gene 2 (shown in green). In addition to acting on downstream targets, gene 2 activates gene 1, forming a positive feedback loop. When the initial activator signal fades, these genes will remain active. B. In an example from *Drosophila* trachea development, *trh* and *dfr* form a feedback loop. *trh* (red) activates *dfr* (green), which completes the loop by in turn activating *trh*. The In this example, *dfr* also feeds back on itself. C. In this variation of the circuit, the intermediary between the first activated gene and downstream target genes is dispensed with altogether: *elt-2* (red) is directly activated by a transient signal from *end-1* and *end-3* and forms an autofeedback loop.

remain locked on, maintaining the new regulatory state of the cell. By constructing the circuit this way, the initiating signal needs only be transient. Once the feedback loop is in place, the downstream genes in developmental subsystems that are constructed in this manner have no further dependence on the initiating transcription factors. A benefit of this strategy is that early regulatory proteins are then free to be enrolled in later specification and differentiation events without conflict.

Many variations on this type of positive feedback loop can be found in the literature (Davidson et al., 2003). In one such example from *Drosophila* development, a positive feedback loop is used to lock down tracheal specification within the initial field of progenitor cells (fig. I.3B). Expression of the genes *trachealess* (*trh*) and *drifter* (*dfr*) define the ten tracheal placodes in the postgastrula embryo, the cells of which will invaginate and migrate to form the trachea (Metzger and Krasnow, 1999). Both genes are activated by an array of A/P and D/V spatial inputs present in the stage 11 embryo (Zelzer and Shilo, 2000). By the beginning of gastrulation, however, these cues have been supplanted by a mutual and feedback circuit, as depicted in figure I.3B. Given the incipient complex migration of these cells, establishing a self-sustaining regulatory state is critical. This feedback loop locks down tracheal specification and in turn provides the regulatory input to critical signaling pathways needed for migration and morphogenesis (Zelzer and Shilo, 2000).

In *C. elegans* endoderm development, an intergenic feedback loop is used to initiate and maintain gut differentiation (fig. I.3C). The activators of this circuit are the GATA transcription factors *end-1* and *end-3*, which drive *elt-2* and *elt-7* (Fukushige et al., 1998; Zhu et al., 1997). These in turn activate the gut specific esterase gene *ges-1*

(Maduro and Rothman, 2002; Marshall and McGhee, 2001). This regulatory mechanism varies slightly from the canonical loop in that *elt-2* completes the circuit itself with an auto-feedback loop. The defining feature of the circuit is the same, however: long after the initial signals from *end-1* and *end-3* fade, the *elt-2* gene maintains gut specification. In fact, throughout the lifetime of the organism, disruption of *elt-2* results in the loss of gut specification (Fukushige et al., 1998). As shown in figure I.3C, the feedback loop is the key to maintaining this persistence.

Another kind of architectural motif often found in gene regulatory networks involves the use of repressors to create boundaries between cells with differing fates. Gradients of positive inputs by themselves rarely suffice to define the sharp boundaries of expression seen for genes that specify cell fates or given progenitor fields; these crisp demarcations are generally imposed by repressors. An excellent example is found in rhombomere (r) specification during mouse hindbrain development. Correct *hox* gene expression is required to establish the identity of each rhombomere, namely *hoxa2* and *hoxb2* in r3, and *hoxb1* and *hoxb2* in r4 (Barrow et al., 2000). As shown in figure I.4, the activation of *hoxa1* and *hoxb1* in the future r4-r6 region simultaneously represses *krox20* there, while causing the expression of a signal that results in the activation of *krox20* in the prospective r3 region (Barrow et al., 2000). *krox20* activates a feedback loop that drives the two hox genes which specify r3 identity, *hoxa2* and *hoxb2*. Thus repression by *hoxb1* sharply divides the boundary between r3 and r4, locking off r3 fate in r4 cells. These network devices, and several others that can be identified, are seen over and over in bilaterian development. They provide explanations at the genomic level for observed developmental specification events. Understanding development as a network of

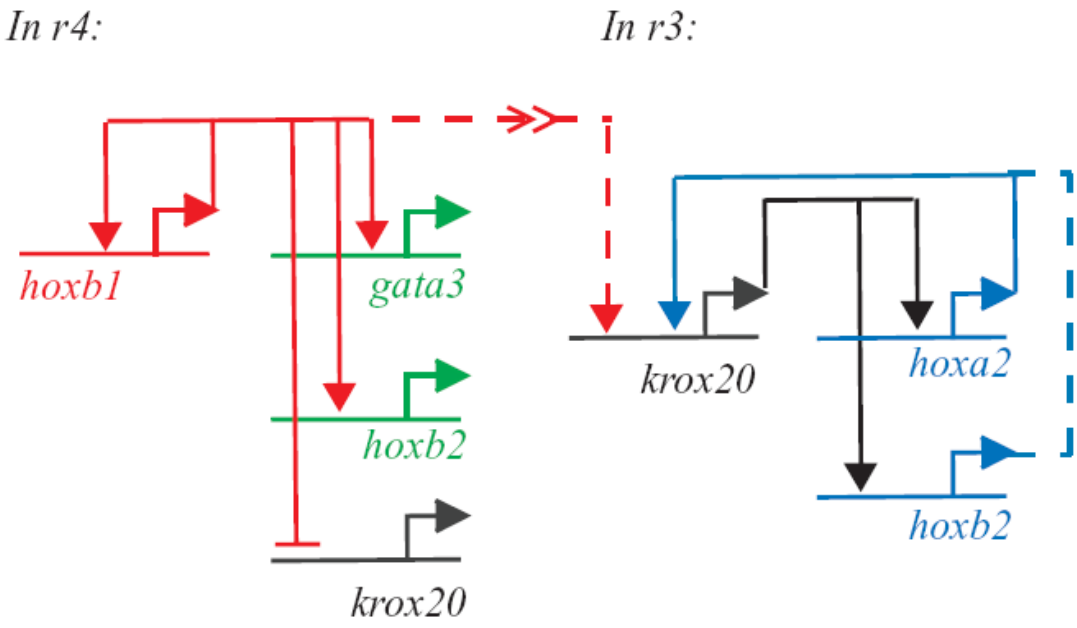


Figure I.4 Repression and boundary formation. Gene regulation in mouse rhombomeres is an example of how repression can give rise to sharp boundaries between regions having different cell fates. *krox20* is instrumental in specifying r3 fate as a conditional input to *hoxa2* and *hoxb2*, but in r4 it is repressed by the same gene (*hoxb1*) that activates r4 specific hox genes.

regulatory genes interacting via *cis*-regulatory modules illuminates how static information written in the DNA translates into the dynamic process of embryogenesis. Furthermore, just as *cis*-regulatory modules draw on a tool kit of AND, OR, NOT, and many other kinds of logic inputs, large developmental networks can be understood as assemblies of smaller sub-circuits, or recurrent architectural motifs.

The *S. purpuratus* endomesoderm gene network

The most extensive gene regulatory network model constructed to date pertains to the development of the endomesoderm of the *S. purpuratus* embryo. The network model

encompasses regulatory events up to 20 - 24 h post-fertilization and just before gastrulation, and includes linkages among about 50 genes, of which over 40 encode transcription factors or signaling molecules (Davidson et al., 2002a; Davidson et al., 2002b). The logic map connecting these many genes specifies how the instructions distributed among the *cis*-regulatory modules of these genes work together to generate an information cascade directing sea urchin development. Each *cis*-regulatory module functions as a node in the network, with each module performing regulatory calculations using inputs from other genes in the network. Several such nodes linked together function as sub-circuits that establish discrete territories in the developing embryo, lock down regulatory states or launch differentiation subroutines. Zooming out one step further, the sum of these sub-circuits amounts to a specific proposition of the genomic code indicating when, where, and why each gene is expressed so as to execute the specification of three prominent domains of the embryo, namely the skeletogenic domain, the remaining mesodermal domain, and the endodermal domain.

At 24 h, the *S. purpuratus* embryo is to the microscopist a still largely unremarkable hollow ball of cells, except that the future skeletogenic cell population has by now ingressed into the blastocoel. However, in terms of the spatial expression of defined regulatory states, by this stage almost all cells in the embryo are already specified, though of course their states of specification will further alter as development proceeds. A recent version of the network model for endomesoderm specification is shown in figure I.5 (the model is continuously updated on our website, http://sugp.caltech.edu/endomes/; see legend for symbolism and details). The model essentially details zygotic *cis*-regulatory interactions at the DNA level. Functions

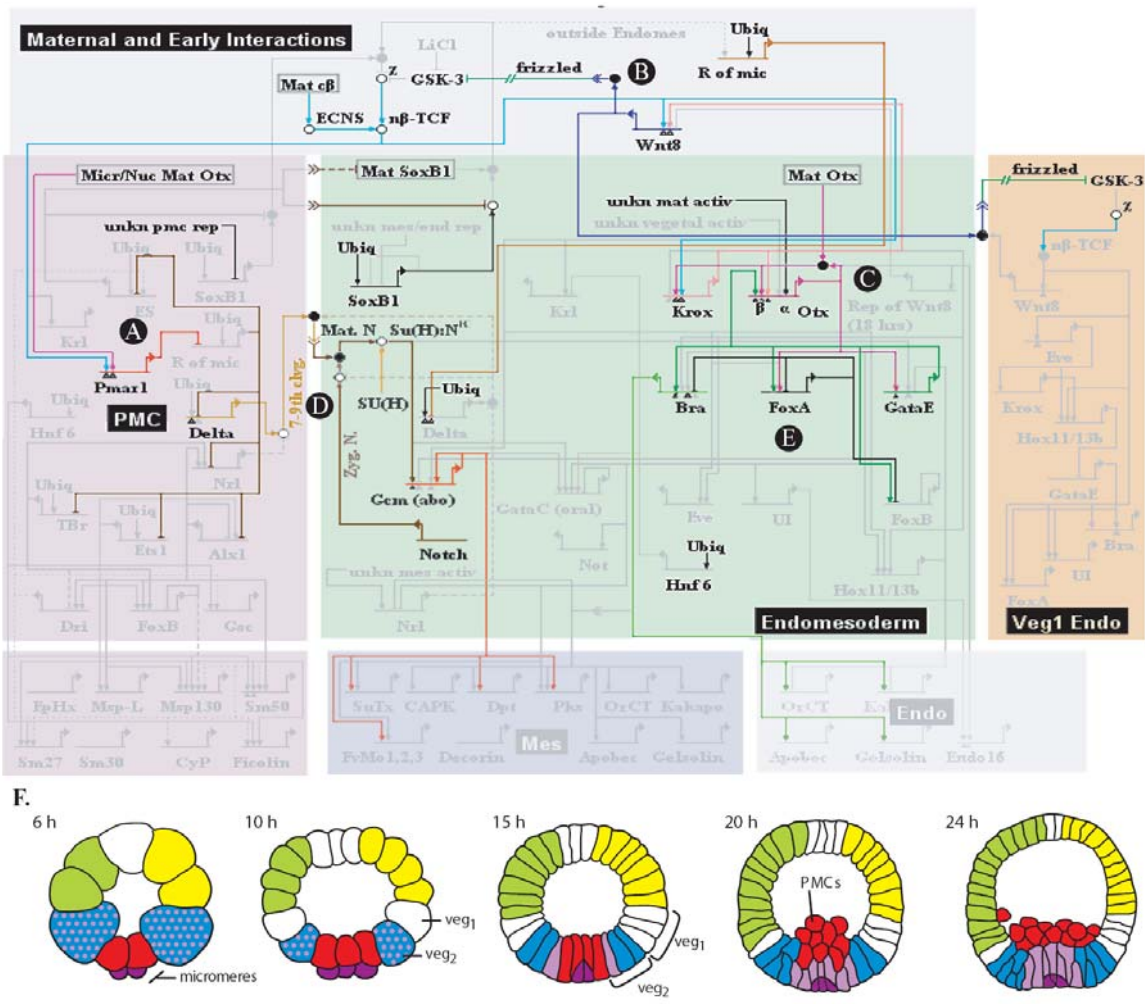


Figure I.5 The *S. purpuratus* endomesodermal gene regulatory network before gastrulation. The diagram was assembled using qPCR data from a variety of perturbation experiments including injection of sea urchin eggs with morpholino anti-sense oligonucleotides, reporter constructs with wild-type and mutated *cis*-regulatory DNA, and engrailed fusion constructs. Each gene is represented by a short horizontal line with a bent arrow. For each gene, the diagram illustrates both upstream regulatory inputs and downstream targets. Activators are connected to their targets by arrows, whereas blunted lines indicate repression. Double arrows signify cell signaling interactions and dashed lines show inferred, indirect interactions. Genes labeled "Repressor" are inferred. "Ubiq" indicates a ubiquitously active positive input. White circles indicate biochemical or protein-protein interactions occurring in the cytoplasm. Circuits described in the text are labeled A-E and depicted in color, while other parts of the network are grayed out. The coloration is not meant to imply that the highlighted circuits are running concurrently. A. Micromere specification occurs via a dual repression circuit. B. Nuclearization of β-catenin drives the initial specification of the

veg_2 lineage and specification is maintained by a Wnt8 feedback loop. C. The veg_2 regulatory state is locked down by progressive regulatory loops involving *krox1*, *otx*, and *gataE*. D. The inner veg_2 cells are specified as mesoderm precursors by a Delta signal originating in the micromeres. E. In the outer veg_2 cells, GataE activates other endoderm genes which will prime the embryo for gastrulation. F. A map of the progressive specification of the sea urchin embryo.

occurring off the DNA are only indicated where necessary to make inputs into model *cis*-regulatory elements intelligible, i.e., to denote maternal inputs or biochemical linkages between signal receptors and the downstream transcription factors they animate. The purpose of the model is the same as the purpose of the individual *cis*-regulatory analyses discussed above: to make explicit the functional significance of each participating element of the genomic DNA regulatory code. It is neither a kinetic transcription model nor a biochemical transcription model. It is not about how these transcriptional systems work, but rather about the structure/function relationships within the DNA that encodes the guiding program for this aspect of development.

S. purpuratus development up to 24 h can be summarized briefly (Davidson et al., 1998). The zygotic regulatory processes can first be tracked at fourth and fifth cleavage, when the small and large micromeres are formed at the vegetal pole of the embryo. The invariant fate of the large micromeres is to serve as the skeletogenic precursor lineage of the embryo. After sixth cleavage, the veg_2 and veg_1 lineages arise as concentric rings of cells surrounding the micromeres. By 15 h the veg_2 cells have begun the process of specification into future endoderm or mesoderm. By 24 h, the skeletogenic precursors have ingressed into the blastocoel, endoderm and mesoderm specification is complete, and the embryo is primed for gastrulation. This apparent morphological simplicity up to 20 h post-fertilization is quite deceptive. In truth, the regulatory gene network depicted in

figure I.5 is launched at the moment of fertilization, guiding the embryo through an ever more complex succession of regulatory states. In figure I.5, early maternal inputs are depicted in the gray box at top, while programs running in the large or skeletogenic micromeres are illustrated in the pink box at left. The central green area includes genes running in endoderm or mesoderm from cleavage through 24 h.

The first regulatory event of note depicted in the model is the specification of the micromeres via a dual repression circuit, an unexpected and almost counterintuitive mechanism for imposing a state of specification on an early embryonic cell lineage. In all other territories of the embryo, primary skeletogenic regulatory and signaling genes, including *delta*, *alx1*, *ets1*, and *tbrain* (*tbr*), are actively repressed by the product of a gene which has yet to be identified. In the micromeres, however, nuclearization of maternal β-catenin and Otx activates *pmar1*, deactivates the gene encoding the repressor, and thus launches the skeletogenic subroutine in only these cells (Oliveri et al., 2002; Oliveri et al., 2003). This linkage is most likely direct, as the *cis*-regulatory module driving *pmar1* expression in the micromeres contains putative TCF and Otx target sites. A few hours later, these initial skeletogenic regulatory genes activate several additional regulatory genes, viz., *deadringer* (*dri*), *foxb*, and *goosecoid*. These genes, together with the initial regulators *tbr*, *alx1* and the ubiquitously expressed activator *hnf6*, constitute the known immediate governors of the terminal skeletogenic genes (see fig. I.5).

The network model also provides an explanation of how the adjacent veg_2 lineage is initially specified (Davidson et al., 2002b). In this domain, an early signal from the micromeres and nuclearization of maternal factors set up the initial endomesodermal regulatory state. Feedback loops are utilized to ensure the forward progress of the

developmental process. For example, maternal β-catenin activates the gene encoding the signaling molecule Wnt8, which in turn results in further β-catenin nuclearization. This circuit creates a self-sustaining "community effect" among veg_2 cells; mediated by the β-catenin/TCF system, these cells are maintained in a common regulatory state (Gurdon et al., 1993).

Shortly after the β-catenin/TCF system is thus locked on in the veg_2 endomesoderm (about 8th cleavage), this input, together with a maternal/early zygotic form of Otx, activates the endomesodermal regulatory gene *krox/blimp11.* A few hours later, *krox/blimp1* in turn drives embryonic *otx* expression via a newly activated zygotic *cis*-regulatory element. Remarkably, there follows the institution of an additional regulatory loop, as the *otx* gene product is now required to activate the *gataE* gene, which then reciprocates by activating the zygotic *otx* gene control element (see fig. I.5). Soon thereafter the *krox1* gene ceases to be expressed in the veg_2 endomesodermal domain. Indeed, its expression is no longer necessary there, as *otx* and *gataE* are now locked in a positive regulatory embrace and no longer require the inputs needed for their initial activation. The net effect of these positive feedback loops is to transfer control of the induced regulatory state to the embryo, and relieve the system of its dependence on maternal and ephemerally expressed early zygotic inputs. Once these feedback loops are in place, veg_2 endomesodermal specification is locked in.

In another coincident specification event, a combination of signaling and repression is used to subdivide the veg_2 lineage into mesoderm and endoderm precursors. The regulatory subroutine running in the micromeres includes among its targets a gene encoding the signal ligand Delta. This gene is expressed and the signal is emitted

between the seventh and ninth cleavages, when it is received in the innermost cells of the veg_2 domain (Amore et al., 2003; Davidson et al., 2002b). The Delta signal provides the spatial cue that specifies the mesoderm, by causing the adjacent cells receiving it to activate the Notch pathway. One immediate effect is the activation of *gcm* in a single ring of cells abutting the micromeres. Once activated, this gene also utilizes an auto-feedback loop to lock itself on, one of the most common regulatory motifs. Its function is to drive a battery of differentiation genes specific to mesoderm pigment cells, in which it continues to be expressed throughout embryogenesis (Ransick et al., 2002a).

In the more outer veg_2 domain, GataE activates many other endomesodermal regulatory genes, its expression having been stabilized, as noted above, by a feedback relationship with the *otx* gene (Davidson et al., 2002b). GataE targets perform several important roles in the ongoing specification of the endoderm. It activates the repressor *foxA*, which will establish the correct boundary for *brachyury* (*bra)* and *foxB* expression during gastrulation. Later, *bra* will directly control a battery of endoderm motility genes required for gastrulation (Rast et al., 2002). In conjunction with a late *wnt8* signal, *gataE* also plays a role in specifying cells of the inner veg_1 domain as endoderm (Ransick and Davidson, 1998). The function of the feedback circuitry upstream of *gataE* is thus ultimately to ensure the stable expression of this centrally important regulator of the endomesoderm.

The logic map for endomesoderm specification in *S. purpuratus* shows explicitly how common regulatory subcircuits have been assembled to produce a unique and complex developmental program. The regulatory network operates progressively (for a display of its temporal behavior, see the website). Its initial inputs are maternal and

cytoplasmic, and it uses these to set in motion the initial tier of zygotic gene expression. These genes are in turn utilized to generate more and more spatially precise cues. The culmination of the specification process is the activation of specific differentiation batteries throughout the embryo. The individual circuits each make a contribution to the system, but the overall logic of the network can truly only be appreciated as a whole.

This perspective has proven useful in understanding specification events in diverse models of development, and several other systems have been described in detail as multigene networks of interacting transcription factors. The *Drosophila* Dorsal gradient network maps interactions between nearly 60 genes, and the system uses a distinctive set of logic circuits that may be specific to syncytial embryos (Levine and Davidson, 2005; Stathopoulos and Levine, 2002). In *C. elegans*, in addition to the Pha-4 network described above, a network of genes directing specification of the C-blastomere lineage has been elaborated, beginning with the homeobox transcription factor Pal-1 (Baugh et al., 2005). Most recently, a provisional gene network describing specification of the *Ciona intestinalis* embryo has been laid out, describing connections between 76 zygotically expressed regulatory genes. As more networks are mapped in detail, interspecies comparisons will shed light on the mechanics of evolution. Knowledge of the *cis*-regulatory modules of different genes in different species, and of the network connections between these modules, will offer insight into how the evolution of regulatory DNA sequence gave rise to the myriad body plans and structures of animals.

A genomic approach to completing the network

While the sea urchin endomesodermal gene networks is one of the best characterized developmental gene regulatory networks, the model is not complete. Indeed, the identities of several key regulators are still not known. In addition, there may be other nodes in the network that are completely missing. If the goal is to fully understand the logic of this network, we must be certain that there are no gaps in our model.

In the past, a difficulty with uncovering additional genes relevant to this network has been that transcription factors are sometimes expressed in only a few cells or at very low levels, meaning even important factors can be rare in EST and macroarray libraries (Davidson, 1986). Overcoming this problem generally involves laborious and time-consuming methods. An example is the method used to identify upstream activators of *endo16*. Embryonic nuclear extracts from the appropriate time points were passed through a DNA column consisting of target sites from the endo16 *cis*-regulatory region. The captured proteins were then digested and partially sequenced, and the corresponding macroarrayed clones were identified using degenerate, complementary probes. However, in addition to being very time consuming, this method is limited to identifying *upstream* regulators of known network genes. To identify downstream targets, a subtractive cDNA assay was developed to deplete housekeeping gene messages and concentrate specifically up-regulated transcripts from perturbed vs. control embryos. A macroarray library of the resulting cDNA pool was then successfully used to identify a number of transiently

expressed network genes (Rast et al., 2002). However, this strategy can never definitively demonstrate that all relevant low-copy transcripts have been found.

In this work, we have made use of the recent sequencing of the *Strongylocentrotus purpuratus* genome to move towards the definitive completion of the endomesodermal gene network by identifying all the transcription factors in the genome. The beauty of this new approach lies in the fact that the DNA binding domains present in transcription factors are generally very conserved between species. Hence, an exhaustive search for these sequence motifs in the genome can be used to generate a reliable, nearly complete list of regulatory genes. Once found, all the uncovered factors can then be assessed for embryonic expression, revealing any still unrecognized players in endomesoderm specification as well as creating a database that will be useful in describing patterning in other parts of the embryo.

A compilation of data on transcription factor usage in sea urchin development will also be interesting in itself. Microarray experiments have become common tools for studying gene usage patterns in organisms with sequenced genomes, and a number of these studies provide interesting comparisons between regulatory gene usage and that of other classes of genes during development. In one such study of *Drosophila melanogaster* gene expression, 4028 assayed genes were sorted by functional class (about one-third of predicted genes), and it was noted which classes were used lightly or heavily during the major life stages of the organism (Arbeitman et al., 2002). Interestingly, transcription factors, signaling molecules, and cell cycle genes were all found to have their overall peak expression usage during embryogenesis, with overall expression levels at their lowest during the larvae, pupae, and adult stages. A similar experiment in mouse

used a microarray incorporating 25,000 unique genes from embryonic and adult tissues to track transcription from embryonic day 8 to birth. Grouping the genes by their gene ontology classification revealed that transcription factors and cell cycle genes were similarly expressed at their highest level during early embryogenesis (Wagner et al., 2005). Microarrays have also been used to examine overall gene expression in *C. elegans* development. One very thorough study looked at gene expression over a range of time points encompassing most specification events in *C. elegans* development (4-cell through 190-cell stages), and found similar biases in transcription factor usage (Baugh et al., 2003). In this study, genes were grouped by functional class and it was asked whether specific classes were overrepresented at various embryonic time points. Again, while the focus was not specifically on regulatory gene usage, as a group these genes are consistently overrepresented among transiently expressed genes at a number of embryonic time points, and under-represented among genes expressed only maternally. These results emphasize the central role the tool kit of signaling and regulatory genes plays in patterning the embryo.

To date, the most comprehensive study of transcription factors in development has been done in *Ciona intestinalis*. Transcription factors and signaling molecules were systematically identified in the *Ciona* genome, and the expression of 352 regulatory and signaling genes was determined by *in situ* hybridization up to the mid-late tailbud stage (Imai et al., 2004; Miwata et al., 2006; Satou and Satoh, 2005). The result of this analysis shows that the majority of these genes are used during development. Strikingly, 74% are expressed as maternal messages in the egg, and 56% are expressed zygotically; only 14 of the genes are not expressed during the period studied. The results of this effort were

then used to lay the foundations for a gene regulatory network describing the patterning of the early ciona embryo (Imai et al., 2006). Since the sea urchin will be only the second organism with such a detailed accounting of transcription factor usage during embryogenesis, it will be interesting to see if similar or different patterns emerge. The comparison between these two organisms will be particularly interesting as the sea urchin develops through a maximally indirect mechanism: the larval structure laid out during embryogenesis is ultimately completely reabsorbed and the adult body plan arises from only a small subset of set-aside cells.

Conclusion

Cis-regulatory architecture lies at the heart of fundamental questions in biology. In a causal sense, *cis*-regulatory and gene network architecture provide the explanation of how development is determined by the regulatory DNA sequence. From emerging developmental gene regulatory networks in several model organisms, it is clear that these networks are built up from certain basic subroutines. With the sequencing of the sea urchin genome, it now becomes practical to fully describe one such system, the sea urchin endomesodermal gene regulatory network. Identifying and characterizing the developmental expression of all the transcription factors in the organism's genome will highlight any players still missing from the network. Furthermore, the compiled statistics on regulatory gene expression will provide further insight into how these genes as a whole are used in development.

References

Amore, G., Yavrouian, R. G., Peterson, K. J., Ransick, A., McClay, D. R., Davidson, E. H., 2003. *Spdeadringer*, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. Dev. Biol. 261**,** 55-81.

Andrioli, L. P. M., Vasisht, V., Theodosopoulou, E., Oberstein, A., Small, S., 2002. Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. Development 129**,** 4931-4940.

Arbeitman, M. N., Furlong, E. E. M., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W., White, K. P., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297**,** 2270-2275.

Arnone, M. I., Davidson, E. H., 1997. The hardwiring of development: Organization and function of genomic regulatory systems. Development 124**,** 1851-1864.

Arnosti, D. N., Barolo, S., Levine, M., Small, S., 1996. The *eve* stripe 2 enhancer employs multiple modes of transcriptional synergy. Development 122**,** 205-214.

Barrow, J. R., Stadler, H. S., Capecchi, M. R., 2000. Roles of Hoxa1 and Hoxa2 in patterning the early hindbrain of the mouse. Development 127**,** 933-944.

Baugh, L. R., Hill, A. A., Claggett, J. M., Hill-Harfe, K., Wen, J. C., Slonim, D. K., Brown, E. L., Hunter, C. P., 2005. The homeodomain protein Pal-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo. Development 132**,** 1843-1854.

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., Hunter, C. P., 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development 130**,** 889-900.

Bolouri, H., Davidson, E. H., 2002a. Modeling DNA sequence-based *cis*-regulatory gene networks. Dev. Biol. 246**,** 2-13.

Bolouri, H., Davidson, E. H., 2002b. Modeling transcriptional regulatory networks. Bioessays 24**,** 1118-1129.

Davidson, E. H. 1986. Gene Activity in Early Development. Academic Press, Orlando.

Davidson, E. H. 2006. Genomic Regulatory Systems. Academic Press, San Diego.

Davidson, E. H., Cameron, R. A., Ransick, A., 1998. Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. Development 125**,** 3269-3290.

Davidson, E. H., McClay, D. R., Hood, L., 2003. Regulatory gene networks and the properties of the developmental process. Proc. Natl. Acad. Sci. USA 100**,** 1475-1480.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L.,

Bolouri, H., 2002a. A genomic regulatory network for development. Science 295**,** 1669-1678.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Schilstra, M. J., Clarke, P. J. C., Rust, A. G., Pan, Z. J., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., Bolouri, H., 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. 246**,** 162-190.

Erwin, D. H., Davidson, E. H., 2002. The last common bilaterian ancestor. Development 129**,** 3021-3032.

Frasch, M., Levine, M., 1987. Complementary patterns of *even-skipped* and *fushi-tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. Genes & Development 1**,** 981-995.

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T., Jaynes, J. B., 1999. Analysis of an *even-skipped* rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. Development 126**,** 2527-2538.

Fukushige, T., Hawkins, M. G., McGhee, J. D., 1998. The GATA-factor *elt-2* is essential for formation of the *Caenorhabditis elegans* intestine. Dev. Biol. 198**,** 286-302.

Gaudet, J., Mango, S. E., 2002. Regulation of organogenesis by the *Caenorhabditis elegans*, FoxA protein Pha-4. Science 295**,** 821-825.

Gray, S., Levine, M., 1996. Transcriptional repression in development. Curr. Opin. Cell Biol. 8**,** 358-364.

Gurdon, J. B., Lemaire, P., Kato, K., 1993. Community effects and related phenomena in development. Cell 75**,** 831-834.

Harding, K., Hoey, T., Warrior, R., Levine, M., 1989. Autoregulatory and gap gene response elements of the *even-skipped* promoter of *Drosophila*. EMBO J 8**,** 1205-1212.

Imai, K. S., Hino, K., Yagi, K., Satoh, N., Satou, Y., 2004. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: Towards a comprehensive understanding of gene networks. Development 131**,** 4047-4058.

Imai, K. S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science 312**,** 1183-1187.

Istrail, S., Davidson, E. H., 2005. Logic functions of the genomic *cis*-regulatory code. Proc. Natl. Acad. Sci. USA 102**,** 4954-4959.

Jiang, J., Levine, M., 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the Dorsal gradient morphogen. Cell 72**,** 741-752.

Levine, M., Davidson, E. H., 2005. Gene regulatory networks for development. Proc. Natl. Acad. Sci. USA 102**,** 4936-4942.

Macdonald, P. M., Ingham, P., Struhl, G., 1986. Isolation, structure, and expression of *even-skipped* – a 2nd pair-rule gene of *Drosophila* containing a homeobox. Cell 47**,** 721-734.

Maduro, M. F., Rothman, J. H., 2002. Making worm guts: The gene regulatory network of the *Caenorhabditis elegans* endoderm. Dev. Biol. 246**,** 68-85.

Marshall, S. D. G., McGhee, J. D., 2001. Coordination of *ges-1* expression between the *Caenorhabditis* pharynx and intestine. Dev. Biol. 239**,** 350-363.

Metzger, R. J., Krasnow, M. A., 1999. Development – Genetic control of branching morphogenesis. Science 284**,** 1635-1639.

Miwata, K., Chiba, T., Horii, R., Yamada, L., Kubo, A., Miyamura, D., Satoh, N., Satou, Y., 2006. Systematic analysis of embryonic expression profiles of zinc finger genes in *Ciona intestinalis*. Dev. Biol. 292**,** 546-554.

Oliveri, P., Carrick, D. M., Davidson, E. H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246**,** 209-228.

Oliveri, P., Davidson, E. H., McClay, D. R., 2003. Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. Dev. Biol. 258**,** 32-43.

Ransick, A., Davidson, E. H., 1998. Late specification of veg(1) lineages to endodermal fate in the sea urchin embryo. Dev. Biol. 195**,** 38-48.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C., Davidson, E. H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246**,** 132-147.

Rast, J. P., Cameron, R. A., Poustka, A. J., Davidson, E. H., 2002. Brachyury target genes in the early sea urchin embryo isolated by differential macroarray screening. Dev. Biol. 246**,** 191-208.

Satou, Y., Satoh, N., 2005. Cataloging transcription factor and major signaling molecule genes for functional genomic studies in *Ciona intestinalis*. Dev. Genes Evol. 215**,** 580-596.

Small, S., Blair, A., Levine, M., 1992. Regulation of *even-skipped* stripe-2 in the *Drosophila* embryo. EMBO J 11**,** 4047-4057.

Small, S., Blair, A., Levine, M., 1996. Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. Dev. Biol. 175**,** 314-324.

Stanojevic, D., Small, S., Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. Science 254**,** 1385-1387.

Stathopoulos, A., Levine, M., 2002. Dorsal gradient networks in the *Drosophila* embryo. Dev. Biol. 246**,** 57-67.

Wagner, R. A., Tabibiazar, R., Liao, A., Quertermous, T., 2005. Genome-wide expression dynamics during mouse embryonic development reveal similarities to *Drosophila* development. Dev. Biol. 288**,** 595-611.

Yuh, C. H., Bolouri, H., Davidson, E. H., 1998. Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. Science 279**,** 1896-1902.

Yuh, C. H., Bolouri, H., Davidson, E. H., 2001. *Cis*-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control. Development 128**,** 617-629.

Yuh, C. H., Davidson, E. H., 1996. Modular *cis*-regulatory organization of *endo16*, a gut-specific gene of the sea urchin embryo. Development 122**,** 1069-1082.

Yuh, C. H., Dorman, E. R., Howard, M. L., Davidson, E. H., 2004. An *otx cis*-regulatory module: A key node in the sea urchin endomesoderm gene regulatory network. Dev. Biol. 269**,** 536-551.

Yuh, C. H., Ransick, A., Martinez, P., Britten, R. J., Davidson, E. H., 1994. Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea-urchin embryo. Mech. Dev. 47**,** 165-186.

Zeller, R. W., Britten, R. J., Davidson, E. H., 1995. Developmental utilization of *Spp3a1* and *Spp3a2* – 2 proteins which recognize the same DNA target site in several sea-urchin gene regulatory regions. Dev. Biol. 170**,** 75-82.

Zelzer, E., Shilo, B. Z., 2000. Cell fate choices in *Drosophila* tracheal morphogenesis. Bioessays 22**,** 219-226.

Zhu, J. W., Hill, R. J., Heid, P. J., Fukuyama, M., Sugimoto, A., Priess, J. R., Rothman, J. H., 1997. *end-1* encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos. Genes & Development 11**,** 2883-2896.

Chapter 1

Identification and Characterization of Homeobox Transcription Factor Genes in *S. purpuratus*, and Their Expression in Embryonic Development

Meredith Howard-Ashby, Stefan C. Materna, C. Titus Brown, Lily Chen, R. Andrew Cameron, and Eric H. Davidson

In press, Developmental Biology.

Abstract

A set of 96 homeobox transcription factors was identified in the *Strongylocentrotus purpuratus* genome using permissive blast searches with a large collection of authentic homeodomain sequences from mouse, human and fly. A phylogenetic tree was constructed to compare the sea urchin homeobox gene family to those of vertebrates, with the result that with the only a few exceptions, orthologs of all vertebrate homeodomain genes were uncovered by our search. QPCR time course measurements revealed that 65% of these genes are expressed within the first 48 hours of development (late gastrula). For genes displaying sufficiently high levels of transcript during the first 24 hours of development (late blastula), whole mount *in situ* hybridization was carried out up to 48 hours to determine spatial patterns of expression. The results demonstrate that homeodomain transcription factors participate in multiple and diverse

developmental functions, in that they are used at a range of time points and in every territory of the developing embryo.

Introduction

Transcription factors are the key players in the gene networks directing development. These networks consist essentially of genes encoding sequence specific regulatory proteins, the targets of which encode other transcription factors, thereby initiating cascades of overlapping directives which ultimately specify the many embryonic territories. To solve the architecture of developmental gene networks requires primary knowledge of which transcription factors are active in the embryo and when and where they are expressed. The availability of the *Strongylocentrotus purpuratus* genome sequence, which has just been obtained by the Human Genome Sequencing Center at Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/; http://www.ncbi.nlm.nih.gov/genome/guide/seaurchin/), has made it possible to identify systematically all the transcription factors encoded in the genome. Thus we sought to find and annotate all genes encoding sequence specific DNA binding proteins predicted by the genome sequence. We then determined whether each is expressed in the early to mid-stage embryo, and, for active genes, established the temporal and spatial modes of expression.

Transcription factors fall into several large families defined by the structures of their DNA binding domains. The largest of these families in *S. purpuratus* is the Zn Finger family, an analysis of which is described in another paper of this series (Materna et al., 2006). The next largest is our present subject, the homeodomain family. Here we consider all subclasses of homeodomain regulatory genes except for the *hox* and *parahox* genes, which are the subject of a separate report (Arnone et al., 2006). Other classes of

transcription factors are dealt with in additional papers (Ets family factors, (Rizzo et al., 2006); Forkhead family factors, (Tu et al., 2006); and all other families (Howard-Ashby et al., 2006).

Materials and methods

Identification of transcription factor sequences

Most of the transcription factors considered here were initially identified from the unassembled sea urchin genome traces and the November, 2004 Baylor University draft genome assembly using a reference database of known transcription factors (excluding zinc fingers). This "rake," was assembled from two sources: nr human, mouse and fly sequences tagged as "transcription factor" and the GO seqdblite databases GO:0003700, GO:0000130, GO:0030528, GO:0003705, GO:0003702, and GO:0003677. Entries were removed if they contained the descriptors "general transcription factor II," "TFII," "TFIII," "protease," "histone," "reverse transcriptase," "nucleosome," "RNA polymerase," "DNA replications," "chromatin," "helicase," "DNase," or "exonuclease." Any nonhomeodomain/nonGATA zinc finger proteins were also removed from the rake database. The final rake contained approximately 4900 protein sequences.

Tblastn (Altschul et al., 1990) of the protein sequences in the rake against the individual traces, as well as the translated Baylor draft assembly (cutoff = e-10) was used to coarsely identify all traces or contigs potentially encoding transcription factors. Blastx of this subset of sequences vs. the rake protein database (cutoff e-12) was then used to

highlight the locations of exons encoding transcription factor specific conserved domains (e.g., bHLH, homeodomain, sox). Finally, the isolated conserved domains were blasted (tblastn) against NCBI's nr database to establish the closest known homologues. To avoid redundancy, efforts were made to group multiple exons from the same protein. Complementary exons from the same large contig as well as complementary exons from smaller contigs with the same closest homologues were assigned one unique number/gene name. PCR of sea urchin cDNA was used to confirm that different exons were in fact part of the same transcript. Our set of newly identified genes was then compared to those in the Baylor GLEAN3 gene models. There were approximately 30 of our genes not present in the GLEAN3 database. Similarly, we added approximately 25 new transcription factors to our data set after finding them among the GLEAN3 models. All of the data obtained in this study were incorporated in the sea urchin genome annotation effort orchestrated by HGSC at Baylor College of Medicine.

Phylogenetic analysis

A phylogenetic tree comparing sea urchin homeobox genes and homologues from multiple other species was constructed in order to name accurately the newly identified genes. Reference homeodomain sequences from *H. sapiens*, *C. elegans, C. briggsae, D. rerio, D. melanogaster*, and *M. mus* were obtained from the supplementary materials of Nam and Nei (Nam and Nei, 2005) and from NCBI by BLAST 2.2.12 (Altschul et al., 1990) search of nr with *S. purpuratus* homeobox sequences. Multiple sequence alignment of the homeodomains was done with CLUSTALW 1.83 for the UNIX operating system.

The tree was constructed with MEGA version 3. (Kumar et al., 2004) using the neighbor joining method and 1000 bootstrap replications. In addition to the data shown here, a more exhaustive tree including sequences from all the above species can be found in the supplementary materials, along with all the homeodomain sequences used.

QPCR data

QPCR was used to determine the expression profile of each identified transcription factor from unfertilized egg to 48 h. mRNA was isolated from egg, 6, 12, 18, 24, 36, and 48 h embryos with the Sigma GenElute Mammalian Total RNA Miniprep Kit, per the manufacturers instructions. Residual DNA was digested with DNase I using the DNA-free kit (Ambion, Austin, TX). cDNA was prepared from 38.5 μL of mRNA sample using the TaqMan Reverse Transcription Reagents Kit (Applied Biosystems, Foster City, CA) in a 100 μL reaction, following the kit instructions. QPCR primers were chosen such that amplicons were preferably between 100 and 140 bp long, though in some cases amplicons were as short as 80 bp or as long as 160 bp. Primer and amplicon sequences can be found online at http://sugp.caltech.edu. To avoid primer inefficiency due to the high rate of polymorphism in the sea urchin genome, primers were chosen to be within the most conserved DNA binding domain of each transcription factor. Amplification reactions were analyzed on an ABI 5700 sequence detection system using SYBR Green chemistry (PE Biosystems, Foster City, CA). All primer pairs were validated by QPCR against a positive (genomic DNA) and negative (water) control. Each 20 μL control reaction contained 10 μL SYBR Green reagent, 2.4 μL forward and reverse

primer mix (5 μM each), 1 μL digested genomic DNA (40 ng) or water, and 6.6 μL water. Template genomic DNA was a mixture of *Kpn*I and *Eco*RI digested genomic DNA. Expression was measured at six time points in triplicate: egg, 6, 12, 18, 24, 36, and 48 h. Each 10 μL reaction included 5 μL SYBR Green reagent, 2.5 μL forward and reverse primer mix (5 μM each), 0.5 μL cDNA, and 3.3 μL. Thermal cycling parameters were 95 °C for 30 s, 60 °C for 1 min, 40 cycles, followed by a denaturation step to verify a single product. All QPCR experiments were performed in triplicate against two preparations of cDNA.

A QPCR experiment measures the number of cycles needed to attain a threshold concentration of QPCR product (C_t). The number of cycles needed for the standard to reach a specified (C_t) can be compared to the C_t for an unknown. A higher C_t for the unknown implies a lower initial concentration in the sample, and vice versa. The threshold value is chosen to fall within the exponential amplification phase, before limiting reagents become a factor in the efficiency of each cycle. Given that ubiquitin sequence domains are present at a constant 87,000 copies/embryo (Nemer et al., 1991; Ransick et al., 2002), and assuming a QPCR amplification rate of 1.9-fold per cycle, the difference in C_t between an unknown and ubiquitin for a given sample can be translated directly into the number of copies per embryo. Our QPCR data were compared to the genome tiling array data as an external control for the identification of unexpressed genes (Samanta et al., 2006). While the tiling data are not quantitative, genes which are not expressed in the early embryo should not give any signal. The two data sets are in strong agreement, with only a few genes giving no or very low QPCR signal showing some

signal according to the tiling array data. Alternately, positive QPCR results were always supported by the tiling array data.

The time course data were plotted on a logarithmic scale to simplify comparison of expression profiles with very different minimum and maximums. Each gene was categorized as to whether expression was maternal only, maternal and zygotic, zygotic only, constant, or null up to 48 h. Genes expressed zygotically were further categorized as to the time by which expression is first activated. Complete time course data can be found online at http://sugp.caltech.edu/ .

Whole mount* in situ *hybridization

In situ probes were designed for genes with zygotic expression within the minimum significant range by 24 h post-fertilization. We attempted to use probes at least 600 bp long, though in some cases shorter probes were used if they gave a positive, specific result. The sequence of the probes was derived either from a sufficiently long exon or multiple exons discovered in our blast searches, or from message sequence identified by blastn against cDNA libraries submitted to NCBI, or from a Genscan prediction of additional exons present on the same contig as a known exon.

All probe sequences were initially amplified using the Expand Hi-Fidelity PCR System (Roche) and sea urchin cDNA and confirmed by sequencing using ABI Prism BigDye Terminator Cycle Sequencing on an ABI 377 sequencer (Applied Biosystems, Foster City, CA). Probes were transcribed either from linearized plasmid after cloning the PCR products, or directly from a PCR fragment made with primers incorporating T7 and

SP6 promoters. The primers used to make these probes can be found at http://sugp.caltech.edu/. Digoxigenin-labeled RNA probes were transcribed using the Roche DIG-labeling mix.

Whole mount *in situ* hybridization was performed as previously described (Otim et al., 2004).

Results

Identifying transcription factor genes

Our strategy was to search for putative sea urchin regulatory proteins by homology to known proteins, taking advantage of the strong conservation of DNA binding domains among even distantly related organisms. A reference database, which we named our "rake," was assembled by extracting human, mouse, and fly transcription factor sequences from NCBI nr and GO-seqdblite databases. We then used tblastn to drag our rake through the genome, pulling out any sea urchin sequence even weakly matching a known transcription factor. The accumulated sequences were then sorted into families using blastx against the rake with a more selective cutoff. Sequences not matching a rake protein better than 1e-12 were discarded, and those retained were associated with the best matching known protein. Within these groupings it was possible to remove redundant sequences manually and also to pair together complementary gene fragments from

Table 1.1. Glean ID and Index numbers of Identified Genes.

Gene Name	Index	Glean ID	Gene Name	Index	Glean ID
Sp-alx1	-	SPU_22817,SPU_25302	Sp-lhx3.4	105	SPU_01975
Sp-alx4	184	SPU_22816	Sp-lim1	44	SPU_06991
Sp-arx	297	SPU_19338	Sp-lmx1	314	SPU_14157
Sp-arxl	298	SPU_17249	Sp-mbx1	270	SPU_11297
Sp-arxl2	389	SPU_21491	Sp-meis	345	SPU_11202
Sp-atbf1	78	SPU_17348	Sp-mox	109	SPU_23868,SPU_25486
Sp-awh	122	SPU_18954	Sp-msx	74	SPU_22049
Sp-barhl	259	SPU_14164	Sp-msxl	395	SPU_20565
Sp-barx	260	SPU_01519,SPU_03920	Sp-not	-	SPU_02129
Sp-brn124	-	SPU_16443	Sp-nk1	265	SPU_12491
Sp-brn3	18	SPU_25632	Sp-nk2.1	266	SPU_00757
Sp-cdx2	300	SPU_24715,SPU_19656	Sp-nk2.2	75	SPU_00756
Sp-chx10	146	SPU_00485	Sp-nk2.5	14	SPU_05472
Sp-cutl	331	SPU_03595	Sp-nk3.2	267	SPU_13047
Sp-dbx1	261	-	Sp-nk6.1	127	SPU_12699
Sp-dlx	309	SPU_02815	Sp-nk7	327	SPU_22573
Sp-emx	150	SPU_02592	Sp-oct1.2	26	SPU_09262
Sp-en	12	SPU_20975	Sp-otp	272	SPU_19290
Sp-eve	257	SPU_12253	Sp-otx	-	SPU_10424
Sp-exd	68	SPU_05435,SPU_23739	Sp-pax1.9	16	SPU_06683
Sp-eyg	321	SPU_19129	Sp-pax258	47	SPU_14539
Sp-eygl	393	SPU_16786	Sp-pax4l	394	SPU_17635,SPU_17636
Sp-gbx	610	SPU_25492	Sp-pax6	296	SPU_06786
Sp-gsc	-	SPU_15982	Sp-paxA	273	SPU_27334
Sp-gsh1	317	SPU_13436	Sp-paxB	274	SPU_18351
Sp-hb9	258	SPU_02816	Sp-paxC	108	SPU_00276
Sp-hbn	324	SPU_23177	Sp-phb1	392	SPU_08112
Sp-hex	263	SPU_27215	Sp-phb2	396	SPU_24093
Sp-hlx	340	SPU_14802	Sp-pbx	-	SPU_23739
Sp-hnf1	56	SPU_08196	Sp-phox2	269	SPU_13464
Sp-hnf6	-	SPU_16449	Sp-pitx1	163	SPU_14461,SPU_24163
Sp-hox1.tlx1	85	SPU_17352	Sp-pitx2	275	SPU_04599
Sp-hox11.13a	97	SPU_02632	Sp-pitx3	84	SPU_06159,SPU_04598
Sp-hox11.13b	256	SPU_02631	Sp-pknox	330	SPU_12122
Sp-hox11.13c	294	SPU_00388	Sp-pmar1	-	SPU_14721
Sp-hox2	293	SPU_12252,SPU_00386	Sp-pou6	618	SPU_10438
Sp-hox3	253	SPU_27568	Sp-prox1	343	SPU_15984
Sp-hox4.5	50.1	SPU_05169	Sp-prx	311	SPU_18951
Sp-hox6	254	SPU_05171	Sp-rough	606	SPU_07242
Sp-hox7	255	SPU_05170,SPU_02634	Sp-rx	151	SPU_16786
Sp-hox8	50.2	SPU_02630,SPU_21309	Sp-shox	310	SPU_19268
Sp-hox9.10	45	SPU_02633	Sp-sip	81	SPU_22242
Sp-irxA	200	SPU_10351	Sp-six1.2	15	SPU_17379
Sp-irxB	299	SPU_11246	Sp-six3	2	SPU_18908
Sp-isl	32	SPU_23730	Sp-six4	21	SPU_17380
Sp-lass6	388	SPU_00948	Sp-tgif	43	SPU_18126
Sp-lbx	115	SPU_14177	Sp-unc4.1	334	SPU_01739,SPU_13704

different contigs. Sequence pairings were confirmed by PCR against sea urchin mRNA and checked against assembled supertigs.

Our search identified a total of 96 homeodomain transcription factors, including those already known. The largest subfamilies are the *paired* class, with 31 members, and the *hox/extended hox* family, with 21 members. We also found 11 atypical homeodomain genes, 12 *nk* class genes, six *lim* homeodomain genes, and 15 members of smaller subfamilies including *dl*, *cut*, *pou*, *barx* and zinc finger homeodomain genes. The complete list of identified genes, with their corresponding gene model numbers assigned by the HGSC at Baylor University, is given in Table 1.1.

Since this search was conducted without any sea urchin genes in our rake database, a set of known sea urchin transcription factors provides a convenient check on the success of our method. Of 20 endomesoderm gene network transcription factors of all types, all but one were identified. However, unpublished work indicates that the gene we missed, *Sp-pmar1*, is in fact missing from the genome traces and assembly. Among homeodomain genes in particular, we successfully identified 10/11 hox genes (Arenas-Mena et al., 2000) and 7/7 pax genes (Czerny et al., 1997). We expect, therefore, that this analysis includes nearly all sea urchin homeobox genes. Of course, were there sea urchin genes encoding transcription factors the DNA binding domains which differ from known DNA binding domains in that they are not strongly conserved across species, these would likely be missed by our search method. Genes which fall across breaks, or lie within gaps in the genome assembly, could also have been missed.

A phylogenetic tree of sea urchin and human homeodomains was constructed in order to identify the new genes uncovered in the search, and determine the subgroup to

which they belong (fig. 1.1). In a few cases where the closest homologue is not a human gene, the appropriate gene from *Drosophila* or zebrafish was included. In general, previously reported sea urchin genes were also included in the tree to provide a full comparison to the vertebrate tool kit of homeodomain transcription factors. We did not include the genes of the *hox* and *parahox* clusters per se, as these have been well studied elsewhere (Arenas-Mena et al., 2000; Arenas-Mena et al., 1998; Arnone et al., 2006; Cameron et al., 2006; Martinez et al., 1999). Because of the number of sequences involved, the neighbor joining method with 1000 bootstrap resamplings was chosen over more computationally intensive tree construction methods such as maximum-parsimony and maximum-likelihood. In addition, neighbor joining has been shown to give the most accurate trees in cases where there are many sequences but the sequence lengths are short (Nam and Nei, 2005).

The tree in figure 1.1 illustrates the common heritage of the homeodomain family in deuterostomes. Figure 1.1A summarizes the overall relationship among the homeodomain subfamilies, which are shown in detail in figure 1.1B-E. Sea urchin genes are highlighted in purple text, with newly identified genes in larger font. Of the sea urchin homeodomain genes included in the tree, 66% cluster to a single clear ortholog or a set of paralogs with bootstrap values of 98 or above, and 93% cluster with bootstrap values of 80 or greater. Thus in the large majority of cases, the homeodomain sequences are highly conserved, and the assignment of cognates is clear cut.

The sea urchin has representatives of all major homeobox gene classes, and in fact contains homologues or sets of paralogs of nearly all human homeobox genes. For

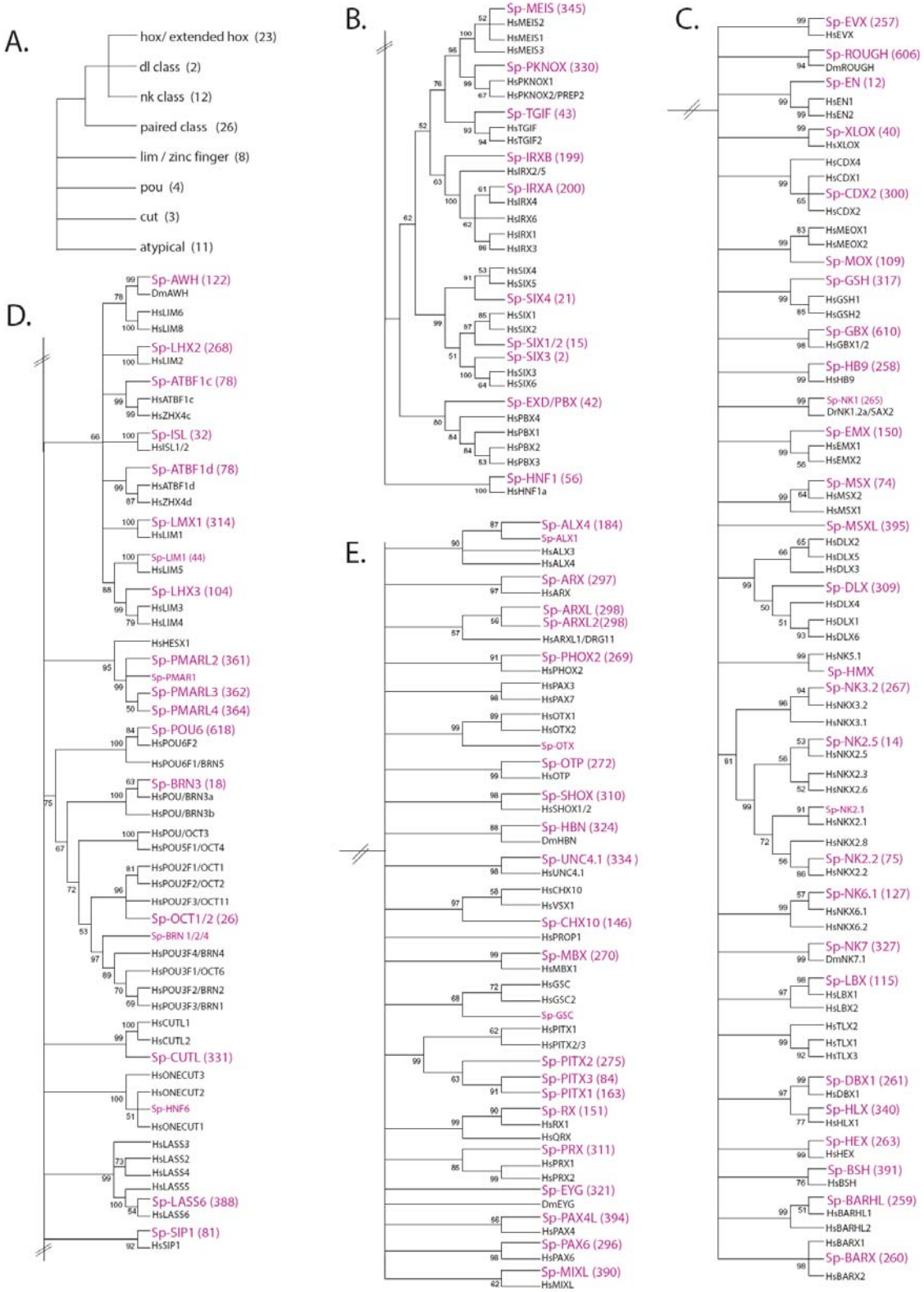

A.
hox/ extended hox (23)
dl class (2)
nk class (12)
paired class (26)
lim / zinc finger (8)
pou (4)
cut (3)
atypical (11)
B.
Sp-MEIS (345)
HsMEIS2
HsMEIS1
HsMEIS3
Sp-PKNOX (330)
HsPKNOX1
HsPKNOX2/PREP2
Sp-TGIF (43)
HsTGIF
HsTGIF2
Sp-IRXB (199)
HsIRX2/5
Sp-IRXA (200)
HsIRX4
HsIRX6
HsIRX1
HsIRX3
HsSIX4
HsSIX5
Sp-SIX4 (21)
HsSIX1
HsSIX2
Sp-SIX1/2 (15)
Sp-SIX3 (2)
HsSIX3
HsSIX6
Sp-EXD/PBX (42)
HsPBX4
HsPBX1
HsPBX2
HsPBX3
Sp-HNF1 (56)
HsHNF1a
C.
Sp-EVX (257)
HsEVX
Sp-ROUGH (606)
DmROUGH
Sp-EN (12)
HsEN1
HsEN2
Sp-XLOX (40)
HsXLOX
HsCDX4
HsCDX1
Sp-CDX2 (300)
HsCDX2
HsMEOX1
HsMEOX2
Sp-MOX (109)
Sp-GSH (317)
HsGSH1
HsGSH2
Sp-GBX (610)
HsGBX1/2
Sp-HB9 (258)
HsHB9
Sp-NK1 (265)
DrNK1.2a/SAX2
Sp-EMX (150)
HsEMX1
HsEMX2
Sp-MSX (74)
HsMSX2
HsMSX1
Sp-MSXL (395)
HsDLX2
HsDLX5
HsDLX3
Sp-DLX (309)
HsDLX4
HsDLX1
HsDLX6
HsNK5.1
Sp-HMX
Sp-NK3.2 (267)
HsNKX3.2
HsNKX3.1
Sp-NK2.5 (14)
HsNKX2.5
HsNKX2.3
HsNKX2.6
Sp-NK2.1
HsNKX2.1
HsNKX2.8
Sp-NK2.2 (75)
HsNKX2.2
Sp-NK6.1 (127)
HsNKX6.1
HsNKX6.2
Sp-NK7 (327)
DmNK7.1
Sp-LBX (115)
HsLBX1
HsLBX2
HsTLX2
HsTLX1
HsTLX3
Sp-DBX1 (261)
HsDBX1
Sp-HLX (340)
HsHLX1
Sp-HEX (263)
HsHEX
Sp-BSH (391)
HsBSH
Sp-BARHL (259)
HsBARHL1
HsBARHL2
HsBARX1
Sp-BARX (260)
HsBARX2
D.
Sp-AWH (122)
DmAWH
HsLIM6
HsLIM8
Sp-LHX2 (268)
HsLIM2
Sp-ATBF1c (78)
HsATBF1c
HsZHX4c
Sp-ISL (32)
HsISL1/2
Sp-ATBF1d (78)
HsATBF1d
HsZHX4d
Sp-LMX1 (314)
HsLIM1
Sp-LIM1 (44)
HsLIM5
Sp-LHX3 (104)
HsLIM3
HsLIM4
HsHESX1
Sp-PMARL2 (361)
Sp-PMAR1
Sp-PMARL3 (362)
Sp-PMARL4 (364)
Sp-POU6 (618)
HsPOU6F2
HsPOU6F1/BRN5
Sp-BRN3 (18)
HsPOU/BRN3a
HsPOU/BRN3b
HsPOU/OCT3
HsPOU5F1/OCT4
HsPOU2F1/OCT1
HsPOU2F2/OCT2
HsPOU2F3/OCT11
Sp-OCT1/2 (26)
Sp-BRN 1/2/4
HsPOU3F4/BRN4
HsPOU3F1/OCT6
HsPOU3F2/BRN2
HsPOU3F3/BRN1
HsCUTL1
HsCUTL2
Sp-CUTL (331)
HsONECUT3
HsONECUT2
Sp-HNF6
HsONECUT1
HsLASS3
HsLASS2
HsLASS4
HsLASS5
Sp-LASS6 (388)
HsLASS6
Sp-SIP1 (81)
HsSIP1
E.
Sp-ALX4 (184)
Sp-ALX1
HsALX3
HsALX4
Sp-ARX (297)
HsARX
Sp-ARXL (298)
Sp-ARXL2(298)
HsARXL1/DRG11
Sp-PHOX2 (269)
HsPHOX2
HsPAX3
HsPAX7
HsOTX1
HsOTX2
Sp-OTX
Sp-OTP (272)
HsOTP
Sp-SHOX (310)
HsSHOX1/2
Sp-HBN (324)
DmHBN
Sp-UNC4.1 (334)
HsUNC4.1
HsCHX10
HsVSX1
Sp-CHX10 (146)
HsPROP1
Sp-MBX (270)
HsMBX1
HsGSC
HsGSC2
Sp-GSC
HsPITX1
HsPITX2/3
Sp-PITX2 (275)
Sp-PITX3 (84)
Sp-PITX1 (163)
Sp-RX (151)
HsRX1
HsQRX
Sp-PRX (311)
HsPRX1
HsPRX2
Sp-EYG (321)
DmEYG
Sp-PAX4L (394)
HsPAX4
Sp-PAX6 (296)
HsPAX6
Sp-MIXL (390)
HsMIXL

Figure 1.1. The phylogenetic trees A-E depict *S. purpuratus* homeobox genes and their closest relatives. The closest human homologues (black text) are shown for each sea urchin gene (purple text) except where there are none. In such cases the closest gene from *Drosophila* or zebrafish is shown. Those genes known from previous studies are indicated in small font. The trees were constructed from homeodomain sequences using the neighbor joining method and 1000 bootstrap iterations. Homeodomain sequences and accession numbers can be found in supporting materials. (A) The master tree shows the relationships between the subfamilies depicted in (B-E); (B) Atypical and *six* class homeodomains; (C) Extended *hox* family, *dl*, and *nk* class genes. Since the phylogeny of the canonical *hox* cluster and *parahox* genes has been well studied, those genes have been omitted from this tree; (D) *lim*, *pou*, *cut*, zinc finger, and miscellaneous homeobox genes; (E) Paired family homeobox genes.

example, the human atypical homeodomain genes *Hsmeis1*, *Hsmeis2*, and *Hsmeis3* have just one sea urchin homolog, *Sp-meis*. In only one case, *Hsvax1/2*, were we unable to find a sea urchin homolog of a vertebrate homeobox gene using our computational search method. There were also only a few sea urchin genes which, inversely, did not have very close homologs in vertebrates, but were instead closer to *Drosophila* genes. Specifically, these are *Sp-hbn*, *Sp-eyg*, *Sp-rough*, and *Sp-nk7*. Finally, two sea urchin homeodomains identified by the genome annotation process were not closely related to any human, mouse, *Drosophila* or *C. elegans* homeodomains. The two paired class homeodomain proteins were named *paired homeodomain1* and *paired homeodomain 2* (*Sp-phd1* and *Sp-phd2*).

An additional phylogenetic tree was constructed to characterize members of the *pax* sub-family. While *pax* genes are generally grouped with homeobox genes, in fact many do not have a canonical homeodomain. *Pax2/5/8/B* supergroup homologues have truncated homeodomains, while *pax1/9* and cniderian *paxA* genes have no homeoboxes at all. For this reason, the *pax* domain is a more useful reference for determining overall *pax*

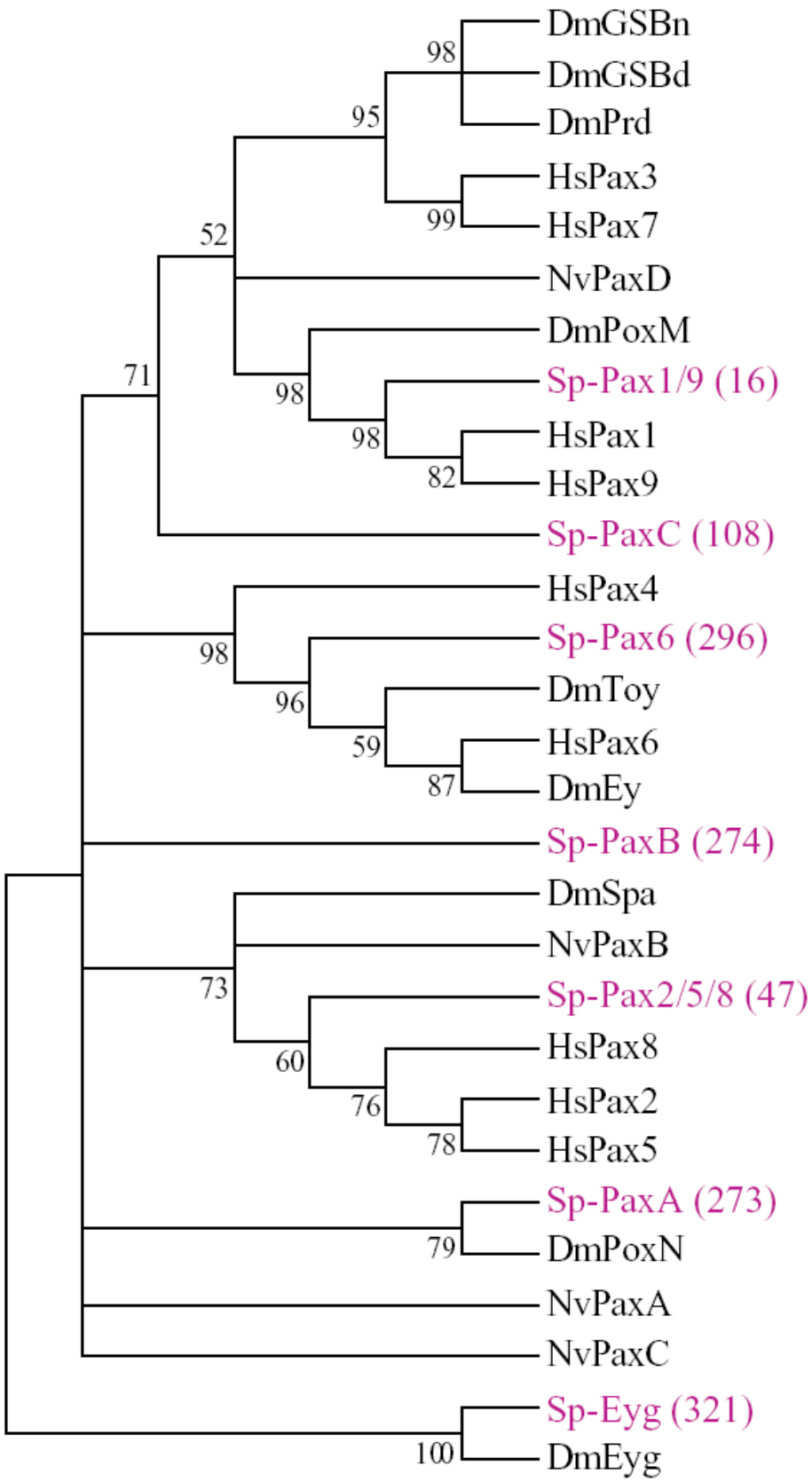

DmGSBn
DmGSBd
DmPrd
HsPax3
HsPax7
NvPaxD
DmPoxM
Sp-Pax1/9 (16)
HsPax1
HsPax9
Sp-PaxC (108)
HsPax4
Sp-Pax6 (296)
DmToy
HsPax6
DmEy
Sp-PaxB (274)
DmSpa
NvPaxB
Sp-Pax2/5/8 (47)
HsPax8
HsPax2
HsPax5
Sp-PaxA (273)
DmPoxN
NvPaxA
NvPaxC
Sp-Eyg (321)
DmEyg
98
95
99
52
71
98
98
82
98
96
59
87
73
60
76
78
79
100

Figure 1.2. The phylogenetic tree of *S. purpuratus*, human, *D. melonagaster*, and *Nematostella vectensis pax* family genes. Sea urchin genes are highlighted in purple, with the newly identified *Sp-eyg* in larger font. The tree was constructed using the pax domain sequences (see Supplementary materials) and a neighbor joining algorithm with 1000 bootstrap iterations.

gene phylogeny. Figure 1.2 shows a neighbor-joining tree of *pax* genes from sea urchin, human, *D. melanogaster*, and the cniderian *Nematostella vectensis*, constructed with 1000 bootstrap iterations. *Sp-pax1/9, Sp-pax2/5/8* and *Sp-pax6* cluster strongly to their vertebrate cognates. *Sp-paxA*, *Sp-paxB*, and *Sp-paxC*, named by reference to *Paracentrotus lividus* genes, do not have clear vertebrate orthologs, though *Sp-paxA* has homology to *Drosophila pox-neuro* and *Sp-paxC* appears to belong to the *pax1/9/3/7* super-group. Finally *Sp-eyg*, in which only the 3' RED part of the paired domain is conserved, is closest to *Dm-eyg*, which also has a truncated paired domain. Another gene, *Sp-pax4-like*, has a homeodomain sequence which appears to be orthologous to that of *Hs-pax4*, but it has no pax domain or octapeptide. However, the predicted sequence of *Sp-pax4-like* begins with the homeobox, so it is possible the N-teminal region of the gene is missing due to an assembly error.

Temporal gene expression patterns

To determine which homeobox genes are active during early development, the expression level of all the newly identified genes was quantified by QPCR. Given the high rate of polymorphisms in S. *purpuratus*, QPCR primers were designed very carefully to assure uniform primer efficiency and consistent results. As much as possible, QPCR primers were chosen to fall within the most conserved part of the protein, the

DNA binding domain. Once an appropriate region was selected, we used a short python script to identify by BLAST, retrieve, and align by clustalw (Higgens et al., 1994) the individual genomic sequencing reads used to assemble that short stretch of the genome. In this way we were able to rapidly identify many SNPs and avoid including these positions in our primers. Once the best locations for primers were mapped, another python script was used to pass this information to Primer3 (Rozen and Skaletsky, 2000), almost fully automating high quality primer design for large data sets.

Primer pairs were validated by QPCR against digested genomic DNA. Primers giving anomalously high or low amplification compared to the standard single copy gene ubiquitin were redesigned. Primer pairs with anomalous denaturation curves, potentially reflecting primer dimerization, were also redesigned. Finally, gene expression was measured quantitatively at six time points: unfertilized egg, 6, 12, 18, 24, 36, and 48 h post-fertilization. All primer validation and quantitative experiments were done in triplicate, and quantitative experiments were repeated using two cDNA preparations.

QPCR allows for the quantitative measurement of transcript levels by comparing the amplification of the target and a known standard. During every PCR cycle a fluorescent reporter dye is used to measure the increasing concentration of the amplicon. Thus, if the cellular copy number of the standard is known, and each PCR cycle produces an amplification of approximately 1.9-fold, the copy number of the unknown at a given time can be easily calculated from the difference in C_ts between the standard and the unknown (see materials and methods).

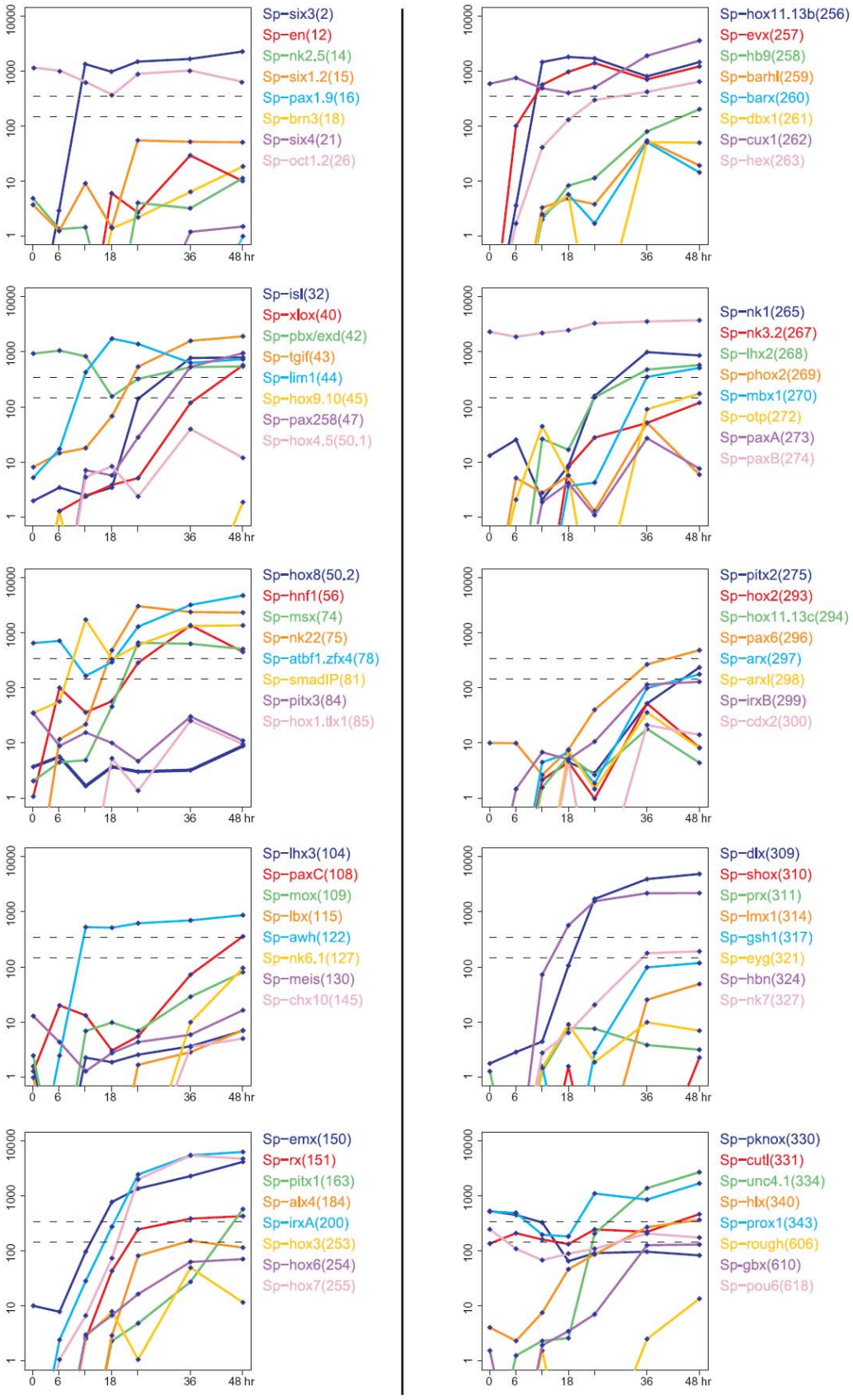

Sp-six3(2)
Sp-en(12)
Sp-nk2.5(14)
Sp-six1.2(15)
Sp-pax1.9(16)
Sp-brn3(18)
Sp-six4(21)
Sp-oct1.2(26)
Sp-hox11.13b(256)
Sp-evx(257)
Sp-hb9(258)
Sp-barhl(259)
Sp-barx(260)
Sp-dbx1(261)
Sp-cux1(262)
Sp-hex(263)
Sp-isl(32)
Sp-xlox(40)
Sp-pbx/exd(42)
Sp-tgif(43)
Sp-lim1(44)
Sp-hox9.10(45)
Sp-pax258(47)
Sp-hox4.5(50.1)
Sp-nk1(265)
Sp-nk3.2(267)
Sp-lhx2(268)
Sp-phox2(269)
Sp-mbx1(270)
Sp-otp(272)
Sp-paxA(273)
Sp-paxB(274)
Sp-hox8(50.2)
Sp-hnf1(56)
Sp-msx(74)
Sp-nk22(75)
Sp-atbf1.zfx4(78)
Sp-smadIP(81)
Sp-pitx3(84)
Sp-hox1.tlx1(85)
Sp-pitx2(275)
Sp-hox2(293)
Sp-hox11.13c(294)
Sp-pax6(296)
Sp-arx(297)
Sp-arxl(298)
Sp-irxB(299)
Sp-cdx2(300)
Sp-lhx3(104)
Sp-paxC(108)
Sp-mox(109)
Sp-lbx(115)
Sp-awh(122)
Sp-nk6.1(127)
Sp-meis(130)
Sp-chx10(145)
Sp-dlx(309)
Sp-shox(310)
Sp-prx(311)
Sp-lmx1(314)
Sp-gsh1(317)
Sp-eyg(321)
Sp-hbn(324)
Sp-nk7(327)
Sp-emx(150)
Sp-rx(151)
Sp-pitx1(163)
Sp-alx4(184)
Sp-irxA(200)
Sp-hox3(253)
Sp-hox6(254)
Sp-hox7(255)
Sp-pknox(330)
Sp-cutl(331)
Sp-unc4.1(334)
Sp-hlx(340)
Sp-prox1(343)
Sp-rough(606)
Sp-gbx(610)
Sp-pou6(618)
1
10
100
1000
10000
0
6
18
36
48 hr

Figure 1.3. Expression time courses of *S. purpuratus* homeobox genes. The graphs show gene expression levels from 0-48 h post fertilization, plotted on a logarithmic scale. The number of copies expressed per embryo was obtained by QPCR experiments done in triplicate (materials and methods). Dashed lines at 150 and 350 copies per embryo indicate an estimated minimum range for biologically significant expression of a transcription factor. The average copy number for low prevalence maternal transcripts in *S. purpuratus* eggs is 1600 copies per embryo (Davidson, 1986), and an arbitrary guideline of >400 copies/embryo should reasonably capture significant mRNAs encoding transcription factors in the egg. Note that in most cases the maternal contents are either well above this threshold or far below it. At later time points, expression as low as 200 copies /embryo can be detected by WMISH if expressed in a small domain of 20 cells at ~10 copies/cell. Allowing for some primer inefficiencies, we used a biological significance guideline of >150-350 copies/embryo. It is interesting to note that the time course presented here for Sp-oct1.2 (26) is somewhat different than has been previously described (Char et al., 1993). The discrepancy can be explained by reference to the transcriptome data, which indicates Sp-oct1.2 likely has alternate splice forms. Char et al. used a probe which measures expression of one splice variant, whereas the primers used to generate these data fall within the homeodomain and do not distinguish between variants.

The results, plotted on a logarithmic scale for easy side by side comparison, are shown in figure 1.3. Individual plots appear in order of the gene numbers assigned in the phylogenetic trees of figures 1.1 and 1.2. Data are shown here only for genes that had not previously been studied quantitatively. The graphs show the mRNA content per embryo for each homeobox gene over developmental time, compared to an arbitrary guideline to the threshold of biological significance, derived as indicated the legend of figure 1.3. Many qualitatively distinct time courses are evident. The majority of these genes are not represented in the maternal mRNA stockpile, and are either activated during embryogenesis or are not activated at all, i.e., up to 48 h (late gastrula) when our observations end. A minor fraction, 11/96 of the genes, is represented significantly in maternal mRNA (>400 transcripts per egg), and a very small group of only three genes is expressed significantly at constant levels throughout (i.e., varying less than threefold in transcript level over time). Thus the one generality that can be made is that expression of

almost all homeobox genes used in the process of embryogenesis is sharply regulated over time; these are not “housekeeping” genes.

In figure 1.4 the expression profiles are grouped by time of initial activation during development, as indicated by a significant rise in transcript level, whether there is significant or insignificant maternal representation. This analysis includes the already known homeobox genes as well as the new ones characterized in Figs. 1-3 (for time course expression data on previously known homeobox genes see references in legend of

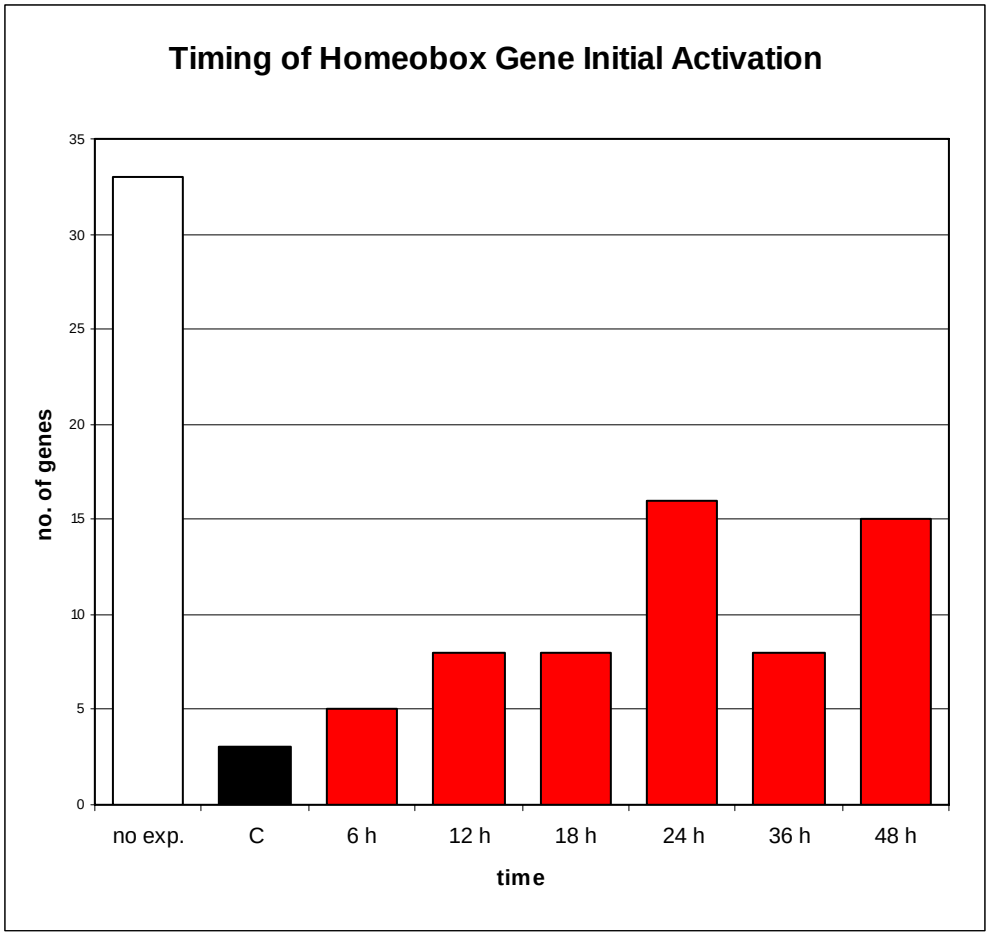


Figure 1.4. Distribution of homeobox gene initial activation times. Each bar represents the number of homeobox transcription factors that are initially activated at each time point. ‘C’ indicates constant expression; “no exp,” no expression by 48 h postfertilization.

fig.1.7). The analysis reveals that homeobox genes are activated at all stages of early development, with a twofold jump in the number of genes activated between 18 and 24 h. This period corresponds approximately to the time of PMC ingression and completion of ectodermal specification, and is just prior to gastrulation. In addition, there is a large subset of 31 genes (30%) which remain unexpressed at 48 h postfertilization, nine of which are *hox* cluster genes as demonstrated previously (Arenas-Mena et al, 2000).

The overall range of expression levels among the various homeobox genes is very broad. Half of all the genes were expressed at no more than 200 copies per embryo, and two-thirds of the homeobox genes had maximal expression below 1000 copies per embryo. However, the remaining third had peak expression ranging as high as 6400 mRNA molecules per embryo by 48 h postfertilization. The most highly expressed of the newly studied genes described in figure 1.3, with greater than 4000 copies per embryo, were *Sp-emx* (#150), *Sp-irxA* (#200), *Sp-hox7* (#255), *Sp-dlx* (#309), *Sp-atbf1* (#78). As with onset of expression, the level of peak expression during early embryogenesis shows no bias by homeodomain sub-family.

Kinetic parameters

Some of the gene expression time courses provided an opportunity to perform simple kinetic analyses. For example, high maternal expression steadily tapering to a very low level suggests that embryonic transcription of the gene is minimal. Fitting an exponential decay function to these data sets provides an estimate of the half-life of the maternal mRNA. Possible zygotic transcription of the gene could lead to a high estimate

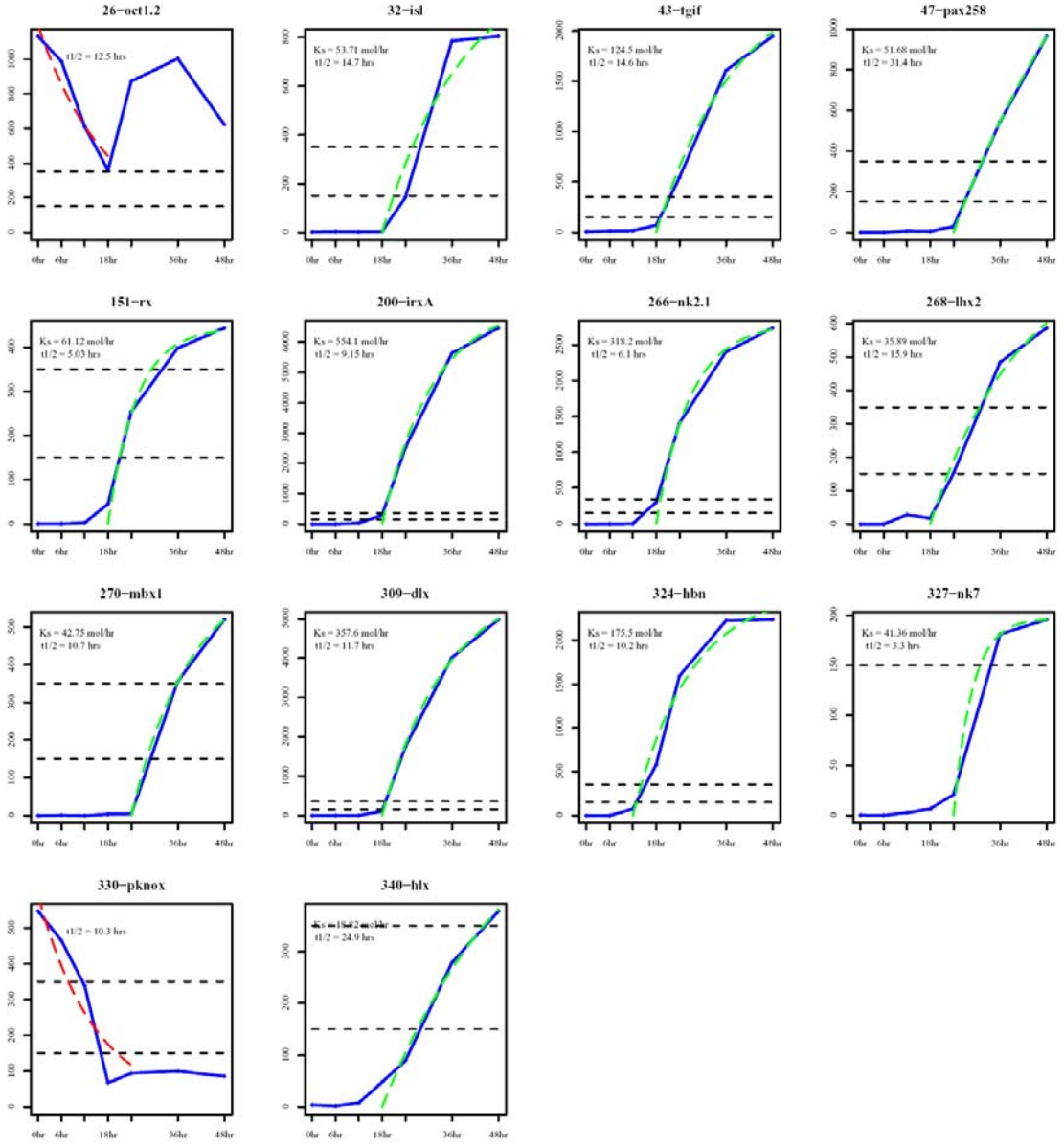


Figure 1.5. Maternal message decay and zygotic transcription and decay kinetics. Accumulation time courses in which maternal expression is followed by a continuous decline in transcript levels were used to estimate maternal mRNA decay rates. A dashed red line indicates the nonlinear least squares fit to the expression $y = C_0 * e^{-kd * t}$, where C_0 is initial concentration, k_d is the decay rate, and concentration y is a function of time t. The rate of decay is given as a half-life, where $t_{1/2} = \ln 2/k_d$. For accumulation time courses lacking maternal expression but displaying sustained and increasing zygotic expression, rates of both message synthesis and decay are calculated. A dashed green line indicates the nonlinear least squares fit to the expression $y = k_s/k_d * (1 - e^{-kd * t})$, where k_s is the rate of message synthesis in molecules per h for the whole embryo. Calculations and graphs were done using the mathematics and graphing program R.

of mRNA half-life in this calculation, but if maternal expression is relatively high, low levels of new transcription will have only a small impact on the calculated rate of decay. In the opposite case, if there is no maternal expression, and the gene is activated during embryogenesis and expressed at a constant rate, the transcript accumulation data will be fit by a simple synthesis and decay function (Davidson, 1986). In this case the rate of

ID	Gene	$t_{1/2}$ (h)	K_s (mol/h)
26	*Sp-oct1/2*	12.5	-
32	*Sp-isl*	14.7	53.7
43	*Sp-tgif*	14.6	124.5
47	*Sp-pax2/5/8*	31.4	51.7
151	*Sp-rx*	5.0	61.1
200	*Sp-irxA*	9.2	554.1
266	*Sp-nk2.1*	6.1	318.2
268	*Sp-lhx2*	15.9	35.9
270	*Sp-mbx1*	10.7	42.8
309	*Sp-dlx*	11.7	357.6
324	*Sp-hbn*	10.2	175.5
327	*Sp-nk7*	3.3	41.4
330	*Sp-pknox*	10.3	-
340	*Sp-hlx*	24.9	18.8
-	average	12.9	152.9

Table 1.2. Decay rates for maternal mRNA and synthesis and turnover rates for zygotic messages.

synthesis and the rate of turnover can both be estimated using a non-linear least squares regression.

Figure 1.5 and table 1.2 display results of analyses of transcript accumulation kinetics for 14 genes, which are adequately fit by one of these two simple models. As expected, there was much variation among the half-lives of different transcripts. *Sp-nk7* had the shortest half-life, at 3.30 h, while *Sp-pax2/5/8* had the longest, at 31.4 h. The average half-life of any message was 12.9 h, about twice the 5.7 hour average half-life for

total polysomal RNA transcript in the *S. purpuratus* blastula-gastrula embryo (Davidson, 1986). The rate of synthesis of different mRNAs was likewise quite variable, ranging from 36 molecules/h (*Sp-lhx2*) to 550 molecules/h (*Sp-irxA*). The average rate of synthesis was about 150 molecules/h per embryo. Much of this apparent variation in rate may of course simply reflect how many cells are expressing a given gene. Note however that even if only 10 cells were expressing the gene per embryo, all of these rates are far below the maximal possible rate of gene expression. This is about 660 molecules/h per cell for any given gene, assuming that both genomic copies are active (Davidson, 1986). As observed earlier, low rates of expression are typical for genes encoding transcription factors (Bolouri and Davidson, 2003).

Spatial patterns of gene expression

Whole mount *in situ* hybridization was used to determine the spatial expression patterns of the more highly expressed genes. Since this method optimally requires >10 copies of a transcript per cell to produce staining, probes were made only for transcripts expressed at 500 copies per embryo or more. Greater sensitivity can be achieved by individual experimental adjustments of the method, but this is incompatible with a screening procedure designed to interrogate many different genes. Furthermore, we focused on early development, selecting only genes expressed at that level by 24 h postfertilization (PMC ingression). For these genes observations were carried out to the 36 h late gastrula stage.

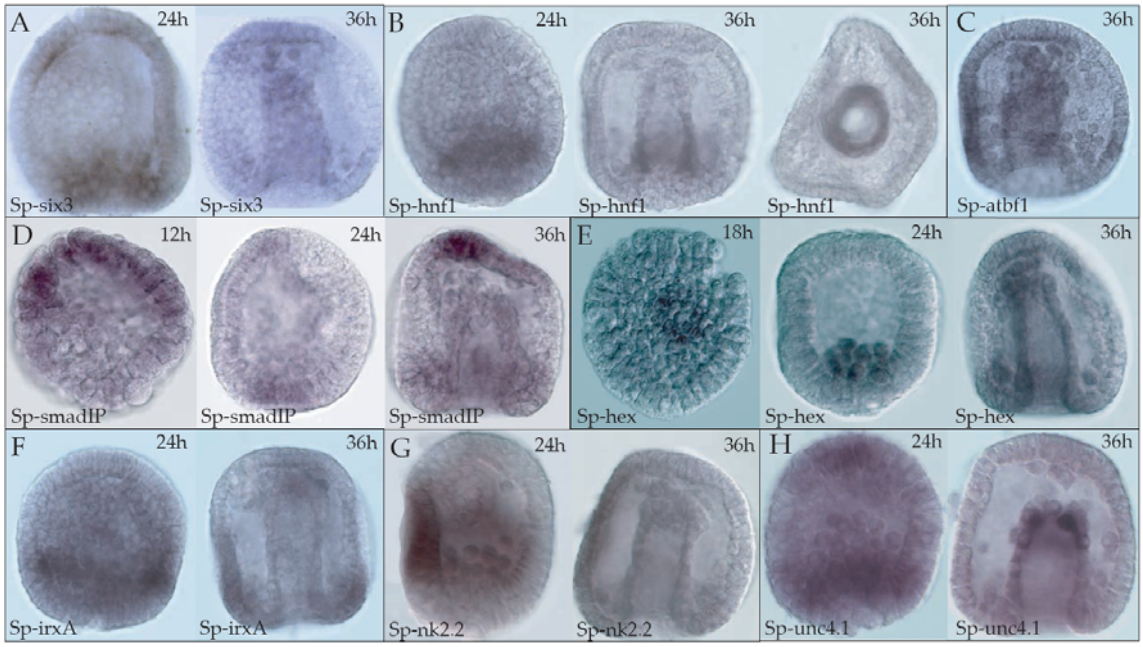


Figure 1.6. Spatial expression of homeobox genes. Panels A-H are whole mount *in situ* hybridizations of previously unstudied homeobox transcription factors which display localized expression patterns. The gene ID and name are displayed in the bottom left corner; the time post-fertilization is indicated in the upper right corner.

Our goal in probe design was to achieve increased throughput by avoiding laborious experimental approaches as much as possible. Thus, while long *in situ* probes are generally preferable, in some cases we used somewhat shorter probes in favor of assessing as many genes as possible. After some trial and error, we found that probes over 500 nucleotides long generally allowed us to detect expression, while shorter probes were insufficiently sensitive. A variety of methods were used to obtain sufficiently long probes given, for most cases, limited knowledge of coding regions. For a few genes, expression was high enough to obtain cDNA sequences by library screening with only the ~150 bp QPCR amplicon. In other cases, a single very long exon containing a recognizable conserved domain was sufficient to design a hybridization probe.

Alternately, positioning one primer in each of two conserved domains generated an adequate probe. Another resource we made use of was the extensive EST library created to aid in assembling the genome. We used Blast to identify ESTs matching our QPCR amplicons and made primers against the coding regions to generate long probes. As a last option, we used Genscan to predict exon positions on the relevant genome sequence contig. The predicted cDNA was then used to generate an *in situ* probe. In all cases, as with designing QPCR primers, we aligned genome sequencing traces to identify variable nucleotide positions and decrease the failure rate of our probe primers.

Expression of the new homeobox genes which are the focus of this study maps to all territories of the developing embryo. *Sp-dlx*, *Sp-pbx/exd, Sp-emx*, and *Sp-awh* are all expressed ubiquitously (data not shown). The localized expression patterns of the remaining genes we tested are shown in figure1.6. In figure1.6A, we see that the atypical homeodomain gene *Sp-six3* is restricted to the mesoderm before gastrulation. After gastrulation, expression continues in the SMCs and is also activated in the apical region. *Sp-hnf1* expression begins in the veg1 territory (fig.1.6B), but by 36 h localizes to the hindgut. Transcripts of *Sp-atbf*, shown in figure1.6C, are first visible in the oral ectoderm at 36 h. Figure 1.6D shows that *Sp-sip* is expressed early in what is most likely the animal half of the embryo. In agreement with the time course data, expression is very faint in the 18 h blastula, fading in the vegetal half but beginning to appear faintly apically. After gastrulation, the marked change in expression is complete, with *Sp-sip* activated only in the apical ectoderm. Figure 6E displays expression of *Sp-hex*, which though initially present at a low level, is expressed in the micromere descendants and the PMCs from 18 h on. Figure 1.6F shows veg_1 expression of *Sp-irxA* at the mesenchyme blastula stage.

After gastrulation, expression includes the SMCs and the oral ectoderm. Figure 1.6G shows that *Sp-nkx2.2* is expressed in the aboral ectoderm. Finally, *Sp-unc4.1* (fig.1.6H) is briefly expressed apically at 24 h. Expression also appears at that time in the endoderm, resolving to the foregut by 36 h. Homeobox genes are thus used to direct the embryo through many different aspects of development, in all territories of the embryo.

Discussion

The repertoire of homeobox genes in S. purpuratus

The newly sequenced *S. purpuratus* genome provides the opportunity of applying the criterion of completeness to our understanding of the transcriptional regulatory apparatus that controls development. Here we report a study of expression of all identifiable homeobox genes during embryogenesis, up to the late gastrula stage. The large majority of these genes had not previously been studied in the sea urchin (fig.1.1). Using Blast to identify sea urchin sequences with homology to known homeobox domains, we recovered 96 *S. purpuratus* homeobox genes, including members from the *paired*, extended *hox*, *nk*, atypical, *lim*, *dl*, *cut*, *pou* and other classes. Phylogenetic analysis of the homeodomain sequences of these genes, supplemented by a similar analysis of pax domains (fig.1.2), revealed that the sea urchin has a close ortholog of nearly every vertebrate homeobox gene or set of paralogs, reflecting their shared deuterostome heritage. In a very few cases, the closest homologue was a *Drosophila* gene, indicating possible vertebrate specific deletions. Thus, as has been seen again and

again in other species, there is a remarkable conservation of the homeobox gene regulatory tool kit across the bilaterians.

Homeobox gene utilization in embryonic development

As described in the results section, embryonic expression of each of the newly analyzed sea urchin homeobox genes was determined by QPCR, and for those adequately expressed, whole mount *in situ* hybridization was used to determine where in the embryo each of these transcription factors is transcribed during development. Figure 1.7 summarizes both the temporal and spatial expression data for the homeodomain transcription factor family of *S. purpuratus* up to the late gastrula stage, grouped by class. The time of initial embryonic expression, including whether the gene is maternally expressed, is color coded in the Category ("Cat.") column. Spatial expression data between 7 and 36 h postfertilization is presented in the following five columns. Grayed out areas indicate that expression was too low to attempt *in situ* hybridization. Results given in blue are new information reported here, whereas data in red are summarized from other published work.

The homeobox gene family is heavily utilized in the processes of early development. Even before fertilization, 13% of homeobox genes are already represented in the egg as maternal transcripts. By 48 h, at only the late gastrula stage, 65% of all homeobox genes in the genome have been already been activated. Homeobox genes are brought into action continuously during early development, as the regulatory state of the embryo increases steadily in complexity. However, we noted a surge in the rate of new

Key:				
	no in situ	6 h	24 h	no/low
	new in situ	12 h	36 h	maternal
	known in situ	18 h	48 h	constant

Homeodomain Transcription Factors (total = 94)

Index	Gene Name	Cat.		7h	12h	18h	24h	36h	E-value
	paired class (31)								
-	Sp-alx1			-	PMC	PMC	PMC	-	-
-	Sp-gsc			-	-	-	PMC, OE	OE	-
-	Sp-pmar1			micromeres		-	-	-	-
-	Sp-otx			-	-	-	veg., OE	gut	-
16	Sp-pax1/9								5e-80
47	Sp-pax2/5/8								1e-30
296	Sp-pax6								3e-63
273	Sp-paxA								4e-31
274	Sp-paxB								1e-26
108	Sp-paxC								8e-36
321	Sp-eyg								1e-14
393	Sp-eygl								4e-18
394	Sp-pax4l								9e-11
146	Sp-chx10								2e-23
151	Sp-rx								4e-21
163	Sp-pitx1								6e-19
275	Sp-pitx2								6e-16
84	Sp-pitx3								1e-41
184	Sp-alx4								3e-20
269	Sp-phox2								3e-16
270	Sp-mbx1								6e-24
272	Sp-otp								2e-19
297	Sp-arx								3e-35
298	Sp-arxl								2e-17
389	Sp-arxl2								2e-12
310	Sp-shox								4e-17
311	Sp-prx								3e-16
324	Sp-hbn			-	-	apical	apical	apical	9e-15
334	Sp-unc4.1			-	-	-	E	foregut	9e-16
392	Sp-phb1								5e-13
396	Sp-phb2								6e-12
	hox / lim (6)								
-	Sp-lim1			-	-	veg1	veg1	veg1	4e-30
32	Sp-isl								6e-35
104	Sp-lhx3								1e-24
122	Sp-awh			ubiq	ubiq	ubiq	ubiq	ubiq	2e-25
268	Sp-lhx2			AO	AO	AO+ apical	AO+ apical	AO+ apical	2e-22
314	Sp-lmx1								4e-16
	atypical class (11)								
2	Sp-six3			-	-	-	M	M, api	2e-61
15	Sp-six1								1e-66
21	Sp-six4								1e-49
42	Sp-pbx/exd			ubiq	ubiq	ubiq	ubiq	ubiq	9e-29
43	Sp-tgif			-	-	PMC	PMC	bpore, SMC	2e-28
56	Sp-hnf1			-	-	-	veg1	hindgut	2e-25
130	Sp-meis								7e-22
200	Sp-irxA			-	-	-	veg1	veg1	3e-18
299	Sp-irxB								5e-31
330	Sp-pknox								2e-16
343	Sp-prox1			-	-	-	-	-	9e-13

Index	Gene Name	Cat.		7h	12h	18h	24h	36h	E-value
	extended hox (21)								
-	Sp-hox1								4e-29
-	Sp-hox2								8e-22
-	Sp-hox3								2e-23
-	Sp-hox4/5								1e-43
-	Sp-hox6								4e-40
-	Sp-hox7			-	-	-	ubiq	gut,OE,SMC	7e-33
-	Sp-hox8								2e-35
-	Sp-hox9/10								3e-41
-	Sp-hox11/13a								5e-19
-	Sp-hox11/13b			-	EM	EM	M	hindgut	4e-21
-	Sp-hox11/13c								1e-07
-	Sp-eve			EM	veg1	veg1	veg1	ABO	2e-19
12	Sp-en								2e-61
40	Sp-xlox								5e-31
109	Sp-mox								9e-18
150	Sp-emx			-	-	ubiq	ubiq	ubiq	3e-22
258	Sp-hb9								4e-17
300	Sp-cdx2								4e-18
317	Sp-gsh1								1e-15
606	Sp-rough								1e-22
610	Sp-gbx								1e-21
	nk class (12)								
-	Sp-hmx			-	-	ectoderm	ectoderm	ectoderm	-
-	Sp-nk1			-	-	-	-	veg1 OE	5e-37
-	Sp-nk2.1			-	-	apical	apical	apical	3e-56
14	Sp-nk2.5								2e-35
75	Sp-nk2.2			-	-	-	OE	OE	2e-38
115	Sp-lbx								1e-28
127	Sp-nk6.1								5e-22
261	Sp-dbx1								1e-23
263	Sp-hex			-	-	micromere	PMC	PMC, gut	2e-26
267	Sp-nk3.2								2e-20
327	Sp-nk7								2e-14
340	Sp-hlx								9e-12
	dl, cut, pou, barx, zinc finger (15)								
-	Sp-hnf6			ubiq	ubiq	ubiq	ubiq	cil band	-
-	Sp-brn1/2/4			-	-	-	-	gut	-
-	Sp-msx			-	-	-	gut, SMC	OE, gut	4e-11
395	Sp-msxl								2e-11
18	Sp-brn3								1e-69
26	Sp-oct1/2			ubiq	-	-	-	-	4e-49
78	Sp-atbf1			-	-	-	-	OE	1e-79
81	Sp-sip			EM	EM	EM	EM	api, SMC	2e-14
94	Sp-cux2								3e-22
259	Sp-barh1								2e-23
260	Sp-barx								8e-16
309	Sp-dlx			-	-	ectoderm, rt.	ectoderm, rt.	ectoderm, rt.	2e-17
331	Sp-cutl								2e-13
388	Sp-lass6								3e-15
618	Sp-pou6								2e-49

Figure 1.7. Spatial and temporal expression of sea urchin homeobox transcription factors. A summary of the expression data for each of the 94 identified homeobox transcription factors. All transcription factors uncovered by our search algorithm and for which QPCR was done were assigned a working ID number (index). Genes previously published have no index number. Newly identified proteins were named according to the closest known homologue, as identified by our phylogenetic tree. The third and fourth columns relate whether the gene is maternally expressed (>400 copies/egg; indicated by tan box) and by what time point (6 h = red; 12 h = orange; 18 h = yellow; 24 h = green; 36 h = blue; 48 h = violet; white = not before 48 h) expression rises to within the minimum range estimated to be significant (150-350 copies/embryo). A black box indicates constant expression varying by less than twofold over the time period studied. Next is given the result of *in situ* staining, if done. Results written in blue are new findings; information in red is cited from previously published work. Finally, the "Eval" column gives the e-value of the top blastx match between the identified gene fragment and nr. Expression data for the following genes was acquired from the sources noted: *Sp-hmx* (Martinez and Davidson, 1997); *Sp-nk1* (Otim et al., 2004) ; *Sp-nk2.1* (Takacs et al., 2004) ; *Sp-alx1* (Ettensohn et al., 2003) ; *Sp-gsc* (Angerer et al., 2001); *Sp-pmar1* (Oliveri et al., 2003); *Sp-otx* (Gan et al., 1995); *Sp-lim1*(P. Oliveri and E. Davidson, unpublished data) ; *hox* genes (Arenas-Mena et al., 1998); *Sp-hox7* (Dobias et al., 1996) ; *Sp-eve* (Davidson et al., 2002); *Sp-hnf6* (Otim et al., 2004); *Sp-brn1/2/4* (Yuh et al., 2005); and *Sp-msx* (Dobias et al., 1997).

homeobox gene activations a surge between 18 and 24 h, just before gastrulation, a time when territorial specification processes are achieving completion throughout the embryo (Davidson, 2006; Ransick and Davidson, 1998).

In a number of cases, the gene expression time courses fit simple kinetic models requiring constant rates of mRNA synthesis and decay. These rates pertain over periods of 24-36 h, and we might infer that during these extended periods of time the regulatory inputs into the control systems of these genes are unchanging, and that the genes are performing one specific regulatory task. Also, as noted above, none of the homeobox genes for which kinetics were obtained appear to be expressed at more than a few percent of the maximum possible transcription rates. While spatial expression data are available for only a few of these genes, *Sp-nk2.1* is present apically (Takacs et al., 2004) in about

50 cells, giving a cellular synthesis rate of perhaps six molecules/ h per cell; i.e., each gene is transcribed only about once every 20 min. *Sp-irxA* is expressed in a narrow band of about 150 cells, yielding a cellular synthesis rate of 3.7 copies/ h per cell. A previous study on the homeobox gene brn1/2/4 (Yuh et al., 2005) yielded very similar data. During its maximum stage of expression in the embryonic midgut, for many hours this essential gene produces only two molecules of mRNA per cell-h, and similarly to those homeobox gene transcripts measured here, its transcripts turn over with the relatively long half-life of about 14 h. For these genes then, the pattern is slow synthesis, with transcripts steadily accumulating because of unusually slow turnover. The reason this suffices for genes encoding transcription factors is that the rates of translation are sufficient to enable the requisite number of transcription factor molecules, about one thousand to a few thousand per cell, to be produced over a period of several hours from a very modest number of mRNAs (Bolouri and Davidson, 2003).

New players in specific embryonic specification processes

Figure 1.7 also summarizes the spatial expression data we were able to obtain for a set of the more highly expressed homeobox genes. Including previously reported data, homeobox gene expression is seen in all embryonic territories, with no family or class bias as to specific domains. The newly reported gene expression patterns shown in figure 1.6 identify genes that may participate in the gene regulatory networks that underlie specification of neurogenic apical plate, oral and aboral ectoderm, primary mesenchyme cells, and endomesoderm.

Sp-hex is the only newly identified gene likely to execute a specific function in the PMC regulatory network. This gene is activated very early, with visible staining of the micromere descendants by 18 h. In other systems the Hex factor may function either as an activator or repressor of transcription, achieving the strongest repression when binding with a corepressor (Kasamatsu et al., 2004; Swingler, 2004). In vertebrates, Hex is involved in hematopoietic specification and differentiation, and in the formation of endoderm derived organs (Guo, 2003).

Sp-six3 is expressed in the veg2 mesoderm before gastrulation and in the apical ectoderm. After gastrulation, expression continues in SMCs delaminating from the tip of the archenteron. Given the lack of connection between the apical and mesodermal domains at 24 h, it seems probable *Sp-six3* functions in at least two distinct regulatory networks. The Six3 factor has a well documented role in the eye specification network (Donner and Maas, 2004; Gehring, 2005), and is also required for vertebrate forebrain specification (Ando et al., 2005). In both settings, it binds a member of the Groucho family and acts as a repressor to define the boundaries of an embryonic territory. It is possible it plays a similar role in the sea urchin mesoderm and apical ectoderm, as Groucho is known to be present in all nuclei throughout sea urchin development (Range et al., 2005).

The *in situ* hybridization experiments also identified several homeobox genes which are likely to belong to the oral/aboral specification network. Here we report the ectodermal expression of *Sp-sip1*, a known repressor and cofactor of the Smad transcription factors, which are activated by BMP signaling (Verschueren et al., 1999). This gene is expressed in the animal half of the embryo by 12 h, and is especially

concentrated in one-half of that domain. *Sp-sip1* could act to modulate the role of an activated Smad factor, turning it into a repressor of oral genes in the aboral ectoderm or vice versa. Curiously, *Sp-sip1* expression drops sharply before PMC ingression, and reappears later apically. *Sp-nkx2.2* and *Sp-atbf1* are also expressed in the oral ectoderm, by 24 and 36 h, respectively.

Three other homeobox genes, *Sp-hnf1*, *Sp-irxA*, and *Sp-unc4.1*, may be assigned to the endomesodermal GRN. *Sp-unc4.1* and *Sp-hnf1* may be involved in both endoderm specification and the deployment of endoderm differentiation genes, as by 36 h their expression is strictly limited to the foregut and hindgut, respectively. *Sp-irxA*, on the other hand, is in the veg_1 territory at 24 h, but not in the portion which forms endoderm. At 36 h, it is expressed in a ring of cells just beyond the blastopore.

In summary, these homeobox gene expression data provide an image of the overall utilization of one of the most prominent of classes of bilaterian regulatory genes in one of the best known of bilaterian embryos. They furthermore identify several probable new components of the gene regulatory networks that control the development of this embryo, a timely addition, as these networks are now rapidly being unraveled.

Acknowledgements

The authors would like to acknowledge Rachel Gray for her assistance in conducting the whole mount *in situ* hybridizations reported in the paper, and Deanna Thomas for her invaluable help with figures. This research was supported by NIH grant HD37105 and by the Office of Science (BER), US Department of Energy, grant DE-FG02-03ER63584.

References

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990. Basic local alignment search tool. J. Mol. Biol. 215**,** 403-410.

Ando, H., Kobayashi, M., Tsubokawa, T., Uyemura, K., Furuta, T., Okamoto, H., 2005. Lhx2 mediates the activity of Six3 in zebrafish forebrain growth. Dev. Biol. 287**,** 456-468.

Angerer, L. M., Oleksyn, D. W., Levine, A. M., Li, X. T., Klein, W. H., Angerer, R. C., 2001. Sea urchin *goosecoid* function links fate specification along the animal-vegetal and oral-aboral embryonic axes. Development 128**,** 4393-4404.

Arenas-Mena, C., Cameron, A. R., Davidson, E. H., 2000. Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. Development 127**,** 4631-4643.

Arenas-Mena, C., Martinez, P., Cameron, R. A., Davidson, E. H., 1998. Expression of the Hox gene complex in the indirect development of a sea urchin. Proc. Natl. Acad. Sci. USA 95**,** 13062-13067.

Arnone, M. I., Rizzo, F., Annunciata, R., Cameron, R. A., Peterson, K. J., 2006. Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: insights into the relationship between clustering and colinearity. Dev. Biol. in press.

Bolouri, H., Davidson, E. H., 2003. Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. Proc. Natl. Acad. Sci. USA 100**,** 9371-9376.

Cameron, A. R., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P. M., Davidson, E. H., Peterson, K. J., Hood, L., 2006. Unusual gene order and organization of the sea urchin Hox cluster. J. Exp. Zool. 306B**,** 45-58.

Char, B. R., Bell, J. R., Dovala, J., Coffman, J. A., Harrington, M. G., Becerra, J. C., Davidson, E. H., Calzone, F. J., Maxson, R., 1993. *Spoct*, a gene encoding the major octamer-binding protein in sea urchin embryos - expression profile, evolutionary relationships, and DNA-binding of expressed protein. Dev. Biol. 158**,** 350-363.

Czerny, T., Bouchard, M., Kozmik, Z., Busslinger, M., 1997. The characterization of novel Pax genes of the sea urchin and *Drosophila* reveal an ancient evolutionary origin of the Pax2/5/8 subfamily. Mech. Dev. 67**,** 179-192.

Davidson, E. H. 1986. Gene Activity in Early Development. Academic Press, Inc., Orlando.

Davidson, E. H. 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, San Diego.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L.,

Bolouri, H., 2002. A genomic regulatory network for development. Science 295**,** 1669-1678.

Dobias, S. L., Ma, L., Wu, H. L., Bell, J. R., Maxson, R., 1997. The evolution of *msx* gene function: Expression and regulation of a sea urchin *msx* class homeobox gene. Mech. Dev. 61**,** 37-48.

Dobias, S. L., Zhao, A. Z. J., Tan, H. Y., Bell, J. R., Maxson, R., 1996. *Sphbox7*, a new Abd-B class homeobox gene from the sea urchin *Strongylocentrotus purpuratus*: Insights into the evolution of Hox gene expression and function. Developmental Dynamics 207**,** 450-460.

Donner, A. L., Maas, R. L., 2004. Conservation and non-conservation of genetic pathways in eye specification. Int. J. Dev. Biol. 48**,** 743-753.

Ettensohn, C. A., Illies, M. R., Oliveri, P., De Jong, D. L., 2003. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. Development 130**,** 2917-2928.

Gan, L., Mao, C.-A., Wikramanayake, A., Angerer, L. M., Angerer, R. C., Klein, W. H., 1995. An Orthodenticle-related protein from *Strongylocentrotus purpuratus*. Dev. Biol. 167**,** 517-528.

Gehring, W. J., 2005. New perspectives on eye development and the evolution of eyes and photoreceptors. J Hered 96**,** 171-184.

Guo, Y., 2003. The homeoprotein Hex is required for hemangioblast differentiation. Blood 102**,** 2428-2435.

Higgens, D. G., Thompson, J. D., Gibson, T. J., 1994. CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22**,** 4673-4680.

Howard-Ashby, M., Materna, S., Brown, C. T., Chen, L., Cameron, A. R., Davidson, E. H., 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev. Biol. in press.

Kasamatsu, S., Sato, A., Yamamoto, T., Keng, V. W., Yoshida, H., Yamazaki, Y., Shimoda, M., Miyazaki, J.-i., Noguchi, T., 2004. Identification of the transactivating region of the homeodomain protein, Hex. J Biochem (Tokyo) 135**,** 217-223.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5**,** 150-163.

Martinez, P., Davidson, E. H., 1997. *Sphmx*, a sea urchin homeobox gene expressed in embryonic pigment cells. Dev. Biol. 181**,** 213-222.

Martinez, P., Rast, J. P., Arenas-Mena, C., Davidson, E. H., 1999. Organization of an echinoderm Hox gene cluster. Proc. Natl. Acad. Sci. USA 96**,** 1469-1474.

Materna, S., Howard-Ashby, M., Gray, R., Davidson, E. H., 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev. Biol. in press.

Nam, J., Nei, M., 2005. Evolutionary change of the numbers of homeobox genes in bilateral animals. Mol. Biol. Evol. 22**,** 1-9.

Nemer, M., Rondinelli, E., Infante, D., Infante, A. A., 1991. Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. Dev. Biol. 145**,** 255-265.

Oliveri, P., Davidson, E. H., McClay, D. R., 2003. Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. Dev. Biol. 258**,** 32-43.

Otim, O., Amore, G., Minokawa, T., McClay, D. R., Davidson, E. H., 2004. SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. Dev. Biol. 273**,** 226-243.

Range, R. C., Venuti, J. M., McClay, D. R., 2005. LvGroucho and nuclear [beta]-catenin functionally compete for Tcf binding to influence activation of the endomesoderm gene regulatory network in the sea urchin embryo. Dev. Biol. 279**,** 252-267.

Ransick, A., Davidson, E. H., 1998. Late specification of veg(1) lineages to endodermal fate in the sea urchin embryo. Dev. Biol. 195**,** 38-48.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C., Davidson, E. H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246**,** 132-147.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A., Arnone, M. I., 2006. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev. Biol. in press.

Rozen, S., Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. *In* "Bioinformatics Methods and Protocols: Methods in Molecular Biology." (S. Krawetz and S. Misener, Eds.), pp. 365-386. Humana Press, Totowa, NJ.

Samanta, M. P., Tongprasit, W., Istrail, S., Cameron, A. R., Tu, Q., Davidson, E. H., Stolc, V., 2006. High resolution transcriptome map of the sea urchin embryo. Dev. Biol. In press.

Swingler, T., 2004. The proline-rich homeodomain protein recruits members of the Groucho/transducin-like enhancer of split protein family to co-repress transcription in hematopoietic cells. J. Biol. Chem. 279**,** 34938-34947.

Takacs, C. M., Amore, G., Oliveri, P., Poustka, A. J., Wang, D., Burke, R. D., Peterson, K. J., 2004. Expression of an *NK2* homeodomain gene in the apical ectoderm defines a new territory in the early sea urchin embryo. Dev. Biol. 269**,** 152-164.

Tu, Q., Brown, C. T., Davidson, E. H., Oliveri, P., 2006. Sea Urchin Forkhead Gene Family: Phylogeny and Embryonic Expression. Dev. Biol. in press.

Verschueren, K., Remacle, J. E., Collart, C., Kraft, H., Baker, B. S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M.-T., Bodmer, R., Smith, J. C., Huylebroeck, D., 1999. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. J. Biol. Chem. 274**,** 20489-20498.

Yuh, C. H., Dorman, E. R., Davidson, E. H., 2005. Brn1/2/4, the predicted midgut regulator of the *endo16* gene of the sea urchin embryo. Dev. Biol. 281**,** 286-298.

Chapter 2

Gene Families Encoding Transcription Factors Expressed in Early Development of *Strongylocentrotus purpuratus*

Meredith Howard-Ashby, Stefan C. Materna, C. Titus Brown, Lili Chen, R. Andrew Cameron, Eric H. Davidson

In press, Developmental Biology

Abstract

All genes encoding transcription factors of the bHLH, Nuclear Receptor, Basic Leucine Zipper, T-box, Smad, Sox, and other smaller families were identified in the *Strongylocentrotus purpuratus* genome by means of a permissive blast search of the genome using a database of known transcription factors. Phylogenetic trees were constructed for the major families, permitting a comparison of the regulatory protein repertoire of the sea urchin and other species. QPCR and whole mount *in situ* hybridization experiments revealed the temporal and spatial expression patterns of these genes during early development. These regulatory genes are initially expressed at a broad range of time points, and the large majority of genes of all families are expressed within the first 48 hours of development. The observations suggest assignment of many

regulatory genes to specific developmental sub-networks, including endomesodermal, oral, aboral and apical.

Keywords: transcription factor, sea urchin, development, bHLH, sox, nuclear receptor, basic zipper, smad

Introduction

Genes encoding transcription factors are the key players in the regulatory networks that specify embryonic territories during development. Developmental regulatory states are set up in the domains of the embryo as differential activation of these genes generates diverse spatial patterns of expression. Knowledge of the specific times and places of transcription factor expression is necessary for experimental solution of regulatory networks, and thereby for understanding how regulatory genes interact to direct the process of development. The *Strongylocentrotus purpuratus* genome sequence has enabled the systematic identification of all the players in embryonic gene regulatory networks. We have made an effort to identify all the sequence specific DNA binding proteins encoded in the sea urchin genome and to establish both temporal and, when possible, spatial patterns of expression throughout early embryogenesis.

Transcription factors are grouped into many families, according to the structure of their DNA binding domains. Genes belonging to the two largest such families, C2H2 Zinc Finger factors and Homeodomain factors, are discussed in companion articles in this volume (Howard-Ashby et al., 2006; Materna et al., 2006). Ets and Forkhead family genes are treated separately in additional dedicated articles (Rizzo et al., 2006; Tu et al., 2006). Here we turn our attention to the remaining families of genes encoding transcription factors, including bHLH, Nuclear Receptor, Basic Leucine Zipper, T-Box, Smad, and Sox factors, as well as other smaller families.

Materials and methods

Identification of transcription factor genes

Sequences encoding transcription factors were located in the both the unassembled genome sequencing reads and the November, 2004 Baylor University Human Genome Sequencing Center draft genome assembly using a tblastn search with a set of reference regulatory proteins from nr and GO seqdblite. For a full description of the method, see materials and methods in chapter 1. Following the blast identification procedure, the gene set was compared to the Baylor HGSC Glean3 gene predictions, and any missed genes were added to our set.

Phylogenetic analysis

For the larger gene families, in which gene identification was not trivial, phylogenetic trees were constructed. All trees included related sequences from human and *Drosophila melanogaster*, and the nuclear receptor tree analysis also includes sequences from *Ciona intestinalis*. The DNA binding domains of each family were aligned manually and Mega 3 (Kumar et al., 2004) was used to generate the trees. For the nuclear receptor family, both the zinc finger and ligand binding domains were concatenated and aligned. The neighbor joining method with 1000 bootstrap replications was used to calculate each tree. For all the families, an initial tree was made to assign sea

urchin genes to specific subfamilies, before calculating separate trees for each of the subfamilies.

QPCR data

Quantitative PCR was used to determine the expression profile of each identified transcription factor during development, from unfertilized egg to 48 h. Observations were made in triplicate. A complete description of both primer selection and the QPCR methodologies can be found in (Howard-Ashby et al., 2006).

The time course data were plotted on a logarithmic scale to simplify comparison of expression profiles of very different magnitudes. The results were compared to data generated by the genome tiling array transcriptome analysis, and found to be in strong agreement (Samanta et al., 2006). Each gene was categorized as to whether expression was maternal only, maternal and zygotic, zygotic only, constant, or null up to 48 h. Genes expressed zygotically were further categorized as to the time by which expression is first activated. Expression between 150 and 250 copies per embryo, or >500 copies in the egg, was deemed to be biologically relevant. Complete time course data and primer sequences can be found online at http://sugp.caltech.edu/supplement/meredith/index.html or in the appendices.

Whole mount in situ *hybridization*

In situ probes were designed for genes with significant zygotic expression by 24 h post-fertilization. We attempted to use probes at least 600 bp long, though some results with shorter probes are reported if they gave a positive, specific result. Coding sequence suitable for making probes was found in a number of ways. In some cases, our original blast searched uncovered a single sufficiently long exon. Alternately, two known conserved domains were bridged to provide a suitable length probe. In other cases, blastn of the whole contig against sea urchin EST and cDNA libraries submitted to NCBI revealed the location of coding sequence. Finally, genscan gene predictions were used to develop probes in the absence of any other evidence of gene structure.

For experimental details of probe construction, and the *in situ* method, see the materials and methods section in chapter 1. The primers used to make probes can be found at http://sugp.caltech.edu/supplement/meredith/index.html or in the appendices.

Results

Identification of previously unknown sea urchin regulatory genes

As described in the companion article on sea urchin homeobox genes (Howard-Ashby et al, 2006), our strategy was to search the genome comprehensively for transcription factors by taking advantage of the sequence conservation among the DNA binding domains of these proteins. A reference database, which we termed our ‘rake,’

was assembled by including the pertinent GO-seqdblite databases as well as human, fly and mouse regulatory proteins from NCBI nr as described in methods and Howard-Ashby et al (2006). Using a permissive tblastn search with our rake of both the unassembled genomic sequencing reads and the draft assembly, we identified sea urchin sequences with apparent homology to genes encoding known transcription factors. The reverse procedure, blastx of these sequences against the rake proteins, effectively sorted the candidate sequences into gene families. Sequences with a blast e-value greater than 1e-12 against the rake proteins were discarded. With the sequences sorted into families it was possible to remove any redundancies and match up sequences belonging to the same genes but corresponding to different conserved domains, or stretched across multiple genomic reads or assembly contigs. Sequence pairings were confirmed by PCR against pooled sea urchin mRNA from multiple embryonic time points and checked against the most recently assembled scaffolds. The remaining genes were tentatively named based on the best match by blastp against the nr database.

The number of sea urchin genes from several major transcription factor families is shown in table 2.1, alongside counts from the *Drosophila melanogaster*, human, *Ciona intestinalis* (Imai et al., 2004), the cnidarian *N. vectensis* (www.stellabase.org), and *C. elegans* (Reece-Hoyes et al., 2005) genomes. In general, the number of sea urchin genes in each family is comparable to the number found in the fly genome and on the order of half those found in the human genome. A similar result was obtained in our analysis of homeobox genes (see chapter 1). One exception is the basic zipper (bzip) family, which includes fewer sea urchin genes than expected, given the number of genes in other sea urchin transcription factor families. Our search method might have been less successful

	Sea urchin	Fly	Human	*C. elegans*	*Ciona*	*N. vectensis*
bzip	13	27	53	32	26	6
smad	4	4	8	7	5	1
sox/hmg	11	12	26	16	21	14
bHLH	47	56	125	42	44	28
NR	33	21	48	274	18	8
t-box	6	8	17	21	8	9

Table 2.1. Number of genes in six transcription factor gene families.

in uncovering bzip genes if sea urchin bzip genes are more divergent than sea urchin genes of other families. Alternatively, many genes from the bzip family could have been lost in this lineage.

Phylogenetic analyses

For the *sox/hmg*, *smad*, *bHLH*, and *nuclear receptor* gene families, assignment of individual orthologs based solely on blast results was ambiguous, and phylogenetic trees were constructed to aid in the systematic naming of the novel sea urchin genes. Trees were constructed by aligning the conserved domains of each family manually and using the neighbor joining method with 1,000 bootstrap replications. The neighbor joining method was chosen as it has been shown to give as accurate trees as other methods in cases where there are many sequences, but the sequence lengths are short, and it has the advantage of being less computationally intensive. A summary of the gene identifications is given in table 2.2, showing the common name, the index number used for the study, and the corresponding gene model number or scaffold number assigned by HGSC at Baylor university.

Gene Name	Index	Glean ID
BHLH		
Sp-Acsc	244	SPU_28148
Sp-Acsc3	387	SPU_22554
Sp-Ahr	226	SPU_05022
Sp-Ap4	336	SPU_03179
Sp-Arnt	209	SPU_00129
Sp-AtoL1	375	SPU_00990
Sp-AtoL2	376	SPU_03681
Sp-Beta3	51	SPU_04028
Sp-bhlhB1	379	SPU_07253
Sp-Bmal	349	SPU_27935
Sp-Clock	188	SPU_17407
Sp-Coe	607	SPU_04702
Sp-E12	52	SPU_16343
Sp-Hand	136	SPU_17287
Sp-Hath6	119	SPU_11315,17983
Sp-Hes (known)	-	SPU_06814
Sp-HesB	377	SPU_06813
Sp-HesC	617	SPU_21608
Sp-Hey	301	SPU_09465
Sp-Hey4	378	SPU_15712
Sp-Hifa1a	197	SPU_01262, C-term
Sp-Id	384	SPU_15374
Sp-Mad	364	SPU_06583
Sp-Max	365	SPU_22163
Sp-Mist	242	SPU_19444,27623
Sp-Mitf	609	SPU_08175
Sp-Mlx	348	SPU_05787
Sp-MlxIPL	380	SPU_08845
Sp-Mnt	386	SPU_26205
Sp-Myc	303	SPU_03166
Sp-MyoD	128	SPU_21119
Sp-MyoD2	129	SPU_06232
Sp-MyoR2	120	SPU_12008
Sp-MyoR3	160	SPU_16445
Sp-Nato3	77	SPU_14401
Sp-NeuroD	6	SPU_24918
Sp-Ngn	49	SPU_07147
Sp-NSCL	381	SPU_09231
Sp-NXF	382	SPU_09413
Sp-Olig3	241	SPU_02627
Sp-Par	137	SPU_16650
Sp-Ptf1a	54	SPU_02677
Sp-Sage	374	SPU_13119, 02448
Sp-Scl	243	SPU_28093
Sp-Sim	605	SPU_13962
Sp-Trh	204	SPU_14249
Sp-Usf	182	SPU_14332

Gene Name	Index	Glean ID
Sox-hmg		
Sp-cic	335	SPU_25292
Sp-bbx	205	SPU_23037
Sp-lef1	251	SPU_03704
Sp-soxB1	249	SPU_22820
Sp-soxB2	198	SPU_25113
Sp-soxC	55	SPU_02603
Sp-soxD	250	SPU_04217
Sp-soxE	46	SPU_16881
Sp-soxF	320	SPU_14170
Sp-soxH	224	SPU_11080
Nuclear Receptor		
Sp-coupTF	-	SPU_23867
Sp-dsf	235	SPU_24486
Sp-e78a	366	SPU_03547,03548
Sp-e78b	338	SPU_18366
Sp-err	367	SPU_04723
Sp-fax1	133	SPU_12586
Sp-fxr	233	SPU_11348,27598
Sp-gcnf	239	SPU_00749
Sp-grf	124	SPU_13305
Sp-hnf4	36	SPU_21192
Sp-nr1AB	368	SPU_28255
Sp-nr1H6a	360	SPU_17404
Sp-nr1H6b	144	SPU_15456
Sp-nr1H6c	143	SPU_04526
Sp-nr1M1	369	SPU_17491
Sp-nr1M2	252	SPU_11576
Sp-nr1M3	175	SPU_13178
Sp-nr1M4	370	SPU_18845
Sp-nr2C	234	SPU_13134
Sp-nr2E6	237	SPU_17375
Sp-nr5A	159	SPU_13843
Sp-nr5B	238	Scaff7192
Sp-nurr1	172	SPU_00255
Sp-pnr	236	SPU_14405
Sp-ppar1	371	SPU_19332
Sp-ppar2	372	SPU_21289
Sp-rar	174	SPU_16523
Sp-reverb	232	SPU_17492
Sp-ror	373	SPU_22678
Sp-rxr	35	SPU_28422
Sp-shr2/tr2.4	155	SPU_08117
Sp-thr	357	SPU_18861,25239
Sp-tll	132	SPU_08936,27487

Other		
Sp-af9	147	SPU_06808
Sp-ap2 (AP2)	154	SPU_16685
Sp-ash1 (trxG)	48	SPU_25482
Sp-ash2 (trxG)	214	SPU_18423
Sp-cp2 (CP2)	316	SPU_14836
Sp-dach (Ski-Sno)	27	SPU_28061
Sp-dmtf (myb)	329	SPU_26633
Sp-dp1 (E2F)	318	SPU_06312
Sp-dri (bright)	-	SPU_05718
Sp-e2f3 (E2F)	123	SPU_06753
Sp-e2f4 (E2F)	339	SPU_28827
Sp-enz1 (pcg)	92	SPU_23366
Sp-enz2 (pcg)	166	SPU_27446, 23366
Sp-fhl2 (lim)	277	SPU_07981
Sp-gataC	-	SPU_27015
Sp-gataE	-	SPU_10635
Sp-gcm (gcm)	-	SPU_06462
Sp-gro	69	SPU_18692
Sp-irf1 (IRF)	307	SPU_10404
Sp-irf4 (IRF)	347	SPU_26877
Sp-ldb2 (lim)	295	SPU_26962
Sp-lmo2 (lim)	312	SPU_13569
Sp-lmo4 (lim)	95	SPU_19586
Sp-mbt1 (pcg)	135	SPU_21123
Sp-mbt2 (pcg)	165	SPU_13689
Sp-mef2 (mads)	352	SPU_16168
Sp-mll3 (trxG)	176	SPU_26465
Sp-mta1 (myb)	285	SPU_07389, 03705
Sp-myb (myb)	284	SPU_00861
Sp-nfIA (NFI)	106	SPU_23339
Sp-nfkB (NFI)	39	SPU_08177
Sp-nsd1 (trxG)	228	SPU_27218
Sp-P3A2	287	SPU_17725
Sp-prkl2 (lim)	279	SPU_23090
Sp-rfx3	70	SPU_07611
Sp-runt1	289	SPU_06917
Sp-runx1	288	SPU_07852
Sp-scml1 (pcg)	164	SPU_26763
Sp-srf (mads)	341	SPU_27774
Sp-Su(H) (IPT)	326	SPU_21566
Sp-tead3	291	Scaffold71849
Sp-tead4	292	SPU_21210
Sp-trx2 (trxG)	356	SPU_15421
Sp-tubby (tulp)	217	SPU_16617
Basic Zipper		
Sp-atf2	354	SPU_26905
Sp-atf6	400	SPU_07749
Sp-creb3L1	402	SPU_12838
Sp-creb3L3	220	SPU_06803
Sp-crem	399	SPU_05358
Sp-fos	398	SPU_21173
Sp-fra2	xx	SPU_21172
Sp-giant	282	SPU_14528
Sp-hlf	280	SPU_04414
Sp-jun	5	SPU_03102
Sp-lztf1	283	SPU_04844,00424
Sp-mafB	281	SPU_25888
Sp-nfIL3	337	SPU_24307
Sp-nfe2-like	7	SPU_08752,11174
Sp-xbp1	401	SPU_08703
Smad		
Sp-smad1.5.8	23	SPU_20722,23107
Sp-smad2.3	11	SPU_17642
Sp-smad4	25	SPU_04287,17971
Sp-smad6.7	290	SPU_01998,18246
Tbox		
Sp-bra	-	SPU_20451
Sp-tbr	-	SPU_25584
Sp-tbx1	142	SPU_06150
Sp-tbx2/3	28	SPU_23386
Sp-tbx20	203	SPU_18392
Sp-tbx6	110	SPU_20346

Table 2.2. Summary of identified genes and corresponding gene model numbers.

Factors of the Sox/Hmg family can be subdivided into two main classes: the sequence-specific DNA binding Sox factors, and the general DNA binding Hmg factors. Here we are concerned only with the former. Canonical Sox transcription factors are grouped into families A-J, according to homology within their DNA binding domains. Of

these, SoxA factors, also known as the Sry subfamily, are vertebrate specific, and SoxH, I, and J are each comprised of just one gene (Bowles et al., 2000). Thus, as shown in figure 2.1A, outside of these the sea urchin genome has nearly the complete expected repertoire of sox family genes, missing only a member of the SoxG family. Recently sox-like genes have been discovered which have hmg boxes but are phylogenetically distinct from both the *hmg* and *sox* class genes (Lee, 2002). The tree in figure 2.1B identifies sea urchin orthologs of these, namely *bobby sox* (*Sp-bbx*(205, SPU_23037)) and *capicua* (*Sp-cic* (335, SPU_25292)), with *tcf/lef* genes included as an out-group.

A phylogenetic analysis of sea urchin *smad* genes is shown in figure2.1C. The Smad family is comprised of four sets of transcription factors with distinct functions. Two of these are R-smads, activated by either BMP or TGFβ signaling systems. Another subset is composed of Co-Smads, which are cofactors needed for R-Smad mediated gene activation. Finally, I-Smads inhibit R-Smads by interfering with their activation (Itoh et al., 2000). Our analysis shows that sea urchin has the complete bilaterian set of Smad factors. Specifically, the genome has an R-Smad for each signaling pathway, a Co-Smad, and an I-Smad.

The *bHLH* gene repertoire of *S. purpuratus*, with 47 members, presents a more complex picture. While the majority of sea urchin genes from this family have clear homology to just one subfamily, there are several apparent deletions and a few genes of unclear phylogeny. The bHLH factors are grouped into seven classes, Groups A-F and the Atonal superfamily, encompassing at least 44 subfamilies of genes (Ledent et al., 2002). Given the size of the *bHLH* gene family, separate trees were made for each class, and the diagram in Figure 2A shows the relationships among the classes. To improve the

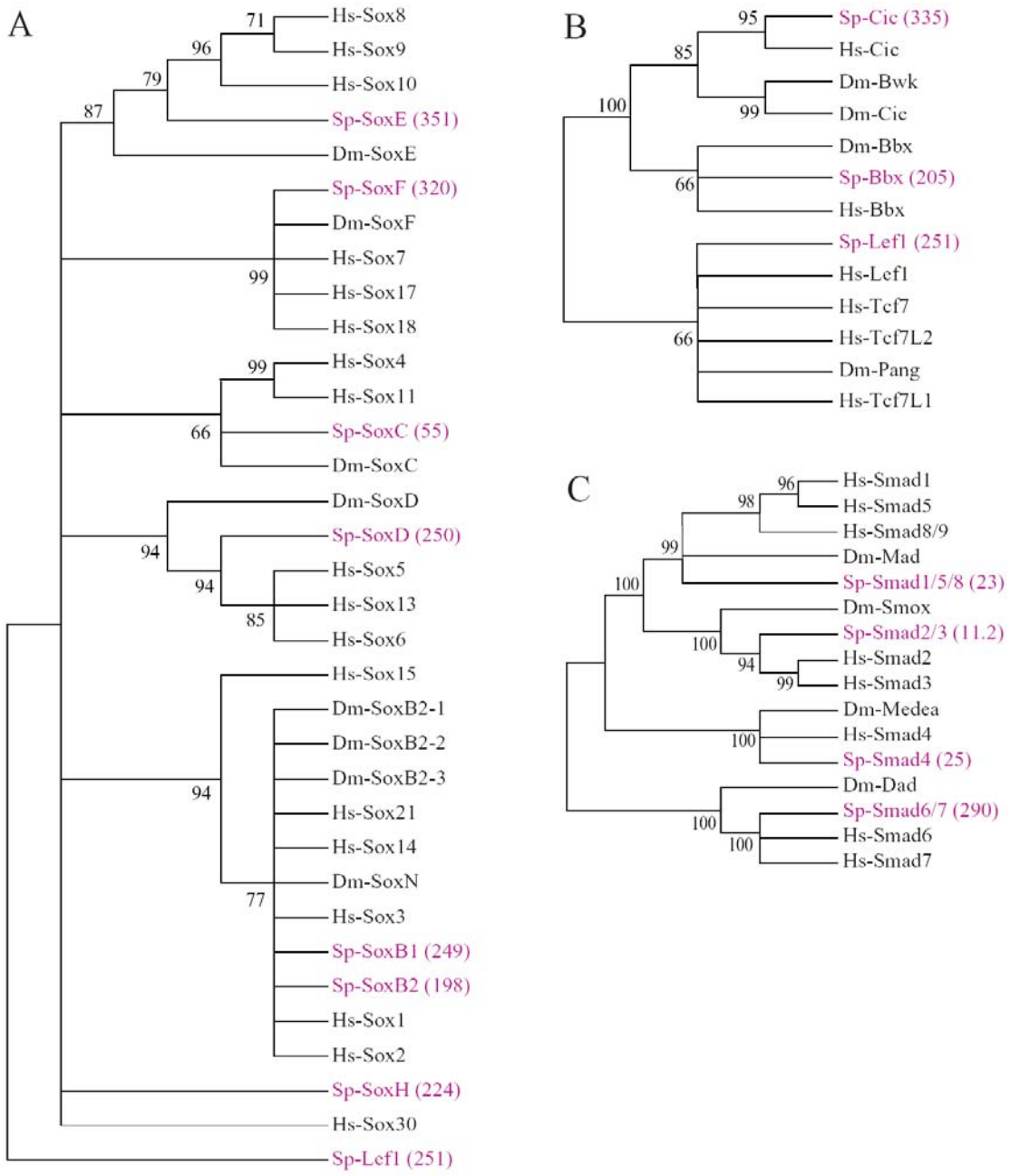


Figure 2.1. The sea urchin Sox/Hmg and Smad families. A phylogenetic analysis of canonical Sox factors is shown in 2.1A, with closely related sox-like genes in 2.1B. Since the sea urchin has no *sry* genes, human *sry* was omitted from this analysis. An analysis of sea urchin Smad factors in 2.1C shows the four genes cluster clearly to the four main functional sets of smad genes. The number in parenthesis following each gene name is an index number to facilitate lookup in the summary figure 2.9.

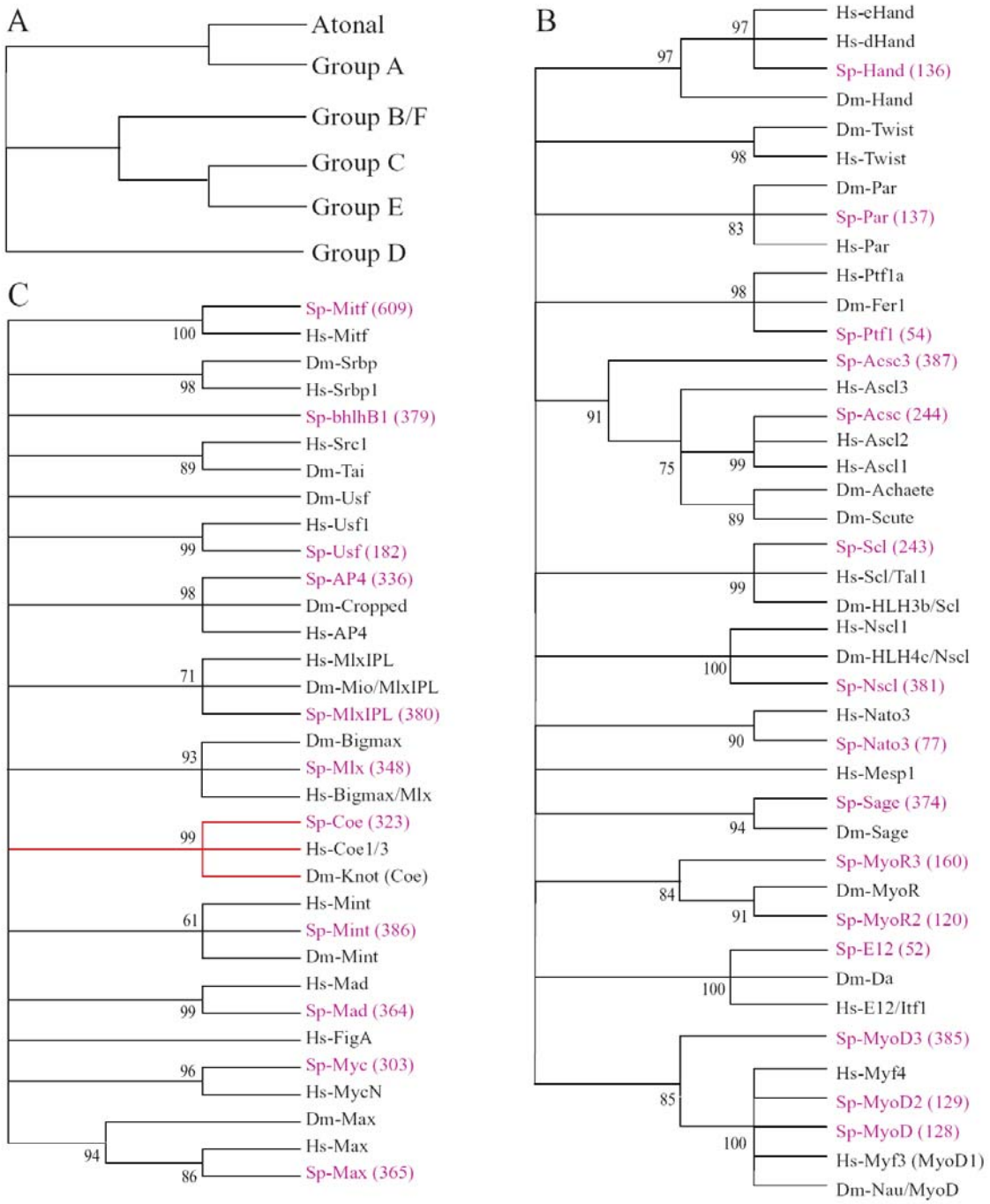


Figure 2.2. The bHLH family structure and phylogenetic trees of genes from Group A and Group B/F. The overall structure of the bHLH gene family is diagrammed in 2.2A. Group F is a single subfamily distinguished by a coe domain and located within the Group B class. The phylogenetic trees of Group A (2.2B) and Groups B/F (2.2C) genes were constructed with the neighbor joining method with 1,000 bootstrap replications. In 2.2C, the Group F gene family is highlighted in red.

clarity of the trees, human and fly paralogs from populous subfamilies were pruned if they provided no additional phylogenetic information about the sea urchin family member.

Phylogenetic analyses of sea urchin Group A and Group B/F bHLH factors are shown in figures 2.2B and 2.2C, respectively. Both Group A and B proteins bind to distinctive DNA sequences termed E-boxes. Group F genes are a single subfamily of Group B which include an additional domain, the coe domain, involved in both dimerization and DNA binding. Two-thirds of *S. purpuratus* Group A genes cluster monophyletically to human or fly genes with strong bootstrap values, with the orthology of the remaining four somewhat less clear. In this family, there is only one apparent deletion, *twist*. *Sp-acsc3* (244, SPU_28148) clearly belongs to the achaete-scute subfamily, and is not a recent duplication of *Sp-acsc*, which clusters more closely to two human orthologs *Hs-acscl1* and *Hs-acscl2*. Relatively recent duplications do appear to have occurred in the MyoD family, as there are three members of this family in the sea urchin. In the Group B/F family, all but one of the genes cluster cleanly to a single subfamily. The exception is *Sp-bhlhB1*, which has a low bootstrap association with both *srbp* and *src* genes (data not shown). Also notable in Group B is the absence of the *figa* gene. The single Group F subfamily gene, *Sp-coe* (607, SPU_04702), is highlighted in figure 2.2C in red.

Continuing through the classes, trees of the Atonal superfamily and Group C are shown in figures 2.3A and 2.3B. The Atonal superfamily is actually a large internal branch of Group A. While most of these sea urchin genes have strong orthology to a

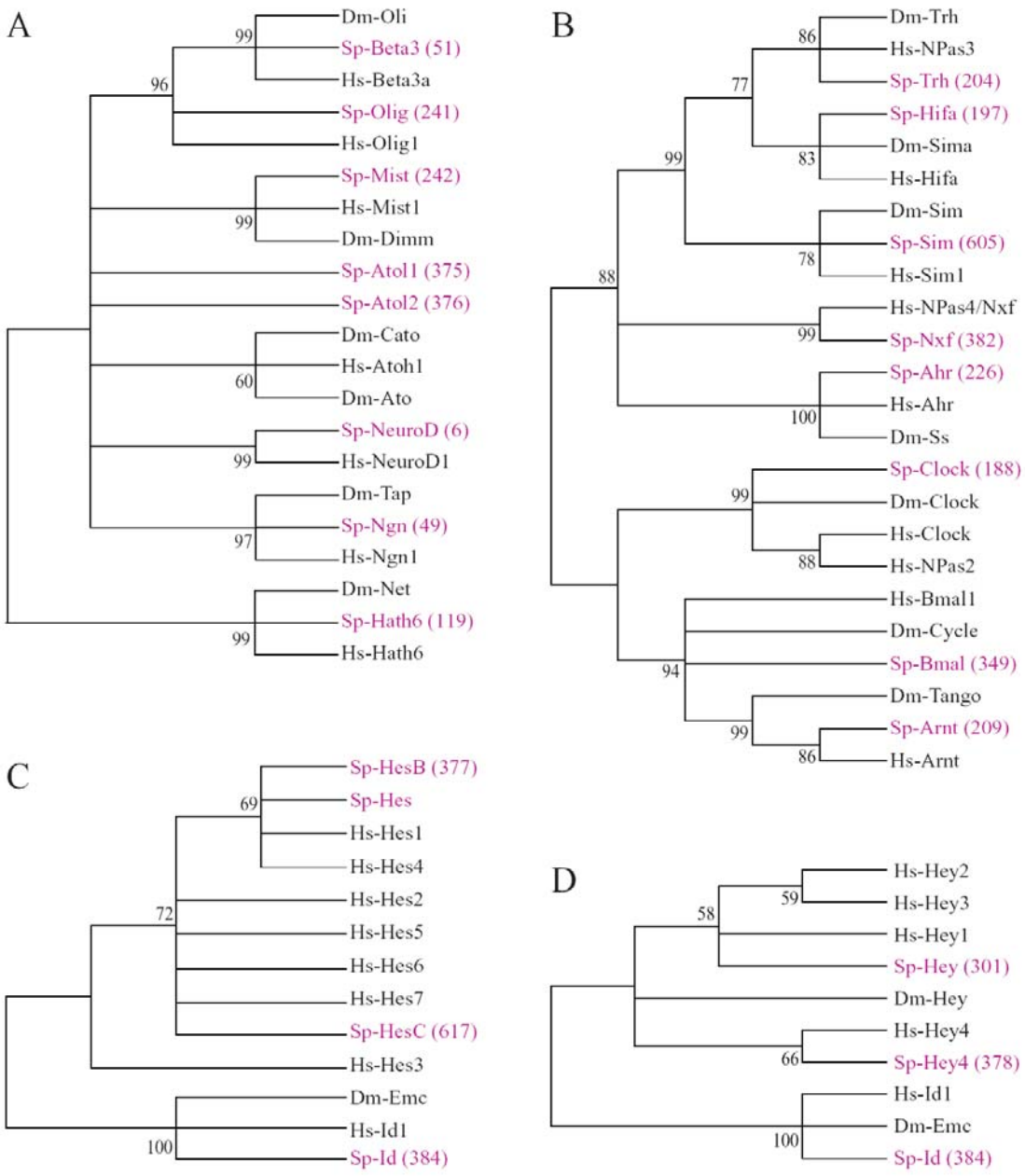


Figure 2.3. The Phylogeny of bHLH Groups C-E and the Atonal superfamily. Phylogenetic trees of sea urchin, human, and fly genes from the Atonal superfamily and Group C are shown in 2.3A and 2.3B, respectively. Because Group D genes appear to have diverged more rapidly, the Hes and Hey subfamilies were analyzed separately. All sea urchin Group E genes were analyzed with the Hes (2.3C) and Hey (2.3D) families, and each tree was then pruned to show the correct assignment. In both cases, the Group D class (Emc/Id subfamily) was included as an out-group. All trees were calculated using the neighbor joining method with 1,000 bootstrap iterations.

single subfamily (fig. 2.3A), the placement of two genes, *Sp-atol1* (375, SPU_00990) and *Sp-atol2* (376, SPU_03681), is ambiguous. In addition, no gene clusters clearly to the Atonal subfamily. One possibility is these two genes belong to the Atonal subfamily, but have been evolving at a rate that obscures their orthology. Group C genes (fig. 2.3B) are characterized by the presence of Pas domains. The sea urchin has a complete repertoire of these genes, with one gene per subfamily and no deletions or duplications.

The final classes are Group D and Group E. Group D bHLH factors, also known as the Hey subfamily, bind to N-box DNA sequences and contain an Orange domain. The phylogenetic relationships within Group D are much less clear, suggesting that these subfamilies are evolving more quickly than other bHLH classes. Trees constructed with all Group D sequences were uninformative due to very low bootstrap values, including between subfamilies. To circumvent this problem, trees were made with just sea urchin and human sequences. Human Hes and Hey sequences were analyzed separately, including all sea urchin Group D genes in both trees, with Group E genes as an out-group. In this way it was determined which sea urchin genes belong in which subfamily. A phylogenetic tree of human and sea urchin Hes genes, with Group E as an out-group, is shown in figure 2.3C. Since it is not possible to discern which human genes are paralogs of sea urchin *hes* genes, the two newly identified genes were named *Sp-hesB* (377, SPU_21608) and *Sp-hesC* (617, SPU_06813). The Hey subfamily structure is depicted in figure 2.3D.

Finally, a phylogenetic analysis was also performed for the *nuclear receptor* gene family. These genes are ligand activated transcription factors which provide direct links

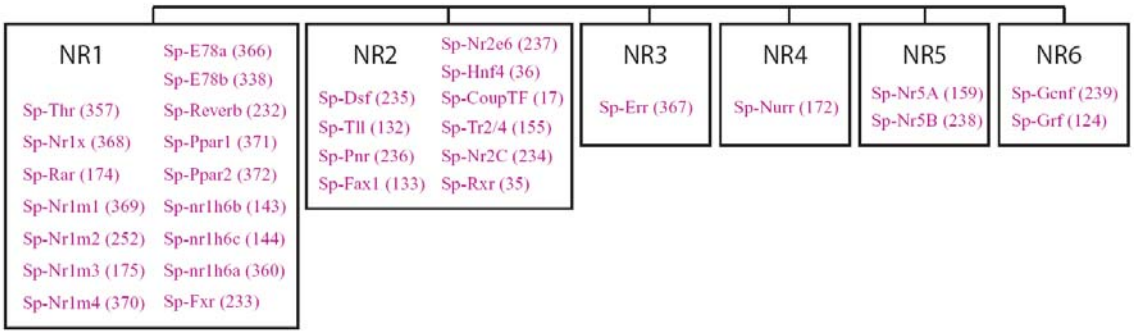


Figure 2.4. The *S. purpuratus* family of nuclear receptors. The nuclear receptor family is divided into six branches NR1-NR6. The 33 sea urchin nuclear receptors belong to the various branches as depicted here. The number in parentheses following each gene name is included to facilitate lookup in the summary figure 2.9.

between a small molecule ligand and gene activation. A tree of the Nuclear Receptor factors divides into six branches, NR1-NR6 (Bertrand et al., 2004). An additional category, NR0, is reserved for those genes which have lost either the ligand binding domain (LBD) or the DNA binding domain (DBD). We identified a total of 33 nuclear receptors in the sea urchin genome, and figure 2.4 shows the distribution of these within the 6 major families. Phylogenetic trees for the subfamilies with more than one sea urchin member were calculated using both the LBD and DBD sequences from urchin, human, fly and ciona.

The individual trees in figure 2.5 show that unlike the other transcription factor families considered here, nuclear receptor genes have evolved sufficiently to make identification of many orthologs within the subfamilies challenging. Within the NR1 family (fig. 2.5A), the identification of *Sp-rar* (174, SPU_16523) and *Sp-thr* (357, SPU_18861, SPU_25239) is very strong. Likewise, the two sea urchin *ppar* genes are clearly the result of a recent duplication. The *Sp-nr1ha* (360, SPU_17404), *Sp-nr1hb*

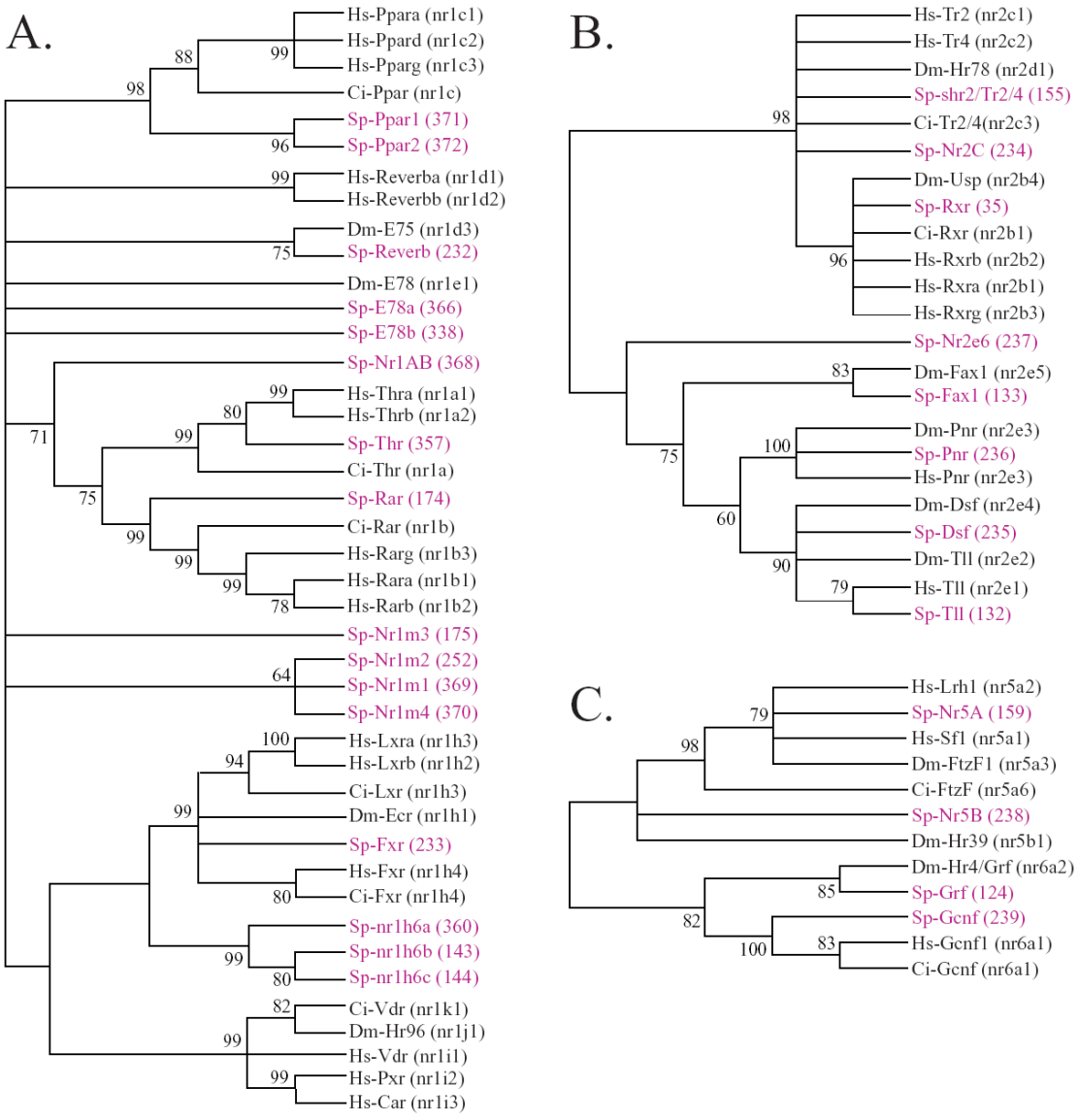


Figure 2.5. Nuclear receptor phylogeny. Phylogenetic trees of nuclear receptor classes with more than one sea urchin member were calculated by the neighbor joining method with sensing 1,000 bootstrap iterations. Both DNA and ligand binding domains from human, fly and ciona genes were used to identify sea urchin genes of types NR1 (2.5A), NR2 (2.5B), and NR5 and NR6 (2.5C).

(144, SPU_15456), *Sp-nr1hc* (143, SPU_04526) genes are also the result of recent duplications, but the orthology of the ancestral gene in unclear. Likewise, the four *Sp-*

nr1m genes likely arise from a series of duplications, but the ancestral gene is unclear beyond the general NR1 classification. Within the NR1H group, *Sp-fxr* (233, SPU_11348, SPU_27598) shows homology to the *lxr* genes though its LBD, and *fxr* through its DBD. The gene was named to reflect stronger homology in the DBD, which is generally more conserved in nuclear receptor proteins (Bertrand et al., 2004). Also of note is the presence of two potential *e78* orthologs, which are not present in chordates. The assignment of these genes is tentative as the C-terminal half of *Sp-E78b* (338, SPU_18366), including the LBD, is missing.

The lineage of sea urchin NR2 family members is somewhat more clear (fig. 2.5B). Three genes, *Sp-fax1* (133, SPU_12586), *Sp-tll* (132, SPU_08936, SPU_27487) and *Sp-pnr* (236, SPU_14405) cluster plainly to either human and/or fly orthologs. *Sp-dsf* (235, 24486), while not monophyletic, is almost certainly an ortholog of *Dm-dsf*. *Sp-rxr* (35, 28422) likewise must derive from the same ancestral gene as other members of the Rxr subfamily. The remaining two genes, however, are of ambiguous lineage.

Finally, a phylogenetic analysis of sea urchin NR5 and NR6 genes is shown in figure 2.5C. *Sp-nr5A* (159, SPU_13843) and *Sp-nr5B* (238, SPU_Scaff7192) were given systematic names corresponding to the two *nr5* genes inferred to be part of the pan-bilaterian nuclear receptor tool kit (Bertrand et al., 2004). An unexpected result is the discovery of two sea urchin members of the *nr6* family (*Sp-grf* (124, SPU_13305) and *Sp-gcnf* (239, SPU_00749)), one clustering to chordate orthologs, the other to a fly ortholog. Finally, no glucocorticoid receptor was found.

Temporal gene expression

Quantitative PCR (QPCR) experiments were undertaken to measure expression of newly identified genes during early development. Given the high rate of polymorphisms in *S. purpuratus*, QPCR primers were designed very carefully to assure uniform primer efficiency and consistent results. As much as possible, QPCR primers were chosen to fall within the conserved DNA binding domain. Since we have only gene predictions and not complete mRNAs for most genes, this has the added benefit of avoiding potential subtle prediction errors in less conserved regions. Having located a suitable target region, the individual genomic sequencing reads used to assemble that short stretch of the genome were retrieved and aligned. In this way we were able to identify at least the SNPs present in the sequenced genomes and avoid including these positions in our primers.

Primer pairs were validated by QPCR against digested genomic DNA. Primers giving anomalously high or low amplification compared to the standard single copy gene ubiquitin were redesigned. Primer pairs with anomalous denaturation curves, potentially reflecting primer dimerization, were also redesigned. Finally, gene expression was measured in triplicate at six time points: unfertilized egg, 6, 12, 18, 24, 36, and 48 h post-fertilization.

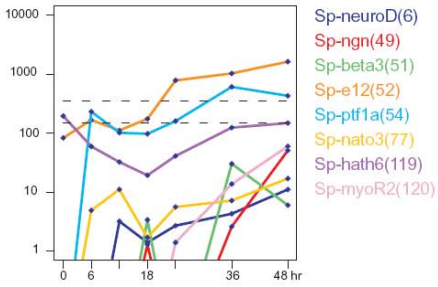
For some genes, high quality primers could not be generated despite numerous attempts, and no expression data are reported for these genes. There are several reasons some genes are problematic. Given that we wish to limit our primers to DNA binding domains, sometimes these sequences are simply not the best suited for primer selection.

In other cases, the selected target regions may be more polymorphic than is apparent from the two phenotypes incorporated in the genome assembly. Alternately, small unrecognized sequencing or assembly errors in the target region may contribute to these difficulties. There are 7 genes for which we do not have expression data: two basic zipper and five bHLH genes.

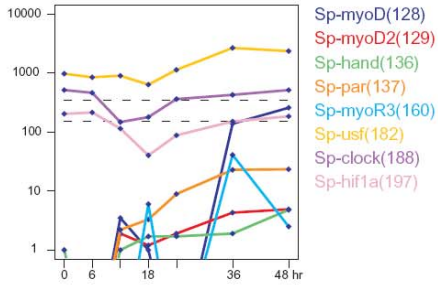
Expression time courses for newly identified genes are shown in figure 2.6, grouped by family and plotted on a logarithmic axis for easy comparison. The graphs show the progression of transcripts per embryo over the first two-thirds of embryonic development. Two dashed guidelines indicate a somewhat arbitrary threshold range of biological significance between 150-350 copies/embryo (see legend for fig. 2.6). This threshold range would be sufficient to capture the first biologically relevant expression of *Sp-pmar1* and *Sp-dri*, both initially expressed in only a few cells, the micromeres (Amore et al., 2003; Oliveri et al., 2002). A glance at the graphs in figure 2.6 is sufficient to note the variety of expression profiles. Even within families, the genes are clearly operating in response to many distinct sets of instructions, and only 12/181 genes have constant expression profiles. Thus the great majority of these genes are not performing

Figure 2.6. Temporal gene expression of *S. purpuratus* transcription factors. The graphs show gene expression levels from 0-48 h post-fertilization, plotted on a logarithmic y-axis. The number of copies expressed per embryo was obtained by QPCR experiments done in triplicate (materials and methods). Dashed lines at 150 and 350 copies per embryo indicate an estimated minimum range for biologically significant expression of a transcription factor. The average copy number for low prevalence maternal transcripts in *S. purpuratus* eggs is 1600 copies per embryo (Davidson, 1986), and an arbitrary guideline of >400 copies/embryo should reasonably capture significant mRNAs encoding transcription factors in the egg. At later time points, expression as low as 200 copies/embryo can be detected by WMISH if expressed in a small domain of 20 cells at ~10 copies/cell. Allowing for some primer inefficiencies, we used a biological significance guideline of >150-350 copies/embryo.

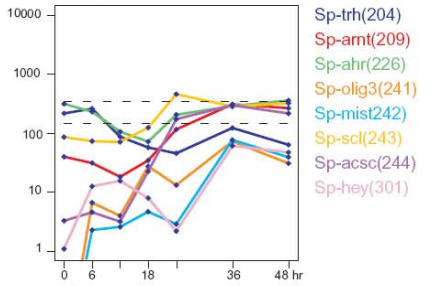
bhlh1



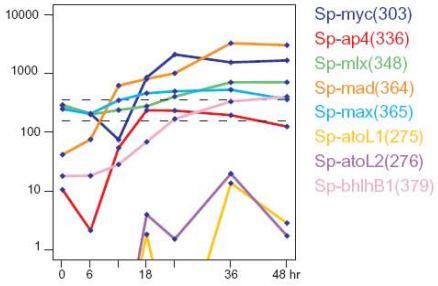
bhlh2



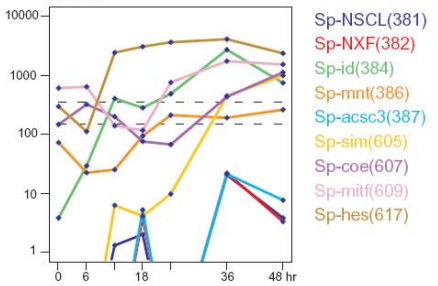
bhlh3



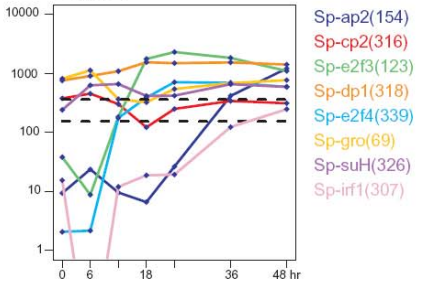
bhlh4



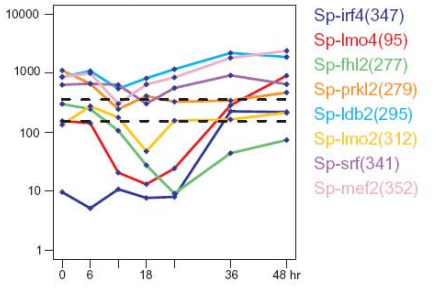
bhlh5



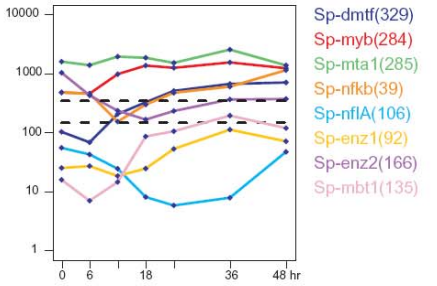
others1



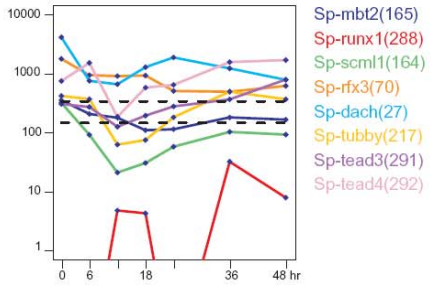
others2



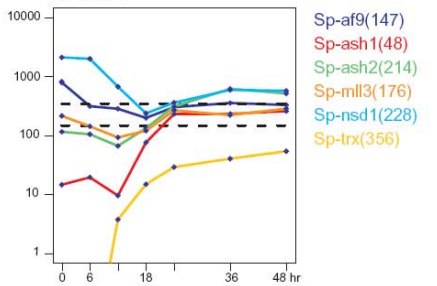
others3

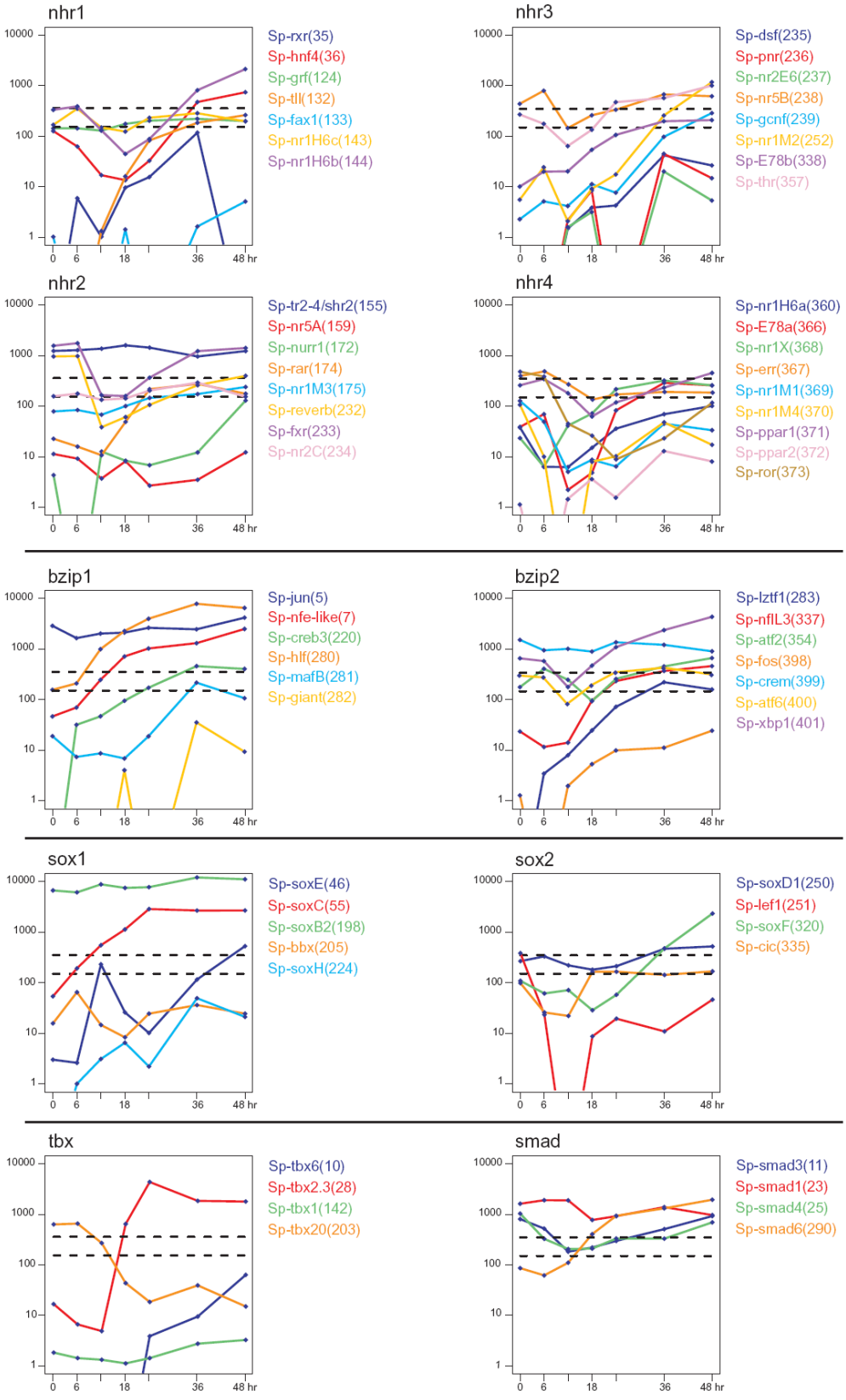


others4



others5




nhr1
Sp-rxr(35)
Sp-hnf4(36)
Sp-grf(124)
Sp-tll(132)
Sp-fax1(133)
Sp-nr1H6c(143)
Sp-nr1H6b(144)
nhr3
Sp-dsf(235)
Sp-pnr(236)
Sp-nr2E6(237)
Sp-nr5B(238)
Sp-gcnf(239)
Sp-nr1M2(252)
Sp-E78b(338)
Sp-thr(357)
nhr2
Sp-tr2-4/shr2(155)
Sp-nr5A(159)
Sp-nurr1(172)
Sp-rar(174)
Sp-nr1M3(175)
Sp-reverb(232)
Sp-fxr(233)
Sp-nr2C(234)
nhr4
Sp-nr1H6a(360)
Sp-E78a(366)
Sp-nr1X(368)
Sp-err(367)
Sp-nr1M1(369)
Sp-nr1M4(370)
Sp-ppar1(371)
Sp-ppar2(372)
Sp-ror(373)
bzip1
Sp-jun(5)
Sp-nfe-like(7)
Sp-creb3(220)
Sp-hlf(280)
Sp-mafB(281)
Sp-giant(282)
bzip2
Sp-lztf1(283)
Sp-nfIL3(337)
Sp-atf2(354)
Sp-fos(398)
Sp-crem(399)
Sp-atf6(400)
Sp-xbp1(401)
sox1
Sp-soxE(46)
Sp-soxC(55)
Sp-soxB2(198)
Sp-bbx(205)
Sp-soxH(224)
sox2
Sp-soxD1(250)
Sp-lef1(251)
Sp-soxF(320)
Sp-cic(335)
tbx
Sp-tbx6(10)
Sp-tbx2.3(28)
Sp-tbx1(142)
Sp-tbx20(203)
smad
Sp-smad3(11)
Sp-smad1(23)
Sp-smad4(25)
Sp-smad6(290)
10000
1000
100
10
1
0
6
18
36
48 hr

housekeeping functions, but rather are likely to be contributing to the cascade of information which specifies the territories of the developing embryo.

In figure 2.7 genes are grouped according the time embryonic activation is first apparent, irrespective of the level of maternal transcripts. This chart includes both new

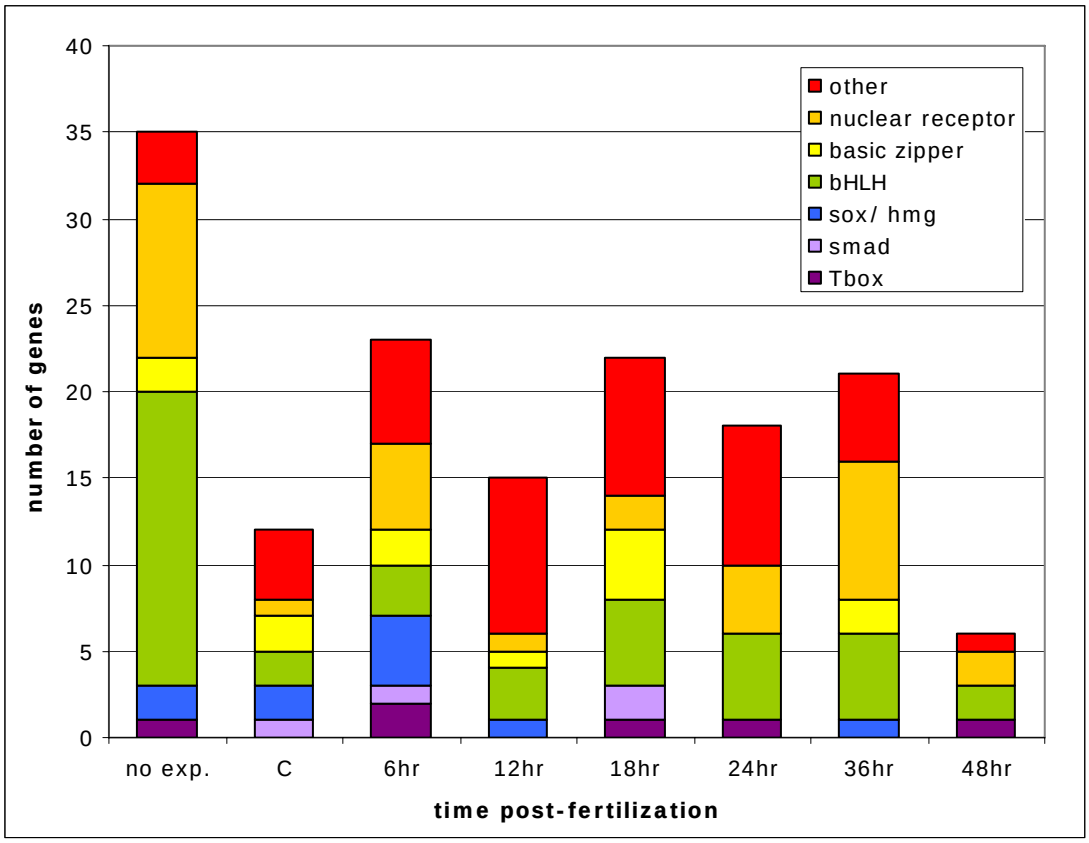


Figure 2.7. The distribution of transcription factor activation during development. The total height of each bar represents the number of transcription factors which are first activated at the indicated time post-fertilization. The bars are further parsed to show the proportion of genes from each family contributing to new gene activation at a given time point. A color key for the different gene families is given in a legend at the top right corner. The number of genes not expressed by 48 h is given in the column labeled 'no exp.;' 'C' gives the number of genes expressed at a constant rate from 0-48 h.

and previously reported transcription factors of the regulatory gene families included in this report. Overall, activation of new transcription factor genes occurs relatively evenly throughout development. This steady rate of new gene activation also applies to the individual families, which are not heavily biased towards any particular time point. Note, however, that a higher proportion of bHLH and nuclear receptor genes are still unexpressed by 48 h. Most striking, though, is that when the embryo is still at the late gastrula stage, only one-fifth of the regulatory genes studied here remain unexpressed.

Spatial gene expression patterns

Whole mount *in situ* hybridization was used to determine the spatial expression patterns of sufficiently active genes. Given the number of genes of interest, probes were made only for transcripts expressed at 500 copies per embryo or more. Furthermore, we focused on early development, studying only genes expressed at that level by 24 h post-fertilization (PMC ingression), though for these genes observations were carried out to the 36 h late gastrula stage. Our strategy for designing probes balanced a need for sufficiently sensitive probes and a desire to use high throughput methods, against a background of limited sequence information.

The expression patterns obtained in this study identified new players in all the major embryonic territories. *In situ* hybridizations of genes that display localized expression patterns are presented in figure 2.8. The basic zipper genes *Sp-jun* (5, SPU_03102) (fig. 2.8A) and *Sp-hlf* (280, SPU_04414) (fig. 2.8B) are expressed in the PMCs and the apical ectoderm, respectively. *Sp-smad4* (25, SPU_04287, SPU_17971)

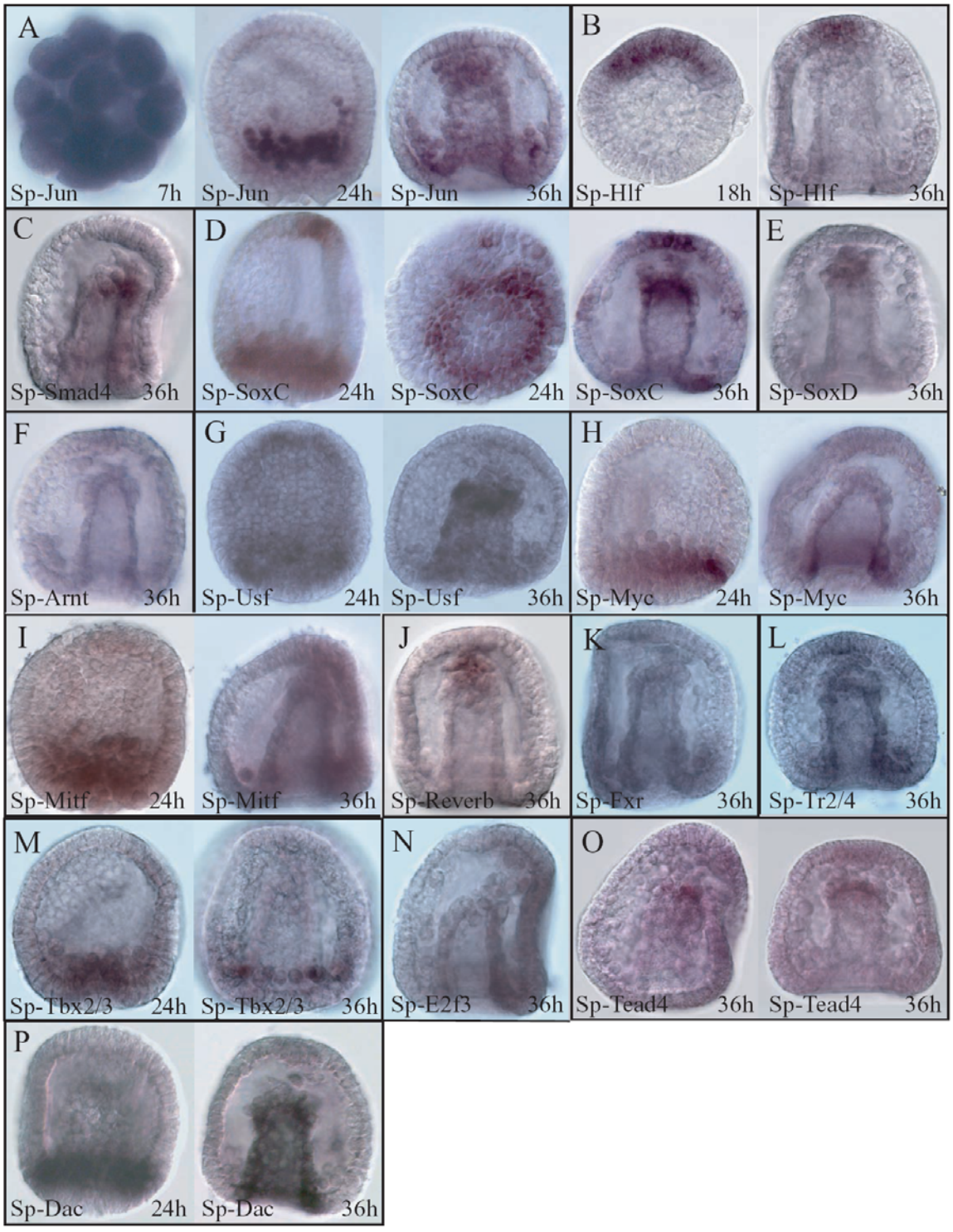

A
Sp-Jun 7h
Sp-Jun 24h
Sp-Jun 36h
B
Sp-Hlf 18h
Sp-Hlf 36h
C
Sp-Smad4 36h
D
Sp-SoxC 24h
Sp-SoxC 24h
Sp-SoxC 36h
E
Sp-SoxD 36h
F
Sp-Arnt 36h
G
Sp-Usf 24h
Sp-Usf 36h
H
Sp-Myc 24h
Sp-Myc 36h
I
Sp-Mitf 24h
Sp-Mitf 36h
J
Sp-Reverb 36h
K
Sp-Exr 36h
L
Sp-Tr2/4 36h
M
Sp-Tbx2/3 24h
Sp-Tbx2/3 36h
N
Sp-E2f3 36h
O
Sp-Tead4 36h
Sp-Tead4 36h
P
Sp-Dac 24h
Sp-Dac 36h

Figure 2.8. Spatial expression of transcription factor genes. Panels A-O are *in situ* hybridizations of previously unstudied homeobox transcription factors displaying localized expression patterns. The gene name is displayed in the bottom left corner; the time postfertilization is indicated in the bottom right corner.

expression (fig. 2.8C) is restricted to the tip of the archenteron, appearing at 36 h. The two *sox* genes studies both showed localized expression patterns. *Sp-soxC* (55, SPU_02603) expression (fig. 2.8D) appears by 24 h in a ring of veg1 cells and in the apical ectoderm, as well as in small patches at the animal-vegetal boundary. With ingression of the archenteron, *Sp-soxC* expression is established in the foregut. *Sp-soxD* (250, SPU_04217) expression (fig. 2.8E) is not visible until 36 h, localizing to the tip of the invaginating gut. Expression of each of the four *bHLH* genes localizes to a distinct territory of the embryo. *Sp-arnt* (209, 00129) expression (fig. 2.8F) is visible in all but either the oral or aboral face of the embryo by 36 h. *Sp-usf* (182, SPU_14332) expression (fig. 2.8G) is restricted to the SMCs and foregut, while *Sp-myc* (303, SPU_03166) (fig. 2.8H) is visible in a ring around the blastopore. Finally, *Sp-mitf* (609, SPU_08175) is seen in the PMCs at 24 h (fig. 2.8I). Expression of nuclear receptor genes is similarly dispersed through the embryo. *Sp-reverb* (232, SPU_17492) (fig. 2.8J) is confined to the tip of the gut, while *Sp-fxr* (233, SPU_11348, SPU_27598) (fig. 2.8K) is visible in all domains but either the oral or aboral face, and *Sp-tr2/4* (155, SPU_08117) is seen in the gut and apical ectoderm (fig. 2.8L). The remaining genes for which localized expression was mapped are from much smaller families. *Sp-e2f3* (123, SPU_06753) is activated in the oral ectoderm and the oral side of the gut by 36 h (fig. 2.8M). *Sp-tead4* (292, SPU_21210) (fig. 2.8N) has a very distinctive expression pattern limited to a thin row of

cells at the tip of the gut. Finally, *Sp-dac* (27, SPU_28061) (fig. 2.8O) is on strongly in the veg1 territory by 24 h, and is established throughout the gut by 36 h.

Discussion

In this work we report the identification and developmental expression of 141 previously unknown sea urchin regulatory genes. For the larger gene families we show detailed phylogenetic analyses. For the unitary gene types and small gene families, the quality of the identifications of the genes is indicated by the high significance values of the best blastx matches to sequences in the nr database (fig.9). The present study, taken together with the accompanying papers on *fox* genes (Tu et al, 2006), *ets* genes (Rizzo et al, 2006), zinc finger genes (Materna et al, 2006), and homeodomain genes (Howard-Ashby et al, 2006) completes the description of the sea urchin regulome. Zinc finger genes are probably not all regulatory in function as this motif occurs in various other kinds of proteins, and zinc finger genes are apparently evolving rapidly in many animal clades (Materna et al, 2006). In contrast to these, the sea urchin genes encoding *ets*, *fox* and homeodomain regulators, and in detail their many subfamilies, are in their DNA binding domains overwhelmingly orthologous to the corresponding gene families and subfamilies of flies, humans, and other bilaterians. The phylogenetic analyses and sequence similarity assessments in this chapter powerfully support the same conclusion for the remainder of the regulatory gene classes. They demonstrate panbilaterian orthology for virtually all other classes of regulatory gene, though in each clade there is a small minority of divergent genes. Because echinoderms are distant from any animal for

Key:

no in situ	6 h	24 h	no/low
new in situ	12 h	36 h	maternal
known in situ	18 h	48 h	constant

Index	Gene Name	Cat.		7h	12h	18h	24h	36h	Best Hit
	Smad (4)								
11	Sp-smad3			-	-	-	-	-	4e-57
23	Sp-smad1			ubiq	-	-	-	-	7e-31
25	Sp-smad4			-	-	-	-	gut tip	3e-26
290	Sp-smad6								8e-56
	Nuclear Hormone Receptors (33)								
-	Sp-coupTF1			oral	oral	oral	OE	OE	8e-79
35	Sp-rxr								1e-84
36	Sp-hnf4								1e-141
124	Sp-grf								2e-30
132	Sp-tll								7e-53
133	Sp-fax1								8e-33
143	Sp-nr1H6c								9e-22
144	Sp-nr1H6b								7e-20
155	Sp-shr2/Tr2.4			ubiq	ubiq	ubiq	gut, apical	gut, apical	1e-33
159	Sp-nr5A								4e-38
172	Sp-nurr1								6e-37
174	Sp-rar								3e-36
175	Sp-nr1M3								6e-23
232	Sp-reverb			ubiq	-	-	-	gut tip	3e-26
233	Sp-fxr			-	-	-	OE	OE	4e-33
234	Sp-nr2C								3e-21
235	Sp-dsf								2e-41
236	Sp-pnr								1e-55
237	Sp-nr2E6								3e-36
238	Sp-nr5B								1e-22
239	Sp-gcnf								2e-35
252	Sp-nr1M2								4e-32
338	Sp-E78b								5e-29
357	Sp-thr			-	-	-	-	-	5e-49
360	Sp-nr1H6a								8e-19
366	Sp-E78a								5e-26
367	Sp-err								3e-29
368	Sp-nr1x								2e-27
369	Sp-nr1M1								5e-31
370	Sp-nr1M4								6e-24
371	Sp-ppar1								9e-49
372	Sp-ppar2								3e-47
373	Sp-ror								1e-11
	Basic zipper (13)								
5	Sp-jun			ubiq	ubiq	ubiq	PMC	-	5e-58
7	Sp-nfe2-like			-	-	-	ubiq	ubiq	2e-38
220	Sp-creb3								1e-38
280	Sp-hlf			-	apical	apical	apical	apical	3e-57
281	Sp-mafB								4e-24
282	Sp-giant								9e-17
283	Sp-lztf1								1e-70
337	Sp-nfIL3			-	-	-	-	-	8e-14
354	Sp-atf2			-	-	-	-	-	7e-62
398	Sp-fos								7e-17
399	Sp-crem								5e-55

Index	Gene Name	Cat.		7h	12h	18h	24h	36h	Best Hit
	Sox/ HMG (10)								
46	Sp-soxE								6e-29
55	Sp-soxC			-	PMC	M	M, apical	apical, ecto, gut	1e-26
198	Sp-soxB2								2e-45
205	Sp-bbx								6e-16
224	Sp-soxH								3e-14
-	Sp-soxB1			non-umere	non-vegetal	non-vegetal	ectoderm	ectoderm	2e-45
250	Sp-soxD			-	-	-	-	gut tip	1e-25
251	Sp-lef1								2e-15
320	Sp-soxF			-	-	-	-	-	8e-16
335	Sp-cic								4e-29
	bHLH (43)								
-	Sp-hes			-	-	-	veg, ecto	OE	-
6	Sp-neuroD								3e-21
49	Sp-ngn								2e-22
51	Sp-beta3								3e-24
52	Sp-E12			-	-	-	ubiq	ubiq	2e-19
54	Sp-ptf1a								3e-21
77	Sp-nato3								1e-15
119	Sp-hath6								5e-17
120	Sp-myoR2								9e-22
128	Sp-myoD								1e-24
129	Sp-myoD2								2e-13
136	Sp-hand								7e-19
137	Sp-paraxis1								4e-18
160	Sp-myoR3								5e-18
182	Sp-usf			ubiq	ubiq	ubiq	M	foregut, SMC	1e-27
209	Sp-arnt			-	-	-	-	non OE or AO	4e-25
188	Sp-clock			-	-	-	-	-	3e-13
197	Sp-hif1a								1e-15
204	Sp-trh								8e-21
226	Sp-ahr								3e-18
241	Sp-olig3								6e-23
242	Sp-mist								3e-10
243	Sp-scl								4e-20
244	Sp-acsc								1e-10
301	Sp-hey								5e-14
303	Sp-myc			-	-	-	E	E	5e-17
336	Sp-ap4								4e-20
348	Sp-mlx			-	-	-	-	-	1e-19
349	Sp-bmal			-	-	-	-	-	3e-18
364	Sp-mad								4e-15
365	Sp-max								9e-22
375	Sp-atol1								6e-08
376	Sp-atol2								3e-11
379	Sp-bhlhB1								2e-25
381	Sp-NSCL								2e-12
382	Sp-NXF								7e-11
384	Sp-id								2e-11
386	Sp-mnt								3e-16
387	Sp-acsc3								2e-08
605	Sp-sim								2e-22
607	Sp-coe								6e-43
609	Sp-mitf			-	-	-	PMC	PMC,OE	1e-22
617	Sp-hesC			-	ubiq	ubiq	ubiq	ubiq	1e-12

Index	Gene Name	Cat.		7h	12h	18h	24h	36h	Best Hit
	T-box (6)								
-	Sp-tbr			ubiq	PMC	PMC	PMC	PMC	-
-	Sp-bra			-	-	veg plate	veg plate	SMC	-
110	Sp-tbx6								4e-39
28	Sp-tbx2/3			-	-	-	PMC, OE	PMC, OE	8e-49
142	Sp-tbx1								9e-15
203	Sp-tbx20			-	-	-	-	-	6e-14
	Other Genes (45)								
154	Sp-ap2 (AP2)								5e-20
-	Sp-dri (bright)			-	PMC	PMC	PMC	OE	-
316	Sp-cp2 (CP2)			-	-	-	-	-	3e-15
123	Sp-e2f3 (E2F)			-	-	-	OE	OE/api/ O gut	5e-25
318	Sp-dp1 (E2F)			ubiq	ubiq	ubiq	ubiq	-	8e-15
339	Sp-e2f4 (E2F)								3e-12
-	Sp-gataE			-	-	veg., non-umere	veg plate	gut, SMC	9e-20
-	Sp-gataC			-	-	-	M	M	-
-	Sp-gcm (gcm)			-	veg2	veg2	SMC	SMC	-
69	Sp-gro			ubiq	-	-	-	-	5e-28
326	Sp-Su(H) (IPT)								5e-14
307	Sp-irf1 (IRF)								3e-19
347	Sp-irf4 (IRF)								6e-11
95	Sp-lmo4 (lim)								2e-20
277	Sp-fhl2 (lim)								9e-15
279	Sp-prkl2 (lim)								2e-40
295	Sp-ldb2 (lim)			-	-	-	-	-	2e-22
312	Sp-lmo2 (lim)								3e-25
341	Sp-srf (mads)			-	-	-	-	-	4e-13
352	Sp-mef2 (mads)								7e-11
329	Sp-dmtf (myb)			-	-	-	-	-	6e-23
284	Sp-myb (myb)			-	-	-	-	OE	2e-90
285	Sp-mta1 (myb)								1e-177
39	Sp-nfkB (NFI)								1e-18
106	Sp- nflA (NFI)								3e-84
92	Sp-enz1 (pcg)								2e-15
166	Sp-enz2 (pcg)			ubiq	-	-	-	-	3e-16
135	Sp-mbt1 (pcg)								1e-28
165	Sp-mbt2 (pcg)								4e-24
287	Sp-P3A2			ubiq	ubiq	OE	OE	OE	-
288	Sp-runx1								6e-126
289	Sp-runt1			-	-	ubiq	vegetal, SMC	gut, OE	-
164	Sp-scml1 (pcg)								6e-32
70	Sp-rfx3			ubiq	ubiq	ubiq	ubiq	ubiq	1e-22
27	Sp-dach (Ski-Sno)			ubiq	ubiq	-	veg1	E	2e-46
217	Sp-tubby (tulp)								5e-16
291	Sp-tead3								6e-17
292	Sp-tead4			-	-	-	-	gut tip	1e-85
147	Sp-af9			ubiq	ubiq	ubiq	ubiq	-	2e-56
48	Sp-ash1 (trxG)								2e-27
214	Sp-ash2 (trxG)								7e-112
176	Sp-mll3 (trxG)			-	-	-	-	-	4e-30
228	Sp-nsd1 (trxG)			ubiq	-	-	-	-	1e-37
356	Sp-trx (trxG)								5e-12

Figure 2.9. Spatial and temporal expression of sea urchin transcription factors. The expression data for each of the identified transcription factors, sorted by family, is summarized. All novel transcription factors uncovered by our search algorithm and for which QPCR was done were assigned a working ID number (index). Genes with previously published expression time courses are indicated by a "-" in the index column. Newly identified proteins were named according to the closest known homologue, as identified by our phylogenetic trees or by blastx of nr if no tree was constructed for the gene family. The third and fourth columns relate whether the gene is maternally expressed (>400 copies/egg; indicated by a tan box) and by what time point (6 h = red; 12 h = orange; 18 h = yellow; 24 h = green; 36 h = blue; 48 h = violet; white = not before 48 h) expression rises to within the minimum range estimated to be significant (150-350 copies/ embryo). A black box indicates constant expression varying by less than twofold over the time period studied. Next is given the result of *in situ* staining, if done. Results written in blue are new findings; information in red is cited from previously published work. A '-' indicates no staining was observed at that stage. Gray boxes indicate no *in situ* was attempted. Finally, the "Eval" column gives the e-value of the top blastx match between the identified gene fragment and nr. Expression data for the following genes has been previously published: *Sp-coupTF* (Vlahou et al., 1996); *Sp-soxB1(Kenny* et al., *2003)*; *Sp-hes* (Minokawa et al., 2004); *Sp-tbr* (Croce et al., 2001); *Sp-bra* (Peterson et al., 1999); *Sp-tbx2/3* (Gross et al., 2003); *Sp-dri* (Amore et al., 2003); *Sp-gataE* (Lee and Davidson, 2004); *Sp-gataC* ; *Sp-gcm* (Ransick et al., 2002); *Sp-myb* (Coffman et al., 1997); *Sp-p3a2* (Zeller et al., 1995); *Sp-runt* (Robertson et al., 2002).

which genomic sequence has so far been available, these studies materially strengthen the concept of a panbilaterian regulome. This idea is now demonstrated with respect to all main branches of the deuterostomes and to ecdysozoans, but its final consummation will await annotated genomic sequence from animals belonging to lophotrochozoan clades.

A summary of both expression timecourses and, when available, spatial expression patterns, is given by family in figure 2.9. For each gene, the time of initial embryonic activation and whether or not there are maternal transcripts, is indicated

together with the spatial expression pattern from 7 to 36 h post-fertilization (grayed out areas indicate that *in situ* hybridization was not attempted). Here we briefly review gene usage by family.

bHLH genes

A total of 47 members of this family were identified in the sea urchin, and expression data are reported for 42 of these. While the majority of sea urchin *bHLH* genes are orthologous to specific human and fly genes, the detailed lineage of the *hes* and *hey* subfamily genes was less clear. Sea urchin *bHLH* genes are activated steadily throughout the developmental interval studied, though usually at a low to modest level of expression. Atypically, however, nearly half remain unexpressed at 48 h. This is much higher than for regulatory genes as a whole, as summarized in figures 8 and 9. The unexpressed *bHLH* genes are largely associated with specific cell differentiation functions, many of them neurogenesis. This process is not advanced in the embryo up to 36 hrs, and it is interesting that some of these same genes, e.g., *Sp-neuroD* (6, SPU_24918), are expressed in the post embryonic larva according to unpublished information (Huelguero and Cameron, 2006). The expression patterns of four *bHLH* genes were mapped to distinct territories of the embryo. *Sp-arnt* (209, SPU_00129) is present in what is probably the oral face of the embryo by 36 h after fertilization. The Arnt factor is the dimerization partner for other members of the bHLH-Pas family including Sim, Hifa, and Ahr, and has been implicated in detection and metabolism of foreign chemicals, and other functions (Kinoshita et al., 2004). The dimerization partner

of *Sp-arnt* in this context is unclear. Of much interest is the expression of *Sp-mitf* (609, SPU_08175) in the PMCs and SMC's (fig. 2.8I). SMC's later differentiate into pigments cells, and Mitf is known to be involved in pigment cell specification in vertebrates (Yajima et al., 2003). It would be interesting to see if other nodes of the specification and differentiation pathway have been conserved.

Nuclear receptor genes

Nuclear receptor genes constitute a large subset of the sea urchin regulome, with 33 family members identified. Phylogenetic analysis of sea urchin nuclear receptors suggests that changes are occurring in this family faster than in other *S. purpuratus* regulatory gene families. Two clusters of sea urchin genes in our phylogenetic tree, the Nr16H genes and Nr1M genes, reflect probable tandem duplication events. In addition, the detailed subfamily affiliations of some genes could not definitively be established, though most fell into known subclasses (fig.5). The *Sp-dsf* (235, SPU_24486) and *Sp-fax1* (133, SPU_12586) genes, and the probable orthologs of *Dm-E78* and *Dm-hr39* are interesting because these genes represent four of five predicted Urbilaterian nuclear receptors lost in chordates (Bertrand et al., 2004). Their presence in echinoderms confirms that these are chordate specific losses, as opposed to deuterostome deletions.

About two-thirds of nuclear receptor genes have been activated by 48 hours post-fertilization. There is very little information on the small molecule ligands that the proteins encoded by these genes might interact with in the sea urchin embryo. The expression patterns of several nuclear receptor genes were mapped to localized territories

of the embryo. The orphan receptor *Sp-reverb* (232, SPU_17492) is expressed in the SMCs delaminating from the tip of the archenteron. The *fxr* genes are implicated in environmental sensing and defense, and it is intriguing that *Sp-fxr* is expressed in the oral part of the ectoderm.

Basic zipper genes

Basic zipper (bzip) transcription factors are long α-helices with DNA sequence-specific basic amino acids in the N-terminal half, and dimerization 'zipper' domains in the C-terminal half. Different basic zipper proteins may form hetero- or homodimers depending on the character of their zipper regions (Vinson et al., 2002). Basic zipper genes can be grouped in to 8 subfamilies on the basis of both their dimerization and DNA recognition domains (Tupler et al., 2001). Comparison of the 14 identified sea urchin bzip factors to the established sets shows that a members of all subfamiles except C/EBP have been identified. Of the 13 bzip genes for which expression data is available, all but 2 have been used in embryogenesis by the 48 h time point.

The expression of two bzip genes can be tentatively assigned to developmental sub-networks on the basis of *in situ* data presented here. *Sp-jun* (5, SPU_03102) is ubiquitous in the very early embryo, but localizes to the PMCs by the time of ingression. The JNK signaling pathway is involved in morphogenesis and cell motility in many settings, including dorsal closure in *Drosophila* and closure of the neural tube in mouse development (Xia and Karin, 2004). Also interesting is the expression of *Sp-hlf* (280,

SPU_04414) in the neurogenic apical ectoderm of the embryo from 18 h. The *hlf* gene is involved in nervous system development in mice (Hitzler et al., 1999).

Sox/hmg genes

Sox/ hmg box genes are minor groove DNA binders that exert their influence on target genes by bending DNA. *Sox* genes are widely expressed in developmental contexts, and indeed 70% are utilized by the late gastrula stage of the sea urchin embryo. The roles of several *sox* family genes in early sea urchin development are already well documented (Kenny et al., 1999; Kenny et al., 2003). Here we report the expression patterns of two additional sox genes, *Sp-soxC* (55, SPU_02603) and *Sp-soxD* (250, SPU_04217). *Sp-soxC* is visible in several territories of the embryo simultaneously, including the blastopore, apical tuft, foregut, and in small patches of ectoderm around the equator of the embryo. The ectodermal and apical expression may indicate a conserved usage of this gene, which is involved in nervous system development in vertebrates (Cheung et al., 2000). *Sp-soxD* is expressed solely in the tip of the gut, in a region overlapping *Sp-soxC* expression.

Smad, t-box, and other transcription factor families

The remaining genes among the newly identified transcription factors all belong to much smaller families in the sea urchin. As a whole this diverse set provides a broad

sample and their very high rate of usage is notable. Of the 55 genes studied, including *smad*, *t-box*, and other genes, 50 are expressed in the developing embryo.

Expression of several of these genes has been mapped to localized parts of the embryo. The Co-smad *Sp-smad4* (25, SPU_04287, SPU_17971) becomes visible at the oral facing tip of the gut by 36 h (fig. 2.8C), though it is unclear with which smad-R is it partnering. Since BMP and TGFβ signaling are involved in specification of the oral and aboral ectoderm during the blastula stage, it can be inferred that *Sp-smad4* also participates in this process (Duboc et al., 2004). Presumably *Sp-smad4* is too diffusely distributed in the embryo at this stage for visible *in situ* staining with our probe. *Sp-tead4* (292, SPU_21210) appears in a very small patch at the tip of the gut at 36 h. Tead/Tef family proteins, also known as *scalloped* in Drosophila, are transcriptional activators. *Sp-dach* (27, SPU_28061) is activated in a band of veg1 cells in the late blastula, and throughout the gut at 36 h. Members of the *ski-sno* family, including *dachshund*, associate with Smad proteins to prevent the antiproliferative effects of TGFβ signaling on cell growth.

The regulome encodes the proteins which directly interpret the genomic *cis*-regulatory instructions for development, and which provide the linkages of gene network architecture. Our knowledge of the repertoire constituting the sea urchin regulome is now close to complete. The functional components of the gene regulatory networks controlling the whole of early development in the sea urchin are now in hand, and the architecture of these networks is accessible to experimental solution.

Acknowledgements

The authors would like to acknowledge Rachel Gray for her invaluable assistance in performing *in situ* hybridization assays; Jongmin Nam for advice and assistance in generating phylogenetic trees; Jed Goldstone and Mark Hahn for their advice on nuclear receptor phylogeny; and Deanna Thomas for her assistance with figures.

References

Amore, G., Yavrouian, R. G., Peterson, K. J., Ransick, A., McClay, D. R., Davidson, E. H., 2003. *Spdeadringer*, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. Dev. Biol. 261**,** 55-81.

Bertrand, S., Brunet, F. G., Escriva, H., Parmentier, G., Laudet, V., Robinson-Rechavi, M., 2004. Evolutionary genomics of the nuclear receptors: From twenty-five ancestral genes to derived endocrine systems. Mol. Biol. Evol. 21**,** 1923-1937.

Bowles, J., Schepers, G., Koopman, P., 2000. Phylogeny of the Sox family of developmental transcription factors based on sequence and structural indicators. Dev. Biol. 227**,** 239-255.

Cheung, M., Abu-Elmagd, M., Clevers, H., Scotting, P. J., 2000. Roles of Sox4 in central nervous system development. Mol. Brain Res. 79**,** 180-191.

Coffman, J. A., Kirchhamer, C. V., Harrington, M. G., Davidson, E. H., 1997. SpMyb functions as an intramodular repressor to regulate spatial expression of *cyIIIa* in sea urchin embryos. Development 124**,** 4717-4727.

Croce, J., Lhomond, G., Lozano, J. C., Gache, C., 2001. *ske-T*, a T-box gene expressed in the skeletogenic mesenchyme lineage of the sea urchin embryo. Mech. Dev. 107**,** 159-162.

Davidson, E. H. 1986. Gene Activity in Early Development. Academic Press, Inc., Orlando.

Duboc, V., Rottinger, E., Besnardeau, L., Lepage, T., 2004. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. Dev. Cell 6**,** 397-410.

Gross, J. M., Peterson, R. E., Wu, S. Y., McClay, D. R., 2003. LvTbx2/3: a T-box family transcription factor involved in formation of the oral/aboral axis of the sea urchin embryo. Development 130**,** 1989-1999.

Hitzler, J. K., Soares, H. D., Drolet, D. W., Inaba, T., O'Connel, S., Rosenfeld, M. G., Morgan, J. I., Look, A. T., 1999. Expression patterns of the hepatic leukemia factor gene in the nervous system of developing and adult mice. Brain Res. 820**,** 1-11.

Howard-Ashby, M., Materna, S., Brown, C. T., Chen, L., Cameron, A. R., Davidson, E. H., 2006. Identification and characterization of homeobox transcription factor genes in *S. purpuratus*, and their expression in embryonic development. Dev. Biol. in press.

Huelguero, E., Cameron, A. R. (2006). Unpublished work.

Imai, K. S., Hino, K., Yagi, K., Satoh, N., Satou, Y., 2004. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131**,** 4047-4058.

Itoh, S., Itoh, F., Goumas, M.-J., Dijke, P. t., 2000. Signaling of transforming growth factor-β family members through Smad proteins. Eur. J. Biochem. 267**,** 6954-6967.

Kenny, A. P., Kozkowski, D. J., Oleksyn, D. W., Angerer, L. M., Angerer, R. C., 1999. SpSoxB1, a maternally encoded transcription factor asymmetrically distributed among early sea urchin blastomeres. Development 126**,** 5473-5483.

Kenny, A. P., Oleksyn, D. W., Newman, L. A., Angerer, R. C., Angerer, L. M., 2003. Tight regulation of SpSoxB factors is required for patterning and morphogenesis in sea urchin embryos. Dev. Biol. 261**,** 412-425.

Kinoshita, K., Kikuchi, Y., Sasakura, Y., Suzuki, M., Fujii-Kuriyama, Y., Sogawa, K., 2004. Altered DNA binding specificity of Arnt by selection of partner bHLH-PAS proteins. Nucleic Acids Res. 32**,** 3169-3179.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5**,** 150-163.

Ledent, V., Paquet, O., Vervoort, M., 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3**,** 1-18.

Lee, C. J., 2002. CIC, a member of a novel subfamily of the HMG-box superfamily, is transiently expressed in developing granule neurons. Mol. Brain Res. 106**,** 151-156.

Lee, P. Y., Davidson, E. H., 2004. Expression of *Spgatae*, the *Strongylocentrotus purpuratus* ortholog of vertebrate GATA4/5/6 factors. Gene Expression Patterns 5**,** 161-165.

Materna, S., Howard-Ashby, M., Gray, R., Davidson, E. H., 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev. Biol. in press.

Minokawa, T., Rast, J. P., Arenas-Mena, C., Franco, C. B., Davidson, E. H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expression Patterns 4**,** 449-456.

Oliveri, P., Carrick, D. M., Davidson, E. H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246**,** 209-228.

Peterson, K. J., Harada, Y., Cameron, R. A., Davidson, E. H., 1999. Expression pattern of Brachyury and Not in the sea urchin: Comparative implications for the origins of mesoderm in the basal deuterostomes. Dev. Biol. 207**,** 419-431.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C., Davidson, E. H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246**,** 132-147.

Reece-Hoyes, J. S., Deplancke, B., Shingles, J., Grove, C. A., Hope, I. A., Walhout, A. J. M., 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 6.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A., Arnone, M. I., 2006. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev. Biol. in press.

Robertson, A. J., Dickey, C. E., McCarthy, J. J., Coffman, J. A., 2002. The expression of SpRunt during sea urchin embryogenesis. Mech. Dev. 117**,** 327-330.

Samanta, M. P., Tongprasit, W., Istrail, S., Cameron, A. R., Tu, Q., Davidson, E. H., Stolc, V., 2006. High resolution transcriptome map of the sea urchin embryo. Dev. Biol. in press.

Tu, Q., Brown, C. T., Davidson, E. H., Oliveri, P., 2006. Sea Urchin Forkhead Gene Family: Phylogeny and Embryonic Expression. Dev. Biol. in press.

Tupler, R., Perini, G., Green, M. R., 2001. Expressing the human genome. Nature 409**,** 832-833.

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., Bonovich, M., 2002. Classification of human B-Zip proteins based on dimerization properties. Mol. Cell. Biol. 22**,** 6321-6335.

Vlahou, A., Gonzalez-Rimbau, M., Flytzanis, C. N., 1996. Maternal mRNA encoding the orphan steroid receptor SpCOUP-TF is localized in sea urchin eggs. Development 122**,** 521-526.

Xia, Y., Karin, M., 2004. The control of cell motility and epithelial morphogenesis by Jun kinases. Trends Cell Biol. 14**,** 94-101.

Yajima, I., Endo, K., Sato, S., Toyoda, R., Wada, H., Shibahara, S., Numakunai, T., Ikeo, K., Gojobori, T., Goding, C. R., Yamamoto, H., 2003. Cloning and functional analysis of ascidian Mitf in vivo: insights into the origin of vertebrate pigment cells. Mech. Dev. 120**,** 1489-1504.

Zeller, R. W., Britten, R. J., Davidson, E. H., 1995. Developmental utilization of Spp3a1 and Spp3a2: Two proteins which recognize the same DNA target site in several sea urchin gene regulatory regions. Dev. Biol. 170**,** 75-82.

Chapter 3

High Regulatory Gene Use in Sea Urchin Embryogenesis: Implications for Bilaterian Development and Evolution.

Meredith Howard-Ashby, Stefan C. Materna, C. Titus Brown, Qiang Tu, Paola Oliveri, R. Andrew Cameron, and Eric H. Davidson

In press, Developmental Biology.

Abstract

A global scan of transcription factor usage in the sea urchin embryo was carried out in the context of the *S. purpuratus* genome sequencing project, and results from six individual studies are here considered. Transcript prevalence data were obtained for over 280 regulatory genes encoding sequence-specific transcription factors of every known family, excluding genes encoding zinc finger factors. This is a statistically inclusive proxy for the total "regulome" of the sea urchin genome. Close to 80% of the regulome is expressed at significant levels by the late gastrula stage. Most regulatory genes must be used repeatedly for different functions as development progresses. An evolutionary implication is that animal complexity at the stage when the regulome first evolved was far simpler than even the last common bilaterian ancestor, and is thus of deep antiquity.

Concepts of the evolutionary origins of bilaterian animals have been transformed by the results of genome sequencing. A most important result is that all bilaterian animals share a common qualitative repertoire of genes encoding sequence-specific transcription factors and signaling system genes. These are the essential constituents of the developmental gene regulatory networks that underlie development of the body plan. The concept of a bilaterian "regulatory tool kit" is now firmly established (Davidson, 2006; Erwin and Davidson, 2002), and the evidence from the new sea urchin genome sequence provides much further support (The Sea Urchin Sequencing Consortium, 2006). Every developmentally utilized signaling system, and with almost no exceptions, every subfamily of every class of transcription factor found in vertebrates and ecdysozoans is also represented in this nonchordate deuterostome genome as well. Essentially the main and sometimes only differences in the regulatory tool kits of bilaterian genomes are in the multiplicity of members of given gene subfamilies. Cnidarians as well share at least a large fraction of this same tool kit (Martindale et al., 2004; Seipel and Schmid, 2005). These are also complex animals, however, which are more similar to bilaterians than once thought, and in geologic time they may have diverged from the bilaterian stem lineage not long before the bilaterians themselves diversified (Peterson et al., 2004). The existence of a shared bilaterian regulatory gene tool kit brings into focus the following question: did the regulatory tool kit, the "regulome," evolve concomitantly with the complex adult body plans of bilaterians (or of cnidarians/ bilaterians)? This would allow the hypothesis that the evolutionary assembly of the tool kit repertoire per se might have been causal with respect to the appearance of animals of the bilaterian grade of morphological complexity. Or, did the regulome predate complex animal forms? This

allows the alternative hypothesis that bilaterian evolution followed from increasingly elegant modes of tool kit utilization, rather than invention and qualitative diversification of the tool kit itself. In mechanistic terms these alternatives at root amount to evolution of animal complexity driven mainly by the appearance of new genes, vs. evolution of animal complexity driven mainly by appearance of new regulatory linkages among preexisting genes.

The sea urchin genome sequence provides a unique opportunity to address this issue. This is the only genome so far sequenced from an organism that utilizes maximum indirect development (Peterson et al., 1997). Here the primary role of the embryo is to produce a larva, which provides a life support system for the postembryonic development of the adult body plan. The body parts of the adult form later develop within the larva, from cell populations that had been set aside from embryological specification and differentiation process. In direct development, on the other hand, the primary object of embryogenesis is construction of the adult plan as immediately as possible. The embryo/larva of indirectly developing form may possess very little similarity to the adult body plan, and are typically far simpler in structure and complexity than any adult bilaterian body plan. Morphological simplicity is an obvious character of the *S. purpuratus* embryo (fig. 3.1). Thus, in contrast to all adult bilaterian forms and all directly developing bilaterian embryos, the sea urchin embryo consists exclusively of single cell thick epithelial layers, and individual mesenchymal cells. It has no mesodermal tissue layers, nor organs, nor body parts formed from mesoderm plus ectoderm or endoderm.

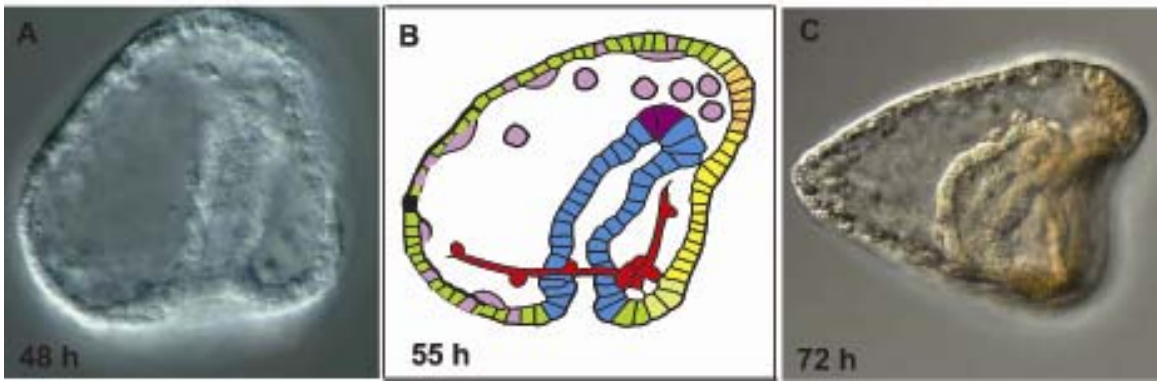


Figure 3.1. Simple morphological and regulatory diversification of the sea urchin embryo. A. Late gastrula, stage at which observations in this chapter end. B. Regulatory complexity of a slightly later embryo, about 800 cells, indicated by the color-coded regulatory states: red, skeletogenic cells; blue, gut endoderm cells, including incipient hindgut, midgut, and foregut; violet, mesenchymal mesodermal cell types, including pigment cells, blastocoelar cells, coelomic pouch cells; yellow, oral ectoderm; orange, neurogenic apical domain; green, aboral ectoderm. The oral ectoderm has several diverse incipient territories within it, including neurogenic ciliated band, stomodaeal, "facial" ectoderm, while the aboral ectoderm is homogeneous. C. Completed embryo/larva able to feed and exist independently in the water column, for comparison; about 1500 cells.

Regulome utilization in embryogenesis

In the course of the *S. purpuratus* genome project all genes encoding recognizable transcription factors were identified and annotated, and their expression during embryonic development was measured quantitatively. Here we have tabulated these gene expression data and reduced them to a common format for analysis. Included are the *forkhead* genes (Tu et al., 2006), the *ets* genes (Rizzo et al., 2006), the *hox* and *parahox* genes (Arnone et al., 2006), all other *homeobox* genes (Howard-Ashby et al., 2006b), the *nuclear hormone receptor* genes, *bhlh*, *smad*, *tbox*, *basic zipper*, and *sox* transcription factor genes, as well as members of other smaller regulatory gene families (Howard-

Ashby et al., 2006a). In addition, prior knowledge was incorporated, particularly the large number of regulatory genes encompassed in the endomesoderm gene regulatory network for *S. purpurarus*.(Davidson, 2006; Levine and Davidson, 2005) Given the genome-wide gene prediction analysis (The Sea Urchin Sequencing Consortium, 2006) and the concordance of an entirely independent search for regulatory genes(Howard-Ashby et al., 2006b), most DNA-binding transcription factors of known families have been identified, except for Zn finger genes. At the very least, the 283 genes included here represent a very large, unbiased sampling of all genes encoding transcription factors in the *S. purpuratus* genome.

Zinc finger genes were specifically excluded because it is difficult at present to generate a comparable high confidence gene set from this class of genomic sequences. Zinc finger motifs have proven difficult to group into subfamilies and to analyze phylogenetically (Knight and Shimeld, 2001). For most genes that encode C_2H_2 Zn finger domains it is impossible to identify clear orthologues known to function as regulatory genes in other species, or even to know whether all such domains identified in the genome have been correctly included in gene models. It is often unclear whether given domains represent splice variants, distinct genes, or assembly errors. Another difficulty is that not all C_2H_2 zinc finger proteins are transcription factors, as proteins including these domains have been demonstrated to function in RNA binding and in protein-protein interactions (Laity et al., 2001; Lu et al., 2003). Illustrating this uncertainty, of the approximately 380 C_2H_2 Zn finger genes identified in *S. purpuratus*, nearly 40 have only one zinc finger domain (Materna et al., 2006), but least two such domains are required for DNA binding specificity. A comprehensive set of true and unique zinc finger regulatory

genes cannot be defined on the basis of genomic sequence and expression data alone. In contrast, identification of most other classes of DNA binding domain in the regulome is unequivocal, given their high conservation and clear orthology across the Bilateria. We therefore took genes encoding all DNA sequence specific transcription factors other than zinc finger factors to be representative of the total regulome, and considered their deployment in embryonic development

Quantitative PCR (QPCR) was used to determine the expression profile of each of the 283 regulatory genes, from fertilization to 48 h post-fertilization (Howard-Ashby et al., 2006a, b; Rizzo et al., 2006; Tu et al., 2006). In addition the spatial patterns of expression were determined for all genes expressed sufficiently to permit *in situ* hybridization (>5-10 copies per cell). The number of regulatory genes in each transcription factor family expressed only maternally; expressed maternally and zygotically at constant levels; activated zygotically during embryogenesis; or remaining silent or expressed at extremely low, insignificant levels by 48 h is collated in table 3.1. The threshold of significant expression was set, conservatively, at 150-350 molecules of mRNA per embryo. From late cleavage onward in the sea urchin embryo the populations expressing given regulatory states are all at least 16 cells, and by gastrula stage the largest territories are 60-200 cells. Thus at 350 mRNAs per embryo there would be 2-20 mRNAs per cell for territorially specific messages. In these embryos the rate of translation is two molecules of protein/mRNA-min (18), and so within a few hours these threshold mRNA concentrations suffice for production of the several hundred to few thousand molecules of transcription factor per cell required for significant target site occupancy (Bolouri and Davidson, 2003; Calzone et al., 1988). Studies on expression of functional genes in the

endomesoderm network show that functionally essential regulatory gene transcript concentrations range from a few to only about 40 molecules of mRNA per cell. The 350 molecule per embryo threshold thus represents a functional level of expression, though close to a minimal one. In any case, however, the great majority of the mRNAs with which we are here concerned are present either at >1,000 molecules or 0-10 molecules per embryo.

Table 3.1. Regulome usage in development by gene family

Family	**Total**	**M**	**Z**	**C**	**-**	**% exp**	**Localized expression**	**Ubiq.[c]**
hox cluster	11	0	2	0	9	**18.2**	2	0
homeobox	85	0	58	3	24	**71.8**	24	4
T-Box	6	0	5	0	1	**83.3**	3	0
smad	4	0	4	0	0	**100**	1	1
forkhead	22	1	20	0	1	**95.5**	20	0
Sox/HMG	10	1	5	2	2	**80.0**	3	0
bHLH	48[a]	0	24	2	17	**59.5**	5	2
Ets	11	0	10	0	1	**90.9**	4	4
bZip	14[a]	0	9	2	2	**84.6**	2	1
nuclear receptor	33	0	22	1	10	**69.7**	4	0
other types	45	1	37	4	3	**93.3**	10	6
all genes	**283**	**3**	**196**	**14**	**70**	**75.3/ 77.6[b]**	**76**	**18**

[a]No expression data is reported for 5 bHLH genes and one bZip gene.
[b]Statistic is recalculated omitting the hox cluster genes.
[c]Only genes with sufficient expression to likely be detectable were examined by *in situ* hybridization.

The majority of all regulatory genes in the sample have been activated by late gastrula stage. More than 80% of members of the *forkhead*, *ets, bZip*, *smad*, *sox*, and many other families are utilized in the embryo by 48 h post-fertilization (table 3.1). The largest family, the non-*hox homeobox* genes, are >70% expressed by late gastrula. Only

the *nuclear receptor* and *bHLH* families are expressed at somewhat lower levels, but the majority of even these have been activated by 48 h. The *hox* genes are a special case. As predicted (Davidson, 1990) and later experimentally demonstrated (Arenas-Mena et al., 1998), the *hox* cluster as such is not utilized until formation of the adult body plan in postembryonic sea urchin development (Arenas-Mena et al., 2000). Only two of the 11 *hox* cluster genes are expressed during embryogenesis. Since the *hox* cluster is utilized as a functional unit, expression of individual *hox* genes cannot be considered as statistically independent events. Overall, 75% of the regulome has already been used at least once by late gastrula stage, when development of this embryo is only two-thirds complete. If the *hox* genes are removed from the calculation, the fraction rises to 77% by 48 h. The cumulative time course of regulome use is plotted in figure 3.2 (green and blue curves).

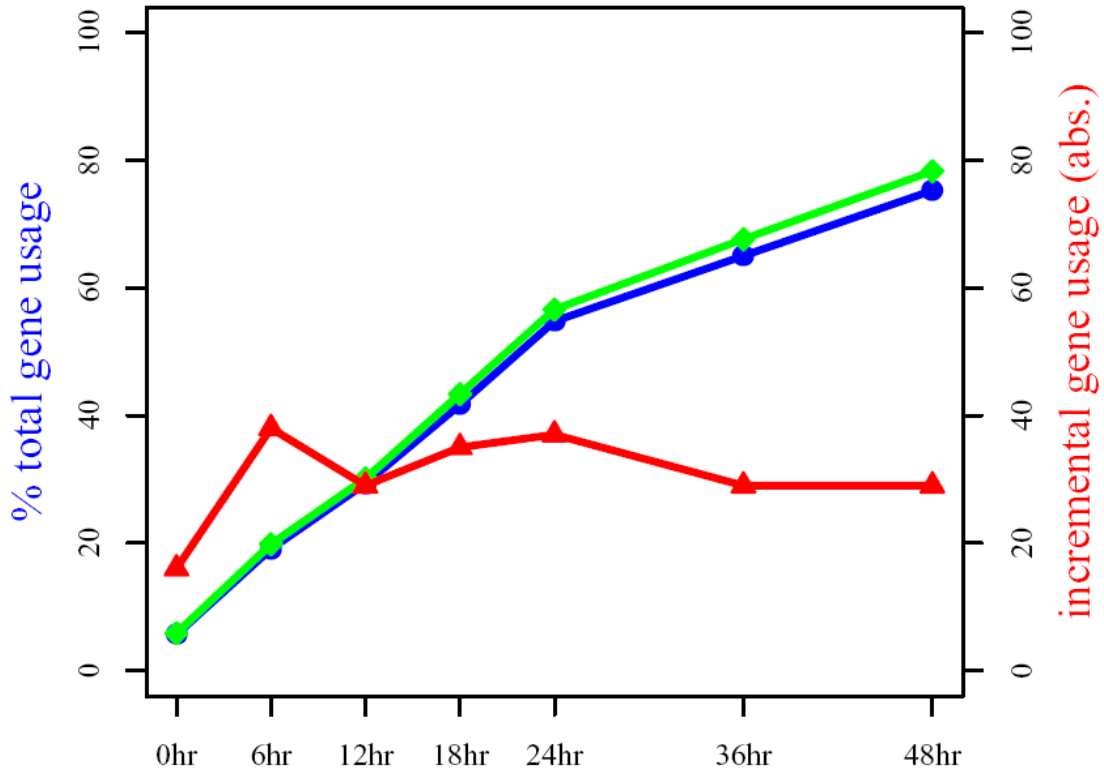


Figure 3.2. Regulatory gene usage in development. Regulome usage is plotted as a function of developmental time. Data were collated from references cited in text. A total of 283 regulatory genes is included in the analysis. The threshold for biological significance was set at 150-350 copies per embryo (see text). Genes were classified as first activated zygotically at 6, 12, 18, 24, 36, or 48 h postfertilization; or not expressed significantly by late gastrula stage. Genes expressed only maternally or at a constant level including maternal expression are included at the 0 h time point. The blue curve is the percentage of all regulatory genes which have been zygotically expressed by the given time after fertilization. The green curve is the same discounting the genes of the *hox* complex (see text). The red line (right ordinate) indicates the number of regulatory genes newly activated in each time interval. Transcript levels in each cDNA sample were measured by comparing the QPCR amplification of the target sequence to that of a standard of known concentration in cDNA prepared from embryos of the appropriate stage (cf. primary references for details). A fluorescent reporter dye is used to measure the increasing concentration of the unknown and standard amplicons at the end of every PCR cycle. If the copy number of the standard is known, given that each PCR cycle produces an amplification of approximately 1.9-fold, the embryonic copy number of the unknown can be calculated from the difference in cycle numbers needed to produce an arbitrary fluorescent signal between standard and unknown (see materials and methods in chapter 1). Ubiquitin, which is present at the same concentration at all developmental time points, rRNA, and other constant sequences were used as the internal standards. Data from the *S. purpuratus* embryonic transcriptome analysis (Samanta et al., 2006) were used for external validation of whether individual genes were truly expressed. For some genes, a slightly different set of time points was used, and the expression at the above time points was extrapolated.

New transcription factors are activated steadily during development (red line in figure 3.2, essentially the experimentally measured derivative of the blue line). Every regulatory gene can be thought of as a node in the gene regulatory network which reads, processes, and transmits spatial and temporal information (Davidson, 2006). A given gene is activated when the correct set of upstream inputs is presented, and the resulting regulatory protein conveys new spatial and temporal cues when it interacts with its *cis*-regulatory targets in downstream genes. Thus figure 3.2 shows that new information processing nodes are being activated continuously, with concomitant increase in the regulatory complexity of the embryo, even though this is yet not apparent

morphologically (fig. 3.1A, B). If the integral percent usage plot is projected forward to 72 h when embryogenesis is complete and the larva becomes capable of feeding (fig. 3.1C), 95% of the regulome will have been used at least once. Measurements on the forkhead transcription factor family did extend out to 72 h (Tu et al., 2006), and indeed 95.5% of these factors are in play by then.

Why is early development so expensive in regulatory apparatus?

The complexity of the regulatory apparatus required to execute a given developmental process is a system level property, which can only be interpreted accurately by means of a system level functional analysis. The endomesoderm gene regulatory network established for this sea urchin species is such an analysis, and it displays the specific roles of over 40 different transcription factors (Davidson, 2006; Davidson et al., 2002; Howard and Davidson, 2004; Levine and Davidson, 2005; Oliveri and Davidson, 2004). This network pertains to only part of the embryo, and to only about half of the developmental period from fertilization to late gastrula. It covers the period from about 6 h after fertilization, when spatially confined zygotic regulatory gene expression begins to dominate the developmental process, to mesenchyme blastula stage. At this point the whole embryo has achieved territorial specification, that is, specific regulatory states have been established in all its territories, but gastrulation has not yet taken place. The endomesoderm network includes the specification of skeletogenic and other mesodermal precursors and of gut endoderm, but it excludes the aboral and oral ectodermal territories, and also the neurogenic apical territory. Between mesenchyme

blastula stage and late gastrula much additional development occurs, including the subdivision of the archenteron into fore-, mid- and hind-gut, and of the oral ectoderm into stomodaeal, lateral and ciliary band subdomains, and the 48 h embryo has significantly more diverse parts than it does at mesenchyme blastula stage. Furthermore, it is a "driver gene network", i.e., it is focused on regulatory genes that are expressed in spatially or temporally specific ways, since these are the regulatory genes that must execute the control logic which specifies cells differentially in space and time.(Davidson, 2006; Yuh et al., 2001) However, ubiquitous regulatory factors that are also necessary for the normal operation of developmentally active *cis*-regulatory modules, as shown explicitly for the *endo16* control system (Yuh et al., 2001; Yuh et al., 2005), and these are not systematically represented in the endomesoderm network. Despite these limitations in coverage, the endomesoderm gene regulatory network includes > 40 sequence specific regulatory genes.

Specific aspects of regulatory gene usage in the sea urchin endomesoderm network, and in other developmental gene regulatory networks (Koide et al., 2005; Loose and Patient, 2004; Stathopoulos and Levine, 2005), illuminate the need for large regulatory apparatus in embryonic development. First, if a regulatory gene is expressed, it will have a function. If its expression is blocked the expression of downstream genes will be affected and therefore the fractions of regulatory genes expressed as shown in figure 3.2 are likely to be directly meaningful. Second, individual regulatory genes at the nodes of developmental gene regulatory networks respond to unique sets of inputs, and the outputs they send onwards have unique sets of destinations; i.e., no two nodes do the same things. Therefore the number of nodes represents the number of *cis*-regulatory input

information processing units the network must encompass. This number is never small. Third, individual developmental jobs the network mediates are each performed by modular subcircuits not used elsewhere in that spatial and temporal stage of development, every one of which consists of several regulatory genes. Such jobs include specification of given territories, such as the prospective skeletogenic or gut territory; or operation of given differentiation gene batteries. The endomesoderm network includes many such subcircuits because there are many such jobs to be done.

In short, developmental gene regulatory networks provide a basis for comprehending the high usage of regulatory genes in development. With respect to the sea urchin embryo, the endomesoderm network by itself would predict by extrapolation to the whole embryo at 48 h, a quantitative requirement for regulatory gene usage consistent with that shown in figure 3.2.

The regulome in development

It is a commonplace that genes encoding given transcription factors are utilized in multiple times and places during the development of an organism, participating in entirely independent processes. Even within the three days required for sea urchin embryogenesis, many specific regulatory genes have been found to be expressed in a succession of diverse domains where they execute distinct and unrelated functions. For example, the *hnf6* gene is initially expressed ubiquitously when it has targets in many parts of the embryo, then it becomes an oral ectoderm regulator, and later is required specifically in ciliated band (Otim et al., 2004); the *deadringer* gene and the *goosecoid*

genes are first utilized in skeletogenic cells and later in oral ectoderm (Amore et al., 2003; Angerer et al., 2001); the diverse regulatory modules of the *otx* gene drive expression in many different domains of the embryo (Yuh et al., 2002); the "early" and "late" modules of the *blimp1/krox* gene respectively control a dynamic pattern of expression in cleavage stage endomesoderm, and later contribute to a dedicated midgut/hindgut regulatory state in the invaginated archenteron (Livi and Davidson, 2006).

Here we see that repeated reutilization must indeed be the overwhelming majority pattern of regulatory gene utilization. This implication follows directly from the finding that most regulatory genes are required for development just to the late gastrula stage. The embryo itself will become significantly more complex after this stage, with the elaboration of its nervous system, the development of the stomodaeum, the ciliated band, the coelomic pouches, the tripartite gut, and so forth. But the development of the adult body plan in postembryonic development dwarfs the whole of the embryonic process in the complexity of its multilayered morphology, and its numerous new cell types. The regulome from which are constituted the many developmental gene regulatory networks required to organize adult body plan development must be the same regulome required to make the gastrula, for there is no more, save the 20% - 25% of regulatory genes not yet deployed by this stage. Some of the regulatory genes not used in the embryo up to gastrula stage have specific roles. For example, a cohort of these genes is expressed specifically in oogenesis (Song et al., 2006); and most of the genes of the *hox* complex are silent until activation in the course of formation of the adult body plan in postembryonic larval development (Arenas-Mena et al., 2000). What is perhaps

unexpected is that such a small fraction of the regulome is dedicated to such "special purposes."

The conclusions, then, are that even simple territorial specification functions require complex networks of many genes of multiple transcription factor families; and that more complex later development is driven by recursive utilization of the same regulatory genes. These same conclusions must inform consideration of early animal evolution as well.

The regulome in evolution

A "minimalist" interpretation of the last common bilaterian ancestor, based on the logic of incontrovertibly shared characters, provides an image of a creature much simpler in morphological organization than any modern bilaterian. It must have had a tripartite through gut, bilateral anterior/ posterior nervous system organization, organ grade internal body parts perhaps including heart (Erwin and Davidson, 2002), and mesodermal layers, used both as major functional and structural components of the body and for developmental signaling interactions with endodermal and ectodermal layers. But such an organism would have been very significantly more complex than embryos or larvae of animals such as the sea urchin: these have neither organ level structures nor mesodermal layers, only a few types of free-wandering mesodermal cells and some muscular sphincters in the gut. Such larvae do possess bilateral anterior/posterior organization and tripartite gut with mouth and anus. Because it had very significantly more diverse morphology, the last common bilaterian ancestor must necessarily have required for its

development a more extensive and elaborated genomic regulatory apparatus, more and deeper networks of regulatory gene interactions encoded in its genome, than does the embryonic phase of modern indirect development.

The palaeontological record of bilaterian origins is famously enigmatic, though in recent years valuable clues have accumulated. Molecular phylogeny based on calibrated protein divergence rates across the Bilateria indicate that bilaterian divergence from a common ancestral lineage probably occurred after the Marinoan Glaciation (Aris-Brosou and Yang, 2003; Douzery et al., 2004; Peterson et al., 2004); the last of the world wide snowball earth episodes which ended about 630 mya, i.e., 70 million years before the beginning of the Cambrian (Peterson and Butterfield, 2005). A variegated assemblage of microfossils from Southwest China dating to about 590 mya, includes a large variety of eggs and embryos that have earmarks of bilaterian forms, such as distinctive patterns of unequal cleavage (Chen et al., 2006; Chen et al., 2000; Dornbos et al., 2005; Xiao and Knoll, 1999). Among these microfossils is a complex, unusually well preserved form that has unmistakable bilaterian structural features (Chen et al., 2004). Later on, by 10 or 15 million years before the beginning of the Cambrian at 542 mya, there appear trace fossils, bore holes in the benthic deposits that were undoubtedly made by bilaterian animals (Bottjer et al., 2000), and also the first macroscopic bilaterian body fossils, such as the complex, mollusk-like *Kimberella* (Fedonkin and Waggoner, 1997).

What was the nature of the Precambrian genomic landscape in which the Bilateria originated; how complex was it? In terms of cellular organization, the simplest current free living bilaterian forms, the larvae of maximally indirectly developing animals, lack distinctive features of the last common bilaterian ancestor and are much less complicated.

It is here entirely irrelevant whether the gene regulatory networks directing the development of such larval forms are themselves evolutionary "simplifications" adaptively derived for the ecological conditions of larval life; or on the other hand, are plesiomorphic survivals of early evolving gene regulatory networks for generation of simple organisms. For, the evidence in figure 3.2 shows that the large majority of the shared bilaterian regulome is required for the mechanism of development of the mere gastrula of an indirectly developing animal. It follows that the development of forms much simpler than the last common bilaterian ancestor must still have required most of the current bilaterian regulome. Therefore, the bilaterian regulome considered in figure 3.2 is thus at least of Upper Neoproterozoic antiquity.

There is yet no evidence as to how deep in time evolutionary assembly of the regulome occurred, or what was the morphology of the form for the development of which it was deployed. If there was an evolutionary stage when the developmental (organismal) complexity of bilaterian ancestors was driven by the assembly of the regulatory tool kit, it was at a remote period, preceding the last common bilaterian ancestor. Ever since, the evolution of animal form has depended mainly on endless reutilization of the same regulome. This of course means endless reorganization of the genomic regulatory apparatus controlling regulatory gene use; primarily evolution of gene regulatory pathways, not evolution of new kinds of regulatory genes.

References

Amore, G., Yavrouian, R. G., Peterson, K. J., Ransick, A., McClay, D. R., Davidson, E. H., 2003. *Spdeadringer*, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. Dev. Biol. 261**,** 55-81.

Angerer, L. M., Oleksyn, D. W., Levine, A. M., Li, X. T., Klein, W. H., Angerer, R. C., 2001. Sea urchin goosecoid function links fate specification along the animal-vegetal and oral-aboral embryonic axes. Development 128**,** 4393-4404.

Arenas-Mena, C., Cameron, A. R., Davidson, E. H., 2000. Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. Development 127**,** 4631-4643.

Arenas-Mena, C., Martinez, P., Cameron, R. A., Davidson, E. H., 1998. Expression of the Hox gene complex in the indirect development of a sea urchin. Proc. Natl. Acad. Sci. U. S. A. 95**,** 13062-13067.

Aris-Brosou, S., Yang, Z. H., 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. Mol. Biol. Evol. 20**,** 1947-1954.

Arnone, M. I., Rizzo, F., Annunciata, R., Cameron, R. A., Peterson, K. J., 2006. Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: insights into the relationship between clustering and colinearity. Dev. Biol. in press.

Bolouri, H., Davidson, E. H., 2003. Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. Proc. Natl. Acad. Sci. U. S. A. 100**,** 9371-9376.

Bottjer, D. J., Hagadorn, J. W., Dornbos, S. Q., 2000. The Cambrian Substrate Revolution. GSA Today 10**,** 1-7.

Calzone, F. J., Theze, N., Thiebaud, P., Hill, R. L., Britten, R. J., Davidson, E. H., 1988. Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. Genes Dev. 2**,** 1074-1088.

Chen, J. Y., Bottjer, D. J., Davidson, E. H., Dornbos, S. Q., Gao, X., Yang, Y. H., Li, C. W., Li, G., Wang, X. Q., Xian, D. C., Wu, H. J., Hwu, Y. K., Tafforeau, P., 2006. Phosphatized polar lobe-forming embryos from the Precambrian of southwest China. Science 312**,** 1644-1646.

Chen, J. Y., Bottjer, D. J., Oliveri, P., Dornbos, S. Q., Gao, F., Ruffins, S., Chi, H. M., Li, C. W., Davidson, E. H., 2004. Small bilaterian fossils from 40 to 55 million years before the Cambrian. Science 305**,** 218-222.

Chen, J. Y., Oliveri, P., Li, C. W., Zhou, G. Q., Gao, F., Hagadorn, J. W., Peterson, K. J., Davidson, E. H., 2000. Precambrian animal diversity: Putative phosphatized embryos from the Doushantuo formation of China. Proc. Natl. Acad. Sci. U. S. A. 97**,** 4457-4462.

Davidson, E. 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, Orlando, FL.

Davidson, E. H., 1990. How embryos work - a comparative view of diverse modes of cell fate specification. Development 108**,** 365-389.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science 295**,** 1669-1678.

Dornbos, S. Q., Bottjer, D. J., Chen, J. Y., Oliveri, P., Gao, F., Li, C. W., 2005. Precambrian animal life: Taphonomy of phosphatized metazoan embryos from southwest China. Lethaia 38**,** 101-109.

Douzery, E. J. P., Snell, E. A., Bapteste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. U. S. A. 101**,** 15386-15391.

Erwin, D. H., Davidson, E. H., 2002. The last common bilaterian ancestor. Development 129**,** 3021-3032.

Fedonkin, M. A., Waggoner, B. M., 1997. The Late Precambrian fossil Kimberella is a mollusc-like bilaterian organism. Nature 388**,** 868-871.

Howard-Ashby, M., Materna, S., Brown, C. T., Chen, L., Cameron, A. R., Davidson, E. H., 2006a. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev. Biol. in press.

Howard-Ashby, M., Materna, S., Brown, C. T., Chen, L., Cameron, A. R., Davidson, E. H., 2006b. Identification and characterization of homeobox transcription factor genes in *S. purpuratus*, and their expression in embryonic development. Dev. Biol. in press.

Howard, M. L., Davidson, E. H., 2004. *cis*-Regulatory control circuits in development. Dev. Biol. 271**,** 109-118.

Knight, R. D., Shimeld, S. M., 2001. Identification of conserved C2H2 zinc-finger gene families in the Bilateria. Genome Biol. 2**,** 0016.0011-0016.0018.

Koide, T., Hayata, T., Cho, K. W. Y., 2005. *Xenopus* as a model system to study transcriptional regulatory networks. Proc. Natl. Acad. Sci. U. S. A. 102**,** 4943-4948.

Laity, J. H., Lee, B. M., Wright, P. E., 2001. Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. 11**,** 39-46.

Levine, M., Davidson, E. H., 2005. Gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. 102**,** 4936-4942.

Livi, C. B., Davidson, E. H., 2006. Expression and function of *blimp1/krox*, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. Dev. Biol. 293**,** 513-525.

Loose, M., Patient, R., 2004. A genetic regulatory network for *Xenopus* mesendoderm formation. Dev. Biol. 271**,** 467-478.

Lu, D., Searles, M. A., Klug, A., 2003. Crystal structure of a zinc-finger - RNA complex reveals two modes of molecular recognition. Nature 426**,** 96-100.

Martindale, M. Q., Pang, K., Finnerty, J. R., 2004. Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea

anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). Development 131**,** 2463-2474.

Materna, S., Ashby-Howard, M., Gray, R., Davidson, E. H., 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev. Biol. in press.

Oliveri, P., Davidson, E. H., 2004. Gene regulatory network controlling embryonic specification in the sea urchin. Curr. Opin. Genet. Dev. 14**,** 351-360.

Otim, O., Amore, G., Minokawa, T., McClay, D. R., Davidson, E. H., 2004. SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. Dev. Biol. 273**,** 226-243.

Peterson, K. J., Butterfield, N. J., 2005. Origin of the Eumetazoa: Testing ecological predictions of molecular clocks against the Proterozoic fossil record. Proc. Natl. Acad. Sci. U. S. A. 102**,** 9547-9552.

Peterson, K. J., Cameron, R. A., Davidson, E. H., 1997. Set-aside cells in maximal indirect development: Evolutionary and developmental significance. Bioessays 19**,** 623-631.

Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J., McPeek, M. A., 2004. Estimating metazoan divergence times with a molecular clock. Proc. Natl. Acad. Sci. U. S. A. 101**,** 6536-6541.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A., Arnone, M. I., 2006. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev. Biol. in press.

Samanta, M. P., Tongprasit, W., Istrail, S., Cameron, A. R., Tu, Q., Davidson, E. H., Stolc, V., 2006. High resolution transcriptome map of the sea urchin embryo. Dev. Biol. in press.

Seipel, K., Schmid, V., 2005. Evolution of striated muscle: Jellyfish and the origin of triploblasty. Dev. Biol. 282**,** 14-26.

Song, J. L., Wong, J. L., Wessel, G., 2006. Oogenesis: Single cell development and differentiation. Dev. Biol. in press.

Stathopoulos, A., Levine, M., 2005. Genomic regulatory networks and animal development. Developmental Cell 9**,** 449-462.

The Sea Urchin Sequencing Consortium, 2006. The genome sequence of the purple sea urchin, *Strongylocentrotus purpuratus*. Science in press.

Tu, Q., Brown, C. T., Davidson, E. H., Oliveri, P., 2006. Sea urchin forkhead gene family: phylogeny and embryonic expression. Dev. Biol. in press.

Xiao, S. H., Knoll, A. H., 1999. Fossil preservation in the Neoproterozoic Doushantuo phosphorite Lagerstatte, South China. Lethaia 32**,** 219-240.

Yuh, C. H., Bolouri, H., Davidson, E. H., 2001. *Cis*-regulatory logic in the *endo16* gene: switching from a specification to a differentiation mode of control. Development 128**,** 617-629.

Yuh, C. H., Brown, C. T., Livi, C. B., Rowen, L., Clarke, P. J. C., Davidson, E. H., 2002. Patchy interspecific sequence similarities efficiently identify positive *cis*-regulatory elements in the sea urchin. Dev. Biol. 246**,** 148-161.

Yuh, C. H., Dorman, E. R., Davidson, E. H., 2005. Brn1/2/4, the predicted midgut regulator of the *endo16* gene of the sea urchin embryo. Dev. Biol. 281**,** 286-298.

Appendix 1: Supplementary Material for Chapter 1.

I. QPCR primers

2-F-Six3/6 AAGAGAGAACGCGGAGTTTG
2-R-Six3/6 AAACCAGTTTCCGACCTGTG

12-en-F CGTCCCAACTCCAGAGACTG
12-en-R TTTTGATCTGGGATTCGCTC

15-F-Six1/2 TGAGCATCTCCACAAGAATGA
15-R-Six1/2 GGTGGTTATGCGGTGAGAAG

16-F-Pax AAGATCCTGGCCCGGTATAA
16-R-Pax GGATCCCGCTGCTTGTATT

18-F-Brn3 CCGCTAGGATGGGTTTAAGA
18-R-Brn3 AAAGCTCTGGCAAACCTGAA

21-F-Six4 CCCTGCCTTCCATACAGAAC
21-R-Six4 TGGGTGCTTTCTCCTAATGC

26-F-Oct1 CACAAGCATCGAGACCAACA
26-R-Oct1 TGTCGACGGTTACAAAACCA

32-F-isl1 CTTGGAGGTCTGCTGATCG
32-R-isl1 ACTGGTCGAGATGACGCAAT

43-tgif-F GCTCTACCTATCTCGCTTGGC
43-tgif-R TGGTGAACTTGTCAGGGTCT

44-F-Lim1 ACAAAGTGCGCAGGATGTCT
44-R-Lim1 CCCGTAGAGAGTTGCTTTCG

45-F-hox9.10 TCGGGTGAGGTACATGTTGA
45-R-hox9.10 AAACAGAAGAACGGGGACAG

47-pax258-F CCAAAGGTGGTGTCGAAGAT
47-pax258-R ATCGAGCTGACACTGGGAAC

50.1-hbox4.5-F
CCAAACGCTCGAACTAGAGAA
50.1-hbox4.5-R CATCCTCCTGTTTTGGAACC

50.2-F-Hbox8 CCACTTCATCCGTCGATTCT
50.2-R-Hbox8 GCTGGAGCTGAGAAAGGAGTT

56-F-hnf1 CGTGCCCTTATTCAAATGCT
56-R-hnf1 CCATGGCAAGTAGCGAAGAT

68-exd-F GACAACATGCTCATAGCCGA
68-exd-R GTTCAATGGCGTTCTCAGGT

74-F-msx2 AGCACAAGACAAACCGGAAG
74-R-msx2 CGTTCGGCTATCGAGAGGTA

75.1-F-nkx2.2 ACACTTGGCGAGCATTATCC
75.1-R-nkx2.2 CGGAGAAGGTAACGGATTCA

78-F-ATBF1 ATGCAATCATTCATGGCTCA
78-R-ATBF1 TGTGTTTGGATGAGGAGCAG

81-smadIP-F TTTACAGACTGCAGCGTCACA
81-smadIPR TGTGAATACGCAAGTGCTCC

84-pitx3-F GAAACGAGAGCGGAATCAAA
84-pitx3-R CTGCTGCCCAGTTGTTGTAG

85-hox1.tlx1-F CATTCACCCGACTCCAAATC
85-hox1.tlx1-R GTCTTCACCTGAGCATCCGT

95-lmo4-F CTGTACGATGTTGCGCGG
95-lmo4-R TGTCCTATGTCACCGAGTTG

105-lhx3-F ATACTGAAGGTGGTGGACCG
105-lhx3-R CTTCCTTGCAGAAAACACCG

108-F-paxC GAGCAAACCCCGTGTGTC
108-R-paxC GAGCAGTCGTTCTCGGATCT

109-F-mox GCTCGACCTAATAGCCAAACC
109-R-mox TGACAGCGATCTCGTATCGT

115-F-lbx CTTCGAACTGGAACGTCGAT
115-R-lbx CCCTCCTGTTCTGAAACCAC

122-F-awh GATCCGCACCACTTTCACC
122-R-awh CCGTGAGTTTTGAAACCAGA

Q127-nk6.1-F,CGCCAGTGTCATCTTCGTCT
Q127-nk6.1-R,GGCCAGGTATTTGGTCTGTT

146-F-chx10
TCAACTTGATGAATTAGAGAAATCG
146-R-chx10 TTACCTGTATTCTGTCCTCGGGTA

150-F-emx CTTCAGGAAACCCAAGAGGA
150-R-emx AGACTCGCTGCCAGTTGTTT

151-rx-F GCTGTCGATCATGGAAGATG
151-rx-R AATGCCCTCTCAAGTTCGTG

163-F-pitx1 CCAACGACGACAGAGAACAC
163-R-pitx1 GGTTCCGTTAAACTCGTCCA

184-F-alx4 CCAACTCGAGGAGATGGAGA
184-R-alx4 TGTACACGAGCTTCGGTCAG

200-irxA-F TATGGAATGGACCTGAACGG
200-irxA-R TATGATCTTTTCGCCCTTGG

206-lmo4-F TGGTCATGAGGACCCAGAAT
206-lmo4-R GTCGTTCTCGCAGACGATG

253-hox3-F TCGAGCTGGAAAAGGAGTTT
253-hox3-R TTTCATTCTTCGATTCTGAAACC

254-hox6-F AGGAGTTCCACTTCAGCCGT
254-hox6-R CGTGTTCCCTCTTCCATTTC

255-hox7-F GGCAGACTTACACCCGCTAC
255-hox7-R TCTGTCGTTCTGTCAATCCG

256-hox11'13b-F
CGAACTAGAGAAGGAGTTCACAA
256-hox11'13b-R TCTTCATTCGCCTGTTCTGG

257-evx-F GGTACCGCACCGCATTTAC
257-evx-R GGTTTCTGGCAGGTTAAGAGC

258-hb9-F ATCCTTGGAAAGACACGGC
258-hb9-R GGGATGTAGCCACTTCGAATC

259-barh1-F
GCATTCACAGATCATCAACTCAA
259-barh1-R CCATGTCTTGACCTGTGTGTC

260-barx-F CTCTCTACCCCTGATCGGTT
260-barx-R TTCCATTTCATTCTTCTGTTTTGA

261-dbx1-F GGCTGTATTCTCCGATGCTC
261-dbx1-R ACCTGTGAATCTTTGAGACCAA

263-hex-F TTCTTGTGGAACCCGTTCAT
263-hex-R CGGGGAGAGGTATTTCTGGT

265-nk1-F ATACCCGGGTTCTGCTTCTT
265-nk1-R AAAACGACGCGATACCTCAG

266-nk2.1-F CATATAGCCCCAAACAGACCA
266-nk2.1-R CTGAGAAGACCGATGGGAAG

Q267-nk3.2-F CTCACACGCGCAGGTCTT
Q267-nk3.2-R CTTGCTGTTCGGTGAGTTTG

268-lhx2-F CATCTGCGATCGGTTTTACC
268-lhx2-R TTTGGCGAAGCAGGATAACT

269-phox2-F CAGGACAACCTTTACCAGTGC
269-phox2-R CCTGTACCCTGGCTTCAGTT

270-mbx1-F ATTCTGGAAGCTCGTTACGG
270-mbx1-R TCATCACAACATCCGGGTAA

Q272-otp-F AACGCCATCGTACTCGATTC
Q272-otp-R CGAGATTCAGTAAGTCCGACC

Q273-paxA-F GTCTGTGTACGGGCACGC
Q273-paxA-R AGTATCTTGGAGACGCAGCC

Q274-paxB-F CCAGTTAGGAGGATGCTTCG
Q274-paxB-R GACACTTTGAGCTGTCGGGA

Q275-pitx2-F ACATTTCACCAGCCAGCAAC
Q275-pitx2-R TCAAGTTACACCACGCACAGA

293-hox2-F GATATGGTTCCAAAATCGGC
293-hox2-R GGTGGATCGTCATCACCTTT

Q294-hox11-13c-F
CGGACAAAACGACGACCATA
Q294-hox11-13c-R AAAGCCTGGCTCTTCGGT

Q296-pax6-F GGCTGCGTCTCGAAGATACT
Q296-pax6-R AGCTACCCGAGGCTTACTCC

Q297-arx-F AGGGCTGAAGCAAGATGTGT
Q297-arx-R TGTTGACGAGCTTTCAGTCG

Q298-arxl-F CGAGCTAAATGGCGTAAAGC
Q298-arxl-R TCTATCATCATTCTCCGGGC

Q299-irxB-F TGGATAGCTCACACACGCTC
Q299-irxB-R GACCAGTAAGTCTCCCAGCG

Q300-cdx2-F GCGTCGTATACACGGACCAT
Q300-cdx2-R GTACCCACCTGTCTTTCCGA

309-F-DLX CCAGCTTACAACTCCAACAGC
309-R-DLX TTACCTGAGTTTGAGTGAGTCCA

310-F-shox GACGGAGCAGGACGAATTT
310-R-shox GGCATAGCTTCCTGCTCAAC

311-F-prx TGGACTCGTGCTTCTGTGAG
311-R-prx AATCGGACCACATTCACCAC

312-F-lmo2 TGCAGAGCTTCCTTCCAAGT
312-R-lmo2 CCGCTACTTCTTAAGAGCCATT

314-F-lmx1 GCAGTATAACTTGCGATCTCTGG
314-R-lmx1 GATGACCGGTACCTCATGAAA

317-F-GSH1 AGGTAGGTGGCGATCTCGATT
317-R-GSH1 AAGAGGATAAGGACGGCATTC

321-F-eyg CAAGCGAAGATGGTTGGATT
321-R-eyg TATACGAACCAACGCCCACT

Q324-hbn-F GGCGATCTAGGACCACCTTC
Q324-hbn-R ACTCGTGATTCGCTGAGGTC

327-F-Nk7 TGTCGGTGACGTTCAGTAGG
327-R-Nk7 AAGAAGAAGGCGAGGACGAC

330-F-PKNOX2
TTGGAACTAGAGAAGGTCAATGAA
330-R-PKNOX2
TGAACTTACCGTTCATCTCTCAA

331-F-Cutl TTGATGACTGTGGATGTTGGA
331-R-Cutl GTGTCTTGTTGACGCCTGAG

334-F-unc4.1 TGAACACGTGACTCAACAAGG
334-R-unc4.1 CAATGGTTGGCAGCTGGAG

340-F-HLX AATGCTTACCTGTGCATCTGA
340-R-HLX TCTTCACAAGTCTGCAGAGGA

343-F-PROSPERO
AGGTACCGGAGGGCTTCTT
343-R-PROSPERO
ATGGTCTTCTTCCAGGATGG

345-F-meis CCCTCTCTGTCCTCTATGACCA
345-R-meis
CACAGGTACATGAACTATGTGACAA

Q388-lass6-F ACAGTCTCCAAGTGCCCAGA
Q388-lass6-R TCTGTCCTGGTTTCGTCTCC

Q389-arxl2-F AACGCTCTTTTCCAAGACACA
Q389-arxl2-R GACAATGCTCACCTGAACCC

Q392-hesxl-F GCAGCCGTACCATTTACACC
Q392-hesxl-R AACCTGGACTCTGGCTTCAG

Q393-eygl-F GACCTTTAACCCGGAACAAC
Q393-eygl-R GGAAGTGATGTCTTGCTGGA

Q394-pax4l-F CTTGGCTCAGATAAACGCACT
Q394-pax4l-R TTTTCGCCCTCCGGTTCT

Q395-msxl-F AAAGGACGGTGCGAAGAAG
Q395-msxl-R CGACAATTCGGCTACATCAA

Q396-phb2-F GCCTACTCGTCCAAGCAACT
Q396-phb2-R GGTGCGTCTGTTCTGAAACC

606-F-rough AATCTGCGCCTTTTCGATT
606-R-rough GCATGCGCAGAGTAGAACTG

610-gbx-F CATTCACGAGCGATCAGTTG
610-gbx-R CTTAACCTGGACCTCGCTCA

618-F-pou6 GCCGACTGAGGTATTCCAGA
618-R-pou6 CTGAAGCCGAGGAGAGACAC

II. QPCR timecourses

2-six3		14-nk2.5		16-pax1.9	
0	0.3	0	4.7	0	0.3
6	2.9	6	1.3	6	0.3
12	1330.5	12	1.4	12	0.0
18	965.1	18	0.6	18	0.2
24	1469.5	24	3.9	24	0.6
36	1644.2	36	3.1	36	0.6
48	2239.1	48	10.9	48	1.0

12-en		15-six1.2		18-brn3	
0	0.0	0	3.6	0	0.0
6	0.1	6	1.2	6	0.2
12	0.8	12	8.8	12	0.4
18	5.8	18	1.4	18	1.4
24	2.6	24	53.0	24	2.2
36	28.3	36	50.0	36	6.4
48	9.7	48	49.1	48	18.5

21-six4	
0	0.4
6	0.1
12	0.2
18	0.0
24	0.6
36	1.2
48	1.5

45-hox9.10	
0	0.1
6	1.3
12	0.7
18	0.6
24	0.8
36	0.9
48	1.9

74-msx	
0	2.1
6	4.6
12	5.0
18	47.1
24	680.6
36	649.7
48	523.8

26-oct1.2	
0	1132.0
6	988.3
12	610.8
18	359.7
24	873.5
36	1004.2
48	623.2

47-pax258	
0	0.0
6	0.2
12	7.2
18	5.8
24	28.6
36	544.9
48	963.1

75-nk2.2	
0	0.6
6	12.0
12	22.5
18	495.8
24	3134.0
36	2456.0
48	2401.4

32-isl	
0	2.0
6	3.5
12	2.5
18	3.5
24	143.1
36	786.1
48	804.6

50.1-hox4.5	
0	0.0
6	0.4
12	5.4
18	8.5
24	2.4
36	40.2
48	12.1

78-atbf1	
0	669.8
6	734.7
12	169.5
18	301.9
24	1330.7
36	3299.9
48	4900.1

40-xlox	
0	0.0
6	1.3
12	2.4
18	3.9
24	5.2
36	120.6
48	581.7

50.2-hox8	
0	3.8
6	5.8
12	1.7
18	3.8
24	3.1
36	3.3
48	9.1

81-smadIP	
0	36.3
6	58.2
12	1784.0
18	339.8
24	614.3
36	1374.7
48	1402.9

43-tgif	
0	8.2
6	14.7
12	18.2
18	69.0
24	544.6
36	1608.5
48	1944.7

56-hnf1	
0	1.1
6	103.4
12	36.9
18	59.0
24	294.5
36	1408.0
48	465.1

84-pitx3	
0	35.6
6	9.1
12	15.9
18	10.3
24	4.8
36	31.0
48	11.3

44-lim1	
0	5.3
6	17.5
12	433.9
18	1768.1
24	1409.5
36	642.3
48	746.7

68-exd	
0	940.2
6	1094.5
12	814.4
18	132.6
24	301.8
36	495.8
48	473.8

85-hox1.tlx1	
0	0.0
6	0.0
12	0.5
18	5.4
24	1.4
36	26.1
48	9.8

95-lmo4

0	151.0
6	139.1
12	20.1
18	12.8
24	23.7
36	279.5
48	890.0

105-lhx3

0	0.2
6	0.7
12	1.0
18	3.3
24	1.5
36	36.5
48	6.9

108-paxC

0	1.3
6	20.4
12	13.4
18	3.1
24	5.6
36	73.6
48	365.0

109-mox

0	2.5
6	0.0
12	7.0
18	10.0
24	7.0
36	29.3
48	82.0

115-lbx

0	1.0
6	0.5
12	0.5
18	0.4
24	1.7
36	2.9
48	7.2

122-awh

0	0.6
6	2.5
12	538.8
18	525.8
24	632.8
36	714.0
48	889.0

127-nk6.1

0	0.2
6	1.3
12	1.1
18	3.8
24	7.8
36	24.9
48	152.5

146-chx10

0	0.0
6	0.2
12	0.3
18	0.8
24	2.6
36	42.2
48	50.8

150-emx

0	10.4
6	8.1
12	100.5
18	797.2
24	1398.8
36	2357.8
48	4279.8

151-rx

0	0.7
6	0.7
12	2.6
18	44.8
24	253.5
36	397.6
48	442.4

163-pitx1

0	0.8
6	0.9
12	0.9
18	2.4
24	5.0
36	28.4
48	592.7

184-alx4

0	0.1
6	0.1
12	0.5
18	3.0
24	84.2
36	158.4
48	118.5

200-irxA

0	0.0
6	2.5
12	29.4
18	284.2
24	2519.3
36	5633.5
48	6463.9

206-lmo4

0	138.9
6	145.3
12	28.2
18	15.1
24	32.3
36	288.3
48	1018.9

253-hox3

0	0.0
6	0.0
12	2.8
18	8.2
24	1.1
36	51.3
48	12.0

254-hox6

0	0.0
6	0.1
12	3.1
18	7.0
24	17.0
36	65.1
48	73.4

255-hox7

0	0.4
6	1.1
12	6.9
18	76.3
24	2052.2
36	5576.1
48	4880.6

256-hox11.13b

0	0.6
6	3.6
12	1449.2
18	1796.9
24	1688.7
36	799.9
48	1449.0

257-evx	
0	0.0
6	100.5
12	562.3
18	963.1
24	1393.1
36	702.6
48	1208.6

258-hb9	
0	0.2
6	0.8
12	2.0
18	8.3
24	11.4
36	79.6
48	202.1

259-barhl	
0	0.0
6	0.0
12	3.3
18	4.8
24	3.8
36	54.4
48	19.2

260-barx	
0	0.0
6	0.0
12	2.3
18	5.7
24	1.7
36	50.7
48	14.4

261-dbx1	
0	0.5
6	0.0
12	2.5
18	5.7
24	0.9
36	50.9
48	49.9

263-hex	
0	0.9
6	1.7
12	41.0
18	130.4
24	297.2
36	418.4
48	640.6

265-nk1	
0	13.3
6	25.8
12	2.1
18	8.7
24	161.2
36	1006.0
48	874.0

266-nk2.1	
0	0.0
6	1.8
12	4.5
18	303.1
24	1406.0
36	2401.6
48	2732.5

267-nk3.2	
0	0.0
6	0.1
12	0.1
18	8.2
24	28.2
36	52.0
48	120.8

268-lhx2	
0	0.0
6	0.0
12	26.7
18	17.0
24	151.8
36	483.6
48	586.0

269-phox2	
0	0.0
6	5.2
12	2.8
18	5.7
24	1.3
36	52.2
48	6.0

270-mbx1	
0	0.0
6	0.9
12	0.0
18	3.7
24	4.3
36	356.4
48	520.0

272-otp	
0	0.0
6	2.1
12	45.1
18	5.8
24	0.8
36	92.1
48	177.4

273-paxA	
0	0.6
6	0.5
12	1.9
18	4.2
24	1.1
36	27.5
48	7.7

274-paxB	
0	2346.6
6	1906.4
12	2240.7
18	2508.7
24	3357.6
36	3616.6
48	3799.9

275-pitx2	
0	0.7
6	0.0
12	2.2
18	4.6
24	2.9
36	53.5
48	242.3

293-hox2	
0	0.0
6	0.8
12	2.2
18	4.6
24	1.0
36	53.5
48	8.5

294-hox11.13c	
0	0.0
6	0.0
12	1.6
18	5.9
24	2.7
36	18.5
48	4.5

296-pax6

0	10.3
6	10.2
12	2.7
18	7.8
24	41.2
36	275.1
48	497.0

297-arx

0	0.0
6	0.5
12	4.6
18	6.7
24	1.9
36	102.0
48	182.1

298-arxl

0	0.0
6	0.0
12	0.6
18	7.6
24	1.5
36	36.9
48	8.3

299-irxB

0	0.6
6	1.5
12	7.0
18	5.2
24	10.9
36	118.3
48	132.9

300-cdx2

0	0.0
6	0.0
12	0.6
18	4.8
24	0.6
36	21.8
48	14.5

309-dlx

0	1.8
6	2.9
12	4.5
18	107.7
24	1749.5
36	4012.0
48	4976.2

310-shox

0	0.0
6	0.1
12	0.1
18	1.6
24	0.5
36	0.1
48	2.3

311-prx

0	1.3
6	0.9
12	1.3
18	8.0
24	7.7
36	3.9
48	3.2

312-lmo2

0	131.2
6	268.8
12	173.6
18	46.3
24	153.3
36	160.5
48	208.8

314-lmx1

0	0.9
6	0.4
12	0.0
18	0.2
24	0.7
36	25.8
48	50.1

317-gsh1

0	0.0
6	0.4
12	1.6
18	0.9
24	2.8
36	100.2
48	120.2

321-eyg

0	0.0
6	0.1
12	1.5
18	9.2
24	1.9
36	10.1
48	7.1

324-hbn

0	0.0
6	0.6
12	73.9
18	580.3
24	1591.9
36	2221.8
48	2234.6

327-nk7

0	0.5
6	0.3
12	2.8
18	6.5
24	20.9
36	181.2
48	195.8

330-pknox

0	547.0
6	464.7
12	338.7
18	67.3
24	94.2
36	99.8
48	85.7

331-cutl

0	140.0
6	216.5
12	166.6
18	137.2
24	253.3
36	230.0
48	479.7

334-unc4.1

0	0.6
6	1.3
12	2.4
18	2.7
24	212.3
36	1423.1
48	2774.2

340-hlx

0	4.2
6	2.4
12	7.8
18	48.0
24	89.2
36	278.8
48	378.2

343-prox1		392-phb1		396-phb2	
0	534.9	0	0.0	0	0.2
6	508.3	6	6.4	6	1.6
12	203.0	12	115.1	12	1.9
18	189.2	18	206.6	18	4.0
24	1137.7	24	150.4	24	4.2
36	881.7	36	304.0	36	24.0
48	1738.2	48	208.0	48	4.3

345-meis		393-eygl		606-rough	
0	66.5	0	0.0	0	0.0
6	23.3	6	0.0	6	0.7
12	5.9	12	0.4	12	1.6
18	11.8	18	3.3	18	0.2
24	16.6	24	0.3	24	0.4
36	27.1	36	17.5	36	2.6
48	124.6	48	9.8	48	14.0

388-lass6		394-pax4l		610-gbx	
0	296.3	0	549.1	0	1.6
6	259.1	6	533.1	6	0.9
12	94.3	12	673.8	12	2.0
18	40.1	18	643.8	18	3.6
24	78.5	24	915.6	24	7.3
36	378.8	36	1383.7	36	131.1
48	524.9	48	743.6	48	134.9

389-arxl2		395-msxl		618-pou6	
0	0.0	0	0.0	0	254.9
6	0.0	6	0.1	6	113.0
12	0.9	12	1.5	12	70.1
18	1.6	18	18.8	18	92.0
24	0.5	24	66.5	24	113.1
36	15.3	36	127.4	36	213.6
48	2.7	48	171.0	48	180.2

III. WMISH primers

W2-six3-F ATTTAGGTGACACTATAGAAGGACGGAGACAGAAACATCG
W2-six3-R TAATACGACTCACTATAGGGGAGTGAGCCGAGTTG

W26-oct1-2F ATTTAGGTGACACTATAGAAGCTGTATGGCAACGACTTCA
W26-oct1-R TAATACGACTCACTATAGGGCCTGTGGGTGGCTGAATTG

W44-lim1-F ATTTAGGTGACACTATAGAATTTGTGCGGGCTGTGAAC
W44-lim1-R TAATACGACTCACTATAGGGCCCCTTCTACTAGTTCTTGGTGAGG

W56-hnf1-F ATTTAGGTGACACTATAGAAGAGGGCGACAACGAAAGC
W56-hnf1-R TAATACGACTCACTATAGGGAACCTGGGAGGGCGACAC

W68-exd-F TGAGTATCAGGGGTGCACAAG
W68-exd-R AGGATCTTTGTGAAACACCCC

W75-nk2.2-F CATTTTCTCTTTATCGTTTTCTTTTTC
W75-nk2.2-R ATGGTTGTGCCATTGAACCT

W78-atbf-F ATTTAGGTGACACTATAGAAAGTGCAAGGTGGCATTTCC
W78-atbf-R TAATACGACTCACTATAGGGTTCTCAAGTTCTTTACTAACATGAAGC

Q79.81-smadIP-F TTTACAGACTGCAGCGTCACA
W79.81-smadIP-R ATTTAGGTGACACTATAGAAGACGCTCTAGCTGGGACTTG

W122-awh-F ATTTAGGTGACACTATAGAACAAGACCTCGAACGCATCG
W122-awh-R TAATACGACTCACTATAGGGTGGCGAGTTTCTCACAGAGG

W150-emx-F ATTTAGGTGACACTATAGAAGCAAGGGGTTAAGAAAAAGG
W150-emx-R TAATACGACTCACTATAGGGATCCAGCATCAACTCGGACT

W200-irxA-F ATTTAGGTGACACTATAGAAGCTGCTGGGGAAGGATATG
W200-irxA-R TAATACGACTCACTATAGGGGCCAAGGCGAGCTGTGAG

W265-nk1-F ATTTAGGTGACACTATAGAAGCTGGTTTTAACCCACCATCC
W265-nk1-R TAATACGACTCACTATAGGGCATGCATGTGCGTAAACATAGG

W266-nk2.1-F ATTTAGGTGACACTATAGAACATATAGCCCCAAACAGACCA
W266-nk2.1-R TAATACGACTCACTATAGGGAAGGAGAAATGGAGCCGTTG

W324-hbn-F ATCAGCATCATCAGCATCCA
W324-hbn-R CACATGGATCTGCAATCTTACTC

Appendix 2: Supplementary Material for Chapter 2,

I. QPCR primers

5-F-Jun CCTTTTCCTCTCAGCCTTGA
5-R-Jun TATCAAGCAGGAACCCTCGT

6-F-NeuroD AGCTACGAGGTTCGTGGTTG
6-R-NeuroD CGCCTCGCAAAGAATTACAT

7-F-Nrf1 ACAAAGTTGCCGCTCAGAAT
7-R-Nrf1 TCAATGCTGTCACGCTCTTT

11.2-F-Smad2/3 TGTCTGCGTGTCTGTTCAACT
11.2-R-Smad2/3 CCATGCATCTCAACCATCAC

14-F-NK2 CGTAAACCCCGTGTTCTCTT
14-R-NK2 GTCGGCGTAAGCTTCAGAAC

23-smad1-F ACCATGGCTTTGAGATGGTC
23-smad1-R AGCACGGGGTAGAGGTAACA

25-F-Smad4 GGATATGGAGATGGCCAGAC
25-R-Smad4 AGGTGACACAATTCGCTCAT

27-F-Dac TGCCACAAGCTTTTGAATTG
27-R-Dac GGATGCGAACCTGTTCTACG

28-F-Tbox2/3 ACTGCCGGTACAAGTTCCAC
28-R-Tbox2/3 GACACATTTCTGCATCCATTG

35-F-RXR AGATGCTCCAGGCATTTGAG
35-R-RXR TGTACGCCTCTCTTGAGGAA

Q36-hnf4-F GGGGAAGCACTACGGAGC
Q36-hnf4-R TCCTCTTGTCCTTGTCCACC

Q39-nfkb-F TGCTGTACAGGAGGAGAGGAG
Q39-nfkb-R CTGTCTATGAACATCGGTTGGA

Q40-xlox-F TTAACGGAACGTCACATCAAA
Q40-xlox-R CTGCTGACGTCGCTACCAT

46-F-SoxE CGGGAAGAGAAAACCTCACA
46-R-SoxE TTTTCCCAGGGTCTTGCTC

48-F-ash1 CATTGCCATAGCGATAACCA
48-R-ash1 TGAGGTGATCAGCGTCAAAG

49-F-Neurogin1 GACATCGTGATCGCTGGTAA
49-R-Neurogin1
AACGCCGACAGCTGAGTAAC

Q51-beta3a-F ATTTGACGAAGAAAGGCGAC
Q51-beta3a-R CGTACGGGATAACGCCAC

52-F-e12 GGCTCTCACCTGAGCAGAAG
52-R-e12 GTTTGGGCCTTATCCTGCTT

54-F-PTF1a TTGAAGGACTTAGGGAACACA
54-R-PTF1a CCCTCGTTCTCGATCATCTC

55-F-SoxC CATGGTTTGGTCACAAATCG
55-R-SoxC TACGGAGATTTCGCCACTTC

61-F-GataE TTGAAGTACAAGCCACAAGCA
61-R-GataE GCTCAAGACGAGAAGGAATTACA

69-F-Gro ACAGCACAAACCACCTCTCC
69-R-Gro CGTGAGTGCAGACGGTAAAA

70-F-RXF3 GGCTCAACAGGTTCAATGGT
70-R-RXF3 CACCGGGTCTAGCTTGTGTT

77-F-nato3 AAGAGTGACGAGGAAGATCAAAA
77-R-nato3 ATGCGATGCCATGCTATTC

92-enz1-F TCCGGAGTGTACCCCTAACA
92-enz1-R CTCACGATGGAGGAAGCAGT

106-F-NFIX GCGTCAAAGGCTTCTCGTAT
106-R-NFIX TTGTCGTTCATGAGCTCCTCT

110-F-tbx2 CAAGTATCAACCCCGTTTTCA
110-R-tbx2 TTCGTTCTGGTAGGCTGTGA

119-F-hath6 AACTCTCGAAGCTGGCCATA
119-R-hath6 TGCAGAGATCGACACAGTCC

120-F-myoR2 GAATGCTCGTGAACGAACTC
120-R-myoR2 GCTTGAAGCAAGTCGAAGTGT

123-F-E2F3 AATCGAGCAGCGGAAGTG
123-R-E2F3 GGATGTTATTCTTGGACTTCTTGG

124-F-grf TATGCTGCGATAAAGCCACA
124-R-grf CATTGTCCATTTCCCACACA

128-F-myoD TTGCCGTTACACAGAAATGC
128-R-myoD GCGAAGGTATTGGCGTACAT

129-F-myoD2 TCAGCGTGCAAGAAGAAAAA
129-R-myoD2 ATTTGAGCAAGTCCGCTGTT

132-F-tll AATCTTAGACATCCCCTGCAAA
132-R-tll TACGTACGGTTCCGTCGAAT

133-F-fax1 CACTACGGCGTCTACTGCTG
133-R-fax1 GTTTCTCCTTGCCTTGTCCA

135-F-mbt1 TGGATAACCGGTTCCTCATC
135-R-mbt1 GTCAGCGCCAATCCATTC

136-F-hand CGCCGACACCAAACTTTCTA
136-R-hand GCCGTTTATCGTCATCTCCA

137-F-par ACCGGACTCATAGCGTCAAC
137-R-par
GGTTGATATAACTTGTCGCTAACCT

142-F-tbx1 TCAATGCATCGCTATCAACC
142-R-tbx1 TGCAGTGAATTGTGTCTCTGG

143-F-nr1H6c GTGAGGGTTGCAAGAGCTTC
143-R-nr1H6c TCCTACAGGCAGGACAGTGA

Q144-nr1H6b-F GACAAGGCAAGCGGTCTG
Q144-nr1H6b-R GGTCCATCACACAGTTTCCA

147-F-af9 CCAACAGGAGAGGGATTCAC
147-R-af9 AGGCTTTGGAAAGCTCTCAT

153-F-dsx CGAGCGGAGAAGCCTTAC
153-R-dsx ATTTAGCGCAGATGCAGTCC

Q154-ap2-F AGGCGATACATTTAGCTCGC
Q154-ap2-R CCGCTAGCACCATTTGTCTT

155-F-tr2.4/shr2 GCAGTTCAAGCTGACCACAC
155-R-tr2.4/shr2
ACCTGAAAGGCCGGTAAACT

159-F-nr5A CCGACTCATGCTGAGGCTAC
159-R-nr5A TTGCTGTGCAACATCTCCAT

160-myoR3-F GTCAAGACCCTTCGAGATGC
160-myoR3-R
CAGATGAGAGATATACGTGGTTGC

164-F-scml1 TGAAACTAGAAGCCCTTGACC
164-R-scml1 CCAACCTCCAAAAGTCGTTC

165-F-mbt2 AGGTTTCAAGGTCGGTCACA
165-R-mbt2 TGTCCCATCCATCAAAGTGA

166-F-enz2 GATGAGACCGTTCTTCACAACA
166-R-enz2 CGATCACCATGGACTTTGC

172.1-F-nurr1 AAGAATGCCAAATACGTGTGC
172.1-R-nurr1 ACCATCCCACAGGCTAGACA

174-F-RAR CGTGCAGAAGAACATGCAAT
174-R-RAR TCTTTGGACATTCCAACTTCAA

175-F-nr1M3 GACGAGGCTTCTGGGATACA
175-R-nr1M3 CTCGCAGTGTCCTTCCTTCT

176-F-all1 GCTCATGACATCGAGAAGCA
176-R-all1 AATGCGGAACATGTAGACG

Q182-Spusf1-F CAATTCCAACACCCGCAT
Q182-Spusf1-R TTGTTGTAACTGTGCCCTCAA

188-F-clock CTGCGAGTGAGAAGAAGAGGA
188-R-clock AGTGTATGGTGGCTCTGAGGA

197-hifa-F CTGCCTTGTCCTCATTGCTA
197-hifa-R CGCAGTAGGTAAACTTCATGTCC

198-soxB2-F CCCTAAAATGCACAACTCCG
198-soxB2-R CCTTCATGTGTAGGGCTCGT

203-tbox20-F ACGTCGTGCCCTTAGACAAC
203-tbox20-R GAACGGTGAATCGGGATG

204-trh-F CGTGGCAAGGAGAACTACGA
204-trh-R GCCTCATGTGAAGATAGCCG

205-bbx-F CTGTAAACGACACCGTCAGG
205-bbx-R CCAGCTGCAGATATTTCTCCTT

209-arnt-F TCTGAGTACCGATTCCAAGC
209-arnt-R ATAGCTGTATTGGTGCAGACGA

214-ash2-F CTCCTCTCGGCTACGACAAG
214-ash2-R AAGAATCCTAACGTGTCGCC

217-tubby-F GGAGGTCACATCCAACATCTG
217-tubby-R CCTGGGCTGGAGATGAAGA

218-mtf1-F AAGCTCACCAGAGAATCCACA
218-mtf1-R GTAAGGGCGTTCCCCTGTAT

220-creb3-F GTCTCAAGACGGTGAGGAGG
220-creb3-R CTGTTCTGCTTGGTGCATGT

223-sin3a-F ACAACACGCTGTTGGATCTG
223-sin3a-R TATCCAGAGTGAAGGCCATGT

224-soxH-F TGACAAATGCCGACATAAGC
224-soxH-R GGATGGTCTTTTCGGTGTTG

226-ahr-F CAACCCTAGTAAGCGGCATC
226-ahr-R GCTCACACTGAGCCTGAGGA

228-trithorax-F
GAAGAATGTCGACGACGGAT
228-trithorax-R
ATGAAACGAGACAGGTTGCC

229-nfya-F CCAGGAGCGGAACTACTTGA
229-nfya-R TTCTCTCCTTGGGAATCCTG

Q232-reverb-F GCCAAACTGATCGAGAGCTG
Q232-reverb-R CACGATCCCTAACGACTTGAA

Q233-fxr-F AAATCTCTTGGAGGAGGGGA
Q233-fxr-R CCTTCTCCTACCCAAGGTCC

Q234-nr2C-F GAAGTGCTACCTCGATTGCC
Q234-nr2C-R AACTGCTCCATAGCCTTTGC

Q235-dsf-F AAGTATGCGGTGACCGTAGC
Q235-dsf-R CTCCTTTCCCTTGTTGCTTG

Q236-pnr-F TTACTGTGCGCTCTGCAATG
Q236-pnr-R CATGATTTCCTGAAGGAGGC

Q237-nr2E6-F GACATTCCATGCCAAGTGTG
Q237-nr2E6-R CGACAGACATAGGCCAGGTT

Q238-nr5B-F TGCAGCTCTCACAAGAATGC
Q238-nr5B-R TGCCAAAAGAGACCCAGAGT

Q239-gcnf-F GGGATCGTATCCTGTGAAGG
Q239-gcnf-R ACACCTGTTCCTCTTCTGTCG

Q241-olig3-F CCATCGTTTTCCAAGTCTGG
Q241-olig3-R GGTATAGCAGCGGTGTTGGT

Q242-mist-F TGCACACACTTAACGATGCC
Q242-mist-R
TTGGCTAGAGTCAGAGTCTCGAT

Q243-scl-F CTCTGCATTTTCCGAGCTTC
Q243-scl-R TCACGGAGCTCCATCAGG

Q244-acsc-F CCATGGATTCGCCAATTTAC
Q244-acsc-R TGCCTCTTCGTCAAGCAATA

Q249-soxB1-F GGCAACAAGAACAACAGCAA
Q249-soxB1-R AATTGTGCATTTTGGGGTTC

Q250-soxD1-F CCAGCAAGCCTCACATCAAG
Q250-soxD1-R
CCAGTATCTTGCTGATGTTGGA

Q251-lef1-F AGCGCAGCCATTAATCAAAT
Q251-lef1-R CTCCAGCCTGGGTATAGCTG

Q252-nr1M2-F GGTGTCACATGGTGGTGAAA
Q252-nr1M2-R AGCTCTGGGAAACCAGGAAT

Q277-fhl2-F TGTGCAGAGTGTGGCAAGTC
Q277-fhl2-R ACGATGGGTTGCTTGCAT

Q278-lmpt-F GCTGGCATCCATTCTGTTTT
Q278-lmpt-R GGATTCTGCATAGTGCCGTT

Q279-prkl2-F AGTACTGCGATTCTTGCGGA
Q279-prkl2-R GCAACGACCTATGGCATGTA

Q280-hlf-F AAAGATCTACGTGCCCGATG
Q280-hlf-R GGATGCCCGGATAACAATTT

Q281-mafB-F GTCGAGATTGAAAAGCCAGC
Q281-mafB-R AGAACTCGGCAGAGTCAGGA

Q282-giant-F AGTCAAACCCGTTCCAGATG
Q282-giant-R TGATTGCGATCTCTTCCTCC

Q283-lztf1-F GTACGGGGTGAGGTGGAAT
Q283-lztf1-R GCTCTGAGATGTCTGCCTGAA

Q284-myb-F CTGAGGCAAGCCATTGAAGT
Q284-myb-R ACCAGGTCAGGGTTCAGGAC

Q285-mta1-F TCAGCTGAAGCATCGTGAAC
Q285-mta1-R AGCAATGTGACTGTGCATTTG

Q287-p3A2-F AGCATCATGGAAGGGATGAC
Q287-p3A2-R GTGTACCACAGCATGGGATG

Q288-runx1-F CAATTGGAGCAGGGAATGAC
Q288-runx1-R ACTTACCACGGCCTGATTTG

Q289-runt1-F AGTTGTTTCGCTGGGAGAGA
Q289-runt1-R CGAGCCACTTGGTTCTTCAT

Q290-smad6-F AAAAATTCGCCAGAAGATCG
Q290-smad6-R CTGTGAACGTCCTGGAGTGA

Q291-tead3-F5 ACGATGCAGAGGGCGTGT
Q291-tead3-R5
CATTTTGCCTTCATCTGATAGAATA

Q292-tead4-F ACTTCATCCACAAACTCAAGCA
Q292-tead4-R
CACCTGTAGAATTGTGAAGTTCTCT

Q295-ldb2-F GACGAACACTTATTCCGCGT
Q295-ldb2-R CGTCTGTTCGCAGTCTAACG

Q301-hey-F TATCGACGCAGATAGCACCA
Q301-hey-R GTGTGTGGGTTGAGGGATGT

303-F-Myc CCCGCCATCCTCACATAAT
303-R-Myc GGAACAGCGCTTTACCACTT

Q307-irf1-F ACCCAGGCTATTCAGGGATT
Q307-irf1-R CTTGTGTGGATAGCCCAAGC

316-F-CP2 ACCGAACATGCACAGATTGA
316-R-CP2 CATTGATAGTCTGTGAGGCAGTG

318-F-DP1 GAACCTGGTGACACGCAAC
318-R-DP1 ATGACCGTTTTCTTGCTGGT

320-F-SoxF TGTAGATCGGGGTTCTGGTC
320-R-SoxF TACTGGACAACGGTGGTGGT

326-F-SuH CATCGTCAGTCGGTACTCGTT
326-R-SuH ACATGCTGTAGCGAGGCATA

329-F-dmtf GCCTCTTGAGACTCCACCAT
329-R-dmtf GAATTGCTGACACAGAAGTGG

335-F-cic TGATCTTCAGCAAGCGTCAT
335-R-cic CTGTTTCTCCTTGGGCTTGA

336-F-AP4 TGCTCAATTTCTCTCCTTCGT
336-R-AP4 CGAGAGATAGCCAATAGCAATG

337-nfil3-F TGAAGACTTGCGAAGTGAGC
337-nfil3-R TGTGTTTGAGGTCCTTGTGG

338-F-E78b AATGTCACTTTTCAGGCAACTG
338-R-E78b TCTTTGGACATTCCTACTGCAA

339-F-E2F4
TCCATTGAATACTGTTCTTTGATTT
339-R-E2F4 GGCTGACACATTAGCAGTACGA

341-F-SRF TTCAAAACACGGCAAGAAAA
341-R-SRF AATGCCTGTTTTCCTTTTGC

347-F-irf4 CTCATCAGCCTCATCGACAG
347-R-irf4 GGAGTCTTCTTGAGGGTCGT

Q348-mlx-F TTAAAGCGATCATGGATGCC
Q348-mlx-R GCTGAATACACATGCCGAAA

352-mef2-F GGTAACGTTCACAAAACGCA
352-mef2-R TGGAAGAGCTTGTTCCCACT

353-F-mbf1 AGAAAGGGGGTCCTGGAAG
353-R-mbf1 TAGGGCTTGGTTCCAGAATG

354-F-ATF2 CGACAAAAGTTCCTGGAGAGG
354-R-ATF2 TGAGAGAATTATTTGTTGCGTTT

356-F-Trx2 TGGATGCAGAAAAGCAAACA
356-R-Trx2 GCACAGTACCCATGCCTTG

357-F-thr TTCGAAGGACGATTCAGAAGA
357-R-thr TAACGGCATTGCTGACATTG

360-F-nr1H6a ACCTGCGAGGGGTGTAAGAC
360-R-nr1H6a ATGCCGACGGAGATACACTT

Q364-mad-F TCGCACAAAGACCAGCTATG
Q364-mad-R TGACGTTGTCTGTGCAAGGT

Q365-max-F AATCCGACTCGAACTCCTCA
Q365-max-R CCCTGAGTGTTCGCCTTTTA

366-E78a-F GATGCAAAGGCTTCTTCAGG
366-E78a-R AGCGACAATGCTGACATCTG

Q367-ERR-F2 GCTGGTAGCAAAGGCGTTAC
Q367-ERR-R2 TCCCCAAGGTGAGTGTTCTC

368-nr1AB-F GCAGAACTGTCCAGCACAAC
368-nr1AB-R AGGATGCACTTCCTGAGTCG

369-nr1M1-F CGGAGTCTCAACCAACATGA
369-nr1M1-R GCCGACACCGAGACATTTT

370-nr1M4-F GATGCAAGGGTTTCTTCAGG
370-nr1M4-R GGCAGGCAGGACATCTATTT

Q371-Ppar1-F3 TAACATCACCCGACCGAGAT
Q371-Ppar1-R3
CAGAAGGGATAGGAGTCTTGGA

Q372-PPAR2-F3
CAGATCGGGAAGGTCTTGTAG
Q372-PPAR2-R3 CCATTAGCTTGGCGTAGAGG

Q373-ROR-F2 CAGAGTCCATACCAGGCTTTG
Q373-ROR-R2 CACATCCCTTCCCACATTCT

Q375-AtoL1-F CGCATGCATCAGCTAAGAGA
Q375-AtoL1-R GGACAGGGTATCGTTGCAGT

Q376-AtoL2-F CGAAAACGTATGCGGAGTCT
Q376-AtoL2-R AACTTTGAGCCAGAAGCAGC

Q379-bhlhB1-F CCCTCTCCTCTTCTCCCAAC
Q379-bhlhB1-R
TGATCTTGAGGTCTTCGATCC

Q381-NSCL-F GGATACGTGTCGAGGCTTTC
Q381-NSCL-R GTAACAGATGGCGAGTCGGA

Q382-NXF-F CGATCAGATCAACTCGGAGA
Q382-NXF-R TGTAGACACAGGCAAGCGAC

Q384-Id-F ACCATGTCCGATTGCTACG
Q384-Id-R CATGCTGTAGTATCTCCACTCG

Q386-Mnt-F GGCGCATCTGAAGGATTG
Q386-Mnt-R CGTGTCAAGACCTGGATGAA

Q387-Acsc3-F GGAGGAATGCGAGGGAAC
Q387-Acsc3-R GAAGCGTCTTGACTTTGGAGA

Q398-fos-F CTGCCTCCAAGTGTCGAAGT
Q398-fos-R CGTTCCGATTCAAGTGCTTT

Q399-creb-F2 AGACCGGCCACATCGTTA
Q399-creb-R2 GCTGCTTCCCTGTTCTTCAT

Q400-ATF6-F3 GGCAGCACACTTTCTTCACTA
Q400-ATF6-R3 CTTTGGAGCCAGGGGTAACT

Q401-XBP1-F2 TCAGTGGTCGTTTTGGATCA
Q401-XBP1-R2 TCGTCAGACTCCACATCAGC

603-F-glass CATTCTGGTGAGCGTCCCTA
603-R-glass GACTGGCAACAGCAGCTACA

605-F-Sim GGAATAGGGCACGCATCTT
605-R-Sim GAGAAGGAGAACGCGGAGT

607-F-Coe CTCACTCCAGACGATCATGC
607-R-Coe AATCAGCCCTAGCGAAGGA

609-F-MITF CCTCCTATTGATGGTCTCCAA
609-R-MITF GGGACCATCCTCAAGTCATC

617-F-HesC CCAGAACAGGGCGAATCTAA
617-R-HesC CGAAGACGGGTTTCAATGTC

II. QPCR timecourses

5-jun		11-smad3		27-dac	
0	2864.6	0	804.9	0	4268.5
6	1639.0	6	522.8	6	775.5
12	2012.6	12	182.5	12	680.8
18	2105.1	18	220.7	18	1326.5
24	2618.4	24	298.2	24	1937.0
36	2447.8	36	509.6	36	1261.8
48	4157.7	48	923.2	48	808.8

6-neuroD		23-smad1		28-tbx2.3	
0	0.1	0	1619.8	0	16.4
6	0.8	6	1903.2	6	6.5
12	3.2	12	1887.4	12	4.8
18	1.4	18	772.7	18	631.4
24	2.7	24	920.2	24	4274.4
36	4.3	36	1390.5	36	1807.6
48	11.1	48	965.8	48	1755.2

7-nfe2		25-smad4		35-rxr	
0	46.2	0	1033.3	0	0.2
6	69.4	6	327.5	6	5.8
12	244.4	12	205.5	12	1.0
18	711.5	18	210.2	18	9.4
24	1022.5	24	332.9	24	15.2
36	1302.1	36	329.9	36	114.9
48	2488.9	48	693.7	48	0.3

36-hnf4	
0	123.8
6	61.2
12	16.6
18	13.2
24	32.0
36	463.6
48	725.8

39-nfkb	
0	487.9
6	464.1
12	156.2
18	303.9
24	481.9
36	612.3
48	1167.1

46-soxE	
0	3.0
6	2.6
12	231.5
18	25.9
24	10.1
36	115.8
48	529.6

48-ash1	
0	14.8
6	19.8
12	9.7
18	77.3
24	235.6
36	235.8
48	263.2

49-ngn	
0	0.5
6	0.8
12	0.8
18	1.3
24	0.5
36	2.6
48	50.9

51-beta3	
0	0.0
6	0.0
12	0.7
18	3.4
24	0.3
36	30.3
48	6.0

52-e12	
0	83.0
6	167.1
12	111.4
18	174.8
24	784.8
36	1029.1
48	1628.5

54-ptf1a	
0	0.3
6	231.5
12	101.6
18	97.3
24	160.2
36	609.0
48	430.0

55-soxC	
0	53.9
6	192.5
12	553.2
18	1133.8
24	2871.3
36	2681.1
48	2688.2

61-gataE	
0	2.2
6	0.8
12	1.3
18	111.4
24	105.4
36	346.1
48	218.7

69-gro	
0	793.6
6	1100.8
12	357.2
18	315.9
24	530.3
36	667.5
48	749.5

70-rfx3	
0	1833.7
6	971.1
12	934.9
18	955.9
24	522.9
36	507.1
48	638.1

77-nato3	
0	0.0
6	4.9
12	11.1
18	1.7
24	5.6
36	7.2
48	17.0

92-enz1	
0	25.5
6	27.5
12	18.7
18	25.0
24	53.8
36	114.2
48	72.4

106-nfIX	
0	56.1
6	43.3
12	24.8
18	8.2
24	5.9
36	8.0
48	48.0

110-tbx6	
0	5.9
6	5.2
12	1.6
18	6.5
24	15.4
36	24.0
48	96.1

119-hath6	
0	195.3
6	59.5
12	32.4
18	19.3
24	41.0
36	124.2
48	148.1

120-myoR2	
0	0.3
6	0.1
12	0.3
18	0.3
24	1.4
36	13.7
48	59.8

123-e2f3	
0	37.1
6	8.5
12	178.8
18	1718.9
24	2247.6
36	1788.2
48	1073.8

124-grf	
0	139.9
6	140.4
12	126.1
18	171.5
24	198.3
36	215.9
48	194.7

128-myoD	
0	0.4
6	0.1
12	3.5
18	1.0
24	0.8
36	139.0
48	259.1

129-myoD2	
0	0.0
6	0.8
12	1.9
18	1.2
24	1.9
36	4.3
48	4.9

132-tll	
0	0.9
6	0.2
12	1.3
18	15.7
24	80.7
36	180.6
48	256.0

133-fax1	
0	1.0
6	0.4
12	0.9
18	1.4
24	0.6
36	1.6
48	5.0

135-mbt1	
0	16.1
6	7.1
12	14.8
18	87.4
24	107.2
36	197.7
48	121.1

136-hand	
0	1.0
6	0.2
12	1.0
18	1.7
24	1.7
36	1.9
48	4.8

137-par	
0	0.5
6	0.3
12	2.2
18	3.3
24	8.9
36	22.9
48	23.4

142-tbx1	
0	1.8
6	1.4
12	1.3
18	1.1
24	1.4
36	2.7
48	3.2

143-nr1H6c	
0	163.8
6	343.5
12	145.3
18	121.1
24	226.8
36	278.8
48	192.9

144-nr1H6b	
0	325.4
6	381.2
12	129.0
18	43.5
24	85.8
36	794.9
48	2084.2

147-af9	
0	805.0
6	319.0
12	288.0
18	202.0
24	306.9
36	361.2
48	334.2

153-dsx	
0	0.3
6	0.1
12	1.6
18	0.2
24	2.0
36	11.8
48	42.3

154-ap2	
0	9.1
6	23.0
12	9.3
18	6.4
24	25.7
36	406.2
48	1183.2

155-tr2.4	
0	1203.2
6	1247.6
12	1335.0
18	1553.7
24	1400.8
36	935.2
48	1198.8

159-nr5A	
0	11.0
6	8.9
12	3.6
18	7.9
24	2.6
36	3.4
48	11.9

160-myoR3	
0	0.0
6	0.1
12	0.8
18	6.0
24	0.6
36	41.0
48	2.5

164-scml1	
0	329.5
6	94.5
12	21.6
18	31.4
24	59.3
36	105.8
48	95.2

165-mbt2	
0	364.0
6	213.2
12	183.7
18	113.0
24	117.3
36	186.4
48	170.9

166-enz2	
0	1056.1
6	435.9
12	239.4
18	169.0
24	236.7
36	370.7
48	378.2

172-nurr1	
0	4.2
6	0.0
12	12.3
18	8.1
24	6.6
36	11.8
48	125.7

174-rar	
0	22.0
6	15.4
12	10.4
18	47.6
24	210.8
36	269.1
48	171.2

175-nr1M3	
0	76.5
6	81.6
12	65.8
18	97.6
24	141.7
36	168.9
48	233.1

176-mll3	
0	218.2
6	144.9
12	94.1
18	121.8
24	271.2
36	224.9
48	289.7

182-usf	
0	979.8
6	850.6
12	904.4
18	639.5
24	1138.0
36	2678.6
48	2364.3

188-clock	
0	517.6
6	464.2
12	148.0
18	179.7
24	363.3
36	428.5
48	516.8

197-hif1a	
0	203.8
6	215.4
12	115.1
18	40.3
24	88.0
36	151.6
48	185.1

198-soxB2	
0	6697.2
6	6137.9
12	8840.7
18	7517.2
24	7770.7
36	12101.0
48	11146.9

203-tbx20	
0	619.7
6	640.8
12	264.8
18	42.6
24	18.1
36	38.5
48	14.7

204-trh	
0	219.2
6	263.3
12	88.0
18	57.9
24	46.0
36	124.3
48	64.3

205-bbx	
0	15.7
6	64.9
12	14.6
18	8.3
24	24.4
36	36.2
48	24.4

209-arnt	
0	40.3
6	31.5
12	18.4
18	34.9
24	117.0
36	312.6
48	268.7

214-ash2	
0	117.2
6	106.6
12	67.5
18	134.6
24	314.8
36	627.7
48	523.5

217-tubby	
0	431.1
6	380.5
12	64.4
18	76.9
24	187.1
36	509.3
48	378.9

218-mtf1	
0	343.1
6	582.0
12	119.0
18	159.3
24	378.9
36	532.4
48	512.4

220-creb3	
0	0.6
6	31.5
12	46.4
18	94.4
24	170.6
36	457.1
48	402.4

223-sin3a	
0	1096.2
6	229.1
12	153.5
18	618.0
24	1132.7
36	850.3
48	1238.8

224-soxH	
0	0.2
6	1.0
12	3.1
18	6.5
24	2.2
36	49.3
48	21.1

226-ahr	
0	318.1
6	234.2
12	107.3
18	72.8
24	208.0
36	292.4
48	363.8

228-trx1	
0	2143.0
6	2032.7
12	682.7
18	241.0
24	365.4
36	602.3
48	582.4

229-nfYa	
0	575.4
6	1093.9
12	330.7
18	92.2
24	274.6
36	712.7
48	629.1

232-reverb	
0	938.0
6	953.3
12	37.3
18	59.2
24	103.1
36	252.6
48	388.1

233-fxr	
0	1522.8
6	1715.8
12	160.5
18	154.1
24	356.3
36	1192.8
48	1377.5

234-nr2C	
0	152.8
6	171.9
12	130.1
18	138.4
24	198.0
36	284.3
48	148.9

235-dsf	
0	0.0
6	0.0
12	1.5
18	3.9
24	4.3
36	42.3
48	26.7

236-pnr	
0	0.0
6	0.0
12	2.2
18	8.5
24	0.6
36	45.4
48	14.9

237-nr2E6	
0	0.0
6	0.0
12	1.6
18	3.2
24	0.6
36	20.3
48	5.4

238-nr5B	
0	441.9
6	803.2
12	146.2
18	262.5
24	338.1
36	679.8
48	624.2

239-gcnf	
0	2.3
6	5.2
12	4.2
18	11.4
24	7.7
36	98.2
48	290.1

241-olig3	
0	0.2
6	6.7
12	4.0
18	27.7
24	13.3
36	73.5
48	31.2

242-mist	
0	0.2
6	2.3
12	2.6
18	4.7
24	2.9
36	77.8
48	39.6

243-scl	
0	86.9
6	74.1
12	71.7
18	125.5
24	464.4
36	287.4
48	325.6

244-acsc	
0	3.3
6	4.6
12	3.2
18	22.8
24	174.9
36	301.8
48	218.6

249-soxB1	
0	14998.4
6	15783.3
12	22847.3
18	10534.7
24	8365.2
36	19074.0
48	22890.4

250-soxD1	
0	268.1
6	333.2
12	221.2
18	181.8
24	213.0
36	471.2
48	524.8

251-lef1	
0	384.9
6	23.3
12	0.2
18	8.7
24	19.5
36	10.9
48	46.5

252-nr1M2	
0	5.6
6	24.6
12	2.1
18	9.2
24	17.7
36	259.1
48	1184.3

277-fhl2	
0	294.1
6	240.9
12	103.7
18	26.8
24	8.9
36	43.0
48	71.9

278-lmpt	
0	215.4
6	179.8
12	86.6
18	21.4
24	4.7
36	53.2
48	87.7

279-prkl2	
0	1091.0
6	649.7
12	239.7
18	400.6
24	320.1
36	336.2
48	459.6

280-hlf	
0	155.5
6	208.2
12	989.5
18	2266.4
24	3957.8
36	7830.3
48	6479.0

281-mafB	
0	18.8
6	7.3
12	8.6
18	6.8
24	18.7
36	215.2
48	106.1

282-giant	
0	0.0
6	0.3
12	0.8
18	4.1
24	0.9
36	36.1
48	9.5

283-lztf1	
0	0.5
6	3.5
12	8.1
18	25.1
24	73.6
36	225.6
48	162.2

284-myb	
0	489.4
6	466.2
12	996.3
18	1396.6
24	1277.0
36	1577.2
48	1256.2

285-mta1	
0	1627.1
6	1414.4
12	1979.9
18	1889.6
24	1549.8
36	2603.4
48	1406.4

287-p3A2	
0	308.6
6	339.8
12	436.3
18	496.2
24	477.2
36	371.6
48	165.1

288-runx1	
0	0.0
6	0.0
12	4.9
18	4.4
24	0.8
36	33.0
48	8.1

289-runt1	
0	22.6
6	2.2
12	110.1
18	253.4
24	829.3
36	2440.2
48	3278.2

290-smad6	
0	86.2
6	61.9
12	110.4
18	406.0
24	934.6
36	1312.0
48	1945.0

291-tead3	
0	314.4
6	279.7
12	129.3
18	199.7
24	286.1
36	377.0
48	816.6

292-tead4	
0	770.6
6	1568.6
12	190.5
18	597.1
24	663.3
36	1619.5
48	1749.9

295-ldb2	
0	847.8
6	1064.4
12	536.5
18	802.2
24	1141.4
36	2155.3
48	1842.3

301-hey	
0	1.1
6	12.7
12	15.7
18	8.0
24	2.2
36	61.7
48	47.9

303-myc	
0	277.1
6	201.3
12	72.8
18	826.4
24	2032.7
36	1492.3
48	1613.2

307-irf1	
0	15.0
6	0.5
12	11.6
18	18.2
24	18.8
36	120.0
48	241.2

316-cp2	
0	368.3
6	443.0
12	294.0
18	118.9
24	243.9
36	332.7
48	304.0

318-dp1	
0	739.9
6	883.1
12	1064.2
18	1508.9
24	1456.8
36	1498.8
48	1381.0

320-soxF	
0	109.3
6	61.5
12	71.6
18	28.6
24	57.9
36	478.7
48	2345.8

326-suH	
0	234.4
6	611.9
12	644.9
18	404.6
24	409.9
36	640.2
48	576.9

329-dmtf	
0	104.8
6	69.5
12	209.4
18	332.9
24	518.7
36	681.1
48	720.1

335-cic	
0	98.6
6	25.9
12	22.1
18	164.7
24	166.0
36	143.0
48	168.4

336-ap4	
0	10.3
6	2.1
12	53.0
18	229.2
24	227.0
36	189.7
48	121.7

337-nfIL3	
0	24.0
6	11.8
12	14.4
18	98.1
24	238.0
36	376.6
48	471.1

338-E78b	
0	10.2
6	20.1
12	20.4
18	54.4
24	107.1
36	199.7
48	210.8

339-e2f4	
0	2.0
6	2.1
12	171.2
18	380.0
24	692.9
36	674.1
48	578.3

341-srf	
0	621.8
6	650.4
12	619.8
18	297.2
24	548.7
36	899.4
48	636.1

347-irf4	
0	9.4
6	5.0
12	10.5
18	7.5
24	7.8
36	221.0
48	214.3

348-mlx	
0	283.5
6	194.7
12	231.4
18	269.1
24	391.8
36	684.4
48	686.8

352-mef2	
0	843.7
6	976.7
12	295.8
18	630.6
24	829.5
36	1779.4
48	2346.5

353-mbf1	
0	1133.3
6	843.4
12	950.0
18	1816.2
24	1851.2
36	684.6
48	611.9

354-atf2	
0	180.3
6	416.4
12	252.4
18	95.4
24	262.7
36	464.3
48	679.2

356-trx2	
0	0.1
6	0.2
12	3.8
18	15.1
24	29.5
36	41.0
48	54.8

357-thr	
0	271.7
6	177.6
12	64.5
18	135.7
24	476.5
36	575.7
48	1016.0

360-nr1H6a	
0	37.7
6	6.3
12	6.2
18	15.0
24	35.8
36	69.1
48	100.7

364-mad	
0	40.6
6	73.7
12	602.2
18	782.5
24	975.8
36	3158.0
48	2921.1

365-max	
0	241.1
6	200.3
12	339.4
18	447.2
24	482.4
36	512.9
48	350.7

366-E78a	
0	38.7
6	69.3
12	2.2
18	4.8
24	82.2
36	288.4
48	256.9

367-Err	
0	394.3
6	486.4
12	268.5
18	134.8
24	169.2
36	190.6
48	184.7

368-nr1AB	
0	23.3
6	6.3
12	40.9
18	71.5
24	216.2
36	320.4
48	256.8

369-nr1M1	
0	126.9
6	49.1
12	5.0
18	8.7
24	6.4
36	44.9
48	33.3

370-nr1M4	
0	106.1
6	10.0
12	0.9
18	7.9
24	10.2
36	47.8
48	17.1

371-Ppar1	
0	249.5
6	333.1
12	174.2
18	60.2
24	114.5
36	224.1
48	438.3

372-Ppar2	
0	1.1
6	0.0
12	1.4
18	3.5
24	1.5
36	12.5
48	7.8

373-Ror	
0	461.6
6	368.9
12	43.0
18	25.0
24	8.5
36	22.1
48	112.1

375-atoL1	
0	0.2
6	0.0
12	0.4
18	1.8
24	0.0
36	13.3
48	2.8

376-atoL2	
0	0.1
6	0.1
12	0.6
18	3.9
24	1.5
36	19.3
48	1.7

379-bhlhB1		387-acsc3		603-glass	
0	17.6	0	0.0	0	0.7
6	17.8	6	0.0	6	0.5
12	27.7	12	0.0	12	1.1
18	66.9	18	5.2	18	0.5
24	163.7	24	0.9	24	4.6
36	323.7	36	21.1	36	387.4
48	390.8	48	7.6	48	638.9

381-NSCL		398-fos		605-sim	
0	0.0	0	1.3	0	0.4
6	0.0	6	0.0	6	0.9
12	1.3	12	2.0	12	6.2
18	2.0	18	5.4	18	4.1
24	0.0	24	10.1	24	9.7
36	20.3	36	11.4	36	447.0
48	3.8	48	24.8	48	987.0

382-NXF		399-creb		607-coe	
0	0.0	0	1553.5	0	146.7
6	0.8	6	965.8	6	320.5
12	0.6	12	1030.8	12	198.6
18	3.9	18	911.2	18	75.4
24	0.6	24	1394.8	24	66.9
36	21.9	36	1239.5	36	431.7
48	3.3	48	923.3	48	1122.3

384-Id		400-Atf6		609-mitf	
0	3.8	0	298.6	0	606.0
6	29.1	6	274.3	6	640.0
12	403.5	12	81.8	12	139.8
18	280.8	18	188.9	18	116.5
24	486.2	24	340.9	24	756.3
36	2710.2	36	418.8	36	1736.4
48	743.4	48	307.5	48	1527.2

386-mnt		401-Xbp1		617-hesC	
0	72.3	0	631.7	0	296.3
6	22.2	6	558.8	6	110.9
12	25.1	12	170.9	12	2424.0
18	93.6	18	454.4	18	3062.3
24	210.3	24	1053.4	24	3629.5
36	190.1	36	2285.9	36	4098.9
48	260.2	48	4191.2	48	2356.1

III. WMISH primers

W5-Jun-F GGAGACTCAGTTCTACGAAGATTCA
W5-Jun-R TTGTTGTGCCAGCATAACTTG

W7-nfe2-F ATTTAGGTGACACTATAGAAAGTCACTCTACAAATGCCACTGC
W7-nfe2-R TAATACGACTCACTATAGGGGATCAACTGGGCTACC

W11-smad3-F ATTTAGGTGACACTATAGAAGCTCAAGGCTGTCGAACG
W11-smad3-R TAATACGACTCACTATAGGGGATACTCTGCACCCCATCC

W25-smad4-2F ATTTAGGTGACACTATAGAACAAGATTCAGTCCTCGTGTCC
W25-smad4-R TAATACGACTCACTATAGGGGAAGGGCACAAGTGATCCTG

W27-dachshund-F CCACCGCCTACTCAGGTTC
W27-dachshund-R AGGTCTTCCTCGTGGTCGT

W28-tbx2/3-F ATTTAGGTGACACTATAGAATCACCGCCTACCAGAACG
W28-tbx2/3-R TAATACGACTCACTATAGGGCCAAAAGCGAAGGGATGG

W52-beta1-F ATTTAGGTGACACTATAGAACACAGCCCACGAAGAAAGG
W52-beta1-R TAATACGACTCACTATAGGGGATGACTGCCCCAAACAGG

W55-SoxC-F GTTCCTCAGAAGAGCTTCGC
W55-SoxC-R GTCGACATGGACGATTGCT

W69-gro-F ATTTAGGTGACACTATAGAATGGCGTACTCATTTCACG
W69-gro-R TAATACGACTCACTATAGGGTGTCCAGACCCCCTGTCC

W70-rfx3-F ATTTAGGTGACACTATAGAACACGGTGACCCTGCAGAC
W70-rfx3-R TAATACGACTCACTATAGGGAGCAATGGGCGTCTCTCC

W117-etv1-F ATTTAGGTGACACTATAGAACCCCCGTCAGGAGATGTTC
W117-etv1-R TAATACGACTCACTATAGGGGTTAGCATGGCTGAGG

W123-e2f3-F ATTTAGGTGACACTATAGAACGCTACGACACATCATTAGGTC
W123-e2f3-R TAATACGACTCACTATAGGGTCGTAGGCATCGAACAGGTC

W155-tr2.4/shr-F ATTTAGGTGACACTATAGAAATGGGCATGGTTTCATCTC
W155-tr2.5/shr-R TAATACGACTCACTATAGGGATGCCGACATCGTTATCTGTG

W166-enz2-F ATTTAGGTGACACTATAGAAGGTGAAAGGTCAAATGACATAATGG
W166-enz2-R TAATACGACTCACTATAGGGTCGATCACCATGGACTTTGC

W182-usf1-F ATTTAGGTGACACTATAGAATCACCACAGAGGCCAAGG
W182-usf1-R TAATACGACTCACTATAGGGTCACCGACGCACACACAC

W188-clock-F ATTTAGGTGACACTATAGAATGTGATATGTATGATGATGGTGAAG
W188-clock-R TAATACGACTCACTATAGGGGGCATCTGCCGAGTTTATCC

W191-smcx-F GTGAGGACCAACCAGTGTGC
W191-smcx-R TGTCCACATGGTGTATGCAG

W203-tbx20-F TAATACGACTCACTATAGGGCACCCCTTCGTTCAGAAGC
W203-tbx20-R ATTTAGGTGACACTATAGAATGATTTTGATAGGCCGTGACAG

W209/186-Arnt-F CCATTTGTTTCTATGATTTTACTTTTG
W209/186-Arnt-R TGTACTCTCCATGATTACATTCCTGT

W228-trx1-F ATTTAGGTGACACTATAGAAGAGTTTCCTGTCCAGTTCTTTGG
W228-trx1-2R TAATACGACTCACTATAGGGATGAAACGAGACAGGTTGCC

W229-nfYa-F ATTTAGGTGACACTATAGAATGGAAGGGAATACAGTAGCTCAGAC
W229-nfYa-R TAATACGACTCACTATAGGGGCGGTAGCTGCACCATTG

W232-reverb-F ATTTAGGTGACACTATAGAATTTCGACGGAGCATCCAG
W232-reverb-2R TAATACGACTCACTATAGGGCACGATCCCTAACGACTTGAA

W233-ecr-F ATTTAGGTGACACTATAGAACGCATGGACCTGGAGTTG
W233-ecr-R TAATACGACTCACTATAGGGCACGTCCCAAATCTCTTGGAG

W250-soxD-F ATTTAGGTGACACTATAGAAGGCAGTGACGATGAAGATATGC
W250-soxD-R TAATACGACTCACTATAGGGGCGATTCCATTTTCCGATTC

W280-hlf-F TAATACGACTCACTATAGGGCAGGAAAACCCCCAGGAG
W280-hlf-R ATTTAGGTGACACTATAGAATTGCATCAATGGGCAATAGG

W290/614-smad6-F TAATACGACTCACTATAGGGTGTGCTGCAATCCCTACC
W290/614-smad6-R ATTTAGGTGACACTATAGAACGACTCCCGGTGCATGAG

W303-SpMyc-F CAGAGGAAGAAATCGATGTGG
W303-SpMyc-R GCGTTATGTGTCCTGGTGTG

Q316-cp2-F ACCGAACATGCACAGATTGA
W316-cp2-R ATTTAGGTGACACTATAGAATGATTCCCCCTACTCCTTGC

W318-dp1-F ATTTAGGTGACACTATAGAACCTGGTGACACGCAACAAG
W318-dp1-R TAATACGACTCACTATAGGGCCTGCTAGGTAAGGTTCCAATG

W320-soxF-F TAATACGACTCACTATAGGGGGCATGGGTGTTTGGTCTG
W320-soxF-R ATTTAGGTGACACTATAGAATGCGACGATGTTCTGTTGC

W339-e2f4-F TAATACGACTCACTATAGGGTGGCTACCCTTGTTGAGACG
W339-e2f4-R ATTTAGGTGACACTATAGAACGAGCCGACTAGCTCTGGTG

W348-mlx-F ATTTAGGTGACACTATAGAATGGCTAACCAAGATTCAAGC
W348-mlx-R TAATACGACTCACTATAGGGTCTGACCTGCTTTCCTCAAGAC

W353-mbf1-F AACTGGAGGTGAGCCTGGTA
W353-mbf1-R GACACAGCTCACATTTATAGGGC

W354-atf2-F TAATACGACTCACTATAGGGTGTGAGGAACTAAACGGGAAGTC
W354-atf2-R ATTTAGGTGACACTATAGAATGGGTTGAGTGTGGCAAATG

W357-thr-F AGGGAAACTGTCGGTTCTGA
W357-thr-R TGAGACCATTGGATAGGATCAA

W607-coe-F ATTTAGGTGACACTATAGAACAACTCCAAACGTTGATGGTC
W607-coe-R TAATACGACTCACTATAGGGTCTGGATCGCCAGGATGC

W609-mitf-F ATTTAGGTGACACTATAGAAGGGACCATCCTCAAGTCATC
W609-mitf-R TAATACGACTCACTATAGGGATGGTGGTGCTGCTGGAG

W617-hesC-F TCCGCCTACTCTTCTTCGTG
W617-hesC-R GTAGCAAACTGCTGGACTGG

Appendix 3: Rake Database Accession Numbers

```
>gi|2315192|emb|CAA72416.1|
>gi|8928574|gb|AAB80759.2|
>gi|20149252|gb|AAM12875.1|
>gi|44843657|emb|CAE77681.1|
>gi|44843665|emb|CAE77675.1|
>gi|44843669|emb|CAE77677.1|
>gi|44843673|emb|CAE77679.1|
>gi|44843675|emb|CAE77680.1|
>gi|44843667|emb|CAE77676.1|
>gi|7619995|dbj|BAA94758.1|
>gi|7684299|dbj|BAA95206.1|
>gi|7684303|dbj|BAA95208.1|
>gi|18389435|dbj|BAB84297.1|
>gi|18249647|dbj|BAB83926.1|
>gi|4126409|dbj|BAA36557.1|
>gi|18249655|dbj|BAB83913.1|
>gi|9279803|dbj|BAB01485.1|
>gi|10119776|dbj|BAB13476.1|
>gi|14041802|dbj|BAB55448.1|
>gi|6561878|dbj|BAA88232.1|
>gi|3551215|dbj|BAA32806.1|
>gi|6939732|dbj|BAA90665.1|
>gi|5921133|dbj|BAA84581.1|
>gi|2627169|dbj|BAA23527.1|
>gi|1842175|dbj|BAA11335.1|
>gi|9295675|gb|AAF85761.1|
>gi|2547046|dbj|BAA22868.1|
>gi|18147668|dbj|BAB83120.1|
>gi|3869123|dbj|BAA34362.1|
>gi|9049977|gb|AAF82570.1|
>gi|9049978|gb|AAF82571.1|
>gi|9049979|gb|AAF82572.1|
>gi|9049980|gb|AAF82573.1|
>gi|9049981|gb|AAF82574.1|
>gi|9049982|gb|AAF82575.1|
>gi|9049983|gb|AAF82576.1|
>gi|9049984|gb|AAF82577.1|
>gi|9049985|gb|AAF82578.1|
>gi|21070381|gb|AAM34281.1|
>gi|17402599|dbj|BAB78731.1|
>gi|18181872|dbj|BAA19053.2|
>gi|6561909|dbj|BAA88247.1|
>gi|15823174|dbj|BAB68738.1|
>gi|15823184|dbj|BAB68743.1|
>gi|15823190|dbj|BAB68746.1|
>gi|17826966|dbj|BAB79292.1|
>gi|7861539|dbj|BAA95683.1|
>gi|7594823|dbj|BAA94695.1|
>gi|40748271|gb|AAR89619.1|
>gi|40748273|gb|AAR89620.1|
>gi|9957633|gb|AAG09441.1|
>gi|9971753|gb|AAG10399.1|
>gi|18698421|emb|CAD23619.1|
>gi|3646272|emb|CAA08816.1|
>gi|4038585|emb|CAA07453.1|
>gi|11125350|emb|CAC15060.1|
>gi|28915|emb|CAA40483.1|
>gi|1039318|emb|CAA82909.1|
>gi|1743256|emb|CAA71047.1|
>gi|30140|emb|CAA34277.1|
>gi|33982|emb|CAA36297.1|
>gi|7711029|emb|CAB90275.1|
>gi|2597929|emb|CAA68843.1|
>gi|35134|emb|CAA79158.1|
>gi|288869|emb|CAA77952.1|
>gi|940537|emb|CAA54441.1|
>gi|1403340|emb|CAA64214.1|
>gi|1504086|emb|CAA64522.1|
>gi|3492787|emb|CAA73816.1|
>gi|1524099|emb|CAA65800.1|
>gi|6580397|emb|CAB63452.1|
>gi|20520631|emb|CAD30828.1|
>gi|20520633|emb|CAD30829.1|
>gi|7649399|emb|CAB89085.1|
>gi|11065699|emb|CAC14279.1|
>gi|1731967|emb|CAA70056.1|
>gi|1731969|emb|CAA70057.1|
>gi|4688673|emb|CAA56147.2|
>gi|1616667|gb|AAB16834.1|
>gi|2285922|emb|CAA72535.1|
>gi|639521|gb|AAB30748.1|
>gi|998548|gb|AAB34055.1|
>gi|998549|gb|AAB34056.1|
>gi|13397845|emb|CAC34575.1|
>gi|7963664|gb|AAF71306.1|
>gi|7963667|gb|AAF71307.1|
>gi|2065249|emb|CAA72115.1|
>gi|1524167|emb|CAA68952.1|
>gi|818010|emb|CAA54697.1|
>gi|1403386|emb|CAA64216.1|
>gi|10281217|gb|AAG15491.1|
>gi|5019546|emb|CAB44493.1|
>gi|4138205|emb|CAA07549.1|
>gi|1946226|emb|CAA73205.1|
>gi|10567755|gb|AAG18573.1|
>gi|14422303|emb|CAC41361.1|
>gi|1842097|gb|AAB47564.1|
>gi|1857752|gb|AAB48487.1|
>gi|1617400|emb|CAA62343.1|
>gi|6179901|gb|AAF05703.1|
>gi|1872475|gb|AAB49282.1|
>gi|1743341|emb|CAA71136.1|
>gi|1333848|emb|CAA80528.1|
```

>gi|1890302|gb|AAB49727.1|
>gi|510150|emb|CAA50470.1|
>gi|1905918|gb|AAB51178.1|
>gi|10803057|gb|AAG23698.1|
>gi|2276329|emb|CAA74671.1|
>gi|51555787|dbj|BAD38649.1|
>gi|1724124|gb|AAB38507.1|
>gi|51971307|dbj|BAD44692.1|
>gi|2114428|gb|AAB58322.1|
>gi|11118687|gb|AAG30415.1|
>gi|2246661|gb|AAB62700.1|
>gi|2245630|gb|AAB64434.1|
>gi|11602838|gb|AAG38875.1|
>gi|6683492|dbj|BAA89208.1|
>gi|1083356|pir||S52074
>gi|85156|pir||S14795
>gi|7446264|pir||JC5273
>gi|11761698|gb|AAG40148.1|
>gi|11761700|gb|AAG40149.1|
>gi|11761702|gb|AAG40150.1|
>gi|12082748|gb|AAG48598.1|
>gi|2654604|gb|AAB87765.1|
>gi|2826523|gb|AAB99738.1|
>gi|2827893|gb|AAB99907.1|
>gi|2827895|gb|AAB99908.1|
>gi|2641240|gb|AAC02971.1|
>gi|2911282|gb|AAC04325.1|
>gi|2944213|gb|AAC05246.1|
>gi|2944214|gb|AAC05247.1|
>gi|3005682|gb|AAC09344.1|
>gi|1575282|gb|AAC09474.1|
>gi|3047334|gb|AAC13696.1|
>gi|2829129|gb|AAC15420.1|
>gi|3108019|gb|AAC15752.1|
>gi|2582203|gb|AAC26143.1|
>gi|3335148|gb|AAC27037.1|
>gi|3335152|gb|AAC27039.1|
>gi|3337302|gb|AAC27373.1|
>gi|3342012|gb|AAC27508.1|
>gi|7512007|pir||T13283
>gi|7512044|pir||T13350
>gi|1079150|pir||A56922
>gi|1079151|pir||A56923
>gi|3551867|gb|AAC35286.1|
>gi|3642656|gb|AAC36518.1|
>gi|3642661|gb|AAC36520.1|
>gi|1161132|gb|AAC37680.1|
>gi|2935700|gb|AAC39639.1|
>gi|2735917|gb|AAC40116.1|
>gi|3138930|gb|AAC40119.1|
>gi|3004987|gb|AAC40156.1|
>gi|3004989|gb|AAC40157.1|
>gi|13173432|gb|AAK14407.1|
>gi|13194526|gb|AAK15458.1|
>gi|13241096|gb|AAB32248.2|
>gi|13241097|gb|AAB32249.2|
>gi|13241098|gb|AAB32250.2|
>gi|1594285|gb|AAC50719.1|
>gi|1561728|gb|AAC50763.1|
>gi|1669494|gb|AAC50816.1|
>gi|12830724|gb|AAK08197.1|
>gi|1832353|gb|AAC50993.1|
>gi|1552996|gb|AAC51108.1|
>gi|1813678|gb|AAC51130.1|
>gi|1732369|gb|AAC51318.1|
>gi|1737212|gb|AAC52046.1|
>gi|13272522|gb|AAK17191.1|
>gi|1669496|gb|AAC52892.1|
>gi|1620598|gb|AAC53108.1|
>gi|1620600|gb|AAC53109.1|
>gi|2209115|gb|AAC53456.1|
>gi|2735135|gb|AAC53559.1|
>gi|13384245|gb|AAK21331.1|
>gi|13384247|gb|AAK21332.1|
>gi|13384249|gb|AAK21333.1|
>gi|13384251|gb|AAK21334.1|
>gi|13384253|gb|AAK21335.1|
>gi|13384255|gb|AAK21336.1|
>gi|3283362|gb|AAC70014.1|
>gi|3901260|gb|AAC78624.1|
>gi|3901266|gb|AAC78626.1|
>gi|3126918|gb|AAC80425.1|
>gi|3152757|gb|AAC80430.1|
>gi|13898847|gb|AAK48898.1|
>gi|13936689|gb|AAK49895.1|
>gi|13926076|gb|AAK49525.1|
>gi|1835587|gb|AAB46595.1|
>gi|1835589|gb|AAB46596.1|
>gi|1835591|gb|AAB46597.1|
>gi|111179|pir||S22543
>gi|53486|emb|CAA41006.1|
>gi|111180|pir||S22539
>gi|53488|emb|CAA41007.1|
>gi|111181|pir||S22544
>gi|53490|emb|CAA41008.1|
>gi|1079117|pir||B56564
>gi|245324|gb|AAB21409.1|
>gi|103419|pir||S19095
>gi|7718|emb|CAA41341.1|
>gi|7428976|pir||TWFF
>gi|8247928|sp|Q61572|FXC1_MOUSE
>gi|3228522|gb|AAC24209.1|
>gi|5921137|dbj|BAA84583.1|
>gi|6016208|sp|O15522|NK28_HUMAN
>gi|2232067|gb|AAC71082.1|
>gi|12382254|gb|AAG53085.1|
>gi|475884|dbj|BAA06218.1|
>gi|1094515|prf||2106212A
>gi|8247930|sp|O60663|LMXB_HUMAN
>gi|21759253|sp|O75444|MAF_HUMAN

>gi|14194481|sp|Q9HBU1|BRX1_HUMAN
>gi|29337108|sp|Q924A0|T7L2_MOUSE
>gi|2981019|gb|AAC06203.1|
>gi|3122917|sp|Q24318|TDP_DROME
>gi|4049922|gb|AAC97879.1|
>gi|4096059|gb|AAC99797.1|
>gi|2944393|gb|AAD03248.1|
>gi|4164572|gb|AAD05558.1|
>gi|20138822|sp|O08609|MLX_MOUSE
>gi|3132524|gb|AAD09205.1|
>gi|433397|gb|AAD12223.1|
>gi|4235356|gb|AAD13185.1|
>gi|14194480|sp|Q9ER42|BRX1_MOUSE
>gi|2745892|gb|AAB94768.1|
>gi|3643195|gb|AAC36725.1|
>gi|2952324|gb|AAC05505.1|
>gi|3183530|sp|Q08369|GAT4_MOUSE
>gi|12643899|sp|Q9ULV5|HSF4_HUMAN
>gi|4324631|gb|AAD16967.1|
>gi|4324633|gb|AAD16968.1|
>gi|1711527|sp|P52172|SRP_DROME
>gi|6094434|sp|O95947|TBX6_HUMAN
>gi|2501114|sp|P70323|TBX1_MOUSE
>gi|2501117|sp|P70324|TBX3_MOUSE
>gi|2501120|sp|P70327|TBX6_MOUSE
>gi|4454797|gb|AAD20979.1|
>gi|2407245|gb|AAB70531.1|
>gi|3122045|sp|O35261|E2F3_MOUSE
>gi|4204923|gb|AAD10852.1|
>gi|17367970|sp|O35738|KLFC_MOUSE
>gi|2507549|sp|P46099|KLF1_MOUSE
>gi|17367666|sp|Q9ESX2|SP6_MOUSE
>gi|4585207|gb|AAD25325.1|
>gi|4585209|gb|AAD25326.1|
>gi|4585211|gb|AAD25327.1|
>gi|4590404|gb|AAD26566.1|
>gi|4809142|gb|AAD30110.1|
>gi|4929495|gb|AAD34020.1|
>gi|3023741|sp|O00358|FXE1_HUMAN
>gi|8134468|sp|O60548|FXD2_HUMAN
>gi|4105432|gb|AAD02417.1|
>gi|5042400|gb|AAD38241.1|
>gi|4731161|gb|AAD28370.1|
>gi|5524693|gb|AAD44343.1|
>gi|45181605|gb|AAS55456.1|
>gi|4809134|gb|AAD30106.1|
>gi|5882288|gb|AAD55267.1|
>gi|5916234|gb|AAD55949.1|
>gi|2145129|gb|AAB58423.1|
>gi|14195676|sp|P97863|NFIB_MOUSE
>gi|28201898|sp|Q9DAC9|SPM1_MOUSE
>gi|8705233|gb|AAF78780.1|
>gi|3135323|gb|AAC16484.1|
>gi|3132610|gb|AAC16322.1|
>gi|13183093|gb|AAK15048.1|
>gi|14581675|gb|AAK56791.1|
>gi|40786789|gb|AAR89905.1|
>gi|5577991|gb|AAD45410.1|
>gi|13477259|gb|AAH05098.1|
>gi|14582798|gb|AAK69649.1|
>gi|14582800|gb|AAK69650.1|
>gi|14582802|gb|AAK69651.1|
>gi|14595643|gb|AAK70869.1|
>gi|28317369|tpe|CAD29859.1|
>gi|4760291|emb|CAB42440.1|
>gi|31620894|emb|CAD35744.1|
>gi|31620898|emb|CAD35747.1|
>gi|31873240|emb|CAD97611.1|
>gi|21732508|emb|CAD38610.1|
>gi|6002919|gb|AAF00199.1|
>gi|4559273|gb|AAD22960.1|
>gi|4559275|gb|AAD22961.1|
>gi|8489093|gb|AAF75586.1|
>gi|2739071|gb|AAB94604.1|
>gi|6318281|gb|AAF06817.1|
>gi|13398525|gb|AAK21915.1|
>gi|13398531|gb|AAK21918.1|
>gi|6466208|gb|AAF12826.1|
>gi|6491915|gb|AAF14073.1|
>gi|6503008|gb|AAF14547.1|
>gi|2511636|emb|CAA56037.1|
>gi|1359830|emb|CAA63440.1|
>gi|5804930|emb|CAA64212.2|
>gi|23495536|dbj|BAC20215.1|
>gi|1869816|emb|CAA68061.1|
>gi|1869969|emb|CAA69399.1|
>gi|5804943|emb|CAA76315.2|
>gi|854182|emb|CAA83435.1|
>gi|24024987|dbj|BAC16749.1|
>gi|12659130|gb|AAK01201.1|
>gi|12659132|gb|AAK01202.1|
>gi|4522024|gb|AAD21787.1|
>gi|3694666|gb|AAC62435.1|
>gi|3387787|gb|AAC28933.1|
>gi|3387788|gb|AAC28934.1|
>gi|6652982|gb|AAF22568.1|
>gi|6652983|gb|AAF22569.1|
>gi|4883932|gb|AAD31712.1|
>gi|1575396|gb|AAB09533.1|
>gi|450593|gb|AAA18518.1|
>gi|15824803|gb|AAL09477.1|
>gi|18700304|dbj|BAB85121.1|
>gi|1568641|gb|AAB09039.1|
>gi|7684294|dbj|BAA95209.1|
>gi|22773783|dbj|BAC11805.1|
>gi|6716870|gb|AAC12973.2|
>gi|6760377|gb|AAF28319.1|
>gi|4838432|gb|AAD30990.1|
>gi|4838434|gb|AAD30991.1|
>gi|6815251|gb|AAF28469.1|

>gi|24371061|dbj|BAC22110.1|
>gi|736381|gb|AAA64491.1|
>gi|4106560|gb|AAD02889.1|
>gi|4097411|gb|AAD09210.1|
>gi|7144550|gb|AAC50051.2|
>gi|7158399|gb|AAF37424.1|
>gi|32478313|gb|AAP83468.1|
>gi|32478315|gb|AAP83469.1|
>gi|48103121|ref|XP_395506.1|
>gi|16506789|gb|AAL23950.1|
>gi|6754002|ref|NP_034426.1|
>gi|11612388|gb|AAG39220.1|
>gi|11612390|gb|AAG39221.1|
>gi|6754002|ref|NP_034426.1|
>gi|391752|dbj|BAA04057.1|
>gi|2160422|dbj|BAA04061.1|
>gi|2160425|dbj|BAA04064.1|
>gi|286023|dbj|BAA02573.1|
>gi|29120006|emb|CAD59560.1|
>gi|7271777|gb|AAF44618.1|
>gi|28569978|dbj|BAC57891.1|
>gi|7963882|gb|AAF71368.1|
>gi|7768755|dbj|BAA95537.1|
>gi|12963247|gb|AAK11181.1|
>gi|31338864|dbj|BAC76998.1|
>gi|7768758|dbj|BAA95539.1|
>gi|531523|emb|CAA85309.1|
>gi|7406868|gb|AAF61816.1|
>gi|7416985|gb|AAF62396.1|
>gi|7341308|gb|AAF61242.1|
>gi|1947098|gb|AAD03155.1|
>gi|18478831|gb|AAL73341.1|
>gi|18478839|gb|AAL73344.1|
>gi|18478843|gb|AAL73345.1|
>gi|7582386|gb|AAF64304.1|
>gi|11493712|gb|AAG35617.1|
>gi|11493716|gb|AAG35619.1|
>gi|11493718|gb|AAG35620.1|
>gi|10441263|gb|AAG16976.1|
>gi|5805398|gb|AAD51979.1|
>gi|28932878|gb|AAO60106.1|
>gi|6180051|gb|AAF05764.1|
>gi|14582216|gb|AAK69408.1|
>gi|25247297|gb|AAN72308.1|
>gi|7677531|gb|AAF67191.1|
>gi|7677537|gb|AAF67193.1|
>gi|7677540|gb|AAF67194.1|
>gi|25808960|gb|AAN74532.1|
>gi|25808962|gb|AAN74533.1|
>gi|7715875|gb|AAF68175.1|
>gi|25989428|gb|AAL58534.1|
>gi|7542466|gb|AAF63466.1|
>gi|17223636|gb|AAK49782.1|
>gi|1388160|gb|AAB02820.1|
>gi|17979652|gb|AAL50340.1|
>gi|17979654|gb|AAL50341.1|
>gi|17979656|gb|AAL50342.1|
>gi|27261725|gb|AAN86032.1|
>gi|15011285|gb|AAK77485.1|
>gi|15011286|gb|AAK77486.1|
>gi|15011290|gb|AAK77488.1|
>gi|15011292|gb|AAK77489.1|
>gi|15011294|gb|AAK77490.1|
>gi|2494231|sp|Q61502|E2F5_MOUSE
>gi|8099214|gb|AAF72066.1|
>gi|8132439|gb|AAF73290.1|
>gi|8099212|gb|AAF72065.1|
>gi|8490223|gb|AAB30577.2|
>gi|19110602|gb|AAL85197.1|
>gi|8705235|gb|AAF78781.1|
>gi|8714511|gb|AAB19475.2|
>gi|8777658|gb|AAF79044.1|
>gi|8777659|gb|AAF79045.1|
>gi|8777835|gb|AAF79116.1|
>gi|19387206|gb|AAL87134.1|
>gi|245518|gb|AAB21441.1|
>gi|23506640|gb|AAN37906.1|
>gi|27370669|gb|AAH38985.1|
>gi|13436005|gb|AAH04834.1|
>gi|13477390|gb|AAH05174.1|
>gi|13937859|gb|AAH07035.1|
>gi|18490183|gb|AAH22329.1|
>gi|14495617|gb|AAH09412.1|
>gi|9909704|emb|CAC04514.1|
>gi|6977877|emb|CAB75345.1|
>gi|9188626|emb|CAB97212.1|
>gi|9188627|emb|CAB97213.1|
>gi|9188628|emb|CAB97214.1|
>gi|9188629|emb|CAB97215.1|
>gi|9188630|emb|CAB97216.1|
>gi|9188631|emb|CAB97217.1|
>gi|9188632|emb|CAB97218.1|
>gi|9188633|emb|CAB97219.1|
>gi|9650707|emb|CAC00685.1|
>gi|9663149|emb|CAC01130.1|
>gi|20071519|gb|AAH27048.1|
>gi|13397837|emb|CAC34632.1|
>gi|6900088|emb|CAB71327.1|
>gi|12697318|emb|CAC28212.1|
>gi|18381013|gb|AAH22131.1|
>gi|11558196|emb|CAC17771.1|
>gi|5689744|emb|CAB51916.1|
>gi|1335343|emb|CAA35714.1|
>gi|20988217|gb|AAH29695.1|
>gi|45708710|gb|AAH32576.1|
>gi|23273562|gb|AAH36022.1|
>gi|23273577|gb|AAH36080.1|
>gi|23273943|gb|AAH35031.1|
>gi|37572298|gb|AAH43029.2|
>gi|29612546|gb|AAH49885.1|

>gi|31657162|gb|AAH53676.1|
>gi|28704080|gb|AAH47511.1|
>gi|33243958|gb|AAH55285.1|
>gi|30931139|gb|AAH52702.1|
>gi|37589103|gb|AAH51699.2|
>gi|12653953|gb|AAH00770.1|
>gi|19684076|gb|AAH25988.1|
>gi|37590620|gb|AAH58894.1|
>gi|28173602|gb|AAF07087.2|
>gi|28175028|gb|AAH43224.1|
>gi|13529587|gb|AAH05504.1|
>gi|14627271|gb|AAH07471.1|
>gi|28394609|gb|AAO38742.1|
>gi|28394611|gb|AAO38743.1|
>gi|28394613|gb|AAO38744.1|
>gi|28394615|gb|AAO38745.1|
>gi|49904701|gb|AAH76630.1|
>gi|28627849|gb|AAN75489.1|
>gi|28627851|gb|AAN75490.1|
>gi|28627853|gb|AAN75491.1|
>gi|27948552|gb|AAO25649.1|
>gi|25044787|gb|AAK93831.1|
>gi|1504084|emb|CAA64521.1|
>gi|34740149|dbj|BAC87712.1|
>gi|3281907|emb|CAA07550.1|
>gi|396174|emb|CAA49156.1|
>gi|21396471|gb|AAM49062.1|
>gi|12314004|emb|CAC12755.1|
>gi|12314005|emb|CAC12756.1|
>gi|12733867|emb|CAC28883.1|
>gi|21518681|gb|AAM60761.1|
>gi|21518684|gb|AAM60762.1|
>gi|21518687|gb|AAM60763.1|
>gi|21518699|gb|AAM60767.1|
>gi|21518704|gb|AAM60769.1|
>gi|21518690|gb|AAM60764.1|
>gi|21518693|gb|AAM60765.1|
>gi|21518696|gb|AAM60766.1|
>gi|21518701|gb|AAM60768.1|
>gi|12581495|gb|AAG59620.1|
>gi|12581486|gb|AAG59617.1|
>gi|22087278|gb|AAM90909.1|
>gi|56398594|emb|CAH59970.1|
>gi|22324549|gb|AAM95607.1|
>gi|22324551|gb|AAM95608.1|
>gi|47480852|gb|AAH69638.1|
>gi|51458332|ref|XP_372757.2|
>gi|51460509|ref|XP_092553.3|
>gi|51461005|ref|XP_040149.2|
>gi|51464004|ref|XP_497908.1|
>gi|51464091|ref|XP_496654.1|
>gi|22478860|gb|AAM97580.1|
>gi|22478862|gb|AAM97581.1|
>gi|22478864|gb|AAM97582.1|
>gi|22478866|gb|AAM97583.1|
>gi|56207082|emb|CAI25051.1|
>gi|56709819|emb|CAI35312.1|
>gi|51465900|ref|XP_379849.2|
>gi|51468129|ref|XP_291729.5|
>gi|51471119|ref|XP_372468.2|
>gi|56206703|emb|CAI25208.1|
>gi|55661454|emb|CAH70546.1|
>gi|51474873|ref|XP_372730.3|
>gi|51475167|ref|XP_044921.5|
>gi|51475209|ref|XP_497759.1|
>gi|51477368|ref|XP_093087.4|
>gi|17486065|ref|XP_066752.1|
>gi|55957868|emb|CAI13221.1|
>gi|11386195|ref|NP_036314.1|
>gi|13786123|ref|NP_112573.1|
>gi|55958352|emb|CAI12792.1|
>gi|55663202|emb|CAH72747.1|
>gi|57162412|emb|CAI40241.1|
>gi|57162415|emb|CAI40244.1|
>gi|57162416|emb|CAI40245.1|
>gi|20127495|ref|NP_006228.2|
>gi|23957692|ref|NP_705832.1|
>gi|23463326|ref|NP_694948.1|
>gi|30578406|ref|NP_849195.1|
>gi|57162467|emb|CAI39786.1|
>gi|57162468|emb|CAI39787.1|
>gi|57162469|emb|CAI39788.1|
>gi|57162520|emb|CAI40404.1|
>gi|57162521|emb|CAI40405.1|
>gi|57162522|emb|CAI40406.1|
>gi|57162523|emb|CAI40407.1|
>gi|55960100|emb|CAI13238.1|
>gi|55960102|emb|CAI13240.1|
>gi|55960103|emb|CAI13241.1|
>gi|55960104|emb|CAI13242.1|
>gi|57162535|emb|CAI39943.1|
>gi|4758228|ref|NP_004415.1|
>gi|42415527|ref|NP_963883.1|
>gi|55665718|emb|CAH72653.1|
>gi|55665720|emb|CAH72655.1|
>gi|55665722|emb|CAH72657.1|
>gi|23308713|ref|NP_008871.3|
>gi|4507107|ref|NP_003077.1|
>gi|4507859|ref|NP_003568.1|
>gi|5453936|ref|NP_006227.1|
>gi|5729728|ref|NP_006483.1|
>gi|5729976|ref|NP_006599.1|
>gi|7657409|ref|NP_055167.1|
>gi|6688760|emb|CAB65259.1|
>gi|51492762|ref|XP_497945.1|
>gi|47078264|ref|NP_997705.1|
>gi|51492896|ref|XP_376652.2|
>gi|56203099|emb|CAI19438.1|
>gi|56203100|emb|CAI19439.1|
>gi|38201649|ref|NP_937988.1|

>gi|38201653|ref|NP_938084.1|
>gi|31455424|emb|CAD92511.1|
>gi|38201651|ref|NP_937989.1|
>gi|10048418|ref|NP_065242.1|
>gi|6755132|ref|NP_035268.1|
>gi|6755602|ref|NP_035569.1|
>gi|6755608|ref|NP_035574.1|
>gi|27777679|ref|NP_035666.1|
>gi|13654264|ref|NP_112475.1|
>gi|22094991|ref|NP_079950.1|
>gi|26006853|ref|NP_080651.1|
>gi|28202073|ref|NP_075993.1|
>gi|18378727|ref|NP_081720.1|
>gi|20467109|gb|AAM22408.1|
>gi|55957717|emb|CAI15167.1|
>gi|55957718|emb|CAI15168.1|
>gi|55957720|emb|CAI15170.1|
>gi|15150793|ref|NP_085032.1|
>gi|15149491|ref|NP_150373.1|
>gi|16716369|ref|NP_444315.1|
>gi|16716509|ref|NP_444432.1|
>gi|17985999|ref|NP_536771.1|
>gi|18875322|ref|NP_573448.1|
>gi|24496776|ref|NP_722482.1|
>gi|55958405|emb|CAI16899.1|
>gi|27369712|ref|NP_766100.1|
>gi|55957984|emb|CAI16037.1|
>gi|55957985|emb|CAI16038.1|
>gi|55957987|emb|CAI16040.1|
>gi|29568079|ref|NP_818774.1|
>gi|56205905|emb|CAI19639.1|
>gi|57208963|emb|CAI40917.1|
>gi|30841039|ref|NP_035228.2|
>gi|31542154|ref|NP_031524.2|
>gi|11276073|ref|NP_068701.1|
>gi|55958372|emb|CAI16532.1|
>gi|55958370|emb|CAI16530.1|
>gi|55958374|emb|CAI16534.1|
>gi|55958375|emb|CAI16535.1|
>gi|55958376|emb|CAI16536.1|
>gi|55958377|emb|CAI16537.1|
>gi|55958378|emb|CAI16538.1|
>gi|55958384|emb|CAI15107.1|
>gi|55958386|emb|CAI15109.1|
>gi|15277319|ref|NP_201579.1|
>gi|34878726|ref|NP_919239.1|
>gi|55959773|emb|CAI15184.1|
>gi|33186878|ref|NP_081780.2|
>gi|34304111|ref|NP_899121.1|
>gi|34328101|ref|NP_031523.2|
>gi|7106250|ref|NP_031467.1|
>gi|34328397|ref|NP_766580.2|
>gi|6679335|ref|NP_032875.1|
>gi|16905095|ref|NP_473409.1|
>gi|55958889|emb|CAI14855.1|
>gi|40538819|ref|NP_035679.2|
>gi|47059067|ref|NP_620394.2|
>gi|6680736|ref|NP_031522.1|
>gi|6679629|ref|NP_031947.1|
>gi|6680207|ref|NP_032262.1|
>gi|6678237|ref|NP_033354.1|
>gi|6678525|ref|NP_033508.1|
>gi|6754838|ref|NP_035036.1|
>gi|7305381|ref|NP_038657.1|
>gi|7949105|ref|NP_058048.1|
>gi|8392932|ref|NP_058560.1|
>gi|8567384|ref|NP_059491.1|
>gi|50818199|gb|AAT81589.1|
>gi|51571541|ref|NP_808520.2|
>gi|51593716|gb|AAH80868.1|
>gi|51593743|gb|AAH80698.1|
>gi|55961062|emb|CAI15371.1|
>gi|30088989|gb|AAP13531.1|
>gi|38049505|ref|XP_136621.2|
>gi|55962111|emb|CAI14960.1|
>gi|51705367|ref|XP_129579.3|
>gi|30143280|gb|AAP15181.1|
>gi|57210047|emb|CAI42372.1|
>gi|57210068|emb|CAI42602.1|
>gi|57284001|emb|CAI43070.1|
>gi|56203171|emb|CAI22453.1|
>gi|56203172|emb|CAI22454.1|
>gi|56202454|emb|CAI22773.1|
>gi|56204436|emb|CAI22546.1|
>gi|56204440|emb|CAI22550.1|
>gi|56204438|emb|CAI22548.1|
>gi|56204441|emb|CAI22551.1|
>gi|56204915|emb|CAI22857.1|
>gi|56204916|emb|CAI22858.1|
>gi|56204917|emb|CAI22859.1|
>gi|51709523|ref|XP_149533.3|
>gi|51709826|ref|XP_290113.3|
>gi|13559512|gb|AAK29784.1|
>gi|22074390|gb|AAK93795.1|
>gi|51710949|ref|XP_144267.3|
>gi|51711252|ref|XP_357518.2|
>gi|38085069|ref|XP_357661.1|
>gi|51717095|ref|XP_132755.2|
>gi|51762066|ref|XP_357814.2|
>gi|51762817|ref|XP_145453.4|
>gi|2146941|pir||JC5256
>gi|51764449|ref|XP_488152.1|
>gi|51764875|ref|XP_357992.2|
>gi|51765984|ref|XP_487127.1|
>gi|51766849|ref|XP_484144.1|
>gi|51767597|ref|XP_127250.4|
>gi|51768610|ref|XP_484409.1|
>gi|51769425|ref|XP_489499.1|
>gi|2586117|gb|AAC53554.1|
>gi|2586118|gb|AAC53555.1|

>gi|2586119|gb|AAC53556.1|
>gi|2586120|gb|AAC53557.1|
>gi|2586121|gb|AAC53558.1|
>gi|4235358|gb|AAD13186.1|
>gi|51827859|ref|XP_488170.1|
>gi|51830914|ref|XP_489510.1|
>gi|44955910|ref|NP_976328.1|
>gi|29748022|gb|AAH50941.1|
>gi|32452058|gb|AAH54777.1|
>gi|24079969|gb|AAN46090.1|
>gi|39985995|gb|AAR36864.1|
>gi|39985997|gb|AAR36865.1|
>gi|39985999|gb|AAR36866.1|
>gi|38649147|gb|AAH63281.1|
>gi|6970608|gb|AAF25007.3|
>gi|6630994|gb|AAF19643.1|
>gi|21955122|ref|NP_660280.1|
>gi|346392|pir||S29334
>gi|53854343|gb|AAU95616.1|
>gi|53854346|gb|AAU95617.1|
>gi|52789285|gb|AAH83084.1|
>gi|39753974|gb|AAR30505.1|
>gi|39753976|gb|AAR30506.1|
>gi|26324129|ref|NP_035591.2|
>gi|4503733|ref|NP_001442.1|
>gi|4557651|ref|NP_001529.1|
>gi|4503437|ref|NP_001942.1|
>gi|4758392|ref|NP_004463.1|
>gi|4827004|ref|NP_005060.1|
>gi|4885237|ref|NP_005242.1|
>gi|31652244|ref|NP_005641.1|
>gi|7108362|ref|NP_033664.1|
>gi|13540471|ref|NP_110383.1|
>gi|24307883|ref|NP_001977.1|
>gi|11386173|ref|NP_068770.1|
>gi|6679701|ref|NP_031986.1|
>gi|11386197|ref|NP_036318.1|
>gi|33468883|ref|NP_033845.1|
>gi|6754238|ref|NP_034599.1|
>gi|45544608|ref|NP_035031.1|
>gi|7110673|ref|NP_035051.1|
>gi|45544613|ref|NP_035168.1|
>gi|33859606|ref|NP_035393.1|
>gi|6755612|ref|NP_035576.1|
>gi|33468925|ref|NP_035580.1|
>gi|6756073|ref|NP_035676.1|
>gi|10938014|ref|NP_005345.2|
>gi|11967991|ref|NP_071899.1|
>gi|9506975|ref|NP_062207.1|
>gi|13376382|ref|NP_079191.1|
>gi|14602433|ref|NP_116569.1|
>gi|15011924|ref|NP_005676.2|
>gi|15559211|ref|NP_005603.2|
>gi|20270033|ref|NP_598718.1|
>gi|41281695|ref|NP_620450.1|
>gi|22212925|ref|NP_667344.1|
>gi|22212927|ref|NP_667345.1|
>gi|21070954|ref|NP_620395.1|
>gi|21361278|ref|NP_005644.2|
>gi|21617855|ref|NP_660133.1|
>gi|22122557|ref|NP_666177.1|
>gi|22267470|ref|NP_671754.1|
>gi|22507365|ref|NP_683752.1|
>gi|23346545|ref|NP_694751.1|
>gi|23308720|ref|NP_694538.1|
>gi|27502386|ref|NP_765975.1|
>gi|27502388|ref|NP_765976.1|
>gi|23956388|ref|NP_705816.1|
>gi|26024343|ref|NP_742165.1|
>gi|27819657|ref|NP_766060.1|
>gi|27545319|ref|NP_002690.2|
>gi|28077091|ref|NP_110378.2|
>gi|27923923|ref|NP_778171.1|
>gi|28893581|ref|NP_796372.1|
>gi|41281895|ref|NP_796373.1|
>gi|28912912|ref|NP_796374.1|
>gi|17998698|ref|NP_536721.1|
>gi|30581117|ref|NP_848511.1|
>gi|31652242|ref|NP_852469.1|
>gi|30039710|ref|NP_835455.1|
>gi|30142703|ref|NP_839985.1|
>gi|31543759|ref|NP_035573.2|
>gi|31543269|ref|NP_032662.2|
>gi|31543849|ref|NP_033359.2|
>gi|6754608|ref|NP_034885.1|
>gi|6755604|ref|NP_035571.1|
>gi|31543007|ref|NP_067434.2|
>gi|17298670|ref|NP_387501.1|
>gi|31542226|ref|NP_542372.2|
>gi|26024345|ref|NP_742166.1|
>gi|31560526|ref|NP_035936.2|
>gi|9910472|ref|NP_064319.1|
>gi|31745180|ref|NP_064568.3|
>gi|14994303|ref|NP_033956.1|
>gi|6679042|ref|NP_032714.1|
>gi|7305369|ref|NP_038655.1|
>gi|8393190|ref|NP_059102.1|
>gi|31981130|ref|NP_067341.2|
>gi|21703728|ref|NP_663340.1|
>gi|38156697|ref|NP_937801.1|
>gi|38156699|ref|NP_937802.1|
>gi|38156701|ref|NP_006713.1|
>gi|38156703|ref|NP_937820.1|
>gi|38156705|ref|NP_937821.1|
>gi|38261962|ref|NP_060649.3|
>gi|38372901|ref|NP_612482.2|
>gi|38505159|ref|NP_006593.2|
>gi|38570158|ref|NP_004552.2|
>gi|38569430|ref|NP_079273.2|
>gi|34304020|ref|NP_899071.1|

>gi|6756071|ref|NP_035896.1|
>gi|8393872|ref|NP_058664.1|
>gi|34536825|ref|NP_035167.2|
>gi|37620161|ref|NP_038864.2|
>gi|39573728|ref|NP_945150.1|
>gi|39573726|ref|NP_945151.1|
>gi|39573732|ref|NP_945152.1|
>gi|40254605|ref|NP_033411.2|
>gi|5453982|ref|NP_006252.1|
>gi|6679627|ref|NP_031946.1|
>gi|42263098|ref|NP_954871.1|
>gi|42518080|ref|NP_003193.2|
>gi|54695916|gb|AAV38330.1|
>gi|45505151|ref|NP_995315.1|
>gi|46370086|ref|NP_996923.1|
>gi|47419938|ref|NP_998813.1|
>gi|31982939|ref|NP_776150.2|
>gi|46048458|ref|NP_996806.1|
>gi|46402511|ref|NP_035678.2|
>gi|46430499|ref|NP_068810.2|
>gi|4505073|ref|NP_002350.1|
>gi|17865632|ref|NP_065800.1|
>gi|48976051|ref|NP_066968.2|
>gi|46849702|ref|NP_032725.1|
>gi|51921327|ref|NP_001004164.1|
>gi|51972228|ref|NP_001004311.1|
>gi|52630419|ref|NP_001005291.1|
>gi|52345409|ref|NP_035273.2|
>gi|4758420|ref|NP_004743.1|
>gi|6997249|ref|NP_002959.1|
>gi|4758848|ref|NP_004843.1|
>gi|5453788|ref|NP_006159.1|
>gi|6679833|ref|NP_032049.1|
>gi|6679847|ref|NP_032050.1|
>gi|6679947|ref|NP_032115.1|
>gi|7110597|ref|NP_032119.1|
>gi|6680211|ref|NP_032264.1|
>gi|6679837|ref|NP_032268.1|
>gi|6680223|ref|NP_032275.1|
>gi|6680245|ref|NP_032290.1|
>gi|6679070|ref|NP_032727.1|
>gi|6679341|ref|NP_032878.1|
>gi|6677689|ref|NP_033062.1|
>gi|6678067|ref|NP_033262.1|
>gi|6678071|ref|NP_033263.1|
>gi|6678155|ref|NP_033310.1|
>gi|6678229|ref|NP_033349.1|
>gi|6678231|ref|NP_033350.1|
>gi|6678247|ref|NP_033358.1|
>gi|6678253|ref|NP_033361.1|
>gi|6753128|ref|NP_033846.1|
>gi|6753706|ref|NP_034226.1|
>gi|6753770|ref|NP_034281.1|
>gi|6753902|ref|NP_034556.1|
>gi|6754562|ref|NP_034855.1|
>gi|6754830|ref|NP_035029.1|
>gi|6753876|ref|NP_036143.1|
>gi|6755732|ref|NP_035675.1|
>gi|6753782|ref|NP_035938.1|
>gi|6754252|ref|NP_036069.1|
>gi|7305399|ref|NP_038661.1|
>gi|7305515|ref|NP_038700.1|
>gi|7305571|ref|NP_038720.1|
>gi|7657098|ref|NP_056573.1|
>gi|7705785|ref|NP_057104.1|
>gi|8393874|ref|NP_058663.1|
>gi|7949068|ref|NP_058092.1|
>gi|21281669|ref|NP_060899.1|
>gi|9055318|ref|NP_061279.1|
>gi|9055288|ref|NP_061293.1|
>gi|9055158|ref|NP_061259.1|
>gi|9507179|ref|NP_062380.1|
>gi|9790013|ref|NP_062790.1|
>gi|9789899|ref|NP_062706.1|
>gi|9910330|ref|NP_064379.1|
>gi|6754610|ref|NP_034886.1|
>gi|54673563|gb|AAH37322.3|
>gi|55249567|gb|AAH44632.1|
>gi|55249580|gb|AAH85619.1|
>gi|11693138|ref|NP_071773.1|
>gi|6756073|ref|NP_035676.1|
>gi|22267470|ref|NP_671754.1|
>gi|28077091|ref|NP_110378.2|
>gi|9910472|ref|NP_064319.1|
>gi|6997249|ref|NP_002959.1|
>gi|9937986|ref|NP_064652.1|
>gi|30582265|gb|AAP35359.1|
>gi|30583333|gb|AAP35911.1|
>gi|14249530|ref|NP_116216.1|
>gi|31322942|gb|AAP22284.1|
>gi|55665218|emb|CAH70915.1|
>gi|13161090|gb|AAK13479.1|
>gi|20269856|gb|AAM18073.1|
>gi|32967486|gb|AAP92423.1|
>gi|20513085|gb|AAM21761.1|
>gi|33090193|gb|AAP93895.1|
>gi|33090195|gb|AAP93896.1|
>gi|33383325|gb|AAM10784.1|
>gi|33413338|tpg|DAA01129.1|
>gi|32879989|gb|AAP88825.1|
>gi|8923944|ref|NP_060014.1|
>gi|6679000|ref|NP_032692.1|
>gi|6678251|ref|NP_033360.1|
>gi|12963705|ref|NP_075991.1|
>gi|46877060|ref|NP_997600.1|
>gi|55743092|ref|NP_068745.2|
>gi|12751477|ref|NP_075555.1|
>gi|28278532|gb|AAH46316.1|
>gi|55715387|gb|AAV59183.1|
>gi|55715400|gb|AAV59195.1|

>gi|55715426|gb|AAV59219.1|
>gi|55715439|gb|AAV59231.1|
>gi|23428950|gb|AAM52484.1|
>gi|32967477|gb|AAP92420.1|
>gi|17227094|gb|AAL38007.1|
>gi|46810275|ref|NP_080772.2|
>gi|22128647|ref|NP_666498.1|
>gi|55741482|ref|NP_055840.1|
>gi|7657154|ref|NP_055386.1|
>gi|55777828|gb|AAH47139.1|
>gi|34558615|gb|AAP69907.1|
>gi|55991528|gb|AAH86635.1|
>gi|56181475|gb|AAV83787.1|
>gi|56181477|gb|AAV83788.1|
>gi|36314596|gb|AAM52485.2|
>gi|56549649|ref|NP_002492.2|
>gi|17136214|ref|NP_476575.1|
>gi|17136236|ref|NP_476587.1|
>gi|17136408|ref|NP_476685.1|
>gi|28573543|ref|NP_476845.2|
>gi|45549118|ref|NP_511069.3|
>gi|17137542|ref|NP_477355.1|
>gi|17530843|ref|NP_511079.1|
>gi|17530873|ref|NP_511100.1|
>gi|17136194|ref|NP_477446.1|
>gi|17647995|ref|NP_524124.1|
>gi|17647553|ref|NP_523833.1|
>gi|17737765|ref|NP_524229.1|
>gi|17738065|ref|NP_524415.1|
>gi|17738215|ref|NP_524513.1|
>gi|15679972|gb|AAH14293.1|
>gi|17864398|ref|NP_524783.1|
>gi|17864496|ref|NP_524846.1|
>gi|13277717|gb|AAH03757.1|
>gi|13529389|gb|AAH05435.1|
>gi|24639496|ref|NP_525062.2|
>gi|19484014|gb|AAH25189.1|
>gi|33392684|gb|AAH55414.1|
>gi|29747918|gb|AAH50922.1|
>gi|21357543|ref|NP_649097.1|
>gi|21358029|ref|NP_649336.1|
>gi|34785450|gb|AAH57564.1|
>gi|45552293|ref|NP_995669.1|
>gi|45552407|ref|NP_995726.1|
>gi|24638820|ref|NP_726651.1|
>gi|24649262|ref|NP_732846.1|
>gi|13374569|ref|NP_076228.1|
>gi|30520037|ref|NP_848782.1|
>gi|37703262|gb|AAR01197.1|
>gi|24582911|ref|NP_723414.1|
>gi|6912580|ref|NP_036523.1|
>gi|51094724|gb|EAL23971.1|
>gi|51094745|gb|EAL23992.1|
>gi|51094804|gb|EAL24050.1|
>gi|51094837|gb|EAL24083.1|
>gi|51094950|gb|EAL24195.1|
>gi|51095075|gb|EAL24318.1|
>gi|37727281|gb|AAO41733.1|
>gi|37727285|gb|AAO41735.1|
>gi|37727287|gb|AAO41736.1|
>gi|37727289|gb|AAO41737.1|
>gi|38048023|gb|AAR09914.1|
>gi|48255909|ref|NP_004600.2|
>gi|51873040|ref|NP_009183.2|
>gi|24233564|ref|NP_705781.1|
>gi|31543818|ref|NP_003713.3|
>gi|32490574|ref|NP_870987.1|
>gi|24308067|ref|NP_056261.1|
>gi|6005892|ref|NP_009040.1|
>gi|19923287|ref|NP_008816.2|
>gi|13938577|gb|AAH07439.1|
>gi|34328189|ref|NP_035395.2|
>gi|4503735|ref|NP_001444.1|
>gi|6857806|ref|NP_006553.1|
>gi|11496978|ref|NP_055038.2|
>gi|38373693|ref|NP_003102.1|
>gi|28329416|ref|NP_783552.1|
>gi|18088411|gb|AAH20712.1|
>gi|39795650|gb|AAH64010.1|
>gi|32250400|gb|AAO38028.1|
>gi|32250402|gb|AAO38029.1|
>gi|6677979|ref|NP_031400.1|
>gi|51873965|gb|AAH78444.1|
>gi|6680209|ref|NP_032263.1|
>gi|55420308|gb|AAV52041.1|
>gi|31560537|ref|NP_035680.2|
>gi|21450629|ref|NP_659204.1|
>gi|57163993|ref|NP_002693.2|
>gi|37499116|gb|AAQ91614.1|
>gi|37222749|gb|AAQ90059.1|
>gi|37723146|gb|AAN77899.1|
>gi|37723148|gb|AAN77900.1|
>gi|38679441|gb|AAR26542.1|
>gi|25992667|gb|AAN77205.1|
>gi|25992717|gb|AAN77230.1|
>gi|25992727|gb|AAN77235.1|
>gi|25989643|gb|AAN10254.1|
>gi|39644486|gb|AAH06225.2|
>gi|39644712|gb|AAH07388.1|
>gi|13376298|ref|NP_079141.1|
>gi|38648687|gb|AAH63252.1|
>gi|32879895|gb|AAP88778.1|
>gi|2136186|pir||I38239
>gi|18201913|ref|NP_003584.2|
>gi|4503929|ref|NP_002042.1|
>gi|28411948|ref|NP_786923.1|
>gi|6754402|ref|NP_034721.1|
>gi|55625186|ref|XP_527063.1|
>gi|47604970|ref|NP_001001320.1|
>gi|6005890|ref|NP_009039.1|

>gi|3172042|dbj|BAA28628.1|
>gi|6754860|ref|NP_035049.1|
>gi|4507389|ref|NP_003189.1|
>gi|4503925|ref|NP_002040.1|
>gi|5803117|ref|NP_003457.1|
>gi|8393633|ref|NP_059035.1|
>gi|1079076|pir||A54590
>gi|13477165|gb|AAH05044.1|
>gi|4758948|ref|NP_004566.1|
>gi|19923223|ref|NP_002493.2|
>gi|4758226|ref|NP_004082.1|
>gi|6755716|ref|NP_035657.1|
>gi|1079148|pir||A56199
>gi|49258082|gb|AAH73990.1|
>gi|4503611|ref|NP_001978.1|
>gi|30582589|gb|AAP35521.1|
>gi|21619723|gb|AAH32549.1|
>gi|13592125|ref|NP_112401.1|
>gi|46430889|emb|CAA60050.2|
>gi|53468|emb|CAA49791.1|
>gi|53470|emb|CAA49792.1|
>gi|31417473|gb|AAH06221.2|
>gi|13928750|ref|NP_113741.1|
>gi|34098946|ref|NP_004550.2|
>gi|5453780|ref|NP_006157.1|
>gi|55958373|emb|CAI16533.1|
>gi|2135327|pir||I37451
>gi|516381|emb|CAA52239.1|
>gi|30582249|gb|AAP35351.1|
>gi|539697|pir||D38095
>gi|619884|emb|CAA87440.1|
>gi|28175825|gb|AAH43050.1|
>gi|33859582|ref|NP_035050.1|
>gi|24586029|ref|NP_724482.1|
>gi|33383326|gb|AAM10785.1|
>gi|12669911|ref|NP_005216.1|
>gi|17530947|ref|NP_511169.1|
>gi|6679843|ref|NP_032267.1|
>gi|45433557|ref|NP_032129.2|
>gi|6754182|ref|NP_034549.1|
>gi|4557673|ref|NP_000200.1|
>gi|6679269|ref|NP_032840.1|
>gi|4507441|ref|NP_003211.1|
>gi|4758054|ref|NP_004370.1|
>gi|54695888|gb|AAV38316.1|
>gi|6679044|ref|NP_032715.1|
>gi|10835246|ref|NP_006554.1|
>gi|4758568|ref|NP_004497.1|
>gi|56204991|emb|CAI20208.1|
>gi|6677709|ref|NP_033071.1|
>gi|55962044|emb|CAI18329.1|
>gi|9625000|ref|NP_000514.1|
>gi|11141881|ref|NP_068581.1|
>gi|12597625|ref|NP_036200.2|
>gi|4507161|ref|NP_003099.1|
>gi|5901980|ref|NP_008962.1|
>gi|45269133|ref|NP_003634.2|
>gi|15011898|ref|NP_002911.2|
>gi|6679213|ref|NP_032808.1|
>gi|5729945|ref|NP_006590.1|
>gi|4505401|ref|NP_002500.1|
>gi|6753768|ref|NP_034280.1|
>gi|6681253|ref|NP_031923.1|
>gi|6754454|ref|NP_034765.1|
>gi|14211949|ref|NP_115996.1|
>gi|4758090|ref|NP_004378.1|
>gi|6678241|ref|NP_033353.1|
>gi|10863967|ref|NP_066993.1|
>gi|56417745|emb|CAI21080.1|
>gi|4885201|ref|NP_005220.1|
>gi|17978475|ref|NP_005797.1|
>gi|30354333|gb|AAH51967.1|
>gi|4885645|ref|NP_005418.1|
>gi|2494684|sp|Q92908|GAT6_HUMAN
>gi|54696188|gb|AAV38466.1|
>gi|5453563|ref|NP_006390.1|
>gi|51770181|ref|XP_484622.1|
>gi|6754404|ref|NP_034722.1|
>gi|4505619|ref|NP_002575.1|
>gi|42544127|ref|NP_973726.1|
>gi|5174715|ref|NP_005988.1|
>gi|4757916|ref|NP_004340.1|
>gi|4505955|ref|NP_000297.1|
>gi|6679425|ref|NP_032926.1|
>gi|6679423|ref|NP_032925.1|
>gi|7106383|ref|NP_035271.1|
>gi|6755474|ref|NP_035485.1|
>gi|6678029|ref|NP_033238.1|
>gi|51095123|gb|EAL24366.1|
>gi|12804507|gb|AAH01664.1|
>gi|34874542|ref|XP_217356.2|
>gi|27436887|ref|NP_002577.2|
>gi|27667054|ref|XP_216745.1|
>gi|4506499|ref|NP_003712.1|
>gi|55778696|gb|AAH86457.1|
>gi|33125891|gb|AAL13158.1|
>gi|6679951|ref|NP_032117.1|
>gi|6005904|ref|NP_009045.1|
>gi|22779860|ref|NP_005241.1|
>gi|22122651|ref|NP_666243.1|
>gi|34098933|ref|NP_035667.1|
>gi|2098744|gb|AAB57686.1|
>gi|21618344|ref|NP_036580.2|
>gi|34559414|gb|AAQ75421.1|
>gi|4885219|ref|NP_005229.1|
>gi|4885221|ref|NP_005230.1|
>gi|6912372|ref|NP_036315.1|
>gi|30583541|gb|AAP36015.1|
>gi|6753746|ref|NP_034257.1|
>gi|34559410|gb|AAQ75419.1|

>gi|19353856|gb|AAH24958.1|
>gi|38093375|ref|XP_285621.2|
>gi|47419907|ref|NP_057653.3|
>gi|12314002|emb|CAC12752.1|
>gi|27777636|ref|NP_003191.1|
>gi|13161078|gb|AAK13475.1|
>gi|17737881|ref|NP_524302.1|
>gi|6005934|ref|NP_009053.1|
>gi|4507847|ref|NP_003358.1|
>gi|40538744|ref|NP_112460.1|
>gi|24648774|ref|NP_732647.1|
>gi|18485518|ref|NP_569725.1|
>gi|14714869|gb|AAH10588.1|
>gi|24644977|ref|NP_731211.1|
>gi|57162470|emb|CAI39789.1|
>gi|46370084|ref|NP_996921.1|
>gi|30582557|gb|AAP35505.1|
>gi|55960084|emb|CAI14533.1|
>gi|25282407|ref|NP_006075.3|
>gi|54696386|gb|AAV38565.1|
>gi|55633927|ref|XP_521481.1|
>gi|311926|emb|CAA51017.1|
>gi|4507399|ref|NP_003190.1|
>gi|55962112|emb|CAI14961.1|
>gi|12083653|ref|NP_073187.1|
>gi|38016909|ref|NP_000536.3|
>gi|34328055|ref|NP_035169.1|
>gi|32567784|ref|NP_005986.2|
>gi|18104954|ref|NP_005985.2|
>gi|30583091|gb|AAP35790.1|
>gi|5032143|ref|NP_005672.1|
>gi|11056038|ref|NP_055403.2|
>gi|55154437|gb|AAH85261.1|
>gi|31981873|ref|NP_033899.2|
>gi|23308601|ref|NP_034788.1|
>gi|8051593|ref|NP_005245.2|
>gi|8051595|ref|NP_057738.1|
>gi|7305415|ref|NP_038919.1|
>gi|6678245|ref|NP_033357.1|
>gi|16507249|ref|NP_443734.1|
>gi|56207083|emb|CAI25052.1|
>gi|2498016|sp|Q61079|SIM2_MOUSE
>gi|6754612|ref|NP_034887.1|
>gi|2498127|sp|Q03347|RUN1_MOUSE
>gi|6678255|ref|NP_033362.1|
>gi|6680742|ref|NP_031526.1|
>gi|6753298|ref|NP_033951.1|
>gi|7705917|ref|NP_057353.1|
>gi|24652319|ref|NP_724882.1|
>gi|13507624|ref|NP_109618.1|
>gi|19548720|gb|AAL90757.1|
>gi|4885073|ref|NP_005162.1|
>gi|7661958|ref|NP_055554.1|
>gi|4507339|ref|NP_003172.1|
>gi|17864246|ref|NP_524677.1|
>gi|6679807|ref|NP_032052.1|
>gi|4502763|ref|NP_001795.1|
>gi|17981751|ref|NP_524992.1|
>gi|5032173|ref|NP_005645.1|
>gi|57527555|ref|NP_001009708.1|
>gi|24286752|gb|AAN46737.1|
>gi|481008|pir||S37616
>gi|18490808|gb|AAH22231.1|
>gi|6679271|ref|NP_032841.1|
>gi|1729777|emb|CAA69220.1|
>gi|14603316|gb|AAH10115.1|
>gi|51095124|gb|EAL24367.1|
>gi|55655872|ref|XP_531526.1|
>gi|27464836|gb|AAO16209.1|
>gi|485451|pir||S28823
>gi|422437|pir||S28820
>gi|4507175|ref|NP_003111.1|
>gi|4507177|ref|NP_003112.1|
>gi|14043975|gb|AAH07921.1|
>gi|157196|gb|AAA28451.1|
>gi|7305023|ref|NP_038536.1|
>gi|54870|emb|CAA35541.1|
>gi|1020401|dbj|BAA10905.1|
>gi|19923352|ref|NP_006158.2|
>gi|4755128|ref|NP_004015.2|
>gi|6654638|ref|NP_000429.2|
>gi|7524356|ref|NP_039230.1|
>gi|6754992|ref|NP_035171.1|
>gi|7524363|ref|NP_039246.1|
>gi|6754990|ref|NP_035170.1|
>gi|17647491|ref|NP_523863.1|
>gi|28573681|ref|NP_788422.1|
>gi|27886561|ref|NP_775188.1|
>gi|6680165|ref|NP_032239.1|
>gi|32879889|gb|AAP88775.1|
>gi|5174545|ref|NP_005911.1|
>gi|6754740|ref|NP_034957.1|
>gi|220519|dbj|BAA14203.1|
>gi|56204010|emb|CAI23240.1|
>gi|4505379|ref|NP_003195.1|
>gi|34328941|ref|NP_620410.3|
>gi|49119071|gb|AAH47878.1|
>gi|23271033|gb|AAH33290.1|
>gi|7446945|pir||T13348
>gi|24583942|ref|NP_523558.2|
>gi|33859480|ref|NP_032322.1|
>gi|7106433|ref|NP_035695.1|
>gi|19526461|ref|NP_035696.1|
>gi|6671541|ref|NP_031468.1|
>gi|56785|emb|CAA47715.1|
>gi|42560248|ref|NP_002692.2|
>gi|19921704|ref|NP_610232.1|
>gi|53482|emb|CAA41004.1|
>gi|297789|emb|CAA77953.1|
>gi|1082825|pir||A54687

>gi|4757918|ref|NP_004341.1|
>gi|33563254|ref|NP_034356.1|
>gi|55742855|ref|NP_001002977.1|
>gi|87016|pir||S05380
>gi|5802980|ref|NP_006847.1|
>gi|4502765|ref|NP_001256.1|
>gi|4507103|ref|NP_003074.1|
>gi|4505963|ref|NP_000298.1|
>gi|88885|pir||A41285
>gi|21687100|ref|NP_660303.1|
>gi|50541959|ref|NP_001002295.1|
>gi|1845570|dbj|BAA11334.1|
>gi|35050|emb|CAA42230.1|
>gi|6677853|ref|NP_033149.1|
>gi|4557755|ref|NP_000239.1|
>gi|11276067|ref|NP_003059.1|
>gi|8617|emb|CAA46889.1|
>gi|20127158|ref|NP_003101.2|
>gi|24653574|ref|NP_523739.2|
>gi|13879280|gb|AAH06612.1|
>gi|5174589|ref|NP_005946.1|
>gi|42518074|ref|NP_963964.1|
>gi|42714657|ref|NP_963965.1|
>gi|18129690|ref|NP_060958.2|
>gi|6678305|ref|NP_033387.1|
>gi|602343|emb|CAA84513.1|
>gi|4505389|ref|NP_002496.1|
>gi|107921|pir||S12788
>gi|107706|pir||S22939
>gi|4507955|ref|NP_003394.1|
>gi|7019549|ref|NP_037483.1|
>gi|37060|emb|CAA40683.1|
>gi|8659574|ref|NP_006512.2|
>gi|23272668|gb|AAH35607.1|
>gi|19263797|gb|AAH25171.1|
>gi|22547195|ref|NP_004167.3|
>gi|21356717|ref|NP_650780.1|
>gi|1168465|sp|P34056|AP2A_MOUSE
>gi|50002|emb|CAA40331.1|
>gi|6755070|ref|NP_035227.1|
>gi|29244208|ref|NP_808401.1|
>gi|32879879|gb|AAP88770.1|
>gi|7662342|ref|NP_055758.1|
>gi|51491831|ref|NP_001003905.1|
>gi|33772632|gb|AAQ54689.1|
>gi|2136301|pir||I59340
>gi|4507447|ref|NP_003214.1|
>gi|1082848|pir||A55237
>gi|55959464|emb|CAI15194.1|
>gi|913967|gb|AAB33999.1|
>gi|5669546|gb|AAD46381.1|
>gi|6678457|ref|NP_033468.1|
>gi|42544172|ref|NP_973728.1|
>gi|7677546|gb|AAF67196.1|
>gi|6857816|ref|NP_006865.1|
>gi|6681247|ref|NP_031919.1|
>gi|6678243|ref|NP_033356.1|
>gi|6679841|ref|NP_032048.1|
>gi|477739|pir||B47746
>gi|477968|pir||C47746
>gi|478081|pir||D47746
>gi|478145|pir||E47746
>gi|478187|pir||F47746
>gi|6755134|ref|NP_035269.1|
>gi|15617374|emb|CAC69875.1|
>gi|6678884|ref|NP_032627.1|
>gi|219455|dbj|BAA01426.1|
>gi|46275824|ref|NP_034582.1|
>gi|6831586|sp|Q61985|NFL1_MOUSE
>gi|6754604|ref|NP_034881.1|
>gi|11139091|gb|AAG31604.1|
>gi|1684822|gb|AAD03154.1|
>gi|17298672|ref|NP_291043.1|
>gi|129650|sp|P24610|PAX3_MOUSE
>gi|53492|emb|CAA41009.1|
>gi|548953|sp|Q06831|SOX4_MOUSE
>gi|6679068|ref|NP_032726.1|
>gi|109950|pir||B40583
>gi|6755130|ref|NP_035267.1|
>gi|6753898|ref|NP_034576.1|
>gi|6679831|ref|NP_032286.1|
>gi|1083353|pir||A54258
>gi|543305|pir||S41873
>gi|4504367|ref|NP_003856.1|
>gi|14714932|gb|AAH10623.1|
>gi|51468067|ref|NP_001003845.1|
>gi|16877941|gb|AAH17194.1|
>gi|15628025|ref|NP_258437.1|
>gi|11230810|ref|NP_068699.1|
>gi|20806530|ref|NP_033950.1|
>gi|24659252|gb|AAH38995.1|
>gi|54697012|gb|AAV38878.1|
>gi|157427|gb|AAA28533.1|
>gi|157429|gb|AAA28534.1|
>gi|423840|pir||D46178
>gi|6042186|gb|AAF02178.1|
>gi|641810|emb|CAA56038.1|
>gi|424034|pir||A46403
>gi|38201614|ref|NP_937848.1|
>gi|24586667|ref|NP_733752.1|
>gi|18104950|ref|NP_542377.1|
>gi|28201860|sp|Q9BZW0|PO5N_HUMAN
>gi|38201717|ref|NP_938195.1|
>gi|4826912|ref|NP_005020.1|
>gi|542553|pir||C36901
>gi|4507169|ref|NP_003103.1|
>gi|55958385|emb|CAI15108.1|
>gi|6755610|ref|NP_035575.1|
>gi|6680205|ref|NP_032261.1|
>gi|200044|gb|AAA39817.1|

>gi|24234711|ref|NP_700476.1|
>gi|24234708|ref|NP_700475.1|
>gi|21361183|ref|NP_000316.2|
>gi|4505007|ref|NP_002307.1|
>gi|17864454|ref|NP_524820.1|
>gi|18201896|ref|NP_542449.1|
>gi|5442096|gb|AAD43250.1|
>gi|5442100|gb|AAD43252.1|
>gi|6760434|gb|AAF28350.1|
>gi|7274372|gb|AAF44742.1|
>gi|7274374|gb|AAF44743.1|
>gi|280601|pir||B43698
>gi|57209341|emb|CAI40795.1|
>gi|21956641|ref|NP_665804.1|
>gi|40254321|ref|NP_796237.2|
>gi|30179903|ref|NP_055402.2|
>gi|27721277|ref|XP_236605.1|
>gi|55957380|emb|CAI14398.1|
>gi|4049635|gb|AAC97604.1|
>gi|4758506|ref|NP_004812.1|
>gi|20380335|gb|AAH27533.1|
>gi|56207887|emb|CAI18708.1|
>gi|1911185|gb|AAB50574.1|
>gi|6755730|ref|NP_035674.1|
>gi|27886526|ref|NP_775322.1|
>gi|33772634|gb|AAQ54690.1|
>gi|31455582|gb|AAN85556.1|
>gi|30583371|gb|AAP35930.1|
>gi|7263926|emb|CAB81658.1|
>gi|4827024|ref|NP_005140.1|
>gi|33150560|gb|AAP97158.1|
>gi|55957250|emb|CAI12665.1|
>gi|18104952|ref|NP_542378.1|
>gi|260574|gb|AAB24289.1|
>gi|10863885|ref|NP_004339.1|
>gi|30061556|ref|NP_005625.2|
>gi|46909569|ref|NP_032118.2|
>gi|5327036|emb|CAB46198.1|
>gi|22477471|gb|AAH36689.1|
>gi|28195386|ref|NP_003097.1|
>gi|6755576|ref|NP_035545.1|
>gi|28827782|ref|NP_055850.1|
>gi|20270039|ref|NP_619727.1|
>gi|17864538|ref|NP_524876.1|
>gi|14669161|gb|AAK71346.1|
>gi|15426512|gb|AAH13362.1|
>gi|25136577|gb|AAN65622.1|
>gi|6941956|gb|AAF32274.1|
>gi|7657299|ref|NP_055368.1|
>gi|18378731|ref|NP_064620.2|
>gi|23346597|ref|NP_694794.1|
>gi|10697003|emb|CAC12700.1|
>gi|19923350|ref|NP_006153.2|
>gi|24430417|dbj|BAC22610.1|
>gi|22748707|ref|NP_689536.1|
>gi|10864025|ref|NP_067035.1|
>gi|13543512|gb|AAH05914.1|
>gi|19923830|ref|NP_065385.2|
>gi|18375619|ref|NP_542986.1|
>gi|30794490|ref|NP_059101.1|
>gi|21361137|ref|NP_002495.2|
>gi|56204607|emb|CAI20506.1|
>gi|21361411|ref|NP_036389.2|
>gi|12331278|emb|CAC24700.1|
>gi|8980369|emb|CAB96873.1|
>gi|21699078|ref|NP_660328.1|
>gi|33469974|ref|NP_877962.1|
>gi|22507329|ref|NP_683754.1|
>gi|31981462|ref|NP_035677.2|
>gi|22122569|ref|NP_666186.1|
>gi|23397423|ref|NP_694878.1|
>gi|31377777|ref|NP_055175.2|
>gi|34328232|ref|NP_062299.2|
>gi|33989348|gb|AAH52041.2|
>gi|39795673|gb|AAH64038.1|
>gi|31982405|ref|NP_031918.2|
>gi|5454112|ref|NP_006277.1|
>gi|24647326|ref|NP_650516.2|
>gi|23491772|dbj|BAC19830.1|
>gi|14211556|gb|AAK55760.1|
>gi|11056048|ref|NP_067063.1|
>gi|51477359|ref|XP_497098.1|
>gi|29746069|ref|XP_293396.1|
>gi|51477313|ref|XP_498398.1|
>gi|55663811|emb|CAH70589.1|
>gi|55665866|emb|CAH70633.1|
>gi|12803265|gb|AAH02447.1|
>gi|56204439|emb|CAI22549.1|
>gi|21618325|ref|NP_004464.2|
>gi|56786157|ref|NP_031374.2|
>gi|15149472|ref|NP_079521.1|
>gi|51702521|ref|NP_005595.2|
>gi|11496974|ref|NP_068351.1|
>gi|25992723|gb|AAN77233.1|
>gi|55960186|emb|CAI17342.1|
>gi|55961323|emb|CAI17406.1|
>gi|55962077|emb|CAI17540.1|
>gi|22779928|ref|NP_036150.1|
>gi|57210045|emb|CAI42694.1|
>gi|38074680|ref|XP_357189.1|
>gi|56204484|emb|CAI23136.1|
>gi|51711146|ref|XP_485662.1|
>gi|51766794|ref|XP_484113.1|
>gi|25047127|ref|XP_204339.1|
>gi|54311320|gb|AAH33891.1|
>gi|22538422|ref|NP_001871.2|
>gi|6031205|ref|NP_006472.1|
>gi|18375621|ref|NP_542987.1|
>gi|13277554|gb|AAH03685.1|
>gi|13161087|gb|AAK13478.1|

>gi|13543913|gb|AAH06101.1|
>gi|25992725|gb|AAN77234.1|
>gi|55631028|ref|XP_519886.1|
>gi|21313208|ref|NP_083895.1|
>gi|17136408|ref|NP_476685.1|
>gi|17647553|ref|NP_523833.1|
>gi|51094649|gb|EAL23900.1|
>gi|19923287|ref|NP_008816.2|
>gi|38648687|gb|AAH63252.1|
>gi|21687100|ref|NP_660303.1|
>gi|11276067|ref|NP_003059.1|
>gi|7662342|ref|NP_055758.1|
>gi|40254321|ref|NP_796237.2|
>gi|6755576|ref|NP_035545.1|
>gi|28827782|ref|NP_055850.1|
>gi|30424679|ref|NP_777277.1|
>UNIPROT|O00146 SPTR:O00146
>UNIPROT|O00287 SPTR:O00287
>UNIPROT|O00409 SPTR:O00409
>UNIPROT|O00712 SPTR:O00712
>UNIPROT|O00716 SPTR:O00716
>UNIPROT|O01667 SPTR:O01667
>UNIPROT|O02289 SPTR:O02289
>UNIPROT|O08537 SPTR:O08537
>UNIPROT|O08755 SPTR:O08755
>UNIPROT|O09100 SPTR:O09100
>UNIPROT|O09102 SPTR:O09102
>UNIPROT|O14503 SPTR:O14503
>UNIPROT|O14529 SPTR:O14529
>UNIPROT|O14593 SPTR:O14593
>UNIPROT|O14753 SPTR:O14753
>UNIPROT|O14813 SPTR:O14813
>UNIPROT|O14901 SPTR:O14901
>UNIPROT|O14948 SPTR:O14948
>UNIPROT|O15178 SPTR:O15178
>UNIPROT|O15187 SPTR:O15187
>UNIPROT|O15266 SPTR:O15266
>UNIPROT|O15415 SPTR:O15415
>UNIPROT|O15419 SPTR:O15419
>UNIPROT|O15499 SPTR:O15499
>UNIPROT|O15516 SPTR:O15516
>UNIPROT|O16011 SPTR:O16011
>UNIPROT|O16117 SPTR:O16117
>UNIPROT|O17427 SPTR:O17427
>UNIPROT|O18381 SPTR:O18381
>UNIPROT|O18660 SPTR:O18660
>UNIPROT|O35137 SPTR:O35137
>UNIPROT|O35185 SPTR:O35185
>UNIPROT|O35233 SPTR:O35233
>UNIPROT|O35261 SPTR:O35261
>UNIPROT|O35275 SPTR:O35275
>UNIPROT|O35437 SPTR:O35437
>UNIPROT|O35473 SPTR:O35473
>UNIPROT|O35819 SPTR:O35819
>UNIPROT|O35905 SPTR:O35905
>UNIPROT|O35906 SPTR:O35906
>UNIPROT|O35908 SPTR:O35908
>UNIPROT|O43186 SPTR:O43186
>UNIPROT|O43245 SPTR:O43245
>UNIPROT|O43316 SPTR:O43316
>UNIPROT|O43439 SPTR:O43439
>UNIPROT|O43593 SPTR:O43593
>UNIPROT|O43638 SPTR:O43638
>UNIPROT|O43918 SPTR:O43918
>UNIPROT|O44080 SPTR:O44080
>UNIPROT|O46339 SPTR:O46339
>UNIPROT|O54790 SPTR:O54790
>UNIPROT|O54845 SPTR:O54845
>UNIPROT|O54962 SPTR:O54962
>UNIPROT|O55087 SPTR:O55087
>UNIPROT|O55140 SPTR:O55140
>UNIPROT|O55170 SPTR:O55170
>UNIPROT|O55199 SPTR:O55199
>UNIPROT|O60435 SPTR:O60435
>UNIPROT|O60479 SPTR:O60479
>UNIPROT|O60519 SPTR:O60519
>UNIPROT|O60548 SPTR:O60548
>UNIPROT|O60663 SPTR:O60663
>UNIPROT|O60682 SPTR:O60682
>UNIPROT|O60861 SPTR:O60861
>UNIPROT|O60870 SPTR:O60870
>UNIPROT|O60925 SPTR:O60925
>UNIPROT|O70192 SPTR:O70192
>UNIPROT|O70273 SPTR:O70273
>UNIPROT|O70494 SPTR:O70494
>UNIPROT|O70608 SPTR:O70608
>UNIPROT|O75081 SPTR:O75081
>UNIPROT|O75461 SPTR:O75461
>UNIPROT|O75478 SPTR:O75478
>UNIPROT|O75528 SPTR:O75528
>UNIPROT|O75603 SPTR:O75603
>UNIPROT|O75618 SPTR:O75618
>UNIPROT|O75683 SPTR:O75683
>UNIPROT|O75698 SPTR:O75698
>UNIPROT|O75717 SPTR:O75717
>UNIPROT|O75971 SPTR:O75971
>UNIPROT|O76906 SPTR:O76906
>UNIPROT|O77215 SPTR:O77215
>UNIPROT|O77238 SPTR:O77238
>UNIPROT|O77459 SPTR:O77459
>UNIPROT|O88181 SPTR:O88181
>UNIPROT|O88443 SPTR:O88443
>UNIPROT|O88470 SPTR:O88470
>UNIPROT|O88573 SPTR:O88573
>UNIPROT|O88609 SPTR:O88609
>UNIPROT|O88621 SPTR:O88621
>UNIPROT|O88880 SPTR:O88880
>UNIPROT|O88942 SPTR:O88942
>UNIPROT|O89087 SPTR:O89087
>UNIPROT|O89088 SPTR:O89088

>UNIPROT|O89090 SPTR:O89090
>UNIPROT|O95361 SPTR:O95361
>UNIPROT|O95367 SPTR:O95367
>UNIPROT|O95416 SPTR:O95416
>UNIPROT|O95443 SPTR:O95443
>UNIPROT|O95480 SPTR:O95480
>UNIPROT|O95619 SPTR:O95619
>UNIPROT|O95644 SPTR:O95644
>UNIPROT|O95718 SPTR:O95718
>UNIPROT|O95935 SPTR:O95935
>UNIPROT|O95936 SPTR:O95936
>UNIPROT|O95947 SPTR:O95947
>UNIPROT|O95983 SPTR:O95983
>UNIPROT|O95997 SPTR:O95997
>UNIPROT|O96004 SPTR:O96004
>UNIPROT|O96680 SPTR:O96680
>UNIPROT|P01101 SPTR:P01101
>UNIPROT|P01106 SPTR:P01106
>UNIPROT|P01108 SPTR:P01108
>UNIPROT|P02833 SPTR:P02833
>UNIPROT|P03069 SPTR:P03069
>UNIPROT|P04053 SPTR:P04053
>UNIPROT|P04150 SPTR:P04150
>UNIPROT|P04197 SPTR:P04197
>UNIPROT|P04553 SPTR:P04553
>UNIPROT|P05084 SPTR:P05084
>UNIPROT|P05554 SPTR:P05554
>UNIPROT|P05627 SPTR:P05627
>UNIPROT|P06401 SPTR:P06401
>UNIPROT|P06798 SPTR:P06798
>UNIPROT|P06843 SPTR:P06843
>UNIPROT|P06876 SPTR:P06876
>UNIPROT|P07248 SPTR:P07248
>UNIPROT|P07261 SPTR:P07261
>UNIPROT|P07269 SPTR:P07269
>UNIPROT|P07270 SPTR:P07270
>UNIPROT|P07272 SPTR:P07272
>UNIPROT|P07548 SPTR:P07548
>UNIPROT|P07664 SPTR:P07664
>UNIPROT|P08046 SPTR:P08046
>UNIPROT|P08152 SPTR:P08152
>UNIPROT|P08235 SPTR:P08235
>UNIPROT|P08651 SPTR:P08651
>UNIPROT|P08970 SPTR:P08970
>UNIPROT|P09016 SPTR:P09016
>UNIPROT|P09017 SPTR:P09017
>UNIPROT|P09067 SPTR:P09067
>UNIPROT|P09079 SPTR:P09079
>UNIPROT|P09083 SPTR:P09083
>UNIPROT|P09084 SPTR:P09084
>UNIPROT|P09086 SPTR:P09086
>UNIPROT|P09416 SPTR:P09416
>UNIPROT|P09450 SPTR:P09450
>UNIPROT|P09629 SPTR:P09629
>UNIPROT|P10070 SPTR:P10070
>UNIPROT|P10071 SPTR:P10071
>UNIPROT|P10083 SPTR:P10083
>UNIPROT|P10084 SPTR:P10084
>UNIPROT|P10105 SPTR:P10105
>UNIPROT|P10158 SPTR:P10158
>UNIPROT|P10180 SPTR:P10180
>UNIPROT|P10275 SPTR:P10275
>UNIPROT|P10276 SPTR:P10276
>UNIPROT|P10588 SPTR:P10588
>UNIPROT|P10589 SPTR:P10589
>UNIPROT|P10826 SPTR:P10826
>UNIPROT|P10827 SPTR:P10827
>UNIPROT|P10828 SPTR:P10828
>UNIPROT|P10961 SPTR:P10961
>UNIPROT|P11308 SPTR:P11308
>UNIPROT|P11420 SPTR:P11420
>UNIPROT|P11473 SPTR:P11473
>UNIPROT|P11474 SPTR:P11474
>UNIPROT|P11536 SPTR:P11536
>UNIPROT|P12383 SPTR:P12383
>UNIPROT|P12524 SPTR:P12524
>UNIPROT|P12525 SPTR:P12525
>UNIPROT|P12980 SPTR:P12980
>UNIPROT|P13053 SPTR:P13053
>UNIPROT|P13056 SPTR:P13056
>UNIPROT|P13096 SPTR:P13096
>UNIPROT|P13097 SPTR:P13097
>UNIPROT|P13098 SPTR:P13098
>UNIPROT|P13297 SPTR:P13297
>UNIPROT|P13378 SPTR:P13378
>UNIPROT|P13469 SPTR:P13469
>UNIPROT|P13574 SPTR:P13574
>UNIPROT|P14651 SPTR:P14651
>UNIPROT|P14652 SPTR:P14652
>UNIPROT|P14653 SPTR:P14653
>UNIPROT|P14859 SPTR:P14859
>UNIPROT|P15036 SPTR:P15036
>UNIPROT|P15066 SPTR:P15066
>UNIPROT|P15173 SPTR:P15173
>UNIPROT|P15307 SPTR:P15307
>UNIPROT|P15314 SPTR:P15314
>UNIPROT|P15407 SPTR:P15407
>UNIPROT|P15408 SPTR:P15408
>UNIPROT|P15806 SPTR:P15806
>UNIPROT|P15918 SPTR:P15918
>UNIPROT|P15923 SPTR:P15923
>UNIPROT|P15976 SPTR:P15976
>UNIPROT|P16220 SPTR:P16220
>UNIPROT|P16241 SPTR:P16241
>UNIPROT|P16455 SPTR:P16455
>UNIPROT|P16951 SPTR:P16951
>UNIPROT|P17096 SPTR:P17096
>UNIPROT|P17208 SPTR:P17208
>UNIPROT|P17433 SPTR:P17433
>UNIPROT|P17481 SPTR:P17481

>UNIPROT|P17482 SPTR:P17482
>UNIPROT|P17483 SPTR:P17483
>UNIPROT|P17509 SPTR:P17509
>UNIPROT|P17542 SPTR:P17542
>UNIPROT|P17544 SPTR:P17544
>UNIPROT|P17676 SPTR:P17676
>UNIPROT|P17679 SPTR:P17679
>UNIPROT|P17861 SPTR:P17861
>UNIPROT|P17947 SPTR:P17947
>UNIPROT|P18146 SPTR:P18146
>UNIPROT|P18494 SPTR:P18494
>UNIPROT|P18846 SPTR:P18846
>UNIPROT|P18847 SPTR:P18847
>UNIPROT|P18911 SPTR:P18911
>UNIPROT|P19091 SPTR:P19091
>UNIPROT|P19360 SPTR:P19360
>UNIPROT|P19484 SPTR:P19484
>UNIPROT|P19532 SPTR:P19532
>UNIPROT|P19538 SPTR:P19538
>UNIPROT|P19541 SPTR:P19541
>UNIPROT|P19544 SPTR:P19544
>UNIPROT|P19793 SPTR:P19793
>UNIPROT|P19838 SPTR:P19838
>UNIPROT|P19880 SPTR:P19880
>UNIPROT|P20050 SPTR:P20050
>UNIPROT|P20105 SPTR:P20105
>UNIPROT|P20263 SPTR:P20263
>UNIPROT|P20264 SPTR:P20264
>UNIPROT|P20265 SPTR:P20265
>UNIPROT|P20267 SPTR:P20267
>UNIPROT|P20482 SPTR:P20482
>UNIPROT|P20719 SPTR:P20719
>UNIPROT|P20787 SPTR:P20787
>UNIPROT|P21190 SPTR:P21190
>UNIPROT|P21272 SPTR:P21272
>UNIPROT|P21952 SPTR:P21952
>UNIPROT|P22149 SPTR:P22149
>UNIPROT|P22361 SPTR:P22361
>UNIPROT|P22449 SPTR:P22449
>UNIPROT|P22712 SPTR:P22712
>UNIPROT|P22736 SPTR:P22736
>UNIPROT|P22816 SPTR:P22816
>UNIPROT|P23204 SPTR:P23204
>UNIPROT|P23409 SPTR:P23409
>UNIPROT|P23441 SPTR:P23441
>UNIPROT|P23511 SPTR:P23511
>UNIPROT|P23512 SPTR:P23512
>UNIPROT|P23611 SPTR:P23611
>UNIPROT|P23708 SPTR:P23708
>UNIPROT|P23759 SPTR:P23759
>UNIPROT|P23760 SPTR:P23760
>UNIPROT|P23769 SPTR:P23769
>UNIPROT|P23771 SPTR:P23771
>UNIPROT|P23899 SPTR:P23899
>UNIPROT|P24610 SPTR:P24610
>UNIPROT|P24898 SPTR:P24898
>UNIPROT|P25208 SPTR:P25208
>UNIPROT|P25302 SPTR:P25302
>UNIPROT|P25579 SPTR:P25579
>UNIPROT|P25583 SPTR:P25583
>UNIPROT|P25611 SPTR:P25611
>UNIPROT|P25656 SPTR:P25656
>UNIPROT|P25799 SPTR:P25799
>UNIPROT|P25800 SPTR:P25800
>UNIPROT|P25992 SPTR:P25992
>UNIPROT|P26343 SPTR:P26343
>UNIPROT|P26367 SPTR:P26367
>UNIPROT|P26687 SPTR:P26687
>UNIPROT|P27540 SPTR:P27540
>UNIPROT|P27699 SPTR:P27699
>UNIPROT|P27782 SPTR:P27782
>UNIPROT|P27797 SPTR:P27797
>UNIPROT|P27889 SPTR:P27889
>UNIPROT|P28033 SPTR:P28033
>UNIPROT|P28069 SPTR:P28069
>UNIPROT|P28159 SPTR:P28159
>UNIPROT|P28324 SPTR:P28324
>UNIPROT|P28347 SPTR:P28347
>UNIPROT|P28700 SPTR:P28700
>UNIPROT|P28702 SPTR:P28702
>UNIPROT|P29374 SPTR:P29374
>UNIPROT|P29375 SPTR:P29375
>UNIPROT|P29506 SPTR:P29506
>UNIPROT|P29617 SPTR:P29617
>UNIPROT|P29774 SPTR:P29774
>UNIPROT|P30051 SPTR:P30051
>UNIPROT|P30561 SPTR:P30561
>UNIPROT|P31249 SPTR:P31249
>UNIPROT|P31268 SPTR:P31268
>UNIPROT|P31271 SPTR:P31271
>UNIPROT|P31273 SPTR:P31273
>UNIPROT|P31277 SPTR:P31277
>UNIPROT|P31310 SPTR:P31310
>UNIPROT|P31311 SPTR:P31311
>UNIPROT|P31316 SPTR:P31316
>UNIPROT|P31360 SPTR:P31360
>UNIPROT|P31361 SPTR:P31361
>UNIPROT|P31629 SPTR:P31629
>UNIPROT|P32027 SPTR:P32027
>UNIPROT|P32114 SPTR:P32114
>UNIPROT|P32115 SPTR:P32115
>UNIPROT|P32182 SPTR:P32182
>UNIPROT|P32183 SPTR:P32183
>UNIPROT|P32242 SPTR:P32242
>UNIPROT|P32243 SPTR:P32243
>UNIPROT|P32314 SPTR:P32314
>UNIPROT|P32608 SPTR:P32608
>UNIPROT|P33122 SPTR:P33122
>UNIPROT|P33200 SPTR:P33200
>UNIPROT|P33242 SPTR:P33242

>UNIPROT|P33244 SPTR:P33244
>UNIPROT|P34056 SPTR:P34056
>UNIPROT|P34161 SPTR:P34161
>UNIPROT|P34707 SPTR:P34707
>UNIPROT|P35227 SPTR:P35227
>UNIPROT|P35233 SPTR:P35233
>UNIPROT|P35396 SPTR:P35396
>UNIPROT|P35428 SPTR:P35428
>UNIPROT|P35452 SPTR:P35452
>UNIPROT|P35453 SPTR:P35453
>UNIPROT|P35548 SPTR:P35548
>UNIPROT|P35583 SPTR:P35583
>UNIPROT|P35584 SPTR:P35584
>UNIPROT|P35638 SPTR:P35638
>UNIPROT|P35680 SPTR:P35680
>UNIPROT|P35711 SPTR:P35711
>UNIPROT|P35869 SPTR:P35869
>UNIPROT|P37231 SPTR:P37231
>UNIPROT|P37238 SPTR:P37238
>UNIPROT|P37242 SPTR:P37242
>UNIPROT|P37243 SPTR:P37243
>UNIPROT|P37244 SPTR:P37244
>UNIPROT|P37275 SPTR:P37275
>UNIPROT|P38128 SPTR:P38128
>UNIPROT|P38532 SPTR:P38532
>UNIPROT|P38533 SPTR:P38533
>UNIPROT|P38699 SPTR:P38699
>UNIPROT|P38749 SPTR:P38749
>UNIPROT|P38830 SPTR:P38830
>UNIPROT|P38845 SPTR:P38845
>UNIPROT|P38889 SPTR:P38889
>UNIPROT|P38907 SPTR:P38907
>UNIPROT|P39001 SPTR:P39001
>UNIPROT|P39521 SPTR:P39521
>UNIPROT|P39720 SPTR:P39720
>UNIPROT|P39943 SPTR:P39943
>UNIPROT|P40424 SPTR:P40424
>UNIPROT|P40425 SPTR:P40425
>UNIPROT|P40426 SPTR:P40426
>UNIPROT|P40427 SPTR:P40427
>UNIPROT|P40489 SPTR:P40489
>UNIPROT|P40630 SPTR:P40630
>UNIPROT|P40645 SPTR:P40645
>UNIPROT|P40646 SPTR:P40646
>UNIPROT|P40763 SPTR:P40763
>UNIPROT|P41161 SPTR:P41161
>UNIPROT|P41164 SPTR:P41164
>UNIPROT|P41212 SPTR:P41212
>UNIPROT|P41225 SPTR:P41225
>UNIPROT|P41738 SPTR:P41738
>UNIPROT|P41778 SPTR:P41778
>UNIPROT|P41969 SPTR:P41969
>UNIPROT|P41971 SPTR:P41971
>UNIPROT|P42128 SPTR:P42128
>UNIPROT|P42224 SPTR:P42224
>UNIPROT|P42226 SPTR:P42226
>UNIPROT|P42227 SPTR:P42227
>UNIPROT|P42228 SPTR:P42228
>UNIPROT|P42229 SPTR:P42229
>UNIPROT|P42230 SPTR:P42230
>UNIPROT|P42232 SPTR:P42232
>UNIPROT|P42571 SPTR:P42571
>UNIPROT|P42582 SPTR:P42582
>UNIPROT|P42586 SPTR:P42586
>UNIPROT|P42838 SPTR:P42838
>UNIPROT|P43429 SPTR:P43429
>UNIPROT|P43634 SPTR:P43634
>UNIPROT|P43680 SPTR:P43680
>UNIPROT|P45448 SPTR:P45448
>UNIPROT|P46153 SPTR:P46153
>UNIPROT|P46684 SPTR:P46684
>UNIPROT|P47239 SPTR:P47239
>UNIPROT|P47242 SPTR:P47242
>UNIPROT|P47974 SPTR:P47974
>UNIPROT|P47988 SPTR:P47988
>UNIPROT|P48382 SPTR:P48382
>UNIPROT|P48432 SPTR:P48432
>UNIPROT|P48985 SPTR:P48985
>UNIPROT|P49116 SPTR:P49116
>UNIPROT|P49335 SPTR:P49335
>UNIPROT|P49640 SPTR:P49640
>UNIPROT|P49698 SPTR:P49698
>UNIPROT|P50220 SPTR:P50220
>UNIPROT|P50222 SPTR:P50222
>UNIPROT|P50539 SPTR:P50539
>UNIPROT|P50549 SPTR:P50549
>UNIPROT|P50553 SPTR:P50553
>UNIPROT|P51145 SPTR:P51145
>UNIPROT|P51448 SPTR:P51448
>UNIPROT|P51449 SPTR:P51449
>UNIPROT|P51480 SPTR:P51480
>UNIPROT|P51608 SPTR:P51608
>UNIPROT|P51692 SPTR:P51692
>UNIPROT|P51774 SPTR:P51774
>UNIPROT|P51825 SPTR:P51825
>UNIPROT|P51827 SPTR:P51827
>UNIPROT|P51843 SPTR:P51843
>UNIPROT|P51974 SPTR:P51974
>UNIPROT|P52168 SPTR:P52168
>UNIPROT|P52172 SPTR:P52172
>UNIPROT|P52631 SPTR:P52631
>UNIPROT|P52632 SPTR:P52632
>UNIPROT|P52756 SPTR:P52756
>UNIPROT|P52909 SPTR:P52909
>UNIPROT|P52946 SPTR:P52946
>UNIPROT|P52947 SPTR:P52947
>UNIPROT|P52951 SPTR:P52951
>UNIPROT|P52952 SPTR:P52952
>UNIPROT|P52960 SPTR:P52960
>UNIPROT|P53147 SPTR:P53147

>UNIPROT|P53539 SPTR:P53539
>UNIPROT|P53566 SPTR:P53566
>UNIPROT|P53568 SPTR:P53568
>UNIPROT|P53740 SPTR:P53740
>UNIPROT|P53762 SPTR:P53762
>UNIPROT|P53805 SPTR:P53805
>UNIPROT|P54198 SPTR:P54198
>UNIPROT|P54274 SPTR:P54274
>UNIPROT|P54841 SPTR:P54841
>UNIPROT|P54843 SPTR:P54843
>UNIPROT|P54846 SPTR:P54846
>UNIPROT|P55055 SPTR:P55055
>UNIPROT|P55197 SPTR:P55197
>UNIPROT|P55315 SPTR:P55315
>UNIPROT|P55316 SPTR:P55316
>UNIPROT|P55317 SPTR:P55317
>UNIPROT|P55771 SPTR:P55771
>UNIPROT|P55895 SPTR:P55895
>UNIPROT|P56177 SPTR:P56177
>UNIPROT|P56179 SPTR:P56179
>UNIPROT|P56222 SPTR:P56222
>UNIPROT|P56223 SPTR:P56223
>UNIPROT|P56672 SPTR:P56672
>UNIPROT|P56721 SPTR:P56721
>UNIPROT|P56915 SPTR:P56915
>UNIPROT|P56931 SPTR:P56931
>UNIPROT|P57082 SPTR:P57082
>UNIPROT|P57086 SPTR:P57086
>UNIPROT|P57760 SPTR:P57760
>UNIPROT|P58012 SPTR:P58012
>UNIPROT|P58334 SPTR:P58334
>UNIPROT|P62508 SPTR:P62508
>UNIPROT|P62509 SPTR:P62509
>UNIPROT|P62510 SPTR:P62510
>UNIPROT|P63015 SPTR:P63015
>UNIPROT|P63058 SPTR:P63058
>UNIPROT|P67809 SPTR:P67809
>UNIPROT|P70262 SPTR:P70262
>UNIPROT|P70279 SPTR:P70279
>UNIPROT|P70300 SPTR:P70300
>UNIPROT|P78424 SPTR:P78424
>UNIPROT|P78426 SPTR:P78426
>UNIPROT|P78962 SPTR:P78962
>UNIPROT|P80205 SPTR:P80205
>UNIPROT|P80206 SPTR:P80206
>UNIPROT|P81133 SPTR:P81133
>UNIPROT|P81269 SPTR:P81269
>UNIPROT|P83870 SPTR:P83870
>UNIPROT|P83871 SPTR:P83871
>UNIPROT|P83949 SPTR:P83949
>UNIPROT|P84022 SPTR:P84022
>UNIPROT|P84025 SPTR:P84025
>UNIPROT|P90953 SPTR:P90953
>UNIPROT|P91664 SPTR:P91664
>UNIPROT|P97303 SPTR:P97303
>UNIPROT|P97360 SPTR:P97360
>UNIPROT|P97367 SPTR:P97367
>UNIPROT|P97471 SPTR:P97471
>UNIPROT|P97474 SPTR:P97474
>UNIPROT|P97510 SPTR:P97510
>UNIPROT|P97782 SPTR:P97782
>UNIPROT|P97832 SPTR:P97832
>UNIPROT|P97875 SPTR:P97875
>UNIPROT|P98149 SPTR:P98149
>UNIPROT|P98177 SPTR:P98177
>UNIPROT|Q00288 SPTR:Q00288
>UNIPROT|Q00322 SPTR:Q00322
>UNIPROT|Q00417 SPTR:Q00417
>UNIPROT|Q00422 SPTR:Q00422
>UNIPROT|Q00444 SPTR:Q00444
>UNIPROT|Q00613 SPTR:Q00613
>UNIPROT|Q00653 SPTR:Q00653
>UNIPROT|Q00899 SPTR:Q00899
>UNIPROT|Q00978 SPTR:Q00978
>UNIPROT|Q01068 SPTR:Q01068
>UNIPROT|Q01069 SPTR:Q01069
>UNIPROT|Q01070 SPTR:Q01070
>UNIPROT|Q01071 SPTR:Q01071
>UNIPROT|Q01147 SPTR:Q01147
>UNIPROT|Q01196 SPTR:Q01196
>UNIPROT|Q01201 SPTR:Q01201
>UNIPROT|Q01295 SPTR:Q01295
>UNIPROT|Q01532 SPTR:Q01532
>UNIPROT|Q01538 SPTR:Q01538
>UNIPROT|Q01543 SPTR:Q01543
>UNIPROT|Q01842 SPTR:Q01842
>UNIPROT|Q01851 SPTR:Q01851
>UNIPROT|Q01860 SPTR:Q01860
>UNIPROT|Q02078 SPTR:Q02078
>UNIPROT|Q02080 SPTR:Q02080
>UNIPROT|Q02248 SPTR:Q02248
>UNIPROT|Q02486 SPTR:Q02486
>UNIPROT|Q02650 SPTR:Q02650
>UNIPROT|Q02780 SPTR:Q02780
>UNIPROT|Q02878 SPTR:Q02878
>UNIPROT|Q02930 SPTR:Q02930
>UNIPROT|Q02953 SPTR:Q02953
>UNIPROT|Q02962 SPTR:Q02962
>UNIPROT|Q03014 SPTR:Q03014
>UNIPROT|Q03052 SPTR:Q03052
>UNIPROT|Q03125 SPTR:Q03125
>UNIPROT|Q03181 SPTR:Q03181
>UNIPROT|Q03188 SPTR:Q03188
>UNIPROT|Q03267 SPTR:Q03267
>UNIPROT|Q03347 SPTR:Q03347
>UNIPROT|Q03410 SPTR:Q03410
>UNIPROT|Q03484 SPTR:Q03484
>UNIPROT|Q03631 SPTR:Q03631
>UNIPROT|Q03701 SPTR:Q03701
>UNIPROT|Q03828 SPTR:Q03828

>UNIPROT|Q03833 SPTR:Q03833
>UNIPROT|Q04207 SPTR:Q04207
>UNIPROT|Q04383 SPTR:Q04383
>UNIPROT|Q04741 SPTR:Q04741
>UNIPROT|Q04743 SPTR:Q04743
>UNIPROT|Q04787 SPTR:Q04787
>UNIPROT|Q04864 SPTR:Q04864
>UNIPROT|Q04886 SPTR:Q04886
>UNIPROT|Q04887 SPTR:Q04887
>UNIPROT|Q04931 SPTR:Q04931
>UNIPROT|Q05066 SPTR:Q05066
>UNIPROT|Q05095 SPTR:Q05095
>UNIPROT|Q05195 SPTR:Q05195
>UNIPROT|Q05215 SPTR:Q05215
>UNIPROT|Q06003 SPTR:Q06003
>UNIPROT|Q06149 SPTR:Q06149
>UNIPROT|Q06219 SPTR:Q06219
>UNIPROT|Q06266 SPTR:Q06266
>UNIPROT|Q06330 SPTR:Q06330
>UNIPROT|Q06416 SPTR:Q06416
>UNIPROT|Q06455 SPTR:Q06455
>UNIPROT|Q06481 SPTR:Q06481
>UNIPROT|Q06630 SPTR:Q06630
>UNIPROT|Q06710 SPTR:Q06710
>UNIPROT|Q06889 SPTR:Q06889
>UNIPROT|Q06945 SPTR:Q06945
>UNIPROT|Q07243 SPTR:Q07243
>UNIPROT|Q07352 SPTR:Q07352
>UNIPROT|Q07644 SPTR:Q07644
>UNIPROT|Q07666 SPTR:Q07666
>UNIPROT|Q07687 SPTR:Q07687
>UNIPROT|Q07802 SPTR:Q07802
>UNIPROT|Q07869 SPTR:Q07869
>UNIPROT|Q08639 SPTR:Q08639
>UNIPROT|Q08775 SPTR:Q08775
>UNIPROT|Q08874 SPTR:Q08874
>UNIPROT|Q08904 SPTR:Q08904
>UNIPROT|Q08945 SPTR:Q08945
>UNIPROT|Q09636 SPTR:Q09636
>UNIPROT|Q10423 SPTR:Q10423
>UNIPROT|Q12224 SPTR:Q12224
>UNIPROT|Q12778 SPTR:Q12778
>UNIPROT|Q12800 SPTR:Q12800
>UNIPROT|Q12837 SPTR:Q12837
>UNIPROT|Q12857 SPTR:Q12857
>UNIPROT|Q12899 SPTR:Q12899
>UNIPROT|Q12946 SPTR:Q12946
>UNIPROT|Q12948 SPTR:Q12948
>UNIPROT|Q12950 SPTR:Q12950
>UNIPROT|Q12951 SPTR:Q12951
>UNIPROT|Q12952 SPTR:Q12952
>UNIPROT|Q12986 SPTR:Q12986
>UNIPROT|Q13028 SPTR:Q13028
>UNIPROT|Q13118 SPTR:Q13118
>UNIPROT|Q13133 SPTR:Q13133
>UNIPROT|Q13207 SPTR:Q13207
>UNIPROT|Q13342 SPTR:Q13342
>UNIPROT|Q13351 SPTR:Q13351
>UNIPROT|Q13422 SPTR:Q13422
>UNIPROT|Q13461 SPTR:Q13461
>UNIPROT|Q13465 SPTR:Q13465
>UNIPROT|Q13469 SPTR:Q13469
>UNIPROT|Q13485 SPTR:Q13485
>UNIPROT|Q13487 SPTR:Q13487
>UNIPROT|Q13761 SPTR:Q13761
>UNIPROT|Q13862 SPTR:Q13862
>UNIPROT|Q13886 SPTR:Q13886
>UNIPROT|Q13890 SPTR:Q13890
>UNIPROT|Q13891 SPTR:Q13891
>UNIPROT|Q13892 SPTR:Q13892
>UNIPROT|Q13901 SPTR:Q13901
>UNIPROT|Q13952 SPTR:Q13952
>UNIPROT|Q14134 SPTR:Q14134
>UNIPROT|Q14186 SPTR:Q14186
>UNIPROT|Q14188 SPTR:Q14188
>UNIPROT|Q14190 SPTR:Q14190
>UNIPROT|Q14211 SPTR:Q14211
>UNIPROT|Q14249 SPTR:Q14249
>UNIPROT|Q14258 SPTR:Q14258
>UNIPROT|Q14267 SPTR:Q14267
>UNIPROT|Q14333 SPTR:Q14333
>UNIPROT|Q14469 SPTR:Q14469
>UNIPROT|Q14488 SPTR:Q14488
>UNIPROT|Q14494 SPTR:Q14494
>UNIPROT|Q14526 SPTR:Q14526
>UNIPROT|Q14541 SPTR:Q14541
>UNIPROT|Q14548 SPTR:Q14548
>UNIPROT|Q14549 SPTR:Q14549
>UNIPROT|Q14561 SPTR:Q14561
>UNIPROT|Q14765 SPTR:Q14765
>UNIPROT|Q14774 SPTR:Q14774
>UNIPROT|Q14814 SPTR:Q14814
>UNIPROT|Q14863 SPTR:Q14863
>UNIPROT|Q14872 SPTR:Q14872
>UNIPROT|Q14897 SPTR:Q14897
>UNIPROT|Q14901 SPTR:Q14901
>UNIPROT|Q14994 SPTR:Q14994
>UNIPROT|Q15270 SPTR:Q15270
>UNIPROT|Q15319 SPTR:Q15319
>UNIPROT|Q15325 SPTR:Q15325
>UNIPROT|Q15329 SPTR:Q15329
>UNIPROT|Q15355 SPTR:Q15355
>UNIPROT|Q15431 SPTR:Q15431
>UNIPROT|Q15562 SPTR:Q15562
>UNIPROT|Q15574 SPTR:Q15574
>UNIPROT|Q15583 SPTR:Q15583
>UNIPROT|Q15631 SPTR:Q15631
>UNIPROT|Q15699 SPTR:Q15699
>UNIPROT|Q15784 SPTR:Q15784
>UNIPROT|Q16236 SPTR:Q16236

>UNIPROT|Q16254 SPTR:Q16254
>UNIPROT|Q16365 SPTR:Q16365
>UNIPROT|Q16464 SPTR:Q16464
>UNIPROT|Q16534 SPTR:Q16534
>UNIPROT|Q16559 SPTR:Q16559
>UNIPROT|Q16621 SPTR:Q16621
>UNIPROT|Q16624 SPTR:Q16624
>UNIPROT|Q16650 SPTR:Q16650
>UNIPROT|Q16676 SPTR:Q16676
>UNIPROT|Q22289 SPTR:Q22289
>UNIPROT|Q22292 SPTR:Q22292
>UNIPROT|Q23976 SPTR:Q23976
>UNIPROT|Q24151 SPTR:Q24151
>UNIPROT|Q24174 SPTR:Q24174
>UNIPROT|Q24206 SPTR:Q24206
>UNIPROT|Q24217 SPTR:Q24217
>UNIPROT|Q24248 SPTR:Q24248
>UNIPROT|Q24255 SPTR:Q24255
>UNIPROT|Q24256 SPTR:Q24256
>UNIPROT|Q24260 SPTR:Q24260
>UNIPROT|Q24432 SPTR:Q24432
>UNIPROT|Q24455 SPTR:Q24455
>UNIPROT|Q24525 SPTR:Q24525
>UNIPROT|Q24573 SPTR:Q24573
>UNIPROT|Q27350 SPTR:Q27350
>UNIPROT|Q27403 SPTR:Q27403
>UNIPROT|Q60591 SPTR:Q60591
>UNIPROT|Q60674 SPTR:Q60674
>UNIPROT|Q60722 SPTR:Q60722
>UNIPROT|Q60765 SPTR:Q60765
>UNIPROT|Q60795 SPTR:Q60795
>UNIPROT|Q60867 SPTR:Q60867
>UNIPROT|Q60929 SPTR:Q60929
>UNIPROT|Q61045 SPTR:Q61045
>UNIPROT|Q61060 SPTR:Q61060
>UNIPROT|Q61079 SPTR:Q61079
>UNIPROT|Q61169 SPTR:Q61169
>UNIPROT|Q61286 SPTR:Q61286
>UNIPROT|Q61312 SPTR:Q61312
>UNIPROT|Q61324 SPTR:Q61324
>UNIPROT|Q61329 SPTR:Q61329
>UNIPROT|Q61340 SPTR:Q61340
>UNIPROT|Q61345 SPTR:Q61345
>UNIPROT|Q61368 SPTR:Q61368
>UNIPROT|Q61442 SPTR:Q61442
>UNIPROT|Q61443 SPTR:Q61443
>UNIPROT|Q61473 SPTR:Q61473
>UNIPROT|Q61478 SPTR:Q61478
>UNIPROT|Q61501 SPTR:Q61501
>UNIPROT|Q61502 SPTR:Q61502
>UNIPROT|Q61723 SPTR:Q61723
>UNIPROT|Q61807 SPTR:Q61807
>UNIPROT|Q61827 SPTR:Q61827
>UNIPROT|Q61927 SPTR:Q61927
>UNIPROT|Q62156 SPTR:Q62156
>UNIPROT|Q62655 SPTR:Q62655
>UNIPROT|Q62735 SPTR:Q62735
>UNIPROT|Q62771 SPTR:Q62771
>UNIPROT|Q62782 SPTR:Q62782
>UNIPROT|Q62814 SPTR:Q62814
>UNIPROT|Q63245 SPTR:Q63245
>UNIPROT|Q63246 SPTR:Q63246
>UNIPROT|Q63248 SPTR:Q63248
>UNIPROT|Q63369 SPTR:Q63369
>UNIPROT|Q63934 SPTR:Q63934
>UNIPROT|Q63943 SPTR:Q63943
>UNIPROT|Q63955 SPTR:Q63955
>UNIPROT|Q64131 SPTR:Q64131
>UNIPROT|Q64163 SPTR:Q64163
>UNIPROT|Q64287 SPTR:Q64287
>UNIPROT|Q64364 SPTR:Q64364
>UNIPROT|Q64731 SPTR:Q64731
>UNIPROT|Q64FM2 SPTR:Q64FM2
>UNIPROT|Q68FF2 SPTR:Q68FF2
>UNIPROT|Q68G54 SPTR:Q68G54
>UNIPROT|Q68G78 SPTR:Q68G78
>UNIPROT|Q68HB9 SPTR:Q68HB9
>UNIPROT|Q6A028 SPTR:Q6A028
>UNIPROT|Q6A056 SPTR:Q6A056
>UNIPROT|Q6LCW3 SPTR:Q6LCW3
>UNIPROT|Q6NSS1 SPTR:Q6NSS1
>UNIPROT|Q6NWW0 SPTR:Q6NWW0
>UNIPROT|Q6P569 SPTR:Q6P569
>UNIPROT|Q6P7T9 SPTR:Q6P7T9
>UNIPROT|Q6P8Q3 SPTR:Q6P8Q3
>UNIPROT|Q6PCM3 SPTR:Q6PCM3
>UNIPROT|Q6PDH9 SPTR:Q6PDH9
>UNIPROT|Q6PF91 SPTR:Q6PF91
>UNIPROT|Q6R3Q6 SPTR:Q6R3Q6
>UNIPROT|Q6S7F2 SPTR:Q6S7F2
>UNIPROT|Q6T3A4 SPTR:Q6T3A4
>UNIPROT|Q6VUP9 SPTR:Q6VUP9
>UNIPROT|Q6ZQD6 SPTR:Q6ZQD6
>UNIPROT|Q7M3M8 SPTR:Q7M3M8
>UNIPROT|Q7RTV0 SPTR:Q7RTV0
>UNIPROT|Q7TMT9 SPTR:Q7TMT9
>UNIPROT|Q7TS99 SPTR:Q7TS99
>UNIPROT|Q7TT22 SPTR:Q7TT22
>UNIPROT|Q7YU77 SPTR:Q7YU77
>UNIPROT|Q80UF1 SPTR:Q80UF1
>UNIPROT|Q80VJ2 SPTR:Q80VJ2
>UNIPROT|Q80VX4 SPTR:Q80VX4
>UNIPROT|Q80W88 SPTR:Q80W88
>UNIPROT|Q80WG4 SPTR:Q80WG4
>UNIPROT|Q80WJ9 SPTR:Q80WJ9
>UNIPROT|Q80WK0 SPTR:Q80WK0
>UNIPROT|Q80WS6 SPTR:Q80WS6
>UNIPROT|Q80X69 SPTR:Q80X69
>UNIPROT|Q80Y36 SPTR:Q80Y36
>UNIPROT|Q80Z36 SPTR:Q80Z36

>UNIPROT|Q80ZG8 SPTR:Q80ZG8
>UNIPROT|Q810B3 SPTR:Q810B3
>UNIPROT|Q812G5 SPTR:Q812G5
>UNIPROT|Q86LS0 SPTR:Q86LS0
>UNIPROT|Q86Y01 SPTR:Q86Y01
>UNIPROT|Q8BFR8 SPTR:Q8BFR8
>UNIPROT|Q8BG65 SPTR:Q8BG65
>UNIPROT|Q8BGV1 SPTR:Q8BGV1
>UNIPROT|Q8BGZ9 SPTR:Q8BGZ9
>UNIPROT|Q8BHD2 SPTR:Q8BHD2
>UNIPROT|Q8BID0 SPTR:Q8BID0
>UNIPROT|Q8BMQ3 SPTR:Q8BMQ3
>UNIPROT|Q8BP10 SPTR:Q8BP10
>UNIPROT|Q8BPE6 SPTR:Q8BPE6
>UNIPROT|Q8BPL7 SPTR:Q8BPL7
>UNIPROT|Q8BPN4 SPTR:Q8BPN4
>UNIPROT|Q8BQ55 SPTR:Q8BQ55
>UNIPROT|Q8BRE2 SPTR:Q8BRE2
>UNIPROT|Q8BRS9 SPTR:Q8BRS9
>UNIPROT|Q8BS64 SPTR:Q8BS64
>UNIPROT|Q8BSM3 SPTR:Q8BSM3
>UNIPROT|Q8BSQ3 SPTR:Q8BSQ3
>UNIPROT|Q8BU17 SPTR:Q8BU17
>UNIPROT|Q8BUN5 SPTR:Q8BUN5
>UNIPROT|Q8BW92 SPTR:Q8BW92
>UNIPROT|Q8BWM0 SPTR:Q8BWM0
>UNIPROT|Q8BXG2 SPTR:Q8BXG2
>UNIPROT|Q8BYH0 SPTR:Q8BYH0
>UNIPROT|Q8BYU5 SPTR:Q8BYU5
>UNIPROT|Q8C0C0 SPTR:Q8C0C0
>UNIPROT|Q8C0F6 SPTR:Q8C0F6
>UNIPROT|Q8C0Q2 SPTR:Q8C0Q2
>UNIPROT|Q8C3F5 SPTR:Q8C3F5
>UNIPROT|Q8C4C0 SPTR:Q8C4C0
>UNIPROT|Q8C6B5 SPTR:Q8C6B5
>UNIPROT|Q8C733 SPTR:Q8C733
>UNIPROT|Q8C8M7 SPTR:Q8C8M7
>UNIPROT|Q8C955 SPTR:Q8C955
>UNIPROT|Q8C9R3 SPTR:Q8C9R3
>UNIPROT|Q8CDA0 SPTR:Q8CDA0
>UNIPROT|Q8CEM2 SPTR:Q8CEM2
>UNIPROT|Q8CF90 SPTR:Q8CF90
>UNIPROT|Q8CFN5 SPTR:Q8CFN5
>UNIPROT|Q8CFY1 SPTR:Q8CFY1
>UNIPROT|Q8CGF9 SPTR:Q8CGF9
>UNIPROT|Q8CH43 SPTR:Q8CH43
>UNIPROT|Q8I7Z8 SPTR:Q8I7Z8
>UNIPROT|Q8IM96 SPTR:Q8IM96
>UNIPROT|Q8IN81 SPTR:Q8IN81
>UNIPROT|Q8IN94 SPTR:Q8IN94
>UNIPROT|Q8INC6 SPTR:Q8INC6
>UNIPROT|Q8INL5 SPTR:Q8INL5
>UNIPROT|Q8INL6 SPTR:Q8INL6
>UNIPROT|Q8INT1 SPTR:Q8INT1
>UNIPROT|Q8IQ98 SPTR:Q8IQ98
>UNIPROT|Q8IQ99 SPTR:Q8IQ99
>UNIPROT|Q8IQT5 SPTR:Q8IQT5
>UNIPROT|Q8IRX4 SPTR:Q8IRX4
>UNIPROT|Q8K120 SPTR:Q8K120
>UNIPROT|Q8K1C8 SPTR:Q8K1C8
>UNIPROT|Q8K1L0 SPTR:Q8K1L0
>UNIPROT|Q8K1S5 SPTR:Q8K1S5
>UNIPROT|Q8K378 SPTR:Q8K378
>UNIPROT|Q8K4J2 SPTR:Q8K4J2
>UNIPROT|Q8K4U6 SPTR:Q8K4U6
>UNIPROT|Q8K557 SPTR:Q8K557
>UNIPROT|Q8K565 SPTR:Q8K565
>UNIPROT|Q8MKX3 SPTR:Q8MKX3
>UNIPROT|Q8ML56 SPTR:Q8ML56
>UNIPROT|Q8MR99 SPTR:Q8MR99
>UNIPROT|Q8NFW5 SPTR:Q8NFW5
>UNIPROT|Q8NFW6 SPTR:Q8NFW6
>UNIPROT|Q8NHD9 SPTR:Q8NHD9
>UNIPROT|Q8NHV9 SPTR:Q8NHV9
>UNIPROT|Q8NHW3 SPTR:Q8NHW3
>UNIPROT|Q8R0B5 SPTR:Q8R0B5
>UNIPROT|Q8R0H1 SPTR:Q8R0H1
>UNIPROT|Q8R0K9 SPTR:Q8R0K9
>UNIPROT|Q8R1T9 SPTR:Q8R1T9
>UNIPROT|Q8R515 SPTR:Q8R515
>UNIPROT|Q8SWV1 SPTR:Q8SWV1
>UNIPROT|Q8SZT1 SPTR:Q8SZT1
>UNIPROT|Q8VBZ1 SPTR:Q8VBZ1
>UNIPROT|Q8VD35 SPTR:Q8VD35
>UNIPROT|Q8VD55 SPTR:Q8VD55
>UNIPROT|Q8VDL9 SPTR:Q8VDL9
>UNIPROT|Q8VH35 SPTR:Q8VH35
>UNIPROT|Q8VHI4 SPTR:Q8VHI4
>UNIPROT|Q8VHL9 SPTR:Q8VHL9
>UNIPROT|Q8VHT7 SPTR:Q8VHT7
>UNIPROT|Q8VI67 SPTR:Q8VI67
>UNIPROT|Q90268 SPTR:Q90268
>UNIPROT|Q90273 SPTR:Q90273
>UNIPROT|Q90415 SPTR:Q90415
>UNIPROT|Q90416 SPTR:Q90416
>UNIPROT|Q91X20 SPTR:Q91X20
>UNIPROT|Q91Y54 SPTR:Q91Y54
>UNIPROT|Q91Y65 SPTR:Q91Y65
>UNIPROT|Q921B7 SPTR:Q921B7
>UNIPROT|Q921C3 SPTR:Q921C3
>UNIPROT|Q921S6 SPTR:Q921S6
>UNIPROT|Q922A5 SPTR:Q922A5
>UNIPROT|Q92481 SPTR:Q92481
>UNIPROT|Q924L0 SPTR:Q924L0
>UNIPROT|Q925C2 SPTR:Q925C2
>UNIPROT|Q92731 SPTR:Q92731
>UNIPROT|Q92754 SPTR:Q92754
>UNIPROT|Q92782 SPTR:Q92782
>UNIPROT|Q92858 SPTR:Q92858
>UNIPROT|Q92886 SPTR:Q92886

>UNIPROT|Q92908 SPTR:Q92908
>UNIPROT|Q92949 SPTR:Q92949
>UNIPROT|Q92988 SPTR:Q92988
>UNIPROT|Q94513 SPTR:Q94513
>UNIPROT|Q94547 SPTR:Q94547
>UNIPROT|Q95RW8 SPTR:Q95RW8
>UNIPROT|Q961D9 SPTR:Q961D9
>UNIPROT|Q96JB3 SPTR:Q96JB3
>UNIPROT|Q99419 SPTR:Q99419
>UNIPROT|Q99453 SPTR:Q99453
>UNIPROT|Q99576 SPTR:Q99576
>UNIPROT|Q99581 SPTR:Q99581
>UNIPROT|Q99583 SPTR:Q99583
>UNIPROT|Q99626 SPTR:Q99626
>UNIPROT|Q99687 SPTR:Q99687
>UNIPROT|Q99718 SPTR:Q99718
>UNIPROT|Q99742 SPTR:Q99742
>UNIPROT|Q99743 SPTR:Q99743
>UNIPROT|Q99801 SPTR:Q99801
>UNIPROT|Q99811 SPTR:Q99811
>UNIPROT|Q99816 SPTR:Q99816
>UNIPROT|Q99856 SPTR:Q99856
>UNIPROT|Q99929 SPTR:Q99929
>UNIPROT|Q99966 SPTR:Q99966
>UNIPROT|Q99967 SPTR:Q99967
>UNIPROT|Q99J30 SPTR:Q99J30
>UNIPROT|Q99JF1 SPTR:Q99JF1
>UNIPROT|Q99K73 SPTR:Q99K73
>UNIPROT|Q99KN2 SPTR:Q99KN2
>UNIPROT|Q99L57 SPTR:Q99L57
>UNIPROT|Q99LQ5 SPTR:Q99LQ5
>UNIPROT|Q99MA9 SPTR:Q99MA9
>UNIPROT|Q99MB7 SPTR:Q99MB7
>UNIPROT|Q99ME7 SPTR:Q99ME7
>UNIPROT|Q99MY0 SPTR:Q99MY0
>UNIPROT|Q99NA7 SPTR:Q99NA7
>UNIPROT|Q99NE9 SPTR:Q99NE9
>UNIPROT|Q99PH9 SPTR:Q99PH9
>UNIPROT|Q99PI0 SPTR:Q99PI0
>UNIPROT|Q99PI2 SPTR:Q99PI2
>UNIPROT|Q9BII5 SPTR:Q9BII5
>UNIPROT|Q9BWW7 SPTR:Q9BWW7
>UNIPROT|Q9BX26 SPTR:Q9BX26
>UNIPROT|Q9BXX3 SPTR:Q9BXX3
>UNIPROT|Q9BYE0 SPTR:Q9BYE0
>UNIPROT|Q9BYU1 SPTR:Q9BYU1
>UNIPROT|Q9BYU3 SPTR:Q9BYU3
>UNIPROT|Q9BZS1 SPTR:Q9BZS1
>UNIPROT|Q9C056 SPTR:Q9C056
>UNIPROT|Q9C0J9 SPTR:Q9C0J9
>UNIPROT|Q9CWU7 SPTR:Q9CWU7
>UNIPROT|Q9CXE4 SPTR:Q9CXE4
>UNIPROT|Q9CYB4 SPTR:Q9CYB4
>UNIPROT|Q9CZK7 SPTR:Q9CZK7
>UNIPROT|Q9D030 SPTR:Q9D030
>UNIPROT|Q9D040 SPTR:Q9D040
>UNIPROT|Q9D297 SPTR:Q9D297
>UNIPROT|Q9D2A5 SPTR:Q9D2A5
>UNIPROT|Q9D4U3 SPTR:Q9D4U3
>UNIPROT|Q9D7L0 SPTR:Q9D7L0
>UNIPROT|Q9DBG9 SPTR:Q9DBG9
>UNIPROT|Q9DBQ6 SPTR:Q9DBQ6
>UNIPROT|Q9DBU5 SPTR:Q9DBU5
>UNIPROT|Q9DBV7 SPTR:Q9DBV7
>UNIPROT|Q9DCZ7 SPTR:Q9DCZ7
>UNIPROT|Q9EP91 SPTR:Q9EP91
>UNIPROT|Q9EPQ8 SPTR:Q9EPQ8
>UNIPROT|Q9EPZ5 SPTR:Q9EPZ5
>UNIPROT|Q9EQD1 SPTR:Q9EQD1
>UNIPROT|Q9EQF7 SPTR:Q9EQF7
>UNIPROT|Q9EQW6 SPTR:Q9EQW6
>UNIPROT|Q9EQX5 SPTR:Q9EQX5
>UNIPROT|Q9ERF6 SPTR:Q9ERF6
>UNIPROT|Q9ES03 SPTR:Q9ES03
>UNIPROT|Q9GZN2 SPTR:Q9GZN2
>UNIPROT|Q9GZR2 SPTR:Q9GZR2
>UNIPROT|Q9H161 SPTR:Q9H161
>UNIPROT|Q9H2C1 SPTR:Q9H2C1
>UNIPROT|Q9H2G9 SPTR:Q9H2G9
>UNIPROT|Q9H2M4 SPTR:Q9H2M4
>UNIPROT|Q9H4I2 SPTR:Q9H4I2
>UNIPROT|Q9H509 SPTR:Q9H509
>UNIPROT|Q9HBD2 SPTR:Q9HBD2
>UNIPROT|Q9HBE0 SPTR:Q9HBE0
>UNIPROT|Q9HBU1 SPTR:Q9HBU1
>UNIPROT|Q9HBZ2 SPTR:Q9HBZ2
>UNIPROT|Q9I9P9 SPTR:Q9I9P9
>UNIPROT|Q9JHE6 SPTR:Q9JHE6
>UNIPROT|Q9JHR9 SPTR:Q9JHR9
>UNIPROT|Q9JJE5 SPTR:Q9JJE5
>UNIPROT|Q9JJW4 SPTR:Q9JJW4
>UNIPROT|Q9JL60 SPTR:Q9JL60
>UNIPROT|Q9JLZ6 SPTR:Q9JLZ6
>UNIPROT|Q9JM73 SPTR:Q9JM73
>UNIPROT|Q9NP08 SPTR:Q9NP08
>UNIPROT|Q9NPC8 SPTR:Q9NPC8
>UNIPROT|Q9NQ87 SPTR:Q9NQ87
>UNIPROT|Q9NQL9 SPTR:Q9NQL9
>UNIPROT|Q9NR55 SPTR:Q9NR55
>UNIPROT|Q9NR83 SPTR:Q9NR83
>UNIPROT|Q9NRJ4 SPTR:Q9NRJ4
>UNIPROT|Q9NS37 SPTR:Q9NS37
>UNIPROT|Q9NZC4 SPTR:Q9NZC4
>UNIPROT|Q9NZG6 SPTR:Q9NZG6
>UNIPROT|Q9NZI6 SPTR:Q9NZI6
>UNIPROT|Q9NZI7 SPTR:Q9NZI7
>UNIPROT|Q9NZR4 SPTR:Q9NZR4
>UNIPROT|Q9PW77 SPTR:Q9PW77
>UNIPROT|Q9QX98 SPTR:Q9QX98
>UNIPROT|Q9QXJ5 SPTR:Q9QXJ5

>UNIPROT|Q9QXT8 SPTR:Q9QXT8
>UNIPROT|Q9QY31 SPTR:Q9QY31
>UNIPROT|Q9QYB2 SPTR:Q9QYB2
>UNIPROT|Q9QZ28 SPTR:Q9QZ28
>UNIPROT|Q9R1E0 SPTR:Q9R1E0
>UNIPROT|Q9TVQ4 SPTR:Q9TVQ4
>UNIPROT|Q9TW49 SPTR:Q9TW49
>UNIPROT|Q9U3V5 SPTR:Q9U3V5
>UNIPROT|Q9UAL9 SPTR:Q9UAL9
>UNIPROT|Q9UBL3 SPTR:Q9UBL3
>UNIPROT|Q9UBP5 SPTR:Q9UBP5
>UNIPROT|Q9UBX0 SPTR:Q9UBX0
>UNIPROT|Q9UD29 SPTR:Q9UD29
>UNIPROT|Q9UD83 SPTR:Q9UD83
>UNIPROT|Q9UFF9 SPTR:Q9UFF9
>UNIPROT|Q9UGM4 SPTR:Q9UGM4
>UNIPROT|Q9UGU0 SPTR:Q9UGU0
>UNIPROT|Q9UGU5 SPTR:Q9UGU5
>UNIPROT|Q9UH92 SPTR:Q9UH92
>UNIPROT|Q9UHB7 SPTR:Q9UHB7
>UNIPROT|Q9UIV1 SPTR:Q9UIV1
>UNIPROT|Q9UIW0 SPTR:Q9UIW0
>UNIPROT|Q9UJU5 SPTR:Q9UJU5
>UNIPROT|Q9UK39 SPTR:Q9UK39
>UNIPROT|Q9UKM6 SPTR:Q9UKM6
>UNIPROT|Q9UKY1 SPTR:Q9UKY1
>UNIPROT|Q9UL17 SPTR:Q9UL17
>UNIPROT|Q9UL49 SPTR:Q9UL49
>UNIPROT|Q9UMS5 SPTR:Q9UMS5
>UNIPROT|Q9UN30 SPTR:Q9UN30
>UNIPROT|Q9UN79 SPTR:Q9UN79
>UNIPROT|Q9UNX5 SPTR:Q9UNX5
>UNIPROT|Q9UPM6 SPTR:Q9UPM6
>UNIPROT|Q9UQR0 SPTR:Q9UQR0
>UNIPROT|Q9V3F3 SPTR:Q9V3F3
>UNIPROT|Q9V5L7 SPTR:Q9V5L7
>UNIPROT|Q9V5Y8 SPTR:Q9V5Y8
>UNIPROT|Q9V6M0 SPTR:Q9V6M0
>UNIPROT|Q9V8S2 SPTR:Q9V8S2
>UNIPROT|Q9V9W8 SPTR:Q9V9W8
>UNIPROT|Q9VAB3 SPTR:Q9VAB3
>UNIPROT|Q9VEZ8 SPTR:Q9VEZ8
>UNIPROT|Q9VF01 SPTR:Q9VF01
>UNIPROT|Q9VGY9 SPTR:Q9VGY9
>UNIPROT|Q9VI61 SPTR:Q9VI61
>UNIPROT|Q9VI63 SPTR:Q9VI63
>UNIPROT|Q9VJJ6 SPTR:Q9VJJ6
>UNIPROT|Q9VKM4 SPTR:Q9VKM4
>UNIPROT|Q9VLA2 SPTR:Q9VLA2
>UNIPROT|Q9VLU0 SPTR:Q9VLU0
>UNIPROT|Q9VML1 SPTR:Q9VML1
>UNIPROT|Q9VMS4 SPTR:Q9VMS4
>UNIPROT|Q9VTX7 SPTR:Q9VTX7
>UNIPROT|Q9VU01 SPTR:Q9VU01
>UNIPROT|Q9VU66 SPTR:Q9VU66
>UNIPROT|Q9VZ09 SPTR:Q9VZ09
>UNIPROT|Q9W0K4 SPTR:Q9W0K4
>UNIPROT|Q9W0K7 SPTR:Q9W0K7
>UNIPROT|Q9W0V9 SPTR:Q9W0V9
>UNIPROT|Q9W4S7 SPTR:Q9W4S7
>UNIPROT|Q9W572 SPTR:Q9W572
>UNIPROT|Q9W7E7 SPTR:Q9W7E7
>UNIPROT|Q9WTJ2 SPTR:Q9WTJ2
>UNIPROT|Q9WTK8 SPTR:Q9WTK8
>UNIPROT|Q9WTL8 SPTR:Q9WTL8
>UNIPROT|Q9WTN3 SPTR:Q9WTN3
>UNIPROT|Q9WU00 SPTR:Q9WU00
>UNIPROT|Q9WU56 SPTR:Q9WU56
>UNIPROT|Q9WUZ3 SPTR:Q9WUZ3
>UNIPROT|Q9WVH4 SPTR:Q9WVH4
>UNIPROT|Q9Y222 SPTR:Q9Y222
>UNIPROT|Q9Y250 SPTR:Q9Y250
>UNIPROT|Q9Y261 SPTR:Q9Y261
>UNIPROT|Q9Y2W7 SPTR:Q9Y2W7
>UNIPROT|Q9Y3Q5 SPTR:Q9Y3Q5
>UNIPROT|Q9Y451 SPTR:Q9Y451
>UNIPROT|Q9Y458 SPTR:Q9Y458
>UNIPROT|Q9Y466 SPTR:Q9Y466
>UNIPROT|Q9Y467 SPTR:Q9Y467
>UNIPROT|Q9Y4A8 SPTR:Q9Y4A8
>UNIPROT|Q9Y5B6 SPTR:Q9Y5B6
>UNIPROT|Q9Y5J3 SPTR:Q9Y5J3
>UNIPROT|Q9Y5K1 SPTR:Q9Y5K1
>UNIPROT|Q9Y5L5 SPTR:Q9Y5L5
>UNIPROT|Q9Y5R5 SPTR:Q9Y5R5
>UNIPROT|Q9Y5R6 SPTR:Q9Y5R6
>UNIPROT|Q9Y5W3 SPTR:Q9Y5W3
>UNIPROT|Q9Y6P7 SPTR:Q9Y6P7
>UNIPROT|Q9Y6Q2 SPTR:Q9Y6Q2
>UNIPROT|Q9Y6X8 SPTR:Q9Y6X8
>UNIPROT|Q9Z0Y9 SPTR:Q9Z0Y9
>UNIPROT|Q9Z1L3 SPTR:Q9Z1L3
>UNIPROT|Q9Z248 SPTR:Q9Z248
>UNIPROT|Q9Z288 SPTR:Q9Z288
>UNIPROT|Q9Z2I5 SPTR:Q9Z2I5
>UNIPROT|Q9Z2K0 SPTR:Q9Z2K0
>gi|2934698|dbj|BAA25018.1|Enz-2
>gi|1168365|sp|P42568|AF9_HUMAN
>gi|34855620|ref|XP_231705.2|enx1
>gi|2130628|gb|AAB58343.1|NFkB1
>gi|15963442|dbj|BAB69473.1|bHLHOU
T
>gi|284412|pir||A41976|hmg
>gi|11967767|emb|CAC19387.1|NK7
>gi|34881391|ref|XP_229086.2|mybl
>gi|15679964|gb|AAH14289.1|cut1
>gi|913312|gb|AAB33046.1|madsMEF2
>gi|10720313|sp|Q24742|TRX
>gi|21614542|ref|NP_005059.2|sim1
>gi|26330670|dbj|BAC29065.1|creb2

>gi|1842175|dbj|BAA11335.1|irfdwb
>gi|26080349|ref|NP_524862.2|bwk-hmg;
>gi|14190146|gb|AAK55553.1|Meis1.1
>gi|18859033|ref|NP_570998.1|bHLHmitf
>gi|28626482|gb|AAO49160.1|deaf1