

Marine organism cell biology and regulatory sequence discovery in comparative functional genomics

David W. Barnes^{1,*}, Carolyn J. Mattingly¹, Angela Parton¹, Lori M. Dowell¹, Christopher J. Bayne^{1,2} and John N. Forrest Jr.^{1,3}

¹Center for Marine Functional Genomics Studies, Mount Desert Island Biological Laboratory, P.O. Box 35, Old Bar Harbour Road, Salisbury Cove, MA 04672, USA; ²Department of Zoology, Marine & Freshwater Biomedical Sciences Center, Oregon State University, Corvallis, OR 97331, USA; ³Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06510, USA; *Author for correspondence (e-mail: dbarnes@mdibl.org)

Received 4 August 2005; accepted in revised form 4 August 2005

Abstract

The use of bioinformatics to integrate phenotypic and genomic data from mammalian models is well established as a means of understanding human biology and disease. Beyond direct biomedical applications of these approaches in predicting structure–function relationships between coding sequences and protein activities, comparative studies also promote understanding of molecular evolution and the relationship between genomic sequence and morphological and physiological specialization. Recently recognized is the potential of comparative studies to identify functionally significant regulatory regions and to generate experimentally testable hypotheses that contribute to understanding mechanisms that regulate gene expression, including transcriptional activity, alternative splicing and transcript stability. Functional tests of hypotheses generated by computational approaches require experimentally tractable *in vitro* systems, including cell cultures. Comparative sequence analysis strategies that use genomic sequences from a variety of evolutionarily diverse organisms are critical for identifying conserved regulatory motifs in the 5′-upstream, 3′-downstream and introns of genes. Genomic sequences and gene orthologues in the first aquatic vertebrate and protovertebrate organisms to be fully sequenced (*Fugu rubripes*, *Ciona intestinalis*, *Tetraodon nigroviridis*, *Danio rerio*) as well as in the elasmobranchs, spiny dogfish shark (*Squalus acanthias*) and little skate (*Raja erinacea*), and marine invertebrate models such as the sea urchin (*Strongylocentrotus purpuratus*) are valuable in the prediction of putative genomic regulatory regions. Cell cultures have been derived for these and other model species. Data and tools resulting from these kinds of studies will contribute to understanding transcriptional regulation of biomedically important genes and provide new avenues for medical therapeutics and disease prevention.

Introduction: challenges to identifying regulatory elements in genomic sequences

The human genome sequencing project and related initiatives for other species have highlighted the value of comparative studies to reveal conserved

sequences of functional significance (Thomas and Touchman 2002; Pennacchio and Rubin 2003). Sequencing and comparative analysis of the mouse and human genomes identified orthologous genes and intraspecies polymorphisms, revealed the extent of sequence similarity and synteny between

the species and provided additional insights about genomic evolution. Another area of interest that stands to benefit greatly from genomic data is that of gene regulation. Understanding in this field is fragmented and, as with any scientific discipline, the concepts and questions that can be conceived and addressed experimentally are dependent on available technological approaches.

Mechanisms by which expression of a single gene is regulated can be extremely complicated. Multiple phosphorylation- or ligand-dependent nuclear receptors that homo- or heterodimerize may be required to achieve activity. Each of these receptors may have different activation specificity or duration, even when acting via the same regulatory DNA sequence such as classical proximal promoter elements. These receptors may also work in combination with other transcription factors that function at sites more distal from the proximal promoter or in introns. Alternatively spliced transcripts represent another complex aspect of gene expression regulation that is influenced by extracellular and intracellular signaling but is not well understood (Stamm et al. 2005). Furthermore, individual genes often are part of a broader, coordinately regulated network of genes that function to elicit a set of cellular responses (Wagner 1999). Through such mechanisms, ligands may, for instance, induce their own metabolism or export, a process that further complicates understanding of gene regulation and that also has critical implications for models of pharmacokinetics and drug efficacy.

Experimental identification of functional genomic sequences depends heavily on cell culture and other techniques of *in vitro* cell biology. Traditionally, identification of gene regulatory regions has been limited by the labor-intensiveness of the requisite strategies. Generally only regions close to the transcriptional start sites have been experimentally tractable for detailed examination, despite evidence that important regulatory regions exist more than 10 kb upstream or downstream from the coding region or in introns of genes (Rowntree et al. 2001).

Applied aspects of vertebrate comparative genomics and predictive cell biology

Identification of specific functional sequences through the generation of deletion constructs is

limited in the sequence size that can be analyzed and is restricted to examination of single genes. Transfection of cells with reporter constructs containing putative proximal promoters may elicit strong activation when treated with receptor-specific ligands in culture, while *in vivo* studies may yield inconsistent results, suggesting that these receptors and transcription factors are a subset of a larger complex of regulators of gene expression. Techniques such as DNASE I hypersensitivity studies and gel shifts are excellent methods for testing functionality of putative regulatory elements; however, they are not efficient for screening candidate sequences. By contrast, computational analyses allow the examination of a significantly larger region when predicting conserved regulatory regions and signals. Computational techniques lend themselves to the identification of patterns or clusters of regulatory motifs and prediction of co-regulated genes, and generate targeted predictions of candidate regulatory sequences and signaling molecules that can then be directly tested in functional and mutagenesis studies (Hughes et al. 2000; Loots et al. 2000; Pennacchio and Rubin 2003; Ovcharenko et al. 2005).

Understanding causative relationships between specific regulatory elements and expression patterns would greatly enhance the ability to predict disease-associated genes (Pennacchio and Rubin 2003). However, even in computational approaches the ability to identify and predict the regulatory functions of non-coding sequences has been limited (Pennacchio and Rubin 2003). Pairwise comparisons have helped to predict functionally conserved regions, however the statistical accuracy of these predictions is increased when more than two sequences are used (Dubchak et al. 2000). Several studies suggest that comparative analysis of multiple evolutionarily diverse organisms facilitate the prediction of functionally important non-coding regions (Dubchak et al. 2000; Matys et al. 2003; Thomas et al. 2003). Comparisons of genome sequences from evolutionarily diverse organisms also elucidate regulatory regions that are specific to a particular species or group of species. Recent additions of the African tree frog (*Xenopus tropicalis*), chicken (*Gallus gallus*) and dog (*Canis familiaris*) to the sequencing pipeline will provide important complementary perspectives for tetrapods, but data from more divergent vertebrate genomes is still

needed. Different rates of molecular evolution, including gene duplication, underscore the importance of the examination of genomic sequences from an evolutionary range of organisms for comparative sequence analyses. When such an approach is taken, the contrast of divergent sequences differentiates between functionally conserved regions and generally conserved regions that simply reflect a lack of divergence time (Dubchak et al. 2000). Using sequences from evolutionarily diverse organisms may provide the necessary divergence to identify functionally conserved regions in genes that have evolved slowly.

The contributions of marine and freshwater organisms to comparative genomics

The increasing availability of genomic data for non-mammalian organisms and similarities to the human genome underscore the value of these organisms as models of a variety of diseases in humans (Aparicio et al. 2002; Dehal et al. 2002; Ballatori et al. 2003). The identification of conserved coding and regulatory regions is enhanced by including divergent sequences in comparative studies (Thomas et al. 2003) because these sequences provide more stringent filters for detecting conserved, and presumably functionally important elements (Dubchak et al. 2000; Thomas et al. 2003; Ahituv et al. 2004). Consequently, sequences from evolutionarily distant marine vertebrates and protovertebrates are being used in comparative studies with increasing frequency. The pufferfish (*Fugu rubripes*) genome sequencing project supported this approach as it led to the discovery of nearly 1000 human genes not previously described (Aparicio et al. 2002). Sequence comparisons of the *Hoxb-4* gene in mouse and *Fugu* identified novel regulatory elements that directed subsets of the full *Hoxb-4* expression pattern in transgenic mice (Aparicio et al. 1995; Amores et al. 2004). In a recent comparative analysis of the *HoxA* cluster in human, horned shark and zebrafish (Chiu et al. 2002), extensive conservation of non-coding sequence motifs was found between the human and shark sequences, whereas zebrafish sequences exhibited significant loss of conservation. The majority of newly identified regulatory elements for this cluster of genes were identical to known

binding sites for regulatory proteins as defined in the transcription factor database, TRANSFAC (<http://www.biobase.de>; Matys et al. 2003), demonstrating the accuracy of this approach (Matys et al. 2003; Santini et al. 2003).

Special physiological attributes exhibited by many evolutionarily diverse aquatic organisms have led to the increased appreciation of their importance as models of human disease. These characteristics have been exploited experimentally to further understanding of human immunology, genomics, stem cell and cancer biology, pharmacology, toxicology and neurobiology. Historically, marine vertebrates provided critical insights into fundamental mechanisms of physiological processes, and the value of these organisms has not been diminished by the advent of molecular approaches. Membrane transporters that are the sites of action of diuretic drugs were first cloned from specialized organs in marine species (Gamba et al. 1993; Xu et al. 1994). Mutagenesis studies in bony fish generated a spectrum of biologically relevant and distinct phenotypes (Naruse et al. 2004; Walter et al. 2004). Large-scale genetic screens produced more than 500 zebrafish mutants, many with phenotypes similar to human disorders (Dooley and Zon 2000). Immunogenetic studies in carp, salmonids and other species are contributing to the growing database of marine vertebrate genomics (Bayne et al. 2001; Fujiki et al. 2001, 2003; Tomana et al. 2002; Ishikawa et al. 2004).

Although increasing numbers of vertebrate genomes are being sequenced, there are still stretches of evolutionary history without representation (Thomas and Touchman 2002). Among aquatic organisms, teleosts are represented by significant EST and genomic sequencing, such as those in zebrafish and pufferfish. Until recently, chondrichthyes, which include elasmobranchs (sharks, rays and skates) were the only major line of gnathostomes, or jawed vertebrates, for which there are no major initiatives to generate genomic sequences. Sequence data from elasmobranchs have provided unique insights into conserved functional domains of genes associated with human liver function, multidrug resistance, cystic fibrosis, G protein coupled receptors, natriuretic peptide receptors, and other biomedically relevant genes (Valentich and Forrest 1991; Henson et al. 1997; Aller et al. 1999; Silva et al. 1999; Greger et al. 1999; Waldegger et al. 1999; Ke et al. 2002;

Wang et al. 2002; Yang et al. 2002; Cai et al. 2001, 2003; Mattingly et al. 2004b).

In March, 2005 the National Human Genome Research Institute of the National Institutes of Health (NIH), USA, announced that it will fund the whole genome sequencing of *Raja erinacea*. In the announcement of the Skate Genome Sequencing Initiative (<http://www.genome.gov/13014443>), NIH states: 'The skate (related to many species of shark and cartilaginous fish) was chosen because it belongs to the first group of primitive vertebrates that developed jaws, an important step in vertebrate evolution. Other innovations in this group of animals include an adaptive immune system similar to that of humans, a closed and pressurized circulatory system, and myelination of the nervous system. Understanding these systems of the skate at a genetic level will help scientists identify the minimum set of genes that create a nervous system or develop a jaw, possibly illustrating how these systems have evolved in humans, and how they sometimes go wrong.'

Chondrichthyes (cartilaginous) fish appeared approximately 450 million years ago. Elasmobranchs comprise most chondrichthyan organisms. They exhibit fundamental vertebrate characteristics, including a recombinatorial immune system (Hinds and Litman 1986; Adelman et al. 2004), and are also the oldest existing vertebrates with circulatory system-related signaling molecules and receptors such as platelet-derived growth factor and adenosine receptors. The specialized rectal gland of *Squalus acanthias*, the spiny dogfish shark, has greatly facilitated the study of cystic fibrosis, sodium and chloride secretion (Devor et al. 1995; Lehrich et al. 1995; Forrest 1996; Henson et al. 1997; Lehrich et al. 1998; Silva et al. 1999; Aller et al. 1999; Greger et al. 1999; Waldegger et al. 1999; Ke et al. 2002; Yang et al. 2002). Unlike many primary cell cultures that dedifferentiate and lose transport polarity immediately after isolation, primary cultures of shark rectal gland tubular cells maintain fully differentiated function and expression of all known receptors, transporters, ion channels and signal transduction pathways *in vitro* (Valentich and Forrest 1991; Devor et al. 1995; Lehrich et al. 1995; Aller et al. 1999; Greger et al. 1999; Waldegger et al. 1999; Ke et al. 2002; Yang et al. 2002; Mattingly et al. 2004b). A comparison of the properties of the cloned shark and human cystic

fibrosis transmembrane regulator (CFTR) has provided insights into structural domains related to functional differences in the normal and mutant proteins (Marshall et al. 1991). The spiny dogfish shark CFTR protein is 72% identical to the human ortholog and comparison of the human and shark CFTR sequences revealed conservation of five cyclic AMP-dependent kinase phosphorylation sites and three residues that, when mutated in the human protein, are associated with cystic fibrosis.

The coding sequences and functions of a number of medically relevant genes are conserved in the spiny dogfish shark and little skate (Cai et al. 2001, 2003; Wang et al. 2002; Yang et al. 2002; Mattingly et al. 2004b). Primary hepatocytes from little skate retain hepatobiliary polarity for at least 8 h and possibly up to several days in culture, offering particular advantages for studies of liver-specific functions (Ballatori et al. 2003). Genomic information applied to existing physiological data in these systems, along with the further development of *in vitro* cell culture systems, will allow the testing of molecular hypotheses and understanding of regulatory mechanisms that are directly applicable to human biology. Targeted sequencing of well-defined genomic regions generated from BAC clones has been extremely useful in providing supporting genomic information in species for which complete genome data are not available (Thomas et al. 2002). In addition to constructing four-fold coverage bacterial artificial chromosome libraries from sperm DNA of dogfish shark and little skate (<http://www.mdibl.org/research/skategenome.shtml>), an expressed sequence tag (EST) sequencing project is underway to substantially increase the availability of sequence data for these model organisms (Mattingly et al. 2004b). Over 10,000 sequences are publicly available through the EST database at the National Center for Biotechnology Information (dbEST; Boguski et al. 1993) and <http://www.mdibl.org/decypher>. These data sets are updated as new sequences become available.

Deriving regulatory information from cross-species comparative analysis of genomic sequences

One of the remarkable findings of the human genome sequencing project was the discovery that

coding regions account for only 5% of the genome (Venter et al. 2001). The remaining sequence consists of repetitive DNA (~40–45%) and extensive non-coding regions for which there is very little functional information. It is within these regions that significant regulatory information presumably is concealed. The major challenge in the post-genomic age is uncovering important functional regions within this non-coding DNA. The rapid development of sequence analysis software tools, increasing availability of genomic data, and recognition of the importance of comparing data from diverse organisms are allowing scientists to make fruitful inroads to understanding genomic structure and gene expression regulation. A brief summary of software tools that are valuable for identifying regulatory information from cross-species comparative analyses follows.

The University of California Santa Cruz Genomic Browser (<http://genome.ucsc.edu/>; Karolchik et al. 2003) is among the most popular databases for querying and retrieving genomic sequences of interest. It currently provides access to genomic assemblies from 23 organisms, including 10 vertebrates, 8 insects, 2 nematodes, a sea squirt, baker's yeast, and the SARS virus. Whether genomic regions are retrieved from an existing database or sequenced locally, there are an increasing number of options for analysis and annotation. Several computational tools allow identification of genes and exon boundaries in genomic sequences. GENSCAN (Burge and Karlin 1997), TWINSKAN (Korf et al. 2001), MZEF (Zhang 1997), and Gene Recognition and Analysis Internet Link (GRAIL; Uberbacher and Mural 1991) are among the most popular. Most gene finding programs were optimized to predict genes in sequences from mammalian models; as a result accuracy is sometimes reduced when using sequences from more divergent organisms or non-vertebrates. These programs use different strategies and are based on current, but incomplete understanding of gene structure. Therefore, predictions from multiple programs should be combined computationally to enhance accuracy and confidence (Rogic et al. 2002).

Quality control strategies can be used to increase accuracy of and confidence in results from gene finding tools. First, the abundance of repetitive elements in genomic sequences can distort predictions of genes and exon locations. To

counter this problem, masking genomic sequences is recommended, before gene analysis, using a program like RepeatMasker (A.F.A. Smit et al. 1996, unpublished). The effectiveness of RepeatMasker with marine and other evolutionarily divergent organisms is not yet clear because interspersed repeats are specific to a species or group of species. Sequences from such organisms should be evaluated under masked and unmasked conditions. Second, aligning ESTs or cDNAs with genomic sequences using programs like Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>; Wheelan et al. 2001) refines predictions of exon and intron boundaries, promoter regions, and splice sites. This approach presents challenges for transcripts that are expressed at very low levels, have significant tissue- or age-specific requirements, or are from species for which minimal sequencing has been done (Schwartz et al. 2000). A new, publicly available resource, the Comparative Toxicogenomics Database (CTD; <http://ctd.mdibl.org/>; Mattingly et al. 2003, 2004a), provides multiple alignment and phylogenetic analysis results with sequences from diverse organisms for biomedically significant genes and proteins. CTD provides access to data valuable for identifying homologous genomic sequences and confirming gene and gene feature predictions. Identification of gene features greatly improves interpretation of subsequent multiple sequence analysis results.

Aligning multiple genomic sequences is becoming widely accepted as a powerful mechanism for identifying important functional regions such as regulatory elements. MultiPipMaker (<http://pipmaker.bx.psu.edu/pipmaker/>; Schwartz et al. 2003) and mVISTA (<http://gsd.lbl.gov/vista/index.shtml>; Bray et al. 2003; Frazer et al. 2004) are two of the most popular web servers that conveniently combine alignment engines with visualization capabilities. MultiPipMaker, and a new server zPicture (<http://zpicture.dcode.org/>; Ovcharenko et al. 2004), use the local alignment program BLASTZ as their alignment engine (Schwartz et al. 2000); VISTA uses AVID, a global alignment program (Bray et al. 2003; Frazer et al. 2004). Local alignment tools find high-scoring, short matching segments and extend these regions based on a scoring threshold. Local alignments may permit greater diversity between sequences by finding regional similarities. Segments of similarity need not be conserved in order or orientation. This

feature may be advantageous for finding conserved transcription factor binding sites, which are very short and prone to reordering (Bray et al. 2003).

High similarity of short regions, however, does not necessarily imply homology (a gene derived from a common ancestral gene) and can lead to false implications of relatedness among sequences that are not homologs. By contrast, global alignments do require that the order and orientation of similar regions is conserved, because similar architecture is often observed in homologous sequences (Bray et al. 2003). MultiPipMaker and VISTA provide visualization options for alignments that include percent similarity and user-submitted annotations (e.g., exon locations, repetitive elements). It is important to note that local and global alignment tools are being refined so rapidly that it is becoming difficult to distinguish between them (Frazer et al. 2004). Readers are referred to two recent reviews of alignment programs for detailed comparisons (Frazer et al. 2003; Pollard et al. 2004).

A major challenge to identifying transcription factor binding sites (TFBSs) is that they tend to be short, degenerate, and occur frequently throughout the genome. Analysis of a single sequence usually leads to an abundance of false positive predictions of TFBSs. Several programs have been developed to respond to this challenge based on two principles. First, functional regulatory elements are often conserved evolutionarily; therefore, identifying TFBSs that are conserved or aligned in multiple sequences may effectively filter false positive predictions (Ovcharenko et al. 2005). Second, gene expression often results from coordinate activation of multiple, proximal regulatory elements; therefore identifying TFBSs in clusters, rather than isolation, may enhance confidence in the functionality of predicted TFBSs. These principles have been leveraged, albeit differently, by rVISTA (<http://gsd.lbl.gov/vista/index.shtml>; Loots et al. 2002; Loots and Ovcharenko 2004), which is a member of the VISTA suite of tools and is also integrated with zPicture, and the newly launched Mulan (<http://mulan.dcode.org/>; Ovcharenko et al. 2005). Both servers use profiles of transcription factor binding sites from the TRANSFAC database (<http://www.biobase.de>; Matys et al. 2003). Because TRANSFAC and other similar resources only contain information for known transcription

factor binding sites, they are inherently incomplete. Furthermore, existing tools do not address other important regulatory features such as properties related to protein–protein interactions and chromatin structure, and clusters of binding sites that may have been reshuffled between organisms over evolutionary time (Loots et al. 2002).

Cell culture of marine genomic model species and experimental verification of predictions from comparative analysis of genomic sequences

Using *in vitro* cell culture systems, the functional significance of conserved, putative regulatory sequences predicted through comparative computational analysis in, for instance, the 5'-upstream region of select genes can be tested experimentally. The availability of sufficient genomic information facilitates targeted studies to evaluate such potential functional regulatory regions. Comparative experimental studies can be designed employing cell lines derived from any species, though mammalian cell lines have thus far been favored (Mather and Barnes 1997; Barnes and Sato 2000). Reporter constructs containing regulatory regions of select genes can be generated using up to 5 kb upstream of the transcriptional start site of the relevant genes; these are inserted upstream of a reporter gene such as that for an enhanced fluorescent protein.

In the last decade, significant progress has been made in development of marine and freshwater organism cell lines with utility for genomic studies. A variety of zebrafish cell lines have been developed, some of which maintain a normal karyotype for extended periods *in vitro* (Barnes and Collodi 2005). Zebrafish cells in culture can be transfected with plasmid DNA using adaptations of approaches common for mammalian cells *in vitro*, and transient transfection methods have identified expression of genes under control of a number of mammalian and fish promoters. In addition, cell cultures from pufferfish (genera *Fugu* and *Tetraodon*) provide a biological complement to the genomic libraries derived to study the molecular biology of these animals, allowing the extension of this model to experiments in functional genomics. Multipassage cell cultures

have been established from embryo and adult tissues of species of both of these pufferfish genera (Barnes and Collodi 2005). One of the *Fugu rubripes* cultures has been maintained for more than 200 population doublings, and flow cytometry showed that the relative amount of DNA present in cultured cells was approximately 15% of that in human cells, as predicted by biochemical analysis. Telomerase, an enzyme associated with indefinite proliferation in mammalian cell cultures, was easily detectable in these cells, suggesting that the cultures are capable of indefinite growth.

Until recently, the lack of *in vitro* culture systems for elasmobranch models has been a major limitation for the use of these species in comparative functional genomics. Cold-water marine animals such as the little skate and dogfish shark are useful models for physiology and genomics, but homologous *in vitro* systems with which investigators can test hypotheses or confirm predictions at a molecular level are essential for widespread use of these models. Expression of elasmobranch genes in heterologous systems such as mammalian cell cultures or *Xenopus* oocytes may be compromised by differences in membrane lipid composition and missing or interfering accessory and messenger proteins. Heterologous systems are also not adaptable to mechanistic studies using genetically altered, dominant negative molecules. Another advantage derived from cell cultures from species such as shark and skate (that may be only seasonably available) is that they make biological material available year-round in any laboratory. Studies on regulation of gene expression in dogfish shark and little skate are beginning to benefit from the expanding genomic databases for these animals. Skate embryo-derived cell cultures also provide new avenues for elasmobranch research in embryology and organogenesis, toxicology, neurobiology, genome regulation and comparative stem cell biology.

Cells have been cultured from the rectal gland, eye, brain, kidney and early embryo of *Squalus acanthias*. Medium is either LDF, developed for zebrafish cells (Barnes and Collodi 2005) or VCM (Valentich and Forrest 1991), a urea-timethylamine oxide (TMAO)-containing medium that supports the short-term culture of shark rectal gland cells. In some cases cells are plated onto

dishes pretreated with collagen. Medium is further supplemented with a variety of cell type-specific peptide growth factors, nutrients and purified proteins at a range of concentrations. These include insulin, transferrin, epidermal growth factor (EGF), basic fibroblast growth factor (FGF), transforming growth factor-beta, vitamins A and E, selenium, mercaptoethanol, dexamethasone, fetal calf serum, shark serum and shark yolk extract. For example shark brain cells can be grown in primary culture in VCM containing EGF, FGF, insulin, transferrin, selenium, a chemically-defined lipid supplement and vitamin E on a fibronectin substratum (Figure 1). Expression of genes from plasmids has been achieved in primary dogfish shark rectal gland cell cultures by lipofection using the CMV promoter to direct expression of the gene of interest (Figure 2). Differentiated function also is maintained, as evidenced by expression of CFTR and vasoactive intestinal peptide receptor (VIPR) mRNA detected by reverse-transcription polymerase chain reaction (RT-PCR) assay (Figures 3 and 4).

Assay for telomerase on cultures of a variety of dogfish shark and little skate cell types showed that all cultures tested were positive, although specific activity varied almost 100-fold among different cell types. Assay for cell proliferation by *in situ* bromo-deoxyuridine (BrdU) incorporation also was carried out on cultures from shark brain, kidney and rectal gland. The assay involves an overnight incubation with BrdU, followed by an immunoassay identifying cells synthesizing DNA during the time of incubation and incorporating the nucleoside analogue. The results showed that cells in the cultures were synthesizing DNA. The most active synthesis was seen in cultures from early shark embryos. Medium was supplemented with insulin, transferrin, selenous acid, EGF, FGF, L-glutamine, chemically defined lipids, non-essential amino acids, heat-inactivated fetal bovine serum (heat treated for 30 min at 56 °C) and shark yolk extract. Conditioned medium from the cells was stimulatory for other shark cell cultures, including the shark rectal gland cells (Figure 5). A normalized c-DNA library has been made from these cells and attempts will be made to identify the cell type by extensive EST analysis.

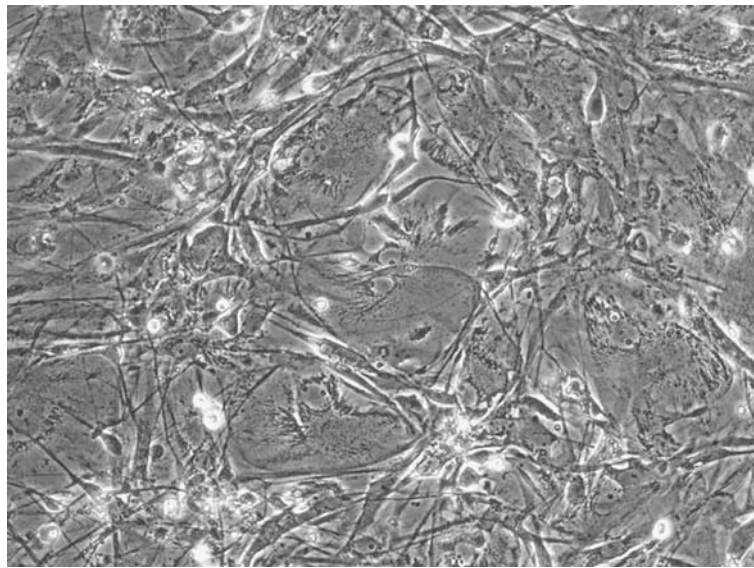


Figure 1. Dogfish shark brain cell culture, third passage.

In addition to the scarcity of cell lines from marine vertebrates, a persistent absence of cell lines from non-arthropod invertebrates has stymied research on many valuable species useful in comparative genomics and a variety of other disciplines (Bayne 1998; Rinkevich 1999). Selection of

the echinoderm *Strongylocentrotus purpuratus* (purple sea urchin) as a model species for genomic sequencing has enhanced the need for cell lines from this species in particular. We recently have explored cell culture using both this species and the closely related *Strongylocentrotus droebachi-*

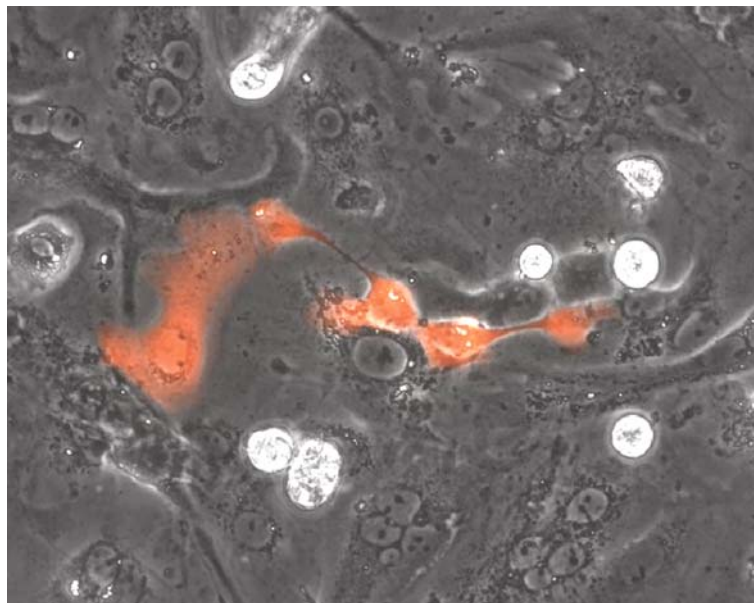


Figure 2. Dogfish shark rectal gland cell culture 10 days after transfection with red fluorescent protein reporter gene under control of the cytomegalovirus early promoter-enhancer.

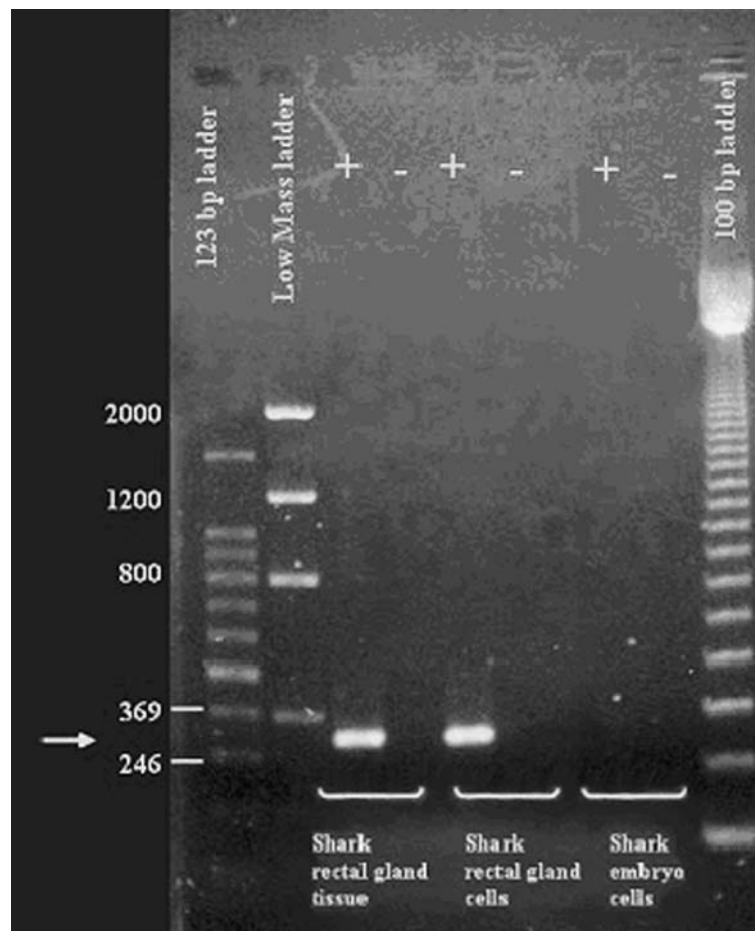


Figure 3. RT PCR identifying expression of VIPR mRNA in shark rectal gland tissue and primary culture. Expected product is 343 bp. No expression was seen in third passage embryo cell cultures. Equal RNA amounts were loaded. (–) Samples had no reverse transcriptase.

ensis (green sea urchin). Cells from Polian vesicles and axial organ (Figures 6 and 7) appear to have proliferated *in vitro* for more than 8 months (Parton and Bayne 2004, 2005). Other cultures yielded thraustochytrid protists that are common parasites of marine invertebrates worldwide (Rinkevich 1999). Basal nutrient culture medium was LDF diluted with 4 volumes of filtered sea water (Kawamura and Fujiwara 1995) with antibiotics (penicillin 200 U/ml, streptomycin 200 μ g/ml, ampicillin 25 μ g/ml), sodium bicarbonate at 0.18 g/l and HEPES at 15 mM. Culture vessels were held at 20 °C in ambient air. Additives included fetal calf serum (heat inactivated) at 1–3%; insulin at 10 μ g/ml; transferrin at 10 μ g/ml; FGF at 50 ng/ml; EGF at 50 ng/ml; β -mercapto-etha-

nol at 55 μ M; chemically defined lipid supplement at 1 μ l/ml; selenium at 10 nM, and L-glutamine at 200 μ M.

Conclusion

Transcriptional regulation involves complex interactions of diverse proteins, or transcription factors, with specific DNA sequences in the non-coding regions of target genes. Furthermore, cells respond to environmental stimuli and to developmental signals by altering expression of gene networks. The limited number of transcription factors suggests that few, if any of these proteins exert their activity exclusively on a single gene;

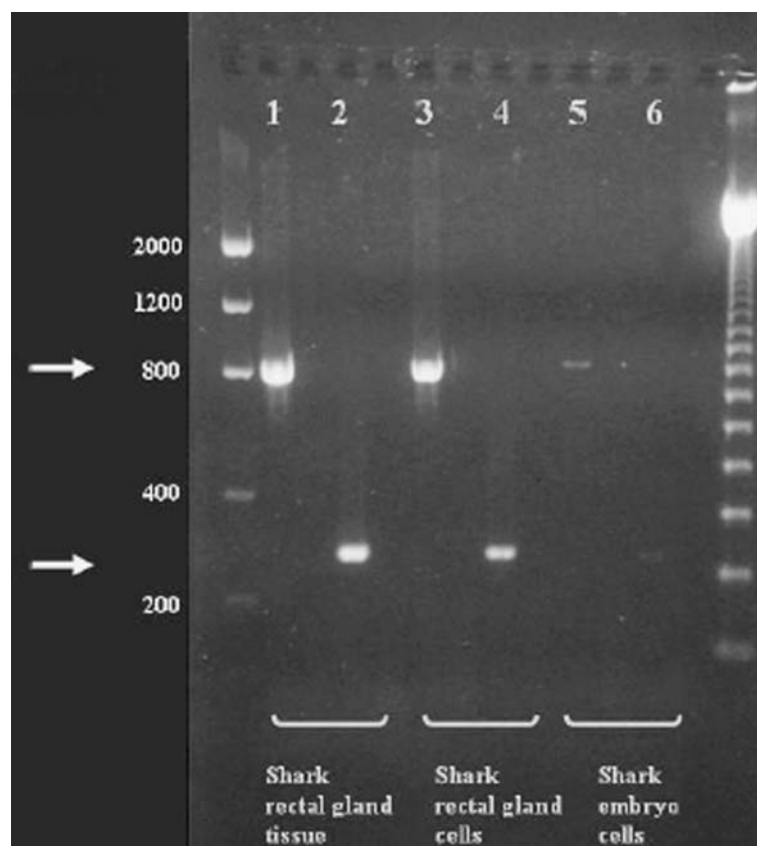


Figure 4. RT PCR identifying expression of CFTR mRNA in shark rectal gland tissue and primary culture. Expected PCR products were 843 and 277 bp in assays with separate primers.

rather, they bind to conserved sites in several genes to coordinate their expression (Wagner 1999; Pennacchio and Rubin 2003). In situations in which there are no clinically acceptable inhibitors or other modulators of clinically relevant proteins, elucidating mechanisms by which these genes are regulated and identifying other coordinately regulated genes may reveal novel strategies by which disease processes may be disrupted or controlled.

Comparative studies provide insight into the possible common ancestor of a gene, trace the accumulation of mutations over time and suggest selective pressures that influence the expression and functions of genes (Pennacchio and Rubin 2003). Comparisons have provided useful predictions about which sequences are minimally essential for function as well as those that may be important on a species-specific level

(Aparicio et al. 1995). Comparative genomic computational approaches continue to identify conserved regions in non-coding sequences. The challenges will be to determine which sequences are functionally significant and to identify coordinately regulated genes and common regulatory pathways. Understanding such networks, the transcription factor binding sites and genes involved in disease states may reveal alternative points of intervention and contribute to a more predictive approach to molecular medicine.

Significant progress has been made in the development of powerful sequence analysis tools. Although often optimized for sequences from more traditional model organisms, they offer great value for comparative studies that include evolutionarily divergent sequences. They should be used cautiously, however, with awareness of

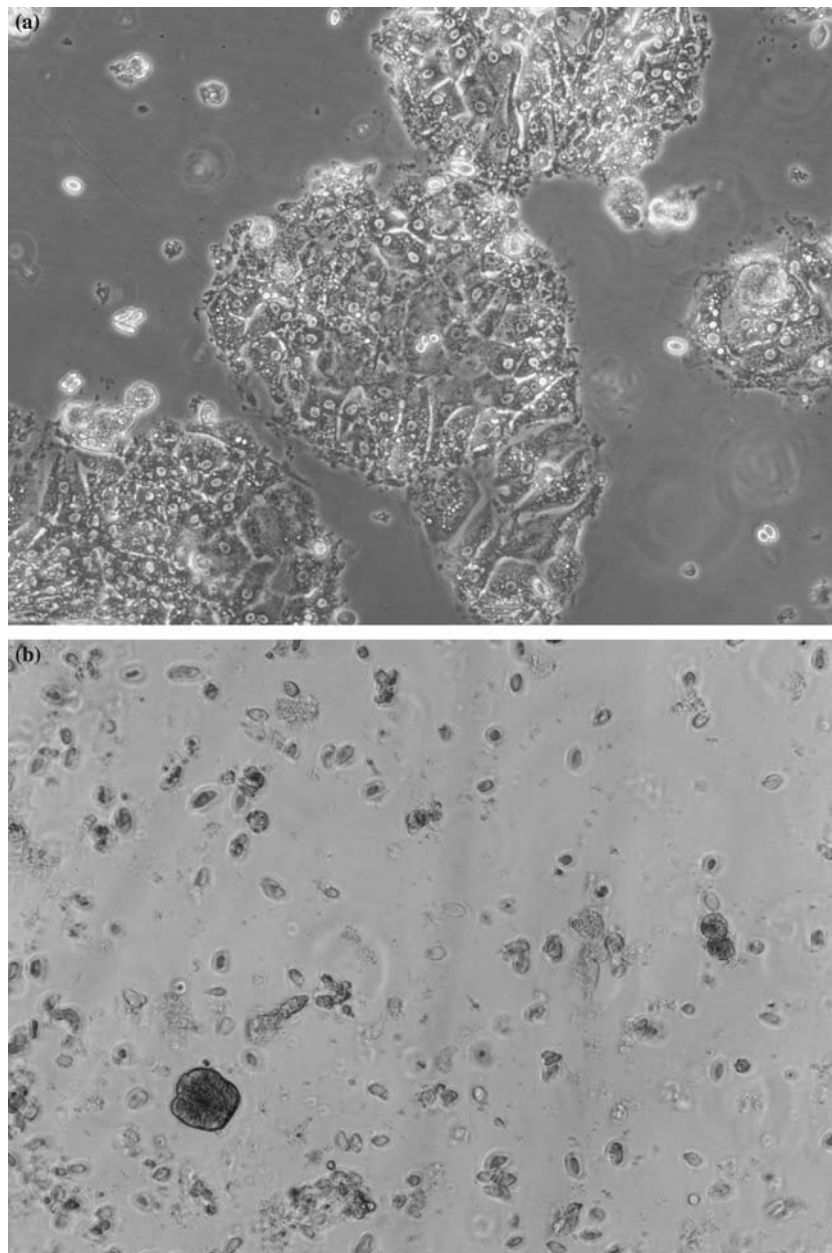


Figure 5. (a) Shark rectal gland cells 20 h after plating on collagen preincubated with conditioned medium from sixth-passage shark embryo-derived cell culture. (b) Shark rectal gland cells 20 h after plating on collagen without conditioned medium.

their limitations, which are largely a reflection of current scientific knowledge and understanding of genome structure. As the diversity of available sequence data continues to increase, it will drive refinements and development of new tools and provide important insights into the evolution and essential mechanisms of gene expression regulation.

Acknowledgments

Supported by funds from the NIH (DK 34208, ES03828, ES11267, RR016463, RO1019732), the Woodcock Foundation, and the Maine Marine Research Fund. The authors thank Drs J. Boyer and C. Bult. D.B. thanks P. Collodi, A. Miller and T. van Zandt.

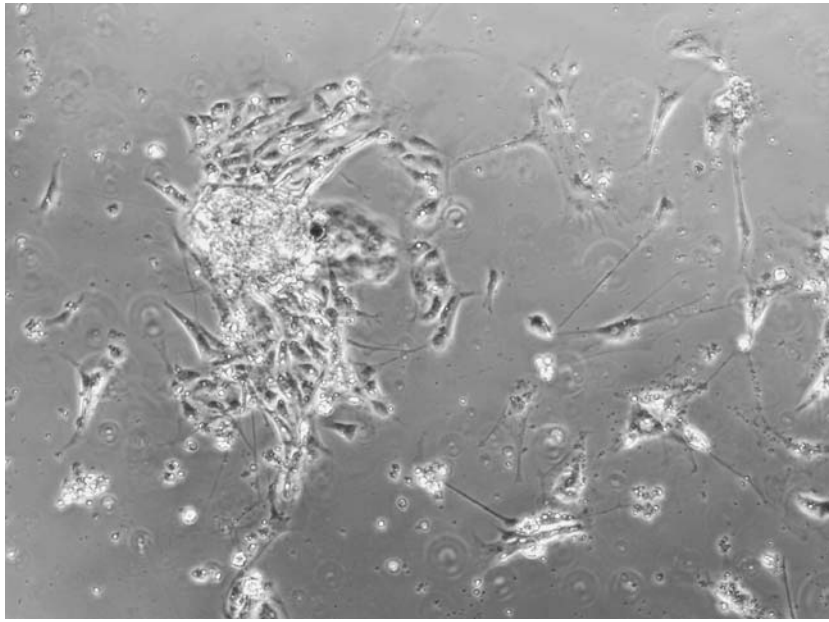


Figure 6. Primary culture of cells from sea urchin (*Stroglyocentrotus*) Polian vesicle.

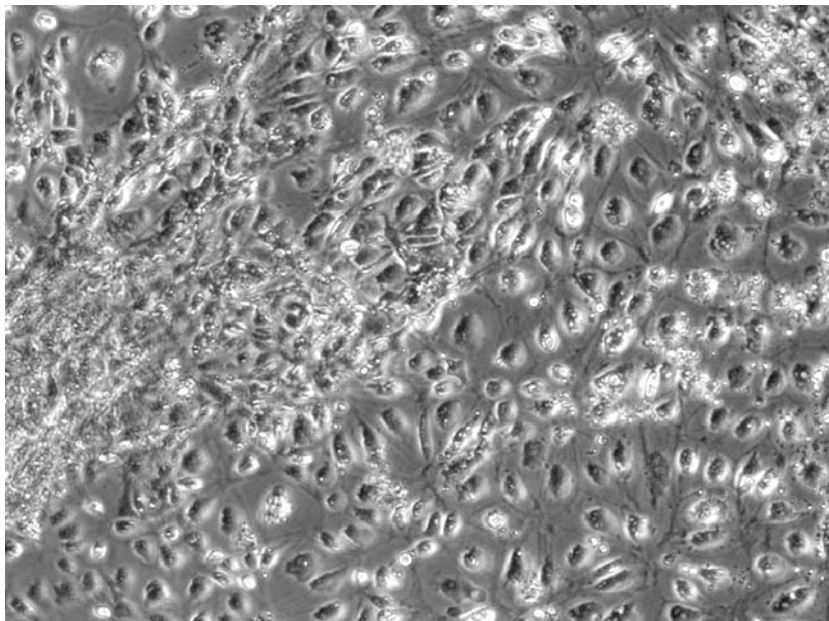


Figure 7. Primary culture of cells from sea urchin (*Stroglyocentrotus*) axial organ.

References

Adelman M.K., Schluter S.F. and Marchalonis J.J. 2004. The natural antibody repertoire of sharks and humans recognizes the potential universe of antigens. *Protein J.* 23(2): 103–118.

Ahituv N., Rubin E.M. and Nobrega M.A. 2004. Exploiting human–fish genome comparisons for deciphering gene regulation. *Hum. Mol. Genet.* 13 Spec No 2:R261–266.

Aller S.G., Lombardo I.D., Bhanot S. and Forrest J.N. Jr. 1999. Cloning, characterization, and functional expression of a CNP receptor regulating CFTR in the shark rectal gland. *Am. J. Physiol.* 276: C442–C449.

Aparicio S., Chapman J., Stupka E., Putnam N., Chia J.M., Dehal P., Christoffels A., Rash S., Hoon S., Smit A., Gelpke M.D., Roach J., Oh T., Ho I.Y., Wong M., Detter C., Verhoeve F., Predki P., Tay A., Lucas S., Richardson P., Smith

- S.F., Clark M.S., Edwards Y.J., Doggett N., Zharkikh A., Tavtigian S.V., Pruss D., Barnstead M., Evans C., Baden H., Powell J., Glusman G., Rowen L., Hood L., Tan Y.H., Elgar G., Hawkins T., Venkatesh B., Rokhsar D. and Brenner S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297: 1301–1310.
- Aparicio S., Morrison A., Gould A., Gilthorpe J., Chaudhuri C., Rigby P., Krumlauf R. and Brenner S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl. Acad. Sci. USA* 92: 1684–1688.
- Ballatori N., Boyer J.L. and Rockett J.C. 2003. Exploiting genome data to understand the function, regulation, and evolutionary origins of toxicologically relevant genes. *Environ. Health Perspect. Toxicogenom.* 111: doi:10.1289/txg.5961.
- Barnes D. and Sato G.H. 2000. In Lazarrini P. (ed.), *Cell Culture Systems in Tissue Engineering*, pp. 111–118.
- Barnes D. and Collodi P. 2005. In: Evans D. and Claiborne J.B. (eds), *Fish Cell Lines and Stem Cells in the Physiology of Fishes*, 3rd edn. CRC Press, in press.
- Bayne C.J. 1998. In: Barnes D.W. and Mather J.P. (eds), *Invertebrate Cell Culture Considerations: Insects, Ticks, Shellfish, and Worms*. *Methods in Cell Biology*, Vol. 57. Academic Press, Chapter 10, pp. 187–201.
- Bayne C.J., Gerwick L., Fujiki K., Nakao M. and Yano T. 2001. Immune-relevant (including acute phase) genes identified in the livers of rainbow trout, *Oncorhynchus mykiss*, by means of suppression subtractive hybridization. *Dev. Comp. Immunol.* 25: 205–217.
- Boguski M.S., Lowe T.M. and Tolstoshev C.M. 1993. dbEST – database for “expressed sequence tags”. *Nat. Genet.* 4: 332–333.
- Bray N., Dubchak I. and Pachter L. 2003. AVID: a global alignment program. *Genome Res.* 13: 97–102.
- Burge C. and Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78–94.
- Cai S.Y., Soroka C.J., Ballatori N. and Boyer J.L. 2003. Molecular characterization of a multidrug resistance-associated protein, Mrp2, from the little skate. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 284: R125–R130.
- Cai S.Y., Wang L., Ballatori N. and Boyer J.L. 2001. Bile salt export pump is highly conserved during vertebrate evolution and its expression is inhibited by PFIC type II mutations. *Am. J. Physiol. Gastrointest. Liver Physiol.* 281: G316–G322.
- Chiu C.H., Amemiya C., Dewar K., Kim C.B., Ruddle F.H. and Wagner G.P. 2002. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. *Proc. Natl. Acad. Sci. USA* 99: 5492–5497.
- Dehal P., Satou Y., Campbell R.K., Chapman J., Degnan B., De Tomaso A., Davidson B., Di Gregorio A., Gelpke M., Goodstein D.M., Harafuji N., Hastings K.E., Ho I., Hotta K., Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen I.A., Nacula S., Nonaka M., Putnam N., Rash S., Saiga H., Satake M., Terry A., Yamada L., Wang H.G., Awazu S., Azumi K., Boore J., Branno M., Chin-Bow S., DeSantis R., Doyle S., Francino P., Keys D.N., Haga S., Hayashi H., Hino K., Imai K.S., Inaba K., Kano S., Kobayashi K., Kobayashi M., Lee B.I., Makabe K.W., Manohar C., Matassi G., Medina M., Mochizuki Y., Mount S., Morishita T., Miura S., Nakayama A., Nishizaka S., Nomoto H., Ohta F., Oishi K., Rigoutsos I., Sano M., Sasaki A., Sasakura Y., Shoguchi E., Shin-i T., Spagnuolo A., Stainier D., Suzuki M.M., Tassy O., Takatori N., Tokuoka M., Yagi K., Yoshizaki F., Wada S., Zhang C., Hyatt P.D., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T., Richardson P., Lucas S., Kohara Y., Levine M., Satoh N. and Rokhsar D.S. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298: 2157–2167.
- Devor D.C., Forrest J.N. Jr., Suggs W.K. and Frizzell R.A. 1995. cAMP-activated Cl-channels in primary cultures of spiny dogfish (*Squalus acanthias*) rectal gland. *Am. J. Physiol.* 268: C70–C79.
- Dooley K. and Zon L.I. 2000. Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* 10: 252–256.
- Dubchak I., Brudno M., Loots G.G., Pachter L., Mayor C., Rubin E.M. and Frazer K.A. 2000. Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* 10: 1304–1306.
- Forrest J.N. Jr. 1996. Cellular and molecular biology of chloride secretion in the shark rectal gland: regulation by adenosine receptors. *Kidney Int.* 49: 1557–1562.
- Frazer K.A., Elnitski L., Church D.M., Dubchak I. and Hardison R.C. 2003. Cross-species sequence comparisons: a review of methods and available resources. *Genome Res.* 13: 1–12.
- Frazer K.A., Pachter L., Poliakov A., Rubin E.M. and Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucl. Acids Res.* 32: W273–W279.
- Fujiki K., Bayne C.J., Shin D.H., Nakao M. and Yano T. 2001. Molecular cloning of carp (*Cyprinus carpio*) C-type lectin and pentraxin by use of suppression subtractive hybridisation. *Fish Shellfish Immunol.* 11: 275–279.
- Fujiki K., Gerwick L., Bayne C.J., Mitchell L., Gauley J., Bols N. and Dixon B. 2003. Molecular cloning and characterization of rainbow trout (*Oncorhynchus mykiss*) CCAAT/enhancer binding protein beta. *Immunogenetics* 55: 53–261.
- Gamba G., Saltzberg S.N., Lombardi M., Miyanoshita A., Lytton J., Hediger M.A., Brenner B.M. and Hebert S.C. 1993. Primary structure and functional expression of a cDNA encoding the thiazide-sensitive, electroneutral sodium-chloride cotransporter. *Proc. Natl. Acad. Sci. USA* 90: 2749–2753.
- Greger R., Bleich M., Warth R., Thiele I. and Forrest J.N. 1999. The cellular mechanisms of Cl-secretion induced by C-type natriuretic peptide (CNP). Experiments on isolated in vitro perfused rectal gland tubules of *Squalus acanthias*. *Pflugers Arch.* 438: 15–22.
- Henson J.H., Roesener C.D., Gaetano C.J., Mendola R.J., Forrest J.N. Jr., Holy J. and Kleinzeller A. 1997. Confocal microscopic observation of cytoskeletal reorganizations in cultured shark rectal gland cells following treatment with hypotonic shock and high external K⁺. *J. Exp. Zool.* 279: 415–424.
- Hinds K.R. and Litman G.W. 1986. Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. *Nature* 320: 546–549.
- Hughes J.D., Estep P.W., Tavazoie S. and Church G.M. 2000. Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 296: 1205–1214.

- Ishikawa J., Imai E., Moritomo T., Nakao M., Yano T. and Tomana M. 2004. Characterisation of a fourth immunoglobulin light chain isotype in the common carp. *Fish Shellfish Immunol.* 16: 369–379.
- Karolchik D., Baertsch R., Diekhans M., Furey T.S., Hinrichs A., Lu Y.T., Roskin K.M., Schwartz M., Sugnet C.W., Thomas D.J., Weber R.J., Haussler D. and Kent W.J. 2003. The UCSC Genome Browser Database. *Nucl. Acids Res.* 31: 51–54.
- Kawamura K. and Fujiwara S. 1995. Establishment of cell lines from multipotent epithelial sheet in the budding tunicate, *Polyandrocarpa misakiensis*. *Cell Struct. Funct.* 20: 97–106.
- Ke Q., Yang Y., Ratner M., Zeind J., Jiang C., Forrest J.N. Jr. and Xiao Y.F. 2002. Intracellular accumulation of mercury enhances P450 CYP1A1 expression and Cl-currents in cultured shark rectal gland cells. *Life Sci.* 21: 2547–2566.
- Korf I., Flicek P. and Duan D. and Brent M.R. 2001. Integrating genomic homology into gene structure prediction. *Bioinformatics* 17(Suppl 1): S140–148.
- Lehrich R.W. and Forrest J.N. Jr. 1995. Tyrosine phosphorylation is a novel pathway for regulation of chloride secretion in shark rectal gland. *Am. J. Physiol.* 269: F594–F600.
- Lehrich R.W., Aller S.G., Webster P., Marino C.R. and Forrest J.N. Jr. 1998. Vasoactive intestinal peptide, forskolin, and genistein increase apical CFTR trafficking in the rectal gland of the spiny dogfish, *Squalus acanthias*. Acute regulation of CFTR trafficking in an intact epithelium. *J. Clin. Invest.* 101: 737–745.
- Loots G.G., Locksley R.M., Blankespoor C.M., Wang Z.E., Miller W., Rubin E.M. and Frazer K.A. 2000. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288: 136–140.
- Loots G.G. and Ovcharenko I. 2004. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucl. Acids Res.* 32: W217–W221.
- Loots G.G., Ovcharenko I., Pachter L., Dubchak I. and Rubin E.M. 2002. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.* 12: 832–839.
- Marshall J., Martin K.A., Picciotto M., Hockfield S., Nairn A.C. and Kaczmarek L.K. 1991. Identification and localization of a dogfish homolog of human cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* 266: 22749–22754.
- Mather J. and Barnes D. 1997. *Cell Culture Methods*, Vol. 57, *Methods in Cell Biology*. Academic Press Inc., NY.
- Mattingly C.J., Colby G.T., Forrest J.N. and Boyer J.L. 2003. The Comparative Toxicogenomics Database (CTD). *Environ. Health Perspect.* 111: 793–795.
- Mattingly C.J., Colby G.T., Rosenstein M.C., Forrest J.N. Jr. and Boyer J.L. 2004a. Promoting comparative molecular studies in environmental health research: an overview of the comparative toxicogenomics database (CTD). *Pharmacogenom. J.* 4: 5–8.
- Mattingly C., Parton A., Dowell L., Rafferty J. and Barnes D. 2004b. Cell and molecular biology of marine elasmobranchs: *Squalus acanthias* and *Raja erinacea*. *Zebrafish* 1: 111–120.
- Matys V., Fricke E., Geffers R., Gossling E., Haubrock M., Hehl R., Hornischer K., Karas D., Kel A.E., Kel-Margoulis O.V., Kloos D.U., Land S., Lewicki-Potapov B., Michael H., Munch R., Reuter I., Rotert S., Saxel H., Scheer M., Thiele S. and Wingender E. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucl. Acids Res.* 31: 374–378.
- Naruse K., Tanaka M., Mita K., Shima A., Postlethwait J. and Mitani H. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* 14: 820–828.
- Ovcharenko I., Loots G.G., Giardine B.M., Hou M., Ma J., Hardison R.C., Stubbs L. and Miller W. 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res.* 15: 184–194.
- Ovcharenko I., Loots G.G., Hardison R.C., Miller W. and Stubbs L. 2004. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.* 14: 472–477.
- Pennacchio L.A. and Rubin E.M. 2003. Comparative genomic tools and databases: providing insights into the human genome. *J. Clin. Invest.* 111: 1099–1106.
- Pollard D.A., Bergman C.M., Stoye J., Celniker S.E. and Eisen M.B. 2004. Benchmarking tools for the alignment of functional noncoding DNA. *BMC Bioinform.* 5: 6.
- Rinkevich B. 1999. Cell cultures from marine invertebrates: obstacles, new approaches and recent developments. *J. Biotechnol.* 70: 133–153.
- Rogic S., Ouellette B.F. and Mackworth A.K. 2002. Improving gene recognition accuracy by combining predictions from two gene-finding programs. *Bioinformatics* 18: 1034–1045.
- Rowntree R.K., Vassaux G., McDowell T.L., Howe S., McGuigan A., Phylactides M., Huxley C. and Harris A. 2001. An element in intron 1 of the CFTR gene augments intestinal expression in vivo. *Hum. Mol. Genet.* 10: 1455–1464.
- Santini S., Boore J.L. and Meyer A. 2003. Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res.* 13: 1111–1122.
- Schwartz S., Elnitski L., Li M., Weirauch M., Riemer C., Smit A., Green E.D., Hardison R.C. and Miller W. 2003. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. *Nucl. Acids Res.* 31: 3518–3524.
- Schwartz S., Zhang Z., Frazer K.A., Smit A., Riemer C., Bouck J., Gibbs R., Hardison R. and Miller W. 2000. PipMaker – a web server for aligning two genomic DNA sequences. *Genome Res.* 10: 577–586.
- Silva P., Solomon R.J. and Epstein F.H. 1999. Mode of activation of salt secretion by C-type natriuretic peptide in the shark rectal gland. *Am. J. Physiol.* 277: R1725–R1732.
- Stamm S., Ben-Ari S., Rafalska I., Tang Y., Zhang Z., Toiber D., Thanaraj T.A. and Soreq H. 2005. Function of alternative splicing. *Gene* 344: 1–20.
- Thomas J.W., Prasad A.B., Summers T.J., Lee-Lin S.Q., Maduro V.V., Idol J.R., Ryan J.F., Thomas P.J., McDowell J.C. and Green E.D. 2002. Parallel construction of orthologous sequence-ready clone contig maps in multiple species. *Genome Res.* 12: 1277–1285.
- Thomas J.W. and Touchman J.W. 2002. Vertebrate genome sequencing: building a backbone for comparative genomics. *Trends Genet* 18: 104–108.
- Thomas J.W., Touchman J.W., Blakesley R.W., Bouffard G.G., Beckstrom-Sternberg S.M., Margulies E.H., Blanchette M., Siepel A.C., Thomas P.J., McDowell J.C., Maskeri B., Hansen N.F., Schwartz M.S., Weber R.J., Kent W.J., Karolchik D.,

- Bruen T.C., Bevan R., Cutler D.J., Schwartz S., Elnitski L., Idol J.R., Prasad A.B., Lee-Lin S.Q., Maduro V.V., Summers T.J., Portnoy M.E., Dietrich N.L., Akhter N., Ayele K., Benjamin B., Cariaga K., Brinkley C.P., Brooks S.Y., Granite S., Guan X., Gupta J., Haghighi P., Ho S.L., Huang M.C., Karlins E., Laric P.L., Legaspi R., Lim M.J., Maduro Q.L., Masiello C.A., Mastrian S.D., McCloskey J.C., Pearson R., Stantripop S., Tiangson E.E., Tran J.T., Tsurgeon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L.S., Young A.C., Zhang L.H., Osoegawa K., Zhu B., Zhao B., Shu C.L., De Jong P.J., Lawrence C.E., Smit A.F., Chakravarti A., Haussler D., Green P., Miller W. and Green E.D. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424: 788–793.
- Tomana M., Ishikawa J., Imai E., Moritomo T., Nakao M. and Yano T. 2002. Characterization of immunoglobulin light chain isotypes in the common carp. *Immunogenetics* 88: 120–129.
- Uberbacher E.C. and Mural R.J. 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* 88: 11261–11265.
- Valentich J.D. and Forrest J.N. Jr. 1991. Cl-secretion by cultured shark rectal gland cells. I. Transepithelial transport. *Am. J. Physiol.* 260: C813–C823.
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman J.R., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M., Subramanian G., Thomas P.D., Zhang J., Gabor Miklos G.L., Nelson C., Broder S., Clark A.G., Nadeau J., McKusick V.A., Zinder N., Levine A.J., Roberts R.J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielson A.E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T.J., Higgins M.E., Ji R.R., Ke Z., Ketchum K.A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G.V., Milshina N., Moore H.M., Naik A.K., Narayan V.A., Neelam B., Nusskern D., Rusch D.B., Salzberg S., Shao W., Shue B., Sun J., Wang Z., Wang A., Wang X., Wang J., Wei M., Wides R., Xiao C., Yan C., Yao A., Ye J., Zhan M., Zhang W., Zhang H., Zhao Q., Zheng L., Zhong F., Zhong W., Zhu S., Zhao S., Gilbert D., Baumhueter S., Spier G., Carter C., Cravchik A., Woodage T., Ali F., An H., Awe A., Baldwin D., Baden H., Barnstead M., Barrow I., Beeson K., Busam D., Carver A., Center A., Cheng M.L., Curry L., Danaher S., Davenport L., Desilets R., Dietz S., Dodson K., Doup L., Ferriera S., Garg N., Gluecksmann A., Hart B., Haynes J., Haynes C., Heiner C., Hladun S., Hostin D., Houck J., Howland T., Ibegwam C., Johnson J., Kalush F., Kline L., Koduru S., Love A., Mann F., May D., McCawley S., McIntosh T., McMullen I., Moy M., Moy L., Murphy B., Nelson K., Pfannkoch C., Pratts E., Puri V., Qureshi H., Reardon M., Rodriguez R., Rogers Y.H., Romblad D., Ruhfel B., Scott R., Sitter C., Smallwood M., Stewart E., Strong R., Suh E., Thomas R., Tint N.N., Tse S., Vech C., Wang G., Wetter J., Williams S., Williams M., Windsor S., Winn-Deen E., Wolfe K., Zaveri J., Zaveri K., Abril J.F., Guigo R., Campbell M.J., Sjolander K.V., Karlak B., Kejariwal A., Mi H., Lazareva B., Hatton T., Narechania A., Diemer K., Muruganujan A., Guo N., Sato S., Bafna V., Istrail S., Lippert R., Schwartz R., Walenz B., Yooseph S., Allen D., Basu A., Baxendale J., Blick L., Caminha M., Carnes-Stine J., Caulk P., Chiang Y.H., Coyne M., Dahlke C., Mays A., Dombroski M., Donnelly M., Ely D., Esparham S., Fosler C., Gire H., Glanowski S., Glasser K., Glodek A., Gorokhov M., Graham K., Gropman B., Harris M., Heil J., Henderson S., Hoover J., Jennings D., Jordan C., Jordan J., Kasha J., Kagan L., Kraft C., Levitsky A., Lewis M., Liu X., Lopez J., Ma D., Majoros W., McDaniel J., Murphy S., Newman M., Nguyen T., Nguyen N., Nodell M., Pan S., Peck J., Peterson M., Rowe W., Sanders R., Scott J., Simpson M., Smith T., Sprague A., Stockwell T., Turner R., Venter E., Wang M., Wen M., Wu D., Wu M., Xia A., Zandieh A. and Zhu X. 2001. The sequence of the human genome. *Science* 291: 1304–1351.
- Wagner A. 1999. Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics* 15: 776–784.
- Waldegger S., Fakler B., Bleich M., Barth P., Hopf A., Schulte U., Busch A.E., Aller S.G., Forrest J.N. Jr., Greger R. and Lang F. 1999. Molecular and functional characterization of s-KCNQ1 potassium channel from rectal gland of *Squalus acanthias*. *Pflügers Arch.* 437: 298–304.
- Walter R.B., Rains J.D., Russell J.E., Guerra T.M., Daniels C., Johnston D.A., Kumar J., Wheeler A., Kelnar K., Khanolkar V.A., Williams E.L., Hornecker J.L., Hollek L., Mamerow M.M., Pedroza A. and Kazianis S. 2004. A microsatellite genetic linkage map for *Xiphophorus*. *Genetics* 168: 363–372.
- Wang L., Soroka C.J. and Boyer J.L. 2002. The role of bile salt export pump mutations in progressive familial intrahepatic cholestasis type II. *J. Clin. Invest.* 110: 965–972.
- Wheeler S.J., Church D.M. and Ostell J.M. 2001. Spidey: a tool for mRNA-to-genomic alignments. *Genome Res.* 11: 1952–1957.
- Xu J.C., Lytle C., Zhu T.T., Payne J.A., Benz E. Jr. and Forbush B. 3rd 1994. Molecular cloning and functional expression of the bumetanide-sensitive Na–K–Cl cotransporter. *Proc. Natl. Acad. Sci. USA* 91: 2201–2205.
- Yang T., Forrest S.J., Stine N., Endo Y., Pasumarthy A., Castrop H., Aller S., Forrest J.N. Jr., Schnermann J. and Briggs J. 2002. Cyclooxygenase cloning in dogfish shark, *Squalus acanthias*, and its role in rectal gland Cl secretion. *Am. J. Physiol. Regulat. Integr. Comp. Physiol.* 283: R631–R637.
- Zhang M.Q. 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci. USA* 94: 565–568.