1-1-2015

Gene Expression in the Choanoderm

Jesús Federico Peña
University of Denver

Gene Expression in the Choanoderm

Abstract

The body plan of sponges (phylum Porifera) is an outlier among modern animals and is thought to have special evolutionary significance. Sponges lack muscles, nerves and a gut. Instead, they are composed of few cell types and simple tissues that function to pump water through an internal canal network where bacterial prey are filtered by a specialized tissue called the choanoderm. The choanoderm is composed of cells with striking similarity to choanoflagellates, the unicellular relatives of animals. Thus, the traditional view is that the sponge choanoderm is a useful model of the first animal epithelial tissues. Using the freshwater sponge *Ephydatia muelleri*, we have performed gene expression analysis of the choanoderm tissue and have begun to develop an experimental method to validate and characterize the function of candidate choanoderm genes. The data suggest that the choanoderm may be the only metazoan tissue not reliant on the classical cadherin/catenin complex for cell adhesion. Yet we find evidence for conserved developmental mechanisms and other structural features such as epithelial polarity and microvillar organization. Finally, we will explore the possibility that genes unique to choanoflagellates and sponges, have conserved functions in the choanoderm tissue. This prediction derives from the hypothesized homology of these putatively ancient cell types.

Document Type

Thesis

Degree Name

M.S.

Department

Biological Sciences

First Advisor

Scott A. Nichols, Ph.D.

Second Advisor

Thomas W. Quinn

Third Advisor

Alysia Mortimer

Keywords

Choanoderm, Sponges, Phylum Porifera, Gene expression analysis

Subject Categories

Biology | Genetics and Genomics | Life Sciences

GENE EXPRESSION IN THE CHOANODERM

A Thesis

Presented to

The Faculty of Natural Sciences and Mathematics

University of Denver

In Partial fulfilment

of the Requirements for the Degree

Master of Science

---

by

Jesús Federico Peña

June 2015

Advisor: Dr. Scott A. Nichols

Author: Jesús Federico Peña
Title: GENE EXPRESSION IN THE CHOANODERM
Advisor: Dr. Scott A. Nichols
Degree Date: June 2015

Abstract

The body plan of sponges (phylum Porifera) is an outlier among modern animals and is thought to have special evolutionary significance. Sponges lack muscles, nerves and a gut. Instead, they are composed of few cell types and simple tissues that function to pump water through an internal canal network where bacterial prey are filtered by a specialized tissue called the choanoderm. The choanoderm is composed of cells with striking similarity to choanoflagellates, the unicellular relatives of animals. Thus, the traditional view is that the sponge choanoderm is a useful model of the first animal epithelial tissues. Using the freshwater sponge *Ephydatia muelleri*, we have performed gene expression analysis of the choanoderm tissue and have begun to develop an experimental method to validate and characterize the function of candidate choanoderm genes. The data suggest that the choanoderm may be the only metazoan tissue not reliant on the classical cadherin/catenin complex for cell adhesion. Yet we find evidence for conserved developmental mechanisms and other structural features such as epithelial polarity and microvillar organization. Finally, we will explore the possibility that genes unique to choanoflagellates and sponges, have conserved functions in the choanoderm tissue. This prediction derives from the hypothesized homology of these putatively ancient cell types.

Acknowledgements

Firstly, I would like to thank my advisor, Dr. Scott Nichols, for his guidance and mentoring throughout my graduate school experience. Not only did he teach me about the broad field of evolutionary developmental biology, but I also learned what it means to do good science.

I would like to thank the members of my committee, Dr. Alysia Mortimer and Dr. Thomas Quinn for their support and input throughout the research process. A special thanks to Dr. Bryan Cowen for his time and effort serving on as the chair of the committee.

This project would not have been possible without the support of the Nichols lab. A special thanks to Maggie Roth for her involvement in developing some of the experiments. Thank you to my lab mates Dr. Klaske Schippers and fellow graduate student Jennyfer Mora for their support and friendship throughout my two years here.

Table of Contents

List of Tables

List of Figures

Introduction

The question of how animals first evolved and diversified is a long standing and controversial topic in evolutionary biology. It is generally accepted that life started as unicellular, but how animals evolved from unicellular life-forms to the diversity we see today is still under question. One of few undisputed facts about early animal evolution is that choanoflagellates are the closest living relatives of modern animals (Ruiz-Trillo et al. 2008). Choanoflagellates are single-celled and colony-forming marine eukaryotes characterized by an apical flagellum surrounded by a microvillar collar (Fig. i-1A; Dayel et al., 2011). Cells with this sort of morphology are typically referred to as collar cells. Collar cells have been reported in diverse animal lineages including cnidarians, echinoderms, and the pilidium larvae of a nemertine (Lyons 1973; Norrevang and Wingstrand 1970; Martinez et al. 1991; Cantell, Franzén, and Sensenbaugh 1982). Other instances of collar cell-like cells in animals are sensory cells of the bilaterian olfactory bulb and the hair cells of the middle and inner ear (Ludeman et al. 2014; Jacobs et al. 2007; Mayer et al. 2009). With distribution of collar cells among metazoan and choanoflagellate lineages, to the exclusion of other eukaryotes, we can infer that their last common ancestor also had cells with microvilli and motile cilia/flagella (King 2004; Sebé-Pedrós et al. 2013).

Some of the first clues about the evolutionary link between animals and choanoflagellates came from comparisons of choanoflagellate morphology to choanocytes, the feeding cells of sponges (Fig. i-1B; James-Clark, 1867, 1871; Kent, 1878). The sponge body plan stands out as an outlier among modern animals (Fig i-1B; Fig. i-2) – so much so that it was initially thought that sponges may be colonial protists (i.e., they may actually be choanoflagellates). They lack muscles, nerves, a gut, and consistent patterns of symmetry. They are composed of few cell types and simple tissues that function to pump water through an internal canal network where bacterial prey are filtered and directly phagocytosed by choanocytes, which collectively make up the feeding tissue known as the choanoderm. Nonetheless, it is now well established that sponges are indeed animals, albeit an early evolutionary branch of animals.

Sponge choanocytes bear a striking resemblance to choanoflagellates in form and function (Fig. i-1A, i-1B). Like choanoflagellates, sponge choanocytes have an apical flagellum surrounded by a microvillar collar. Both cell types use their flagella to manipulate water currents. Choanoflagellates draw water through the microvillar collar to capture bacterial prey, whereas sponge choanocytes use the microvillar collar to slow the flow of water so that prey can be phagocytosed directly (Mah, Christensen-Dalsgaard, and Leys 2014). Only recently have phylogenetic studies confirmed the relationship between choanoflagellates and animals, placing them as sister groups (Ruiz-Trillo et al. 2008). The hypothesized homology between choanoflagellates and sponge choanocytes was predictive of their phylogenetic connection with animals.

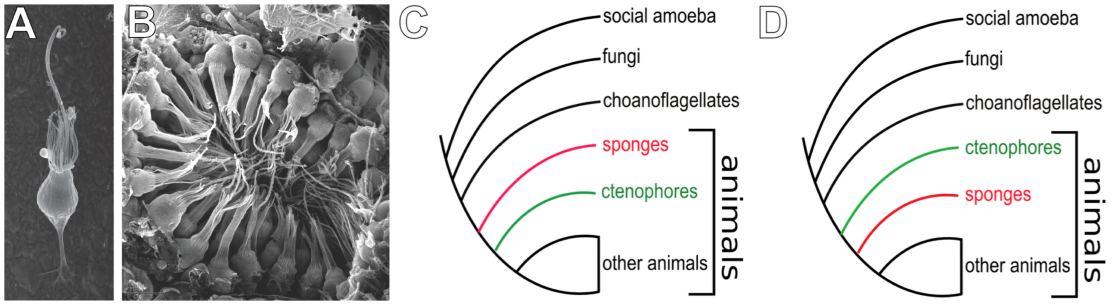

Figure i-1 Choanoflagellate, choanoderm, and phylogeny of animals. (A) Choanoflagellate with apical flagella surrounded by a microvillar collar (Dayel et al. 2011). (B) A sponge chooanocytes organized into a simple epithelium (choanoderm) that lines the water canal system. The traditional view is that sponges are the earliest branching animals (C) given the homology between choanocytes and choanoflagellates. Recent studies suggest that ctenophores (D), not sponges, are the earliest evolutionary branch of animals.

It has been hypothesized, and widely accepted, that because of the homology between choanoflagellates and sponge choanocytes, the sponge body plan may represent an early stage in animal evolution (Fig. i-1C; Collins, 1998; King, 2004; Maldonado, 2004; Medina, Collins, Silberman, & Sogin, 2001; Nielsen, 2008). Recently, this view has been challenged, citing subtle structural and functional differences between choanoflagellates and choanocytes (Maldonado 2004; Mah, Christensen-Dalsgaard, and Leys 2014; Dunn, Leys, and Haddock 2015). Another challenge to the evolutionary significance of sponge choanocytes and the antiquity of the sponge body plan is the idea that ctenophores rather than sponges are the earliest evolutionary branch of animals (Fig. i-1D).

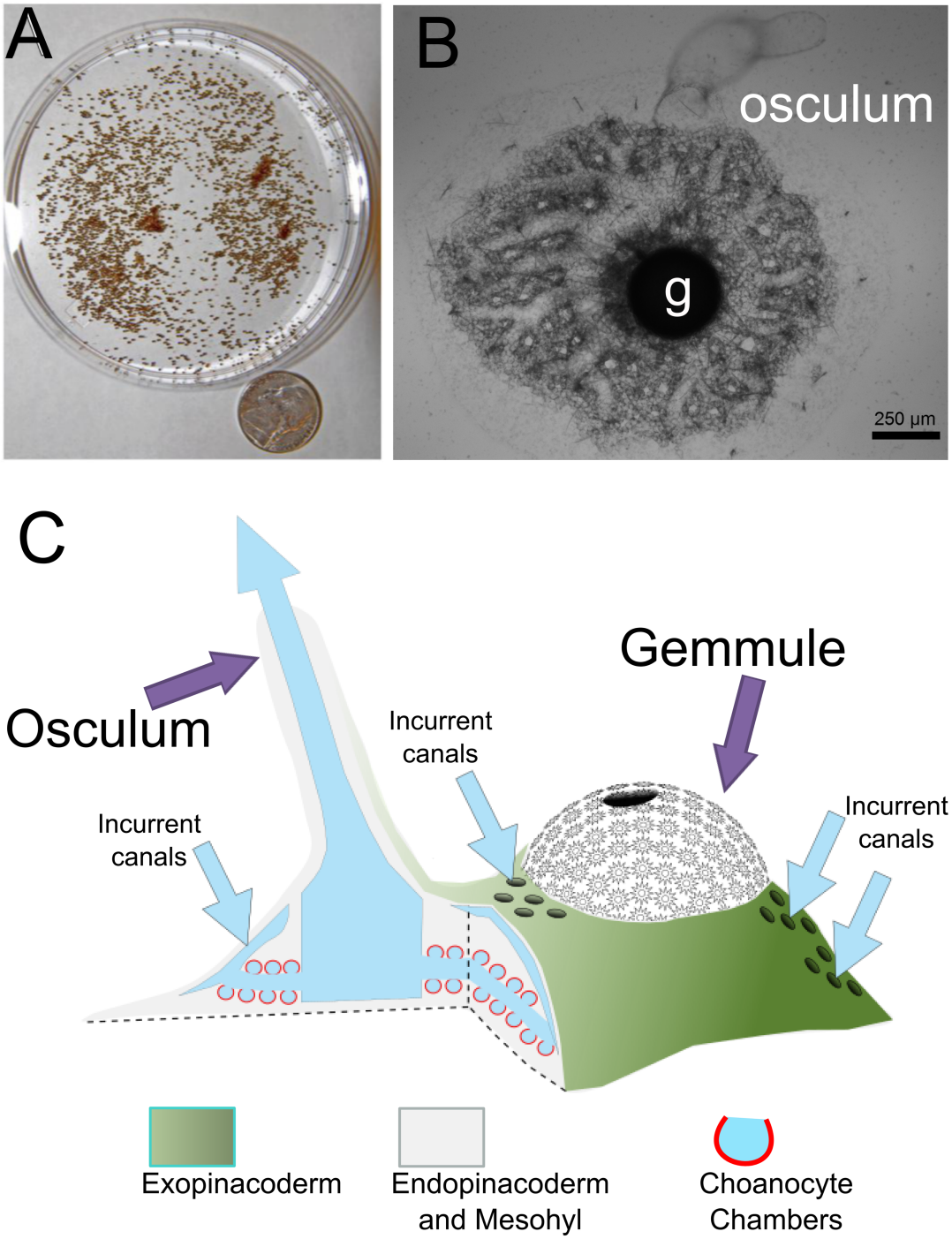

Figure i-2 Gemmules and the body plan of *Ephydatia muelleri*. (A) The sponge *E. muelleri* forms overwintering spores that can be stored at 4ºC. (B) Upon placement at room temperature the gemmules hatch and develop into a fully differentiated sponge; g=gemmule. (C) Schematic of a typical *E. muelleri* juvenile with a cross-section revealing internal canal system lined by choanocyte chambers.

This claim is supported by recent studies using genome-scale datasets (Moroz et al. 2014; Ryan et al. 2013; Dunn et al. 2008). Ctenophores have muscles, nerves and a gut, suggesting that these features are ancestral to all animals and that sponges are secondarily simplified rather than ancestrally simple.

To independently evaluate the evolutionary significance of cytological similarities between choanoflagellates and sponge choanocytes, this study explores the molecular basis of choanocyte structure and developmental induction of the sponge choanoderm. Ultimately this will inform our understanding of the sponge body plan in the context of animal evolution and how the choanoderm relates to more typical animal epithelia.

Chapter 1: Differential gene expression analysis of the choanoderm

Introduction

Despite its evolutionary significance, little is known about either the gene regulatory networks that control choanoderm development, or the adhesion and polarity genes that contribute to choanoderm organization and choanocyte structure. Previous studies have sought to identify choanocyte markers of freshwater sponges through proteomic approaches despite the difficult nature of isolating choanocyte chambers (Funayama, Nakatsukasa, Hayashi, et al. 2005; Funayama et al. 2010).

The present study focuses on the freshwater sponge *Ephydatia muelleri*. Like many sponges, *E. muelleri* produces an overwintering spore called a gemmule (Fig. i-2A). The gemmule houses thesocytes—dormant sponge stem cells—inside a spicule coat. When gemmules are placed at room temperature, thesocytes become active archeocytes which migrate out of the gemmule and differentiate into exopinacocytes (Rozenfeld & Rasmont 1977)—cells that line the external epithelia—and sclerocytes which produce spicules. The endopinacocytes begin to assemble the internal canal system (Rozenfeld & Rasmont 1977). The last cells to differentiate are the choanocytes, which line the canal system (Fig. i-2C; Funayama, Nakatsukasa, Hayashi, et al., 2005; Rozenfeld & Rasmont, 1977).

It has been previously reported that applying the drug hydroxyurea (HU) to developing sponges prevents cell differentiation and even canal development (Rozenfeld and Rasmont 1977). Hydroxyurea blocks DNA synthesis machinery and ultimately prevents cell division (Koç et al. 2004). We present here a method using HU to specifically prevent the development of choanocyte chambers. We compare choanocyte depleted sponges to normal sponges using RNA-sequencing and bioinformatics techniques to quantify and identify genes that are downregulated in the sponge choanoderm. Through analysis of the homology and domain architecture of these downregulated, putative choanoderm genes, we identify new gene candidates with possible functions in choanoderm development, structure and function. The long-term goal of this study is to provide a platform for analyzing the proposed homology of choanoflagellate and choanocyte cell structure, and for identifying elements of homology between sponge tissues and bilaterian epithelial tissues. This mechanistic approach will provide new lines of evidence, independent of phylogenetic and ultrastructural arguments, for interpreting the significance of the sponge body plan for our understanding of early animal evolution.

Methods

Living materials

Ephydatia muelleri gemmules were collected from Red Rock Lake, Colorado, USA (Em-CO); Beavertail Lake, Vancouver Island, British Columbia, Canada (Em-BTL); and Nanaimo River, Vancouver Island, British Columbia, Canada (Em-NR). The gemmules were stored in ultrapure milliQ® water, in the dark at 4°C.

Cultivation for hydroxyurea (HU) treatment

Gemmules were washed three times with ultrapure milliQ® water and placed on a coverslip in a petri dish containing 20 ml of autoclaved lake water at room temperature. Control and HU-treated gemmules were grown in 6-well culture plate format, with three biological replicates corresponding to Em-CO, Em-BTL, and Em-NR. Our preliminary studies indicated that early HU treatment prevents gemmule hatching or interferes with normal differentiation of tissues in addition to the choanoderm. In order to fine-tune the timing of HU addition until just before choanocyte differentiation, we established an "indicator" culture of gemmules 24 hours before starting our experimental cultures. When the indicator sponge developed choanocyte chambers, the experimental groups were then treated with hydroxyurea (100 μg/mL). The experimental groups were one day younger andhad not yet developed choanocyte chambers. Control sponges were untreated. Hydroxyurea was refreshed every day until sponges were harvested. The experiment is outlined in Table 1.1.

Harvesting and RNAseq

RNA was isolated from HU sponges and control sponges with TRIzol reagent (Invitrogen) following the manufacturer's protocol. Presence and quality of total RNA was confirmed by formaldehyde-agarose gel electrophoresis. The concentration of isolated RNA was measured using a nanodrop spectrophotometer. Samples were multiplexed and sequenced (single-end 100bp reads) in a single flow-cell lane by the Genomics and Microarray Core (University of Colorado Denver).

Mapping

Short single-end RNA-seq reads were trimmed using CLC Genomics Workbench 7.0.4 (Qiagen). The read files contained Phred scale quality scores which were used to trim sequence ends based on quality. The Q score is first converted to a base-calling error probability which is used to set the limit for which bases should be trimmed. Here a quality trim limit of 0.05 was used. For every base, the Workbench calculated the running sum of the value 0.05-P_{error}. If a sum dropped below 0, it is set to 0. Untrimmed regions of reads would end in the highest value of the running sum and start at the last zero value before the highest score; regions before and after this portion are trimmed. Additionally, ambiguous nucleotides were trimmed and discarded. Adapter sequences were also trimmed for each sample.

Trimmed sequences were mapped to a publicly available *de novo* assembled *Ephydatia muelleri* transcriptome (Hemmrich and Bosch 2008). The transcriptome was assembled with Trinity using default parameters except *–kmer_coc_2* as opposed to the default of 1. This helped reduce the noise of contigs. The option "One reference sequence per transcript" was selected. This option treats each sequence as a transcript and is often used with RNA-Seq data. The mapping parameters were set to default: mismatch cost=2; insertion cost=3; deletion cost=3; length fraction=0.8; similarity fraction=0.8. The maximum number of hits for a read was set to 10. Hits that match to multiple distinct places are randomly assigned to one of those places based on the number of unique matches that the gene already has. If a read matched to more than 10 distinct places, it was discarded. Strand specificity was set to Both, Forward, and Reverse; a higher proportion of reads mapped with higher unique specificity when strand specificity was set to Reverse, so these mappings were used for Corset analyses.

Corset analysis

In an attempt to improve the accuracy of read counts for differential gene expression (DGE) using a reference transcriptome, we used the program Corset (version 1.03) that is designed to cluster RNA transcripts that presumably derive from a single genomic DNA locus (Davidson and Oshlack 2014). Mapped reads were analyzed with experimental groups identified (-g option) to improve the power it has when splitting

differentially expressed paralogues. Corset analysis was done on the University of Denver High Performance Cluster.

Differential gene expression analysis

The Corset output was processed using edgeR, a bioconductor package in R (Robinson, McCarthy, and Smyth 2010; McCarthy, Chen, and Smyth 2012; Robinson and Smyth 2008; Zhou, Lindsay, and Robinson 2014; Robinson and Smyth 2007). Statistical testing was performed for differences between control group RNAseq data and HU treated group RNAseq data. The cluster-level count data was converted to an edgeR object. First, a group variable was created to direct edgeR to separate samples by group (control vs. HU-treated). Using the function *DGEList*(), supplied with group variable and the cluster-level count data, creates the edgeR object.

Once converted, edgeR is used to calculate normalization factors based on the trimmed mean of M-values normalization method. TMM normalization can effectively estimate relative RNA production levels from RNA-seq data and can estimate scale factors between samples that can be incorporated into downstream statistical methods for differential expression. This normalization corrects for the different compositions of the samples and generates effective library sizes. A multidimensional scaling plot was generated from normalized samples to measure sample similarity in two dimensions.

Count data obtained from RNA-seq experiments is analyzed using negative binomial models due to higher variation in data. The mean of counts for each gene

corresponds to the abundance of that gene in the RNA sample. EdgeR models the mean of a gene as the library size multiplied by concentration. The dispersion parameter determines how the variance of each gene is modelled. The first dispersion to be calculated was the common dispersion. Under the common dispersion model, each gene is assigned the same value for dispersion when modelling its variance. The next dispersion that was calculated was the tagwise dispersion. Under the tagwise dispersion model, each gene gets assigned a unique dispersion estimate. Following this step, normalized counts were obtained in order to generate histograms to display differential expression of specific clusters.

The *exactTest*() function was executed on the edgeR object to perform pair-wise tests for differential expression between the two groups. The function *topTags*() takes the output from *exactTest*() and adjusts the raw p-values using the false discovery rate correction and returned the top differentially expressed genes.

BLAST2GO analysis

Sequences associated with downregulated clusters were extracted using a Python script. Downregulated cluster IDs were obtained from edgeR. Clusters and associated sequences were obtained as a Corset output. The python script compared these two files and wrote a new file containing the cluster ID and the name of the associated sequence. A biopython script was used to extract whole sequences from the *E. muelleri* transcriptome. The input for this extraction was the transcriptome as a fasta file and a

text file with the names of downregulated sequences. The output was a fasta file containing downregulated sequences of interest. These extracted sequences were analyzed by a BLAST search against the *E. muelleri* predicted proteome using CLC Genomics Workbench. The top hits from the protein BLAST were extracted with the same biopython script this time using an *E. muelleri* reference proteome as the input fasta. The top hit protein sequences were analyzed by a BLAST search against nr protein database using CLC Genomics Workbench.

The blastp result was converted to a Blast2GO project using the Blast2GO plug-in for CLC Genomics Workbench (Conesa et al. 2005; Conesa and Götz 2008; Götz et al. 2008; Götz et al. 2011). Gene ontology terms associated with blast hits were retrieved by executing mapping function through the BLAST2GO plug-in. The mapping step links all BLAST hits to functional information stored in the Gene Ontology database, where each GO term is associated with an evidence code. Gene ontology annotations are all associated with evidence codes which indicate how the annotation is supported; that is, evidence codes link GO terms to previous work and analyses done on a particular gene product which support the GO assignment.

The annotation function was used to assign GO terms from the GO pool generated by the mapping step to the query sequence. Annotation applies an annotation rule on the ontology terms in the pool. This rule searches for the most specific annotations with a certain level of reliability. An annotation score is computed for each GO term obtained from the mapping step. The annotation score takes into account two terms: direct and

abstraction. The direct term represents the highest hit similarity of a GO term weighted by a factor corresponding to its evidence code. The abstraction term provides the possibility of abstraction. This term multiplies the number of total GOs unified at the node by a GO weight factor that controls the possibility and strength. The annotation rule then selects the lowest term per branch that lies over an annotation cut-off. Annotation was limited to GO terms obtained from hits with an e-value less than 1.0E-8. Additionally, to determine what level of abstraction allowed for more informative annotations, the annotation cut-off was varied (55, 30, and 20) while the GO-weight was set to 5. Hsp-Hit Coverage CutOff and EC-weight were left to default settings. Following annotation, InterProScan was executed to retrieve domain and motif information (Jones et al. 2014). The GO terms obtained by IPScan were transferred to the sequences and merged with already existing GO terms. Sequences were then sorted by e-value. A list of downregulated genes was manually generated with an e-value cut-off of 1.0e-10.

Identification of genes restricted to choanoflagellates and sponges

Downregulated protein sequences which were obtained from the dataset as previously described. A phmmer search was performed on all the sequences. Query sequences with less than 1000 hits were examined closely to determine if the protein was restricted to choanoflagellates and sponges.

Results

In support of a previous study by Rasmont and Rozenfeld (1977), our results confirm that hydroxyurea—an inhibitor of DNA replication (Koç et al. 2004)—can be applied during *E. muelleri* germination to inhibit differentiation of choanocytes. Gemmulation and cell differentiation in sponges follows a predictable sequence. The first cells to exit the gemmule are the archeocytes, a putative sponge stem cell. These cells then differentiate into the other cell types of the sponge, with choanocytes being the smallest, and therefore the last to differentiate. Rasmont and Rozenfeld (1977) reported more widespread effects on development such as the lack of a canal system and osculum. Their treated sponges developed as a hollow dome with archeocytes lining the floor of it (Rozenfeld and Rasmont 1977). In our studies, we found that if hydroxyurea (HU) is applied within 12-24 hours prior to choanocyte differentiation (Table 1.1), that these HU-treated sponges develop all of the detectable features of the untreated control sponges, except that they lack choanocytes. For example, spicules are found in both HU treated and untreated sponges, as is an organized canal system and an osculum (Fig. 1.1). When ink is added to the water, it is the choanocyte chambers which first catch and concentrate the ink, clearing the water almost completely. This produces a distinct pattern in untreated sponges (Fig. 1.1A',A"), but in the HU treated sponges, we do not see the same pattern (Fig. 1.1B'-B").

We reasoned that we could exploit the effects of HU on *E. muelleri* development to identify genes that are normally expressed in choanocytes. To be conservative, we

performed these experiments on biological replicates from each of the three geographically isolated populations: Red Rock Lake, Colorado (Em-CO); Beavertail Lake, Canada (Em-BTL); and Nanaimo River, Canada (Em-NR). The relative similarity of expression between samples based on the 500 most heterogeneous genes is greater in terms of locality; that is, sponges clustered based on geographic origin (Fig. 1.4). There was a less pronounced separation by treatment condition. This suggests that between geographically isolated populations there is already a difference in relative gene expression levels. Polymorphism between populations combined with the fact that all reads were mapped to the Colorado *E. muelleri* transcriptome could also influence this clustering. Essentially, not all the reads from the Nanaimo River sponges and the Beavertail Lake sponges mapped to the transcriptome. This is supported by the fact that Em-CO, whether control (Em-CO-C; Table 1.2; Fig. 1.2) or HU-treated (Em-CO-HU; Table 1.3; Fig. 1.3) had the least amount of its fragments uncounted; that is, a smaller percentage of reads were discarded during the mapping step. Potentially, a more pronounced clustering based on treatment would be observed if sponges all came from the same population. Nevertheless, the extent of variation exhibited between sponges from different locations underscores that the genes that were detected as significantly downregulated in all samples are likely to be biologically meaningful. Using a false discovery rate of 1%, we found a total of 879 transcripts (corresponding to ~1% of the Corset clustered transcriptome; Fig. 1.5; Table 1.4) to be significantly downregulated in hydroxyurea treated sponges.

To classify and evaluate the biological categories of the downregulated transcripts, we used BLAST2GO (Conesa and Götz 2008; Conesa et al. 2005; Götz et al. 2008; Götz et al. 2011). This approach relies on performing a BLAST search of all the candidate sequences. Following the BLAST search, BLAST2GO will map the hits to their gene ontologies using the gene ontology database. This step generates a pool of gene ontologies that could potentially be assigned to the query sequence. It is worth noting that not all queries are mapped because hits may not be associated with gene ontologies in the GO database; so at this step some sequences will be excluded from downstream annotation. The annotation step assigns specific GO terms from the pool generated in the mapping step. It has previously been reported that the default parameters of the annotation step are the recommended settings, but more permissive parameters can be set when the sequence similarities as reported by BLAST2GO are low (Götz et al. 2008). Even under permissive parameters, not all sequences were annotated (~21%). We augmented the annotation by performing an InterProScan which compares protein sequences and identifies domains and functional sites in order to functionally characterize the new sequence (Jones et al. 2014). Once this was done we proceeded to do manual curation by looking at the BLAST hits and comparing them to the GO terms and predicted domain architectures. In some cases we also used phmmer to verify our comparisons (http://hmmer.org/). Downregulated transcript clusters were ultimately associated with gene products, summarized in Table 1.6.

Normalized counts were obtained for genes associated with microvillar and ciliary structures (Fig. 1.6; Fig. 1.7) as well as for genes implicated in the classical cadherin catenin adhesion complex (Fig. 1.8). By comparing normalized counts for these genes between control and hydroxyurea-treated sponges, we see that some of these components are significantly downregulated while others are not, consistent with edgeR differential gene expression analyses.

Additionally, 24 transcript clusters were identified as restricted to choanoflagellates and sponges through phmmer searches (Table 1.5). A number of these genes have signal peptides as well as other domains implicated in signaling. One gene (m.236078) has a cadherin domain. Of the 24 genes identified, 23 are downregulated in choanocyte depleted sponges (Fig. 1.9). There is one gene (m.6183) which is upregulated in choanocyte deficient sponges (Fig. 1.9). The upregulated gene has a signal peptide as well as a transmembrane region (Table 1.5).

Table 1.1 Hydroxyurea treatment plan

	Day 0	Day 1	Day 2	Day 3	…	Day 5	…	Day 8
Indicator	Plated		Hatched		…	Choanocytes detected	…	
Control		Plated		Hatched	…		…	Harvested
Hydroxyurea		Plated		Hatched	…	HU added	…	Harvested

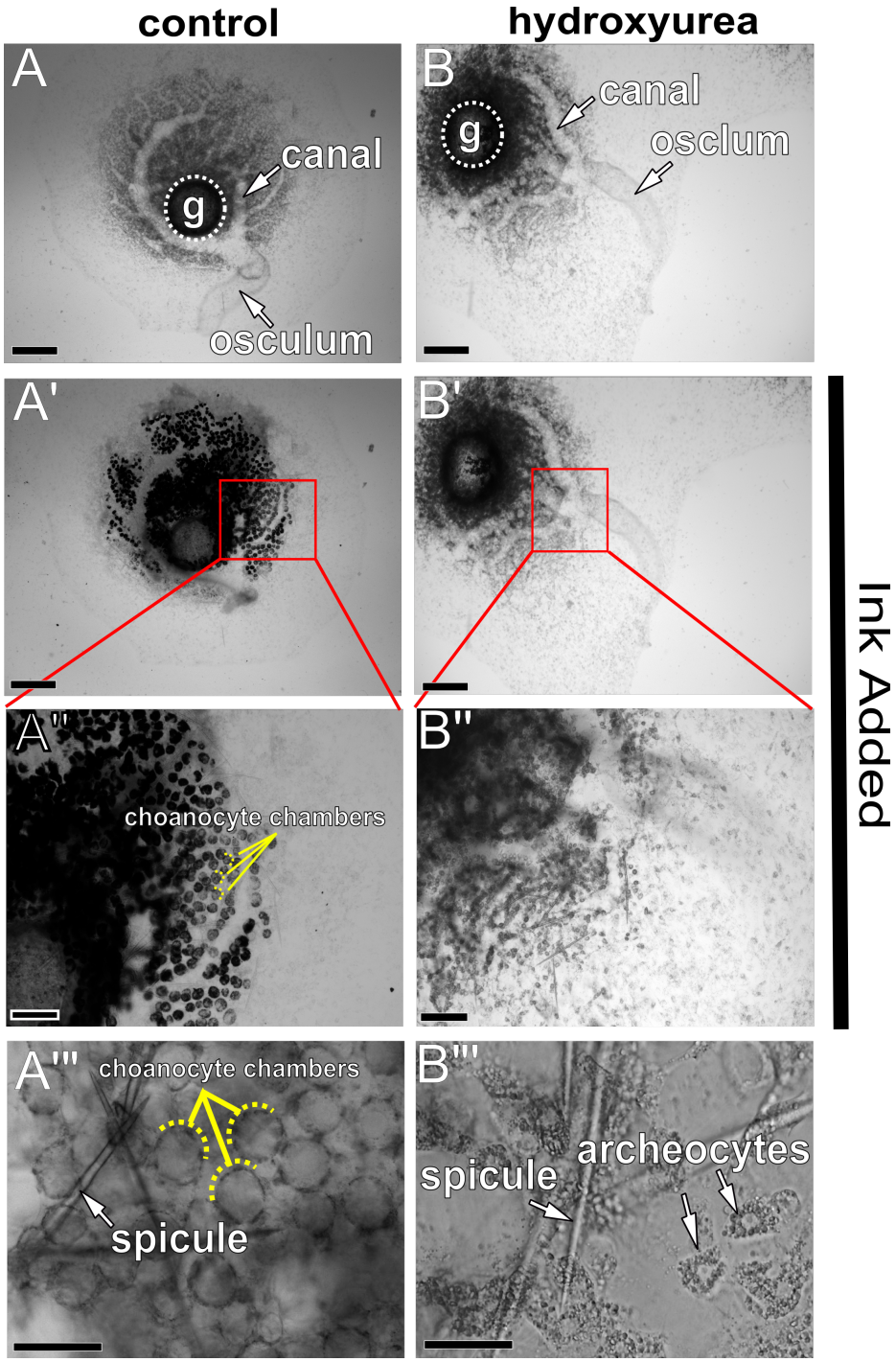

Figure 1.1 Untreated and hydroxyurea treated sponge morphology. Control (A) and hydroxyurea treated (B) sponges develop water canals and an osculum. Ink shows choanocyte chambers in control sponges (A', A") but not in HU-treated sponges (B', B"). Both control (A"') and HU-treated (B"') sponges develop spicules and archeocytes. Scale bars: 250 µm (A-B'), 100 µm (A", B"), 25 µm (A"', B"').

Table 1.2 CLC Genomics Workbench read mapping summary for control samples			
Sample	Em-BTL-C	Em-NR-C	Em-CO-C
Counted fragments	26,827,639 (85.19%)	21,117,814 (90.77%)	29,351,889 (93.03%)
Unique fragments	22,462,085 (71.33%)	17,633,865 (75.79%)	24,275,678 (76.94%)
Non-specifically	4,365,554 (13.86%)	3,483,949 (14.97%)	5,077,211 (16.09%)
Uncounted fragments	4,663,682 (14.81%)	2,147,804 (9.23%)	2,200,099 (6.97%)
Total fragments	31,491,321 (100%)	23,265,618 (100%)	31,552,988 (100%)

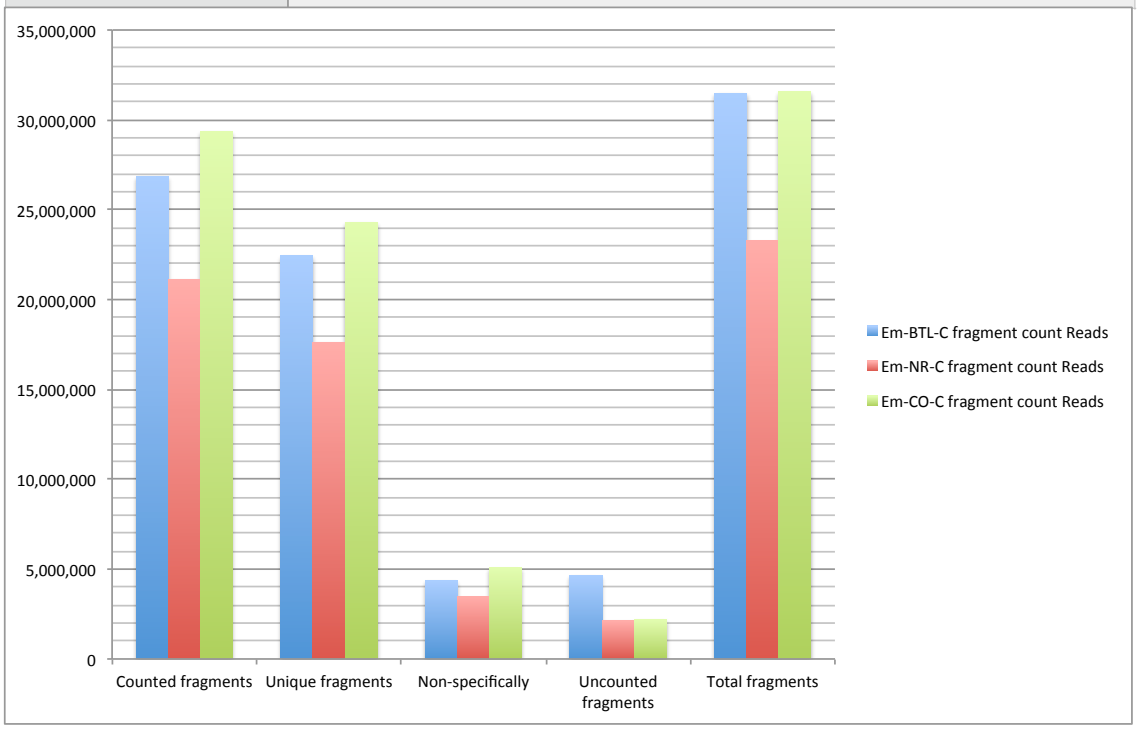

Figure 1.2 Read mapping summary of control samples.

Table 1.3 CLC Genomics Workbench read mapping summary for HU-treated samples			
Sample	Em-BTL-HU	Em-NR-HU	Em-CO-HU
Counted fragments	26,182,932 (84.33%)	25,290,543 (91.87%)	27,162,188 (93.75%)
Unique fragments	22,438,826 (72.27%)	21,329,957 (77.48%)	22,686,820 (78.31%)
Non-specifically	3,744,106 (12.06%)	3,960,586 (14.39%)	4,475,368 (15.45%)
Uncounted fragments	4,865,864 (15.67%)	2,237,935 (8.13%)	1,809,488 (6.25%)
Total fragments	31,048,796 (100%)	27,528,478 (100%)	28,971,676 (100%)

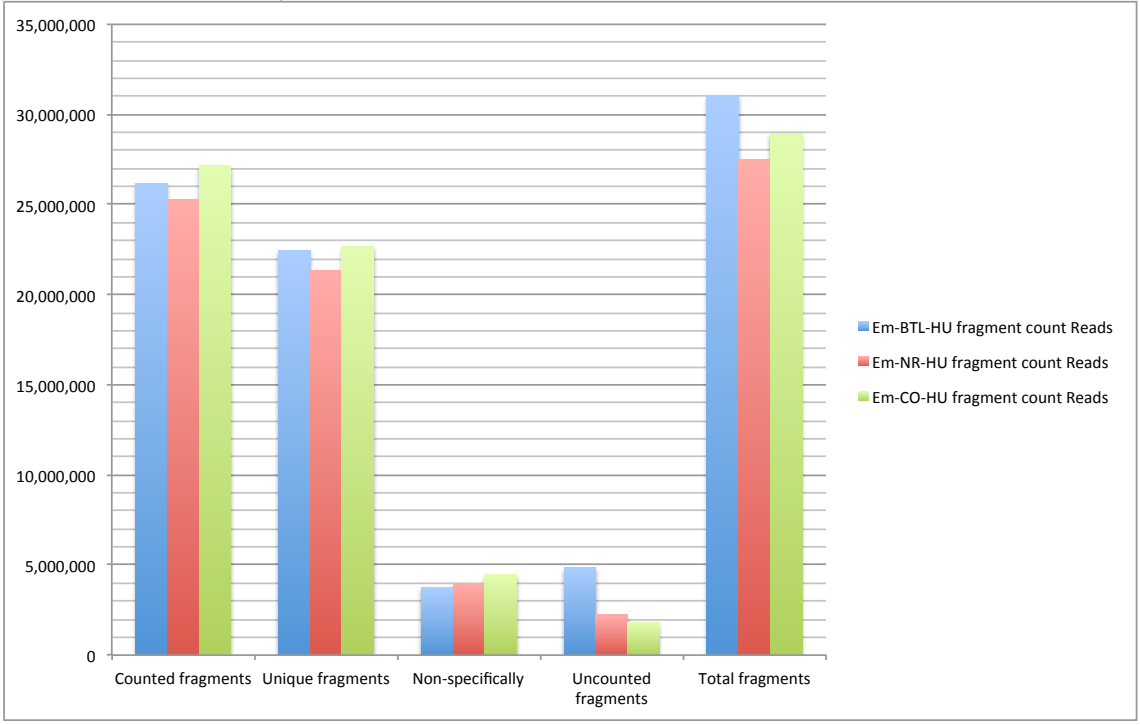

Figure 1.3 Read mapping summary of hydroxyurea samples.

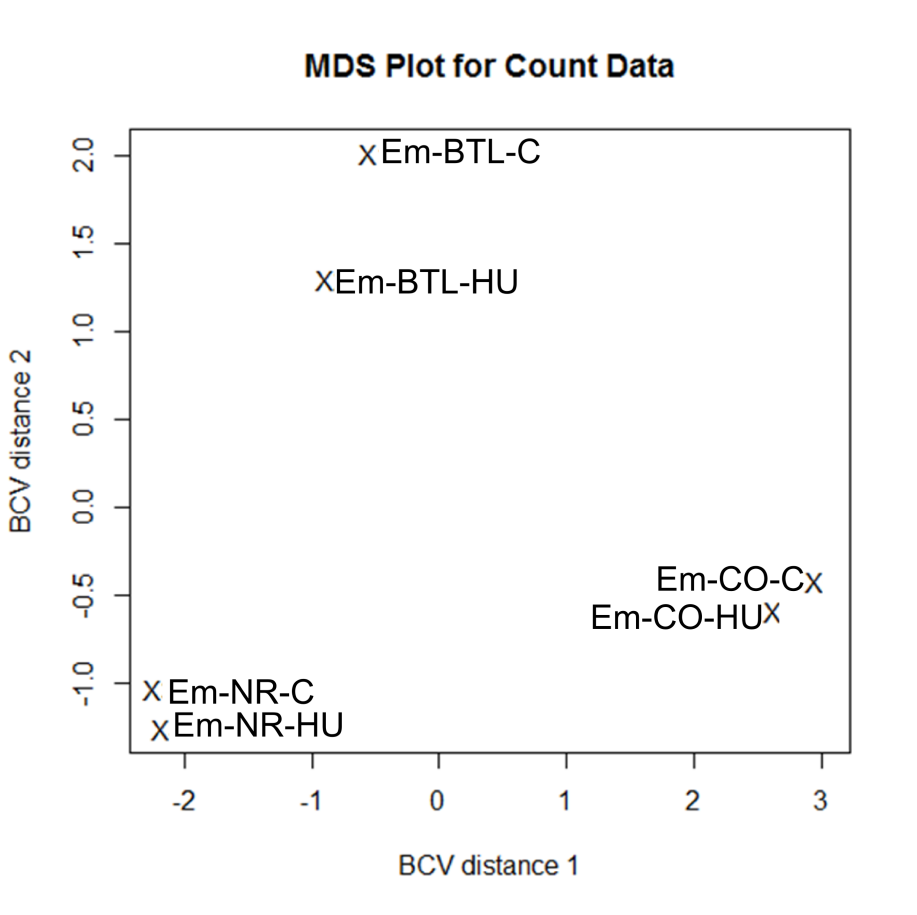

Figure 1.4 The multi-dimensional scaling plot projects sample similarity in 2-dimensions. Em-CO: Red Rock Lake population; Em-BTL: Beavertail Lake population; Em-NR: Nanaimo River population. C: control; HU: hydroxyurea treated

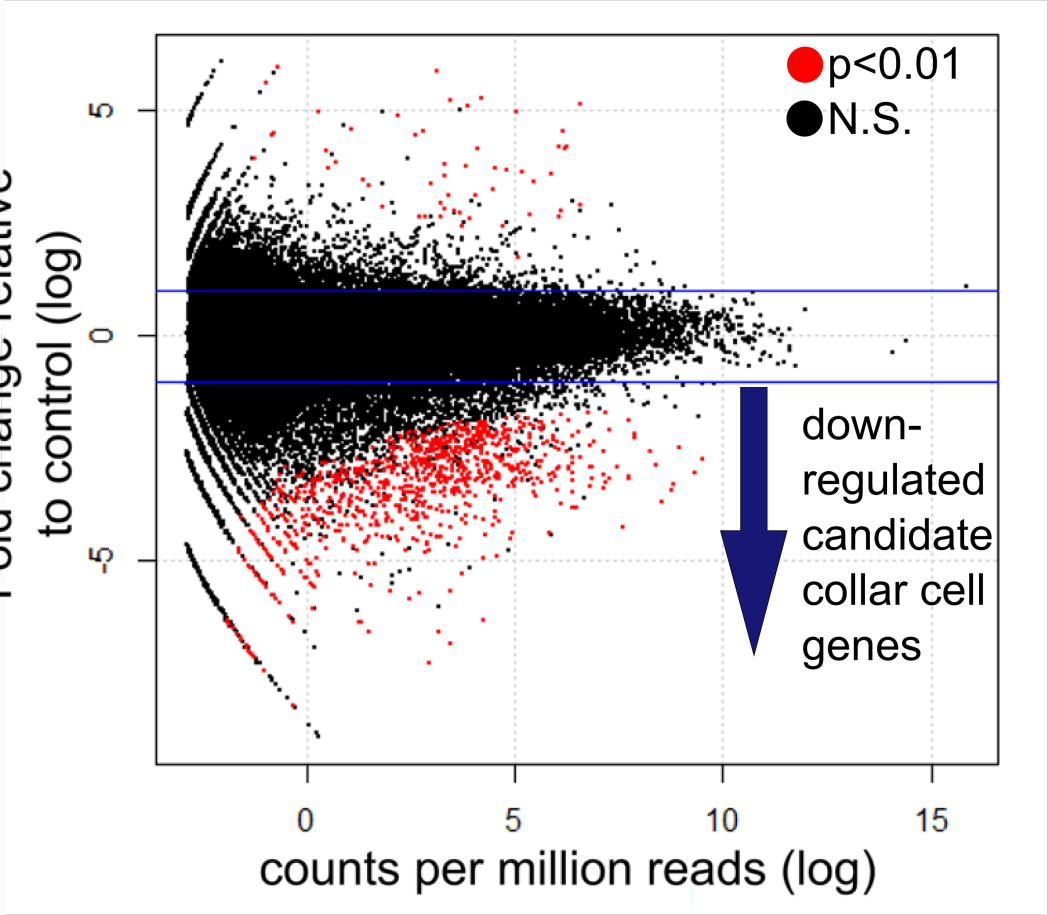

FIGURE 1.5 Microarray plot for differentially expressed gene clusters. The microarray plot shows the relationship between counts per million reads and fold-change across the genes clusters. Differentially expressed gene clusters are shown in red (P<0.01). Non-differentially expressed genes are in black. The blue lines denote biological significance.

Table 1.4 Differentially expressed clusters at different false discovery rates

	FDR<1%	FDR<0.1%
Upregulated clusters	54	37
Downregulated clusters	879	481
NS	77,018	77,433

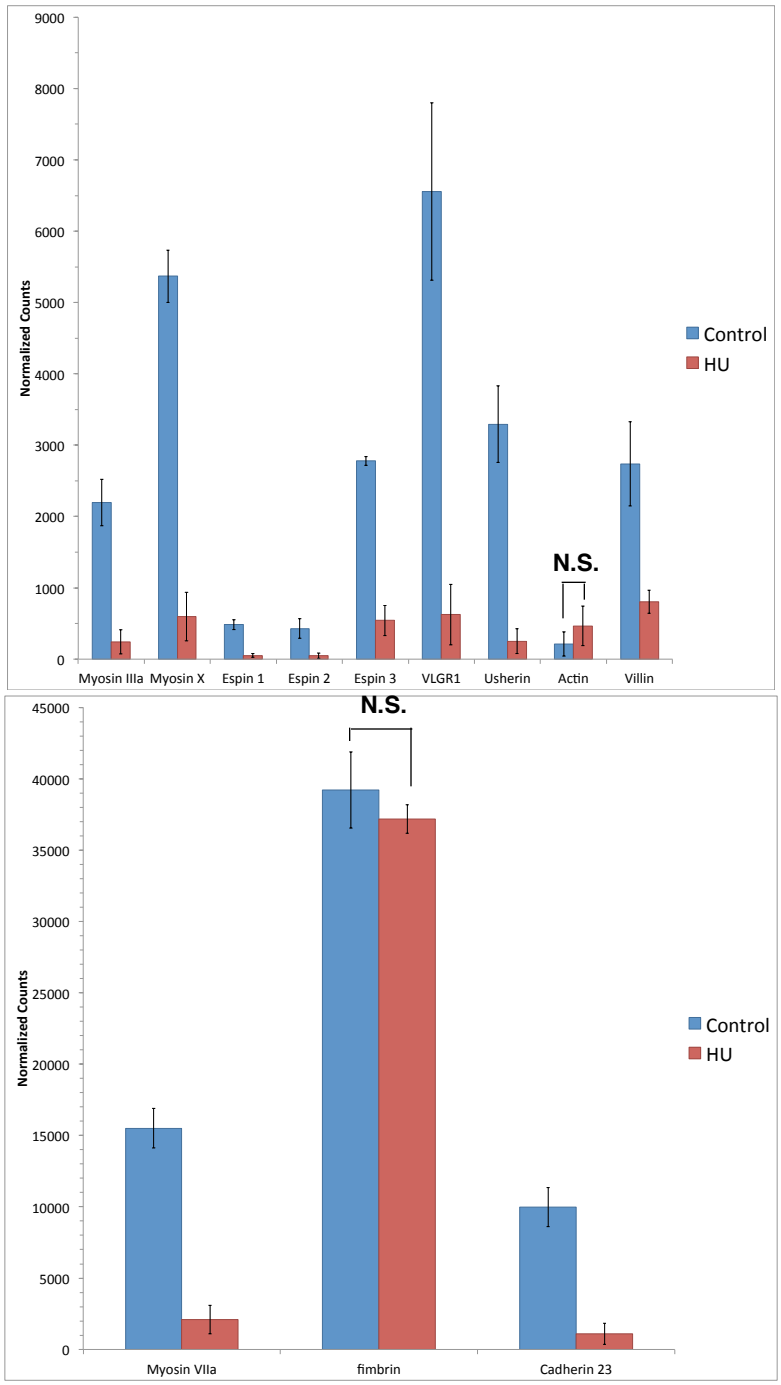

Figure 1.6 Mean normalized counts of microvillar genes in control and hydroxyurea treated sponges. Except where stated, all genes are significantly differentially expressed in HU-treated sponges. Error bars represent standard deviation from the mean.

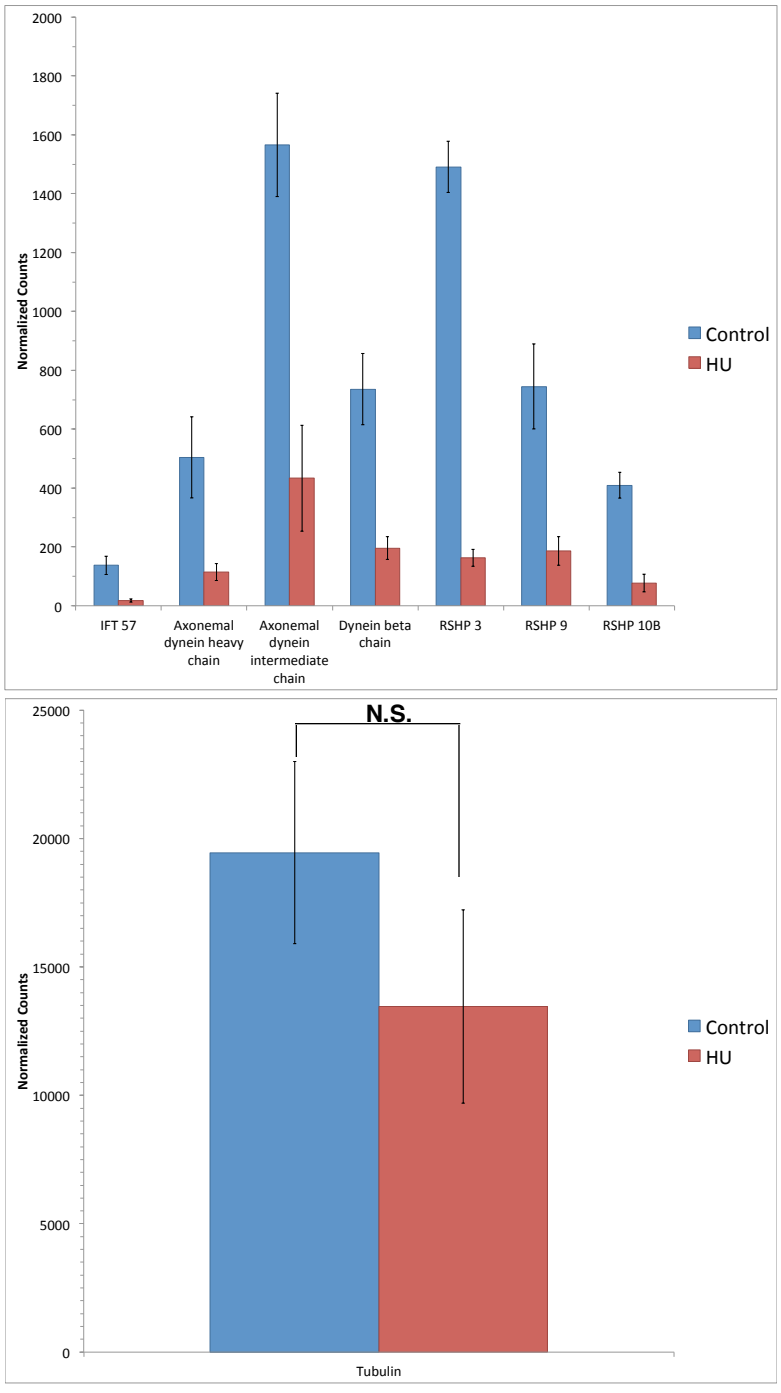

Figure 1.7 Mean normalized counts of ciliary genes in control and hydroxyurea treated sponges. Except where stated, all genes are significantly differentially expressed in HU-treated sponges. Error bars represent standard deviation from the mean.

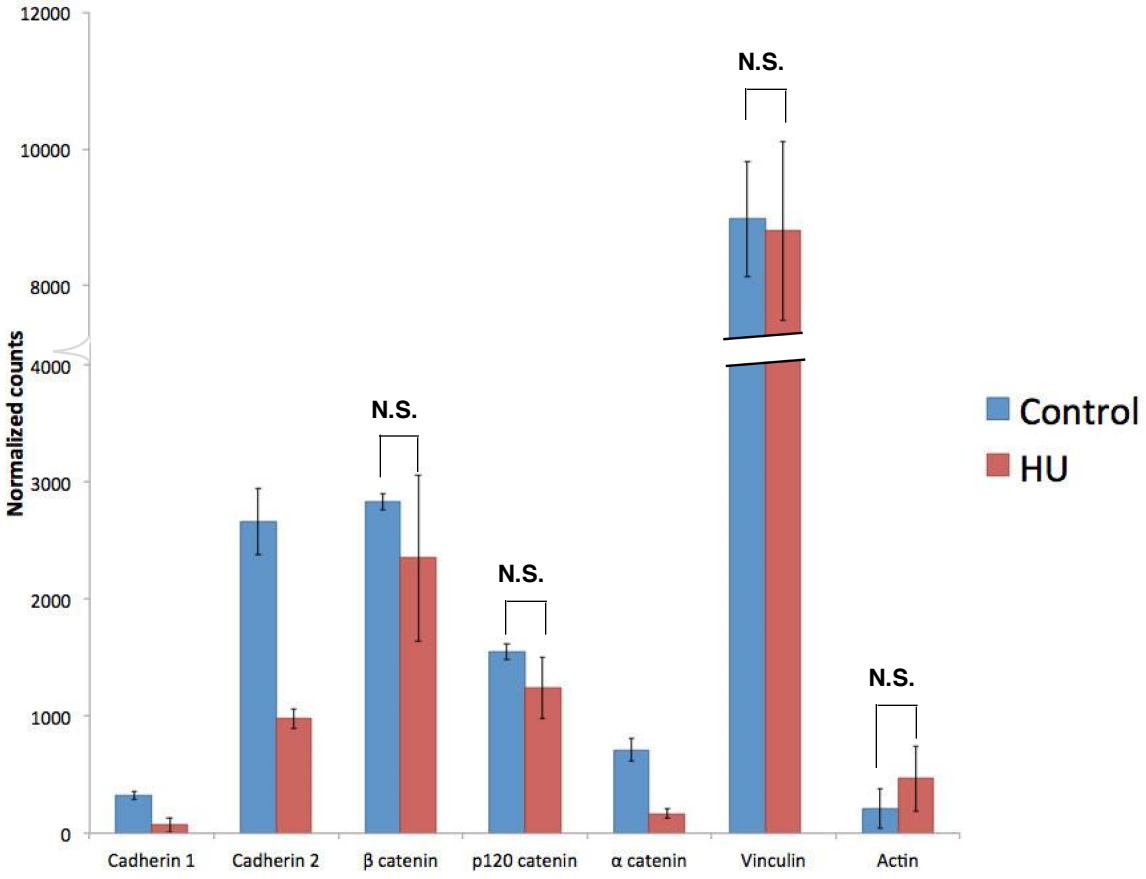

Figure 1.8 Mean normalized counts of genes associated with the classical cadherin catenin adhesion complex in control and hydroxyurea treated sponges. Except where stated, all genes are significantly differentially expressed in HU-treated sponges. Error bars represent standard deviation from the mean.

Table 1.5 Gene products restricted to sponges and choanoflagellates		
Accession	**Length**	**Domain architecture**
m.19803	491	EF hand
m.164262	252	Transmembrane x2
m.170264	536	Concavalin A-like lectin/glucanase
m.232397	4940	None
m.232402	995	IPT/TIG x3, Calx-beta
m.232409	1076	None
m.244211	3379	Signal peptide, IPT/TIG x11, transmembrane
m.277222	5907	transmembrane
m.292216	842	None
m.9857	217	Signal peptide
m.19803	517	Signal peptide, Receptor L, transmembrane
m.29736	491	EF hand
m.156173	523	PH, coiled-coil
m.206651	1051	Coiled-coil x6, SAM, Ras association, PDZ
m.236078	1028	Cadherin
m.8558	856	PH, costar x2, LIM
m.192346	400	Beta-1 integrin binding
m.41140	713	PDZ
m.254520	348	Signal peptide, transmembrane
m.45730	153	None
m.72689	425	PH, PTB
m.175777	410	Coiled-coil
m.178438	854	Sfi1 spindle body protein
m.6183	559	Signal peptide, transmembrane

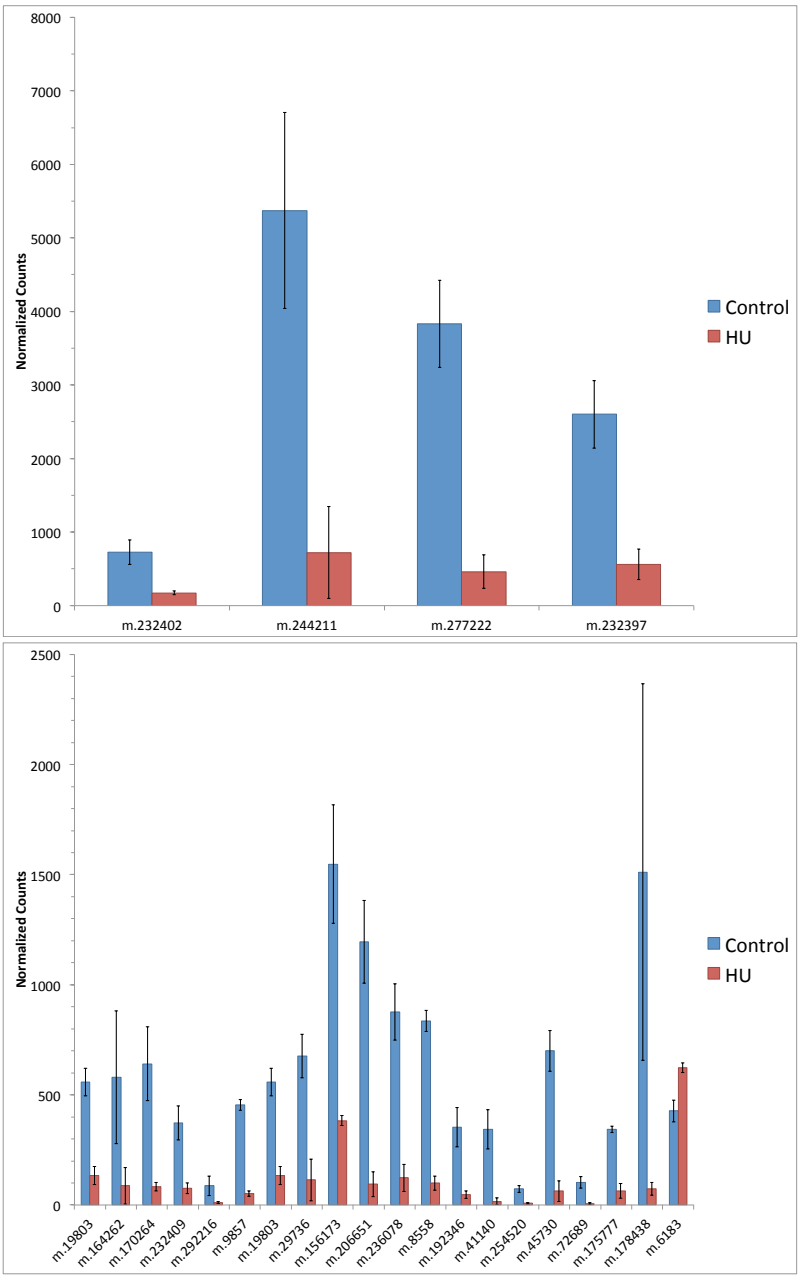

Figure 1.9 Mean normalized counts of genes identified only in sponges and choanoflagellates. The counts correspond to control and hydroxyurea treated sponges. Accession IDs correspond to the predicted proteome of *E. muelleri*. All genes are differentially expressed in hydroxyurea treated sponges. Error bars represent standard deviation from the mean.

Table 1.6 Summary of downregulated candidate genes identified by BLAST2GO (e-value< 1.0e-10)	
cilia/flagella-specific	IFT-52/57/81/105/172, CCDC-39/40/113, RSPH-3/9/10B, NME5, tectorin, PITG-05447, ODF-3, axonemal dynein, Cep89, PTPRQ
other microtubule associated	GAS-8, Serine/Threonine-protein kinase Nek8, TTC-16/30A/28, Kif-15, protein polyglycylase TTL 10, kinesin-like, Futsch, battenin, CCDC-87, KIF25, kif3b, kif5b, alp11
microvillar specific	VLGR1, spectrin, usherin, espin, whirlin, calmodulin, SLC9A3R1
other actin associated	Myosin-VIIa, myosin IIIa, myosin-X, filamin-a/b/c, SHANK3, socius, FGD6, INF2, ankyrin, WASp, EPB-41, zeta-sarcoglycan, SWAP70, inverted-formin, talin
cadherin/catenin related	Protocadherin Fat-4, Lefftyrin, coherin, alpha-catenin, cadherin 23
other adhesion related	Kifc3, fibropellin, PTPRS, laminin-beta, attractin-like 1, fibrillin, adhesin-like, hemicentin-1/2, ephrin type-b receptor 1, invadolysin, dystrophin, hedgling, ELFN3, C-type lectin, latrophilin, delphilin, HSPG2, FRMPD2, sponge aggregation factor 3, contactin, tetraspanin-5, MEGF11, zonadhesin , integrin-beta, VW Factor A, connexin 32
polarity (apical/basal & planar)	Prickle 2, crumbs, flamingo, RTK-like orphan receptor, alp11, DLG1
epithelial development	Hensin, Cfap-57, Plexin-A2, fibrocystin-L, NOTCH1, inversin, TGM1, TBATA, TBX2b, KRIT1, MIB-1, malcavernin
epithelial/mesenchyme transition	HUNK, EGF-like, invadolysin, krueppel-like, SCUBE2, zinc finger transcription factor
known epithelial expression	Anoctamin, Krit-1 trichohyalin-like
Cell proliferation/ differentiation	NOX-5, cohesin, jagged-1/2, CCDC-135, NOTCH1, Protein polybromo, EGF-like, TOE1, NME5, EGR-1, MELK, tetraspanin-5, Delta

Discussion

In general, manual curation allowed us to be more confident in some of the classifications of downregulated transcripts. Something that emerged from these data was a strong enrichment of microvillar/ciliary genes (Table 1.6). All eukaryotic cilia and flagella are cored by doublet microtubules arranged in a circle and in motile cilia there are an additional two singlet microtubules in the center (Kleene and Van Houten 2014; Mayer et al. 2009). Dynein, kinesins, and radial spoke head proteins are all associated with microtubules in motile cilia/flagella (Mohri et al. 2012; Paradela et al. 2005), all of which are in our data set (Fig. 1.7; Table 1.6). Alongside this, we see proteins that are implicated in the structure of microvilli (Fig 1.6; Table 1.6). In contrast to cilia/flagella, microvilli are non-motile and restricted to the holozoan lineage. Much like cilia/flagella, these are projections from the cell body but are cored by actin microfilament bundles (Gonobobleva and Maldonado 2009; Sebé-Pedrós et al. 2013; Brown and McKnight 2010). While we do not find actin to be significantly downregulated (Fig. 1.6), we see an enrichment in actin associated genes (Fig. 1.6; Table 1.6). Genes involved with cross-linking actin filaments of microvilli, for example *espin*, are significantly downregulated (Fig. 1.6). This is consistent with previous studies on the nature of microvilli and filopodia in holozoans (Sebé-Pedrós et al. 2013). Nucleation promoting factors such as WASP as well as unconventional myosins are also seen in our data set and are consistent with previous analyses of filopodial and microvillar structures (Sebé-Pedrós et al. 2013). The presence of these ciliary/flagellar- and microvillar-associated genes in the dataset

supports the idea that the general approach was successful, since choanocytes are characterized by a microvillar collar and apical flagellum. Taken together, it can be inferred that downregulated genes reflect genes normally expressed in choanocytes.

In addition to structural proteins of cilia and microvilli, we have evidence for downregulation of proteins specifically associated with mechanosensory hair cells. The idea that choanocytes may serve a mechanosensory function is not new given their structural similarities to hair cells (Jacobs et al. 2007). Mechanosensory hair cells and their associated proteins have been found in conserved contexts such as cnidarian cnidocytes, fish lateral line, and mammalian cochlea (Jacobs et al. 2007; Suli et al. 2012; Michel et al. 2005; McGee et al. 2006). Cnidocytes are characterized by a cyst and tubule which inverts upon deflection of the cnidocil. These structures can be used for prey capture as well as adhesion to a substrate (Anderson and Bouchard 2009). The zebrafish lateral line has mechanosensory hair cells which help the fish align itself with water currents (Suli et al. 2012). The hair cells of the cochlea are involved in auditory signaling. Sound induces vibrations on the basilar membrane which mechanically stimulates hair cells to send an electrical signal that is processed as sound (Schwander, Kachar, and Müller 2010). The structures responsible for mechanosensation in these cells are actin-based stereocilia and tubulin based kinocilia/cnidocil (Anderson and Bouchard 2009; Schwander, Kachar, and Müller 2010; Suli et al. 2012). During hair cell development, proteins that act as lateral links between stereocilia are transiently expressed. These proteins include usherin, VLGR1, and cadherin 23 (McGee et al. 2006;

Michel et al. 2005; Schwander, Kachar, and Müller 2010). Our data set shows downregulation of VLGR1, usherin, and cadherin 23 (Fig. 1.6; Table 1.6). This is particularly interesting when considering that lateral links have been reported in the microvilli of choanoflagellates and sponge choanocytes (Mah, Christensen-Dalsgaard, and Leys 2014; Dayel and King 2014). Mice that lack VLGR1 lose organization of the stereocilia and are profoundly deaf (McGee et al. 2006). The role VLGR1 is playing is not one directly involved in signal transduction but in developing the morphology of the stereocilia. In the context of choanocytes or choanoflagellates, VLGR1 may be playing a role in maintaining microvillar structure. Unlike VLGR1, cadherin 23 is transiently expressed during hair cell development as well as in the mature hair cell (McGee et al. 2006; Michel et al. 2005). During development, cadherin 23 aids in maintaining the structure of the stereocilia and it progressively becomes restricted to the top portions, eventually becoming part of the tip link complexes where it is involved in mechanosensory transduction (Michel et al. 2005; Selvakumar, Drescher, and Drescher 2013). One possible mechanosensory role of choanocytes could be in signaling for the initial events of the sponge 'sneeze' reaction. It has been suggested that ciliated cells in the osculum can sense changes in water flow (Ludeman et al. 2014). Though they lack innervation, signal from the osculum could be propagated through calcium waves which would be detected by the Calx-beta domains of VLGR1, potentially regulating flagellar beating. Another role for mechanosensation in the choanoderm is to regulate the orientation of choanocyte chambers to maintain the directionality of flow.

Another category of genes we are interested in are adhesion genes. Sponge tissues generally lack features of bilaterian epithelia such as electron dense cell-cell junction (Leys, Nichols, and Adams 2009). There is evidence for the presence of classical cadherin catenin adhesion (CCA) complex proteins and that sponge cadherin 1 binds with β-catenin in a yeast two-hybrid screen (Nichols et al. 2012). Here we have also identified α-catenin and other adhesion molecules as a downregulated (Fig. 1.8; Table 1.6). However, β-catenin is not downregulated in hydroxyurea treated sponges (Fig. 1.8), suggesting that β-catenin is not significantly more expressed in the choanoderm than in other sponge tissues. This opens the possibility that the choanoderm could be the only known animal tissue that uses CCA complex-independent methods for adhesion.

Adhesion mechanisms like the CCA complex are common features of bilaterian epithelia. Another characteristic of animal epithelia is cell polarity (Tyler 2003). The gene *crumbs* has been implicated in regulating apical/basal polarity in metazoans (Chen et al. 2010). From an ultrastructural perspective, the choanoderm has apical/basal polarity, and consistent with this observation is the fact that crumbs and other apical/basal genes are downregulated. Unlike apical/basal polarity, planar polarity in the choanoderm is less obvious, yet the gene prickle 2, which in *Drosophila* is a core planar cell polarity gene (Mrkusich, Flanagan, and Whitington 2011), is down regulated. In fact, from an ultrastructural perspective, we do not see planar polarity in any sponge tissue. A case for planar polarity in sponges can be made if we consider the orientation of cells relative to

the flow of water. Because the planar polarity proteins are downregulated in the choanoderm, we can hypothesize that the mechanosensory role of the choanoderm in regulating the orientation of cells relative to the flow of water is linked to expression of genes like *prickle*. Other epithelial developmental genes that are downregulated have roles in duct/lumen architecture of bilaterian organs. An example is fibrocystin, which is involved in maintaining duct/lumen architecture in kidneys (Zhang et al. 2004). Another example is hensin, which can induce epithelial polarity (Watanabe et al. 2005). Since these genes are downregulated, we might consider the sponge aquiferous system analogous to duct-lumen structures in bilaterians.

We conducted a phmmer search with our downregulated genes in an attempt to identify genes restricted to sponges and choanoflagellates and identified 24 genes and characterized domain architecture where possible (Table 1.5). As previously mentioned, microvillar links have been reported in sponges and choanoflagellates. The uncharacterized genes are strong candidates for being microvillar links. Many of the genes have signal peptides which could support the idea that signaling from the collar regulates phagocytosis.

The choanoderm has long been considered to be an ancient animal tissue due to the structural similarities of choanoflagellates and choanocytes. Here, we examined gene expression in the sponge choanoderm by comparing RNA-sequencing data of sponges that develop normally and those that develop without a choanoderm. Our results are consistent with the idea that the sponge choanoderm ultrastructurally divergent from

other animal epithelia but is still under control of common developmental mechanisms and is evolutionarily homologous to other animal epithelia. Structurally, choanocytes are homologous to collar cells throughout metazoans from mechanosensory hair cells to enterocytes in the gut of bilaterians. The choanoderm seems to represent a transitional tissue in metazoans since it shares character traits with choanoflagellates and metazoans. Confirming that the genes reported here are actually expressed in choanocytes is a step towards elucidating their function and teasing apart their evolutionary history. In the next chapter I will discuss *in situ* hybridization as a technique for validating choanocyte candidate gene expression. Further studies to characterize the genes expressed in choanocytes should include detailed descriptions of domain architecture as well as functional studies such as the pharmacologic inhibition of particular genes.

Chapter 2: Optimizing whole-mount *in situ* hybridization for *Ephydatia muelleri* tissue

Introduction

Having identified candidate choanocyte genes as described in the previous chapter, the next step was to experimentally validate expression dynamics. A technique commonly used in developmental biology is whole-mount *in situ* hybridization (WISH), which allows visualization of gene expression dynamics in different tissues and throughout development. In the variation of WISH used in this study, digoxigenin-labelled RNA probes are used. The digoxigenin tag is targeted by an antibody conjugated with an alkaline phosphatase (Jin and Lloyd 1997). This conjugated enzyme will react with a combination of NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) which results in the deposition of an insoluble black-purple precipitate (Pearson et al. 2009). A reliable WISH protocol provides sensitive and accurate detection of gene expression without compromising morphology.

The technique has been used to detect expression of transcription factors in the marine demosponge *Amphimedon queenslandica* as well as cell markers in the freshwater demosponge *Ephydatia fluviatilis* (Larroux et al. 2006; Funayama, Nakatsukasa, Hayashi, et al. 2005; Funayama, Nakatsukasa, Kuraku, et al. 2005; Funayama et al. 2010). In this study, the focus was to optimize tissue permeabilization, probe specificity, signal

detection, and post-detection treatments. To accomplish this, three protocols were tested and modified. The first probe used was *annexin* as this was the target of previous WISH of freshwater sponges (Funayama, Nakatsukasa, Hayashi, et al. 2005). Other probes were designed to target two genes which were identified in our differential gene expression analysis of the choanoderm. A fourth probe was designed against a gene which we determined to be restricted to sponges and choanoflagellates, and downregulated in choanocyte depleted sponges. The goal of this project was to develop a robust and sensitive *in situ* hybridization protocol for *E. muelleri* tissue to characterize gene expression dynamics.

Methods

Candidate sequence identification

Three genes from the candidate list generated from the Blast2GO analyses were selected. The three genes were VLGR1, usherin, and cadherin 23. A fourth, uncharacterized gene restricted to choanoflagellates and sponges was also included ("choanogene"). The protein sequences that Blast2GO associated with these genes were Blasted against the *E. muelleri* transcriptome (Hemmrich and Bosch 2008). Once identified, the nucleic acid sequences were translated using ExPASy translate to find the appropriate reading frame (Gasteiger et al., 2003). The translated sequences were then run through phmmer (http://hmmer.org/) to confirm their identities.

Candidate gene amplification, cloning, and transformation

Gene specific primers were designed to amplify a 700-1000 bp region near the 3' end of each sequence. Primers were designed with Web Primer, and the best pair was selected (Table 2.1). A previously constructed *E. muelleri* cDNA library was used as the starting template. Quality of amplification was confirmed with agarose gel electrophoresis.

The PCR products of VLGR1, Cadherin 23, and choanogene were cloned into pCR II vector with Dual Promoter TA Cloning Kit (Life Technologies) following the manufacturer's instructions. Dh5-alpha cells were transformed with the construct and plated on LB agar in the presence of kanamycin and ampicillin.

Several colonies from each plate were selected for colony screen PCR with M13 primers. Following the PCR, the presence of the insert was confirmed by agarose gel electrophoresis. For each gene, a successfully transformed colony was picked and grown in a liquid culture overnight. A plasmid miniprep was performed for each overnight culture with QIAPrep® Spin Miniprep Kit (Qiagen). The isolated plasmids were sent to the DNA Sequencing and Analysis Core (University of Colorado Denver) for sequencing.

The orientation and presence of the gene insert relative to the promoters in the vector was determined by analyzing the chromatogram obtained from sequencing with CLC Genomic Workbench 7.0.4 (Qiagen**)**. The original colony was grown overnight in a larger volume to perform a plasmid midiprep with NucleoBond ® Xtra Midi (Macherey-Nagel).

In situ RNA hybridization probe synthesis

The pCR II TOPO vector has an SP6 promoter and a T7 promoter flanking the region of the insert. To synthesize anti-sense dig-labelled RNA probes, plasmids were digested with EcoRV (New England Biolabs). Synthesis of sense dig-labelled RNA probes required a restriction digest with HindIII-HF (New England Biolabs). Digesting with two different restriction enzymes allows for in vitro transcription with one of the two promoters. Here, transcription with SP6 polymerase gave rise to antisense RNA probes whereas transcription with T7 polymerase gave rise to sense RNA probes.

Restriction digests were carried out overnight at 37ºC. Quality of the digest was assessed with agarose gel electrophoresis.

Digested plasmids were phenol/chloroform extracted twice then chloroform extracted once. The digested plasmids were precipitated with ethanol and sodium acetate. In vitro transcription of dig-labelled RNA probes was done with SP6 polymerase for antisense and T7 polymerase for sense. Dig-labelled RNA probes were precipitated with 2.5 volumes of 100% absolute ethanol and 1/10th volume of 3 M sodium acetate (pH 5.2) at -20ºC overnight then resuspended in RNAse free water. The quality of RNA probes was assessed with agarose gel electrophoresis.

In some cases, probes were hydrolyzed to a length of 200 bp with 0.2 M bicarbonate buffer at pH 10 at 65ºC for 35 minutes. Hydrolysis was stopped with 0.2% glacial acetic acid, 40 mM sodium acetate, and 1 µg/µl glycogen with 440 µl of 100% ethanol. This mix was stored in -20ºC overnight. The RNA precipitate was centrifuged and the pellet was washed and resuspended in RNAse free water.

Additionally, probes were tested by setting up a dilution series for each and crosslinking samples from each dilution to polyamide membrane (GE Healthcare). Crosslinking was done using UV Stratalinker 1800 (Stratagene) on the autocrosslink setting. The membranes were washed three times in 2x saline sodium citrate (20X SSC: 3M NaCl, 0.3 M sodium citrate; pH 7.0) before being incubated in 1:2000 AP-coupled anti-DIG Fab fragments. The membranes were washed in PBS four times before being equilibrated with alkaline phosphatase buffer D (100 mM Tris pH 9.5, 50 mM $MgCl_2$,

150 mM NaCl, 0.1% Tween 20). Following equilibration, the AP buffer D was replaced with AP buffer D containing 160 µl of NBT/BCIP solution. They were left to develop and then rinsed with tap water.

Protocol 2: Whole-mount *in situ* RNA hybridization for *Drosophila* embryo

The protocol for *Drosophila* embryos from Draizen, Ewer, and Robinow (1999) was modified as described here.

Cultivation and fixation. Gemmules were grown in autoclaved lake water and 100 µg/ml ampicillin in 24-well plate format. After hatching, sponges were cultured for 3 days then fixed in 4% paraformaldehyde / 0.05% glutaraldehyde in phosphate buffered saline (1X PBS: 8.0g/L NaCl, 0.2g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4; pH 7.4) overnight at 4ºC. The sponges were washed four times in PBS.

Hybridization. Following the final wash step, PBS was replaced by a 1:1 mixture of PBS and hybridization buffer B (50% formamide, 5x SSC); this was left for 10 minutes at room temperature. The PBS:hybridization buffer B mix was replaced with hybridization buffer B and incubated for 10 minutes at room temperature. The hybridization buffer B was replaced with hybridization buffer (Cold Spring Harbor recipe: 1x Denhardt's solution, 5mM EDTA pH 8, 50% formamide, 5x SSC, 100 µg/ml heparin, 100 µg/ml yeast tRNA, 0.1% Tween 20) and left to incubate for 1 hour at 55ºC. Probes were added to hybridization buffer (1:500) and denatured by heating for 5 minutes at 80ºC. Probes were either hydrolyzed or whole. Hybridization buffer was removed

from sponges and replaced with 1:500 probe:hybridization buffer mix. Hybridization was left overnight at 55ºC.

Antibody incubation. The probe:hybridization buffer mix was removed and sponges were washed 6 times in hybridization buffer B at 55ºC. The hybridization buffer B was replaced with 1:1 PBS:hybridization buffer B for 20 minutes at room temperature. Following the incubation, the samples were washed 3 times in PBS. The samples were incubated in 2% (w/v) blocking reagent (Roche) in Tween-20/maleic acid buffer (100 mM maleic acid, 150 mM NaCl, 0.1% tween-20) at room temperature for 1 hour. The blocking solution was replaced with 1:2000 AP-coupled anti-DIG Fab fragments in 2% blocking reagent. This was left overnight at 4ºC.

Detection. The samples were washed in PBS 4 times then equilibrated with AP buffer D. Following equilibration, the AP buffer D was replaced with development buffer D (AP buffer D,7 µl NBT/ 13 µl BCIP / ml). Staining was done in the dark and monitored until purple precipitate was observed in antisense groups. The samples were rinsed in PBS then fixed in 4% paraformaldehyde at room temperature for 15 minutes. After fixation, samples were rinsed in PBS. The PBS was replaced with 70% glycerol.

Protocol 1: Whole-mount *in situ* RNA hybridization for *Ephydatia*

This in situ hybridization protocol from Funayama et al. (2005) was modified as described here.

Cultivation and fixation. Gemmules were grown on Hybri-Slips (Sigma-Aldrich) in a petri dish with autoclaved lake water and 100 µg/ml ampicillin. After hatching,

sponges were cultured for 3 days then transferred to 24-well plates. Once transferred they were cultured for one day after which they were fixed in 4% paraformaldehyde / 0.05% glutaraldehyde in PBS overnight at 4ºC. Sponges were washed 3 times in PBS. Optimization for this step involved washing once in PBS with 0.5% triton-x 100.

Permeabilization and acetylation. After washes, sponges were treated with 7.5 µg/ml Proteinase K at 37ºC for 10 minutes. The reaction was stopped with 2 mg/ml glycine. Glycine was replaced with 0.1 M triethanolamine. Triethanolamine was removed and replaced with 0.1 M triethanolamine in 1.5 µl/ml acetic anhydride. The mix was replaced with 0.1 M triethanolamine in 3 µl/ml acetic anhydride. Acetylation steps were meant to inactivate endogenous RNases. The mix was removed and sponges were fixed in 4% paraformaldehyde / 0.05% glutaraldehyde in PBS for 1 hour at room temperature. Permiabilization with Proteinase K and acetylation were omitted in optimizations of this protocol.

Hybridization. The sponges were washed in PBS 5 times. After washes, prehybridization was carried out in hybridization buffer overnight at 51ºC. Hybridization buffer was replaced with new hybridization buffer containing 0.2 ng/µl denatured probe (hydrolyzed or unhydrolyzed). The hybridization step was left overnight at 51ºC.

Antibody incubation. After hybridization, the probe and hybridization buffer were replaced with pre-warmed hybridization buffer. This was left for 10 minutes at 51ºC. The sponges were washed twice for ten minutes at 51ºC in 50% formamide/4x SSC/0.1% Tween-20. They were then washed twice at 51ºC for 10 minutes in 25% formamide/2x

SSC/0.1% Tween-20. Finally, three 15 minute washes with 2x SSC/0.1% Tween-20 were done at room temperature. Blocking was then done for 1 hour at room temperature with 2% blocking reagent (Roche) in tween-20/maleic acid buffer. After the incubation, the blocking solution was replaced with 1:5000 ap-coupled anti-digoxigenin Fab fragments in 2% blocking reagent/tween-20/maleic acid buffer and was left overnight at 4ºC.

Detection. The samples were washed 6 times for 30 minutes in maleic acid buffer. The sponges were then equilibrated with alkaline phosphatase buffer E (100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris pH 9.5, 0.1% Tween 20, 1 mM levamisole). Following equilibration, the AP buffer E was replaced with development buffer E (AP buffer E, 7 µl NBT/ 13 µl BCIP / ml). The staining reaction proceeded in the dark until sponges became a dark purple. The reaction was stopped by washing with PBS. Sponges were mounted in 100% glycerol.

Ethanol washes. In an optimization trial, after stopping the development reaction, the PBS was replaced with 100% ethanol for 60 minutes at room temperature. The ethanol was then replaced with 50% ethanol and left for 10 minutes at room temperature. Sponges were mounted in 100% glycerol.

Protocol 3: Whole-mount *in situ* RNA hybridization for planarians

Cultivation and fixation. We modified a whole mount in situ hybridization for planarians by Pearson et al. (2009). Gemmules were grown on Hybri-Slips (Sigma-Aldrich) in a petri dish with autoclaved lake water and 100 µg/ml ampicillin. After

hatching, sponges were cultured for 3 days then transferred to 24-well plates. Once transferred they were cultured for one day after which they were fixed in 4% paraformaldehyde / 0.05% glutaraldehyde in PBS for 15 minutes at room temperature. The fixative was removed and the sponges were rinsed with PBS/0.3% triton-x 100.

Reduction and hybridization. PBS/0.3% triton-x 100 was replaced with reduction solution (50 mM DTT, 1% Tween-20, 0.5% SDS, in PBS). Reduction was carried out in a 37ºC water bath for 5 minutes with intermittent agitation. In planarians, the reduction step aids in permeabilization to improve probe penetration. The samples were rinsed with PBS then incubated in a 1:1 PBS and hybridization buffer mix for 10 minutes at room temperature. The mix was replaced with hybridization buffer and left for 2 hours at 55ºC. After prehybridization, the hybridization buffer was replaced with hybridization buffer containing 0.2 ng/µl unhydrolyzed denatured probe. The hybridization reaction was carried out over night at 55ºC.

Antibody incubation. After hybridization, the probe/hybridization buffer mix was removed. Samples were washed with a 1:1 hybridization buffer and 2x SSC + 0.1% Tween-20 mix. This was done twice for 30 minutes at 55C. The samples were then washed twice for 30 minutes at 55C with 2x SSCC + 0.1% Tween 20. The final two 30 minute washes were done with 0.2x SSC + 0.1% Tween 20. The samples were returned to room temperature and washed twice for 10 minutes in maleic acid buffer + 0.1% tween 20. After washing, the solution was replaced with 2% blocking reagent in maleic acid buffer + tween 20 and kept at 4ºC overnight. The blocking reagent was removed and

replaced with 1:5000 AP-coupled anti-digoxigenin Fab fragments in 2% blocking reagent/maleic acid buffer + tween 20 and was left overnight at 4 C.

Detection. The antibody solution was removed and the samples were rinsed with maleic acid buffer + tween 20, 7 times, 20 minutes each. The tissue was equilibrated with alkaline phosphatase buffer P (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween 20, brought up tp volume with 10% polyvinyl alcohol solution). After equilibrating for 10 minutes at room temperature AP buffer P was replaced with development buffer P (AP buffer P, 4.5 µl/ml NBT, 3.5 µl/ml BCIP). Development was carried out in the dark until samples developed the purple precipitate.

Development was stopped by replacing development buffer P with PBS. Post-fixation was done at room temperature for 10 minutes using 4% paraformaldehyde. Afterwards, samples were rinsed with PBS. The PBS was replaced with 100% ethanol for 20 minutes at room temperature. The ethanol was then replaced with 50% ethanol and left for 5 minutes at room temperature. The samples were rinsed with PBS then mounted in glycerol mounting media (80% glycerol, 10 mM Tris pH 7.4, 1 mM EDTA).

Results

Probe synthesis

Due to the nature of library construction, the templates for *in situ* probes were amplified from a region near the 3' end of the sequence (Fig. 2.1-5). The candidate genes were identified from our Blast2GO analyses. VLGR1 has been implicated in the development of auditory hair bundles, where it acts as a transient ankle link (McGee et al. 2006). Usherin and cadherin 23 are also transiently expressed in developing hair cells (Schwander, Kachar, and Müller 2010; Michel et al. 2005). The fourth candidate, "choanogene", has been identified as a gene restricted to choanoflagellates and sponges (Table 1.5: m.244211). Amplified regions were of similar size (Table 2.1). Of the four candidate genes, usherin was the only one to not be amplified by PCR (Fig. 2.6). The amplicons were cloned into pCR®II-TOPO vector with their 3' ends oriented towards the SP6 promoter. (Fig. 2.7). DH5-alpha cells were successfully transformed with the plasmid carrying one of the three inserts (Fig. 2.8). Plasmids were digested prior to the *in vitro* transcription reaction (Fig. 2.9). Compared to the control, which shows three bands in the lane, the digested plasmids show only one band. Antisense and sense probe synthesis was verified by gel electrophoresis (Fig. 2.10A). For each probe, there is only a single band. Probes appeared on dot blots with intensity of dot directly related to probe concentration (Fig. 2.10B).

In situ hybridization

In following a previously described sponge protocol, the acetylation step destroyed the majority of tissue. When that step was removed, proteinase K destroyed the tissue. A protocol used on *Drosophila* embryos was used. The annexin sense probe did not produce a staining pattern (Fig. 2.11A). Ubiquitous staining was seen with the annexin antisense probe (Fig. 2.11B). Annexin expression has been previously reported in choanocyte chambers and archeocytes (Funayama, Nakatsukasa, Hayashi, et al. 2005).

Rather than optimize the *Drosophila* protocol, we returned to the sponge protocol and repeated it without the acetylation or proteinase K steps, but with hydrolyzed probes. While tissue integrity was maintained, no staining was seen (data not shown). We switched back to full length probes. Again, omitting acetylation and proteinase K treatment improved tissue integrity (Fig. 2.12). When post-treated with ethanol, background staining was significantly reduced. Regardless of the probe used, spicule staining was observed. The sense probe showed faint choanoderm staining. The antisense probes for VLGR1 and choanogene showed staining of choanocyte chambers (Fig. 2.12B, D). The cadherin 23 antisense probe showed strong staining of the pinacoderm (Fig. 2.12C).

A third protocol was also tested. This protocol was originally developed for planarians and includes a reduction step which aids in probe penetration. Another difference was that the development buffer was made with polyvinyl alcohol. Proteinase K treatment was omitted as before. Unlike the sponge protocol, the sense probe shows

little to no staining (Fig. 2.13A), although in a later trial sense probe staining was significant (Fig. 2.14A). There is very strong signal of VLGR1 and choanogene in the choanoderm (Fig. 2.13B, D). Like the sponge protocol, cadherin 23 signal is seen in the pinacoderm but not the choanoderm (Fig. 2.13C). Pinacocyte cell boundaries can also be seen.

To improve signal to noise using the planarian protocol, post-hybridization treatment with RNase A was done. The choanogene sense and antisense probes were used. The sense probe had faint signal in the choanoderm (Fig. 2.14A). Choanogene signal in the choanoderm was very strong (Fig. 2.14B). Post-hybridization, tissue was treated with different concentrations of RNase A. At the lowest concentration, choanogene antisense signal was much fainter than the sense signal (Fig. 2.14B). The signal appears to get fainter with increasing concentration of RNase A (Fig. 2.14).

Table 2.1 List of primers and amplicon size				
Gene	Amplicon size (bp)	Primer name	Orientation	Primer sequence
VLGR1	884	SN288F	Forward	TCACTGACATCTATGGCCTCA
		SN288R	Reverse	TTCTGACACGAGAGATGCCTT
Usherin	885	SN289F	Forward	TGTCGTTGTCCCGGCGTT
		SN289R	Reverse	TGCGTGATGTCGGGTGTG
Cadherin 23	894	SN290F	Forward	TGGCACATATCCATCTCTGTC
		SN290R	Reverse	GAAGCACTGGCAGATTGCTTT
Choanogene	897	SN291F	Forward	ATTCCAGAGGACAAGCCAGTA
		SN290R	Reverse	TGCCTTAACATCTTTGTCCG
Annexin	974	SN264F	Forward	GGTGGTCACGGAACTGTCAA
		SN264R	Reverse	TTAGTTGGGACCAACAATGGC

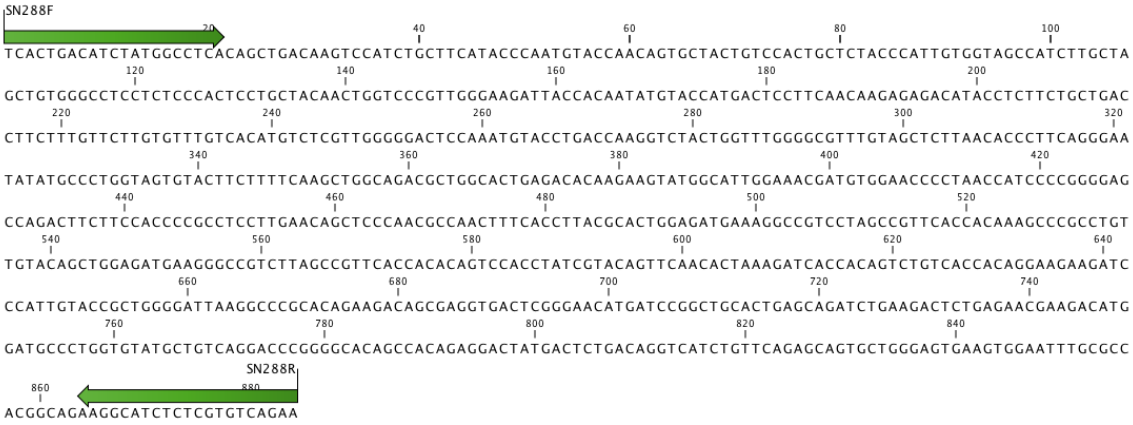

Figure 2.1 *VLGR1* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp49556_c0_seq1, data set available on Compagen.

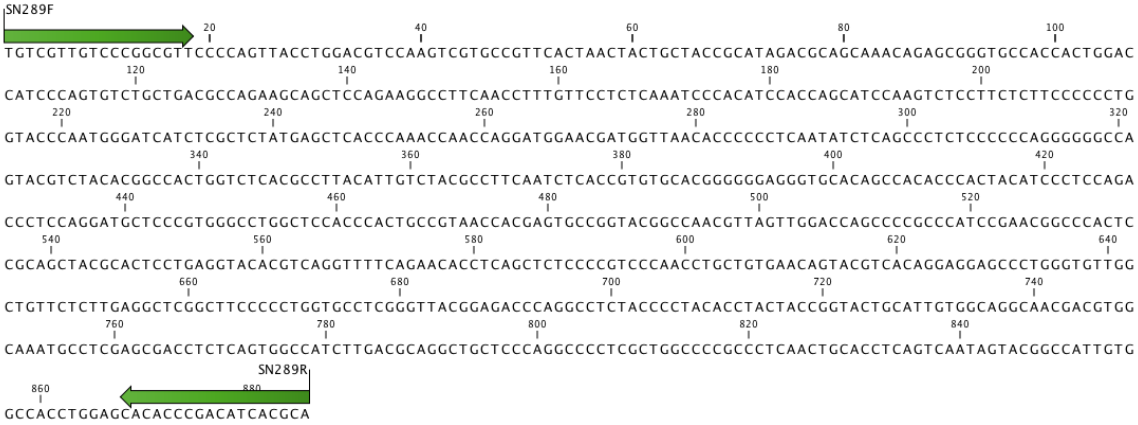

Figure 2.2 *Usherin* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp63844_c0_seq2, data set available on Compagen.

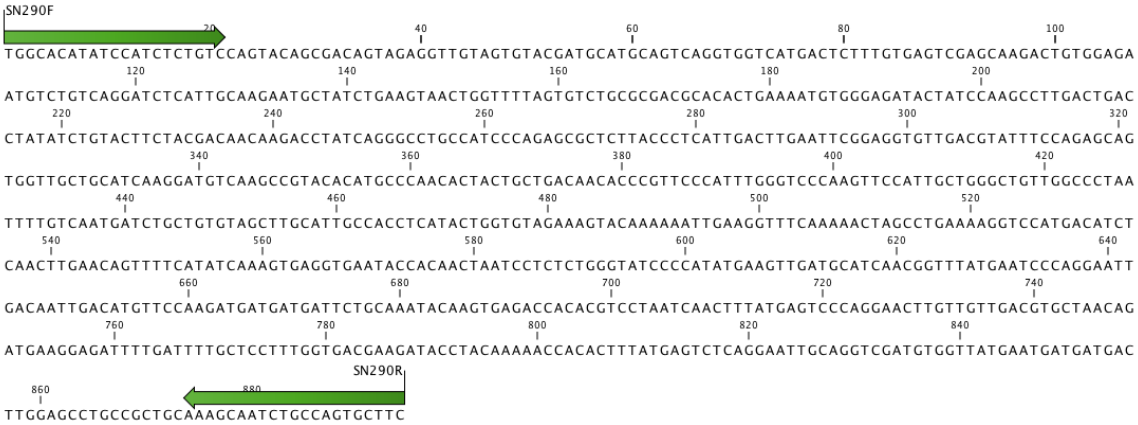

Figure 2.3 *Cadherin* 23 fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp46992_c0_seq3, data set available on Compagen.

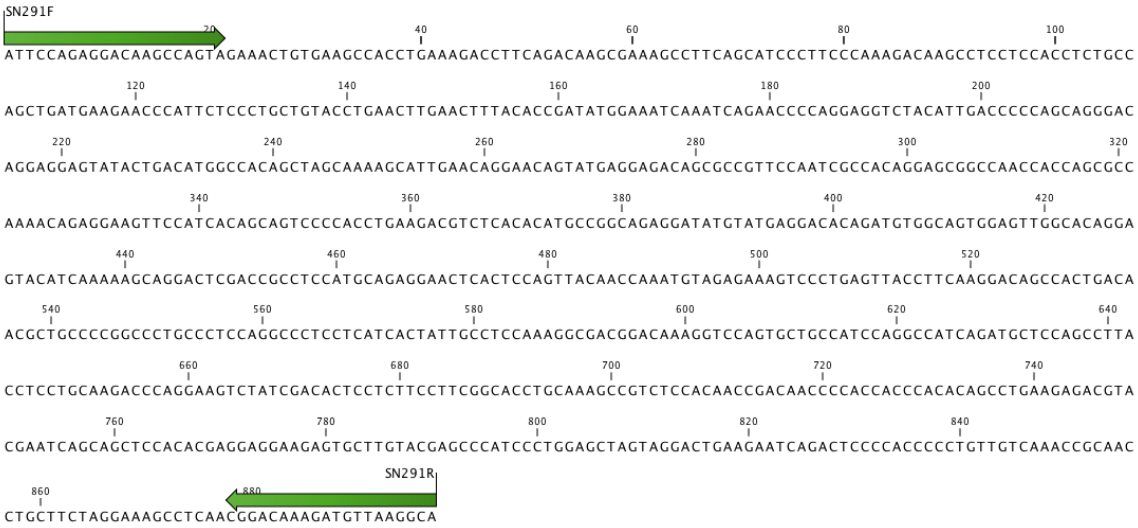

Figure 2.4 *Choanogene* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp68328_c0_seq1, data set available on Compagen.

Figure 2.5 *Annexin* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp66863_c0_seq2, data set available on Compagen.

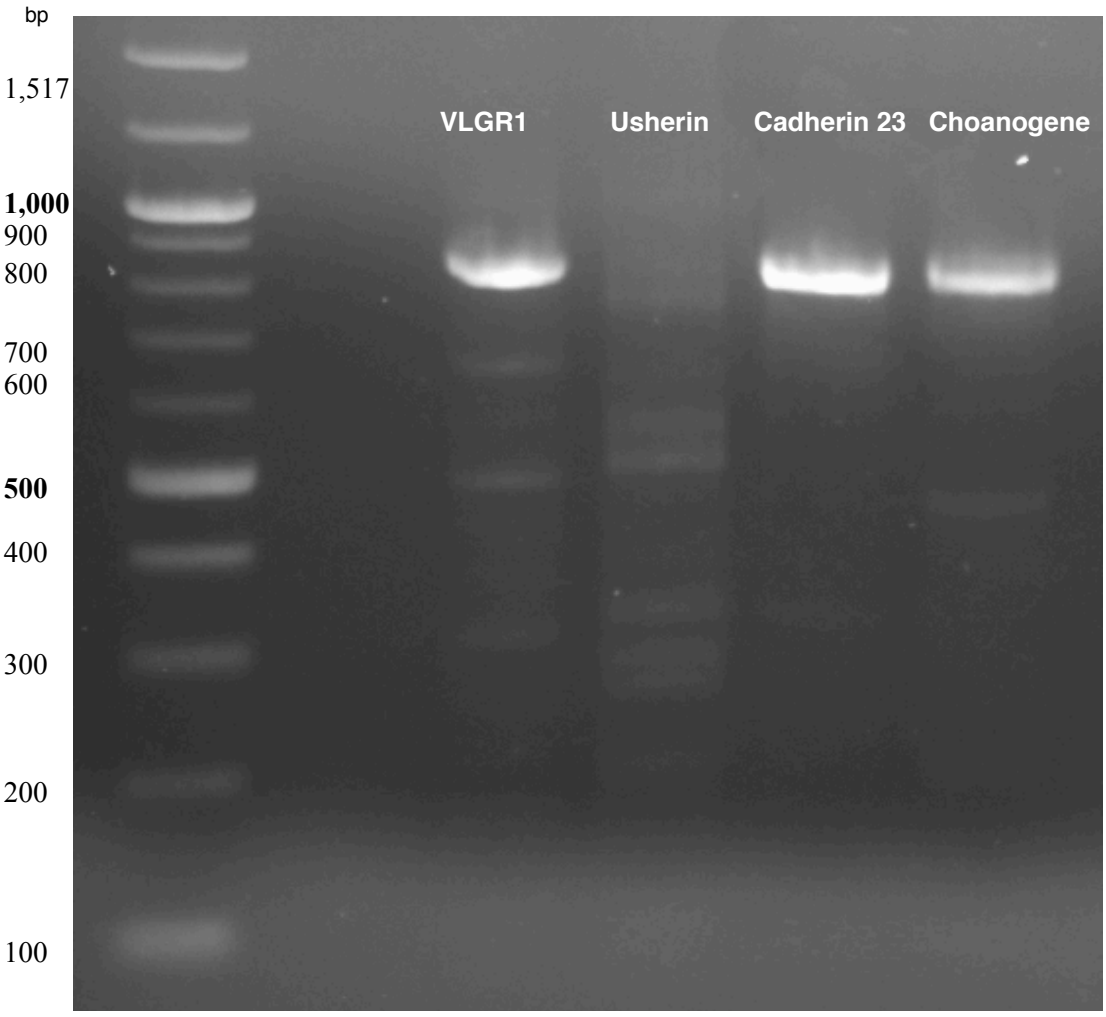

Figure 2.6 Products of PCR on 1% agarose gel. Lane 1 shows 100 basepair ladder (New England BioLabs)

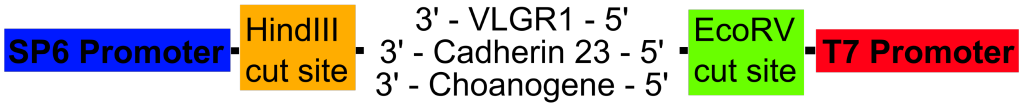

Figure 2.7 Schematic of inserts relative to RNA polymerase promoters in pCR®II-TOPO dual promoter vector

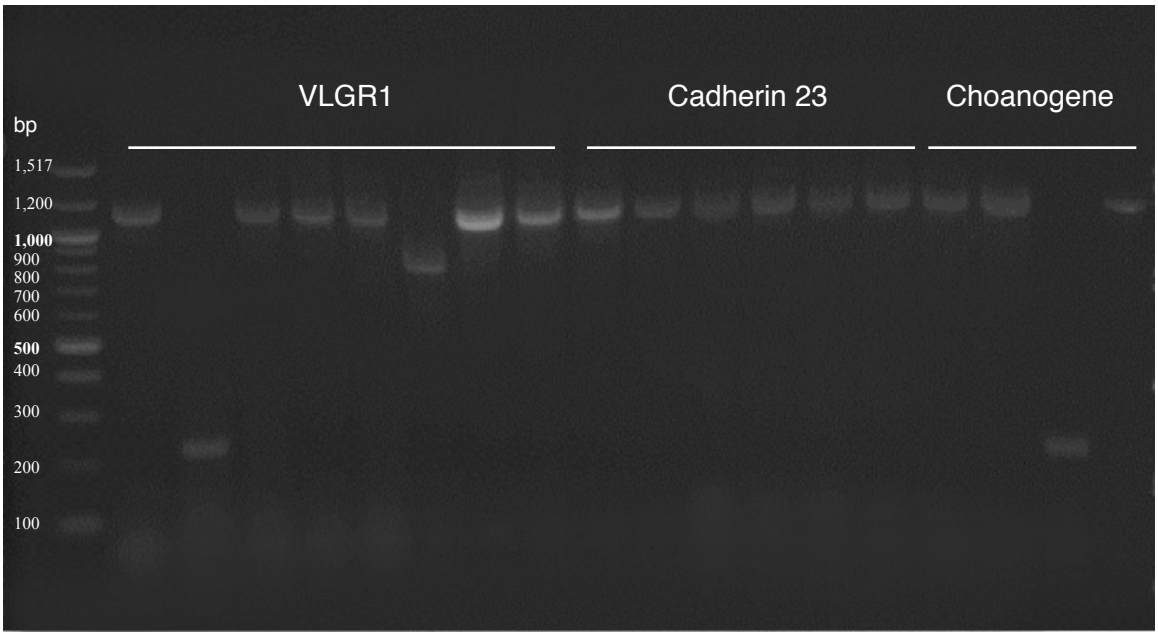

Figure 2.8 Colony screen PCR visualized on 1% agarose gel. The first lane shows 100 bp ladder (New England BioLabs).

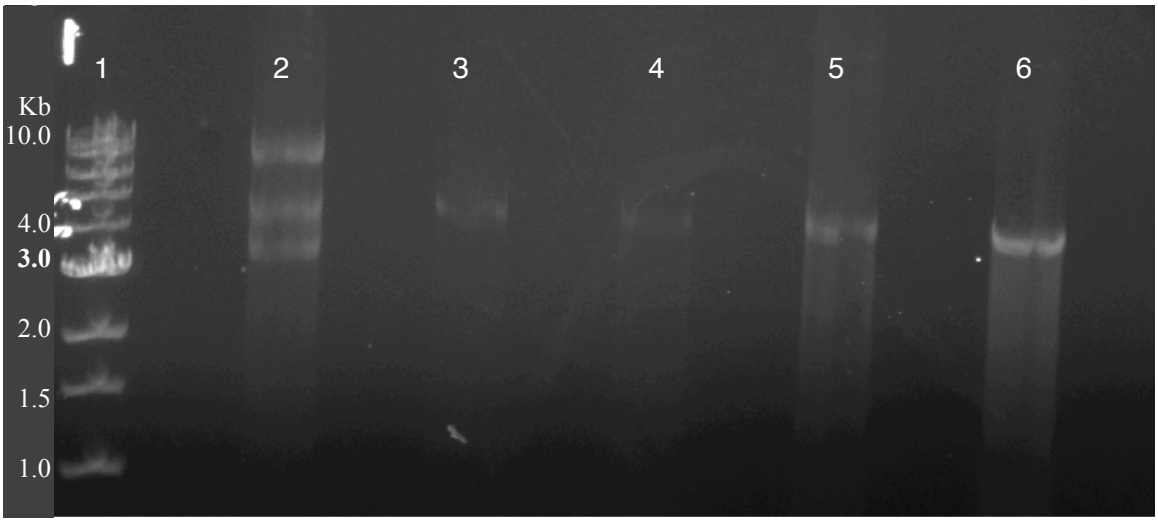

Figure 2.9 Plasmid digest visualized on 1% agarose gel. The first lane shows 1 kilobase ladder (New England BioLabs). Lane 2: undigested choanogene. Lane 3: choanogene sense template. Lane 4: cadherin 23 antisense template. Lane 5: VLGR1 antisense template. Lane 5: choanogene antisense template.

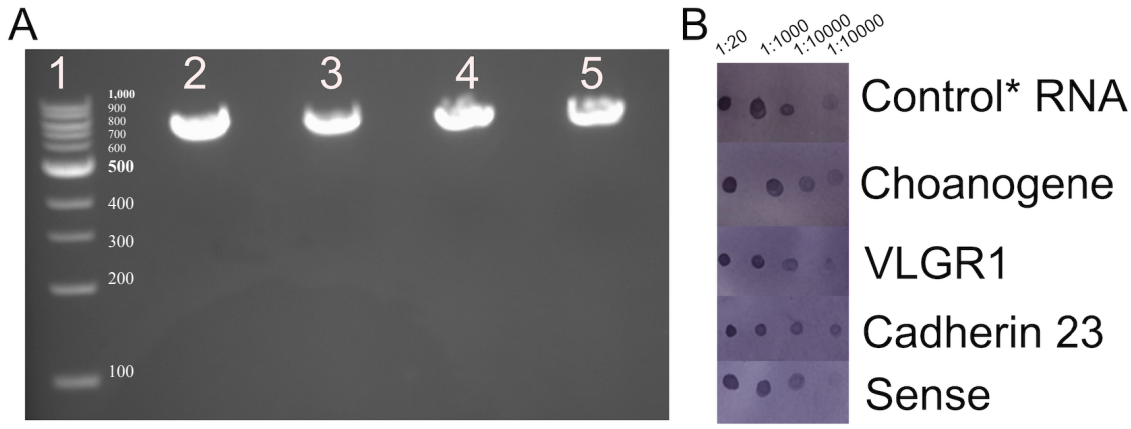

Figure 2.10 Quality of DIG-labelled RNA probes. Probes visualized on a 1% agarose gel (A). Lane 1: 100 bp ladder (New England BioLabs). Lane 2: choanogene antisense probe. Lane 3: VLGR1 antisense probe. Lane 4: cadherin 23 antisense probe. Lane 5: choanogene sense probe. Dot blots of DIG-labelled RNA probes (B). *DIG-labelled control RNA with concentration of 0.1 µg/µl (Roche).

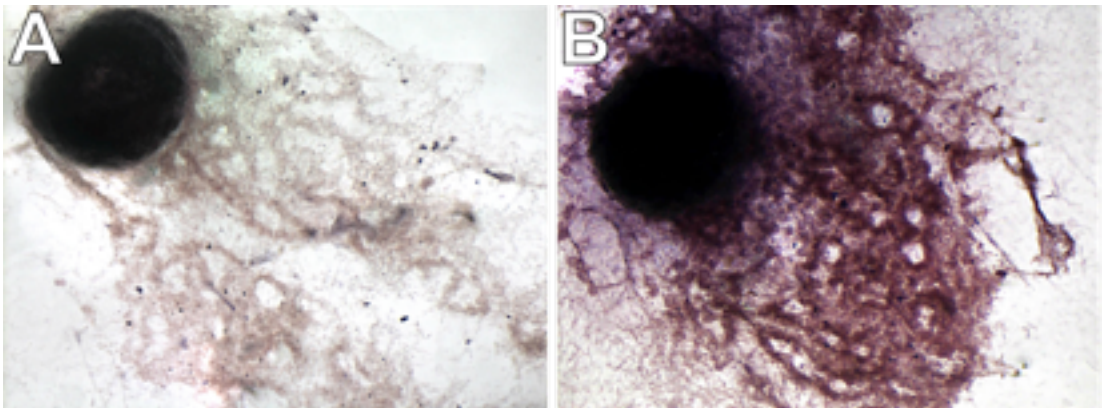

Figure 2.11 Whole-mount *in situ* hybridization using Protocol 2. Annexin sense (A) does not show clear staining. Annexin antisense (B), previously reported to be expressed in choanocytes and archeocytes, displaying ubiquitous staining of tissue.

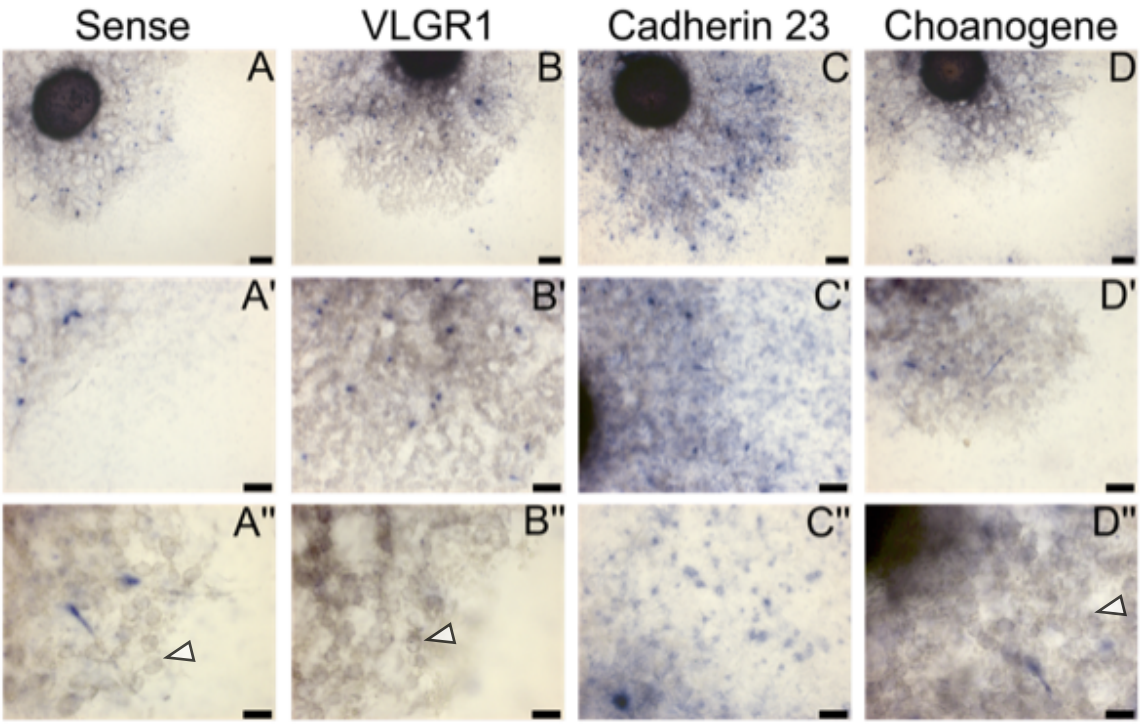

Figure 2.12 Whole-mount *in situ* hybridization using sponge protocol. Choanogene sense strand (A,A',A") shows faint staining in the aquiferous system. VLGR1 (B, B', B") shows faint staining in the aquiferous system. Cadherin 23 shows staining in the pinacoderm (C, C', C"). Choanogene shows staining in the aquiferous system (D, D', D"). Arrowheads pointing to choanocyte chamber. Scale bars= 200 µm (A, B, C, D), 100 µm (A', B', C', D'), 50 µm (A", B", C", D").

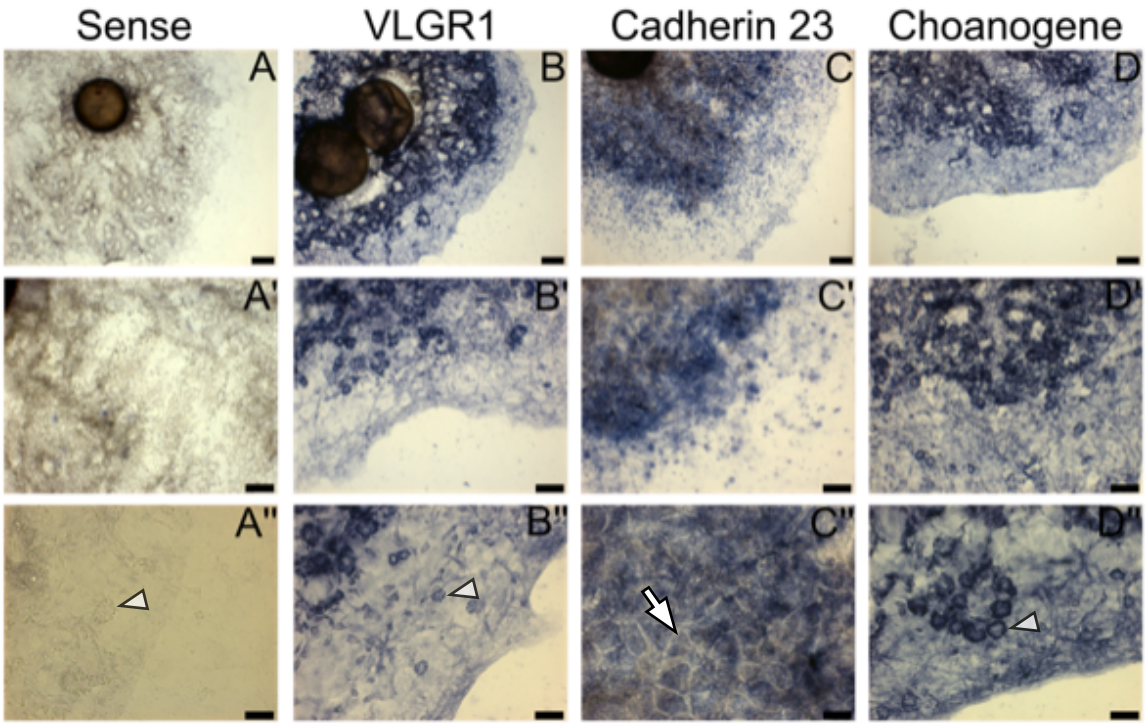

Figure 2.13 Whole-mount *in situ* hybridization using planarian protocol. Choanogene sense strand (A,A',A") shows faint staining in the aquiferous system. VLGR1 (B, B', B") shows strong staining in the aquiferous system. Cadherin 23 shows strong staining in the pinacoderm (C, C', C"). Choanogene shows strong staining in the aquiferous system (D, D', D"). Arrowheads pointing to choanocyte chambers; arrow pointing to pinacocyte. Scale bars= 200 µm (A, B, C, D), 100 µm (A', B', C', D'), 50 µm (A", B", C", D").

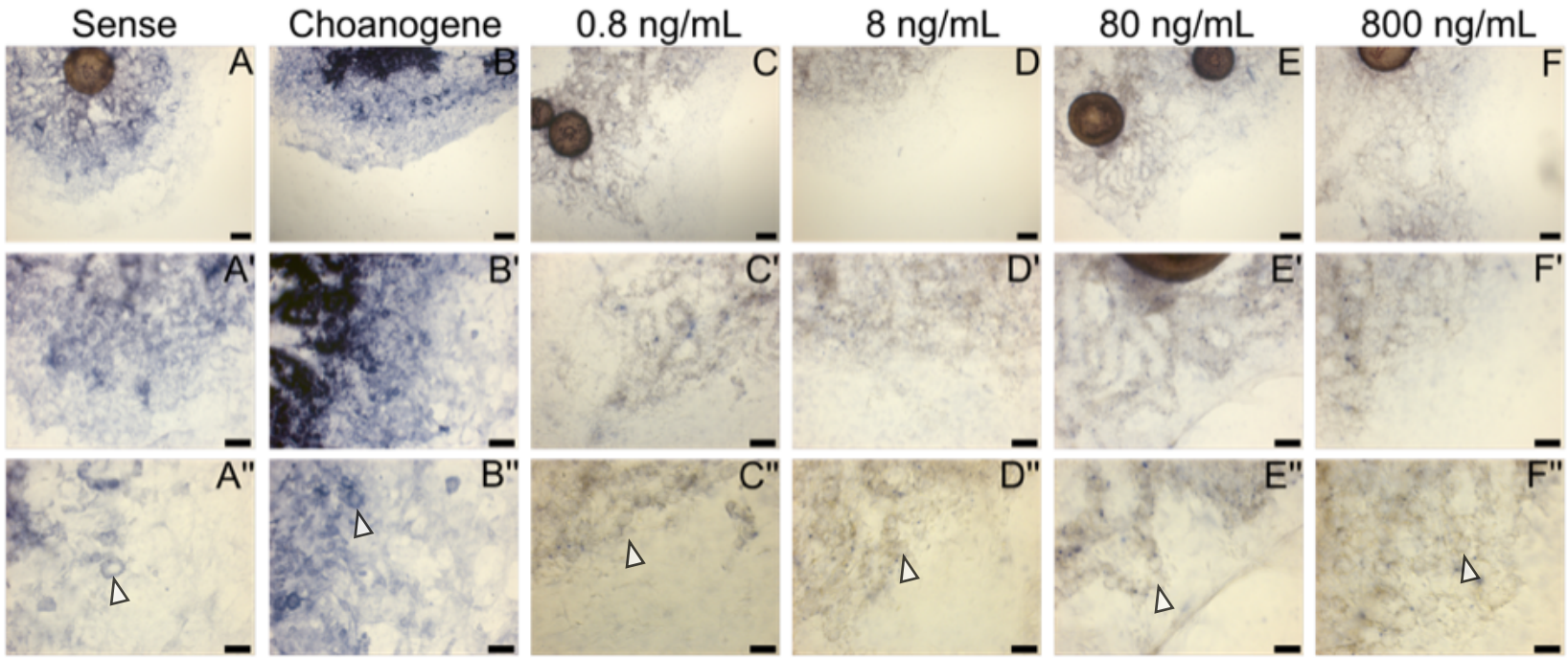

Figure 2.14 Whole-mount *in situ* hybridization treated with varying concentrations of RNase A (Sigma-Aldrich) post-hybridization. Sense choanogene probe (A, A', A") shows staining throughout the tissue. Choanogene antisense probe (B, B', B") shows strong staining as before. Staining throughout tissue is reduced with 0.8 ng/mL RNase A (C, C', C"). Treatment with 8 ng/mL RNase A also reduces staining (D, D', D"). Treatment with 80 ng/mL RNase A reduces staining (E, E', E"). Treatment with 800 ng/mL RNase A also reduces staining (F, F', F"). Arrowheads pointing to choanocyte chambers. Scale bars= 250 µm (A-F), 100 µm (A'-F'), 50 µm (A"-F").

Table 2.2 Comparison of Protocols 1-3 and the Optimal Protocol

	Fixation	Permeabilization	Hybridization	Post-hybridization	Development	Post-treatment	Outcome
Protocol 1	4% PFA 0.05% Glut.	Proteinase K Acetylation		NONE	Water based development buffer		Tissue destroyed
Protocol 2	4% PFA 0.05% Glut.	NONE			Water based development buffer		Ubiquitous staining
Modified Protocol 1	4% PFA 0.05% Glut.	0.5% Triton x-100			Water based development buffer	Ethanol	Some staining
Protocol 3	4% PFA 0.05% Glut.	0.5% Triton x-100 Reduction solution			PVA based development buffer	Ethanol	Strong staining
Modified Protocol 3	4% PFA 0.05% Glut.	0.5% Triton x-100 Reduction solution		RNase treatment	PVA based development buffer	Ethanol	No staining
Optimal Protocol	4% PFA 0.05% Glut.	0.5% Triton x-100 Reduction solution	Stringent hybridization	RNase treatment	PVA based development buffer	Ethanol	

Discussion

In an attempt to validate gene expression dynamics in *E. muelleri*, we worked towards developing a robust *in situ* hybridization protocol. Our general approach was to use whole-mount *in situ* hybridization with dig-labelled RNA probes. Detection relied on the NBT/BCIP color development substrate. For our first attempt we selected the gene annexin, which has been reported to be expressed in choanocytes (Funayama, Nakatsukasa, Hayashi, et al. 2005). This same study used whole-mount *in situ* hybridization, so we followed their protocol (Protocol 1). Two key steps in this protocol prior to hybridization were acetylation and treatment with proteinase K. Acetylation is used to inactivate RNases and decrease background signal (Pernthaler and Amann 2004; Hayashi et al. 1978). Short treatment with proteinase K aids in tissue permeabilization, so the riboprobe can access RNA inside cells (Pearson et al. 2009). Both acetylation and proteinase K treatment resulted in extensive damage to tissues of interest (data not shown). Prior to either of these steps, tissue was intact and structures such as choanocyte chambers and canals were maintained through the various post-fixation wash steps. These observations led us to believe that our fixative was effective and that it was harsh conditions of acetylation and protease treatment which destroyed the tissue. Therefore fixation was not a target of optimization.

Given the fragility of sponge tissue after fixation, we opted for what seemed to be a more gentle protocol (Protocol 2), developed for *Drosophila* embryos (Draizen, Ewer, and Robinow 1999). Unlike the sponge protocol, there was neither an acetylation step

nor a proteinase K permeabilization step so tissue was kept intact throughout. When we developed the tissue, the sense and antisense annexin probes had different staining patterns (Fig. 2.11). The sense probe staining was extremely faint and nonspecific (Fig. 2.11A). In comparing the staining pattern of the antisense annexin probe, we see a strong case for detection of annexin expression (Fig. 2.11B), however, this staining pattern is very different from what has been previously reported (Funayama, Nakatsukasa, Hayashi, et al. 2005). Although we would expect that if it were truly ubiquitous and non-specific staining, the sense probe would have identical staining. The difficulties in interpreting these results prompted us to return to the sponge protocol.

Rather than continue with annexin probes, we transitioned to using candidate choanocyte genes identified with our hydroxyurea/RNA-seq studies (see previous chapter). We focused on cadherin 23 and VLGR1 because they are associated with microvillar structure, function, and development in other metazoans (McGee et al. 2006; Michel et al. 2005; Selvakumar, Drescher, and Drescher 2013; Assad, Shepherd, and Corey 1991; Suli et al. 2012; Anderson and Bouchard 2009). Validating expression of these two genes in choanocytes could support novel mechanosensory roles for the choanoderm. We included a third gene which we refer to as "choanogene." Homologues of sponge choanogene are found only in choanoflagellates based on our own phmmer searches.

In the second iteration of the Protocol 1, we omitted acetylation and proteinase K treatment steps. Since proteinase treatment enhances probe penetration, we reasoned

excluding it would reduce permeability, so we chose to hydrolyze the dig-labelled RNA probes assuming that a smaller probe would be able to penetrate tissue more effectively. Despite this, we did not see any staining (data not shown). Reasoning that reduced tissue permeability would still present a barrier, we included a single wash step with PBS and 0.5% triton X-100 after fixation and before pre-hybridization. Using full length probes, we found that this additional wash/permeabilization step resulted in tissue staining. The sense probe shown in Figure 2.12 has slightly weaker signal intensity when compared to the antisense probes (Fig. 2.12B-D). The staining patterns of VLGR1 and choanogene are comparable in that they both primarily show up in the choanoderm and the aquiferous system. What seems to lend support towards the robustness of this procedure is that cadherin 23 has a very different staining pattern than the other two probes. According to our results, cadherin 23 signal is localized to the basopinacoderm – the tissue that interfaces with the substrate. It has been reported that the cells of the basal surface of the sponge are responsible for its ability to crawl (Bond and Harris 1988). Taken together, it's possible that cadherin 23 is playing a role in basal pinacocyte motility. Alternatively, cadherin 23 could be playing a role in adhesion to the substrate.

To evaluate whether the results of the Protocol 1 could be accepted or even improved, we tested a third protocol, Protocol 3. This protocol was originally developed as a formaldehyde-based *in situ* hybridization for planarians (Pearson et al. 2009). In adapting Protocol 3 for sponge tissue, we excluded steps involved in mucolysis and removal of pigment. From our previous *in situ* trials, we also omitted acetylation and

proteinase K treatment but included one wash step with 0.5% triton X-100 in PBS. A key step in Protocol 3 is a reduction step which calls for a solution containing DTT, a reducing agent, and two different detergents. This step has been shown to increase probe penetration in previously impermeable tissue of the planarian *Schmidtea mediterranea* (Pearson et al. 2009). DTT may be targeting the exoplasmic domains of proteins with disulphide bridges (Yang et al. 2006). One other difference in this protocol is the use of 10% polyvinyl alcohol in the development buffer instead of water. This is a crowding agent that artificially increases the concentration of the NBT/BCIP development substrate. Additionally we included post-treatment with ethanol to remove non-specific background staining. This protocol resulted in very distinct staining patterns. Importantly, there was little to no staining in the sense-strand control, suggesting that the other staining patterns were specific. The staining patterns of the antisense probes resembled those of the sponge protocol antisense probes, except they are much more intense (Fig. 2.13B-D). Both VLGR1 and choanogene show choanoderm expression as predicted by our RNA-seq studies. Cadherin 23 however shows expression in the basal pinacoderm as in the sponge protocol.

We focused on optimizing tissue integrity, probe penetrance, and detection in both Protocol 1 and Protocol 2. Next, we wanted to verify specificity as well as decrease background non-specific staining. An RNAse treated choanogene sample was included in the first attempt of the Protocol 3 which resulted in no signal (data not shown). To follow up on this result, Protocol 3 was repeated with an RNase treatment step just after

hybridization to cleave unbound probe and increase signal specificity (Fig. 2.14; Pearson et al. 2009; H. Yang et al. 1999). We varied the concentration of RNase in PBS and saw that signal was abolished even at low concentrations (Fig 2.14B), indicating that all probe binding was actually non-specific (Fig. 2.14). No sponge specific protocols have reported the use of RNase post-hybridization. This calls into question the validity of *in situ* hybridizations performed on sponge tissue in other studies.

In situ hybridization in *E. muelleri* remains a significant challenge and requires further optimization, perhaps at other steps in the process. As of now we have determined that fixation with 4% paraformaldehyde and 0.05% glutaraldehyde in PBS maintains tissue integrity. Reducing membrane proteins aids with permeabilization without compromising tissue integrity. Post-hybridization treatment with RNase should be used as an additional control to gauge non-specific binding. The use of polyvinyl alcohol seems to help in the intensity of substrate development. Post-treatment with ethanol has also been useful to remove non-specific background staining. The next step for optimization will be to focus on the hybridization parameters. One of these parameters is the stringency of hybridization. The idea is that varying the stringency would affect the stability of bound probe and seeing if this is the reason that staining disappears when treated with RNase.

Additionally, the optimized protocol will be adapted to use fluorescent detection to enable high-resolution, cell-level discrimination of gene expression patterns. Choanocytes are very small cells, ranging from 2 to 8 μm in width and 2 to 10 μm in

width (Mah, Christensen-Dalsgaard, and Leys 2014). The sponge body is also thick enough that confocal imaging is required to get an accurate depiction of internal structures at higher magnifications.

Literature Cited

Anderson, Peter a V, and Christelle Bouchard. 2009. "The Regulation of Cnidocyte Discharge." *Toxicon* 54 (8). Elsevier Ltd: 1046–53. doi: 10.1016/j.toxicon.2009.02.023.

Assad, J A, G M Shepherd, and D P Corey. 1991. "Tip-Link Integrity and Mechanical Transduction in Vertebrate Hair Cells." *Neuron* 7 (6): 985–94. doi:10.1016/0896-6273(91)90343-X.

Bond, C, and A K Harris. 1988. "Locomotion of Sponges and Its Physical Mechanism." *The Journal of Experimental Zoology* 246 (3): 271–84. doi: 10.1002/jez.1402460307.

Brown, Jeffrey W., and C. James McKnight. 2010. "Molecular Model of the Microvillar Cytoskeleton and Organization of the Brush Border." *PLoS ONE* 5 (2). doi:10.1371/journal.pone.0009406.

Cantell, C. E., Å Franzén, and T. Sensenbaugh. 1982. "Ultrastructure of Multiciliated Collar Cells in the Pilidium Larva of Lineus Bilineatus (Nemertini)." *Zoomorphology* 101: 1–15.

Chen, Chiao-Lin, Kathleen M Gajewski, Fisun Hamaratoglu, Wouter Bossuyt, Leticia Sansores-Garcia, Chunyao Tao, and Georg Halder. 2010. "The Apical-Basal Cell Polarity Determinant Crumbs Regulates Hippo Signaling in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 107 (36): 15810–15. doi: 10.1073/pnas.1004060107.

Collins, a G. 1998. "Evaluating Multiple Alternative Hypotheses for the Origin of Bilateria: An Analysis of 18S rRNA Molecular Evidence." *Proceedings of the National Academy of Sciences of the United States of America* 95 (26): 15458–63. doi:10.1073/pnas.95.26.15458.

Conesa, Ana, and Stefan Götz. 2008. “Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics.” *International Journal of Plant Genomics* 2008. doi:10.1155/2008/619832.

Conesa, Ana, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. 2005. “Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research.” *Bioinformatics* 21 (18): 3674–76. doi:10.1093/bioinformatics/bti610.

Davidson, Nadia M, and Alicia Oshlack. 2014. “Corset: Enabling Differential Gene Expression Analysis For.” *Genome Biology* 15 (7): 410. doi:10.1186/PREACCEPT-2088857056122054.

Dayel, Mark J, Rosanna a Alegado, Stephen R Fairclough, Tera C Levin, Scott a Nichols, Kent McDonald, and Nicole King. 2011. “Cell Differentiation and Morphogenesis in the Colony-Forming Choanoflagellate Salpingoeca Rosetta.” *Developmental Biology* 357 (1). Elsevier Inc.: 73–82. doi:10.1016/j.ydbio.2011.06.003.

Dayel, Mark J, and Nicole King. 2014. “Prey Capture and Phagocytosis in the Choanoflagellate Salpingoeca Rosetta.” *PloS One* 9 (5): e95577. doi:10.1371/journal.pone.0095577.

Draizen, Troy a., John Ewer, and Steven Robinow. 1999. “Genetic and Hormonal Regulation of the Death of Peptidergic Neurons in the Drosophila Central Nervous System.” *Journal of Neurobiology* 38 (4): 455–65. doi:10.1002/(SICI)1097-4695(199903)38:4<455::AID-NEU2>3.0.CO;2-F.

Dunn, Casey W, Andreas Hejnol, David Q Matus, Kevin Pang, William E Browne, Stephen A Smith, Elaine Seaver, et al. 2008. “Broad

Phylogenomic Sampling Improves Resolution of the Animal Tree of Life." *Nature* 452 (7188): 745–49. doi:10.1038/nature06614.

Dunn, Casey W., Sally P. Leys, and Steven H.D. Haddock. 2015. "The Hidden Biology of Sponges and Ctenophores." *Trends in Ecology & Evolution* 30 (5). Elsevier Ltd: 282–91. doi:10.1016/j.tree.2015.03.003.

Funayama, Noriko, Mikiko Nakatsukasa, Tetsutaro Hayashi, and Kiyokazu Agata. 2005. "Isolation of the Choanocyte in the Fresh Water Sponge, Ephydatia Fluviatilis and Its Lineage Marker, Ef Annexin." *Development, Growth & Differentiation* 47 (4): 243–53. doi:10.1111/j.1440-169X. 2005.00800.x.

Funayama, Noriko, Mikiko Nakatsukasa, Shigehiro Kuraku, Katsuaki Takechi, Mikako Dohi, Naoyuki Iwabe, Takashi Miyata, and Kiyokazu Agata. 2005. "Isolation of Ef Silicatein and Ef Lectin as Molecular Markers for Sclerocytes and Cells Involved in Innate Immunity in the Freshwater Sponge Ephydatia Fluviatilis." *Zoological Science* 22 (10): 1113–22. http://www.ncbi.nlm.nih.gov/pubmed/16286723.

Funayama, Noriko, Mikiko Nakatsukasa, Kurato Mohri, Yoshiki Masuda, and Kiyokazu Agata. 2010. "Piwi Expression in Archeocytes and Choanocytes in Demosponges: Insights into the Stem Cell System in Demosponges." *Evolution and Development* 12 (3): 275–87. doi:10.1111/ j.1525-142X.2010.00413.x.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. 2003. "ExPASy: The proteomics server for in-depth protein knowledge and analysis." *Nucleic Acids Research* 31 (13): 3784–3788. http://doi.org/ 10.1093/nar/gkg563

Gonobobleva, Elisaveta, and Manuel Maldonado. 2009. “Choanocyte Ultrastructure in Halisarca Dujardini (Demospongiae, Halisarcida).” *Journal of Morphology* 270 (5): 615–27. doi:10.1002/jmor.10709.

Götz, Stefan, Roland Arnold, Patricia Sebastián-León, Samuel Martín-Rodríguez, Patrick Tischler, Marc André Jehl, Joaquín Dopazo, Thomas Rattei, and Ana Conesa. 2011. “B2G-FAR, a Species-Centered GO Annotation Repository.” *Bioinformatics* 27 (7): 919–24. doi:10.1093/bioinformatics/btr059.

Götz, Stefan, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. 2008. “High-Throughput Functional Annotation and Data Mining with the Blast2GO Suite.” *Nucleic Acids Research* 36 (10): 3420–35. doi:10.1093/nar/gkn176.

Hayashi, S., I. C. Gillam, A. D. Delaney, and G. M. Tener. 1978. “Acetylation of Chromosome Squashes of Drosophila Melanogaster Decreases the Background in Autoradiographs from Hybridization with [125I--Labeled RNA.” *The Journal of Histochemistry and Cytochemistry* 33 (August): 683–85.

Hemmrich, Georg, and Thomas C G Bosch. 2008. “Compagen, a Comparative Genomics Platform for Early Branching Metazoan Animals, Reveals Early Origins of Genes Regulating Stem-Cell Differentiation.” *BioEssays* 30 (10): 1010–18. doi:10.1002/bies.20813.

Jacobs, Dave K., Nagayasu Nakanishi, David Yuan, Anthony Camara, Scott a. Nichols, and Volker Hartenstein. 2007. “Evolution of Sensory Structures in Basal Metazoa.” *Integrative and Comparative Biology* 47 (5): 712–23. doi:10.1093/icb/icm094.

James-Clark, Henry. 1867. “On the Spongiae Ciliatiae as Infusora Flagellata.” *Annals and Magazine of Natural History* 1: 133–42.

———. 1871. “No Note on the Infusoria Flagellata and the Spongiae Ciliatae.” *American Journal of Science* 1: 113–14. doi:10.2475/ajs.s3-1.2.113.

Jin, L, and R V Lloyd. 1997. “In Situ Hybridization: Methods and Applications.” *Journal of Clinical Laboratory Analysis* 11 (1): 2–9. doi:10.1002/(SICI)1098-2825(1997)11:1<2::AID-JCLA2>3.0.CO;2-F.

Jones, Philip, David Binns, Hsin Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” *Bioinformatics* 30 (9): 1236–40. doi:10.1093/bioinformatics/btu031.

Kent, W. S. 1878. “Notes on the Embryology of Sponges.” *Annals and Magazine of Natural History* 5: 139–56.

King, Nicole. 2004. “The Unicellular Ancestry of Animal Development.” *Developmental Cell* 7 (3): 313–25. doi:10.1016/j.devcel.2004.08.010.

Kleene, S. J., and J. L. Van Houten. 2014. “Electrical Signaling in Motile and Primary Cilia.” *BioScience* 64 (12): 1092–1102. doi:10.1093/biosci/biu181.

Koç, Ahmet, Linda J. Wheeler, Christopher K. Mathews, and Gary F. Merrill. 2004. “Hydroxyurea Arrests DNA Replication by a Mechanism That Preserves Basal dNTP Pools.” *Journal of Biological Chemistry* 279 (1): 223–30. doi:10.1074/jbc.M303952200.

Larroux, Claire, Bryony Fahey, Danielle Liubicich, Veronica F. Hinman, Marie Gauthier, Milena Gongora, Kathryn Green, Gert Wörheide, Sally P. Leys, and Bernard M. Degnan. 2006. “Developmental Expression of

Transcription Factor Genes in a Demosponge: Insights into the Origin of Metazoan Multicellularity." *Evolution and Development* 8 (2): 150–73. doi:10.1111/j.1525-142X.2006.00086.x.

Leys, Sally P, Scott a Nichols, and Emily D M Adams. 2009. "Epithelia and Integration in Sponges." *Integrative and Comparative Biology* 49 (2): 167–77. doi:10.1093/icb/icp038.

Ludeman, Danielle a, Nathan Farrar, Ana Riesgo, Jordi Paps, and Sally P Leys. 2014. "Evolutionary Origins of Sensation in Metazoans: Functional Evidence for a New Sensory Organ in Sponges." *BMC Evolutionary Biology* 14: 3. doi:10.1186/1471-2148-14-3.

Lyons, K. M. 1973. "Collar Cells in Planula and Adult Tentacle Ectoderm of the Solitary Coral Balanophyllia Regia (Anthozoa Eupsammiidae)" 145: 57–74.

Mah, Jasmine L, Karen K Christensen-Dalsgaard, and Sally P Leys. 2014. "Choanoflagellate and Choanocyte Collar-Flagellar Systems and the Assumption of Homology." *Evolution & Development* 16 (1): 25–37. doi: 10.1111/ede.12060.

Maldonado, Manuel. 2004. "Choanoflagellates , Choanocytes , and Animal Multicellularity." *Invertebrate Biology* 123 (1): 1–22. doi:10.1111/j. 1744-7410.2004.tb00138.x.

Martinez, A, J Lopez, A C Villaro, and P Sesma. 1991. "Choanocyte-Like Cells in the Digestive-System of the Starfish Marthasterias-Glacialis (Echinodermata)." *Journal of Morphology* 208: 215–25. <Go to ISI>:// A1991FW20200006.

Mayer, Ulrich, Alexander Küller, Philipp C Daiber, Inge Neudorf, Uwe Warnken, Martina Schnölzer, Stephan Frings, and Frank Möhrlen. 2009.

"The Proteome of Rat Olfactory Sensory Cilia." *Proteomics* 9 (2): 322–34. doi:10.1002/pmic.200800149.

McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. "Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation." *Nucleic Acids Research* 40 (10): 4288–97. doi: 10.1093/nar/gks042.

McGee, Joann, Richard J Goodyear, D Randy McMillan, Eric a Stauffer, Jeffrey R Holt, Kirsten G Locke, David G Birch, et al. 2006. "The Very Large G-Protein-Coupled Receptor VLGR1: A Component of the Ankle Link Complex Required for the Normal Development of Auditory Hair Bundles." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 26 (24): 6543–53. doi:10.1523/JNEUROSCI. 0693-06.2006.

Medina, M, a G Collins, J D Silberman, and M L Sogin. 2001. "Evaluating Hypotheses of Basal Animal Phylogeny Using Complete Sequences of Large and Small Subunit rRNA." *Proceedings of the National Academy of Sciences of the United States of America* 98 (17): 9707–12. doi: 10.1073/pnas.171316998.

Michel, Vincent, Richard J. Goodyear, Dominique Weil, Walter Marcotti, Isabelle Perfettini, Uwe Wolfrum, Corné J. Kros, Guy P. Richardson, and Christine Petit. 2005. "Cadherin 23 Is a Component of the Transient Lateral Links in the Developing Hair Bundles of Cochlear Sensory Cells." *Developmental Biology* 280 (2): 281–94. doi:10.1016/j.ydbio. 2005.01.014.

Mohri, Hideo, Kazuo Inaba, Sumio Ishijima, and Shoji a. Baba. 2012. "Tubulin-Dynein System in Flagellar and Ciliary Movement."

Proceedings of the Japan Academy, Series B 88 (8): 397–415. doi: 10.2183/pjab.88.397.

Moroz, Leonid L, Kevin M Kocot, Mathew R Citarella, Sohn Dosung, Tigran P Norekian, Inna S Povolotskaya, Anastasia P Grigorenko, et al. 2014. "The Ctenophore Genome and the Evolutionary Origins of Neural Systems." *Nature* 510 (7503): 109–14. doi:10.1038/nature13400.

Mrkusich, Eli M., Dustin J. Flanagan, and Paul M. Whitington. 2011. "The Core Planar Cell Polarity Gene Prickle Interacts with Flamingo to Promote Sensory Axon Advance in the Drosophila Embryo." *Developmental Biology* 358 (1). Elsevier B.V.: 224–30. doi:10.1016/ j.ydbio.2011.07.032.

Nichols, Scott Anthony, Brock William Roberts, Daniel Joseph Richter, Stephen Robert Fairclough, and Nicole King. 2012. "Origin of Metazoan Cadherin Diversity and the Antiquity of the Classical Cadherin/β-Catenin Complex." *Proceedings of the National Academy of Sciences of the United States of America* 109 (32): 13046–51. doi: 10.1073/pnas.1120685109.

Nielsen, Claus. 2008. "Six Major Steps in Animal Evolution: Are We Derived Sponge Larvae?" *Evolution & Development* 10 (2): 241–57. doi:10.1111/ j.1525-142X.2008.00231.x.

Norrevang, A., and K. G. Wingstrand. 1970. "On the Occurrence and Structure of Choanocyte-like Cells in Some Echinoderms." *Acta Zoologica*.

Paradela, Alberto, Susana B Bravo, Mauricio Henríquez, Gloria Riquelme, Francisco Gavilanes, José M González-Ros, and Juan P Albar. 2005. "Proteomic Analysis of Apical Microvillous Membranes of Syncytiotrophoblast Cells Reveals a High Degree of Similarity with

Lipid Rafts." *Journal of Proteome Research* 4 (6): 2435–41. doi:10.1021/pr050308v.

Pearson, Bret J., George T. Eisenhoffer, Kyle a. Gurley, Jochen C. Rink, Diane E. Miller, and Alejandro Sánchez Alvarado. 2009. "Formaldehyde-Based Whole-Mount in Situ Hybridization Method for Planarians." *Developmental Dynamics* 238 (2): 443–50. doi:10.1002/dvdy.21849.

Pernthaler, Annelie, and Rudolf Amann. 2004. "Simultaneous Fluorescence In Situ Hybridization of mRNA and rRNA in Environmental Bacteria Simultaneous Fluorescence In Situ Hybridization of mRNA and rRNA in Environmental Bacteria." *Applied and Environmental Microbiology* 70 (9): 5426–33. doi:10.1128/AEM.70.9.5426.

Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics (Oxford, England)* 26 (1): 139–40. doi:10.1093/bioinformatics/btp616.

Robinson, Mark D., and Gordon K. Smyth. 2007. "Moderated Statistical Tests for Assessing Differences in Tag Abundance." *Bioinformatics* 23 (21): 2881–87. doi:10.1093/bioinformatics/btm453.

———. 2008. "Small-Sample Estimation of Negative Binomial Dispersion, with Applications to SAGE Data." *Biostatistics* 9 (2): 321–32. doi:10.1093/biostatistics/kxm030.

Rozenfeld, F., and R. Rasmont. 1977. "Hydroxyurea: An Inhibitor of the Differentiation of Choanocytes in Fresh-Water Sponges and a Possible Agent for the Isolation of Embryonic Cells." *Differentiation* 7 (1-3). International Society of Differentiation: 53–60. doi:10.1111/j.1432-0436.1977.tb01496.x.

Ruiz-Trillo, Iñaki, Andrew J Roger, Gertraud Burger, Michael W Gray, and B Franz Lang. 2008. “A Phylogenomic Investigation into the Origin of Metazoa.” *Molecular Biology and Evolution* 25 (4): 664–72. doi:10.1093/molbev/msn006.

Ryan, Joseph F, Kevin Pang, Christine E Schnitzler, Anh-Dao Nguyen, R Travis Moreland, David K Simmons, Bernard J Koch, et al. 2013. “The Genome of the Ctenophore Mnemiopsis Leidyi and Its Implications for Cell Type Evolution.” *Science (New York, N.Y.)* 342 (6164): 1242592. doi:10.1126/science.1242592.

Schwander, Martin, Bechara Kachar, and Ulrich Müller. 2010. “The Cell Biology of Hearing.” *Journal of Cell Biology* 190 (1): 9–20. doi:10.1083/jcb.201001138.

Sebé-Pedrós, Arnau, Pawel Burkhardt, Núria Sánchez-Pons, Stephen R Fairclough, B Franz Lang, Nicole King, and Iñaki Ruiz-Trillo. 2013. “Insights into the Origin of Metazoan Filopodia and Microvilli.” *Molecular Biology and Evolution* 30 (9): 2013–23. doi:10.1093/molbev/mst110.

Selvakumar, Dakshnamurthy, Marian J. Drescher, and Dennis G. Drescher. 2013. “Cyclic Nucleotide-Gated Channel ??-3 (CNGA3) Interacts with Stereocilia Tip-Link Cadherin 23+Exon 68 or Alternatively with Myosin VIIa, Two Proteins Required for Hair Cell Mechanotransduction.” *Journal of Biological Chemistry* 288 (10): 7215–29. doi:10.1074/jbc.M112.443226.

Suli, Arminda, Glen M. Watson, Edwin W. Rubel, and David W. Raible. 2012. “Rheotaxis in Larval Zebrafish Is Mediated by Lateral Line Mechanosensory Hair Cells.” *PLoS ONE* 7 (2): 1–6. doi:10.1371/journal.pone.0029727.

Tyler, Seth. 2003. “Epithelium--the Primary Building Block for Metazoan Complexity.” *Integrative and Comparative Biology* 43 (1): 55–63. doi: 10.1093/icb/43.1.55.

Watanabe, Seiji, Shuichi Tsuruoka, Soundarapandian Vijayakumar, Gunter Fischer, Yixin Zhang, Akio Fujimura, Qais Al-Awqati, and George J Schwartz. 2005. “Cyclosporin A Produces Distal Renal Tubular Acidosis by Blocking Peptidyl Prolyl Cis-Trans Isomerase Activity of Cyclophilin.” *American Journal of Physiology. Renal Physiology* 288 (1): F40–47. doi:10.1152/ajprenal.00218.2004.

Yang, H, I B Wanner, S D Roper, and N Chaudhari. 1999. “An Optimized Method for in Situ Hybridization with Signal Amplification That Allows the Detection of Rare mRNAs.” *The Journal of Histochemistry and Cytochemistry : Official Journal of the Histochemistry Society* 47 (4): 431–46. doi:10.1177/002215549904700402.

Yang, Jun, Hongtao Chen, Iontcho R Vlahov, Ji-Xin Cheng, and Philip S Low. 2006. “Evaluation of Disulfide Reduction during Receptor-Mediated Endocytosis by Using FRET Imaging.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (37): 13872–77. doi:10.1073/pnas.0601455103.

Zhang, Ming-Zhi, Weiyi Mai, Cunxi Li, Sae-youll Cho, Chuanming Hao, Gilbert Moeckel, Runxiang Zhao, et al. 2004. “PKHD1 Protein Encoded by the Gene for Autosomal Recessive Polycystic Kidney Disease Associates with Basal Bodies and Primary Cilia in Renal Epithelial Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (8): 2311–16. doi:10.1073/pnas.0400073101.

Zhou, Xiaobei, Helen Lindsay, and Mark D. Robinson. 2014. “Robustly Detecting Differential Expression in RNA Sequencing Data Using

Observation Weights.” *Nucleic Acids Research* 42 (11). doi:10.1093/nar/gku310.

Appendix 1

Table A.1 RNA concentration from TRIzol extraction		
Sample	Concentration	260/280 Ratio
Em-BTL-C	1480 ng/µl	2.0
Em-BTL-HU	589 ng/µl	2.0
Em-NR-C	931 ng/µl	2.0
Em-NR-HU	380 ng/µl	1.98
Em-CO-C	711 ng/µl	2.16
Em-CO-HU	743 ng/µl	2.0

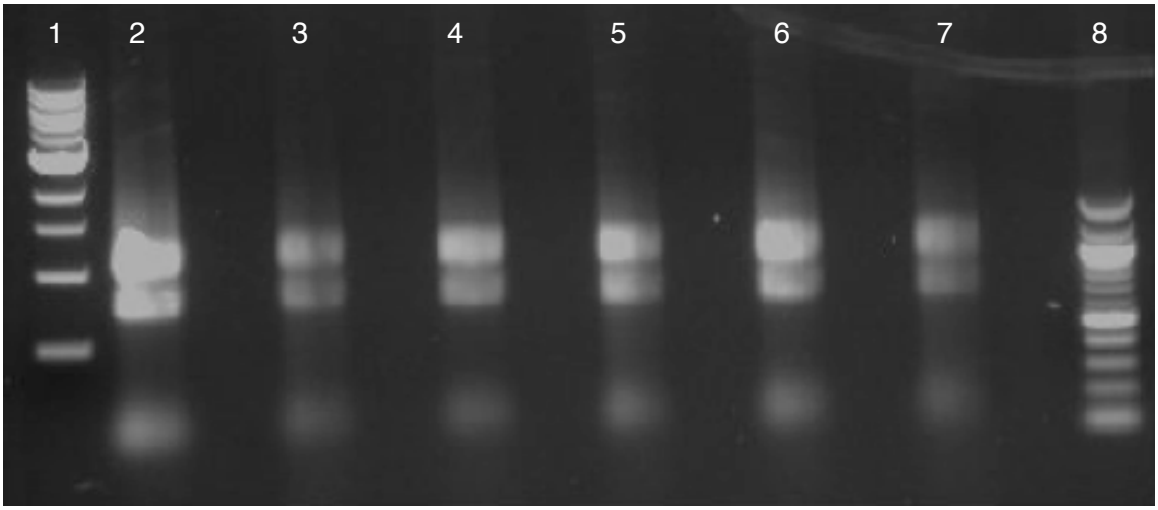

Figure A.1 RNA isolated with TRIzol visualized on a 1% agarose formaldehyde gel. The first lane shows 1 kb ladder and lane 8 shows 100 bp ladder (New England Biolabs). Lane 2: Em-BTL-C, Lane 3: Em-BTL-HU, Lane 4: Em-NR-C, Lane 5: Em-NR-HU, Lane 6: Em-CO-C, Lane 7: Em-CO-HU.

Table A.2 RNAseq run quality statistics

Sample ID	Em-BTL-C	Em-BTL-HU	Em-NR-C	Em-NR-HU	Em-CO-C	Em-CO-HU
Yield (Mbases)	1,607	1,584	1,187	1,405	1,610	1,479
%PF*	100	100	100	100	100	100
# Reads	31,514,217	31,065,437	23,283,522	27,542,226	31,571,758	28,999,119
% raw clusters per	17.85	17.59	13.19	15.6	17.88	16.42
% Perfect index	99.46	99.28	99.4	99.19	99.4	99.45
% one mismatch	0.54	0.72	0.6	0.81	0.6	0.55
% ≥Q30 bases (PF)	97.19	97.11	97.11	97.13	97.15	97.15
Mean quality score (PF)	37.7	37.67	37.68	37.68	37.69	37.69

*Percentage of clusters that passed filtering

Table A.3 Summary of RNAseq run quality statistics for control group				
Sample ID	Mean	Standard Deviation	Maximum	Minimum
Yield (Mbases)	1,468	243.4	1,610	1,187
%PF	100.0	0.0	100	100
# Reads	28,789,832.3	4,768,691.4	31,571,758	23,283,522
% raw clusters per lane	16.3	2.7	17.88	13.19
% Perfect index reads	99.4	0.0	99.46	99.40
% one mismatch reads (index)	0.6	0.0	0.60	0.54
% ≥Q30 bases (PF)	97.2	0.0	97.19	97.11
Mean quality score (PF)	37.7	0.0	37.70	37.68

Table A.4 Summary of RNAseq run quality statistics for HU-treated group				
Sample ID	Mean	Standard Deviation	Maximum	Minimum
Yield (Mbases)	1,489.3	89.9	1,584	1,405
%PF	100.0	0.0	100	100
# Reads	29,202,260.7	1,770,368.3	31,065,437	27,542,226
% raw clusters per lane	16.5	1.0	17.59	15.60
% Perfect index reads	99.3	0.1	99.45	99.19
% one mismatch reads (index)	0.7	0.1	0.81	0.55
% ≥Q30 bases (PF)	97.1	0.0	97.15	97.11
Mean quality score (PF)	37.7	0.0	36.69	37.67

Table A.5 Normalization of raw count data			
Sample ID	Library size	Normalization factor	Effective Library Size
Em-BTL-C	26,821,746	0.9998731	26,818,342
Em-BTL-HU	26,177,350	0.8950515	23,430,076
Em-NR-C	21,112,184	1.0431572	22,023,327
Em-NR-HU	25,285,262	1.0123698	25,598,036
Em-CO-C	29,342,058	1.0147581	29,775,091
Em-CO-HU	27,152,719	1.0426909	28,311,893

Downregulated gene accession ID

comp68397_c0_seq23
comp42749_c0_seq1
comp51379_c0_seq2
comp66724_c0_seq1
comp66688_c0_seq8
comp67932_c2_seq21
comp65681_c0_seq1
comp67188_c0_seq1
comp219645_c0_seq1
comp25505_c0_seq1
comp23605_c0_seq1
comp66101_c0_seq14
comp68572_c3_seq18
comp42165_c0_seq1
comp49514_c0_seq1
comp69560_c0_seq1
comp210916_c0_seq1
comp58872_c0_seq1
comp56961_c0_seq1
comp62810_c1_seq4
comp68360_c0_seq2

comp68368_c0_seq38
comp62877_c0_seq4
comp51165_c0_seq2
comp47644_c0_seq1
comp68275_c0_seq5
comp67296_c0_seq1
comp78308_c0_seq1
comp68156_c0_seq2
comp61147_c0_seq1
comp39881_c0_seq1
comp65654_c0_seq5
comp68328_c0_seq1
comp68464_c0_seq47
comp59914_c0_seq1
comp62776_c0_seq3
comp60720_c0_seq5
comp70813_c0_seq1
comp65565_c0_seq2
comp23295_c0_seq1
comp61010_c0_seq4
comp37951_c0_seq2
comp69492_c0_seq1
comp58018_c0_seq2
comp63516_c0_seq3
comp67378_c0_seq12
comp67114_c0_seq2
comp58004_c0_seq8
comp23695_c0_seq2
comp54618_c0_seq1
comp66145_c0_seq2
comp54521_c0_seq2
comp56532_c1_seq1
comp62374_c0_seq2
comp54823_c0_seq2
comp62268_c0_seq4
comp68572_c3_seq22
comp63799_c0_seq2
comp61704_c1_seq3
comp32203_c0_seq1
comp68749_c0_seq3
comp40343_c0_seq1

comp64722_c0_seq4
comp76174_c0_seq1
comp46883_c0_seq1
comp53802_c0_seq1
comp65767_c1_seq1
comp66203_c0_seq6
comp65461_c0_seq19
comp67842_c2_seq15
comp38927_c0_seq2
comp65638_c0_seq1
comp50750_c0_seq1
comp64542_c0_seq3
comp53184_c0_seq1
comp26980_c0_seq2
comp74597_c0_seq1
comp38895_c0_seq3
comp64731_c0_seq1
comp65834_c0_seq11
comp66992_c0_seq23
comp36527_c0_seq1
comp55727_c0_seq1
comp66675_c0_seq2
comp66688_c0_seq6
comp68453_c1_seq5
comp60464_c0_seq3
comp60768_c1_seq2
comp67315_c0_seq1
comp62941_c1_seq2
comp67378_c0_seq7
comp78595_c0_seq1
comp67548_c0_seq14
comp54521_c0_seq1
comp28841_c0_seq1
comp60887_c0_seq1
comp58411_c0_seq2
comp39935_c0_seq1
comp65432_c0_seq1
comp63764_c0_seq1
comp52768_c0_seq1
comp65511_c0_seq10
comp42857_c0_seq1

comp66265_c0_seq24
comp58769_c0_seq1
comp66862_c0_seq39
comp72009_c0_seq1
comp66862_c0_seq26
comp41306_c0_seq1
comp66645_c1_seq10
comp49103_c0_seq1
comp50429_c0_seq2
comp66848_c0_seq1
comp56722_c0_seq5
comp62105_c0_seq1
comp85908_c0_seq1
comp66296_c0_seq31
comp67721_c3_seq1
comp64170_c0_seq2
comp68156_c1_seq1
comp64637_c0_seq6
comp62145_c0_seq18
comp58764_c0_seq6
comp62063_c0_seq1
comp62880_c0_seq1
comp67702_c3_seq2
comp52352_c0_seq3
comp40152_c0_seq1
comp53411_c0_seq3
comp51364_c0_seq1
comp57630_c0_seq1
comp63974_c0_seq8
comp62938_c0_seq1
comp218384_c0_seq1
comp36222_c0_seq2
comp65002_c0_seq8
comp68368_c0_seq23
comp58357_c0_seq2
comp67247_c1_seq4
comp64731_c0_seq5
comp73300_c0_seq1
comp62338_c0_seq1
comp61434_c0_seq2
comp56096_c0_seq1

comp67635_c1_seq8
comp66602_c0_seq14
comp66498_c0_seq29
comp55702_c0_seq1
comp67586_c0_seq20
comp64699_c0_seq2
comp77221_c0_seq1
comp67093_c1_seq2
comp64531_c0_seq4
comp61588_c0_seq1
comp63198_c0_seq1
comp27422_c0_seq1
comp67141_c0_seq3
comp65935_c3_seq1
comp65834_c0_seq4
comp58758_c0_seq1
comp64492_c0_seq1
comp73634_c0_seq1
comp81818_c0_seq1
comp64983_c0_seq1
comp71372_c0_seq1
comp202345_c0_seq1
comp68660_c0_seq8
comp68766_c0_seq1
comp60420_c1_seq1
comp63283_c0_seq1
comp66101_c0_seq23
comp68813_c0_seq1
comp39932_c0_seq1
comp66862_c0_seq1
comp118640_c0_seq1
comp48074_c0_seq1
comp68453_c1_seq11
comp61344_c0_seq2
comp34150_c0_seq1
comp51753_c0_seq1
comp68004_c0_seq2
comp62113_c0_seq2
comp1385_c1_seq1
comp38126_c1_seq1
comp263528_c0_seq1

comp60679_c0_seq2
comp66862_c0_seq14
comp60661_c0_seq3
comp66602_c0_seq68
comp64308_c1_seq2
comp65910_c0_seq15
comp54257_c0_seq2
comp155541_c0_seq1
comp71825_c0_seq1
comp67466_c0_seq5
comp34623_c0_seq1
comp22170_c0_seq1
comp68462_c0_seq2
comp62064_c0_seq3
comp66498_c0_seq20
comp67136_c0_seq2
comp65799_c1_seq39
comp56143_c0_seq1
comp68593_c2_seq1
comp64020_c0_seq11
comp67053_c0_seq2
comp77389_c0_seq1
comp65037_c0_seq35
comp68192_c0_seq7
comp68453_c1_seq14
comp98478_c0_seq1
comp56144_c0_seq2
comp65910_c0_seq16
comp56972_c0_seq1
comp68453_c1_seq3
comp54266_c0_seq1
comp66928_c0_seq13
comp66836_c0_seq2
comp68635_c2_seq15
comp65037_c0_seq24
comp60491_c0_seq2
comp59307_c1_seq1
comp50339_c0_seq1
comp56554_c0_seq3
comp67686_c2_seq4
comp60185_c0_seq1

comp65511_c0_seq3
comp67164_c1_seq1
comp64722_c0_seq1
comp65517_c0_seq2
comp44119_c0_seq2
comp66602_c0_seq6
comp78474_c0_seq1
comp232016_c0_seq1
comp49931_c0_seq2
comp55248_c0_seq2
comp62941_c0_seq3
comp22862_c0_seq1
comp65236_c0_seq9
comp54662_c0_seq1
comp43565_c0_seq1
comp58737_c0_seq1
comp429991_c0_seq1
comp68220_c0_seq51
comp63974_c0_seq5
comp60043_c0_seq5
comp36780_c0_seq1
comp47748_c0_seq1
comp67079_c0_seq9
comp58065_c0_seq2
comp50049_c0_seq1
comp69111_c0_seq1
comp40100_c0_seq1
comp32920_c0_seq1
comp64647_c0_seq6
comp67767_c1_seq1
comp64161_c0_seq1
comp52058_c0_seq2
comp65158_c0_seq3
comp67698_c0_seq4
comp67247_c1_seq5
comp67171_c0_seq4
comp163003_c0_seq1
comp232495_c0_seq1
comp65511_c0_seq5
comp58683_c0_seq2
comp40447_c0_seq1

comp429604_c0_seq1
comp56388_c0_seq1
comp67548_c0_seq16
comp66371_c0_seq39
comp60326_c0_seq1
comp54473_c0_seq4
comp64637_c0_seq1
comp68453_c1_seq8
comp50611_c0_seq1
comp53749_c0_seq1
comp64910_c0_seq15
comp65820_c1_seq1
comp58958_c0_seq1
comp44436_c0_seq1
comp64391_c0_seq1
comp66862_c0_seq23
comp22312_c0_seq1
comp68192_c0_seq14
comp61713_c0_seq2
comp51364_c1_seq1
comp17139_c0_seq1
comp68214_c0_seq3
comp68572_c3_seq24
comp119126_c0_seq1
comp66926_c0_seq13
comp60816_c0_seq2
comp37652_c0_seq1
comp66101_c0_seq1
comp66375_c0_seq2
comp66928_c0_seq17
comp19962_c0_seq1
comp54438_c0_seq4
comp68660_c0_seq3
comp66548_c0_seq2
comp65654_c0_seq2
comp74415_c0_seq1
comp68749_c0_seq2
comp61284_c0_seq3
comp66498_c0_seq53
comp66688_c0_seq2
comp50030_c0_seq4

comp67709_c1_seq4
comp60466_c0_seq1
comp36589_c0_seq1
comp49672_c0_seq1
comp64731_c0_seq8
comp48557_c0_seq1
comp68660_c0_seq2
comp62913_c0_seq1
comp59529_c0_seq2
comp67315_c0_seq2
comp66862_c0_seq30
comp66440_c0_seq2
comp67007_c0_seq9
comp63764_c0_seq2
comp66318_c1_seq2
comp65047_c0_seq1
comp68762_c0_seq10
comp67698_c0_seq5
comp71342_c0_seq1
comp38126_c0_seq1
comp39059_c0_seq1
comp64308_c0_seq1
comp61791_c0_seq1
comp67217_c0_seq1
comp61219_c0_seq6
comp56904_c0_seq2
comp67967_c0_seq1
comp66861_c0_seq3
comp59815_c0_seq3
comp66498_c0_seq31
comp71510_c0_seq1
comp53784_c0_seq4
comp27431_c0_seq1
comp56736_c0_seq1
comp38449_c1_seq1
comp67954_c1_seq12
comp68795_c0_seq9
comp61263_c0_seq8
comp37414_c1_seq1
comp66296_c0_seq9
comp64056_c0_seq1

comp62809_c1_seq1
comp64910_c0_seq19
comp65236_c0_seq4
comp37554_c0_seq1
comp67325_c0_seq35
comp49750_c0_seq1
comp65442_c0_seq1
comp62557_c0_seq5
comp79150_c0_seq1
comp65566_c0_seq4
comp65022_c1_seq8
comp50380_c0_seq1
comp66926_c0_seq7
comp304022_c0_seq1
comp58065_c0_seq3
comp34829_c0_seq1
comp66498_c0_seq3
comp60441_c0_seq1
comp66699_c0_seq7
comp66862_c0_seq44
comp70278_c0_seq1
comp32792_c0_seq1
comp98101_c0_seq1
comp63062_c0_seq4
comp65565_c0_seq1
comp66132_c0_seq1
comp68433_c0_seq28
comp68006_c0_seq5
comp61167_c0_seq3
comp66234_c0_seq9
comp66624_c2_seq1
comp61125_c1_seq1
comp66675_c0_seq3
comp80920_c0_seq1
comp65548_c0_seq3
comp61902_c1_seq1
comp59139_c0_seq1
comp66391_c0_seq20
comp74870_c0_seq1
comp67175_c0_seq2
comp67967_c0_seq10

comp58164_c0_seq1
comp61344_c0_seq3
comp68572_c3_seq12
comp65935_c2_seq10
comp57910_c0_seq2
comp67566_c0_seq3
comp53338_c0_seq1
comp65148_c0_seq12
comp68453_c1_seq4
comp68214_c0_seq1
comp65669_c0_seq2
comp58578_c1_seq2
comp21960_c0_seq1
comp30882_c0_seq1
comp55311_c0_seq2
comp64071_c0_seq9
comp62539_c0_seq4
comp36222_c0_seq1
comp66893_c0_seq1
comp66136_c0_seq8
comp65432_c0_seq10
comp67682_c0_seq3
comp59564_c0_seq1
comp67967_c0_seq4
comp64195_c0_seq3
comp67171_c0_seq1
comp30785_c0_seq1
comp67253_c0_seq8
comp22366_c0_seq1
comp61425_c0_seq3
comp77617_c0_seq1
comp36268_c0_seq2
comp64731_c0_seq3
comp49690_c0_seq1
comp68770_c0_seq1
comp58065_c0_seq1
comp66992_c0_seq8
comp158713_c0_seq1
comp64576_c0_seq1
comp66290_c0_seq30
comp55377_c0_seq2

comp67417_c0_seq5
comp61219_c0_seq9
comp60677_c0_seq4
comp55243_c0_seq1
comp54573_c0_seq2
comp55076_c0_seq1
comp56995_c1_seq1
comp41481_c0_seq1
comp62843_c0_seq2
comp62877_c1_seq1
comp54425_c0_seq2
comp4279_c0_seq1
comp68474_c1_seq10
comp67655_c0_seq22
comp64520_c0_seq1
comp33561_c0_seq2
comp68207_c0_seq3
comp58983_c0_seq1
comp184096_c0_seq1
comp67164_c1_seq17
comp70399_c0_seq1
comp62374_c0_seq1
comp68792_c0_seq6
comp83948_c0_seq1
comp49304_c0_seq1
comp52244_c0_seq1
comp61249_c0_seq1
comp64845_c0_seq2
comp21570_c0_seq1
comp41527_c0_seq1
comp41481_c1_seq1
comp50527_c0_seq2
comp68397_c0_seq19
comp50123_c0_seq1
comp22317_c1_seq1
comp64729_c0_seq3
comp66675_c0_seq5
comp64168_c1_seq19
comp67491_c0_seq8
comp58381_c0_seq1
comp65511_c0_seq1

comp52058_c0_seq3
comp78420_c0_seq1
comp54635_c0_seq7
comp66979_c0_seq10
comp18936_c0_seq1
comp67932_c2_seq11
comp65511_c0_seq2
comp56103_c0_seq1
comp23480_c0_seq1
comp63357_c0_seq1
comp51598_c0_seq3
comp63283_c1_seq5
comp65918_c0_seq1
comp68693_c0_seq30
comp21984_c0_seq2
comp68006_c0_seq1
comp67566_c0_seq2
comp47748_c0_seq2
comp67274_c0_seq1
comp49042_c0_seq1
comp36456_c0_seq1
comp65763_c1_seq11
comp66984_c0_seq8
comp63252_c0_seq8
comp67553_c0_seq1
comp49997_c0_seq2
comp64470_c0_seq3
comp55324_c0_seq5
comp63502_c0_seq4
comp54902_c0_seq3
comp68571_c0_seq29
comp67636_c1_seq1
comp65214_c0_seq1
comp64464_c2_seq2
comp47748_c0_seq3
comp2189_c0_seq1
comp56538_c0_seq1
comp68411_c1_seq1
comp54292_c0_seq2
comp67548_c0_seq18
comp38668_c0_seq1

comp68037_c0_seq9
comp64183_c0_seq10
comp66326_c0_seq6
comp63599_c0_seq1
comp62219_c0_seq3
comp58847_c0_seq1
comp56916_c0_seq1
comp63858_c0_seq1
comp65935_c2_seq3
comp64267_c0_seq2
comp28438_c0_seq1
comp50948_c0_seq1
comp58004_c0_seq2
comp51758_c0_seq1
comp68572_c3_seq3
comp66716_c0_seq3
comp67315_c0_seq3
comp55257_c0_seq2
comp64104_c0_seq5
comp64293_c0_seq6
comp62394_c0_seq4
comp265972_c0_seq1
comp62531_c0_seq8
comp37567_c0_seq1
comp68590_c1_seq3
comp66675_c0_seq9
comp64169_c0_seq3
comp184748_c0_seq1
comp21627_c0_seq2
comp66805_c0_seq1
comp61664_c1_seq4
comp63387_c0_seq1
comp68698_c1_seq26
comp68192_c0_seq8
comp67825_c0_seq1
comp261334_c0_seq1
comp65763_c1_seq4
comp68192_c0_seq15
comp54694_c0_seq2
comp48538_c0_seq1
comp65143_c2_seq1

comp64900_c0_seq1
comp56506_c0_seq1
comp68764_c1_seq1
comp57732_c0_seq1
comp64078_c0_seq1
comp12562_c0_seq1
comp61263_c0_seq2
comp55032_c0_seq1
comp65037_c0_seq6
comp66132_c1_seq1
comp64281_c1_seq4
comp57555_c0_seq5
comp52553_c0_seq2
comp66599_c0_seq2
comp62741_c0_seq1
comp60177_c0_seq2
comp61434_c0_seq1
comp64267_c0_seq1
comp67682_c0_seq2
comp66101_c0_seq6
comp110932_c0_seq1
comp64303_c0_seq15
comp66675_c0_seq12
comp55970_c0_seq2
comp68572_c3_seq25
comp67003_c0_seq2
comp38076_c0_seq1
comp23695_c0_seq1
comp66468_c0_seq6
comp56560_c0_seq1
comp71727_c0_seq1
comp65991_c1_seq2
comp63844_c0_seq2
comp62145_c0_seq15
comp53530_c0_seq2
comp66814_c1_seq5
comp56201_c0_seq1
comp63713_c0_seq2
comp66288_c0_seq6
comp66371_c0_seq11
comp57101_c0_seq1

comp62206_c0_seq1
comp444385_c0_seq1
comp56778_c0_seq1
comp43619_c0_seq1
comp17454_c0_seq1
comp68523_c0_seq4
comp53083_c0_seq2
comp335969_c0_seq1
comp67450_c0_seq4
comp72828_c0_seq1
comp52754_c0_seq1
comp65511_c0_seq13
comp33460_c0_seq1
comp61283_c0_seq1
comp22052_c0_seq1
comp67210_c0_seq29
comp65722_c0_seq2
comp65834_c0_seq10
comp64643_c0_seq7
comp68572_c2_seq1
comp327696_c0_seq1
comp30693_c0_seq1
comp15182_c0_seq1
comp66624_c0_seq10
comp121447_c0_seq1
comp67948_c0_seq2
comp50274_c0_seq2
comp62041_c0_seq2
comp67635_c1_seq15
comp20241_c0_seq1
comp81725_c0_seq1
comp67253_c0_seq2
comp56667_c0_seq4
comp65717_c0_seq4
comp62941_c0_seq2
comp10040_c0_seq1
comp62877_c0_seq3
comp65743_c0_seq3
comp35869_c0_seq1
comp64424_c0_seq1
comp64699_c0_seq8

comp30973_c0_seq1
comp55495_c0_seq3
comp55325_c0_seq1
comp63241_c0_seq2
comp66926_c0_seq2
comp517042_c0_seq1
comp68453_c1_seq15
comp68474_c1_seq7
comp68474_c1_seq8
comp74757_c0_seq1
comp76688_c0_seq1
comp49156_c0_seq1
comp68220_c0_seq30
comp59952_c0_seq3
comp67247_c1_seq10
comp47644_c0_seq2
comp83717_c0_seq1
comp68474_c0_seq1
comp67548_c0_seq4
comp65037_c0_seq43
comp65897_c0_seq1
comp67655_c0_seq2
comp64168_c1_seq8
comp64447_c0_seq2
comp64182_c0_seq1
comp66862_c0_seq21
comp61263_c0_seq5
comp64170_c0_seq1
comp56677_c0_seq2
comp54431_c0_seq6
comp56576_c0_seq4
comp66716_c0_seq2
comp62080_c0_seq15
comp66370_c0_seq1
comp21491_c0_seq1
comp41417_c0_seq1
comp37414_c0_seq1
comp57770_c0_seq6
comp38449_c2_seq1
comp65669_c0_seq5
comp20363_c0_seq1

comp51838_c0_seq3
comp65767_c0_seq1
comp62751_c0_seq2
comp63062_c0_seq2
comp56916_c1_seq1
comp2066_c0_seq1
comp54471_c0_seq1
comp67378_c0_seq10
comp62219_c0_seq2
comp63672_c0_seq3
comp68474_c1_seq4
comp58448_c0_seq2
comp54878_c0_seq1
comp44083_c0_seq1
comp65275_c0_seq3
comp57499_c0_seq1
comp67721_c3_seq10
comp67315_c0_seq5
comp55698_c0_seq1
comp68571_c0_seq8
comp552193_c0_seq1
comp66862_c0_seq46
comp63474_c0_seq1
comp62394_c0_seq3
comp49597_c0_seq1
comp66862_c0_seq29
comp20959_c0_seq1
comp68453_c1_seq7
comp66924_c0_seq2
comp67718_c0_seq1
comp68474_c1_seq9
comp71905_c0_seq1
comp52054_c0_seq1
comp65515_c0_seq10
comp67461_c0_seq13
comp68573_c1_seq4
comp66862_c0_seq41
comp66101_c0_seq20
comp49591_c0_seq1
comp47614_c0_seq1
comp67664_c0_seq9

comp38810_c0_seq1
comp42291_c0_seq2
comp68761_c0_seq1
comp68059_c0_seq2
comp44461_c0_seq1
comp62878_c0_seq3
comp63241_c0_seq1
comp67840_c2_seq9
comp63239_c0_seq1
comp57855_c0_seq1
comp39624_c0_seq1
comp42627_c0_seq1
comp67655_c0_seq8
comp48184_c0_seq1
comp68660_c0_seq10
comp68816_c0_seq1
comp66862_c0_seq6
comp66862_c0_seq10
comp65610_c0_seq1
comp60883_c0_seq2
comp287798_c0_seq1
comp37158_c0_seq1
comp38030_c0_seq1
comp63108_c0_seq5
comp68644_c1_seq15
comp64153_c0_seq3
comp110759_c0_seq1
comp62111_c2_seq4
comp67164_c1_seq24
comp67210_c0_seq47
comp42460_c0_seq1
comp65445_c0_seq2
comp50845_c0_seq1
comp45413_c0_seq1
comp54723_c0_seq1
comp62412_c0_seq4
comp60816_c0_seq3
comp50295_c0_seq1
comp64977_c0_seq1
comp22114_c0_seq1
comp65753_c0_seq4

comp67312_c0_seq4
comp51072_c0_seq2
comp65511_c0_seq8
comp66862_c0_seq12
comp68025_c1_seq9
comp57910_c0_seq1
comp64889_c0_seq2
comp61721_c0_seq1
comp64056_c0_seq6
comp36010_c0_seq1
comp61902_c0_seq1
comp61639_c0_seq1
comp76526_c0_seq1
comp58782_c0_seq2
comp35870_c0_seq1
comp68572_c3_seq9
comp61675_c0_seq1
comp68795_c0_seq7
comp33070_c0_seq1
comp68526_c0_seq1
comp65050_c0_seq10
comp40440_c1_seq1
comp65327_c0_seq4
comp21779_c0_seq1
comp66801_c0_seq10
comp34051_c0_seq1
comp68360_c0_seq3
comp68143_c2_seq2
comp634986_c0_seq1
comp52369_c1_seq2
comp54312_c0_seq1
comp22092_c0_seq1
comp62557_c0_seq1
comp21842_c0_seq1
comp68220_c0_seq12
comp21625_c0_seq1
comp65483_c0_seq2
comp66692_c1_seq2
comp68572_c3_seq15
comp62145_c0_seq7
comp68156_c0_seq1

comp17171_c0_seq1
comp68192_c0_seq16
comp23582_c0_seq1
comp60350_c0_seq2
comp56554_c0_seq2
comp49556_c0_seq1
comp61508_c0_seq1
comp67164_c1_seq4
comp66101_c0_seq4
comp67548_c0_seq19
comp64492_c0_seq2
comp64303_c0_seq3
comp330564_c0_seq1
comp219387_c0_seq1
comp32423_c0_seq1
comp68368_c0_seq67
comp66782_c0_seq4
comp68128_c0_seq6
comp53411_c0_seq5
comp62810_c1_seq7
comp68795_c0_seq3
comp65511_c0_seq15
comp54334_c0_seq2
comp32989_c0_seq1
comp62002_c0_seq3
comp59035_c0_seq2
comp67378_c0_seq18
comp57039_c0_seq1
comp36327_c0_seq1
comp66511_c0_seq2
comp49797_c1_seq2
comp20715_c0_seq1
comp61743_c0_seq1
comp56667_c0_seq2
comp48570_c0_seq1
comp61703_c0_seq3
comp68192_c0_seq6
comp68572_c3_seq16
comp54823_c0_seq3
comp62145_c0_seq8
comp65158_c0_seq9

comp58357_c0_seq1
comp39461_c0_seq1
comp62220_c0_seq7
comp49974_c0_seq3
comp66675_c0_seq7
comp66101_c0_seq27
comp60932_c0_seq2
comp64637_c0_seq5
comp71920_c0_seq1
comp40917_c0_seq1
comp28170_c0_seq1
comp66805_c0_seq3
comp154202_c0_seq1
comp67553_c0_seq6
comp61762_c0_seq11
comp66326_c0_seq1
comp58249_c0_seq1
comp22317_c0_seq1
comp23464_c0_seq1
comp64170_c0_seq3
comp65397_c0_seq2
comp67204_c0_seq2
comp65386_c0_seq1
comp49422_c0_seq1