

University of Denver

Digital Commons @ DU

---

Electronic Theses and Dissertations

Graduate Studies

---

1-1-2015

## Gene Expression in the Choanoderm

Jesús Federico Peña  
*University of Denver*

Follow this and additional works at: <https://digitalcommons.du.edu/etd>



Part of the [Biology Commons](#), and the [Genetics and Genomics Commons](#)

---

### Recommended Citation

Peña, Jesús Federico, "Gene Expression in the Choanoderm" (2015). *Electronic Theses and Dissertations*. 513.

<https://digitalcommons.du.edu/etd/513>

This Thesis is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact [jennifer.cox@du.edu](mailto:jennifer.cox@du.edu), [dig-commons@du.edu](mailto:dig-commons@du.edu).

# 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

Introduction	1
Chapter 1: Differential gene expression analysis of the choanoderm	6
Introduction	6
Methods	8
Results	15
Discussion	31
Chapter 2: Optimizing whole-mount in situ hybridization for <i>Ephydatia muelleri</i> tissue	37
Introduction	37
Methods	39
Results	48
Discussion	61
Literature Cited	67
Appendix 1	79

## List of Tables

Table 1.1 Hydroxyurea treatment plan	19
Table 1.2 CLC Genomics Workbench read mapping summary for control samples	21
Table 1.3 CLC Genomics Workbench read mapping summary for HU-treated samples	22
Table 1.4 Differentially expressed clusters at different false discovery rates	24
Table 1.5 Gene products restricted to sponges and choanoflagellates	28
Table 1.6 Downregulated candidate genes identified by BLAST2GO (e-value < 1.0e-10)	30
Table 2.1 List of primers and amplicon size	51
Table 2.2 Comparison of Protocols 1-3 and the Optimal Protocol	60

## List of Figures

Figure i-1 Choanoflagellate, choanoderm, and phylogeny of animals	3
Figure i-2 Gemmules and the body plan of <i>Ephydatia muelleri</i>	4
Figure 1.1 Untreated and hydroxyurea treated sponge morphology	20
Figure 1.2 Read mapping summary of control samples	21
Figure 1.3 Read mapping summary of hydroxyurea samples	22
Figure 1.4 MDS Plot for count data	23
Figure 1.5 Microarray plot for differentially expressed gene clusters	24
Figure 1.6 Mean normalized counts of microvillar genes in control and hydroxyurea treated sponges	25
Figure 1.7 Mean normalized counts of ciliary genes in control and hydroxyurea treated sponges	26
Figure 1.8 Mean normalized counts of genes associated with the classical cadherin catenin adhesion complex in control and hydroxyurea treated sponges	27
Figure 1.9 Mean normalized counts of genes identified only in sponges and choanoflagellates	29
Figure 2.1 VLGR1 fragment amplified from <i>E. muelleri</i> cDNA library	51
Figure 2.2 <i>Usherin</i> fragment amplified from <i>E. muelleri</i> cDNA library.	52
Figure 2.3 <i>Cadherin 23</i> fragment amplified from <i>E. muelleri</i> cDNA library	52
Figure 2.4 <i>Choanogene</i> fragment amplified from <i>E. muelleri</i> cDNA library	53
Figure 2.5 <i>Annexin</i> fragment amplified from <i>E. muelleri</i> cDNA library	53

Figure 2.6 Products of PCR on 1% agarose gel	54
Figure 2.7 Schematic of inserts relative to RNA polymerase promoters in pCR®II-TOPO dual promoter vector	54
Figure 2.8 Colony screen PCR visualized on 1% agarose gel	55
Figure 2.9 Plasmid digest visualized on 1% agarose gel	55
Figure 2.10 Quality of DIG-labelled RNA probes	56
Figure 2.11 Whole-mount <i>in situ</i> hybridization using Protocol 2	56
Figure 2.12 Whole-mount <i>in situ</i> hybridization using Protocol 1	57
Figure 2.13 Whole-mount <i>in situ</i> hybridization using Protocol 3	58
Figure 2.14 Whole-mount <i>in situ</i> hybridization treated with varying concentrations of RNase A	59

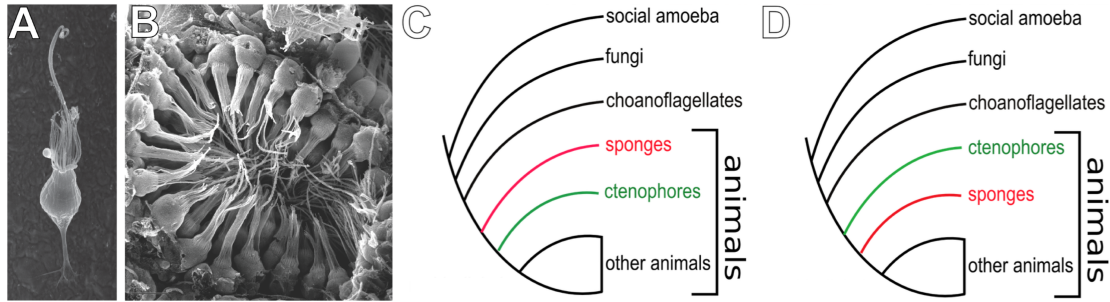


## **Introduction**

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).

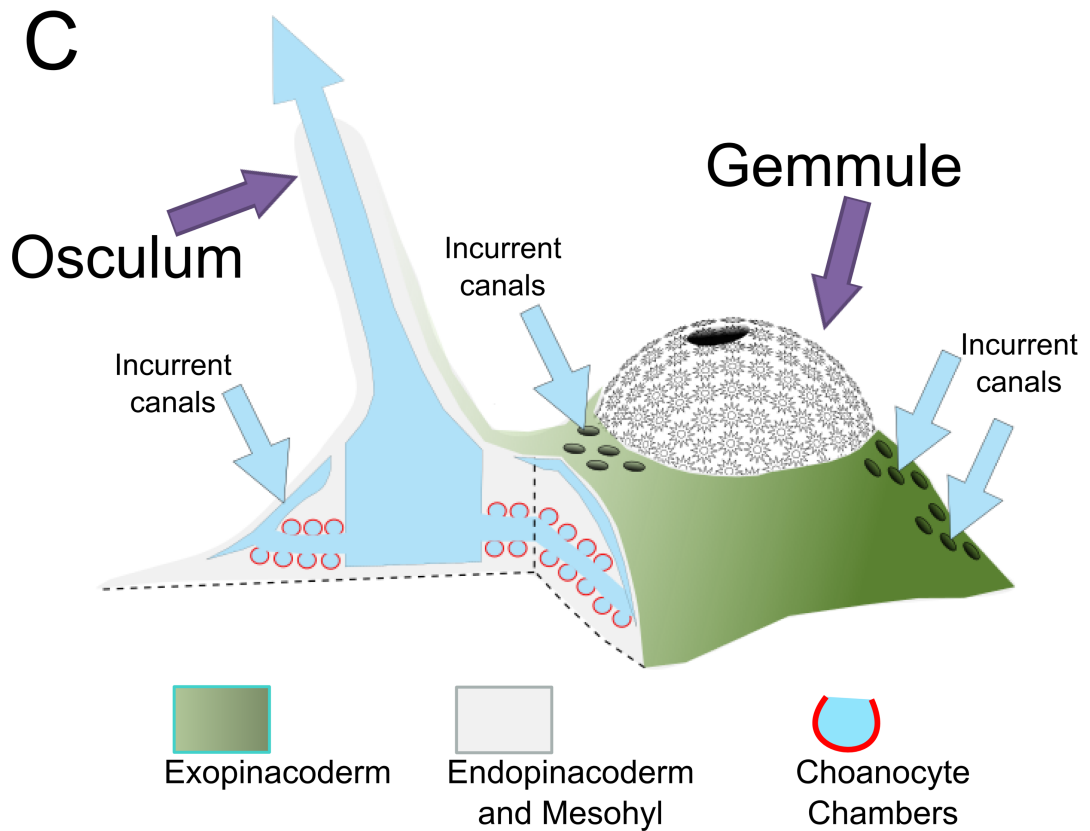
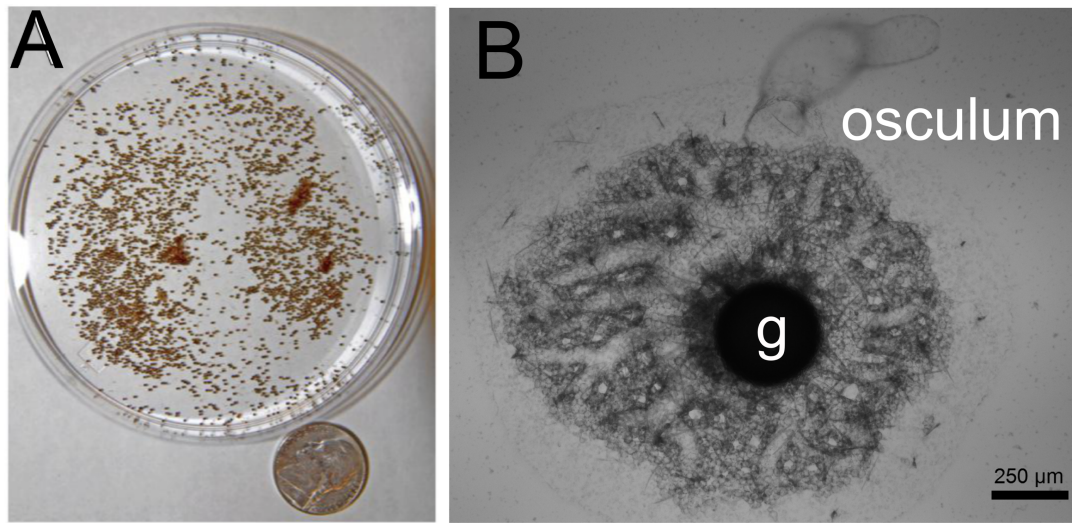
Some of the first clues about the evolutionary link between animals and choanoflagellates came from comparisons of choanoflagellate morphology to choanocytes, the feeding cells of sponges (Fig. i-1B; James-Clark, 1867, 1871; Kent, 1878). The sponge body plan stands out as an outlier among modern animals (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 choanocytes 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 4°C. (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 and had 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  $0.05 - P_{\text{error}}$ . 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 (Hemrich 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.



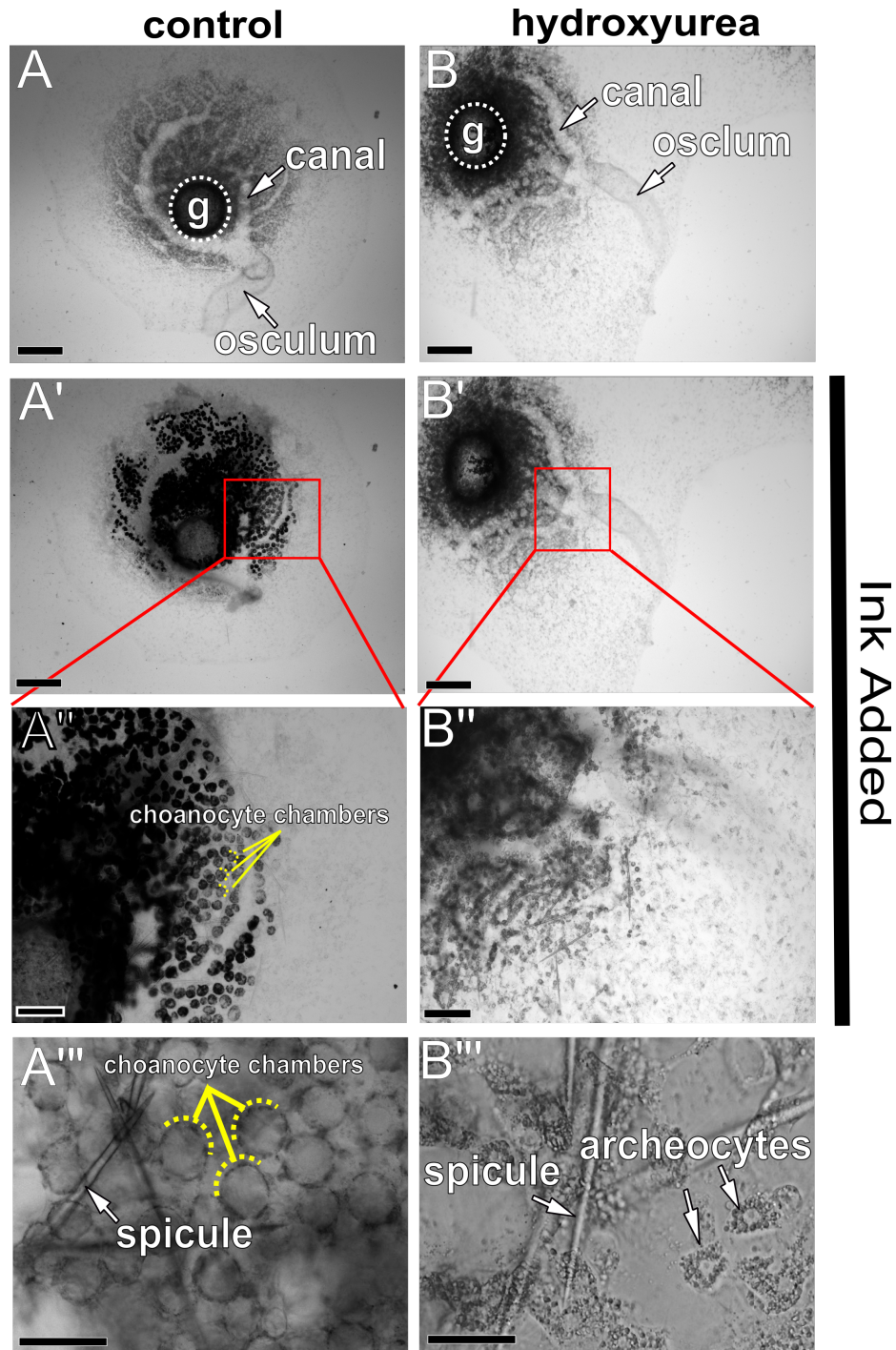
To classify and evaluate the biological categories of the downregulated transcripts, we used BLAST2GO (Conesa and Götz 2008; Conesa et al. 2005; Götz et al. 2008; Götz et al. 2011). This approach relies on performing a BLAST search of all the candidate sequences. Following the BLAST search, BLAST2GO 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.

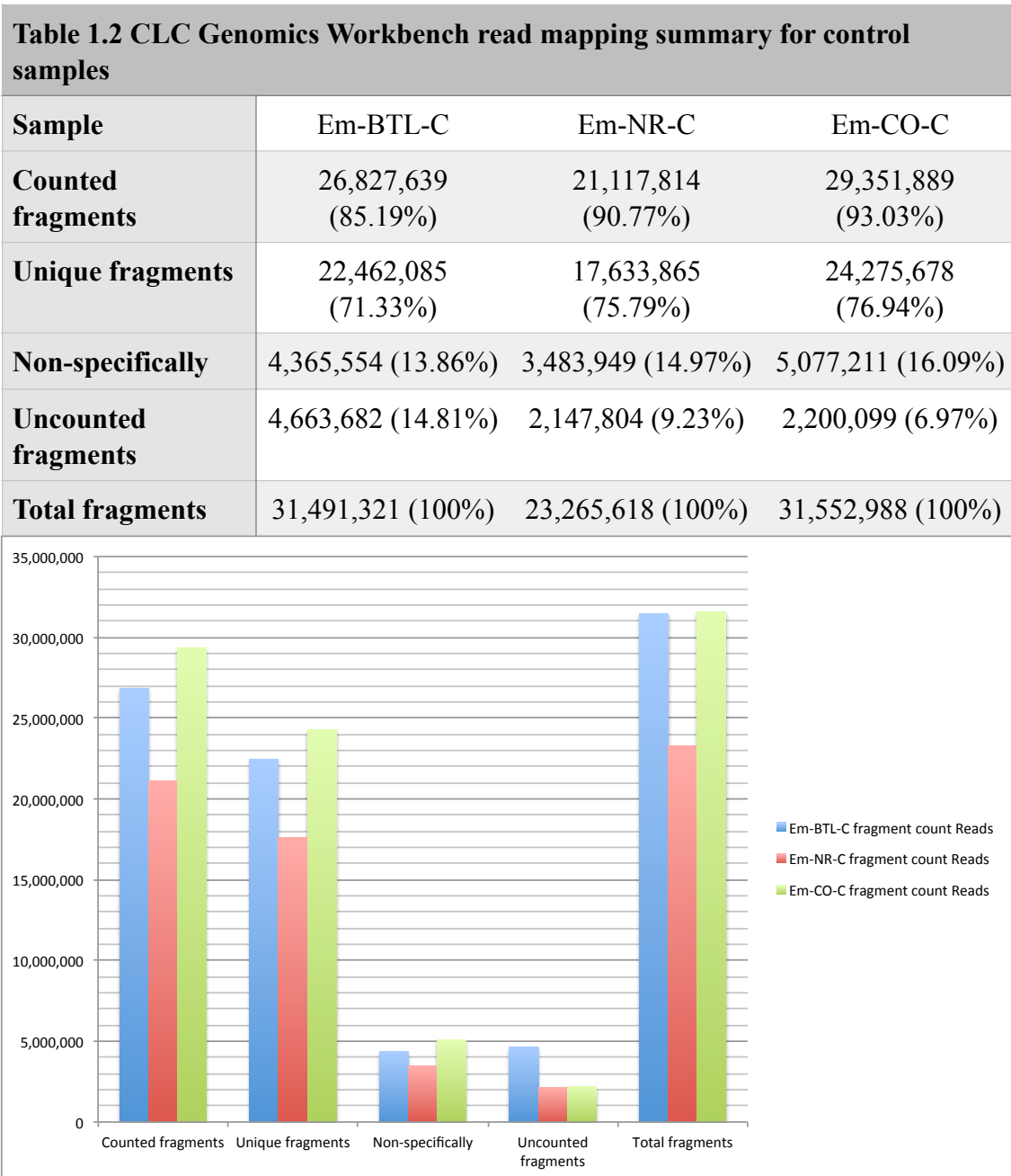
Additionally, 24 transcript clusters were identified as restricted to choanoflagellates and sponges through phmmer searches (Table 1.5). A number of these genes have signal peptides as well as other domains implicated in signaling. One gene (m.236078) has a cadherin domain. Of the 24 genes identified, 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
<b>Indicator</b>	Plated		Hatched		...	Choanocytes detected	...	
<b>Control</b>		Plated		Hatched	...		...	Harvested
<b>Hydroxyurea</b>		Plated		Hatched	...	HU added	...	Harvested



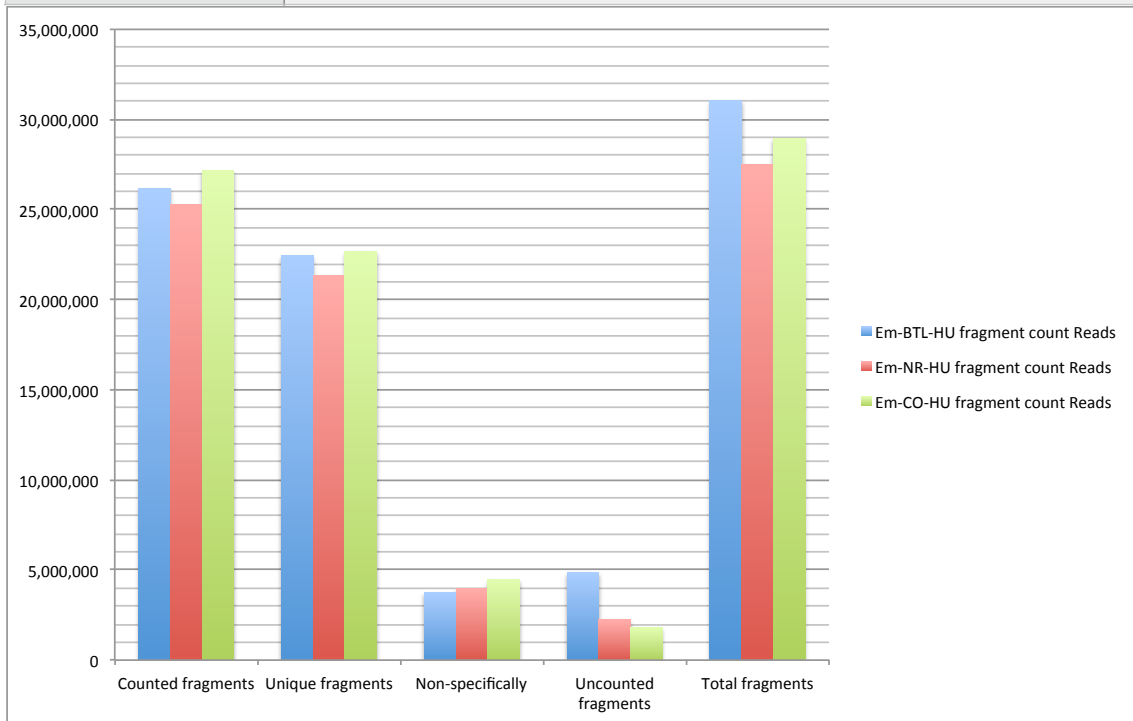
**Figure 1.1** Untreated and hydroxyurea treated sponge morphology. Control (A) and hydroxyurea treated (B) sponges develop water canals and an osculum. Ink shows choanocyte chambers in control sponges (A', A'') but not in HU-treated sponges (B', B''). Both control (A''') and HU-treated (B''') sponges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



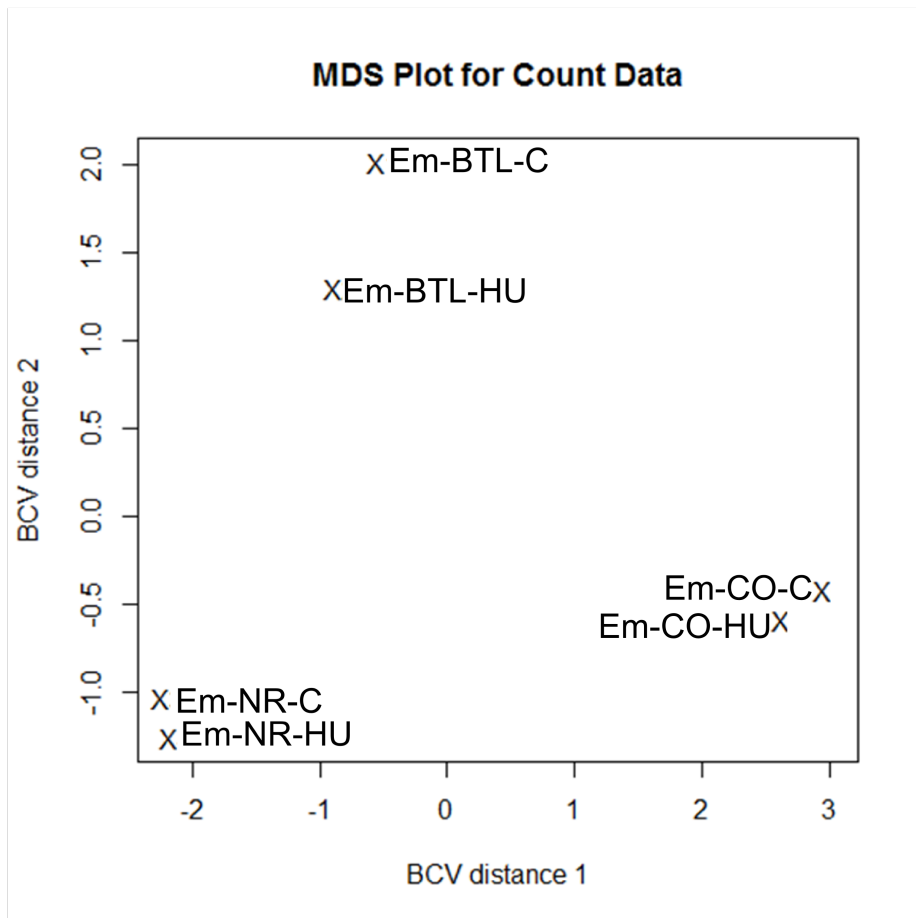
**Figure 1.2** Read mapping summary of control samples.

**Table 1.3 CLC Genomics Workbench read mapping summary for HU-treated samples**

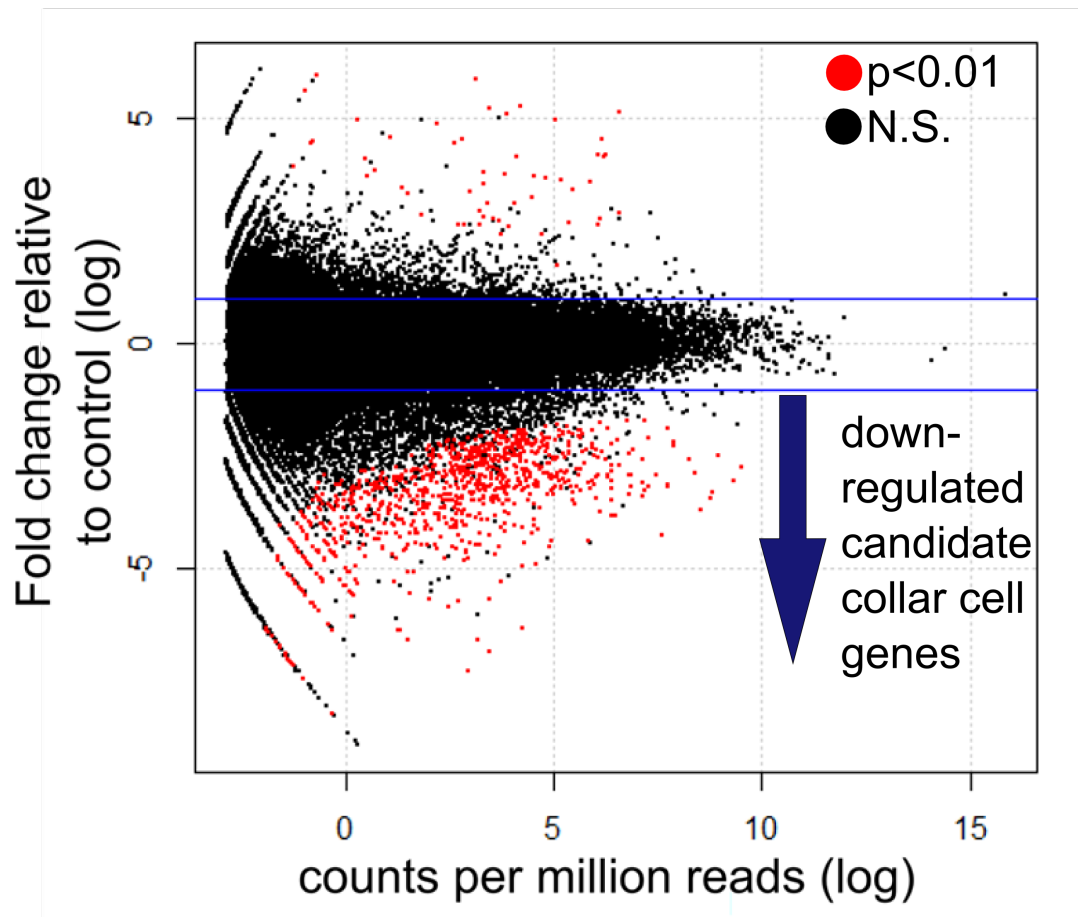
Sample	Em-BTL-HU	Em-NR-HU	Em-CO-HU
<b>Counted fragments</b>	26,182,932 (84.33%)	25,290,543 (91.87%)	27,162,188 (93.75%)
<b>Unique fragments</b>	22,438,826 (72.27%)	21,329,957 (77.48%)	22,686,820 (78.31%)
<b>Non-specifically</b>	3,744,106 (12.06%)	3,960,586 (14.39%)	4,475,368 (15.45%)
<b>Uncounted fragments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



**Figure 1.3** Read mapping summary of hydroxyurea samples.



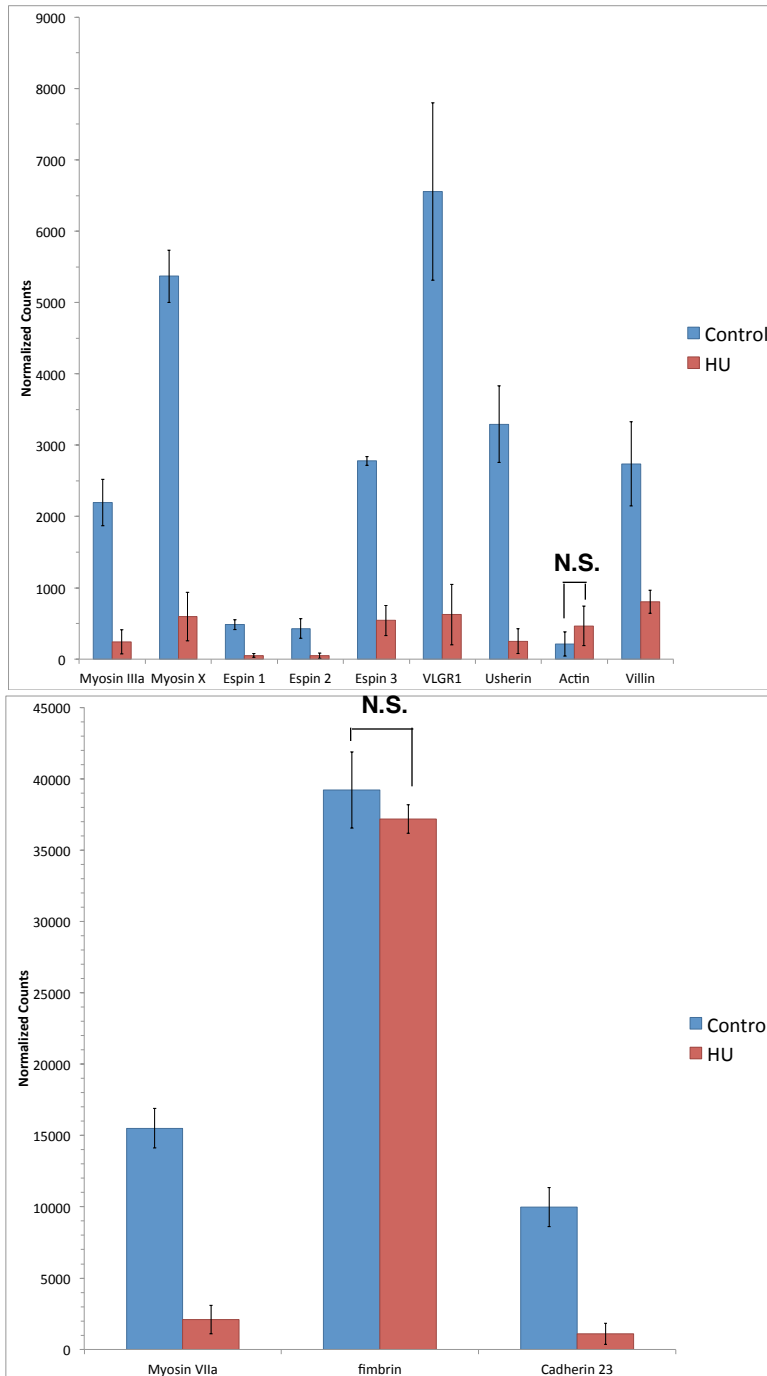
**Figure 1.4** The multi-dimensional scaling plot projects sample similarity in 2-dimensions. Em-CO: Red Rock Lake population; Em-BTL: Beavertail Lake population; Em-NR: Nanaimo River population. C: control; HU: hydroxyurea treated



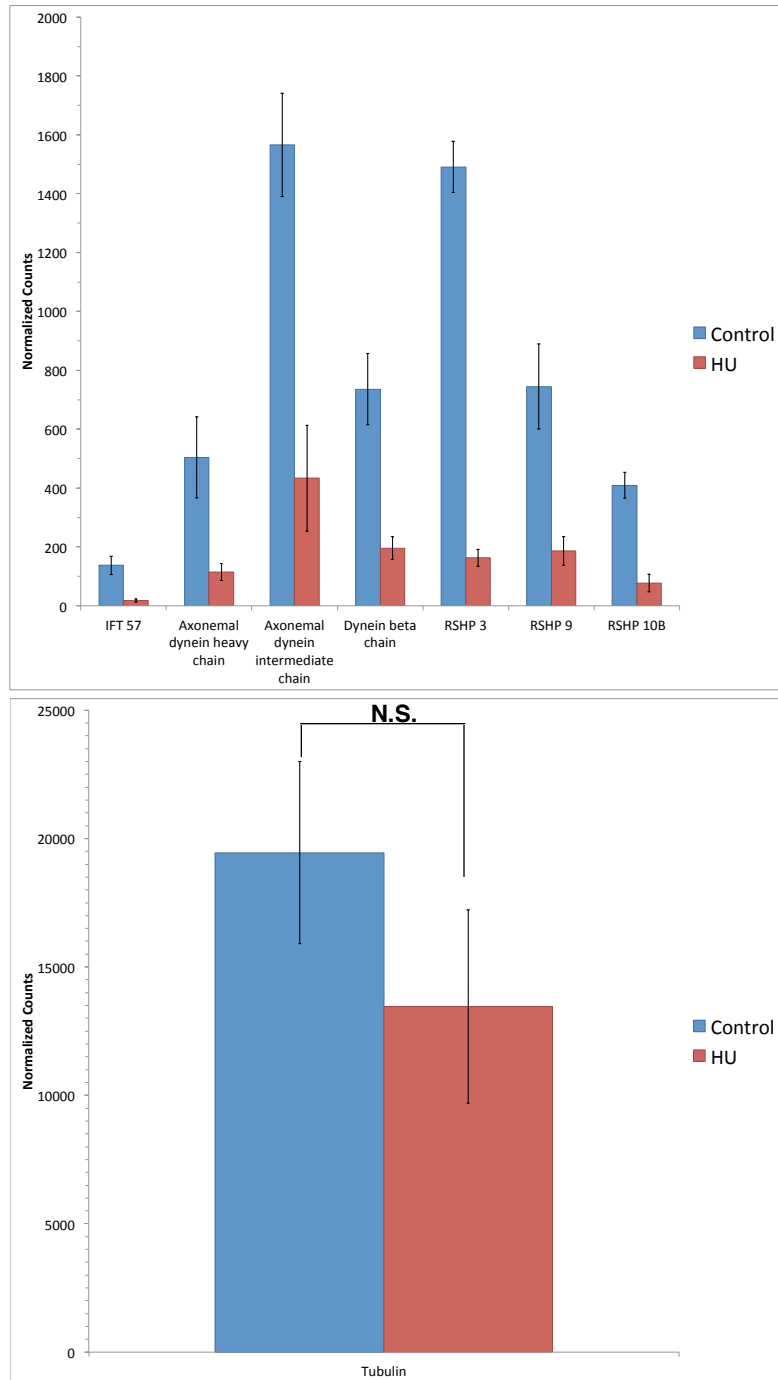
**FIGURE 1.5** Microarray plot for differentially expressed gene clusters. The microarray plot shows the relationship between counts per million reads and fold-change across the genes clusters. Differentially expressed gene clusters are shown in red ( $P < 0.01$ ). Non-differentially expressed genes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

<b>Table 1.4 Differentially expressed clusters at different false discovery rates</b>		
	FDR < 1%	FDR < 0.1%
<b>Upregulated clusters</b>	54	37
<b>Downregulated clusters</b>	879	481
<b>NS</b>	77,018	77,433

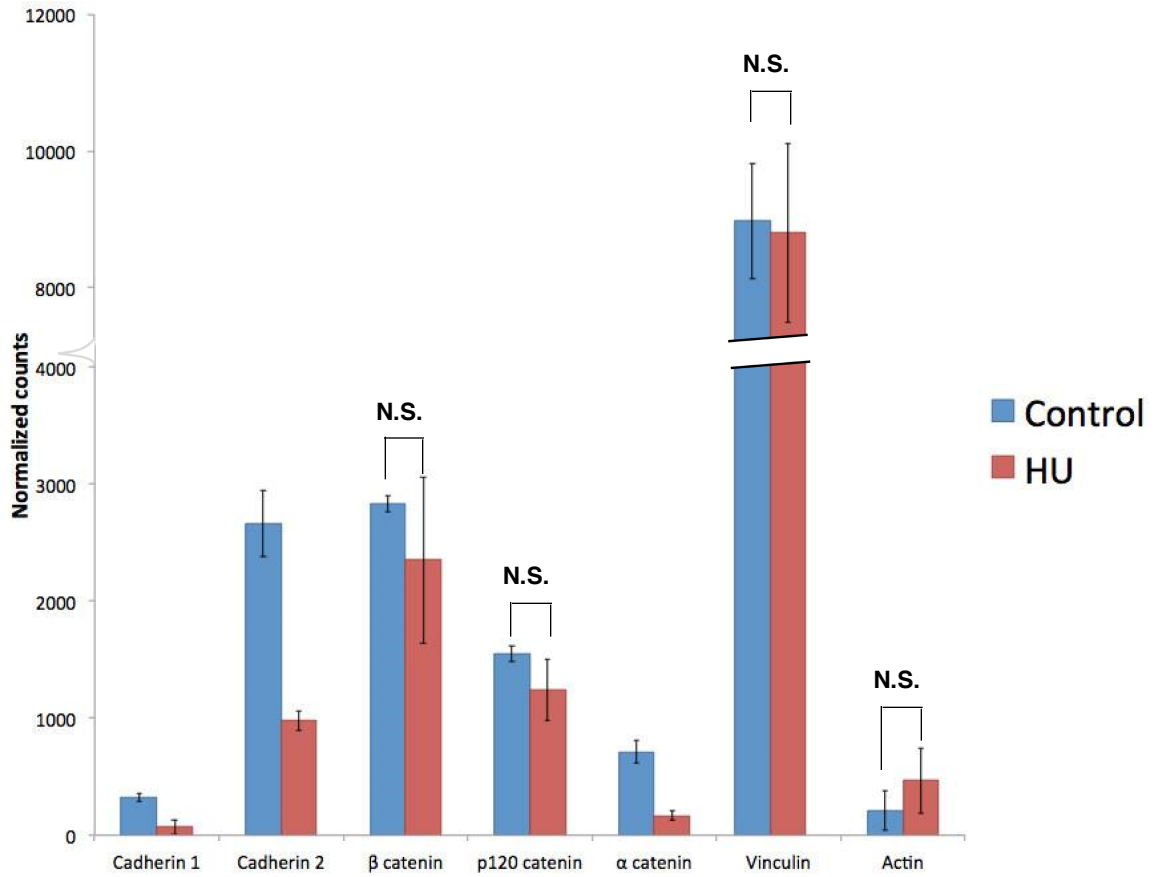




**Figure 1.6** Mean normalized counts of microvillar genes in control and hydroxyurea treated sponges. Except where stated, all genes are significantly differentially expressed in HU-treated sponges. Error bars represent standard deviation from the mean.

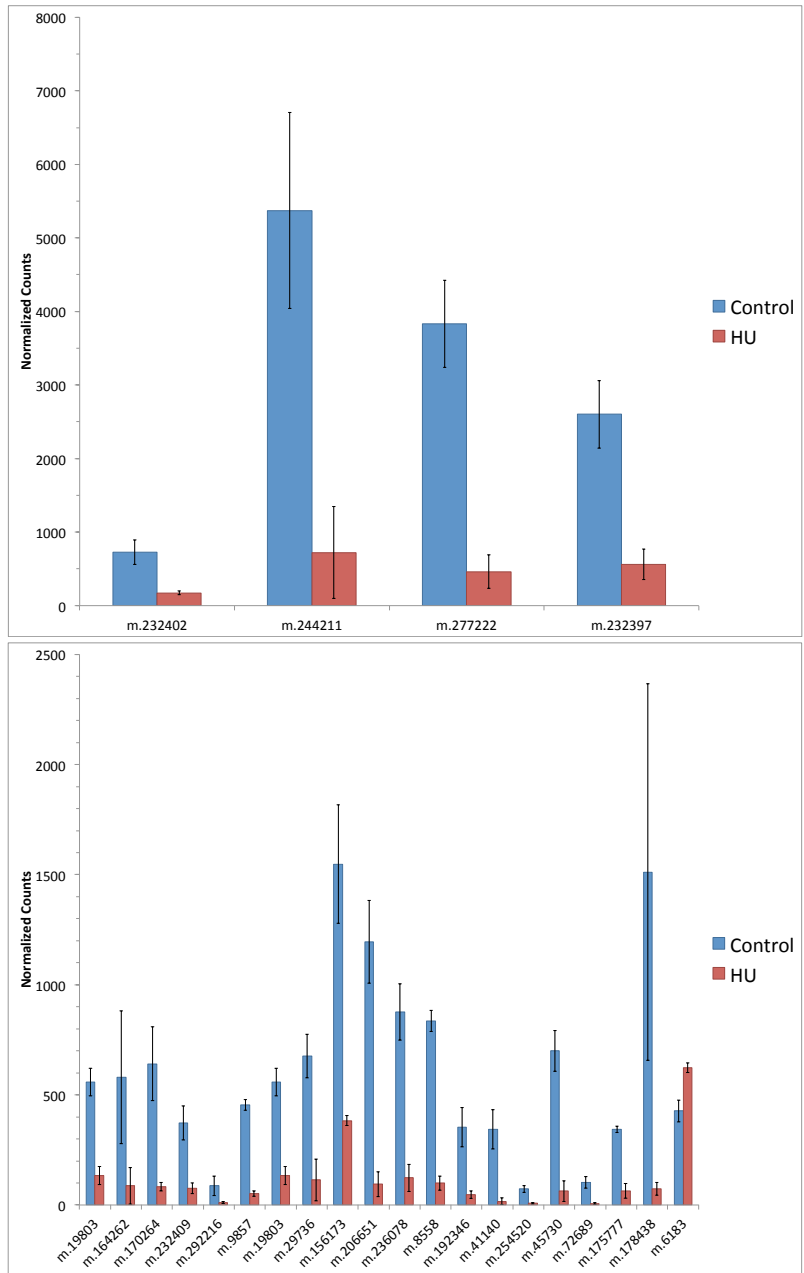


**Figure 1.7** Mean normalized counts of ciliary genes in control and hydroxyurea treated sponges. Except where stated, all genes are significantly differentially expressed in HU-treated sponges. Error bars represent standard deviation from the mean.



**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.

<b>Table 1.5 Gene products restricted to sponges and choanoflagellates</b>		
<b>Accession</b>	<b>Length</b>	<b>Domain architecture</b>
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)**

<b>cilia/flagella-specific</b>	IFT-52/57/81/105/172, CCDC-39/40/113, RSPH-3/9/10B, NME5, tectorin, PITG-05447, ODF-3, axonemal dynein, Cep89, PTPRQ
<b>other microtubule associated</b>	GAS-8, Serine/Threonine-protein kinase Nek8, TTC-16/30A/28, Kif-15, protein polyglycyclase TTL 10, kinesin-like, Futsch, battenin, CCDC-87, KIF25, kif3b, kif5b, alp11
<b>microvillar specific</b>	VLGR1, spectrin, usherin, espin, whirlin, calmodulin, SLC9A3R1
<b>other actin associated</b>	Myosin-VIIa, myosin IIIa, myosin-X, filamin-a/b/c, SHANK3, socius, FGD6, INF2, ankyrin, WASp, EPB-41, zeta-sarcoglycan, SWAP70, inverted-formin, talin
<b>cadherin/catenin related</b>	Protocadherin Fat-4, Lefty1, coherin, alpha-catenin, cadherin 23
<b>other adhesion related</b>	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
<b>polarity (apical/basal &amp; planar)</b>	Prickle 2, crumbs, flamingo, RTK-like orphan receptor, alp11, DLG1
<b>epithelial development</b>	Hensin, Cfap-57, Plexin-A2, fibrocystin-L, NOTCH1, inversin, TGM1, TBATA, TBX2b, KRIT1, MIB-1, malcavernin
<b>epithelial/mesenchyme transition</b>	HUNK, EGF-like, invadolysin, krueppel-like, SCUBE2, zinc finger transcription factor
<b>known epithelial expression</b>	Anoctamin, Krit-1 trichohyalin-like
<b>Cell proliferation/differentiation</b>	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  $\beta$ -catenin in a yeast two-hybrid screen (Nichols et al. 2012). Here we have also identified  $\alpha$ -catenin and other adhesion molecules as a downregulated (Fig. 1.8; Table 1.6). However,  $\beta$ -catenin is not downregulated in hydroxyurea treated sponges (Fig. 1.8), suggesting that  $\beta$ -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 Whittington 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 37°C. 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 -20°C 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 65°C 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 -20°C 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<sub>2</sub>,

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<sub>2</sub>HPO<sub>4</sub>, 0.24 g/L KH<sub>2</sub>PO<sub>4</sub>; pH 7.4) overnight at 4°C. 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 55°C. Probes were added to hybridization buffer (1:500) and denatured by heating for 5 minutes at 80°C. 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 55°C.

*Antibody incubation.* The probe:hybridization buffer mix was removed and sponges were washed 6 times in hybridization buffer B at 55°C. 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 4°C.

*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 4°C. 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 37°C 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. Permeabilization 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 51°C. 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 51°C.

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

*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<sub>2</sub>, 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 37°C 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 55°C. 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 55°C.

*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 SSC + 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 4°C 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 MgCl<sub>2</sub>, 0.1% Tween 20, brought up to 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).

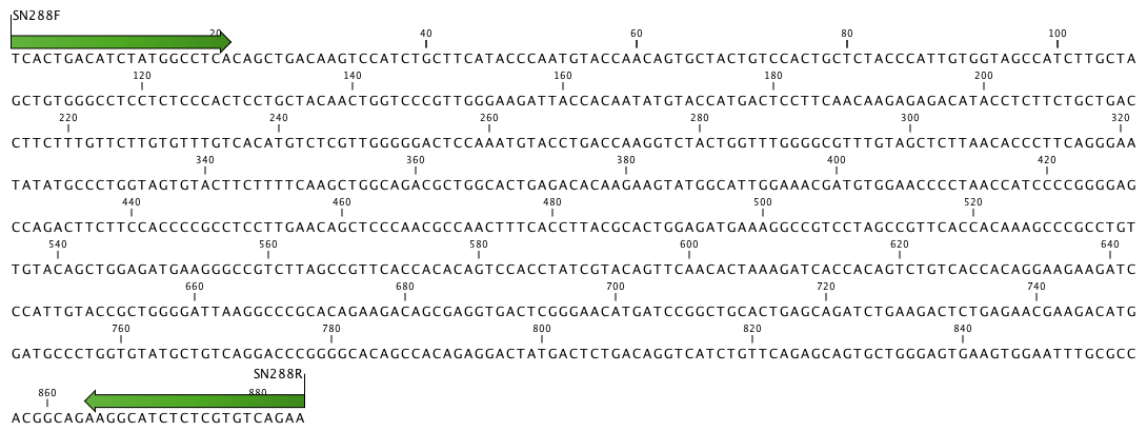
A third protocol was also tested. This protocol was originally developed for planarians and includes a reduction step which aids in probe penetration. Another difference was that the development buffer was made with polyvinyl alcohol. Proteinase K treatment was omitted as before. Unlike the sponge protocol,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

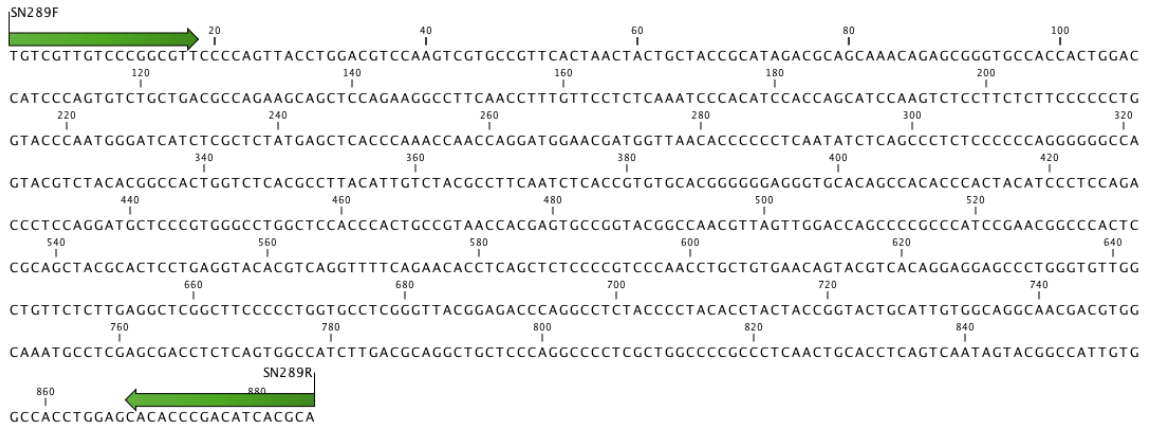
To improve signal to noise using the planarian protocol, post-hybridization treatment with RNase A was done. The choanogene sense and antisense probes were used. The sense probe had faint signal in the choanoderm (Fig. 2.14A). Choanogene signal in the choanoderm was very strong (Fig. 2.14B). Post-hybridization, 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**

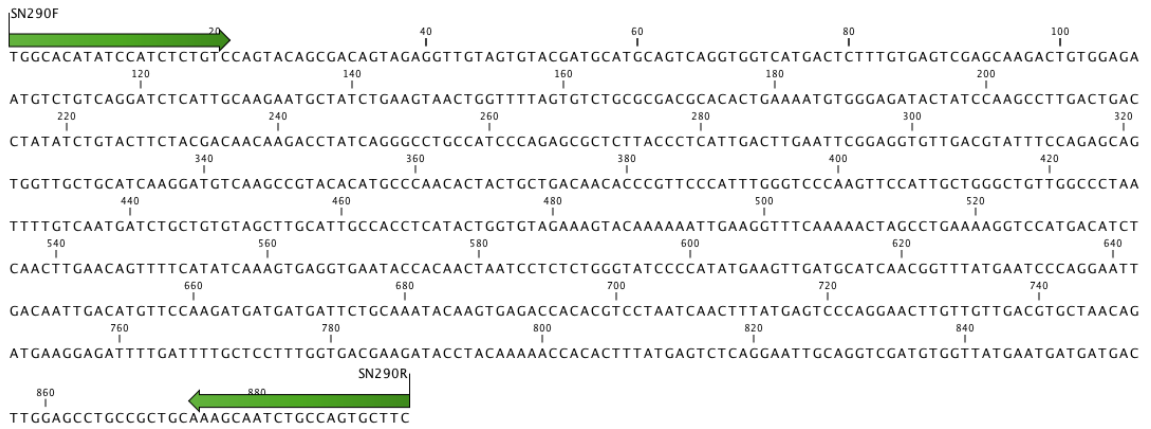
Gene	Amplicon size (bp)	Primer name	Orientation	Primer sequence
VLGR1	884	SN288F	Forward	TCACTGACATCTATGGCCTCA
		SN288R	Reverse	TTCTGACACGAGAGATGCCTT
Usherin	885	SN289F	Forward	TGTCGTTGTCCCGGCGTT
		SN289R	Reverse	TGCGTGATGTCGGGTGTG
Cadherin 23	894	SN290F	Forward	TGGCACATATCCATCTCTGTC
		SN290R	Reverse	GAAGCACTGGCAGATTGCTTT
Choanogene	897	SN291F	Forward	ATTCCAGAGGACAAGCCAGTA
		SN290R	Reverse	TGCCTAACATCTTTGTCCG
Annexin	974	SN264F	Forward	GGTGGTCACGGAAGTGTCAA
		SN264R	Reverse	TTAGTTGGGACCAACAATGGC



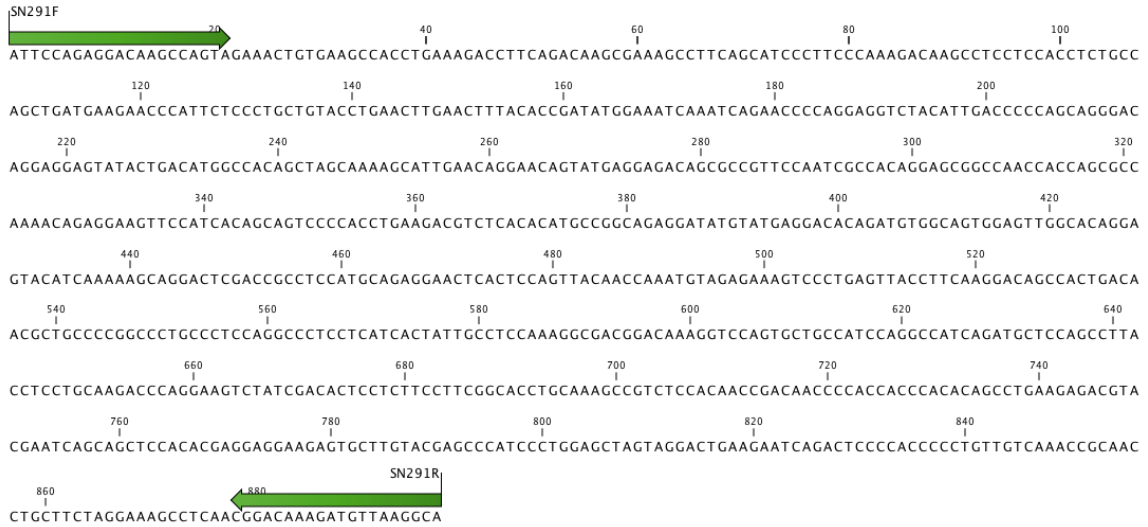
**Figure 2.1** *VLGR1* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp49556\_c0\_seq1, data set available on Compagen.



**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.



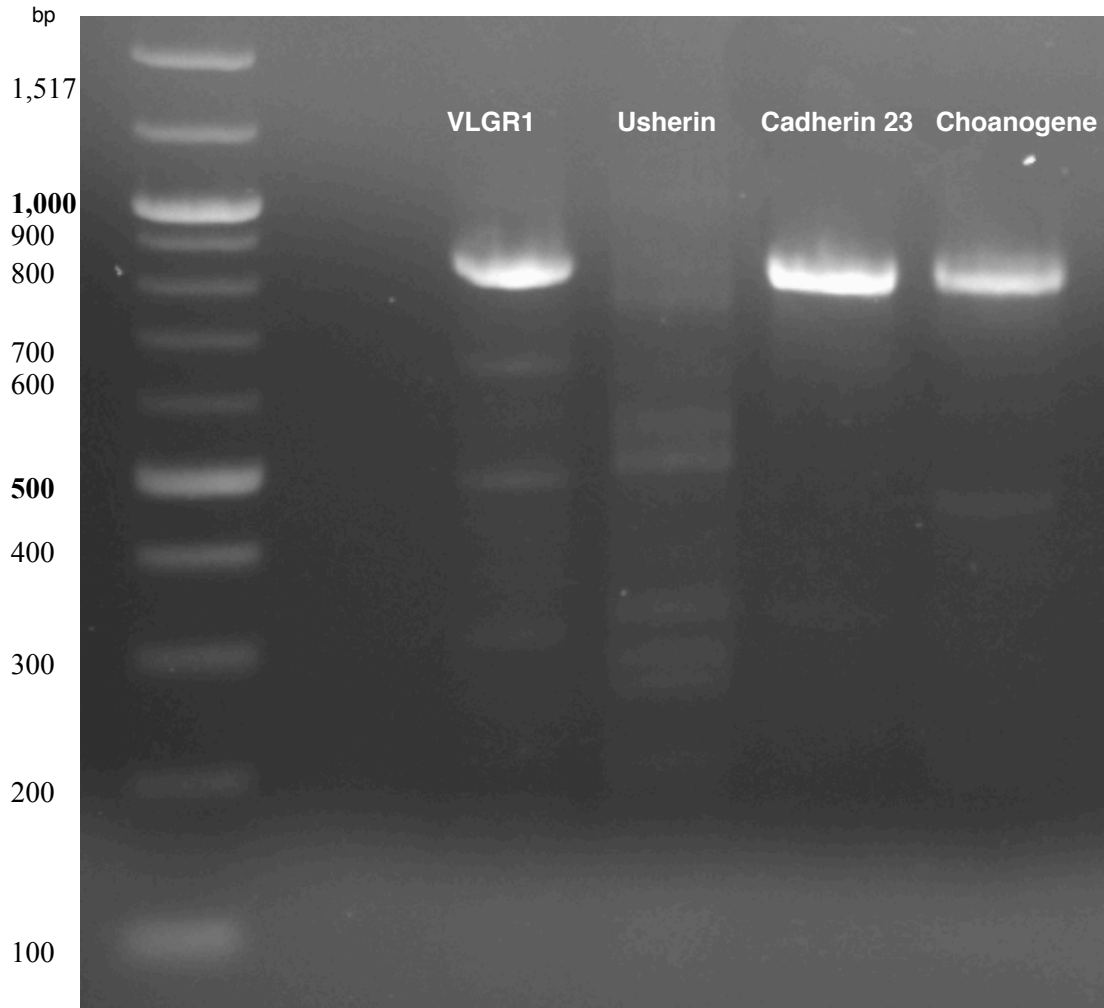
**Figure 2.3** *Cadherin 23* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp46992\_c0\_seq3, data set available on Compagen.



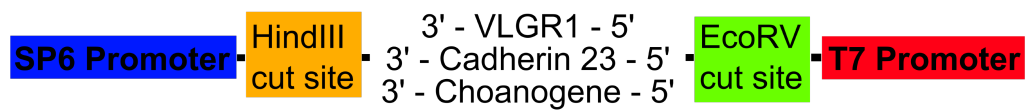
**Figure 2.4** *Choanogene* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp68328\_c0\_seq1, data set available on Compagen.



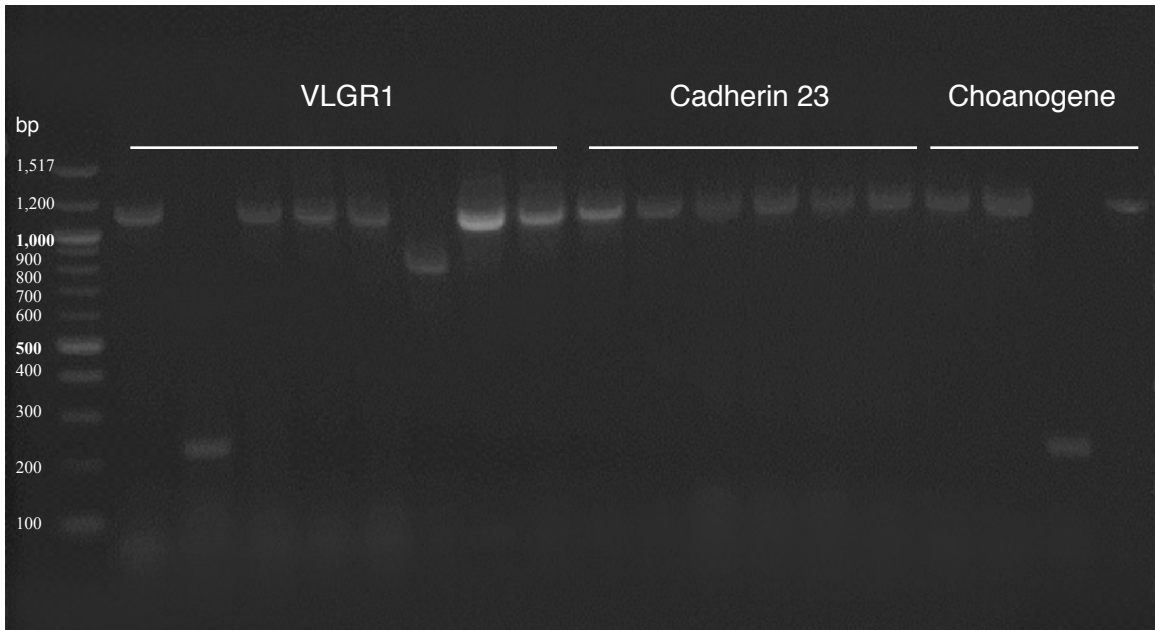
**Figure 2.5** *Annexin* fragment amplified from *E. muelleri* cDNA library. Green arrows represent primer binding sites. Sequence obtained from *E. muelleri* transcriptome, Accession: comp66863\_c0\_seq2, data set available on Compagen.



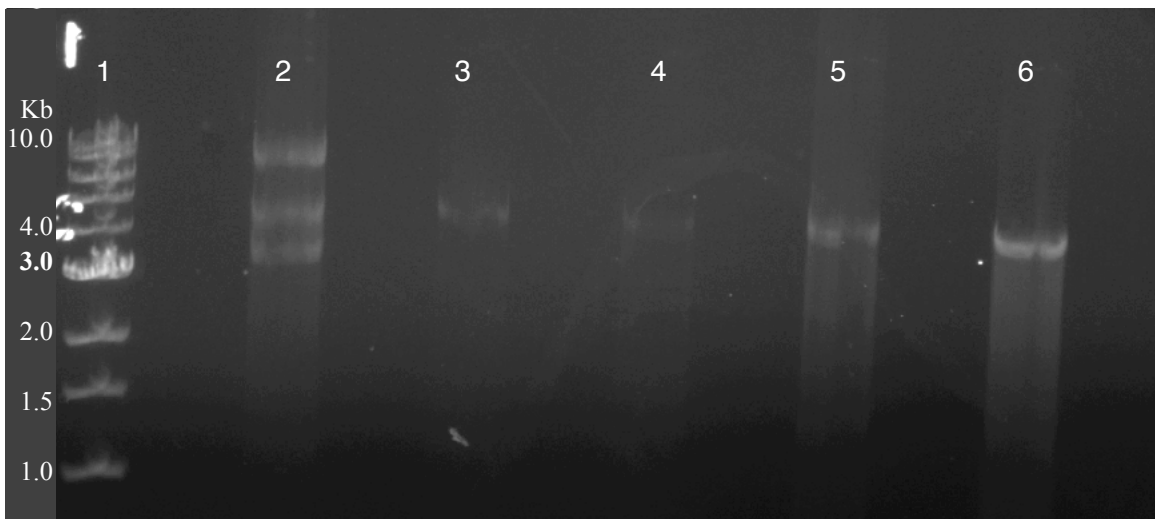
**Figure 2.6** Products of PCR on 1% agarose gel. Lane 1 shows 100 basepair ladder (New England BioLabs)



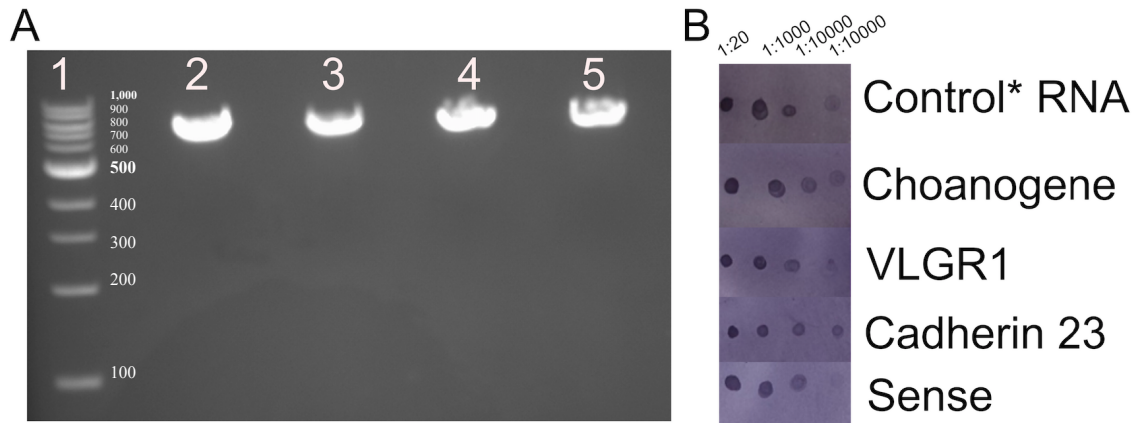
**Figure 2.7** Schematic of inserts relative to RNA polymerase promoters in pCR®II-TOPO dual promoter vector



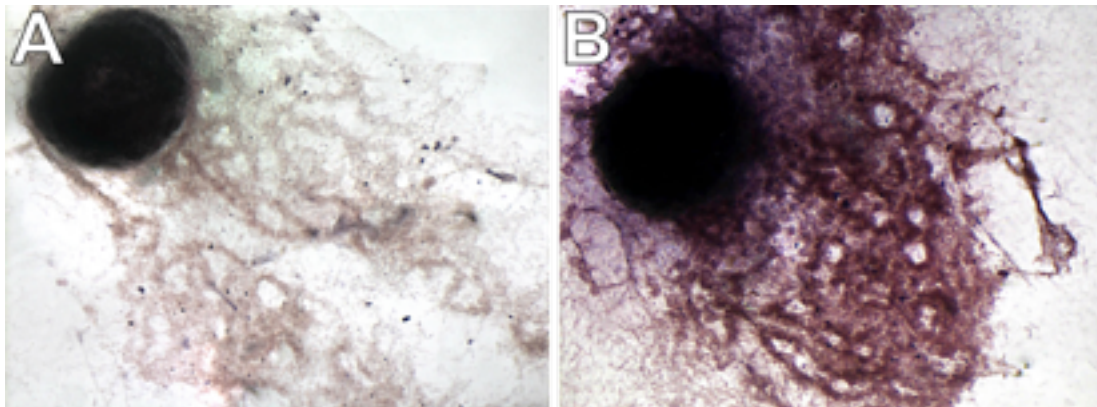
**Figure 2.8** Colony screen PCR visualized on 1% agarose gel. The first lane shows 100 bp ladder (New England BioLabs).



**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.

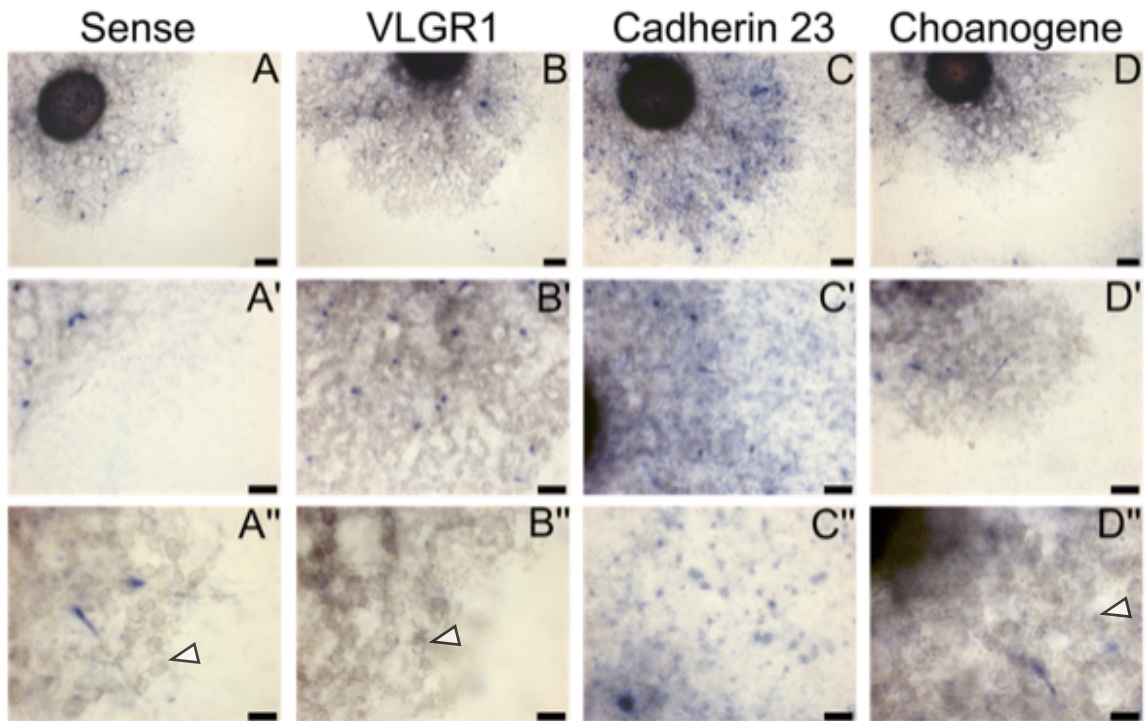


**Figure 2.10** Quality of DIG-labelled RNA probes. Probes visualized on a 1% agarose gel (A). Lane 1: 100 bp ladder (New England BioLabs). Lane 2: choanogene antisense probe. Lane 3: VLGR1 antisense probe. Lane 4: cadherin 23 antisense probe. Lane 5: choanogene sense probe. Dot blots of DIG-labelled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

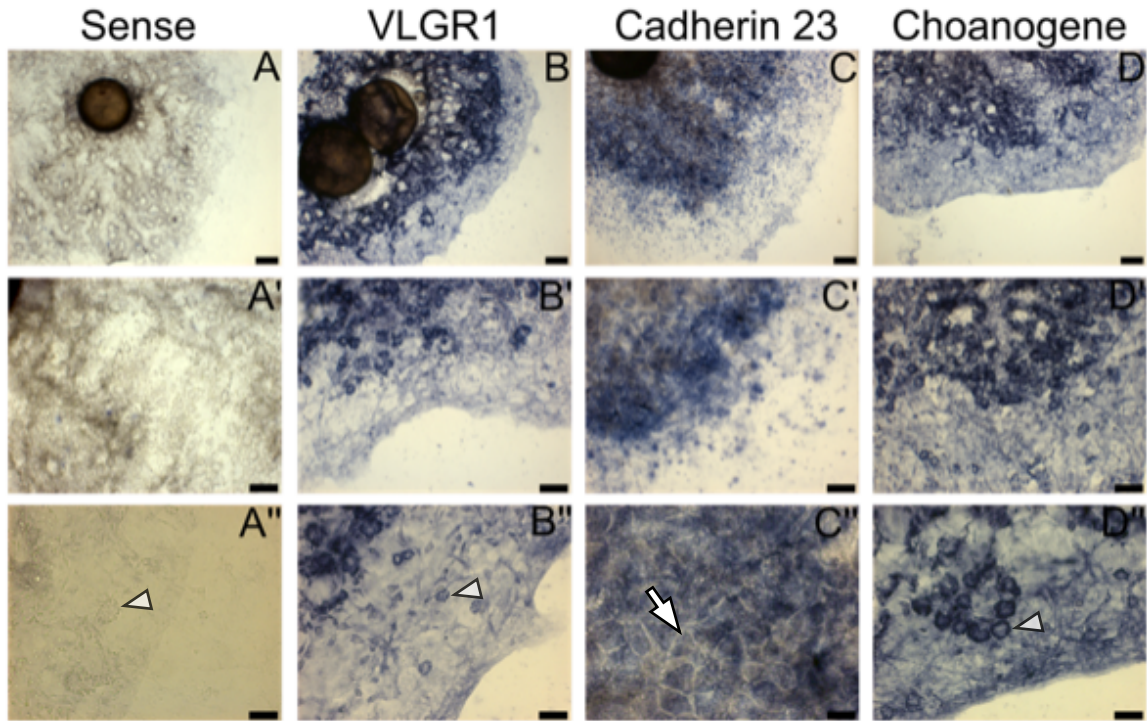


**Figure 2.11** Whole-mount *in situ* hybridization using Protocol 2. Annexin sense (A) does not show clear staining. Annexin antisense (B), previously reported to be expressed in choanocytes and archeocytes, displaying ubiquitous staining of tissue.

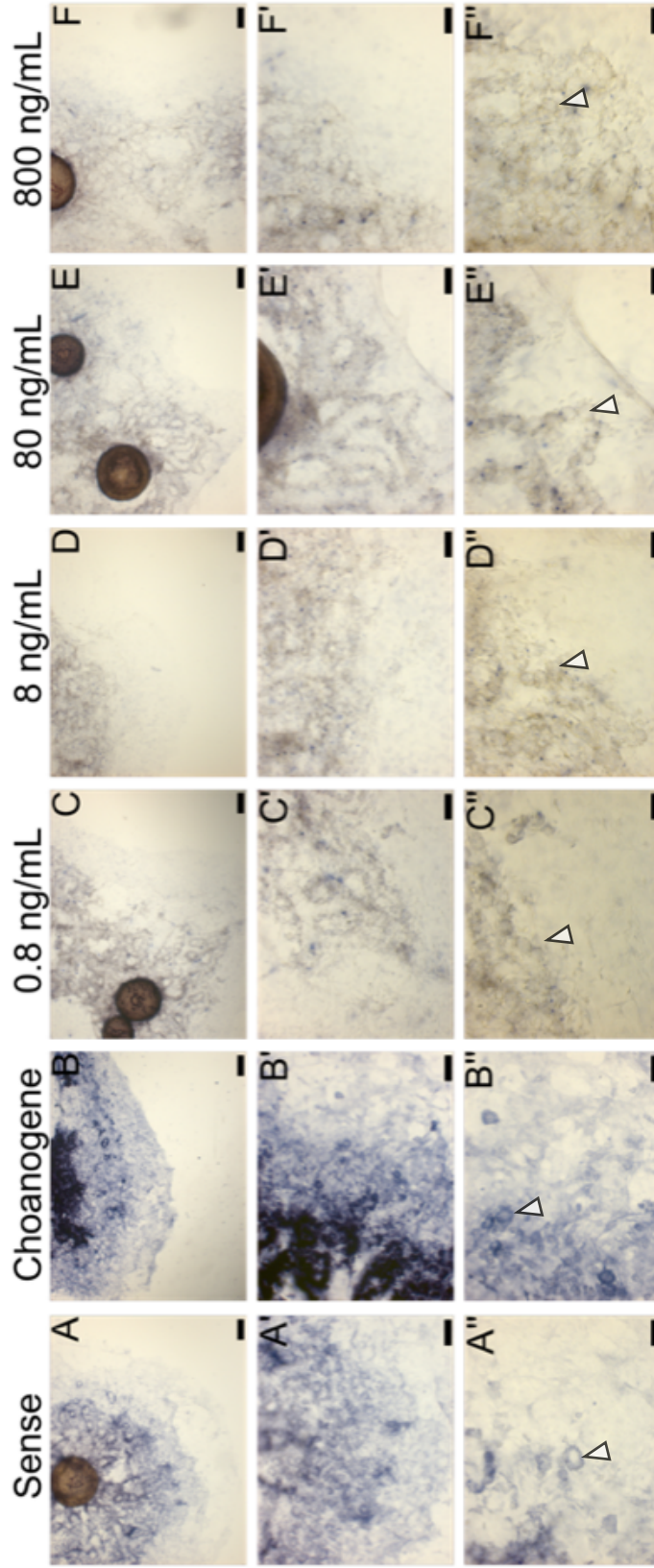




**Figure 2.12** Whole-mount *in situ* hybridization using sponge protocol. Choanogene sense strand (A, A', A'') shows faint staining in the aquiferous system. VLGR1 (B, B', B'') shows faint staining in the aquiferous system. Cadherin 23 shows staining in the pinacoderm (C, C', C''). Choanogene shows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

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

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

## 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 phmmr 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  $\mu\text{m}$  in width and 2 to 10  $\mu\text{m}$  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 Hejnl, 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.
- Hemrich, 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 Ciliatae 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/ $\beta$ -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 from 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.

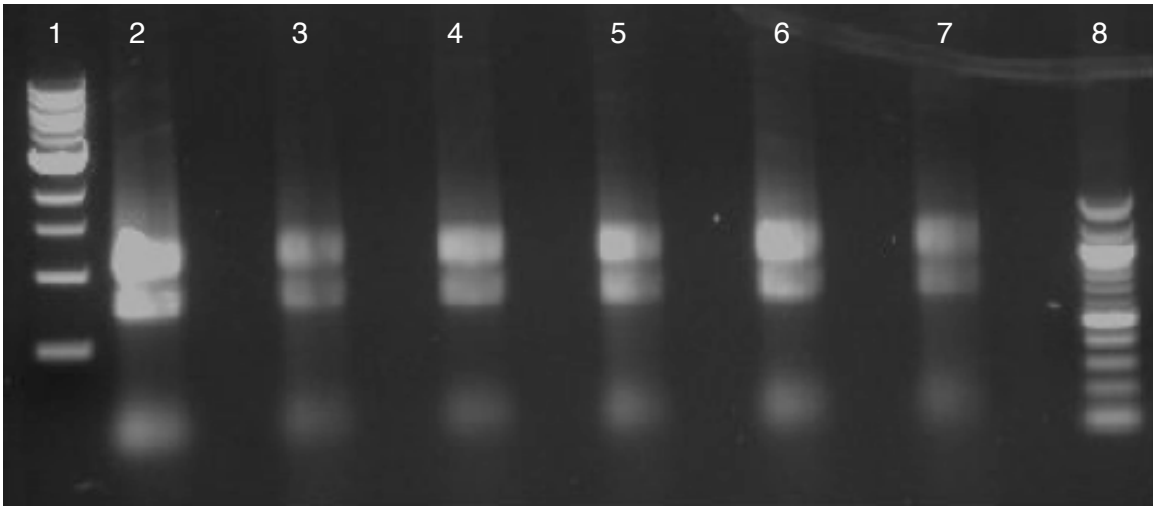
- Ruiz-Trillo, Iñaki, Andrew J Roger, Gertraud Burger, Michael W Gray, and B Franz Lang. 2008. "A Phylogenomic Investigation into the Origin of Metazoa." *Molecular Biology and Evolution* 25 (4): 664–72. doi:10.1093/molbev/msn006.
- Ryan, Joseph F, Kevin Pang, Christine E Schnitzler, Anh-Dao Nguyen, 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 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/ $\mu$ l	2.0
Em-BTL-HU	589 ng/ $\mu$ l	2.0
Em-NR-C	931 ng/ $\mu$ l	2.0
Em-NR-HU	380 ng/ $\mu$ l	1.98
Em-CO-C	711 ng/ $\mu$ l	2.16
Em-CO-HU	743 ng/ $\mu$ l	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**

<b>Sample ID</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Maximum</b>	<b>Minimum</b>
<b>Yield (Mbases)</b>	1,468	243.4	1,610	1,187
<b>%PF</b>	100.0	0.0	100	100
<b># Reads</b>	28,789,832.3	4,768,691.4	31,571,758	23,283,522
<b>% raw clusters per lane</b>	16.3	2.7	17.88	13.19
<b>% Perfect index reads</b>	99.4	0.0	99.46	99.40
<b>% one mismatch reads (index)</b>	0.6	0.0	0.60	0.54
<b>% <math>\geq</math>Q30 bases (PF)</b>	97.2	0.0	97.19	97.11
<b>Mean quality score (PF)</b>	37.7	0.0	37.70	37.68

**Table A.4 Summary of RNAseq run quality statistics for HU-treated group**

<b>Sample ID</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Maximum</b>	<b>Minimum</b>
<b>Yield (Mbases)</b>	1,489.3	89.9	1,584	1,405
<b>%PF</b>	100.0	0.0	100	100
<b># Reads</b>	29,202,260.7	1,770,368.3	31,065,437	27,542,226
<b>% raw clusters per lane</b>	16.5	1.0	17.59	15.60
<b>% Perfect index reads</b>	99.3	0.1	99.45	99.19
<b>% one mismatch reads (index)</b>	0.7	0.1	0.81	0.55
<b>% <math>\geq</math>Q30 bases (PF)</b>	97.1	0.0	97.15	97.11
<b>Mean quality score (PF)</b>	37.7	0.0	36.69	37.67

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