Published in final edited form as:
*Dev Biol*. 2014 November 1; 395(1): 144–153. doi:10.1016/j.ydbio.2014.08.015.

# Identification of differentially expressed genes during development of the zebrafish pineal complex using RNA sequencing

**Sataree Khuansuwan** and **Joshua T. Gamse***
Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

## Abstract

We describe a method for isolating RNA suitable for high-throughput RNA sequencing (RNA-seq) from small numbers of fluorescently labeled cells isolated from live zebrafish (*Danio rerio*) embryos without using costly, commercially available columns. This method ensures high cell viability after dissociation and suspension of cells and gives a very high yield of intact RNA. We demonstrate the utility of our new protocol by isolating RNA from fluorescence activated cell sorted (FAC sorted) pineal complex neurons in wild-type and *tbx2b* knockdown embryos at 24 hours post fertilization. Tbx2b is a transcription factor required for pineal complex formation. We describe a bioinformatics pipeline used to analyze differential expression following high-throughput sequencing and demonstrate the validity of our results using *in situ* hybridization of differentially expressed transcripts. This protocol brings modern transcriptome analysis to the study of small cell populations in zebrafish.

### Keywords

pineal; parapineal; zebrafish; epithalamus; transcriptome analysis; RNA-seq; bioinformatics; tbx2b; FACS; cell dissociation; pineal complex

## Introduction

Gene expression profiling is an excellent starting point to study a large variety of biological mechanisms associated with specific developmental stages, mutations, and/or disease conditions (Fodor et al., 2013; Jiang et al., 2014; Stockhammer et al., 2010; Vesterlund et al., 2011; Wells et al., 2013). RNA sequencing (RNA-seq) has become an established method for transcriptome profiling, especially with recent reduction of the associated costs. As compared to a more traditional method, such as an oligonucleotide microarray, RNA-seq offers greater dynamic range and the ability to detect novel transcripts (Ozsolak and Milos, 2011), even at the single cell level (Grindberg et al., 2013; Tang et al., 2009; Treutlein et al.,



* Corresponding author: josh.gamse@vanderbilt.edu.

2014; Vesterlund et al., 2011). The ability to use fluorescent markers and cell sorting for transcriptome profiling is a powerful advance that can overcome the limitations of using whole embryos or tissues, where changes in tightly regulated or localized transcripts can be masked by more broadly expressed genes. Isolating a small number of viable cells and obtaining good quality RNA suitable for downstream applications such as RNA-seq present its own challenges. Furthermore, the task of performing bioinformatic analysis on a large data set resulting from a single RNA-seq experiment can be intimidating when generating and analyzing transcriptome profiling data for the first time. This paper describes the methods that we used to isolate cells and extract high quality RNA from the developing zebrafish pineal complex, which consists of a relatively small population of neurons in the brain, and a description of the bioinformatic pipeline used to perform differential expression analysis between control and knockdown animals. Many of the techniques described will also be broadly applicable to other tissues or model systems.

The zebrafish pineal complex resides in the epithalamus and consists of the pineal and parapineal organs. The zebrafish pineal organ, which is made up of projection neurons and photoreceptors, is an important neuroendocrine organ that secretes melatonin in response to circadian stimuli (Cahill, 1996; Fodor et al., 2013; Gamse et al., 2002; Jiang et al., 2014; Stockhammer et al., 2010; Wells et al., 2013). Cell fate decisions between projection neurons and photoreceptors in the pineal organ is determined by the coordinated actions of BMP and Notch signaling (Cau et al., 2008; Quillien et al., 2011). However, the molecular mechanisms controlling photoreceptor subtypes differentiation in the pineal organ remain to be elucidated. Adjacent to the pineal organ is the parapineal organ, a small cluster of 10-12 neurons that typically resides on the left side of the brain (Concha et al., 2000; 2003; Gamse et al., 2003; Ozsolak and Milos, 2011). The unilateral placement of the parapineal organ is integral to the development of left/right asymmetry in the zebrafish brain (Concha et al., 2003; Gamse et al., 2003). Thus, understanding the molecular mechanisms involved in pineal complex formation will elucidate pineal photoreceptor differentiation and is key to understanding left/right asymmetry formation in the zebrafish brain.

The T-box containing transcription factor 2b (Tbx2b) plays important roles during cell fates determination in the zebrafish central nervous system. We have shown that Tbx2b is essential for both parapineal migration and specification, as well as proper formation of the pineal organ (Snelson et al., 2008b). Tbx2b is also necessary for inducing a bipotential population of pineal complex precursors that can give rise to either pineal cone photoreceptors or parapineal cells (Clanton et al., 2013). In the zebrafish retina, Tbx2b is required for UV cone specification (Alvarez-Delfin et al., 2009) and proper neuronal differentiation along the dorsal axis (Gross and Dowling, 2005). Thus, identification of Tbx2b responsive genes will elucidate its role in neuronal specification, differentiation, and migration. We sought to better understand pineal complex formation by comparing the transcriptomes of pineal complex cells between wild-type controls and *tbx2b* morphants during development by using RNA-seq. We have identified several genes that are under the control of Tbx2b and verified these results using *in situ* hybridization. Intriguingly, many of the Tbx2b responsive genes seem to have more traditional roles in photoreceptor differentiation and function. Given the known roles of Tbx2b during parapineal

development, these results imply that parapineal neurons might be a specialized form of photoreceptor. These data shed light on how cell identity within the pineal complex is defined.

## Materials and Methods

### Zebrafish

Zebrafish were raised at 28.5°C on a 14/10 light/dark cycle. Embryos obtained from natural mating were staged according to hours post-fertilization (hpf). The wild-type strain AB* (Walker, 1999) and the transgenic strain Tg[*flh*BAC:*kaede*]$^{vu376}$ (Clanton et al., 2013) were used.

### Morpholino injections

Embryos at 1-2 cell stage were pressure injected with 1 nL of a *tbx2b* splice blocking morpholino (MO), 5’ AAAATATGGGTACATACCTTGTCGT 3’, (Snelson et al., 2008b) diluted with nuclease-free water (Ambion) to 6 mg/mL.

### Tissue preparation

Using a dissecting microscope with a fluorescent filter set (Leica MZ16F), embryos from Tg[*flh*BAC:*kaede*]$^{vu376}$ and AB* crosses were sorted to separate carriers and non-carriers of the transgene. Embryos were dechorionated using 1 mg/mL Pronase solution (Roche) diluted in egg water at 28.5°C for 5-15 minutes prior to the desired developmental stage. Petri dishes were coated with 1% agarose to avoid injury to the embryos. After most chorions had broken open, Pronase solution was rinsed out with several washes of egg water. Embryos were then placed back into the 28.5°C incubator until they reach the desired developmental stage. In order to obtain as many stage-matched embryos as possible, we used embryos from clutches fertilized at 1-2 hours intervals and staggered the dissection time accordingly.

Embryos were anesthetized with Tricaine methanesulfonate (Acros Organics) prior to dissection at 24 hpf. To prevent collection of notochord cells, (which, like pineal cells, are fluorescently labeled by the Tg[*flh*BAC:*kaede*]$^{vu376}$ transgene) we manually dissected out the head tissue from embryos (Fig. 1) in Ringer's solution using fine forceps (Dumont #5, Fine Science Tools). For deyolking of embryos, refer to Manoli and Driever (2012) for instructions. We routinely collected about 200 heads per experimental condition. Dissected head tissues were then placed into 1.5 mL Eppendorf tube containing 1.0 mL of cold Ringer's solution without calcium or magnesium and containing EDTA (see recipe, Supplementary Materials and Methods) and placed on ice. As more tissue and Ringer's solution was added during the collection process, Ringer's solution was removed as necessary to keep the total volume constant. Similar to the heads, the trunks from dissected Tg[*flh*BAC:*kaede*]$^{vu376}$ embryos were placed into a 1.5 mL Eppendorf tube containing 1.0 mL of cold Ringer's solution on ice, and later used as a positive fluorophore control for fluorescence activated cell (FAC) sorting. Kaede-negative embryos were anesthetized and placed directly into a 1.5 mL Eppendorf tube containing 1.0 mL of cold Ringer's solution on ice, and later used as a negative control for FAC sorting.

### Tissue dissociation

After all tissue samples had been collected, each sample was washed 3 times with 1 mL of Dulbecco's PBS (D-PBS, Gibco). Care was taken to ensure that no samples were lost during washes. After the last D-PBS wash, as much DPBS was removed as possible. Accumax cell dissociation enzyme mix (Innovative Cell Technologies) was added to each sample at a ratio of 0.5 mL of Accumax enzyme mix per 200 heads or 0.75 mL per 200 trunks. Additionally, 5 μL of 10KU/ml DNase I (Sigma) was added to each tube to alleviate cellular clumping during dissociation. Each tube was placed into a 37°C water bath for 30-60 minutes. During this time, dissociation was facilitated with gentle pipetting. Pipette tips with a filter barrier were always used to avoid contamination. Larger tissues (e.g. trunk tissue) took longer to dissociate. After the tissues were fully dissociated, each tube was removed from the water bath and placed on ice. Then, a volume of washing solution (see recipe) equal to the amount of Accumax solution used was added to each tube. With gentle pipetting as previously, no clumps of tissue were observed. In both of our sample conditions, heat shock proteins were activated at similar levels (Supp. Fig. 1). Researchers are advised to look for alternative dissociation enzyme that would work at room temperature should their gene(s) of interest is activated as a response to heat shock condition.

### Sample preparation for fluorescence-activated cell sorting

Each cell suspension sample was filtered through a 35 μM nylon mesh directly into a round bottom 12 × 75 mm Falcon collection tube (Becton Dickinson) (Supp. Fig. 2). In order to obtain as much sample as possible, the tube used for dissociation was then rinsed once with 0.5 mL of full-strength washing solution and twice with diluted washing solution (see recipe in Supplementary Materials and Methods); all of these rinses were subsequently filtered through the same nylon mesh into the collection tube. The dissociated cells were then pelleted using a tabletop centrifuge (Beckman X15R) at 1100 rpm (282 g) for 10 minutes at 4°C. The supernatant was removed and the pellet resuspended thoroughly with an appropriate amount of diluted washing solution to achieve a density of no greater than $5\times10^6$ cells/ml, which is appropriate for low pressure (100 μM nozzle) FAC sorting. The minimum sorting volume used was 300 μL. For the experimental sample (cells from transgenic embryo heads), the vital dye propidium iodide (PI, Sigma) was added at 1 μL per 1 million cells. Typically, we added 1 μL per 1 mL of cell suspension. The samples were kept on ice prior to flow cytometry. Kaede positive cells were sorted directly into Trizol-LS (Life Technologies) using the BD FACS Aria II (VUMC Flow Cytometry Core, Nashville, TN). Our protocol yields an average of 83.1 percent viable cells (N=8), ranging from 68.8-95.0 percent. On average, we collected 226 Kaede-positive cells per head (Fig. 2).

For each FAC sorting experiment, the following controls were used: (1) negative, unstained sample (i.e. Kaede-negative embryos, no PI added); (2) PI compensation control (i.e. Kaede-negative embryos with PI viability dye added); (3) positive fluorophore control (i.e. Kaede-positive trunks, no PI added). Prior to running the experimental samples, we adjusted the gating of the FAC sorter using our positive compensation control with PI added so as not to lose any experimental sample.

### RNA extraction

All equipment used for RNA isolation was pretreated with RNAse Away (Molecular BioProducts). When required, the ratio of the Trizol-LS volume to the sample volume were brought up to 3:1 with addition of nuclease-free water. Cells were lysed by pipetting the mixture several times. RNA was extracted using Trizol-LS similar to the manufacturer's protocol, with one modification: we used Phase Lock Gel-Heavy tubes (PLG, Fisher) during the phase separation step.

Each PLG tube was prepared for use by centrifugation at 14,000 rpm (18,000 g; Beckman Coulter Microfuge 18 centrifuge) for 2 minutes at room temperature. 800 µL of cell lysate (in Trizol-LS) was added to a pre-spun PLG tube and incubated for 5 minutes at room temperature to allow for complete dissociation of nucleoprotein complexes. Chloroform was added to the PLG tube at a ratio of 200 µL per 1mL of Trizol-LS plus sample mixture initially used. The tube was then shaken vigorously by hand (without vortexing) for 15 seconds. Phase separation was achieved via centrifugation at 13,200 rpm (16,100 g; Eppendorf 5415D centrifuge) for 15 minutes at 4°C. The clear, aqueous top phase containing RNA, atop the PLG, was then transferred to a fresh tube via decanting or by using a pipette. Ice-cold isopropanol was added to the aqueous phase at a ratio of 500 µL per 1 mL Trizol-LS plus sample initially used, along with 3 µL of 20mg/mL mussel glycogen (Roche, used as a carrier for the precipitation of nucleic acids). Samples were mixed by repeated inversion and incubated at room temperature for 10 minutes. To allow for precipitation of RNA, samples were incubated at −20°C for at least 1 hour (up to 16 hours). RNA was pelleted via centrifugation at 13,200 rpm (16,100 g; Eppendorf 5415D centrifuge) for 15 minutes at 4°C, resulting in a visible RNA pellet. The supernatant was discarded and the pellet was washed twice with 1 mL of ice-cold 75% ethanol. The location of the pellet was marked and the pellet was allowed to air-dry at room temperature; after drying the pellet became translucent. The pellet was resuspended with 20 µL of nuclease-free water plus 1 µL of RNase inhibitor (40U/µL RNasin, Promega) at 55°C for 10 minutes. The sample was then gently pipetted to ensure complete resuspension of RNA. Because the subsequent RNA-seq includes a poly-A selection, there was no need to treat samples with DNase. RNA was stored at −80°C until needed.

RNA quantity was determined using the RNA assay program on a Qubit fluorometer (Life Technologies). Each FAC-sorted Kaede-positive cell yields an average of 3.5 pg of RNA, ranging from 1.3-5.8 pg of RNA/cell (n=11). These values are similar to yields obtained from neuronal cells in another study (Saxena et al., 2012). RNA quality and integrity were analyzed by capillary electrophoresis using the RNA Pico chip on an Agilent 2100 Bioanalyzer. Our protocol typically yielded total RNA with RIN integrity values ranging from 6.0-8.5, with an average of 7.2 (n=14) (Fig. 3). We observed improved RNA integrity values when we spent less time completing the RNA extraction process (Fig. 3, Supp. Fig. 3). In our hands, the use of a commercially available column (Purelink RNA Micro Kit, Invitrogen) led to a significant loss of total RNA (Supp. Fig. 3), despite following the manufacturer's protocol precisely.

### cDNA library preparation and RNA-seq

cDNA sequencing libraries preparation and sequencing was performed by the Genome Sciences Resource at Vanderbilt University Medical Center (Nashville, TN) using the Illumina TruSeq RNA Sample Preparation Kit as previously described (Elmore et al., 2012; Venkateswaran et al., 2013). High-throughput RNA-seq was performed on four libraries (two control groups and two *tbx2b* knockdown groups) twice. Using the multiplexing strategy of the TruSeq protocol, each technical replicate, consisting of four libraries, was performed on one lane of the Illumina HiSeq 2000 utilizing version 3 chemistry. On average, 40 million 51-101 bp single-end reads were generated for each group, which should be sufficient to detect differentially expressed genes. A study performed in human and mouse tissues has shown that sequencing beyond the depth of 3 million reads does not increase transcript detection (Ramsköld et al., 2009). Others have concluded that cDNA sequencing with 30-40 million 25 bp reads is sufficient to detect major splice isoforms for abundant and moderately abundant transcripts in mouse and that 1X coverage of the transcriptome can be achieved with 40 million 25 bp reads (Mortazavi et al., 2008).

### Read mapping, differential expression analysis, and data visualization

Bioinformatic analysis was performed on a 64-bit MacBook Pro (Apple) running Mac OS X 10.8.5 with 8 GB of RAM. Raw 51 or 101 bp single-end reads generated from the Illumina HiSeq 2000 were first assessed for sequence quality using the FastQC software version 0.10.0 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Quality scores across all bases fell into the “very good quality” category (>28, Sanger/Illumina 1.9 encoding). Adapter sequences were removed using cutadapt software version 1.4.1 (Martin, 2011) (https://code.google.com/p/cutadapt/). Reads were trimmed to 40 or 80 bp, respectively, to remove bases of lower quality using trimmomatic software version 0.32 (Bolger et al., 2014). Each sequencing data set was independently mapped to the zebrafish genome with a bowtie2 index generated from Danio_rerio.Zv9.70 (Ensembl) downloaded from Illumina's iGenomes collection (https://support.illumina.com/sequencing/sequencing_software/igenome.ilmn) using the TopHat2 program (Kim et al., 2013) with novel splice discovery disabled. HiSeq 2000 libraries were aligned with the following options: “-p 2 -G genes.gft --no-novel-juncs” where genes.gft contains the Ensembl coding transcripts in GTF format downloaded from Illumina's iGenomes collection. On average, 89% of trimmed reads were mapped (83.7% uniquely mapped) to the genome.

Differentially expressed genes between control and *tbx2b* morphant pineal cell samples were identified by using Cuffdiff 2 program (Trapnell et al., 2013) with the following options: “-b genome.fa -p 2 -u -c 5 genes.gtf” where genome.fa is a multifasta file downloaded from Illumina's iGenomes collection. We used a 1.8-fold change in expression and an FDR adjusted p value (q value) of ≤ 0.05 as cut offs.

CummeRbund software was used to generate volcano plot and bar plots for expression levels for genes of interest (Goff et al., 2012). Volcano plot was generated using the following options: “alpha=0.05, showSignificant=T, xlimits=c(-12,12)”.

For additional information on how to install and use TopHat, Cufflinks, and CummeRbund, see Trapnell et. al. (2012). For read mapping, differential expression analysis, and visualization of individual Tbx2b exons, see Supplementary Methods.

### In situ hybridization

Whole-mount RNA *in situ* hybridization was performed as previously described (Gamse et al., 2003), using reagents from Roche Applied Bioscience. Embryos were hybridized to probe at 70°C in hybridization solution containing 50% formamide (Roche). Hybridized probes were detected using alkaline phosphatase-conjugated antibodies (Roche) and visualized by 4-nitro blue tetrazolium (NBT; Roche) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP; Roche). RNA probes (Table 1) were labeled using digoxygenin-UTP.

### Imaging

Bright field images were obtained using Qcapture software (Qimaging) with a Retiga EXi *Fast* 1394 cooled monochrome-12 bit CCD camera (Qimaging) attached to a RGB color filter (Qimaging) mounted on a Leica DM6000B microscope with a 20X objective.

### RT-PCR

First strand cDNA was generated using 1 g total RNA, 200 ng random hexamers (Applied Biosystem), 1 μL of 10 mM dNTP (NEB), 4 μL of 5x first strand buffer (Invitrogen), 2 μL of 0.1 M DTT (Invitrogen), 1 μL of RNase inhibitor (Promega) and 1 μL of Superscript III reverse transcriptase (Invitrogen) in a 20 μL reaction. PCR was performed using 2 μL of cDNA from the RT reaction, with 30 rounds of amplification with annealing temperature of 60 °C and 90 seconds extension time. The following primers were used: *β-actin* 5’ CCATGGATGAGGAAATCGCTGC 3’ and 5’ GTCACACCATCACCAGAGTCC 3’ ; *flh* 5’ GTACTGGCGAAAGCAGCAGTT 3’ and 5’ AGCAGATGCCAACAGAAAGC 3’; *no tail* 5’ TGGAAATACGTGAACGGTGA 3’ and 5’ TCTGAATCCCACCGACTTTC 3’; *tbx2b* 5’ TGTGACGAGCACTAATGTCTTCCTC 3’ and 5’ GCAAAAAGCATCGCAGAACG 3’.

## Results

### FACS yields an enriched population of pineal complex cells

To isolate pineal cells, we used the Tg[*flh*BAC:*kaede*]$^{vu376}$ transgenic line that expresses the fluorophore Kaede at high levels in the pineal complex and the notochord at 24 hpf (Fig. 1A,B). To minimize the collection of unwanted Kaede-positive notochord cells, we resected head tissue, and then FAC sorted and collected dissociated Kaede-positive cells from the head tissue of transgenic embryos. RT-PCR for specific pineal and notochord markers was performed on cDNA synthesized from the FAC sorted Kaede-positive cells to ensure that we enriched for the desired pineal cell population and excluded notochord cells. As controls, we performed RT-PCR on cDNA synthesized from whole Tg[*flh*BAC:*kaede*]$^{vu376}$ embryos and Kaede-negative cells from head tissue (Supp. Fig. 4A,B, respectively). FAC sorted Kaede-positive cells from dissected head tissue showed high expression of *flh*, a marker present in the pineal complex and the notochord with barely detectable levels of the notochord-specific marker *no tail*. As expected, *tbx2b* was readily detected in the Kaede-positive cells.

### RNA-seq from small numbers of FAC sorted cells identifies pineal-specific genes affected by Tbx2b loss-of-function

With the ability to selectively enrich for pineal complex cells, we sought to identify pineal specific genes that are under the control of the transcription factor Tbx2b. In order to identify genes regulated by Tbx2b, we FAC sorted pineal cells as above in the presence or absence of an antisense morpholino that targets *tbx2b* (Snelson et al., 2008b). This morpholino is predicted to block inclusion of exon 3 from *tbx2b* primary transcripts, resulting in a frameshift and premature truncation of translated protein. To test the efficacy of the morpholino, we aligned RNA-seq reads for individual exons of *tbx2b* and found that the presence of exon 3 was reduced by 94.5% in *tbx2b* morphants, as compared to non injected controls (NIC) (aligned reads ratio of morphants:NIC=0.054, N=2). Other exons of *tbx2b* were unaffected (aligned reads ratio of morphants:NIC=1.007, N=2) (Fig. 4).

Comparison of expression profiles in the presence or absence of *tbx2b* knockdown resulted in the identification of 83 genes that were significantly up (20/83) or down (63/83) regulated ($q \leq 0.05$) by more than 1.8 fold across all biological and technical replicates (Fig. 5). By searching published literature, as well as Zfin.org and Biomart (Ensembl.org), expression data were available for 57 of 83 identified genes. Of these, 14/57 had pineal-specific expression patterns (Table 2, see Supplementary Table 1 for a complete list) and 14/57 had broad expression patterns in the brain (Supplementary table 2). There has been no documented role for these genes during pineal complex development. At 24 hpf, approximately 284 genes showed specific expression in the pineal complex (BioMart, Ensembl.org using Prim5 and epiphysis as search terms). Therefore, Tbx2b function is required for appropriate expression of at least 4.9% of pineal specific genes during parapineal development.

To validate our findings, we used *in situ* hybridization to examine the expression of 6 of the 14 pineal-specific genes in both NIC and *tbx2b* morphant embryos at 24 hpf. Of 6 genes tested, we found 100% concordance between our *in situ* data and our RNA-seq results (Fig. 6A). We also performed *in situ* hybridization experiments for these 6 genes in *tbx2b*$^{c144}$ mutants at 24 hpf (Fig. 6B). Genes that showed the highest fold reduction in *tbx2b* knockdown condition from RNA-seq reads (Fig. 6C, Table 2) also showed the highest reduction in expression as observed by *in situ* hybridization. By 36 hpf, the parapineal organ has migrated away from the midline and is located to the left of the pineal organ (Gamse et al., 2003). Of the 6 differentially expressed genes tested, *otx5* and *rdh5* are expressed in the migrating parapineal cells (Fig. 7, 36 hpf). At a later stage, *otx5* is maintained in differentiated parapineal cells (Fig. 7, 2dpf).

## Discussion

High-throughput cDNA sequencing (RNA-seq) has emerged as a powerful tool for gene expression profiling. This technique is replacing microarrays because RNA-seq provides a larger dynamic range, requires less starting material, and allows detection of novel transcripts. Despite the high sensitivity, however, this technique can still be difficult to perform when analyzing small numbers and/or dispersed populations of cells. Additionally, researchers are faced with the challenge of analyzing large data sets generated by these

experiments. Using 24 hpf zebrafish embryos, we have described the methods that we used to isolate a small number of pineal complex cells from each embryo and extract high quality RNA suitable for RNA-seq experiments. We also described how we utilized a simple bioinformatics pipeline using readily available open-source programs to obtain differential gene expression data between two experimental conditions; in this case, the pineal transcriptomes between wild-type and *tbx2b* morphants. We were able to identify 83 genes with 1.8 fold or more significant change in expression. Focusing on genes that have specific expression in the pineal complex, we have identified 14 genes that were differentially and significantly regulated. We verified 6 of these findings via *in situ* hybridization.

Our experiments enabled the identification of 14 genes that have restricted expression within the pineal complex and are differentially expressed between *tbx2b* morphants and WT. Intriguingly, these genes have not been shown to participate in pineal complex development. However, most of these genes do participate in photoreceptor differentiation and function within the retina. This is of potential interest, as pineal and retinal photoreceptors share similar features and may have evolved from a common ancestral cell type (Mano and Fukada, 2007). *rlbp1b* is expressed in the zebrafish Müller glia cells and has been shown to be important for normal cone vision (Collery et al., 2008; Fleisch et al., 2008). Loss of *rpe65a* or rod arrestins, results in altered rod outer segment morphology (Schonthaler et al., 2007) or reduced photoresponse recovery of the rod photoreceptor (Renninger et al., 2011; Xu et al., 1997), respectively.Furthermore, *rdh5* and *rbp4l* are important for retinal cell differentiation (Nadauld et al., 2006; Nagashima et al., 2009; 2010). In mouse, rhodopsin regeneration is dependent on Rgr, the *rgra* homolog (Chen et al., 2001). Lastly, recoverin, encoded by the *rcv1 gene*, participates in the visual cycle (Stryer, 1991). A role for Tbx2b in pineal photoreceptor formation and function is unexpected as mutation of *tbx2b* did not lead to a reduction in Arr3a or rhodopsin expression in red/green cone cells or rod outer segments (Clanton et al., 2013; Snelson et al., 2008b). However, our results are consistent with known roles of Tbx2b in promoting particular cone fates in the zebrafish retina (Alvarez-Delfin et al., 2009). Together, these findings suggest novel roles of Tbx2b in pineal photoreceptor formation and function. Additionally, these results suggest that the parapineal may be a specialized form of photoreceptor.

It has been shown that prenylation of G protein gamma subunits is necessary to facilitate directed migration of primordial germ cells (PGC) (Mulligan et al., 2010; Mulligan and Farber, 2011). The parapineal organ also migrates directionally. Therefore, it is reasonable to predict that the molecular mechanisms controlling parapineal migration are similar to those controlling PGC migration. Our current study has identified two genes affected by Tbx2b that encode G protein gamma subunits, *gngt1* and *gngt2a.* We are currently investigating the roles of these two genes in parapineal migration.

Although we have shown the power of FAC sorting plus RNA-seq to reveal novel gene targets of Tbx2b in the developing pineal complex, we are aware of experimental limitations. The zebrafish pineal complex anlage is composed of approximately 30-32 pineal complex cells by 24 hpf (Masai et al., 1997; Quillien et al., 2011), but we isolated an average of 226 cells/embryo. Since the number of cells that we FAC sorted from each Tg[*flh*BAC:*kaede*]$^{vu376}$ head exceeds the number of pineal cells, we are likely not working

with a 100% pure population; rather, we are working with a highly enriched population of pineal cells. Having ruled out the inclusion of significant numbers of Kaede expressing notochord cells (Supp. Fig. 4A,B), we conclude that many non-pineal cells from the brain have been collected and subsequently analyzed. Nevertheless, we did achieve significant enrichment of our desired cell types with an approximate 17-fold enrichment of pineal cells as compared to using total head tissue. Kaede-positive pineal cells normally constitute about 0.85% of total population of head tissues. Post FAC sorting, pineal cells constitute about 14.2% of the sorted population. One constraint on the ability to better enrich the desired population is the transgene used to label pineal cells, *flh:Kaede*. In preparation for FAC sorting experiments, one should fully characterize the expression of the transgene used to mark cells. Using high resolution imaging, we found weakly expressing Kaede-positive cells in the telencephalon. In any given embryo, these cells express Kaede at a much lower level than cells in the pineal complex. However, the absolute level of transgene expression varies from embryo to embryo. Cells with a low absolute level of Kaede expression could be pineal cells from embryos with overall low transgene expression, or telencephalic cells from embryos with overall high transgene expression. Collecting only the highest expressing population of Kaede-positive cells (pineal cells from embryos with high expression of the transgene) would have resulted in extremely low yields. Therefore, a transgenic line with less embryo-to-embryo variation is needed in order to obtain a more pure population of pineal complex cells. Alternatively, the use of two transgenes with different fluorophores with expression that overlaps only in the tissue to be isolated (in this case, the pineal) could be used.

Despite collection of extraneous cells by our FAC sorting protocol, it was nevertheless successful in identifying differentially expressed genes in the pineal, as confirmed by *in situ* hybridization. It is also important to note that in the head, Tbx2b is only expressed in the pineal complex (which is Kaede positive) and in the eyes (which are Kaede negative), and as a transcription factor it is predicted to act cell autonomously. Therefore, although we are capturing some telencephalic cells in our sample, we do not expect that gene expression in those cells would be significantly changed by loss of Tbx2b function, and therefore would not affect our list of differentially expressed genes.

We used cuffdiff to identify differentially expressed genes, using two biological replicates. However, increasing the number of biological replicates (3 or greater) is strongly suggested, as it should yield fewer false positives and false negatives. Additionally, there are other widely used tools available for differential expression analysis including DESeq (Anders and Huber, 2010) and EdgeR (Robinson et al., 2009).

Tbx2b has been shown to be crucial for the development of the pineal and parapineal organs (Snelson et al., 2008b); its role in parapineal formation makes it important for the establishment of left-right asymmetries in the zebrafish brain (Gamse, 2005; Gamse et al., 2003). Identifying downstream targets of Tbx2b should pinpoint the genes involved in pineal complex development including the specification and migration of parapineal cells. We used *tbx2b* morphants because of the challenges of rapidly identifying a sufficient number of *tbx2b* homozygous mutant embryos, as they are morphologically indistinguishable from their siblings at 24 hpf. However, we are convinced that we have

successfully knocked down Tbx2b using the morpholino. Not only does treatment with this concentration of morpholino phenocopy *tbx2b*$^{c144}$ parapineal defects (Snelson et al., 2008b), but we find that it almost completely eliminates the presence of exon 3 in *tbx2b* transcripts, leading to a premature truncation of nearly all Tbx2b protein in the treated embryos (this study). Our current study looked at the 24 hours post fertilization stage of development, which is just prior to parapineal and pineal photoreceptor differentiation as well as parapineal migration. Our results have yielded genes that may be involved in pineal photoreceptor formation and function, as well as genes that may have roles in parapineal migration. In order to identify genes responsible for parapineal specification, we believe this experiments should to be repeated at an earlier time point, such as 15-18 hpf when parapineal specification occurs (Snelson et al., 2008a).

Aside from its roles in parapineal and UV cone development, Tbx2b has also been shown to be involved in other important biological processes including neuronal differentiation in the dorsal retina (Gross and Dowling, 2005), atrioventricular canal formation in the heart (Chi et al., 2008), and notochord formation (Fong et al., 2005). While our differential expression data came from comparing the transcriptomes between WT and *tbx2b* morphant pineal complex cells, it will be interesting to explore our list for Tbx2b responsive genes that also function in retina, heart, and/or notochord formation. In particular, such genes could lead to insight into human health, since Tbx2 is a putative modifier of two human syndromes caused by mutation in related T-box genes, ulnar-mammary syndrome (Tbx3) and 22q11.2 deletion syndrome (Tbx1) (Jerome-Majewska et al., 2005; Mesbah et al., 2012).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the following: Leah Potter and Mark Magnuson's laboratory (VUMC) for sharing protocols, suggestions, and reagents relating to cell dissociation; John Gibbons, David Rinker, Xiaofan Zhou and Antonis Rokas's laboratory for guidance on transcriptome data analysis; Genome Sciences Resource at VUMC (VANTAGE) for performing cDNA library construction and RNA-seq; VUMC Flow Cytometry Shared Resource, which is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485) and the Vanderbilt Digestive Disease Research Center (DK058404), for performing cell sorting experiments; James G. Patton for suggestions regarding manuscript preparation. We would also like to thank Erin Booton, Brittany Parker, Elleena Benson and Qiang Guan for fish care and maintenance. This work was supported by NIH grants HD054534 and EY024354.

## References

Alvarez-Delfin K, Morris AC, Snelson CD, Gamse JT, Gupta T, Marlow FL, Mullins MC, Burgess HA, Granato M, Fadool JM. Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. Proc Natl Acad Sci USA. 2009; 106:2023–2028. doi:10.1073/pnas.0809439106. [PubMed: 19179291]

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 doi:10.1093/bioinformatics/btu170.

Cahill GM. Circadian regulation of melatonin production in cultured zebrafish pineal and retina. Brain Res. 1996; 708:177–181. [PubMed: 8720875]

Cau E, Quillien A, Blader P. Notch resolves mixed neural identities in the zebrafish epiphysis. Development. 2008; 135:2391–2401. doi:10.1242/dev.013482. [PubMed: 18550717]

Chen P, Hao W, Rife L, Wang XP, Shen D, Chen J, Ogden T, Van Boemel GB, Wu L, Yang M, Fong HK. A photic visual cycle of rhodopsin regeneration is dependent on Rgr. Nat. Genet. 2001; 28:256–260. doi:10.1038/90089. [PubMed: 11431696]

Chi NC, Shaw RM, De Val S, Kang G, Jan LY, Black BL, Stainier DYR. Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. Genes Dev. 2008; 22:734–739. doi:10.1101/gad.1629408. [PubMed: 18347092]

Clanton JA, Hope KD, Gamse JT. Fgf signaling governs cell fate in the zebrafish pineal complex. Development. 2013; 140:323–332. doi:10.1242/dev.083709. [PubMed: 23250206]

Collery R, McLoughlin S, Vendrell V, Finnegan J, Crabb JW, Saari JC, Kennedy BN. Duplication and divergence of zebrafish CRALBP genes uncovers novel role for RPE- and Muller-CRALBP in cone vision. Invest Ophthalmol Vis Sci. 2008; 49:3812–3820. doi:10.1167/iovs.08-1957. [PubMed: 18502992]

Concha ML, Burdine RD, Russell C, Schier AF, Wilson SW. A nodal signaling pathway regulates the laterality of neuroanatomical asymmetries in the zebrafish forebrain. 2000; 28:399–409.

Concha ML, Russell C, Regan JC, Tawk M, Sidi S, Gilmour DT, Kapsimali M, Sumoy L, Goldstone K, Amaya E, Kimelman D, Nicolson T, Gründer S, Gomperts M, Clarke JDW, Wilson SW. Local tissue interactions across the dorsal midline of the forebrain establish CNS laterality. Neuron. 2003; 39:423–438. [PubMed: 12895418]

Elmore MH, Gibbons JG, Rokas A. Assessing the genome-wide effect of promoter region tandem repeat natural variation on gene expression. G3 (Bethesda). 2012; 2:1643–1649. doi:10.1534/g3.112.004663. [PubMed: 23275886]

Fleisch VC, Schonthaler HB, Lintig von, J. Neuhauss SCF. Subfunctionalization of a retinoid-binding protein provides evidence for two parallel visual cycles in the cone-dominant zebrafish retina. J Neurosci. 2008; 28:8208–8216. doi:10.1523/JNEUROSCI.2367-08.2008. [PubMed: 18701683]

Fodor E, Zsigmond Á, Horváth B, Molnár J, Nagy I, Tóth G, Wilson SW, Varga M. Full transcriptome analysis of early dorsoventral patterning in zebrafish. PLoS ONE. 2013; 8:e70053. doi:10.1371/journal.pone.0070053. [PubMed: 23922899]

Fong SH, Emelyanov A, Teh C, Korzh V. Wnt signalling mediated by Tbx2b regulates cell migration during formation of the neural plate. Development. 2005; 132:3587–3596. doi:10.1242/dev.01933. [PubMed: 16033799]

Gamse JT. Directional asymmetry of the zebrafish epithalamus guides dorsoventral innervation of the midbrain target. Development. 2005; 132:4869–4881. doi:10.1242/dev.02046. [PubMed: 16207761]

Gamse JT, Shen Y-C, Thisse C, Thisse B, Raymond PA, Halpern ME, Liang JO. Otx5 regulates genes that show circadian expression in the zebrafish pineal complex. Nat. Genet. 2002; 30:117–121. doi:10.1038/ng793. [PubMed: 11753388]

Gamse JT, Thisse C, Thisse B, Halpern ME. The parapineal mediates left-right asymmetry in the zebrafish diencephalon. Development. 2003; 130:1059–1068. [PubMed: 12571098]

Goff L, Trapnell C, Kelley D. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. 2012:1–65.

Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, Araúzo-Bravo MJ, Lee J, Fishman M, Robbins GE, Lin X, Venepally P, Badger JH, Galbraith DW, Gage FH, Lasken RS. RNA-sequencing from single nuclei. Proc Natl Acad Sci USA. 2013; 110:19802–19807. doi:10.1073/pnas.1319700110. [PubMed: 24248345]

Gross JM, Dowling JE. Tbx2b is essential for neuronal differentiation along the dorsal/ventral axis of the zebrafish retina. Proc Natl Acad Sci USA. 2005; 102:4371–4376. doi:10.1073/pnas.0501061102. [PubMed: 15755805]

Jerome-Majewska LA, Jenkins GP, Ernstoff E, Zindy F, Sherr CJ, Papaioannou VE. Tbx3, the ulnar-mammary syndrome gene, and Tbx2 interact in mammary gland development through a p19Arf/p53-independent pathway. Dev Dyn. 2005; 234:922–933. doi:10.1002/dvdy.20575. [PubMed: 16222716]

Jiang L, Romero-Carvajal A, Haug JS, Seidel CW, Piotrowski T. Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. Proc Natl Acad Sci USA. 2014; 111:E1383–92. doi:10.1073/pnas.1402898111. [PubMed: 24706903]

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. doi:10.1186/gb-2013-14-4-r36. [PubMed: 23618408]

Mano H, Fukada Y. A median third eye: pineal gland retraces evolution of vertebrate photoreceptive organs. Photochem Photobiol. 2007; 83:11–18. doi:10.1562/2006-02-24-IR-813. [PubMed: 16771606]

Manoli M, Driever W. Fluorescence-Activated Cell Sorting (FACS) of Fluorescently Tagged Cells from Zebrafish Larvae for RNA Isolation. Cold Spring Harbor Protocols. 20122012 pdb.prot069633–pdb.prot069633. doi:10.1101/pdb.prot069633.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011; 17:10–12. doi:10.14806/ej.17.1.200.

Masai I, Heisenberg CP, Barth KA, Macdonald R, Adamek S, Wilson SW. floating head and masterblind regulate neuronal patterning in the roof of the forebrain. Neuron. 1997; 18:43–57. [PubMed: 9010204]

Mesbah K, Rana MS, Francou A, van Duijvenboden K, Papaioannou VE, Moorman AF, Kelly RG, Christoffels VM. Identification of a Tbx1/Tbx2/Tbx3 genetic pathway governing pharyngeal and arterial pole morphogenesis. Hum Mol Genet. 2012; 21:1217–1229. doi:10.1093/hmg/ddr553. [PubMed: 22116936]

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 2008; 5:621–628. doi:10.1038/nmeth.1226. [PubMed: 18516045]

Mulligan T, Blaser H, Raz E, Farber SA. Prenylation-deficient G protein gamma subunits disrupt GPCR signaling in the zebrafish. Cell. Signal. 2010; 22:221–233. doi:10.1016/j.cellsig.2009.09.017. [PubMed: 19786091]

Mulligan T, Farber SA. Central and C-terminal domains of heterotrimeric G protein gamma subunits differentially influence the signaling necessary for primordial germ cell migration. Cell. Signal. 2011; 23:1617–1624. doi:10.1016/j.cellsig.2011.05.015. [PubMed: 21699975]

Nadauld LD, Chidester S, Shelton DN, Rai K, Broadbent T, Sandoval IT, Peterson PW, Manos EJ, Ireland CM, Yost HJ, Jones DA. Dual roles for adenomatous polyposis coli in regulating retinoic acid biosynthesis and Wnt during ocular development. Proc Natl Acad Sci USA. 2006; 103:13409–13414. doi:10.1073/pnas.0601634103. [PubMed: 16938888]

Nagashima M, Mawatari K, Tanaka M, Higashi T, Saito H, Muramoto K-I, Matsukawa T, Koriyama Y, Sugitani K, Kato S. Purpurin is a key molecule for cell differentiation during the early development of zebrafish retina. Brain Res. 2009; 1302:54–63. doi:10.1016/j.brainres.2009.09.020. [PubMed: 19748496]

Nagashima M, Saito J, Mawatari K, Mori Y, Matsukawa T, Koriyama Y, Kato S. A hypoplastic retinal lamination in the purpurin knock down embryo in zebrafish. Adv. Exp. Med. Biol. 2010; 664:517–524. doi:10.1007/978-1-4419-1399-9_59. [PubMed: 20238054]

Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011; 12:87–98. doi:10.1038/nrg2934. [PubMed: 21191423]

Quillien A, Blanco-Sanchez B, Halluin C, Moore JC, Lawson ND, Blader P, Cau E. BMP signaling orchestrates photoreceptor specification in the zebrafish pineal gland in collaboration with Notch. Development. 2011; 138:2293–2302. doi:10.1242/dev.060988. [PubMed: 21558377]

Ramsköld D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 2009; 5:e1000598. doi:10.1371/journal.pcbi.1000598. [PubMed: 20011106]

Renninger SL, Gesemann M, Neuhauss SCF. Cone arrestin confers cone vision of high temporal resolution in zebrafish larvae. Eur. J. Neurosci. 2011; 33:658–667. doi:10.1111/j.1460-9568.2010.07574.x. [PubMed: 21299656]

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009; 26:139–140. doi:10.1093/bioinformatics/btp616. [PubMed: 19910308]

Saxena A, Wagatsuma A, Noro Y, Kuji T, Asaka-Oba A, Watahiki A, Gurnot C, Fagiolini M, Hensch TK, Carninci P. Trehalose-enhanced isolation of neuronal sub-types from adult mouse brain. BioTechniques. 2012; 52:381–385. doi:10.2144/0000113878. [PubMed: 22668417]

Schonthaler HB, Lampert JM, Isken A, Rinner O, Mader A, Gesemann M, Oberhauser V, Golczak M, Biehlmaier O, Palczewski K, Neuhauss SCF, Lintig von, J. Evidence for RPE65-independent vision in the cone-dominated zebrafish retina. Eur. J. Neurosci. 2007; 26:1940–1949. doi: 10.1111/j.1460-9568.2007.05801.x. [PubMed: 17868371]

Snelson CD, Burkart JT, Gamse JT. Formation of the asymmetric pineal complex in zebrafish requires two independently acting transcription factors. Dev Dyn. 2008a; 237:3538–3544. doi:10.1002/dvdy.21607. [PubMed: 18629869]

Snelson CD, Santhakumar K, Halpern ME, Gamse JT. Tbx2b is required for the development of the parapineal organ. Development. 2008b; 135:1693–1702. doi:10.1242/dev.016576. [PubMed: 18385257]

Stockhammer OW, Rauwerda H, Wittink FR, Breit TM, Meijer AH, Spaink HP. Transcriptome analysis of Traf6 function in the innate immune response of zebrafish embryos. Mol Immunol. 2010; 48:179–190. doi:10.1016/j.molimm.2010.08.011. [PubMed: 20851470]

Stryer L. Visual excitation and recovery. J Biol Chem. 1991; 266:10711–10714. [PubMed: 1710212]

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods. 2009; 6:377–382. doi:10.1038/nmeth.1315. [PubMed: 19349980]

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 2013; 31:46–53. doi:10.1038/nbt.2450. [PubMed: 23222703]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562–578. doi:10.1038/nprot.2012.016. [PubMed: 22383036]

Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature. 2014 doi:10.1038/nature13173.

Venkateswaran A, Sekhar KR, Levic DS, Melville DB, Clark TA, Rybski WM, Walsh AJ, Skala MC, Crooks PA, Knapik EW, Freeman ML. The NADH Oxidase ENOX1, a Critical Mediator of Endothelial Cell Radiosensitivity, is Crucial for Vascular Development. Cancer Res. 2013 doi: 10.1158/0008-5472.CAN-13-1981.

Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J. The zebrafish transcriptome during early development. BMC Dev Biol. 2011; 11:30. doi:10.1186/1471-213X-11-30. [PubMed: 21609443]

Walker C. Haploid screens and gamma-ray mutagenesis. Methods Cell Biol. 1999; 60:43–70. [PubMed: 9891330]

Wells QS, Becker JR, Su YR, Mosley JD, Weeke P, D'Aoust L, Ausborn NL, Ramirez AH, Pfotenhauer JP, Naftilan AJ, Markham L, Exil V, Roden DM, Hong CC. Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. Circ Cardiovasc Genet. 2013; 6:317–326. doi:10.1161/CIRCGENETICS.113.000011. [PubMed: 23861363]

Xu J, Dodd RL, Makino CL, Simon MI, Baylor DA, Chen J. Prolonged photoresponses in transgenic mouse rods lacking arrestin. Nature. 1997; 389:505–509. doi:10.1038/39068. [PubMed: 9333241]

## Highlights

- Optimized protocols for live cell dissociation and RNA isolation are described
- RNA-sequencing has been adapted for very small numbers of FAC-sorted neurons
- A streamlined and open-source bioinformatics pipeline is utilized for RNA-seq data
- Transcriptomes of WT and *tbx2b*-depleted zebrafish pineal complex were compared
- Differentially expressed genes in the pineal complex were identified and validated

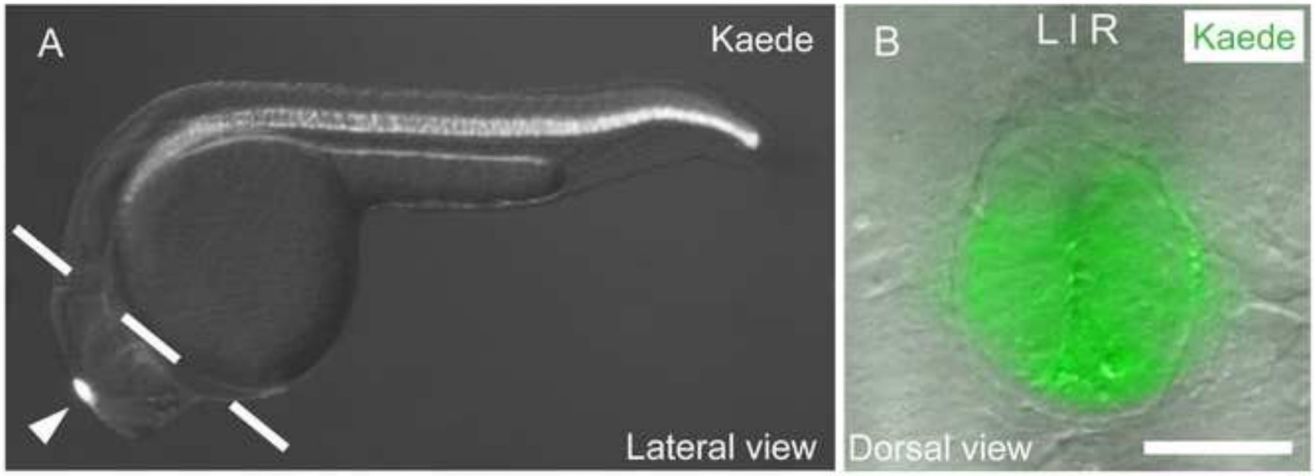


**Figure 1. Expression pattern of Tg[flhBAC:kaede]$^{vu376}$ at 24 hpf**

(A) Kaede is expressed in the notochord and the pineal complex (arrow head). Head tissues are isolated by manual dissection along the dashed line. (B) Dorsal view of the Tg[*flhBAC*:*kaede*]$^{vu376}$ epithalamus showing the fluorescent Kaede channel overlaid on a DIC image. Scale bar: 30 μM

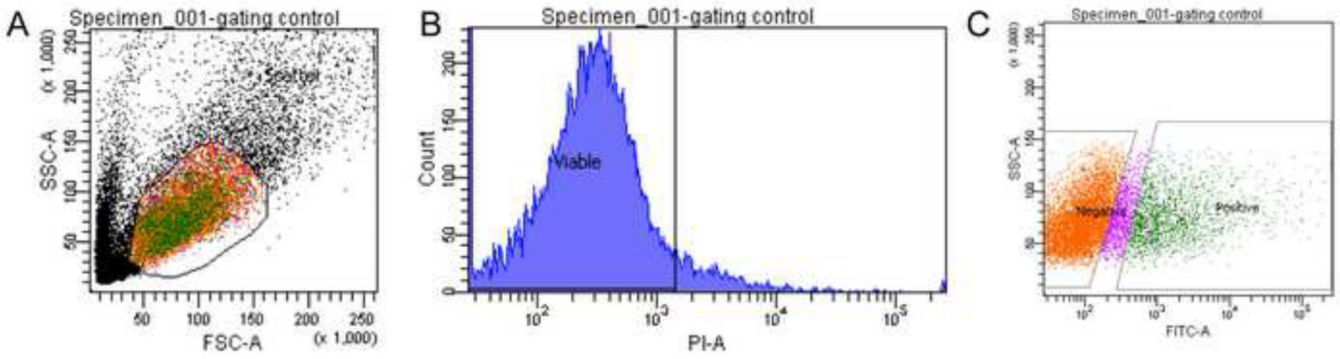


**Figure 2. Fluorescence activated cell sorting of Kaede positive cells from head tissues dissociated from 24hpf Tg[flhBAC:kaede] $^{vu376}$ embryos**

(A) Cells first undergo forward scatter (FSC-A) and side scatter (SSC-A) analysis to be separated based on cell size and internal granularity, respectively. Only the cells from the selected colored region are subsequently sorted for viability. (B) Viable cells do not take up propidium iodine (PI-A). (C) Intact, viable cells then get sorted based on fluorescent intensity.

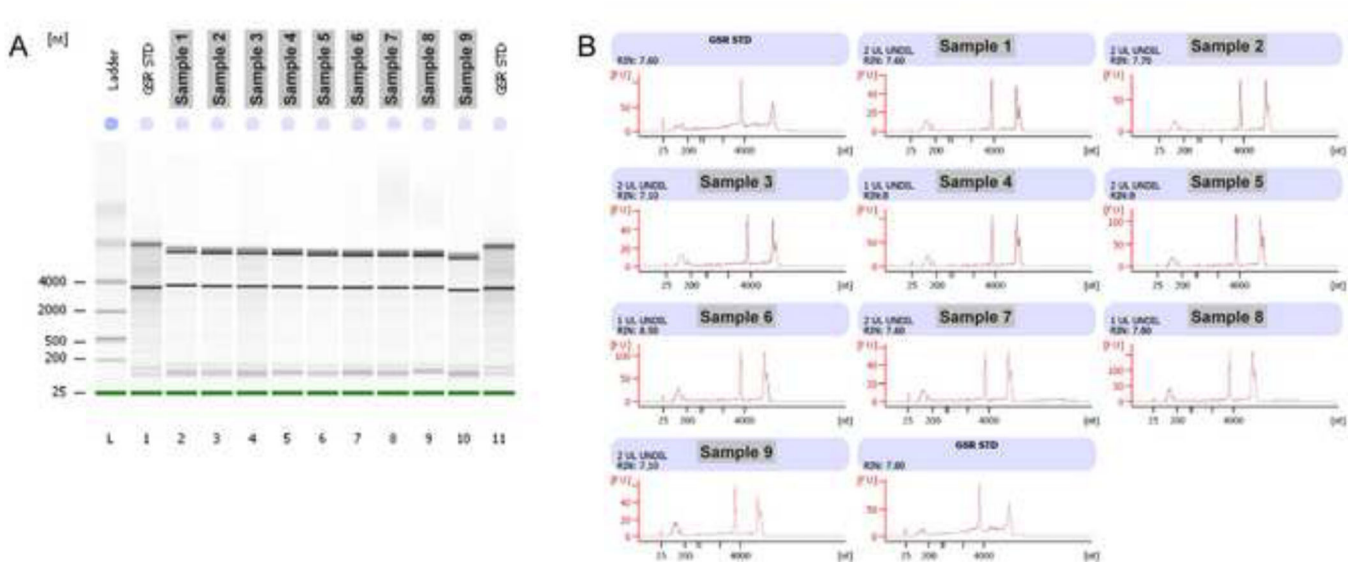


**Figure 3.**

An example of Bioanalyzer reports from an RNA Pico chip for nine different FAC sorted samples extracted using our protocol and as quickly as possible. (A) L: ladder; Lane 1 and 11: Standard controls; Lane 2-10: sample 1-9. (B) Sharp ribosomal RNA peaks indicate that the quality of RNA extracted is consistently good (with RIN of 7.10 or greater), meaning that RNA is not significantly degraded.

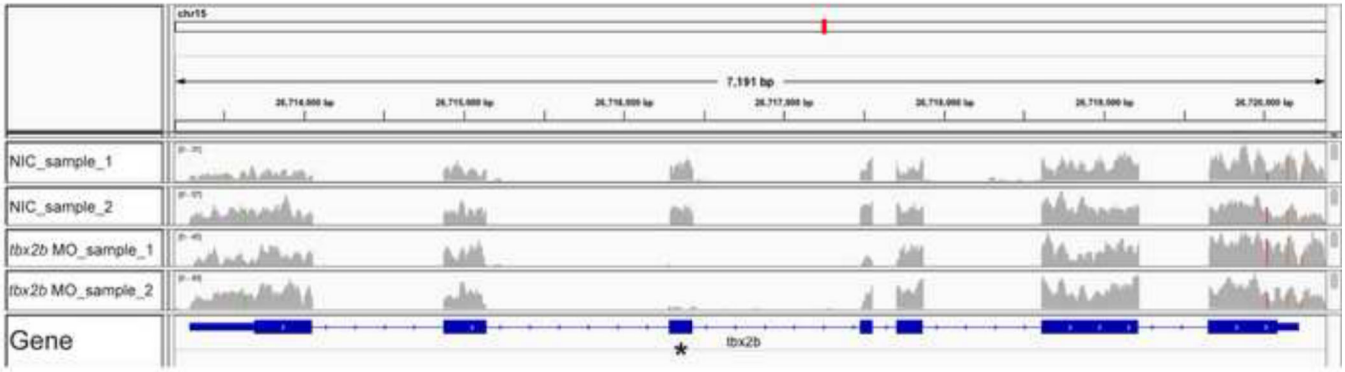


**Figure 4.**

The read coverage of Tbx2b gene viewed using the Integrative Genomics Viewer application shows a decrease in sequencing reads that align to exon 3 (asterisk) in *tbx2b* morphant (MO) samples as compared to non-injected control (NIC) samples.

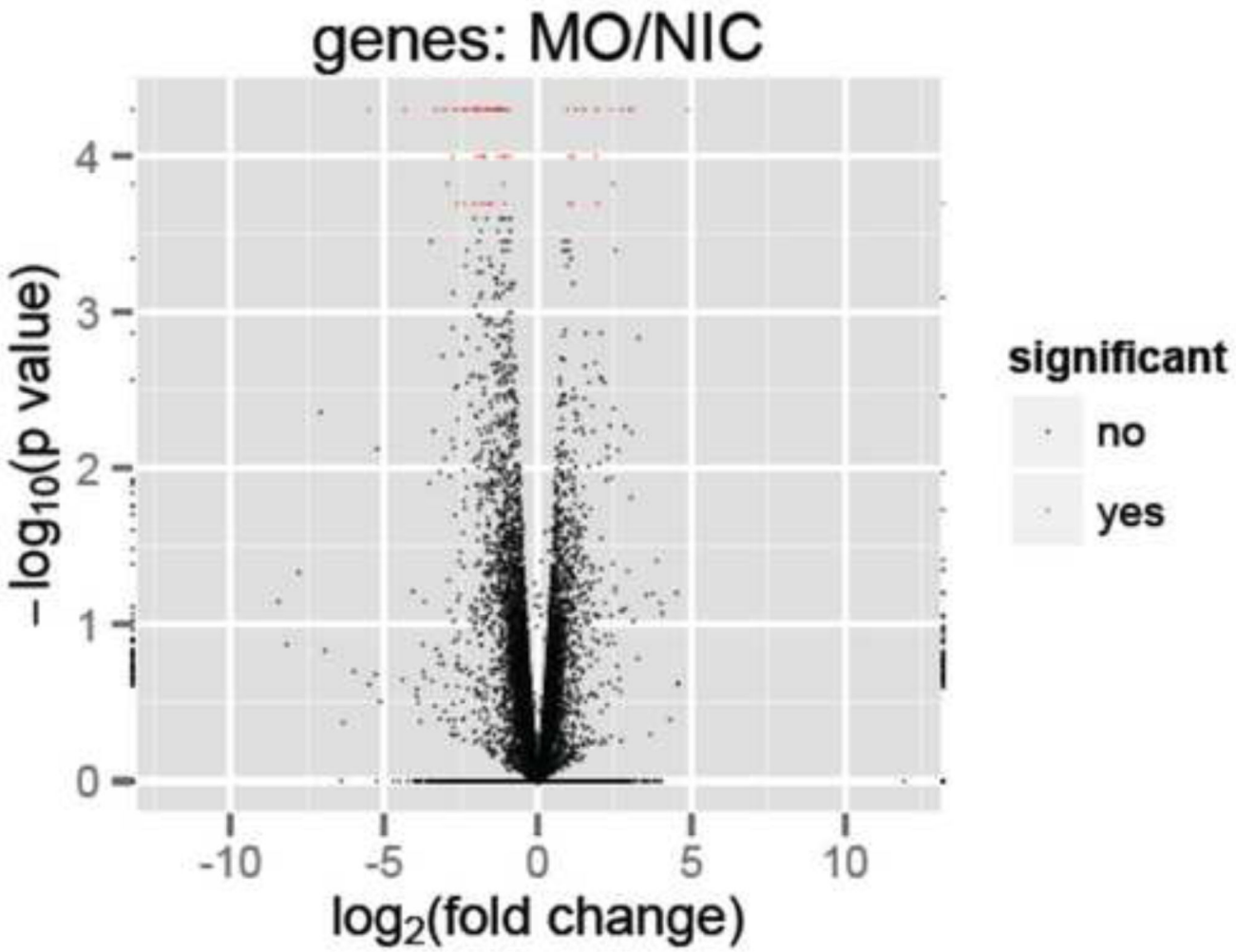


**Figure 5.**
Volcano plot reveal genes that are significantly different (represented by red marks, q 0.05) between control and *tbx2b* morphant conditions.

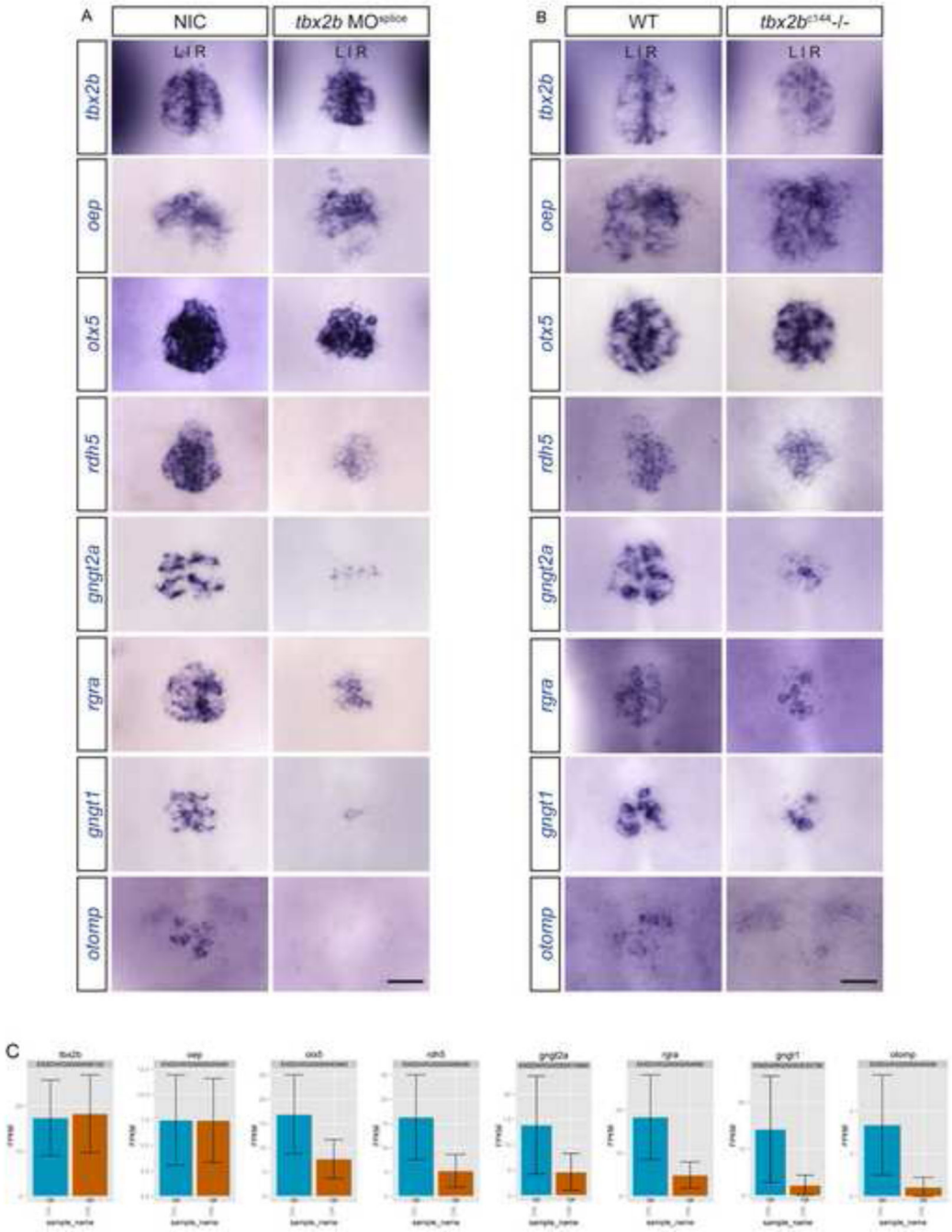


**Figure 6. Differentially expressed genes identified from RNA sequencing experiments were validated using *in situ* hybridization**

At 24 hpf, expression of *otx5, rdh5, gngt2a, rgra, gngt1,* and *otomp,* shown to be downregulated in *tbx2b* morphant pineal transcriptome as compared to controls by 2.2, 3.1, 3.0, 3.7, 6.4, and 7.8 fold, respectively, are reduced in *tbx2b* morphant (A) and *tbx2b*$^{c144}$–/– (B) pineal complex as compared to NIC or WT. *tbx2b* and *oep,* display no significant change in FPKM values from RNA seq experiments (C), and are used as controls. (A and B) Dorsal view of the epithalamus at 24 hpf. Scale bar: 30 μM

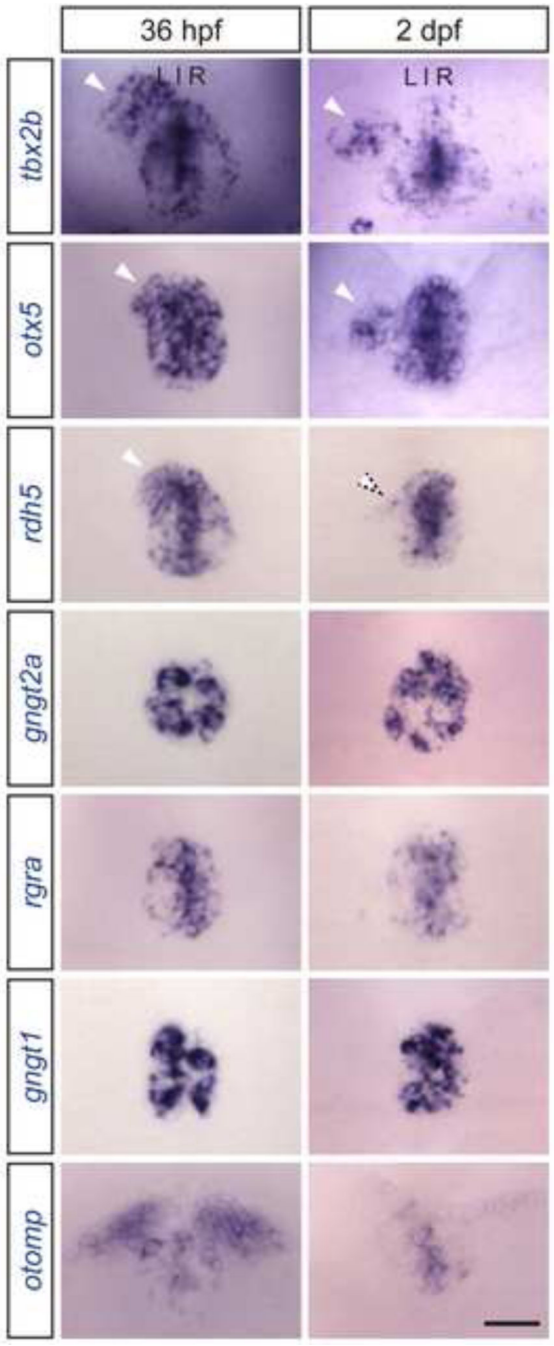


**Figure 7. Tbx2b responsive genes display distinct expression patterns**

At 36 hpf, *tbx2b*, *otx5*, and *rdh5* are expressed throughout the pineal and parapineal organs (arrowheads). At 2 dpf, *tbx2b* and *otx5* expression is maintained in differentiated parapineal cells (arrowheads), but *rdh5* is downregulated (dotted arrowhead). *rgra*, *gngt1*, and *gngt2a* are only expressed in the pineal organ. *otomp* is expressed in only a small population of pineal cells and is also expressed in the habenula nuclei. *tbx2b* expression is maintained

throughout parapineal development and is used as a control. Images are dorsal views of the epithalamus of wild-type embryos at indicated stages. Scale bar: 30 μM

## Table 1

Templates and enzymes used to synthesize antisense RNA probe

| Plasmid name | Reference | Enzyme used to linearized plasmid | RNA polymerase |
|---|---|---|---|
| pBS-SK-*oep*5A1 | Zhang et al., 1998 | HindIII | T3 |
| pCR4-TOPO-*otomp/zomp-1* | Murayama et al., 2005 | NotI | T3 |
| pBS-*otx5* | Gamse et al., 2002 | EcoRI | T7 |
| pCRII-TOPO-*rdh5* | Nadauld et al., 2006 | EcoRV | Sp6 |
| pME18S-FL3-*rgra* | Open Biosystems (Clone ID 8123753) | PCR amplified.<br>Forward primer:<br>5′ TGCTCTAAAAGCTGCGGAAT 3′<br>Reverse primer:<br>5′ TAATACGACTCACTATAGGGGGGAGGTGTGGGAGGTTTT 3′ | T7 |
| pCRII-*tbx2b* | Snelson et al., 2008b | BamHI | T7 |

## Table 2

Differentially expressed genes between *tbx2b* morphants and controls with pineal specific expression

| Gene Symbol | Description | Ensembl Gene ID | (*tbx2b* MO/NIC)[a] | *p*-value[b] | *q*-value[c] |
|---|---|---|---|---|---|
| *otomp* | otolith matrix protein | ENSDARG00000040306 | −7.84 | 5.00E-05 | 0.0176 |
| *gngt1* | guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 1 | ENSDARG00000035798 | −6.41 | 5.00E-05 | 0.0176 |
| *rlbp1b* | retinaldehyde binding protein 1b | ENSDARG00000045808 | −5.37 | 5.00E-05 | 0.0176 |
| *saga* | S-antigen; retina and pineal gland (arrestin) | ENSDARG00000012610 | −4.14 | 5.00E-05 | 0.0176 |
| *sagb* | S-antigen; retina and pineal gland (arrestin) b | ENSDARG00000038378 | −4.00 | 5.00E-05 | 0.0176 |
| *rgra* | retinal G protein coupled receptor a | ENSDARG00000054890 | −3.73 | 5.00E-05 | 0.0176 |
| *rcv1* | recoverin | ENSDARG00000019902 | −3.60 | 5.00E-05 | 0.0176 |
| *rdh5* | retinol dehydrogenase 5 (11-cis/9-cis) | ENSDARG00000008306 | −3.10 | 5.00E-05 | 0.0176 |
| *gngt2a* | guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 2a | ENSDARG00000010680 | −2.98 | 2.00E-04 | 0.0457 |
| *pde6g* | phosphodiesterase 6G, cGMP-specific, rod, gamma | ENSDARG00000022820 | −2.70 | 5.00E-05 | 0.0176 |
| *rbp4l* | retinol binding protein 4, like | ENSDARG00000044684 | −2.45 | 5.00E-05 | 0.0176 |
| *otx5* | orthodenticle homolog 5 | ENSDARG00000043483 | −2.19 | 1.00E-04 | 0.0287 |
| *rpe65a* | retinal pigment epithelium-specific protein 65a | ENSDARG00000007480 | −2.06 | 1.00E-04 | 0.0287 |
| *fsta* | Follistatin a | ENSDARG00000052846 | 2.87 | 5.00E-05 | 0.0176 |

[a] Fold change of FPKM values in *tbx2b* morphants versus non-injected controls; negative values denote reduction; positive values denote increase

[b] The uncorrected *p*-value of the test statistic

[c] The FDR-adjusted *p*-value of the test statistic