Gene 519 (2013) 82–90

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

A methodical microarray design enables surveying of expression of a broader range of genes in *Ciona intestinalis*

Hiromi Matsumae [a], Mayuko Hamada [b], Manabu Fujie [b], Yoshihito Niimura [a], Hiroshi Tanaka [a], Takeshi Kawashima [b,*]

[a] Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo 113-8510, Japan
[b] Marine Genomics Unit and Genome Sequencing Center, Okinawa Institute of Science and Technology Graduate University, Onna, Kunigami, Okinawa 904-0495, Japan

ARTICLE INFO

Article history:
Accepted 24 January 2013
Available online 4 February 2013

Keywords:
Microarray design
Gene expression analysis
Ciona intestinalis
Tissue specific analysis
Cross-platform analysis

ABSTRACT

We provide a new oligo-microarray for *Ciona intestinalis*, based on the NimbleGen 12-plex × 135k format. The array represents 106,285 probes, which is more than double the probe number of the currently available 44k microarray. These probes cover 99.2% of the transcripts in the KyotoHoya (KH) models, published in 2008, and they contain 81.1% of the entries in the UniGene database that are not included in the KH models. In this paper, we show that gene expression levels measured by this new 135k microarray are highly correlated with those obtained by the existing 44k microarray for genes common to both arrays. We also investigated gene expression using samples obtained from the ovary and the neural complex of adult *C. intestinalis*, showing that the expression of tissue-specific genes is consistent with previous reports. Approximately half of the highly expressed genes identified in the 135k microarray are not included in the previous microarray. The high coverage of gene models by this microarray made it possible to identify splicing variants for a given transcript. The 135k microarray is useful in investigating the functions of genes that are not yet well characterized. Detailed information about this 135k microarray is accessible at no charge from supplemental materials, NCBI Gene Expression Omnibus (GEO), and http://marinegenomics.oist.jp.

1. Introduction

Tunicates are key organisms for understanding the evolution of vertebrates because they form a sister clade of the vertebrates. *Ciona intestinalis* has become a well-explored tunicate, having several useful features for studies in developmental biology (Satoh et al., 2003). The draft genome sequence of *C. intestinalis* was determined in 2002, and its 160 Mb genome was estimated to contain approximately 16,000 protein-coding genes (Dehal et al., 2002). The genome sequence greatly accelerated genomic studies of *C. intestinalis*. Following the whole genome sequence, a draft physical map (Shoguchi et al., 2006), a collection of Expressed Sequence Tags (ESTs), and full/partial cDNA sequences (Satou et al., 2003; Satou et al., 2005) were then created. These genomic data have been incorporated into databases such as Ghost (http://ghost.zool.kyoto-u.ac.jp/cgi-bin/gbrowse/kh/) (Satou et al., 2005) and ANISEED (http://www.aniseed.cnrs.fr/) (Tassy et al., 2010).*C. intestinalis* genomic data are also accessible via CIPRO (http://cipro.ibio.jp/), a large-scale protein database (Endo et al., 2011), and CiABA, an anatomical atlas (http://bioinfo.s.chiba-u.jp/ciaba/).

These genomic resources have facilitated the construction of cDNA microarrays (Azumi et al., 2003; Ishibashi et al., 2003) and oligo-DNA microarrays (Christiaen et al., 2008; Hamada et al., 2011; Ishibashi et al., 2005; Yamada et al., 2005) for *C. intestinalis* on the Agilent and Affymetrix platforms. Using these microarrays, gene expression profiles have been analyzed for different stages of the *C. intestinalis* life cycle (Azumi et al., 2007), for maternal transcripts in eggs and early embryos (Hamaguchi et al., 2007; Yamada et al., 2005), for 11 internal organs in adults (Shoguchi et al., 2011), and for neuropeptide- and hormone-encoding genes in larval brain (Hamada et al., 2011). Currently, for studying *C. intestinalis* gene expression, an Agilent microarray with 44,000 (44k) probes is the most widely used platform (Aoyama et al., 2008).

In 2008, new evidence-based gene models, designated KyotoHoya (KH; "hoya" means ascidian in Japanese) models, were published (Satou et al., 2008). The KH models were constructed by considering both gene models (Dehal et al., 2002) and EST data available at the time. The KH models were not available when the Agilent 44k microarray was designed.

We were motivated by the development of the KH models and newly accessible microarray technology to design novel microarrays for *C. intestinalis*. For this purpose, we employed NimbleGen 12-plex microarrays with 135,000 (135k) probes per array. This platform

Abbreviations: ATP, Adenosine-5′-triphosphate; bp, base pair(s); cDNA, complementary DNA; cRNA, complementary RNA; NC, neural complex; oligo, oligonucleotide(s); oligo dT, oligodeoxythymidylate; 3′-end, three prime end; 5′-end, five prime end.

* Corresponding author. Tel.: +81 98 966 8653; fax: +81 98 966 8622.
E-mail address: takeshik@oist.jp (T. Kawashima).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.01.042

triples the number of probes previously available with Agilent microarrays. We chose probes from two independent gene collections, the KH models (Satou et al., 2008) and UniGene entries (Sayers et al., 2012).

This paper describes our design strategy for the new NimbleGen microarray. In order to validate its performance, we compared results from the NimbleGen 135k microarray with those from the Agilent 44k microarray using samples from *C. intestinalis* neural complex and ovary.

Some new design features of the 135k microarray allowed us to detect highly expressed UniGene clusters that were previously undetected. It was also to detect different splicing variants in neural complex and ovary. This methodically designed microarray enabled us to survey the expression of a broader range of genes in *C. intestinalis* than was previously possible.

2. Materials and methods

2.1. Probe design

We employed the NimbleGen 135k × 12-plex format for the newly developed *C. intestinalis* microarray. Probe sequences on the chip are available as GPL14377 on the Gene Expression Omnibus (GEO) website of the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/geo/) and the Okinawa Institute of Science and Technology (OIST), Marine Genomics Unit (http://marinegenomics.oist.jp). We have also provided these sequences as: Supplementary_probefiles.xls.

In this paper, we categorized the probes into five groups: KH(u), KH(l), KH(n), and CKH that are obtained from the KH models, and UG that are obtained from UniGene, but which do not correspond to any of the KH models (Table 1). KH(u) and UG consist of probes that are unique in nucleotide sequence among KH transcripts and UniGene clusters, respectively. Each of the probes in KH(l) is specific to a locus with multiple splicing variants. Probes in KH(n) are "non-unique" probes, while CKH probes were designed for quality control of microarray experiments. Fig. 1 illustrates the process of probe design.

2.1.1. KH(u) (unique)

The KH(u) group contains 37,635 unique (transcript-specific) probes representing 13,597 transcripts from the KH models. We designed up to three probes for each transcript. Each probe sequence in KH(u) is unique among all the transcripts from the KH models.

a) KH(u) probes were created for all of the 24,025 transcripts in the KH models (Fig. 1). We first collected all continuous 60-bp sequences within a 1500-bp region at the 3′-end. These sequences were generated using a sliding 60-bp window, that moved in 1-bp increments along each transcript. We used only the 3′-end, because probes bind more weakly as one approaches the 5′-terminus of a transcript, a phenomenon known as 3′-end bias (Kreil et al., 2006). (Roche, the vendor of NimbleGen microarrays, strongly recommends that users design probes within 1500-bp of the 3′-end.)
b) We next examined the uniqueness of each 60-bp sequence against all of the 24,025 KH model transcripts, using Bowtie ver. 0.12.5, (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2009). This program gives the number of exact matches in these transcript sequences to a given query sequence. We extracted 60-bp sequences having only one exact match among the 24,025 KH transcripts. We used exact match algorithms for probe design instead of similarity search programs, because a probe sequence with a mismatch reduces signal intensity (Gresham et al., 2008).
c) We chose some of the unique sequences obtained above as candidates for probe sequences. First, Roche technicians removed sequences that were inappropriate for technical reasons. Probes can comprise as many as 37 nucleotides on the NimbleGen platform. For example, a 60-bp sequence including 38 adenines is considered an inappropriate sequence for probe design because of its repetitiveness. Secondly, we selected sequences with a GC content of 40–55%, because a GC content in this range optimizes the efficiency of hybridization between the target cDNA and the probe in an oligo-microarray (Letowski et al., 2004).
d) Next, we selected representative probes for each KH transcript using the following criteria. When the number of probe candidates for a given KH transcript was three or fewer, we chose all of them as KH(u) probes. When four or more probe candidates remained for a given KH transcript, we selected the three best from among them, using the following rules. The first probe selected was the one closest to the 3′-end of the transcript. The second probe had to be at least 10-bp from the first probe to avoid choosing a highly similar sequence. If it proved impossible to choose a completely separate probe, we selected the second closest candidate to the 3′-end. The same rules were employed for selection of the third probe (see Supplementary Fig. 1).

2.1.2. KH(l) ("l" designates "locus-specific")

In total, 43,272 locus-specific probes were selected from 4360 loci, covering 11,890 transcripts from the KH models (Table 1). Each probe in this group corresponds to a gene locus with multiple splicing variants (transcripts).

We first identified sequences common to all transcripts derived from a given locus. We then constructed a multiple alignment with ClustalW, using all transcripts derived from each locus (Larkin et al., 2007). Then a "locus-specific" sequence that was common to all transcripts at that locus and was located within 1500-bp from the 3′-end, was extracted for each locus. Among the extracted sequences, we chose "locus-specific" probe sequences in the following manner. Using Bowtie, we identified 60-bp, unique sequences among the locus-specific sequences, as in sections a–b) in KH(u). After eliminating inappropriate sequences, we then determined at least one "locus-specific" probe from among the candidates for each locus-specific sequence (see Fig. 1).

2.1.3. KH(n) ("n" represents "non-unique")

After selection of the KH(u) and KH(l) probes, some KH transcripts still lacked probes. The probes for this group of transcripts were different from the KH(u) and KH(l) probes. Sometimes a 60-bp probe is not long enough to guarantee uniqueness for some transcripts. Therefore, the probes in KH(n) are non-unique, and each of them corresponds to multiple KH transcripts. In total, 3095 probes were chosen for KH(n).

The process of probe design for KH(n) was as follows (Fig. 1). Non-unique 60-bp probe candidates were determined with Bowtie after excluding inappropriate sequences. We chose up to three probes for each of the unassigned-transcripts as described above in d) for KH(u).

Table 1
Summary of the microarray design.

Gene models (probe groups)		Number of loci	Number of transcripts	Number of probes
KH		**15,084**[a]	**23,842**[a]	**89,566**
	KH(u)	12,416	13,597	37,635
	KH(l)	4360	11,890	43,272
	KH(n)	294	1073	3095
	CKH	53	70	**5564**
UG		–	**5573**	**16,719**
Total		**15,084**	**29,415**	**106,285**

Total numbers were calculated as the sum of the figures in bold.
[a] Some transcripts or loci are included in two or more probe groups; therefore, totals represent the number of unique transcripts or loci, not the sum of all models.

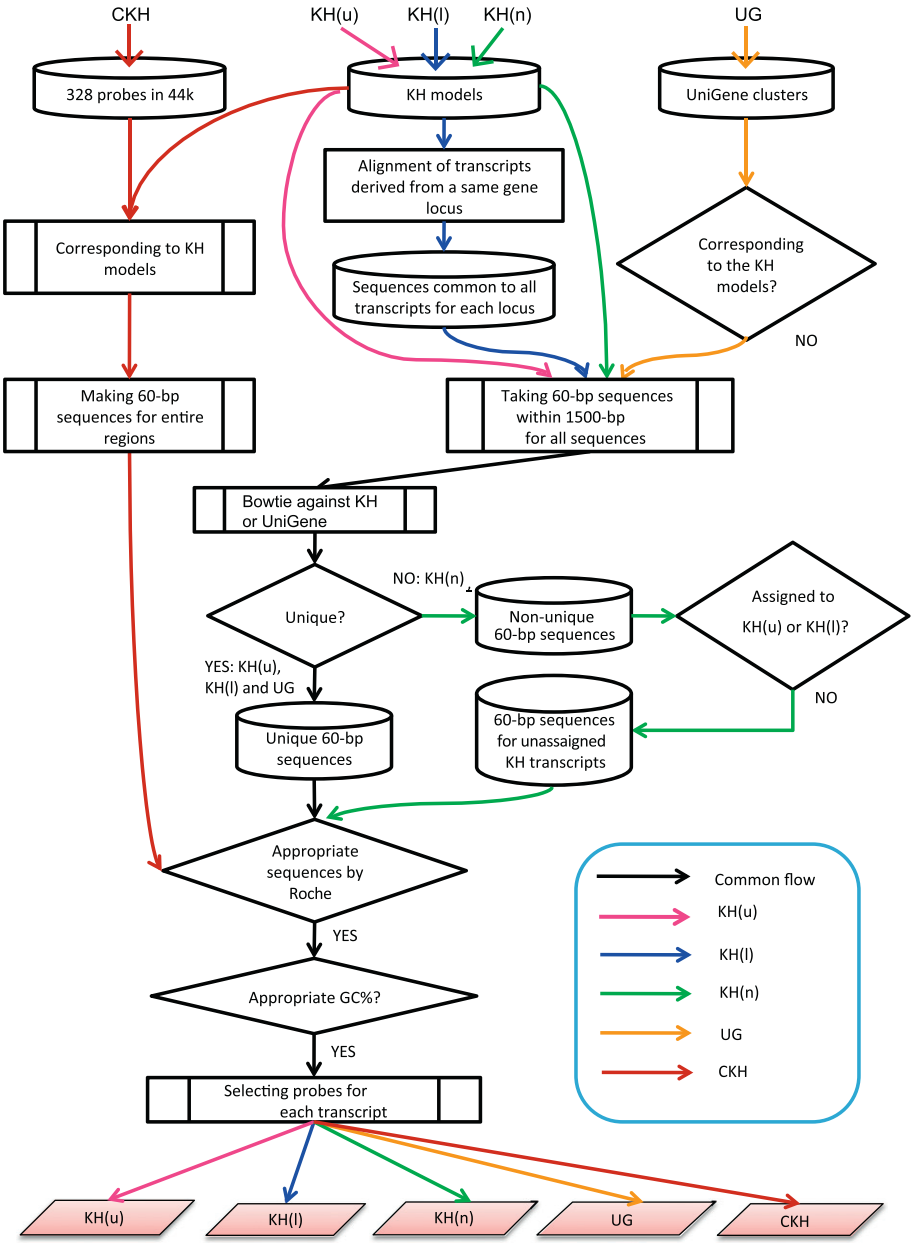

Fig. 1. Workflow of probe design. Overview of the workflow in probe selection for KH(u), KH(l), KH(n) and UG. As shown within the box, each arrow indicates a flow of each group. Only a black arrow indicates the flows in common with three or four groups. Details of this workflow were described in Materials and methods.

After we designed probes for the KH models, 183 transcripts (corresponding to 118 loci) remained unrepresented by any probes (see the Supplementary Information file: SI_unassignedKH.pdf).

2.1.4. UG ("UniGene")

This group contains 16,719 probes constructed from the UniGene clusters (Table 1). From BLASTN searches (Altschul et al., 1990), 7334 sequences (23.8%) of 30,774 UniGene clusters were not highly similar to any KH gene model. Then, we determined unique 60-bp sequences for 5573 of 7334 UniGene-specific transcripts. UG probes for 5573 transcripts were designed using the same procedures described in steps a–d for KH(u) (Fig. 1). Finally, three or fewer probes were chosen for each UniGene sequence.

2.1.5. CKH ("control KH")

Probes in this group were designed as a control for checking the 3′-end bias. CKH probes were constructed from "CKH genes", which will be explained in detail below. In total, 5564 CKH probes were selected. (See the list in Supplementary Table 1.)

We identified CKH genes in various *C. intestinalis* tissues corresponding to 328 probes from the previous 44k study (GSE24941; Shoguchi et al., 2011). First, we used 33 gene expression profiles obtained from muscle, endostyle, intestine, blood cells, heart, neural complex, branchial sac, esophagus, stomach, ovary, and testis. We chose probes showing similar signal intensities among all tissues. In addition, we classified the signal intensities of the probes as high, middle or low.

Cutoff values for high, middle and low were determined from probes falling within the range of percentiles 98.6–100, 38–62 (50 ± 12), and 1–25 in all replicates, respectively, resulting in approximately 100 probes for each level. For example, "high" CKH probes are present within the percentile 98.6–100 for all of the 33 expression profiles. The range of percentile is much smaller for high than that for middle or low. The reason is that highly expressed genes are generally common to any

expression profiles, while probes with middle and low expression levels are highly diverse among expression profiles. As a result, we determined 120, 86 and 122 probes in the 44k for high, middle, and low, respectively.

We then identified 388 KH transcripts corresponding to those 328 probes, using Bowtie. CKH probes were designed for these 388 KH transcripts. We selected CKH probes from the entire region of each transcript, rather than only the 3′-end region, because we intended to use these probes to examine the 3′-end bias. These 60-bp sequences were chosen without considering probe uniqueness and were chosen from anywhere along a given transcript, not just the 3′-end (Fig. 1).

2.2. Naming scheme for probe IDs

We assigned an ID to each of the 135k microarray probes as follows: KH(u), KH(l), KH(n), and UG probe IDs have four fields separated by underscores, e.g., kh_KH.C2.161.v1.A.ND8-1_u_1. The first field "kh" designates probes designed from KH models (for UniGene clusters, "ug" is used). The second field is an accession ID in KH models or UniGene clusters (the #-sign in a UniGene accession was eliminated; for example, ug_CinS20248498_u_1 represents a probe for Cin#S20248498). The third field indicates probe uniqueness: "u" designates a unique probe for KH(u) or UG, "l" is a locus-specific probe for KH(l), and "ov" is a non-unique probe for KH(n). The fourth field shows the probe numbering for each transcript.

An ID for CKH probes comprised three fields; e.g., c1_KH.C8.459.v2.A.SL1-1_1. The first field of a CKH probe indicates an expression level, where "c1," "c38," and "c98" represent low, middle, and high expression levels, respectively. The second field indicates an accession ID from the KH models, and the third field is a probe number for each transcript.

2.3. Samples

Mature *C. intestinalis* adults were obtained from the National BioResource Project in Japan. Animals were kept in fresh seawater at 21 °C under constant lighting to prevent light-induced spawning. Eight pairs of ovaries and neural complexes (eight replicates for each tissue) were sampled from eight individuals and quickly saturated with an adequate amount of RNALater Reagent (Ambion), then held in the reagent at 4 °C overnight. Samples were stored at −30 °C until RNA extraction.

2.4. RNA extraction

Total RNA was extracted using Trizol Reagent (Invitrogen) and precipitated with isopropyl alcohol and high salt (1.2 M NaCl and 0.8 M sodium citrate) to eliminate polysaccharides that are abundant in marine invertebrate tissues. For neural complex, a processing step with 4 μl GenTLE (Takara Bio) as carrier was added, because of the smaller amount of RNA precipitated from one tissue. Total RNA was cleaned up with RNeasy Micro (Qiagen) to avoid genomic DNA contamination, according to the manufacturer's instructions. The condition of the total RNA was assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and a BioAnalyzer (Agilent Technologies).

2.5. Microarray experiments

Total RNA from each of the eight replicates was equally divided into two aliquots for Agilent and NimbleGen Microarray experiments. Eight replicates were analyzed by NimbleGen microarray, while two of the remaining eight replicates were used for Agilent Microarray Analysis. We synthesized labeled cRNA targets with Cyanine-3 CTP from 200 ng of total RNA for Agilent microarrays using a Quick Amp Labeling Kit (Agilent Technologies) for labeling and amplification. The labeled cRNA targets (1.65 μg for each sample) were then used in a hybridization reaction using an In Situ Hybridization Kit Plus (Agilent Technologies). Hybridized microarrays were washed according to the manufacturer's protocol. Arrays were scanned using an Agilent Technologies G2565BA microarray scanner system and Feature Extraction 10.7 software. Reverse transcription and cDNA amplification steps for NimbleGen microarrays were carried out with an Ovation Pico WTA system SL (NuGEN) using 30 ng of total RNA because of the small amount of total RNA obtained from the samples. Labeling reactions and microarray hybridization were carried out using 1 μg cDNA according to the instructions in the NimbleGen Array User's Guide, Gene Expression Arrays, version 51. Hybridized microarrays were scanned with a NimbleGen MS 200 Microarray Scanner and analyzed with version 1.0 software.

2.6. Microarray normalization

Microarrays for the Agilent 44k chips were normalized by R 2.12.2 and limma (ver. 3.6.9) (Bioconductor) for background correlation and quantile normalization (Smyth et al., 2005). For the NimbleGen 135k microarrays, RMA normalization (Irizarry et al., 2003) and gene summarization were carried out using NimbleScan software. Both raw and normalized expression data for Agilent 44k and NimbleGen 135k arrays were recorded in NCBI GEO: GSE33498 and GSE31902, respectively.

2.7. Comparison between 135k and 44k microarrays

The previous version of a customized microarray, designated as *Ciona intestinalis* 44k Oligoarray ver. 2, was designed on the Agilent 44k×4-plex format (Agilent Technologies; NCBI GEO accession no. GPL14686; Hamada et al., 2011; Shoguchi et al., 2011). This chip contains 42,034 oligonucleotide probes representing 19,964 genes.

We first identified the transcripts commonly present in both 44k and 135k microarrays using Bowtie (Langmead et al., 2009). We analyzed the correspondence of 42,034 probe sequences to 23,842 KH transcripts covered in the 135k microarray. We also determined 44k probes corresponding to 5573 UniGene sequences (Fig. 2). A probe corresponding to a given transcript sometimes differed between the two microarrays.

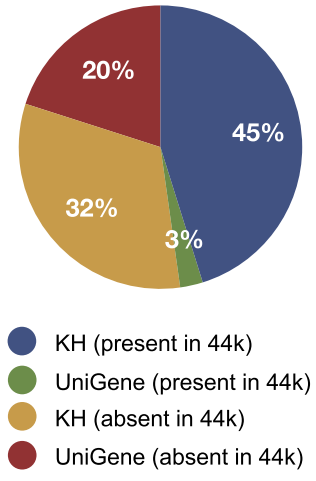

Fig. 2. Composition of transcripts in 135k microarray. Blue and green pie chart areas indicate KH models and UniGene clusters correspond to probes in the 44k microarray. Yellow and red areas represent KH models and UniGenes which were absent in the 44k. To compare common entries between the 135k and the 44k, we summarized the total number of KH loci and transcripts. Total number of KH models included 4360 loci for KH(l) and 14,670 transcripts for KH(u) and KH(n) (Table 1). UniGene clusters included 5573 sequences (Table 1). Finally, total numbers of KH present in 44k, UniGene present in 44k, KH absent in 44k, and UG absent in 44k are 11,115, 639, 7915, and 4934, respectively.

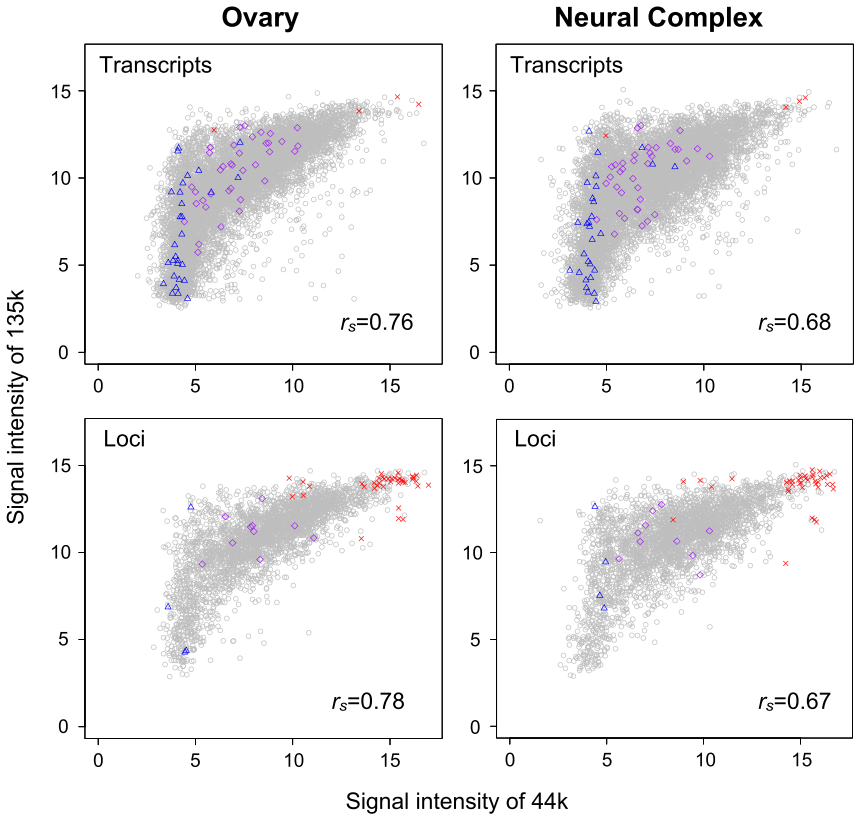

Fig. 3. Correlations of signal intensities between 135k and 44k microarrays. Correlations for the locus-specific probes, corresponding to 2994 loci (top) and unique probes for 7100 transcripts (bottom) for ovary (left) and neural complex (right). X- and Y-axes represent the signal intensities for the 44k and 135k microarrays, respectively. Red, yellow, and blue dots indicate KH transcripts or loci that are ubiquitously expressed at the high, medium, and low levels. Spearman's rank correlation coefficients (r_S) are also shown.

2.8. Statistical analysis

All data were analyzed using R. Spearman's rank test, and correlation coefficients were calculated with the cor.test from R Project for Statistical Computing (http://www.r-project.org). The average signal intensity of probes (Supplementary Fig. 2), which corresponded to transcript-specific and locus-specific probes in the CKH groups for each region (within and beyond 1500-bp), was calculated for genes expressed in common for both the microarrays (Fig. 3) using the KH transcripts, as described above.

Genes preferentially highly expressed in ovary and neural complex (Tables 2 and 3; Supplementary Tables. 2 and 3) were detected by the Rank Products method (Breitling et al., 2004; Jeffery et al., 2006) for R, version 2.22.0. Rank Products provides statistical thresholds with differentially expressed genes, although the results obtained from this method are quite similar to those of fold change analysis. Supplemental Tables 2 and 3 show the top 200 genes expressed preferentially in ovary and neural complex, respectively. Here, genes showing high deviation in expression levels among the eight replicates (i.e., SD was greater than or equal to 1) were excluded. We also identified high expression of genes in both tissues by signal intensity cutoff at 10 (Supplementary Table 4). This cutoff value was based on the median values of signal intensity in the 135k microarray. Median values calculated from distributions of signal intensity in eight replicates for each tissue were less than 10 (Fig. 4). Therefore, we inferred that genes showing signal intensities greater than or equal to 10 in the two tissues were highly expressed in both.

2.9. Gene annotations

Functional annotations for the gene models were retrieved from the original papers and from the Ghost Database (Satou et al., 2005) and Aniseed (Tassy et al., 2010).

3. Results

3.1. Probe design

We designed a newly customized microarray for *C. intestinalis* on the NimbleGen 135k. Details of the procedure for probe design are given in the Materials and methods and in Fig. 1. Probes were designed from the KH models (Satou et al., 2008) and UniGene clusters. This microarray included 106,285 probes selected from 29,415 transcripts (23,842 out of 24,025 transcripts in the KH models and 5573 out of 30,774 UniGene entries) (Table 1). Overall, the newly developed 135k microarray covered 99.2% of the KH transcripts and 81.1% of the UniGene clusters that were not included in the KH models. Among transcripts represented by probes on the 135k microarray, 45% corresponded to both KH models and the probes on the 44k microarray (Fig. 2), while 32% covered only KH models. Together with transcripts from the UniGene clusters, 55% were newly designed probes in this microarray.

We designed five probe groups: KH(u), KH(l), KH(n), UG, and CKH (Materials and methods; Fig. 1). The KH(u) group included "unique" probe sequences that were present only in a single transcript among the 29,415 transcripts from KH models. KH(l) contained "locus-specific" probes that were common to all transcripts derived from a given gene locus, but present only in that locus. KH(n) included "non-unique" probes that corresponded to several transcripts. CKH probes were designed from CKH genes from various tissues with the aim of quality validation. UG probes were chosen from the UniGene clusters that were not included in KH models.

3.2. Comparison of gene expression profiles between 44k and 135k microarrays

We validated the quality of the 135k microarray by comparing gene expression profiles between the previous 44k Agilent microarray and

Table 2
Examples of genes preferentially expressed in ovary to neural complex.

KH accession	Aniseed ID	Description	EST egg[a]	EST gonad[a]	EST testis[a]	Reference
KH.C7.685	aniseedV3_7824	Weakly similar to jagged-1b precursor (jagged1b) (jagged3)	✓			
KH.S2047.1.v1.A.SL1-1	aniseedV3_9237	–		✓		Shoguchi et al. (2011)
KH.C1.1271.v1.A.ND1-1	aniseedV3_3300	Highly similar to transcription factor AP-2 alpha	✓	✓		Nishikata et al. (2001) and Yamada et al. (2005)
KH.C11.95.	aniseedV3_125	Ci-clstr01137	✓	✓		Shoguchi et al. (2011)
KH.S2101.1.v1.A.ND1-1	aniseedV3_1106	Highly similar to FP2/Notch1		✓		
KH.C3.72.v1.A.ND1-1	aniseedV3_1106	Notch1		✓		Shoguchi et al. (2011)
KH.C5.225.v1.A.ND1-1	aniseedV3_7193	Notch1 precursor	✓			
KH.C3.896.v1.C.ND1-1	aniseedV3_1175	Highly similar to neurogenic locus notch homolog protein 1	✓	✓		Shoguchi, et al. (2011)
KH.C1.755.	aniseedV3_275	Ci-pem	✓	✓		Nishikata et al. (2001) and Yamada et al. (2005)
KH.C1.1261.v1.A.ND1-1	aniseedV3_5913	Ferritin 1 heavy chain homolog		✓		
KH.C2.36.v2.A.ND1-1	aniseedV3_37	Ferritin 3 heavy chain	✓	✓	✓	
KH.C2.36.v3.A.ND1-1	aniseedV3_37	Ferritin 3 heavy chain	✓	✓	✓	
KH.C6.199.v1.A.SL1-1	aniseedV3_9412	Ci-connexin-related-13			✓	
KH.L20.1.v1.A.ND1-1	aniseedV3_5510	Ci-GATAa	✓	✓		D'Ambrosio et al. (2003)
KH.C1.1256.v1.A.SL1-1	aniseedV3_1903	Similar to dorsal–ventral patterning tolloid-like protein 1 precursor (mini fin protein)	✓	✓		Yamada et al., 2005 and Shoguchi et al. (2011)
KH.C7.99.	aniseedV3_9325	Abd-B/Hoxd12	✓			
KH.C11.102.	aniseedV3_1334	Ci-piwi-like/piwi				Satou et al. (2001)
KH.C11.656.v1.A.ND1-1	aniseedV3_6342	piwi-like/piwi	✓	✓		
KH.L135.2.v1.A.ND2-2	aniseedV3_532	Ci-VH/Ci-vasa	✓	✓	✓	Fujimura and Takamura (2000)
KH.C2.731.v1.A.ND1-1	aniseedV3_8619	Highly similar to alpha-tectorin		✓		Shoguchi et al. (2011)
KH.C10.90.	aniseedV3_209	HEBP2/complexin-2	✓	✓		Shoguchi et al. (2011)
KH.C5.500.	aniseedV3_477	Zar1	✓	✓		Nishikata et al. (2001) and Shoguchi et al. (2011)
KH.C5.450.v1.A.ND1-1	aniseedV3_477	Zar1	✓	✓		Nishikata et al. (2001)
KH.C1.1047.v1.A.ND1-1	aniseedV3_13688	–		✓		Shoguchi et al. (2011)
KH.C3.563.v1.C.SL1-1	aniseedV3_3435	Ci-CA8/carbonic anhydrase-related protein			✓	
KH.C1.685.v1.A.ND1-1	aniseedV3_7374	Similar to protein gooseberry/retinal homeobox protein Rx3	✓	✓		Imai et al. (2004)
KH.C5.188.v1.A.ND1-1	aniseedV3_12195	Cd69	✓			Shoguchi et al. (2011)
KH.C6.25.	aniseedV3_7380	Wnt/Wnt8	✓			Imai et al. (2004)

[a] Genes with evidence of ESTs in egg, gonad, or testis cDNA libraries is marked. Note, the full list of the preferentially expressed genes is shown in Supplementary Table 2.

our newly designed 135k NimbleGen microarray. We used RNA extracted from ovary or neural complex of the same individual for both microarrays in each experiment. To compare signal intensities between 135k and 44k microarrays, we identified the 44k probes corresponding to KH(u) and KH(l). These included 13,118 and 5661 44k probes corresponding to 7100 KH transcripts and 2994 KH loci, respectively. Spearman's rank correlation coefficients for signal intensities of these common transcripts and loci between 135k and 44k microarrays are 0.67 or greater ($p<2.2\times10^{-16}$), suggesting similar expression estimates for a given transcript or locus between the two microarrays (Fig. 3).

Expression levels of CKH genes are also conserved in the two different microarrays (Fig. 3). The transcripts or loci encoding these CKH genes were categorized into three classes: high, medium, and low, according to their expression levels obtained from the 44k microarray (Supplementary Table 1; Materials and methods). We identified 53 KH loci and 70 KH transcripts in CKH, corresponding to 328 probes of the 44k microarray. Highly expressed CKH transcripts or loci identified by the 44k microarray (represented by red crosses in Fig. 3) also showed high expression levels in the 135k microarray, again confirming the similar expression levels detected by both microarrays.

However, we found that some CHK genes for low, which were identified by the 44k microarray, showed high expression in 135k (Fig. 3). Therefore, we conducted quantitative Real-Time (qRT-PCR) experiments to see whether gene expressions of such genes are actually high or low (Supplementary Materials and Methods). We performed qRT-PCR validation experiments for 9 genes for which signal intensities in neural complex are clearly different between the 44k and 135k microarrays (Supplementary Fig. 3 and Supplementary Materials and Methods). We then compared the fold changes of the expressions of 9 genes among the three measurements, 44k, 135k, and qRT-PCR (Supplementary Table 5). The results showed that the expression levels obtained by the qRT-PCR experiments are closer to those by the 135k microarray than the 44k microarray, suggesting that expression levels obtained by the 135k microarray is reliable.

3.3. Highly expressed genes detected by the 135k microarray

Next, we investigated the signal intensities for the 135k probes representing the KH models that were not covered by the previous 44k microarray. Remarkably, many genes that were not covered by 44k array showed high signal intensity. Overall, signal intensities of the probes that are not covered by the 44k are similar to those covered by the 44k (Fig. 4). These data suggest that a considerable number of genes not covered by the 44k are covered by the new array.

Then, we focused on the top 200 genes that are highly expressed genes identified with the 135k microarray (Supplementary Tables. 2 and 3). Differential expression pattern between the top 200 genes from ovary and from neural complex are clearly discernible.

To focus on tissue-specific expression of genes, we summarized the well-annotated genes with supportive evidence (representative references and the presence of expression data in tissue-specific EST libraries) from the 200 preferentially expressed genes (Table 2, Table 3). We considered ESTs from eggs, gonads, and testes as evidence supporting gene expression in ovary, because in *Ciona* ovarian tissue adheres to testicular tissue. Therefore, the new 135k microarray successfully reproduced results from these previous reports.

3.4. Precise recognition of splicing variants

The 135k microarray produced a number of probes that were not created by the 44k microarray, illustrating that it can detect differences in expression levels among splicing variants that are derived from the same locus. Here, we demonstrate an example of differential

Table 3
Examples of genes preferentially expressed in neural complex to ovary.

KH accession	Aniseed ID	Description	EST NC[a]	Reference (adults NC)
KH.C3.553.v1.A.SL1-1	aniseedV3_7440	ci-six1/2|Six1/2|SIX1|SIX2|so		* Larva only
KH.C1.337.v1.A.SL1-1	aniseedV3_12598	Ci-nAChR-A7/8-1		* Larva only
KH.C1.498.v1.A.SL1-1	aniseedV3_13124	VAChT		* Larva only
KH.C1.498.v2.A.SL1-2	aniseedV3_13124	VAChT		* Larva only
KH.C9.462.	aniseedV3_5604	Ci-Nav1/paralytic/sodium channel protein type 9 subunit alpha	✓	Shoguchi et al. (2011)
KH.C10.27.v1.A.SL1-1	aniseedV3_5431	–	✓	Shoguchi et al. (2011)
KH.C11.495.v1.A.SL1-1	aniseedV3_8522	Rhodopsin-like GPCR, potential chemoreceptor 1/somatostatin receptor/ angiotensin receptor		Deyts et al. (2006)
KH.L10.11.v1.A.ND1-1	aniseedV3_12022	Highly similar to rhodopsin, GQ-coupled/melatonin-related receptor (H9)	✓	
KH.L17.15.v1.A.SL1-1	aniseedV3_7880	COUP/NR2F1		* Larva only?
KH.L17.15.v2.A.SL1-1	aniseedV3_7880	COUP/NR2F1		* Larva only?
KH.C3.635.v1.C.ND1-1	aniseedV3_10628	Calcitonin receptor	✓	
KH.C1.604.v1.A.ND1-1	aniseedV3_12093	Highly similar to contactin 4 precursor (brain-derived immunoglobulin superfamily protein 2) (BIG-2)	✓	
KH.S1035.3.v2.A.ND1-1	aniseedV3_1553	Highly similar to brain protein I3	✓	
KH.S1104.4.v1.A.SL1-1	aniseedV3_6922	Ci-gnrh-X	✓	Kawada et al. (2009)
KH.C13.190.v1.B.ND1-1	aniseedV3_13727 aniseedV3_1867	Highly similar to otogelin	✓	* Savignyi
KH.C2.734.v1.A.SL1-1	aniseedV3_10347	Ci-synaptotagmin7|SYT7	✓	* Larva only
KH.C1.1124.v1.A.ND1-1	aniseedV3_2274	SLC6A14	✓	
KH.S966.4.v1.A.SL1-1	aniseedV3_4575	SLC6A7	✓	
KH.C11.488.	aniseedV3_6240	Ci-Pax1/9		* Larva only
KH.C1.99.v1.A.SL1-1	aniseedV3_1226	SoxB1	✓	* Larva only
KH.C4.57.v1.A.SL1-1	aniseedV3_65	Ci-TnT-a|Ci-TnT-b	✓	Ohshiro et al. (2010)
KH.C12.301.v1.A.SL1-1	aniseedV3_3533	Ci-KIAA1263 protein (fragment)|follistatin-related protein 5 precursor (follistatin-like 5)	✓	Shoguchi et al. (2011)

[a] Genes with evidence of ESTs in neural complex cDNA libraries are marked. The full list of the preferentially expressed genes is shown in Supplementary Table 3.

expression among splicing variants, selected from the preferentially expressed top-200 genes. A gene locus named KH.C1.2 encodes six splice variants in KH models that are similar to inosine-5′-monophosphate dehydrogenase (IMPDH) (Fig. 5). The six probes corresponding to only this locus were designed from five out of 17 exons in this gene. The exon, Ex1, corresponding to two probes (pink lines in Fig. 5), is shared among three of the six variants. Ex1 showed preferential expression in the ovary compared with the neural complex. Other exons also showed similar expression patterns between the two tissues (Figure, pink lines). Splicing variants v4, v5, and/or v6, which contain Ex1, appeared to be expressed specifically in the ovary.

4. Discussion

In this study, we constructed a new 135k microarray with probes that cover almost all sequences in the KH models and UniGene clusters. We also demonstrated the high quality of the 135k microarray with a cross-platform comparison between the 44k and the 135k microarrays (Fig. 3), and by investigating highly expressed genes (Tables 2 and 3).

KH models are the best models currently available and do not contain any genes without strong evidence of ESTs (Satou et al., 2008). However, we found that a considerable number of UG probes, which are not included in the KH models, are rather highly expressed in ovary or neural complex, or both ("Ovary," "Neural Complex," and "Both" include 25, 11, and 726 UniGene clusters, respectively, Supplementary Tables. 2, 3 and 4). This suggests that some genes not covered by the KH models might be functional, meaning that these probes from the 135k microarray would be informative for analyzing *C. intestinalis* genes that are not yet well investigated.

Analysis of differential expression between ovary and neural complex accords well with previous reports (Tables 2, 3 and 4). Preferentially-expressed genes in ovary or neural complex were detected by the 135k microarray These genes in ovary include well-known maternally expressed genes in eggs and gonads, such as Ci-Pem and Ci-vasa (Nishikata et al., 2001; Shoguchi et al., 2011; Yamada et al., 2005) (Table 2). The absence of some maternally expressed genes in Table 2, such as cyclin, histone H1, and ß-catenin, might be seemingly contradictory data. However, these genes are known to be expressed in several tissues. In fact, they are highly expressed not only in ovary but also in neural complex (Table 4; Supplementary Table 4).

The list of preferentially-expressed genes in neural complex included Ci-six, Ci-nAChR, VAChT, Ci-GABAAR, and Ci-gnrh-X (Deyts et al., 2006; Ogasawara et al., 2002; Shoguchi et al., 2011) (Table 3). Kamesh et al. (2008) reported that G-protein coupled receptors (GPCRs), two latrophilin receptors (LPHN1 and LPHN2) and Ci-flamingo (also called cadherin EGF LAG seven-pass G-type receptor, Celsr) are expressed in mammalian brain, while only the LPHN2 ortholog is expressed in the neural complex of adult C. intestinalis, based on EST data. Our results showed the expression of these genes in both neural complex and ovary (Table 4). However, these genes are thought to be neural complex-specific, rather than ubiquitously expressed genes, because eggs generally express a much larger number of genes than other tissues. Therefore, our results are consistent with those of Kamesh et al. (2008).

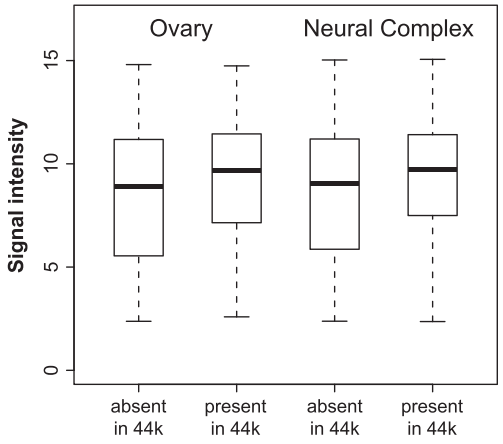

Fig. 4. Signal intensities of probes did not discriminate between correspondences to 44k. Comparison of signal intensities among probes present or absent in 44k. Left and right distributions are signal intensities from ovary and neural complex, respectively. Four distributions include expression of probes corresponding to both KH and UG.

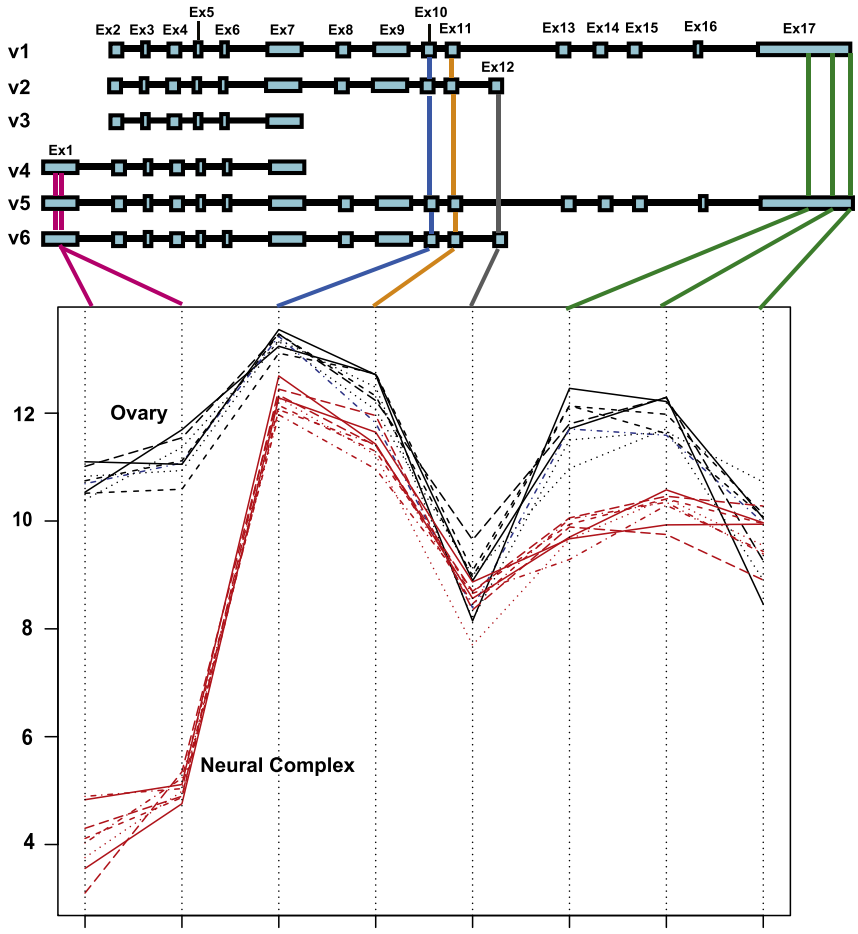

Fig. 5. Comparison of expression in probe levels. a: Structure of six variants derived from KH.C1.2(v1–v6). Exons are numbered Ex1–Ex17. b: Signal intensity of five exons corresponding to v4, v5, and v6. X and Y axes indicate exons and signal intensity, respectively. Ovary and neural complex are colored black and red, respectively. Each tissue has eight replicates (each line indicates a replicate).

Previously, microarrays have been used mainly for examining developmental processes of *C. intestinalis* (Hamada et al., 2011; Hamaguchi et al., 2007; Yamada et al., 2005). However, the functions of many *C. intestinalis* genes still remain elusive. These unexplored genes are expected to be involved in the regulation of tunicate-specific physiological functions. For example, genes related to circadian clock are highly conserved between humans and insects. However, many orthologs of clock-related genes are missing in the tunicate lineage, even though some genes show circadian clock-like behavior and expression patterns (Minamoto et al., 2010). Some clock-related genes may be present in the tunicate lineage, and these may be represented among the highly expressed genes that were observed in the newly designed 135k microarray. Many genes remain to be studied by microarray analysis.

CKH genes from *C. intestinalis*, presented in Supplementary Table 1, were validated in the two different microarrays and therefore represent a useful resource for investigating tissue-specific or lineage-specific phenomena. Notably, highly expressed CKH transcripts mainly consist of highly conserved housekeeping or maintenance genes, such as ribosomal proteins, glyceraldehyde 3-phosphate dehydrogenase, actin, heat shock proteins, and ATP synthase subunits (Supplementary Table 1). This result accords with that of an earlier survey of ubiquitously-expressed genes in human tissues (Warrington et al., 2000).

Moreover, we proposed the usage of CKH probes for technical validation of the 3′-end bias. We used a mixture of oligo-dT and random primers for reverse transcription (Materials and methods). As mentioned above, 3′-end bias is observed when oligo-dT primers are used for reverse transcription, but not random primers (Stangegaard et al., 2006). Therefore, we confirmed the absence of 3′-end bias by comparing the average signal intensities of the CKH probes for each transcript between two regions: within 1500-bp of the 3′-end and outside the 1500-bp region. Signal intensities for the two regions were strongly correlated with each other (Spearman's rank correlation coefficients were $r_s = 0.69$ and 0.80 for ovary and neural complex, respectively), indicating no 3′-end bias (Supplementary Fig. 2).

In summary, this new large-scale microarray design will be of particular value for investigating the functions of unstudied genes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.01.042.

Acknowledgments

We thank Soichi Ogishima for his advice in designing probes from the viewpoint of computational analysis. We also thank the staff of the National BioReseource Center in Japan for supplying animals, and we thank Keisuke Nakashima and Takeshi Noda in OIST for maintaining tunicates in the laboratory. We are grateful to Nori Satoh and Marine Genomics Unit members in OIST for their general support and encouragement. Moreover, we thank the research community of *C. intestinalis* for making the KH models public. Design of probes and training experiments for NimbleGen microarrays were supported by Roche Diagnostics in Japan. We appreciate the administrators at the Genome Server in Department of Bioinformatics at TMDU, Super Computers in Human Genome Center at the University of Tokyo, and the Cstone server at OIST. We also thank S. Aird in OIST for critical reading of the manuscript. This work was supported by KAKENHI (C) 22510209.

Table 4
Examples of genes highly expressed in both ovary and neural complex.

KH accession	Aniseed ID	Description	Reference
KH.C11.125.v1.A.ND1-1	aniseedV3_4240	Highly similar to semaphorin 4G	Shoguchi et al. (2011)
KH.C5.8.v2.A.SL1-1	aniseedV3_6526	Synapsin	Shoguchi et al. (2011)
KH.C4.440	aniseedV3_5311	Myotubularin-related protein 5	Shoguchi et al. (2011)
KH.L116.85.v1.A.ND1-1	aniseedV3_14	beta-Tubulin at 56D/tubulin beta-2C chain	Nishikata et al. (2001)
KH.C9.606.v1.A.SL1-1	aniseedV3_21	beta-Tubulin at 56D/tubulin beta-2C chain	Nishikata et al. (2001)
KH.C4.213	aniseedV3_12	Cyclin|cyclin-B	Nishikata et al. (2001)
KH.L172.2	aniseedV3_294	ci-MRas	Nishikata et al. (2001)
KH.C11.177/KH.C11.121	aniseedV3_55	GAPDH	
KH.C3.79	aniseedV3_569	ci-Calmodulin	
KH.S1251.1	aniseedV3_1068	GSTK1	
KH.L147.22.v1.A.SL1-1	aniseedV3_2420	Ci-MEK1/2	
KH.C5.108.v1.B.ND1-1	aniseedV3_45	EEF2	
KH.C2.593.v1.A.SL1-1	aniseedV3_124	Histone H1	
KH.C1.26.	aniseedV3_708	Ci-CREB1/ATF-c	
KH.C1.412.	aniseedV3_1601	CK II beta	
KH.C8.318.	aniseedV3_2418	CK I alpha	
KH.C6.13.	aniseedV3_4574	Ci-flamingo|Celsr1	Kamesh et al. (2008)
KH.C8.782.v1.A.SL1-1	aniseedV3_10038	Cirl|LPHN1	Kamesh et al. (2008)
KH.L59.10.	aniseedV3_3848	Lphn2	Kamesh et al. (2008)

The full list of the highly expressed genes in both tissues is shown in Supplementary Table 4.

Author's contributions: M.F. determined ubiquitously expressed genes in the previous arrays. H.M., T.K., M.H., and H.T. designed this project. H.M. calculated probe selection. H.M. sampled animal tissues. H.M. and M.H. performed microarray experiments and qRT-PCR. H.M. executed microarray analysis. T.K. maintained the *Ciona intestinalis* genome browser to provide probe data. H.M., N.Y., and T.K. wrote the paper.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 5, 403–410.

Aoyama, M., et al., 2008. A novel biological role of tachykinins as an up-regulator of oocyte growth: identification of an evolutionary origin of tachykininergic functions in the ovary of the ascidian, *Ciona intestinalis*. Endocrinology 149, 4346–4356.

Azumi, K., et al., 2003. Construction of a cDNA microarray derived from the ascidian *Ciona intestinalis*. Zoolog. Sci. 20, 1223–1229.

Azumi, K., et al., 2007. Gene expression profile during the life cycle of the urochordate *Ciona intestinalis*. Dev. Biol. 308, 572–582.

Breitling, R., Armengaud, P., Amtmann, A., Herzyk, P., 2004. Rank Products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 573, 83–92.

Christiaen, L., et al., 2008. The transcription/migration interface in heart precursors of *Ciona intestinalis*. Science 320, 1349–1352.

D'Ambrosio, P., Fanelli, A., Pischetola, M., Spagnuolo, A., 2003. Ci-GATAa, a GATA-class gene from the ascidian Ciona intestinalis: isolation and developmental expression. Dev. Dyn. 226, 145–148.

Dehal, P., et al., 2002. The draft genome of *Ciona intestinalis*: insights into Chordate and Vertebrate origins. Science 298, 2157–2167.

Deyts, C., Casane, D., Vernier, P., Bourrat, F., Joly, J.S., 2006. Morphological and gene expression similarities suggest that the ascidian neural gland may be osmoregulatory and homologous to vertebrate peri-ventricular organs. Eur. J. Neurosci. 24, 2299–2308.

Endo, T., et al., 2011. CIPRO 2.5: *Ciona intestinalis* protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. Nucleic Acids Res. 39, D807–D814 (Database issue).

Fujimura, M., Takamura, K., 2000. Characterization of an ascidian DEAD-box gene, Ci-DEAD1: specific expression in the germ cells and its mRNA localization in the posterior-most blastomeres in early embryos. Dev. Genes Evol. 210, 64–72.

Gresham, D., Dunham, M.J., Botstein, D., 2008. Comparing whole genomes using DNA microarrays. Nat. Rev. Genet. 9, 291–302.

Hamada, M., et al., 2011. Expression of neuropeptide- and hormone-encoding genes in the *Ciona intestinalis* larval brain. Dev. Biol. 352, 202–214.

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

Hamaguchi, M., Fujie, M., Noda, T., Satoh, N., 2007. Microarray analysis of zygotic expression of transcription factor genes and cell signaling molecule genes in early *Ciona intestinalis* embryos. Dev. Growth Differ. 49, 27–37.

Irizarry, R.A., et al., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249–264.

Ishibashi, T., Nakazawa, M., Ono, H., Satoh, N., Gojobori, T., Fujiwara, S., 2003. Microarray analysis of embryonic retinoic acid target genes in the ascidian *Ciona intestinalis*. Dev. Growth Differ. 45, 249–259.

Ishibashi, T., Usami, T., Fujie, M., Azumi, K., Satoh, N., Fujiwara, S., 2005. Oligonucleotide-based microarray analysis of retinoic acid target genes in the protochordate, *Ciona intestinalis*. Dev. Dyn. 233, 1571–1578.

Jeffery, I.B., Higgins, D.G., Culhane, A.C., 2006. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 7, 359.

Kamesh, N., Aradhyam, K.G., Manoj, N., 2008. The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. BMC Evol. Biol. 8, 129.

Kreil, D.P., Russell, R.R., Russell, S., 2006. Microarray oligonucleotide probes. Methods Enzymol. 410, 73–98.

Kawada, T., et al., 2009. A novel inhibitory gonadotropin-releasing hormone-related neuropeptide in the ascidian, Ciona intestinalis. Peptides 30, 2200–2205.

Langmead, B., Trapnel, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

Larkin, M.A., et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Letowski, J., Brousseau, R., Masson, L., 2004. Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarray. J. Microbiol. Methods 57, 269–278.

Minamoto, T., et al., 2010. Circadian clock in *Ciona intestinalis* revealed by microarray analysis and oxygen consumption. J. Biochem. 147, 175–184.

Nishikata, T., et al., 2001. Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis*. Dev. Biol. 238, 315–331.

Ogasawara, M., et al., 2002. Gene expression profiles in young adult *Ciona intestinalis*. Dev. Genes Evol. 212, 173–185.

Ohshiro, K., Obinata, T., Dennisson, J.G., Ogasawara, M., Sato, N., 2010. Troponin in both smooth and striated muscles of Ascidian Ciona intestinalis functions as a Ca^{2+}-dependent accelerator of actin−myosin interaction. Biochemistry 49, 9563–9571.

Satoh, N., Satou, Y., Davidson, B., Levine, M., 2003. *Ciona intestinalis*: an emerging model for whole-genome analyses. Trends Genet. 19, 376–381.

Satou, Y., et al., 2001. Gene expression profiles in Ciona intestinalis tailbud embryos. Development. 128, 2893–2904.

Satou, Y., Kawashima, T., Kohara, Y., Satoh, N., 2003. Large scale EST analyses in *Ciona intestinalis*: its application as Northern blot analyses. Dev. Genes Evol. 213, 314–318.

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., Satoh, N., 2005. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. Zoolog. Sci. 22, 837–843.

Satou, Y., et al., 2008. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. Genome Biol. 9, R152.

Sayers, E.W., et al., 2012. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 40, D13–D25.

Shoguchi, E., et al., 2006. Chromosomal mapping of 170 BAC clones in the ascidian *Ciona intestinalis*. Genome Res. 16, 297–303.

Shoguchi, E., Hamada, M., Fujie, M., Satoh, N., 2011. Direct examination of chromosomal clustering of organ-specific genes in the Chordate *Ciona intestinalis*. Genesis 49, 662–672.

Smyth, G.K., Michaud, J., Scott, H.S., 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067–2075.

Stangegaard, M., Dufva, I.H., Dufva, M., 2006. Reverse transcriptin using random prentadecamer primers increases yield and quality or resulting cDNA. Biotechniques 40, 649–657.

Tassy, O., et al., 2010. The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. Genome Res. 20, 1459–1468.

Warrington, J.A., Nair, A., Mahadevappa, M., Tsyganskaya, M., 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol. Genomics 2, 143–147.

Yamada, L., Kobayashi, K., Satou, Y., Satoh, N., 2005. Microarray analysis of localization of maternal transcripts in eggs and early embryos of the ascidian, *Ciona intestinalis*. Dev. Biol. 284, 536–550.