Published in final edited form as:

Gene Expr Patterns. 2008 February ; 8(3): 171–180.

A Web Based Resource Characterizing the Zebrafish Developmental Profile of over 16,000 transcripts

Ming Ouyang^{1,#}, Aaron T. Garnett², Tina M. Han², Kotaro Hama³, Amy Lee^{3,4}, Yun Deng^{3,4}, Nancy Lee⁴, Hsing-Yin Liu⁴, Sharon L. Amacher², Steven A. Farber^{3,*}, and Shiu-Ying Ho^{4,*}

¹ Department of Pharmacology, UMDNJ, Piscataway, NJ

² Department of Molecular and Cell Biology and Center for Integrative Genomics, Berkeley, CA

³ Department of Embryology, Carnegie Institution, Baltimore, MD

⁴ Department of Biochemistry and Molecular Biology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA

Abstract

Using a spotted 65-mer oligonucleotide microarray, we have characterized the developmental expression profile from mid-gastrulation (75% epiboly) to 5 days post-fertilization (dpf) for >16,000 unique transcripts in the zebrafish genome. Microarray profiling data sets are often immense, and one challenge is validating the results and prioritizing genes for further study. The purpose of the current study was to address such issues, as well as to generate a publicly available resource for investigators to examine the developmental expression profile of any of the over 16,000 zebrafish genes on the array. On the chips, there are 16,459 printed spots corresponding to 16,288 unique transcripts and 172 β -actin (AF025305) spots spatially distributed throughout the chip as a positive control. We have collected 55 microarray gene expression profiling results from various zebrafish laboratories and created a Perl/CGI-based software tool (<http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>) for researchers to look for the expression patterns of their gene of interest. Users can search for their genes of interest by entering the accession numbers or the nucleotide sequences and the expression profiling will be reported in the form of expression intensities versus time-course graphical displays. In order to validate this web tool, we compared seventy-four genes' expression results between our web tool and the *in situ* hybridization results from Thisse *et al.* (2004) as well as those reported by Mathavan *et al.* (2005). The comparison indicates the expression patterns are 80% and 75% in agreement between our web resource with the *in situ* database (Thisse *et al.* 2004) and with those reported by Mathavan *et al.* (2005), respectively. Those genes that conflict between our web tool and the *in situ* database either have high sequence similarity with other genes or the *in situ* probes are not reliable. Among those genes that disagree between our web tool and those reported by Mathavan *et al.* (2005), 93% of the genes are in agreement between our web tool and the *in situ* database, indicating our web tool results are quite reliable. Thus, this resource provides a user-friendly web based platform for researchers to determine the developmental profile of their gene of interest and to prioritize genes identified in microarray analyses by their developmental expression profile.

*Corresponding authors: farber@ciwemb.edu and sho@mail.jci.tju.edu.

#Current address: Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Microarray profiling is a powerful technology that enables researchers to compare gene expression across samples (Lee and Saeed, 2007). In order to study gene expression in zebrafish embryos at various developmental stages, we have printed zebrafish DNA microarray chips at the Kimmel Cancer Center (Thomas Jefferson University) comprising more than 16,000 unique transcripts from the Zebrafish oligo library from Compugen/Sigma-Genosys. The Kimmel Cancer Center microarray profiling service has been used by many investigators in the zebrafish community to investigate a variety of interesting biological problems and proven to be an useful resource for the community (Linney *et al.*, 2004, Sumanas *et al.*, 2005, Covassin L *et al.*, 2006, Kassen *et al.*, 2007, Kreiling, *et al.*, 2007, Robu *et al.*, 2007). Using a spotted 65-mer oligonucleotide microarray, we have characterized the developmental expression profile from mid-gastrulation (75% epiboly) to 5 days post-fertilization (dpf) for 16,288 unique transcripts in the zebrafish genome, representing about a third to one half of all genes in the genome, according to current predictions (zebrafish genome project zv7, http://www.sanger.ac.uk/Projects/D_rerio/Zv7_assembly_information.shtml). We have collected 55 independent RNA sets of zebrafish gene expression profiles spanning 75% epiboly to 5 dpf from several zebrafish laboratories (Amacher, Farber, Ho) and created a useful web resource tool (<http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>) for researchers. One of the challenges in analyzing microarray data is prioritizing and validating positives for further study; our web resource tool allows researchers to more quickly refine their microarray data. In addition, the database provides “virtual data” for those interested in studying groups of genes with distinct developmental expression profiles.

2. Results and Discussion

Our Perl/CGI-based software tool (<http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>) provides an easy-to-use web based platform for users to determine the developmental profile of their gene of interest. The established web tool has three basic functionalities. First, the user can determine which probe on the microarray chip corresponds to the DNA sequence of interest. The web tool will look up the closest matching probe(s) and report the corresponding GenBank Accession Number(s). Second, the user can examine the developmental profile of the transcript by entering the GenBank Accession Number. The example in Figure 1 shows the developmental expression profile for BE201395, a gene with similarity to *myosin regulatory light chain 2*. The tool displays three figures: the raw intensities, the normalized intensities, and the intensities relative to β -actin. The scanned and digitized data are processed as follows: the local background intensities are subtracted from the spot intensities, and the resulting values are called “Raw Intensities” (Figure 1A). The raw intensities are divided by the median intensity of the chip, and the resulting values are called “Normalized Intensities” (Figure 1B). The normalized intensities are divided by the median of the β -actin intensities of the chip, and the resulting values are called “Intensities Relative to β -actin” (Figure 1C). All three kinds of intensities are reported as mean and standard error of the replicates, with time-course graphical display. Since the β -actin expression levels are slightly increased after 24 hpf during the developmental stages, the intensities relative to β -actin are slightly different than raw intensities or normalized intensities. Finally, the top 30 positively correlated and the top 30 negatively correlated transcripts are displayed in the search results (Fig. 1 D). Another functionality of the tool is to perform a BLAST search of the NCBI nonredundant (nr) database (data not shown).

In addition to obtaining expression profiles of individual genes, users may also use the web interface to obtain information about groups of genes with similar developmental expression profiles. By comparing the profiles of all genes on the microarray, genes expressed at similar

times during development were clustered into 10 groups (called “nodes”) using the program Cluster 3.0 (de Hoon *et al.*, 2004, Eisen *et al.*, 1998) using K-means clustering with 1000 runs and an uncentered Pearson correlation metric. The results were visualized using the program Maple Tree (<http://rana.lbl.gov/EisenSoftware.htm>) and are displayed on the website (Fig. 2). For instance, genes in Node 2 are expressed at high levels during gastrulation, declining to basal levels after the first day of development, whereas genes in Node 5, such as the *amylase α2a* gene (BM103972) that encodes a pancreas-enriched enzyme (Fig. 3), are not expressed until late larval stages. BE201395, the *myosin regulatory light chain 2*-like gene profiled in Fig. 1A–C, belongs to Node 8, representing genes with peak expression at 48 hpf. We anticipate that users will use node information to prioritize genes within their microarray datasets for further study.

Investigators can use the website to obtain gene expression levels at developmental stages from 75% epiboly to 5 dpf. To validate the gene expression profiling, we chose 74 genes (34 genes with gene names starting with the letter a, and 40 genes at random) and compared their expression profiles on the microarray with the *in situ* hybridization expression data reported by Thisse *et al.* (2004) and archived by ZFIN

(<http://zfin.org/cgi-bin/webdriver?Mival=aa-pubview2.apg&OID=ZDB-PUB-040907-1>). Basically, we assigned a score of 1 to those expression profiles that correlate, a score of 0 to those that differ, and no score (n) to those time points with no *in situ* hybridization data (Table 1). The summarized scores were divided by the total number of time points. Of the genes in this survey, 80% show an agreement between our microarray data and the *in situ* hybridization results from Thisse *et al.* (2004). Among these 74 genes, 10 genes exhibit an average score less than or equal to 0.5. Further analysis of the 10 low-scoring genes revealed that 5 of them have more than one hit on the microarray chip with E values from 0 to e^{-6} range. These results suggest that the 5 low-scoring genes (zgc:92191, zgc:56709, acvr2b, zgc:64120, zgc:77800) likely have high sequence similarity to other genes. The other 5 low-scoring genes (zgc:113518, zgc:101040, zgc:92530, zgc:92313, zgc:85659) either display weak expression in the *in situ* hybridization results or the probe has not been validated. Of the 64 high-scoring genes, 34 genes (53%) have developmental expression profiles that fully correlate when microarray and *in situ* hybridization data are compared. For example, the microarray data indicate a muscle-expressed gene related to *myosin regulatory light chain 2* (zgc:103639; BE201395) is not expressed before the 18-somite stage (Fig. 1 A–C). The *in situ* hybridization results also indicate that BE201395 is not expressed before 20 somites

(http://zfin.org/cgi-bin/webdriver?Mival=aa-fxallfigures.apg&OID=ZDB-PUB-051025-1&fxallfig_probe_zdb_id=ZDB-GENE-051025-1). The expression of BE201395 peaks at 2 dpf and is lower at 5 dpf in our array data (Fig. 1 A–C) and correlates well with the *in situ* hybridization data which shows robust expression from late segmentation stages (20–25 somites) to high-pec (42 hpf) and long-pec stages (48 hpf), with potentially weaker expression at 5 dpf stage. Likewise, the array data indicate that the enzyme *amylase α2a* (BM103972) is initially expressed at 4 dpf, increasing substantially by 5 dpf (Fig. 3 A–C), which agrees nicely with *in situ* hybridization data showing pancreas-specific expression at 5 dpf and no expression at early developmental stages. We chose the above two examples because their expression patterns have been documented across all the developmental stages; expression patterns for many of the genes having ‘perfect’ scores in Table 1 are not available for later developmental stages (i.e. 5 dpf) in the *in situ* hybridization database. We also compared these 74 genes’ expression patterns with those reported by Mathavan *et al.* (2005), which used the same Compugen oligo set for their microarrays. We found that 75% of the genes are well-matched between the two databases (Table 2). We further compared the low-scored 15 genes (≤ 0.5) in the table 2 with those in the table 1, only one gene (acvr2b) has low score in both tables (table 1 and table 2), suggesting our web tool results are relatively reliable. In addition, we have performed quantitative RT-PCR on a randomly selected set of genes and compared these results with data obtained using the web tool. Our quantitative

RT-PCR of *β-actin* (AF025305) showed a similar pattern with the retrieved web data which showed slightly increased levels after 24 hpf (Fig. 4A). The quantitative RT-PCR profiles of the selected Node 2 gene (BM104515, *s100 calcium binding protein A1*), Node 3 gene (AW421995, *fat-free* gene), Node 4 gene (BG738899, *vacuolar protein sorting 26* homolog B) and Node 5 gene (AF254642, *fatty acid binding protein 10, liver type*) also parallel the microarray data, thus further validating our web resource (Fig. 4B–E). Furthermore, our web resource provides additional time points at later developmental stages (i.e. 3 to 5 dpf) and displays easy-to-read virtual data for a gene of interest. We hope that we can add more data to this resource and extend an invitation to other Compugen oligo set users to contact us by email so we can determine if their data can be added. In conclusion, we have established a user-friendly web-based resource that enables researchers to obtain information regarding zebrafish gene expression patterns during early development.

3. Experimental Procedure

3.1. Total RNA isolation

Zebrafish (*Danio rerio*) embryos were collected and staged from 75% epiboly to 5 days post fertilization (dpf). Total RNA was extracted by TRIZOL (Invitrogen).

3.2. Zebrafish Microarray

The customized Gene Expression microarrays are transcriptional profiles designed to monitor the mRNA levels of multiple genes. The technology of Zebrafish microarray used by the KCC/TJU for expression profiling is a single color system similar to Affymetrix's technology. The microarrays utilize nucleic acid hybridization of a biotin-labeled complementary RNA (cRNA) target with DNA oligonucleotide probes attached to a gel matrix.

3.3. Array Design

The commercial Zebrafish oligo library from Compugen/Sigma-Genosys consists of 16,399 oligos (65nt) representing 16,288 unique gene clusters. The oligos are modified with a 5' amino C6 linker. The library includes 172 zebrafish *β-actin* oligos distributed over the entire library as internal controls (approximately 4 per 384 well plate). Amersham CodeLink™ activated glass slides were prepared using a hydrophilic polymer contain N-hydroxysuccinimide ester groups covalently attached to the silane basecoat by the vendor. The oligos were resuspended in 50 mM phosphate buffer, pH 8.0, at 20 mM concentration. The individual oligo-probes were printed on CodeLink™ activated slides under 45% humidity by GeneMachine OmniGrid™ 100 Microarrayer in 4 × 12 pin configuration and 20 × 20 spot configuration of each subarray. The spot diameter was 100 μm and distance from center to center was 200 μm. The printed Zebrafish microarrays were further immobilized via covalent coupling methods at 70% humidity overnight. Samples were hybridized after additional 100 mM Taurine/Bicine blocking and 4X SSC/0.1% SDS washing steps.

3.4. Target Preparation Protocol

The biotin-labeled cRNA target is prepared by a linear amplification method. Poly (A) RNA from 5ug of total RNA is primed for reverse transcription by a DNA oligonucleotide containing a T7 RNA polymerase promoter 5' to a d(T)24 sequence. After second-strand cDNA synthesis, the cDNA serves as the template in an *in vitro* transcription (IVT) reaction to produce the target cRNA. The IVT is performed in the presence of biotinylated nucleotides to label the target cRNA: Total RNA (5 ug) was incubated with T7 primer at 70°C for 10 minutes before addition of Superscript reverse transcriptase reaction mix and further incubation at 37°C for 1 hour. The first-strand cDNA was added to polymerase mix and incubated in a TropiCooler incubator at 16°C for 2 hours. Double-strand cDNA was precipitated in isopropanol at –80°C for 20 min,

centrifuged at 14k rpm for 20 min, and pellet resuspended in ddH₂O after air-drying for 30 minutes. *In vitro* transcription was performed using the Ambion MEGAscript T7 kit. Biotinylated UTP, NTPs and enzyme reaction mix were added to dried cDNA and incubated at 37°C overnight. Synthesized cRNA was purified using the Qiagen RNeasy kit. Concentration and quality of cRNA was measured by an Eppendorf Biophotometer. The target labeling by the KCC/TJU standard procedure results in a 50 – 200 fold linear amplification of the input poly(A) RNA.

3.5. Hybridization Procedures and Parameters

Biotin-labeled cRNA, 10 µg of each target, was used for hybridization on each KCC/TJU 17K Zebrafish oligo expression microarray on a TECAN HS4800 Hybridization Station. The microarrays were hybridized in 6X SSPE with 50% formamide at 37°C for 20 hours, washed in 0.75X TNT at 46°C for 1 hour, followed by blocking in TNB at RT for 30 minutes, and then processed by using direct detection method of the biotin-containing transcripts by Streptavidin-Alexa 647 conjugate in TNB buffer (1:500) at RT for 30 minutes. Stained chips were washed in 1X TNT wash buffer at room temperature for 60 minutes with changing fresh TNT buffer every 15 minutes.

3.6. Array Data processing

Processed chips were scanned by using a Perkin Elmer ScanArray® XL5000, software version 3.1, with the laser set to 635 nm, at power 90 and PMT 70-50 setting, and a scan resolution of 10 microns. Images were quantified by PerkinElmer QuantArray® Software 3.0. The fixed circle quantification method was used and output was given as total intensities. Confidence calculation was weighted average. The median normalized data from control and testing samples were compared to each other in fold change for further identification of differentially expressed genes in Excel spreadsheet format.

3.7. Web-based resource establishment

The profiling of zebrafish development is conducted at seven time points: 75% epiboly, 18 somite, 24 hpf, 48 hpf, 3 dpf, 4 dpf, and 5 dpf. The number of independent RNA samples that were applied to the microarrays for each selected developmental time point were 3, 4, 8, 13, 11, 8, and 8, respectively. For each of these 55 microarrays, the scanned and digitized data are processed as follows: First, the local background intensities are subtracted from the spot intensities resulting in raw intensities. Second, the raw intensities are divided by the median intensity of the chip resulting in normalized intensities. Third, the normalized intensities are divided by the median of the *β-actin* intensities, and the resulting values are the intensities relative to *β-actin*. Then, the replicates at each time point are combined by calculating the mean and standard error for each transcript. We constructed a web tool to facilitate access to and visualization of the data. The tool is implemented using HTML/Perl/CGI, and the URL of the tool is <http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>. In addition to cluster analysis, we use Pearson correlation in gene to gene time-course profiles to capture detailed information of co-regulation during the various developmental stages. The positively correlated genes are those genes that have the same expression patterns to the query gene; whereas the negatively correlated genes are those have opposite expression patterns.

3.8 Quantitative PCR

The SYBR Green master mix reagents (Applied Biosystems) were used on the ABI PRISM 7900HT or the DNA Engine OPTICOM according to the manufacturers instructions. The transcripts were amplified using the following forward and reverse primers. AF025305: 5'-ATGGATGATGAAATTGCCGC-3' and 5'-AGTTGGTGACAATACCGTGC-3'; BM104515: 5'-TACAAAGCTA AGCAAATCTGAGCT-3' and 5'-ATTGCATGCCAC

AGTGAGA-3'; AW421995:5'-ATGAGTTCAGCCACAACAC-3' and 5'-AAAGTCATTCTTCATTTTCTG-3'; BG738899:5'-ACCTCATTGTTTCATCAGCTG-3' and 5'-ACCTCATTGTTTCATCAGCTG-3'; AF254642: 5'-ATCCAGCAGAACGGCAGC-3' and 5'-ACGATGCACTTGAGCTTCTT-3'. The threshold cycle (C_T value) was determined for each transcript at designated time point. The relative expression levels of each transcript were calculated by the comparative C_T method (Livak and Schmittgen, 2001).

Acknowledgements

The authors want to thank the Microarray Facility at Kimmel Cancer Center (Thomas Jefferson University) for the microarray experiments (Drs. Chang-Gong Liu and Jennifer Ma) and Jennifer Anderson at the Carnegie Institution for helpful comments on the manuscript. This work was supported by an American Heart Association SDG award (0730083N) to S.Y.H., a US National Institute of Health Grant (DK060369) to S.A.F., and a March of Dimes grant (1FY05-118) to S.L.A.. T.M.H. was supported by an American Cancer Society Postdoctoral Fellowship and A.T.G. was supported by the UC Berkeley Center for Integrative Genomics.

References

- Covassin L, Amigo JD, Suzuki K, Teplyuk V, Straubhaar J, Lawson ND. Global analysis of hematopoietic and vascular endothelial gene expression by tissue specific microarray profiling in zebrafish. *Dev Biol* 2006;299:551–62. [PubMed: 16999953]
- de Hoon MJL, Imoto S, Nolan J, Miyano S. Open Source Clustering Software. *Bioinformatics* 2004;20(9):1453–1454. [PubMed: 14871861]
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *PNAS* 1998;95:14863–14868. [PubMed: 9843981]
- Kassen SC, Ramanan V, Montgomery JE, TBurket C, Liu CG, Vihtelic TS, Hyde DR. Time course analysis of gene expression during light-induced photoreceptor cell death and regeneration in albino zebrafish. *Dev Neurobiol* 2007;67:1009–31. [PubMed: 17565703]
- Kreiling JA, Creton R, Reinisch C. Early Embryonic Exposure to Polychlorinated Biphenyls Disrupts Heat-Shock Protein 70 Cognate Expression in Zebrafish. *J Toxicol Environ Health A* 2007;70:1005–1013. [PubMed: 17497411]
- Lee NH, Saeed AI. Microarray: an overview. *Methods Mol Biol* 2007;353:265–300. [PubMed: 17332646]
- Linney E, Dobbs-McAuliffe B, Sajadi H, Malek RL. Microarray gene expression profiling during the segmentation phase of zebrafish development. *Comp Biochem Physiol C Toxicol Pharmacol* 2004;138(3):351–62. [PubMed: 15533793]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ Method. *Methods* 2001;25:402–408. [PubMed: 11846609]
- Mathavan S, Lee SG, mark A, Miller LD, Murthy KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan Y, Korzh V, Gong Z, Liu ET, Lufkin T. Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet* 2005;1:260–76. [PubMed: 16132083]
- Robu ME, Larson JD, Nasevicius A, Beiraghi S, Brenner C, Farber SA, Ekker SC. p53 activation by knockdown technologies. *PLoS Genet* 2007;3(5):e78. [PubMed: 17530925]
- Sumanas S, Joriniak T, Lin S. Identification of novel vascular endothelial-specific genes by the microarray analysis of the zebrafish cloche mutants. *Blood* 2005;106:534–541. [PubMed: 15802528]
- Thisse B, Heyer V, Lux A, Alunni A, Degraeve A, Seiliez I, Kirchner J, Parkhill JP, Thisse C. Spatial and Temporal Expression of the Zebrafish Genome by Large-Scale In Situ Hybridization Screening. *Meth Cell Biol* 2004;77:505–519.

A**Cluster analysis**

BE201395 belongs to Node 8 , Genes Peaking at 48 Hours

BE201395 raw intensities

	epi75	18S	24h	48h	day3	day4	day5
Expression level	19315	23051	3.17E+06	4.25E+06	3.53E+06	4.26E+06	2.42E+06
Standard error	1012.3	553.513	683193	318499	517289	497919	785370

B**BE201395 normalized intensities**

	epi75	18S	24h	48h	day3	day4	day5
Expression level	0.334774	0.450676	54.0617	71.3083	64.0917	70.4725	33.5199
Standard error	0.0215527	0.0497228	13.7704	2.53072	7.87584	8.94892	7.59211

C

BE201395 intensities relative to Beta Actin

	epi75	18S	24h	48h	day3	day4	day5
Expression level	0.00528927	0.00731232	0.597909	0.915501	0.756891	0.847103	0.413979
Standard error	0.000587927	0.00141055	0.137456	0.0667166	0.12752	0.106227	0.125379

D

Correlated expression

(numbers in parentheses are Pearson correlation coefficients)

Transcripts with expression profiles that are positively highly correlated to BE201395

BM024802 (0.998542), BI878376 (0.993987), BG308798 (0.993314), BM183855 (0.992428), AF327410 (0.986471), BM036676 (0.985086), BI880905 (0.980730), BG306132 (0.980473), BG306411 (0.977155), AI964223 (0.976948), AW076610 (0.975524), BF717356 (0.973503), U67845 (0.972107), BM185112 (0.970985), BM026584 (0.970895), BI846505 (0.970683), BI430332 (0.969506), BM071221 (0.968545), BI879201 (0.968404), BE558108 (0.968228), BI845046 (0.968203), BM070916 (0.967818), BG883309 (0.966485), BI864557 (0.965928), AW282069 (0.965706), AI943132 (0.965236), BI671039 (0.965204), BI326217 (0.965135), AI793755 (0.964887), AW116132 (0.964698)

Transcripts with expression profiles that are negatively highly correlated to BE201395

BI886160 (-0.980352), BI673604 (-0.977951), AI877588 (-0.971639), AI722668 (-0.965029), BG799622 (-0.956245), AW343766 (-0.955705), BI839927 (-0.955622), AW171194 (-0.939923), BI867202 (-0.937391), AI959516 (-0.937057), AW203025 (-0.935983), BI709117 (-0.928819), BM071094 (-0.928133), AW116451 (-0.926740), AI558921 (-0.925108), BM101527 (-0.924611), BI878842 (-0.923293), BM036285 (-0.922846), BG303928 (-0.922528), AF317837 (-0.922495), BG303926 (-0.922360), BI884761 (-0.920829), BI705531 (-0.919357), AI396688 (-0.919284), BI888567 (-0.918480), AW826518 (-0.916023), BE201395 (-0.915856), AI477305 (-0.914237), BM103959 (-0.913748), AF371368 (-0.913611)

Figure 1.

Web Tool “output” for zebrafish microarray analyses. To demonstrate the functions of the web resource, the expression profile of zgc:103639 (BE201395) retrieved from the web (<http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>) and used as an example. A. BE201395 raw intensities display. B. BE201395 normalized intensities display. C. BE201395 intensities relative to β -actin. D. Transcripts with expression profiles that are highly positively or negatively correlated to BE201395.

Figure 2.

Gene Clusters Display. To identify genes expressed at similar times during development, K-means clustering was performed on the expression profiles of each gene in the array. The intensity values for each spot were normalized to the median intensity on the array. The mean of these values across the replicates for each gene at each time point was given to the program Cluster 3.0. The genes were clustered into 10 groups using 1000 runs and an uncentered Pearson correlation metric. The results were visualized using the program Maple Tree.

A
Cluster analysis

BM103972 belongs to Node 5 , Genes Expressed Late in Development

B**BM103972 normalized intensities**

	epi75	18S	24h	48h	day3	day4	day5
Expression level	0.703032	0.808975	0.670076	0.52291	0.651909	12.6415	46.5507
Standard error	0.0336759	0.0915218	0.044524	0.0353933	0.0371936	3.41206	10.9057

C**BM103972 intensities relative to Beta Actin**

	epi75	18S	24h	48h	day3	day4	day5
Expression level	0.0111484	0.0128885	0.00737263	0.00665974	0.00704257	0.143007	0.548154
Standard error	0.00129417	0.00196244	0.000846339	0.000549956	0.000400205	0.0292376	0.0947396

D**Correlated expression**

(numbers in parentheses are Pearson correlation coefficients)

Transcripts with expression profiles that are positively highly correlated to BM103972

AW019025 (0.998812), BE016623 (0.998534), BG727383 (0.998342), AI522688 (0.998332), AW018954 (0.998041), BM154820 (0.997767), AI497473 (0.997516), AF327373 (0.997198), BF156623 (0.996961), AW232238 (0.996554), BI982101 (0.996426), BM154195 (0.996289), BF937964 (0.995961), BM183654 (0.995937), BI707054 (0.995813), BI980613 (0.995812), AW203087 (0.995673), AW018623 (0.995420), BM184897 (0.995333), BI980438 (0.995320), AI584397 (0.995315), AI522773 (0.995204), AI721596 (0.994931), BG728611 (0.994466), AW018969 (0.994217), BM096461 (0.994164), AF189238 (0.994086), U57390 (0.994078), AW019487 (0.994021), BI983181 (0.993624)

Transcripts with expression profiles that are negatively highly correlated to BM103972

AI617819 (-0.973768), BI888372 (-0.955733), AW019608 (-0.953393), AI721960 (-0.947774), BI880263 (-0.945421), AW344170 (-0.943115), AI545062 (-0.942661), BG306109 (-0.939178), AI667208 (-0.937799), BM102339 (-0.937674), AW510252 (-0.933889), AA605677 (-0.932401), BE693150 (-0.929399), BE558061 (-0.929139), BG305622 (-0.929058), BI474802 (-0.928914), BM025885 (-0.928276), BI672630 (-0.927037), BI890446 (-0.925807), BI881975 (-0.921197), BM025903 (-0.918859), BI878723 (-0.918792), AW567307 (-0.912057), AW344134 (-0.911131), AF164440 (-0.910651), AW567328 (-0.910138), BM023731 (-0.904391), BM155568 (-0.901596), BM035598 (-0.896980), BF156732 (-0.896121)

Figure 3.

The expression profile of pancreatic enzyme gene, *amylase $\alpha 2a$* (BM103972) retrieved from the web (<http://serine.umdj.edu/~ouyangmi/cgi-bin/zebrafish/profile.htm>). A. BM103972 raw intensities display. B. BM103972 normalized intensities display. C. BM103972 intensities relative to β -actin. D. Transcripts with expression profiles that are highly positively or negatively correlated to BM103972.

Figure 4. Comparison of array results with quantitative RT-PCR. Using the expression level of the 75% epiboly stage as reference, (A) the quantitative RT-PCR of β -actin (AF025305) at designated developmental stages have similar pattern to those obtained from the web resource which showed slightly increased expression after 24 hpf. The quantitative RT-PCR intensities for the selected (B) Node 2 gene (BM104515), (C) Node 3 gene (AW421995), (D) Node 4 gene (BG738899), and (E) Node 5 gene (AF254642) are closely correlated with the microarray data.

Table 1

Comparison of Microarray Data with ZFIN Expression Data

Gene Name	Accession number	Developmental Stage							score
		epi75	18s	24hr	2d	3d	4d	5d	
zic2a (wu:fb26a03)	AF151535	1	1	1	n	1	n	n	1.00
zic1 (zgc:136736)	AF052435	1	1	1	1	n	n	1	1.00
ikzf1 (zmf1a1)	AF092175	1	1	1	1	n	n	n	1.00
zgc:114150 (im:6894309)	BE605395	n	0	1	1	n	n	n	0.67
barh2 (zgc:113948)	BI325578	1	1	1	1	n	n	1	1.00
zgc:113518	AA494741	n	n	0	0	n	n	n	0.00
zgc:113055	BI839803	n	1	1	1	n	n	n	1.00
zgc:112167 (wu:fa01a02)	AA495302	1	0	1	1	n	n	n	0.75
zgc:110692 (sb:cb637)	AI793426	1	1	1	1	n	n	n	1.00
zgc:110355 (wu:fe22a07)	BI885942	0	1	1	n	n	n	n	0.67
spata18 (zgc:110352)	BI841395	n	1	1	0	n	n	n	0.67
zgc:109976 (wu:fr86d05)	BE017604	n	0	1	1	n	n	n	0.67
zgc:103639	BE201395	1	1	1	1	n	n	1	1.00
zgc:101687	BE202096	1	1	1	1	n	n	n	1.00
zgc:101040	AI877806	n	1	0	0	n	n	n	0.33
zgc:100829 (wu:fr81b03)	BM035967	1	1	0	1	n	n	n	0.75
zgc:92530	AI497360	0	0	0	0	n	n	1	0.20
zgc:92470	BG306838	n	1	1	1	n	n	n	1.00
zgc:92313 (wu:fd15b06)	BI890296	0	0	1	1	n	n	n	0.50
zgc:92191	AI617763	n	0	0	0	n	n	n	0.00
bel7a (zgc:92023)	BG985470	n	1	1	1	n	n	n	1.00
zgc:91930	BG308877	n	1	1	1	n	n	n	1.00
zgc:86889	BI890361	n	1	1	1	n	n	n	1.00
zgc:86841	AI942924	n	1	1	0	n	n	n	0.67
zgc:86661	BG892130	n	n	1	1	n	n	0	0.67
zgc:85763 (wu:fc83b02)	BG985435	1	1	1	1	n	n	0	0.80
tpbp3 (zgc:77934)	BE016153	1	1	1	0	n	n	n	0.75
zgc:77462 (wu:fi58a11)	AW281850	1	1	1	1	n	n	n	1.00
zgc:77775	BG883995	1	1	1	1	n	n	n	1.00
zgc:77759	AW076700	n	1	1	1	n	n	n	1.00
zgc:77106 (wu:fi24h01)	BM103127	1	1	1	1	n	n	n	1.00
zgc:73134 (wu:fc89h04)	BI671166	n	1	1	1	n	n	n	1.00
zgc:73153	BG304888	1	0	1	1	n	n	0	0.60
zgc:56425 (wu:fi40h07)	BM184151	1	1	1	1	n	n	0	0.80
zgc:55345	BM005190	n	1	1	0	n	n	n	0.67
zgc:55879 (wu:fi36f02)	BI877696	1	1	1	1	n	n	n	1.00
zgc:56382 (wu:fb09a02)	BM101640	n	1	1	1	n	n	1	1.00
zgc:56709	BI843480	0	1	1	0	n	n	n	0.50
zgc:55262 (wu:fi67g01)	AF282675	n	1	1	1	n	n	n	1.00
zgc:63471 (wu:fi06a10)	AW116003	1	1	1	1	n	n	0	0.80
abat (wu:fi82a01)	AI882774	n	0	1	1	n	n	n	0.67
abce1 (zgc:111906)	AW133635	0	1	1	1	n	n	n	0.75
acat2 (wu:fb10a06)	AF143488	n	n	1	1	n	n	n	1.00
ache (zgc:92550)	AJ251640	1	1	1	1	n	n	n	1.00
acp5 (wu:fb19f01)	AI882824	1	1	0	1	n	n	n	0.75
acs14 (zgc:66186)	AI722213	0	1	1	1	n	n	n	0.75
acta1 (wu:fb83f06)	AF180887	0	1	1	1	n	n	n	0.75
acvr2b (wu:fi97d11)	AF069500	0	0	1	0	n	n	n	0.25
alas2 (cb1063)	AF095747	n	1	1	1	n	n	n	1.00

Gene Name	Accession number	Developmental Stage**							score
		epi75	18s	24hr	2d	3d	4d	5d	
aldh2l (zgc:85659)	AW344274	0	0	1	1	n	n	n	0.50
aldh3d1 (wu:fb17d04)	AF254955	0	1	1	1	n	n	n	0.75
aldh7a1 (wu:fb35e06)	B1879264	n	1	1	1	n	n	n	1.00
alox12 (zgc:64120)	A1558352	0	0	1	1	n	n	n	0.50
aldoaa (wu:fa28b10)	AW116838	0	1	1	1	n	n	n	1.00
amy2a (zgc:77877)	BM103972	1	1	1	1	n	n	1	1.00
angptl3 (wu:fb60e11)	AF379604	1	1	1	1	n	n	1	1.00
anp32b (wu:fa28c10)	B1673416	1	1	1	1	n	n	n	1.00
anxa1a (wu:fa01d11)	AW128428	1	1	0	1	n	n	n	0.75
anxa5 (wu:fa98f06)	Y16043	0	1	1	1	n	n	n	0.75
anxa6	B1891593	1	1	1	1	n	n	n	1.00
anxa11a (sb:cb401)	B1889313	1	1	1	0	n	n	n	0.75
anxa11b (eb755)	A1957567	1	1	0	n	n	n	n	0.67
anxa13 (wu:fb40a08)	AJ306452	1	1	1	1	n	n	n	1.00
apg3l (zgc:64156)	B1887098	n	n	1	1	n	n	n	1.00
apoa4 (zgc:122983)	A1477980	1	1	1	0	n	n	1	0.80
apoeb (wu:fe95h12)	AJ236882	1	1	1	1	n	n	1	1.00
aqp3 (zgc:55525)	AW422269	1	1	1	1	n	n	1	1.00
aren1l (wu:fa10d11)	BM082411	1	1	1	1	n	n	n	1.00
ard1a (zgc:63981)	BE693122	n	1	1	0	n	n	n	0.67
arhgef1 (zgc:158261)	BG883520	n	1	1	1	n	n	n	1.00
arl3l1 (zgc:73049)	BG305579	0	1	1	1	n	n	1	0.80
arl3l2 (zgc:73116)	BG306391	0	1	1	1	n	n	n	0.75
arl4a (zgc:77800)	BM071892	0	0	1	1	n	n	n	0.50
arl4l (zgc:77711)	BG308473	n	1	1	0	n	n	n	0.67
Average score***									0.80

**

A score of 0, 1 or "n" is assigned to indicate whether the data from each source correlate at that stage. A score of 1 is assigned to expression profiles that correlate, a score of 0 to those that differ, and no score (n) to those time points with no *in situ* hybridization data.

Indicates the average agreement between the two databases.

Table 2

Comparison of Microarray Data with Mathavan *et al.* (2005)

Gene Name	Accession number	epi75	Developmental Stage [^]			
			18s	24hr	2d	score
zic2a (wu:fb26a03)	AF151535	1	1	1	1	1.00
zic1 (zgc:136736)	AF052435	1	1	1	0	0.75
ikzf1 (zmf1a1)	AF092175	0	1	1	0	0.50
zgc:114150 (im:6894309)	BE605395	1	1	1	0	0.75
barhl2 (zgc:113948)	BI325578	0	0	0	1	0.25
zgc:113518	AA494741	1	1	0	0	0.25
zgc:113055	BI839803	1	1	0	1	0.75
zgc:112167 (wu:fa01a02)	AA495302	1	1	1	0	0.75
zgc:110692 (sb:cb637)	AI793426	1	1	1	1	1.00
zgc:110355 (wu:fe22a07)	BI885942	1	1	0	1	0.75
spata18 (zgc:110352)	BI841395	0	1	1	1	0.75
zgc:109976 (wu:fr86d05)	BE017604	1	1	1	0	0.75
zgc:103639	BE201395	0	1	1	0	0.50
zgc:101687	BE202096	0	1	1	1	0.75
zgc:101040	AI877806	1	1	0	1	0.75
zgc:100829 (wu:ff81b03)	BM035967	1	1	1	1	1.00
zgc:92530	AI497360	1	1	1	0	0.75
zgc:92470	BG306838	1	1	1	1	1.00
zgc:92313 (wu:fd15b06)	BI890296	0	1	1	1	0.75
zgc:92191	AI617763	1	1	1	0	0.75
bel7a (zgc:92023)	BG985470	1	1	1	1	1.00
zgc:91930	BG308877	1	1	1	0	0.75
zgc:86889	BI890361	1	1	1	1	1.00
zgc:86841	AI942924	0	1	1	1	0.75
zgc:86661	BG892130	1	1	1	1	1.00
zgc:85763 (wu:fc83b02)	BG985435	1	1	1	1	1.00
tpbp3 (zgc:77934)	BE016153	1	1	1	0	0.75
zgc:77462 (wu:ff58a11)	AW281850	0	0	0	1	0.25
zgc:77775	BG883995	1	1	0	1	0.75
zgc:77759	AW076700	1	1	1	1	1.00
zgc:77106 (wu:ff24h01)	BM103127	0	0	0	1	0.25
zgc:73134 (wu:fc89h04)	BI671166	1	1	1	0	0.75
zgc:73153	BG304888	1	1	0	0	0.50
zgc:56425 (wu:ff40h07)	BM184151	1	1	1	1	1.00
zgc:55345	BM005190	1	1	1	1	1.00
zgc:55879 (wu:ff36f02)	BI877696	1	1	1	0	0.75
zgc:56382 (wu:fb09a02)	BM101640	0	1	1	1	0.75
zgc:56709	BI843480	1	1	0	1	0.75
zgc:55262 (wu:ff67g01)	AF282675	1	1	1	1	1.00
zgc:63471 (wu:ff06a10)	AW116003	1	1	1	1	1.00
abat (wu:ff82a01)	AI882774	0	0	1	0	0.25
abce1 (zgc:111906)	AW133635	0	0	0	0	0.00
acat2 (wu:fb10a06)	AF143488	1	1	1	0	0.75
ache (zgc:92550)	AJ251640	1	1	1	0	0.75
acp5 (wu:fb19f01)	AI882824	1	1	1	1	1.00
acs14 (zgc:66186)	AI722213	0	1	1	0	0.50
acta1 (wu:fb83f06)	AF180887	1	1	1	0	0.75
acvr2b (wu:ff97d11)	AF069500	0	0	0	0	0.00
alas2 (cb1063)	AF095747	0	1	1	0	0.50

Gene Name	Accession number	Developmental Stage [^]				
		epi75	18s	24hr	2d	score
aldh2l (zgc:85659)	AW344274	1	1	1	0	0.75
aldh3d1 (wu:fb17d04)	AF254955	1	1	1	1	1.00
aldh7a1 (wu:fb35e06)	B1879264	1	1	1	1	1.00
alox12 (zgc:64120)	A1558352	1	1	1	1	1.00
aldoaa (wu:fa28b10)	AW116838	0	0	1	1	0.50
amy2a (zgc:77877)	BM103972	1	1	1	1	1.00
angptl3 (wu:fb60e11)	AF379604	1	1	0	0	0.50
anp32b (wu:fa28c10)	B1673416	1	1	1	1	1.00
anxa1a (wu:fa01d11)	AW128428	1	1	1	1	1.00
anxa5 (wu:fa98f06)	Y16043	1	1	1	1	1.00
anxa6	B1891593	0	0	0	1	0.25
anxa11a (sb:cb401)	B1889313	1	1	1	1	1.00
anxa11b (eb755)	A1957567	1	1	1	0	0.75
anxa13 (wu:fb40a08)	AJ306452	1	1	1	0	0.75
apg3l (zgc:64156)	B1887098	0	0	1	0	0.25
apoa4 (zgc:122983)	A1477980	1	1	1	1	1.00
apoeb (wu:fe95h12)	AJ236882	1	1	1	1	1.00
aqp3 (zgc:55525)	AW422269	1	1	0	1	0.75
aren1l (wu:fa10d11)	BM082411	0	0	1	0	0.25
ard1a (zgc:63981)	BE693122	1	1	1	1	1.00
arhgef1 (zgc:158261)	BG883520	1	1	1	1	1.00
arl31l (zgc:73049)	BG305579	1	1	1	1	1.00
arl3l2 (zgc:73116)	BG306391	0	1	1	1	0.75
arl4a (zgc:77800)	BM071892	1	1	1	0	0.75
arl4l (zgc:77711)	BG308473	1	1	1	1	1.00
Average score ^{^^}						0.75

[^] A score of 1 is assigned to expression profiles that correlate and a score of 0 to those that differ to those time points with Mathavan *et al.* (2005).

^{^^} Indicates the average agreement between the two databases.