Genomics Data 5 (2015) 206–209

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Neurotranscriptome profiles of multiple zebrafish strains

Ryan Y. Wong [a,*], John Godwin [b]

[a] Department of Biology, University of Nebraska at Omaha, Omaha, NE 68182, USA
[b] Department of Biological Sciences, W.M. Keck Center for Behavioral Biology, Box 7617, North Carolina State University, Raleigh, NC 27695-7617, USA

ARTICLE INFO

Article history:
Received 31 May 2015
Accepted 1 June 2015
Available online 6 June 2015

Keywords:
RNA-sequencing
Brain
Gene expression
Sex
Zebrafish

ABSTRACT

Behavioral displays or physiological responses are often influenced by intrinsic and extrinsic mechanisms in the context of the organism's evolutionary history. Understanding differences in transcriptome profiles can give insight into adaptive or pathological responses. We utilize high throughput sequencing (RNA-sequencing) to characterize the neurotranscriptome profiles in both males and females across four strains of zebrafish (*Danio rerio*). Strains varied by previously documented differences in stress and anxiety-like behavioral responses, and generations removed from wild-caught individuals. Here we describe detailed methodologies and quality controls in generating the raw RNA-sequencing reads that are publically available in NCBI's Gene Expression Omnibus database (GSE61108).

Specifications	
Organism/cell line/tissue	Zebrafish brain
Sex	Male and female
Sequencer or array type	Illumina Genome Analyzer IIx
Data format	Raw: FASTQ files
Experimental factors	Strain (AB, Scientific Hatcheries, LSB, HSB), sex (male, female)
Experimental features	RNA-sequencing analysis of male and female zebrafish brains in four different strains (AB, Scientific Hatcheries, LSB, HSB), 17 weeks post-fertilization
Consent	All procedures approved by the North Carolina State University Institutional Animal Care and Use Committee
Sample source location	Experiments were conducted at North Carolina State University, Raleigh, North Carolina, USA

1. Direct link to deposited data

The raw FASTQ files can be accessed through the Gene Expression Omnibus.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61108.

2. Experimental design, materials and methods

In this study we analyzed the whole-brain transcriptome profiles of male and female zebrafish (*Danio rerio*) in four different strains [1,2]. In brief, 17 week old zebrafish were quickly sacrificed and whole-brains were removed and processed for RNA-sequencing. Sequencing reads were subsequently aligned, analyzed, and quantified using open-source software. We also conducted technical and biological validation and replication of the RNA-sequencing results using quantitative reverse-transcriptase PCR (for overview of procedures, see Fig. 1).

2.1. Animal subjects

Zebrafish cohorts were generated and reared using previously described methods [3]. All fish were kept in mixed sex 100-liter tanks. Tanks were on a custom-built recirculating filtration system with water temperature kept at 28 °C and on a 12:12 light:dark cycle. Fish were fed twice daily with commercial feed (Tetramin). The AB and Scientific Hatcheries (SH) zebrafish strains originated from commercial suppliers (Zebrafish International Resource Center and Scientific Hatcheries, respectively). Although the AB and SH strains were bred in laboratory conditions for many generations at their respective stock centers, these strains were maintained in our laboratory for four and one generations, respectively. The two other strains (High Stationary Behavior (HSB); Low Stationary Behavior (LSB)) of zebrafish originated from approximately 200 wild caught individuals and were six generations removed from the wild (see [3] for additional selective breeding details).

2.2. Tissue collection

We collected whole brains from 160 individual zebrafish (n = 20 for each sex for each strain) that were 17 weeks post-fertilization. Between 09:00–12:00 we quickly removed fish from their home tanks, deeply anesthetized with tricaine methanesulfonate, followed by decapitation.

* Corresponding author.
E-mail address: rwong@unomaha.edu (R.Y. Wong).

http://dx.doi.org/10.1016/j.gdata.2015.06.004

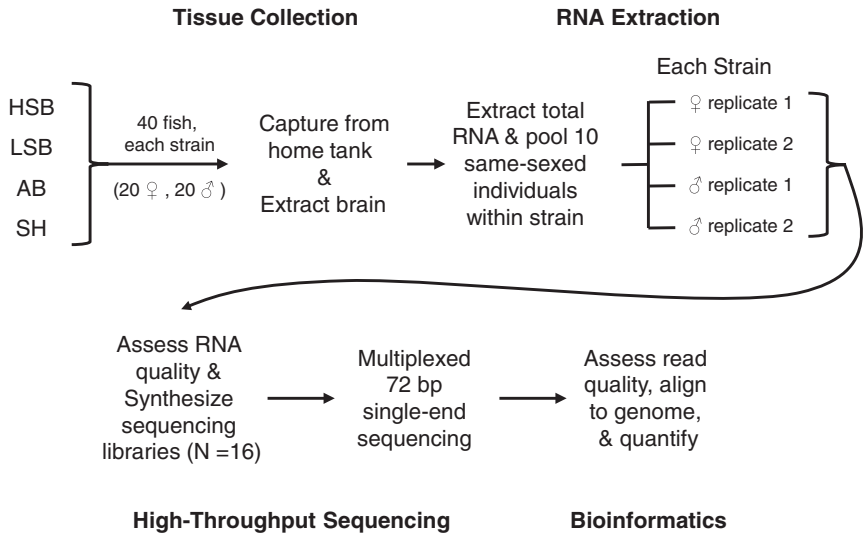

Fig. 1. Workflow for collecting and processing the neurotranscriptome in each zebrafish strain.

Whole brains were removed within 3 min of being caught and placed in RNAlater (Ambion). After storing the samples at 4 °C overnight, we removed all RNAlater and stored brains at −80 °C until RNA extraction. Sex was assigned by observation of testes or ovaries on dissection.

2.3. RNA isolation

We extracted total RNA using column purification (RNeasy Plus Mini Kit, Qiagen). Brains were homogenized for 3 min at maximum speed with 50–100 μl of zirconium oxide beads (Bullet Blender, Next Advance) in 0.6 ml of Buffer RLT (Qiagen) with 2-mercaptoethanol (Sigma). We then added 100 μl of chloroform, mixed, and incubated at room temperature for 5 min. We subsequently centrifuged the samples at 12,000 × *g* for 15 min at 4 °C. The supernatant was transferred to the RNeasy genomic DNA column (Qiagen) and then we proceeded according to the manufacturer's instructions. All samples were eluted with 30 μl of DEPC-treated water (Ambion).

2.4. RNA-sequencing library preparation and sequencing

For each strain we pooled 1 μg of total RNA from 10 same sex individuals into a biological replicate. This generated four biological replicates for each strain (two biological replicates for each sex). We analyzed the quantity and quality of the RNA for the 16 samples with a 2100 Bioanalyzer (Agilent). All samples were of high quality (RIN > 8.0, Table 1). Using 1 μg of total RNA from the pooled samples we generated cDNA libraries following the manufacturer's protocol (TruSeq RNA Sample Prep V2, Illumina). We ligated a unique Illumina Index adapter to each biological replicate to allow for multiplexing. After cDNA library synthesis we submitted samples to the Genomic Sciences Laboratory at North Carolina State University for 72 bp single-end RNA sequencing (Illumina GAIIx). We followed a balanced block design [4] and multiplexed all 16 samples and ran them across 16 lanes.

2.5. Data processing

With reads that passed default quality controls (Illumina), we combined across lanes for each biological replicate. Total read counts varied between 34–65 million reads (Table 2). We utilized the open source software GSNAP [5] to align the reads to the zebrafish genome. We first built GSNAP genomic and GSNAP known and novel splice site databases using the Zv9 (release 71) *D. rerio* genome and gene sets, respectively (Ensembl). For each biological replicate we successfully aligned over 99% of the reads (assessed by SAMtools [6]) to the zebrafish genome using the default GSNAP parameters (Table 2).

2.6. Validation and replication with quantitative reverse-transcriptase PCR

We performed both technical validation of RNA-sequencing libraries and independent biological replication (HSB and LSB strains) through

Table 1
RNA characteristics of biological replicates as measured by a 2100 Bioanalyzer (Agilent).

Sample name in GSE61108	Strain	Sex	RNA concentration (ng/μl)	RNA integrity number
AB female rep1	AB	Female	69.88	8.5
AB female rep2	AB	Female	70.58	8.5
AB male rep1	AB	Male	67.76	8.5
AB male rep2	AB	Male	73.16	8.6
SH female rep1	Scientific Hatcheries	Female	66.12	8.6
SH female rep2	Scientific Hatcheries	Female	90.58	8.5
SH male rep1	Scientific Hatcheries	Male	118.38	8.7
SH male rep2	Scientific Hatcheries	Male	105.76	8.7
LSB female rep1	Low Stationary Behavior	Female	88.8	8.4
LSB female rep2	Low Stationary Behavior	Female	69.88	8.4
LSB male rep1	Low Stationary Behavior	Male	54.14	8.7
LSB male rep2	Low Stationary Behavior	Male	57.52	8.4
HSB female rep1	High Stationary Behavior	Female	71.48	8.5
HSB female rep2	High Stationary Behavior	Female	89.62	8.7
HSB male rep1	High Stationary Behavior	Male	82.3	8.5
HSB male rep2	High Stationary Behavior	Male	75.62	8.5

Table 2
Sequenced library characteristics.

Sample name in GSE61108	Strain	Sex	Read count	Reads aligning to zebrafish genome (%)
AB female rep1	AB	Female	63,333,522	99.2863542
AB female rep2	AB	Female	48,539,389	99.194584
AB male rep1	AB	Male	52,400,106	99.2919213
AB male rep2	AB	Male	42,245,840	99.2484941
SH female rep1	Scientific Hatcheries	Female	65,493,707	99.257547
SH female rep2	Scientific Hatcheries	Female	60,919,457	99.2179313
SH male rep1	Scientific Hatcheries	Male	59,127,323	99.2323989
SH male rep2	Scientific Hatcheries	Male	44,528,195	99.2321562
LSB female rep1	Low Stationary Behavior	Female	43,983,674	99.1853227
LSB female rep2	Low Stationary Behavior	Female	49,943,584	99.2696399
LSB male rep1	Low Stationary Behavior	Male	57,927,979	99.2060158
LSB male rep2	Low Stationary Behavior	Male	53,242,350	99.2381685
HSB female rep1	High Stationary Behavior	Female	55,298,353	99.2311091
HSB female rep2	High Stationary Behavior	Female	34,150,835	99.2296909
HSB male rep1	High Stationary Behavior	Male	60,575,809	99.2129366
HSB male rep2	High Stationary Behavior	Male	44,987,343	99.2242418

quantitative reverse-transcriptase PCR (qPCR). We quantified the reads for each protein-coding gene by using the "union" mode in HTSeq [7] in all of our RNA-sequencing libraries. Read counts were then normalized to the library size in edgeR [8]. We selected eight genes (*msmo1*, *oxt*, *gabbr1a*, *comta*, *sell*, *prodha*, *hsd11b2*, *gapdh*) for technical validation and 14 genes (*msmo1*, *oxt*, *gabbr1a*, *comta*, *sell*, *prodha*, *hsd11b2*, *gapdh*, *cyp19a1b*, *dio2*, *pmchl*, *cfos*, *gabbr1b*, *igf1*) for independent biological replication (see [1,2] for detailed primer characteristics and qPCR reaction parameters).

After normalizing each gene's expression to *ef1a*, an endogenous reference gene [9], we confirmed a significant correlation between gene expression measured by RNA-sequencing and qPCR. Using the same material from cDNA libraries that were submitted for RNA-sequencing, we found a significant correlation between normalized read count (RNA-sequencing quantification) and cycle threshold (qPCR quantification) for the eight genes examined (technical validation; $n = 64$, Spearman's $\rho = -0.278$ $p = 0.026$; Fig. 2). Using independent samples ($n = 9$ for each sex in each of the LSB and HSB strains), we similarly observed a significant correlation between expression measurements from the two techniques (RNA-sequencing and qPCR) for 14 genes (independent biological replication; $n = 56$, Spearman's $\rho = -0.406$ $p = 0.002$). Of note, we also observed consistent patterns of differential gene expression between sexes and stress coping styles (see [1,2] for details).

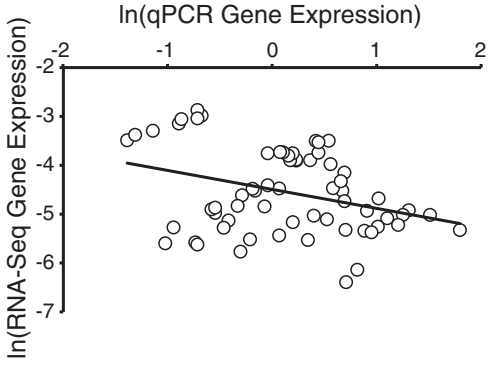

Fig. 2. Technical validation of RNA-sequencing results using qPCR. Each point represents a gene expression value for one of eight genes in each of the biological replicates in the HSB and LSB strains. Gene expression was normalized to an endogenous reference, *ef1a*, as measured in their respective quantification methods.

3. Conclusions

Zebrafish are a model system utilized in many developmental, toxicological, neuroscience, and biomedical studies [10–15]. Understanding and accounting for genomic and transcriptomic variation will provide important additional insights. Here we describe in detail the procedures and methodologies in sequencing the whole-brain transcriptome of both male and female adult zebrafish in four different strains. The high quality RNA-sequencing results, which have been both technically and biologically validated, are available through the NCBI's GEO database (GSE61108). This dataset should be of use to studies in a variety of contexts (e.g. evolution, neuroscience, genetics, bioinformatics, and biomedicine).

Acknowledgements

We thank Brad Ring and John Davis for the assistance with fish husbandry. We are grateful to Cory Dashiell, Melissa Lamm, Melissa McLeod, Katie Robertson, Christopher Gabriel, and Noffisat Oki for the helpful discussions and technical assistance. We are grateful to Barrie Robison for the generous donation of the Scientific Hatcheries line of zebrafish. This study received support from the National Institutes of Health (1R21MH080500) to JG. This study is a contribution of the W.M. Keck Center for Behavioral Biology at North Carolina State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] R.Y. Wong, M.M. McLeod, J. Godwin, Limited sex-biased neural gene expression patterns across strains in zebrafish (*Danio rerio*). BMC Genomics 15 (1) (2014) 905.
[2] R.Y. Wong, M.S. Lamm, J. Godwin, Characterizing the neurotranscriptomic states in alternative stress coping styles. BMC Genomics 16 (2015) 425.
[3] R.Y. Wong, F. Perrin, S.E. Oxendine, Z.D. Kezios, S. Sawyer, L.R. Zhou, S. Dereje, J. Godwin, Comparing behavioral responses across multiple assays of stress and anxiety in zebrafish (*Danio rerio*). Behaviour 149 (10–12) (2012) 1205–1240.
[4] P.L. Auer, R.W. Doerge, Statistical design and analysis of RNA sequencing data. Genetics 185 (2) (2010) 405–416.
[5] T.D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26 (7) (2010) 873–881.
[6] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, Genome Project Data Processing S: the Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16) (2009) 2078–2079.
[7] S. Anders, P.T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31 (2) (2015) 166–169.
[8] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1) (2010) 139–140.

[9] A.T. McCurley, G.V. Callard, Characterization of housekeeping genes in zebrafish: male–female differences and effects of tissue type, developmental stage and chemical treatment. BMC Mol. Biol. 9 (2008) 102.
[10] R.Y. Wong, S.E. Oxendine, J. Godwin, Behavioral and neurogenomic transcriptome changes in wild-derived zebrafish with fluoxetine treatment. BMC Genomics 14 (2013) 348.
[11] K. Howe, M.D. Clark, C.F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J.E. Collins, S. Humphray, K. McLaren, L. Matthews, et al., The zebrafish reference genome sequence and its relationship to the human genome. Nature 496 (7446) (2013) 498–503.
[12] R.T. Peterson, C.A. Macrae, Systematic approaches to toxicology in the zebrafish. Annu. Rev. Pharmacol. Toxicol. 52 (2012) 433–453.
[13] D.J. Grunwald, J.S. Eisen, Headwaters of the zebrafish — emergence of a new model vertebrate. Nat. Rev. Genet. 3 (9) (2002) 717–724.
[14] A.M. Stewart, O. Braubach, J. Spitsbergen, R. Gerlai, A.V. Kalueff, Zebrafish models for translational neuroscience research: from tank to bedside. Trends Neurosci. 37 (5) (2014) 264–278.
[15] R.F. Oliveira, Mind the fish: zebrafish as a model in cognitive social neuroscience. Frontiers in neural circuits 7 (2013) 131.