



**National Center for
Research Resources**

NATIONAL INSTITUTES OF HEALTH

Realizing the Scientific Potential of Transcriptomics in Aquatic Models

Portland, OR

September 19, 2010

Final Workshop Report

The views expressed in written conference materials or publication and by speakers and moderators at HHS-sponsored conferences do not necessarily reflect the official policies of the Department of Health and Human Services, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government

Table of Contents

A.	Executive Summary	3
B.	Goals of the Workshop	4
C.	Summary of Presentations and Discussion	5
	Session 1: Comparative Aquatic Transcriptomics & Human Disease: New Data Highlighting Comparative Transcriptomics.	5
	Session 2: Status of and Recent Developments of High-Throughput <i>de novo</i> Sequencing: Technologies-Projects Under Way or Recently Completed.	5
	Session 3: The Current Status of Transcriptomic Resources: Examples of Where Transcriptomics Would Be Impacting.	5
D.	Recommendations	6
E.	Conclusions	7
	Appendix A: Workshop Agenda	8
	Appendix B: Participant Roster	11
	Appendix C: NIH Officials	14

A. Executive Summary

On September 19, 2010, the National Center for Research Resources (NCRR) sponsored a workshop entitled “Realizing the Scientific Potential of Transcriptomics in Aquatic Models,” in Portland, Oregon. The workshop was held one day prior to the “5th Aquatic Animal Models for Human Disease Conference” held in Corvallis, Oregon, September 20 – 22 and thus enjoyed a rich list of attendees representing various aquatic model systems. This report summarizes the September 19 workshop and presents recommendations stemming from a round-table discussion.

Determining patterns of gene expression under a given set of parameters to better understand how organisms adapt to environmental perturbations and uncover alternative strategies employed to sustain life has been a goal of significant scientific research for many years. In order to assist an experimental approach to disease-related research, labor-intensive and expensive genomic resources were developed for several biological models used by the largest numbers of research laboratories (e.g., rodents, zebrafish, sea urchin, *Drosophila*, etc.). However, “next generation” high-throughput *de novo* sequencing technology (e.g., 454Ti, Illumina, etc.) has substantially reduced labor, time and cost of developing genomic resources shifting this paradigm.

Transcriptomics is the study of the complete set of expressed genes — both qualitatively and quantitatively — in cells under various conditions of health and disease. Transcriptomic analyses using experimental organisms that do not have existing genomic resources has become feasible only in the last year due to the enormous decrease in next-generation sequencing costs and marked improvements in software for sequence assembly. Now, even large and complex mixtures of transcripts from virtually any organism, either laboratory-raised or wild, may be completely sequenced within a few days and at modest cost. This application of transcriptomics (i.e., RNAseq) is benefitting scientists using many varied aquatic models by providing the high depth of sequence coverage required for *de novo* transcriptome reference libraries to be assembled and used for virtually any organism. However, understanding genetic expression mechanisms on a global scale — and interpreting a vast amount of sequence data — requires enormous computers and new software algorithms for transformation into biological understanding. This problem of post-sequence data processing and analysis prevents wide-spread application of these new and powerful methodologies. These issues are not limited to research using aquatic models, but due to the relatively small size of the aquatic models, community efforts to overcome limitations in data processing capabilities have not been addressed. Thus, strategic identification of solutions that would serve to significantly accelerate the adoption of transcriptomic methods and the quality of investigator-driven inquiry is both timely and reasonable.

The 26 participants (listed in Appendix B) were interested in the further development of transcriptomics. It was generally accepted by all attendees that promotion of transcriptomic applications within the enormous variety of natural models represented by the aquatic research community will result in rapid access to new knowledge. These models may serve to illuminate existing problems that may have been taken to the limits of experimental discovery by employing the handful of more widely popular research organisms.

This final report — stemming from workshop presentations and the subsequent round-table discussion — was developed by a committee composed of the session chairs. This report was

This final report — stemming from workshop presentations and the subsequent round-table discussion — was developed by a committee composed of the session chairs. This report was presented to the aquatic models community at the 5th Aquatic Animal Models for Human Disease Conference (September 20 – 22, 2010) and was accepted at large by the aquatic biomedical community.

B. Goals of the Workshop

The central purpose of this workshop was to assess the current nature and immediate future of transcriptomic expression profiling resources that may advance the utility of aquatic models to address relevant issues in human disease research. Consistent with this central purpose, the workshop was divided into three sessions of invited speakers with expertise and experience in the session topic. In the first session, entitled “Comparative Aquatic Transcriptomics & Human Disease: New Data Highlighting Comparative Transcriptomics,” six speakers highlighted novel findings of applying transcriptomic methods to their various aquatic models. The second session, entitled “Status of and Recent Developments of High-Throughput *de novo* Sequencing: Technologies-Projects Under Way or Recently Completed,” brought four well-regarded experts from the technical side of next-generation sequencing to provide their vision of where the technology is now, and where it will be in the near future. In the third session, “The Current Status of Transcriptomic Resources: Examples of Where Transcriptomics Would Be Impacting,” six speakers representing various aquatic models documented that removal of current hurdles to post-sequence data analyses would allow them to conquer impacting research problems heretofore unapproachable by standard methods. These presentations served as valuable examples of where improvements to current research support structure could remove limitations in the application of transcriptome experimental designs. This served as the segue to a round table discussion entitled “Recommendations of How to Realize Potential of Transcriptomics in Aquatic Models,” in which workshop attendees expressed their own thoughts on the centralized workshop issues and presented comments enlisted from investigators who could not attend.

The goals of this workshop were to delineate and discuss:

1. The benefit of massively parallel sequencing applications to scientists using aquatic models having a high potential for novel and impacting research discoveries;
2. The prospects of new technological developments that will eclipse current ultra-high throughput technologies and result in enhanced research capability while concurrently serving to further complicate post-sequence data analyses;
3. The hurdles that inhibit application of transcriptomic experimental design and analyses by investigators using aquatic models to investigate human disease related questions; and
4. Specific recommendations that would serve to overcome these hurdles and promote more rapid research progress.

C. Summary of Presentations and Discussion

Session 1: Comparative Aquatic Transcriptomics & Human Disease: New Data Highlighting Comparative Transcriptomics

In the first session, six established aquatic model scientists presented data highlighting the use of transcriptomics, its application to various aquatic models, and the novel findings with these applications. Novel comparative techniques discussed included Restriction-site Associated DNA (RAD), expression RAD (eRAD), RNA sequencing (RNAseq), and bioinformatic analyses and transcriptomics. Data demonstrating the usefulness of each novel technique listed above on various aquatic model systems was presented and further discussed during question & answer sessions following each presentation.

Session 2: Status of and Recent Developments of High-Throughput *de novo* Sequencing: Technologies-Projects Under Way or Recently Completed

In the second session, four technical experts presented on the use of current sequencing technologies and the development of next generation of sequencing technologies, while providing their vision of where the technology will be in the near future. These technologies included current sequencing instruments (i.e., 454, Illumina, and single-molecule sequencing) and data analysis methodology/software (i.e., graphical computations, IMAGE, etc). Again, examples of use in aquatic model systems were included, as well as examples of where these technologies would be beneficial.

Next generation sequencing technology has fundamentally changed our approach to gene expression experimentation. The *de novo* assembly of these reads remains a formidable challenge. Over the last year, the assembly process has fundamentally changed to accommodate the different read types and the sheer volume of reads seen with next generation platforms. Assembly quality remains a moving target often defined differently depending on the user community. Based on community feedback in this session the lower contiguity seen in next generation assemblies is acceptable as a starting point for many types of analyses while admittedly deficient for others.

Session 3: The Current Status of Transcriptomic Resources: Examples of Where Transcriptomics Would Be Impacting

In the third session, six speakers representing various aquatic models described the potential of the transcriptomic techniques to enable them to conquer impacting research problems. The first two presenters described preliminary transcriptome analysis tools used in aquatic models comparative studies. The remaining four speakers described aquatic model systems where little to no sequence data is available (i.e., killifish, *Aplysia*, damselfish, etc.) and the expected impacts of using transcriptome type analysis with these model systems.

A critical issue discussed here was the tremendous phylogenetic diversity of fishes and other aquatic models that have provided many unique model systems, each of which is selected based on the advantages provided by heritable, toxicological and/or infectious syndromes or other attributes that contribute unique ability to model human diseases. For example, explicit links to diseases such as heritable and induced melanoma development, craniofacial and other bone

developmental diseases, anemias, unique cancer-causing virus-like elements, appendage regeneration studies, rare resistance to heavy metal genotoxicity compared to humans, and the genetics of complex traits, were presented. However, it was generally realized that investigators will only unleash the full potential of these models if they are able to access state-of-the-art gene expression and genomic technologies. Constraining these technologies to only a few models with very large user bases will result in lost potential to take advantage of access to such a great diversity of model systems.

D. Recommendations

Throughout the workshop, participants engaged in discussions about the latest developments in transcriptomics and how their individual research programs would be impacted and their scientific findings more impacting with the addition of transcriptome applications. In the final workshop discussion, participants engaged in an open, round-table discussion moderated by the workshop organizer. The initial discussion centered on aquatic model systems not represented by workshop participants, but where the current attendees had knowledge of those using the models and felt cognizant enough to indicate how these models may also be affected by application of transcriptomic methods. Then, discussion centered on workshop models presented that would benefit from further development of transcriptome resources. A thorough discussion ensued concerning identification of transcriptome resources needed to enhance scientific understanding of individual aquatic models.

The major recommendations include:

- To establish single or multiple transcriptome service site(s) providing guidance for planning (i.e., best practices), analyzing (i.e., best software), and training (i.e., hosting workshops/training sessions) external scientists. The mission of these sites would be primarily dedicated to aquatic models research and other model systems that do not have large following or user bases. It is these smaller scientific models that may have the most impact in application of these transcriptome technologies.
- It was recognized that genome sequencing and assembly had dropped in cost to a level where many of the aquatic model genomes could be sequenced and made available to the scientific users both quickly and for a low level of support. This should be a priority for an entity or unit to see these genomes (12 – 14 models) sequenced, assembled, and annotated, over the next two to three years. The expected scientific impact of this activity is seen as great, because it supplies a reference genome to very many research investigators, enabling them to immediately apply transcriptome expression studies.
- To facilitate and encourage cross-training of students and post doctoral associates in computer science fields into programs and bioinformatic programs at various levels (i.e., B.S., M.S., etc.) and with interest in exploring aquatic species. The stimulation of expertise and a trained workforce in these areas, and in the many unique and varied aquatic models research foci is seen as having substantial long term impact on the scientific infrastructure.

E. Conclusions

The transcriptomics workshop blended a diverse group of aquatic biomedical model experts to assess the scientific potential of utilizing transcriptomics in aquatic models of human disease and to evaluate mechanisms to overcome the substantial hurdles for individual labs to capitalize on the new methods. Suggestions included the establishment of centers to facilitate the design and bioinformatic analysis of transcriptomic experiments by high-throughput sequencing, to encourage more whole genome sequencing for aquatic medical models, and to stimulate the entry of trained computer scientists into bioinformatic programs. Discussions and recommendations are intended to support the advancement of research on aquatic models that may impact human health.

Appendix A: Workshop Agenda

Realizing the Scientific Potential of Transcriptomics in Aquatic Models Portland, OR September 19, 2010

- 8:00 **Welcome Remarks**
Michael Chang, Health Scientist Administrator, NCRR
- 8:10 – 8:20 **Opening Remarks: *The Challenges/Bottlenecks and Liberating Power of Transcriptome Studies***
Ron Walter, Texas State University
- Session 1:**
- 8:20 – 10:30 ***Comparative Aquatic Transcriptomics & Human Disease: New Data Highlighting Comparative Transcriptomics***
Chair: John Postlethwait
- Speakers:**
- John Postlethwait*** (U. Oregon): *Comparative genomics for any fish: RAD-tag meiotic mapping coupled to RNA-seq*
- Julian Catchen*** (U. Oregon): *Bioinformatics interfaces of transcriptomes*
- Manfred Schartl*** (U. Wurzburg): *Transcriptomics of pigment cell tumors from transgenic medaka*
- Craig Albertson*** (Syracuse U.): *New genes from new species: African cichlids as models for craniofacial development and disease*
- Bill Cresko*** (U. Oregon): *Quantifying allele-specific expression*
- Ron Walter*** (Texas State U.): *Transcriptomics in Xiphophorus: recent developments and promise for novel discoveries*
- 10:30 – 11:00 **Coffee Break**

Session 2:

11:00 – Noon ***Status of and Recent Developments of High-Throughput de novo Sequencing: Technologies-Projects Under Way or Recently Completed***

Chair: Wes Warren

Speakers:

Wes Warren (Washington U., GSC): *Whole genome assemblies: disclosure for transcriptome analyses*

Jeff Boore (Genome Project Solutions, CA): *Evolution as the organizing principle for genomic data*

Devin Locke (Washington U., GSC): *Sequencing platforms & transcriptomics*

Randall Voss (U. Kentucky): *Using transcriptomics to enable a salamander model for regenerative biology*

Session 3 (Working Lunch):

Noon – 3:15 ***The Current Status of Transcriptomic Resources: Examples of Where Transcriptomics Would Be Impacting***

Chair: Mike Schmale

Speakers:

Seth Kullman (North Carolina State U.): *Comparative approaches to transcript profiling in medaka*

Yingjia Shen (Texas State U.): *XiphBrowser: A platform for visualization of RNAseq data*

Andrew Whitehead (Louisiana State U.): *Comparative transcriptomics reveals mechanisms of acclimation and adaptation in killifish*

John Wise (U. Southern Maine): *Comparative genotoxicity of hexavalent chromium in whale and human cells (Video recorded)*

Lynne Fieber (U. Miami): *Changes in gene expression with aging in Aplysia - a use for RNAseq?*

Mike Schmale (U. Miami): *Can we understand interactions between viral, mitochondrial and nuclear genomes leading to carcinogenesis in damselfish using RNAseq technologies?*

10:30 – 11:00 **Coffee Break**

Round-Table Discussion

3:30 – 5:00 ***Recommendations of How to Realize Potential of Transcriptomics in Aquatic Models***

Moderator: Ron Walter

Appendix B: Participant Roster

R. Craig Albertson, Ph.D.

Syracuse University
Department of Biology
256 Life Sciences Complex
Syracuse, NY 13244
Phone: (315) 443-8106
E-mail: rcalbert@syr.edu

Angel Amores, Ph.D.

University of Oregon
1425 E. 13th Avenue
Eugene OR 97403
Phone: (541) 346-4495
E-mail amores@uoregon.edu

Jeffrey L. Boore, Ph.D.

Genome Project Solutions
1024 Promenade Street
Hercules, CA 94547
Phone: (877) 867-0146
E-mail:
jlboore@GenomeProjectSolutions.com

Rachell Booth, Ph.D.

Dept. of Chemistry & Biochemistry
401B Centennial Hall
Texas State University
601 University Dr.
San Marcos, TX 78666-4616
Phone: (512) 245-2327
E-mail: rbooth@txstate.edu

Julian M. Catchen, Ph.D.

Institute of Neuroscience
Center for Ecology and Evolutionary Biology
University of Oregon
Eugene, OR 97403
Phone: (541) 346-4495
E-mail: jcatchen@uoregon.edu

William Cresko, Ph.D.

312 Pacific Hall
Center for Ecology and Evolutionary Biology
University of Oregon
Eugene, OR 97403
Phone: (541) 346-4779
E-mail: wcresko@uoregon.edu

Lynne Fieber, Ph.D.

Division of Marine Biology and Fisheries
Rosenstiel School of Marine and
Atmospheric Science
University of Miami
4600 Rickenbacker Cswy.
Miami, FL 33149
Phone: (305) 421-4140
E-mail: lfieber@rsmas.miami.edu

Tzintzuni Garcia, Ph.D.

Dept. of Chemistry & Biochemistry
419 Centennial Hall
Texas State University
San Marcos, TX 78666-4616
Phone: (512) 245-0357
E-mail: tzintzuni@gmail.com

David E. Hinton, Ph.D.

Nicholas School of the Environment
and Earth Sciences
Box 90328
A333B LSRC, Duke University
Durham, NC 27708-0328
Phone: (919) 613-8038
E-mail: dhinton@duke.edu

Seth W. Kullman, Ph.D.

North Carolina State University
Department of Environmental and
Molecular Toxicology
Box 7633
Raleigh, NC 27695-7633
Phone: (919) 515-4378
E-mail: swkullma@ncsu.edu

Devin Locke, Ph.D.

Genome Sequencing Center
Washington University School of Medicine
4444 Forest Park Blvd.
St Louis, MO 63108
Phone: (314) 286-1860
E-mail: dlocke@watson.wustl.edu

Kiyoshi Naruse, Ph.D.

National Institute for Basic Biology,
Laboratory of Bioresources
Nishigonaka 38, Myodaiji, Okazaki
444-8585
Aichi, Japan Phone:
0564-55-7581
E-mail: naruse@nibb.ac.jp

John H. Postlethwait, Ph.D.

Institute of Neuroscience
1254 University of Oregon
1425 E. 13th Avenue
Eugene, OR 97403
Phone: (541) 346-4538
E-mail: jpostle@uoneuro.uoregon.edu

Prof. Dr. Manfred Scharl

Physiologische Chemie I
Universität Würzburg
Biozentrum, Am Hubland
D-97074 Würzburg, Germany
Phone: 49-(0)931-888-4148
E-mail:
phch1@biozentrum.uni-wuerzburg.de

Michael C. Schmale, Ph.D.

Division of Marine Biology and Fisheries
Rosenstiel School of Marine and
Atmospheric Science
University of Miami
4600 Rickenbacker Cswy.
Miami, FL 33149
Phone: (305) 421-4140
E-mail: mschmale@rsmas.miami.edu

Yingjia Shen, Ph.D.

Dept. of Chemistry & Biochemistry
419 Centennial Hall
Texas State University
San Marcos, TX 78666-4616
Phone: (512) 245-0357
E-mail: ys14@txstate.edu

Stephen Randal Voss, Ph.D.

Department of Biology
University of Kentucky
Lexington, KY 40506
Phone: (859) 257-9888
E-mail: srvoss@uky.edu

Ronald B. Walter, Ph.D.

Dept. of Chemistry & Biochemistry
419 Centennial Hall
Texas State University
601 University Dr.
San Marcos, TX 78666-4616
Phone: (512) 245-0357
E-mail: RWalter@txstate.edu

Wes Warren, Ph.D.

Genome Sequencing Center
Washington University School of Medicine
4444 Forest Park Blvd.
St Louis, MO 63108
Phone: (314) 286-1899
E-mail: wwarren@watson.wustl.edu

Andrew Whitehead , Ph.D.

Department of Biological Sciences
202 Life Sciences Building
Louisiana State University
Baton Rouge, LA 70803
Phone: (225) 578-8210
E-mail: andreww@lsu.edu

John Pierce Wise, Sr., Ph.D.

Professor of Toxicology and Molecular
Epidemiology
Department of Applied Medical Sciences
University of Southern Maine
96 Falmouth St.
Portland, ME 04104-9300
Phone: (207) 228-8050
E-mail: John.Wise@usm.maine.edu

Appendix C: NIH Officials

Michael Chang, Ph.D.

Health Scientist Administrator
Division of Comparative Medicine
National Center for Research Resources
6701 Democracy Boulevard
Bethesda, MD 20892-4874
Phone: (301) 435-0750
E-Mail: changmic@mail.nih.gov

Miguel A. Contreras, Ph.D.

Health Scientist Administrator
Division of Comparative Medicine
National Center for Research Resources
6701 Democracy Boulevard
Bethesda, MD 20892-4874
Phone: (301) 594-9410
E-Mail: contre1@mail.nih.gov

Mahadev Murthy, Ph.D.

Digestive and Renal Program
National Institute on Aging
GWY – Gateway Building, 2C231
7201 Wisconsin Ave.
Bethesda, MD 20892-9205
Phone: (301) 402-7749
E-mail: mmurthy@nia.nih.gov

Lee Slice, Ph.D.

Scientific Review Officer
Office of Review
National Center for Research Resources
6701 Democracy Boulevard
Bethesda, MD 20892-4874
Phone: (301) 435-0807
E-Mail: slicelw@mail.nih.gov