

SHIRAZ and CABERNET: Leveraging Automation, Crowdsourcing, and Ontologies to Improve the Accuracy and Throughput of Zebrafish Histological Phenotype Annotations

Brian Canada¹, Georgia Thomas², John Schleicher³, James Z. Wang³, Keith C. Cheng²

¹University of South Carolina Beaufort, Bluffton, SC, USA

²Penn State College of Medicine, Hershey, PA, USA

³The Pennsylvania State University, University Park, PA, USA

Abstract. One of the goals of the Zebrafish Phenome Project is to systematically annotate the cellular-level morphological phenotypes associated with each gene in the zebrafish genome. Here, we offer demonstrations of two complementary software tools designed to help achieve that objective: SHIRAZ, a content-based image retrieval system designed for automated high-throughput annotation of histological phenotypes in the larval zebrafish, and CABERNET, a “crowdsourcing” application for histology image tagging that enables multiple domain experts to achieve consensus on ontology-compliant phenotype annotations. Potentially, such “consensus annotations” not only can be used to improve the accuracy of ground truth data for training SHIRAZ, but they can also be imported directly into PATO-compatible phenomic databases such as the Zebrafish Information Network.

Keywords: content-based image retrieval, crowdsourcing, phenotype annotation, histology, Zebrafish Phenome Project, Phenotype and Trait Ontology, PATO

The Zebrafish Phenome Project [1] aims to produce a comprehensive, ontology-compliant workup of the morphological, behavioral, and physiological phenotypes associated with mutation, environmental, and toxicological effects on each of the 20,000-25,000 genes in the zebrafish genome. Researchers attending the March 2010 Zebrafish Phenome Project community meeting concurred with our (the Cheng lab’s) opinion that annotated high-resolution imaging at the cellular level – such as by histology – was critical for morphological screening, particularly during larval development. While the technology for producing high-resolution histological imaging exists for zebrafish [2], the qualitative aspects of current histological assessments can result in intra- and inter-observer variability caused by differences in training, ability, timing, fatigue, and experience. Consequently, image datasets associated with high-throughput projects such as the Zebrafish Phenome Project will require analysis by *automated* approaches if reproducible cellular-level phenotypes are to be obtained for each gene in a reasonable timeframe.

Recently, we introduced a working prototype of a content-based image retrieval system, called SHIRAZ (System of Histological Image Retrieval and Annotation for Zoomorphology; online demo available at: <http://shiraz.ist.psu.edu>) [3], which has been designed to be capable of automatically annotating high-resolution images depicting histological abnormalities in the larval zebrafish, with a pilot application involving the developing eye at 5dpf. When a “query” eye image is uploaded to SHIRAZ, its automatically-extracted profile of texture features is compared to *feature signatures* associated with a set of ground truth “annotation concepts.” These concepts were derived from our Phenotype and Trait Ontology (PATO)-compliant knowledge base of manually-characterized semi-quantitative phenotype abnormality scores (completed for approximately 100 cloned mutants), ranging from 0 to 4 in order of increasing abnormality [Cheng lab, unpublished data]. SHIRAZ presents its results as a list of predicted annotation concepts, ranked by overall feature similarity to the query image.

While the current SHIRAZ prototype constitutes a viable proof-of-concept for an automated high-throughput histological analysis laboratory workflow, its accuracy is limited by the reliability of the ground truth annotations used in training its classification model. The present model is based on phenotype scores recorded and self-validated by a single observer with an appropriate level of domain knowledge. However, other experts, with varying levels of knowledge and experience, may interpret these same phenotypes differently. If multiple observers can achieve a consensus on the abnormality score of a given phenotype, then one can be more confident in the accuracy of the annotation, thereby improving ground truth data quality. One potential mechanism for achieving this consensus is by *crowdsourcing*, in which large-scale tasks are distributed to communities of human workers. For example, in the popular “ESP Game” for crowdsourced image labeling [4], two randomly paired “players” are presented with an image and, without communicating, must try to guess possible words to label that image. If both players successfully guess the same word, their “game score” increases, and the agreed-upon label is saved in a database for later use. The ESP Game therefore provides a mechanism for recording accurate, non-trivial image annotations in exchange for user entertainment.

To help improve the accuracy of the ground truth annotations used in SHIRAZ, we have proposed our own crowdsourcing application, called CABERNET (Crowdsourcing the Annotation of Bio-images for Education, Retrieval, and Network-Enabled Telephenotyping; current demo at: <http://shiraz.ist.psu.edu/cabernet>). CABERNET uses a controlled interface in which a registered user chooses a zebrafish histology “virtual slide” (powered by the Penn State Zebrafish Atlas [5]) and selects an abnormality score for each ontology-compliant

(e.g., PATO-based) phenotype to be characterized. Each user’s “game score” is dynamically updated based on the degree to which his or her chosen abnormality scores match those selected by other users. Abnormality scores with the highest consensus will be used as ground truth for later re-training of SHIRAZ to help improve its prediction accuracy. Meanwhile, these ontology-compliant “consensus annotations” will potentially form a rich phenotype dataset that, with minimal (if any) modification, can be directly imported into phenomic databases such as ZFIN [6]. Therefore, we expect that phenotypic annotations generated using both SHIRAZ and CABERNET can expedite the progress of the Zebrafish Phenome Project and help to close the genome-phenome knowledge gap.

References

1. Zebrafish Phenome Project 2010 Meeting, <http://www.blsmeetings.net/zebrafish/>
2. Mohideen, M.A., Beckwith, L.G., Tsao-Wu, G.S., Moore, J.A., Wong, A.C., Chinoy, M.R., Cheng, K.C.: Histology-based screen for zebrafish mutants with abnormal cell differentiation. *Dev. Dyn.* 228, 414–423 (2003)
3. Canada, B.A., Thomas, G.K., Cheng, K.C., Wang, J.Z.: SHIRAZ: an automated histology image annotation system for zebrafish phenomics. *Multimed. Tools Appl.* 51, 401–440 (2011)
4. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: *Proc. of SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, pp. 319–326, ACM, New York (2004)
5. Penn State Zebrafish Atlas, <http://www.zfatlas.psu.edu>
6. ZFIN: The Zebrafish Model Organism Database, <http://www.zfin.org>