Application Notes

ZFIQ: a software package for zebrafish biology

Tianming Liu¹,², Jingxin Nie¹,²,³, Gang Li¹,²,³, Lei Guo³, Stephen TC Wong¹,²,⁴
¹The Methodist Hospital Research Institute and Department of Radiology, The Methodist Hospital
²Weill Cornell Medical College; ³School of Automation, Northwestern Polytechnic University, China

Associate Editor: Dr. Chris Stoeckert

ABSTRACT

Summary: Rapid development, transparency and small size are the outstanding features of zebrafish that make it an increasingly important vertebrate system for developmental biology, functional genomics, disease modeling and drug discovery. Zebrafish has been regarded as ideal animal specie for studying the relationship between genotype and phenotype, for pathway analysis and systems biology. However, the tremendous amount of data generated from large numbers of embryos has led to the bottleneck of data analysis and modeling. The ZFIQ (zebrafish image quantitator) software provides streamlined data processing and analysis capability for developmental biology and disease modeling using zebrafish model.

Availability: ZFIQ is available for download at http://www.cbi-platform.net.

Contact: stwong@tmhs.org or tliu@tmhs.org

Supplementary information: Additional documentation for this software package is referred to http://www.cbi-platform.net/document.htm. Application examples of this software are referred to http://www.cbi-platform.net/download.htm

1 OVERVIEW

The overall architecture of the ZFIQ platform is illustrated in Figure 1a. The major four components of the ZFIQ platform, represented by the blue boxes, include: 1) Pipeline construction environment. The ZFIQ platform allows developers or users to dynamically construct data processing and analysis pipelines in an interactive graphical environment. The pipeline construction is driven by the data quantification problem raised by the zebrafish research community. The data types depend on specific applications. Current supported data types include image data and text data. 2) Pipeline pool. Once the components and parameters of a data quantification pipeline are tested and validated, the pipeline is archived in the pipeline pool, which is the aggregation of users’ knowledge and experience of data quantification. 3) Algorithm and method repository. Currently, the algorithms and method repository has already integrated most of the data analysis routines (over 400) in the Intel OpenCV library [1], and in-house built zebrafish data quantification methods. 4) Pipeline running environment. This environment provides users with graphic tools to perform large scale of data quantification studies, and to output the quantification results for specific statistical modeling and analysis. The ZFIQ platform runs on the Microsoft Windows system and accepts most of popular formats of input images.

(a)

Figure 1. (a) Overall architecture of the ZFIQ platform. (b). An example of zebrafish nuclei segmentation result produced by ZFIQ. The algorithm is referred to [11].

2 FUNCTIONALITIES

2.1 Pipeline Construction Environment

There are four windows in the pipeline construction environment, as shown in Figure 1b. Here, a constructed pipeline for cell nuclei segmentation is used as an example. Specifically, the left window in Figure 1b lists all of the integrated computational algorithms in the ZFIQ platform. The user can drag-and-drop any of these algorithms, or called filters, into the middle window of pipeline construction window, and link them together into a pipeline. In particular, the ZFIQ platform provides automatic checking mechanism to make sure that the links that connects algorithm components are legal in terms of parameter types. At the same time, the execution order of the pipeline is rendered in the execution sequence window on the top right. The execution sequence can be adjusted by user’s drag-and-drop operation in this window, if there are independent

*To whom correspondence should be addressed.
executions in parallel. Finally, the parameters in the pipeline can be interactively edited in the parameter editing window, as shown in the right bottom of Figure 1b. This functionality provides the capability to adaptively fine-tune some parameters of the pipeline for a specific dataset. In this sense, a pipeline not only means a sequence of data analysis algorithms, but a set of parameters associated with these algorithms.

2.2 Algorithm and Method Repository
The ZFIQ algorithm and method repository has integrated: 1) the data analysis routines (over 400 functions) in the Intel OpenCV library [1], and 2) zebrafish data quantification methods and tools developed by the CBI (Center for Biomedical Informatics) and its collaborators. The ZFIQ platform focuses on data quantification problems in the zebrafish biology community. In particular, we are interested in the following common zebrafish data quantification problems: 1) zebrafish cell quantification and pattern analysis, 2) zebrafish neuron and axon quantification, 3) zebrafish somite quantification, 4) zebrafish gene expression quantification, and 5) zebrafish morphological analysis. Currently, we have integrated over 30 algorithms and tools developed by CBI and its collaborators into the ZFIQ platform [2, 4].

2.3 Pipeline Pool
So far, the ZFIQ platform has over 50 pipelines for zebrafish cell detection, zebrafish cell quantification, and zebrafish neuron detection, as well as for general data manipulation and processing. Although these pipelines are already extensively tested and validated using datasets from our collaborating zebrafish research labs, other users might need to fine-tune the parameters according to their own datasets. The pipeline pool is under the folder /CBI_platform/pipelines after the installation of ZFIQ.

2.4 Pipeline Running Environment
The pipeline running environment provides users with different modes to run the selected pipeline, depending on the amount of dataset to be analyzed. For example, the running environment includes graphical tools that allow users to select large amounts of datasets for batch mode running using a specific pipeline. This support is particularly useful when large datasets generated from zebrafish experiments are to be processed and quantified. The analyzed results and relevant meta-data can be output to a structured database for statistical analysis and modeling of the relationship between genotypes and phenotypes. The pipeline running environment also provides users graphical tool to fine-tune the parameters specific for their zebrafish data quantification applications.

3 APPLICATION
The ZFIQ platform has been applied for zebrafish data quantification in study of Alzheimer’s disease genes using zebrafish models, zebrafish retina development, and zebrafish somitogenesis. We emphasize that this paper focuses on the demonstration of the software platform and that the specific algorithm details and biological questions represented by the datasets are addressed in other publications [2,3,4,10-12].

4 CONCLUSION
Developing a data analysis system that can deal with the diversity of zebrafish datasets and data quality is a challenge. The development of the ZFIQ software package aims to fill the void. By building ZFIQ on top of the public domain OpenCV platform, we welcome the computation community to contribute their data quantification pipelines to the ZFIQ, and, in turn, expect the zebrafish research community to contribute their quantification problems and associated datasets to the development. The growth of the pipeline pool of the ZFIQ package would depend on the fruitful interaction of the two communities. In addition, the development of ZFIQ would be complementary to available tools for other model systems such as C. elegans and fly [13, 14].

ACKNOWLEDGEMENTS
We would like to thank Dr. Weiming Xia and Dr. Jarema Malicki of Harvard Medical School and Dr. Scott Holley of Yale University for providing the zebrafish datasets discussed in this paper. The research was funded by the Bioinformatics Research Center Program Grant (PI: Dr. STC Wong) from Harvard Neurodiscovery Center (formally known as HCNR), Harvard Medical School and The Bioinformatics Program Grant (PI: Dr. STC Wong) from the Methodist Hospital Research Institute.

REFERENCES