Toxicology and Applied Pharmacology 314 (2017) 109–117

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/taap

A data-driven weighting scheme for multivariate phenotypic endpoints recapitulates zebrafish developmental cascades

Guozhu Zhang [a], Kyle R. Roell [a], Lisa Truong [c], Robert L. Tanguay [c], David M. Reif [a,b,*]

[a] Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA
[b] Department of Biological Sciences, Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, USA
[c] Department of Environmental and Molecular Toxicology, Sinnhuber Aquatic Research Laboratory, Oregon State University, Corvallis, OR, USA

ARTICLE INFO

Article history:
Received 23 July 2016
Revised 4 November 2016
Accepted 20 November 2016
Available online 22 November 2016

Keywords:
Zebrafish
High-dimensional
Bayesian
Developmental cascade
ToxRefDB
Risk assessment
Multiple endpoints
Multivariate
Scoring

ABSTRACT

Zebrafish have become a key alternative model for studying health effects of environmental stressors, partly due to their genetic similarity to humans, fast generation time, and the efficiency of generating high-dimensional systematic data. Studies aiming to characterize adverse health effects in zebrafish typically include several phenotypic measurements (endpoints). While there is a solid biomedical basis for capturing a comprehensive set of endpoints, making summary judgments regarding health effects requires thoughtful integration across endpoints. Here, we introduce a Bayesian method to quantify the informativeness of 17 distinct zebrafish endpoints as a data-driven weighting scheme for a multi-endpoint summary measure, called weighted Aggregate Entropy (wAggE). We implement wAggE using high-throughput screening (HTS) data from zebrafish exposed to five concentrations of all 1060 ToxCast chemicals. Our results show that our empirical weighting scheme provides better performance in terms of the Receiver Operating Characteristic (ROC) curve for identifying significant morphological effects and improves robustness over traditional curve-fitting approaches. From a biological perspective, our results suggest that developmental cascade effects triggered by chemical exposure can be recapitulated by analyzing the relationships among endpoints. Thus, wAggE offers a powerful approach for analysis of multivariate phenotypes that can reveal underlying etiological processes.

1. Introduction

There are tens of thousands of compounds currently in commerce and the environment worldwide, and while the number is growing rapidly, the toxicity information for humans or other species is still limited to a relatively small number of chemicals (Wambaugh et al., 2013). A major focus of developing alternative toxicity testing methods is to reduce the cost, complexity, labor, time, throughput, and animal welfare issues in traditional animal assays while retaining useful toxicological profiles (Basketter et al., 2012). High-throughput *in vitro* screening assays, such as ToxCast, were developed for chemicals in order to find targeted receptors and expedite toxicity testing (Judson et al., 2010). However, these assays do not provide systemic organismal responses for outcomes such as developmental toxicity. Thus, developing new cost-effective, high-throughput methods to evaluate the hazard information of these compounds is critical.

Zebrafish (*Danio rerio*), a small, vertebrate organism, has been widely used in toxicological research due to benefits such as *ex vivo* development and optical clarity of the embryo, suitability for high-throughput screening (HTS), cost-effectiveness, and rapid sexual maturation of only 3 months (Delvecchio et al., 2011; Truong et al., 2011). The genomic similarity between zebrafish and humans is approximately 70% (Howe et al., 2013), making it an ideal model to aid in understanding toxicity translatable to human health. Moreover, the developmental stages of zebrafish are characterized in fine detail (Kimmel et al., 1995). This allows studies of developmental progression perturbed by exposure to environmental stressors, where diverse behavioral and morphological endpoints can be assessed (Kokel et al., 2010; Noyes et al., 2015; Truong et al., 2014). Analysis across time points and

Abbreviations: AC50, half-maximal activity concentration; AOP, Adverse Outcome Pathway; AggE, Aggregate Entropy; EZ, embryonic zebrafish; HTS, high-throughput screening; Hpf, hours post fertilization; LD50, 50% lethal dose; MoA, mechanisms of action; POD, point of departure; ROC, Receiver Operating Characteristic; ToxRefDB, Toxicity Reference Database; wAggE, weighted Aggregate Entropy; TN, true negative; TP, true positive; FN, false negative; FP, false negative; MORT, mortality; YSE, yolk sac edema; AXIS, body axis; EYE, eye; SNOU, snout; JAW, jaw; OTIC, otic vesicle; PE, pericardial edema; BRAI, brain; SOMI, somite; PFIN, pectoral fin; CFIN, caudal fin; PIG, pigmentation; CIRC, circulation; TRUN, truncated body; SWIM, swim bladder; NC, notochord & bent tail; TR, touch response.

* Corresponding author at: Bioinformatics Research Center, Center for Human Health and the Environment, Department of Biological Sciences, North Carolina State University, Ricks Hall 356, Box 7566, 1 Lampe Drive, Raleigh, NC 27695, USA.
E-mail addresses: gzhang6@ncsu.edu (G. Zhang), krroell@ncsu.edu (K.R. Roell), lisa.truong@oregonstate.edu (L. Truong), robert.tanguay@oregonstate.edu (R.L. Tanguay), dmreif@ncsu.edu (D.M. Reif).

http://dx.doi.org/10.1016/j.taap.2016.11.010
0041-008X/

endpoint types can develop or refine Adverse Outcome Pathways (AOPs), inform risk assessment, and build predictive models for systems toxicology (Reif et al., 2015).

Traditional methods of identifying an inflection point along the curve to determine the effective concentration, such as LD50 (50% lethal dose) or AC50 (half-maximal activity concentration), are concentration-dependent and require major assumptions that are highly sensitive to common sources of noise (Beam and Motsinger-Reif, 2014). For example, the response data are typically expected to be monotonic, which is easier to achieve using *in vitro* cell line models, since the phenotypes are singular measurements of fold-change, percent inhibition, or cell death. Bayesian approaches have been applied to fit curves for scenarios where information can be borrowed across large chemical or assay sets (Wilson et al., 2014); however, curve-fitting may not be appropriate for developmental toxicity *in vivo*, largely because it is difficult to assure homogeneity across doses. This happens for several reasons: 1) manifestation of competing AOPs by different concentrations of chemical; 2) censoring by mortality; and 3) developmental cascade effects. Disentangling these factors is analytically challenging, as evidenced by the high mutual information shared across endpoints (Zhang et al., 2016). Moreover, the majority of the chemicals remain inactive or in constant response, presenting another challenge in identifying concentration-dependence of potential hazards (Truong et al., 2014; Zhang et al., 2016). In order to address these challenges, Aggregate Entropy (AggE) was designed as a concentration-independent method to interpret the overall effect as a point of departure (POD) without differentially weighting specific endpoints (Zhang et al., 2016).

Although several approaches have been used to aggregate information from multiple endpoints into a summary score, there is no consensus on how endpoints should be weighted (Shaw et al., 2016). Most published weighting schemes are heuristics based upon theoretical biological impact and are heavily weighted toward catastrophic endpoints such as lethality or inability to hatch (Harper et al., 2015; Liu et al., 2013; Padilla et al., 2012). In contrast, we take the opposite approach by deriving weights from observed data, then using empirical wAggE weights to explore biological underpinnings. First, we utilize a Bayesian method to quantify the severity of 17 distinct zebrafish endpoints (YSE, AXIS, EYE, SNOU, JAW, OTIC, PE, BRAI, SOMI, PFIN, CFIN, PIG, CIRC, TRUN, SWIM, NC, and TR). Second, we show that wAggE provides superior performance in terms of the ROC curve in identifying significant morphological effects. Third, we explore whether this weighting scheme reveals developmental cascade effects wherein early phenotypes can predict those occurring at later developmental stages. Fourth, we compare developmental scoring in zebrafish and mammalian results from the U.S. EPA's Toxicity Reference Database (ToxRefDB). Finally, we compare wAggE to a logistic-based curve-fitting method.

2. Materials and methods

2.1. Materials and analysis pipeline

The experimental data are described in Truong et al., 2014. Fig. 1 shows a consensus timeline that includes experimental conditions, key early developmental stages and landmarks (Kimmel et al., 1995), and morphological assessments. The data structure and details about AggE are provided in Zhang et al., 2016. ToxRefDB data were downloaded from https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data (toxrefdb_v1, October 2014). All analysis was implemented using custom R code (R core team, 2016).

2.2. Weighted Aggregate Entropy

In AggE, each biological state (18 assessed endpoints and No Observed Adverse Outcome) of an embryo is scored independently before summarizing across biological states and screened embryos. Briefly, let $X1 \ldots X18$ represent 18 assessed endpoints of an embryo, with 1 indicating present and 0 indicating absent. $X19$ represents No Observed Adverse Outcome (NOAE) with a value of $X19 = 19 - (X1 + \ldots + X18)$. The score, which is Shannon's entropy in nats unit, of this embryo is equal to $E = \frac{X1}{19} * \log(\frac{X1}{19}) + \ldots + \frac{X19}{19} * \log(\frac{X19}{19})$. Thus, by assigning weight to each biological state, for each chemical at a given concentration, the wAggE can be written as:

$$wAggE = \sum_{All\ screened\ embryos} w1 * \frac{X1}{19} * log\left(\frac{X1}{19}\right) + \ldots + w19 * \frac{X19}{19} * log\left(\frac{X19}{19}\right)$$

where $w1 \ldots w19$ are the weighting factors for each biological state.

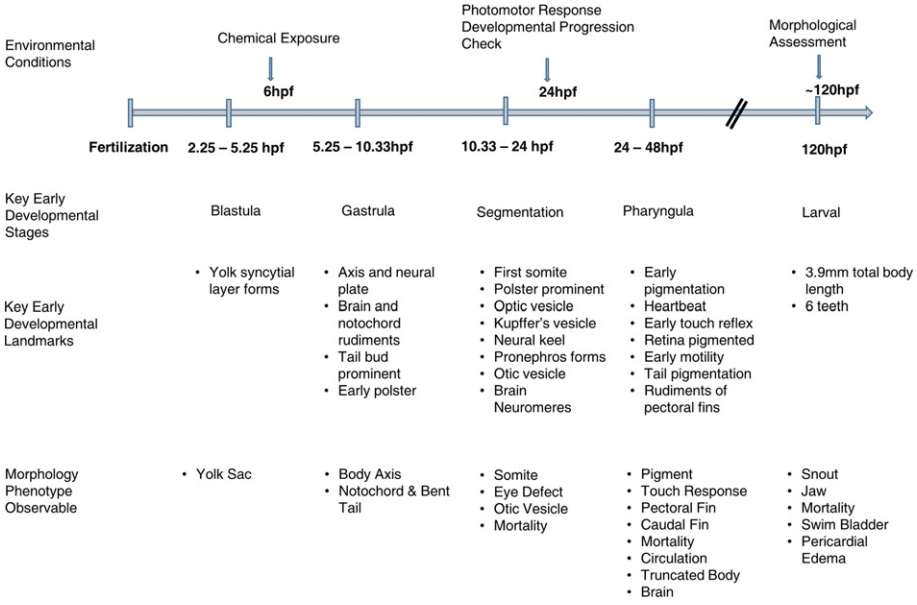

Fig. 1. Zebrafish developing timeline from fertilization to 120 hpf. Zebrafish key early developmental stages and associated landmarks. Environmental conditions prior to phenotypic assessments (18 distinct endpoints) are indicated on the top. Timeline of observable phenotypes are listed on the bottom. Only phenotypes that match our data are included.