**Springer Protocols**

Hao Zhu
Menghang Xia  *Editors*

# High-Throughput Screening Assays in Toxicology

Humana Press

# METHODS IN MOLECULAR BIOLOGY

# High-Throughput Screening Assays in Toxicology

Editors

## Hao Zhu

*Department of Chemistry, Rutgers Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

## Menghang Xia

*National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

Humana Press

*Editors*
Hao Zhu
Department of Chemistry
Rutgers Center for Computational
   and Integrative Biology
Rutgers University
Camden, NJ, USA

Menghang Xia
National Center for Advancing
   Translational Sciences
National Institutes of Health
Bethesda, MD, USA

Printed on acid-free paper

# Preface

The use of traditional animal models for chemical toxicity testing is often costly, time-consuming, low throughput, and sometimes inconsistently predictive of human toxicity. Cell-based High-Throughput Screening (HTS) assays used to measure the toxicity of environmental compounds have been widely applied as an alternative to traditional animal tests of chemical toxicity. Current HTS assays provide the environmental and toxicological scientific community with rich toxicology data that has the potential to be integrated into chemical toxicity research. While classic animal models have been standardized in most regulatory agencies, industries, and institutes, HTS protocols are still being developed, optimized, and updated in modern toxicology research. The US ToxCast and Tox21 initiatives from several government agencies including the National Toxicology Program (NTP) based at the National Institute of Environmental Health Sciences (NIEHS), the Environmental Protection Agency (EPA), the National Center for Advancing Translational Sciences (NCATS), and the Food and Drug Administration (FDA) represent a paradigm shift in compound toxicity testing from traditional animal studies to in vitro cell-based HTS assays that can be used to evaluate large amounts of environmental chemicals. This book will focus on introducing recently developed HTS assay protocols, many involved in the ToxCast and/or Tox21 initiatives, and the relevant HTS data analysis techniques. This is divided into the following three major parts:

1. In vitro HTS assays
2. In vivo HTS assays
3. Computational techniques to analyze HTS data

The book's authors have many years of experience in applying HTS assays to chemical toxicity evaluations and aim to present their HTS techniques to scientists at every level of pursuing chemical toxicology research. We hope that this book will serve as a valuable reference resource for translating new HTS techniques into standardized chemical toxicity assessment tools and will advance modern toxicology research to a new era where HTS techniques can partially replace the prevailing animal models.

*Camden, NJ, USA*                                                                                      *Hao Zhu*
*Bethesda, MD, USA*                                                                         *Menghang Xia*

# Acknowledgements

We sincerely thank Dr. John Walker for his help in the whole process of writing all chapters as well as the Springer team who provided the necessary assistance to develop this book.

This book would not have been completed without the willingness of all the authors who shared their scientific expertise and contributed their valuable knowledge, time, and effort in writing these chapters. We sincerely thank them for allowing us to edit, organize, and use their contributions to make this book available to our community.

H.Z.: I would like to thank my father Dr. Keqin Zhu who guides me in my current career path, and Ning and Stella who always encourage me to keep challenging myself in my career life and to pursue science topics of interest. Without them, nothing is possible.

M.X. dedicates this to Wei, Rena, Roman, and Lily for their love and support.

*Camden, NJ, USA*                                                                                  *Hao Zhu*
*Bethesda, MD, USA*                                                                        *Menghang Xia*

# Contents

# Contributors

PATRICK ALLARD • *The Allard Laboratory, University of California, Los Angeles, Los Angeles, CA, USA*

MATIAS S. ATTENE-RAMOS • *National Center for Advancing Translational Sciences, NIH, Bethesda, MD, USA*

JESSICA CAMACHO • *Society & Genetics, Environmental Health Sciences, and Molecular Toxicology Inter-Departmental Program, University of California Los Angeles, Los Angeles, CA, USA*

CARLOS CORIANO • *Department of Oncology, University of Wisconsin-Madison, Madison, WI, USA*

JUI-HUA HSIEH • *Kelly Government Solutions, Supporting NTP, NC, USA*

CHIA-WEN (AMY) HSU • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

RUILI HUANG • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

THAI KHUC • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

MARLENE T. KIM • *Department of Chemistry, Rutgers Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

YITONG LIU • *Division of Applied Toxicology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Laurel, MD, USA*

ZACHARY LUNDBY • *Society & Genetics, Environmental Health Sciences, and Molecular Toxicology Inter-Departmental Program, University of California, Los Angeles, Los Angeles, CA, USA*

CAITLIN LYNCH • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

KANA NISHIHARA • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

EMILY POWELL • *Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*

DANIEL P. RUSSO • *Department of Chemistry, Rutgers Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

SRILATHA SAKAMURU • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

ALEXANDER SEDYKH • *Multicase Inc., Beachwood, OH, USA*

SAMPADA A. SHAHANE • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

SUNITA J. SHUKLA • *Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA*

MICHAEL T. SIMONICH • *Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA; The Sinnhuber Aquatic Research Laboratory and the Environmental Health Sciences Center, Oregon State University, Corvallis, OR, USA*

ROBERT L. TANGUAY • *Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA; The Sinnhuber Aquatic Research Laboratory and the Environmental Health Sciences Center, Oregon State University, Corvallis, OR, USA*

LISA TRUONG • *Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA; The Sinnhuber Aquatic Research Laboratory and the Environmental Health Sciences Center, Oregon State University, Corvallis, OR, USA*

HONGBING WANG • *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD, USA*

WENYI WANG • *Department of Chemistry, Rutgers Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

YUHONG WANG • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

MENGHANG XIA • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

WEI XU • *Department of Oncology, University of Wisconsin-Madison, Madison, WI, USA*

JINGHUA ZHAO • *National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA*

HAO ZHU • *Department of Chemistry, Rutgers Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

# Part I

In Vitro Toxicological High-Thoughput Screening Methods

# Chapter 1

# Monitoring Ligand-Activated Protein–Protein Interactions Using Bioluminescent Resonance Energy Transfer (BRET) Assay

**Carlos Coriano, Emily Powell, and Wei Xu**

## Abstract

The bioluminescent resonance energy transfer (BRET) assay has been extensively used in cell-based and in vivo imaging systems for detecting protein–protein interactions in the native environment of living cells. These protein–protein interactions are essential for the functional response of many signaling pathways to environmental chemicals. BRET has been used as a toxicological tool for identifying chemicals that either induce or inhibit these protein–protein interactions. This chapter focuses on describing the toxicological applications of BRET and its optimization as a high-throughput detection system in live cells. Here we review the construction of BRET fusion proteins, describe the BRET methodology, and outline strategies to overcome obstacles that may arise. Furthermore, we describe the advantage of BRET over other resonance energy transfer methods for monitoring protein–protein interactions.

**Key words** Bioluminescent Resonance Energy Transfer (BRET), Protein–protein interactions, Screening assay, Imaging assay

## 1 Introduction

Proteins facilitate almost every biological process in cells. While many proteins perform their functions independently, the vast majority of cellular processes involve the interaction of proteins with each other [1]. Characterizing these protein–protein interactions is vital to understanding their biological importance. It is well known that toxicants can mimic endogenous ligands that can lead to the activation or disruption of protein–protein interactions. An important challenge when studying protein–protein interactions is in the context of living cells, where the proteins function in their endogenous spatial, organizational, and biological networks. *B*ioluminescent *R*esonance *E*nergy *T*ransfer (BRET) is a method that can measure the dynamics of protein–protein interactions in the intact cell in real time [2]. Thus, BRET is an ideal tool for studying protein–protein interactions [3].

BRET is observed naturally in the marine organisms *Renilla reniformis* and *Aequorea victoria*. This phenomenon has been adapted in the laboratory setting to interrogate protein–protein interactions. BRET is non-radiative energy transfer between a bioluminescent donor molecule and a fluorescent acceptor molecule. Proteins of interest are fused to the donor or the acceptor fluorophore, respectively. BRET is defined by the Förster Resonance Energy Transfer (RET) phenomenon, where bioluminescent energy resulting from an oxidative luciferase enzyme is transferred when in proximity (1–100 Å) of an acceptor, resulting in excitation and increased emission of the fluorescent acceptor. Resonance energy transfer occurs when the two proteins of interest interact at a distance of less than 10 nm, resulting in a decrease in luciferase emission and a concomitant increase in fluorescence. BRET monitors changes in the ratio of the acceptor and donor light emission. This dependency on the distance of donor and acceptor molecules makes it a reliable technique to study protein–protein interactions.

Since the development of BRET to study protein interactions [2], it has been successfully adopted to monitor many different types of protein–protein interactions in bacterial [2], plant [4, 5], and mammalian cells [6] and even in whole animal imaging [7]. It has commonly been employed as a screening tool to identify chemical compounds that inhibit or induce protein–protein interactions [8, 9]. The BRET assay has been used to monitor the bioactivity of compounds towards receptors such as the induction of receptor homodimerization and heterodimerization [10], GPCR ligand binding [11], the activation of tyrosine kinase receptors [12], and receptor oligomerization [13]. It can also be used as a biosensor assay to monitor cellular signaling, such as secondary messenger signaling and activation of intracellular kinases [14]. Highly optimized high-throughput BRET assays have also been used for screening xenoestrogens in non-purified extracts [15].

Cell-based assays using BRET technology that detect toxicants capable of disrupting normal protein–protein interactions are valuable because they are targeted tests that can often be performed in a high-throughput manner with lower costs than in vivo animal studies. The ability to translate BRET from the benchtop to an automated high-throughput assay has gained popularity because of its several advantages compared to other assays [15–17]. Because BRET utilizes a bioluminescent chemical donor substrate, it does not require an external excitation light source and thus eliminates issues of autofluorescence and photobleaching associated with other RET techniques that use fluorescent donors including Fluorescent Resonance Energy Transfer (FRET). Instead of an external excitation light source, BRET occurs after enzymatic oxidation of a chemical substrate, and can be employed as a microplate assay with the appropriate filter setup for bioluminescence and fluorescence light emission readout [3, 18]. A major advantage of BRET over other experiments that measure protein–protein interactions is that BRET can monitor proteins in intact live cells.

Over time many combinations of BRET fusion proteins have been developed, and each presents its advantages and drawbacks. The evolution of the BRET technology and details on the advantages of the different BRET methodologies have been extensively reviewed previously [8, 18, 19]. When choosing which donor and acceptor proteins to use, it is important to consider that the BRET assay depends on the degree of spectral overlap between donor and acceptor proteins. The peak emission of the donor luciferase is dictated by inherent properties of the luciferase in combination with the substrate selected for the assay. The excitation spectrum of the acceptor fluorophore must therefore overlap with the emission spectrum of the donor, but the emission of the acceptor must be spectrally distinct from emission of the donor. The two most used BRET variations are the originally described BRET methodology (now referred to as BRET1) and the second generation BRET2. BRET1 utilizes *Renilla* luciferase (Rluc) as its donor and Yellow Fluorescent Protein (YFP) or one of its variants [8] as the acceptor and coelenterazine h as the Rluc substrate. Coelenterazine h causes Rluc to emit at a peak wavelength of 470 nm, which is ideal to excite YFP at its peak excitation of 515 nm. YFP then emits at a peak wavelength of 530 nm. There is a substantial overlap in their emission spectra of Rluc and YFP, which results in a high background signal. To account for this background a correction factor is used to calculate the BRET signal [10], which requires the additional transfection condition of the Rluc fusion in the absence of the YFP fusion. BRET2 replaces the YFP acceptor molecule with a green fluorescent protein (GFP2 or GFP10) and utilizes the substrate DeepBlueC which results in a shift of the RLuc emission to a peak wavelength of 395 nm. This allows for an enhanced spectra separation between the donor and acceptor [20].

However, the major advantage of BRET1 over BRET2 for high-throughput screening is that Coelenterazine h is stable with half-life of 25 min, which improves assay stability, decreases sample-to-sample variation, and omits the use of injectors. In contrast, the rapid signal decay of DeepBlueC requires the use of automated injectors for the delivery of substrate and results in increased assay variability. In addition, DeepBlueC has a low quantum yield (more than 100X lower). Thus, higher protein expression is required for BRET2 than BRET1 to achieve sufficiently high luminescence levels for efficient energy transfer.

Other BRET fusion partners have recently been developed. For example the enhanced BRET2 assay (eBRET2) uses Rluc8, a mutant Rluc with enhanced luminescence intensity [21]. BRET3 combines Rluc8 with the more far-shifted fluorescent protein, mOrange; this red-shifted BRET makes it suitable for live tissue imaging [7]. A recently developed NanoBRET assay utilizes the Nanoluc luciferase (Nluc), a smaller (19 kDa) engineered energy donor along with a red-emitting fluorophore [22]. Extended

BRET (eBRET) uses the Rluc substrate EnduRen™ (Promega Inc.) which is slowly metabolized in live cells into coelenterazine h and allows for monitoring protein–protein interactions over time [23]. An alternative Promega substrate, ViviRen™, generates about a three times brighter output than coelenterazine but it is short lived. Another improvement of BRET has been the development of the donor substrate to improve quantum output. The coelenterazine analog coelenterazine-$v$ can be used in combination with any of the Rluc variants and it is able to slightly shift the emission spectra (~35 nm) of the Rluc and the Rluc mutants allowing for better red shifted output for in vivo BRET. However, this substrate is not available commercially because of the difficulty in purification [24]. A series of red shifted coelenterazine analogs were engineered by adding a C-8-bonded S heteroatom to low light emitting coelenterazines [25]. Levi et al., synthesized a group of "protected" DeepBlueC, which were tested in BRET2 and showed higher light output with an improved substrate half-life [26]. Deciding which BRET system to use is based on the laboratory and the protein of interest being studied.

BRET assays have been applied to study Estrogen Receptor (ER) dimerization [10, 27, 28]. ERs regulate the physiological effects of estrogens. There are two main estrogen receptors, ERα and ERβ, and they follow the classical nuclear receptor transcription factor mechanism of ligand-induced chaperone protein dissociation, receptor dimerization, and initiation of gene transcription, thus regulating several processes related to cell proliferation and cell growth. Ligand binding can lead to homodimerization (ERβ/β or ERα/α) or heterodimerization (ERα/β). BRET assays have been used to study whether antiestrogens were able to induce ERα homodimerization [10, 29]. Given the critical roles that ERs play in regulating cell growth in response to estrogens, there has been significant effort put forth to understand and predict the impact of xenoestrogens on ER signaling. The Xu Lab at the University of Wisconsin, Madison developed a BRET1 assay that utilizes *Renilla* Luciferase (RLuc) as the donor moiety and yellow florescent protein (YFP) as an acceptor. This assay was utilized to monitor the dimerization of all three dimer pairs (ERα/α, ERβ/β, and ERα/β) [10], has been used to screen the dimerization ability of xenoestrogens [30], and has been optimized for high-throughput screening of chemicals that induce ER dimerization [9, 15].

This chapter describes Bioluminescent Resonance Energy Transfer assays as high-throughput screening tools to identify compounds that interfere with or induce protein–protein interactions with emphasis on studies aimed at the identification of xenoestrogens that interfere with or induce the dimerization of the estrogen receptors. The assay can be adapted for studying other protein–protein interactions in the living cells.

## 2    Reagents, Materials, and Equipment

- LT1 Transfection ReagentTrypsin.
- HEK293 cells (ATCC, CRL-1573).
- Expression vectors containing the donor (Rluc) and acceptor (YFP) moieties.
- Dulbecco's Modified Eagle's Medium (DMEM) supplemented with fetal bovine serum (FBS) and DMEM phenol-red free medium supplemented with charcoal-stripped FBS.
- Phosphate buffered saline (PBS).
- Opti-MEM: a reduced serum medium.
- Coelenterazine h (Promega).
- Automated robotic system (Beckman Biomek FX).
- A microplate reader with the appropriate filters for detecting luminescence and fluorescence.
  - For 96-well plate: Victor Wallac V plate reader (PerkinElmer).
  - For 384-well plate: Safire 2 plate reader (Tecan).

## 3    Methods

*Safety considerations*

Always follow basic laboratory safety rules:

(a) Always wear clean laboratory outer garment.
(b) Wear protective glasses to avoid splashes to the eyes.
(c) Wear gloves when handling any reagents or samples
(d) Cell culture should be carried out inside a biosafety cabinet.

### 3.1    Development and Characterization of BRET Fusion Protein Plasmids

While all BRET assays possess the common theme of a bioluminescent donor and fluorescent acceptor protein, different mutational variants of donor and acceptor may be chosen based on the desired application and biological system in question. For consideration in this protocol, BRET1 is described, which utilizes Rluc as the donor and YFP as the acceptor (Fig. 1).

#### 3.1.1    Construction of BRET Fusion Plasmids

To generate the fusion proteins, the cDNA of the proteins of interest is cloned into a eukaryotic expression vector containing the donor or acceptor proteins.
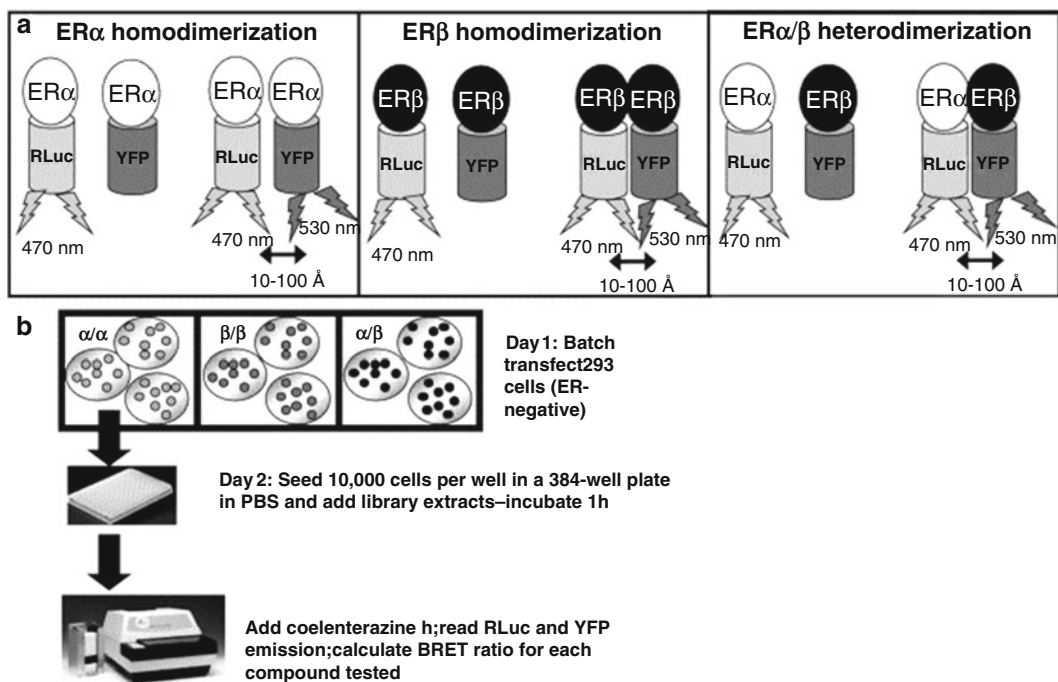
**Fig. 1** BRET assay methodology. (**a**) Schematic representing ligand-dependent dimerization and resonance energy transfer between RLuc and YFP fusions via BRET. (**b**) Schematic representing the BRET assay 96-well format protocol. Adapted from ref. [15]

*3.1.2 Design of Fusion Constructs for Use in BRET Assays*

(a) Using standard molecular biology techniques employing PCR and fragment replacement strategies, proteins of interest are cloned into expression vectors containing the donors and acceptor Rluc and YFP at either the N- or C-terminus. Each construct should be designed to include a 10–14 amino acid flexible linker between the receptor and Rluc or YFP (*see* **Notes 1** and **2**).

Full length ERs have not been crystallized, and thus the dimerized conformations of these proteins are difficult to predict. Therefore, the following constructs should be made to express the fusion proteins under the control of the CMV promoter at both the N- and C-terminus. All permutations of N- and C-terminal fusions should be tested for maximal signal output. To be able to study all combination of dimer pairs ERα and ERβ were fused to either Rluc or YFP (*see* **Note 3**). For example, using the vector backbone pCMX:

pCMX-ERα-RLuc.

pCMX-YFP-ERβ.

pCMX-Pl2 (negative control).

ER-RLuc denotes that RLuc is fused to the C terminus of ER, whereas YFP-ER indicates that YFP is fused to the N terminus of ER.

**3.2 Characterization and Assessment of Fusion Construct Functionality**

To characterize the fusion protein of interest many different methods may be utilized including:

1. Sequencing of all constructs to verify that the proteins of interest are in frame with the Rluc/YFP sequence.

2. Relative protein expression levels of each construct by Western blotting using antibodies towards proteins of interest.

3. Functional assays to ensure that the proteins of interest retain their biological activity despite addition of the YFP or RLuc. Whole cell ligand binding assays may be performed to demonstrate the retained ability of the fusion constructs to bind its endogenous ligand E2. Because ERs are transcription factors, retained transcriptional activity of the fusion proteins can be assessed by estrogen response element (ERE)-luciferase reporter assays. The ability of the fusion constructs to bind directly to an ERE can be assessed by electrophoretic mobility shift assays [10]. Subcellular localization of YFP fusion proteins can be visualized by fluorescence microscopy. Luminescence and fluorescence of the fusion proteins should also be confirmed. YFP can be assessed by direct visualization using a fluorescence microscope or excitation at the appropriate wavelength on a plate reader with fluorescence capabilities, and the RLuc can be assessed on a luminescence-capable plate reader following addition of its substrate coelenterazine *h* (*see* **Note 4**).

**3.3 Transfection and Expression of Fusion Vectors**

Once the fusion constructs have been developed and validated, the plasmids are transfected into the host cells to carry out the BRET assay. The most commonly used mammalian cell line for transient transfection of BRET fusion proteins is the HEK293 cell line. Cell line selection will depend on the laboratory, system, and proteins of interest being studied.

(a) Grow and maintain HEK293 cells in DMEM media supplemented with 10% FBS.

(b) Split and seed ~$1.5 \times 10^5$ HEK293 cells onto 6-well plates in phenol red-free DMEM supplemented with 10% charcoal-stripped FBS (*see* **Notes 5** and **6**).

(c) DNA mixes should be diluted in Opti-MEM medium without serum, mixed gently, and incubated with a transfection agent such as LT1 following manufacturer's instructions.

- In this case, HEK293 cells were transfected in 6 different conditions (*see* **Notes 7** and **8**):
  *Mock-transfected cells:*
  - Using the empty vector as a negative control to establish background levels of luminescence and fluorescence.

*With a single BRET fusion plasmid:*

– Single BRET fusion plasmid of pCMX-ERα-RLuc.

– Single BRET fusion plasmid of pCMX-RLuc-ERβ.

*Co-transfected with RLuc and YFP BRET fusions:*

– pCMX-ERα-RLuc+pCMX-YFP-ERβ    for    ERα/ERβ heterodimers;

– pCMX-ERα-RLuc+pCMX-ERα-YFP    for    ERα homodimers.

– pCMX-RLuc-ERβ+pCMX-YFP-ERβ    for    ERβ homodimers.

(d) Add the DNA-LT1 solution to cells in a 6-well plate. Mix gently by rocking the dish back and forth.

(e) Incubate the cells at 37 °C and allow 24–48 h for protein expression. Cells transfected with YFP containing plasmids can be visualized with a florescent microscope to verify transfection efficiency.

### 3.4 Seeding Transfected Cells for High-Throughput Screening

After transfection and allowing enough time for protein expression (typically 24-72 h), the cells can be seeded for the BRET assay.

*3.4.1 Harvesting Transfected Cells*

(a) Cells are trypsinized and counted using a cell counter.

(b) Cells are seeded at 11,000 cells per well on a 384-well white-bottom plates in 40 µL PBS. Alternatively, this assay may be scaled up to a 96-well format by proportionally increasing cell numbers and volumes (*see* **Note 9**).

*3.4.2 Screening of Chemicals*

(a) An automated pipetting system can be used to add the chemicals to the cell suspension. For this protocol 0.2 µL of 1 mM library compound was added to each well using the Biomek FX Robot such that the final concentration per well was 5 µM (*see* **Note 10**).

(b) Cells should also be treated with vehicle alone as a negative control and 17-β Estradiol as a positive control. All treatments should be done in minimum replicates of three.

(c) Cell suspensions are incubated with chemicals and controls for 1 h in the dark at room temperature. Timing may need to be optimized based on kinetics of the cell line utilized.

(d) The RLuc substrate coelenterazine *h* is then added to a final concentration of 5 µM. The plates are then gently shaken on a plate shaker for 10 s at 300 rpm (*see* **Note 11**).

### 3.5 Measurement of the BRET Signal

Measurement of the BRET signal is read on a microplate reader that contains the appropriate filters and has the capacity to simultaneously or sequentially detect the two distinct wavelengths.

(a) Immediately after the addition of the RLuc substrate, emission of RLuc and YFP is read on a microplate reader. Rluc emission is read at 460 nm followed immediately by YFP emission at 535 nm at 0.1 s per wavelength read per well. Each RLuc and YFP emission measurement should be taken consecutively per well before moving to the next well (*see* **Note 12**).

**3.6 Calculating BRET Ratios and Data Analysis**

Corresponding emission values are used to calculate the BRET ratio as described in [10].

*3.6.1 Calculating BRET Ratio*

(a) The BRET ratio is calculated as a ratio of YFP emission to RLuc emission with subtraction of background and a correction factor (CF) for emission spectra isolation. The broad tail of the RLuc emission spectrum "leaks" into the range of detection for the peak emission of YFP at 530 nm. The CF, obtained by expressing the RLuc-ER construct in the absence of the YFP fusion construct, allows for the separation of the contributions of YFP and RLuc to the emission intensity measured at 530 nm.

$$\textbf{BRET ratio} = \left[\frac{\overbrace{530\ emission}^{\substack{\text{RLuc and YFP fusions}\\\text{cotransfected}}} - \overbrace{530\ emission}^{\substack{\text{Mock Transfected}}}}{\underbrace{460\ emission} - \underbrace{460\ emission}}\right] - \left[\frac{\overbrace{530\ emission}^{\substack{\text{RLuc transfected alone}}} - \overbrace{530\ emission}^{\substack{\text{Mock Transfected}}}}{460\ emission - 460\ emission}\right]$$

(b) An increase in the BRET ratio is indicative of a specific interaction between the two tagged proteins.

# 4  Notes

1. The stop codon between the cDNA of the protein of interest and the fusion protein should be removed to guarantee that one single protein of interest and fusion protein is produced [31].

2. A peptide linker sequence can be used to add flexibility between the fusion protein and protein of interest. In cases where the donor and acceptor proteins might not be in proximity when protein–protein interaction occurs, the linker could facilitate the alignment of donor and acceptor dipoles [3, 32].

3. It is important to take into account the relative orientation of the RLuc and YFP fusion to ensure that there is no disruption of protein structure or function. In some occasions the proper combination of fusion proteins at both the amino and carboxyl end must be used to account for the spatial orientation of the protein–protein interaction that allows for optimal and

efficient energy transfer to occur and maximal BRET signal output [3, 10].

4. To confirm the correct interaction of the BRET constructs, molecular simulation can be performed to identify the interaction interface if the structure of the proteins being investigated is known. Directed mutation of this interface provides a negative control [33].

5. The use of phenol red media should be avoided at the time of transfection and thereafter because phenol red absorbs light thus disrupting the luminescence signal. Phenol red also possesses a low level of estrogenic activity, thus decreasing assay sensitivity.

6. Similarly, FBS should also be stripped of hormone activity using charcoal and dextran to ensure that the ligand binding pocket of ERs is empty, thus increasing assay sensitivity.

7. Several optimization procedures should be done to achieve a reliable BRET assay. It is essential to establish the optimal donor to acceptor ratio for reliable BRET signals. The desirable concentration can be determined by titration; maintain the levels of donor expression constant and increase the levels of acceptor protein. Increasing the level of acceptor fusions should lead to saturation of the BRET signal. The lowest ratio of donor and acceptor that does not increase background signal should be chosen.

8. Competition assays can be performed to determine the specificity of the fusion proteins to each other. These assays verify that increases in BRET ratios are not the result of random collisions between donor and acceptor moieties. Competition assays are performed by maintaining the donor and acceptor proteins at constant levels and titrating either increasing amounts of the untagged protein of interest, which would be fused with the acceptor protein, or another known non-interacting protein. BRET signal should be disrupted if using the untagged protein of interest, but if using a non-interacting protein the BRET signal should not change. Once BRET has been optimized it is a straightforward technique with a high degree of reproducibility.

9. White microplates are used to avoid light absorption.

10. The color of the chemicals tested can influence the BRET signal. If the compound is florescent and shares the same spectral properties as the acceptor, it can promote a change in the BRET signal [17]. The effect of the color on the BRET signal can be determined by incubating just the donor without the acceptor and the chemical; if an increase in the BRET signal is observed, this indicates that the color of the compounds is interfering with the BRET reading.

11. Coelenterazine h is best reconstituted in ethanol, as it has low solubility in water and is unstable in DMSO.

12. Single cell and live animal BRET measurements can be achieved using red-shifted BRET vectors and a charge-coupled device (CCD) camera. BRET from live animal tissues is complicated by tissue attenuation of photons, and therefore additional mathematical calculations are required [34].

## 5 Summary and Discussion

Here we describe a BRET system that was developed to monitor the formation of estrogen receptor dimerization and to identify potential ER dimer selective small molecule compounds [9, 10]. This assay measures energy transfer between ER proteins fused to *Renilla* luciferase (Rluc, the donor) or yellow fluorescent protein (YFP, the acceptor) and YFP emission indicates dimerization. Using the BRET assay, ERα/α and ERβ/β homodimerization were measured in response to ligand, as well as ERα/β heterodimerization, revealing ligands that selectively induce heterodimerization or homodimerization.

BRET-based studies provide robust insight into the identification of chemicals that induce protein–protein interactions in a physiologically relevant environment in real time. BRET can be applied to study a vast array of protein systems. The optimization of high-throughput BRET for screening assays leverages automation to quickly facilitate the discovery of new chemicals which are biologically and biochemically relevant. This presents a useful alternative to other low-throughput methods of examining protein–protein interactions or ligand binding.

Like any biological assay, BRET has innate disadvantages. For example, in the ER-BRET assay described here, the ERβ/β condition shows a higher background than the ERα/α and ERα/β condition. This may be due to ligand induction of conformational changes which affect the position of the donor or acceptor; alternatively, this high background may due to ligand-independent dimerization. This may be circumvented by using another BRET system with a better spectral resolution. Indeed, considerable efforts have been made to improve the BRET technology with the development of new fusion tags, substrates, and equipment. Assays like BRET3 have a clean separation emission signal that circumvent the need for calculation of the correction factor portion of the BRET ratio, which minimize the variability and improve signal-to-noise ratio [35].

An additional concern is that the fusion proteins can cause steric hindrance, thus inhibiting protein functionality. Thus, thorough characterization of fusion constructs is essential to assay integrity. This concern has led to the recent development of smaller fusion partners [22]. Another limitation is the fact that transfection of fusion plasmids can lead to non-physiologically relevant

expression of proteins; this concern can be resolved by using the titration method described above or by engineering stable expression of fusion proteins in cells.

Despite these caveats, the versatility of BRET still makes it a powerful high-throughput screening tool. In most cases, these caveats can be controlled for or avoided with careful optimization. BRET has the advantage of monitoring protein–protein interactions in live, intact cells in real time and can be applied to investigate almost any type of protein–protein interaction with high reliability.

## References

1. Phizicky EM, Fields S (1995) Protein-protein interactions: methods for detection and analysis. Microbiol Rev 59(1):94–123

2. Xu Y, Piston DW, Johnson CH (1999) A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. Proc Natl Acad Sci U S A 96(1):151–6

3. Pfleger KD, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). Nat Methods 3(3):165–74

4. Subramanian C et al (2006) A suite of tools and application notes for in vivo protein interaction assays using bioluminescence resonance energy transfer (BRET). Plant J 48(1):138–52

5. Xu X et al (2007) Imaging protein interactions with bioluminescence resonance energy transfer (BRET) in plant and mammalian cells and tissues. Proc Natl Acad Sci U S A 104(24):10264–9

6. Angers S et al (2000) Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET). Proc Natl Acad Sci U S A 97(7):3684–9

7. De A et al (2009) BRET3: a red-shifted bioluminescence resonance energy transfer (BRET)-based integrated platform for imaging protein-protein interactions from single live cells and living animals. FASEB J 23(8):2702–9

8. Bacart J et al (2008) The BRET technology and its application to screening assays. Biotechnol J 3(3):311–24

9. Powell E et al (2012) Identification of estrogen receptor dimer selective ligands reveals growth-inhibitory effects on cells that co-express ERα and ERβ. PLoS One 7(2):e30993

10. Powell E, Xu W (2008) Intermolecular interactions identify ligand-selective activity of estrogen receptor alpha/beta dimers. Proc Natl Acad Sci U S A 105(48):19012–7

11. Stoddart LA et al (2015) Application of BRET to monitor ligand binding to GPCRs. Nat Methods 12(7):661–3

12. Boute N, Jockers R, Issad T (2002) The use of resonance energy transfer in high-throughput screening: BRET versus FRET. Trends Pharmacol Sci 23(8):351–4

13. Szidonya L, Cserzo M, Hunyady L (2008) Dimerization and oligomerization of G-protein-coupled receptors: debated structures with established and emerging functions. J Endocrinol 196(3):435–53

14. Salahpour A et al (2012) BRET biosensors to study GPCR biology, pharmacology, and signal transduction. Front Endocrinol 3:105

15. Powell E et al (2010) Identification and characterization of a novel estrogenic ligand actinopolymorphol A. Biochem Pharmacol 80(8):1221–9

16. Hamdan FF et al (2005) High-throughput screening of G protein-coupled receptor antagonists using a bioluminescence resonance energy transfer 1-based beta-arrestin2 recruitment assay. J Biomol Screen 10(5):463–75

17. Couturier C, Deprez B (2012) Setting up a bioluminescence resonance energy transfer high throughput screening assay to search for protein/protein interaction inhibitors in mammalian cells. Front Endocrinol 3:100

18. Pfleger KD, Seeber RM, Eidne KA (2006) Bioluminescence resonance energy transfer (BRET) for the real-time detection of protein-protein interactions. Nat Protoc 1(1):337–45

19. Borroto-Escuela DO et al (2013) Bioluminescence resonance energy transfer methods to study G protein-coupled receptor-receptor tyrosine kinase heteroreceptor complexes. Methods Cell Biol 117:141–64

20. Bertrand L et al (2002) The BRET2/arrestin assay in stable recombinant cells: a platform to screen for compounds that interact with G

protein-coupled receptors (GPCRS). J Recept Signal Transduct Res 22(1-4):533–41

21. Kocan M et al (2010) Enhanced BRET technology for the monitoring of agonist-induced and agonist-independent interactions between GPCRs and β-arrestins. Front Endocrinol 1:12

22. Machleidt T et al (2015) NanoBRET-A novel BRET platform for the analysis of protein-protein interactions. ACS Chem Biol 10(8):1797–804

23. Pfleger KD et al (2006) Extended bioluminescence resonance energy transfer (eBRET) for monitoring prolonged protein-protein interactions in live cells. Cell Signal 18(10): 1664–70

24. Loening AM, Wu AM, Gambhir SS (2007) Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. Nat Methods 4(8):641–3

25. Giuliani G et al (2012) New red-shifted coelenterazine analogues with an extended electronic conjugation. Tetrahedron Lett 53(38):5114–5118

26. Levi J et al (2007) Bisdeoxycoelenterazine derivatives for improvement of bioluminescence resonance energy transfer assays. J Am Chem Soc 129(39):11900–1

27. Michelini E et al (2004) Development of a bioluminescence resonance energy-transfer assay for estrogen-like compound in vivo monitoring. Anal Chem 76(23):7069–76

28. Koterba KL, Rowan BG (2006) Measuring ligand-dependent and ligand-independent interactions between nuclear receptors and associated proteins using Bioluminescence Resonance Energy Transfer (BRET). Nucl Recept Signal 4:e021

29. Lupien M et al (2007) Raloxifene and ICI182,780 increase estrogen receptor-alpha association with a nuclear compartment via overlapping sets of hydrophobic amino acids in activation function 2 helix 12. Mol Endocrinol 21(4):797–816

30. Sievers CK et al (2013) Differential action of monohydroxylated polycyclic aromatic hydrocarbons with estrogen receptors α and β. Toxicol Sci 132(2):359–67

31. Pfleger KD, Eidne KA (2003) New technologies: bioluminescence resonance energy transfer (BRET) for the detection of real time interactions involving G-protein coupled receptors. Pituitary 6(3):141–51

32. Deriziotis P et al (2014) Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. J Vis Exp 87

33. Borroto-Escuela DO et al (2010) Characterization of the A2AR-D2R interface: focus on the role of the C-terminal tail and the transmembrane helices. Biochem Biophys Res Commun 402(4):801–7

34. De A et al (2013) Evolution of BRET biosensors from live cell to tissue-scale in vivo imaging. Front Endocrinol 4:131

35. De A, Loening AM, Gambhir SS (2007) An improved bioluminescence resonance energy transfer strategy for imaging intracellular events in single cells and living subjects. Cancer Res 67(15):7175–83

# Chapter 2

# Mitochondrial Membrane Potential Assay

**Srilatha Sakamuru, Matias S. Attene-Ramos, and Menghang Xia**

## Abstract

Mitochondrial function, a key indicator of cell health, can be assessed by monitoring changes in mitochondrial membrane potential (MMP). Cationic fluorescent dyes are commonly used tools to assess MMP. We used a water-soluble mitochondrial membrane potential indicator (m-MPI) to detect changes in MMP in HepG2 cells. A homogenous cell-based MMP assay was optimized and performed in a 1536-well plate format to screen several compound libraries for mitochondrial toxicity by evaluating the effects of chemical compounds on MMP.

**Key words** Mitochondrial membrane potential (MMP), Mitochondrial membrane potential indicator (m-MPI), Mitochondrial toxicity, 1536-well plate format, Mesoxalonitrile 4-trifluoromethoxyphenyl hydrazone (FCCP)

## 1 Introduction

Mitochondria, commonly referred to as power houses of the cell, play a vital role in cellular physiology. The majority of the cellular energy (ATP) in eukaryotic cells is generated in the mitochondria through oxidative phosphorylation [1], during which electrons are transferred from electron donors to electron acceptors such as oxygen. The mitochondrial electron transport chain creates an electrochemical gradient through a series of redox reactions. This electrochemical gradient drives the synthesis of ATP [2] and generates the mitochondrial membrane potential (MMP), which is a key parameter for evaluating mitochondrial function [3].

Mitochondrial dysfunctions have been associated with various disorders such as cancer, cardiovascular diseases, diabetes, and neurodegenerative diseases [4]. The toxicity of xenobiotic compounds can have either a direct or a secondary effect on mitochondrial function. Many of these compounds reduce MMP by perturbing a variety of macromolecules in the mitochondria, and therefore affecting different mitochondrial functions. A decrease in the
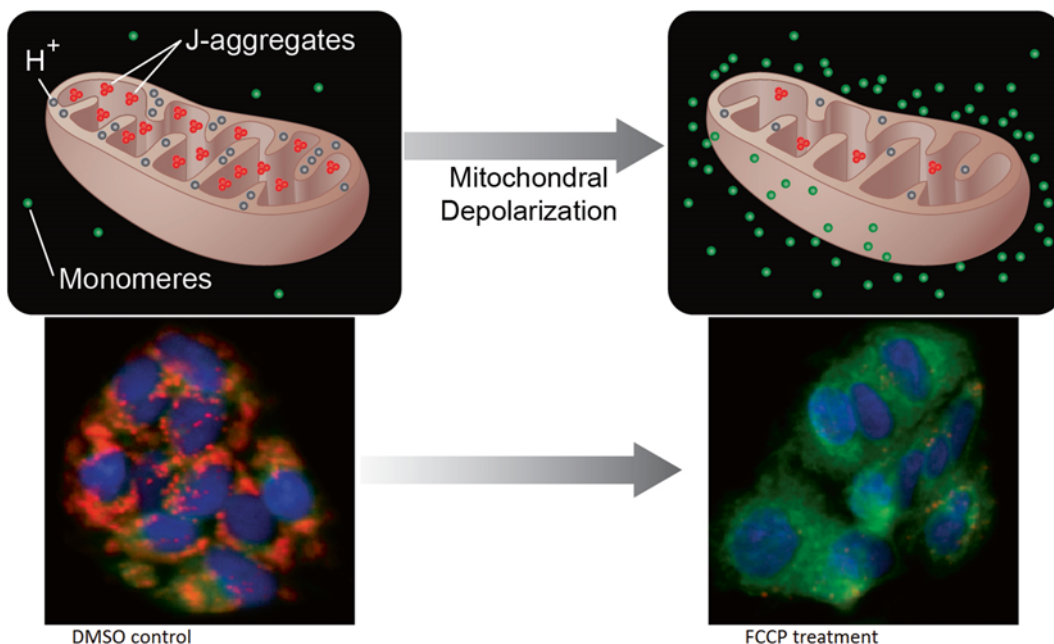
**Fig. 1** MMP assay principle: In the healthy cells, m-MPI dye accumulates in mitochondria as aggregates showing *red* fluorescence. When mitochondrial potential collapses after FCCP treatment, the m-MPI dye remains in cytoplasm with *green* fluorescence. Figure reproduced from Ref. [6]

MMP may also be linked to apoptosis [5]. Thus these organelles are an ideal target for in vitro toxicity studies.

Several cell membrane permeable fluorescent dyes, such as 3, 3′-dihexyloxacarbocyanine iodide [DiOC6(3)], rhodamine-123 (Rh123), tetramethyl rhodamine methyl and ethyl esters (TMRM and TMRE), and JC-1, are currently available to measure changes in MMP. Based on the assay optimization of our previous study [6], we selected the water-soluble m-MPI indicator to determine mitochondrial toxicity by screening the compound libraries against HepG2 cells in a 1536-well plate format. In healthy cells, m-MPI accumulates in the mitochondria as red fluorescent aggregates (emission at 590 nm). When MMP depolarizes and cells become less healthy, m-MPI aggregates are converted to green fluorescent monomers (emission at 535 nm) and remain in the cytoplasm (Fig. 1). So the red/green fluorescence ratio can be used in determining the mitochondrial function of cells.

## 2    Materials

### 2.1    Equipment

1. Purifier Logic + Class II, Type A2 Biosafety Cabinet for cell operations.

2. Steri-Cult $CO_2$ Incubator for culturing cells at 37 °C under a humidified atmosphere and 5 % $CO_2$.

3. Multidrop™ Combi Reagent Dispenser (Thermo Scientific, Waltham, MA) for dispensing cells into 1536-well plates by using an 8-tip dispense cassette.

4. Pintool workstation (Wako Automation, San Diego, CA) for transferring 23 nL of compounds from a compound plate to an assay plate.

5. BioRAPTR Flying Reagent Dispenser™ (FRD) workstation (Beckman Coulter, Inc., Brea, CA) for dispensing reagent into a 1536-well plate.

6. EnVision® Multilabel Plate Reader (Perkin Elmer, Shelton, CT) for reading fluorescence intensity.

7. ViewLux uHTS Microplate Imager (Perkin Elmer) for reading luminescence intensity.

8. ImageXpress Micro Widefield High Content Screening system (Molecular Devices, Sunnyvale, CA) for imaging purposes.

*2.2 Reagents/ Supplies*

1. Human HepG2 (hepatocellular carcinoma) cell line was purchased from ATCC.

2. Culture medium for HepG2 cells: 1000 mL of Eagle's Minimum Essential Medium, 100 mL of fetal bovine serum, and 10 mL of 10,000 U/mL penicillin–streptomycin.

3. Trypsin–EDTA (0.05 %).

4. DPBS without calcium and magnesium.

5. Mitochondrial Membrane Potential Indicator (m-MPI) (Codex BioSolutions, Inc., Gaithersburg, MD).

6. CellTiter-Glo® Luminescent Cell Viability Assay (Promega Corporation, Madison, WI).

7. Hoechst 33342.

8. Mesoxalonitrile 4-trifluoromethoxyphenylhydrazone, FCCP, (positive control compound for the assay, CAS Registry Number, CASRN, 370-86-5).

9. Tetraoctyl ammonium bromide (positive control for cytotoxicity assay, CASRN 14866-33-2).

10. 1536-well black wall/clear bottom, white wall/solid bottom and clear polystyrene microplates for MMP assay, cytotoxicity assay and compound storage respectively.

# 3  Methods

*3.1 Cell Culture*

1. HepG2 cells obtained as a frozen stock, were thawed in culture medium by adding 1 mL of frozen stock to 9 mL of medium, and were centrifuged for 4 min at 900 rpm.

2. The seeding density for thawing was $2.0 \times 10^6$ cells per T-75 flask.

3. HepG2 cells were grown at 37 °C, 5 % $CO_2$ and 95 % humidity in T225 flasks in Steri-Cult $CO_2$ Incubator.

4. For the expansion, the culture medium was aspirated and the monolayer was rinsed twice with DPBS, followed by the addition of 7 mL of Trypsin–EDTA solution.

5. The cells were detached from the surface by incubation for 3–4 min at 37 °C with Trypsin–EDTA, and resuspended with culture medium. The cells were then centrifuged for 4 min at 900 rpm.

6. The seeding density for expansion was $4.0 \times 10^6$ cells per T225 flasks.

*3.2 Quantitative High-Throughput Screening (qHTS) Protocol of MMP and Cell Viability Multiplex Assay [6, 7]*

1. The human HepG2 cells were harvested from the 80–90 % confluent culture flasks by using Trypsin–EDTA for detachment of the cells from the culture flask. The cells were centrifuged and the resulting pellet was resuspended with culture medium.

2. The cells were plated at 2000 cells/well in 5 μL of the culture medium into a 1536-well black wall/clear bottom plate using Multidrop Combi (*see* **Note 1**).

3. The assay plates were incubated overnight at 37 °C for cell adhesion.

4. The adhered cells were treated with 23 nL of test compounds and positive control using a Pintool. The test compounds were transferred to columns 5–48, the positive control compound (FCCP) was transferred to columns 1–3 (Dose titration in column 1 at a start final concentrations of 11.5 μM with 1:1.5 dilutions; Columns 2 and 3 with 6.9 and 3.5 μM FCCP, respectively), and DMSO only was transferred to column 4.

5. The assay plates were incubated at 37 °C for 1 h or 5 h.

6. After the respective incubation times, 5 μL of 2x m-MPI dye-loading solution (10 μL of m-MPI stock solution added to 5 mL of MMP assay buffer, mixed well by vortexing) was added to each well using FRD (*see* **Note 2**).

7. The assay plates were incubated at 37 °C for 30 min.

8. Fluorescence intensity (485 nm excitation and 535 nm emission for green fluorescent monomers, 540 nm excitation and 590 nm emission for red fluorescent aggregates) was measured using an Envision plate reader.

9. Data were expressed as the ratio of 590 nm/540 nm emissions, an indicator of MMP. The positive control, FCCP, concentration-dependently decreases MMP with $IC_{50}$s of 44 and 116 nM for 1 and 5 h treatment, respectively (Fig. 2).

10. Right after MMP assay, 2 μL of CellTiter-Glo® reagent was added to each well using FRD.
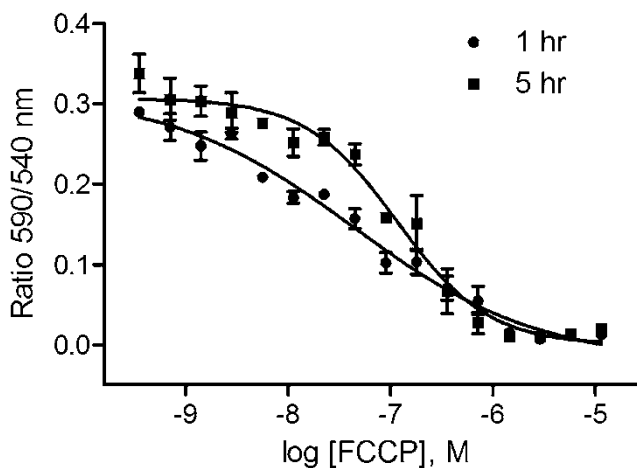
**Fig. 2** Concentration-response curves of FCCP after 1 h or 5 h treatment

11. The assay plates were incubated at room temperature for 30 min.

12. Luminescence intensity was measured using Viewlux plate reader.

***3.3   Imaging Based MMP Assay [6, 8]***

1. The human HepG2 cells were plated at 2000 cells/well in 5 μL of the culture medium into a 1536-well black wall/clear bottom plate using Multidrop Combi.

2. The assay plates were incubated overnight at 37 °C for cell adhesion.

3. The adhered cells were treated with the test compounds and positive control (FCCP, 6.9 and 3.5 μM).

4. The treated plates were incubated at 37 °C for 1 or 5 h.

5. After the respective incubation times, 5 μL of 2x m-MPI dye-loading solution with 0.3 μg/mL of Hoechst 33342 were added to each well using FRD.

6. The assay plates were incubated at 37 °C for 30 min.

7. The fluorescence intensities (482 nm excitation and 536 nm emission for green fluorescent monomers; 543 nm excitation and 593 nm emission for red fluorescent aggregates; 377 nm excitation and 447 nm emission for Hoechst 33342) were measured using an ImageXpress Screening System.

8. Imaging was processed and analyzed with the MetaXpress® and PowerCore® software using the Multi Wavelength Cell Scoring algorithm. The mean of average fluorescence intensity from each positive cell was calculated per well for both green and red fluorescent colors.

9. Data were expressed as ratio of 593 nm/536 nm emissions.

## 4    Notes

1. The black wall/clear bottom assay plates should not be touched at the bottom as the fluorescence intensity is read from the bottom of the plate.

2. For proper mixing of the m-MPI dye with the buffer, the buffer should be taken out from 4 °C a couple of hours prior to the assay in order to reach room temperature.

**References**

1. Chance B, Williams GR (1956) The respiratory chain and oxidative phosphorylation. Adv Enzymol Relat Subj Biochem 17:65–134

2. Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature 191:144–148

3. Chen LB (1988) Mitochondrial membrane potential in living cells. Annu Rev Cell Biol 4:155–181

4. Pieczenik SR, Neustadt J (2007) Mitochondrial dysfunction and molecular pathways of disease. Exp Mol Pathol 83(1):84–92

5. Lemasters JJ, Qian T, He L, Kim JS, Elmore SP, Cascio WE, Brenner DA (2002) Role of mitochondrial inner membrane permeabilization in necrotic cell death, apoptosis, and autophagy. Antioxid Redox Signal 4:769–781

6. Sakamuru S, Li X, Attene-Ramos MS, Huang R, Lu J, Shou L, Shen M, Tice RR, Austin CP, Xia M (2012) Application of a homogenous membrane potential assay to assess mitochondrial function. Physiol Genomics 44(9):495–503

7. Attene-Ramos MS, Huang R, Michael S, Witt KL, Richard A, Tice RR, Simeonov A, Austin CP, Xia M (2015) Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. Environ Health Perspect 123(1):49–56

8. Attene-Ramos MS, Huang R, Sakamuru S, Witt KL, Beeson GC, Shou L, Schnellmann RG, Beeson CC, Tice RR, Austin CP, Xia M (2013) Systematic study of mitochondrial toxicity of environmental chemicals using quantitative high throughput screening. Chem Res Toxicol 26(9):1323–1332

# Chapter 3

## Using β-Lactamase and NanoLuc Luciferase Reporter Gene Assays to Identify Inhibitors of the HIF-1 Signaling Pathway

**Thai Khuc, Chia-Wen (Amy) Hsu, Srilatha Sakamuru, and Menghang Xia**

### Abstract

The hypoxia-inducible factor 1 (HIF-1) is a transcriptional factor involved in the regulation of oxygen within cellular environments. In hypoxic tissues or those with inadequate oxygen concentrations, activation of the HIF-1 transcription factor allows for subsequent activation of target gene expression implicated in cell survival. As a result, cells proliferate through formation of new blood vessels and expansion of vascular systems, providing necessary nourishment needed of cells. HIF-1 is also involved in the complex pathophysiology associated with cancer cells. Solid tumors are able to thrive in hypoxic environments by overactivating these target genes in order to grow and metastasize. Therefore, it is of high importance to identify modulators of the HIF-1 signaling pathway for possible development of anticancer drugs and to better understand how environmental chemicals cause cancer. Using a quantitative high-throughput screening (qHTS) approach, we are able to screen large chemical libraries to profile potential small molecule modulators of the HIF-1 signaling pathway in a 1536-well format. This chapter describes two orthogonal cell based assays; one utilizing a β-lactamase reporter gene incorporated into human ME-180 cervical cancer cells, and the other using a NanoLuc luciferase reporter system in human HCT116 colon cancer cells. Cell viability assays for each cell line are also conducted respectively. The data from this screening platform can be used as a gateway to study mode of action (MOA) of selected compounds and drug classes.

**Key words** Hypoxia inducible factor 1, Hypoxia response elements, Beta-lactamase, Luciferase, NanoLuc, Reporter gene, Cancer, Genome editing, Drug, Quantitative high-throughput screening, qHTS, Fluorescence resonance energy transfer, FRET

## 1 Introduction

Hypoxia-inducible factor (HIF-1), an oxygen-sensitive transcription factor, is the major regulator of cell development and survival in hypoxic, low oxygen tension environments. This transcription factor maintains oxygen homeostasis within cells by activating an array of anti-apoptotic genes implicating angiogenesis, erythropoiesis, glucose metabolism, and vascular expansion [1]. HIF-1 is a heterodimeric basic helix–loop–helix protein [2] consisting of a

hypoxic responsive HIF-1α subunit and a constitutively expressed HIF-1β subunit, known as the aryl hydrocarbon receptor nuclear translocator (ARNT) [3]. Under normoxic conditions, HIF-1α is quickly degraded through the ubiquitin–proteasome pathway [4, 5] under direct control of prolyl hydroxylases (PHDs), preventing formation of the transcription complex. However, in decreasing oxygen concentrations, HIF-1α levels increase exponentially [2] by reduction of prolyl hydroxylase activity [6]. The available HIF-1α forms dimeric complexes with HIF-1β. The transcriptionally active complex is able to translocate into the nucleus to bind DNA regulatory sequences known as hypoxia response elements (HREs) downstream to promoter regions [1] of target genes that are activated with assistance of transcriptional coactivators such as p300 and the cAMP response element (CREB)-binding protein (CBP) [7].

HIF-1 activity is also involved in the proliferation, metastasis, and pathophysiology associated with cancer. Intratumor hypoxia and overexpression of HIF-1 occurs as a result of increased oxygen consumption in tumor microenvironments. As a result, cancer cells utilize the HIF-1 signaling pathway to activate the same target genes, increasing their aggressiveness and persistence even in hypoxic conditions. Therefore, the hypoxic environment commonly associated with solid tumors can be exploited as a target for potential anticancer drug developments and therapies [8]. Several small molecule inhibitors of HIF-1 are currently in clinical use as anticancer drugs including topotecan (Hycamtin) and vorinostat (Zolinza) [9]. We have utilized a pair of orthogonal cell-based reporter gene assays using a quantitative high-throughput screening approach to rapidly identify small molecule modulators of the HIF-1 signaling pathway [10, 11]. In the first assay, an HRE driven β-lactamase reporter gene (HRE-bla) was stably transfected to human ME-180 cervical cancer cells, and the expressing beta-lactamase can cleave a chemically modified substrate that undergoes fluorescence resonance energy transfer (FRET). The second assay uses a HIF-1α- NanoLuc luciferase reporter allele stably integrated in human HCT116 colon cancer cells to enable sensitive detection of HIF-1 activity. Both assays have been successfully miniaturized into 1536-well plate formats to profile large chemical libraries including the Tox21 10K chemical collection. Subsequent cell viability assays were conducted for each cell line respectively.

## 2    Materials

### 2.1  β-Lactamase Reporter Gene Assay

*2.1.1  Cell Line and Culture Condition*

1. CellSensor® HRE-bla ME-180 cell line (Life Technologies, Carlsbad, CA).

2. 0.25 % Trypsin–EDTA, phenol red.

3. Dulbecco's Phosphate-Buffered Saline (DPBS).

4. Trypan Blue Solution, 0.4%.

5. Thaw Medium: DMEM medium supplement with 10% dialyzed fetal bovine serum (FBS), 0.1 mM nonessential amino acids (NEAA), and 50U/mL penicillin and 50 μg/mL streptomycin.

6. Cell Culture Medium: DMEM medium supplement with 10% dialyzed fetal bovine serum, 0.1 mM nonessential amino acids, 50U/mL penicillin and 50 μg/mL streptomycin, and 5 μg/mL blasticidin.

7. Assay Medium: Opti-MEM medium supplement with 0.5% dialyzed fetal bovine serum, 0.1 mM nonessential amino acids, and 50 U/mL penicillin and 50 μg/mL streptomycin.

8. Cellometer Auto Cell Counter (Nexcelom Bioscience, Lawrence, MA).

9. $CO_2$ incubator with variable oxygen control.

*2.1.2 Assay Reagents and Detection*

1. Solution A containing LiveBLAzer™ FRET B/G Substrate (CCF4-AM), Solution B, and Solution C (Life Technologies, Carlsbad, CA).

2. Solution D (Life Technologies, Carlsbad, CA).

3. CellTiter-Glo® Luminescent Cell Viability assay (Promega, Madison, WI).

4. 1536-well tissue culture treated black clear bottom microplates.

5. Assay and compound metal lids.

6. PinTool workstation (Wako Automation, San Diego, CA).

7. BioRAPTR flying reagent dispenser (Beckman Coulter, Pasadena, CA).

8. Envision plate reader (PerkinElmer, Waltham, MA).

9. ViewLux plate reader (PerkinElmer, Waltham, MA).

10. Envision bottom mirror: Beta-Lactamase Dual Enh. D425 nm/D490 nm #661, excitation filter: Photometric 405 nm, emission filter: FITC 535 nm and second emission filter: Umbelliferone 460 nm (PerkinElmer, Waltham, MA).

## 2.2 NanoLuc Luciferase Reporter Gene Assay

*2.2.1 Cell Line and Culture Condition*

1. X-Man® HIF1A NanoLuc HCT116 protein reporter cell line (Horizon Discovery, Cambridge, UK).

2. Thaw, Cell Culture, and Assay Medium: RPMI 1640 medium supplement with 10% Hyclone defined fetal bovine serum, 100U/mL penicillin and 100 μg/mL streptomycin, and 300 μg/mL G418.

3. 0.05% Trypsin–EDTA, phenol red.

4. Dulbecco's Phosphate-Buffered Saline (DPBS).

5. Trypan Blue Solution, 0.4%.

6. Cellometer Auto Cell Counter.

7. $CO_2$ incubator with variable oxygen control.

*2.2.2 Assay Reagents and Detection*

1. Nano-Glo® Luciferase assay (Promega, Madison, WI).

2. CellTiter-Glo® Luminescent Cell Viability assay.

3. 1536-well tissue culture treated white wall/solid bottom plates.

4. BioRAPTR flying reagent dispenser.

5. Assay and compound metal lids.

6. PinTool workstation.

7. ViewLux plate reader.

# 3  Methods

*3.1  Thawing Cells*

1. Add 9 mL of pre-warmed Cell Culture Medium in a 15 mL conical tube.

2. Thaw a vial of cells in a 37 °C water bath for 1–2 min (*see* **Note 1**).

3. Transfer thawed cells to the conical tube and gently mix the content.

4. Centrifuge the tube for 4 min at $200 \times g$ at room temperature.

5. Aspirate supernatant and resuspend cell pellet with 10 mL of Cell Culture Medium.

6. Count total cell number and transfer $2 \times 10^6$ HIF-1α-NanoLuc HCT116 to a T75 flask and $1 \times 10^6$ HRE-bla ME-180 cells to a T75 flask with a final medium volume of 10 mL (*see* **Note 2**).

7. Incubate the flask in a humidified incubator at 37 °C, 5 % $CO_2$, and 20 % $O_2$ until 80–90 % confluence (*see* **Notes 3** and **4**).

*3.2  Culturing Cells*

1. Aspirate Cell Culture Medium from a T75 flask of cells.

2. Rinse cell layer with $Ca^{2+}/Mg^{2+}$-free DPBS (*see* **Note 5**).

3. Add 2 mL of 0.05 % Trypsin–EDTA to the HIF-1α-NanoLuc HCT116 cell flask and 0.25 % Trypsin–EDTA to the HRE-bla ME-180 cell flask (*see* **Note 5**).

4. Incubate the flask in a humidified incubator at 37 °C, 5 % $CO_2$ until all cells are detached as verified by a tissue culture microscope (*see* **Note 6**).

5. Add 4 mL of Cell Culture Medium to the flask to terminate Trypsin-EDTA action.

6. Transfer the detached cells to a 50 mL conical tube.

7. Centrifuge the tube for 4 min at $200 \times g$ at room temperature.

8. Aspirate supernatant and resuspend the cell pellet with 10 mL of Cell Culture Medium.

9. Count total cell number and passage cells at 1:10–1:15 ratios twice per week.

10. Incubate the flask in a humidified incubator at 37 °C, 5 % $CO_2$, and 20 % $O_2$ until 80–90 % confluence.

**3.3 Plating Cells and Compound Treatment**

1. Aspirate Cell Culture Medium from a flask of cells.

2. Rinse cell layer with $Ca^{2+}/Mg^{2+}$-free DPBS.

3. Add 2 mL per T75 flask or 6 mL per T225 flask of 0.05 % Trypsin–EDTA to the HIF-1α-NanoLuc HCT116 cell flask and 0.25 % Trypsin–EDTA to the HRE-bla ME-180 cell flask.

4. Incubate the flask in a humidified incubator at 37 °C, 5 % $CO_2$ until all cells are detached as verified by a tissue culture scope.

5. Transfer the detached cells to a 50 mL conical tube and measure cell density.

6. Centrifuge the tube for 4 min at $200 \times g$ at room temperature.

7. Aspirate supernatant and resuspend the cell pellet with 10 mL of pre-warmed assay medium.

8. Pass the cell suspension through a cell strainer and collect the flow through.

9. Count cell number and dilute to $3 \times 10^5$ HIF-1α-NanoLuc HCT116 cells per mL and $5 \times 10^5$ HRE-bla ME-180 cells per mL.

10. Dispense 5 μL of 1500 HIF-1α-NanoLuc HCT116 cells to each well in a 1536-well white wall/solid bottom plate and 5 μL of 2500 HRE-bla ME-180 cells to each well in a 1536-well black wall/clear bottom plate using BioRAPTR flying reagent dispenser (*see* **Note 7**).

11. Place a porous metal lid on top of the assay plate and incubate in a humidified incubator at 37 °C, 5 % $CO_2$, and 20 % $O_2$ for 6 h.

12. Transfer 23 nL of compound solutions from a control plate and a sample compound plate to the corresponding wells in the assay plate using a Pintool (*see* **Notes 8** and **9**).

13. Place the porous metal lid back to the assay plate and incubate in a humidified incubator set at 37 °C, 5 % $CO_2$, and 20 % $O_2$ (normoxia) or 1 % $O_2$ (hypoxia) for 18 h for HIF-1α-NanoLuc HCT116 cells and 17 h for HRE-bla ME-180 cells.

**3.4 β-lactamase Reporter Gene Assay Multiplexed with Viability Assay**

1. Add 12 μL of Solution A to 120 μL of Solution B and vortex (*see* **Note 10**).

2. Add 20 μL of Solution D to 1848 μL of Solution C and vortex (*see* **Note 10**).

3. Combine solutions from above steps and vortex to create CCF4 β-lactamase detection mix (*see* **Note 10**).

4. Add 1 μL of the CCF4 β-lactamase detection mix to each 1536 well using BioRAPTR flying reagent dispenser.

5. Place a nonporous lid on the assay plate and incubate in the dark at room temperature for 2.5 h to allow cells to cleave the FRET substrate (*see* **Note 11**).

6. Read assay plate using an Envision fluorescence plate reader with excitation filters of 405/8 nm and emission filters of 460/25 and 535/20 nm.

7. Calculate the 460 nm to 535 nm fluorescence emission intensity ratio for each well.

8. Add 3 μL of CellTiter-Glo Cell viability assay detection reagent to each well using BioRAPTR flying reagent dispenser.

9. Place a nonporous lid on the assay plate and incubate at room temperature in the dark for 30 min.

10. Record luminescence intensity values of CellTiter-Glo viability assay on a ViewLux plate reader (*see* **Note 12**).

*3.5  NanoLuc Luciferase Reporter Gene Assay and Viability Assay*

1. Thaw Nano-Glo luciferase assay buffer at 4 °C or room temperature (*see* **Note 13**).

2. Mix 1 volume of Nano-Glo luciferase assay substrate with 50 volumes of Nano-Glo Luciferase Assay buffer to prepare Nano-Glo luciferase assay detection reagent (*see* **Note 14**).

3. Add 4 μL of Nano-Glo luciferase assay detection reagent or 4 μL of CellTiter-Glo Cell viability assay detection reagent to each well in the assay plate using BioRAPTR flying reagent dispenser (*see* **Note 15**).

4. Place a nonporous lid on the assay plate and incubate at room temperature in the dark for 30 min (*see* **Note 16**).

5. Record luminescence intensity values of Nano-Glo or CellTiter-Glo viability assay on a ViewLux plate reader.

*3.6  Data Analysis*

1. Normalize raw plate reads of the β-lactamase reporter gene assay, Nano-Glo luciferase assay, and CellTiter-Glo values to hypoxia mimetic $CoCl_2$ control wells (100 % activation for normoxia mode), HIF-1 inhibitor topotecan control wells (100 % inhibition for hypoxia mode), viability control tetraoctylammonium bromide (TOAB) wells (0 % viability), and DMSO-only wells (0 % activation/inhibition and 100 % viability).

2. Fit concentration–response curves of normalized data points to determine $IC_{50}$ and efficacy values using a four-parameter Hill equation in GraphPad Prism software (Figs. 1, 2, 3, and 4).
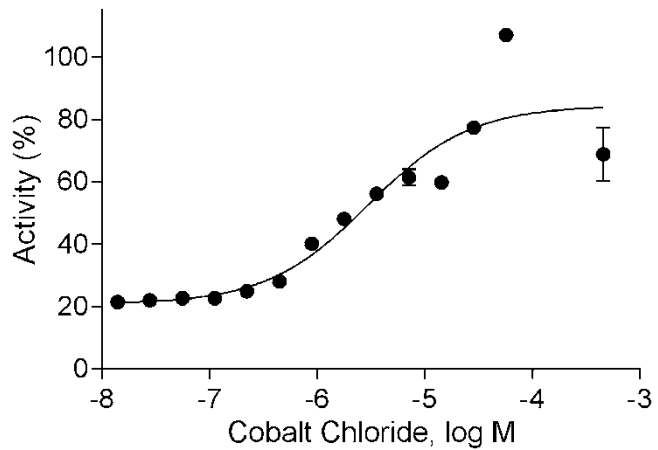
**Fig. 1** Concentration–response curves of cobalt chloride in HRE-bla normoxia mode assay. Each data point is presented as mean ± SD from duplicates
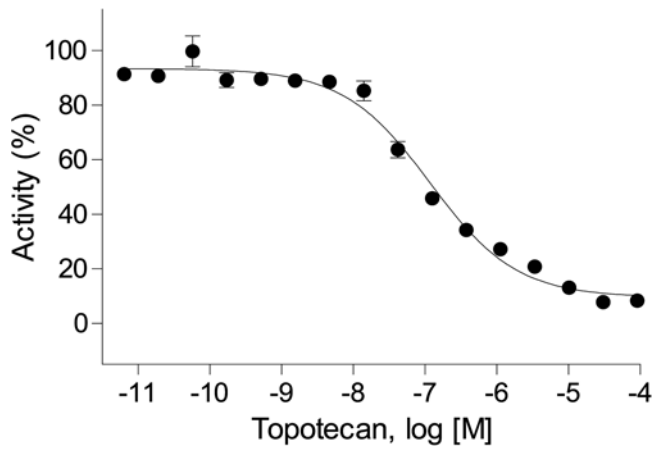


**Fig. 2** Concentration–response curves of topotecan in HRE-bla hypoxia mode assay. Each data point is presented as mean ± SD from duplicates
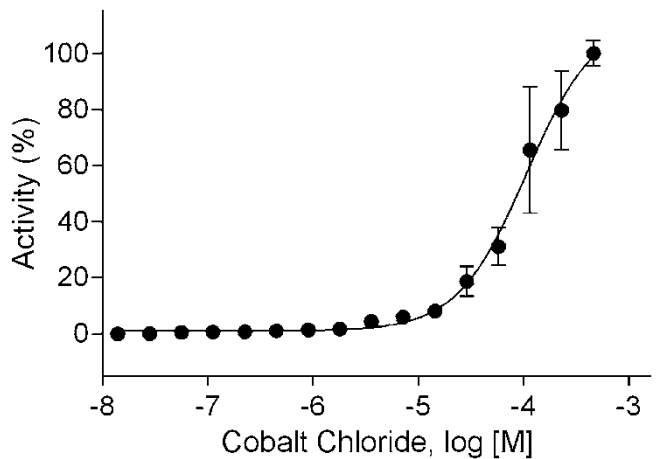


**Fig. 3** Concentration–response curves of cobalt chloride in HIF-1α-NanoLuc normoxia mode assay. Each data point is presented as mean ± SD from duplicates

**Fig. 4** Concentration–response curves of topotecan and YC-1 in HIF-1α-NanoLuc hypoxia mode assay. Each data point is presented as mean ± SD from duplicates

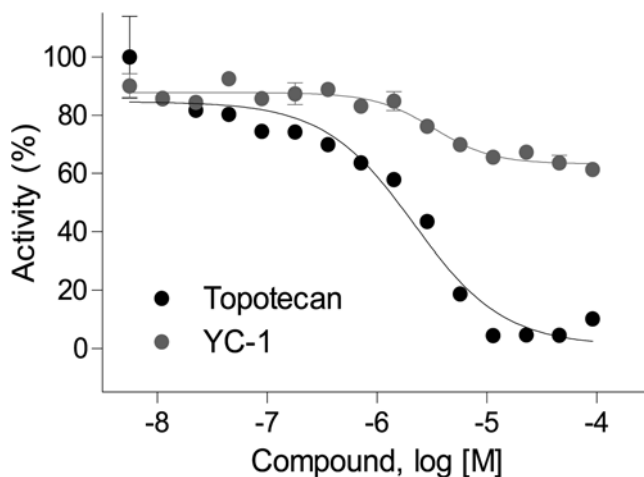## 4   Notes

1. When thawing cells in water bath do not overthaw cells. Leaving a small ice pellet will ensure not overthawing and increase cell viability.

2. To ensure accurate cell counts for plating, it is a good idea to count three times and take the average cell density.

3. Gently shake cell flask and place on a flat surface inside the incubator to ensure even cell growth.

4. Do not allow cells to reach 100 % confluence which might affect assay performance.

5. When washing cells with DPBS or adding Trypsin–EDTA, make sure to cover all areas of the flask.

6. To ensure increased cell viability, do not leave cells in trypsin for more than 4 min. It is also helpful to gently tap the flask to help cells detach.

7. Be sure to wash the BioRaptr dispensing tips with 70 % ethanol and distilled water before use to avoid contamination.

8. Be sure to wash the PinTool workstation tips using and dimethyl sulfoxide (DMSO) and methanol before use to avoid contamination.

9. Allow assay plates cool down to room temperature prior to addition of compound solutions or detection reagents; therefore, there is less evaporation which might affect assay performance.

10. Gently mix to avoid bubbling.

11. Be sure the nonporous metal lid is on the assay plate properly to prevent evaporation and light interference.

12. Adjust exposure time and make sure luminescence intensity values are within the linear dynamic range of the plate reader used.

13. Use fresh Nano-Glo Assay Reagent in each experiment for best assay performance. At room temperature, the reagent will be 10% to 50% less active in approximately 8 h. The reagent will slightly lose activity (<10%) over 2 days of storage at 4 °C.

14. To recover most volume of Nano-Glo luciferase assay substrate, spin down the vial of Nano-Glo luciferase assay substrate before opening it.

15. It is feasible to use CellTiter-Blue viability assay reagent for multiplexing with the NanoLuc assay.

16. Wait for at least 3 min to collect NanoLuc luminescence readings. The half-life of Nano-Glo luminescence is approximately 120 h. We found the 1536-well HIF-1α-NanoLuc assay reached the maximum luminescence intensity between 30 and 45 min after adding detection reagent.

## References

1. Xia M, Bi K, Huang R, Cho MH, Sakamuru S, Miller SC, Li H, Sun Y, Printen J, Austin CP, Inglese J (2009) Identification of small molecule compounds that inhibit the HIF-1 signaling pathway. Mol Cancer 8:117. doi:10.1186/1476-4598-8-117

2. Jiang BH, Semenza GL, Bauer C, Marti HH (1996) Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O2 tension. Am J Phys 271(4 Pt 1):C1172–C1180

3. Wang GL, Semenza GL (1995) Purification and characterization of hypoxia-inducible factor 1. J Biol Chem 270(3):1230–1237

4. Huang LE, Gu J, Schau M, Bunn HF (1998) Regulation of hypoxia-inducible factor 1alpha is mediated by an O2-dependent degradation domain via the ubiquitin-proteasome pathway. Proc Natl Acad Sci U S A 95(14):7987–7992

5. Salceda S, Caro J (1997) Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. J Biol Chem 272(36):22642–22647

6. van Uden P, Kenneth NS, Rocha S (2008) Regulation of hypoxia-inducible factor-1alpha by NF-kappaB. Biochem J 412(3):477–484. doi:10.1042/bj20080476

7. Lando D, Peet DJ, Whelan DA, Gorman JJ, Whitelaw ML (2002) Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. Science 295(5556):858–861. doi:10.1126/science.1068592

8. Huang W, Huang R, Attene-Ramos MS, Sakamuru S, Englund EE, Inglese J, Austin CP, Xia M (2011) Synthesis and evaluation of quinazolin-4-ones as hypoxia-inducible factor-1alpha inhibitors. Bioorg Med Chem Lett 21(18):5239–5243. doi:10.1016/j.bmcl.2011.07.043

9. Hu Y, Liu J, Huang H (2013) Recent agents targeting HIF-1alpha for cancer therapy. J Cell Biochem 114(3):498–509. doi:10.1002/jcb.24390

10. Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt KL, Inglese J, Tice RR, Austin CP (2009) Identification of chemical compounds that induce HIF-1alpha activity. Toxicol Sci 112(1):153–163. doi:10.1093/toxsci/kfp123

11. Hsu CW, Huang R, Khuc T, Shou D, Bullock J, Grooby S, Griffin S, Zou C, Little A, Astley H, Xia M (2015) Identification of approved and investigational drugs that inhibit hypoxia-inducible factor-1 signaling. Oncotarget 7(7):8172–8183

# Quantitative High-Throughput Luciferase Screening in Identifying CAR Modulators

**Caitlin Lynch, Jinghua Zhao, Hongbing Wang, and Menghang Xia**

## Abstract

The constitutive androstane receptor (CAR, NR1I3) is responsible for the transcription of multiple drug metabolizing enzymes and transporters. There are two possible methods of activation for CAR, direct ligand binding and a ligand-independent method, which makes this a unique nuclear receptor. Both of these mechanisms require translocation of CAR from the cytoplasm into the nucleus. Interestingly, CAR is constitutively active in immortalized cell lines due to the basal nuclear location of this receptor. This creates an important challenge in most in vitro assay models because immortalized cells cannot be used without inhibiting the high basal activity. In this book chapter, we go into detail of how to perform quantitative high-throughput screens to identify hCAR1 modulators through the employment of a double stable cell line. Using this line, we are able to identify activators, as well as deactivators, of the challenging nuclear receptor, CAR.

**Key words** Constitutive Androstane Receptor (CAR), Cytochrome P450 2B6 (CYP2B6), Luciferase, Quantitative high-throughput screening (qHTS)

## 1 Introduction

The constitutive androstane receptor (CAR, NR1I3) is a well-known transcription factor found mainly in the liver and intestine which modulates the expression of drug metabolism genes, such as oxidation and conjugation enzymes, while also regulating certain transporters [1, 2]. Through this gene regulation, CAR has the potential to play a major role in drug–drug interactions. Drugs which activate CAR, for example, can induce cytochrome P450 (CYP) 2B6 and CYP3A4 which will increase the clearance of any drug metabolized by these two enzymes. Alternately, when CAR is deactivated and CYP2B6 and CYP3A4 protein expression is inhibited, a previously prescribed drug can become toxic to the body due to the decreased metabolism. Recently, it has also been implicated that CAR plays a pivotal role in energy metabolism [3, 4]. Therefore, it is not only important to identify compounds causing

potential drug–drug interactions, but also drugs which can have a therapeutic effect through hCAR activation.

One of the unique features of CAR is the differing initial localization found when comparing human primary hepatocytes (HPH) and immortalized cell lines. This nuclear receptor has basal localization in the cytoplasm until activation occurs. Once stimulated, CAR will translocate into the nucleus and begin the activation process. However, in immortalized cell lines, CAR constantly resides inside the nucleus without stimulation. When performing in vitro assays, this nuclear localization results in high constitutive activity and difficulty to increase the signal further even after xenobiotic stimulation [5–7]. Because of the high cost and low availability of HPH, it is not an ideal option to use them in high-throughput screenings (HTS). Using immortalized cells is an excellent choice when working with HTS because of their easy culturing and relatively cheap characteristics.

Due to the high basal activity of hCAR in immortalized cells as stated previously, a low concentration of a known deactivator can be co-treated with a test compound to decrease the constitutive activity and allow for the prediction of an activator. The most well-known deactivators of CAR are 1-(2-chlorophenylmethylpropyl)-3-isoquinoline-carboxamide (PK11195), meclizine, and clotrimazole [8–10]. However, along with deactivating CAR, meclizine has inconsistently been reported as a hCAR inverse agonist, an mCAR agonist, and also to have no effect on hCAR [8, 11]. There is also conflicting clotrimazole data. This drug's deactivation effects came into question when confirmation could not be completed, along with varying results in different cell lines [12, 13]. Deactivation of hCAR by PK11195 has also been shown to be an activator of PXR in HPH. This activation of PXR overrides its deactivation of CAR in HPH, proving PK11195 to be a viable option of hCAR deactivation only in immortalized cell lines [9].

The luciferase reporter gene assay is a common technique utilized to determine modulation of receptors [14, 15]. A basic way to perform this assay is to transfect an expression plasmid containing the nuclear receptor along with a vector containing the promoter region of its target gene with a downstream luciferase reporter into a cell line of choice. However, when performing a HTS using 1536-well plates, the amount of cells per well is an important aspect of each experiment. Because transfection rates are not 100%, it is difficult to use transient transfection in a HTS. Therefore, stably transfecting the cells is an important first step in this protocol.

There are significant issues to overcome when identifying hCAR modulators using in vitro methods. This newly adopted quantitative HTS approach overcomes many of the difficulties residing throughout the luciferase assay [16]. However, there are still limitations for this HTS. For instance, this assay is more likely to identify direct activators than indirect activators. Further studies should be completed to confirm actual hCAR modulation.

Here, we outline step-by-step instructions to generate this HepG2-CYP2B6-hCAR stable cell line alongside the quantitative HTS luciferase method.

## 2  Materials

### 2.1  Equipment

1. GLOMAX® 20/20 Single-tube Luminometer (Promega) for the luciferase assay.
2. $CO_2$ incubator MCO-17AIC for all cell work.
3. 1536-well plates for the high-throughput screen.
4. Multidrop Combi (Thermo Fisher Scientific Inc) for plating cells in the large screen.
5. Pintool station (Kalypsys) to transfer compounds into the 1536-well assay plates.
6. Bioraptr Flying Reagent Dispenser workstation (Beckman Coulter) to add activator/deactivator into the assay plates.
7. ViewLux plate reader (PerkinElmer) to read the luminescent and fluorescent signals created by the ONE-Glo and CellTiter Fluor reagents (Promega).

### 2.2  Reagents and Solutions

1. Collagen solution: Using distilled water, prepare 130 μg/mL MCDI    (N-Cyclohexyl-N-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate) solution. Make a 100 μg/mL solution of collagen type I from rat tail in MCDI solution.
2. Seeding medium for HepG2 cells: 500 mL DMEM, 50 mL fetal bovine serum (FBS), 5 mL penicillin–streptomycin.
3. Transfection medium for HepG2 cells: 500 mL DMEM, 50 mL FBS.
4. Culture medium for the HepG2-CYP2B6-hCAR cells: DMEM, 10% FBS, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.6–1 mg/mL geneticin, 10 μg/mL blasticidin.
5. CITCO (6-(4-Chlorophenyl)imidazo[2,1-b][1,3]thiazole-5-carbaldehyde O-(3,4-dichlorobenzyl)oxime.
6. PK11195(1-(2-chlorophenyl)-N-methyl-N-(1-methylpropyl)-3-isoquinolinecarboxamide.
7. Opti-MEM® I reduced-serum medium.
8. pEF6/V5-hCAR expression plasmid.
9. pGL4.17[luc2/Neo]-CYP2B6-2.2 kb construct containing both the PBREM and XREM.
10. ONE-Glo luciferase reagent (Promega) to create a luminescent signal during the assay.
11. CellTiter-Fluor reagent (Promega) used for determining the viability of cells.

## 3 Methods

***3.1 Generation of hCAR-CYP2B6-HepG2 Cell Line***

1. Pre-coated collagen dishes are used to plate the HepG2 cells so that a clear monolayer occurs. Table 1 describes the volume of cold collagen solution which should be added to each dish and well or collagen-coated flasks and dishes may be purchased. Once the solution is in the dish/flask/plate, incubate at 37 °C with 5% $CO_2$ overnight. The next day, take the items out of the incubator and remove the liquid. Add PBS to each well and seal each lid with Parafilm. Once wrapped, store the items in 4 °C until ready for using (*see* **Note 1**).

2. Warm the seeding medium to 37 °C in a water bath.

3. Plate cells in a pre-coated 6-well plate at a seeding density of $5 \times 10^5$ cells/well using the pre-warmed medium. Incubate cells for 5 h, or until attached, at 37 °C and 5% $CO_2$.

4. Check confluency of the cells; they should be about 70–90% confluent to proceed to the next step.

5. Use Table 2 to combine the proper amounts of each reagent to the wells of plated HepG2 cells. Let the mix sit for 5–25 min at room temperature before adding to the well.

6. Replace the seeding medium in the well with 1.75 mL of transfection medium.

7. Add 250 μL of the transfection mix into each well and incubate at 37 °C and 5% $CO_2$ overnight.

8. Trypsinize two of the wells and reseed cells into about ten 10 cm² dishes using the culture medium for HepG2-CYP2B6-hCAR cells for selection. The confluency should be around 10–20% to allow optimal space for colony growth.

9. Change the medium, including blasticidin and geneticin, on every dish every 3–4 days. Allow cells to grow for about 2 weeks until the colonies can be seen with the naked eye.

10. Identify the single colonies on each dish (about 5–10 colonies) and isolate them using a cloning cylinder. Trypsinize each

**Table 1**

**Preparation of collagen-coated plates**

|  | Collagen solution (mL) | 1× PBS (mL) |
|---|---|---|
| 10 cm² Dish | 10 | 10 |
| 6-well plate | 2 | 2 |
| 48-well plate | 0.25 | 0.25 |

**Table 2**

**Transfection Mix**

|  | 6-well plate (µL/well) |
| --- | --- |
| Opti-MEM® | 227 |
| CYP2B6 (100 ng/µL) | 12 |
| hCAR1 (100 ng/µL) | 6 |
| Lipofectamine® 2000 | 5 |
|  | 250 µL/well |

    colony by adding 100 µL of trypsin inside the cylinder and allow it to sit for about 2 min.

11. Acquire two pre-coated 48-well plates (*see* **Note 2**) and add 500 µL of culture medium into each well of plate 1.

12. Gently pipette the trypsin inside each cylinder and pipette all liquid from inside one cylinder into one of the wells with medium on plate 1. Mix the well by gently pipetting up and down and transfer 250 µL of the contents into the corresponding empty well in plate 2.

13. Repeat **steps 10–12** until each colony has been plated in corresponding wells for both plates.

14. Treat every well on plate 1 with 1 µM CITCO and incubate at $37\,°C/5\%\,CO_2$ for 24 h. Incubate plate 2 with culture medium only until the luciferase assay on plate 1 is complete.

15. After plate 1 is treated for 24 h, rinse each well with PBS and add 75 µL lysis buffer to create the cell lysate (*see* **Note 3**).

16. Combine 25 µL of firefly from a Dual-Luciferase® Assay System Kit (Promega) and 15 µL of the lysate in a 1.5 mL tube and read the luminescence value using a GLOMAX luminometer.

17. Once the data is collected, keep any colonies in plate 2 which have a value above 1000 so that further tests can be completed.

18. Once plate 2 is confluent, transfer the cells which have a luminescence value above 1000 to a 6-well plate and allow the colony to grow until confluent using culture medium.

19. For each colony, using culture medium, plate 6 wells of a pre-coated 48-well plate at a concentration of $5 \times 10^4$ cells/well, and keep the rest of the cells from that colony in one well of a 6-well plate to allow the colony to grow. Incubate these plates at $37\,°C$ and $5\%\,CO_2$.

20. Once the cells have attached to the bottom of the plate, treat 3 of the wells with 0.1% DMSO and the other three wells with

1 μM CITCO in culture medium. Allow the cells to incubate at 37 °C/5 % CO$_2$ for 24 h.

21. Repeat **steps 19** and **20** for all colonies to be analyzed.

22. Calculate the average of the three DMSO treated wells and divide each CITCO value by this average. Complete this calculation for each colony separately. Then, determine the average of these three newly calculated numbers to identify the fold induction for each colony. The colony with the highest fold induction and smallest standard deviation was chosen to use.

23. The selected colony should be kept growing in culture medium in a collagen-coated flask. This newly generated double stable cell line should be used to perform the quantitative high-throughput screen in the next assay.

### 3.2 Assay Optimization in a 1536-Well Plate Format

1. Using culture medium, plate the double stable HepG2-CYP2B6-CAR cells into a 1536-well plate at a density of 2500 cells/well in 4 μL using the Multidrop Combi.

2. Allow plates to incubate at 37 °C/5 % CO$_2$ for 4–5 h or until the cells have attached to the bottom.

3. Create a positive control plate. For the agonist mode, the first column should be 16 duplicate concentrations of a dose response curve of CITCO. The starting concentration is 20 mM (final assay concentration of 77 μM) with a 1:2 serial dilution. The second column should have the top 16 wells be CITCO at a single dose of 13 mM (final assay concentration of 50 μM), while the bottom 16 wells are 10.4 mM (final assay concentration of 40 μM) CITCO. The third column should have the top 16 wells contain DMSO and the bottom 16 wells comprise of 20 mM (final assay concentration of 77 μM) tetra-octylammonium bromide for cytotoxicity control. The fourth column should be comprised solely of DMSO. For the antagonist mode, the first column should be 16 duplicate concentrations of a dose response curve of PK11195. The starting concentration is 20 mM (final assay concentration of 77 μM) with a 1:2 serial dilution. The second column should have the top 16 wells be PK11195 at a single dose of 7.8 mM (final assay concentration of 30 μM), while the bottom 16 wells are 5.2 mM (final assay concentration of 20 μM) PK11195. The third column should have the top 16 wells contain DMSO and the bottom 16 wells comprise of 20 mM (final assay concentration of 77 μM) tetraoctylammonium bromide. The fourth column should be comprised solely of DMSO.

4. Create a plate with only DMSO in columns 5–48.

5. Once the cells have attached, use the Pintool station to transfer 23 nL of the positive control or DMSO into the assay plates.

6. Use PK11195 to reduce the basal level of CAR expression and test several concentrations of PK11195 (final 0, 0.5, 0.75, 1, and 1.5 μM) for agonist mode in every well. For the antagonist mode, test several CITCO concentrations (final 0, 25, 50, and 100 nM) in every well.

7. Incubate the assay plates at 37 °C/5 % $CO_2$ for 23 h.

8. Add 1 μL of Cell-Titer Fluor to each well and incubate at 37 °C/5 % $CO_2$ for 1 h.

9. Read the fluorescence of each plate by using a ViewLux plate reader.

10. Immediately following the fluorescence reading, add 4 μL of ONE-Glo reagent and allow the plates to incubate at room temperature for 30 min.

11. Read the luminescence of each plate by using a ViewLux plate reader.

12. Use the following calculations to acquire the CV, S/B ratio, and z-factors for each plate.

CV = (Standard deviation of replicates/mean of replicates) × 100 %

z-factor = 1 − 3 × [(st. dev. Positive control + st. dev. Negative control)/(mean positive control − mean negative control)].

S/B = signal/background.

To determine the optimal concentration to use, look at every factor (Table 3) as well as the graphical representation of the concentration curves, as shown in Fig. 1.

**3.3 Quantitative High-throughput Screen**

1. Create new positive control plates using the same methods as in the previous section.

2. Acquire compound plates, where the test compounds are in columns 5–48. Each compound has fifteen 2.236-fold serial dilutions so that half maximal activation ($EC_{50}$s) and half-maximal inhibition ($IC_{50}$s) can be obtained.

3. Add activator or deactivator to each plate as optimized in Subheading 3.2 (*see* **Note 4**).

4. Perform qHTS as previously determined to be the optimal assay condition in Subheading 3.2.

5. Normalize the agonist mode plates to 57 μM CITCO equaling 100 %, and 38 μM PK11195 equaling 100 % for the antagonist mode plates. DMSO should also be set to equal 0 % activity. The $EC_{50}$s, $IC_{50}$s, and efficacy (maximal response value) for each compound will be calculated by in house data pipeline (*see* Chapter 12 in this book).

**Table 3**

**Optimization Factors**

| | Agonist Mode | | | | | Antagonist Mode | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0 µM | 0.5 µM | 0.75 µM | 1 µM | 1.5 µM | 0 nM | 25 nM | 50 nM | 100 nM |
| S/B | 2.14 | 3.81 | 4.30 | 5.10 | 5.87 | 7.18 | 8.00 | 7.68 | 7.92 |
| CV (%) | 5.04 | 5.53 | 10.00 | 7.43 | 10.06 | 6.42 | 6.13 | 2.96 | 6.29 |
| Z-factor | 0.53 | 0.60 | 0.64 | 0.73 | 0.54 | 0.78 | 0.83 | 0.93 | 0.90 |
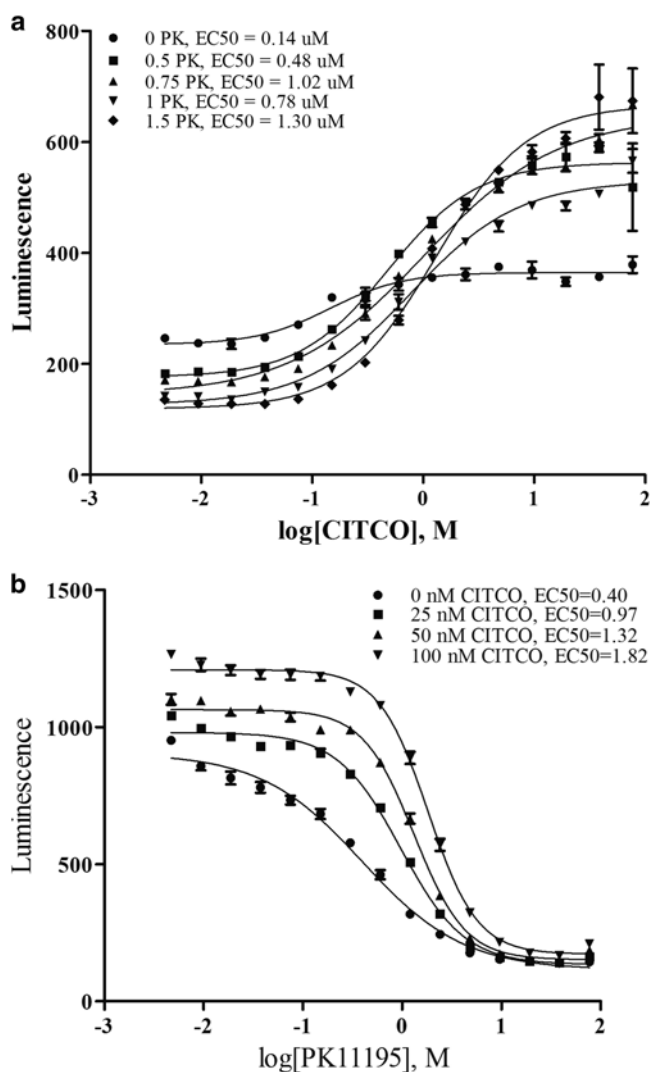


**Fig. 1** Concentration curves of hCAR1 optimization. A double stable HepG2. CYP2B6.CAR cell line was used to perform a luciferase assay in agonist and antagonist mode. Varying concentrations of PK11195 and CITCO were used to identify the $EC_{50}$ of CITCO and PK11195 respectively for the agonist mode (**a**) and antagonist mode (**b**)

## 4   Notes

1. Once coated, the plates can be kept in 4 °C up to 2 months.

2. Use plate 1 to perform the luciferase assay to determine if the cells have both hCAR1 and CYP2B6 transfected properly into them and use plate 2 to keep colony growing.

3. The experiment can either proceed immediately to the next step or the entire plate with lysate can be stored in a –20 °C freezer until the assay is ready to be completed.

4. For example, in the agonist mode, add 1 µL of 4.5 µM PK11195, diluted in culture medium, to make a final concentration of 0.75 µM PK11195 inside each well using the Bioraptr. For the antagonist mode, add 1 µL of 300 nM CITCO, diluted in culture medium, to make a final concentration of 50 nM CITCO inside each well also using the Bioraptr.

## References

1. Honkakoski P, Zelko I, Sueyoshi T, Negishi M (1998) The nuclear orphan receptor CAR-retinoid X receptor heterodimer activates the phenobarbital-responsive enhancer module of the CYP2B gene. Mol Cell Biol 18(10): 5652–5658

2. Wang H, LeCluyse E (2003) Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. Clin Pharmacokinet 42(15):1331–1357. doi:10.2165/00003088-200342150-00003

3. Lynch C, Pan Y, Li L, Heyward S, Moeller T, Swaan PW, Wang H (2014) Activation of the constitutive androstane receptor inhibits gluconeogenesis without affecting lipogenesis or fatty acid synthesis in human hepatocytes. Toxicol Appl Pharmacol 279(1):33–42, doi: http://dx.doi.org/10.1016/j.taap.2014.05.009

4. Dong B, Saha PK, Huang W, Chen W, Abu-Elheiga LA, Wakil SJ, Stevens RD, Ilkayeva O, Newgard CB, Chan L, Moore DD (2009) Activation of nuclear receptor CAR ameliorates diabetes and fatty liver disease. 106. doi:10.1073/pnas.0909731106

5. Chen T, Tompkins LM, Li L, Li H, Kim G, Zheng Y, Wang H (2010) A single amino acid controls the functional switch of human constitutive androstane receptor (CAR) 1 to the xenobiotic-sensitive splicing variant CAR3. J Pharmacol Exp Ther 332(1):106–115

6. Li H, Chen T, Cottrell J, Wang H (2009) Nuclear translocation of adenoviral-enhanced yellow fluorescent protein-tagged-Human Constitutive Androstane Receptor (hCAR): a novel tool for screening hCAR activators in human primary hepatocytes. Drug Metab Dispos 37(5): 1098–1106. doi:10.1124/dmd.108.026005

7. Zelko I, Sueyoshi T, Kawamoto T, Moore R, Negishi M (2001) The peptide near the C terminus regulates receptor CAR nuclear translocation induced by xenochemicals in mouse liver. Mol Cell Biol 21(8):2838–2846. doi:10.1128/mcb.21.8.2838-2846.2001

8. Huang W, Zhang J, Wei P, Schrader WT, Moore DD (2004) Meclizine is an agonist ligand for mouse Constitutive Androstane Receptor (CAR) and an inverse agonist for human CAR. Mol Endocrinol 18(10):2402–2408. doi:10.1210/me.2004-0046

9. Li L, Chen T, Stanton JD, Sueyoshi T, Negishi M, Wang H (2008) The peripheral benzodiazepine receptor ligand 1-(2-Chlorophenyl-methylpropyl)-3-isoquinoline-carboxamide is a novel antagonist of human constitutive androstane receptor. Mol Pharmacol 74(2): 443–453. doi:10.1124/mol.108.046656

10. Moore LB, Parks DJ, Jones SA, Bledsoe RK, Consler TG, Stimmel JB, Goodwin B, Liddle C, Blanchard SG, Willson TM, Collins JL, Kliewer SA (2000) Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. J Biol Chem 275(20):15122–15127. doi:10.1074/jbc.M001215200

11. Lau AJ, Yang G, Rajaraman G, Baucom CC, Chang TKH (2011) Differential effect of

meclizine on the activity of human pregnane X receptor and constitutive androstane receptor. J Pharmacol Exp Ther 336(3):816–826. doi:10.1124/jpet.110.175927

12. Mäkinen J, Frank C, Jyrkkärinne J, Gynther J, Carlberg C, Honkakoski P (2002) Modulation of mouse and human phenobarbital-responsive enhancer module by nuclear receptors. Mol Pharmacol 62(2):366–378. doi:10.1124/mol.62.2.366

13. Toell A, Kröncke K-D, Kleinert H, Carlberg C (2002) Orphan nuclear receptor binding site in the human inducible nitric oxide synthase promoter mediates responsiveness to steroid and xenobiotic ligands. J Cell Biochem 85(1):72–82. doi:10.1002/jcb.10104

14. Freitas J, Miller N, Mengeling BJ, Xia M, Huang R, Houck K, Rietjens IMCM, Furlow JD, Murk AJ (2014) Identification of thyroid hormone receptor active compounds using a quantitative high-throughput screening platform. Curr Chem Genomics Transl Med 8:36–46. doi:10.2174/2213988501408010036

15. Shukla SJ, Sakamuru S, Huang R, Moeller TA, Shinn P, VanLeer D, Auld DS, Austin CP, Xia M (2011) Identification of clinically used drugs that activate pregnane X receptors. Drug Metab Dispos 39(1):151–159. doi:10.1124/dmd.110.035105

16. Lynch C, Zhao J, Huang R, Xiao J, Li L, Heyward S, Xia M, Wang H (2015) Quantitative High-Throughput Identification of Drugs as Modulators of Human Constitutive Androstane Receptor. Sci Rep 5. doi:10.1038/srep10405. http://www.nature.com/srep/2015/150520/srep10405/abs/srep10405.html#supplementary-information

# Transactivation and Coactivator Recruitment Assays for Measuring Farnesoid X Receptor Activity

## Chia-Wen (Amy) Hsu, Jinghua Zhao, and Menghang Xia

## Abstract

The farnesoid X receptor (FXR) is a nuclear receptor responsible for homeostasis of bile acids, lipids, and glucose. Compounds that alter endogenous FXR signaling can be used as therapeutic candidates or identified as potentially hazardous compounds depending on exposure doses and health states. Therefore, there is an increasing need for high-throughput screening assays of FXR activity to profile large numbers of environmental chemicals and drugs. This chapter describes a workflow of FXR modulator identification and characterization. To identify compounds that modulate FXR transactivation at the cellular level, we first screen compounds from the Tox21 10 K compound library in an FXR-driven beta-lactamase reporter gene assay multiplexed with a cell viability assay in the same well of the 1536-well plates. The selected compounds are then tested biochemically for their ability to modulate FXR-coactivator binding interactions using a time-resolved fluorescence resonance energy transfer (TR-FRET) coactivator assay. The assay results from the workflow can be used to prioritize compounds for more extensive investigations.

**Key words** Farnesoid X receptor, Nuclear receptor, Drugs, Environmental chemicals, High-throughput screening, HTS, Reporter gene, Beta-lactamase, Coactivator recruitment, Fluorescence resonance energy transfer, FRET

## 1 Introduction

Farnesoid X receptor (FXR) or bile acid receptor (BAR) is a nuclear receptor modulated by primary and secondary bile acids, including chenodeoxycholic acid (CDCA), dexoycholic acid (DCA), and lithocholic acid (LCA)[1]. FXR is mainly expressed in the liver and intestine to maintain metabolic balance of bile acids, glucose, and lipids. Upon activation, FXR translocates into the nucleus and forms a heterodimer with retinoid X receptor (RXR). The FXR-RXR heterodimer is able to bind FXR-responsive elements (FXREs) and regulate expression of FXRE-containing genes. The key FXR target genes include the bile salt excretion pump (BSEP/ABCB11) and small heterodimer partner (SHP) genes. BSEP promotes ATP-dependent and bile acid-dependent transport of bile acids

across the canalicular membrane of hepatocytes. SHP suppresses expression of a rate-limiting bile acid synthesis enzyme cholesterol 7-alpha-monooxygenase (CYP7A1) and apical sodium dependent bile acid transporter (ASBT), resulting in reduced levels of bile acids in hepatocytes and elevated flow of bile acids to the colon. Abnormal FXR transcriptional activity by overexpression, genetic knockouts, and overexposure to FXR agonists or FXR antagonists has been linked to liver injury and liver cancer. Several synthetic FXR agonists are under development or clinical investigation as treatments for numerous metabolic disorders or syndromes, including non-alcoholic fatty liver disease (NAFLD), obesity, primary biliary cirrhosis (PBS), and type 2 diabetes mellitus [2]. Inhibition of FXR activity by knockout, gene silencing, or FXR antagonists can be used to treat obstructive cholestasis, Barrett's epithelia, and certain types of cancer overexpressing FXR [3].

Several in vitro FXR assays have been developed to measure FXR transactivation activity in various cell backgrounds and FXR binding interactions using isolated FXR protein. Transient transfection of human hepatocytes with an FXRE-driven luciferase reporter gene construct is a commonly used method to identify FXR-active compounds [4]. Other available FXR transactivation assays include a beta-galactosidase reporter gene-coupled chemiluminescent assay [5] and a beta-lactamase reporter gene-coupled fluorescence resonance energy transfer (FRET) assay [6]. To identify compounds that can disrupt CDCA-FXR binding interactions, fluorescent analogs of CDCA are used to develop FXR competitive binding assays using fluorescence polarization (FP) [7] and time-resolved FRET (TR-FRET) [8] technologies. In addition, FXR coactivator recruitment assays that use glutathione-S-transferase (GST)-tagged FXR protein and a biotinylated SRC1-derived coactivator peptide are designed in ALPHAScreen, surface plasmon resonance (SPR), time-resolved fluorescence (TRF), and TR-FRET formats [9, 10]. A simplified version of FXR TR-FRET coactivator recruitment assay is developed using a fluorescein-labeled coactivator peptide [6]. Some of the FXR assays are optimized in 96-, 384-, or 1536-well formats and used for compound screening.

Here we describe a high-throughput screening approach to identify FXR-active compounds and to characterize mode of action (MOA) of FXR ligands. Two commercially available FXR assays (Fig. 1), based on human FXR-LBD, are miniaturized in 1536-well plate formats. Compounds are first tested with an FXR-bla assay, a beta-lactamase-based reporter gene cell line (Fig. 1a). To eliminate potential false positives and negatives, autofluorescence and cytotoxicity of each test compound are measured using autofluorescence and viability assays. The FXR-modulating compounds identified from the FXR-bla screens are then characterized as either
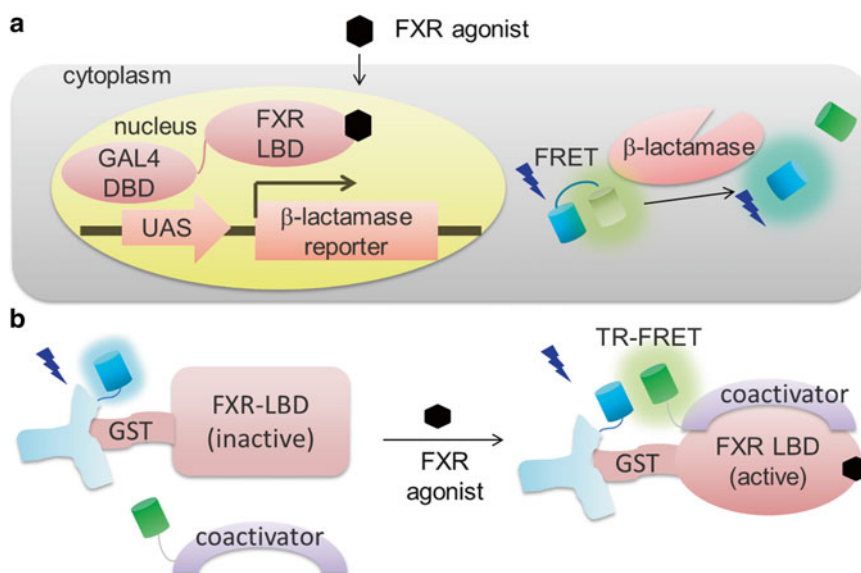
**Fig. 1** Assay principle of FXR-bla transactivation and FXR TR-FRET coactivator assays. (**a**) FXR-bla transactivation assay. *DBD* DNA-binding domain. *FRET* fluorescence resonance energy transfer. *LBD* ligand-binding domain. *UAS* upstream activator sequence. (**b**) FXR TR-FRET coactivator assay. *GST* glutathione-S-transferase. *TR-FRET* time-resolved FRET

FXR agonists or FXR antagonists in an FXR coactivator recruitment assay (Fig. 1b). This approach has been successfully employed to profile the Tox21 10 K chemical collection for FXR activity [6].

## 2  Materials

### 2.1  Beta-Lactamase Reporter Gene Assay

1. Control compounds and solvents: dimethyl sulfoxide (DMSO), chenodeoxycholic acid (CDCA), (Z)-guggulsterone, and tetraoctylammonium bromide (TOAB).

2. Cell line: GeneBLAzer FXR-UAS-bla HEK293T stable cell line (FXR-bla) acquired in cryogenic vials.

3. Medium components listed in Table 1 for FXR-bla cell line: Dulbecco's Modified Eagle Medium with GlutaMax (DMEM with GlutaMax), phenol red-free DMEM, dialyzed fetal bovine serum (FBS), charcoal-stripped FBS, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.3), nonessential amino acids (NEAA), sodium pyruvate, penicillin–streptomycin, hygromycin B, and zeocin.

4. 0.05 % Trypsin–EDTA solution.

5. Dulbecco's Phosphate-Buffered Saline (DPBS).

6. Recovery cell culture freezing medium.

**Table 1**
**Cell culture and assay medium for FXR-bla cells**

| Component | FXR-bla Thaw medium | FXR-bla Growth medium | FXR-bla Assay medium |
|---|---|---|---|
| DMEM with GlutaMAX and sodium pyruvate | 90% | 90% | – |
| Phenol red-free DMEM with HEPES | – | – | 90% |
| Dialyzed FBS | 10% | 10% | – |
| Charcoal-stripped FBS | – | – | 2% |
| HEPES (pH 7.3) | 25 mM | 25 mM | – |
| NEAA | 0.1 mM | 0.1 mM | 0.1 mM |
| Sodium pyruvate | – | – | 1 mM |
| Penicillin | 100 U/mL | 100 U/mL | 100 U/mL |
| Streptomycin | 100 µg/mL | 100 µg/mL | 100 µg/mL |
| Hygromycin B | – | 100 µg/mL | – |
| Zeocin | – | 100 µg/mL | – |

7. T75 flasks.

8. Hemacytometer.

9. Cellometer Auto Cell Counter (Nexcelom Bioscience, Lawrence, MA).

10. 1536-well cycloolefin compound plates.

11. 1536-well polystyrene assay plates, black wall/clear bottom, cell culture treated.

12. Assay and compound microplate lids.

13. LiveBLAzer-FRET B/G loading kit (Life Technologies, Carlsbad, CA) containing LiveBLAzer-FRET B/G substrate (CCF4-AM), Solution A, Solution B, and Solution C.

14. CellTiter-Glo luminescent cell viability assay (Promega, Madison, WI).

15. Vortex mixer.

16. $CO_2$ incubator.

17. BioRAPTR flying reagent dispenser (Beckman Coulter, Pasadena, CA).

18. PinTool workstation (Wako Automation, San Diego, CA).

19. Plate centrifuge.

20. Envision plate reader with a bottom-read mirror, an excitation filter of 405/8 nm, and emission filters of 535/25 nm and 450/25 nm (PerkinElmer, Waltham, MA).

21. ViewLux plate reader (PerkinElmer, Waltham, MA).

*2.2 TR-FRET Coactivator Assay*

1. Control compounds and solvents: DMSO, chenodeoxycholic acid (CDCA), and (Z)-guggulsterone.

2. 1536-well compound plates.

3. 1536-well polystyrene assay plates, black wall/ solid bottom.

4. Assay and compound microplate lids.

5. LanthaScreen TR-FRET FXR coactivator assay (Life Technologies, Carlsbad, CA) containing glutathione S-transferase (GST)-tagged FXR ligand binding domain (LBD) protein, fluorescein-labeled SRC2-2 coactivator peptide, Tb-labeled anti-GST antibody (goat), dithiothreitol (DTT), and assay buffer.

6. 0.75% bovine serum albumin (BSA) in Phosphate Buffered Saline (PBS) buffer.

7. Mantis single-channel low volume liquid dispenser (Formulatrix, Bedford, MA).

8. Pintool workstation (Wako Automation, San Diego, CA).

9. Plate centrifuge and microcentrifuge.

10. ThermoMixer C (Eppendorf North America, Hauppauge, NY).

11. Envision plate reader with a LANCE/DEFIA D400/630 dual mirror, an excitation filter of 340/60 nm, and emission filters of 495/10 nm and 520/25 nm (PerkinElmer, Waltham, MA).

# 3  Methods

*3.1  Cell Thawing*

1. Add 9 mL of pre-warmed thaw medium in a 15 mL conical tube.

2. Thaw a cryopreserved vial of FXR-bla cells at 37 °C in a water bath for 1–2 min.

3. Transfer cell suspension to the conical tube and mix the contents by pipetting.

4. Centrifuge cells at $200 \times g$ for 4 min at room temperature.

5. Aspirate supernatant and resuspend the cell pellet in 15 mL of thaw medium.

6. Count cells using a hemacytometer.

7. Transfer $2 \times 10^6$ cells to a T75 flask and place the flask in a humidified incubator set at 37 °C, 5% $CO_2$.

| | |
|---|---|
| ***3.2  Cell Propagation of FXR-bla Cells*** | 1. Aspirate medium from the T75 flask and briefly rinse cells with 10 mL of DPBS. |

1. Aspirate medium from the T75 flask and briefly rinse cells with 10 mL of DPBS.

2. Add 3 mL of 0.05 % Trypsin–EDTA to the T75 flask and swirl the flask for even distribution.

3. Check cell detachment on a light microscope after 2–3 min of trypsin addition. Cells usually detach after 2–5 min exposure to Trypsin–EDTA.

4. Add 4 mL of Growth Medium to stop trypsin action and transfer the content to a 15 mL conical tube.

5. Centrifuge cells at $200 \times g$ for 4 min at room temperature.

6. Aspirate supernatant and resuspend the cell pellet with 15 mL of growth medium.

7. Count cells using a hemacytometer.

8. Transfer the cells to a T75 flask and place the flask in a humidified incubator set at 37 °C, 5 % $CO_2$.

9. Maintain cells under 95 % confluence and passage cells every 2 or 3 days.

***3.3  Storage of Cell Lines***

1. Count cell numbers and viability of resuspended FXR-bla cells on Cellometer.

2. Centrifuge cells at $200 \times g$ for 4 min at room temperature.

3. Remove supernatant with house vacuum and resuspend cell pellet with Recovery cell culture freezing medium to $2 \times 10^6$ cells/mL.

4. Transfer 1 mL of cell suspension in each pre-labeled cryogenic tube and place the tubes in a cell freezing container.

5. Place cell freezing container in a –80 °C freezer overnight.

6. Move the cryogenic tubes to a –150 °C freezer for long-term storage.

***3.4  FXR-bla and Cell Viability Assays***

1. Prepare 5 μL of control compounds in DMSO in column 1–4 of a 1536-well cycloolefin storage plate and 5 μL of test compounds in DMSO in column 5–48 of a second plate.

2. Harvest cells from culture in growth medium and resuspend in assay medium (*see* **Note 1**) at density of $1 \times 10^6$ cells/mL.

3. Dispense 5 μL of FXR-bla cells (*see* **Note 2**) at 1000 cells/μL to a 1536-well black clear bottom assay plate using BioRAPTR FRD.

4. Place the plate in a humidified incubator set at 37 °C, 5 % $CO_2$ for 5 h.

5. Transfer 23 nL of control or test compounds from the compound plates (*see* **Notes 3** and **4**) to the assay plate on a Pintool

workstation. For antagonist mode, an extra 1 μL of CDCA at a final concentration of 50 μM was added to sample wells.

6. Place the plate in a humidified incubator set at 37 °C, 5 % $CO_2$ for 16 h.

7. Add 1 μL of 6× CCF4-AM substrate mixtures (*see* **Note 5**) to each well using BioRAPTR FRD. To prepare 6× CCF4-AM substrate mixtures, 6 μL of Solution A was first mixed with 60 μL of Solution B on a vortex mixer, followed by addition of 934 μL Solution C to the mixture and vortexing.

8. Incubate the plate in the dark at room temperature for 2 h.

9. Excite the plate at 405 nm and collect fluorescence emission intensity values at 460 nm and 530 nm of each well on an Envision plate reader (*see* **Note 6**).

10. Add 4 μL of CellTiter-Glo reagent to each well using BioRAPTR FRD.

11. Incubate the plate in the dark at room temperature for 30 min.

12. Collect luminescence intensity values of each well on a ViewLux plate reader.

13. Percentage of activity of the compounds is calculated by normalizing the raw data to DMSO wells (0 % FXR activity, 100 % cell viability), CDCA wells (100 % FXR activity), and TOAB wells (0 % cell viability).

14. The compound half maximum effective/inhibitory concentration ($EC_{50}$ or $IC_{50}$) and maximum response (efficacy) values were calculated using a four-parameter Hill equation in GraphPad Prism software (Fig. 2a, 2b).
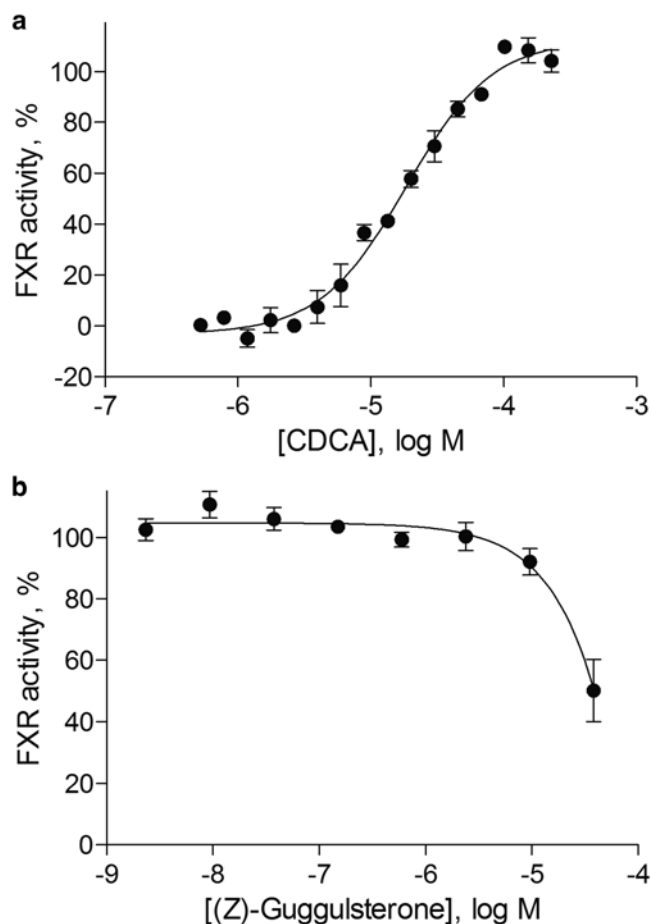
*3.5 FXR Coactivator Assay*

1. Prepare 5 μL of control compounds in DMSO in column 1-4 of a 1536-well cycloolefin storage plate and 5 μL of test compounds in DMSO in column 5–48 of a second plate.

2. Thaw protein components on ice and non-protein components on ThermoMixer C at room temperature, followed by centrifugation of protein components at 4 °C, 15,668 × $g$ for 10 min.

3. Dispense 4 μL of FXR reagent containing 7.5 nM GST-tagged FXR-LBD and 0.15 % BSA (*see* **Note 7**) to a 1536-well black wall/ solid bottom assay plate using Mantis dispenser. For counterscreen experiments, FXR reagent is used as positive controls and FXR-free reagent is used in negative control wells and sample wells.

4. Centrifuge the plate at 1,829 × $g$ at room temperature for 1 min (*see* **Note 8**).

5. Transfer 23 nL of control and test compounds from the compound plates (*see* **Notes 1** and **2**) to the assay plate on a Pintool workstation.
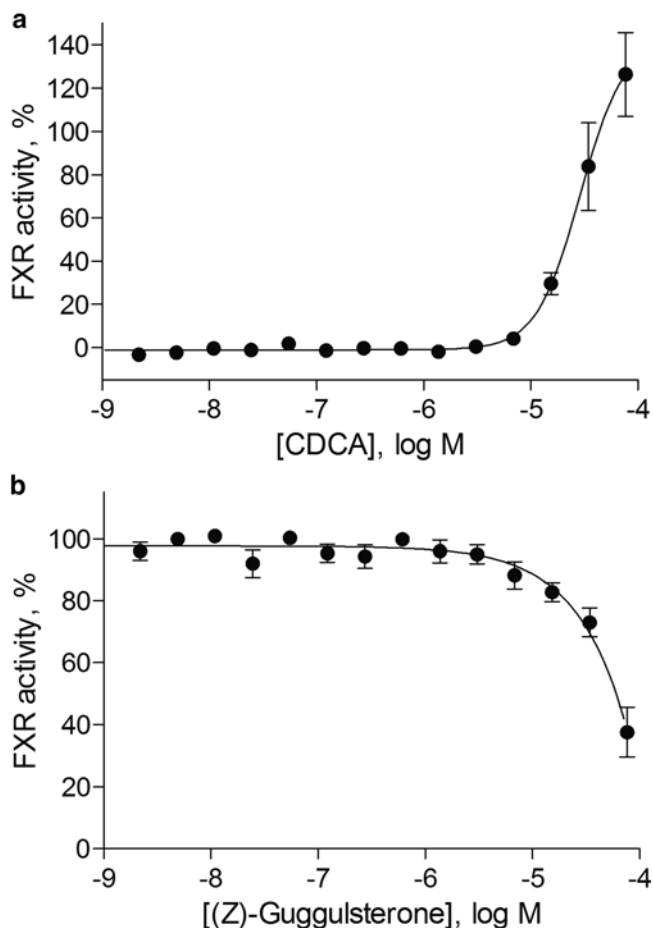
**Fig. 2** Concentration-response curves of CDCA and guggulesterone in FXR-bla assay. (**a**). CDCA, an FXR agonist control with an $EC_{50}$ value of 33.55 $\mu$M. (**b**) Guggulsterone, an FXR antagonist control with an $IC_{50}$ value of 63.12 $\mu$M

6. Add 2 $\mu$L of detection reagent containing 15 nM Tb-labeled anti-GST antibody and 1.5 mM of fluorescein-labeled SRC2-2 coactivator peptide (*see* **Note 3**) to each well using Mantis.

7. Centrifuge the plate at 1,829 $\times$ *g* at room temperature for 1 min (*see* **Note 8**).

8. Place the plate in the dark at room temperature for 30 min.

9. Excite the samples at 340 nm and collect fluorescence intensity values of each well at 495 nm and 520 nm on an Envision plate reader.

10. Percentage of compound activity is calculated by normalizing the raw data to DMSO wells (0 % FXR activity) and CDCA wells (100 % FXR activity).

**Fig. 3** Concentration-response curves of CDCA and guggulesterone in FXR coactivator assay. (**a**) CDCA, an FXR agonist control with an $EC_{50}$ value of 18.47 $\mu$M. (**b**) Guggulsterone, an FXR antagonist control with an $IC_{50}$ value of 38 $\mu$M

11. The compound half maximum effective/inhibitory concentration ($EC_{50}$ or $IC_{50}$) and maximum response (efficacy) values were calculated using a four-parameter Hill equation in GraphPad Prism software (Fig. 3a, b).

# 4    Notes

1. Selection markers, such as hygromycin and zeocin, are important to ensure assay robustness of stable reporter gene cell line by suppressing growth of non-resistant cells. The stable cell lines are generated by transfecting parental cells with plasmids containing a reporter expression construct and an antibiotic-resistant gene. However, the antibiotics are only used in growth medium because they may exert environmental stress during cell recovery from cryopreserving and cell-based assays.

2. A cell strainer can be used to prepare assay cells and minimize well-to-well variation of FXR-bla cells by generating more uniform single-cell suspension and enhancing percentage of viable cells.

3. Compound plates should be centrifuged at $200 \times g$ at room temperature for 15–60 s prior to compound transfer. If a titration series of compound collection is designed in multiple plates in which each compound is plated in the same well location with varying concentration, assay plates should be prepared in the following order: DMSO control, the lowest compound concentration, the middle concentration series from low to high, the highest compound concentration, and the second DMSO control compound plates. Plate pattern recognition algorithms generated by the two DMSO plates are used to normalize raw plate reads, which might be affected by environmental factors (e.g., uneven temperature or air flow in cell incubator) and instrumental bias in liquid handlers or plate readers.

4. The final DMSO concentrations in the beta-lactamase reporter gene and TR-FRET assays range from 0.38 to 0.45 % in 1536-well formats. If the assays are conducted in lower density formats (i.e., 96-wells and cuvettes), DMSO sensitivity should be tested during assay optimization.

5. The detection reagents of the beta-lactamase reporter gene and TR-FRET assays should be stored and used in dark or low light environment to reduce photobleaching. Since the CCF4-AM and fluorescein-labeled peptides are prepared in dry DMSO, the two reagents need to be thawed at room temperature.

6. The fluorescence signals of the beta-lactamase assay are read from the bottom of assay plates. Manually touching the bottom of assay plates might affect fluorescence assay readouts.

7. BSA is crucial to the FXR TR-FRET coactivator assay in 1536-well plates. BSA helps reduce nonspecific binding interactions between the plate surface and other protein-based assay components (i.e., FXR-LBD and anti-GST antibody). If lyophilized powder of BSA is used, the resulting BSA solution should be filtered with a 0.22 micron syringe filter prior to prepare the FXR-LBD/BSA solution.

8. Centrifugation of FXR TR-FRET coactivator assay plates after adding all assay components is critical to minimize air bubbles that might affect protein folding and assay performance. Centrifugation of Tb-labeled anti-GST antibody at 4 °C prior to preparing the detection reagent can also help minimize coefficient of variation (CV) of the FXR TR-FRET coactivator assay.

# References

1. Kuipers F, Bloks VW, Groen AK (2014) Beyond intestinal soap–bile acids in metabolic control. Nat Rev Endocrinol 10(8):488–498. doi:10.1038/nrendo.2014.60

2. Merk D, Steinhilber D, Schubert-Zsilavecz M (2012) Medicinal chemistry of farnesoid X receptor ligands: from agonists and antagonists to modulators. Future Med Chem 4(8):1015–1036. doi:10.4155/fmc.12.47

3. Lamers C, Schubert-Zsilavecz M, Merk D (2014) Medicinal chemistry and pharmacological effects of Farnesoid X Receptor (FXR) antagonists. Curr Top Med Chem 14(19):2188–2205

4. Zheng ZH, Lv GP, Si SY, Dong YS, Zhao BH, Zhang H, He JG (2007) A cell-based high-throughput screening assay for Farnesoid X receptor agonists. Biomed Environ Sci 20(6):465–469

5. Wu X, Glickman JF, Bowen BR, Sills MA (2003) Comparison of assay technologies for a nuclear receptor assay screen reveals differences in the sets of identified functional antagonists. J Biomol Screen 8(4):381–392. doi:10.1177/1087057103256466

6. Hsu CW, Zhao J, Huang R, Hsieh JH, Hamm J, Chang X, Houck K, Xia M (2014) Quantitative high-throughput profiling of environmental chemicals and drugs that modulate farnesoid X receptor. Sci Rep 4:6437. doi:10.1038/srep06437

7. Han KC, Kim JH, Kim KH, Kim EE, Seo JH, Yang EG (2010) Identification of farnesoid X receptor modulators by a fluorescence polarization-based interaction assay. Anal Biochem 398(2):185–190. doi:10.1016/j.ab.2009.11.008

8. Yu DD, Lin W, Chen T, Forman BM (2013) Development of time resolved fluorescence resonance energy transfer-based assay for FXR antagonist discovery. Bioorg Med Chem 21(14):4266–4278. doi:10.1016/j.bmc.2013.04.069

9. Glickman JF, Wu X, Mercuri R, Illy C, Bowen BR, He Y, Sills M (2002) A comparison of ALPHAScreen, TR-FRET, and TRF as assay methods for FXR nuclear receptors. J Biomol Screen 7(1):3–10. doi:10.1089/108705702753520288

10. Fujino T, Sato Y, Une M, Kanayasu-Toyoda T, Yamaguchi T, Shudo K, Inoue K, Nishimaki-Mogami T (2003) In vitro farnesoid X receptor ligand sensor assay using surface plasmon resonance and based on ligand-induced coactivator association. J Steroid Biochem Mol Biol 87(4-5):247–252

# Chapter 6

# Cell-Based Assay for Identifying the Modulators of Antioxidant Response Element Signaling Pathway

## Jinghua Zhao, Sunita J. Shukla**, and Menghang Xia

### Abstract

The antioxidant response element (ARE) signaling pathway plays an important role in the amelioration of cellular oxidative stress. Thus, assays that detect this pathway can be useful for identifying chemicals that induce or inhibit oxidative stress signaling. The focus of this chapter is to describe a cell-based ARE assay in a quantitative high-throughput screening (qHTS) format to test a large collection of compounds that induce nuclear factor erythroid 2-related factor (Nrf2)/ARE signaling. The assay is described through cell handling, assay preparation, and instrument usage.

**Key words** Antioxidant response element (ARE), Reactive oxygen species (ROS), Nuclear factor erythroid 2-related factor (Nrf2), Quantitative high-throughput screening (qHTS), β-lactamase (bla), Fluorescence Resonance Energy Transfer (FRET)

## 1 Introduction

Oxidative stress, an imbalance of reactive oxygen species (ROS), and antioxidant defenses plays roles in chemical-induced toxicity, cancer, and age-related diseases [1–3]. ROS can cause toxic effects through the production of peroxides and free radicals that damage all components of the cell, including proteins, lipids, and DNA, ROS also acts as cellular messengers in redox signaling that can cause disruptions in normal cellular signaling mechanisms [4]. The induction of many cytoprotective enzymes in response to ROS is mediated by antioxidant response elements (ARE), which activate the nuclear factor erythroid 2-related factor (Nrf2).

ARE is a cis-acting enhancer located in the 5′ flanking region of many phase II detoxification genes. Nrf2 is a member of the cap 'n' collar family of transcription factors. Under oxidative stress, Nrf2 is released from keap1 and quickly translocated to the nucleus where it

---

** The views expressed in the chapter do not necessarily represent the views of the Food and Drug Administration or the United States.
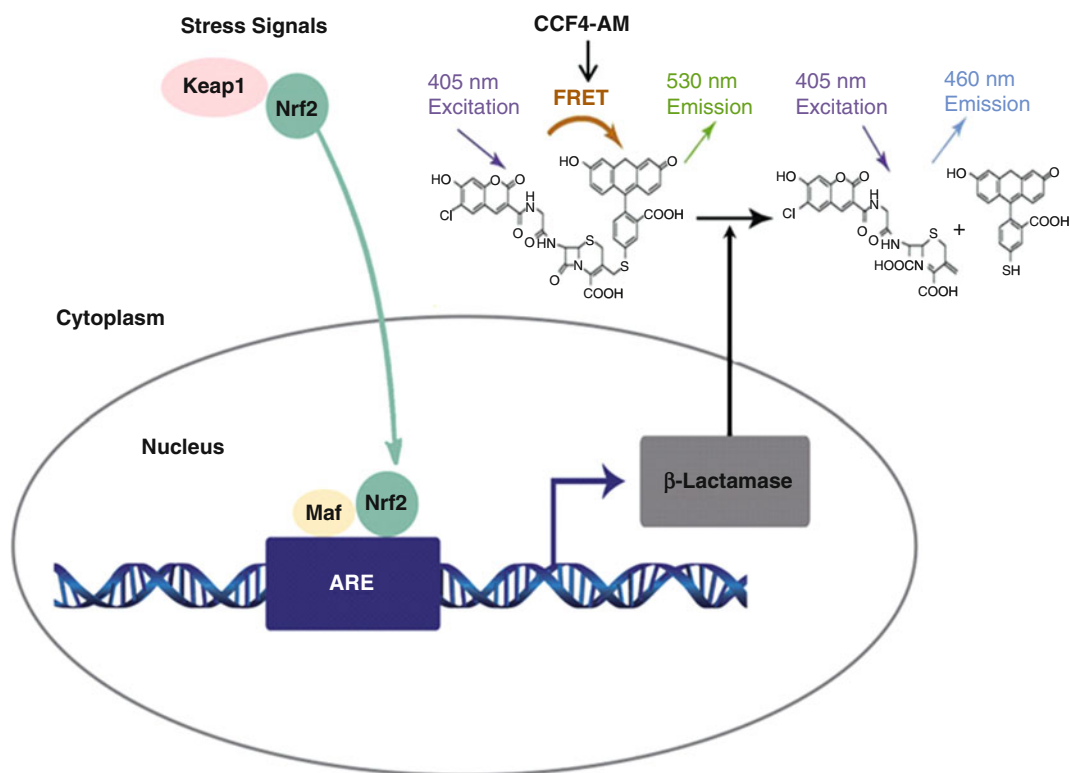
**Fig. 1** Schematics of ARE-*bla* reporter gene assays

binds to the ARE (Fig. 1). The Nrf2/ARE transcriptional pathway plays an important role in the regulation of genes that control the expression of proteins critical to the detoxification and elimination of ROS and electrophiles [5, 6]. Several studies suggest the protective role of Nrf2/ARE activation against environmentally induced oxidative stress and the Nrf2/ARE pathway as a potential therapeutic target [7, 8]. Here we describe a cell-based ARE β-lactamase (bla) reporter gene assay (ARE-bla assay) using Fluorescence Resonance Energy Transfer (FRET) technology to identify the compounds that modulate the ARE signaling pathway. The ARE-bla cell line contains a β-lactamase reporter gene under control of ARE stably integrated into HepG2 cells. The assay uses a FRET-based fluorescent substrate, CCF4-AM, for detection. Once inside the cells, CCF4-AM is hydrolyzed by cytoplasmic esterases to form a polar molecule (CCF4). Upon excitation at 405 nm, the energy is transferred to the fluorescein moiety by FRET resulting in emission at 530 nm. In the presence of β-lactamase expression, the CCF4 substrate is cleaved at β-lactam ring by β-lactamase, which leads to excitation at 405 nm and emission at 460 nm [9] (Fig. 1). Thus, the fluorescence measured in these cells quantitatively corresponds to the activity change of ARE signaling. The ARE-bla assay has been optimized in a quantitative high-throughput screening (qHTS) platform to detect the compounds that activate this pathway. This chapter describes a protocol for this assay in a 1536-well plate format.

## 2  Materials

### 2.1  Cell Line and Cell Culture Condition

1. All of the cell-culture-related medium and components (Table 1) are purchased from Life Technologies.
2. Cell line: A CellSensor® ARE-*bla* HepG2 cell line (Life Technologies) that contains three stably integrated copies of the ARE derived from the reduced form of human nicotinamide adenine dinucleotide phosphate (NADPH) quinone oxidoreductase 1 gene (*NQO1*) driving the expression of a downstream beta-lactamase reporter gene [7]. The cells were maintained in culture medium at 37 °C under a humidified atmosphere and 5 % $CO_2$.
3. Culture media: DMEM medium supplement with 10 % dialyzed fetal bovine serum, 0.1 mM nonessential amino acids, 25 mM HEPES, 100 U/mL penicillin, 100 µg/mL streptomycin, and 5 µg/mL blasticidin.
4. Thaw media: same as culture medium but without blasticidin.
5. Assay media: DMEM medium supplement with 1 % dialyzed fetal bovine serum, 0.1 mM nonessential amino acids, 25 mM HEPES, 100 U/mL penicillin, and 100 µg/mL streptomycin.
6. Freezing Medium: Recovery™ Cell Culture Freezing Medium from Life Technologies.
7. 0.25 % trypsin–EDTA.
8. Dulbecco's phosphate-buffered saline (DPBS) without calcium and magnesium.

### 2.2  Assay Reagents and Chemicals

1. Reagents for fluorescence detection: LiveBLAzer™ FRET B/G Loading Kit (Life Technologies) including Solution A (CCF4-AM), Solution B, and Solution C.

**Table 1**
**Cell culture and assay medium**

| Components | Culture medium | Assay medium | Thaw medium |
|---|---|---|---|
| DMEM | 90 % | 99 % | 90 % |
| Dialyzed FBS | 10 % | 1 % | 10 % |
| NEAA | 0.1 mM | 0.1 mM | 0.1 mM |
| HEPES | 25 mM | 25 mM | 25 mM |
| Penicillin | 100 U/mL | 100 U/mL | 100 U/mL |
| Streptomycin | 100 µg/mL | 100 µg/mL | 100 µg/mL |
| Blasticidin | 5 µg/mL | – | – |

2. Solution D (Life Technologies) containing an anion transport inhibitor that is used in conjunction with CCF4-AM-based assay to prevent active cellular export of the FRET-based substrate after loading. It was used in combination with Solution A, B, and C to create CCF4-AM substrate mixture.

3. Dimethyl sulfoxide (DMSO) was used to dissolve compounds and as a vehicle control for basal signal.

4. β-Naphthoflavone was used as a positive control for ARE-bla Assay.

*2.3 Supplies and Equipment*

1. T225 cell culture flasks.

2. Disposable, sterile centrifuge tubes.

3. 1536-well assay plates: black wall/clear bottom, cell culture treated and with low fluorescence background.

4. Cell strainer: a receptacle with a 40 μm nylon filter that is used to remove clumped cells from cell suspensions.

5. Lids for assay and compound plate: these reusable lids are made from stainless steel and contain a rubber gasket that sits around the top outer edge. The cellular assay lid contains small evenly placed holes that allow air exchange necessary for cell-based assays. The weight of the lid allows the gasket to form a strong barrier around the plate, virtually eliminating edge effects.

6. PinTool workstation (Wako Automation): the PinTool performs transfer of 23 nL of compound from a 1536-well compound plate to a 1536-well assay plate (*see* **Note 1**).

7. BioRAPTR FRD workstation (Beckman Coulter): a liquid handling system that can transfer of 0.2–10 μL of up to four different reagents or cells simultaneously into a 1536 well plate.

8. Multidrop Combi dispenser: a high speed dispenser capable of dispensing one reagent or cells using eight-channel detachable dispensing cassettes.

9. CyBi-well Vario pipettor, a 96, 384, and 1536 channel simultaneous pipettor: It requires the use of disposable tips, which is used for the preparation of positive control plate in a 1536-well plate format.

10. Cellometer Auto Cell Counter (Nexcelom Bioscience) is used to count viable cells.

11. EnVision plate reader (PerkinElmer) covers a wide range of fluorescence, absorbance, and luminescence readouts commonly used in high-throughput assays. The EnVision reader includes two detectors enabling simultaneous dual wavelength reading that is well suited for β-lactamase reporter assays. For an ARE-bla assay, reading requires a dual bottom mirror and compatible filter sets (Table 2).

**Table 2**
**Filter selection for ARE-bla assay**

| Filters | Excitation | Emission |
| --- | --- | --- |
| Channel 1 (green) | 409/20 nm | 530/30 nm |
| Channel 2 (blue) | 409/20 nm | 460/40 nm |
| Ratio (Ch2/Ch1) | 460 nm/530 nm | |

## 3   Methods

### 3.1   Cell Culture

1. Remove the cryovial containing the frozen cells from liquid nitrogen storage and immediately place it into a 37 °C water bath with gentle agitation for 1–2 min.

2. Place the vial into a laminar flow hood. Before opening the vial, wipe the outside of the vial with 70% ethanol.

3. Transfer the thawed cells into a 15 mL conical sterile centrifuge tube with 9 mL pre-warm thaw medium.

4. Centrifuge the cell supernatant for 4 min at $200 \times g$.

5. Carefully aspirate supernatant without disturbing the cell pellet

6. Gently resuspend cell pellet in thaw medium.

7. Transfer the desired amount of the cells to a T225 tissue culture flask.

8. Place the flask in an incubator with a humidified atmosphere of 5% $CO_2$ until passage. Maintain the cells at 30–90% confluence prior to passage. During the first passage, switch to culture medium.

9. After 48–72 h or cells with 80–90% confluence, aspirate medium and rinse once with DPBS, followed by the addition of 10 mL of 0.25% trypsin–EDTA and swirl to coat the cells evenly around flask.

10. Place the flask in incubator at 37 °C for 2–3 min or until the cells detach.

11. Add 10 mL of culture medium to deactivate trypsin.

12. Transfer the cells to a 50 mL conical tube and centrifuge at $200 \times g$ for 4 min.

13. Carefully aspirate supernatant and resuspend cell pellet in the culture medium.

14. Count cells using a Cellometer auto cell counter.

15. Transfer the cell suspension to a T225 tissue culture flask.

16. Incubate until next passage or assay. Cells should be passaged at least twice a week. Please not let cells grow over 90% confluence.

### 3.2 Preparation of Positive Control Plate

1. Assay-specific controls (positive and negative) are located on an additional 1536-well compound plate in columns 1–4. These controls will transfer simultaneously with the test compounds to the assay plate. The final concentration of DMSO in the assay is less than 0.5 %.

2. To make a positive control plate for ARE-bla assay, 5 μL of β-naphthoflavone stock solution in DMSO is dispensed into each well in columns 1, 2, and 3 (Fig. 2). CyBi-well Vario pipettor can be used to prepare the control plate.

### 3.3 ARE-bla Assay

1. Harvest the ARE-bla HepG2 cells and resuspend the cell pellet in assay medium (*see* **Note 2**).

2. Place the cells on a cell strainer to remove clumped cells before counting.

3. Count cell number and determine cell viability. Cell viability of 95 % or greater will have a better window of signal to basal.

4. Prepare cell stock in assay medium at density of $0.4 \times 10^6$ cells/mL.

5. Dispense 5 μL of cells prepared at step 4 into each well of a 1536-well black, clear-bottom, tissue culture-treated assay plates using a Multidrop Combi dispenser or BioRAPTR FRD.
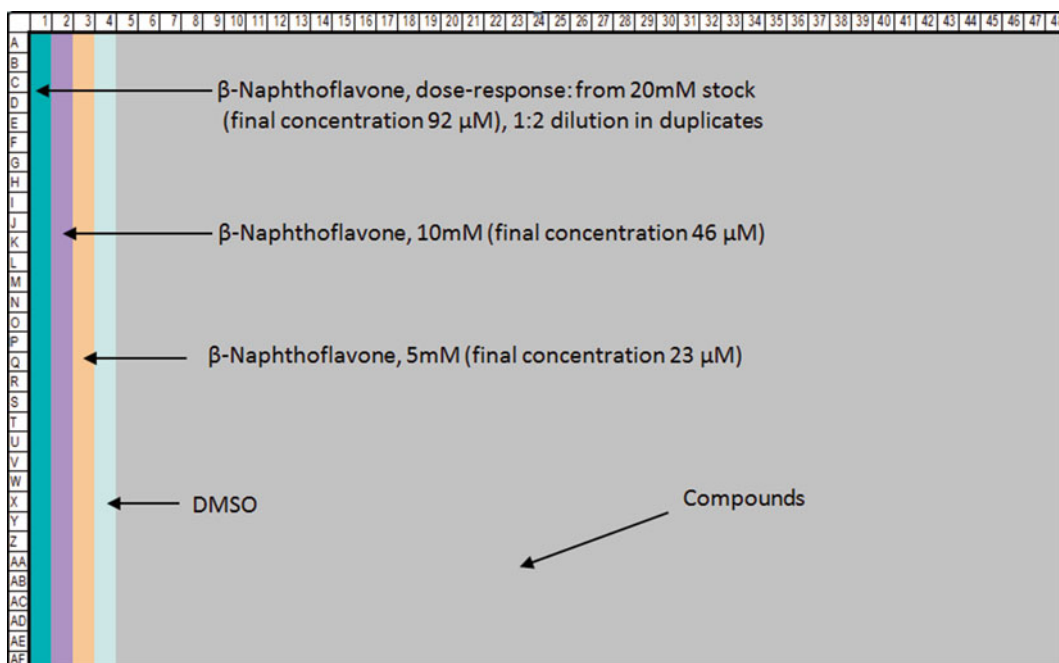


**Fig. 2** Control plate map in 1536-well format. Column 1 contains a 16-concentration titration ranging from 92 μM–2.8 nM in duplicates. Columns 2 and 3 contain replicates for the β-naphthoflavone: 46 μM in column 2 and 23 μM in column 3 in 32 replicates for each concentration. Additionally, DMSO, as a negative control, is dispensed into column 4

6. Place a pre-cleaned plate lid over the plate and incubate assay plates at 37 °C under a humidified atmosphere and 5 % $CO_2$ for 5 h to allow cells to attach.

7. Transfer 23 nL of compounds and controls to the assay plate by a PinTool.

8. Incubate the assay plates at 37 °C, 5 % $CO_2$ for 16 h.

9. Freshly prepare 6× CCF4-AM substrate mixture prior to assay termination: add 6 µL of 1 mM of CCF4-AM substrate to 60 µL of Solution B and mix, and then add 874 µL of Solution C and 60 µL of Solution D to the combined solution and vortex.

10. Dispense 1 µL of 6× CCF4-AM Substrate Mixture to each well using a BioRAPTR dispenser after 16 h compound treatment.

11. Incubate the plates in the dark at room temperature for 2 h for fluorescence development.

12. Measure fluorescence intensity at 460 and 530 nm emission and 405 nm excitation using an EnVision plate reader. Data is expressed as the ratio of the 460 nm/530 nm emissions (*see* **Note 3**).

13. Percentage of activity of the compounds is calculated by normalizing the raw data to DMSO wells (0 % activity) and maximal β-napthoflavone wells (100 % activity).

14. The compound half maximum effective concentration ($EC_{50}$) and maximum response (efficacy) values were calculated using a four-parameter Hill equation in GraphPad Prism software (Fig. 3).
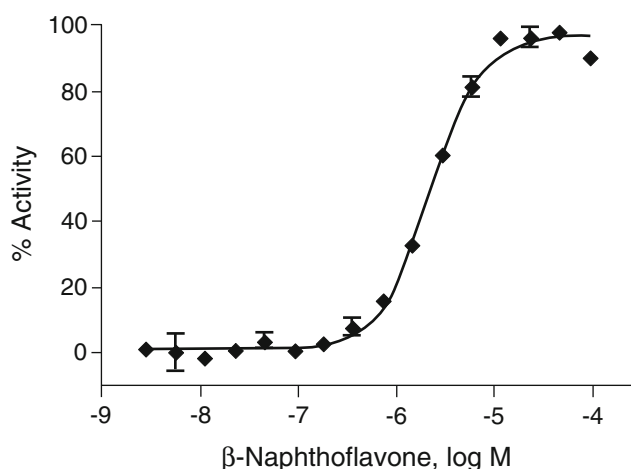


**Fig. 3** β-Naphthoflavone stimulated β-lactamase activity with $EC_{50}$ of 2.1 µM in ARE-bla HepG2 cells in a 1536-well plate format

## 4    Notes

1. To avoid cross-contamination, it is important to wash pins with appropriate reagents (e.g., DMSO and methanol) before and after use.

2. For screening, the cells that have passed at least one passage after thawing are recommended.

3. The ratiometric readouts from dual emissions (460 and 530 nm) can minimize well-to-well and plate-to-plate variation caused by differences in plating cell density.

## References

1. Hu R, Saw CL, Yu R, Kong AN (2010) Regulation of NF-E2-related factor 2 signaling for cancer chemoprevention: antioxidant coupled with antiinflammatory. Antioxid Redox Signal 13(11):1679–1698. doi:10.1089/ars.2010.3276

2. Saw CL, Kong AN (2011) Nuclear factor-erythroid 2-related factor 2 as a chemopreventive target in colorectal cancer. Expert Opin Ther Targets 15(3):281–295. doi:10.1517/14728222.2011.553602

3. Tufekci KU, Civi Bayin E, Genc S, Genc K (2011) The Nrf2/ARE pathway: a promising target to counteract mitochondrial dysfunction in Parkinson's disease. Parkinsons Dis 2011:314082. doi:10.4061/2011/314082

4. Al-Dalaen SM (2014) Review article: oxidative stress versus antioxidants. Am J Biosci Bioeng 2(5):60. doi:10.11648/j.bio.20140205.11

5. Kensler TW, Wakabayashi N, Biswal S (2007) Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu Rev Pharmacol Toxicol 47:89–116. doi:10.1146/annurev.pharmtox.46.120604.141046

6. Nguyen T, Nioi P, Pickett CB (2009) The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. J Biol Chem 284(20):13291–13295. doi:10.1074/jbc.R900010200

7. Shukla SJ, Huang R, Simmons SO, Tice RR, Witt KL, Vanleer D, Ramabhadran R, Austin CP, Xia M (2012) Profiling environmental chemicals for activity in the antioxidant response element signaling pathway using a high throughput screening approach. Environ Health Perspect 120(8):1150–1156. doi:10.1289/ehp.1104709

8. Wu RP, Hayashi T, Cottam HB, Jin G, Yao S, Wu CC, Rosenbach MD, Corr M, Schwab RB, Carson DA (2010) Nrf2 responses and the therapeutic selectivity of electrophilic compounds in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 107(16):7479–7484. doi:10.1073/pnas.1002890107

9. Zlokarnik G, Negulescu PA, Knapp TE, Mere L, Burres N, Feng L, Whitney M, Roemer K, Tsien RY (1998) Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. Science 279(5347):84–88

# Chapter 7

## Study Liver Cytochrome P450 3A4 Inhibition and Hepatotoxicity Using DMSO-Differentiated HuH-7 Cells

### Yitong Liu

### Abstract

Metabolically competent, inexpensive, and robust in vitro cell models are needed for studying liver drug-metabolizing enzymes and hepatotoxicity. Human hepatoma HuH-7 cells develop into a differentiated in vitro model resembling primary human hepatocytes after a 2-week dimethyl sulfoxide (DMSO) treatment. DMSO-treated HuH-7 cells express elevated cytochrome P450 3A4 (CYP3A4) enzyme gene expression and activity compared to untreated HuH-7 cells. This cell model could be used to study CYP3A4 inhibition by reversible and time-dependent inhibitors, including drugs, food-related substances, and environmental chemicals. The DMSO-treated HuH-7 model is also a suitable tool for investigating hepatotoxicity. This chapter describes a detailed methodology for developing DMSO-treated HuH-7 cells, which are subsequently used for CYP3A4 inhibition and hepatotoxicity studies.

**Key words** HuH-7, DMSO, CYP3A4, Inhibition, Hepatotoxicity

## 1 Introduction

Metabolism is a key function of liver, which is carried out by drug-metabolizing enzymes, such as cytochromes P450. Over the years, several in vitro cell models have been developed to mimic liver function, but many lack drug-metabolizing enzyme activities, which is true, for example, with HepG2 cells [1]. Other models, such as primary human hepatocytes and HepaRG cells, express drug-metabolizing enzyme activities, but are expensive, scarce, or require higher levels of maintenance [2]. Therefore, a cost-effective, metabolically competent, and robust in vitro cell model is needed for medium to high-throughput screening purposes.

Human hepatoma HuH-7 cells were derived from a Japanese male with well differentiated hepatocellular carcinoma in 1982 [3]. Studies have shown that HuH-7 cells could be induced by DMSO and differentiate into primary human hepatocyte-like cells [4, 5]. The DMSO-treated HuH-7 cells resemble primary human hepatocyte characteristics, such as, polygonal shape and binucleated cells
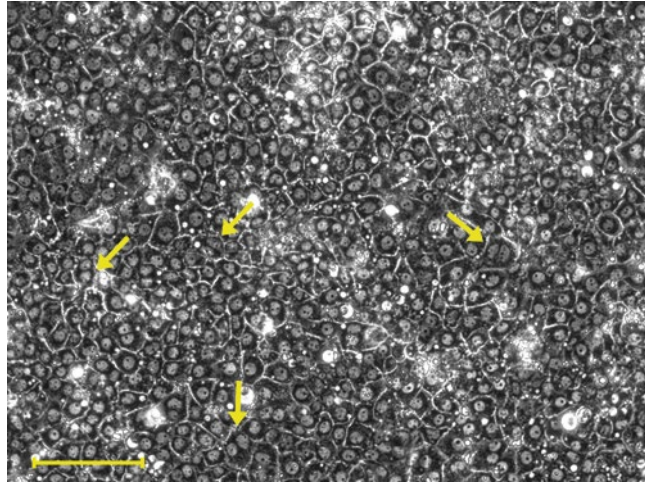
**Fig. 1** Phase contrast photomicrographs of DMSO-treated HuH-7 cells with *arrows* pointing to binucleated cells. Bar represents 100 μm
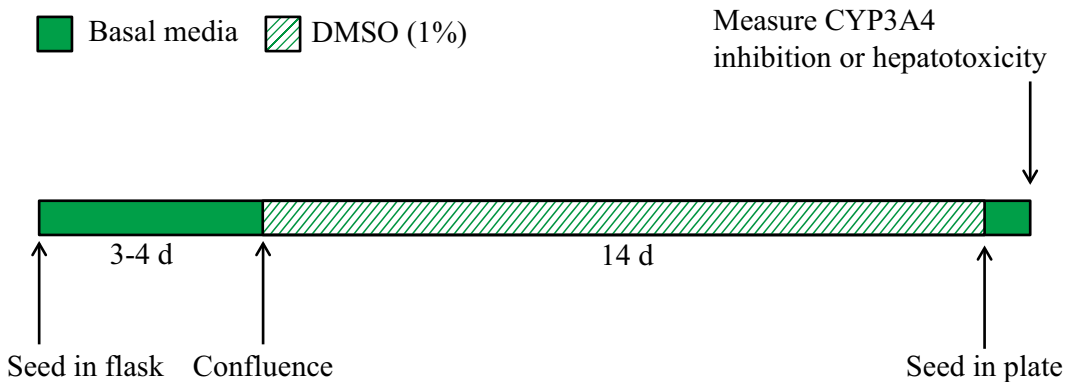


**Fig. 2** Development of DMSO-treated HuH-7 cells. Timeline begins two to three passages after initial thawing of cells

(Fig. 1). Furthermore, DMSO-treated HuH-7 cells express the functional CYP 3A4 enzyme, the most abundant liver enzyme which mediates the metabolism of more than 50% of all marketed drugs [5]. Inhibition of CYP3A4 activity could cause drug–drug or food–drug interactions, which may lead to adverse effects [6].

This chapter describes a protocol to establish DMSO-treated HuH-7 cells for measuring CYP3A4 inhibition and hepatotoxicity. A general timeline for developing DMSO-treated HuH-7 cell model is illustrated in Fig. 2. The cells are passaged for two or three times after thawing, then seeded in a flask for confluence and DMSO treatment. Finally, the cells are assayed for both reversible and time-dependent CYP3A4 inhibition, as well as tested for hepatotoxicity using control compounds.

## 2　Materials

| | |
|---|---|
| ***2.1　Equipment*** | 1. BMG FLUOstar Omega multimode microplate reader. |

1. BMG FLUOstar Omega multimode microplate reader.

2. New Brunswick Galaxy 48R $CO_2$ incubator.

3. Inverted microscope.

4. Countess, automated cell counter; counting slides; Trypan blue stain 0.4 %.

5. 96-well, black, clear flat bottom, tissue culture-treated polystyrene microplates, with lids, sterile.

6. 96-well, white, opaque flat bottom, untreated polystyrene microplates, nonsterile.

7. Vacuum filter/storage bottle system, 0.22 μm pore, cellulose acetate membrane filter, sterile.

8. 8-channel pipette and tips; multichannel reservoir.

***2.2　Cells and Culture Medium***

1. Human hepatoma HuH-7 cells, Health Science Research Resources Bank, Japan Health Sciences Foundation.

2. Dulbecco's Minimal Essential Medium (DMEM), low glucose, GlutaMAX Supplement, pyruvate.

3. 100× MEM nonessential amino acids and 1 M HEPES.

4. Fetal bovine serum, Atlanta Biologicals Premium Select.

***2.3　Reagents and Solutions***

1. Rat tail collagen type I.

2. Glacial acetic acid.

3. Dulbecco's phosphate-buffered saline (DPBS, no calcium, no magnesium) and Versene (0.02 % EDTA).

4. Trypsin–EDTA (0.5 %), without phenol red.

5. Promega P450-Glo CYP3A4 (luciferin-IPA substrate) and CellTiter-Glo assays; beetle luciferin, potassium salt.

6. DMSO, ketoconazole, troleandomycin, salicylamide, nitrofurantoin, and nefazodone hydrochloride.

## 3　Methods

***3.1　Cell Culture***

1. Coat tissue culture surface using rat tail collagen type I (Table 1). Dissolve collagen in 0.02 N acetic acid to make a 100 μg/mL solution. Coat tissue culture surfaces at 10 μg/cm². Incubate coated vessels at room temperature for at least 2 h and rinse with DPBS. Use fresh or air dry and store at 2–8 °C for up to 1 month.

2. Prepare basal medium (Table 2) and 1 % DMSO supplemented basal medium. Measure and mix all components under sterile

**Table 1**
**Collagen coating**

| Growth chamber | Surface area (cm²) | Collagen concentration (µg/mL) | Volume (mL) | Collagen density (µg/cm²) |
|---|---|---|---|---|
| 96-well plate | 0.32 | 100 | 0.032 | 10 |
| T25 flask | 25 | 100 | 2.5 | 10 |
| T75 flask | 75 | 100 | 7.5 | 10 |

**Table 2**
**Basal medium**

| Basal medium component | Concentration | Amount per 500 mL |
|---|---|---|
| DMEM | 88% | 440 mL |
| 100× MEM nonessential amino acids | 1× | 5 mL |
| 1 M HEPES | 10 mM | 5 mL |
| Fetal bovine serum | 10% | 50 mL |

conditions, followed by filter sterilization using a 0.22 µM filter. Medium is stable for 1 month if stored at 2–8 °C. 1% DMSO supplemented basal medium is prepared fresh for medium change after cells reach confluence.

3. Thaw and recover HuH-7 cells into T25 or T75 flasks using basal medium. Passage cells two or three times when reach 85–90% confluence.

4. When desired cell amount is reached, remove the basal medium from flask using aspiration and rinse twice with Versene solution.

5. Add 0.025% trypsin–EDTA reagent (prepared using 0.5% trypsin–EDTA and DPBS) and rinse the flask quickly. Remove extra trypsin reagent and leave 1 mL in the flask. Incubate the flask (37 °C) until cell detachment (3–5 min).

6. Use a microscope to observe. When cells are completely detached, add 3–4 mL prewarmed (37 °C) basal medium to stop trypsinization. Failure to add medium promptly can result in over-trypsinization and significant cell death.

7. Determine the viable cell concentration (cells/mL) using a cell counter. Adjust cell concentration using basal medium as needed.

8. Seed HuH-7 cells at $6.3 \times 10^4$ cells/cm² onto collagen coated T25 or T75 flasks using basal medium. Wait 10 min at room temperature before returning to the incubator. Replenish medium twice per week until 100% confluence, which takes about 3–4 days after seeding.

9. Upon cell confluence, switch basal medium to 1% DMSO supplemented basal medium (fresh prepared each time). Replenish medium twice per week for 2 weeks using 1% DMSO supplemented basal medium.

10. After the 2-week DMSO treatment, seed DMSO-treated HuH-7 cells at $3.1 \times 10^5$ cells/cm² onto a collagen coated clear bottom 96-well plate using basal medium. Use plated cells for assays within 48 h after seeding (*see* **Note 1**).

*3.2 CYP3A4 Reversible Inhibition Assay*

1. Reagent preparation.

   Serum-free medium, prepare according to Table 2, leave out fetal bovine serum, and compensate volume with DMEM.

   Luciferin-IPA substrate, thaw 3 mM stock solution, protect from light.

   Salicylamide, Phase II conjugation inhibitor, prepare 3 M in DMSO fresh each time.

   Beetle luciferin, prepare 2 mM stock in H2O, serial dilute using serum-free medium (0.2–80 nM, *see* **Note 2**).

   Luciferin detection reagent, equilibrate to room temperature.

   Ketoconazole, prepare 2 mM stock in DMSO and conduct serial dilution using DMSO (0.01–2 mM).

   Dissolve luciferin-IPA substrate (3 μM), salicylamine (3 mM) and various concentrations of ketoconazole (0.01–2 μM) using serum-free medium (*see* **Note 3**) and prewarm to 37 °C. Preparation without ketoconazole serves as the control.

2. CYP3A4 reversible inhibition.
   Wash DMSO-treated HuH-7 cells on 96-well plates with serum-free medium and incubate with ketoconazole (0.01–2 μM) simultaneously with CYP3A4 substrate luciferin-IPA (3 μM, Km value) at 37 °C (100 μL/well) for 30 min. Include beetle luciferin standards (100 μL) in blank wells.

3. Assay detection.
   Terminate reaction by transferring 50 μL of incubation medium from each well to a separate white flat bottom plate with wells containing 50 μL of luciferin detection reagent at room temperature. Mix well and incubate the plate for 20 min, then measure luminescence using a plate reader (Fig. 3a).

4. Calculate $IC_{50}$, the half-maximal inhibitory concentration of ketoconazole determined using a nonlinear regression Hill model.
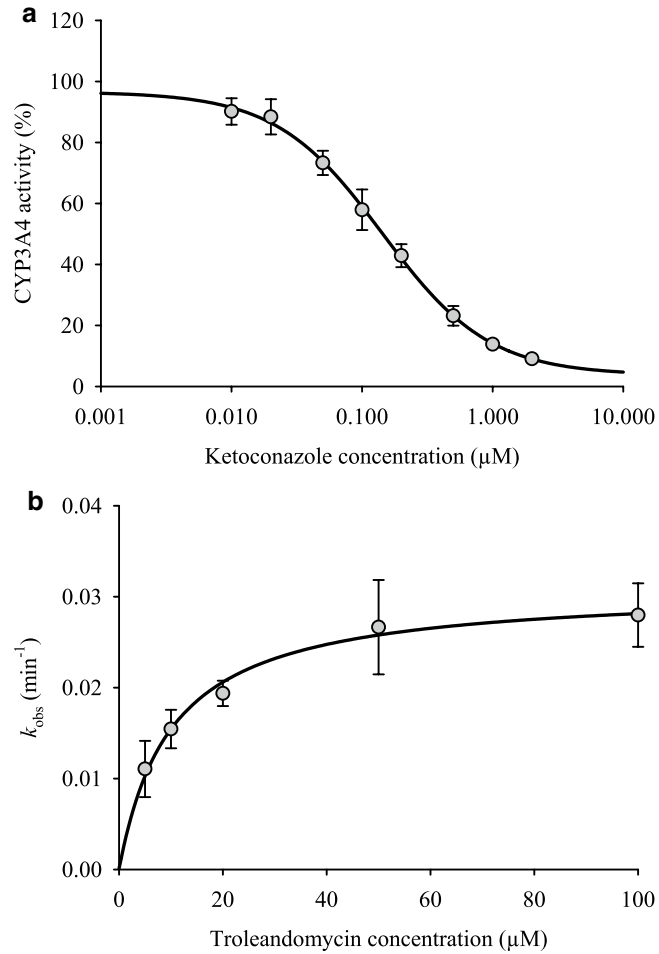
**Fig. 3** (**a**) CYP3A4 reversible inhibition by ketoconazole and (**b**) time-dependent inhibition by troleandomycin in DMSO-treated HuH-7 cells. Figure reproduced from reference [5]

**3.3 CYP3A4 Time-Dependent Inhibition Assay**

1. Reagent preparation.

    Serum-free medium, prepare according to Table 2, leave out fetal bovine serum, and compensate volume with DMEM.

    Luciferin-IPA substrate, thaw 3 mM stock solution, protect from light.

    Salicylamide, Phase II conjugation inhibitor, prepare 3 M in DMSO fresh each time.

    Beetle luciferin, prepare 2 mM stock in $H_2O$, serial dilute using serum-free medium (0.2–80 nM, *see* **Note 2**).

    Luciferin detection reagent, equilibrate to room temperature.

    Troleandomycin, prepare 10 mM in DMSO and conduct serial dilution using DMSO (0.2–10 mM). Dissolve each concentration using serum-free medium (5–100 µM) and prewarm

to 37 °C. Incubation without troleandomycin serves as the control.

Dissolve luciferin-IPA substrate (3 μM) and salicylamide (3 mM) using serum-free medium and prewarm to 37 °C.

2. CYP3A4 time-dependent inhibition: preincubation.
Wash DMSO-treated HuH-7 cells on 96-well plates with serum-free medium and incubate with troleandomycin (5–100 μM, 100 μL/well) for 0, 15, 30, and 60 min at 37 °C.

3. CYP3A4 time-dependent inhibition: substrate incubation.
At different time points, wash cells with serum-free medium and incubate with CYP3A4 substrate luciferin-IPA (3 μM, Km value) at 37 °C (100 μL/well) for an additional 30 min. Include beetle luciferin standards (100 μL) in blank wells.

4. Assay detection.
Terminate reaction by transferring 50 μL of incubation medium from each well to a separate white flat bottom plate with wells containing 50 μL of luciferin detection reagent at room temperature. Mix well and incubate the plate for 20 min, then measure luminescence using a plate reader (Fig. 3b).

5. Calculate Parameters.
Obtain inactivation rate constant $k_{obs}$ by plotting the natural logarithm of the remaining CYP3A4 activity (%) against preincubation time with troleandomycin [I].

Calculate kinetic parameters $k_{inact}$ and $K_I$ by fitting data to the following equation using a nonlinear regression,

$$k_{obs} = \frac{k_{inact}[I]}{K_I + [I]}$$

**3.4 Hepatotoxicity Assay**

1. Reagent preparation.
Nitrofurantoin and nefazodone, prepare 100 mM in DMSO, protect from light. Conduct serial dilution using basal medium (1–100 μM) on treatment day.
CellTiter-Glo assay, reconstitute and equilibrate at room temperature on assay day (24 h after treatment day).

2. Treatment.
Incubate DMSO-treated HuH-7 cells with nitrofurantoin or nefazodone (100 μL/well) for 24 h at 37 °C. Incubation without chemicals serves as the control.

3. Toxicity assay.
After treatment, equilibrate HuH-7 cells at room temperature for 30 min. Add 100 μL of CellTiter-Glo reagent to each well and mix well. Incubate the plate for additional 10 min, then measure luminescence using a plate reader (Fig. 4).

4. Calculate $EC_{50}$, the half-maximal cytotoxic concentration of hepatotoxicants determined using a nonlinear regression Hill model.
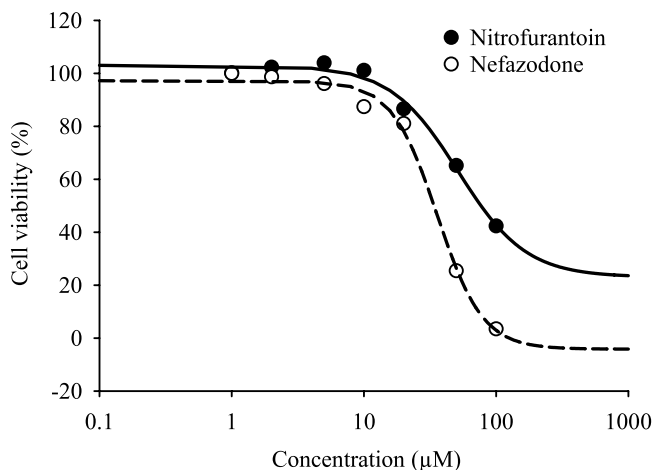
**Fig. 4** Cytotoxicity of nitrofurantoin and nefazodone (24 h) in DMSO-treated HuH-7 cells. Figure reproduced from reference [5]

## 4  Notes

1. DMSO-treated HuH-7 cells maintain CYP3A4 activity in basal medium for at least 48 h after plating.

2. Add 100 μL/well of beetle luciferin standards (0.2–80 nM) to blank wells on the 96-well plate, proceed together with samples for inhibition studies and luminescence detection. When calculate enzyme activity, use beetle luciferin standard curve range 0.1–40 nM.

3. Prepare luciferin-IPA and salicylamide in serum-free medium, then dissolve various concentrations of ketoconazole using a multichannel reservoir.

## References

1. Lin J, Schyschka L, Muhl-Benninghaus R, Neumann J, Hao L, Nussler N, Dooley S, Liu L, Stockle U, Nussler AK, Ehnert S (2012) Comparative analysis of phase I and II enzyme activities in 5 hepatic cell lines identifies Huh-7 and HCC-T cells with the highest potential to study drug metabolism. Arch Toxicol 86(1):87–95. doi:10.1007/s00204-011-0733-y

2. Donato MT, Jover R, Gomez-Lechon MJ (2013) Hepatic cell lines for drug hepatotoxicity testing: limitations and strategies to upgrade their metabolic competence by gene engineering. Curr Drug Metab 14(9):946–968

3. Nakabayashi H, Taketa K, Miyano K, Yamane T, Sato J (1982) Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. Cancer Res 42(9): 3858–3863

4. Choi S, Sainz B Jr, Corcoran P, Uprichard S, Jeong H (2009) Characterization of increased drug metabolism activity in dimethyl sulfoxide (DMSO)-treated Huh7 hepatoma cells. Xenobiotica 39(3):205–217. doi:10.1080/00498250802613620

5. Liu Y, Flynn TJ, Xia M, Wiesenfeld PL, Ferguson MS (2015) Evaluation of CYP3A4 inhibition and hepatotoxicity using DMSO-treated human hepatoma HuH-7 cells. Cell Biol Toxicol 31(4-5):221–230. doi:10.1007/s10565-015-9306-9

6. Hisaka A, Ohno Y, Yamamoto T, Suzuki H (2010) Prediction of pharmacokinetic drug-drug interaction caused by changes in cytochrome P450 activity using in vivo information. Pharmacol Ther 125(2):230–248. doi:10.1016/j.pharmthera.2009.10.011

# Chapter 8

# Determination of Histone H2AX Phosphorylation in DT40 Cells

**Kana Nishihara, Sampada A. Shahane, and Menghang Xia**

## Abstract

Visualization of DNA damage response protein recruitment to DNA damage sites enables measurement of the DNA damage. DNA double-strand breaks (DSBs) and blocked replication forks induce the phosphorylation of H2AX at serine 139 (γH2AX), and accumulate γH2AX which can then be detected as foci. The detection of γH2AX foci by immunostaining with antibodies that recognize γH2AX is an indicator of DSBs presence. This chapter describes the measurement of γH2AX immunostaining using a high-content imaging platform in chicken DT40 B-lymphocyte cell lines.

**Key words** γH2AX, High-contents imaging, DSB, Immunostaining

## 1 Introduction

DNA damage response is crucial to maintain the homeostasis of cells. Damage that remains unrepaired or incorrectly repaired may lead to genetic mutations, instability, and increased risk of carcinogenesis. One of the most serious sources of damage in cells, DNA double-strand breaks (DSBs) are often induced by a various sources, including ionizing radiation and exposure to DNA-damaging chemical or environmental stress [1]. On the occurrence of DSB, cells initiate DNA response signaling and recruit DNA damage repair proteins to affected DNA sites to repair the altered DNA.

After the formation of DSBs, H2AX is rapidly phosphorylated on a serine residue to create γH2AX [2]. H2AX is a variant form of histone H2A and is ubiquitously distributed throughout the genome. Its sequence is conserved well among species [3]. In the initial response to DSBs, H2AX is phosphorylated on serine 139 by three kinases: ataxia telangiectasia mutated (ATM), ataxia telangiectasia and Rad3 related (ATR), and DNA-dependent protein kinase (DNA-PK). The γH2AX triggers the recruitment of various proteins involved in DNA repair [3]. H2AX is phosphorylated in megabase regions of surrounding the DNA break site [4]. Large numbers of γH2AX

molecules can be visualized as foci in nuclear region by immunostaining with antibodies that recognize γH2AX. Monitoring of γH2AX foci formation is useful for detecting the incidence of DSBs.

DT40 Chicken B-lymphocyte cells are widely used to make and analyze the DNA-repair gene knockout clones because of their high efficacy in targeted integration [5, 6]. DT40 cells have a short cell doubling time (~8 h), a long S phase (about 70% of the cell cycle), and a lack of a G1/S checkpoint. DT40 cells are sensitive to chemicals that produce DSBs by disrupting replication forks because of their long S phase and their lack of a G1/S check point [7].

Here, we describe the methods of γH2AX immunostaining for high-content imaging analysis in 384-well plate format using the chicken DT40 B-lymphocyte cell line [8].

## 2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C) and analytical grade reagents. Prepare all solutions at room temperature unless indicated otherwise.

1. DT40 cells (provided by S. Takeda, Kyoto University, Japan).

2. Culture medium: Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS), 1% chicken serum, 50 μM β-mercaptoethanol, 100 U/mL penicillin and 100 μg/mL streptomycin. Store at 4 °C.

3. Collagen I coated 384-well black wall/clear bottom plate (Corning Incorporated, Tewksbury, MA, USA).

4. Positive control compounds, adriamycin (CASRN (Chemical Abstract Services Registry Number) = 25316-40-9) and melphalan (CASRN = 148-82-3).
   Chemicals are dissolved in dimethyl sulfoxide (DMSO) and prepared as 20 mM stock solutions prior to use.

5. Hank's balanced salt solution (HBSS).

6. 10 mg/mL Hoechest 33342 solution in water.

7. Fixing solution: 12% Paraformaldehyde and 0.3% Hoechst 33342 in HBSS, 10 mg/mL of Hoechst 33342 solution in water is used. 32% Paraformaldehyde stock solution is used. Add 13 mL of 32% paraformaldehyde solution and 105 μL of 10 mg/ml Hoechst solution to 22 mL of HBSS.

8. Permeabilization solution: 0.1% IGEPAL® (Sigma-Aldrich) in HBSS. Add 100 μL of IGEPAL to 100 mL of HBSS and mix well using Vortex.

9. Blocking solution: 3% Bovine serum albumin (BSA) in HBSS. Dissolve 3 g BSA in 100 mL HBSS.

10. Anti-phospho-Histone H2AX antibody (EMD Millipore). Add 60 μL of anti-phospho-Histone H2AX antibody to 60 mL of blocking buffer. Keep on ice before use.

11. Alexa Fluor 594 goat–anti-mouse IgG secondary antibody (Life Technologies). Add 60 μL of Alexa Fluor 594 goat–anti-mouse IgG secondary antibody to 60 mL of blocking buffer. Keep on ice before use.

12. 384-well plate sealing film.

13. Pipettes.

14. 8-channel aspirator.

15. ImageXpress Micro Widefield High-Content Screening System (Molecular Devices).

## 3   Methods

Carry out all procedures at room temperature unless otherwise specified.

### 3.1   Compound Treatment

1. Plate the cells (6000–8000 cells/well/25 μL) into collagen I coated 384-well black wall/clear bottom plate and incubate overnight at 37 °C under a humidified atmosphere and 5% $CO_2$. To allow the cell attachment on the bottom of the wells, incubate the plates overnight (*see* **Note 1**).

2. Next day, add 25 μL/well of the compound on top at desired concentration into the wells and incubate at 37 °C for 24 h or time optimized for cell line of interest. Prepare 2× final concentrations of compounds and add into assay plate, to get 1× final concentration in assay plate.

### 3.2   Fixation and Antibody Staining

1. Add 25 μL/well of fixing solution on top, and incubate for 10 min. Final concentration of paraformaldehyde is 4% and Hoechst is 0.1%.

2. Remove liquid gently using aspirator.

3. Wash with HBSS (50 μL/well).

4. Add 25 μL/well of permeabilization solution for 15 min.

5. Remove liquid gently using aspirator.

6. Wash twice with HBSS (50 μL/well).

7. Add 50 μL/well of blocking solution and incubate for 30 min at 37 °C.

8. Remove blocking buffer and add 25 μL/well of anti-phospho-Histone H2AX antibody (1:1000 dilution, diluted in blocking buffer) and incubate for 1 h at 37 °C.

9. Wash three times with HBSS (50 μL/well).

10. Add 25 μL/well of Alexa Fluor 594 goat–anti-mouse IgG secondary antibody (1:1000 dilution, diluted in blocking buffer) and incubate for 45 min at 37 °C.

11. Wash three times with HBSS (50 μL/well).

12. Add 50 μL/well of HBSS and seal the plate.

*3.3  Imaging Readout*    1. Take images using a 40× objective in ImageXpress Micro Widefield High-Content Screening System using DAPI (acquire nuclear images), Texas Red (acquire γH2AX foci) filter sets. Analyze at least 100 cells per compound treatment using the Transfluor module, a proprietary analysis protocol with MetaXpress software (Figs. 1 and 2).
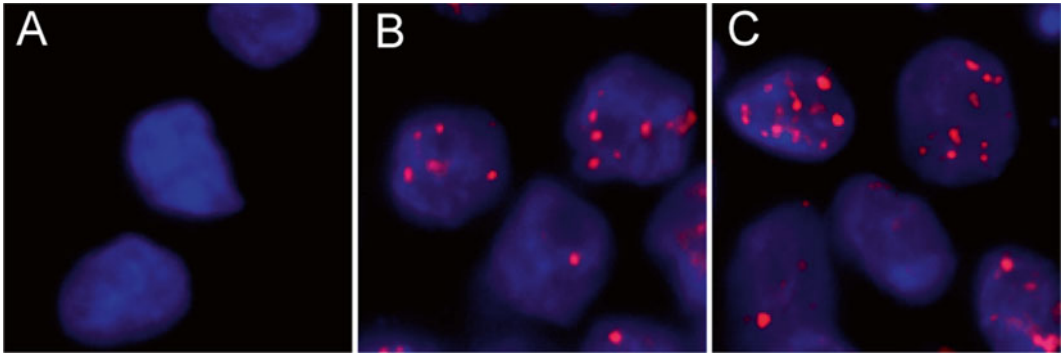


**Fig. 1** Induction of γH2AX foci in wild-type DT40 cells. (**a–c**) Representative images of nuclei (*blue*) with γH2AX (*red*) foci. (**a**) Without compound treatment (DMSO, control). (**b**) Treatment with 0.13 μM adriamycin for 24 h. (**c**) Treatment with 30.7 μM melphalan for 24 h
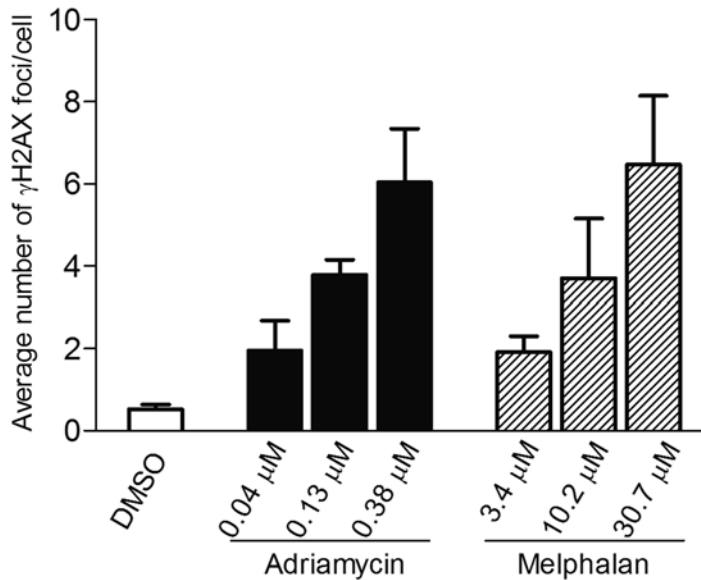


**Fig. 2** Induction of γH2AX foci by Adriamycin and Melphalan in a concentration-dependent manner. The average number of γH2AX foci per cell was measured after the compound treatment. Error bars represent SD from three independent experiments

| Hoechst (DAPI) | γH2AX (Texas Red) | Segmentation and Identification |
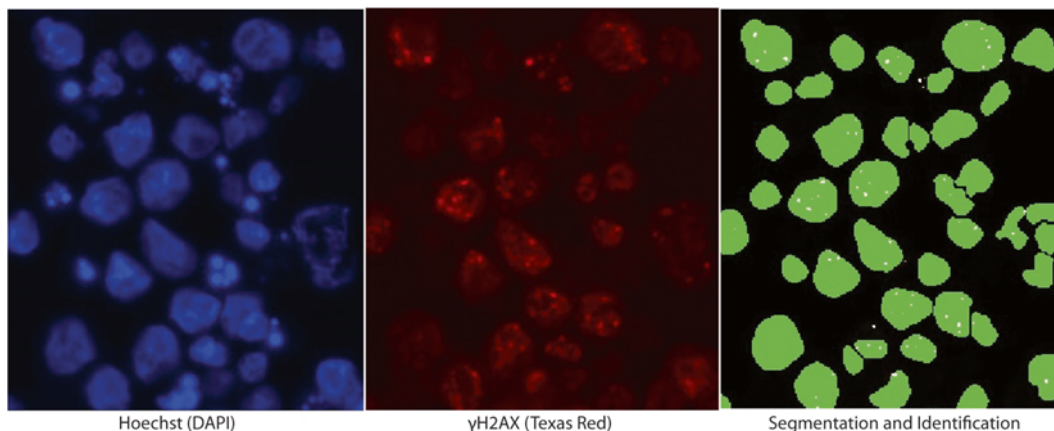
**Fig. 3** Segmentation: Transfluor analysis module correctly identifies the nucleus and H2AX foci

*3.4*  ***Image Analysis***    Molecular Devices Systems proprietary MetaXpress Transfluor analysis module works in a way that it identifies Nuclei based on the size and intensity of DAPI channel as nuclei are stained with Hoechst. Total number of nuclei in a well will represent number of cells in the well. Identification of foci is based on the size and intensity of Texas Red channel as foci are stained with red rH2AX antibody. Thus the program quantifies the number of foci per nucleus. In the results, the average number of foci per cell (Fig. 3) was used to calculate the activity.

## 4  Notes

1. Maintaining DT40 cells in culture under $1 \times 10^6$ cells/ml is critical for cell health. Due to the nature of the suspension cells, DT40 cells need to adhere to collagen coated surface as a crucial step in this assay in order to reduce the assay variability.

### References

1. Geric M, Gajski G, Garaj-Vrhovac V (2014) gamma-H2AX as a biomarker for DNA double-strand breaks in ecotoxicology. Ecotoxicol Environ Saf 105:13–21. doi:10.1016/j.ecoenv.2014.03.035

2. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem 273(10):5858–5868

3. Bonner WM, Redon CE, Dickey JS, Nakamura AJ, Sedelnikova OA, Solier S, Pommier Y (2008) GammaH2AX and cancer. Nat Rev Cancer 8(12):957–967. doi:10.1038/nrc2523

4. Sedelnikova OA, Rogakou EP, Panyutin IG, Bonner WM (2002) Quantitative detection of (125)IdU-induced DNA double-strand breaks with gamma-H2AX antibody. Radiat Res 158(4):486–492

5. Buerstedde JM, Takeda S (1991) Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell 67(1):179–188

6. Yamazoe M, Sonoda E, Hochegger H, Takeda S (2004) Reverse genetic studies of the DNA damage response in the chicken B lymphocyte line DT40. DNA Repair (Amst) 3(8-9):1175–1185. doi:10.1016/j.dnarep.2004.03.039

7. Evans TJ, Yamamoto KN, Hirota K, Takeda S (2010) Mutant cells defective in DNA repair pathways provide a sensitive high-throughput assay for genotoxicity. DNA Repair (Amst) 9(12):1292–1298. doi:10.1016/j.dnarep.2010.09.017

8. Nishihara K, Huang R, Zhao J, Shahane SA, Witt KL, Smith-Roe SL, Tice RR, Takeda S, Xia M (2015) Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. Mutagenesis 31:69–81. doi:10.1093/mutage/gev055 [pii]

# High-Throughput and High-Content Micronucleus Assay in CHO-K1 Cells

## Sampada A. Shahane, Kana Nishihara, and Menghang Xia

## Abstract

Visualization of micronuclei induction by chemicals and drugs enables measurement of possible compound genotoxicity. A loss of entire chromosome or a fragment of chromosome can lead to formation of micronuclei (MNi). The in vitro micronucleus assay can be conducted using nuclear dyes with high-content imaging platforms. This chapter describes the cytochalasin block method of measuring micronuclei in CHO-K1 cell lines.

**Key words** Micronuclei, High-content imaging, Genotoxicity

## 1 Introduction

The study of DNA damage at chromosomal level is a vital part of genetic toxicology. DNA damage may result from chromosome loss or from mal-segregation of a chromosome, leading to an important event in aging and carcinogenesis [1]. Chromosome loss and mal-segregation are probably caused by defects in a spindle or a centromere, but they may be a consequence of under-condensation of a chromosome structure before metaphase [2–4]. Evaluation of chromosome aberrations in metaphases is often used in the cytogenetic research [5]. Although this approach provides the most detailed analysis, the complexity and laboriousness of this method, in addition to the artifactual loss of chromosomes from metaphase preparations, call for the development of simpler systems to measure chromosomal damage [1].

Heddle in 1973 [6] and Schmid in 1975 [7] independently proposed measurement of micronuclei (MNi) in vivo in dividing cell populations, such as the bone marrow and peripheral blood erythrocytes, as an alternative and simpler method of assessing genotoxicity that was also more robust [6–8]. Although these assays are now one of the best established in vivo cytogenetic assays

in the field of genotoxicity, it has limited applicability to other cell types. To overcome this limitation, Fenech et al. developed a method that can measure MNi in a variety of nucleated cells [1]. MNi are ideally scored in the binucleated stage of cells, as MNi are expressed in cells that are completing nuclear division [9, 10]. In addition, several methods based on stathmokinesis, flow cytometry, and DNA labeling approaches have been developed. Among these methods, cytokinesis-block micronucleus (CBMN) method is found to be most favored in vitro micronucleus assay (IVMN) because of its ease, speed, reproducibility, and ability to concurrently estimate mitotic delay and lack of uncertainty regarding its effect on baseline genetic damage [9–12]. So far, CBMN has been proven to be an effective tool for the study of cellular and nuclear dysfunction caused by aging, deficiency, or excess of micronutrient, genotoxin exposure, and genome maintenance [13]. This method is also helpful in the emerging fields of nutrigenomics and toxicogenomics because nutrient status may affect the sensitivity of the cells to exogenous genotoxins [14].

For many years being part of several recommended regulatory battery tests for testing genotoxicity, in vitro micronucleus assay is routinely used as a rapid screening test [15]. For example, OECD guidelines require that the micronuclei frequency be measured in binucleated cells [16]. CBMN is a very useful approach of detecting micronuclei in binucleated cells, making this method an excellent candidate for automated image analysis [12]. Various high-content systems have been developed using several automated scoring methodologies owing to faster delivery of results and reduced variability among scorers [17]. The experiment procedure is the same for both manual and automated scoring. In manual scoring, a trained operator reads the slides under a microscope. Automated scoring, on the other hand uses proprietary image analysis software specially designed for the particular imaging system [18].

Recent publications indicate that the automated IVMN assay on cultured cells is a powerful genotoxicity assay with cellular imaging [18, 19]. In 2013, Tilmant et al., using CHO-K1 cells in automated micronucleus assay, found 91 % predictivity with a sensitivity of 94 % and a specificity of 85 % [20]. Here, we describe the CBMN-based method of micronuclei detection using automated high-content imaging analysis coupled with fluorescent microscopy in a 384-well plate format in CHO-K1 cell line [21].

## 2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C) and analytical grade reagents. Prepare all solutions at room temperature (RT) unless indicated otherwise.

**2.1 Sample Preparation**

1. Chinese hamster ovary (CHO-K1) cells.

2. Culture medium (CM): F-12K Nutrient Mixture supplemented with 10% FBS and 100 U/ml penicillin and 100 μg/ml streptomycin. Store at 4 °C.

3. Collagen I coated 384-well black wall/clear bottom plate.

4. Positive control compounds: cyclophosphamide (CP) (CASRN (Chemical Abstract Services Registry Number = 6055-19-2), mitomycin C (MMC) (CASRN = 50-07-7) and staurosporine (CASRN = 62996-74-1).

5. 1 mg/ml cytochalasin B.

6. Ultrapure distilled water.

7. 20% Aroclor 1254-induced Sprague Dawley male rat liver S9 mix (MUTAMYME™ S9 Mix, Moltox).

**2.2 Fluorescent Staining and Imaging**

1. Hank's balanced salt solution (HBSS).

2. Fixing solution: 8% Paraformaldehyde (from 32% stock) and 0.2% Hoechst 33342 (from 10 mg/ml stock) 0.04% Red Cell Mask (from 10 mg/ml stock) and 0.4% CellEvent™ Caspase-3/7 Green Detection Reagent (2 mM stock) in HBSS.

3. 384-well plate sealing film.

4. Pipettes.

5. 8-channel aspirator.

6. ImageXpress Micro Widefield High-Content Screening System (Molecular Devices).

# 3 Methods

Carry out all procedures at RT unless otherwise specified.

**3.1 Compound Treatment**

1. Cell dispensing: Calculate the cells preferably in single cell suspension by using the cell strainer and adjust the cell density for +S9 and −S9 conditions. Plate 4500 cells/well/25 μl for +S9 condition or 750 cells/well/25 μl for −S9 condition into collagen I coated 384-well black wall/clear bottom plate and incubate the assay plates for 4 h to allow proper attachment of cells to the well at 37 °C under a humidified atmosphere and 5% $CO_2$.

Getting a single cell suspension is critical for imaging assays as it facilitates identification of individual cell and then normalizing data accordingly.

In +S9 condition the cell density is higher than in −S9 condition as per OECD guidelines to combat the probable S9 related toxicity.

2. After 4 h incubation, add 25 μl/well of the compound (e.g., MMC and CP) with or without 4% S9 on top at desired concentrations into the wells and incubate at 37 °C for 24 h or time optimized for cell line of interest.

Prepare the stock solutions of 1.2 mM mitomycin C (MMC) and 30 mM cyclophosphamide (CP) in ultrapure distilled water, and 40 mM staurosporine in dimethyl sulfoxide (DMSO). Stock solutions of MMC and CP are stored at –80 °C and staurosporine is stored at –20 °C. MMC is used as a positive control for –S9 condition, CP is used as a positive control for +S9 condition and staurosporine is used as a positive control for apoptosis. Using apoptotic marker aids in identification of healthy binucleated cells.

Prepare 20% S9 by adding 10 ml of ice-cold, sterile, purified water to MUTAZYME™ Aroclor 1254-induced Sprague Dawley male rat liver S9 lyophilized with NADPH-regenerating system cofactors and phosphate buffer.

S9 is an exogenous metabolic activation system needed by some test compounds to exert their genotoxic effect using in vitro test systems. S9 by itself is toxic to the cells and the amount of S9 used in an assay needs to be optimized. Also the above S9 is lyophilized with NADPH-regenerating system cofactors and phosphate buffer at the optimal concentrations. Two percent of S9 in the final assay volume is optimal with less cytotoxic effect to the cells.

Prepare 2× final concentrations of compounds and add into assay plate to get 1× final concentration in assay plate.

S9 mix: Prepare 4% S9 by adding 2 ml stock of 20% S9 in 8 ml CM.

MMC: Prepare 2.4 μM (800 ng/ml) (final concentration, 400 ng/ml) MMC by diluting 1.2 mM (400 μg/ml) stock of MMC to intermittent 60 μM MMC by adding 10 μl of stock MMC to 190 μl of 4% S9 mix (for +S9) or CM (for –S9). Then add 10 μl of 60 μM MMC to 240 μl of 4% S9 mix (for +S9) or CM (for –S9).

CP: Prepare 71.6 μM (final concentration, 35.8 μM) CP by diluting 30 mM stock of CP to intermittent 1.5 mM CP by adding 10 μl of stock CP to 190 μl of 4% S9 mix (for +S9) or CM (for –S9). Then add 10 μl of 1.5 mM CP to 200 μl of 4% S9 mix (for +S9) or CM (for –S9).

Staurosporine: Prepare 184 μM (final concentration, 91 μM) staurosporine by diluting 40 mM stock of staurosporine to intermittent 4 μM staurosporine by adding 5 μl of stock to 45 μl of 4% S9 mix (For +S9) or CM (for –S9). Then add 13.5 μl of 4 μM staurosporine to 280 μl of 4% S9 mix (for +S9) or CM (for –S9).

3. After 4 h incubation, remove compounds with S9, followed by washing three times with CM very gently, and then adding 25 μl CM/well. The assay plate is incubated overnight.

For –S9 condition, the assay plate stays in the incubator for compound treatment overnight.

### 3.2 Cytochalasin B Treatment

1. Remove medium gently and add cytochalasin B (final concentration, 3 μg/ml).

Prepare 3 μg/ml cytochalasin B by diluting 1 mg/ml stock of cytochalasin B to intermittent 100 μg/ml cytochalasin B by adding 100 μl of stock to 900 μl of CM. Then add 600 μl of 100 μg/ml cytochalasin B to 19.4 ml CM.

2. Incubate for 24 h at 37 °C under a humidified atmosphere and 5 % $CO_2$ (*see* **Note 1**).

### 3.3 Fixation and Staining

1. Add 25 μl/well fixing solution on top after 24 h. The various components of fixing solution and the recipe are as follows:

HCS CellMask: Stock 10 mg/ml HCS CellMask (HCS CellMask™ Red stain, Life Technologies, Catalog number H32712)—Dissolve the entire content of dye vial in 25 μl DMSO. Store at –20 °C.

HCS CellMask™ Stains are available in either blue, green, orange, red, or deep red detection spectra. We used HCS CellMask™ Red stain to label the whole cell to accommodate multi wavelength assay since Hoechst used in the assay is blue and CellEvent Caspase-3/7 is green.

CellEvent Caspase-3/7: 2 mM stock solution in DMSO (CellEvent™ Caspase-3/7 Green Detection Reagent, Catalog number C10423, Life Technologies) is used to measure activated caspase 3/7. Although the mechanism by which apoptosis is initiated varies depending on cell type and initiating event, activation of caspases is one of the markers of programmed cell death. The reagent consisting of the DEVD peptide sequence conjugated to a nucleic acid–binding dye acts in a way that the substrate is intrinsically nonfluorescent and only in the presence of activated caspase 3/7, produces a fluorogenic response indicative of apoptosis. The fluorescence emission maximum of the dye is approximately 520 nm. Additionally this no wash protocol preserves fragile apoptotic cells typically lost during wash steps. Lastly, the fluorescent signal from the CellEvent™ Caspase-3/7 reagent survives fixation and permeabilization, providing the flexibility to perform end-point assays and probe for other proteins of interest using immunocytochemistry. CellEvent™ Caspase-3/7 Green Detection Reagent is used to differentiate the healthy cells from unhealthy cells as required by OECD guidelines for our automated micronucleus assay.

Add 5.5 ml of 32 % paraformaldehyde solution, 4.4 µl of HCS CellMask Red, 44 µl CellEvent Caspase-3/7 green dye and 22 µl of 10 mg/ml Hoechst solution to 16.5 ml of HBSS to make total 22 ml fixing solution.

Final concentration of paraformaldehyde is 4 %, HCS CellMask™ Red stain is 0.02 %, CellEvent™ Caspase-3/7 Green Detection Reagent is 0.2 % and Hoechst is 0.1 %.

2. Incubate at RT for 30 min.

3. Wash with HBSS two times. During the wash step, be very gentle and careful.

4. Add 50 µl HBSS/well, seal, and image or store at 4 °C.

**3.4   Imaging Readout**    Acquire images using a 20× objective in ImageXpress Micro Widefield High-Content Screening System with DAPI (acquire nuclear images), Texas Red (acquire whole cell) and FITC (acquire apoptotic cells) filter sets. Acquire at least 1050 binucleated cells per compound treatment using the Micronucleus module, a proprietary protocol with MetaXpress software.

**3.5   Image Analysis**    As described previously [21], the Molecular Devices Systems proprietary MetaExpress Micronucleus analysis module identified stained nuclei using Hoechst based on the size, intensity, and distance from adjacent cell. The total number of cells in a well accounted for all the cells in the well. Main nuclei in a well were classified as mononucleated, binucleated, and multinucleated. Micronuclei also stained with Hoechst were identified based on size, intensity, and distance from main nuclei. A small nuclear mass attached to main nucleus was not depicted as micronucleus. The image analysis software provided information on the number of micronuclei in mononucleated, binucleated, and multinucleated cells, respectively. The percentage of micronuclei in present study represented healthy binucleated cells containing micronuclei (Figs. 1 and 2).

**3.6   Data Analysis**    As described previously [21], the number of micronucleated binucleated cells/1050 binucleated cells in treated cultures was compared with the number of micronucleated binucleated cells/1050 binucleated cells in the corresponding vehicle control culture. Data was expressed as the mean percentage of micronucleated binucleated cells from three replicate cultures ± standard deviation. Statistical significance of the frequency of micronucleated cells in the treated cultures at each dose level compared with the control value was determined using a one-tailed $t$-test. For cytotoxicity assessment, the nuclear division index (NDI) from MetaXpress software was used. NDI was defined as: $(M_1 + 2M_2 + 3M_3 + 4M_4)/N$, in which $M_1$–$M_4$ represented the number of cells with 1–4 nuclei and N was the total number of viable cells (excluding apoptotic

CellMask Red (Texas Red)          Hoechst (DAPI)          Segmentation & Identification
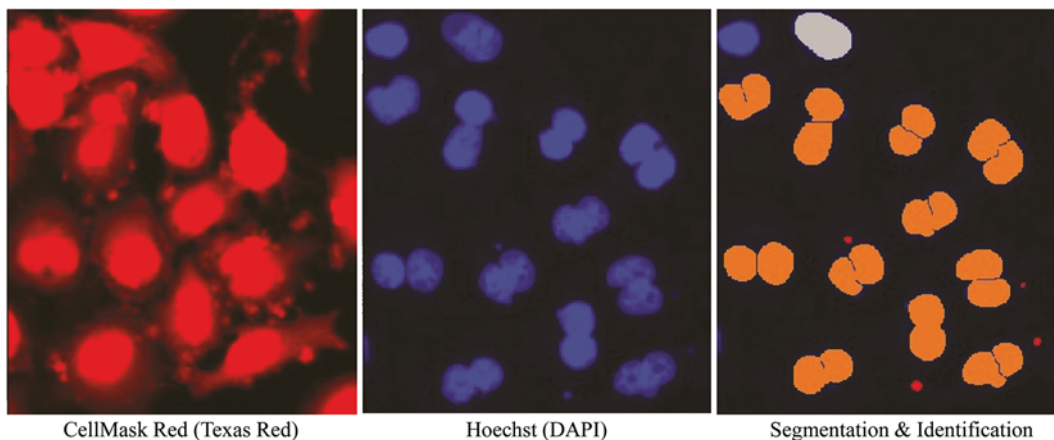
**Fig. 1** Segmentation in micronucleus analysis module correctly identifies the whole cell, main nucleus, and micronucleus represented by CellMask Red and Hoechst respectively
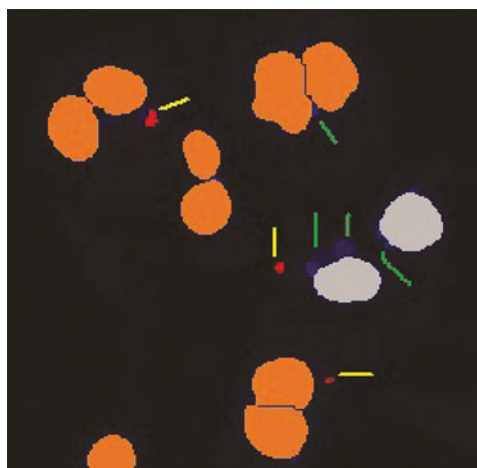


**Fig. 2** Identification of micronuclei according to OECD guidelines. Note the red micronuclei identified by a *yellow line*. Nuclear masses attached to the main nuclei; identified by *green lines* not identified as micronuclei

cells). The percentage of cytotoxicity (% Cytotoxicity) was defined as: $100-100\{(NDI_T-1)/(NDI_D-1)\}$ where, $NDI_T = NDI$ of treated cells; $NDI_D = NDI$ of DMSO control (Figs. 3 and 4).

## 4    Notes

1. Optimization of cytochalasin B treatment time is very crucial in getting healthy binucleated cells. Also while using vacuum to wash in sample preparation steps, make sure to use slowest possible force to avoid mechanical damage of the sample.

0 μM MMC                   0.6 μM MMC                   1.2 μM MMC
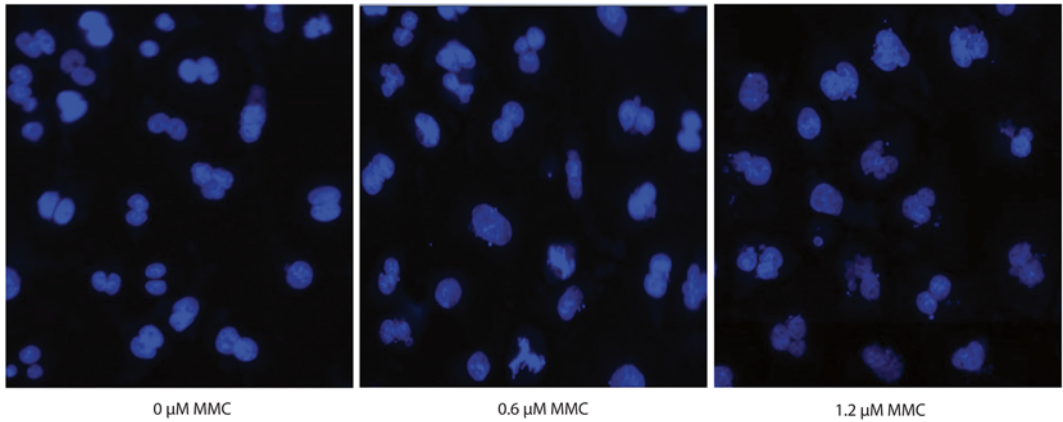
**Fig. 3** Induction of micronuclei in CHO-K1 cells. Representative images of nuclei (*blue*) with micronuclei. Note that without S9 treatment the micronuclei induction increases with higher MMC concentration
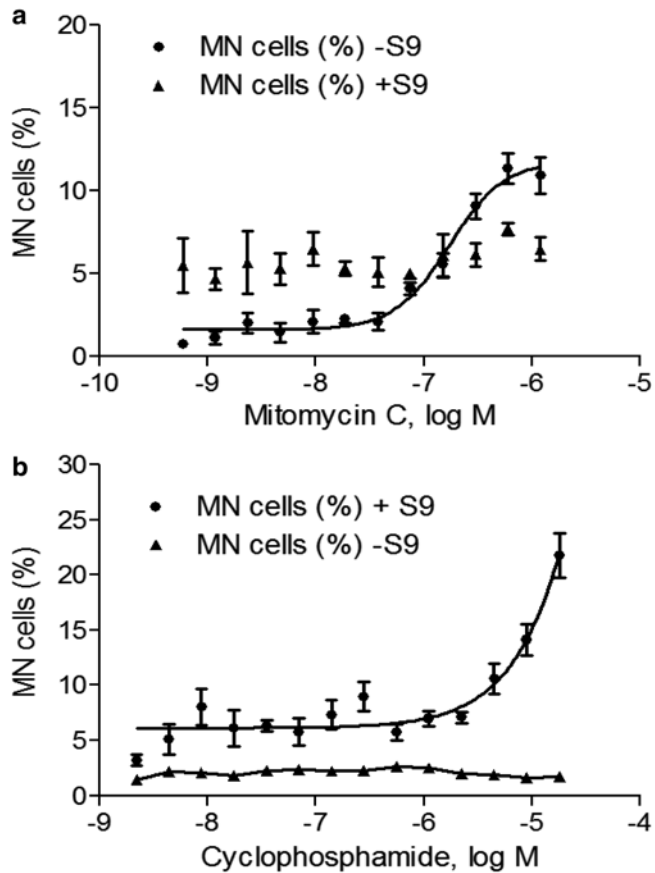


**Fig. 4** Induction of Micronuclei by MMC and CP in CHO-K1 cells. (**a**) MMC concentration response curves with and without S9 mix. (**b**) CP concentration response curves with and without S9 mix

## References

1. Fenech M (2008) The micronucleus assay determination of chromosomal level DNA damage. Methods Mol Biol 410:185–216

2. Evans HJ (1990) Cytogenetics: overview. Prog Clin Biol Res 340B:301–323

3. Dellarco VL, Mavournin KH, Tice RR (1985) Aneuploidy and health risk assessment: current status and future directions. Environ Mutagen 7(3):405–424

4. Guttenbach M, Schmid M (1994) Exclusion of specific human chromosomes into micronuclei by 5-azacytidine treatment of lymphocyte cultures. Exp Cell Res 211(1):127–132. doi:10.1006/excr.1994.1068, S0014-4827(84)71068-8 [pii]

5. Natarajan AT (2002) Chromosome aberrations: past, present and future. Mutat Res 504(1-2):3–16, doi:S0027510702000751 [pii]

6. Heddle JA (1973) A rapid in vivo test for chromosomal damage. Mutat Res 18(2):187–190, doi:0027-5107(73)90035-3 [pii]

7. Schmid W (1975) The micronucleus test. Mutat Res 31(1):9–15, doi:0165-1161(75)90058-8 [pii]

8. Miller B, Albertini S, Locher F, Thybaud V, Lorge E (1997) Comparative evaluation of the in vitro micronucleus test and the in vitro chromosome aberration test: industrial experience. Mutat Res 392(1-2):45–59, 187–208 doi:S0165-1218(97)00044-X [pii]

9. Fenech M, Morley A (1985) Solutions to the kinetic problem in the micronucleus assay. Cytobios 43(172–173):233–246

10. Fenech M, Morley AA (1985) Measurement of micronuclei in lymphocytes. Mutat Res 147(1-2):29–36, doi:0165-1161(85)90015-9 [pii]

11. Fenech M, Morley AA (1986) Cytokinesis-block micronucleus method in human lymphocytes: effect of in vivo ageing and low dose X-irradiation. Mutat Res 161(2):193–198, doi:0027-5107(86)90010-2 [pii]

12. Kirsch-Volders M (1997) Towards a validation of the micronucleus test. Mutat Res 392(1-2): 1–4

13. Fenech M (2007) Cytokinesis-block micronucleus cytome assay. Nat Protoc 2(5):1084–1104. doi:10.1038/nprot.2007.77, nprot.2007.77 [pii]

14. Beetstra S, Thomas P, Salisbury C, Turner J, Fenech M (2005) Folic acid deficiency increases chromosomal instability, chromosome 21 aneuploidy and sensitivity to radiation-induced micronuclei. Mutat Res 578(1-2):317–326. doi:10.1016/j.mrfmmm.2005.05.012, S0027-5107(05)00235-6 [pii]

15. Marzin D (1997) The position of the in vitro micronucleus test within the battery of screening for genotoxic potential determination and the regulatory guidelines. Mutat Res 392(1-2):175–181, doi:S0165-1218(97)00055-4 [pii]

16. OECD guidelines for the testing of chemicals, section 4 Test No. 487: in vitro mammalian cell micronucleus test (2014). OECD

17. De Sanctis S, De Amicis A, Di Cristofaro S, Franchini V, Regalbuto E, Mammana G, Lista F (2014) Cytokinesis-block micronucleus assay by manual and automated scoring: calibration curves and dose prediction. Health Phys 106(6):745–749. doi:10.1097/HP.0000000000000052, 00004032-201406000-00017 [pii]

18. Diaz D, Scott A, Carmichael P, Shi W, Costales C (2007) Evaluation of an automated in vitro micronucleus assay in CHO-K1 cells. Mutat Res 630(1-2):1–13. doi:10.1016/j.mrgentox.2007.02.006, S1383-5718(07)00049-6 [pii]

19. Westerink WM, Schirris TJ, Horbach GJ, Schoonen WG (2011) Development and validation of a high-content screening in vitro micronucleus assay in CHO-k1 and HepG2 cells. Mutat Res 724(1-2):7–21. doi:10.1016/j.mrgentox.2011.05.007, S1383-5718(11)00132-X [pii]

20. Tilmant K, Gerets HH, De Ron P, Cossu-Leguille C, Vasseur P, Dhalluin S, Atienzar FA (2013) The automated micronucleus assay for early assessment of genotoxicity in drug discovery. Mutat Res 751(1):1–11. doi:10.1016/j.mrgentox.2012.10.011, S1383-5718(12)00347-6 [pii]

21. Nishihara K, Huang R, Zhao J, Shahane SA, Witt KL, Smith-Roe SL, Tice RR, Takeda S, Xia M (2016) Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. Mutagenesis 31:69–81. doi:10.1093/mutage/gev055 [pii]

# Part II

**In Vivo Toxicological High-Thoughput Screening Methods**

# Chapter 10

# Better, Faster, Cheaper: Getting the Most Out of High-Throughput Screening with Zebrafish

## Lisa Truong, Michael T. Simonich, and Robert L. Tanguay

## Abstract

The field of toxicology is undergoing a vast change with high-throughput (HT) approaches that rapidly query huge swaths of chemico-structural space for bioactivity and hazard potential. Its practicality is due in large part to switching from high-cost, low-throughput mammalian models to faster and cheaper alternatives. We believe this is an improved approach because the immense breadth of the resulting data sets a foundation for predictive structure–activity-based toxicology. Moreover, rapidly uncovering structure-related bioactivity drives better decisions about where to commit resources to drill down to a mechanism, or pursue commercial leads. While hundreds of different in vitro toxicology assays can collectively serve as an alternative to mammalian animal model testing, far greater efficiency and ultimately more relevant data are obtained from the whole animal. The developmental zebrafish, with its well-documented advantages over many animal models, is now emerging as a true biosensor of chemical activity. Herein, we draw on nearly a decade of experience developing high-throughput toxicology screens in the developmental zebrafish to summarize the best practices in fulfilling the better, faster, cheaper goals. We include optimization and harmonization of dosing volume, exposure paradigms, chemical solubility, chorion status, experimental duration, endpoint definitions, and statistical analysis.

**Key words** Zebrafish, High-throughput screening, Toxicity testing

## 1 Introduction

The field of toxicology is challenged with tens of thousands of synthetic chemicals that are on the market and released to the environment with little or no safety information. These synthetic chemicals span an enormous range of physico-chemical properties, which makes it nearly impossible to address the backlog of unknown bioactivity using traditional toxicology approaches. Refocusing toxicity testing from high-cost, low-throughput mammalian models to alternative systems has been underway for some time. The key is to use a model, preferably a whole animal model, which is able to rapidly detect different modes of bioactivity and thus uncover molecular response pathways. Alternative model selection is guided by the 3Rs (Replacement, Reduction, and Refinement)

that minimize the use of animals to advance science. Current efforts to address this need are being led by multiple federal agencies such as the US Environmental Protection Agency—National Center for Computational Toxicology (NCCT), National Toxicology Program, National Center for Advancing Translational Sciences (NCATS), and US Food and Drug Administration (FDA). Together, the federal agencies developed a partnership termed "Tox21" to test a large set of compounds (>10,000) that previously had little or no toxicological data.

The Tox21 research initiative relied on more than a hundred different in vitro, high-throughput assays that were target-specific and mechanism-based to conduct toxicity pathway profiling followed by intensive computational approaches to interpret the findings. The first two phases of Tox21 provided the opportunity to optimize the HTS assays and test > 10,000 compounds from which it was learned that the biological coverage of the assays and relevance of the data were lacking. In phase 3, high-content assays were refined to include nonmammalian whole animal models, e.g., worms and zebrafish.

EPA-NCCT developed the ToxCast program to assess a large number of chemicals in a diverse set of in vitro assays [1]. The main difference from the Tox21 program was that ToxCast focused on evaluating a significantly smaller set of chemicals in a larger, biologically relevant assay space ToxCast Phase 1 consisted of ~300 well studied chemicals with existing toxicity information evaluated in ~600 in vitro HTS assays, which allowed for multidimensional signatures to predict animal toxicity using the traditional toxicity data to gauge accuracy. In the next phase, an additional ~700 chemicals with little to no traditional toxicity data were tested in the same assays. Phase 3 is currently ongoing, and consists of 1001 chemicals that will be evaluated in a refined set of assays identified in Phase 2 to provide insight on the mode of action of the chemicals.

There are several potential limitations in collecting in vitro data and then translating it to human hazard potential. The most obvious is that an in vitro assay queries only a small fraction of biological space, thus, very many different cell types and assays are used which increases the cost of chemical assessments. The net effect is that these large in vitro efforts may result in data with substantial uncertainty regarding translation to higher levels of organization. Whole animal models, that span embryonic development, express and present nearly all the potential cell types and gene products that are interacting in concert. This unique life stage offers an ideal time to determine if a test chemical has the inherent structure to interact with and perturb any of the carefully orchestrated signaling events necessary for normal development. If sufficient perturbations occur, the normal developmental plan will be disrupted resulting in a chemical-induced phenotype. The purpose

of screening in zebrafish thus is to more rapidly identify chemically induced phenotypes. To increase throughput, most of these phenotypes are simply visual assessments of major organ defects [2], but have also been extended to include motion-tracking behavioral assays in 96-well plate format [3, 4]. Using this general approach, developing zebrafish can therefore be considered as a sensitive biosensor of chemical activity.

## 2 Experimental Considerations for Optimal Zebrafish Use in High-Throughput Screens

As a biosensor amenable to high-throughput screens, the model is still only as good as the techniques and assays employed. In one effort at standardization, formulation of the OECD Fish Embryo Acute Toxicity Test (FET) [5] is aimed at acute toxicity of chemicals to embryonic zebrafish. Newly fertilized eggs are exposed to test chemicals for 96 h, with renewals every 24 h. Six concentrations and a control are tested in 24-well plates with 20 embryos per concentration, and every 24 h, four apical endpoints are recorded: (1) coagulation of fertilized embryos, (2) lack of somite formation, (3) lack of tailbud detachment from yolk sack, and (4) lack of a heartbeat. From these measured endpoints, the concentration associated with 50 % lethality ($LC_{50}$) is computed. However, this approach is laborious, has a limited number of endpoint and may be less sensitive because the chorion is intact for approximately the first half of the assay duration (48 h of development). Less apparent, but essential details for adoption of a standardized protocol are also lacking. These include standardization of the number of embryos per well, multi-well plate format and plastic type, as well as an optimized statistical analysis. The OECD standardization is thus a guideline for moderate throughput chemical screening in the developmental zebrafish.

The goal of high-throughput screening using a whole animal is to evaluate chemicals more rapidly and at a lower cost so the results can be utilized more immediately. To implement and harness the power of the zebrafish, there are design considerations that can help to make this model amenable to efficient HTS (Table 1). These considerations are detailed in the following sections.

### 2.1 Optimizing and Harmonizing Dose Volumes

The small size of the embryonic zebrafish allows for the use of multi-well plates. Each of the multi-well plates differs by the number of wells each plate holds (i.e., 6, 12, 24, 48, 96, 384) and subsequently the maximum volume per well. The number of wells per plate is an important factor to take into consideration as the volume per well for an individual well of a 6-well plate can hold up to 16.8 mL but less than 300 µL in 96-well plate [6]. Well diameter can be as large as 34.8 mm down to only 6.4 mm in a 96 well

**Table 1**

**Experimental factors to consider**

| Experimental design topics | Available option |
| --- | --- |
| Plate type | 6, 12, 24, 48, 96 well plates |
| Dose volumes | 300 μL to 16.8 mL |
| Exposure paradigms: chemical delivery | Liquid handler<br>Manual pipet<br>Digital dispensing |
| Exposure paradigm: solution renewal | Continuous or static |
| Chemical solubility solutions | Solvent selection<br>Sonication of samples |
| Chorion status | Chorion in place or dechorionate |
| Experimental duration | 96 or 120 h post fertilization |
| Endpoint definitions | Apical or detailed endpoints |
| Data management | Excel sheet or laboratory information management system |
| Statistical analysis | Binary data or rank scores for each endpoint |

format. Plate type and well format used for HTS could certainly influence the outcome of zebrafish behavioral assays that have a swimming activity readout. The OECD guidelines for a FET require the use of 24-well plates with at least 2 mL of chemical. But others have successfully reported using 96-well plates [2, 7] and with a robust behavioral component. The volume typically used in the 96-well plate exposures by our group is 100 μL which strikes a balance between minimal test chemical consumption and maximal water volume to support development and swim activity by 5 days post fertilization (dpf).

**2.2 Chorion Status** The chorion, an acellular matrix surrounding the embryo, can be removed enzymatically as early as 4 h post fertilization (hpf) [2]. When reared at 28 °C an embryo hatches out of its chorion between 48 and 72 hpf [8]. In our view, for HT screening, it is ideal to remove the chorions to maximize chemical bioavailability as the chorion can impact chemical partitioning and toxicity [9, 10] despite the observation that the chorion has pores which are approximately 0.6–0.7 μm [11]. The presence of the chorion therefore is a critical concern when evaluating the bioactivity of chemicals and its presence could result in an increased false negative rate.

**2.3 Exposure Paradigms** With embryonic zebrafish developing so rapidly, it is challenging to initiate exposures immediately after fertilization. It is also difficult to identify viable fertilized embryos until approximately 3–4 hpf thus, many assays begin their exposures starting at varying life

stages, but we have found that exposure commencement in the 6–8 hpf window with embryos dechorionated at 4 hpf is readily achievable in a HTS environment. The 6–8 hpf embryos are mature enough to withstand gentle handling while still at a sufficiently early life stage for assessments of important developmental events. Exposure commencements at 24 hpf are certainly less technically challenging, but are too late as many critical stages of tissue differentiation and primary organogenesis would be missed [12].

Once viable embryos are in the well, there are important considerations in determining the best way to deliver chemicals and to encourage solubility in the test medium.

### 2.3.1 Liquid Handler/ Manual Pipetting/Digital Dispensing

With advancements in liquid handling technology, test chemical delivery to the experimental chamber has evolved. The utility of manual pipetting has always been constrained by the need to work with relatively large volumes (>5 μL) for consistency in serial dilutions and dispensing into multi-well plates. Indeed, the need for serial dilution is solely a function of volume constraints with traditional pipetting, and serial diluting itself is an effective means of propagating error. Moreover, plastic (polypropylene) pipette tips are suspected to adsorb hydrophilic and low surface tension chemicals, further increasing uncertainty about how much compound was actually delivered to the test chamber [13]. Enter the age of digital dispensing platforms which use different approaches such as inkjet technology [14] or acoustic vibration [15] as the motive force to rapidly dispense single droplet streams of test solution. Droplet size is currently as small as 13 pL with one commercial platform [14, 16] and will almost certainly be driven lower by improving technology and growing market demand. The net result of sequentially transferring such small volumes of millimolar range solutions is that serial diluting is obviated. Reproducibility of chemical delivery dramatically increases while chemical sorption losses and pipettable volume excesses shrink. Because the dispensing is software-controlled, additional advantages such as complex mixtures and randomized plate layouts become trivial to execute, and a traceable record of all dispense events is automatically stored. Of course there are limitations. Entry level pricing is $25K, consumable costs are not trivial, and efficient mixing, taken for granted with manual pipetting of traditional volumes, requires additional steps to execute in the picoliter realm.

### 2.3.2 Continuous vs. Static Exposure

Static and continual renewal regimens have both emerged in the zebrafish chemical screening literature. When we consider that the embryonic zebrafish has a functioning liver by 48 hpf and a phase I and II metabolism similar to humans [17, 18], it is apparent that renewal regimen can profoundly influence the outcome. For instance, continual renewal reduces concerns about chemical liability over the typical 5 day experiment. But continuous renewal

maintains a high and possibly nonrepresentative exposure to the parent compound with the potential for false positives, and necessitates potentially confounding embryo handling to remove and replenish the test solution. There are tools available that facilitate the process such as 96-well plate inserts to lift all embryos at the same time into a new plate [7], but require large exposure volumes to ensure the entire embryo is covered, and when transferring embryos into a new solution tray, chemical carryover and cross contamination is a concern. The availability and cost of these inserts and plates somewhat reduce their practicality.

Static renewal has the advantages of minimizing test chemical consumption, manipulation and labor costs. A disadvantage is that, with little or no prior knowledge of the individual chemistries in a large library, false negatives will occur because the concentrations of some labile compounds may be reduced by the time their potential targets are developmentally expressed. There is a gray area between the pros and cons. For instance, zebrafish complete organogenesis within 48 hpf [8], so that the test chemical is potentially undergoing metabolism and degradation early in the experiment. Parent compound, metabolite and byproducts are thus players throughout the experiment. In a continual renewal, this complexity is removed daily, potentially missing toxicity associated, especially, with degradation. But one could similarly argue that a static exposure is biased toward, or even confounded with, the toxicity of metabolites and degradation products, obscuring the original goal of characterizing the parent toxicity. What we do know is that the chemical bioactivity associated with each regimen often differs. Our position is that the choice is a matter of throughput. Static exposure supports HTS while a continual renewal regimen supports moderate throughput screening.

**2.4 Chemical Solubility**

To increase solubility, chemicals are typically delivered in a biocompatible solvent such as dimethyl sulfoxide (DMSO) or ethanol. Embryonic zebrafish develop phenotypically normal to 5 dpf in ≤1 % ethanol or DMSO [19]. The most common solvent by far for HTS is DMSO, though its hygroscopic nature is especially troublesome because rapid absorption of water by DMSO can accelerate degradation and precipitation of test compounds [20]. Keeping compounds in DMSO dry is emerging as a major consideration around library storage and handling, and new commercial platforms to monitor sample hydration and more efficiently store samples under a dry atmosphere are readily searchable online. Once dispensed into an aqueous environment where the vehicle is only 1 % of the bath composition, maintaining solubility and hence, bioavailability of hydrophilic test chemicals can be a challenge. We have empirically determined that for PAHs, possibly a worst case solubility scenario, gentle mixing (235 rpm) of the exposure plate on an orbital mixer overnight routinely shifts the concentration

response downward by five to tenfold. This is relative to the same exposure mixed thoroughly for 15 s at 235 rpm, immediately after digital dispensing, then left undisturbed overnight. After confirming that the overnight mixing motion had no effect on zebrafish developmental morphology or behavioral endpoints, we have instituted the overnight mixing for all HTS exposures in our lab.

**2.5 Experimental Duration**

We have benchmarked much of the discussion so far to a 5 day (120 hpf) screen. The current OECD guideline for the zebrafish embryo test is to conduct the experiment until 96 hpf. At this stage, embryos have hatched, completed organogenesis, are metabolically highly active, and about 3.7 mm in length [8]. In many countries, the 96 hpf embryo is not considered a "living organism," and not regulated as an animal used for experimental and other scientific purposes in the European Union (EU). For these reasons, the FET test is terminated at 96 hpf. However, it is possible to go until 120 hpf where larval behavioral endpoints offer more information with which to detect chemical bioactivity. Visual observation of morphology changes is also easier because the larva is bigger. Thus, the heart and circulation are easily observed, and the brain, eye, snout, and jaw are more distinguished [8].

**2.6 Endpoint Definitions**

There are numerous developmental endpoints that can be adversely affected by chemical exposure in a HTS. We routinely assay 22 visual endpoints and 3 behavioral endpoints that we have determined provide a rich profile of chemical bioactivity and insights into potential mechanisms of toxicity [3]. The OECD guideline stipulates only four endpoints that primarily bin outcomes as either normal, dead, or abnormal. Endpoint scoring can range from simple presence/absence binary data to a scale of severity scores for each endpoint. Severity scores offer the advantage of tracking dose-response effects more accurately, but are inherently more variable across multiple scorers [7]. To circumvent this, one approach is to automate, as much as practical, the measurements of basic descriptors (e.g., length and width) [21] or to restrict the focus by screening in a transgenic reporter line for abnormalities in just the labeled tissue (e.g., vascular system: [22]). We have found that simple presence/absence data for as many endpoints as practical strikes the best balance among data quality, throughput and cost-efficiency [3, 4, 23, 24].

**2.7 Statistical Analysis**

The field of toxicology has established several standardized metrics of toxicity, such as the concentration that caused 50 % lethality in the embryos ($LC_{50}$), maximum concentration causing no mortality within the test period (NOEC), minimum concentration causing 100 % mortality ($LC_{100}$), and lowest concentration that causes a significant effect when compared to control (LOEC). The statistical analysis to calculate these readouts is by probit analysis, logistic

regression models, geometric means, and ANOVAs. In addition to these readouts, for HTS data, the lowest effect level (LEL) is utilized to describe the lowest level that induced any effect, which is similar to the LOEC, however, the method to compute this readout is by using a fisher's exact or binomial test [3] for each endpoint rather than overall. This should be applied to each endpoint due to the fact that the endpoints are highly correlated, which makes it difficult to discern the LOEC, but not the LEL. If more endpoints are used for high-throughput zebrafish screening, the current readouts established in the OECD guidelines are inappropriate because they do not account for the fact that as lethality increases as a function of concentration, the number of viable embryos to evaluate diminishes, prohibiting the use of logistic regression models, and probit analysis. The most appropriate statistical test is Fishers exact, and a binomial test for binary data. An alternative is a summation of the endpoints (including mortality), and fitting a regression model to the data to compute the OECD guideline readouts. A limitation to this aggregation method is determining what weight each endpoint should receive. Regardless of the statistical method, the current readouts used for zebrafish HTS can be optimized for greater relevance and efficiency to better handle the large volume of data.

**2.8  Data Management**

HTS generates "big data" that are essentially impossible to curate or share without electronic database storage and management through a lab information management system (LIMS). The system should track assay plates by barcodes permanently associated to plate layout, test chemicals, experimental dates and results, and other useful metadata such as hi-res digital images of the animals. A database enables immediate and nearly effortless checks on test chemical scheduling, data integrity and reproducibility. Moreover, the necessary organizational constraints of well-constructed database tables automatically ensures that third party downstream reduction and analysis tools can easily read and process the data.

## 3  Conclusions and Future Perspectives

As the use of zebrafish for HTS grows in popularity, the field will benefit greatly from continuing to embrace the developmental zebrafish as a vertebrate biosensor of chemical bioactivity while shedding the limited view that it is just fish model for developmental toxicity. High-throughput screens are still very young in this model, but have already, and will continue to give us great discoveries. The considerations and recommendations outlined here are primarily from a successful toxicological perspective, but they are equally amenable to therapeutics discovery. Going forward, embedding high-throughput transcriptomics into vertebrate HTS

will be the next huge change in this field. The technology is there, but the cost is still high enough that inclusion of transcriptomics is often a choice rather than a given. But the continually declining cost and increasing power of the technology may change that soon. Once these efforts are widely integrated, truly predictive toxicology will be a reality.

## References

1. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5–12. doi:10.1093/toxsci/kfl103, kfl103 [pii]

2. Truong L, Harper SL, Tanguay RL (2011) Evaluation of embryotoxicity using the zebrafish model. Methods Mol Biol 691:271–279. doi:10.1007/978-1-60761-849-2_16

3. Truong L, Reif DM, St Mary L, Geier MC, Truong HD, Tanguay RL (2014) Multidimensional in vivo hazard assessment using zebrafish. Toxicol Sci 137(1):212–233. doi:10.1093/toxsci/kft235

4. Reif DM, Truong L, Mandrell D, Marvel S, Zhang G, Tanguay RL (2015) High-throughput characterization of chemical-associated embryonic behavioral changes predicts teratogenic outcomes. Arch Toxicol. doi:10.1007/s00204-015-1554-1

5. OECD Test No. 236: Fish Embryo Acute Toxicity (FET) test. OECD

6. Corning (2006) Surface areas and recommended medium volumes for Corning cell culture vessels. Corning, New York; Corning Incorporated

7. Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, Sipes N, Kleinstreuer N, Knudsen T, Dix DJ, Reif DM (2012) Zebrafish developmental screening of the ToxCast Phase I chemical library. Reprod Toxicol 33(2):174–187. doi:10.1016/j.reprotox.2011.10.018

8. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203(3):253–310. doi:10.1002/aja.1002030302

9. Kim KT, Tanguay RL (2014) The role of chorion on toxicity of silver nanoparticles in the embryonic zebrafish assay. Environ Health Toxicol 29:e2014021. doi:10.5620/eht.e2014021

10. Panzica-Kelly JM, Zhang CX, Augustine-Rauch KA (2015) Optimization and performance assessment of the chorion-off [dechorinated] zebrafish developmental toxicity assay. Toxicol Sci 146(1):127–134. doi:10.1093/toxsci/kfv076

11. Lee KJ, Nallathamby PD, Browning LM, Osgood CJ, Xu XH (2007) In vivo imaging of transport and biocompatibility of single silver nanoparticles in early development of zebrafish embryos. ACS Nano 1(2):133–143. doi:10.1021/nn700048y

12. Westerfield M (2007) The Zebrafish book. a guide for the laboratory use of zebrafish (Danio rerio), 5th edn. University of Oregon Press, Eugene, OR

13. Fries E, Zarfl C (2012) Sorption of polycyclic aromatic hydrocarbons (PAHs) to low and high density polyethylene (PE). Environ Sci Pollut Res Int 19(4):1296–1304. doi:10.1007/s11356-011-0655-5

14. Jones R, Zheng W, McKew JC, Chen CZ (2013) An alternative direct compound dispensing method using the HP D300 digital dispenser. J Lab Automat 18(5):367–374. doi:10.1177/2211068213491094

15. Hadimioglu B, Stearns R, Ellson R (2015) Moving liquids with sound: the physics of acoustic droplet ejection for robust laboratory automation in life sciences. J Lab Automat. doi:10.1177/2211068215615096

16. HP (2012) Reinventing drug titration for IC50 determinations. Corvallis, Oregon: Tecan Publisher. In: HP D300 Digital Dispenser—White Paper

17. Goldstone J, McArthur A, Kubota A, Zanette J, Parente T, Jonsson M, Nelson D, Stegeman J (2010) Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. BMC Genomics 11(1):643

18. Vliegenthart AD, Tucker CS, Del Pozo J, Dear JW (2014) Zebrafish as model organisms for studying drug-induced liver injury. Br J Clin Pharmacol 78(6):1217–1227. doi:10.1111/bcp.12408

19. Chen TH, Wang YH, Wu YH (2011) Developmental exposures to ethanol or

dimethylsulfoxide at low concentrations alter locomotor activity in larval zebrafish: implications for behavioral toxicity bioassays. Aquat Toxicol 102(3-4):162–166. doi:10.1016/j.aquatox.2011.01.010

20. Ellson R, Stearns R, Mutz M, Brown C, Browning B, Harris D, Qureshi S, Shieh J, Wold D (2005) In situ DMSO hydration measurements of HTS compound libraries. Comb Chem High Throughput Screen 8(6):489–498

21. George S, Xia T, Rallo R, Zhao Y, Ji Z, Lin S, Wang X, Zhang H, France B, Schoenfeld D, Damoiseaux R, Liu R, Bradley KA, Cohen Y, Nel AE (2011) Use of a high-throughput screening approach coupled with *in vivo* zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. ACS Nano 5(3):1805–1817. doi:10.1021/nn102734s

22. Yozzo KL, Isales GM, Raftery TD, Volz DC (2013) High-content screening assay for identification of chemicals impacting cardiovascular function in zebrafish embryos. Environ Sci Technol 47(19):11302–11310. doi:10.1021/es403360y

23. Kim KT, Truong L, Wehmas L, Tanguay RL (2013) Silver nanoparticle toxicity in the embryonic zebrafish is governed by particle dispersion and ionic environment. Nanotechnology 24(11):115101. doi:10.1088/0957-4484/24/11/115101

24. Knecht AL, Goodale BC, Truong L, Simonich MT, Swanson AJ, Matzke MM, Anderson KA, Waters KM, Tanguay RL (2013) Comparative developmental toxicity of environmentally relevant oxygenated PAHs. Toxicol Appl Pharmacol 271(2):266–275. doi:10.1016/j.taap.2013.05.006

# Fast Functional Germline and Epigenetic Assays in the Nematode *Caenorhabditis elegans*

## Zachary Lundby, Jessica Camacho, and Patrick Allard

## Abstract

Germ cells are unique in their ability to transfer traits and genetic information from one generation to the next. The proper development and integrity of their genome are therefore of utmost importance for the health of organisms and survival of species. Many features of mammalian germ cells, including their long development span and difficulty of access, present challenges for their study in the context of toxicity assays. In light of these barriers, the model system *Caenorhabditis elegans* shows great potential given its ease of manipulation and genetic tractability which can be easily adapted for high-throughput analysis. In this chapter, we discuss the advantages of examining germ cell processes in *C. elegans*, and describe three functional germline assays for the examination of chemical impact on germline maintenance and function including assays probing germ cell differentiation, germline apoptosis, and germline epigenetic regulation.

**Key words** *C. elegans*, Germline, Meiosis, Toxicity, Apoptosis

## 1 Introduction

The nematode *Caenorhabditis elegans* is arguably one of the most valuable contemporary genetic model organisms. Discoveries in the free-living roundworm include the identification of the apoptotic cascade, RNA interference, and microRNAs to only name a few [1]. *C. elegans* was first isolated and described by Emile Maupas from Algerian soil in the early 1900s but it was not until the 1940s that the organism was studied in a laboratory setting by Victor Nigon and Ellsworth Dougherty [2]. Soon after, Sydney Brenner focused its attention on the nematode, rightly arguing that its ease of maintenance and tractability would enlighten the fields of genetics and developmental biology [3]. It is precisely these features, as well as many others described below, that make *C. elegans* a powerful model system not only for the aforementioned fields but also for the field of toxicology [4].

*C. elegans* worms are easily and cheaply grown and maintained in the laboratory setting. The most common method of *C. elegans*

culture is on regular plastic petri dishes filled with nematode growth medium (NGM) [5]. The agar media is then covered in a layer of bacteria, most commonly OP50 or HB101 strains, which will serve as food source for the lifetime of the worms. Alternatively, the worms can also be grown in liquid media composed of a salt buffer and bacteria. Either growth media allows for scalability and versatility in worm culture depending on the desired application. Worms can be grown in flasks, tubes, or plates of all sizes including 96-well and 384-well plates, a useful format for high-throughput screening [6, 7].

In these media, and under normal temperature conditions (15–25 °C), the worms will develop as embryos, hatch, and then proceed through four larval stages to reach adulthood in 3–4 days. Once the worms are adult, they will lay between 200 and 300 embryos for another 4 days after which the worms will cease to reproduce. Importantly, the worms are self-fertilizing hermaphrodites. They make sperm as they transition from their last larval stage (L4) into adulthood at which point, their germline will only produce oocytes. As these oocytes pass through the spermatheca, located next to the uterus, they are fertilized and initiate embryonic development within the confines of the uterus before being laid into the media (Fig. 1) [8].
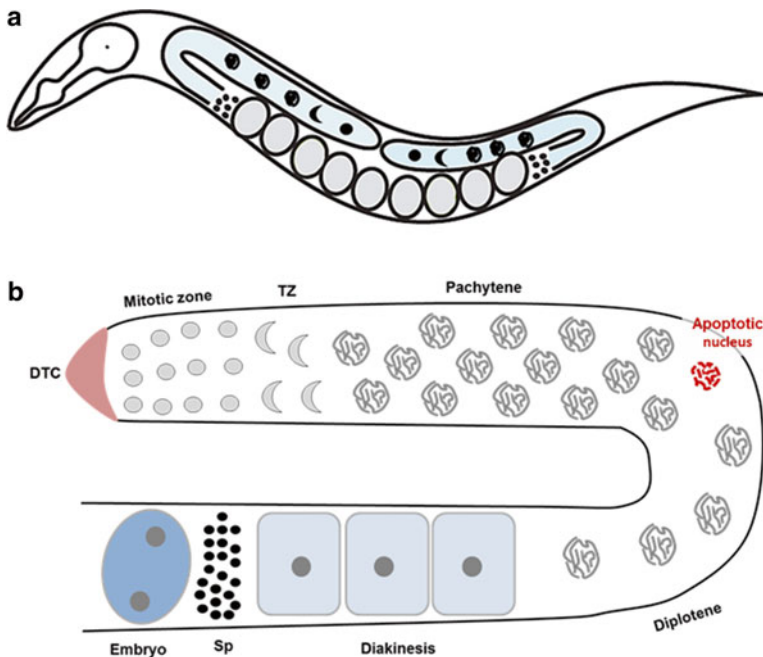


**Fig. 1** Worm and germline morphological features. (**a**) *Caenorhabditis elegans* nematodes are round worms displaying two gonadal arms (*lightly shaded in blue*) that open into one common uterus where embryos develop before being deposited into the media through the vulva. (**b**) Schematized representation of one gonadal arm. The distal tip cell emits a Notch ligand that maintains germ cells in mitosis. Once far enough away from the ligand, the germ cells will enter and progress through the different phases of prophase I of meiosis. Apoptotic nuclei, an outcome of defective meiotic recombination or synapsis, are visible in late pachytene. TZ = Transition zone (leptotene and zygotene). *Sp* = spermatheca

In the assays described in this chapter (Fig. 2), several other features of *C. elegans* are mobilized. In particular, assessment of toxicity is simplified by the transparency of the worm's cuticle, allowing for the observation of many endpoints without dissection of the worm itself. This is particularly useful for the monitoring of germ cell differentiation in the gonad of *C. elegans*. In the nematode, germline nuclei differentiate continuously from the distal mitotic zone, through transition zone (leptotene and zygotene of prophase I), pachytene, diplotene, and diakinesis of prophase I. Each stage of meiotic differentiation harbors specific nuclear morphology which can be appreciated without the need for dissection by simple fixation of whole worms and staining by DAPI [9]. Thus, the kinetics of meiotic differentiation as well as germline nuclear loss can be quickly monitored in *C. elegans* following exposure.

The other two assays rely on the number of genetic strains that are available in *C. elegans* available from the *Caenorhabditis* Genetics Center (CGC). We are making use of two specific strains: the first carries a transgene coding for CED-1::GFP which surrounds engulfed apoptotic nuclei. The expression of CED-1::GFP therefore easily labels apoptotic nuclei in the germline of *C. elegans* [10]. This assay is particularly informative as germline apoptosis is a hallmark of defective processes of meiotic differentiation including defective chromosomal synapsis and meiotic recombination. Induction of germline apoptosis is therefore used as a proxy for defective germline function [11–13].

Similarly, the third assay relies on a GFP-based reporter which is specifically silenced in the germline of *C. elegans* but not in its somatic tissues [14–17]. Furthermore, this silencing is controlled through epigenetic means involving the recruitment of repressive histone modifications and the loss of active marks [16, 17]. We describe here the application of this reporter approach for the screening of chemicals with epigenetic dysregulation effects.

## 2 Materials

1. M9 (stored at ambient temperature).
   3 g KH2PO$_4$, 6 g Na2HPO$_4$, 5 g NaCl, 1 mL 1 M MgSO$_4$, H$_2$O to 1 L.

2. NMG media plates.
   3 g NaCl, 2.5 g Bactopepton, 20 g Sigma Agar, add H$_2$O to 1 L, 1 mL 5 mg/mL cholesterol, 1 mL 1 M CaCl$_2$, 1 mL 1 M MgSO$_4$, autoclave for 40 min, 25 mL 1 M pH6 KH$_2$PO$_4$.

3. Positively charged slides.

4. Cover slips.

5. 1.5 mL Microcentrifuge tubes.
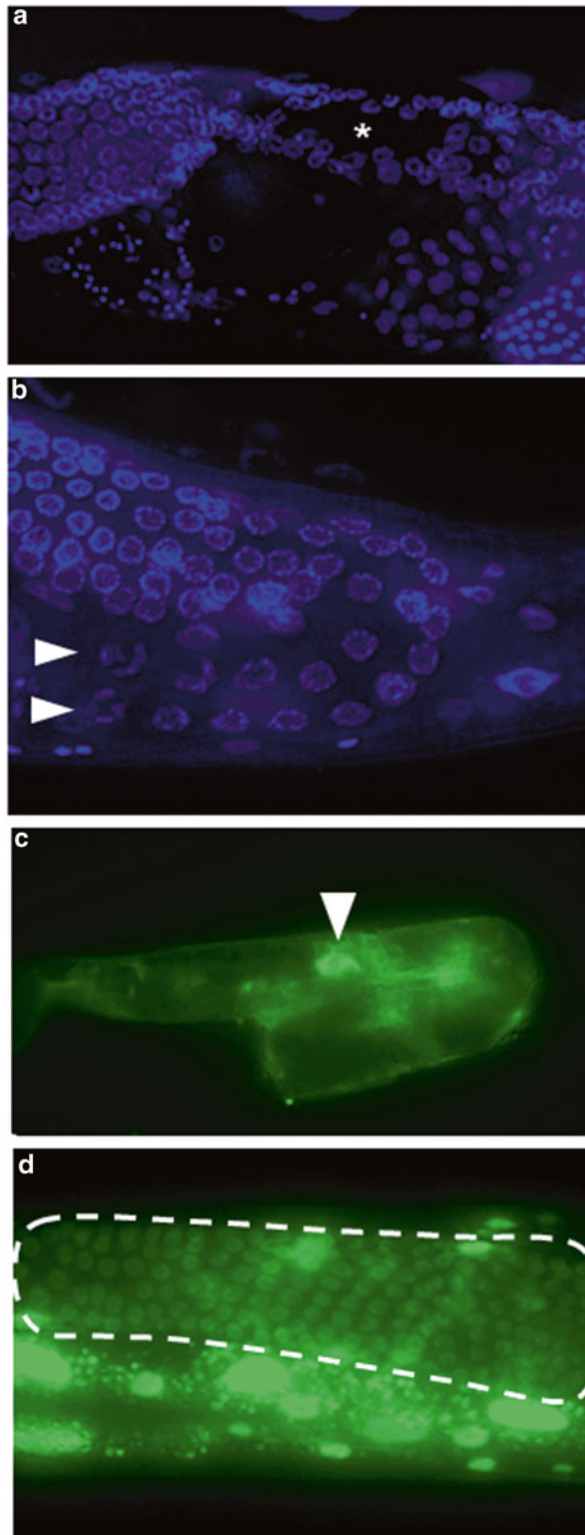
**Fig. 2** Several tools to examine germline toxicity. (**a**) Example of DAPI staining of the worm germline showing a large gap. (**b**) Example of defective germline morphology with two rows of diplotene-stage nuclei instead of the normal sole row of nuclei. (**c**) Use of the CED-1::GFP apoptotic reporter showing an apoptotic nucleus within the germline. (**d**) Epigenetic assay. The worm shown here was exposed for 48 h to the HDAC inhibitor valproic acid at 100 µM. Dashed line surrounds the germline nuclei expressing GFP from the normally epigenetically repressed let-858::gfp transgene

6. Fluorescence microscope.

7. Positive controls (stored at –20 °C unless otherwise noted).

    (a) BPA—CAS 80-05-7.

    (b) Vinclozolin—CAS 50471-44-8.

    (c) Sodium butyrate—CAS 156-54-7.

8. Dimethyl sulfoxide—CAS 67-68-5 (stored at ambient temperature).

9. 15 mL Conical vials.

10. Glass Pasteur pipette.

11. Pipette 1000, 10, 1 μL.

12. Worm pick.

13. NL2507 worms (CGC).

14. Ced-1::GFP worms (CGC).

15. Wild-type worms.

16. 60 mm vented plates.

17. OP50 bacteria.

18. Immersion oil.

19. DAPI stain (Sigma-Aldrich).

20. Carnoy's solution.

21. Kimwipes.

22. Nail polish.

Note that this is but a brief summary of the steps for each of our methods of analysis. For more detailed information, please visit Parodi DA, Damoiseaux R, Allard P(2015) Comprehensive assessment of germline chemical toxicity using the nematode *Caenorhabditis elegans.* Journal of visualized experiments: JoVE 96 available in PubMed #25741987.

# 3    Methods

## 3.1    Procedure for Caenorhabditis elegans Growth and Exposure

This procedure focuses on establishing a viable population of worms that can be used in any of the experiments listed later in this chapter. Following this procedure, one will possess chemically exposed worms at the proper larval stage to immediately begin the epigenetic, DAPI, or apoptotic analysis.

1. Dilute all chemicals of interest to 100 mM with DMSO directly in stock bottle and transfer to 1.5 mL microcentrifuge tubes.

2. Store at –20 °C.

3. Synchronize worms by bleaching and plate worms until they reach L4 stage.

4. Label three 15 mL conical vials one through three.

5. Fill tubes 1 and 2 with 7 mL of M9.

6. Wash 3–4 plates of worms using 1 mL of M9.

7. Dispense on each plate twice to collect worms and add them to the first conical vial.

8. Allow L4 worms to settle to the bottom of the first conical vial.

9. During this process, start on the following:

    (a) Thaw out chemical stocks by placing them in the fume hood.

    (b) Add 50 μL of OP50 bacteria to the third conical vial per each exposure group.

    (c) Add 500 μL of M9 to the third conical vial per each exposure group.

10. Once the L4 worms have settled, extract them with a glass Pasteur pipette and add them to the second conical vial.

11. After they have settled once more, transfer them to the third and final conical vial using the glass Pasteur pipette once more.

12. Count the worms from the third conical vial.

13. Count 5 μL worth of worms on a cover slip and multiply by 100.

14. This figure should not exceed 400 worms.

15. Label the 1.5 mL tubes with the appropriate markings to designate the chemicals that will be used.

16. Add 500 μL of the solution from the third conical vial to each of the 1.5 mL microcentrifuge tubes.

17. Transfer the 1.5 mL tubes to the fume hood.

18. Add 0.5 μL of chemical of interest to each tube to reach a final concentration of 100 μM.

19. Dispense 0.5 μL of DMSO (0.1 % final) directly from the stock bottle as the negative control into the appropriately marked 1.5 mL microcentrifuge tube.

20. Dispense 0.5 μL of any of the aforementioned positive controls into the appropriately marked 1.5 mL tubes.

21. Vortex each of the 1.5 mL tubes upon completion to ensure complete distribution.

22. Place the 1.5 mL tubes in a rotator for 24 h at 20 °C.

23. Note that rotation is necessary for proper growth.

24. At the end of 24 h, remove the 1.5 mL tubes from the rotator for 5 min to aerate.

25. Add an additional 50 μL of OP50 bacteria to each tube.

26. Place back in rotator for another 24 h.

27. At the end of the final 24 h, label petri dishes corresponding to each of the chemicals used.

28. Retrieve the 1.5 mL tubes and allow the worms to settle to the bottom.

29. Extract the worm pellet and plate on the proper petri dishes.

30. Allow the worms to settle for 1–2 h on the petri dishes at 20 °C.

Note that after this step each of the different types of analysis diverges. Be aware that a 24 h exposure does not require any of the steps after **step 24**.

*3.2 DAPI Analysis*

Chemical disruption can result in morphological errors in the germline of the worm that may manifest itself in a variety of ways [18, 19]. The goal of this type of analysis is to discover what, if any, morphological errors have occurred.

1. After placing N2 worms in 10 μL of M9 on positively charged slide, use a Kimwipe to absorb as much of the M9 liquid as possible while being careful not to extract the worms as well.

2. Add Carnoy's solution dropwise until the worms have been completely dehydrated.

3. Rehydrate the worms with a humidifying apparatus by suspending the slide between two objects with water underneath.

4. Add M9 directly to the dehydrated worms and cover the apparatus for 1 h.

5. At the end of the hour, absorb the M9 once more with Kimwipes.

6. Add a single drop of DAPI stain and distribute it evenly with a worm pick.

7. Add a cover slip over the solution and seal the edges with nail polish.

8. Store at 0 °C when not being examined.

9. Place slide under Nikon H600L microscope and identify worms at 10× with bright-field illumination.

10. Add immersion oil after the worms have been identified.

11. Switch to the software package and while retaining focus on the worm of interest, rotate to the 100× objective lens.

12. Additional magnification may be necessary to see the germline clearly; our lab recommends 400x using the "4×" option on the software package.

13. The nuclei in the germline will be plainly visible. Gaps in nuclei, uneven distribution of nuclei, extension of the mitotic or transition zones, and other gross errors in morphology are all evidence of toxicity. Areas of apoptosis $n > 5$ are "major" errors while areas of apoptosis $n < 5$ are "minor" errors.

**3.3 Apoptotic Analysis**

The goal of this type of analysis is to look specifically for cell death in the germline of the worm. Apoptotic cells are marked by the surrounding expression of the engulfment marker CED-1::GFP transgene which will be visible through the microscope [10]. While apoptosis is a process that occurs even in a healthy organism, a number of apoptotic nuclei exceeding the baseline can be used as an indicator of toxicity of the chemical in question.

1. Select for the CED-1::GFP worms with the best wild-type movement.

2. Following the same steps as the "epigenetic screen," examine the germline looking specifically for expression that appears as bright circles or a network of circles.

3. Distribute worms evenly across the liquid and add a cover slip for analysis.

4. Using the Nikon H600L microscope, identify worms at 10x under bright-field illumination.

5. Switch to FITC illumination and deactivate the standard microscope light source.

6. Nuclei positive for apoptosis will be surrounded by a GFP-positive ring (*see* Fig. 2).

**3.4 Epigenetic Analysis**

The goal of this type of analysis is to uncover changes in epigenetic expression, specifically epigenetic desilencing, resulting from chemical exposure. Such desilencing is revealed by the de-repression of a repetitive GFP transgene via the removal of repressive histone marks and the addition of activating histone marks [16, 17]. Furthermore, de-repression of the germline in the P0 generation has the possibility of being transmitted to successive generations.

1. Select NL2507 worms with the best rolling phenotype from the aforementioned petri dishes and add to 10 μL of M9 placed directly in the middle of a positively charged slide.

2. Distribute worms evenly across the liquid and add a cover slip for analysis.

3. Using the Nikon H600L microscope, identify worms at 10x under bright-field illumination.

4. Switch to FITC illumination and deactivate the standard microscope light source.

5. It may also be helpful to work in a darkroom for better identification.

6. Somatic cells will automatically fluoresce but the germline is the area of interest.

7. Redirect the image from the microscope lens to the software program.

8. If nothing is visible in the germline, it is negative for expression.

9. If individual nuclei are visible, it is positive for expression.

10. Expression should be plainly evident without magnifying further or adjusting the contrast. Please see attached images as a reference (Fig. 2).

## 4    Conclusion

The complementary of the three assays described above allows to obtain a broad picture of toxicity elicited by chemical exposure in *C. elegans*. Mechanistic follow-ups are needed to examine the germline pathways perturbed by the exposure. This can be easily performed by examining the resistance or sensitivity of mutants involved in germline processes. The use of the nematode as a reproductive model therefore offers the possibility to quickly reveal mechanisms of toxicity that may have taken much longer in other species.

## References

1. Macrae R (2014) On the shoulders of worms. Trends Genet 30(11):475–475

2. de Chadarevian S (1998) Of worms and programmes: Caenorhabditis elegans and the study of development. Stud Hist Phil Biol Biomed Sci 29(1):81–105

3. Brenner S (2009) In the beginning was the worm …. Genetics 182(2):413–415

4. Leung MC, Williams PL, Benedetto A, Au C, Helmcke KJ, Aschner M, Meyer JN (2008) Caenorhabditis elegans: an emerging model in biomedical and environmental toxicology. Toxicol Sci 106(1):5–28

5. Brenner S (1974) The genetics of Caenorhabditis elegans. Genetics 77(1):71–94

6. Parodi DA, Damoiseaux R, Allard P (2015) Comprehensive assessment of germline chemical toxicity using the nematode Caenorhabditis elegans. J Vis Exp 96

7. Boyd WA, McBride SJ, Rice JR, Snyder DW, Freedman JH (2010) A high-throughput method for assessing chemical toxicity using a Caenorhabditis elegans reproduction assay. Toxicol Appl Pharmacol 245(2):153–159

8. Greenstein D (2005) Control of oocyte meiotic maturation and fertilization. WormBook 1-12

9. Colaiacovo MP (2006) The many facets of SC function during C. elegans meiosis. Chromosoma 115(3):195–211

10. Zhou Z, Hartwieg E, Horvitz HR (2001) CED-1 is a transmembrane receptor that mediates cell corpse engulfment in C. elegans. Cell 104(1):43–56

11. Bhalla N, Dernburg AF (2005) A conserved checkpoint monitors meiotic chromosome synapsis in Caenorhabditis elegans. Science 310(5754):1683–1686

12. Gartner A, Boag PR, Blackwell TK (2008) Germline survival and apoptosis. WormBook 1-20

13. Gartner A, MacQueen AJ, Villeneuve AM (2004) Methods for analyzing checkpoint responses in Caenorhabditis elegans. Methods Mol Biol 280:257–274, doi: 1-59259-788-2:257

14. Kelly WG (2014) Transgenerational epigenetics in the germline cycle of Caenorhabditis elegans. Epigenetics Chromatin 7(1):6

15. Li T, Kelly WG (2011) A role for Set1/MLL-related components in epigenetic regulation of the Caenorhabditis elegans germ line. PLoS Genet 7(3), e1001349

16. Schaner CE, Kelly WG (2006) Germline chromatin. WormBook 1–14

17. Kelly WG, Xu S, Montgomery MK, Fire A (1997) Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics 146(1): 227–238

18. Allard P, Kleinstreuer NC, Knudsen TB, Colaiacovo MP (2013) A C. elegans screening platform for the rapid assessment of chemical disruption of germline function. Environ Health Perspect 121(6):717–724

19. Parodi DA, Sjarif J, Chen Y, Allard P (2015) Reproductive toxicity and meiotic dysfunction following exposure to the pesticides Maneb, Diazinon and Fenarimol. Toxicol Res 4(3):645–654

# Part III

**In Silico High-Thoughput Screening Toxcity Data Analysis**

# Chapter 12

## A Quantitative High-Throughput Screening Data Analysis Pipeline for Activity Profiling

### Ruili Huang

### Abstract

The US Tox21 program has developed in vitro assays to test large collections of environmental chemicals in a quantitative high-throughput screening (qHTS) format, using triplicate 15-dose titrations to generate over 50 million data points to date. Counter screens are also employed to minimize interferences from non-target-specific assay artifacts, such as compound auto fluorescence and cytotoxicity. New data analysis approaches are needed to integrate these data and characterize the activities observed from these assays. Here, we describe a complete analysis pipeline that evaluates these qHTS data for technical quality in terms of signal reproducibility. We integrate signals from repeated assay runs, primary readouts, and counter screens to produce a final call on on-target compound activity.

**Key words** HTS, Concentration response, In vitro assay, Activity profile, Tox21

## 1 Introduction

The US Tox21 program [1–4], a collaboration among the National Institute of Environmental Health Sciences (NIEHS)/National Toxicology Program (NTP), the US Environmental Protection Agency's (EPA) National Center for Computational Toxicology (NCCT), the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (NCATS), and the US Food and Drug Administration (FDA), is aimed at developing alternative testing methods that can quickly and efficiently assess the toxic potential of tens of thousands of environmental chemicals. Working toward this goal, the Tox21 program has successfully developed various cell-based assays to serve as in vitro models for toxicity assessment [4–7]. These assays have been miniaturized and validated in a 1536-well plate format at NCATS for quantitative high-throughput screening (qHTS) [8]. These assays are currently being screened against a collection of ~10,000 compounds (Tox21 10 K) composed of environmental chemicals and approved drugs as triplicate 15-dose titrations, generating over 50 million data points to date [9–11].

The use of qHTS to produce high quality and biologically relevant data is critical in correlation to in vivo activity, low dose extrapolation, and risk assessment. However, as with any technology, these assays are not immune to noise or artifacts that may interfere with the true biological activity. When a signal is observed in an assay, it is important to be able to distinguish a true biological effect from an artifact. In addition to "noise" and experimental variations, common artifacts found with fluorescence or luminescence-based reporter assays, such as those employed by Tox21, include compound auto fluorescence [12], interference with the assay reporter gene [13], and cytotoxicity [6]. Compound auto fluorescence often interferes with agonist mode assays, in which an increase in signal indicates activity. Compound interaction with the assay reporter gene itself could be mistaken for either agonist activity, when the compound activates the reporter gene, or antagonist activity, when the compound inhibits the reporter gene. Luciferase and β-lactamase reporters are commonly used in Tox21 and other HTS assays. Finally, cell-based antagonist mode assays are often confounded with cytotoxicity interference because both cell death and inhibition of the target of interest result in a decrease in assay signal.

To minimize compound or assay technology-dependent artifacts, all compounds in the Tox21 10 K library are tested for auto fluorescence at wavelengths used for assay readouts and for luciferase activity. In addition, each assay is multiplexed with cell viability measurements to identify cytotoxicity interference. The challenge is then to devise methods to (1) evaluate these qHTS data for technical quality, e.g., signal reproducibility, and (2) integrate signals from repeated assay runs, primary readouts and counter screens to produce a final call on on-target compound activity. Here, we describe a complete qHTS data analysis pipeline developed at NCATS that begins with plate level raw data processing, followed by concentration response curve fitting and classification, data reproducibility evaluation, and assignment of activity outcomes to compounds through integration of data from multiple readouts and counter screens. This approach has been applied to all the qHTS data generated from the Tox21 assay validation runs and 10 K screens, and can be adapted to analyze other qHTS data generated in a similar fashion.

## 2    qHTS Data Pipeline

### 2.1    Plate Level Data Processing, Curve Fitting and Classification

During the execution of the screen, quality metrics, such as CV, S/B, and Z-factor [14], are calculated using raw fluorescence or luminescence reads from each plate to monitor gross assay performance. These metrics are recorded for each plate and examined. "Failed plates," identified by abnormally poor values, are inspected visually and, if necessary, excluded from further data analysis. Upon completion of an assay run, raw plate reads for each titration point are first normalized relative to the positive control compound (100% for agonist mode and -100% for antagonist mode assays) and DMSO-only wells (0%) placed in the first four columns of

each plate as follows: % Activity $= ((Vcompound - V_{DMSO})/(Vpos - V_{DMSO})) \times 100$, where Vcompound denotes the compound well values, Vpos denotes the median value of the positive control wells, and $V_{DMSO}$ denotes the median values of the DMSO-only wells, and then corrected using compound-free control plates (i.e., DMSO-only plates) at the beginning and end of the compound plate stack to remove background patterns and subtle abnormalities such as tip effects or blotting from cell dispenses [15].

Corrected plate data are pivoted to form concentration–response series, which are subsequently fit to a four-parameter Hill equation [16] yielding concentrations of half-maximal activity (AC50) and maximal response (efficacy) values [17]. Concentration-response curves are designated as Class 1–4 based on efficacy, the number of data points observed above background activity, and the quality of fit [18]. Curve classes are heuristic measures of data confidence. The qHTS curve classification scheme has been recently amended to better suit the needs of toxicology research (Table 1) [6]. The most problematic concentration responses are automatically assigned curve class 5 based on considerations like the direction of activity (observing alternately both increases and decreases in signal over a short concentration range) and unusually large signal at low sample concentrations (activity at zero concentration is estimated to be >3SD of control) [6]. Class 5 curves and other cases in which an inconsistency between the highest compound activity and the curve class assigned is identified, such as assigning a compound with

**Table 1**
**Amended qHTS curve classification**

| Curve class | Description | Efficacy | *p*-value[a] | Asymptotes | Inflection |
|---|---|---|---|---|---|
| 1.1 | Complete curve | >6SD[b] | <0.05 | 2 | Yes |
| 1.2 | Complete curve | ≤6SD; >3SD | <0.05 | 2 | Yes |
| 1.3 | Complete curve | >6SD | ≥0.05 | 2 | Yes |
| 1.4 | Complete curve | ≤6SD; >3SD | ≥0.05 | 2 | Yes |
| 2.1 | Incomplete curve | >6SD | <0.05 | 1 | Yes |
| 2.2 | Incomplete curve | ≤6SD; >3SD | <0.05 | 1 | Yes |
| 2.3 | Incomplete curve | >6SD | ≥0.05 | 1 | Yes |
| 2.4 | Incomplete curve | ≤6SD; >3SD | ≥0.05 | 1 | Yes |
| 3 | Single point activity | >3SD | NA | 1 | No |
| 4 | Inactive | ≤3SD | ≥0.05 | 0 | No |
| 5[c] | Inconclusive | NA | NA | NA | NA |

[a]*p*-value is derived from a F-test that measures the quality of curve fit
[b]SD is the standard deviation of sample activities at the lowest tested concentration and values of the DMSO control wells
[c]Class 5 is a special class for samples with activity at zero concentration (zero activity; extrapolated) exceeding 6SD or with zero activity > 3SD and the difference between the maximal change in activity observed in the tested concentration range and zero activity is <3SD

a positive response a negative curve class, are manually inspected to correct the curve class, if necessary. Adjustment is normally done by masking or by unmasking data points improperly masked by the automated curve fitting process to adjust the curve fit (Fig. 1). To facilitate analysis and activity profiling, each curve class is further combined with an efficacy cutoff and converted to a numerical curve rank such that more potent and efficacious compounds with higher quality curves are assigned a higher rank (Table 2). Curve ranks should be viewed as a numerical measure of compound activity.
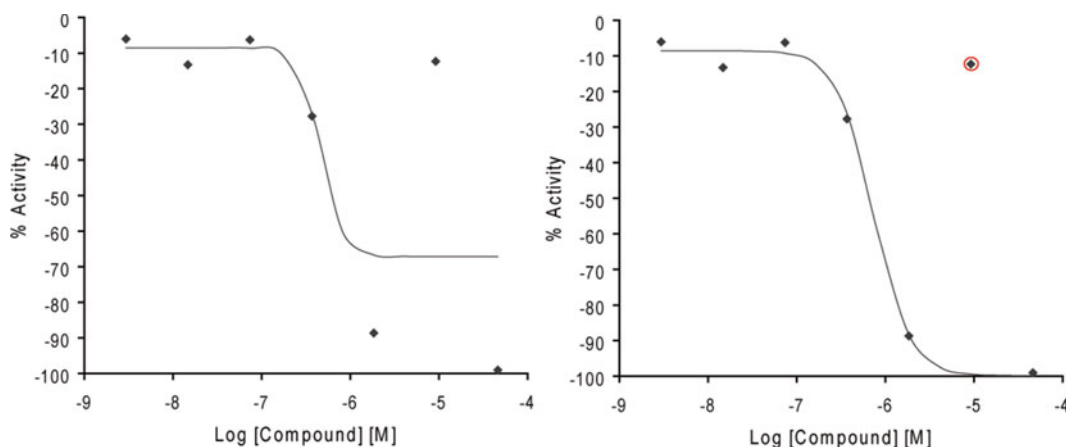


**Fig. 1** Example of outlier masking. Curves are manually inspected to mask or unmask data points improperly masked by the automated curve fitting process to adjust the curve fit when necessary

**Table 2**
**Definition of curve rank as a numeric measure of compound activity**

| Curve class | Efficacy | Curve rank | Activity category |
|---|---|---|---|
| 1.1 | | 9 | Agonist |
| 1.2 | >50% | 8 | Agonist |
| 2.1 | | 7 | Agonist |
| 1.2 | ≤50% | 6 | Agonist |
| 2.2 | >50% | 5 | Agonist |
| 2.2 | ≤50% | 4 | Inconclusive |
| 1.3 | | 3 | Inconclusive |
| 1.4 | | 3 | Inconclusive |
| 2.3 | | 2 | Inconclusive |
| 2.4 | | 2 | Inconclusive |
| 3 | | 2 | Inconclusive |
| 5 | | 1 | Inconclusive |

(continued)

**Table 2  (continued)**

| Curve class | Efficacy | Curve rank | Activity category |
|---|---|---|---|
| 4 | | 0 | Inactive |
| –2.3 | | –2 | Inconclusive |
| –2.4 | | –2 | Inconclusive |
| –3 | | –2 | Inconclusive |
| –1.3 | | –3 | Inconclusive |
| –1.4 | | –3 | Inconclusive |
| –2.2 | ≤50% | –4 | Inconclusive |
| –2.2 | >50% | –5 | Antagonist |
| –1.2 | ≤50% | –6 | Antagonist |
| –2.1 | | –7 | Antagonist |
| –1.2 | >50% | –8 | Antagonist |
| –1.1 | | –9 | Antagonist |

### 2.2 Assay Performance Measured by Reproducibility

After manual curation, the "clean" curve fitting results from the replicate assay runs are assessed for activity reproducibility to determine the final assay performance. Each sample curve is first assigned an activity outcome based on its curve class as follows: inactive (class 4), active agonist/antagonist (class 1.1, 2.1; class 5 due to super potency (AC50 < lowest test concentration)), agonist/antagonist (class 1.2, 2.2), inconclusive agonist/antagonist (all other non-5 classes), no call (other cases of class 5). Each activity outcome category (excluding the "no call" category, which is treated as missing data) is then assigned a score: active agonist (3), agonist (2), inconclusive agonist (1), active antagonist (–3), antagonist (–2), inconclusive antagonist (–1), inactive (0). The pair-wise activity outcome score differences for all replicate curves of each sample are then averaged and the % of inactive calls for the sample calculated to determine the final reproducibility call of the sample: active match (average score difference <1.1, %inactive call <25%), inactive match (average score difference <1.1, %inactive call >50%), mismatch (average score difference >2.5), inconclusive (all other cases). Each assay is assigned a performance score as follows: reproducibility score = 2 × %active match + %inactive match – %inconclusive – 2 × %mismatch.

### 2.3 Identification of Auto Fluorescence and Cytotoxicity Artifacts

In fluorescence-based agonist mode assays, auto fluorescent compounds can show the same phenotype as those of agonists. Two approaches are used to identify potential auto fluorescent artifacts to distinguish them from true agonists. One approach is using the auto fluorescence detection counter screen [12] measured at the same

wavelength as the assay readout (e.g., 460 nm for β-lactamase assays). Any compound with the agonist phenotype in the assay signal channel that also shows activation in the auto fluorescence counter screen with an AC50 difference <3-fold is flagged as a potential auto fluorescent false positive. The second approach is examining the activity of each compound in all the assays screened having the same reporter (e.g., β-lactamase). Compounds with the agonist phenotype in the assay signal channel (e.g., the 460 nm channel of β-lactamase assays) that also had an >4 average curve rank in the signal channel of all the assays with the same reporter are considered promiscuously active in such reporter gene assays and potentially auto fluorescent. The compounds identified by either approach are assigned the "inconclusive agonist (fluorescent)" activity outcome category. The auto fluorescence counter screen data on the Tox21 10 K have been made publicly available in PubChem [19] (assay IDs 720678, 720680, 720679, 720681, 720682, 720683, 720687, 720675, 720674, 720685, 720686, 720684). The same methods could be applied to luciferase reporter gene assays to identify compounds that are promiscuous luciferase stabilizers. Luciferase counter screens are also currently underway as an auxiliary approach to identify such compounds in the Tox21 10 K library.

In antagonist mode assays, cytotoxic compounds can show the same inhibitory phenotype as those of antagonists that need to be distinguished from true antagonists and flagged as cytotoxicity-related false positive responses. For this reason, each antagonist mode assay screened for Tox21 is accompanied with a cell viability readout that serves as the counter screen. Any compound with the antagonist phenotype in the assay signal channel that also shows inhibition in the cell viability counter screen with an AC50 difference <3-fold or $p > 0.05$ ($t$-test) is flagged as a potential cytotoxic false positive. As an alternative to the cell viability counter screen, the control channel in assays with multiple channel readouts (e.g., the 530 nm readout of β-lactamase assays) can be used to identify potential cytotoxic compounds. Either activation or inhibition shown in this channel can be an indication of cytotoxicity [6]. The effectiveness of using the control readout to identify potential cytotoxicity artifacts has been compared with the cell viability counter screen. The two approaches achieve similar specificity in correctly distinguishing true antagonists from cytotoxic artifacts, while filtering with the cell viability counter screen results in better sensitivity compared to the control readout [11].

**2.4 Compound Activity Assignment**

Compounds are assigned one of the following activity outcome categories: active agonist, inconclusive agonist (due to poor curve quality), inconclusive agonist (due to auto fluorescence), active antagonist, inconclusive antagonist (due to poor curve quality), inconclusive antagonist (due to cytotoxicity), inconclusive (activity direction cannot be determined), or inactive. The antagonist outcome labels in agonist mode assays are for compounds that show inhibition, which does not necessarily reflect true antagonism but rather might reflect increased cytotoxicity or promiscuous

reporter gene inhibition. The agonist outcome labels in antagonist mode assays are for compounds that show activation, which does not necessarily reflect true agonism but rather may reflect compound auto fluorescence or promiscuous reporter gene activation. To generate these assignments, curve ranks from all replicates of a compound are first averaged for each of the assay readouts, and the activity outcome of the compound in the assay readout is assigned based on the compound's average curve rank and reproducibility call as shown in Table 3. For luminescence assays with a single readout that is run in agonist mode, such as the estrogen receptor alpha luciferase reporter gene assay (BG1-ER-luc) [11], this activity outcome is assigned as the final activity outcome for a compound. For the same assay run in antagonist mode in complex with a cell viability counter screen, an activity outcome is assigned to both the antagonist mode readout and the cell viability readout first, and the final assay activity outcome for a compound is determined according to Table 4 (c). For fluorescence-based assays with multiple channel readouts (signal, control and ratio), such as the estrogen receptor alpha β-lactamase reporter gene assay (ER-bla) [11] or the mitochondrial membrane potential assay [9], the final activity outcome of a compound is determined based on its multi-channel activity as shown in Tables 4 (a) and (b). For antagonist mode assays, the cell viability counter screen data are used to flag potential cytotoxic artifacts. For agonist mode assays, potential artifacts produced by compounds that auto fluoresce in the signal channel (e.g., 460 nm readout in the ER-bla assay) are flagged

**Table 3**
**Compound single channel activity outcome assignments based on curve rank and reproducibility**

| Curve rank | Reproducibility call | Activity outcome |
|---|---|---|
| > –1 and <1 | Inactive match | Inactive |
| > –1 and <1 | Inconclusive | Inconclusive |
| ≥1 | Mismatch | Inconclusive agonist |
| ≥1 | Active match | Active agonist |
| >4 | Inconclusive | Active agonist |
| ≥1 and ≤4 | Inconclusive | Inconclusive agonist |
| ≤–1 | Mismatch | Inconclusive antagonist |
| ≤–1 | Active match | Active antagonist |
| < –4 | Inconclusive | Active antagonist |
| ≥ –4 and ≤–1 | Inconclusive | inconclusive antagonist |

**Table 4**
**Compound activity outcome assignments based on multi-channel assay readouts (*a*) multi-channel fluorescence agonist mode assay (*b*) multi-channel fluorescence antagonist mode assay (*c*) luminescence antagonist mode assay with cytotoxicity counter screen**

**(a)**

| Ratio outcome[a] | Signal channel outcome | Same reporter assay promiscuity | Auto fluorescence outcome | Activity outcome |
|---|---|---|---|---|
| Inactive | N/A | N/A | N/A | Inactive |
| Inconclusive | N/A | N/A | N/A | Inconclusive |
| Active agonist | Agonist | Average curve rank ≤4 | Inactive or AC50 fluor/AC50 signal ≥3 | Active agonist |
| Inconclusive agonist | Agonist | Average curve rank ≤4 | Inactive or AC50 fluor/AC50 signal ≥3 | Inconclusive agonist |
| Agonist | Agonist | Average curve rank >4 | Agonist and AC50 fluor/AC50 signal <3 | Inconclusive agonist (fluorescent) |
| Active antagonist | Antagonist | N/A | N/A | Active antagonist |
| Inconclusive antagonist | Antagonist | N/A | N/A | Inconclusive antagonist |

[a]Ratio = signal channel/control channel
Abbreviations: *AC50 fluor* = AC50 in the auto fluorescence assay, *AC50 signal* = AC50 in the ratio channel

**(b)**

| Ratio outcome[a] | Signal channel outcome | Cell viability outcome | Other conditions | Activity outcome |
|---|---|---|---|---|
| Inactive | N/A | N/A | N/A | Inactive |
| Inconclusive | N/A | N/A | N/A | Inconclusive |
| Active agonist | Agonist | Inactive or agonist | N/A | Active agonist |
| Active agonist | Agonist | Antagonist | AC50 viability/AC50 signal ≥3 ($p < 0.05$) | Active agonist |
| Inconclusive agonist | Agonist | N/A | N/A | Inconclusive agonist |

**Table 4**
**(continued)**

**(b)**

| | | | | |
|---|---|---|---|---|
| Agonist | Agonist | Antagonist | AC50 viability/AC50 signal <3 or $p \geq 0.05$ | Inconclusive agonist (cytotoxic) |
| Active antagonist | Antagonist | Inactive or agonist | N/A | Active antagonist |
| Active antagonist | Antagonist | Antagonist | AC50 viability/AC50 signal $\geq 3$ ($p < 0.05$) | Active antagonist |
| Inconclusive antagonist | Antagonist | N/A | N/A | Inconclusive antagonist |
| Antagonist | Antagonist | Antagonist | AC50 viability/AC50 signal <3 or $p \geq 0.05$ | Inconclusive antagonist (cytotoxic) |

[a]Ratio = signal channel/control channel
Abbreviations: *AC50 viability* = AC50 in the cell viability assay, *AC50 signal* = AC50 in the ratio channel

**(c)**

| Signal channel outcome | Cell viability outcome | Other conditions | Activity outcome |
|---|---|---|---|
| Inactive | N/A | N/A | Inactive |
| Inconclusive | N/A | N/A | Inconclusive |
| Active agonist | Inactive or agonist | N/A | Active agonist |
| Active agonist | Antagonist | AC50 viability/AC50 signal $\geq 3$ ($p < 0.05$) | Active agonist |
| INCONCLUSIVE agonist | N/A | N/A | inconclusive agonist |
| Agonist | Antagonist | AC50 viability/AC50 signal <3 or $p \geq 0.05$ | Inconclusive agonist (cytotoxic) |
| Active antagonist | Inactive or agonist | N/A | Active antagonist |
| Active antagonist | antagonist | AC50 viability/AC50 signal $\geq 3$ ($p < 0.05$) | Active antagonist |
| Inconclusive antagonist | N/A | N/A | Inconclusive antagonist |
| Antagonist | antagonist | AC50 viability/AC50 signal <3 or $p \geq 0.05$ | Inconclusive antagonist (cytotoxic) |

Abbreviations: *AC50 viability* = AC50 in the cell viability assay, *AC50 signal* = AC50 in the antagonist mode assay
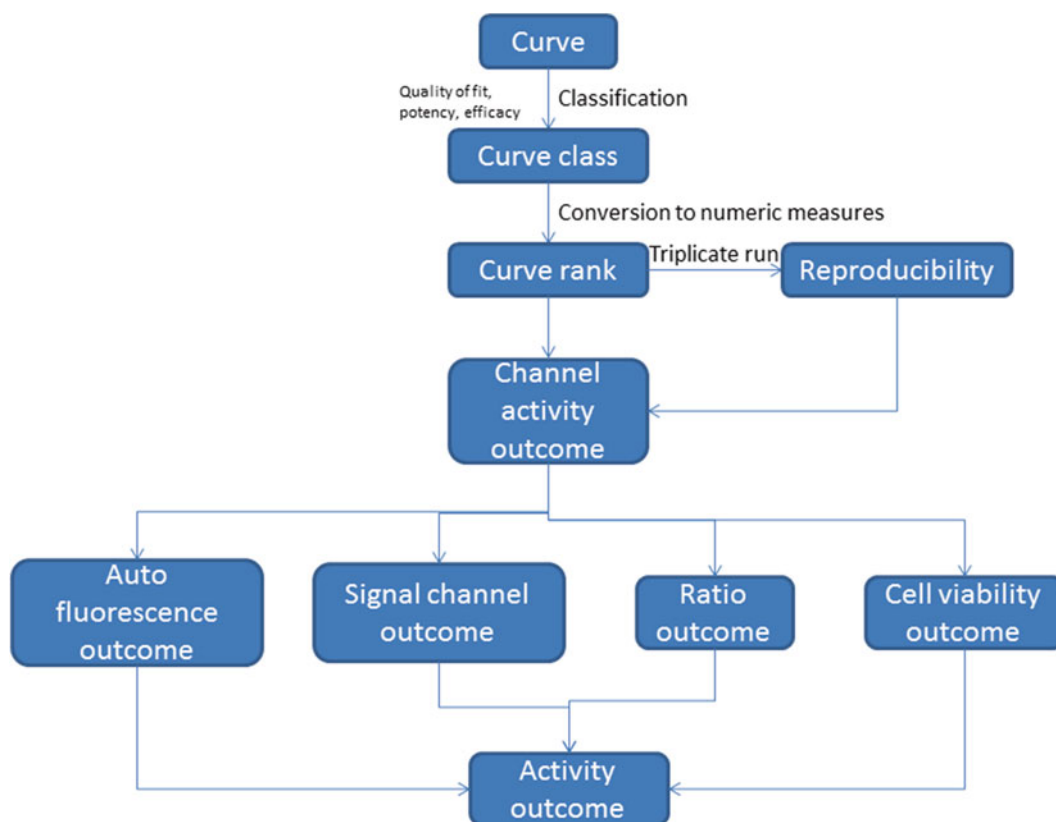
**Fig. 2** Compound activity assignment process. This *flowchart* shows the process of assigning a final assay activity outcome to a compound based on its qHTS data from replicate assay runs, multiple readouts, and counter screens

using both the compound auto fluorescence profiling data and the promiscuous compound activity shown in the signal readout of all the assays screened in Tox21 that have the same reporter (e.g., β-lactamase). The complete activity assignment process is illustrated in Fig. 2.

## 3   Data Sharing

As soon as the initial data parsing and assessment at the NCATS are complete, the concentration response data, curve fitting results, the raw plate reads, the assay conditions, and the sample mapping information are shared with the Tox21 partners through a suite of databases and software tools custom built by the NCATS for the Tox21 program (http://tripod.nih.gov/tox/). Within the first 6 months of data generation, the assay data are only made available to the Tox21 partners through the aforementioned controlled-access site

in which the data are further scrutinized for quality and utility. The data are then released to the public domain in a number of public databases including PubChem (http://pubchem.ncbi.nlm.nih.gov/), CEBS (http://tools.niehs.nih.gov/cebs3/ui/) and ACToR (http://actor.epa.gov). The high-quality concentration response data generated on a wide spectrum of pathways and phenotypic toxicity endpoints provide a valuable resource for predictive toxicity modeling. These data can not only serve as in vitro signatures that can be used to predict in vivo toxicity endpoints [20, 21] and to prioritize chemicals for more in depth toxicity testing [22] that helps fulfill the goals of the Tox21 program, but also provide rich training data sets for the QSAR (quantitative structure–activity relationship) modeling community to build more robust models [23, 24] such as the ones in the recent Tox21 Data Challenge hosted by NCATS in 2014 (https://tripod.nih.gov/tox21/challenge/) [25].

# References

1. Collins FS, Gray GM, Bucher JR (2008) Toxicology. Transforming environmental health protection. Science 319(5865):906–907. doi:10.1126/science.1154619, 319/5865/906 [pii]

2. Kavlock RJ, Austin CP, Tice RR (2009) Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal 29(4):485–487. doi:10.1111/j.1539-6924.2008.01168.x, discussion 492–487 RISK1168 [pii]

3. NRC (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington, DC

4. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765. doi:10.1289/ehp.1205784

5. Attene-Ramos MS, Huang R, Sakamuru S, Witt KL, Beeson GC, Shou L, Schnellmann RG, Beeson CC, Tice RR, Austin CP, Xia M (2013) Systematic study of mitochondrial toxicity of environmental chemicals using quantitative high throughput screening. Chem Res Toxicol 26(9):1323–1332. doi:10.1021/tx4001754

6. Huang R, Xia M, Cho MH, Sakamuru S, Shinn P, Houck KA, Dix DJ, Judson RS, Witt KL, Kavlock RJ, Tice RR, Austin CP (2011) Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ Health Perspect 119(8):1142–1148. doi:10.1289/ehp.1002952

7. Shukla SJ, Huang R, Austin CP, Xia M (2010) The future of toxicity testing: a focus on in vitro methods using a quantitative high throughput screening platform. Drug Discov Today 15(23-24):997–1007

8. Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, Austin CP, Shinn P, Simeonov A, Tice RR, Xia M (2013) The Tox21 robotic platform for the assessment of environmental chemicals – from vision to reality. Drug Discov Today 18(15-16):716–723. doi:10.1016/j.drudis.2013.05.015, S1359-6446(13)00161-X [pii]

9. Attene-Ramos MS, Huang R, Michael S, Witt KL, Richard A, Tice RR, Simeonov A, Austin CP, Xia M (2015) Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. Environ Health Perspect 123(1):49–56. doi:10.1289/ehp.1408642

10. Hsu CW, Zhao J, Huang R, Hsieh JH, Hamm J, Chang X, Houck K, Xia M (2014) Quantitative high-throughput profiling of environmental chemicals and drugs that modulate farnesoid X receptor. Sci Rep. doi:10.1038/srep06437

11. Huang R, Sakamuru S, Martin MT, Reif DM, Judson RS, Houck KA, Casey W, Hsieh JH, Shockley KR, Ceger P, Fostel J, Witt KL, Tong W, Rotroff DM, Zhao T, Shinn P, Simeonov A, Dix DJ, Austin CP, Kavlock RJ, Tice RR, Xia M (2014) Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Sci Rep. doi:10.1038/srep05664, srep05664 [pii]

12. Simeonov A, Jadhav A, Thomas CJ, Wang Y, Huang R, Southall NT, Shinn P, Smith J,

Austin CP, Auld DS, Inglese J (2008) Fluorescence spectroscopic profiling of compound libraries. J Med Chem 51(8):2363–2371. doi:10.1021/jm701301m

13. Thorne N, Auld DS, Inglese J (2010) Apparent activity in high-throughput screening: origins of compound-dependent assay interference. Curr Opin Chem Biol 14(3):315–324. doi:10.1016/j.cbpa.2010.03.020, S1367-5931(10)00046-3 [pii]

14. Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4(2):67–73

15. Southall NT, Jadhav A, Huang R, Nguyen T, Wang Y (2009) Enabling the large scale analysis of quantitative high throughput screening data. In: Seethala R, Zhang L (eds) Handbook of drug screening, 2nd edn. Taylor and Francis, New York, pp 442–463

16. Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol Lond 40:4–7

17. Wang Y, Jadhav A, Southal N, Huang R, Nguyen DT (2011) A grid algorithm for high throughput fitting of dose-response curve data. Curr Chem Genomics 4:57–66. doi:10.2174/1875397301004010057

18. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A 103(31): 11473–11478.doi:10.1073/pnas.0604348103, 0604348103 [pii]

19. PubChem (2013) Tox21 phase II data. http://www.ncbi.nlm.nih.gov/pcassay?term=tox21. Accessed 16 Nov 2013

20. Martin MT, Knudsen TB, Reif DM, Houck KA, Judson RS, Kavlock RJ, Dix DJ (2011) Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol Reprod 85(2):327–339. doi:10.1095/biolreprod.111.090977, biolreprod.111.090977 [pii]

21. Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, Chandler KJ, Dix DJ, Kavlock RJ, Knudsen TB (2011) Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. Toxicol Sci 124(1):109–127. doi:10.1093/toxsci/kfr220, kfr220 [pii]

22. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM, Dix DJ (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485–492. doi:10.1289/ehp.0901392

23. Huang R, Southall N, Xia M, Cho MH, Jadhav A, Nguyen DT, Inglese J, Tice RR, Austin CP (2009) Weighted feature significance: a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Toxicol Sci 112(2):385–393. doi:10.1093/toxsci/kfp231, kfp231 [pii]

24. Sun H, Xia M, Austin CP, Huang R (2012) Paradigm shift in toxicity testing and modeling. AAPS J 14(3):473–480. doi:10.1208/s12248-012-9358-1

25. Huang R, Xia M, Nguyen D-T, Zhao T, Sakamuru S, Zhao J, Shahane SA, Rossoshek A, Simeonov A (2016) Tox21 challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front Environ Sci 3:85. doi:10.3389/fenvs.2015.00085

# Chapter 13

## Correction of Microplate Data from High-Throughput Screening

**Yuhong Wang and Ruili Huang**

### Abstract

High-throughput screening (HTS) makes it possible to collect cellular response data from a large number of cell lines and small molecules in a timely and cost-effective manner. The errors and noises in the microplate-formatted data from HTS have unique characteristics, and they can be generally grouped into three categories: run-wise (temporal, multiple plates), plate-wise (background pattern, single plate), and well-wise (single well). In this chapter, we describe a systematic solution for identifying and correcting such errors and noises, mainly basing on pattern recognition and digital signal processing technologies.

**Key words** HTS, Data correction, Concentration response, In vitro assay

---

## 1 Introduction

High-throughput screening (HTS) is a method of identifying new biologically active compounds on a massive trial-and-error basis [1–4]. In this respect, HTS opens a new, broad avenue for toxicological research as an auxiliary tool to traditional animal-based testing for environmental and chemical prioritization in more in-depth toxicological studies [5]. At the same time, HTS advances the technology available for conducting massively parallel, automated, and relatively inexpensive biological experiments.

However, HTS data are generated from a complicated system over an extended run time; they tend to be noisy due to various factors such as the variations of experimental conditions and the instability of chemical compounds. While normalization assures comparability of signals between different plates, it does not guarantee that the measurements in different wells of each single plate are comparable. For example, signal gradients are quite frequent in HTS due to inhomogeneity of parameters, such as temperature, humidity, concentrations, or cell numbers, in different wells across a plate—factors that even the most elaborate optimization cannot completely eradicate. These noises need to be quickly identified

and excluded from subsequent processing by thorough quality control. This is of utmost importance at every stage of screening and analysis because occasional errors may obscure and invalidate the data resulting in false-positive or -negative findings.

The noises in the microplate-formatted data from HTS have unique characteristics; they can be generally grouped into three categories: run-wise (temporal, multiple plates), plate-wise (background pattern, single plate), and well-wise (single well). The uneven dispensing of reagents is an example of plate-wise noises. The uneven evaporation-caused edge effects across plates of a batch run are an example of run-wise noises.

In this chapter, we describe a systematic solution for identifying and correcting noises in microplate-formatted data, mainly accomplished through pattern recognition and digital signal processing algorithms. We also discuss its limitations and possible improvements.

## 2    Experimental/Design and Notations

At the National Institutes of Health (NIH), National Center of Advancing Translational Sciences (NCATS), most experiments are performed on 1536-well microplates in a quantitative high-throughput screening (qHTS) format [6]. In order to make data comparable across different plates in a batch run and assist the statistical correction of noises, we adopt two designs. First, positive, negative, or neutral control samples are placed in the first four columns. Compounds are placed in columns 5 through 48, where each well contains a different compound. In qHTS, compounds are tested as a concentration-response series such that each plate represents a different concentration [5]. These control wells are used to normalize primary readings and assure global comparability of signals between different plates. Second, control or DMSO plates are placed among titration series throughout a batch run (Fig. 1). As described below, these control plates are used to detect and correct run-wise noises of gradient nature.

At NCATS, neutral wells are normalized to 0, antagonist controls to −100 %, and agonist controls to 100 %. For convenience of discussion, we denote a well at row $i$ and column $j$ in a plate x as x[i,j]. For 1536-well microplates, $i$ ranges from 0 to 31, and $j$ from 0 to 47. In computational terms, plate x is represented as a two-dimensional matrix with an index starting at 0, a common convention used in modern computer languages.

For HTS data, normalized x is expected to be a sparse matrix due to the nature of low hit rates in typical HTS experiments. In other words, the activity values in the majority of the wells are expected to be insignificant.
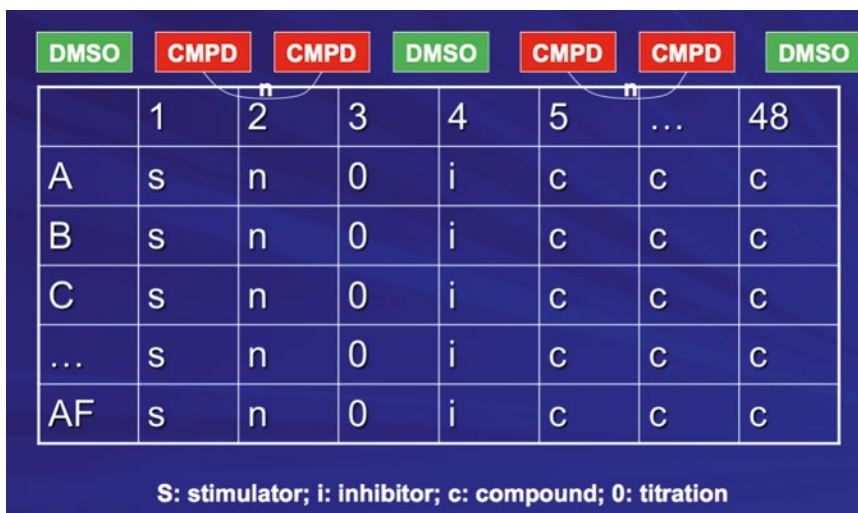
**Fig. 1** qHTS plate design at NCATS

Correction is performed on normalized microplate-formatted data and consists of two consecutive steps. First, each plate is corrected or "pre-cleaned" using the run-wise algorithm. Second, the pre-cleaned plates are then visually inspected to see if further plate-wise corrections are necessary. Variations on this recommended procedure are of course possible.

## 3  Correction of Plate-Wise Noises

There are four types of common noises in individual microplates. The first is the edge effect (Fig. 2). During plate incubation, solvent and media commonly evaporate faster from wells that are closer to the perimeters of the plate, with the outer and corner wells being the most affected. The result is a variation in cell growth across the plate, while any media components, such as salt, can become concentrated to the point where they are harmful to the cells. A volume loss as small as 10 % can concentrate media components and metabolites enough to alter cell physiology, consequently impacting the viability of downstream data and causing heterogeneous or biased results to occur.

The second common source of noise is the dispense pattern (Fig. 3) caused by either blocked or leaking tips. Depending upon the setup, cells can be dispensed more or less than the programmed number in consecutive wells or be dispensed in other patterns such as alternating wells.

The third type of noise or assay artifact is signal flare (Fig. 4). This noise is usually caused by compounds generating abnormally bright light. Such light can affect or pollute neighboring wells.
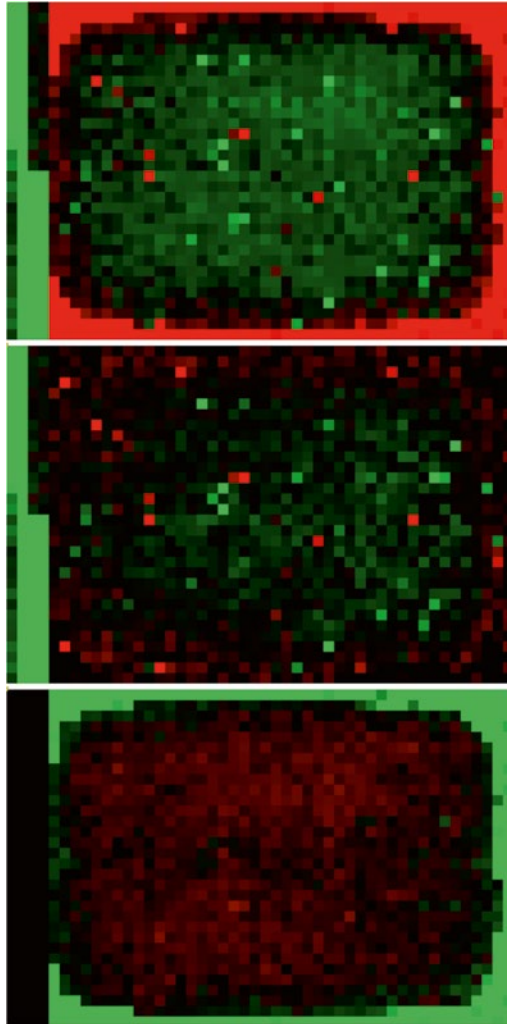
**Fig. 2** Example of edge effect. *Above*: normalized data; *middle*: corrected data; *bottom*: correction factor or the difference between corrected and normalized data

The last common source of noise is background pattern (Fig. 5). There are various causes of background patterns, such as an uneven plate, unevenly dispensed cells/reagents, or even worn-out lighting source. Unlike the three types of noises described above, the background pattern tends to be continuous across a large section of the plate at a certain gradient.

The easiest way of dealing with noisy plates is to exclude them from downstream analyses. However, such practice will not only waste expensive reagents but also delay a project. Plus, as seen in numerous experiments, the number of affected wells is usually small, and data from the majority of the wells in a plate remain usable.

Further examination of these four types of noises shows that edge effects and dispense patterns have common characteristics or
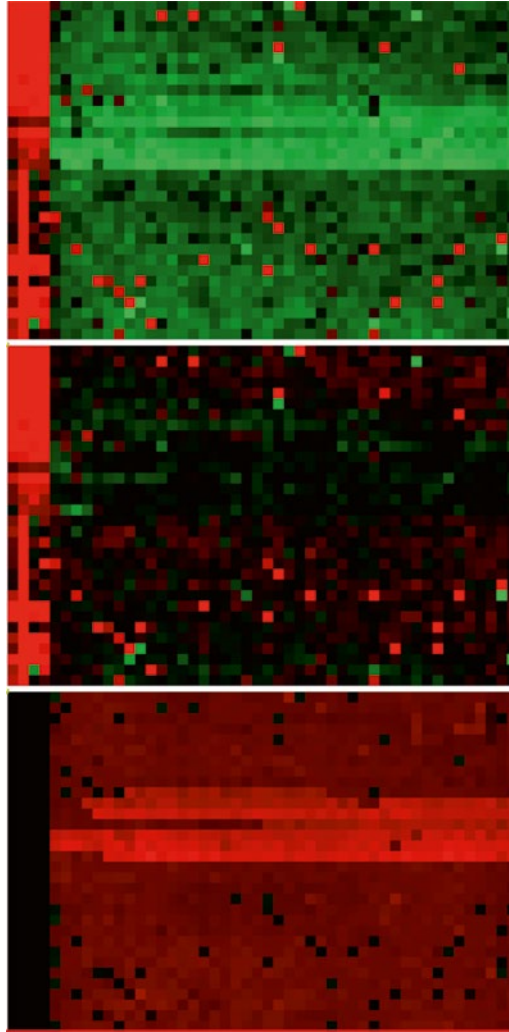
**Fig. 3** Example of dispense pattern. *Above*: normalized data; *middle*: corrected data; *bottom*: correction factor or the difference between corrected and normalized data

clear patterns; that is, significantly altered signals are often seen in either consecutive or alternating wells.

We designed three plate-wise correction algorithms. The first is called block algorithm, which can be used to correct dispense patterns, edge effects, and signal flares. The second and third are called discrete cosine transform (DCT) and median algorithms for correcting signal flares and background patterns.

*3.1 Block Algorithm*

The block algorithm consists of two steps: block identification and data correction. For block identification, we defined three parameters: minimal signal or signal cutoff (SC), minimum number of wells (NW), and interval between wells (IW). The default values for SC, NW, and IW are 25 %, 3, and 1, respectively.
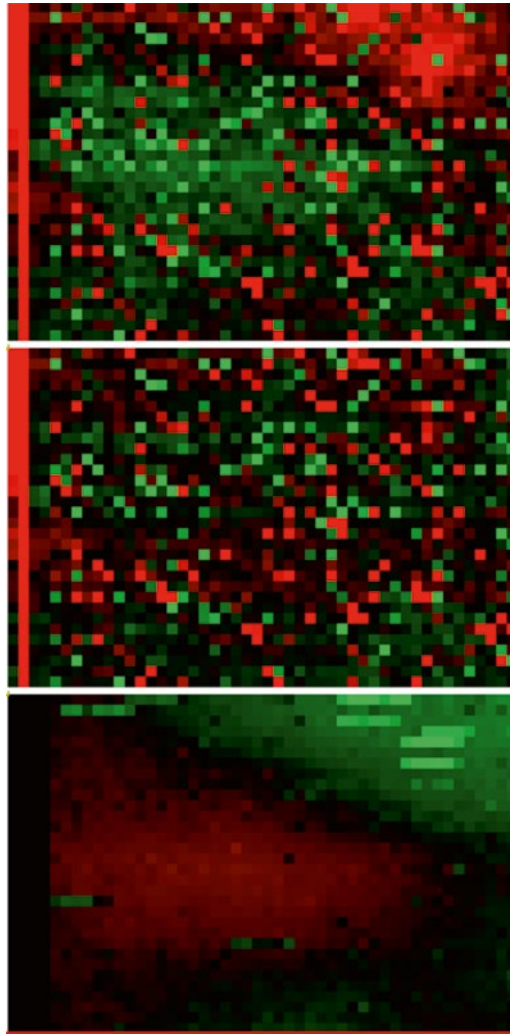
**Fig. 4** Example of signal flare. *Above*: normalized data; *middle*: corrected data; *bottom*: correction factor. Two signal flares are detected: *top right* and *left middle*

SC and NW are self-evident, but IW may need further clarification. Let us assume a detected block of three wells. When IW = 1, the detected block of 3 wells is like A1, A2, and A3 for row-wise dispensing pattern. When IW = 2, the detected block of three wells is like A1, A3, and A5, in other words, every other well. Patterns of IW = 1 are more common than those of IW = 2.

Here is the pseudo code for detecting row-wise blocks for an agonist mode or activation-type assay.

Start from row A.

Start from column 1.
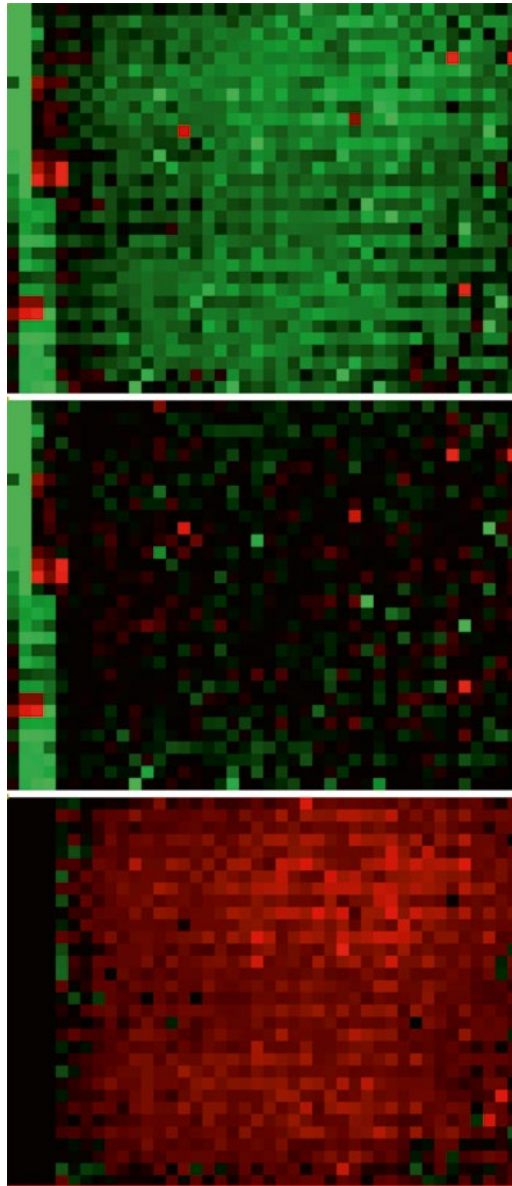
**Fig. 5** Example of background pattern. *Above*: normalized data; *middle*: corrected data; *bottom*: correction factor or the difference between corrected and normalized data

Look for block of wells meeting three criteria as defined by the three parameters of SC, NW, and IC:

x[i,j] > SC.

Number of wells > NW.

Interval between wells = IC.

Extend the block until the next well of IW wells away from the last identified one is less than SC.

Continue until the last column (column 48 for 1536-well plate).

Continue until last row (row AF for 1536-well plate).

Repeat by starting from column 1 and row A to identify column-wise blocks.

The pseudo code for detecting block for an inhibition assay is very similar except, instead of checking x[i,j] > SC, we use x[i,j] < -SC.

For data correction, we defined another parameter—multiplier of standard deviation (MS). The correction is performed for each identified block, and a pseudo code is given below.

Calculate median value (PM) of the entire plate.

For each identified block (either row-wise or column wise):

Calculate median value of the block (BM).

Calculate the standard deviation of the block (BSD).

For each well in the block:

If |x[i,j] – BM| > MS*BSD,

x[i,j] = x[i,j] – BM.

Otherwise

x[i,j] = PM.

Figure 3 shows one example of applying the block algorithm on a plate with dispensing pattern.

*3.2 Discrete Cosine Transform Algorithm*

A *discrete cosine transform (DCT)* expresses a finite sequence of data points in terms of a sum of cosine functions of different frequencies. In particular, a DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT) but uses only real numbers. The mathematical equations of DCT and the inverse DCT are given as follows:

$$X[u,v] = \frac{2}{\sqrt{MN}} C[u]C[v] \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} x[i,j] \cos\left[\frac{(2i+1)u\pi}{2M}\right] \cos\left[\frac{(2j+1)v\pi}{2N}\right]$$

$$x[i,j] = \frac{2}{\sqrt{MN}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} C[u]C[v]X[u,v] \cos\left[\frac{(2i+1)u\pi}{2M}\right] \cos\left[\frac{(2j+1)v\pi}{2N}\right]$$

$$C[\gamma] = \frac{1}{\sqrt{2}} \text{ for } \gamma = 0; 1 \text{ for } \gamma > 0$$

For 1536-well microplate, $M = 32$ and $N = 48$. x[i,j] are called real data points; X[u,v], real spectra or the coefficients of cosine functions of different frequencies.

In typical applications of lossy compression of audio (e.g., MP3) and images (e.g., JPEG), high-frequency components or spectra are discarded. The idea behind is that noises in images tend to be of high frequency.

In HTS data, the usual hit signals are of random nature; the noises in signal flare and background pattern are of low-frequency nature. Thus, in this DCT algorithm, original real data of $32 \times 48$ format are transformed to spectral space first. The lowest m and n frequency components are discarded. The spectra data are then inversely transformed.

DCT transform: x[i,j] - > X[u,v].

Discard low spectra: X[u,v] = 0 for u < MU and v < MV.

Inverse DCT transform: X[u,v] - > x'[i,j].

MU and MV are optional parameters with default values of [1, 1].
x'[i,j] is the corrected data.

Figure 5 shows one example of applying DCT algorithms to correct background pattern.

**3.3 Median Algorithm**

Some background patterns are quite evenly distributed. In that case, the median algorithm is both simple and effective. In this algorithm, median signal of all wells of a plate is calculated and then subtracted from all wells. We use median instead of arithmetic mean because median is a more robust estimate of experimental signals. An arithmetic mean is a non-resistant statistic that can be greatly influenced by the commonly observed outlier signals in HTS data.

# 4    Correction of Run-Wise Noises and Errors

While normalization assures global comparability of signals between different plates, it does not guarantee that the measurements in different wells of each single plate are comparable in the presence of gradients. These signal gradients are quite frequent in HTS due to inhomogeneity of parameters, such as differences in temperature, humidity, cell number in different wells across a plate, and other factors. Statistical methods can identify and remove common gradient patterns in groups of plates. Such correction also adds discriminatory power to the assay results because it renders results, independent of plate location, perfectly comparable and reduces noise.

Experimental artifacts are often associated with specific conditions or settings that prevail on a certain day or in a certain batch. Furthermore, such conditions can be separated into two

groups: random and gradient. One example of the gradient variation is evaporation. During a batch run, if plates are not well sealed, the solvent tends to gradually evaporate, causing a gradual and one-directional change in signal reading.

In HTS data, random variations are a common phenomenon, and gradient variations are less frequent. At NCATS, in order to assist in the statistical correction of such gradients, control or DMSO plates are placed among titration series throughout a batch run (Fig. 1). We developed two algorithms to utilize these control data to correct run-wise noises of gradient nature: median algorithm and pattern algorithm.

### 4.1 Median Algorithm

In this algorithm, the median values of all wells in the control plates (Fig. 1) are calculated and then subtracted from corresponding wells of assay plates. The pseudo codes for the 7 plates shown in Fig. 1 are given below:

Start from row A.

    Start from column 1.

        Calculate median of well A1 for 1st, 4th, and 7th DMSO plates.

        Subtract the median from 2nd, 3rd, 5th, and 6th compound plates.

    Repeat until column 48.

Repeat until row AF.

Median algorithm, like the one for individual plate, is very simple and effective for run-wise, random variations.

### 4.2 Pattern Algorithm

In the pattern algorithm, we first try to identify any time-dependent and significantly linear gradient pattern for each well in the control plates. In other words, we try to examine the correlation between signal and time. If the correlation is statistically significant, we perform a least square linear fitting of the signal, and the fitting parameters are used to calculate the correction factors for compound plates based upon their measure time. If no significant correlation is observed for a well, we default to the median algorithm for its correction.

In the linear regression equation,

$$x[i, j]_t = a + bt$$

we have three parameters: $a$, $b$, and $p$-value. The p-value tests the null hypothesis that the coefficient $b$ is equal to zero (no effect). A low $p$-value ($<0.05$) indicates that we can reject the null hypothesis. In this algorithm, we use the commonly used cutoff of 0.05. When the $p$-value $<0.05$, we will calculate correction factor for the

well [i,j] of a compound plate according to the same equation and the plate's measurement time. The correction factor is then subtracted from the well of the plate. If the *p*-value >= 0.05, the correction will be performed using the median algorithm.

The following is the pseudo code, again assuming seven plates (Fig. 1).

Start from row A.

Start from column 1.

> Perform linear regression for data from well A1 for 1st, 4th, and 7th DMSO plates using the above equation. Calculate the median at the same time.
>
> Subtract the median from well A1 of 2nd, 3rd, 5th, and 6th compound plates if *p*-value >= 0.05.
>
> Otherwise calculate correction factor delta for well A1 of 2nd, 3rd, 5th, and 6th compound plates:

$$\delta_t = a + bt$$

Subtract the delta from well A1 of 2nd, 3rd, 5th, and 6th compound plates.

Repeat until column 48.

> Repeat until row AF.
>
> One example of applying pattern algorithm to nine plates, with first two and last two plates as DMSO control plates, is given in Fig. 6.

## 5  Conclusion

At NCATS, ~470,000 qHTS plates have been processed and corrected using a combination of algorithms for plate-wise and run-wise corrections. These algorithms have been proven to be effective
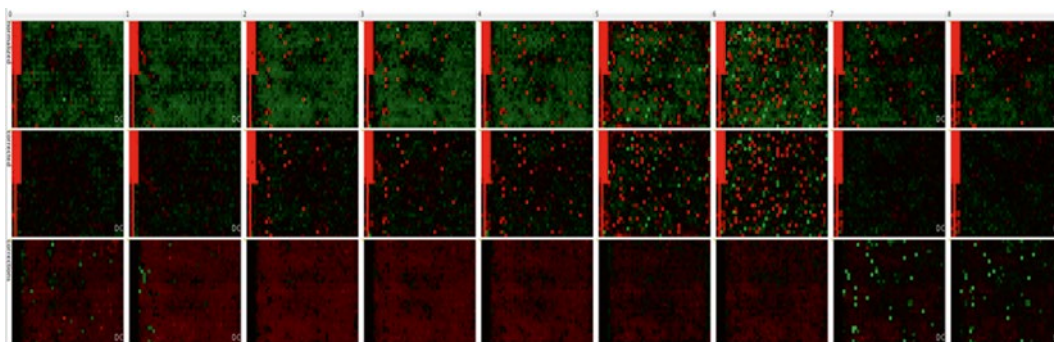


**Fig. 6** Run-wise correction using pattern algorithm. *Above*: normalized data; *middle*: corrected data; *bottom*: correction factor

and efficient for both cellular and enzymatic assays and have saved a significant amount in terms of time, cost, and resources.

Nevertheless, the current algorithms have two major limitations. First, we do not have a robust and objective measure of the quality of the correction procedure. The quality of the correction results is largely determined by apparently subjective human judgment, even though in practice human judgment seems adequate for our routine data analysis. Secondly, manual interventions and visual inspections are still necessary during the correction process. A number of parameters need to be optimized interactively. Unlike the other parts in the NCATS qHTS data processing pipeline, the correction process is still not fully automated.

To improve and automate the current correction procedure, we are currently experimenting with a number of statistical measurements of plate noise and an iterative process for parameter optimization in order to find the best combination of algorithms for a fully automated correction procedure.

## References

1. Cox B, Denyer JC, Binnie A, Donnelly MC, Evans B, Green DV, Lewis JA, Mander TH, Merritt AT, Valler MJ, Watson SP (2000) Application of high-throughput screening techniques to drug discovery. Prog Med Chem 37:83–133

2. Spencer RW (1998) High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. Biotechnol Bioeng 61(1):61–67. doi:10.1002/(SICI)1097-0290(199824)61:1<61::AID-BIT11>3.0.CO;2-C [pii]

3. Böhm H-J, Schneider G (eds) (2000) Virtual screening for bioactive molecules. WILEY-VCH, New York

4. Hertzberg RP, Pope AJ (2000) High-throughput screening: new technology for the 21st century. Curr Opin Chem Biol 4(4):445–451, doi:S1367-5931(00)00110-1 [pii]

5. Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, Austin CP, Shinn P, Simeonov A, Tice RR, Xia M (2013) The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. Drug DiscovToday.doi:10.1016/j.drudis.2013.05.015, S1359-6446(13)00161-X [pii]

6. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A 103(31):11473–11478. doi:10.1073/pnas.0604348103, 0604348103 [pii]

# Chapter 14

# CurveP Method for Rendering High-Throughput Screening Dose–Response Data into Digital Fingerprints

## Alexander Sedykh

## Abstract

The nature of high-throughput screening (HTS) puts certain limits on optimal test conditions for each particular sample, therefore, on top of usual data normalization, additional parsing is often needed to account for incomplete read outs or various artifacts that arise from signal interferences.

CurveP is a heuristic, user-tunable, curve-cleaning algorithm that attempts to find a minimum set of corrections, which would give a monotonic dose–response curve. After applying the corrections, the algorithm proceeds to calculate a set of numeric features, which can be used as a fingerprint characterizing the sample, or as a vector of independent variables (e.g., molecular descriptors in case of chemical substances testing). The resulting output can be a part of HTS data analysis or can be used as input for a broad spectrum of computational applications, such as Quantitative Structure-Activity Relationship (QSAR) modeling, computational toxicology, bio- and cheminformatics.

**Key words** Nonparametric fitting, Monotonicity, Heuristics

## 1 Introduction

In 1980s, high-throughput screening (HTS) had gained popularity in drug design as a fast and inexpensive way of identifying chemical substances with bioactive potential [1]. Typically, a large library of substances was tested at a single concentration; then positive hits would be selected for a follow up screening required to validate selected candidates further (e.g., low or medium throughput confirmatory tests). In recent years, this technique has evolved into quantitative high-throughput screening (qHTS) [2] that attempts to capture, in one screening, entire dose–response relationship of the substance with a given biological target. As such, qHTS found application not only in medicinal chemistry, but also in computational toxicology, where the current trend is to replace slow and expensive animal tests with faster and simpler in vitro assays.

Current qHTS protocols feature several enhancements, such as cytotoxicity counter screening, interference screens (e.g., autofluoresence), and replicate runs with plate position reshuffling of samples [3]. However, two major qHTS issues, such as signal artifacts and incomplete profiles, persist, as these are rooted in the nature of the high-throughput technology. Artifacts, and signal noise in general, arise from multiple factors of biological and equipment origin and can significantly distort actual dose–response curves. Moreover, each substance is usually tested at the same, predefined concentration range (e.g., same number of serial dilutions of the initial stock solution). For too potent or too weak compounds, this would produce incomplete, partial dose–response curves. Such issues thwart the analysis of resulting experimental data, as these cannot be always well accounted for by common curve-modeling methods, such as Hill model or isotonic regression.

This work describes CurveP [4], a nonparametric algorithm that attempts to address the mentioned issues by seeking a solution that minimizes the number of corrected measurements. CurveP also introduces curve fingerprint metric, amenable to be used as a biological information to be related to higher-order effects in complex biological systems (e.g., in vivo toxicity).

## 2    Method

Typical qHTS data for a single sample consists of $n$ response measurements ($r_1$, $r_2$, …, $rn$) at $n$ test concentrations ($c_1$, $c_2$, … $cn$). Measurements are usually given in % relative to the positive (100%) and negative (0%) controls. Concentration values are usually given in decimal logarithm transformation ($\log_{10} c$). An example of a dose–response curve is shown in Fig. 1a.

CurveP assumes each dose–response curve has to exhibit monotonic behavior: increasing, decreasing, or flat (constant), and tries to find a minimal set of correctable test points to achieve that.

### 2.1   User-Controlled Parameters of CurveP

The algorithm's behavior is controlled by several variables, which are modifiable by the user:

**THR**—baseline threshold, in % of response. THR is a response level, below which any signal is considered not different from the baseline (zero level signal). Based on prior studies [4] its default value was set to 15%. This parameter helps to suppress background noise.

**MXDV**—is a maximum allowed deviation from monotonicity. If adjacent test responses violate monotonic pattern, but the absolute difference of their responses is equal or less than MXDV, in % of response, such violation is ignored. This parameter prevents overcorrections of small local violations in monotonicity, thus shifting focus to global behavior of the curve. The default value of MXDV is 5%.
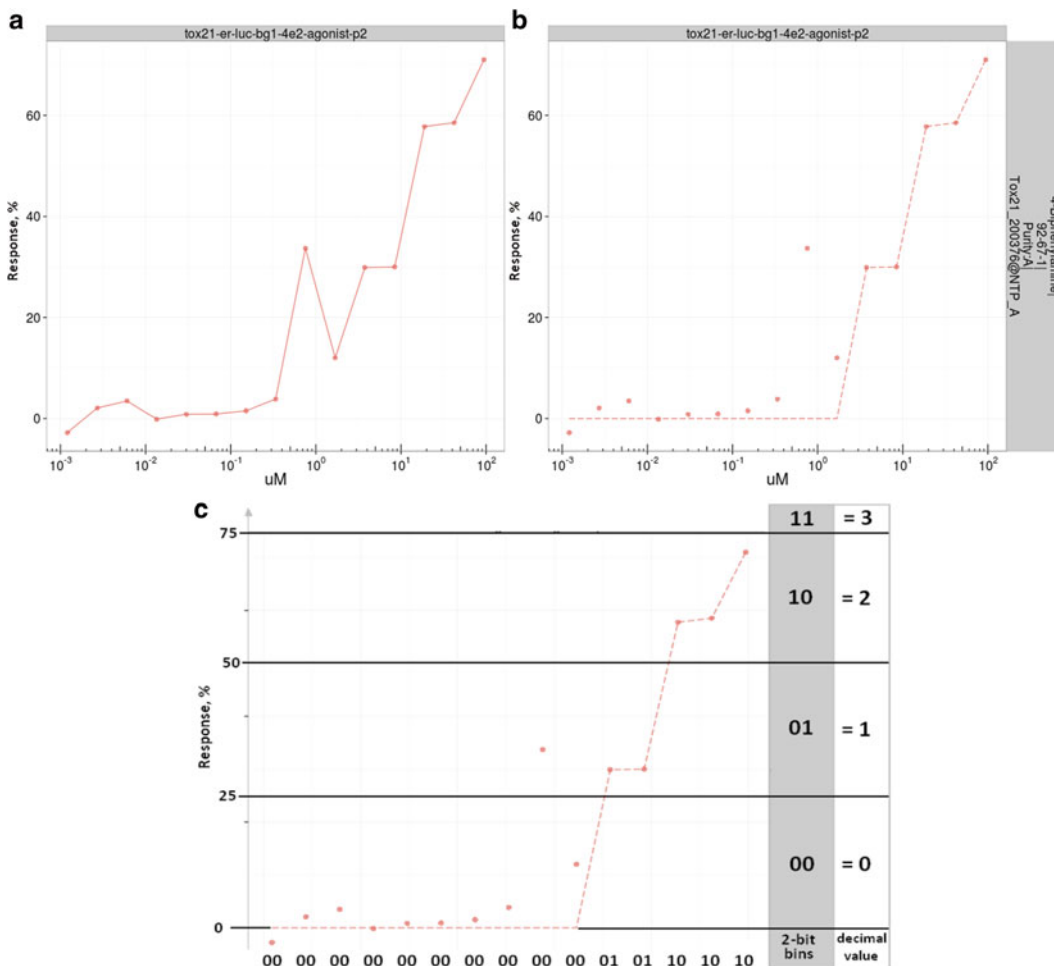
**Fig. 1** Dose–response curves processing by CurveP shown on example of 4-Aminobiphenyl (CAS RN: 92-67-1) tested in estrogen receptor agonism assay from NIEHS Tox21 curve browser (https://hsiehjh.shinyapps.io/tox21curvebrowser), with original qHTS data (**a**), resulting cleaned curve (**b**), and CurveP digital fingerprint (**c**) calculated as $0\ldots0101101010_2 = 2^8 + 2^6 + 2^5 + 2^3 + 2^1 = 362_{10}$

**BYHI/BYLO**—flags are used when several alternative corrective solutions are found, tiebreaks are then resolved by trusting data from higher/lower test concentrations. The default mode is BYLO, as it is assumed, there is more noise at higher test concentrations.

**RNG**—maximal response (can be positive or negative), specifies which range of the responses to analyze (important for bidirectional assays). Responses outside of the [0…RNG] interval are reset to its closest boundary.

**USHAPE**—sets minimum number of test points that must support a U-shaped curve in order for its detection and resolution.

**CRO**—sets minimum signal at the first test concentration that would trigger detection and resolution of carry-over cases.

**BSHIFT**—minimum number of test points at the constant part of the curve to detect baseline-shift cases.

Once a cleaned curve is produced, it can be converted to various metrics, including CurveP fingerprint (Fig. 1c), which individually or jointly can be utilized as biological properties (i.e., qHTS descriptors) of tested substances.

*2.2 Algorithmic Description of CurveP*

1. Read responses ($r_{1,...,n}$), and test concentrations ($c_{1,...,n}$). Log-transform $ci$ values, if needed.

2. Read mask of invalid points: ($m_{1,...,n}$) flags. These points are treated as missing data during analysis.

3. Reset $|ri| <$ THR to 0. Reset $|ri| > |$RNG$|$ to RNG.

4. Correct "blips": If $|r_1| > 0$, but $r_2$ and $r_3 = 0$, then reset $r_1$ to 0.

5. Define curve's direction as ($r_1 - rn$). This will be >0 for decreasing curves, <0 for rising curves, and 0 for flat, constant curves. If $|r_1 - rn| <$ THR, and either std($r_{1,...,n}$) < MXDV or max($r_1, rn$) < THR, consider curve as flat and go to **step 9**.

6. Count points that violate global direction:

Find all $ri$, where ($|r_1 - rn| +$ MXDV $< |ri - r_1|$). If less than two, go to **step 9**.

7. If std($r_{1,...,n}$) > MXDV, analyze for U-shaped curves with "low-high-low" or "high-low-high" response profiles using the USHAPE parameter. If found, invalidate appropriate portion and establish monotonicity direction.

This step can correct some carry-over cases that fall into "high-low-high" profile.

Go to **step 9.**

8. Treat flat curves with mean($r_{1,...,n}$) ≥ THR as potential baseline-shift or carry-over cases: Invalidate $ri$, if $|ri -$ mean($r_{1,...,n}$)$| >$ MXDV, then go to **step 12**.

9. Monotonicity direction has been established. If a flat curve, reset to baseline, then go to **step 12**.

10. Scan $r_{1,...,n}$ for monotonicity violations large than MXDV. Find a smallest set of points to restore monotonicity, once corrected. If multiple alternative solutions exist, pick the one with less correction at higher (lower) test concentrations as defined by the BYHI (BYLO) mode.

11. Extrapolate missing and invalidated data points from the remaining ones that are adjacent.

12. Detect remaining carry-over cases and related artifacts. Skip this step, if CRO = 0 or $r_1$ = 0.

If the curve is not flat, but goes in reverse to the expected [0...RNG] direction, then mark it as a carry-over and correct the points, but only if $r_1$ < CRO or RNG < 0 (negative range is typical for inhibition assays). This filter allows to pass very potent samples in activation-type assays, i.e., those with RNG > 0 and $r_1$ > CRO. Go to **step 13**.

Analyze remaining cases of curves that are flat or monotonic in the regular, expected direction.

If $r_1$ > CRO, mark as very potent, go to **step 13**.

If flat, mark as either baseline-shift or carry-over case (insufficient data to detect which).

If monotonic, detect and correct for baseline-shift using the BSHIFT parameter.

13. Print out adjusted dose–response data.

14. Calculate curve fingerprint. Use two bits to represent signal at each test concentration:

0 (binary 00), if $|ri| <= |RNG|/4$

1 (binary 01), if $|RNG|/4 < |ri| <= |RNG|/2$

2 (binary 10), if $|RNG|/2 < |ri| <= 3|RNG|/4$

3 (binary 11), if $3|RNG|/4 < |ri|$

Concatenate bits, from lowest to highest test concentrations, for all $r_{1,...,n}$ responses. This gives a string of $2n$ bits, which can be interpreted as a single integer value – the CurveP fingerprint (*see* Fig. 1c).

15. Calculate other curve parameters. Estimate test concentration at which certain default response thresholds are achieved (e.g., $EC_5$, $EC_{10}$, $EC_{20}$). Estimate point-of-departure (POD) as a lowest concentration at which signal is expected to reach THR. For metrics, whose estimation is impossible (such as in case of $EC_{50}$ for flat, no-response curves), missing value flags are reported instead.

The CurveP program is open source (written in C++) and available at Github (https://github.com/sedykh/curvep). It can be compiled for both, Windows and Linux platforms.

## 3   CurveP Applications

Several studies have applied CurveP for noise-cleaning and utilized its output metrics as biological descriptors in computational modeling or analysis [3–8]. In our earlier work we compared the impact

of curve cleaning on the utility of resulting biological descriptors [4]. We found that descriptors based on the original, "as-is," qHTS data were too noisy to show any improvement in machine learning models from k-nearest neighbors and random forest methods [4]. Our modeling experiments [4, 8] indicated that the level of noise in response is usually at 5–25 % of the positive control signal, and depends on the nature of the assay [3].

There are several ways for rendering resulting cleaned data into a set of numerical variables. Simplest is to use cleaned responses directly as descriptor values, one for each test concentration. This guarantees no missing values in resulting descriptors, but may have significant degree of inter-correlation between descriptors of the same assay. It also introduces model's dependency on a particular set of test concentrations, which would require imputation of responses in the input data, if those test concentration are different (e.g., come from alternative sources). Alternatively, a set of concentration-based metrics could be used as descriptors, such as the effective concentration (EC), at which a certain predefined level of response is achieved (e.g., $EC_{10}$, $EC_{20}$, $EC_{50}$). The EC*xx*-like values are imputed from cleaned curves and should be independent (except for the imputation bias) of actual test concentrations used in screening. However, EC metrics can have missing values for flat, no-response curves, so a default value will need to be agreed on and uniformly added (e.g., some very high concentration) to denote those.

Finally, when a large, diverse panel of assays is available, a single descriptor per assay may suffice for biological space representation. In this case, CurveP fingerprint (Fig. 1c), weighted area-under-curve [3], or other traditional curve metrics can be employed.

CurveP fingerprint (further, simply, CurveP) is a convenient metric for comparing entire dose–response curves. Curves with the same or close CurveP values can be expected to follow similar dose–response pattern. Flat, no-response curves would have all bits set to zero, so their overall CurveP value will be exactly zero as well. Curves with more potent responses will have higher CurveP values. This is because bit-assigning order gives higher weight to responses from lower test concentrations, as well as because at each test concentration higher bits in the bin denote higher response (Fig. 1c).

CurveP's major limitation is its tacit dependency on a given order of test concentrations. Those are treated as a predefined sequence of measurements, and in case of irregularly spaced test concentrations or changes in their number or range for different samples, the comparisons of resulting CurveP values will be biased.

As can be seen from Fig. 1c, CurveP yields a nonnegative integer value that can be very large for highly potent samples. The numerical scale of CurveP can thus span many orders of magnitude, which is inconvenient in a metric, which therefore should be log-normalized [7]. Applying a logarithm with base 4 (preferable, so as to correspond to 2-bit bins of response) to a $(1 + CurveP)$ value would give

log-CurveP that can be interpreted (when rounded to the next highest integer) as the rank-order of the test concentration (from highest to lowest), at which substantial (>25 %) signal was detected. For example, CurveP = 362 (Fig. 1c) would have $\log_4(1 + CurveP)$ close to 4.252, which corresponds to the fifth highest test concentration (counted from right) as the lowest concentration with significant signal. Thus, CurveP can be interpreted as a kind of point-of-departure metric, and the calculated rank-order can be further converted to the corresponding test concentration [5].

## 4  Conclusions

In our opinion, data cleaning and processing of ordinary dose–response curves (i.e., curves with a clear trend and relatively small amount of noise) can be successfully accomplished by many different methods, parametric or not. A real hurdle comes from the difficult cases, when the curve's pattern is broken in nonobvious ways due to flukes, interference effects, and technical artifacts. Most curve-fitting methods, such as Hill equation or isotonic regression, would need some prior handling of outliers in order to accurately fit the curve to the given responses of the remaining data points.

On the other hand, CurveP, as a heuristic method, does not require any predefined analytical shape of the curve to rely on. On assumption of curve's monotonicity it searches for a minimum set of outliers and then corrects them, thus minimizing not the difference of fitted vs. observed responses, but instead, the number of corrections applied.

## References

1. Pereira DA, Williams JA (2007) Origin and evolution of high throughput screening. Br J Pharmacol 152:53–61

2. Inglese J, Auld DS, Jadhav A et al (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A 103: 11473–11478

3. Hsieh JH, Sedykh A, Huang R et al (2015) A data analysis pipeline accounting for artifacts in Tox21 quantitative high-throughput screening assays. J Biomol Screen 20:887–897

4. Sedykh A, Zhu H, Tang H et al (2011) Use of in vitro HTS-derived concentration–response Data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect 119:364–370

5. Lock EF, Abdo N, Huang R et al (2012) Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. Toxicol Sci 126:578–588

6. Low Y, Sedykh A, Fourches D et al (2013) Integrative chemical-biological read-across approach for chemical hazard classification. Chem Res Toxicol 26:1199–1208

7. Sprague B, Shi Q, Kim MT et al (2014) Design, synthesis and experimental validation of novel potential chemopreventive agents using random forest and support vector machine binary classifiers. J Comput Aided Mol Des 28:631–646

8. Sedykh A, Low Y, Lock E et al. (2012) Using population-based dose-response cytotoxicity data for in silico prediction of rat acute toxicity. Abstracts of Papers, 51th SOT National meeting, San Francisco, CA, March 11–15

# Chapter 15

## Accounting Artifacts in High-Throughput Toxicity Assays

## Jui-Hua Hsieh

### Abstract

Compound activity identification is the primary goal in high-throughput screening (HTS) assays. However, assay artifacts including both systematic (e.g., compound auto-fluorescence) and nonsystematic (e.g., noise) complicate activity interpretation. In addition, other than the traditional potency parameter, half-maximal effect concentration ($EC_{50}$), additional activity parameters (e.g., point-of-departure, POD) could be derived from HTS data for activity profiling. A data analysis pipeline has been developed to handle the artifacts and to provide compound activity characterization with either binary or continuous metrics. This chapter outlines the steps in the pipeline using Tox21 glucocorticoid receptor (GR) β-lactamase assays, including the formats to identify either agonists or antagonists, as well as the counter-screen assays for identifying artifacts as examples. The steps can be applied to other lower-throughput assays with concentration-response data.

**Key words** HTS, qHTS, Concentration-response data, Tox21, Assay artifacts, Data analysis pipeline, Point-of-departure

## 1  Introduction

In drug discovery field, the term HTS usually means the experiment with over 1 million of compounds tested at a single concentration (e.g., 10 μM) against a particular drug protein target [1]. Secondary assays with compounds tested in either the same and/ or orthogonal assays with more concentrations are needed to validate the HTS hits. On the contrary, in the field of toxicity, HTS usually means the experiment with fewer than 10K compounds tested at multiple concentrations (i.e., concentration-response data) in various types of assays ranging from receptor binding to cellular processes. ToxCast®[2] and Tox21[3] are the two initiatives that are performing HTS with either a larger number of chemicals (10K) but fewer assays (<100) or a smaller number of chemicals (<2000) with more assays (>700), respectively. Both of them have generated rich concentration-response data for analysis. However, the data also pose challenge on how to systematically and accurately provide compound activity characterization.

One of the major reasons to use multiple concentrations (or called as quantitative high-throughput screening, qHTS) in toxicological studies is to identify weakly active compounds in assays. However, the weakly actives could be challenging for traditional data analysis methods, where concentration-response data with maximal effect (i.e., 100% relative to the positive control) are preferred and thus half-maximal effect concentration ($EC_{50}$) can be more accurately estimated and more rationally compared across chemicals [4]. In addition, noise/outliers (could be nonsystematic artifacts or true responses) are common in qHTS data [5], which could also be challenging for data analysis methods where monotonicity assumption is usually required. Systematic HTS artifacts, including auto-fluorescence, quenching, luciferase inhibition, and compound aggregation, have been noted and characterized in a large screening library in previous studies [6–9]. In addition, cytotoxicity is reported as the most common systematic artifacts in Tox21 inhibition-type assays [10]. In this chapter, a data analysis pipeline [10], especially for Tox21 assay data, is described, that takes above issues into account in order to derive activity calls. Two metrics are implemented in addition to the traditional $EC_{50}$ and maximal effect ($E_{max}$): POD (point-of-departure) and wAUC (weighted area under the curve). The POD is the lowest concentration where its response equivalent to the assay-specific noise threshold. It allows more straight-forward potency comparison across assays than $EC_{50}$, especially when $E_{max}$ varies. The wAUC represents the total effect and is designed to allow value comparison when testing concentration range is different. The wAUC also shows better reproducibility than the other potency metrics [10]. The Curvep program [11] implements these activity metrics as well as the functions to identify nonsystematic artifacts in qHTS such as compound activity carryover (Fig. 2a) or baseline shift. In the pipeline, protocols of handling fluorescence/quenching/cytotoxicity were also implemented by taking the advantage of availability of counter-screens with comparable assay condition in Tox21. Herein, Tox21 glucocorticoid receptor (GR) assays in agonist mode and antagonist mode, as well as counter-screens for auto-fluorescence and cytotoxicity are used as an example to demonstrate the influence of artifacts on signal reproducibility and number of actives.

## 2    Materials

*2.1    Assays*        The Tox21 GR assays (both agonist mode and antagonist mode) are based on β-lactamase reporter gene technology. The activation of the reporter gene due to the binding of ligand-bound GR to the glucocorticoid response element can be detected by the shift of fluorescence emission (from green color to blue color). The details of assay protocols can be found in PubChem (https://pubchem.
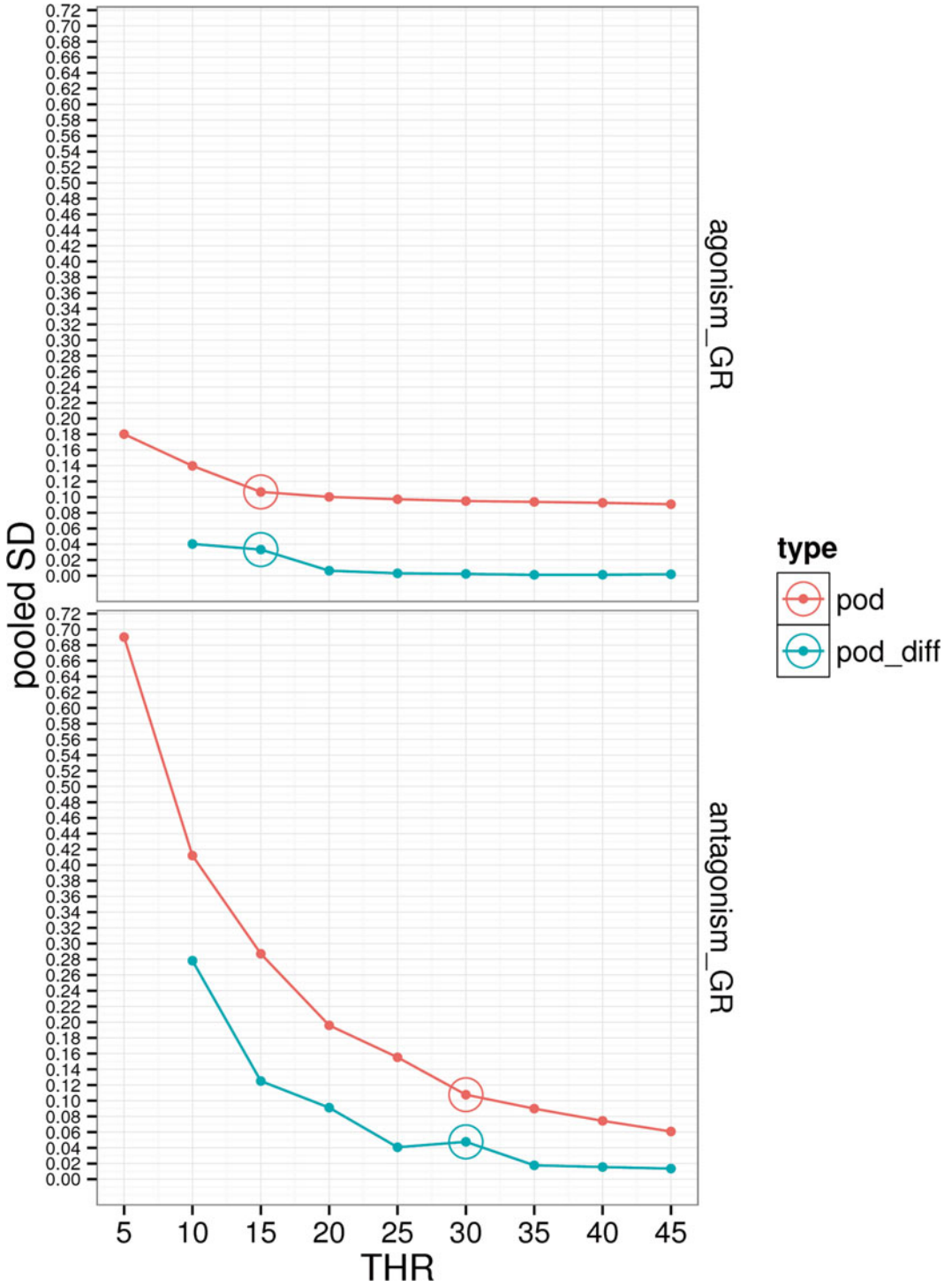
**Fig. 1** The pooled SD as the function of baseline threshold in GR assays (agonist mode and antagonist mode); the circle represents the threshold that was selected in the assays

ncbi.nlm.nih.gov/, AID: 720719 and 720725). The protocols of counter-screen assays, including auto-fluorescence and cell viability assays, can be also found in PubChem (AID: 720678 and 720693). The β-lactamase report gene assay has two assay readouts: background readout (channel 1, ch1, or green channel) and gene expression readout (channel 2, ch2, or blue channel), which are used to calculate a ratio (ch2/ch1) for analysis [12].

**2.2  Computer Programs**

Curvep is a qHTS noise filter algorithm written in C++ and can be downloaded from Github (https://github.com/sedykh/curvep). For detailed algorithm of Curvep, please refer to the previous chapter. An R Shiny based graphical interface was developed for Curvep and can be downloaded from Github (https://github.com/moggces/qHTSPipelineGUI).

# 3  Methods

**3.1  Raw Read Normalization (See Note 1)**

Generally, in Tox21 qHTS assays, the raw plate reads for each titration point were normalized relative to the reads in the wells with assay-specific positive control and DMSO-only wells as follows: % Response $= [(V_{compound} - V_{DMSO})/(V_{pos} - V_{DMSO})] \times 100$, where $V_{compound}$ denotes the compound well values, $V_{pos}$ denotes the median value of the positive control wells, and $V_{DMSO}$ denotes the median values of the DMSO-only wells. The data set was then corrected using the DMSO-only compound plates at the beginning and end of the compound plate stack by applying an NCATS in-house pattern correction algorithm. The response data as input for Curvep must be presented as percentage with baseline set as 0 % and the molar concentration ($M$) is transformed by logarithm with base 10.

**3.2  Preferred Response Direction Determination (See Note 2)**

It is preferable to adjust the directionality of the response data that matches with the intuition. For example, the directionality of green channel readout and cell viability assay should be "down," suggesting the readouts record compound inhibition effects. On the other hand, the directionality of ratio data and blue channel readout in the agonist mode of assays should be "up." The Curvep determines the data directionality (up or down) based on the sign of input parameter Range ($RNG$).

**3.3  Noise Threshold Identification (See Note 3)**

Assay-dependent noise threshold (*THR* parameter) was applied when analyzing concentration-response data using Curvep. *THR* is the key factor for calculation point-of-departure (POD) since POD is defined as the concentration where the response equivalent to the *THR*. The optimal threshold is defined as the *THR* that can sufficiently reduce POD variance and further increasing *THR* can only reduce POD variance to a certain degree. A range of *THR*s (5–45 %

with 5 % increment) were applied and a cutoff of 0.02 log10 unit of pooled SD of POD was suggested. In Fig. 1, the noise threshold finding process is demonstrated using GR data as an example. For the GR assay (agonist mode), when increasing the *THR* from 15 to 20 %, the degree of SD difference between 15 and 20 % is less than 0.02 (0.006); thus 15 % was selected. Similarly in the GR assay (antagonist mode), when increasing the *THR* from 30 to 35 %, the degree of SD difference between 30 and 35 % is less than 0.02 (0.017); thus 30 % was selected. The pooled POD variance/SD was calculated using duplicates in the Tox21 library, where 88 duplicates were intentionally plated on each plate for quality control.

***3.4  Other Parameters in Curvep (See* Note 4*)***

The carryover threshold (*CRO*) was set as 80 % for all the readouts except for the ratio and blue channel data in the agonist mode of the GR assay, where 60 % was applied. The *USHAPE* parameter was applied using the default value 4. *BYHI* parameter was turned on. The maximum deviation (*MXDV*) parameter (5 %, default value) and other parameters were universally applied in Tox21 assays.

***3.5  Response Masking (See* Note 5*)***

The Curvep was applied using parameters described above on both preferred and nonpreferred response direction. In addition, extra masking techniques were applied on the preferred response direction of major readouts to capture potential signals that may be missed by using default parameters only: (1) the response at the highest tested concentration was masked; (2) for blue channel data, the Curvep-treated green channel data were used as input mask. To maximize the signals that can be captured, the relative precedence of data used in data collapsing is as follows: data generated based on default parameters > data generated based on masking using green data (if applied) > data based on masking at the highest tested concentration (if applied).

***3.6  Systematic Artifacts (See* Note 6*)***

Triplicate testing was performed in Tox21 qHTS assays. The median value (2 out of 3) was used to collapse the curve activity parameters, including POD, wAUC, $EC_{50}$, and $E_{max}$. Some systematic artifacts are known and handled in the pipeline: a compound is flagged as auto-fluorescent if it has significant signals in both GR agonist mode assay and auto-fluorescence assay (Fig. 2d); a compound is flagged with BLA normalization issue if the directions of signals in the blue channel readout and ratio data do not correspond (ratio/ch2 conflict, Fig. 2e); a compound is flagged as cytotoxic if the wAUC fold change of ratio data vs cell viability data is smaller than 6 (or 4 for the weak signals, see below "Hit calling" for the definition of weak signals) (Fig. 2f). The number of hits before and after the systematic artifacts filtering using GR assays is shown in Table 1. The quenching (or ratio/ch2 conflict) and cytotoxicity issues are the two most common artifacts in Tox21
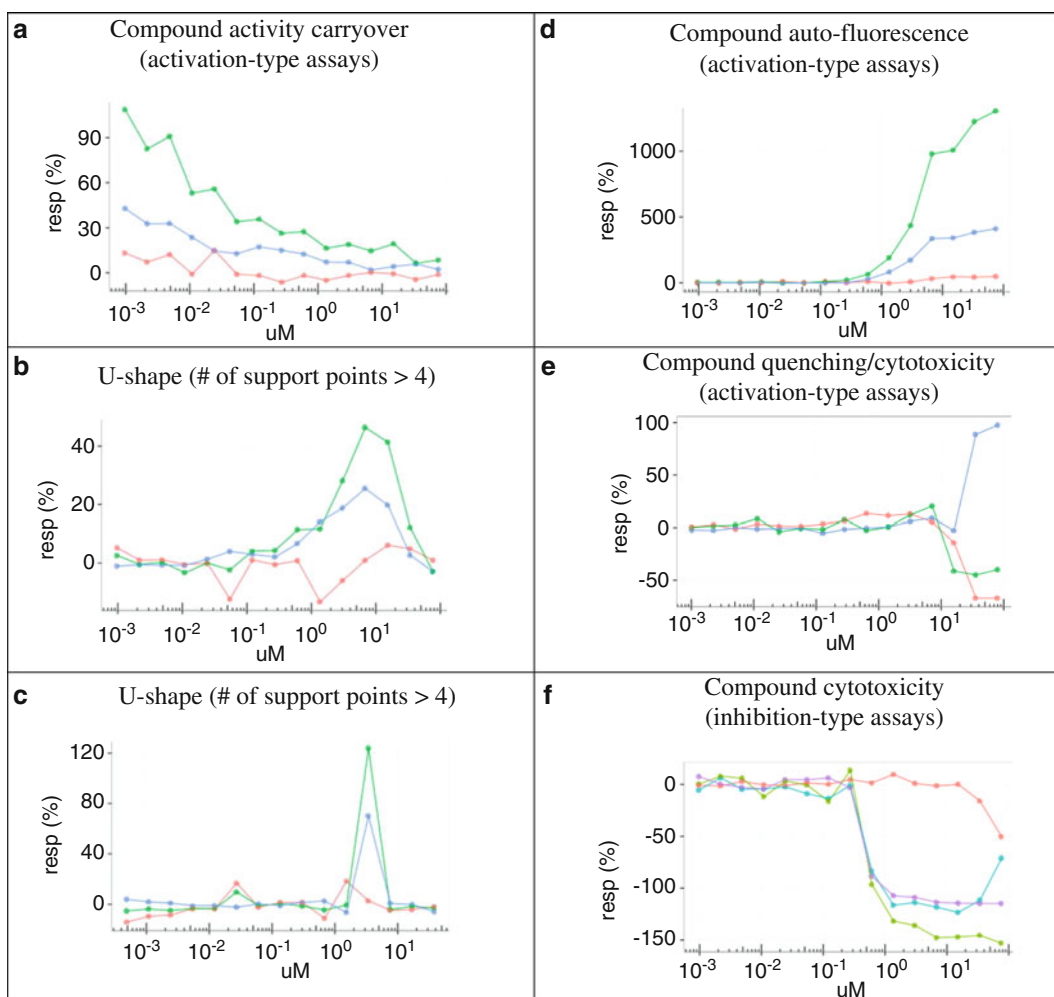
**Fig. 2** Types of potential assay artifacts; *blue*: ratio; *green*: ch2; *red*: ch1; *purple*: cell viability data if available; more examples of curves could be found in the reference [10]

GR agonist mode assay and antagonist mode assay, respectively; affecting about 60 or 44 % of potential hits in either agonist mode or antagonist mode assay.

*3.7 Hit Calling (See Note 7)*

The wAUC value of the curves in the major readout of the assay, which have $EC_{50} > 10\ \mu M$, are collected. The 50% percentile ($T50_{\%}$) and 25% percentile ($T25_{\%}$) of distribution are used as thresholds to stratify signals in the assays: if $wAUC \leq T50_{\%}$, the signal is considered as weak; if $wAUC \leq T25_{\%}$, the signal is considered as very weak. In addition to the flags for systematic artifacts, for the very weak signals, if not all triplicates are active, the compounds is flagged as "weak\noisy." Compounds without flags and $wAUC > 0$ are considered as actives.

**Table 1**
**The assigned groups of compounds in Tox21 GR assays**

| Category (based on ratio data only) | # | Flag | # |
|---|---|---|---|
| Agonism_GR ($n = 10496$) | | | |
| Inactive | 90.47 (9496) | No flag | 88.09 (9246) |
| | | Quenching/ratio normalization | 2.38 (250) |
| Significant signal (preferred direction) | 6.15 (645) | **No flag** | **2.20 (231)** |
| | | Fluorescent | 0.22 (23) |
| | | Quenching/ratio normalization | 3.73 (391) |
| | | Cytotoxicity | NA |
| Weak\noisy | 2.08 (218) | Weak\noisy | 2.08 (218) |
| Significant signal (nonpreferred direction) | 1.31 (137) | No flag | 1.03 (108) |
| | | Fluorescent | 0.17 (18) |
| | | Quenching/ratio normalization | 0.10 (11) |
| | | Cytotoxicity | NA |
| Antagonism_GR ($n = 10496$) | | | |
| Inactive | 81.64 (8569) | No flag | 80.82 (8483) |
| | | Quenching/ratio normalization | 0.82 (86) |
| Significant signal (preferred direction) | 10.78 (1131) | **No flag** | **5.82 (611)** |
| | | Fluorescent | 0.11 (12) |
| | | Quenching/ratio normalization | 0.12 (13) |
| | | Cytotoxicity | 4.72 (495) |
| Weak\noisy | 3.01 (316) | Weak\noisy | 3.01 (316) |
| Significant signal (nonpreferred direction) | 4.57 (480) | No flag | 2.72 (286) |
| | | Fluorescent | 0.20 (21) |
| | | Quenching/ratio normalization | 1.65 (173) |
| | | Cytotoxicity | NA |

*bolded text: hits in the assays after accounting for artifacts

# 4  Notes

1. The raw read normalization should depend on how assays are designed. For lower-throughput assays, usually only DMSO-only wells are available. The normalization method will be different from what was described in this chapter and should be tailored specifically.

2. For activation-type assays, it is quite common to have maximal response higher than 100%. It is suggested to set *RNG* as $10^6$ or $-10^6$ when running Curvep.

3. *THR* is the most important parameter in Curvep. On one hand, if *THR* is too low, too much noise is kept in the data (which can be detected by number of warnings ["remark"] reported by Curvep). On the other hand, if the *THR* is too high, some true signals could be removed. There are many different approaches to determine the noise threshold in addition to the one described in the Methods. For example, 3 SD of responses in DMSO-only wells or 10 mean absolute deviation (MAD) of responses in the lowest concentration could be applied [13, 14]. Alternatively, the optimal threshold could be the threshold that noise has been sufficiently filtered so that signal reproducibility is acceptable based on a certain degree of measurement [10]. For the approach described in the Methods, the data with as few as 10 compounds with triplicate testing have been applied to derive the threshold [15]. In addition, bootstrap sampling has been applied to generate hypothetical curves when data at each concentration are based on pooling from multiple plates [15]. When calculating the pooled variance across chemicals, the highest tested concentration was set as POD if the curve is flat (inactive) after Curvep.

4. It is found that maximum deviation (*MXDV*) parameter has limited effect on the Curvep results in qHTS assays [5]. Thus, the default value 5% was applied. Although Curvep corrects the response pattern based on monotonicity assumption, it allows certain degree of nonmonotonicity in the data, which may result from cytotoxicity. *USHAPE* = 4 was suggested to identify "u-shape" curve (Fig. 2b,c) constructed by at least four points except for the very low-noise assays, where *USHAPE* = 3 might be applied. The robot pin-tool compound activity carryover across assay plates is a unique nonsystematic artifact in Tox21 qHTS assays (Fig. 2a), which often presents monotonic property that can be confused with the true signal induced by very strong agonists in agonist mode assays due to receptor saturation. *CRO* parameter helps to differentiate these two cases, where any monotonic curves will be set as inactive if the response at lowest concentration is below the *CRO*. *CRO* = 60% (80%) were applied for agonist (antagonist) mode data except for the very low-noise assays (30%). In addition, using the plate sequence information in the qHTS assays can also help differentiate the carryover and the strong agonists. *BYHI* parameter provides a more conservative estimation on POD than *BYLO* (i.e., higher POD value).

5. Many of the issues (i.e., nonmonotonicity and systematics artifacts) presented in this chapter are due to compound tested concentration range lying in the cytotoxicity zone of the testing assays. Unlike low-throughput assays where compound testing concentration range can be optimized individually, compounds usually are tested in the same concentration in all the assays. In the blue channel data, the degree of nonmonotonicity is more severe than in the ratio data since in the ratio data, background (green channel) is used to smooth the effect (blue channel). Similar masking techniques can be applied on the single-channel activation-type assays (e.g., agonist mode in luciferase reporter gene assays, AID: 743122). Since there is no background channel in the luciferase reporter gene assays, a potent-active curve may be transverse both directionality due to cytotoxicity. The Curvep-treated "down" direction data can be used as input mask. Also, at the highest concentration, an abrupt decrease of response can be seen sometimes, which may be due to compound solubility or cytotoxicity issue. Masking the response at the highest concentration might help to capture these "U-shape" but not so potent signals.

6. The fold change of $EC_{50}$ between assays and counter-screens (either cell viability assays or auto-fluorescence assays) have been used to filter the artifacts [16]. In addition, studies have been conducted to filter cytotoxic compounds based on the $Z$-score value calculated using assay $EC_{50}$ value and the distribution of $EC_{50}$ values from various cell viability assays, either in different cell lines or by different assay technologies [17]. The ratio normalization (ratio/ch2 conflict) issue is common in β-lactamase assays [10] and can be related to either quenching effect or cytotoxicity. Some known artifacts are not considered in the current pipeline such as frequent hitters [18] in β-lactamase-specific assays (e.g., anti-tubulin inhibitors), Pan Assay Interference Compounds (PAINS) [19], compound aggregation [9], etc.

7. The goal of hit calling in assays is to identify the signals that can be reproduced in an independent assay which conducts in another time points. Depending on the goal of the study, hit calls can be determined with different criteria and gray zone between hits and nonhits. For example, for toxicological studies, false positives are more tolerable; thus the gray zone could be smaller. In general, compounds with higher activity values (e.g., POD or wAUC) tend to be more reproducible.

## References

1. Macarron R, Banks MN, Bojanic D et al (2011) Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov 10:188–195. doi:10.1038/nrd3368

2. Judson RS, Houck KA, Kavlock RJ et al (2009) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118:485–492. doi:10.1289/ehp.0901392

3. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121:756–765. doi:10.1289/ehp.1205784

4. Shockley KR (2014) Quantitative high-throughput screening data analysis: challenges and recent advances. Drug Discov Today. doi:10.1016/j.drudis.2014.10.005

5. Sedykh A, Zhu H, Tang H et al (2011) Use of in vitro HTS-derived concentration–response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect 119:364–370. doi:10.1289/ehp.1002476

6. Simeonov A, Jadhav A, Thomas CJ et al (2008) Fluorescence spectroscopic profiling of compound libraries. J Med Chem 51:2363–2371. doi:10.1021/jm701301m

7. Imbert P-E, Unterreiner V, Siebert D et al (2007) Recommendations for the reduction of compound artifacts in time-resolved fluorescence resonance energy transfer assays. Assay Drug Dev Technol 5:363–372. doi:10.1089/adt.2007.073

8. Thorne N, Shen M, Lea WA et al (2012) Firefly luciferase in chemical biology: a compendium of inhibitors, mechanistic evaluation of chemotypes, and suggested use as a reporter. Chem Biol 19:1060–1072. doi:10.1016/j.chembiol.2012.07.015

9. Feng BY, Simeonov A, Jadhav A et al (2007) A high-throughput screen for aggregation-based inhibition in a large compound library. J Med Chem 50:2385–2390. doi:10.1021/jm061317y

10. Hsieh J-H, Sedykh A, Huang R et al (2015) A data analysis pipeline accounting for artifacts in Tox21 quantitative high-throughput screening assays. J Biomol Screen 1087057115581317. doi: 10.1177/1087057115581317

11. Sedykh A (2015) Curvep. https://github.com/sedykh/curvep. Accessed 3 June 2016.

12. Qureshi SA (2007) Beta-lactamase: an ideal reporter system for monitoring gene expression in live eukaryotic cells. BioTechniques 42:91–96

13. Huang R, Xia M, Cho M-H et al (2011) Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ Health Perspect 119:1142–1148. doi:10.1289/ehp.1002952

14. Filer DL, Kothiya P, Setzer WR et al (2015) EPA pipeline overview. https://www.epa.gov/sites/production/files/2015-08/documents/pipeline_overview.pdf. Accessed 25 March 2016

15. Behl M, Hsieh J-H, Shafer TJ, Mundy WR, Rice JR, Boyd WA et al Use of alternative assays to identify and prioritize organophosphorus flame retardants for potential developmental and neurotoxicity. Neurotoxicology and Teratology; doi:10.1016/j.ntt.2015.09.003

16. Huang R, Sakamuru S, Martin MT et al (2014) Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Sci Rep. doi:10.1038/srep05664

17. Judson RS (2014) ToxCast data processing overview. http://epa.gov/comptox/download_files/chemical_prioritization/Judson_CoP_Dec2014.pdf.Accessed 14 Aug. 2015

18. Schorpp K, Rothenaigner I, Salmina E et al (2014) Identification of small-molecule frequent hitters from alphascreen high-throughput screens. J Biomol Screen 19:715–726. doi:10.1177/1087057113516861

19. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740. doi:10.1021/jm901137j

# Accessing the High-Throughput Screening Data Landscape

## Daniel P. Russo and Hao Zhu

## Abstract

The progress of high-throughput screening (HTS) techniques is changing the chemical data landscape by producing massive biological data from tested compounds. Public data repositories (e.g., PubChem) receive HTS data provided by various institutes and this data pool is being updated on a daily basis. The goal of these data sharing efforts is to let users quickly obtain the biological data of target compounds. Without a universal chemical identifier, the repositories (e.g., PubChem) provide users various methods to query and retrieve chemical properties and biological data by several different chemical identifiers (e.g., SMILES, InChIKey, and IUPAC name). The major challenge for most users, especially computational modelers, is obtaining the biological data for a large dataset of compounds (e.g., thousands of drug molecules) instead of a single compound. This chapter aims to introduce the steps to access the public data repositories for target compounds with specific emphasis on the automatic data downloading for large datasets.

**Key words** Compounds, Chemical identifier, Biological data, PubChem

## 1  Introduction

In the past decade, the use of high-throughput screening (HTS) in drug discovery and chemical toxicity evaluations has greatly facilitated the progress of cheminformatics studies. The massive data generated from HTS studies provides scientists a new vision of the biological effects induced by the compounds being tested. As a consequence, data sharing efforts to make these data easily available to communities have been undertaken within the same period. Data repositories, such as PubChem, ChEMBL, BindingDB, and the Comparative Toxicogenomics Database (CTD), have become daily-used scientific tools for many scientists. With a chemical identifier (e.g., InChIKey), the biological data for a target compound can be obtained by clicking a button on the web based search portals of these repositories.

The PubChem project (https://pubchem.ncbi.nlm.nih.gov), initiated and hosted by the National Center for Biotechnology Information (NCBI) of the National Institutes of Health (NIH), is the largest public chemical data source. The goal of PubChem is to

provide information for the biological activities of small molecules, and as of September 2015, it has received over 60 million unique chemical structures and 1 million biological assays from over 350 contributors [1]. The PubChem data repository consists of three primary databases: Substance, Compound, and BioAssay. The PubChem Substance database, indexed by the PubChem substance identifier (SID), contains chemical structures, synonyms, registration IDs, descriptions, related links, database cross-reference links to PubMed, protein 3D structures, and biological screening results. The PubChem BioAssay Database contains experimental testing results of the chemical substances described within the PubChem Substance database. Each HTS assay has a unique assay identifier (AID) in this database. The PubChem Compound database contains validated chemical depiction information that is provided to describe substances in the PubChem Substance database. The chemical identifier, PubChem compound ID (CID), records unique structural information and, in turn, allows target compounds to be queried within the PubChem Compound database. Compound records are supplemented with textual structural identifiers such as Simplified Molecular-Input Line-Entry Systems (SMILES), the IUPAC International Chemical Identifier (InChI), and a more compact version, the InChI key [2–5].

PubChem HTS data were obtained from various sources including university, industry, or government laboratories. One of the initial missions of PubChem is to function as the repository to host the data generated by the HTS projects supported by the NIH's Molecular Libraries Program. A full listing of data sources can be found at http://pubchem.ncbi.nlm.nih.gov/sources#assay. The types of HTS data from these sources include results obtained from binding assays, functional cell assays, and Absorption, Distribution, Metabolism, Exclusion and Toxicity (ADMET) assays. The HTS data points can be qualitative (e.g., active or inactive), quantitative (e.g., the half-maximal effective concentration of a drug), or both. PubChem stores the HTS data of a compound in two fields: the *activity outcome* and the *active concentration*. The *activity outcome* either identifies the relevant compound as a chemical probe (i.e., a positive control of a HTS assay) or qualitatively transforms the experimental data into one of the following categories: active, inactive, unspecified/inconclusive, or untested. On the other hand, the *active concentration* stores the HTS data quantitatively as a concentration value in μM units as well-defined biological endpoints, such as the half-maximal activity response (e.g., IC50 and EC50).

There are two methods to obtain HTS data by accessing the data sharing repositories, such as PubChem. The data can be obtained by querying manually with individual compounds' textual chemical identifiers. However, if the goal is to download all relevant HTS data for a large set of compounds, automatic data extraction is needed. This chapter will use PubChem as an example to show how to obtain HTS data for target compounds, especially for a large set of compounds.

## 2   Materials

To extract the HTS data for target compounds via public data repositories (i.e., PubChem), the following software needs to be downloaded/installed on the computer:

– A web browser (e.g., Mozilla FireFox, Google Chrome, Microsoft Internet Explorer, and Apple Safari).

– Microsoft Excel® or other spreadsheet program.

– A programming package (e.g., Java, Python, Perl, and C#).

– A file archiver that supports .gz decompression such as WinZip or 7-zip (Windows users only).

## 3   Methods

### 3.1   Accessing HTS Data Manually Through the PubChem Portal

Similar to popular internet search portals (e.g., Google®), PubChem provides users a manual search function by which queries can be made using various chemical identifiers. Each unique compound in the PubChem Compound database has an individual page listing standardized chemical information and properties, including a list of all submitted biological testing results. For a target compound existing in the PubChem database, its biological testing data can be exported and downloaded as a comma-separated values (CSV) file and managed using Microsoft Excel®. Figure 1 shows the screenshot of the plain text file for the biological data for aspirin, with PubChem CID 2244 (downloaded from PubChem on February 15, 2016). The biological data of a compound is summarized by including not only the bioassay identifier (AID) and the associated testing results, but also detailed information of the bioassays and the definitions of the activities. This file can be obtained by inputting various identifiers of aspirin to their appropriate categories. Figure 2 shows a screenshot of the homepage to the PubChem search function. More information on the appropriate search option for a given identifier can be found in the Subheading 4. The biological data of a single target compound can be accessed by the following steps:

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AID | Panel ID | SID | CID | Activity | AC Value (micromolar) | AC Name | BioAssay Name | GI | Target Name |
| 2 | 157039 | 0 | 103164874 | 2244 | Unspecified | 7.76E+10 | IC50 | Antiaggregatory effect against arachidonic acid induced platelet aggregation in the human platelet rich plasma(PRP) | | |
| 3 | 89416 | 0 | 103164874 | 2244 | Unspecified | 7.00E+07 | IC50 | Inhibition of collagen induced platelet aggregation in platelet-rich human plasma (PRP) | | |
| 4 | 376764 | 0 | 103164874 | 2244 | Unspecified | 49700 | IC50 | Inhibition of sheep placental COX2 by liquid scintillation counter | | |
| 5 | 313117 | 0 | 103164874 | 2244 | Unspecified | 20000 | Ki | Inhibition of cyclooxygenase | | |
| 6 | 376763 | 0 | 103164874 | 2244 | Unspecified | 18800 | IC50 | Inhibition of COX1 from ram seminal vesicles by liquid scintillation counter | | |
| 7 | 162666 | 0 | 103164874 | 2244 | Unspecified | 18000 | IC50 | Tested for inhibitory activity against Prostaglandin G/H synthase 2 from human | | |
| 8 | 349142 | 0 | 103164874 | 2244 | Unspecified | 7530 | Ki | Inhibition of human carbonic anhydrase 1 esterase activity by noncompetitive Lineweaver-Burke plot | 115449 | Carbonic anhydrase 1 |
| 9 | 625305 | 0 | 103164874 | 2244 | Unspecified | 6000 | IC50 | Antiinflammatory activity against TNF-alpha-induced VCAM-1 protein expression in HAEC by ELISA | | |
| 10 | 399402 | 0 | 103164874 | 2244 | Unspecified | 5000 | IC50 | Inhibition of sheep placental cotyledons COX2 assessed as PGE2 production | 3914304 | Prostaglandin G/H synthase 2 |

**Fig. 1** Example of a subset of the biological testing results for aspirin (PubChem CID 2244) downloaded in plain text format

**Fig. 2** The PubChem search tool interface as of February, 2016

| | |
|---|---|
| Step 1 | Open a web browser and visit the PubChem Compound search tool at: https://pubchem.ncbi.nlm.nih.gov/search/search.cgi |
| Step 2 | Select the appropriate search tab |
| Step 3 | Enter the correct information (e.g., chemical name as shown in Fig. 2) and click "Search". Using a unique identifier (e.g., PubChem CID) will result in the desired compound. Otherwise, manually analyzing the search results (i.e., a list of compounds containing the input information) is required |
| Step 4 | From the compound summary page, scroll down to "BioAssay Results". Click "Refine/ Analyze" and select "Go To Bioactivity Analysis Tool" from the pull-down menu |
| Step 5 | On the Bioactivity Analysis Tool page, click "Download Table" |

The resulting bioassay information for that compound will be automatically retrieved as a plain text file.

*3.2 Retrieving PubChem HTS Data Through Web Services*

If the goal is to download the HTS data for a large dataset (e.g., consisting of more than 1000 compounds), automatic querying is needed by executing a coding script. To this end, PubChem offers specialized data retrieval services through a programmatic interface: PubChem Power User Gateway (PUG). The PUG provides quick access to PubChem data retrieval functions. Information on all the available PUG services can be found in the reference [6] as well as within the PubChem portal (https://pubchem.ncbi.nlm.nih.gov/pug/pughelp.html). The most broadly applicable function to retrieve HTS data for a large chemical dataset is PUG-REST. PUG-REST, which uses a Representational State Transfer (REST)-style interface, allows users to construct Uniform Resource Identifiers (URLs) to retrieve data from PubChem. Through this way, PUG-REST is easily integrated with all programming/ scripting languages that can post URLs (e.g., Java, Python, Perl, and C#). Using PUG-REST, multiple records can be accessed

automatically to fulfill the request of retrieving HTS data for large datasets. More information on all the features available through PUG-REST can also be found on the PubChem portal (https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html).

To construct a URL for PUG-REST data retrieval, a text URL string needs to be created and it contains four parts: base, input, operation, and output. The construction of this kind of URL is shown in Figure 3. The input section of the URL describes the target database (BioAssay, Compound, or Substance) to be queried, the category of the identifier, along with the identifier information. The operation section designates the information to be retrieved ("assay-summary" in this case to retrieve HTS data). The output section specifies the format of the output file. Figure 3 also shows several examples of URLs used to retrieve HTS data through PUG-REST.

Inputting a constructed URL into a web browser will result in the display of bioassay information for this target compound in the desired format. Within a programming script, individual URLs can be constructed for each compound in a large dataset in an automated fashion. In this case, the HTS data for all compounds in this dataset will be retrieved in the preferred output format.

*3.3 Downloading Master HTS Database from PubChem*

While the PubChem database is exceptionally large, sometimes it is necessary to transfer all the HTS data from PubChem to a local server for further analysis. File Transfer Protocols (FTP), which are common tools to share/transfer files over the internet, can be used to realize this goal. PubChem offers the download of all three databases through an FTP site. Most operating systems (e.g., Windows, Mac, and Linux) support access to FTP sites, allowing for easy file transfers from the data server (e.g., PubChem) to the user's local computer. Using an FTP, the entire PubChem BioAssay database can be accessed and downloaded in four formats: Abstract Syntax Notation (ASN), CSV, JavaScript Object Notation (JSON), and Extensible Markup Language (XML). The overall BioAssay database is large and requires all entries to be distributed into compressed folders (in .zip format) with each folder containing a maximum of 1000 assays and their associated assay data. Folders are named after the set of AIDs it contains, where an individual file in these folders is



| Base | Input | | Operation | Output |
|---|---|---|---|---|
| http://pubchem.ncbi.nlm.nih.gov/rest/pug/ | compound/cid/2244/ | | assaysummary/ | CSV/ |
| | compound/name/aspirin/ | | | ASNB/ |
| | compound/smiles/CC(=O)OC1=CC=CC=C1C(=O)O/ | | | XML/ |
| | compound/inchikey/BSYNRYMUTXBXSQ-UHFFFAOYSA-N/ | | | JSON/ |
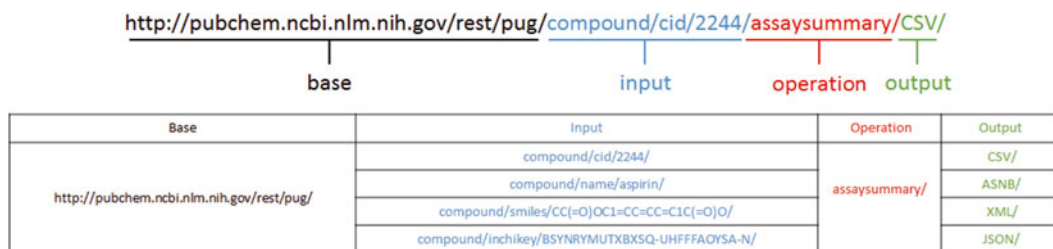
**Fig. 3** Various queries to retrieve PubChem HTS data for aspirin using PUG-REST

named after the AID information it contains. For example, within the folder "0000001_0001000" all the HTS data corresponding to AIDs 1-1000 can be found. Likewise, the file in this folder, "1.xml", contains data from assay with AID 1. For storage efficiency, these files are compressed using the GZip algorithm (.gz format). While both Linux and Mac users will find support for GZip decompression, Windows users are required to download a file archiver program that supports GZip decompression (e.g., WinZip or 7zip). The following steps are needed to access these data:

| Step 1 | Connect to the PubChem FTP site containing the BioAssay database files |
| --- | --- |
| | For Windows users: open a Windows Explorer window by clicking the "Start" button and then clicking "Computer". In the address bar, type ftp://ftp.ncbi.nlm.nih.gov/pubchem/Bioassay/ |
| | For Mac users: Click "Go" then click "Connect to Server…". In the "Server Address" field type: ftp://ftp.ncbi.nlm.nih.gov/pubchem/Bioassay/. When asked to enter name and password, choose to connect as "Guest" |
| | For Linux users: Open any file manager window. Click "File" and then click "Connect to Server…". In the "Server" field type ftp://ftp.ncbi.nlm.nih.gov/pubchem/Bioassay/ |
| Step 2 | Select the folder containing the desired format. For example, the "XML" folder contains the bioassay database in XML format |
| Step 3 | Copy the compressed folders to the desired directory on the local computer |

It is important to realize that PubChem updates its databases frequently.

## 4  Notes

The PubChem search portal accepts a variety of search options. For example, exact queries (under the "Name/Text" tab) can be performed based on general identifiers, such as CID, IUPAC name, etc. Alternatively, queries can also be performed based on chemical structure information. Under the "Identity/Similarity" search option, structural identifiers (i.e., SMILES or InChI) can be used. Commonly used chemical structure file formats, such as Structure Data File (SDF) format, can also be accepted under this search tab. If no identifier is available for a compound, there is also an option to draw the chemical structure using the drawing tool.

## 5  Conclusions

HTS data can be accessed through public repositories for target compounds. Since there are several existing methods to obtain HTS data, the individual researcher can choose the suitable way to

get the HTS data based on the actual needs (e.g., the size of the target dataset). Meanwhile, the HTS data retrieval procedure needs necessary computer skills and background knowledge of the data sources, such as what have been described for PubChem data acquired in this chapter.

## References

1. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, et al. (2015) PubChem Substance and Compound databases. Nucleic Acids Res 44:D1202–D1213.

2. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36

3. Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci 29:97–101

4. Weininger D (1990) SMILES. 3. DEPICT. Graphical depiction of chemical structures. J Chem Inf Comput Sci 30:237–243

5. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I (2013) InChI - the worldwide chemical structure identifier standard. J Cheminformatics 5:7

6. Kim S, Thiessen PA, Bolton EE, Bryant SH (2015) PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. Nucleic Acids Res 43:W605–W611

# Chapter 17

## Curating and Preparing High-Throughput Screening Data for Quantitative Structure-Activity Relationship Modeling

### Marlene T. Kim, Wenyi Wang, Alexander Sedykh, and Hao Zhu

### Abstract

Publicly available bioassay data often contains errors. Curating massive bioassay data, especially high-throughput screening (HTS) data, for Quantitative Structure-Activity Relationship (QSAR) modeling requires the assistance of automated data curation tools. Using automated data curation tools are beneficial to users, especially ones without prior computer skills, because many platforms have been developed and optimized based on standardized requirements. As a result, the users do not need to extensively configure the curation tool prior to the application procedure. In this chapter, a freely available automatic tool to curate and prepare HTS data for QSAR modeling purposes will be described.

**Key words** QSAR, Data curation, Chemical structures, Computational modeling

## 1 Introduction

A typical high-throughput screening (HTS) data set can contain over 10,000 compounds (e.g., Antioxidant Response Element assay data listed as PubChem AID 743219). Although they are potential resources for developing Quantitative Structure-Activity Relationship (QSAR) models, normally these public HTS data sets cannot be used directly for modeling purposes due to the presence of duplicates, artifacts, and other issues. There are public chemical data repositories such as PubChem, ChemSpider, and ChEMBL that contain lots of HTS data available for download, but the original data stored in these resources still need further curation. However, HTS data sets are so large that it is very inefficient, and usually ineffective, to process all the data points manually. The assistance of automated tools is highly recommended.

Chemical structure curation and standardization is an integral step in QSAR modeling. This step is essential since it is likely the same compounds will be represented differently among different sources. For example, organic compounds could be drawn with implicit or explicit hydrogens, in aromatized or Kekulé form, as

well as in different tautomeric forms. These differences in chemical structure representations could influence the computed chemical descriptor values for the same compound and greatly affect the usefulness and quality of the resulting QSAR models. Furthermore, the existence of inorganic compounds and mixtures, which are not suitable for traditional QSAR modeling studies, also limits the use of public HTS data.

Another issue with HTS data is that it is very common for it to have an unbalanced distribution of activities, where there are substantially more inactive than active compounds. This unbalanced distribution of activities (i.e., low active ratio) could result in biased QSAR model predictions. Data sampling, an approach that selects and analyzes a subset of the overall data, can resolve this issue. The specific data sampling method that will be discussed in this chapter is down-sampling, since it is most relevant to HTS data processing. Down-sampling is an approach that ignores most of the data points that are in the largest activity category. This will allow you to select a sample of the inactive compounds from the data set to balance the distribution of activities for modeling. Furthermore, smaller data sets are easier to manage and, in most cases, more informative since it captures the most important elements of the data.

In this chapter, an automatic data curation process that can standardize/harmonize chemical structures and down-sample the results of a large HTS data set will be described. The approaches to construct the modeling and validations sets, including balancing the HTS activity via down-sampling, were configured using Konstanz Information Miner (KNIME ver. 2.10.1) (www.knime.org) workflows that utilize the two most common approaches for selecting a sample size: random and rational selection methods. These processes utilize basic statistical approaches [1] and will transform an original public HTS data set into a curated format suitable for QSAR model development and other relevant in silico modeling efforts. The quantitative high-throughput screening (qHTS) Antioxidant Response Element assay data obtained from PubChem (PubChem AID 743219) will be used to illustrate this data curation process.

## 2  Materials

Automated procedures to curate chemical structures and down-sample the large data set (i.e., HTS data) will be described in this chapter. All of the workflows were developed and executed in the open-source platform KNIME. The output files of the workflows are curated data sets with standardized structures that are ready to be processed by QSAR modeling tools. The workflows can be downloaded as a zip file at https://github.com/zhu-lab.

## 3 Methods

*3.1 Prepare an Input File for the Curation Workflow*

An input file should be a tab delimited multiple column *txt* file (*FileName.txt*) with a header to each column, where one column must contain the structure information as a SMILES code [3]. The input file (a sample file was provided within the zip file) should have at least three columns: *ID*, *SMILES*, and *activity*. If needed, other useful features of compounds (e.g., compound names) could also be included as extra columns.

*3.2 Prepare the Curation Workflow*

Install the KNIME software. It can be downloaded from www. knime.org. Download the curation workflow (https://github. com/zhu-lab/curation-workflow) and extract the zip file into a computer directory.

*3.3 Configure the Workflow*

In the *File* menu bar of KNIME, select "*Import KNIME workflow…*" to import the structure standardizer workflow into KNIME. Now in the pop-up window (Fig. 1), click on "*Source: Select root directory*," find the computer directory that the zip file was extracted to. Select the destination directory, which will be where the output files will be saved to. In "*Workflows:*" select the "*Structure Standardizer*" workflow and click "*Finish*" (Fig. 2).
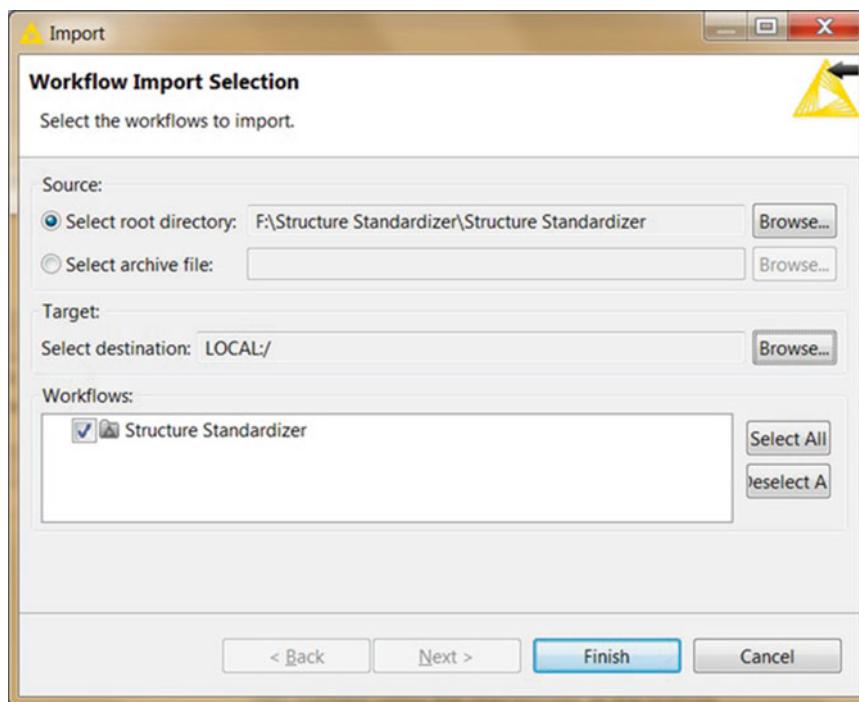


**Fig. 1** The KNIME "Workflow Import Selection" window
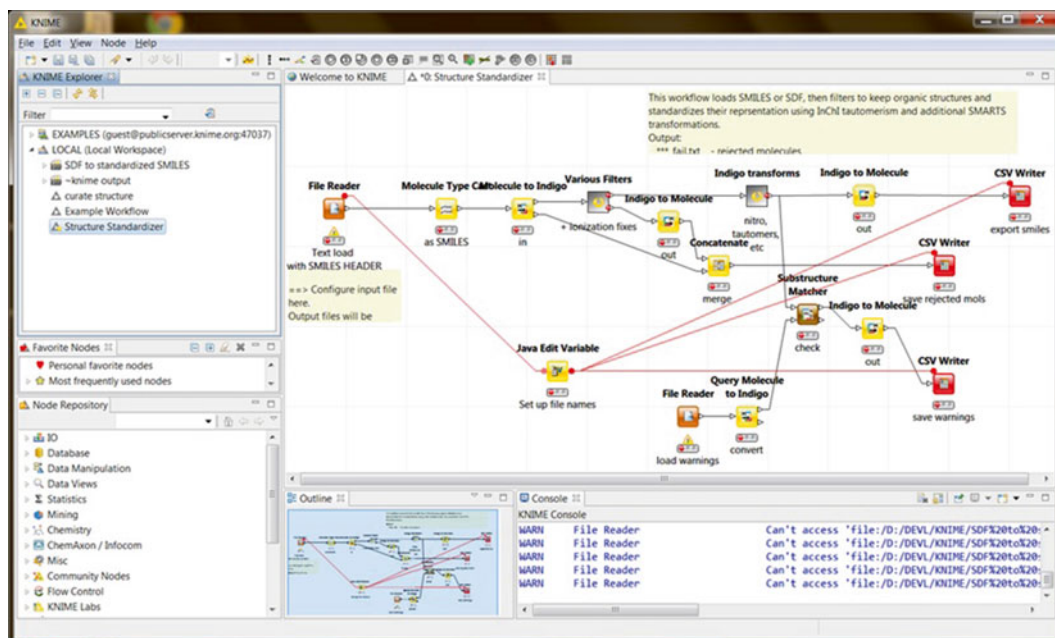
**Fig. 2** A window of the KNIME "Structure Standardizer" workflow

**3.4  Set
Up Parameters
and Run the Workflow**

To open the workflow, double click on the Structure Standardizer in the "*KNIME Explorer*" window under "*LOCAL (Local Workspace)*" located in the top left side bar. At this time, the workflow will show up in the main space, which is called the workflow editor (*see* **Note 1**) Right click the "*File Reader*" node and select "*Configure.*" In the pop-up windows, input the valid file location of the input file that has been prepared in the previous step. Make sure the headers of the input file are read correctly. Click "*OK*" to save the changes and close the configuration window. Next, right click the "*Java Edit Variable*" node in the bottom left and change the variable *v_dir* to the directory of the folder where all the files are extracted in the second step. Then, configure sub-workflows individually by double clicking on each node. Within each sub-workflow, configure the Java Edit Variable node the same as described above. After closing the sub-workflow windows, the yellow lights on all the nodes should be on, indicating that the workflow is ready to be used. Click on the green "*double-arrow button*" located in the top menu bar to execute the whole workflow and the green lights on all nodes should be on. Three output files should have been generated in the same folder as the input file (*FileName_fail.txt*, *FileName_std.txt*, and *FileName_warn.txt*) (*see* **Note 2**). (Or the files will be in a folder directory substituting all spaces with % 20 if spaces are in the directory (e.g., if input file is in F:\Structure Standardizer\output, then output file would be in F:\Structure%20Standardizer\output).) The standardized

compounds will be in canonical SMILES format (*see* **Note 3**). The file *FileName_std.txt* is the data set curated for modeling purposes.

### 3.5 Preparing the Chemical Descriptor File

With the chemical structures curated, the chemical descriptors can be calculated by using various descriptor generators, such as RDKit (http://www.rdkit.org/), Molecular Operating Environment® (MOE) (https://www.chemcomp.com/MOE-Molecular_ Operating_Environment.htm), and Dragon® (http://www.talete. mi.it/products/dragon_description.htm) (*see* **Note 4**).

### 3.6 Preparing the Modeling and Validation Set Files

To develop a predictive QSAR model, the compound classifications in the modeling set need to be balanced (*see* **Note 5**) To this end, the inactive compounds of HTS data need to be downsampled to be similar to the number of actives in the modeling set. There are two methods that can be applied for this purpose: random and rational selection.

The random selection approach will randomly select an equal number of inactive compounds compared to the actives. Figure 3a shows a KNIME workflow that could be used to randomly select
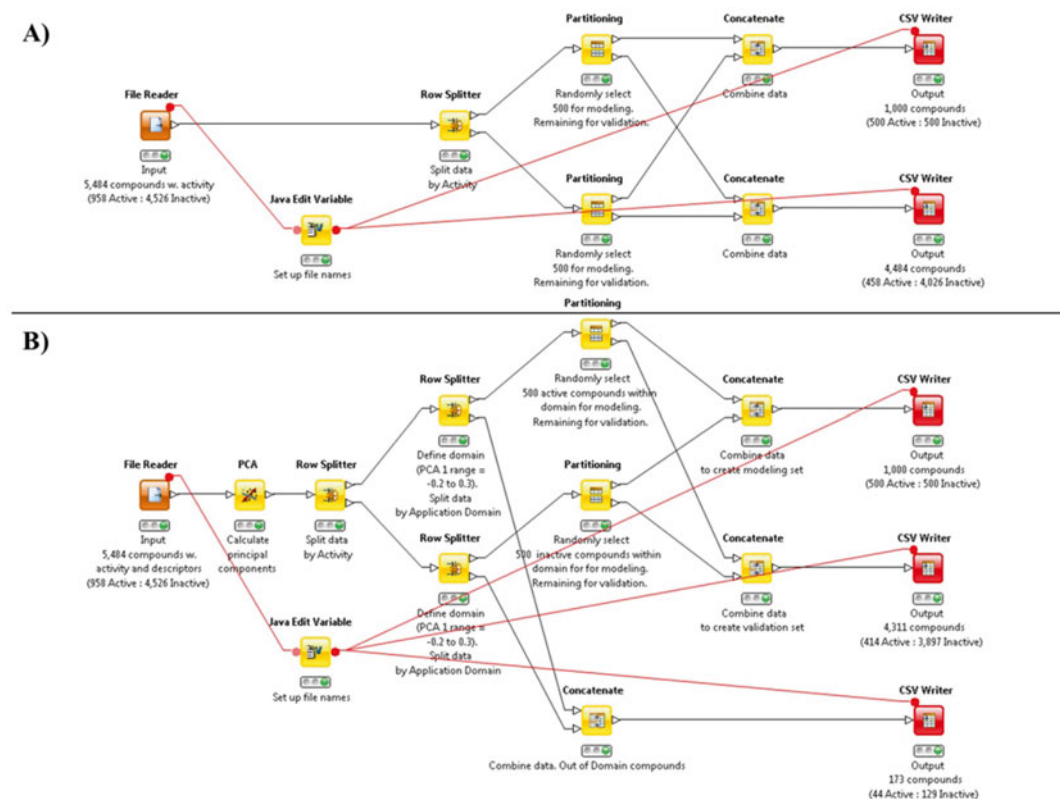


**Fig. 3** Example of KNIME workflow for selecting compounds and partitioning data set into modeling and validation sets using (**a**) random and (**b**) rational selection approaches

compounds and partition the data set into modeling and validation sets (*see* **Note** 6). This workflow ensures that the relationships between each compound selected for the model development and validation purposes were not explicitly selected. To run the workflow, first input the curated file (e.g., the file *FileName_std.txt*) from previous step with a minimum of two columns for the *ID* and *activity* in the "*File Reader*" node. Then, right click on the *activity* column header to open the "*Column Properties*" and set the "*Type*" as "*String.*" The workflow has already been configured to randomly select 500 active and 500 inactive compounds; however, the numbers of active/inactive compounds can be changed. Click on the green "*double-arrow button*" located in the top menu bar to execute the whole workflow. Two files will be generated in the destination directory: *ax_input_modeling.txt* and *ax_input_intValidating.txt*. The *ax_input_modeling.txt* file contains the 500 active and 500 inactive compounds randomly selected to balance the distribution of activities in the modeling set. The *ax_input_intValidating.txt* file contains the remaining compounds (e.g., 458 active and 4026 inactive compounds from the sample data set) that could be used for validation purposes.

Compared to random selection, rational selection is also frequently used in down-sampling (*see* **Note** 7). Figure 3b shows a KNIME workflow that could be used to rationally select compounds for QSAR model development, based on the threshold defined using principal component analysis (*see* **Note** 8), and partition the data set into modeling and validation sets. The rational selection approach uses a quantitatively defined threshold of similarity to select inactive to active compounds. In this case, inactive compounds that share the same descriptor space of active compounds will be selected and successively define the applicability domain in the resulting QSAR models [2]. The KNIME workflow described here differs slightly from the random selection workflow described above in that it allows one to quantitatively define the similarity threshold using PCA. To run the workflow, first input the curated file (e.g., *FileName_std.txt*) from the previous step with columns for the *ID*, *activity*, and *descriptors* into the "*File Reader*" node. Then right click on the *activity* column header to open the "*Column Properties*" and set the "*Type*" to "*String.*" The workflow has already been configured to select 500 active and 500 inactive compounds and the numbers of active/inactive compounds can be changed. Click on the green "*double-arrow button*" located in the top menu bar to execute the whole workflow. Three files will be generated in the destination directory: *ax_input_ratl_modeling.txt*, *ax_input_ratl_intValidating.txt*, and *ax_input_ratl_outAD.txt*.

**3.7 Verification: Visualizing the Chemical Space Covered by the Data Set Using Principal Components**

After the modeling and validation sets are created, the chemical space (*see* **Note** 9) can be visualized. The chemical space of a data set can be shown in a 3-D plot using the first three principal components (of the descriptor space) generated from MOE chemical descriptors (*see* **Note** 10). In Fig. 4a, the chemical space using the first three principal components of the entire ARE data set 7034

compounds was plotted. Then, 500 active and 500 inactive compounds organized the chemical space for the modeling set, as shown in Fig. 4b. The inactive compounds were selected based on the chemical similarity to the actives, so the chemical space occupied by the modeling set is clearly different from the whole data set. Therefore, the predictions of resulting QSAR models should be considered reliable within the chemical space (i.e., the applicability domain) of modeling set.

A principal component analysis was performed in KNIME on all the active and inactive compounds in the ARE data set of 5484 compounds (Fig. 5). From the scatter plot of principal components 1 versus 2, it was noticeable that most of the compounds clustered at principal component 1 values between –0.2 and 0.3. Therefore, the applicability domain of the resulting model can be defined as any compound that falls within this range. To adjust this applicability domain in the KNIME workflow, adjust both "*Row*
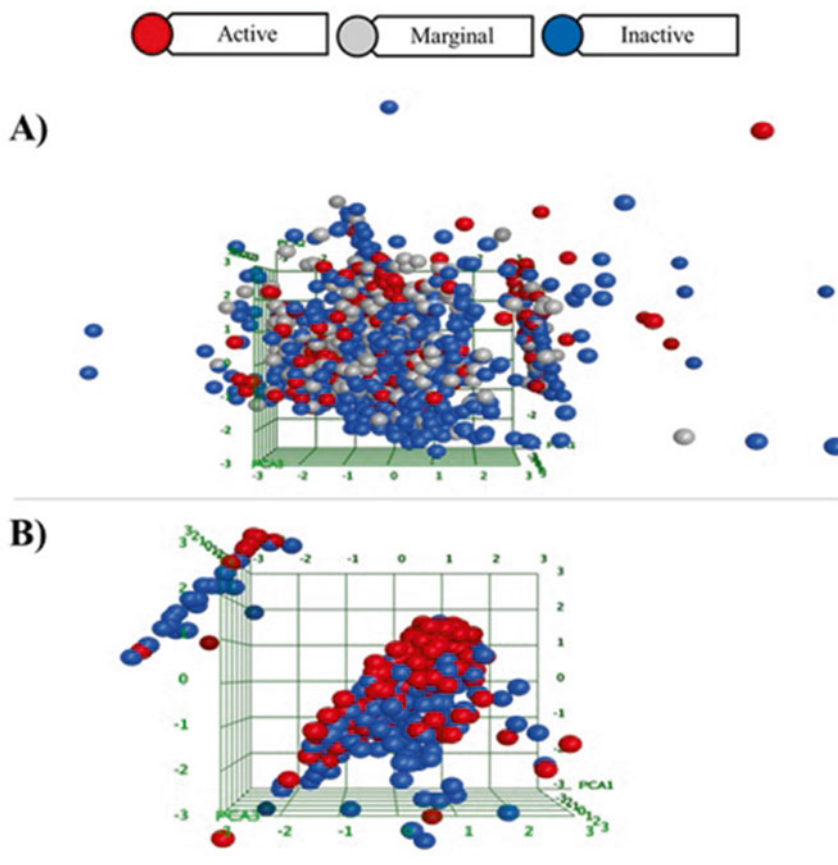


**Fig. 4** 3-D plots of ARE data set using (**a**) all 7034 data points, and (**b**) modeling set using principal components 1–3 generated using 10 MOE descriptors
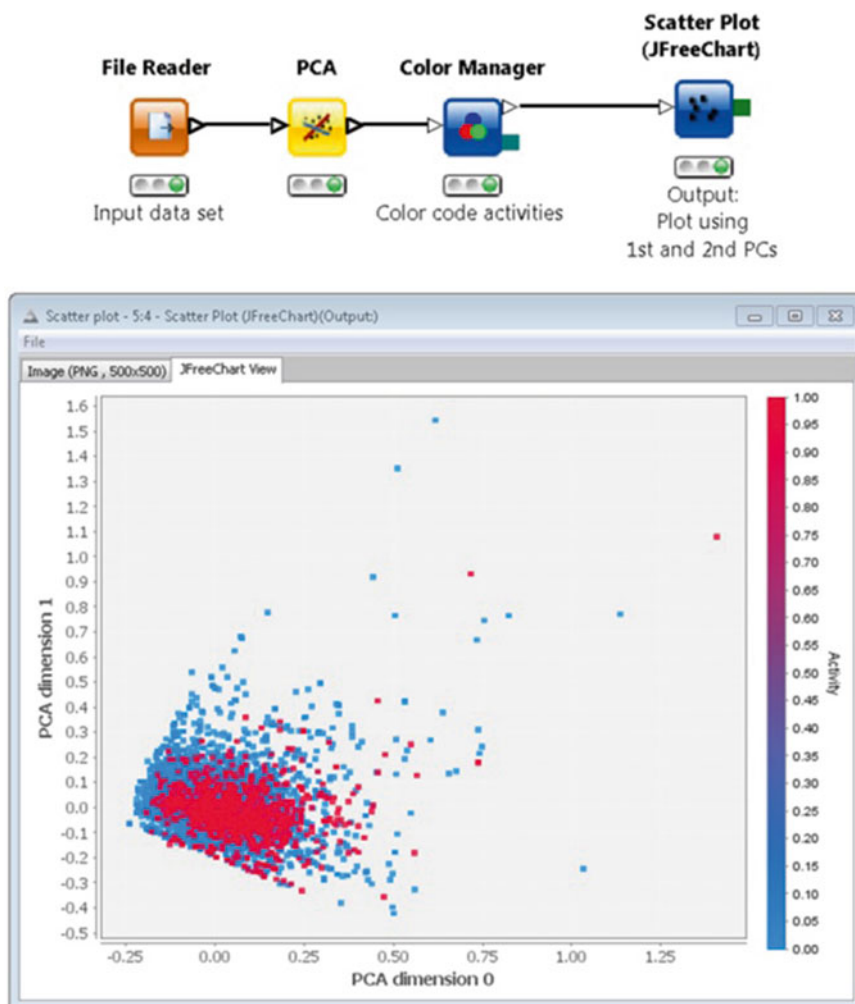
**Fig. 5** Example of KNIME workflow for visualizing the chemical space of all active and inactive compounds

*Splitter*" nodes by right clicking the node, under "*use range checking,*" adjust the "*lower bound*" and "*upper bound.*" Under this condition 500 active and 500 inactive compounds within the range of −0.2 and 0.3 will be selected for the modeling set, while the others will be placed into the validation set. Compounds that were out of domain will be placed into the *ax_input_ratl_outAD.txt* file.

## 4    Notes

1. If you cannot find these windows, go to the "*View*" in the menu bar and select "*Reset Perspective…*"

2. Description of the three output files:

*FileName_fail.txt* contains compounds that could neither be standardized nor be used in the QSAR modeling (e.g., mixtures, inorganics, and large molecules like polypeptides).

*FileName_std.txt* contains the remaining structurally standardized compounds in which the SMILES are curated as the canonical format.

*FileName_warn.txt* contains compounds with potential problems that require further review. For example, compounds with positive/negative charges need to be compared to their original structures to decide the correct structure information. These compounds with warnings will not be removed from the data set and are included in the *FileName_std.txt file*.

3. Compounds in this file are curated, standardized, and represented in canonical form by removing metals, de-isomerizing tautomer, neutralizing salts and charges, de-aromatizing rings, and fulfill the requirements of QSAR modeling. For more information please look into the commented *.smk* files.

4. The descriptor values of the whole data set need to be normalized between 0 and 1 before QSAR model development. Furthermore, if there are too many descriptors (e.g., the number of resulting Dragon descriptors is normally over 1000), it is necessary to reduce the number of descriptors to save computational time for model development. Performing a pairwise comparison between any two descriptor values is one way to find correlated and redundant descriptors. This can be done by constructing a scatter plot for every pair of descriptors and determining the Pearson's product-moment coefficient for every pair [1].

5. After the descriptor file is generated and optimized, it is needed to balance active/inactive classification ratio in the modeling set and prepare the activity file for modeling purpose. Normally the number of inactives is much larger than the number of actives in HTS data sets. For example, the ARE data set contains 958 active and 4,526 inactive compounds (Fig. 6).

6. KNIME also has an "*Equal Size Sampling*" node that automatically down-samples the data set and it can be substituted into the workflow. However, it does not partition the data set into modeling and validation sets.

7. It has been reported that there is little difference in the QSAR model performance resulting from these two methods (Martin et al. 2012). This method ensures that the test set will have structurally similar analogs in the modeling set, but this cannot be guaranteed for external set compounds. However, rational selection approach may be advantageous when the applicability domain of the QSAR model needs to be clearly defined (Golbraikh et al. 2003).
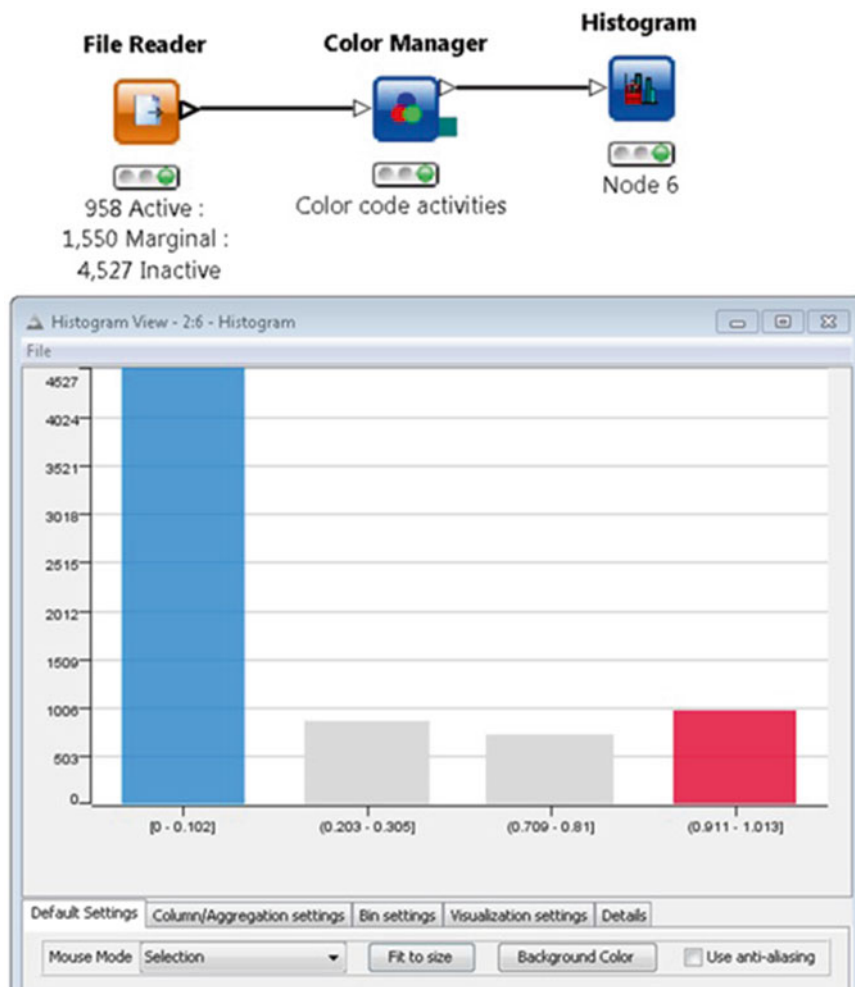
**Fig. 6** Example of KNIME histogram plot workflow and the resulting histogram plot showing the frequency of activity values 0 (inactive, *blue*), 0.25 to 0.50 (marginal, *gray*), and 1 (active, *red*)

More information on random and rational selection and applicability domain can be found at: Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, et al. 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52:2570–8; doi:10.1021/ci300338w.

Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models. J. Comput. Aided. Mol. Des. 17:241–53; doi:10.1023/A:1025386326946.

8. Principal component analysis is a statistical method that reduces the dimensions of descriptors in a data set by finding groups of

descriptor combinations. It also provides one of the most informative statistics about the data. The first principal component covers the largest amount of variance in the data set. Each consecutive principal component will cover another portion of the variance, but less than the previous one. Therefore, the combination of all principal components represents the total variance in the data set. And the total number of principal components is less than the number of descriptors. All these calculations can be done in software such as KNIME and MOE.

Typically the first three principal components can be used to analyze the diversity of the chemical space and the overall relationships in the model. For example, in the sample descriptor file there are 10 descriptors calculated for the whole data set. A principal component analysis was performed to generate six principal components. Principal components 1 and 2 are plotted in a scatter plot to show the chemical space. Figure 3 shows the KNIME node that can be used to generate the principal components and the scatter plot of principal components 1 and 2 using all active and inactive compounds ($n = 5484$). Similar compounds will be clustered together and dissimilar compounds will be dispersed. In this case, the modeling set shows that the active and inactive compounds share the same chemical space. If active and inactive compounds occupy different spaces in the scatter plot, QSAR models will not be able to be developed.

More information on principal components can be found at: Izenman AJ. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning.* 1st ed. Springer Publishing Company, Incorporated.

9. The chemical space indicates the applicability domain of resulting QSAR models.

10. The MOE descriptors used in this study were FCharge, PC+, PC-, TPSA, Weight, a_acc, a_don, density, logP(o/w), and logS.

## 5  Summary

Publicly available HTS data contains chemical structure errors and unbalanced activity distributions that need to be addressed before the data can be modeled. Due to its size, curating the data for QSAR modeling purpose requires automated computational tools. Furthermore, the activity distribution in HTS data is usually heavily skewed towards inactive compounds, which leads to biased predictions. To avoid biased predictions in the resulting QSAR models, the number of inactive and active compounds selected for modeling needs to be balanced. Down-sampling using either random or rational selection approaches mitigates this issue and results in a

sample data set suitable for QSAR modeling. The technology described in this chapter enables one to use automated approaches to curate and prepare the public HTS data for modeling purposes.

## References

1. Daniel WW (2009) Biostatistics: a foundation for analysis in the health sciences, 9th edn. Wiley, Hoboken, NJ

2. Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13:3494–3504. doi:10.2174/138161207782794257

3. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36. doi:10.1021/ci00057a005

# INDEX