# Interface

## Modes (button)

### One chemical 🡸🡺 multiple assays (parallel)

### One chemical 🡸🡺 multiple assays (overlay)

### Multiple chemicals 🡸🡺 one assay

## Pathways (Check)

### ATAD5

### P53

### DSB

### SRF

### TR antagonism

### ER antagonism (Hek293)

## Compound loader (by CAS)

## Curve display options (check)

### Runs

#### Run1

#### Run2

#### Run3

### Cytotoxicity

### Ch1

### Ch2

## Curve plotting methods ()

### Raw

### Curvep

### Hill

## Curves (tab)

## About (tab)

# Implementation

1. Preparation
2. MySQL database (future)
   1. R DBI package
3. CEBS data
4. Load into RData with new columns
   1. Curve readout (with compound plate ID. Only applicable to EPA, NTP plates, mostly should be based on my Library Column): ratio.[1-3]; ch2.[1-3]; ch1.[1-3]; 653.[1-3]; 657.[1-3]; 100.[1-3]; via.[1-3]; luc.[1-3]
   2. Pathways: atad5, p53, … (should be informative. Because it will be shown on the plot)
   3. New CAS; StructureID; New Chemical.Name; uniqueID: Tox21\_ID@Cmpd\_Library
   4. Save into RData
5. Loading
   1. After pathways are selected. Load the data (could be really slow…)
      1. 83-79-4 (3 sources, and “active” in TR)
      2. Tox21\_303646@EPA\_C (mistake compounds, 4:2:3)
   2. Slice the data based on CAS and pathways and append them
   3. Filter based on options
      1. Mode#2 only allows
   4. Select CAS, uniqueID, Chemcial.Name, readout, pathway, c[0-14], r[0-14], curvepr[0-14],
   5. Melt c[0-14] as x ; melt r[0-14] as y; NA don’t remove!; resp\_type <- raw
   6. Append another data: Melt c[0-14] as x ; melt curvepr [0-14] as y; NA don’t remove!; resp\_type <- curvep
   7. CAS + uniqueID + Chemical.Name + readout + pathway + x + y + resp\_type
6. Plotting
   1. mode#1
      1. title: the CAS
      2. ggplot(temp, aes(x=x, y=raw, color=readout)) + geom\_line() + facet\_grid(uniqueID ~ pathway) + geom\_point()
   2. mode#2
      1. title: the CAS
      2. filter ch2 and ch1 data
      3. make a new readout (path\_ readout) : pathway@ readout
      4. ggplot(aes(x=x, y=y, color= path\_ readout)) + geom\_line() + facet (uniqueID ~)
   3. mode#3
      1. title: pathway
      2. make a new uniqueID (cas\_uniqueID): CAS|uniqueID
      3. ggplot(aes(x=x, y=y, color= readout)) + geom\_line() + facet (cas\_uniqueID ~)

# Potential to do

1. order based on wAUC

# Required columns in the input (general plotting)

1. Chemical.ID (such as CAS or some others)
   1. so the user needs to be responsible for the input
2. conc[0-9]+ (uM)
3. resp[0-9]+
4. pathway
5. readout

basic\_cols <- c('CAS', 'uniqueID', 'Chemical.ID', 'Chemical.Name', 'Tox21.ID', 'Sample.ID', 'StructureID', 'readout', 'pathway')

1. It is allowed multiple inputs to combine together
2. The Chemical.ID and Chemical.Name columns are needed. These two columns (a) are used in the plots.
3. Additional IDs (e.g., CAS, Tox21.ID, NCGC.ID) can be used for searching. (Thus, you may duplicate CAS as Chemical.ID)
4. conc[0-9]+ (b) (after log transformation)
5. resp[0-9]+ [-100% - 100%] (c)
6. pathway (d) & readout (e) (e.g., via.1, thus batch information is included)
7. a, b, c, d, e construct the plot
8. parent column is used to specify the data region that can be used to couple with the parent pathway (e.g., cytotoxicity data are coupled with many pathways in high-content assay)
9. Currently only one mask is allowed. Mask.Flags is the primary column. If Mask.Flags unavailable, curvep\_mask is used.
10. Three curve plotting methods are used: raw, curvep (curvep\_r[0-9]+ columns are needed), hill 4-point(Zero.Activity, Inf.Activity, LogAC50, Hill.Coef)
11. Other columns will be kept but not used

# Optional columns in the input (general plotting)

1. curvep\_r[0-9]+
2. Mask.Flags
3. Hill related: (Zero.Activity, Inf.Activity, LogAC50 (I feel it’s better to keep them as uM), Hill.Coef

# Auto-fluorescence data

Pathway is set as autofluo\_hek293 or autofluo\_hepg2

Readout is only [cell|medi]\_[red|blue|green] no batch information

Since there are some new Tox21 IDs are related to the old ones (\_1 in the end), I populated the data in load\_cebs\_file by randomly pick one (if more than one) based on the old tox21 ID. Also, I created a new function get\_relevant\_cmpd\_library() so that I can modify the Tox21AgencyID and populate data for all the pathways. Also I fixed the mask columns

# To do:

1. Print pdf … only finish the parallel

Auto data

Pathway main is null ; medi.]blue|green|read] cell.[blue\green|red] readout

Need to edit cmpd\_library and tox21agencyid , tox21id after loading the data /create a mapping file… (done)

Make sure options will include non .1, .2, .3 (done)

But it will be only available in assay overlapping mode

Subset autofluo pathway data ; lapply(1: nrow(df) ; foreach Tox21.ID find the relevant autofluo data ; edit Cmpd library, Tox21AgencyID