

CDD BioAssay Express

Bringing bioassay protocols to the world of informatics, using semantic annotations

Alex M. Clark

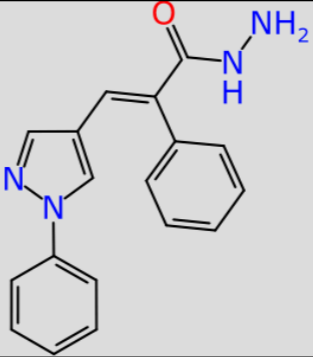
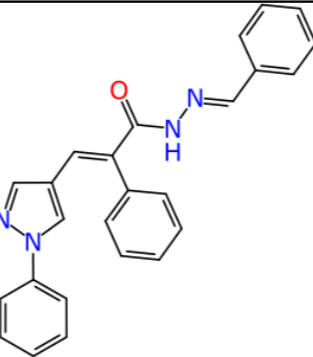
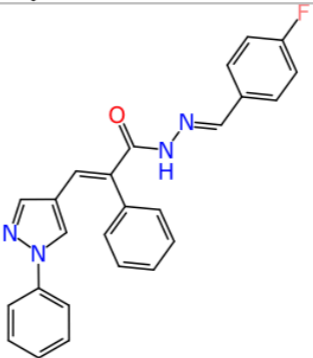
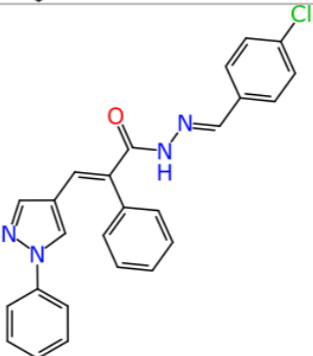
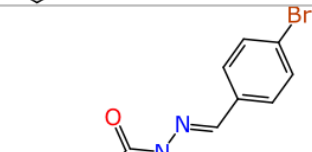


CDD VAULT[®]
Complexity Simplified

SAR

- ◆ The **activity column**:
 - ▷ TCruzi (chagas)
 - ▷ IC50
 - ▷ nM
- ◆ Underdefined in a typical SDfile or table
- ◆ Only alternative is a verbose publication

Molecular Datasheet - chagas.ds

	Molecule	TCruzi:IC50	R	Units	A
1		2.81e+06		nM	
2		190000		nM	
3		2.44e+06		nM	
4		8.59e+06		nM	
5					

Assay Informatics

- ◆ Measurement of an activity has many details:
 - ▷ **target**, **organism**, **cell line**, **measurement type**, **units**, **experimental design**, **instrumentation**, etc.
- ◆ Usually only available as **text**...
- ◆ ... needs to be **informatics**.
- ◆ Searching for experiments, comparing datasets, spotting trends, evaluating completeness



Cell-Free Homogeneous Primary HTS to Identify Inhibitors of GSK3beta Activity

- (1) Dispense 1 uL/well of CABPE, 0.5 uL of ATP, and 1 uL of positive control GW8510 or AB in respective wells according to plate design to 1536-well assay ready plates (Aurora 29847) that contain 2.5 nL/well of 10 mM compound using BioRAPTR (Beckman) to start the reaction. Incubate at room temperature for 60 minutes.
- (2) Add 2.5 uL/well of ADP-glo (Promega, V9103) with BioRAPTR, incubate at room temperature for 40 minutes
- (3) Add 5 uL/well of ADP-glo (Promega, V9103) with Combi nL (Thermo), incubate at room temperature for 30 minutes

Assay Informatics

- ◆ Measurement of an activity has many details:
 - ▷ **target**, **organism**, **cell line**, **measurement type**, **units**, **experimental design**, **instrumentation**, etc.

- ◆ Usually only available as **text**...

http://www.bioassayontology.org/bao#BAO_0002909
http://www.drugtargetontology.org/dto/DTO_03300101
http://purl.obolibrary.org/obo/NCBITaxon_9606

- ◆ ... needs to be **informatics**.

- ◆ Searching for experiments, comparing datasets, spotting trends, evaluating completeness

Cell-Free Homogeneous Primary HTS to Identify Inhibitors of GSK3beta Activity

(1) Dispense 1 uL/well of CABPE, 0.5 uL of ATP, and 1 uL of positive control GW8510 or AB in respective wells according to

the assay protocol (536-well assay ready plates (Aurora 29847) that contain 2.5 nL/well of 10 mM DMSO using a 100 nL pipette (Beckman) to start the reaction in the wells at room temperature for

(2) Add 2.5 uL/well of ADP-glo (Promega, V9103) with BioRAPTR, incubate at room temperature for 40 minutes

(3) Add 5 uL/well of ADP-glo (Promega, V9103) with Combi nL (Thermo), incubate at room temperature for 30 minutes

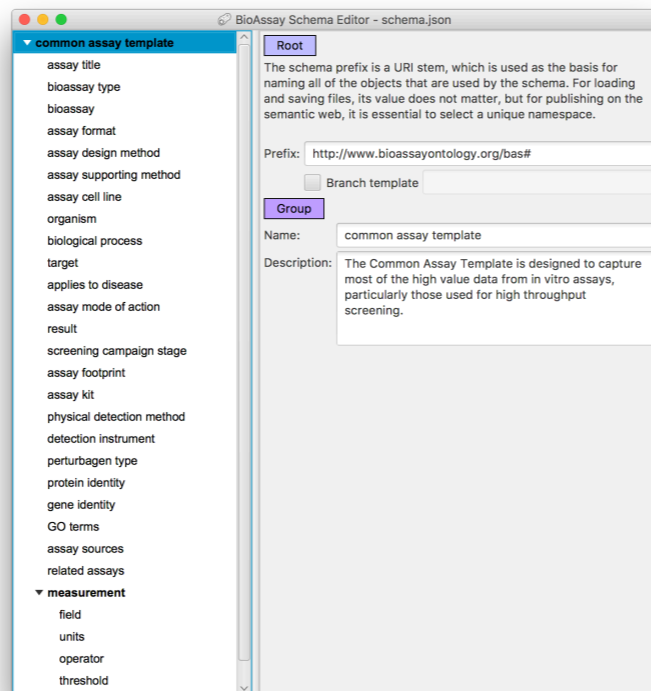
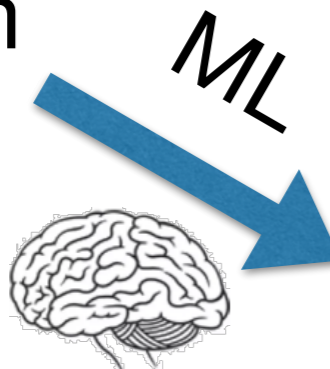
The Mission

- ◆ To make the world's bioassay protocol data **machine readable**
- ◆ Much progress toward making this goal an actual reality...
- ◆ With the help of ontology markup:
 - ▷ BAO, DTO, CLO, GO, UO, UniProt...

Ontologies BAO, DTO, CLO, UO, UniProt

Legacy Data

PubMed
PubChem
MLPCN
Tox21
ChEMBL



BioAssay Schema Editor - schema.json

Root

The schema prefix is a URI stem, which is used as the basis for naming all of the objects that are used by the schema. For loading and saving files, its value does not matter, but for publishing on the semantic web, it is essential to select a unique namespace.

Prefix:

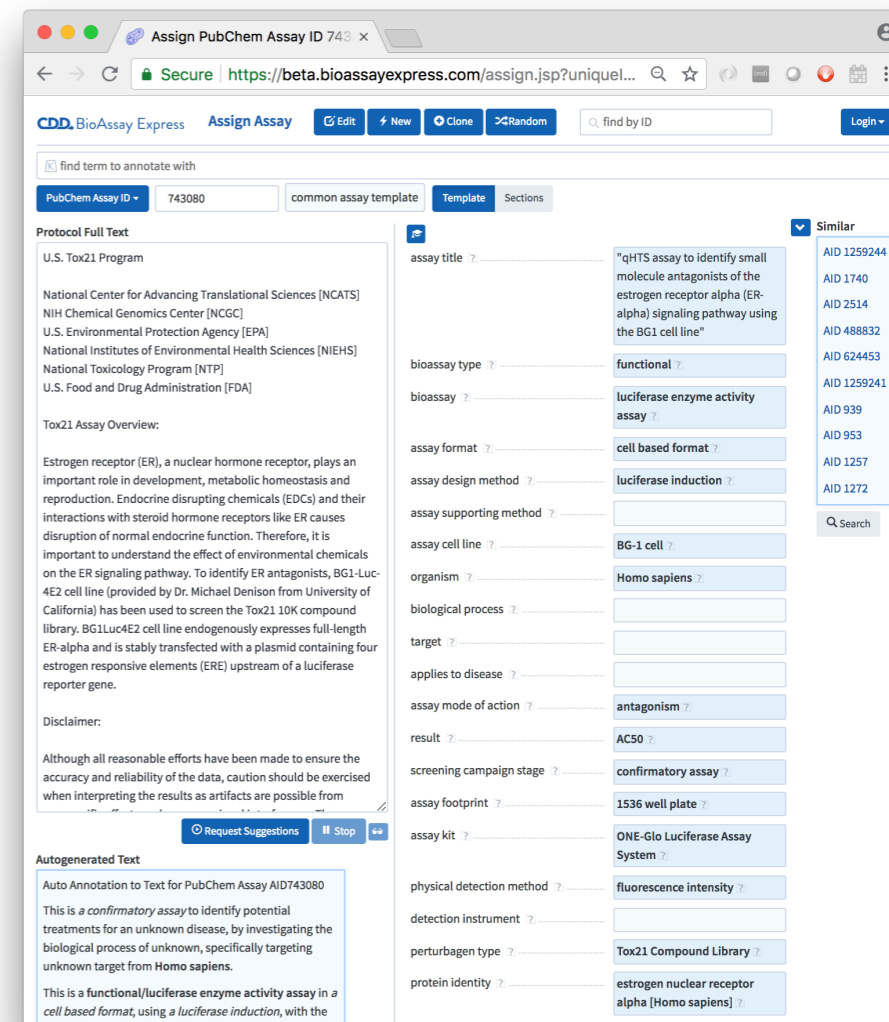
Branch template

Group

Name:

Description:

- assay title
- bioassay type
- bioassay
- assay format
- assay design method
- assay supporting method
- assay cell line
- organism
- biological process
- target
- applies to disease
- assay mode of action
- result
- screening campaign stage
- assay footprint
- assay kit
- physical detection method
- detection instrument
- perturbagen type
- protein identity
- gene identity
- GO terms
- assay sources
- related assays
- measurement
 - field
 - units
 - operator
 - threshold



Assign PubChem Assay ID 743080

Secure | <https://beta.bioassayexpress.com/assign.jsp?uniqueid...>

CDD BioAssay Express Assign Assay Edit New Clone Random find by ID Login

PubChem Assay ID: 743080 common assay template Template Sections

find term to annotate with

Protocol Full Text

U.S. Tox21 Program

National Center for Advancing Translational Sciences [NCATS]
NIH Chemical Genomics Center [NCGC]
U.S. Environmental Protection Agency [EPA]
National Institutes of Environmental Health Sciences [NIEHS]
National Toxicology Program [NTP]
U.S. Food and Drug Administration [FDA]

Tox21 Assay Overview:

Estrogen receptor (ER), a nuclear hormone receptor, plays an important role in development, metabolic homeostasis and reproduction. Endocrine disrupting chemicals (EDCs) and their interactions with steroid hormone receptors like ER causes disruption of normal endocrine function. Therefore, it is important to understand the effect of environmental chemicals on the ER signaling pathway. To identify ER antagonists, BG1-Luc-4E2 cell line (provided by Dr. Michael Denison from University of California) has been used to screen the Tox21 10K compound library. BG1Luc4E2 cell line endogenously expresses full-length ER-alpha and is stably transfected with a plasmid containing four estrogen responsive elements (ERE) upstream of a luciferase reporter gene.

Disclaimer:

Although all reasonable efforts have been made to ensure the accuracy and reliability of the data, caution should be exercised when interpreting the results as artifacts are possible from

Autogenerated Text

Auto Annotation to Text for PubChem Assay AID743080

This is a *confirmatory* assay to identify potential treatments for an unknown disease, by investigating the biological process of unknown, specifically targeting unknown target from Homo sapiens.

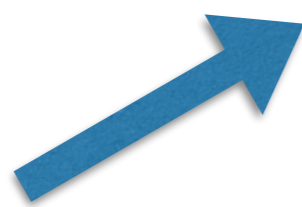
This is a *functional/luciferase enzyme activity* assay in a *cell based format*, using a *luciferase induction*, with the

Similar

- assay title ? "qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line"
- bioassay type ? functional ?
- bioassay ? luciferase enzyme activity assay ?
- assay format ? cell based format ?
- assay design method ? luciferase induction ?
- assay supporting method ?
- assay cell line ? BG-1 cell ?
- organism ? Homo sapiens ?
- biological process ?
- target ?
- applies to disease ?
- assay mode of action ? antagonism ?
- result ? AC50 ?
- screening campaign stage ? confirmatory assay ?
- assay footprint ? 1536 well plate ?
- assay kit ? ONE-Glo Luciferase Assay System ?
- physical detection method ? fluorescence intensity ?
- detection instrument ?
- perturbagen type ? Tox21 Compound Library ?
- protein identity ? estrogen nuclear receptor alpha [Homo sapiens] ?

New Data

ELN



Common Assay Template

Fully Annotated Assays

Common Assay Template

- ◆ A recipe for how to use ontologies
- ◆ Capture most important information for most assays

assay title ?	>	"GSK3beta Surface Plasmon Resonance (SPR) Assay Measured in Biochemical System Using Microfluidics - 2133-08_Inhibitor_Dose_DryPowder_Activity" x	assay footprint ?	>	Sensor chip CM5 ? x 384 well plate ? x
bioassay type ?	>	binding type ? x	assay kit ?	>	
bioassay ?	>	binding assay ? x	physical detection method ?	>	
assay format ?	>	biochemical format ? x	detection instrument ?	>	Biacore 3000 SPR System (GE Healthcare) ? x
assay design method ?	>	binding assessment method ? x	perturbagen type ?	>	compound library ? x
assay supporting method ?	>	single perturbagen ? x	protein identity ?	>	glycogen synthase kinase-3 beta isoform 1 [Homo sapiens] ? x
assay cell line ?	>	SF-21 cell ? x	gene identity ?	>	GSK3B: glycogen synthase kinase 3 beta ? x
organism ?	>	Homo sapiens ? x	GO terms ?	>	
biological process ?	>		assay sources ?	>	Broad Institute (Harvard-MIT) ? x
target ?	>	Kinase ? x GSK3B ? x	related assays ?	>	AID 2119 x
applies to disease ?	>		measurement ? * ? ?		
assay mode of action ?	>	modulation ? x inhibition ? x	field ?	>	KD_(nM) x
result ?	>		units ?	>	nanomolar ? x
screening campaign stage ?	>		operator ?	>	
			threshold ?	>	

Legacy Curation

- ◆ Text-centric data: millions of assays...
 - ▷ private collections (**ELNs**)
 - ▷ published literature (**PubMed**)
 - ▷ assay databases (**PubChem**)
- ◆ Most accessible sources: **MLPCN, Tox21**
- ◆ Data comes in as plain text + several fields + SAR

Hybrid machine learning

1. Text extraction
 2. Probabilistic:
 - ▷ natural language models
 - ▷ correlation models
 3. Axiomatic rules
- ◆ Each of these *accelerates* the human curator

Text extraction

◆ Optimised for low false positives

qHTS Assay for Small Molecule Inhibitors of the **Human** hERG Channel Activity

NIH Chemical Genomics Center [NCGC]
National Institutes of Environmental Health Sciences [NIEHS]
National Toxicology Program [NTP]

NCGC Assay Overview:

We have developed a **1536-well** cell-based assay for quantitative high throughput screening (qHTS) to determine in vitro hERG channel blockage as a measure of cardio **toxicity** with small molecules. This particular assay uses the **U2OS** cell line which is derived from human **osteosarcoma**.

...

bioassay type ?	functional ✓
bioassay ?	cell viability assay ✓
assay format ?	cell based format ✓
assay design method ?	luciferase enzyme activity measurement method ✓
assay supporting method ?	silencing RNA delivery method ✓
assay cell line ?	U2-OS cell ✓ osteosarcoma cell ✓
organism ?	Homo sapiens ✓
biological process ?	cell proliferation ✓
target ?	Protein ✓
applies to disease ?	viral infectious disease ✓
assay mode of action ?	cytotoxicity ✓
result ?	raw activity ✓
screening campaign stage ?	compound toxicity assay ✓
assay footprint ?	1536 well plate ✓
assay kit ?	assay kit ✓
physical detection method ?	fluorescence method ✓
detection instrument ?	ViewLux CCD Imager ✓

Text extraction

◆ Optimised for low false positives

qHTS Assay for Small Molecule Inhibitors of the **Human** hERG Channel Activity

NIH Chemical Genomics Center [NCGC]
National Institutes of Environmental Health Sciences [NIEHS]
National Toxicology Program [NTP]

NCGC Assay Overview:

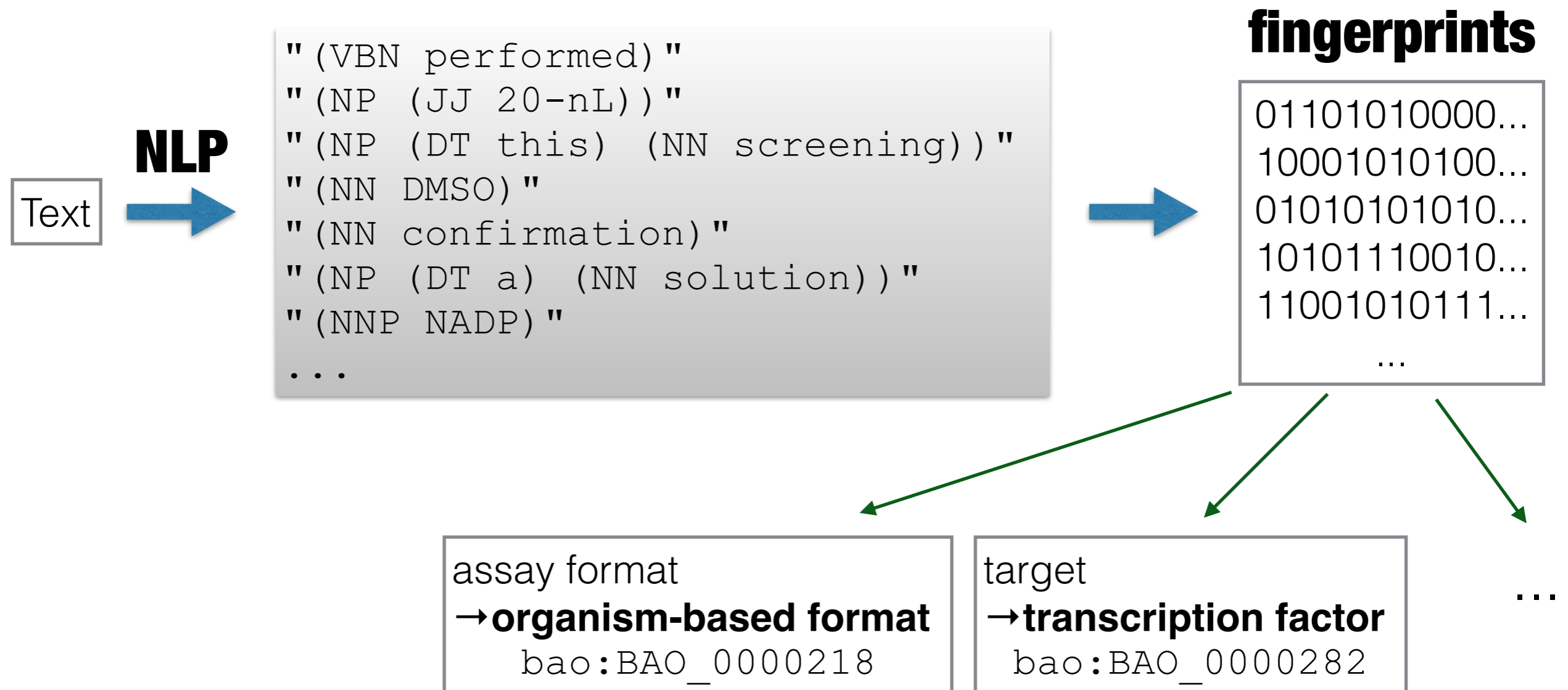
We have developed a **1536-well** cell-based assay for quantitative high throughput screening (qHTS) to determine in vitro hERG channel blockage as a measure of cardio **toxicity** with small molecules. This particular assay uses the **U2OS** cell line which is derived from human **osteosarcoma**.

...

bioassay type ?	functional ✓
bioassay ?	cell viability assay ✓
assay format ?	cell based format ✓
assay design method ?	luciferase enzyme activity measurement method ✓
assay supporting method ?	silencing RNA delivery method ✓
assay cell line ?	U2-OS cell ✓ osteosarcoma cell ✓
organism ?	Homo sapiens ✓
biological process ?	cell proliferation ✓
target ?	Protein ✓
applies to disease ?	viral infectious disease ✓
assay mode of action ?	cytotoxicity ✓
result ?	raw activity ✓
screening campaign stage ?	compound toxicity assay ✓
assay footprint ?	1536 well plate ✓
assay kit ?	assay kit ✓
physical detection method ?	fluorescence method ✓
detection instrument ?	ViewLux CCD Imager ✓

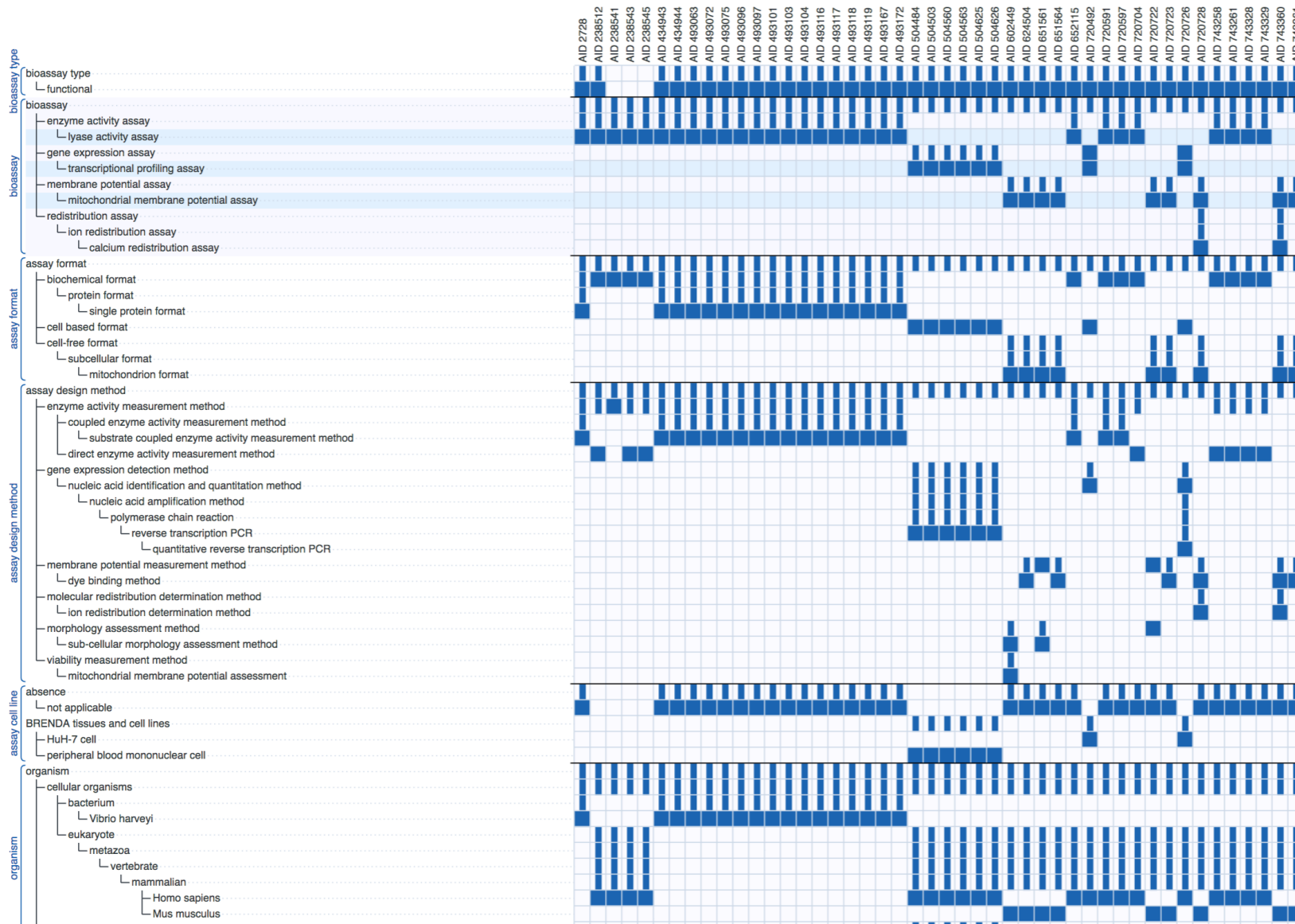
Natural language models

- ◆ Use **NLP** to create **fingerprints** for each text document
- ◆ One Bayesian model for each possible annotation term



Correlation models

◆ Similar principle: existing annotations are fingerprints



one model per term

◆ Combine *natural language* + *correlation* = **ranking**

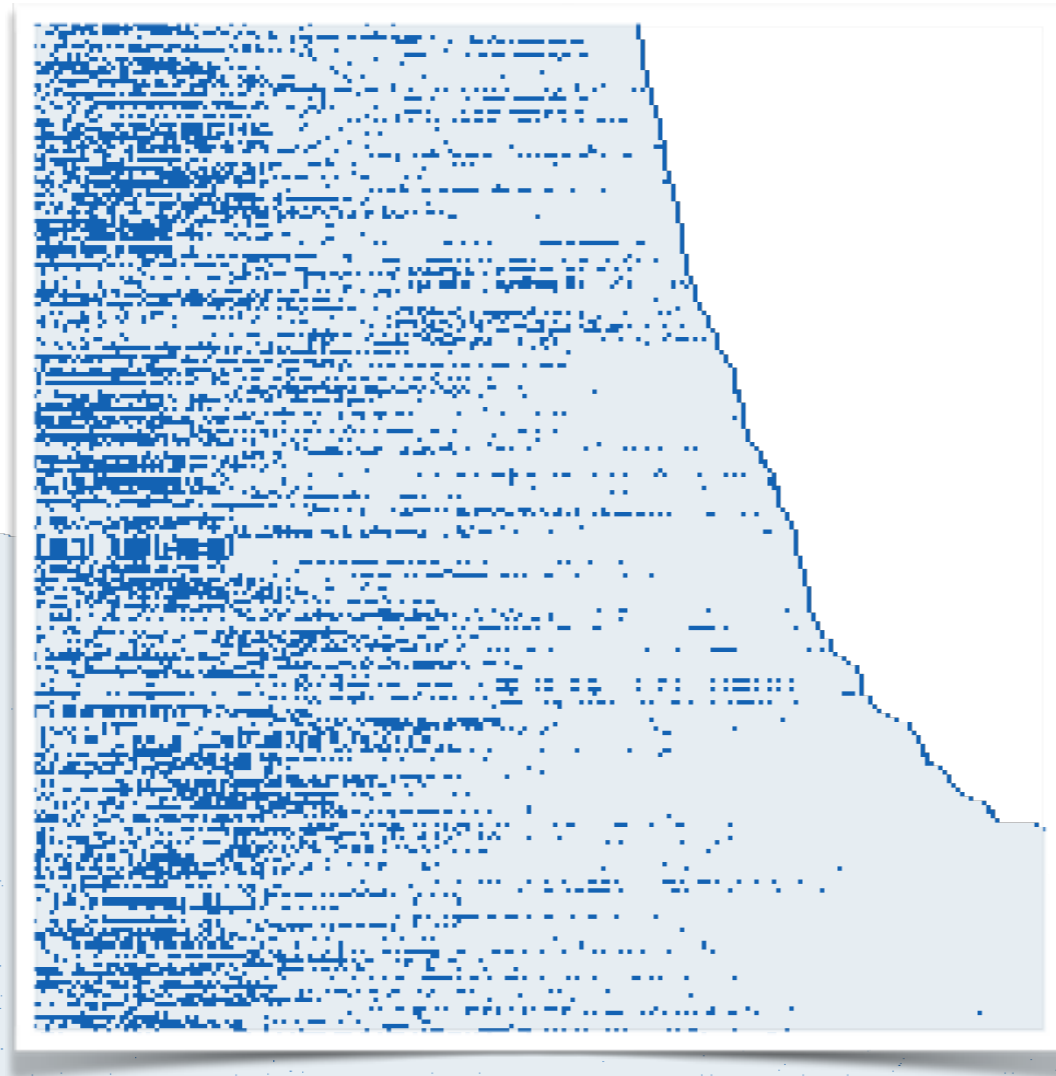
Metrics

Metrics

◆ NLP only: 96.9% \pm 4.4

Metrics

◆ NLP only: 96.9% \pm 4.4



Metrics

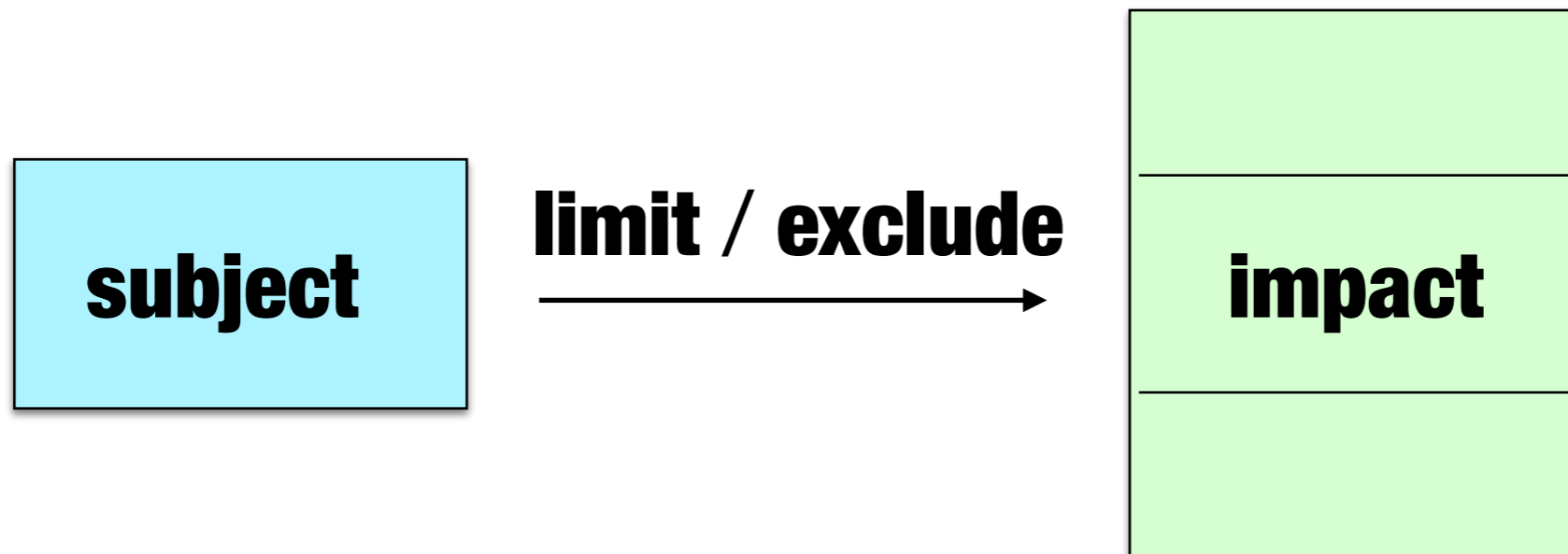
- ◆ NLP only: 96.9% \pm 4.4
- ◆ + correlation: 97.5% \pm 3.8

Metrics

- ◆ NLP only: 96.9% \pm 4.4
- ◆ + correlation: 97.5% \pm 3.8
- ◆ + text mining: 97.6% \pm 3.8

Axioms

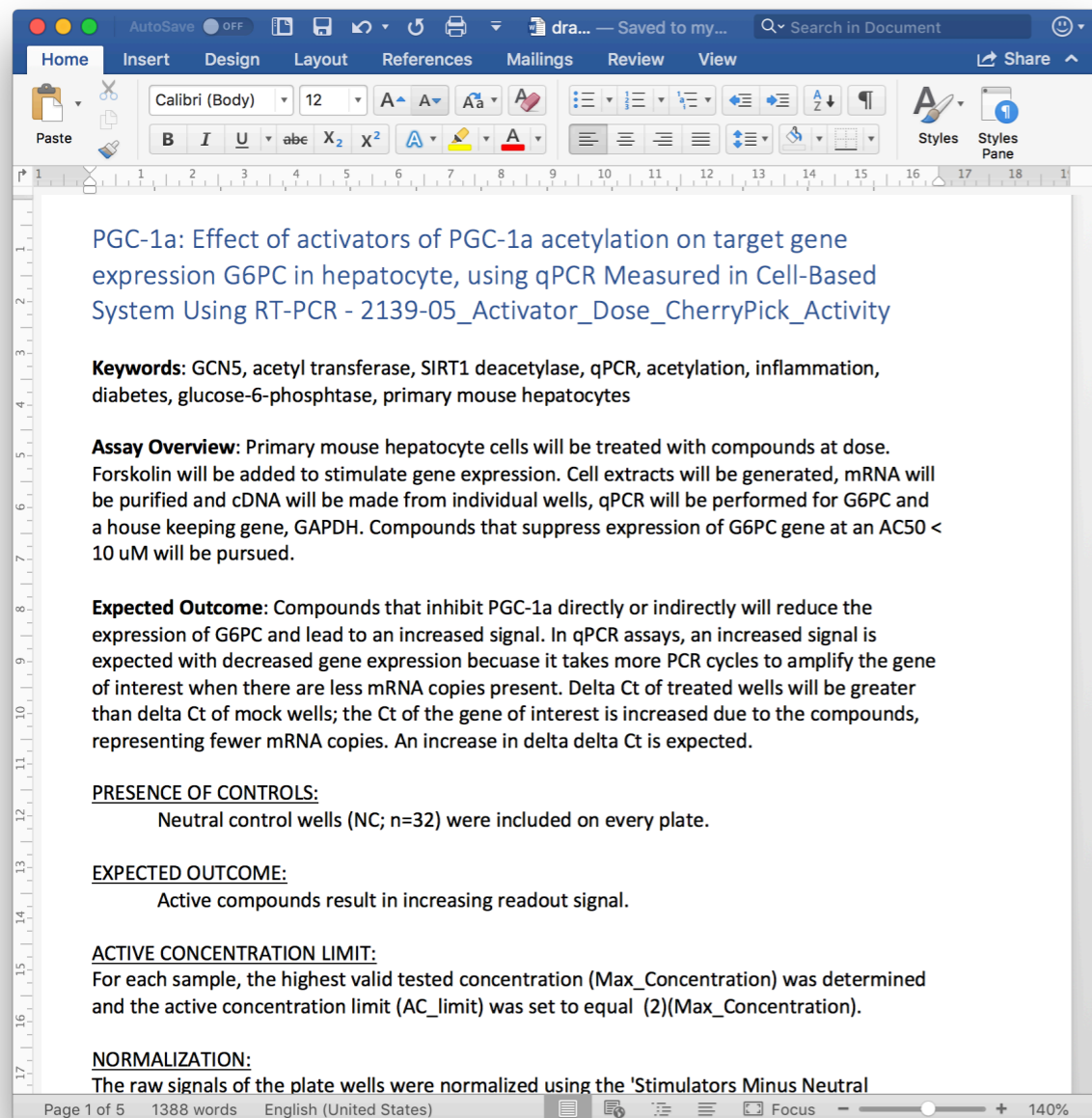
- ◆ Ontologies like BAO & DTO have *axioms*
- ◆ Convert these into **rules**:



- ◆ Use to constrain predictions at each step
- ◆ Like correlation models, except absolute

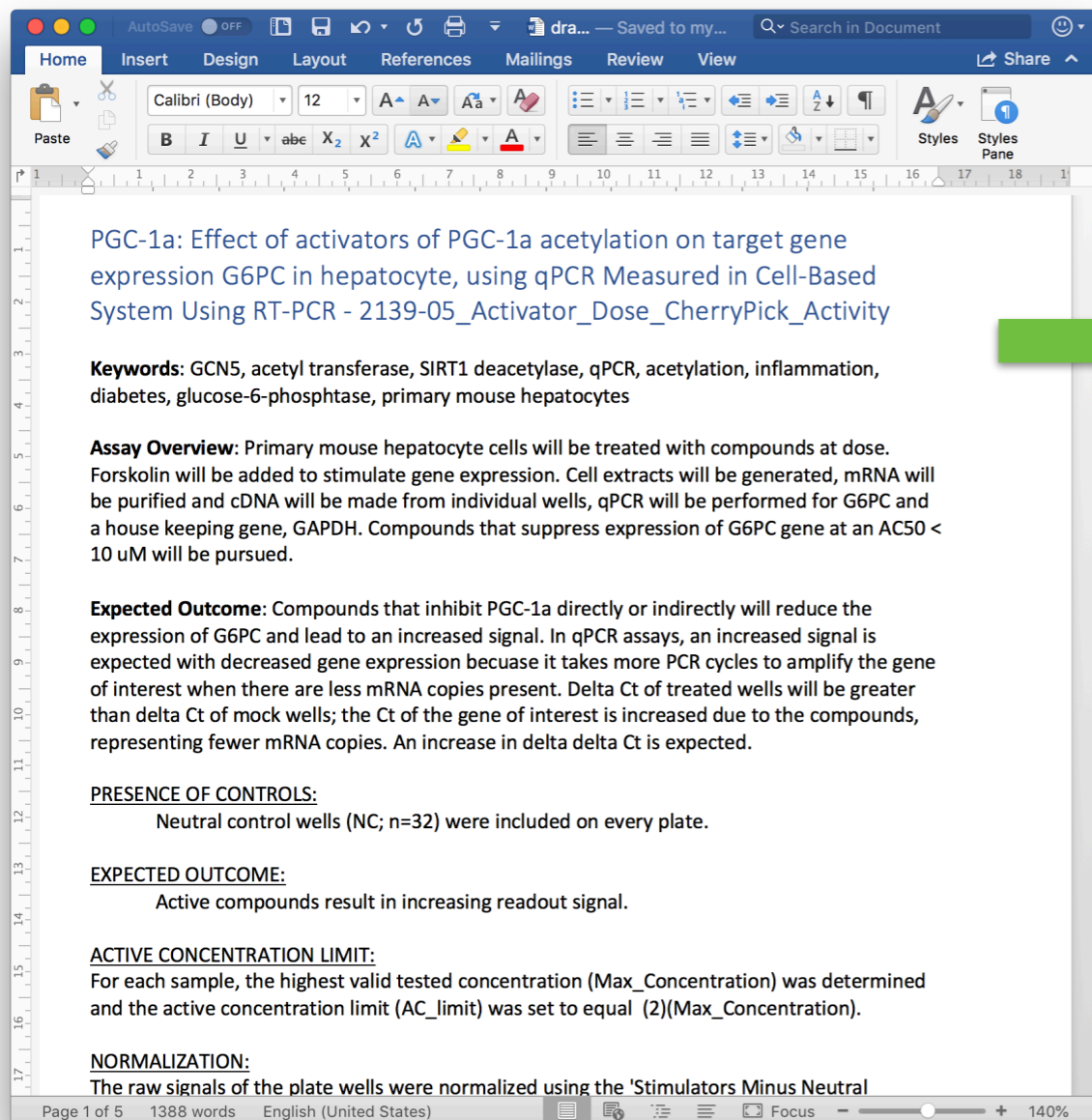
New data: ELN

◆ Paste text or drag a document...



New data: ELN

◆ Paste text or drag a document...



PGC-1a: Effect of activators of PGC-1a acetylation on target gene expression G6PC in hepatocyte, using qPCR Measured in Cell-Based System Using RT-PCR - 2139-05_Activator_Dose_CherryPick_Activity

Keywords: GCN5, acetyl transferase, SIRT1 deacetylase, qPCR, acetylation, inflammation, diabetes, glucose-6-phosphatase, primary mouse hepatocytes

Assay Overview: Primary mouse hepatocyte cells will be treated with compounds at dose. Forskolin will be added to stimulate gene expression. Cell extracts will be generated, mRNA will be purified and cDNA will be made from individual wells, qPCR will be performed for G6PC and a house keeping gene, GAPDH. Compounds that suppress expression of G6PC gene at an AC50 < 10 uM will be pursued.

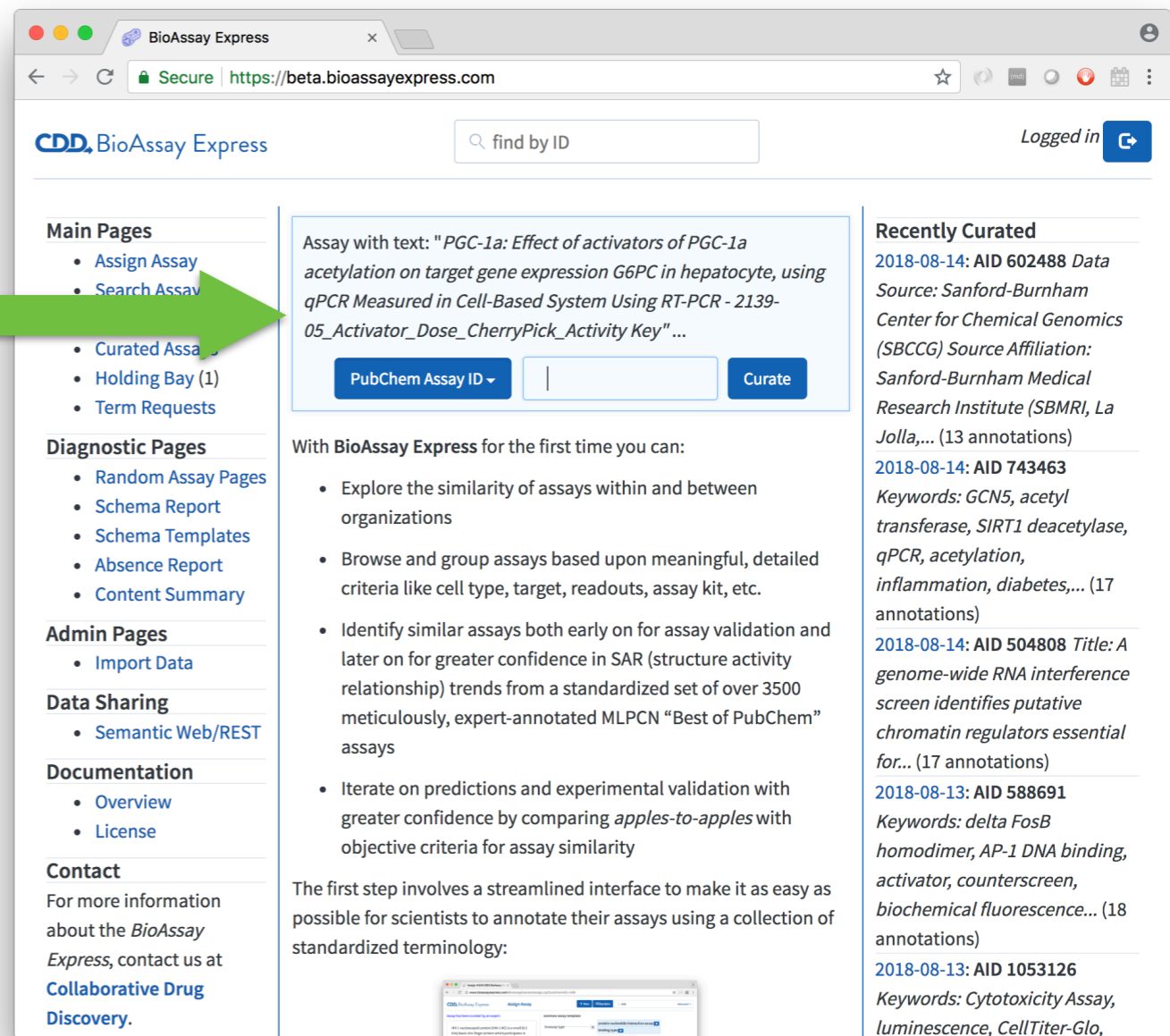
Expected Outcome: Compounds that inhibit PGC-1a directly or indirectly will reduce the expression of G6PC and lead to an increased signal. In qPCR assays, an increased signal is expected with decreased gene expression because it takes more PCR cycles to amplify the gene of interest when there are less mRNA copies present. Delta Ct of treated wells will be greater than delta Ct of mock wells; the Ct of the gene of interest is increased due to the compounds, representing fewer mRNA copies. An increase in delta delta Ct is expected.

PRESENCE OF CONTROLS:
Neutral control wells (NC; n=32) were included on every plate.

EXPECTED OUTCOME:
Active compounds result in increasing readout signal.

ACTIVE CONCENTRATION LIMIT:
For each sample, the highest valid tested concentration (Max_Concentration) was determined and the active concentration limit (AC_limit) was set to equal (2)(Max_Concentration).

NORMALIZATION:
The raw signals of the plate wells were normalized using the 'Stimulators Minus Neutral



CDD, BioAssay Express

find by ID

Logged in

Main Pages

- Assign Assay
- Search Assay
- Curated Assays
- Holding Bay (1)
- Term Requests

Diagnostic Pages

- Random Assay Pages
- Schema Report
- Schema Templates
- Absence Report
- Content Summary

Admin Pages

- Import Data

Data Sharing

- Semantic Web/REST

Documentation

- Overview
- License

Contact

For more information about the *BioAssay Express*, contact us at [Collaborative Drug Discovery](#).

Assay with text: "PGC-1a: Effect of activators of PGC-1a acetylation on target gene expression G6PC in hepatocyte, using qPCR Measured in Cell-Based System Using RT-PCR - 2139-05_Activator_Dose_CherryPick_Activity Key" ...

PubChem Assay ID Curate

With BioAssay Express for the first time you can:

- Explore the similarity of assays within and between organizations
- Browse and group assays based upon meaningful, detailed criteria like cell type, target, readouts, assay kit, etc.
- Identify similar assays both early on for assay validation and later on for greater confidence in SAR (structure activity relationship) trends from a standardized set of over 3500 meticulously, expert-annotated MLPCN "Best of PubChem" assays
- Iterate on predictions and experimental validation with greater confidence by comparing *apples-to-apples* with objective criteria for assay similarity

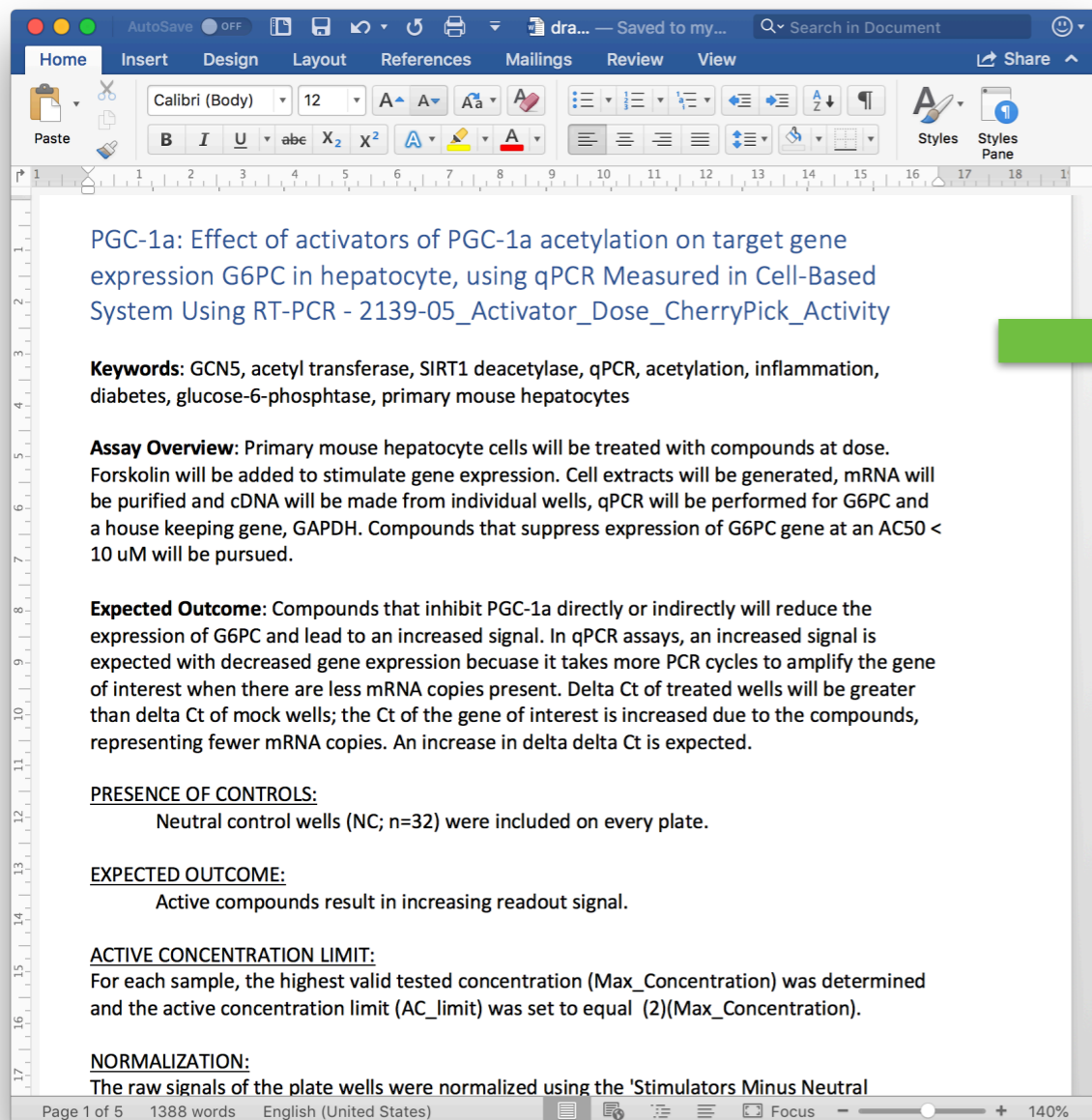
The first step involves a streamlined interface to make it as easy as possible for scientists to annotate their assays using a collection of standardized terminology:

Recently Curated

- 2018-08-14: AID 602488 Data Source: Sanford-Burnham Center for Chemical Genomics (SBCCG) Source Affiliation: Sanford-Burnham Medical Research Institute (SBMRI, La Jolla,... (13 annotations)
- 2018-08-14: AID 743463 Keywords: GCN5, acetyl transferase, SIRT1 deacetylase, qPCR, acetylation, inflammation, diabetes,... (17 annotations)
- 2018-08-14: AID 504808 Title: A genome-wide RNA interference screen identifies putative chromatin regulators essential for... (17 annotations)
- 2018-08-13: AID 588691 Keywords: delta FosB homodimer, AP-1 DNA binding, activator, counterscreen, biochemical fluorescence... (18 annotations)
- 2018-08-13: AID 1053126 Keywords: Cytotoxicity Assay, luminescence, CellTiter-Glo,

New data: ELN

◆ Paste text or drag a document...



PGC-1a: Effect of activators of PGC-1a acetylation on target gene expression G6PC in hepatocyte, using qPCR Measured in Cell-Based System Using RT-PCR - 2139-05_Activator_Dose_CherryPick_Activity

Keywords: GCN5, acetyl transferase, SIRT1 deacetylase, qPCR, acetylation, inflammation, diabetes, glucose-6-phosphatase, primary mouse hepatocytes

Assay Overview: Primary mouse hepatocyte cells will be treated with compounds at dose. Forskolin will be added to stimulate gene expression. Cell extracts will be generated, mRNA will be purified and cDNA will be made from individual wells, qPCR will be performed for G6PC and a house keeping gene, GAPDH. Compounds that suppress expression of G6PC gene at an AC50 < 10 uM will be pursued.

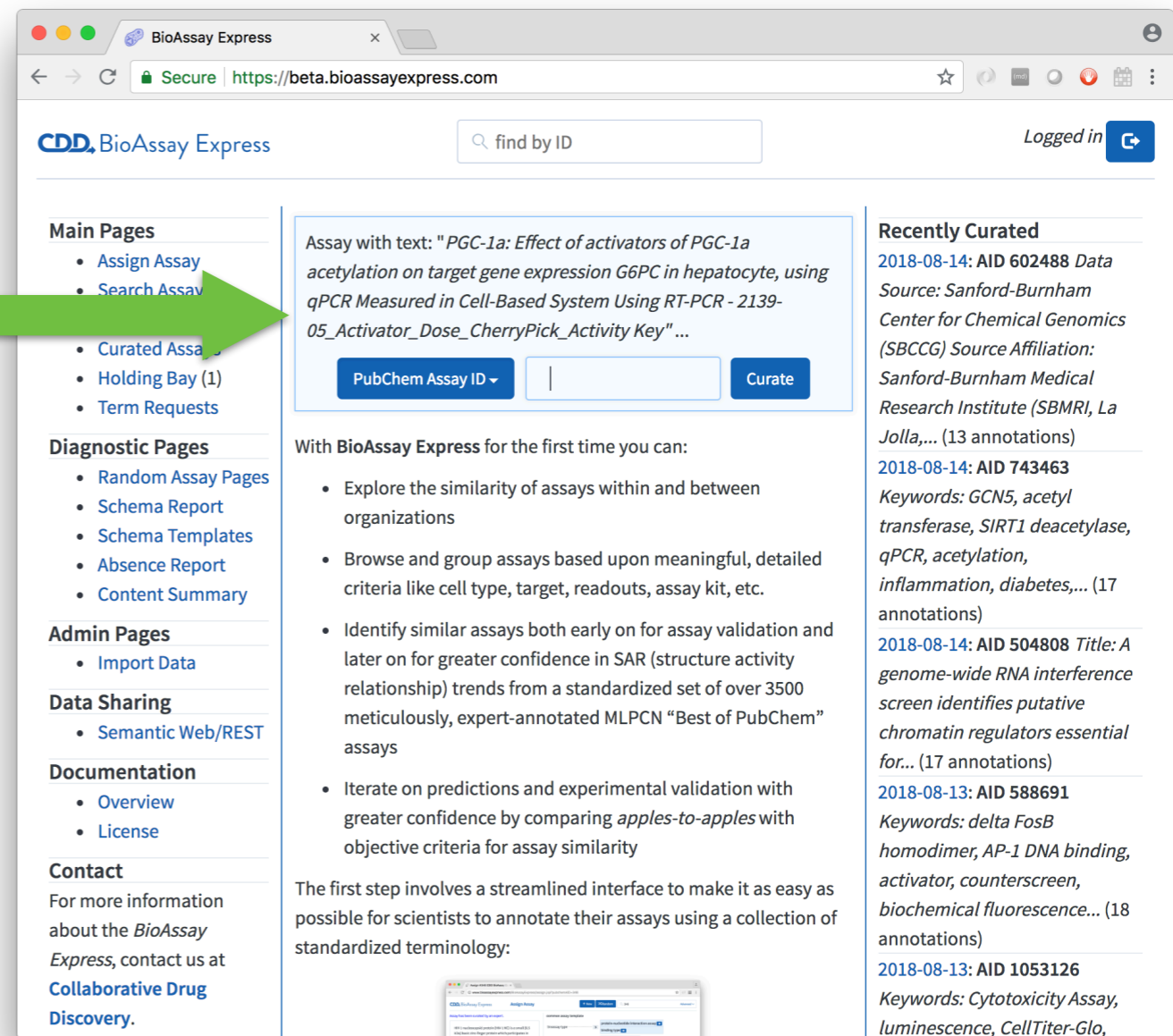
Expected Outcome: Compounds that inhibit PGC-1a directly or indirectly will reduce the expression of G6PC and lead to an increased signal. In qPCR assays, an increased signal is expected with decreased gene expression because it takes more PCR cycles to amplify the gene of interest when there are less mRNA copies present. Delta Ct of treated wells will be greater than delta Ct of mock wells; the Ct of the gene of interest is increased due to the compounds, representing fewer mRNA copies. An increase in delta delta Ct is expected.

PRESENCE OF CONTROLS:
Neutral control wells (NC; n=32) were included on every plate.

EXPECTED OUTCOME:
Active compounds result in increasing readout signal.

ACTIVE CONCENTRATION LIMIT:
For each sample, the highest valid tested concentration (Max_Concentration) was determined and the active concentration limit (AC_limit) was set to equal (2)(Max_Concentration).

NORMALIZATION:
The raw signals of the plate wells were normalized using the 'Stimulators Minus Neutral



CDD, BioAssay Express

find by ID

Logged in

Main Pages

- Assign Assay
- Search Assay
- Curated Assays
- Holding Bay (1)
- Term Requests

Diagnostic Pages

- Random Assay Pages
- Schema Report
- Schema Templates
- Absence Report
- Content Summary

Admin Pages

- Import Data

Data Sharing

- Semantic Web/REST

Documentation

- Overview
- License

Contact

For more information about the *BioAssay Express*, contact us at [Collaborative Drug Discovery](#).

Assay with text: "PGC-1a: Effect of activators of PGC-1a acetylation on target gene expression G6PC in hepatocyte, using qPCR Measured in Cell-Based System Using RT-PCR - 2139-05_Activator_Dose_CherryPick_Activity Key" ...

PubChem Assay ID

With BioAssay Express for the first time you can:

- Explore the similarity of assays within and between organizations
- Browse and group assays based upon meaningful, detailed criteria like cell type, target, readouts, assay kit, etc.
- Identify similar assays both early on for assay validation and later on for greater confidence in SAR (structure activity relationship) trends from a standardized set of over 3500 meticulously, expert-annotated MLPCN "Best of PubChem" assays
- Iterate on predictions and experimental validation with greater confidence by comparing *apples-to-apples* with objective criteria for assay similarity

The first step involves a streamlined interface to make it as easy as possible for scientists to annotate their assays using a collection of standardized terminology:

Recently Curated

- 2018-08-14: AID 602488 *Data Source: Sanford-Burnham Center for Chemical Genomics (SBCCG) Source Affiliation: Sanford-Burnham Medical Research Institute (SBMRI, La Jolla,...* (13 annotations)
- 2018-08-14: AID 743463 *Keywords: GCN5, acetyl transferase, SIRT1 deacetylase, qPCR, acetylation, inflammation, diabetes,...* (17 annotations)
- 2018-08-14: AID 504808 *Title: A genome-wide RNA interference screen identifies putative chromatin regulators essential for...* (17 annotations)
- 2018-08-13: AID 588691 *Keywords: delta FosB homodimer, AP-1 DNA binding, activator, counterscreen, biochemical fluorescence...* (18 annotations)
- 2018-08-13: AID 1053126 *Keywords: Cytotoxicity Assay, luminescence, CellTiter-Glo,*

◆ Or re-use...

Assign Assay

Edit

New

Clone

Random

Autogenerated text

Auto Annotation to Text for PubChem Assay AID743463

This is *a confirmatory assay* to identify potential treatments for an unknown disease, by investigating the biological process of **regulation of protein acetylation**, specifically targeting unknown target from **Mus musculus**.

This is a **functional/gene expression assay** in *a cell based format*, using *a gene expression detection method*, with the assay kit *Cell to CT RT Kit (Ambion)*. The cell line *hepatocyte* was used. It was conducted in *384 well plates*, with the detection method of *fluorescence method*, using *a LightCycler 480 Real-Time PCR detection instrument*. Results are reported as *Ct*, in units of unknown. This assay tested the mode of action of *inhibition* by unknown perturbagens.

- ◆ Annotations-to-text is much easier
- ◆ Transliteration template:

```
{
  "schemaURI": ["http://www.bioassayontology.org/bas#"],
  "boilerplate":
  [
    "<p>",
    "Auto Annotation to Text for PubChem Assay ",
    {"field": "uniqueID"},
    "</p>",

    "<p>",
    "This is ",
    {
      "ifany": "bao:BA0_0000210",
      "then": {"term": "bao:BA0_0000210", "article": "indefinite", "style": "italic"},
      "else": "an <i>unknown stage assay</i>"
    },
    " to identify potential treatments for ",
    {"term": "bao:BA0_0002848", "style": "bold", "empty": "an unknown disease"},
    ", by investigating the biological process of ",
    {"term": "bao:BA0_0002009", "style": "bold", "empty": "unknown"},
    ", specifically targeting ",
    {"term": "bao:BA0_0000211", "style": "bold", "empty": "unknown target"},
    " from ",
    {"term": "bao:BA0_0002921", "style": "bold", "empty": "unknown organism"},
    ".</p>",

    "<p>",

    "This is a ",
    {"terms": [{"bao:BA0_0002854"}, {"bao:BA0_0002855"}], "sep": "/", "style": "bold", "empty": "?"},
    " in ",
    {"term": "bao:BA0_0000205", "style": "italic", "article": "indefinite"},
    ", using ",
    {"term": "bao:BA0_0095009", "style": "italic", "article": "indefinite"},

    {
      "ifany": "bao:BA0_0002663",
      "then":
      {
        "ifbranch": ["bat:Absence", "bao:BA0_0002663"],
        "else":
        [
          " , with the assay kit ",
```

Extending ontologies

- ◆ Science is never complete: diction evolves
- ◆ Curators can request a *provisional term*

The screenshot shows a web browser window with the URL <https://www.bioassayexpress.com>. The page title is "New Assay - CDD BioAssay Express". A modal window titled "bioassay type Tree" is open, displaying a hierarchical list of bioassay types. The modal includes a search bar, a "Hide unmatched" checkbox, and a checked "Request Terms" checkbox. The list is organized into two main categories: "absence (7)" and "bioassay type (5)".

Category	Sub-category	Item	Request
absence (7)	-	ambiguous	* +
		dubious	* +
		missing	* +
		needs checking	* +
		not applicable	* +
		requires term	* +
		unknown	* +
bioassay type (5)	-	ADMET	* +
		binding type	* +
		functional	* +
		phenotypic	* +
		physicochemical	* +

Extending ontologies

- ◆ Science is never complete: diction evolves
- ◆ Curators can request a *provisional term*

The screenshot displays the CDD BioAssay Express web application. In the background, a browser window shows the URL <https://www.bioassayexpress.com>. The main content area features a 'bioassay type Tree' with a search bar and a list of categories. The 'absence (7)' category is expanded, showing sub-terms like 'ambiguous', 'dubious', 'missing', 'needs checking', 'not applicable', 'requires term', and 'unknown'. The 'bioassay type (5)' category is also expanded, showing 'ADMET', 'binding type', 'functional', 'phenotypic', and 'physicochemical'. Each term has a search icon and a plus sign. In the foreground, a modal window titled 'bioassay type' is open. It contains a form with the following fields: 'Parent URI: bao:BAO_0000008', 'Label: phenotypic', 'Description: Measurement of phenotypic expression.', and 'Explanation:'. At the bottom right of the modal are 'Cancel' and 'Request' buttons.

Upgrading ontologies

- ◆ Next step: submit *provisional terms* to **OntoloBridge**
- ◆ Joint collaboration between CDD and:
 - ▷ Stanford (BioPortal, CEDAR)
 - ▷ University of Miami (BAO, DTO)
- ◆ Ontology owner receives new additions via API
- ◆ Experts vet all new terms, outcome made available
- ◆ Users of BAE are unimpeded...

Detailed templates

- ◆ The **Common Assay Template** is a summary
- ◆ Can insert extra branch groups
- ◆ Ultimate goal:
 - ▷ capture whole experiment
 - ▷ replace text entirely
- ◆ Syncing with ontology maintainers...

★ measurement

- measurement details

★ assay components

- buffer concentration
- carrier protein concentration
- chelator concentration
- detergent concentration
- metal salt concentration
- reducing agent concentration
- solvent concentration

★ control details

★ organism details

- cell line details
- microbe/virus details

★ protocol details

- antibody details
- assay molecule
- biochemical assay
- coupled substrate incubation
- enzyme reaction
- ligand incubation
- perturbagen incubation
- substrate incubation

★ screened entity details

Content expansion

- ◆ Public data originally from **PubChem** subset:
 - ▷ **MLPCN**: thousands of screening assays
 - ▷ **Tox21**: hundreds of assays from EPA
- ◆ Detailed text, and compounds + measurements (SAR)
- ◆ Done most of them: where's the next goldmine?

PubMed

- ◆ PubMed is massive (1.7M open access), but:
 - ▷ only some have relevant assays
 - ▷ no compounds or other useful metadata
- ◆ Current project:
 - ▷ rank assays based on likelihood of assay description
 - ▷ pick out relevant text
- ◆ Hundreds of thousands of candidate assays for BAE
- ◆ Overlap with ChEMBL: ~600 (1.7M \cap 70K) - metadata, SAR

Crowdsourcing

◆ Anyone can authenticate on beta.bioassayexpress.com

The screenshot shows the BioAssay Express website interface. At the top, there is a browser window with the address bar showing "Secure | https://beta.bioassayexpress.com". The website header includes the "CDD BioAssay Express" logo, a search bar with the text "find by ID", and a "Login" button. A dropdown menu is open next to the "Login" button, showing options for "ORCID" and "Google".

Main Pages

- Assign Assay
- Search Assays
- Explore Assays
- Curated Assays
- Holding Bay (2)
- Term Requests

Paste or drag to start annotating a new assay protocol, or start with the assay identifier.

PubChem Assay ID Curate

With **BioAssay Express** for the first time you can:

- Explore the similarity of assays within and between organizations

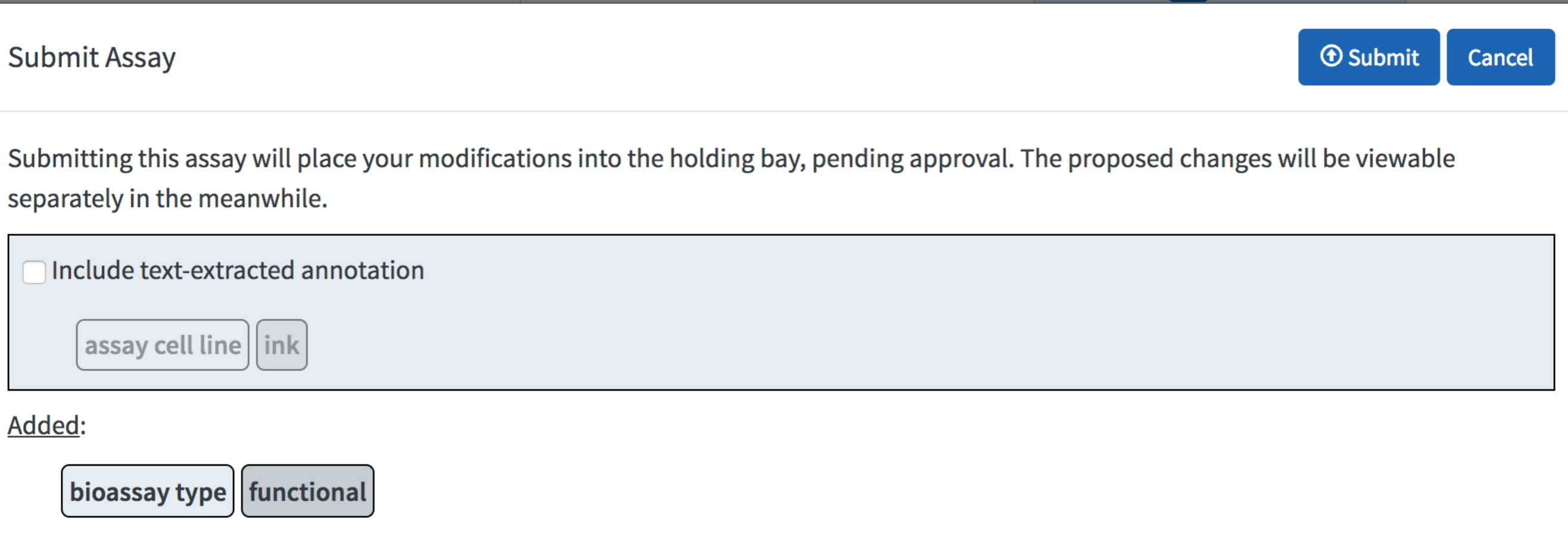
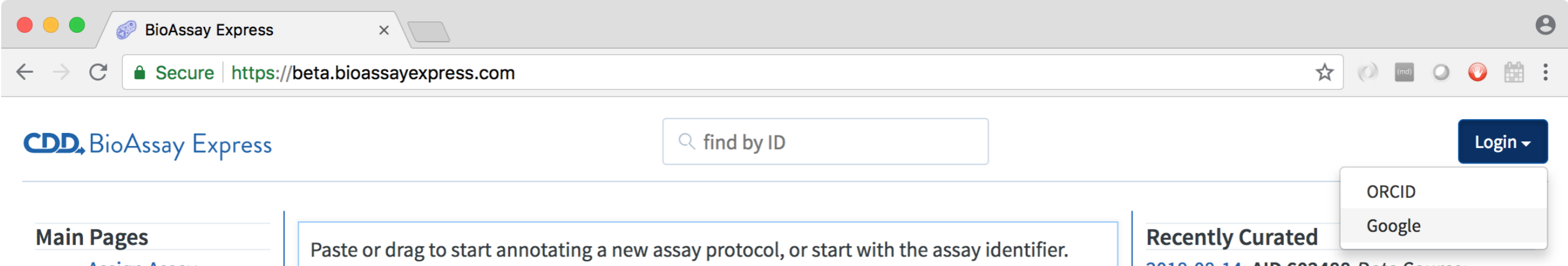
Recently Curated

2018-08-14: AID 602488 *Data Source: Sanford-Burnham Center for Chemical Genomics (SBCCG) Source Affiliation: Sanford-Burnham Medical Research Institute (SBMRI, La Jolla,... (13 annotations)*

2018-08-14: AID 743463 *Keywords: GCN5,*

Crowdsourcing

◆ Anyone can authenticate on beta.bioassayexpress.com



Integration

- ◆ Basic functionality integrated into **CDD Vault**
- ◆ ELN users can insert an **assay**
- ◆ Annotations are semantic, like standalone BAE

The screenshot displays a web browser window with the URL <https://app.collaborativedrug.com/entry/2286>. The page title is "Entry: 2286 - Dose Response". The user is logged in as "Alex Clark" in the "Alex Clark Sandbox" environment. The main content area shows the title "Dose Response of round 1 SAR compounds for Cdc42 probe extension project tested by multiplex of eight GTPase proteins under the condition of Mg Buffer, non-chelator buffer" with ID: 2286. Below the title is a rich text editor with a toolbar and a "Saved" indicator. The editor contains an "Assay Annotation" section with two paragraphs of text. Below the text is a "Target" section with a table of semantic annotations.

Category	Annotation
Bioassay Type	- binding type
Mode Of Action	- inhibition
Applies To Disease	- cancer
Organism	- Homo sapiens
Target	- protein
Biological Process	- cell proliferation - regulation of signal transduction - regulation of apoptotic process
Protein Identity	- GTPase [Homo sapiens] - Rac1 protein [Homo sapiens] - ras protein, partial [Homo sapiens]
Gene Identity	- CDC42: cell division cycle 42 - HRAS: HRas proto-oncogene, GTPase

Method

Method	Assay Supporting Method	Physical Detection Method
Bioassay	Assay Supporting Method	Physical Detection Method

Integration

- ◆ Basic functionality integrated into **CDD Vault**
- ◆ ELN users can insert an **assay**
- ◆ Annotations are semantic, like standalone BAE

Entry: 2286 - Dose Response

Collaborative Drug...covery, Inc. [US] | <https://app.collaborativedrug...>

CDD,VAULT - Alex Clark Sandbox

Explore Data | **ELN** | Import Data | Reports | Settings | Help | Log out

Alex Clark

Assay Annotation

TARGET | METHOD | OTHER

bioassay type
binding type

mode of action
inhibition

applies to disease
cancer

organism
Homo sapiens

target
protein

biological process
cell proliferation | regulation of signal transduction | regulation of apoptotic process

protein identity
GTPase [Homo sapiens] | Rac1 protein [Homo sapiens] | ras protein, partial [Homo sapiens]

gene identity
CDC42: cell division cycle 42 | HRAS: HRas proto-oncogene, GTPase

Go Terms

Submit or cancel

Project: Alex Clark S

Normal

Ass

This is apopt

This is conduc
MLSM

Target

Bioass
- binc

Mode
- inhi

Applie
- can

Organi
- Homo sapiens

Method

Bioassay | Assay Supporting Method | Physical Detection Method

ID: 2286

Saved

Acknowledgments

◆ BioAssay Express Team

- ⇒ Peter Gedeck
- ⇒ Hande Küçük
- ⇒ Barry Bunin
- ⇒ Jason Harris
- ⇒ Charlie Read
- ⇒ Samantha Jeschonek



◆ Collaborators

- ⇒ René Witte & Bahar Sateli (Knowlet)
- ⇒ Julia Davies (UCSF)
- ⇒ Mark Musen & John Graybeal (Stanford)
- ⇒ Stephan Schürer (U. of Miami)

◆ More information

<http://www.bioassayexpress.com>

<http://collaborativedrug.com>

Funding
NIH SBIR