ToxCast Assay Annotation Version 1.0 Data User Guide

OCTOBER 2014

**Compiled and Edited by the
ToxCast Assay Annotation Team:**
Jimmy Phuong
Lisa Truong
Nisha Sipes
Kristin Connors
Keith Houck
Richard Judson
Matthew Martin

**National Center for Computational Toxicology
U.S. Environmental Protection Agency
Research Triangle Park, NC 27709**

Disclaimer

The purpose of this document is to assist new data users in understanding how to use the annotation terms within the ToxCast High-Throughput Screening (HTS) and High Content Screening (HCS) data. Although it may be useful for the readers to have some prior knowledge of ToxCast assays from previous ToxCast publications, this data users guide is intended to provide sufficient exposure to annotation concepts and terminologies used. For detailed assay protocol information, it is recommended that the reader consult ToxCast publications or the US Environmental Protection Agency (EPA) - National Center for Computational Toxicology (NCCT).

We recommended that data users first obtain the latest version of the ToxCast screening data and assay annotation files prior to evaluating the ToxCast assay annotation. The October 2014 version of this document was released with two attached comma-separated value files: an annotation binary map and a reproducibility summary. Please consult the EPA ToxCast website for additional resources [http://www.epa.gov/ncct/toxcast/].

Use-cases with their suggested solutions are introduced to lead the data users using practical search criteria. As a quick-search alternative solution, a sample R script that is applicable across the use-cases is provided in the appendix.

The intent of the annotation term glossary is to represent the terminology and context of usage for each annotation term. Unless otherwise represented, the definitions are taken from BioAssay Ontology (BAO) version 1.6 (www.bioassayontology.org).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC50 ....Concentration at 50% maximal activity
ACEA ....ACEA Biosciences
APR ....Apredica
ATG....................Attagene
BAO....................BioAssay Ontology
BSK ....BioSeek
ESR1..................Estrogen Receptor Alpha
GO ....Gene Ontology
GPCR.................G-Protein Coupled Receptor
HCS ....High-Content Screening
HTS ....High-Throughput Screening
NCCT ....National Center for Computational Toxicology
NVS ....NovaScreen
OT......................Odyssey Thera
Tox21.................Toxicity Testing in the 21st Century

ToxCast Assay Annotation Summary

Overview

The EPA's ToxCast Program seeks to use high-throughput, *in vitro* biological assays to characterize potential health hazards of chemicals with limited to no toxicity information for use in chemical prioritization for more in-depth study. These chemical screening assays range from simple, single-readout to complex, multiplexed and multi-parametric technologies. The ToxCast Assay Annotation captures both technological and biological descriptions of each assay, which includes its measured readouts and analyzed results. This information is essential for proper interpretation of screening results and will facilitate higher level analyses, e.g. incorporation in to Adverse Outcome Pathways (AOP) models. Specific use-cases will be described for merging the annotations to the ToxCast chemical screening data as well as ways for searching the annotations by biology or technology.

Table 1: General content readouts by assay source

Assay Source	Assay	Assay Component	Assay Endpoint	Content readout type	
				Single	Multiplexed
ACEA Biosciences (ACEA)	1	1	2	•	
Apredica (APR)	3	30	60		•
Attagene (ATG)	3	83	83	•	•
BioSeek (BSK)	8	87	174		•
NovaScreen (NVS)	275	275	420	•	
Odyssey Thera (OT)	17	17	17	•	
Tox21	35	65	65	•	
Subtotal	342	558	821		

The EPA's ToxCast Program released data for 821 *assay endpoints on* October of 2014 (Table1). These *assay endpoints* are analyses with outputs (e.g. AC50 values) derived from 558 *assay components* (i.e., raw data readouts). The *assay components* are generated from 342 *assays* (i.e., experimental protocols); therefore, the 821 *assay endpoints* are derived from the 342 *assays*, where each *assay endpoint* comes from a single *assay component* and a single *assay*. To word the relationship in the opposite direction, an *assay* can generate one or many *assay components*, and one or more *assay endpoints* from each *assay component*. With few exceptions, the complete ToxCast Phase I and II chemical sets have been tested across these 342 *assays*, to yield the 821 *assay endpoints* of data.

As HTS data is generated from new HTS assays, it is necessary to communicate the technological and biological features of the assay, though this often encounters the challenge of

1

consistent vocabulary usage across all assays. Using controlled vocabulary principally from the BAO version 1.6 (http://bioassayontology.org/) along with other publicly available ontologies, the ToxCast Assay Annotations provide the structure necessary to capture and deliver descriptive information as new vocabulary enter common use and the annotations essential to understand the ToxCast HTS assay data. The structure takes the annotations and organizes them as attributes of the *assay* (i.e., experiments), the *assay components* (i.e., readouts), or the *assay endpoints* (i.e., analyzed readouts) enabling aggregation and differentiation of the data generated through ToxCast and Tox21.

Assay Annotation Structure

Overview

The ToxCast Assay Annotation Version 1.0 contains 38 annotations. An annotation is a descriptive feature (e.g., cell short name) that captures a particular aspect or annotation term (e.g., HepG2). A short description of the annotations is provided in Table 2. The individual annotations are developed from the BAO 1.6 framework and then further structured under the three following criteria: the annotations are 1) separated into four types of information: *i* - assay identification information such as the *assay source*, annotation level (Figure 1A) and peripheral catalog information)*, ii* - design information such as the technology, format, and objective aspects (Figure 1B-D) that decompress the assay's innovations, *iii* - target information such as the target of technological measurement and the biologically intended target (Figure 1F), and *iv* - analysis information about how the data were processed and analyzed (Figure 1G); then 2) arranged to display dependencies between annotations; then 3) each annotation was assigned as a feature of one of three levels: 'assay' level, 'assay component' level, or 'assay endpoint' level, which are analogous to the experiment, the measured readouts, or the analyzed data, respectively. Using these criteria, the assignment puts an annotation as a feature of a certain level, enabling data users to use the annotations to filter down to the *assay endpoints* of interest or to approach the screening data with questions directed at any of the four types of information.

Figure 1: The assay annotation structure. The 38 annotations can be grouped into A) assay identification information, B) design information, C) target information, and D) analysis information. Relationships between annotations are either one-to-many (solid arrow) or conditional (dashed arrow), where certain dependencies may not be applicable.
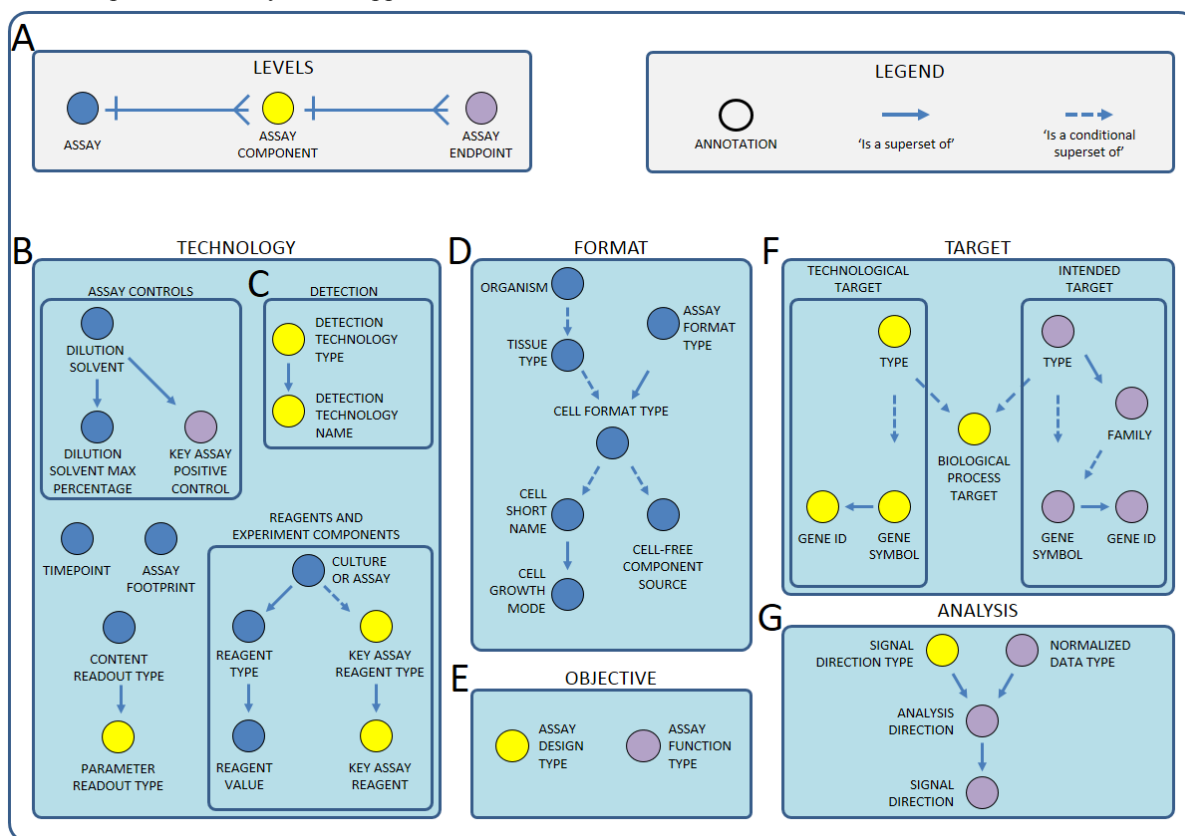
Annotation relationships

Most of the 38 annotations are related to at least one other annotation (Figure 1). The solid arrows depicted in Figure 1 point to annotations that are dependent on the parent annotation; thereby, the parent annotation is a superset grouping of the dependent annotation. Five annotations (i.e., *assay design type, detection technology type, assay format type, technological target type, intended target type,* and *intended target family*) were further separated into subtypes to clarify what concepts are considered from the bigger grouping. The dashed arrows depict points in the annotations that are conditional and may not need to be characterized for all assays due to the limitations from the assay technology or target. Annotations that are not applicable to be characterized are given a placeholder 'NA' or '0'. For instance, a cell-based assay will have a 'NA' annotation for *cell-free component source*.

Table 2: The 38 annotations and a short description of the concepts captured.

Annotation name	Short description of the annotation	Example
assay source name	a short name for the entity that conducted the assay	'BSK'
assay source long name	the long name for the entity that conducted the assay	'BioSeek'
assay name	a short name for the assay	'BSK_3C'
assay component name	a short name containing the assay and a component readout	'BSK_3C_IL8'
assay endpoint name	a short name containing the assay, the component readout, and an analysis applied	'BSK_3C_IL8_up'
timepoint hr	the duration length to conduct the test portion of the assay	'24'
organism id	the NCBI taxonomy id for the organism or cellular derivative used for the assay	'9606'
organism*	the organism related to the target of the assay	'human'
tissue	the organ-level, anatomical entity of the protein or cell used in the assay	'vascular'
cell format	the cellular or subcellular format of the assay	'primary cell'
cell free component source	the cellular or sample tissue source of the assayed gene protein	NA
cell short name	the name of the cell line or primary cell used	'umbilical vein endothelium'
cell growth mode*	the growth mode of a cell line	'Adherent'
assay footprint	the physical format, such as plate density, in which an assay is performed	'microplate: 96-well plate'
assay format type* ¥	the conceptual biological and/or chemical features of the assay system	'cell-based'
content readout type*	the throughput and information content generated	'multiplexed'
assay design type* ¥	the method that a biological or physical process is translated into a detectable signal	'binding reporter'
detection technology type ¥	the type of detection signal	'Fluorescence'
detection technology	the name of the detection technology method	'ELISA'
key positive control	the designated positive control	'NA'
dilution solvent	the solvent used as the negative control and to solubilize the test chemical	'DMSO'
dilution solvent percent max	the maximal amount of the dilution solvent tolerable for a particular assay	'0.2'
key assay reagent type	the type of key determinant substance of the assay	'antibody'
key assay reagent	the name of the key determinant substance of the assay	'IL-8 antibody'
parameter readout type	the data is a multiparametric calculation or a single detected readout	'single'
assay function type	the purpose of the analyzed readout in relation to others from the same assay	'signaling'
normalized data type	the fold induction or percent activity scale in which the assay data is displayed	'log10_fold_induction'
signal direction type	the expected direction of the detected signal in relation to the negative control	'both'
analysis direction	the analyzed positive (upward) or negative (downward) direction	'positive'
signal direction	the direction observed of the detected signal in relation to what was expected of it	'gain'
biological process target	the biological process or processes investigated by the assay	'regulation of gene expression'
technological target type ¥	the measured chemical, molecular, cellular, or anatomical entity	'protein'
technological target gene id	the Entrez gene ID for the molecular target measured	'3576'
technological target gene symbol	the Entrez gene symbol for the molecular target measured	'CXCL8'
intended target type ¥	the objective chemical, molecular, cellular, pathway or anatomical entity	'protein'
intended target gene id	the Entrez gene ID for the molecular target that is the objective	'3576'
intended target gene symbol	the Entrez gene symbol for the molecular target that is the objective	'CXCL8'
intended target family ¥	the target family of the objective target for the assay	'cytokine'
Culture or Assay	the culture or assay condition for reagent annotations	
Reagent Name Value type	the type of substance or function served by the reagent	
Reagent Name Value	the name of the reagent	

* The descriptions are adapted in part from the BAO definitions; ¥ This annotation is separated to capture types and subtypes

A key addition to the ToxCast Assay Annotation is the separation of *technological target* information and *intended target* information (Figure 1F). The technological target is mapped to the 'assay component' level as it represents the specific target of each assay readout and is assigned solely based on technological parameters. In contrast, the intended target is mapped to the 'assay endpoint' level as it represents the biological intention of the assay and considers the signal direction relative to the negative controls, and may consider the positive control and other parameters.

A good example of the distinctions and the relationships between technological and intended target is the ACEA_T47D assay, which has a single assay component at the 80 hour time point. The assay utilizes the xCELLigence Real Time Cell Analysis technology to measure cell density by electrical impedance measurements (Xing et al. 2006). For that reason, the *technological target type* is 'cellular' and specifically increased or decreased cell proliferation (as well as potential confounding signals involving cell morphology changes or other cellular-level treatment-related changes). The assay uses the T47D cell line, which is responsive to estrogens, so the intention of the ACEA_T47D assay is to measure increased cell proliferation as a result of the presence of estrogen mimicking chemicals or decreased cell proliferation due to the presence of cytotoxic chemicals (Rotroff et al. 2013). This is evident by the positive controls used in the experiment and in all downstream data normalization with 17beta-estradiol used to benchmark increased cell proliferation and MG-132 used to benchmark decreased cell proliferation. Therefore, the ACEA_T47D_80hr_positive (signal direction: gain) intended target was mapped to the estrogen receptor pathway, whereas the ACEA_T47D_80hr_negative (signal direction: loss) intended target was mapped to cell cycle/cytotoxicity. The distinctions at the *assay component* and *assay endpoint* levels between technological and intended target information ensures clear communication for potential downstream applications of the assay data for modeling and chemical prioritization.

Use-Cases for the Assay Annotations

## *Merging annotations with the ToxCast chemical screening data*

The ToxCast Assay Annotation files have been released as two spreadsheets, the 'study design' information and the other containing 'target info' information. Both files share the *assay endpoint* names (row) by different annotations (columns). Moreover, the ToxCast chemical screening data is released as a separate file containing the chemical identifiers (rows) by the *assay endpoint* names (column).

When using the spreadsheet format, it is recommended that the user start from the annotation file to find the *assay endpoints* of interest before progressing to filter the screening data files. To find the chemical screening results for each *assay endpoint*, limit the screening data file columns to the *assay endpoints* of interest.

Search by target genes

Context: A user has a gene target of interest. The user wants to find out whether the gene symbol is represented in ToxCast, which *assay endpoints* target it, and what annotations are associated to these *assay endpoints*. Figure 2 provides a logical workflow for this use-case.

Figure 2: Use-case workflow for searching by gene
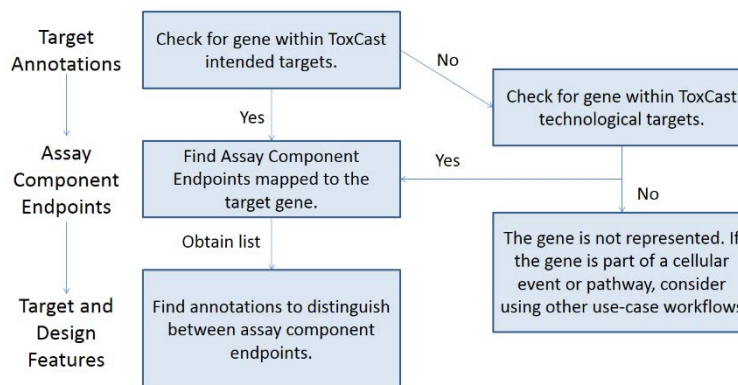
Solution: The user can use the target annotations from the 'target info' file to map back to the *assay endpoints* of interest.

Using the registered NCBI Entrez gene symbol or gene id for the target of interest (http://www.ncbi.nlm.nih.gov/gene/), the user can use these terms to filter the *intended target gene symbol* or *intended target gene id* annotations, respectively. First, the gene must be checked

to see if it is an *intended target gene symbol*; if not, check if it is a *technological target*. If the gene is found in either annotation, obtain the list of *assay endpoints*.

Incorporating other 'target info' and 'study design' information, the user can use any number of annotations to identify the features of the *assay endpoints*. Some recommended annotations include the *intended target type*, *technological target type*, *assay design type*, and *signal direction*.

A target, particularly a gene, can have several forms. The *intended target type* captures the objective form of the intended target, while the *technological target type* provides the measured form of the target used in the experimental methods (Figure 1F). Table 3 depicts the representation of technological versus intended targets, where boxed values represent technological targets that were measured to understand changes in a different *intended target type*. To provide more specificity, we've characterized subtypes to better define each target type. Currently, a gene may be annotated as a 'protein' or 'pathway' *intended target type* (shown in Carolina Blue), where the corresponding *technological target types* may range from 'DNA', 'RNA', 'protein', or 'cellular' measurements. For targets that are 'cellular' features, the subtypes identify the 'nucleus', 'mitochondria', or 'lysosome' as the focal point organelle; 'cellular' may also be a subtype to identify the whole cell or whole cellular response as the focus, as seen in cytotoxicity or cell number measurements.

Table 3: Technological and intended target types. The counts are the number of targets separated by the *technological target type*, *intended target type* and *assay source*.

Intended	cellular					pathway				protein				dna	chemical
Technological	ACEA	APR	ATG	BSK	Tox21	ACEA	APR	OT	Tox21	ATG	BSK	NVS	Tox21	APR	Tox21
cellular (28)	1	12	1	12	1	1	0	0	0	0	0	0	0	0	0
protein (686)	0	6	0	8	0	0	29	31	16	0	154	420	22	0	0
RNA (138)	0	0	0	0	0	0	0	0	0	138	0	0	0	0	0
DNA (24)	0	12	0	0	0	0	6	0	0	0	0	0	0	6	0
chemical (32)	0	0	0	0	20	0	0	0	0	0	0	0	0	0	12

The *assay design type* annotation captures the method by which the technological target is measured (Figure 1E). Table 4 shows the annotation terms under the *assay design type/subtype* ordered by the number of *assay endpoint* mapped to those terms. Each *assay design type/subtype* highlights how chemicals may affect targets of interest at varying levels of biological relevance. For example, an *assay endpoint* measuring a gene using a 'binding reporter' *assay design type* highlights functional protein changes like changes to the gene protein's enzyme-substrate affinity. For the same gene target, another *assay endpoint* that uses an 'inducible reporter' *assay design type would* highlight the changes to the transcriptional responses. The majority of ToxCast assays were characterized as 'binding reporters', 'enzyme reporters' or 'inducible reporters', but *assay endpoints* have also been characterized as 'conformation reporters', 'morphology reporters', or 'growth reporters'.

Table 4: ToxCast assay design types.
Values are counts of *assay endpoints*

Assay design type	subtype	Totals
Enzyme reporter	Enzyme activity	296
Binding reporter	ELISA	149
	Radioligand binding	120
	Protein fragment complementation	14
	FRET	8
	Fluorescent polarization	1
Inducible reporter	mRNA induction	84
	Beta lactamase induction	23
	Luciferase induction	10
	Fluorescent protein induction	2
Viability reporter	ATP content	20
	Protein content	24
	DNA content	12
	Cell number	7
Morphology reporter	Cell phenotype	26
Background reporter	Artifact detection	10
Membrane potential reporter	Dye binding	7
Conformation reporter	Protein conformation	6
Growth reporter	Real-time cell-growth kinetics	2

The *signal direction* (Figure 1G) indicates whether the *assay endpoint* provides 'gain' or 'loss' of signal data. This annotation interprets the *signal direction type* and *analyzed direction* annotations together, and reflects whether chemicals found to be produce a hit (i.e., AC50 $\leq$ 1000 μM) causes a gain or loss of signal. This concept also extends to 'cellular' and 'chemical' target types.

Applicability: This approach may be useful for users who start with a pre-determined list of chemical-gene or disease-gene interactions, and are looking to develop hypotheses about how chemicals may interact through the candidate genes.
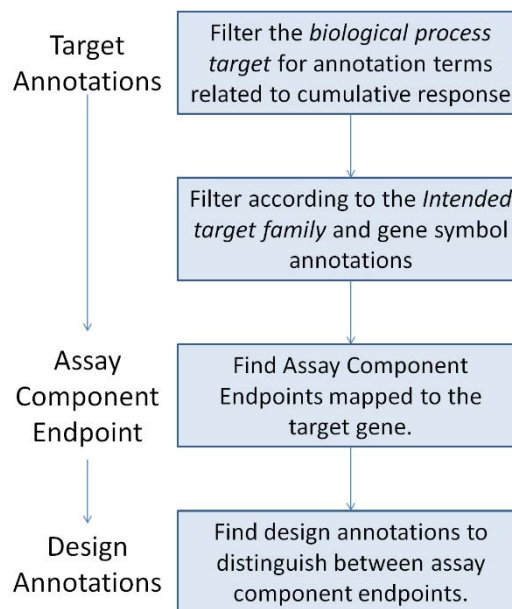
Limitations: There are thousands of known genes. Gene symbols, such as ESR1 for estrogen receptor alpha, can be a conserved targets across species but some other gene symbols may be species-specific thus presenting obstacles to mapping conserved targets across species. In some assays, genes of interest were transfected from one organism to be expressed in the cells of a different species. To clarify, though the *organism* annotations captures the species that the cellular and gene protein extracts were obtained from, it is not meant to represent the species genome. Therefore, it is imperative that the user have the correct NCBI Entrez gene symbol to search with, and to match the *organism* annotation if endogenous gene expression is desired.

The target type may play a role in the responsiveness of a gene to chemical perturbation. For example, a chemical may be specifically targeting the gene of interest in a complexed form (e.g. dimerized protein). Looking for activity through the mRNA transcript expression may come secondary to the initial chemical actions. The target type, other biological factors, as well as instrument limitations may explain why AC50 values vary between different *assay endpoints*.

Search for targets that are cumulative responses

Context: A chemical may be active in single gene-protein cell-free assays; but the user wants to know the system-level responses. This includes cell cycle events, morphological responses, and other regulatory processes that have a cascade of interactions. The user may not know what annotation terms to use, but will decide as they get to the *assay endpoints* of interest. Figure 3 provides a logical workflow for this use-case.

Figure 3: Use-case workflow for searching for cumulative response targets.

Target Annotations → Filter the *biological process target* for annotation terms related to cumulative response

Filter according to the *Intended target family* and gene symbol annotations

Assay Component Endpoint → Find Assay Component Endpoints mapped to the target gene.

Design Annotations → Find design annotations to distinguish between assay component endpoints.

Solution: Without knowing what keyword to look for, the user can begin by filtering through the *biological process target* annotations within the 'target info' file (Figure 3). As a feature of the assay component level, each assay component is currently characterized with a single *biological process target* to represent the higher biological concept as they relate to the measured readout (Figure 1F). From there, the *intended target family* and *intended target gene symbol* annotations can help narrow down the cumulative responses. The resulting *assay endpoints* can then be used to filter the 'study design' file to differentiate between factors of the experiments and components therein.

At present, there are 12 unique annotation terms for the *biological process target* annotations across all of ToxCast. Eight of these annotation terms relate to cumulative responses i.e., 'cell cycle', 'cell death', 'cell morphology', 'cell proliferation', 'mitochondrial depolarization', 'oxidative phosphorylation', 'protein stabilization', 'regulation of gene expression', and 'regulation of transcription factor activity'. Each of these annotation terms represents a different facet of biological response. Some are more complex in nature and require other target annotations to understand the context of how the *biological process target* is probed, whereas 'regulation of catalytic activity', 'receptor activity', and 'receptor binding' *biological process targets* are considered to be more indicative of single-gene mediated activities.

Figure 4: Filtering the *assay endpoints* based on *biological process target* annotations
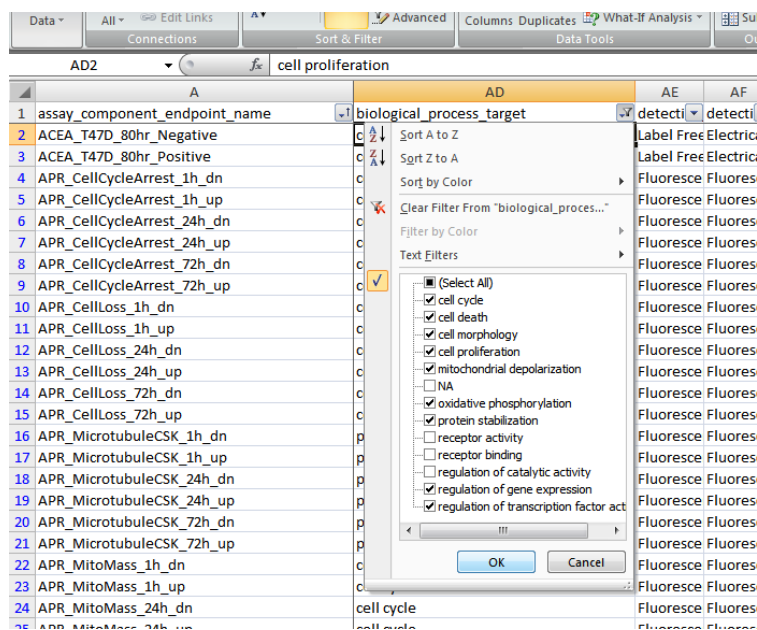
After selecting the *biological process target* of interest, the *intended target family* can be used to separate cellular responses from gene-mediated pathway responses. Together, 'cell cycle' and 'cell morphology' *intended target families* encompass cellular life-cycle events such as 'cytotoxicity', 'apoptosis', 'arrest', 'stress response', 'cell conformation' and 'organelle conformation' subfamilies. These

intended target families and subfamilies are displayed in Figure 5. In addition, with the exception of 'background measurement', other *intended target families* and subfamilies display the genetic groupings for each characterized *intended target gene symbol*. The technological and intended target gene annotations could further describe the context that the pathway was explored.

At this point, the *assay endpoints* are limited to a subset of interest by target information. This list of *assay endpoints* can be further subset by the 'study design' file (e.g., for the *cell format type, organism* and *tissue* annotations).

The *cell format type* is meant to annotate the state of the cellular material (e.g. cell line, tissue-based cell-free, or primary cell co-culture) as it is used during chemical exposure. Associated with the *cell format type* are the *tissue type* and *organism* annotations (Figure 1D). The *tissue type* annotation attempts to link the biological material used to an anatomical entity (e.g. brain, liver, and kidney), while the *organism* annotation further links the cellular-biological material to the species of origin. *Organism* is accompanied with *Organism ID*, which uses the NCBI Taxonomy ID (http://www.ncbi.nlm.nih.gov/taxonomy/) for the respective organism. For cell-based experiments, the *cell short name* annotations should be consulted since differences in sensitivity may be seen within cells of the same tissue type, e.g., human primary hepatocyte, HepG2 cells, and HepaRG cells are all derived from the liver. For primary cells, also consider the *Organism* annotations.

Application: This approach can be useful for scientists looking to identify chemical activity at different levels of biological relevance. The *biological process target* offers different categories to focus the higher biological concept from the *assay component*. Counter to the example given, the *biological process target* can also identify *assay endpoints* mapped to single-gene responses. Other target annotations will be necessary to better understand the biological context. The user can also bypass the first step within the workflow and use the *intended target family* annotations to filter to a target of interest, if cumulative versus single-gene is not a criteria.

Limitations: The *biological process target* annotation terms follow Gene Ontology (GO) biological processes. Because the GO biological processes are only partially represented in the ToxCast assays, the hierarchical connections among the terms used are not well-represented. To

better understand how the *biological process target* is investigated, examine the other target annotations.

The current target annotations have a bias towards gene-oriented assays. For some assays, such as the ATG CIS-factorial assay, there is not an agreed upon naming or identification convention for the transcriptional response elements. Therefore, intended target gene symbols downstream of the response element are currently characterized in its representation.

We recognize the need to have more definitions for assays that target gene or pathway 'agonist' and 'antagonist' activities as well as for those that do not probe gene targets. The APR assays are an example of where protein technological targets can be measured using different probes and antibodies to understand correlated, cellular morphological changes. While the readouts can be understood independently, the APR assays can be looked at across readouts to understand cellular arrest during different segments of the cell cycle pathways.

The *intended target family* and subfamily annotations attempt to represent the individual targets across *assay endpoints*. These families and subfamilies pertain to gene families and include morphological and cell cycle concepts. These annotation terms are adapted from ontology terms. The source ontologies can have convoluted, tree-like connections that are difficult to understand. Therefore, these annotations were developed to be simplified groupings for communication purposes but precise enough to reflect the biology; they serve only to associate assays by their targets using annotation terms that are understandable. On the other hand, the annotation terms are controlled vocabulary and have the possibility of being connected back for additional information from the source ontology.

The *Organism* annotations should not be assumed to reflect the gene target's genome of origin. Some assays may use cells that express gene proteins through recombinant transfection. It is advisable that the user check the 'target gene ID' of their genes of interest with the NCBI Entrez Gene database (http://www.ncbi.nlm.nih.gov/gene/).

Figure 5: Intended target family annotation terms. The node size and font size of each node is relative to the number of *assay endpoints* mapped to that *intended target family* and *subfamily*.
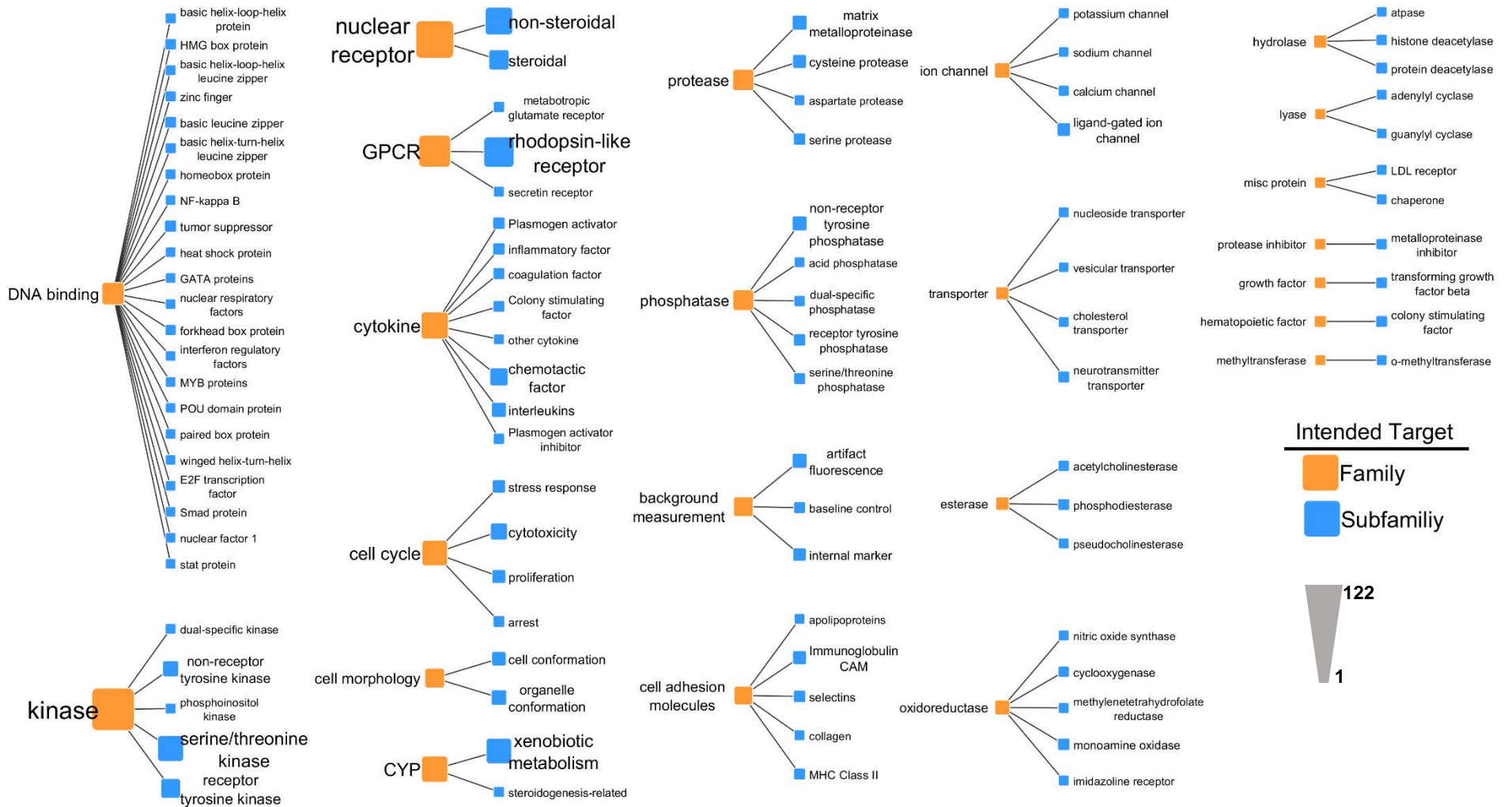
Search by technology

Context: The user is interested in a specific kind of laboratory technology. They want to find *assay endpoints* within ToxCast that either used data from the same technology or is very similar.

Solution: The user can first look through *detection technology name,* which annotates the general name or acronym for the technology used. Alternatively, the *detection technology type/subtype* annotation groups the technologies using annotation terms representative of their detection signal, which can filter down to *detection technology names* that are relatable. Thereafter, *key assay reagent type, key assay reagent*, *technological target gene symbol* and the *intended target gene symbol* annotations can provide additional insight onto how the technology probes the target.

Table 5: ToxCast detection technology types.
Values are counts of assay endpoints

Detection technology type	subtype	Totals
Fluorescence	Fluorescence intensity	601
	FRET: TR-FRET	8
	Fluorescence polarization	1
Radiometry	Scintillation counting	136
Spectrophotometry	Absorbance	40
Luminescence	Bioluminescence	28
	Chemiluminescence	1
Microscopy	Optical microscopy: Fluorescence microscopy	4
Label-Free Technology	Electrical Sensor: Impedance	2

ToxCast HTS assays currently has 6 *detection technology types* as shown in Table 5. The majority of assay technologies use some form of 'fluorescence' to quantify changes, but 'radiometry', 'spectrophotometry', and 'luminescence' are also applied for specific uses. With the exception of the 'label-free technologies', these detection technologies can depend on a key reagent to quantitate assay readouts. For instance, 'fluorescence' technologies may rely on emissions generated from a type of fluorophore or fluorescent antibody, while 'radiometry' technologies rely on radiolabeled ligands. We recognize that each *detection technology* can be unique, capturing the *detection technology type* can be a critical piece of information for comparing the study designs across the same target (Figure 1C).

When comparing the *assay design type* annotations, the 'fluorescence' technologies are observed in nearly every *assay design type* annotation term. Hence, it is recommended that *assay design types*, *intended target gene symbol*, and *signal direction* annotations be used to select the methods, objective target, and response provided in the resulting *assay endpoints*. Data users may also want to peruse the *key assay reagent* and *key assay reagent type* annotations.

Application: As new technologies are developed, users may be interested to see how the design and data generated may compare. This approach may also be useful to compare assay results with non-related technologies.

Limitations: Gene proteins can have various functions or play a role in a variety of responses. However, not all gene targets or cellular targets are measurable using the same technological methods. If the data user were to begin filtering through ToxCast by using the *detection technology* annotations, though the target annotations are still available to them, this approach may sacrifice the biological targets that have limited technological applications.

Annotation term glossary

Unless specified, definitions were adapted from BAO v1.6
¥: the definition was extracted from BAO v2.0
‡: the definition was extracted from Gene Ontology (GO)
φ: the definition was extracted from Medical Subject Headings (MeSH)
^: the definition was extracted from G-Protein Coupled Receptor BioAssays Ontology (GPCR-BAO)
§: proposed new annotation term and definition

Analysis direction: the direction for data fitting relative activity

§**Positive analysis direction**: The data were analyzed for <u>increases</u> in activity relative to the negative controls.

§**Negative analysis direction**: The data were analyzed for <u>decreases</u> in activity relative to the negative controls.

Assay Design Type: the method that a biological or physical process is translated into a detectable signal

§**Background Reporter**: The technology that is used to identify and understand background noise and artificial signals generated due to non-target related interference. These could arise from inherent physicochemical properties of the test substance such as autofluorescence and agglomeration, or obstruction to the test environment such as degradation or rearrangement of the test protein.

Binding Reporter: The technology used to quantitate the interactions between two molecules, e.g. perturbagen-protein, protein-protein, protein-DNA, etc. The technologies used are radioligand binding, resonance energy transfer, fluorescence polarization, protein fragment complementation assay, and several label-free methods, including surface plasmon resonance, impedance, nuclear magnetic resonance, X-ray diffraction, isothermal titration calorimetry, mass spectrometry, etc.

Conformation Reporter: The technology relates to measurement of protein and nucleic acid conformational properties. Proteins undergo changes to their tertiary structure upon heating, binding to perturbagens, other proteins and substrates. Nucleic acids have characteristic secondary structures, which include the B-family of structures, A-form, Z-form, guanine quadruplexes, cytosine quadruplexes, triplexes, etc. Information about conformational properties of protein and nucleic acids can be obtained by circular dichroism spectroscopy, thermal shift, fluorescence resonance energy transfer (FRET), nuclear magnetic resonance (NMR), X-ray crystallography, etc.

Enzyme Reporter: The technology used to measure the effect of a perturbagen on the enzyme activity, which can be measured either directly or coupled to a secondary signal.

§**Growth Reporter**: The technology used to measure the effects of a perturbagen on the rate of cell proliferation.

Inducible Reporter: Inducible reporter technologies involve the use of a reporter gene to study the effect of perturbagens on gene expression. Some reporter genes can be easily detected, e.g,. green fluorescence protein, and others function as selectable markers, e.g., antibiotic resistance.

Membrane Potential Reporter: All living cells maintain a resting membrane potential, which is the difference in voltage between the interior and exterior of the cell. Electrically active cells, such as neurons, muscle and endocrine cells undergo changes in membrane potential due to depolarization and hyperpolarization as a result of opening and closing of ion channels. Action potentials are large and rapid changes in membrane potential which are generated as a result of a large depolarization event. In neurons, action potentials are used to transmit signals across neurons, while in muscle and endocrine cells they lead to intracellular

events, such as muscle contraction and hormone release, respectively.

In addition, mitochondria possess a trans-membrane potential as a result of their respiratory function. Membrane potential can be measured either by using electrophysiological techniques or by staining with dyes whose binding is dependent on the membrane potential.

Morphology Reporter: The technology used to measure the size and shape of cells and organelles (phenotype). These are detected by microscopy.

Viability Reporter: Compound cytotoxicity is an important parameter to measure when developing potential human therapeutics. Cytotoxicity is determined as a measure of radioisotope (3H thymidine or 51Cr) release, lactate dehydrogenase release from damaged cells, tetrazolium salt and alamar blue reduction, fluorescent dyes that selectively stain live or dead cells, and decrease in ATP content. The viability reporter technologies measure cell life or death by quantitating ATP content, caspase activity, membrane potential, and so on.

Assay Design Subtypes: the subtype method through which a biological or physical process is translated into a detectable signal

§**Artifact Detection**: The technology designed to detect interference signals generated due to the test substance inherent chemical properties or through non-target related interactions [a subset of '**Background Reporter**']

ATP Content: ATP levels are an indicator of viable cell number. ATP values higher than controls (untreated cells) indicate proliferation and cultures with ATP concentrations lower than controls indicate cytotoxicity. ATP levels can be measured using different commercial kits. [a subset of '**Viability Reporter**']

Beta-Lactamase Induction: Beta lactamase gene is attached to the regulatory sequence of a gene of interest. The effect of a perturbagen on the expression of the gene of interest could be easily monitored by the detection of the product of beta lactamase enzyme. [a subset of '**Inducible Reporter**']

Cell Number: Cells are counted to determine their viability. Cultured animal cells are counted either manually using hemocytometer, or by an automated cell counter, e.g. Coulter counter or Vi-CELL. The latter can distinguish live and dead cells by the trypan blue cell exclusion method. Yeast and bacterial cell growth is determined by measuring absorbance in a spectrophotometer, or by viable plate count. [a subset of '**Viability Reporter**']

Cell Phenotype: The technologies used to quantitate the cell and organelle size and shape. This is detected by microscopy. [a subset of '**Morphology Reporter**']

DNA Content: Cellular nuclei can be by stained with Hoechst, DAPI, etc, which is a measure of cell viability and proliferation. [a subset of '**Viability Reporter**']

Dye Binding: Most often, the mitochondrial membrane potential which is called delta-psi is quantitated by staining with specific dyes whose binding is dependent on the potential. [a subset of '**Mitochondrial Membrane Reporter**']

ELISA: Enzyme linked immunosorbent assay or ELISA is used to detect the presence of a particular protein, either an antigen or antibody in the sample such as a body fluid. [a subset of '**Binding Reporter**']

Enzyme Activity: An enzyme is a protein that acts as a catalyst for specific biochemical reaction, converting specific substrates into chemically distinct products. Enzyme activity assays quantify the kinetics of the product formation or substrate depletion in the reaction, which result in fluorescence, luminescence or colorimetric outputs. [a subset of '**Enzyme Reporter**']

§**Fluorescent Polarization:** When linear polarized light is used to excite fluorophores, only those aligned with the plane of polarization will be excited. Fluorophores attached to small, rapidly rotating molecules result in low fluorescence polarization, but binding of the low molecular weight tracer to a large, slowly rotating molecule results in high fluorescence polarization. [a subset of '**Binding Reporter**']

Fluorescent Protein Induction: Fluorescent proteins are commonly used as reporter genes to study the effect of perturbagens on gene expression. Examples include green fluorescent protein (GFP) from the jellyfish Aequorea victoria and the various modifications of it created by amino acid substitutions to either optimize the codon usage in mammalian cells or to change the spectral properties of the protein. [a subset of '**Inducible Reporter**']

§**FRET**: Fluorescence resonance energy transfer (FRET) is based on the transfer of energy between two fluorophores, a donor and an acceptor, when in close proximity. The energy transfer is directed from higher-energy donor fluorophore to lower-energy acceptor fluorophore of labeled protein pairs. [a subset of '**Binding Reporter**']

Luciferase Induction: Luciferase gene is attached to the regulatory sequence of a gene of interest. The effect of a perturbagen on the expression of the gene of interest could be easily measured by the detection of light emitted as a product the luciferase reaction. [a subset of '**Inducible Reporter**']

¥**mRNA Induction** (also known as nucleic acid amplification method): A method that uses purified enzymes to isolate and then replicate specific nucleic acids to levels where they can be detected. Examples are PCR, TMA, NASBA, etc. [a subset of '**Inducible Reporter**']

Protein Conformation: This technology relates to measurement of protein conformational properties using one of the following methods: circular dichroism spectroscopy, thermal shift, fluorescence resonance energy transfer (FRET), nuclear magnetic resonance (NMR), X-ray crystallography, etc. [a subset of '**Conformation Reporter**']

§**Protein Content:** Cellular proteins can be tagged using sulforhodamine B stain, where decreased protein content is a measure of decreased cell viability. [a subset of '**Viability Reporter**']

Protein Fragment Complementation: This is used to study the interaction of two proteins. An enzyme or fluorescent protein is rationally dissected into two fragments and fused to two test proteins, whose interaction is being studied. Binding of the two test proteins results in the reconstitution of the enzyme or fluorescent protein from the two fragments. [a subset of '**Binding Reporter**']

Radioligand Binding: An analysis of receptor-ligand interactions in screening applications. It utilizes a radiolabeled ligand and a source of receptor (membranes, soluble/purified). [a subset of '**Binding Reporter**']

§**Real-Time Cell-Growth Kinetics**: It may also be referred to as a 'kinetic assay', these technologies monitor the rates at which cell proliferation occurs in real-time by measuring several time points over a period of time [a subset of '**Growth Reporter**']

Assay Format Type: the conceptual biological and/or chemical focus of the assay system

Biochemical: An *in vitro* format used to measure the activity of a biological macromolecule, either a purified protein or a nucleic acid. It is most often a homogeneous assay type, but can be heterogeneous if a solid phase such as beads is used to immobilize the macromolecule.

§**Cell-Based**: Assays using intact cells to measure cellular changes in response to exposure with the test substances.

§**Organism-Based**: Assays using an intact whole-organism to measure changes in response to exposure with the test substances.

Assay format subtype: the subset of the conceptual biological and/or chemical focus of the assay system

Cell-Based Format: Involves the use of living cells of eukaryotic origin and is a heterogeneous assay type. [a subset of '**Cell-Based**' format type]

Physicochemical Format: Assays in which the physical and chemical properties of perturbagens are measured, namely, aqueous solubility, octanol water partition, cell permeability models, for e.g. Caco2 cells, parallel artificial membrane permeability assay (PAMPA), etc. [a subset of '**Cell-Based**' format type and '**Cell-Based Format**' subtype]

Protein Complex Format: Two or more proteins interact to form a stable complex. Each of the members of this complex has a distinct role in mediating the overall function of the complex, which can be as a chaperone (DnaK, DnaJ and grpE), an enzyme (mitochondrial respiratory complexes), a signaling complex (G-proteins), etc. It can be viewed as a quaternary structure of the proteins. [a subset of '**Cell-Based**' format type and '**Cell-Based Format**' subtype]

Organism-Based Format: Involves the use of a living organism and is a heterogeneous assay type. [a subset of '**Organism-Based**' format type]

Protein single format: one protein sequence [a subset of '**Biochemical**' format type]

Assay function type: the purpose of the analyzed readout in relation to others from the same assay

§**Background Control**: A measure of the normal baseline. This could include measuring the expression levels of house-keeping genes.

Binding: Binding assays are performed to understand the interaction between two molecules, e.g. perturbagen-protein, protein-protein, protein-DNA, etc.

§**Enzymatic Activity**: An assay that measures the modulation of an enzyme activity by a perturbagen. An enzyme is a protein that acts as a catalyst for a specific biochemical reaction, converting specific substrates into chemically distinct products.

Reporter Gene: A gene that is attached to a regulatory sequence of another gene of interest and introduced into cultured cells, animals or plants. Certain genes function as reporters because they are easily identified and measured, or because they are selectable markers. Common reporter genes are luciferase, green fluorescent protein (GFP), beta-galactosidase and chloramphenicol acetyltransferase (CAT).

Signaling: Signaling pathway is a mechanism that converts a mechanical/chemical stimulus that impinges on the cell into a specific cellular response. Signal transduction starts with a signal to a receptor, and ends with a change in cell function which is mediated through a network of macromolecules.

Viability: A measure of the number of living cells in a culture. Viability is determined by measuring tetrazolium salt, alamar blue reduction, staining with selective fluorescent dyes, or ATP content. ATP levels are detected using a luminescence based kit such as CellTiter-Glo (Promega). ATP values higher than controls (untreated cells) indicate proliferation and cultures with ATP concentrations lower than controls indicates cytotoxicity.

Biological process target: the biological process or processes investigated by the assay

‡**Cell Cycle**: The progression of biochemical and morphological phases and events that occur in a cell during successive cell replication or nuclear replication events. Canonically, the cell cycle comprises the replication and segregation of genetic material followed by the division of the cell, but in endocycles or syncytial cells nuclear replication or nuclear division may not be followed by cell division.

‡**Cell Death**: Any biological process that results in permanent cessation of all vital functions of a cell. A cell should be considered dead when any one of the following molecular or morphological criteria is met: (1) the cell has lost the integrity of its plasma membrane; (2) the cell, including its nucleus, has undergone complete fragmentation into discrete bodies (frequently referred to as "apoptotic bodies\"); and/or (3) its corpse (or its fragments) have been engulfed by an adjacent cell in vivo.

‡**Cell Proliferation**: The multiplication or reproduction of cells, resulting in the expansion of a cell population.

‡**Mitochondrial Depolarization**: The process in which the membrane potential of the mitochondria changes in the depolarizing direction from the resting potential, from negative to positive.

‡**Oxidative Phosphorylation**: The phosphorylation of ADP to ATP that accompanies the oxidation of a metabolite through the operation of the respiratory chain. Oxidation of compounds establishes a proton gradient across the membrane, providing the energy for ATP synthesis.

‡**Protein Stabilization**: Any process involved in maintaining the structure and integrity of a protein and preventing it from degradation or aggregation.

‡**Receptor Binding**: Interacting selectively and non-covalently with one or more specific sites on a receptor molecule, a macromolecule that undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function.

‡**Regulation of Catalytic Activity**: Any process that modulates the activity of an enzyme.

‡**Regulation of DNA repair**: Any process that modulates the frequency, rate or extent of DNA repair.

‡**Regulation of Gene Expression**: Any process that modulates the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's coding sequence is converted into a mature gene product or products (proteins or RNA). This includes the production of an RNA transcript as well as any processing to produce a mature RNA product or an mRNA (for protein-coding genes) and the translation of that mRNA into protein. Some protein processing events may be included when they are required to form an active form of a product from an inactive precursor form.

‡**Regulation of Transcription Factor Activity** (also known as '**Transcription Factor Binding Transcription Factor Activity**'): Interacting selectively and non-covalently with a specific transcription factor, which may be a single protein or a complex, in order to modulate transcription. A protein binding transcription factor may or may not also interact with the template nucleic acid (either DNA or RNA) as well.

Cell format type: the cellular or subcellular format of the assay

§**Cell-Free**: The assay used purified proteins of endogenous expression or from recombinant transfection that were extracted from cell lines or lysates of cellular origin.

Cell Line: The assay used an immortalized cell line of known strain lineage.

Primary Cell: The assay used cells that are cultured from a subject (tissue or tumor) and which are not immortalized. They can undergo only a limited number of cell divisions prior to reaching senescence.

§**Primary Cell Co-Culture**: The assay used cells obtained from a donor cultured together with another cellular strain.

§**Tissue-Based Cell-Free**: The assay used purified proteins extracted from tissue of endogenous expression.

Content readout type: the throughput and information content generated

Single Readout: Typically uses a plate reader and generates one or two readouts.

Multiparametric Readout: This is typically performed in HCS, where multiple parameters are measured simultaneously from a single cell/well, by the detection of several fluorescent labels e.g., neurite length, nuclear size, protein translocation, etc.

Multiplexed Readout: The readout is measured from multiple targets from one well.

Detection technology type: the type of detection signal measured

Fluorescence: Detection techniques that use the principles of fluorescence, whereby incident light excites a fluorophore which then emits light at lower energy (higher wavelength). The emitted light is typically from the visible portion of the UV-Visible spectrum.

Label Free Technology: It allows the detection of binding interactions and cell-based reactions without the need to use conventional labels, e.g., fluorescent probes. The advantages include the ability to measure a) functional activity without modifying the binding partners with labels, b) binding interactions independent of functional activity, and c) cell-Based GPCR assays without the need for engineering cell-lines to over-express given receptors.

Luminescence: The emission of light from a substance that occurs from an electronically excited state which is reached either by a physical, mechanical, or chemical mechanism.

Microscopy: The field of using microscopes to see objects that cannot be seen with the unaided eye.

Radiometry: Detection technology using radioactive tracers, e.g., filter assay, SPA, etc.

Spectrophotometry: The use of a spectrophotometer to measure the amount of light that a sample absorbs. The instrument operates by passing a beam of light through a sample and measuring the intensity of light reaching a detector.

Detection technology subtypes: the subtypes of detection signal measured

Absorbance: The absorbance of a beam of collimated monochromatic radiation in a homogeneous isotropic medium is proportional to the absorption path length and to the concentration of the absorbing species. Absorbance is a logarithmic measure of the amount of light (at a specific wavelength) that is absorbed when passing through a sample. $A = -\log_{10}(I/I_0)$, where A is absorbance, I is the intensity of incident light and I_0 is the intensity of light transmitted after passing through the sample. [a subset of '**Spectrophotometry**' detection technology type]

Bioluminescence: Bioluminescence refers to the production and emission of light by a living organism. Bioluminescence is a naturally occurring form of chemiluminescence whereby energy is released by a

chemical reaction in the form of light emission. The chemical reaction can occur either inside or outside the cell, and can be detected using a plate reader or a charge-coupled device (CCD) camera. [a subset of '**Luminescence**' detection technology type]

Chemiluminescence: Chemiluminescence is the generation of light as a result of a chemical reaction using synthetic compounds and usually involving a highly oxidized species such as peroxide. [a subset of '**Luminescence**' detection technology type]

Electrical Sensor:Impedance: Impedance biosensors measure the electrical opposition of an interface in AC steady state with constant DC bias conditions. This is accomplished by imposing a small sinusoidal voltage at a particular frequency and measuring the resulting current; the process can be repeated at different frequencies. The current-voltage ratio gives the impedance. [a subset of '**Label Free Technology**' detection technology type]

Fluorescence Intensity: Fluorescence intensity is the measure of fluorescence from a sample. [a subset of '**Fluorescence**' detection technology type]

FRET:TR-FRET: Time resolved-fluorescence resonance energy transfer (TR-FRET) unites TRF (Time-Resolved Fluorescence) and FRET (Fluorescence Resonance Energy Transfer) principles. This combination brings together the low background benefits of TRF with the homogeneous assay format of FRET. Introducing a time delay (50-150 micro seconds) between the system excitation and fluorescence measurement allows the signal to be cleared of all non-specific short lived emissions. TR-FRET uses lanthanide donors which are fluorophores with long emission half-lives. [a subset of '**Fluorescence**' detection technology type and '**Fluorescence Intensity**' subtype]

Optical Microscopy:Fluorescence Microscopy: Optical or light microscopy involves passing visible light transmitted through or reflected from the sample through a single or multiple lenses to allow a magnified view of the sample. Fluorescence microscopy is used to image the distribution of molecules based on the properties of fluorescence emission. Excitation of fluorescent molecules with a specific wavelength results in the production of an image by the secondary fluorescence emission at longer wavelengths. [a subset of '**Microscopy**' detection technology type]

Scintillation Counting: It involves the incorporation of radiolabeled precursors into uniform distribution with a liquid medium capable of converting the kinetic energy of nuclear emissions into light energy. A scintillation counter is used to measure ionizing radiation. The scintillation counter consists of a sensor, called a scintillator, made of transparent crystal or an organic liquid that fluoresces when struck by ionizing radiation. A photomultiplier is then used to measure the light emitted by the sensor and is attached to an electronic apparatus that amplifies and quantifies the radiation signal. [a subset of '**Radiometry**' detection technology type]

Intended target family: the representative genetic family or biological process of the target

§**Background measurement**: The objective target is a representative measurement of the baseline noise.

Cell Adhesion Molecules (also referred to **Adhesion,** a subclass of protein): The proteins located on the cell surface which are involved with the binding with other cells or the extracellular matrix. They belong to 5 families: Immunoglobulin superfamily, integrins, cadherins, selectins and the lymphocyte homing receptors.

φ**Cell Cycle** (also known as '**Cell Physiological Processes**'): Cellular functions, mechanisms, and activities.

φ**Cell Morphology** (also known as '**Cytology**'): Used for cellular appearance of unicellular and multicellular organisms.

^φ**CYP** (also known as **'Cytochrome P-450 Enzyme System'**): A superfamily of hundreds of closely related HEMEPROTEINS found throughout the phylogenetic spectrum, from animals, plants, fungi, to bacteria. They include numerous complex monooxygenases (MIXED FUNCTION OXYGENASES). In animals, these P-450 enzymes serve two major functions: (1) biosynthesis of steroids, fatty acids, and bile acids; (2) metabolism of endogenous and a wide variety of exogenous substrates, such as toxins and drugs (BIOTRANSFORMATION). They are classified, according to their sequence similarities rather than functions, into CYP gene families (>40% homology) and subfamilies (>59% homology). For example, enzymes from the CYP1, CYP2, and CYP3 gene families are responsible for most drug metabolism.

^φ**Cytokine**: Non-antibody proteins secreted by inflammatory leukocytes and some non-leukocytic cells, that act as intercellular mediators. They differ from classical hormones in that they are produced by a number of tissue or cell types rather than by specialized glands. They generally act locally in a paracrine or autocrine rather than endocrine manner.

^φ**DNA Binding** (also known as '**Transcription Factor**'): Endogenous substances, usually proteins, which are effective in the initiation, stimulation, or termination of the genetic transcription process.

> ^φ**Nuclear Receptors**: Nuclear receptors are localized in the cytosol and mediate the activity of hormones and certain vitamins. They are all transcription factors, which upon activation by ligand binding, translocate to the nucleus, bind specific DNA sequence and activate gene expression. [a subclass of '**DNA Binding**' family and '**Zinc Finger**' subfamily]

^**GPCR** (also known as '**G-Protein Coupled Receptors**'): G Protein-Coupled Receptors (GPCRs), comprise the largest integral membrane protein family with nearly 1000 members. They integrate extracellular signals from hormones, growth factors, neurotransmitters, etc to downstream cellular responses. They are important pharmacological targets, being associated with a number of diseases. Upon ligand binding extracellularly, the GPCR signals through a downstream heterotrimeric G protein complex, consisting of the α, β and γ subunits.

^φ**Growth Factors**: The role played by a protein in initiating cell division via cell signaling, subsequent to binding to its cognate cell surface receptor.

^φ**Hydrolase**: Any member of the class of enzymes that catalyze the cleavage of the substrate and the addition of water to the resulting molecules, e.g., ESTERASES, glycosidases (GLYCOSIDE HYDROLASES), lipases, NUCLEOTIDASES, peptidases (PEPTIDE HYDROLASES), and phosphatases (PHOSPHORIC MONOESTER HYDROLASES). EC 3.

> ^φ**Esterases**: A class of enzymes that catalyze the formation of carbon-oxygen-carbon moieties on the substrate. [a subclass of '**Hydrolases**' target family]

> ^φ**Phosphatases** (also known as '**Phosphoric Monoester Hydrolases**'): A group of hydrolases which catalyze the hydrolysis of monophosphoric esters with the production of one mole of orthophosphate. EC 3.1.3. [a subclass of '**Hydrolases**' target family]

> ^φ**Proteases** (also known as '**Peptide Hydrolases**'): Hydrolases that specifically cleave the peptide bonds found in PROTEINS and PEPTIDES. Examples of sub-subclasses for this group include EXOPEPTIDASES and ENDOPEPTIDASES. [a subclass of '**Hydrolases**' target family]

^φ**Kinases** (also known as '**Phosphotransferases**'): A rather large group of enzymes comprising not only those transferring phosphate but also diphosphate, nucleotidyl residues, and others. These have also been subdivided according to the acceptor group. (From Enzyme Nomenclature, 1992) EC 2.7.

^φ**Lyases**: A class of enzymes that catalyze the cleavage of C-C, C-O, and C-N, and other bonds by other means than by hydrolysis or oxidation. (Enzyme Nomenclature, 1992) EC 4.

^φ**Methyltransferases**: A subclass of enzymes of the transferase class that catalyze the transfer of a methyl group from one compound to another. (Dorland, 28th ed) EC 2.1.1.

[§]**Misc proteins**: Proteins that do not have a clear gene grouping.

^φ**Non-steroidal hormones** (also known as '**Peptide Hormones**'): Hormones synthesized from amino acids. They are distinguished from INTERCELLULAR SIGNALING PEPTIDES AND PROTEINS in that their actions are systemic.

^φ**Oxidoreductases**: The class of all enzymes catalyzing oxidoreduction reactions. The substrate that is oxidized is regarded as a hydrogen donor. The systematic name is based on donor:acceptor oxidoreductase. The recommended name will be dehydrogenase, wherever this is possible; as an alternative, reductase can be used. Oxidase is only used in cases where O_2 is the acceptor. (Enzyme Nomenclature, 1992, p9)

^φ**Protease inhibitors**: Compounds [or proteins] which inhibit or antagonize biosynthesis or actions of proteases (ENDOPEPTIDASES).

Transporters (also known as '**Membrane Transport Proteins**'): Membrane proteins whose primary function is to facilitate the transport of molecules across a biological membrane. Included in this broad category are proteins involved in active transport (BIOLOGICAL TRANSPORT, ACTIVE), facilitated transport and ION CHANNELS. Any of a group of enzymes which mediate the phenomenon of active transport (McGraw-Hill Dictionary of Scientific and Technical Terms, 5th ed). Although proteins in this class are sometimes referred to an enzymes they have not been accepted as an enzyme class by the Enzyme Commission.

    ^φ**Ion Channels**: Gated, ion-selective glycoproteins that traverse membranes. The stimulus for ION CHANNEL GATING can be due to a variety of stimuli such as LIGANDS, a TRANSMEMBRANE POTENTIAL DIFFERENCE, mechanical deformation or through INTRACELLULAR SIGNALING PEPTIDES AND PROTEINS. [a subclass of '**Transporters**' target family]

Intended target subfamily

^φ**Acetylcholinesterase:** An enzyme that catalyzes the hydrolysis of ACETYLCHOLINE to CHOLINE and acetate. In the CNS, this enzyme plays a role in the function of peripheral neuromuscular junctions. EC 3.1.1.7. [a subset of '**Esterases'** target family]

^φ**Acid phosphatases**: An enzyme that catalyzes the conversion of an orthophosphoric monoester and water to an alcohol and orthophosphate. EC 3.1.3.2. [a subset of '**Phosphatases**' target family]

^φ**Adenylyl Cyclase:** An enzyme of the lyase class that catalyzes the formation of CYCLIC AMP and pyrophosphate from ATP. EC 4.6.1.1. [a subset of '**Lyase**' target family]

^φ**Apolipoproteins:** Protein components on the surface of LIPOPROTEINS. They form a layer surrounding the hydrophobic lipid core. There are several classes of apolipoproteins with each playing a different role in lipid transport and LIPID METABOLISM. These proteins are synthesized mainly in the LIVER and the INTESTINES. [a subset of '**Cell Adhesion Molecules**']

^φ**Arrest** (also known as '**Cell Cycle Checkpoints**'): Cellular mechanisms that ensure that all events occuring during specific times in the CELL CYCLE are completed and without mistakes prior to progression to the next phase of the cell cycle. The halting of progression of a cell through the CELL CYCLE when events in the CELL DIVISION process are not completed accurately or damage to cell components in the process are detected. Regulatory signaling systems that control the progression through the CELL CYCLE. They ensure that the cell has completed, in the correct order and without mistakes, all the processes required to

replicate the GENOME and CYTOPLASM, and divide them equally between two daughter cells. If cells sense they have not completed these processes or that the environment does not have the nutrients and growth hormones in place to proceed, then the cells are restrained (or "arrested") until the processes are completed and growth conditions are suitable. [a subset of '**Cell Cycle**' target family]

φ**Artifact Fluorescence** (also known as '**Artifacts**'): Any visible result of a procedure which is caused by the procedure itself and not by the entity being analyzed. Common examples include histological structures introduced by tissue processing, radiographic images of structures that are not naturally present in living tissue, and products of chemical reactions that occur during analysis. [a subset of '**Background Measurement**' target family]

φ**Aspartate Proteases** (also known as '**Aspartic Acid Proteases**'): A subclass of peptide hydrolases that depend on an ASPARTIC ACID residue for their activity. [a subset of '**Proteases**' target family]

φ**ATPases** (also known as '**Adenosine Triphosphatases**'): A group of enzymes which catalyze the hydrolysis of ATP. The hydrolysis reaction is usually coupled with another function such as transporting Ca(2+) across a membrane. These enzymes may be dependent on Ca(2+), Mg(2+), anions, H+, or DNA. [a subset of '**Hydrolases**' target family]

Baseline Control (also known as '**Background control**'): The role played by the reading obtained from wells to which the sample or one of the reagents was not added. This is subtracted from all the experimental readings prior to further analysis. [a subset of '**Background Measurement**' target family]

φ**Basic Helix-Loop-Helix Protein** (also known as '**Basic Helix-Loop-Helix Transcription Factors**'): A family of DNA-binding transcription factors that contain a basic HELIX-LOOP-HELIX MOTIF. [a subset of '**DNA Binding**' target family]

φ**Basic Helix-Loop-Helix Leucine Zipper Protein** (also known as '**Basic Helix-Loop-Helix Leucine Zipper Transcription Factors**'): A family of transcription factors that contain regions rich in basic residues, LEUCINE ZIPPER domains, and HELIX-LOOP-HELIX MOTIFS. [a subset of '**DNA Binding**' target family]

§**Basic Helix-Turn-Helix Protein:** Transcription factors that possess the helix-turn-helix motif. [a subset of '**DNA Binding**' target family]

φ**Basic Leucine Zipper** (also known as '**Basic-Leucine Zipper Transcription Factors**'): A large superfamily of transcription factors that contain a region rich in BASIC AMINO ACID residues followed by a LEUCINE ZIPPER domain. [a subset of '**DNA Binding**' target family]

φ**Calcium Channel**: Voltage-dependent cell membrane glycoproteins selectively permeable to calcium ions. They are categorized as L-, T-, N-, P-, Q-, and R-types based on the activation and inactivation kinetics, ion specificity, and sensitivity to drugs and toxins. The L- and T-types are present throughout the cardiovascular and central nervous systems and the N-, P-, Q-, & R-types are located in neuronal tissue. [a subset of '**Ion Channels**' target family]

φ**Cell Conformation** (also known as '**Cell Shape**'): The quality of surface form or outline of cells. [a subset of '**Cell Morphology**' target family]

φ**Chemotactic Factor** (also known as '**Chemokines**'): Class of pro-inflammatory cytokines that have the ability to attract and activate leukocytes. They can be divided into at least three structural branches: C; (CHEMOKINES, C); CC; (CHEMOKINES, CC); and CXC; (CHEMOKINES, CXC); according to variations in a shared cysteine motif. [a subset of '**Cytokines**' target family]

^φ**Cholesterol transporters** (also known as '**Fatty Acid Transport Proteins**'): A broad category of membrane transport proteins that specifically transport FREE FATTY ACIDS across cellular membranes. They play an important role in LIPID METABOLISM in CELLS that utilize free fatty acids as an energy source. [a subset of '**Transporters**' target family]

^φ**Coagulation Factor** (also known as '**Blood Coagulation Factor**'): Endogenous substances, usually proteins, that are involved in the blood coagulation process. [a subset of '**Cytokines**' target family]

> ^φ**Plasmogen Activator**: A heterogeneous group of proteolytic enzymes that convert PLASMINOGEN to FIBRINOLYSIN. They are concentrated in the lysosomes of most cells and in the vascular endothelium, particularly in the vessels of the microcirculation. [a subset of '**Cytokines**' target family and '**Coagulation Factor**' subfamily]

> ^φ**Plasmogen Activator Inactivator**: Important modulators of the activity of plasminogen activators. The inhibitors belong to the serpin family of proteins and inhibit both the tissue-type and urokinase-type plasminogen activators. [a subset of '**Cytokines**' target family and '**Coagulation Factor**' subfamily]

^φ**Collagen:** A polypeptide substance comprising about one third of the total protein in mammalian organisms. It is the main constituent of SKIN; CONNECTIVE TISSUE; and the organic substance of bones (BONE AND BONES) and teeth (TOOTH). A topical hemostatic agent prepared from collagen. [a subset of '**Cell Adhesion Molecules**' target family]

^φ**Colony-Stimulating Factors:** Glycoproteins found in a subfraction of normal mammalian plasma and urine. They stimulate the proliferation of bone marrow cells in agar cultures and the formation of colonies of granulocytes and/or macrophages. The factors include INTERLEUKIN-3; (IL-3); GRANULOCYTE COLONY-STIMULATING FACTOR; (G-CSF); MACROPHAGE COLONY-STIMULATING FACTOR; (M-CSF); and GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR; (GM-CSF). [a subset of '**Cytokine**' target family]

^φ**Cyclooxygenase** (also known as '**Prostaglandin-Endoperoxide Synthase**'): Enzyme complexes that catalyze the formation of PROSTAGLANDINS from the appropriate unsaturated FATTY ACIDS, molecular OXYGEN, and a reduced acceptor. [a subset of '**Oxidoreductases**' target family]

^φ**Cytotoxicity** (also known as **Cell Death**'): The termination of the cell's ability to carry out vital functions such as metabolism, growth, reproduction, responsiveness, and adaptability. [a subset of '**Cell Cycle**' target family]

> **Apoptosis**: Apoptosis that is triggered via cell stress and mitochondrial damage. One of the mechanisms by which CELL DEATH occurs (compare with NECROSIS and AUTOPHAGOCYTOSIS). Apoptosis is the mechanism responsible for the physiological deletion of cells and appears to be intrinsically programmed. It is characterized by distinctive morphologic changes in the nucleus and cytoplasm, chromatin cleavage at regularly spaced sites, and the endonucleolytic cleavage of genomic DNA; (DNA FRAGMENTATION); at internucleosomal sites. This mode of cell death serves as a balance to mitosis in regulating the size of animal tissues and in mediating pathologic processes associated with tumor growth. Apoptosis that is triggered via CELL SURFACE RECEPTORS such as TUMOR NECROSIS FACTOR RECEPTORS and DEATH DOMAIN RECEPTORS. [a subset of '**cell cycle**' target family and the '**Cytotoxicity**' subfamily]

^φ**Cysteine Proteases**: A subclass of peptide hydrolases that depend on a CYSTEINE residue for their activity. [a subset of '**Proteases**' target family]

^φ**DNA Conformation** (also known as '**Nucleic Acid Conformation**'): The spatial arrangement of the atoms of a nucleic acid or polynucleotide that results in its characteristic 3-dimensional shape. [a subset of '**cell morphology**' target family]

§**Dual-Specific kinase**: A class of enzymes that can act as a tyrosine and serine/threonine kinases [a subset of '**Kinase**' target family]

^{φ}**Dual-Specific Phosphatases** (also known as '**Dual-Specificity Phosphatases**'): A sub-class of protein tyrosine phosphatases that contain an additional phosphatase activity which cleaves phosphate ester bonds on SERINE or THREONINE residues that are located on the same protein. [a subset of '**Phosphatases**' target family]

^{φ}**GATA Proteins** (also known as '**GATA Transcription Factors**'): A family of transcription factors that contain two ZINC FINGER MOTIFS and bind to the DNA sequence (A/T)GATA(A/G). [a subset of '**DNA Binding**' target family]

^{φ}**Guanylate Cyclase**: An enzyme that catalyzes the conversion of GTP to 3',5'-cyclic GMP and pyrophosphate. It also acts on ITP and dGTP. (From Enzyme Nomenclature, 1992) EC 4.6.1.2. [a subset of '**Lyase**' target family]

^{φ}**Heat Shock Protein:** Proteins which are synthesized in eukaryotic organisms and bacteria in response to hyperthermia and other environmental stresses. They increase thermal tolerance and perform functions essential to cell survival under these conditions. [a subset of '**DNA Binding**' target family]

^{φ}**Histone Deacetylase**: Deacetylases that remove N-acetyl groups from amino side chains of the amino acids of HISTONES. The enzyme family can be divided into at least three structurally-defined subclasses. Class I and class II deacetylases utilize a zinc-dependent mechanism. The sirtuin histone deacetylases belong to class III and are NAD-dependent enzymes. [a subset of '**Hydrolase**' target family]

^{φ}**HMG Box Protein** (also known as '**High Mobility Group Proteins**'): A family of low-molecular weight, non-histone proteins found in chromatin. [a subset of '**DNA Binding**' target family]

^{φ}**Homeobox Protein** (also known as '**Homeodomain Proteins**'): Proteins encoded by homeobox genes (GENES, HOMEOBOX) that exhibit structural similarity to certain prokaryotic and eukaryotic DNA-binding proteins. Homeodomain proteins are involved in the control of gene expression during morphogenesis and development (GENE EXPRESSION REGULATION, DEVELOPMENTAL). [a subset of '**DNA Binding**' target family]

^{φ}**Paired Box Protein:** A family of transcription factors that control EMBRYONIC DEVELOPMENT within a variety of cell lineages. They are characterized by a highly conserved paired DNA-binding domain that was first identified in DROSOPHILA segmentation genes. [a subset of '**DNA Binding**' target family and '**Homeobox Protein**' subfamily]

^{φ}**POU domain protein** (also known as '**POU Domain Factors**'): A family of transcription factors characterized by the presence of a bipartite DNA-binding domain known as the POU domain. The POU domain contains two subdomains, a POU-specific domain and a POU-homeodomain. The POU domain was originally identified as a region of approximately 150 amino acids shared between the Pit-1, Oct-1, Oct-2, and Unc-86 transcription factors. [a subset of '**DNA Binding**' target family and '**Homeobox Protein**' subfamily]

§**Inflammatory Factor**: Cytokines that participate in inflammation [a subset of '**Cytokine**' target family]

^{φ}**Imidazoline receptor:** Receptors of CLONIDINE and other IMIDAZOLINES. Activity of the ligands was earlier attributed to ADRENERGIC ALPHA-2 RECEPTORS. Endogenous ligands include AGMATINE, imidazoleacetic acid ribotide, and harman. [a subset of '**Oxidoreductases**' target family]

§**Immunoglobulin CAM**: Cell adhesion molecules that belong to the Immunoglobulin superfamily. [a subset of '**Cell Adhesion Molecules**' target family]

^φ**Insulin**: Peptide hormones that cause an increase in the absorption of GLUCOSE by cells within organs such as LIVER, MUSCLE and ADIPOSE TISSUE. During normal metabolism insulins are produced by the PANCREATIC BETA CELLS in response to increased GLUCOSE. Natural and chemically-modified forms of insulin are also used in the treatment of GLUCOSE METABOLISM DISORDERS such as DIABETES MELLITUS. [a subset of '**Non-Steroidal Hormones'** target family]

^φ**Interfuron Regulatory Factors**: A family of transcription factors that share an N-terminal HELIX-TURN-HELIX MOTIF and bind INTERFERON-inducible promoters to control GENE expression. IRF proteins bind specific DNA sequences such as interferon-stimulated response elements, interferon regulatory elements, and the interferon consensus sequence. [a subset of '**DNA Binding**' target family]

^φ**Interleukins**: Soluble factors which stimulate growth-related activities of leukocytes as well as other cell types. They enhance cell proliferation and differentiation, DNA synthesis, secretion of other biologically active molecules and responses to immune and inflammatory stimuli. [a subset of '**Cytokine**' target family]

^φ**Internal Marker** (also known as '**Genes, Essential**'): Those genes found in an organism which are necessary for its viability and normal function. Constitutively and evenly expressed genes involved in routine cellular metabolisms. [a subset of '**Background Measurement**' target family]

^φ**Ligand-Gated Ion Channel**: A subclass of ion channels that open or close in response to the binding of specific LIGANDS. [a subset of '**Ion Channels**' target family]

^φ**Matrix Metalloproteinase**: A family of zinc-dependent metalloendopeptidases that is involved in the degradation of EXTRACELLULAR MATRIX components. [a subset of '**Proteases**' target family]

^**Metabotropic Glutamate Receptor** (also known as '**G-Protein Coupled Receptor Class C**'): The class C G protein-coupled receptors are a class of G protein-coupled receptors that include the metabotropic glutamate receptors and several additional receptors. [a subset of '**GPCR'** target family]

^φ**Metalloproteinase inhibitor** (also known as '**Matrix Metalloproteinase Inhibitors**'): Compounds [or proteins] that inhibit the enzyme activity or activation of MATRIX METALLOPROTEINASES. [a subset of '**Protease inhibitors**' target family]

^φ**Methylenetetrahydrofolate reductase** (NADPH2): A flavoprotein amine oxidoreductase that catalyzes the reversible conversion of 5-methyltetrahydrofolate to 5,10-methylenetetrahydrofolate. This enzyme was formerly classified as EC 1.1.1.171. [a subset of '**Oxidoreductases**' target family]

^φ**MHC Class II** (also known as '**Histocompatibility Antigens Class II**'): Large, transmembrane, non-covalently linked glycoproteins (alpha and beta). Both chains can be polymorphic although there is more structural variation in the beta chains. The class II antigens in humans are called HLA-D ANTIGENS and are coded by a gene on chromosome 6. In mice, two genes named IA and IE on chromosome 17 code for the H-2 antigens. The antigens are found on B-lymphocytes, macrophages, epidermal cells, and sperm and are thought to mediate the competence of and cellular cooperation in the immune response. The term IA antigens used to refer only to the proteins encoded by the IA genes in the mouse, but is now used as a generic term for any class II histocompatibility antigen. [a subset of '**Cell Adhesion Molecules**' target family]

^φ**Monoamine Oxidase**: An enzyme that catalyzes the oxidative deamination of naturally occurring monoamines. It is a flavin-containing enzyme that is localized in mitochondrial membranes, whether in nerve terminals, the liver, or other organs. Monoamine oxidase is important in regulating the metabolic degradation of catecholamines and serotonin in neural or target tissues. Hepatic monoamine oxidase has a crucial defensive role in inactivating circulating monoamines or those, such as tyramine, that originate in the gut and are absorbed into the portal circulation. (From Goodman and Gilman's, The Pharmacological Basis of Therapeutics, 8th ed, p415) EC 1.4.3.4. [a subset of '**Oxidoreductases**' target family]

φ**MYB proteins** (also known as '**Proto-Oncogene Proteins c-myb**'): Cellular DNA-binding proteins encoded by the myb gene (GENES, MYB). They are expressed in a wide variety of cells including thymocytes and lymphocytes, and regulate cell differentiation. Overexpression of myb is associated with autoimmune diseases and malignancies. [a subset of '**DNA Binding**' target family]

φ**Neurotransmitter transporters** (also known as '**Neurotransmitter Transport Proteins**'): Membrane transport proteins found predominately in NEURONS and neuroendocrine cells that facilitate neurotransmitter transport. They include two distinct families of proteins that transport NEUROTRANSMITTERS across the PLASMA MEMBRANE and that transport NEUROTRANSMITTERS into SECRETORY VESICLES. [a subset of '**Transporters**' target family]

φ**NF-Kappa B**: Ubiquitous, inducible, nuclear transcriptional activator that binds to enhancer elements in many different cell types and is activated by pathogenic stimuli. The NF-kappa B complex is a heterodimer composed of two DNA-binding subunits: NF-kappa B1 and relA. [a subset of '**DNA Binding**' target family]

φ**Nitric Oxide Synthase**: An NADPH-dependent enzyme that catalyzes the conversion of L-ARGININE and OXYGEN to produce CITRULLINE and NITRIC OXIDE. [a subset of '**Oxidoreductases**' target family]

φ**Non-Receptor Tyrosine Kinase** (also known as '**Protein-Tyrosine Kinase**'): Protein kinases that catalyze the PHOSPHORYLATION of TYROSINE residues in proteins with ATP or other nucleotides as phosphate donors. [a subset of '**Kinase**' target family]

φ**Non-Receptor Tyrosine Phosphatases** (also known as '**Protein Tyrosine Phosphatases, Non-Receptor**'): A subcategory of protein tyrosine phosphatases that occur in the CYTOPLASM. Many of the proteins in this category play a role in intracellular signal transduction. [a subset of '**Phosphatases**' target family]

§**Non-Steroidal [nuclear receptors]**: nuclear receptors that bind non-steroidal hormones as their natural ligands [a subset of '**Nuclear Receptor**' target family]

φ**Nuclear Factor I** (also known as '**NFI Transcription Factors**'): Transcription factors that were originally identified as site-specific DNA-binding proteins essential for DNA REPLICATION by ADENOVIRUSES. They play important roles in MAMMARY GLAND function and development. [a subset of '**DNA Binding**' target family]

φ**Smad Proteins:** A family of proteins that are involved in the translocation of signals from TGF-BETA RECEPTORS; BONE MORPHOGENETIC PROTEIN RECEPTORS; and other surface receptors to the CELL NUCLEUS. They were originally identified as a class of proteins that are related to the mothers against decapentaplegic protein, Drosophila and sma proteins from CAENORHABDITIS ELEGANS. [a subset of '**DNA Binding**' target family and '**Nuclear Factor I**' subfamily]

φ**Nuclear Respiratory Factors**: A family of transcription factors that control expression of a variety of nuclear GENES encoding proteins that function in the RESPIRATORY CHAIN of the MITOCHONDRIA. [a subset of '**DNA Binding**' target family]

φ**Nucleoside Transporters** (also known as '**Nucleoside Transport Proteins**'): Proteins involved in the transport of NUCLEOSIDES across cellular membranes. [a subset of '**Transporters**' target family]

φ**O-methyltransferase** (also known as '**Protein O-Methyltranferase**'): An enzyme that catalyzes the transfer of methyl groups from S-adenosylmethionine to free carboxyl groups of a protein molecule forming methyl esters. EC 2.1.1.-. [a subset of '**Methyltransferase**' target family]

φ**Organelle Conformation** (also known as '**Organelle Shape**'): The quality of surface form or outline of Organelles. [a subset of '**Cell Morphology**' target family]

§**Other Cytokine**: Other known cytokines that do not clearly identify or directly participate in inflammation, coagulation, chemotaxis, or function as interleukins. [a subset of '**Cytokine**' target family]

φ**Phosphodiesterase** (also known as '**Phosphoric Diester Hydrolases**'): A class of enzymes that catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. EC 3.1.4. [a subset of '**Esterases**' target family]

φ**Phosphoinositol Kinase**: Phosphotransferases that catalyzes the conversion of 1-phosphatidylinositol to 1-phosphatidylinositol 3-phosphate. Many members of this enzyme class are involved in RECEPTOR MEDIATED SIGNAL TRANSDUCTION and regulation of vesicular transport with the cell. Phosphatidylinositol 3-Kinases have been classified both according to their substrate specificity and their mode of action within the cell. [a subset of '**Kinase**' target family]

φ**Potassium Channel**: Cell membrane glycoproteins that are selectively permeable to potassium ions. At least eight major groups of K channels exist and they are made up of dozens of different subunits. [a subset of '**Ion Channels**' target family]

φ**Proliferation** (also known as '**Cell Proliferation**'): All of the processes involved in increasing CELL NUMBER including CELL DIVISION. [a subset of '**Cell Cycle**' target family]

φ**Protein Deacetylase** (also known as '**Peptide Hydrolases**'): Hydrolases that specifically cleave the peptide bonds found in PROTEINS and PEPTIDES. Examples of sub-subclasses for this group include EXOPEPTIDASES and ENDOPEPTIDASES. [a subset of '**Hydrolase**' target family]

φ**Pseudocholinesterase:** An enzyme that non-specifically hydrolyses choline esters. [a subset of '**Esterases**' target family]

φ**Receptor Tyrosine Kinase**: A class of cellular receptors that have an intrinsic PROTEIN-TYROSINE KINASE activity. [a subset of '**Kinase**' target family]

φ**Receptor Tyrosine Phosphatases** (also known as '**Receptor-like Protein Tyrosine Phosphatases**'): A subcategory of protein tyrosine phosphatases that are bound to the cell membrane. They contain cytoplasmic tyrosine phosphatase domains and extracellular protein domains that may play a role in cell-cell interactions by interacting with EXTRACELLULAR MATRIX components. They are considered receptor-like proteins in that they appear to lack specific ligands. [a subset of '**Phosphatases**' target family]

^**Rhodopsin-like Receptor** (also known as '**G-Protein Coupled Receptor Class A**'): Rhodopsin-like receptors are a widespread protein family that includes hormones, neurotransmitters, and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins. Although their activating ligands vary widely in structure and character, the amino acid sequences of the receptors are very similar and are believed to adopt a common structural framework comprising 7 transmembrane (TM) helices. [a subset of '**GPCR**' target family]

^**Secretin Receptor** (also known as '**G-Protein Coupled Receptor Class B**'): Class B of the GPCRs, the secretin-receptor family GPCRs, include vasoactive intestinal peptide receptors and receptors for secretin, calcitonin and parathyroid hormone/parathyroid hormone-related peptides. These receptors activate adenylyl cyclase and the phosphatidyl-inositol-calcium pathway. The receptors in this family have 7 transmembrane helices, like rhodopsin-like GPCRs. However, there is no significant sequence identity between these two GPCR families and the secretin-receptor family has its own characteristic 7TM signature. The secretin-receptor family GPCRs exist in many animal species, but have not been found in plants, fungi or prokaryotes. Three distinct sub-families (B1-B3) are recognized. [a subset of '**GPCR**' target family]

^φ**Selectins**: Transmembrane proteins consisting of a lectin-like domain, an epidermal growth factor-like domain, and a variable number of domains that are homologous to complement regulatory proteins. They are important cell adhesion molecules which help LEUKOCYTES attach to VASCULAR ENDOTHELIUM. [a subset of '**Cell Adhesion Molecules**' target family]

^φ**Serine Proteases**: Peptide hydrolases that contain at the active site a SERINE residue involved in catalysis. [a subset of '**Proteases**' target family]

^φ**Serine/threonine Kinase** (also known as '**Protein-Serine Threonine Kinases**'): A group of enzymes that catalyzes the phosphorylation of serine or threonine residues in proteins, with ATP or other nucleotides as phosphate donors. [a subset of '**Kinase**' target family]

^φ**Serine/threonine Phosphatases** (also known as '**Phosphoprotein Phosphatases**'): A group of enzymes removing the SERINE- or THREONINE-bound phosphate groups from a wide range of phosphoproteins, including a number of enzymes which have been phosphorylated under the action of a kinase. (Enzyme Nomenclature, 1992) [a subset of '**Phosphatases**' target family]

^φ**Sodium Channels:** Ion channels that specifically allow the passage of SODIUM ions. A variety of specific sodium channel subtypes are involved in serving specialized functions such as neuronal signaling, CARDIAC MUSCLE contraction, and KIDNEY function. [a subset of '**Ion Channels**' target family]

^φ**Steroidogenesis-related** (also known as '**Steroid Hydroxylases**'): Cytochrome P-450 monooxygenases (MIXED FUNCTION OXYGENASES) that are important in steroid biosynthesis and metabolism. [a subset of '**CYP**' target family]

[§]**Steroidal [nuclear receptor]**: nuclear receptors that bind steroid hormones as their natural ligands [a subset of '**Nuclear Receptor**' target family]

^φ**Stat protein** (also known as '**STAT Transcription Factors**'): A family of transcription factors containing SH2 DOMAINS that are involved in CYTOKINE-mediated SIGNAL TRANSDUCTION. STAT transcription factors are recruited to the cytoplasmic region of CELL SURFACE RECEPTORS and are activated via PHOSPHORYLATION. Once activated they dimerize and translocate into the CELL NUCLEUS where they influence GENE expression. They play a role in regulating CELL GROWTH PROCESSES and CELL DIFFERENTIATION. STAT transcription factors are inhibited by SUPPRESSOR OF CYTOKINE SIGNALING PROTEINS and PROTEIN INHIBITORS OF ACTIVATED STAT. [a subset of '**DNA Binding**' target family]

^φ**Stress Response** (also known as '**Stress, Physiological**'): The unfavorable effect of environmental factors (stressors) on the physiological functions of an organism. Prolonged unresolved physiological stress can affect HOMEOSTASIS of the organism, and may lead to damaging or pathological conditions. [a subset of '**Cell Cycle**' target family]

^φ**Transforming Growth Factor Beta**: A factor synthesized in a wide variety of tissues. It acts synergistically with TGF-alpha in inducing phenotypic transformation and can also act as a negative autocrine growth factor. TGF-beta has a potential role in embryonal development, cellular differentiation, hormone secretion, and immune function. TGF-beta is found mostly as homodimer forms of separate gene products TGF-beta1, TGF-beta2 or TGF-beta3. Heterodimers composed of TGF-beta1 and 2 (TGF-beta1.2) or of TGF-beta2 and 3 (TGF-beta2.3) have been isolated. The TGF-beta proteins are synthesized as precursor proteins. [a subset of '**Growth Factor**' target family]

^φ**Tumor Suppressor** (also known as '**Tumor Suppressor Proteins**'): Proteins that are normally involved in holding cellular growth in check. Deficiencies or abnormalities in these proteins may lead to unregulated cell growth and tumor development. [a subset of '**DNA Binding**' target family]

Vesicular Transporter (also known as '**Vesicular Transport Proteins**'): A broad category of proteins involved in the formation, transport and dissolution of TRANSPORT VESICLES. They play a role in the intracellular transport of molecules contained within membrane vesicles. Vesicular transport proteins are distinguished from MEMBRANE TRANSPORT PROTEINS, which move molecules across membranes, by the mode in which the molecules are transported. [a subset of '**Transporter**' target family]

φ**Winged Helix-Turn-Helix** (also known as '**Winged-Helix Transcription Factors**'): A subfamily of HELIX-TURN-HELIX DNA-binding proteins that contain a variable length loop adjacent to the HTH motif. The loop connects two anti-parallel strands and forms a wing when bound to DNA. [a subset of '**DNA Binding**' target family]

φ**E2F Transcription Factor**: A family of basic helix-loop-helix transcription factors that control expression of a variety of GENES involved in CELL CYCLE regulation. E2F transcription factors typically form heterodimeric complexes with TRANSCRIPTION FACTOR DP1 or transcription factor DP2, and they have N-terminal DNA binding and dimerization domains. E2F transcription factors can act as mediators of transcriptional repression or transcriptional activation. [a subset of '**DNA Binding**' target family and '**Winged Helix-Turn-Helix**' subfamily]

φ**Forkhead box protein** (also known as '**Forkhead Transcription Factor**'): A subclass of winged helix DNA-binding proteins that share homology with their founding member fork head protein, Drosophila. [a subset of '**DNA Binding**' target family and '**Winged Helix-Turn-Helix**' subfamily]

φ**Xenobiotic Metabolism** (also known as '**Aryl Hydrocarbon Hydroxylases**'): A large group of cytochrome P-450 (heme-thiolate) monooxygenases that complex with NAD(P)H-FLAVIN OXIDOREDUCTASE in numerous mixed-function oxidations of aromatic compounds. They catalyze hydroxylation of a broad spectrum of substrates and are important in the metabolism of steroids, drugs, and toxins such as PHENOBARBITAL, carcinogens, and insecticides. Monooxygenases that oxidize exogenous substrates. [a subset of '**CYP**' target family]

§**Zinc Finger**: Transcription factors that possess the zinc finger motif. In the classic zinc finger, one zinc atom is bound to two cysteines and two histidines. In between the cysteines and histidines are 12 residues which form a DNA binding fingertip. [a subset of '**DNA Binding**' target family]

Normalized Data Type: the fold induction or percent activity scale in which the assay data is displayed

§**Percent Activity**: The raw readout data was normalized as a percent change from the negative control with regards to the positive control as the 100% maximal effect.

§**Fold Induction**: The raw readout data was normalized to fold inductions with regards to the negative control.

§**Log$_2$ Fold Induction**: The raw readout data was normalized to find the two-fold inductions with regards to the negative control.

§**Log$_{10}$ Fold Induction**: The raw readout data was normalized to find the ten-fold inductions with regards to the negative control.

Reagent Name Value Types: the type of substance or function served by the reagent

Antibody: A protein (immunoglobulin) produced by the immune system of an organism in response to exposure to a foreign molecule (antigen) and characterized by its specific binding to a site of that molecule (antigenic determinant or epitope).

Coupled substrate: An enzyme substrate whose concentration is regulated by the activity of a different enzyme. e.g., the concentration of ATP is dependent on a kinase, with ATP being its substrate, which in turn determines the rate of a coupled enzyme luciferase, since ATP is one of its substrates too.

Inducer: A molecular entity that is required to initiate a biological process. e.g., interferon-gamma is an inducer of the transcription factor STAT-1, which brings about the activation of STAT-1 upon addition to cells.

Ligand: A soluble molecule, such as a hormone or neurotransmitter that binds specifically and reversibly to another chemical entity, such as a receptor.

Stain (also known as dye): Molecules used as optical probes, which have fluorescence/colorimetric properties.

Substrate: The substance on which the enzyme acts to generate a product, e.g. ATP is the substrate for ATPase, which hydrolyzes it to ADP and Pi.

Signal Direction Type: the expected direction of the detected signal in relation to the negative control

Both: Increase in perturbation could increase or decrease the signal.

Gain: Increase in perturbation is correlated to increase in signal e.g. activation.

Loss: Increase in perturbation is correlated to decrease in signal e.g. inhibition.

None: Increase in perturbation is not expected to change the signal e.g. background control.

Signal Direction: the direction observed of the detected signal

§**Gain:** Increase in perturbation of this endpoint results in increases in signal e.g. activation.

§**Loss**: Increase in perturbation of this endpoint results in decreases in signal e.g. inhibition.

Target Types: the abstract chemical or biological target groupings

§**Cellular**: The target of focus is an organelle, a cellular response, or some aspect within cell biology.

§**Chemical**: The target of focus is an endogenous or exogenous chemical entity.

§**DNA**: The target of focus is the physical feature, structural conformation, or base changes to deoxyribonucleic nucleic acid (DNA).

§**Pathway**: The target of focus is the set of enzymatic reactions contributing to a measureable biological events.

§**Protein**: The target of focus is a protein, interaction between proteins, or a specific function of proteins.

§**RNA**: The target of focus is the physical feature, structural conformation, or expression changes to a form of ribonucleic acid (RNA).

Target Subtypes: the subtypes to the abstract chemical entity, gene form, or biological target groupings

§**ATP**: The focus is the free adenosine triphosphate (ATP) levels. [a subset of '**Chemical**' target type]

§**Cellular**: The focus is the cell morphology or cellular response. [a subset of '**Cellular**' target type]

§**DNA-unspecified**: The focus is the change to the overall structural feature of the deoxyribonucleic acid (DNA) within each cell, without regard to specific gene regions. [a subset of '**DNA**' target type]

§**Enzyme**: The focus is the enzymatic functionality of the target protein. [a subset of '**Protein**' target type]

§**Lysosome**: The focus is the structural or functional abundance of the organelle, lysosome. [a subset of '**Cellular**' target type]

§**Mitochondria**: The focus is the structural or functional abundance of the organelle, mitochondria. [a subset of '**Cellular**' target type]

§**mRNA**: The focus is the relative abundance of the ribonucleic acid (DNA), with regards to specific gene targets. [a subset of '**DNA**' target type]

§**Nucleus**: The focus is the structure, size, and shape of the nucleus. [a subset of '**Cellular**' target type]

§**Pathway-specified**: The focus is the gene functions or cellular responses indicative of activation of a pathway of interest. [a subset of '**Pathway**' target type]

§**Protein-specified**: The focus is the protein expression levels of a specified gene target. [a subset of '**Protein**' target type]

§**Protein-unspecified**: The focus is the level of free or total proteins without regard to the specificity to certain genes. [a subset of '**Protein**' target type]

§**Physical feature**: The focus is the inherent physicochemical properties of the test substance, typically a chemical. [a subset of '**Chemical**' target type]

§**Receptor**: The focus is the receptor-ligand function of the target gene protein. [a subset of '**Protein**' target type]

§**Transcription factor**: The characterized genes are measured as the downstream gene products for understanding the changes in activity from the upstream transcription factor engine. [a subset of '**Protein**' target type]

§**Transporter**: The focus is the transporter functionality of the target gene protein. [a subset of '**Protein**' target type]

Appendix

Appendix1: find assay endpoints by some term
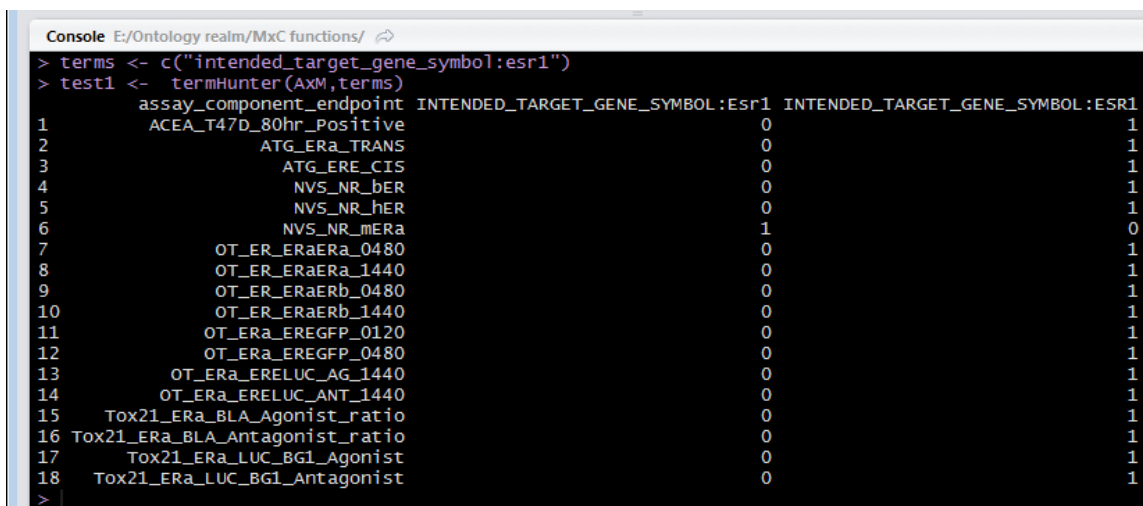
For a quick search by any annotation term, below is an R script intended to be run with the file "ToxCast_Annotation_Binary.csv". The script encodes for a function called TermHunt. It requires two parameters: 'x', which should be a data frame, and 'y', which is some string to be looked up. This function requires that the 'reshape2' library be initiated.

The file "ToxCast_Annotation_Binary.csv" needs to be read-in as a data frame to be used as the parameter 'x' within the TermHunt function. The parameter 'y' can be any string imaginable, and it would be looked up to find corresponding fragments from the variable names within 'x'. The variable names from "ToxCast_Annotation_Binary.csv" should be the annotation names (e.g. detection_technology_type) and the annotation terms (e.g. fluorescence) separated by a semicolon. The function will narrow down *assay endpoints* to those that have a hit value of '1'. The more specific 'y' is, the more precise the output. For example, 'gene' is less precise compared with 'intended_target_gene', which is also less precise compared with 'intended_target_gene_symbol:esr1'.

```
# TermHunt <- function (x,y) {
#  x <- melt(x, id="assay_component_endpoint", na.rm=TRUE)
#  z <- x[grep(  gsub(", ", "|", toupper(toString(y))),  toupper(x[,"variable"])),]
#  z <- z[z["value"]==1,]
#  z <- dcast(z, assay_component_endpoint~variable, function(x) 1, fill=0)
#  z[with(z, order(z[,c(2)],decreasing=FALSE)),]
#  print(z)
# }

# Example1
Annotationbinary <- read.csv("ToxCast_Annotation_Binary.csv")
term <- c("intended_target_gene_symbol:esr1")
Test1 <-  TermHunt ( Annotationbinary, term)
```

The function can also be used with multiple terms. NOTE: as long as one of those terms finds an annotation term, and as long as the *assay endpoint* corresponds to one of those terms, they will be part of the output.

```
# Example2
term <- c('cell death', 'viability reporter')
Test2 <-  TermHunt ( Annotationbinary, term)
names(Test2)
```

```
Console E:/Ontology realm/MxC functions/
[3]  BIOLOGICAL_PROCESS_TARGET:cell death
> Test2
                      assay_component_endpoint ASSAY_DESIGN_TYPE:viability reporter BIOLOGICAL_PROCESS_TARGET:cell death
1                            APR_CellLoss_1h_dn                                   1                                   1
2                            APR_CellLoss_1h_up                                   1                                   1
3                           APR_CellLoss_24h_dn                                   1                                   1
4                           APR_CellLoss_24h_up                                   1                                   1
5                           APR_CellLoss_72h_dn                                   1                                   1
6                           APR_CellLoss_72h_up                                   1                                   1
7                      APR_Hepat_CellLoss_1hr_dn                                  1                                   0
8                      APR_Hepat_CellLoss_1hr_up                                  1                                   0
9                    APR_Hepat_CellLoss_24hr_dn                                   1                                   0
10                   APR_Hepat_CellLoss_24hr_up                                   1                                   0
11                   APR_Hepat_CellLoss_48hr_dn                                   1                                   0
12                   APR_Hepat_CellLoss_48hr_up                                   1                                   0
13                   APR_Hepat_DNADamage_1hr_dn                                   1                                   0
14                   APR_Hepat_DNADamage_1hr_up                                   1                                   0
15                  APR_Hepat_DNADamage_24hr_dn                                   1                                   0
16                  APR_Hepat_DNADamage_24hr_up                                   1                                   0
17                  APR_Hepat_DNADamage_48hr_dn                                   1                                   0
18                  APR_Hepat_DNADamage_48hr_up                                   1                                   0
19                      APR_OxidativeStress_1h_dn                                 1                                   0
20                      APR_OxidativeStress_1h_up                                 1                                   0
21                     APR_OxidativeStress_24h_dn                                 1                                   0
22                     APR_OxidativeStress_24h_up                                 1                                   0
23                     APR_OxidativeStress_72h_dn                                 1                                   0
24                     APR_OxidativeStress_72h_up                                 1                                   0
25                             APR_p53Act_1h_dn                                   1                                   1
26                             APR_p53Act_1h_up                                   1                                   1
27                            APR_p53Act_24h_dn                                   1                                   1
28                            APR_p53Act_24h_up                                   1                                   1
29                            APR_p53Act_72h_dn                                   1                                   1
30                            APR_p53Act_72h_up                                   1                                   1
31                           ATG_XTT_Cytotoxicity                                 1                                   0
32                     BSK_3C_Proliferation_down                                  1                                   0
33                       BSK_3C_Proliferation_up                                  1                                   0
34                               BSK_3C_SRB_down                                  1                                   1
35                                 BSK_3C_SRB_up                                  1                                   1
36                               BSK_4H_SRB_down                                  1                                   1
37                                 BSK_4H_SRB_up                                  1                                   1
38                             BSK_BE3C_SRB_down                                  1                                   1
39                               BSK_BE3C_SRB_up                                  1                                   1
40                 BSK_CASM3C_Proliferation_down                                  1                                   0
41                   BSK_CASM3C_Proliferation_up                                  1                                   0
42                           BSK_CASM3C_SRB_down                                  1                                   1
43                             BSK_CASM3C_SRB_up                                  1                                   1
44                 BSK_hDFCGF_Proliferation_down                                  1                                   0
45                   BSK_hDFCGF_Proliferation_up                                  1                                   0
46                           BSK_hDFCGF_SRB_down                                  1                                   1
47                             BSK_hDFCGF_SRB_up                                  1                                   1
48                            BSK_KF3CT_SRB_down                                  1                                   1
49                              BSK_KF3CT_SRB_up                                  1                                   1
50                             BSK_LPS_SRB_down                                   1                                   1
51                               BSK_LPS_SRB_up                                   1                                   1
52               BSK_SAg_PBMCCytotoxicity_down                                    0                                   1
53                 BSK_SAg_PBMCCytotoxicity_up                                    0                                   1
54                     BSK_SAg_Proliferation_down                                 1                                   0
55                       BSK_SAg_Proliferation_up                                 1                                   0
56                              BSK_SAg_SRB_down                                  1                                   1
57                                BSK_SAg_SRB_up                                  1                                   1
58                            Tox21_AhR_viability                                 1                                   0
59               Tox21_AR_BLA_Antagonist_viability                               1                                   0
60         Tox21_AR_LUC_MDAKB2_Antagonist_viability                             1                                   0
61                  Tox21_ARE_BLA_agonist_viability                              1                                   0
62           Tox21_Aromatase_Inhibition_viability                                1                                   0
63                 Tox21_ELG1_LUC_Agonist_viability                              1                                   0
64               Tox21_ERa_BLA_Antagonist_viability                              1                                   0
65          Tox21_ERa_LUC_BG1_Antagonist_viability                              1                                   0
66                           Tox21_ESRE_BLA_viability                            1                                   0
67                Tox21_FXR_BLA_agonist_viability                                1                                   0
68               Tox21_FXR_BLA_antagonist_viability                             1                                   0
69               Tox21_GR_BLA_Antagonist_viability                               1                                   0
```