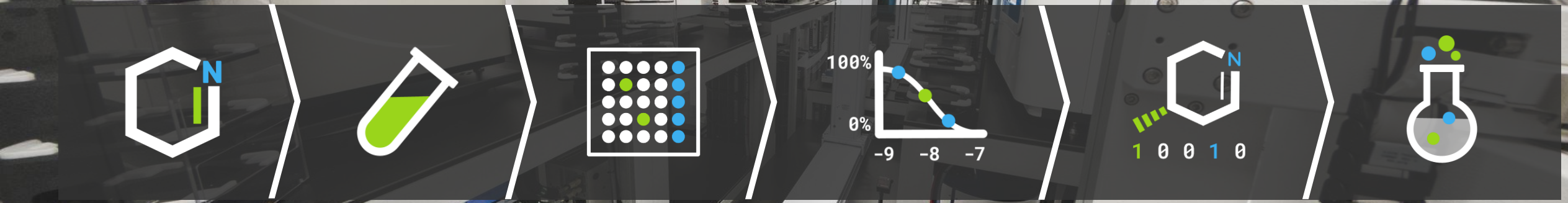
DATA PLATFORMS FOR CHEMICAL BIOLOGY

DEVELOPED AT CZ-OPENSCREEN

Ctibor Škuta



CZ OPENSCREEN AT IMG

1st Czech Cheminformatics Meeting 21/02/2023

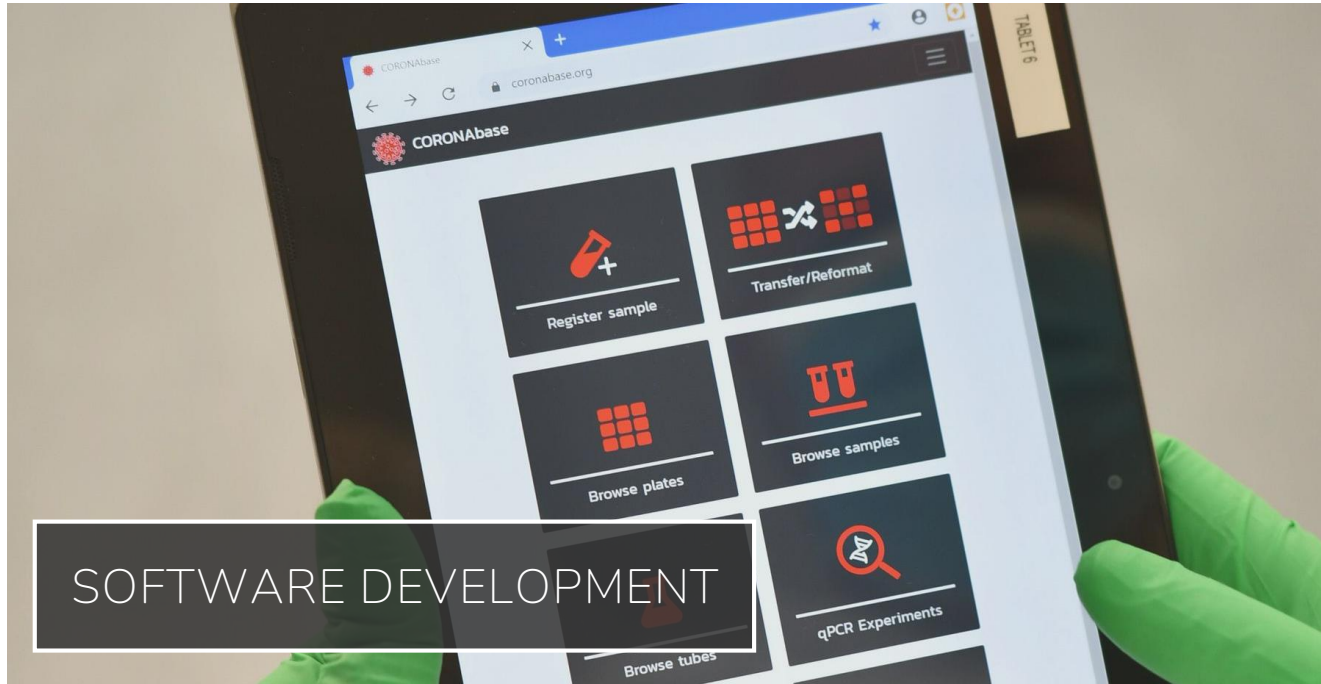
CZ-OPENSCREEN

NATIONAL INFRASTRUCTURE FOR CHEMICAL BIOLOGY

openscreen.cz

HTS/HCS SCREENING

COMPOUND MANAGEMENT



SOFTWARE DEVELOPMENT

SOFTWARE DEVELOPMENT AT CZ-OPENSCREEN

A hub for the integration of high-quality bioactive compound sets enabling their analysis and comparison

probes-drugs.org

An intuitive, cross-platform animal tracking database with advanced visualization, reporting and management features

zebrabase.org

European Chemical Biology Database (ECBD) is a central data hub for data generated within the EU-OPENSCREEN network

ecbd.eu

A comprehensive, cross-platform database for PCR diagnostics, which has been developed and employed at the IMG during the pandemic

coronabase.org

A laboratory information management system (LIMS) - from sample acquisition to results reporting

not public yet

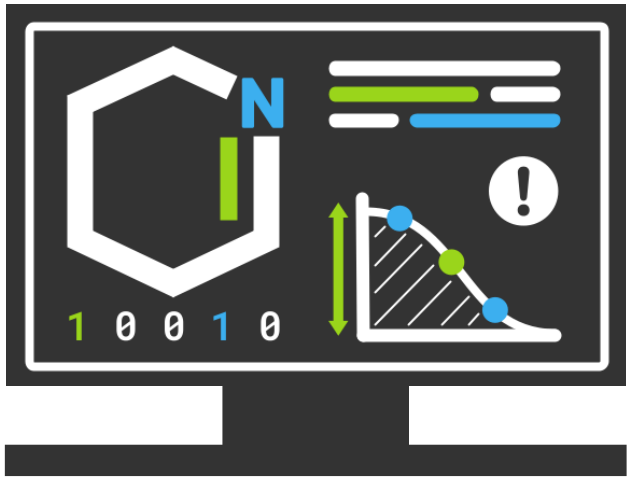
A cross-platform inventory application for biological samples management

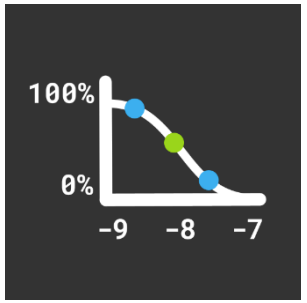
not public yet

Other independent libraries: InCHlib, ChemSpace.js, RackScanner, printwes...

DATA PLATFORMS IN CHEMICAL BIOLOGY

- Compound-centric systems with „chemical intelligence“
- Based on their purpose/users contain various sets of functions/data
- Public data integration platform (P&D) X Data hub for a large research consortium (ECBD) X Laboratory information management system (ScreenX)
- Common functionality – browse/detail views, filtering system, structure-based queries, data export





OUR TECHNOLOGY STACK

Python/Django, JavaScript/Vue/Quasar, RDkit, PostgreSQL/RDkit cartridge, HTML, CSS, etc.

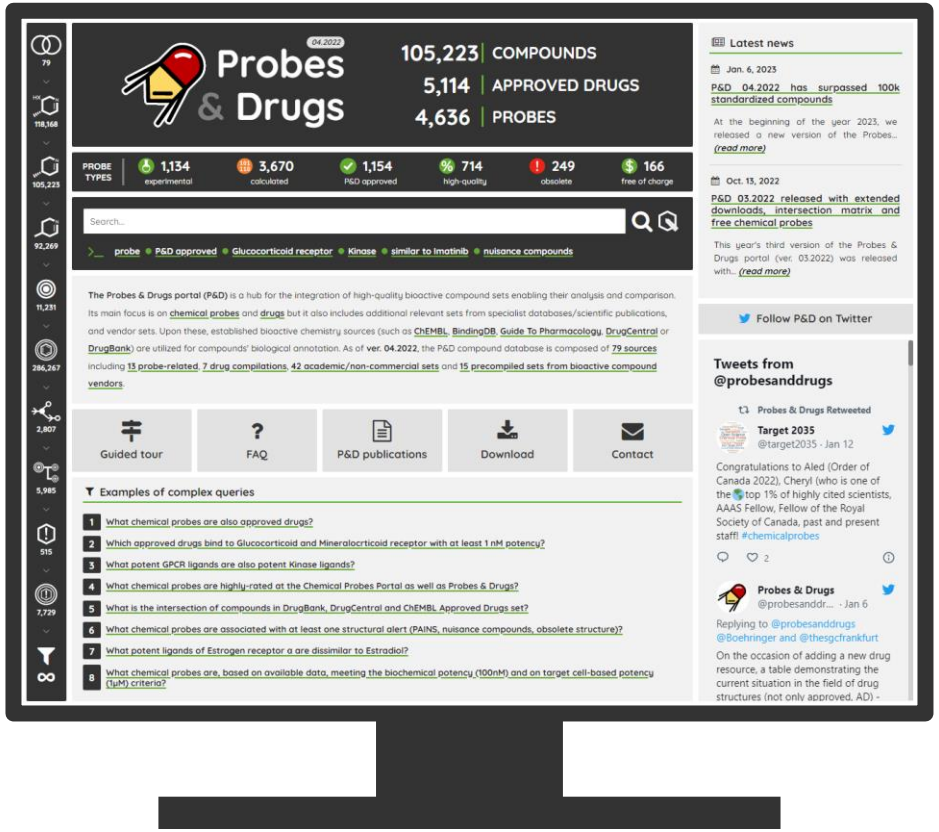
A SHORT HISTORY OF THE PROBES & DRUGS PORTAL

- Data platform for bioactive compounds selection for a wide scientific community
- Started as a CZ-OPENSCREEN in-house project that would simplify our bioactive library updates selection
- High-quality bioactive compound sets scattered across databases, web-pages, publications, vendor sets
- Our goal was to put all of the relevant sets together and provide a user-friendly interface with tools that would allow anyone to work with the set
- P&D released in 2017 with ~30,000 compounds representing 29 sets

Skuta C, Popr M, Muller T, Jindrich J, Kahle M, Sedlak D, Svozil D, Bartunek P. (2017) Probes & Drugs Portal: an interactive, open data resource for chemical biology. Nature Methods 14(8), pp.759-760.

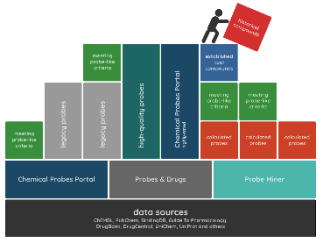
PROBES & DRUGS PORTAL IN 2023

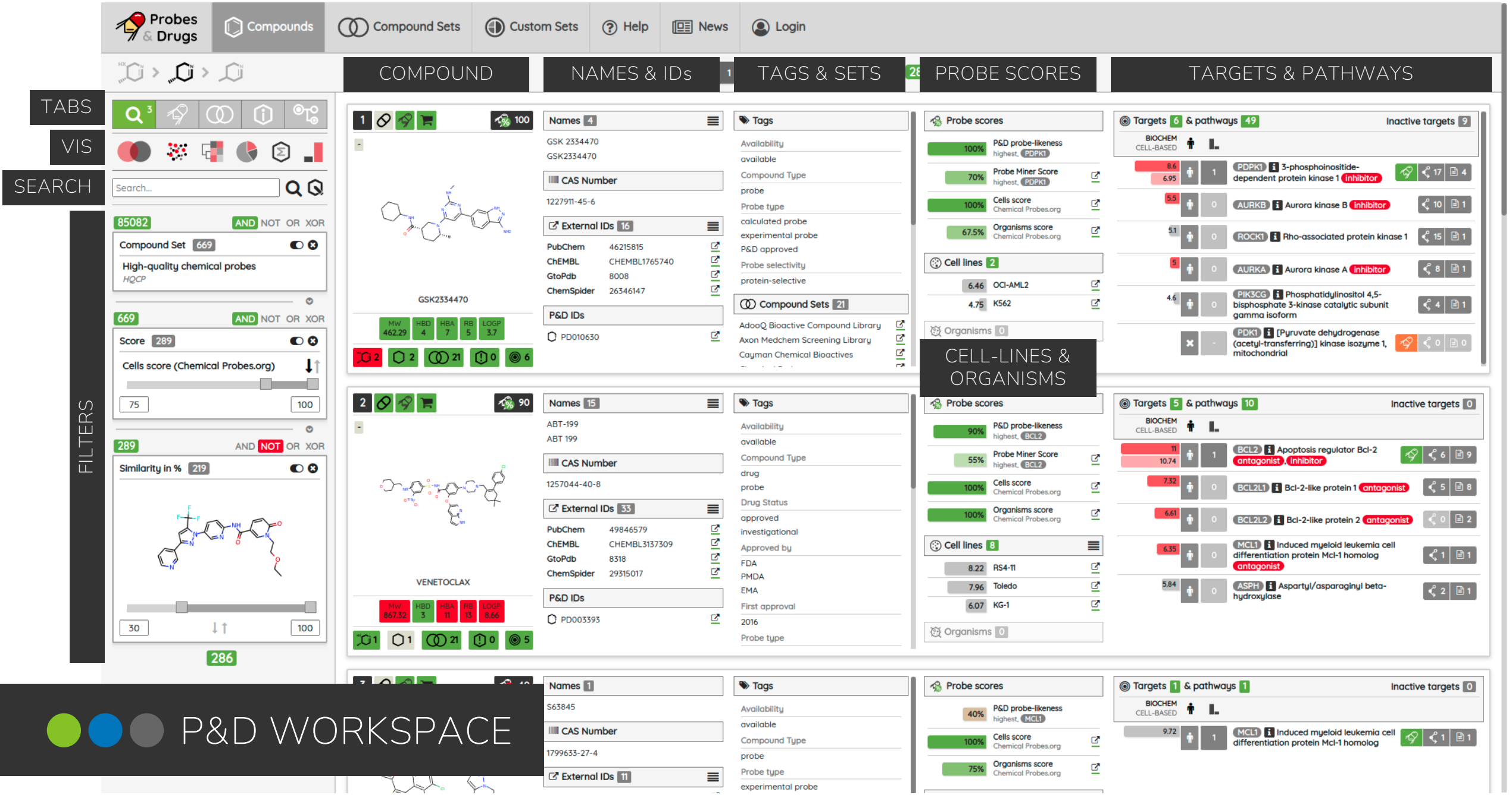
- P&D integrates 79 compound sources together with established public data resources (ChEMBL, BindingDB, GtoPDB, DrugBank, Drugcentral, Reactome and others)
- P&D probe-likeness score integrated together with scores from other probe resources Chemical Probes Portal and Probe Miner
- Live set of High-quality Chemical Probes (updated with each new P&D version)

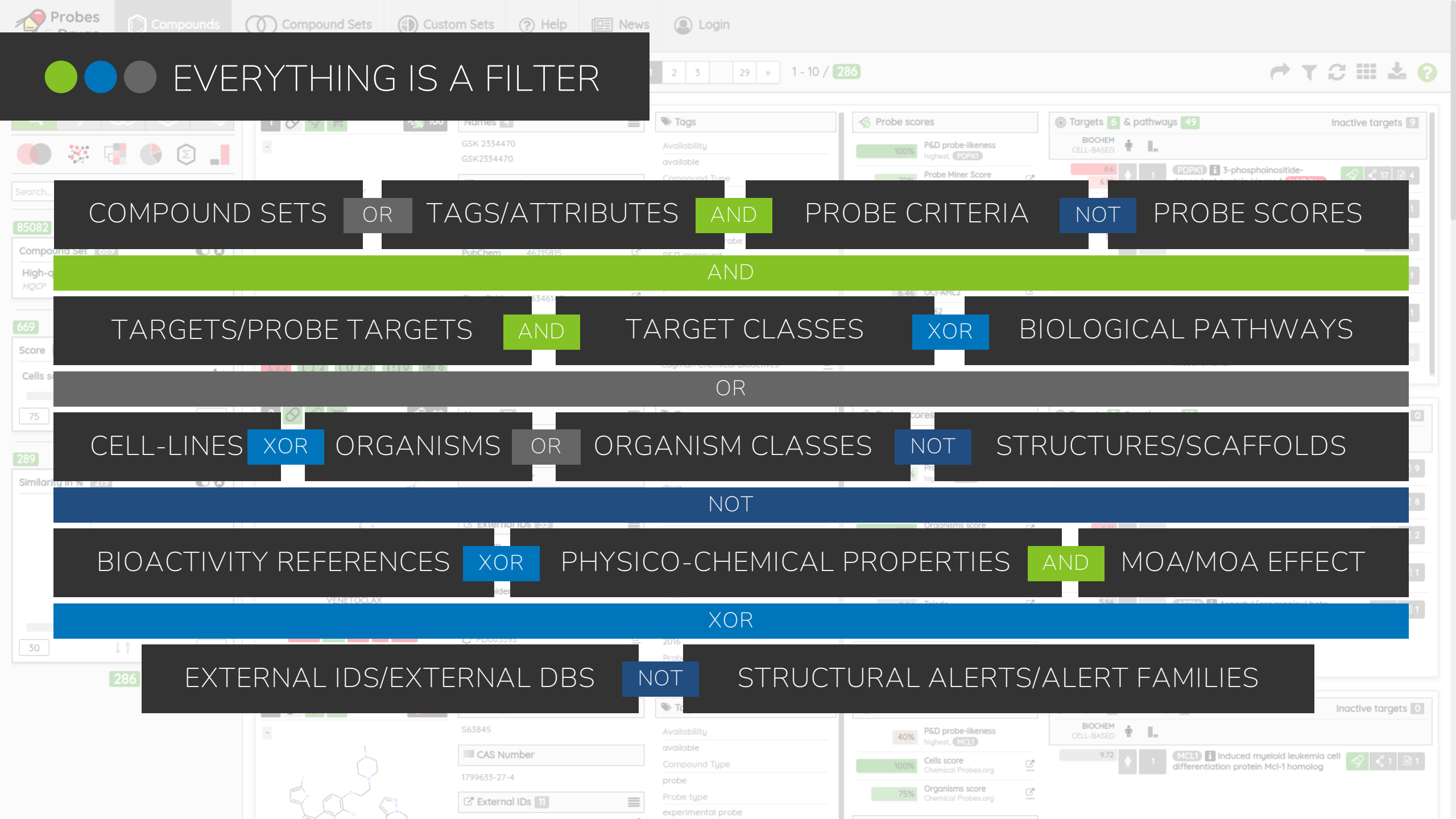


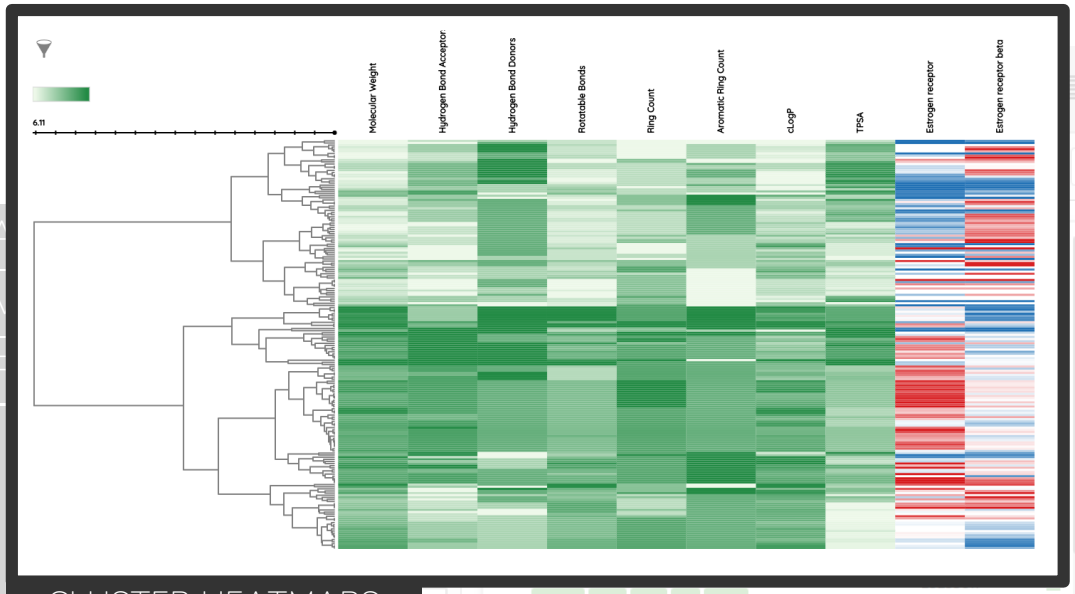
Currently, the most comprehensive resource in the field of chemical probes

Škuta, C.; Southan, C.; Bartůněk, P. "Will the Chemical Probes Please Stand Up?" *RSC Med. Chem.* 2021, *12* (8), 1428–1441.

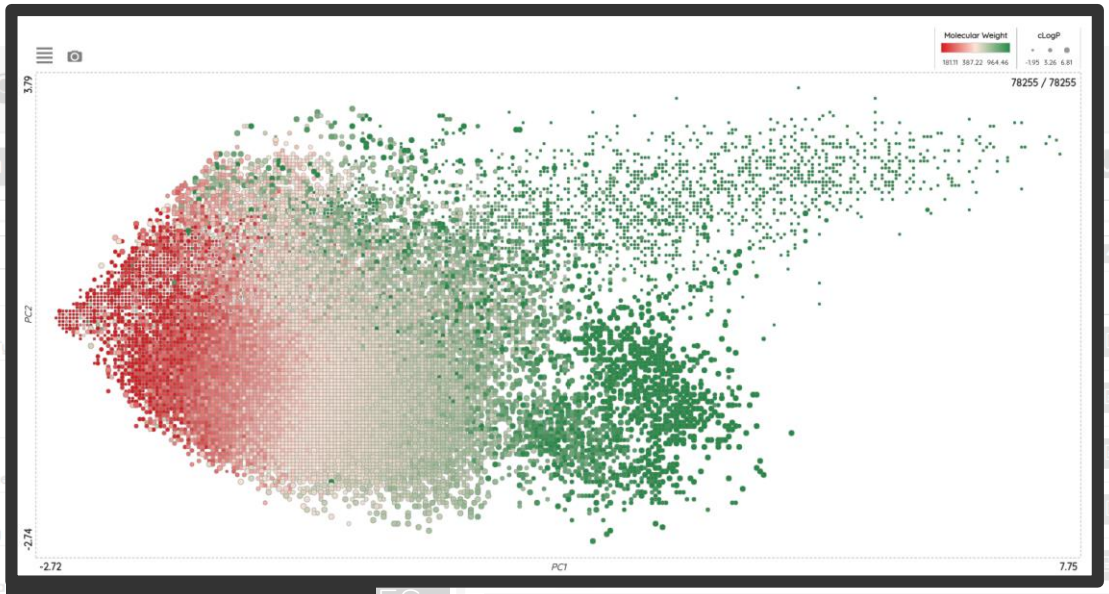



Probes & Drugs
Compounds
Compound Sets
Custom Sets
Help
News
Login
COMPOUND
NAMES & IDs
TAGS & SETS
PROBE SCORES
TARGETS & PATHWAYS
TABS
VIS
SEARCH
FILTERS
Search...
85082
AND NOT OR XOR
Compound Set 669
High-quality chemical probes
HQCP
669
Score 289
Cells score (Chemical Probes.org)
75
100
289
Similarity in % 219
30
100
286
GSK2334470
Names 4
GSK 2334470
GSK2334470
CAS Number
1227911-45-6
External IDs 16
PubChem 46215815
ChEMBL CHEMBL1765740
GtoPdb 8008
ChemSpider 26346147
P&D IDs
PD010630
Tags
Availability
available
Compound Type
probe
Probe type
calculated probe
experimental probe
P&D approved
Probe selectivity
protein-selective
Compound Sets 21
AdooQ Bioactive Compound Library
Axon Medchem Screening Library
Cayman Chemical Bioactives
Probe scores
100% P&D probe-likeness highest. PDPK1
70% Probe Miner Score highest. PDPK1
100% Cells score Chemical Probes.org
67.5% Organisms score Chemical Probes.org
Cell lines 2
6.46 OCI-AML2
4.75 K562
Organisms 0
CELL-LINES & ORGANISMS
Targets 6 & pathways 49
Inactive targets 9
BIOCHEM
CELL-BASED
PDPK1 3-phosphoinositide-dependent protein kinase 1 inhibitor
AURKB Aurora kinase B inhibitor
ROCK1 Rho-associated protein kinase 1
AURKA Aurora kinase A inhibitor
PIK3CG Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform
PDK1 [Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial
VENETOCLAX
Names 15
ABT-199
ABT 199
CAS Number
1257044-40-8
External IDs 33
PubChem 49846579
ChEMBL CHEMBL3137309
GtoPdb 8318
ChemSpider 29315017
P&D IDs
PD003393
Tags
Availability
available
Compound Type
drug
probe
Drug Status
approved
investigational
Approved by
FDA
PMDA
EMA
First approval
2016
Probe type
Probe scores
90% P&D probe-likeness highest. BCL2
55% Probe Miner Score highest. BCL2
100% Cells score Chemical Probes.org
100% Organisms score Chemical Probes.org
Cell lines 8
8.22 RS4-11
7.96 Toledo
6.07 KG-1
Organisms 0
Targets 5 & pathways 10
Inactive targets 0
BIOCHEM
CELL-BASED
BCL2 Apoptosis regulator Bcl-2 antagonist inhibitor
BCL2L1 Bcl-2-like protein 1 antagonist
BCL2L2 Bcl-2-like protein 2 antagonist
MCL1 Induced myeloid leukemia cell differentiation protein Mcl-1 homolog antagonist
ASPH Aspartyl/asparaginyl beta-hydroxylase
S63845
Names 1
CAS Number
1799633-27-4
External IDs 11
Tags
Availability
available
Compound Type
probe
Probe type
experimental probe
Probe scores
40% P&D probe-likeness highest. MCL1
100% Cells score Chemical Probes.org
75% Organisms score Chemical Probes.org
Targets 1 & pathways 1
Inactive targets 0
BIOCHEM
CELL-BASED
MCL1 Induced myeloid leukemia cell differentiation protein Mcl-1 homolog
P&D WORKSPACE


EVERYTHING IS A FILTER
COMPOUND SETS
OR
TAGS/ATTRIBUTES
AND
PROBE CRITERIA
NOT
PROBE SCORES
AND
TARGETS/PROBE TARGETS
AND
TARGET CLASSES
XOR
BIOLOGICAL PATHWAYS
OR
CELL-LINES
XOR
ORGANISMS
OR
ORGANISM CLASSES
NOT
STRUCTURES/SCAFFOLDS
NOT
BIOACTIVITY REFERENCES
XOR
PHYSICO-CHEMICAL PROPERTIES
AND
MOA/MOA EFFECT
XOR
EXTERNAL IDS/EXTERNAL DBS
NOT
STRUCTURAL ALERTS/ALERT FAMILIES



CLUSTER HEATMAPS

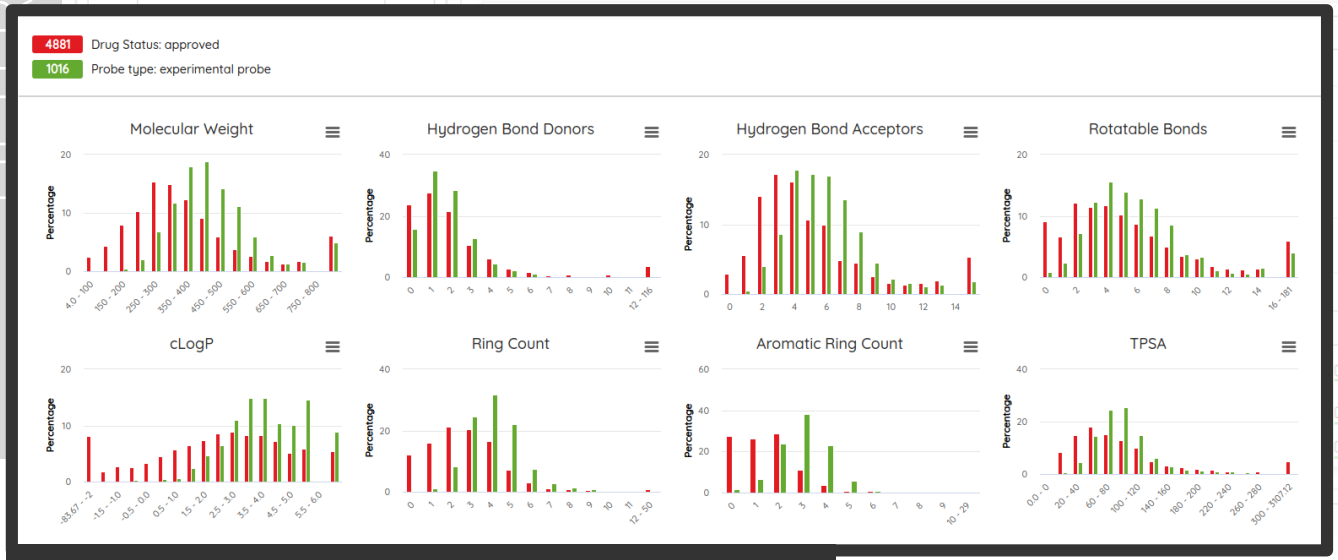


CHEMICAL SPACE

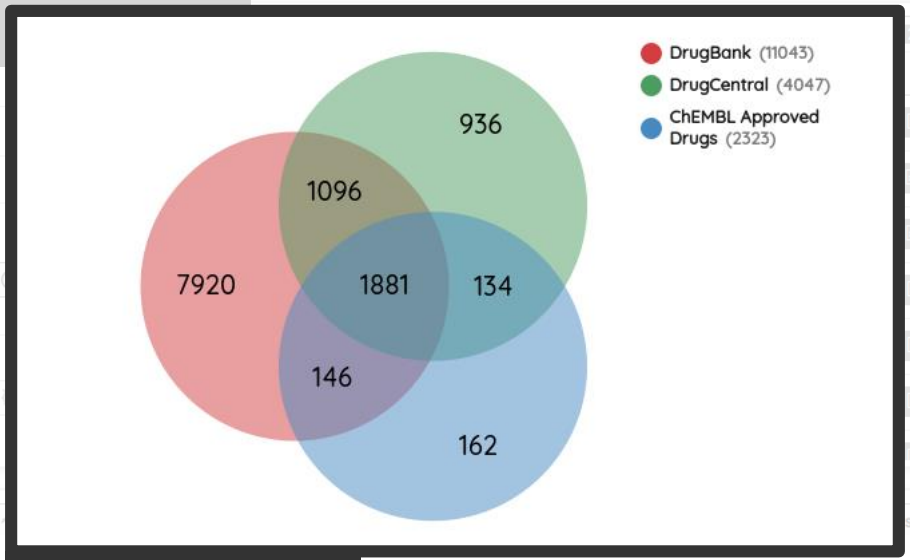
NAMES & IDs

TAGS & SETS

ORGANISM



PHYSICAL-CHEMICAL PROPERTIES DISTRIBUTION

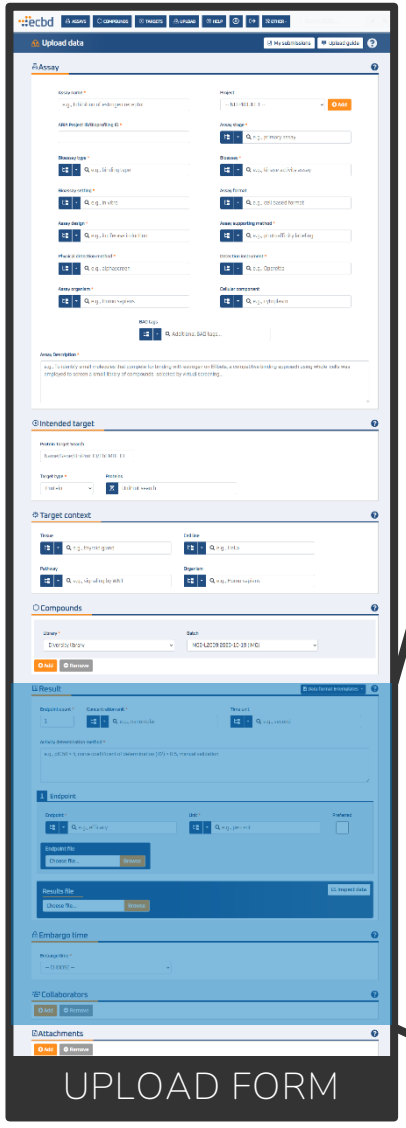


VENN DIAGRAMS

EUROPEAN CHEMICAL BIOLOGY DATABASE (ECBD)

- Central repository for data generated within the EU-OPENSCREEN network
- Web interface for data upload, browse, analysis and export
- Public and private data (under embargo up to 3 years)
- Data in ECBD must be FAIR (Findable, Accessible, Interoperable and Reusable)
- Ontologies, established IDs and formats
- Web UI, exports, database dumps (public, private) and API
- Creative Commons license
  Attribution 4.0 International (CC BY 4.0)

ECBD DATA UPLOAD



UPLOAD FORM

ONTOLOGY-BASED FIELDS

Result — Data format & templates

Endpoint count * | Concentration unit * | Time unit

1 | e.g., nanomolar | e.g., second

Activity determination method *

> 0.5, manual validation

1 Endpoint

Endpoint * | Unit * | Preferred

e.g., efficacy | e.g., percent

Endpoint file: Choose file... Browse

Results file: Choose file... Browse | Inspect data

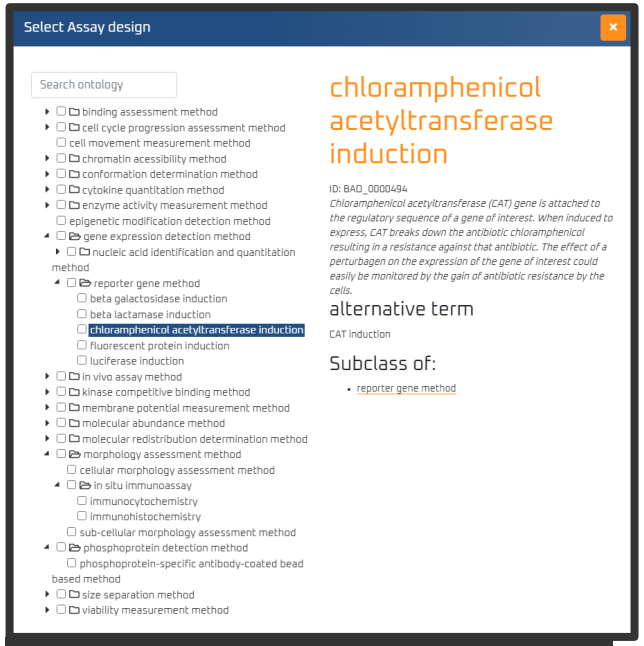
Embargo time

Embargo time *

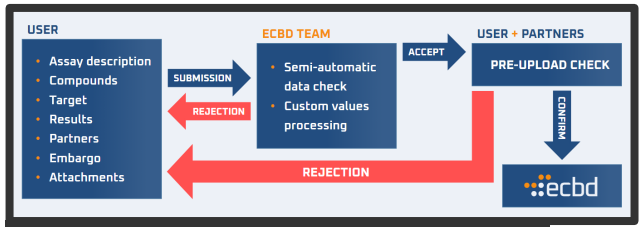
-- CHOOSE --

Collaborators

Add | Remove



ONTOLOGY TERM SELECTION



DATA UPLOAD PROCESS

ECBD DATA

UPLOAD FORM

Select Assay design

Search ontology

- binding assessment method
- cell cycle progression assessment method
- cell movement measurement method
- chromatin acessibility method
- conformation determination method
- cytokine quantitation method
- enzyme activity measurement method
- epigenetic modification detection method
- gene expression detection method
  - nucleic acid identification and quantitation method
  - reporter gene method
    - beta galactosidase induction
    - beta lactamase induction
    - **chloramphenicol acetyltransferase induction**
    - fluorescent protein induction
    - luciferase induction
- in vivo assay method
- kinase competitive binding method
- membrane potential measurement method
- molecular abundance method
- molecular redistribution determination method
- morphology assessment method
  - cellular morphology assessment method
  - in situ immunoassay
    - immunocytochemistry
    - immunohistochemistry
  - sub-cellular morphology assessment method
- phosphoprotein detection method
  - phosphoprotein-specific antibody-coated bead based method
- size separation method
- viability measurement method

chloramphenicol acetyltransferase induction

ID: BAO_0000494

Chloramphenicol acetyltransferase (CAT) gene is attached to the regulatory sequence of a gene of interest. When induced to express, CAT breaks down the antibiotic chloramphenicol resulting in a resistance against that antibiotic. The effect of a perturbagen on the expression of the gene of interest could easily be monitored by the gain of antibiotic resistance by the cells.

alternative term

CAT induction

Subclass of:

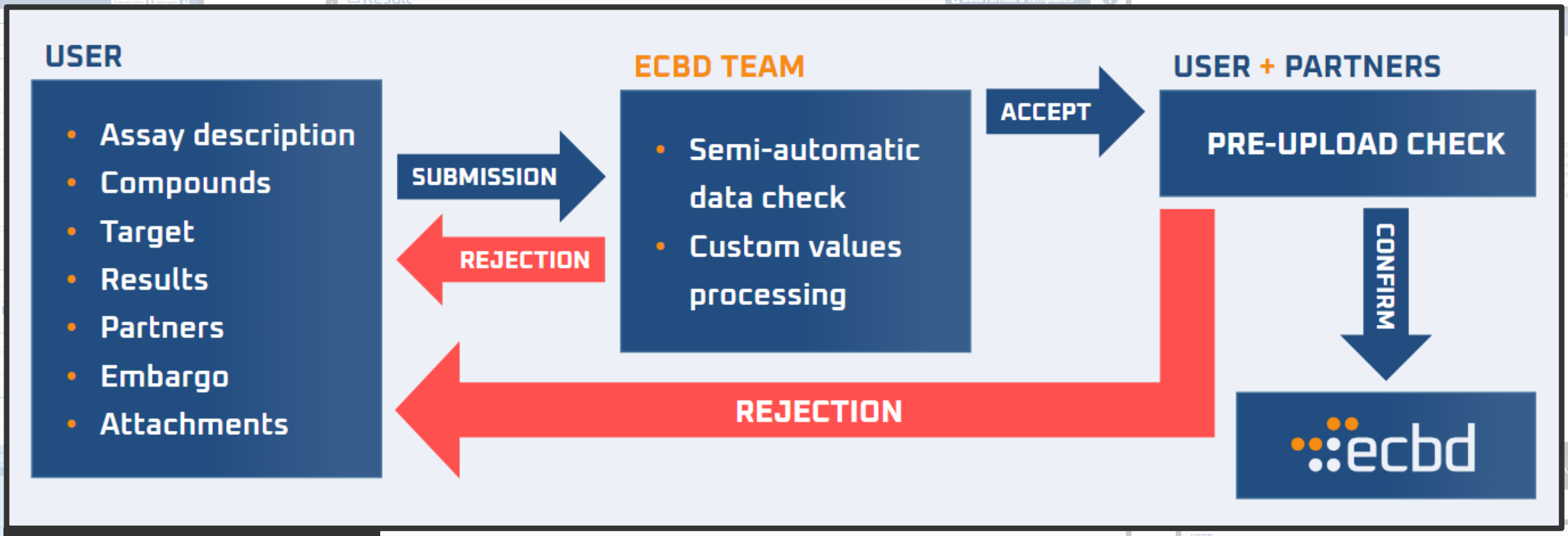
- reporter gene method

ONTOLOGY TERM SELECTION

NTOLOGY TERM SELECTION

ATA UPLOAD PROCESS

ECBD DATA UPLOAD



DATA UPLOAD PROCESS

UPLOAD FORM

DATA UPLOAD PROCESS

Assays

1 4

Search in assays

CURRENT FILTERS 2

Assay stage 3
primary assay

Bioassay 2
cell growth assay

ASSAY STAGE 3
- confirmatory assay 2
- primary assay 3
- alternate assay conditions 1

BIOASSAY TYPE 2

BIOASSAY 4
- cell growth assay 2
- protease activity assay 3
- signal transduction assay 1
- Fluorescence imaging cell growth inhibition assay 2

ASSAY ORGANISM 4
- Viruses 1

Assay EOS300008

Name	Single dose inhibition of SARS-CoV-2 cytopathic effect in VeroE6
Stage	primary assay
Type	functional
Description	To measure inhibition of the SARS-CoV-2 cytopathic effect, 384-well imaging plates (Greiner 781092) were spotted with test compounds and controls (16 ... (*read more*)

Compounds 2464

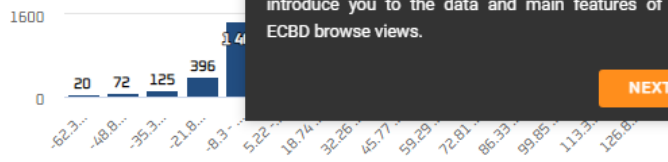
Activity	✓	?	×	...
	126	0	2337	2

Target

Name	Organism Severe acute respiratory syndrome-related coronavirus
Type	Organism

Activities 2686

Endpoint: percent inhibi...



Assay EOS300033

Name	Dose-dependent inhibition of SARS-CoV-2 cytopathic effect in VeroE6
Stage	confirmatory assay
Type	functional
Description	To measure inhibition of the SARS-CoV-2 cytopathic effect, 384-well imaging plates (Greiner 781092) were spotted with test compounds and controls (16 ... (*read more*)

Compounds 134

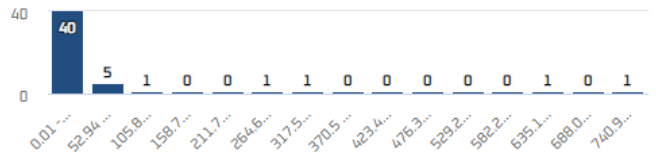
Activity	✓	?	×	...
	19	0	115	0

Target

Name	Organism Severe acute respiratory syndrome-related coronavirus
Type	Organism

Activities 50

Endpoint: IC50 relative



Assay EOS300038

Name	GPR35 antagonism
Stage	primary assay
Type	functional
Description	Although it has been described as a G protein-coupled receptor (GPCR) activated by either kynurenic acid, or by the chemokine CXCL17, GPR35 remains an... (*read more*)

Compounds 92313

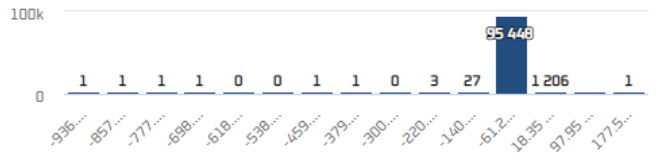
Activity	✓	?	×	...
	0	0	92312	2

Target

Name	G-protein coupled receptor 35
Type	Protein

Activities 98600

Endpoint: percent inhibition



Assay EOS300040

Description	The activity of 3-CLpro (GenBank : MN908947.3 Residues 3264-3569) was measured by a Förster resonance energy transfer

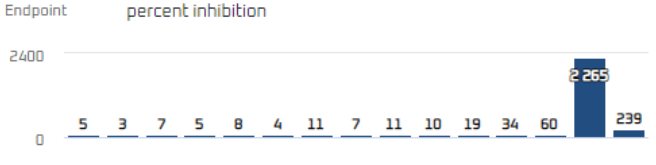
Compounds 2465

Activity	✓	?	×	...
	64	0	2400	2

Target

Activities 2688

Endpoint: percent inhibition



ECBD guide

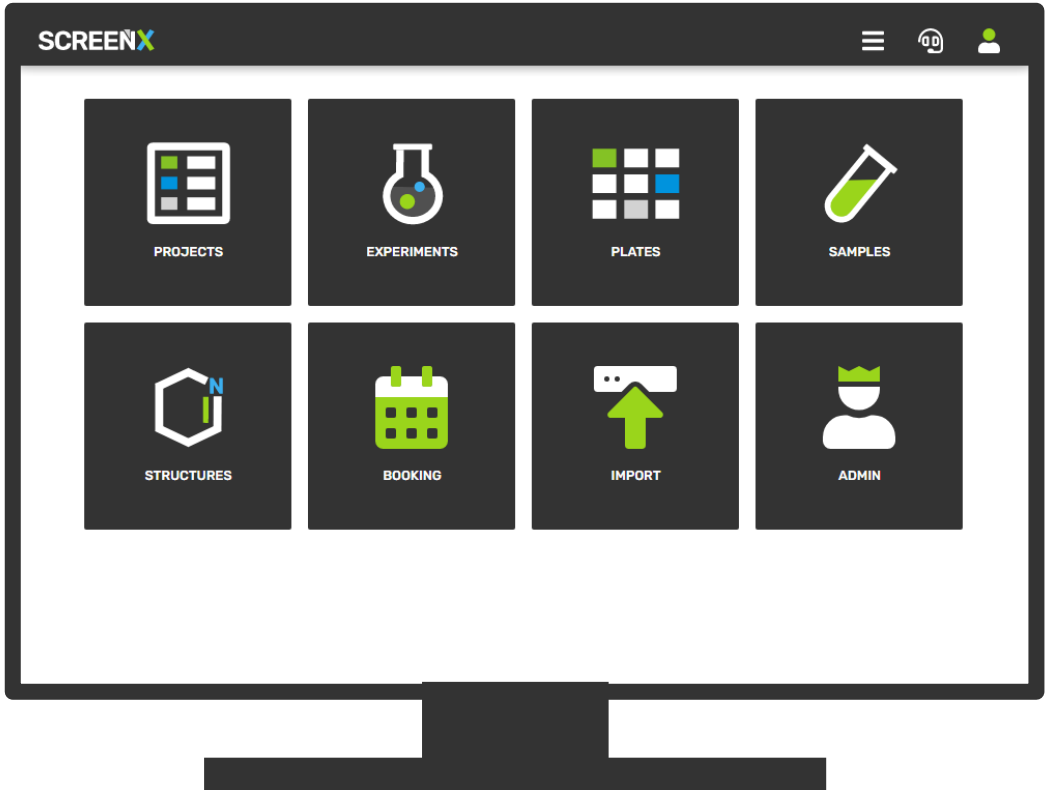
Welcome to the ECBD guide! This guided tour will introduce you to the data and main features of the ECBD browse views.

NEXT

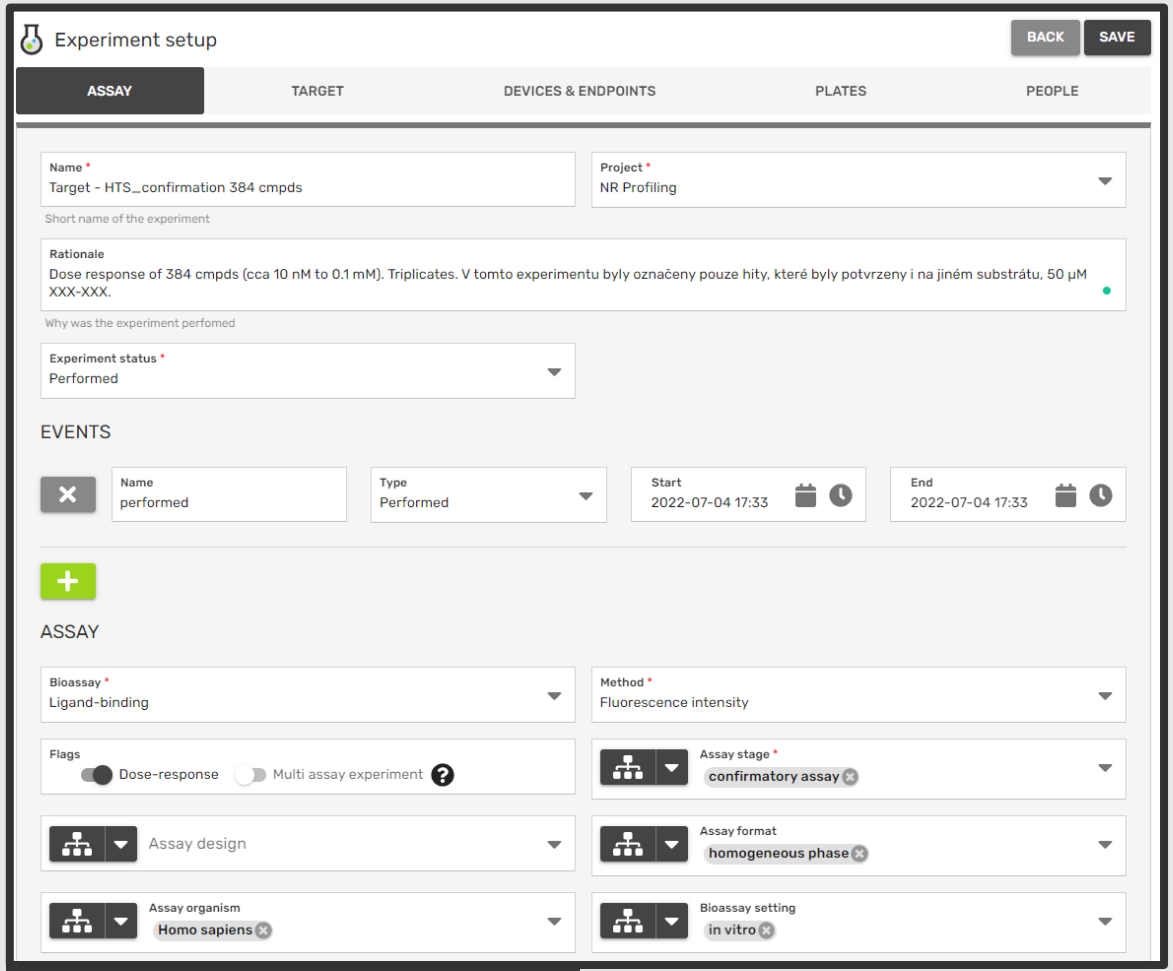
ECBD WORKSPACE

SCREENX

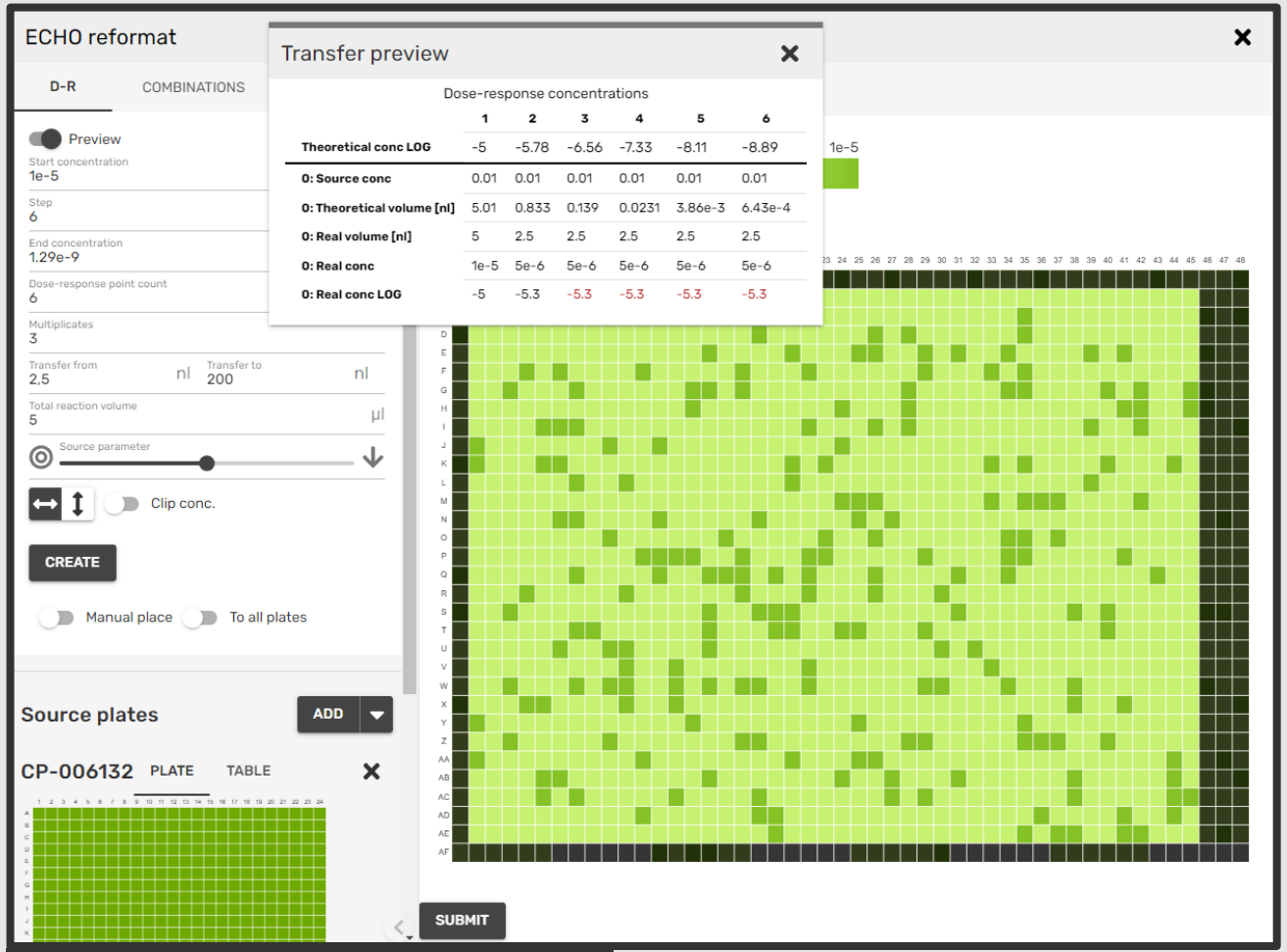
- Laboratory information management system (LIMS)
- In an ideal case (in a distant future), it should be able to track everything that is happening in the lab
- Compound management, plate reformatting, instrument instructions, project/experiment definition, data normalizations/analysis, reporting, imports/exports, booking system, attachments, user roles



EXPERIMENT SETUP

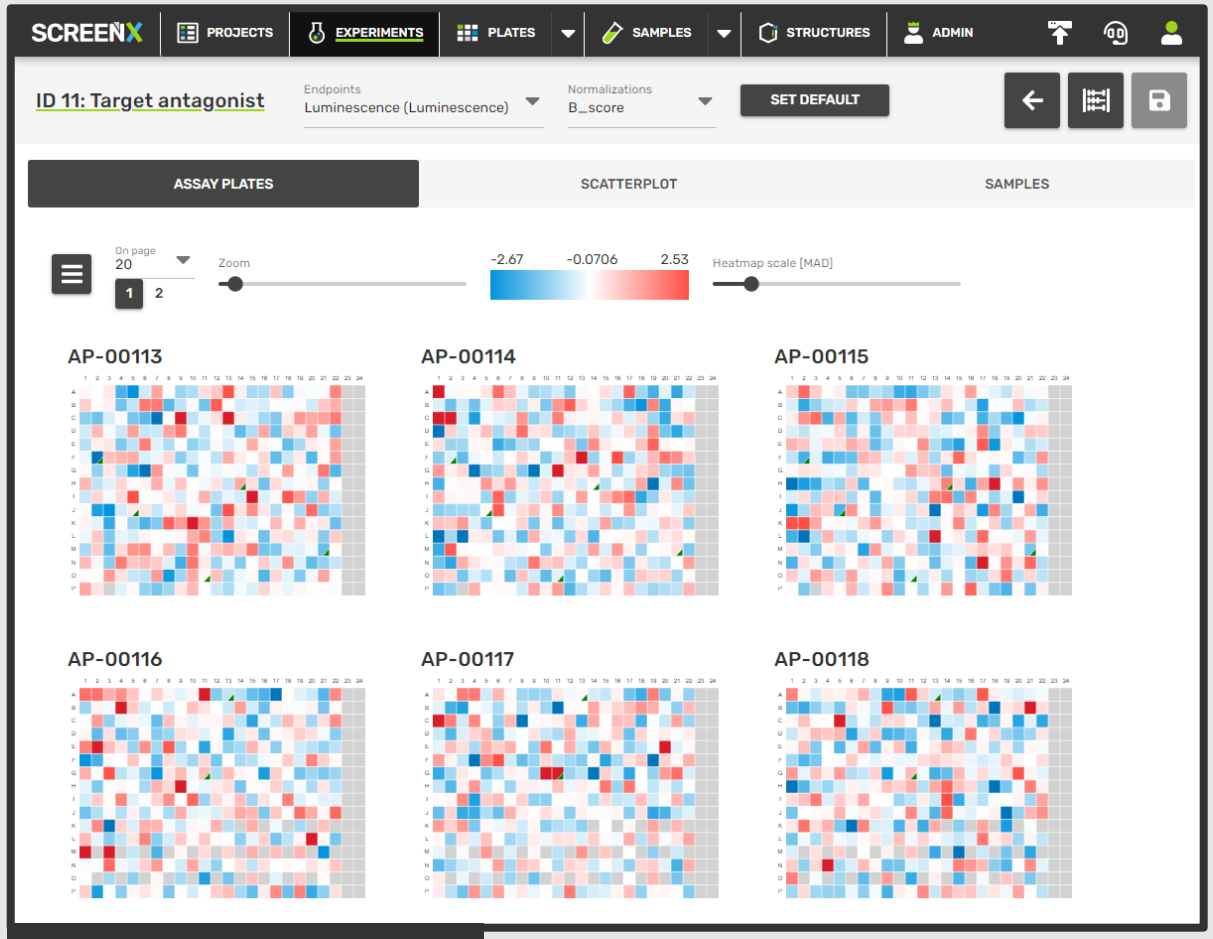


MULTI-TAB DESCRIPTION FORM

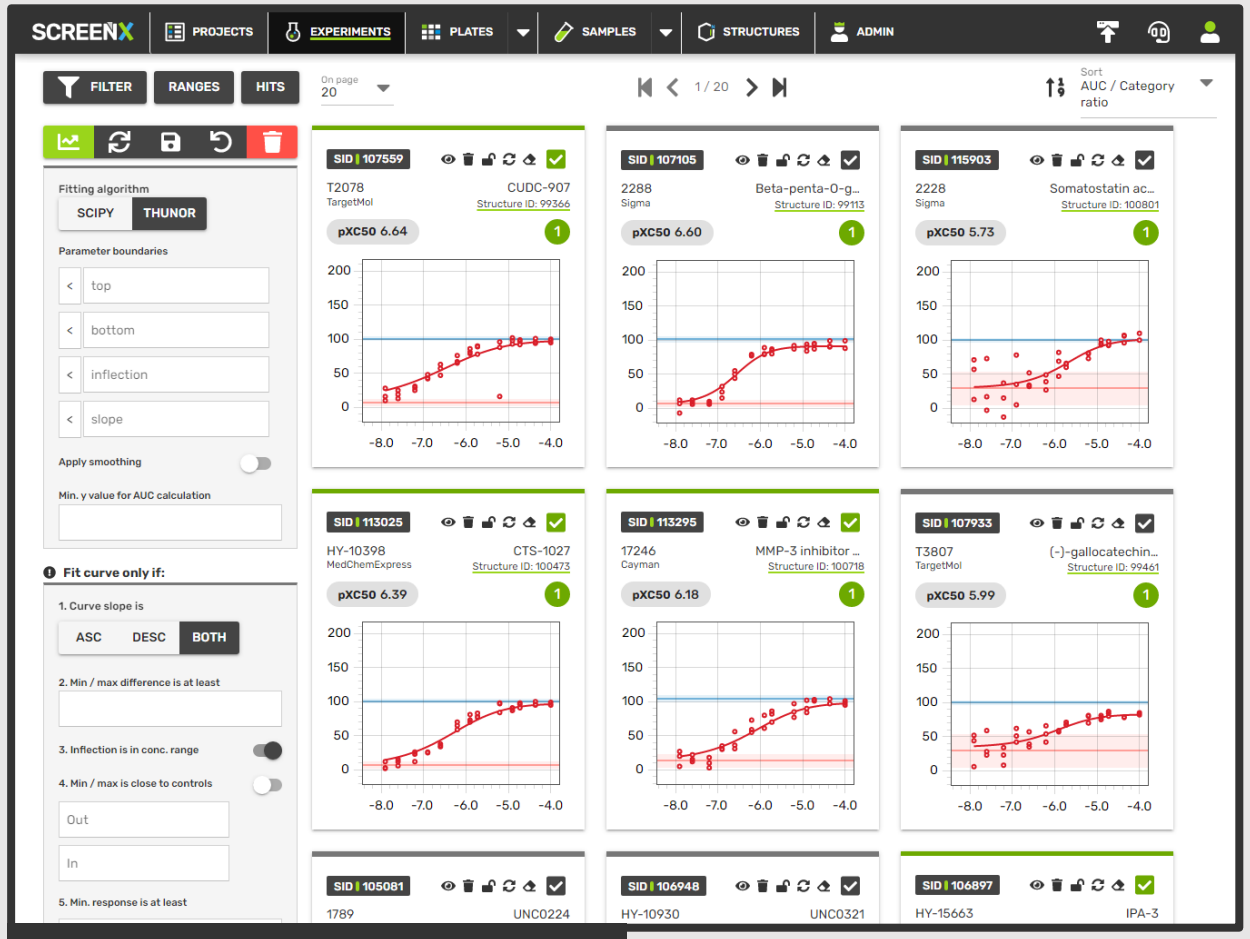


ASSAYPLATES REFORMATTING

EXPERIMENT ANALYSIS



ASSAYPLATE HEATMAPS



DOSE-RESPONSE CURVE FITTING

THE TEAM

CONTACT

- skutac@img.cas.cz
- cz-openscreen@img.cas.cz
- openscreen.cz/en/contact

openscreen.cz

CZ OPENSCREEN