Rancho BioSciences GiNAS Curation

Yoshiyuki Tokiwa, PhD
Yulia Skovpen, PhD
Laura Brovold, PhD

April 2018

Agenda

- What does Rancho BioSciences do?
  - Example curation projects
- InXight/GiNAS project – compound curation
  - Curation Interface
- PK Data for G-SRS

Rancho BioSciences Services

Bioinformatics Analysis Services

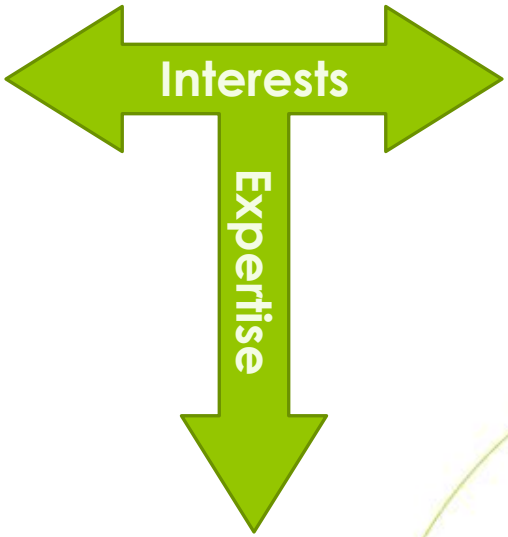
Data Curation Services

IT Solutions

Hosting Services: tranSMART, etc…

Reducing Friction – Business Model

- Fee for Service. No requirement for pre-committed hours – because most customers are too dynamic for this model.
  - FTE or project based
- We have a large team
  - Can commit people to your projects quickly.



IT
Bioinformatics

"Rancho BioSciences take care of data quality, normalization, and curation as well as bioinformatics analysis and knowledge mining for multiple teams at CHDI. Colleagues often approach me to engage Rancho's services now; they've become an integral part of how we work." - Jeff Aaronson, CHDI

How do people benefit from organized, clean data?

- Statisticians/Data analyst – Less time looking for data
  - "I spend 75% of my time looking for data. This would save so much time." - Statistician, Mayo Clinic
- Researchers – More precise models and approaches
  - *"Harmonizing disparate datasets across domains for imaging, clinical, bioassays (genomic and proteomics), and multidimensional outcomes allows us to build more precise models and approaches for improving the diagnosis and treatment of TBI."*

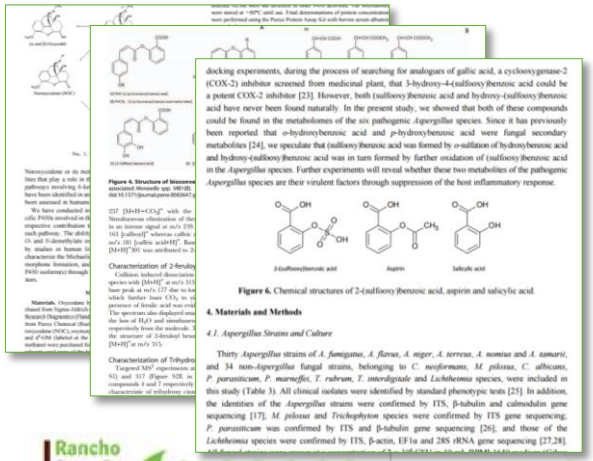
    - Mary Vassar, UCSF
- Publications on new methodologies
  - http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178006
  - https://www.ncbi.nlm.nih.gov/pubmed/29020921?dopt=Abstract

Chemical report data conversion to a machine readable format

- Created a tool which converts chemical reports into a spreadsheet
  - Text mining
  - Chemical structure conversion into SMILEs
    - exact structures
    - uncertain metabolite structures

We have a very high opinion of Rancho capabilities
-Pharma customer

- Generated spreadsheet is integrated into an internal database



docking experiments, during the process of searching for analogues of gallic acid, a cyclooxygenase-2 (COX-2) inhibitor screened from medicinal plant, that 3-hydroxy-4-(sulfooxy)benzoic acid could be a potent COX-2 inhibitor [23]. However, both (sulfooxy)benzoic acid and hydroxy-(sulfooxy)benzoic acid have never been found naturally. In the present study, we showed that both of these compounds could be found in the metabolomes of the six pathogenic *Aspergillus* species. Since it has previously been reported that *o*-hydroxybenzoic acid and *p*-hydroxybenzoic acid were fungal secondary metabolites [24], we speculate that (sulfooxy)benzoic acid was formed by *o*-sulfation of hydroxybenzoic acid and hydroxy-(sulfooxy)benzoic acid was in turn formed by further oxidation of (sulfooxy)benzoic acid in the *Aspergillus* species. Further experiments will reveal whether these two metabolites of the pathogenic *Aspergillus* species are their virulent factors through suppression of the host inflammatory response.

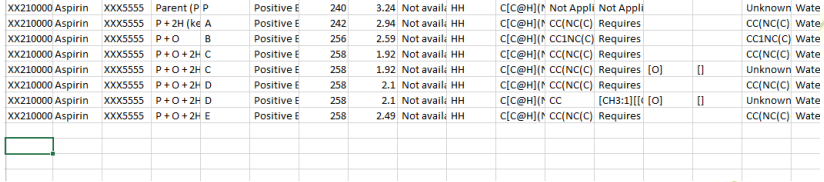
2-(sulfooxy)benzoic acid | Aspirin | Salicylic acid

Figure 6. Chemical structures of 2-(sulfooxy)benzoic acid, aspirin and salicylic acid.

4. Materials and Methods

4.1. Aspergillus Strains and Culture

Thirty *Aspergillus* strains of *A. fumigatus*, *A. flavus*, *A. niger*, *A. terreus*, *A. nomius* and *A. tamarii*, and 34 non-*Aspergillus* fungal strains, belonging to *C. neoformans*, *M. pilosus*, *C. albicans*, *P. parasiticum*, *P. marneffei*, *T. rubrum*, *T. interdigitale* and *Lichtheimia* species, were included in this study (Table 3). All clinical isolates were identified by standard phenotypic tests [25]. In addition, the identities of the *Aspergillus* strains were confirmed by ITS, β-tubulin and calmodulin gene sequencing [17]; *M. pilosus* and *Trichophyton* species were confirmed by ITS gene sequencing; *P. parasiticum* was confirmed by ITS and β-tubulin gene sequencing [26]; and those of the *Lichtheimia* species were confirmed by ITS, β-actin, EF1α and 28S rRNA gene sequencing [27,28].



Study Coc	Compoun	Compoun	Tentative	Peak ID	Ion Mode	Ions Foun	RT(min)	Product I	Species N	SMILES fo	SMILES fo	SMILES fo	Atoms Ad	Atoms Su	SMILES fo	UPLC
XX210000	Aspirin	XXX5555	Parent (P	P	Positive E	240	3.24	Not availa	HH	C[C@H](N	Not Appli	Not Appli			Unknown	Wate
XX210000	Aspirin	XXX5555	P + 2H (ke	A	Positive E	242	2.94	Not availa	HH	C[C@H](N	CC(NC(C)	Requires			CC(NC(C)	Wate
XX210000	Aspirin	XXX5555	P + O	B	Positive E	256	2.59	Not availa	HH	C[C@H](N	CC1NC(C)	Requires			CC1NC(C)	Wate
XX210000	Aspirin	XXX5555	P + O + 2H	C	Positive E	258	1.92	Not availa	HH	C[C@H](N	CC(NC(C)	Requires			CC(NC(C)	Wate
XX210000	Aspirin	XXX5555	P + O + 2H	C	Positive E	258	1.92	Not availa	HH	C[C@H](N	CC(NC(C)	Requires	[O]	[]	Unknown	Wate
XX210000	Aspirin	XXX5555	P + O + 2H	D	Positive E	258	2.1	Not availa	HH	C[C@H](N	CC(NC(C)	Requires			CC(NC(C)	Wate
XX210000	Aspirin	XXX5555	P + O + 2H	D	Positive E	258	2.1	Not availa	HH	C[C@H](N	CC	[CH3:1][[(	[O]	[]	Unknown	Wate
XX210000	Aspirin	XXX5555	P + O + 2H	E	Positive E	258	2.49	Not availa	HH	C[C@H](N	CC(NC(C)	Requires			CC(NC(C)	Wate

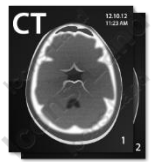
Rancho bioSciences | www.RanchoBioSciences.com

Use Case: Cleanup List of Drugs

- On Ontologies side, used **SNOMED-CT**
- On Vocabularies side, used **NDF-RT** XML
- Neither proved enough, because the data contained brand names (i.e. Advil for Ibuprofen)
- Added **FDA label data** (map Advil > Ibuprofen)

A	B	C	D	E
Advil Pm	ADVIL PM	IBUPROFEN	C0020740	1
Aerius	AGARICUS	AGARICUS	C0001775	0.63
Aerobid	AEROBID	FLUNISOLIDE	C0060501	1
Afeditab	AFEDITAB CR	NIFEDIPINE	C0028066	0.73
Afrin	AFRINOL	PSEUDOEPHEDRINE	C0033798	0.71

Example of data portal - UCSF



4,000
CT scans

CT imaging software

CT image analysis

Radiologists

CT analysis annotation

Query/export data

Study coordinators and other relevant parties

Monitoring, quality control and query/export data

Sponsors

Monitoring and query/export data

Ability to query TRACK-TBI database and export annotation data subsets in a format that can be imported into QG database

TRACK-TBI
Transforming Research and Clinical Knowledge in Traumatic Brain Injury
International Traumatic Brain Injury Research Initiative

MySQL

Cloud app
Angular.js / Node.js

Database

Data entry form

Dashboard and quality control module

Query and export data

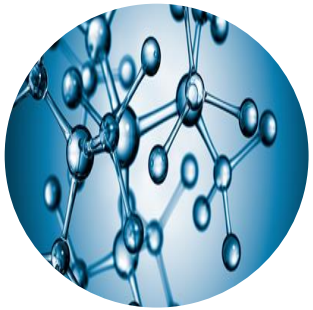
Imaging Annotation & Management Tool

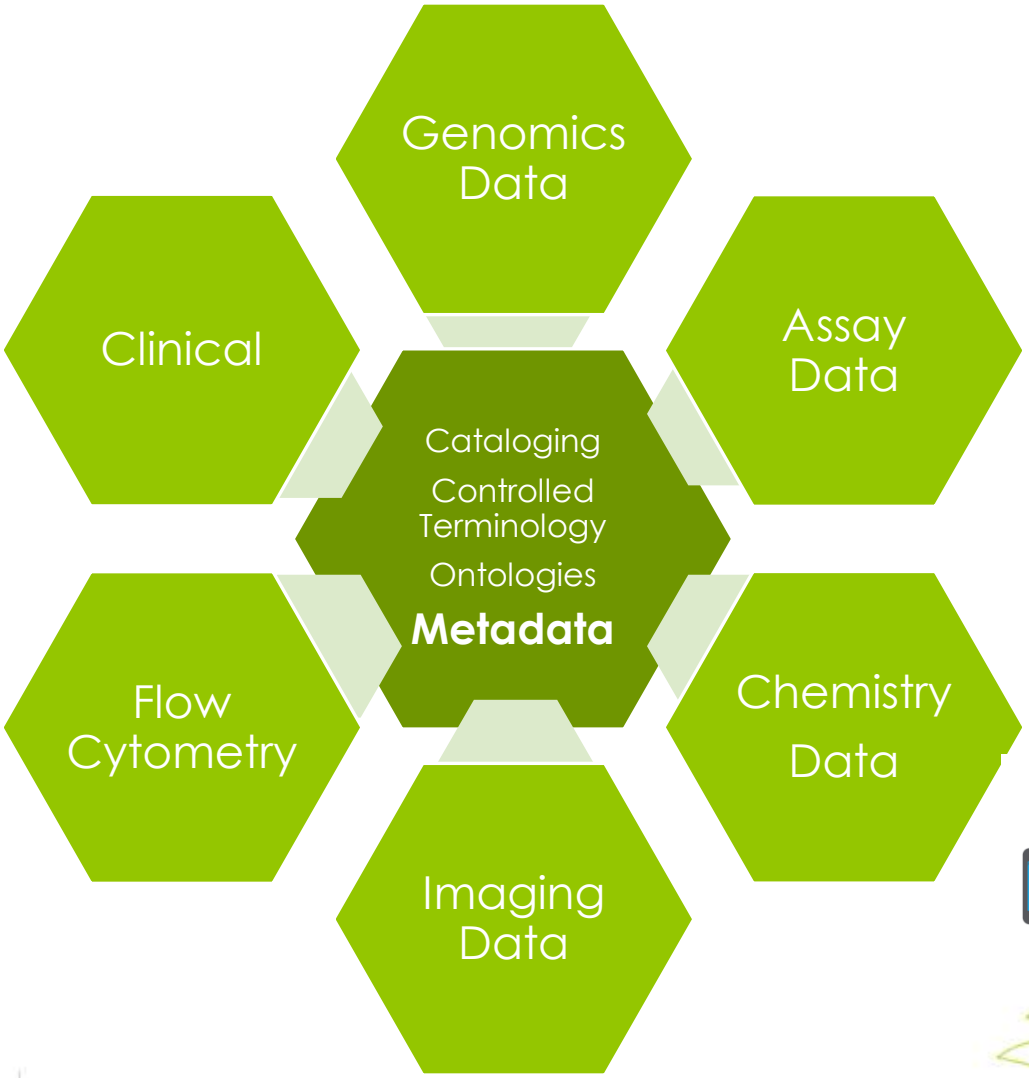
to be developed by Rancho BioSciences

mongoDB.

Imaging Annotation dedicated **database**:
- **Flexible** (data model must allow easy creation of new annotation fields) and **scalable**
- MongoDB based on the structure defined by Esther Yuh

Pharma Cataloging Projects



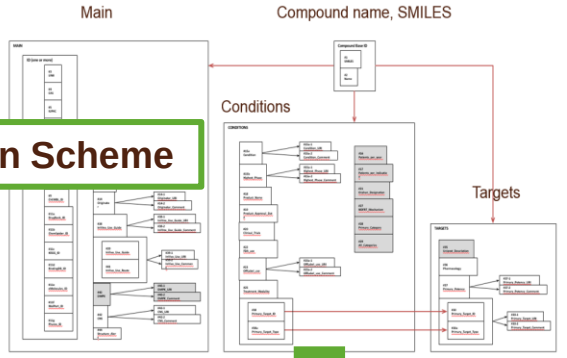


InXight Compounds Curation

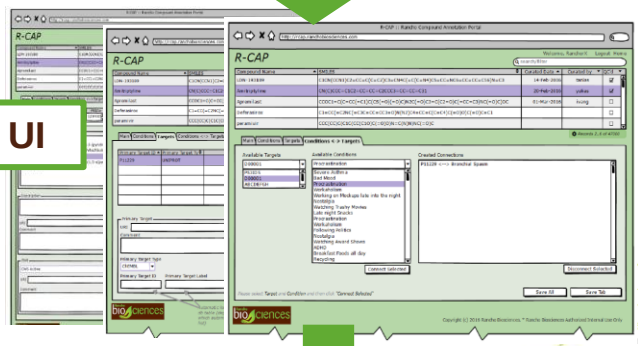
Compounds Curation and Annotation

- More than 9,000 compounds curated
  - Manual and automated approach
  - Curation Interface
    - > 70 Fields for curation (can add new and modify)
    - Built-in ontologies
    - Built-in QC elements
    - More than 10 curators can work simultaneously
  - Curated data is included in **Inxight Drugs**
  - https://drugs.ncats.io/ginas/app

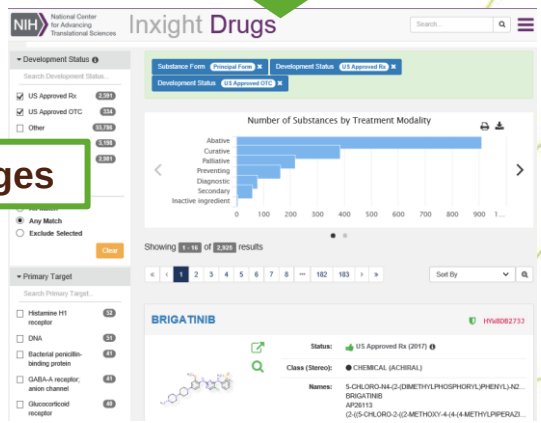
Curation Scheme



Automation - Curation UI



Drug Pages



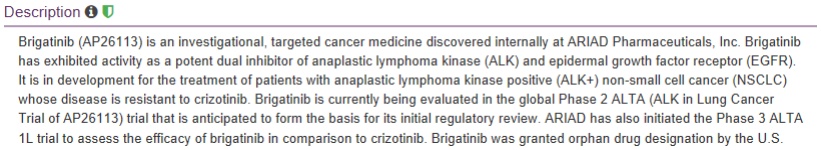
Rancho BioSciences | www.RanchoBioSciences.com

Automation – Curation Interface

Rancho Curation Interface

Comprehensive Manual Curation Effort for Inxight Drugs

- Over **9000** curated substances
- **40 to 200+** total fields per substance
- Concise **descriptions**


Description

Brigatinib (AP26113) is an investigational, targeted cancer medicine discovered internally at ARIAD Pharmaceuticals, Inc. Brigatinib has exhibited activity as a potent dual inhibitor of anaplastic lymphoma kinase (ALK) and epidermal growth factor receptor (EGFR). It is in development for the treatment of patients with anaplastic lymphoma kinase positive (ALK+) non-small cell cancer (NSCLC) whose disease is resistant to crizotinib. Brigatinib is currently being evaluated in the global Phase 2 ALTA (ALK in Lung Cancer Trial of AP26113) trial that is anticipated to form the basis for its initial regulatory review. ARIAD has also initiated the Phase 3 ALTA 1L trial to assess the efficacy of brigatinib in comparison to crizotinib. Brigatinib was granted orphan drug designation by the U.S.

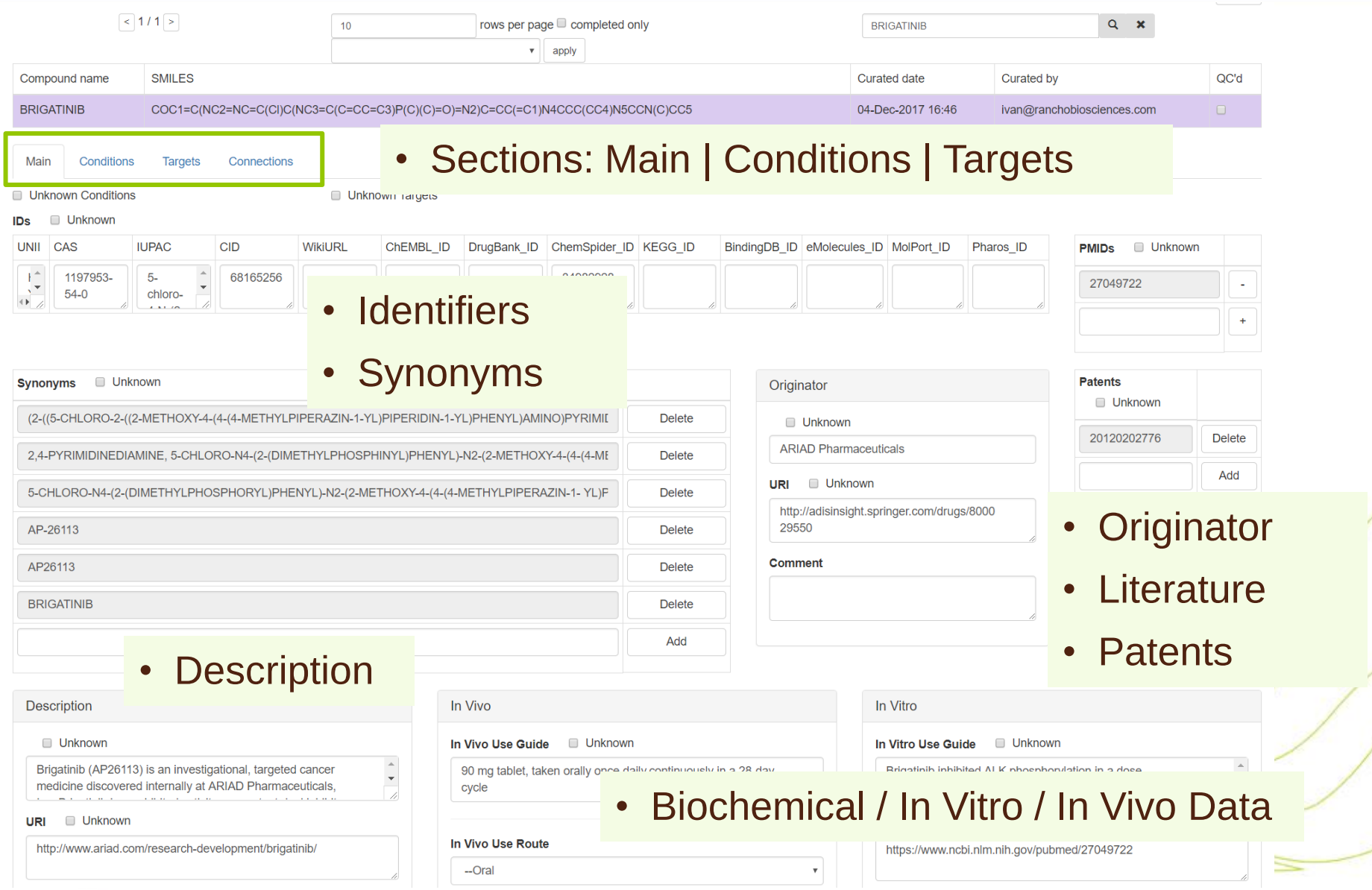
- **Conditions** / **Targets**
- Development status / Approved products / Off-label indications
- **Sources** for all data
- Data source for Inxight Drugs (portal for drug development information)
- Example: https://drugs.ncats.io/ginas/app/drug/HYW8DB273J

Rancho Data
GSRS
…

→ NCATS Stitcher →

Inxight Drugs
https://drugs.ncats.io

Rancho Curation Interface



- Sections: Main | Conditions | Targets
- Identifiers
- Synonyms
- Originator
- Literature
- Patents
- Description
- Biochemical / In Vitro / In Vivo Data

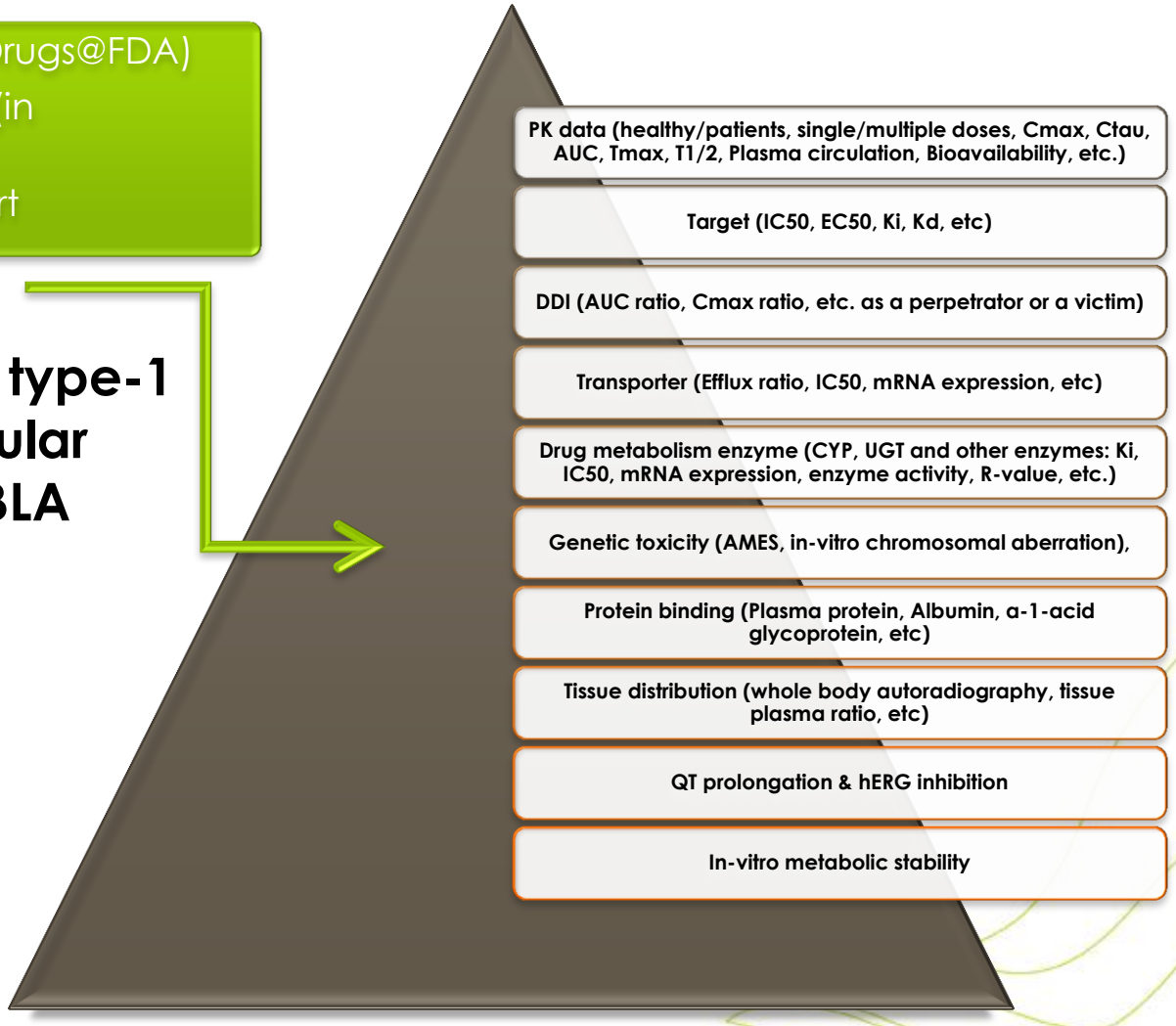
PK Data for G-SRS

PK data curation for G-SRS

Data Sources

- NDA review reports (Drugs@FDA)
- PMDA review reports (in Japanese)
- EMA assessment report

Focusing on type-1 (new molecular entity) and BLA



PK data curation process

- NDA clinical, non-clinical pharmacology, and/or Microbiology review report(s)
  - Pick up PK parameters/other clinical non-clinical data from TEXT
  - Data on PK data tables
  - Clinical/non-clinical pharmacology data tables (plasma distribution, elimination, protein binding, tissue distribution, ΔQTc or ΔΔQTc, hERG (IC50), transporter, DDI, CYP, AMES test, in-vitro chromosomal aberration, etc.)
  - Targets are directly input in G-SRS
  - Population PK parameter table (final model: Vc, Vp, CL, and simulated AUC/Cmax if available)
- Goal: Create a standardized table to give to pharma/biotech to this will be “automated”

Curation table

- PK data (ex. VABORBACTAM, component of VABOMERE)

PK - Absorption, Distribution & Elimination													
Study No.	Study com	Substance	Metabolit	Substance	Substance	Product na	Product ID	Reference	reference	Clinically l	Dosage Ac	Recomme	Species
Study 402,	Phase 1, (s	VABORBACTAM		1148818A/	1C75676F8V			https://w	102-107 (112-117)				Human
								https://w	102-107 (112-117)				Human
								https://w	102-107 (112-117)				Human
								https://w	102-107 (112-117)				Human

Type of Ar	Populatio	Special po	C	Route of A	Dosage fo	Dosage sti	Dose_low	Dose_high	Dosage_ti	Dosage Ur	Dose_radi	Dose_RI-U	Timing of	Dose Freq
	Adult	Healthy	m	IV infusion (3-h)	250					mg				Once ever
	Adult	Healthy	m	IV infusion (3-h)	1000					mg				Once ever
	Adult	Healthy	m	IV infusion (3-h)	1500					mg				Once ever
	Adult	Healthy	m	IV infusion (3-h)	2000					mg				Once ever

				AUCtau_mean							AUCtau_Ratio			Cmax						Cmax_Rat
C	N_total	N_male	N_female	N	AUCtau_m	AUCtau_%	AUCtau_m	AUCtau_lo	AUCtau_h	AUCtau_u	A	AUCtau_R	AUCtau_R	A	Cmax_me	Cmax_%C	Cmax_mo	Cmax_low	Cmax_hig	Cmax_uni	Cmax_Rat
	6				16.3					mcg*h/mL					4.81					mcg/mL	
	5				74.6					mcg*h/mL					21.3					mcg/mL	
	6				118					mcg*h/mL					33.4					mcg/mL	
	6				145					mcg*h/mL					40.9					mcg/mL	

Rancho team

Laura Brovold, Business Development
Laura.Brovold@ranchobiosciences.com

Yulia Skovpen, Data Curator
Yulia.Skovpen@ranchobiosciences.com

Yoshiyuki Tokiwa, Data Curator
Yoshiyuki.Tokiwa@fda.hhs.gov

Ivan Grishagin, Bioinformatics Scientist
Ivan.Grishagin@ranchobiosciences.com

Danny Katzel, Senior Software Engineer
Daniel.Katzel@ranchobiosciences.com