CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

DEMYSTIFYING MACHINE LEARNING (AI) IN DRUG DISCOVERY

Moderated by **John Overington** of *Medicines Discovery Catapult*
Panel: **Jeff Warrington**, of *Atomwise* and **John Griffin** of *Integral Health*

September 10, 2020

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Today's Introduction by…

Frank Cole

Head of Sales

Collaborative Drug Discovery Inc.

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Today's Moderator and Panelists

John Overington, Ph.D.
Chief Informatics Officer,
Medicines Discovery Catapult

John Griffin, Ph.D.
Vice President, Integral Health

Jeff Warrington, Ph.D.
Senior Scientist, Atomwise

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Questions?

Have a question to ask our panel?
Open the ZOOM Q&A and type in your question at anytime !

We'll be sure to save time for them later!

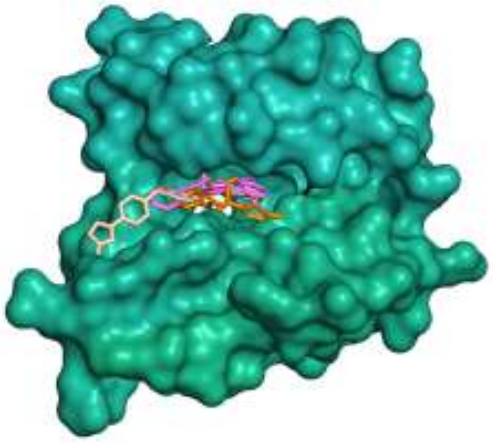
CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Atomwise

CDD Webinar: Demystifying Machine Learning (AI) in Drug Discovery

Jeff Warrington, PhD.
Senior Scientist
Atomwise Inc.

Traditional Docking Approaches



- Based on equations which approximate physical interactions (e.g., Coulomb, Van Der Waals).
- To improve accuracy, additional statistically derived terms are sometimes added to the physical potential:
  - Desolvation penalty
  - Hydrophobic enclosure
  - Hydrogen bond motifs
- These equations rely on human intuition about the most relevant physical interactions.

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] +$$

$$\sum_{nb\ pairs} \frac{q_i \cdot q_j}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{nb\ pairs} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right) +$$

$$\sum_{H\ bonds} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}} \right) \cdot \cos^2(\theta_{Don-H\cdots Acc}) \cdot \cos^n(\omega_{H\cdots Acc-LP}) +$$

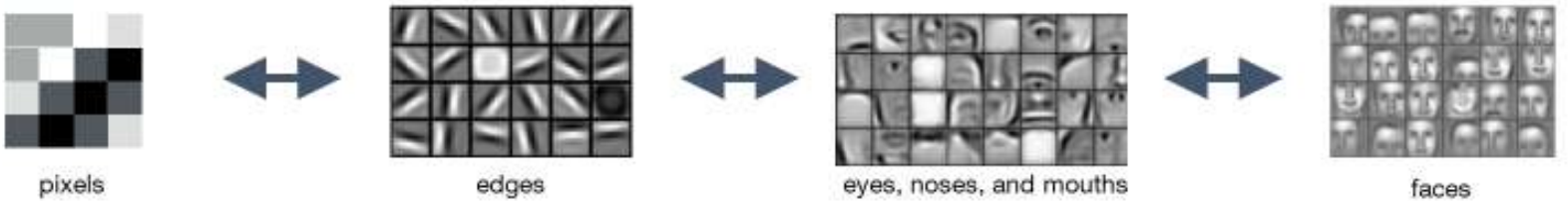
$$\sum_{metal\ pairs} \frac{q_i^{CT} \cdot q_j^{CT}}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{metal\ pairs} \left(\frac{E}{r_{ij}^{12}} - \frac{F}{r_{ij}^{10}} \right) +$$

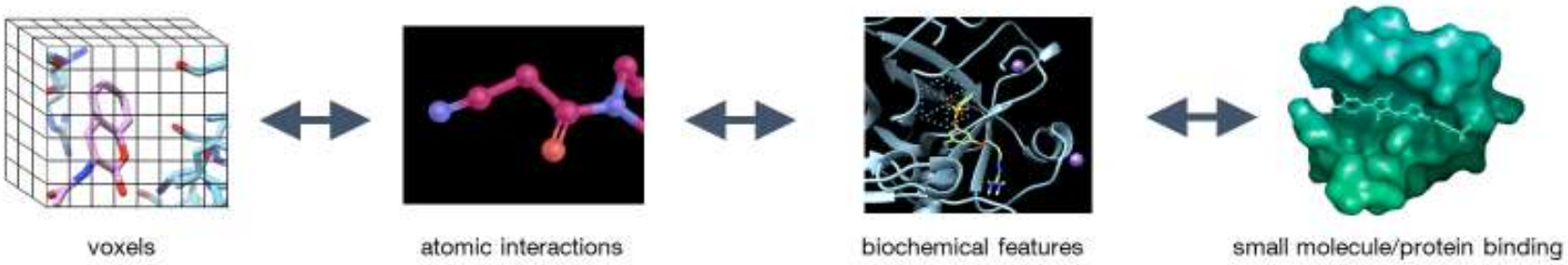
$$(E_{MC} + E_{LFS}) \cdot \prod_{angles} \cos^2(\Psi_{Lig-Met-Lig'} - \Psi_{eq}) \cdot \frac{1}{\pi} \sum_{ligands} \cos^n(\omega_{Met\cdots Lig-LP})$$

Building off prior achievements in image recognition …

AtomNet: a 3D-CNN for ligand discovery

AtomNet applies similar technology to that used in image recognition for predicting small molecule/protein binding.





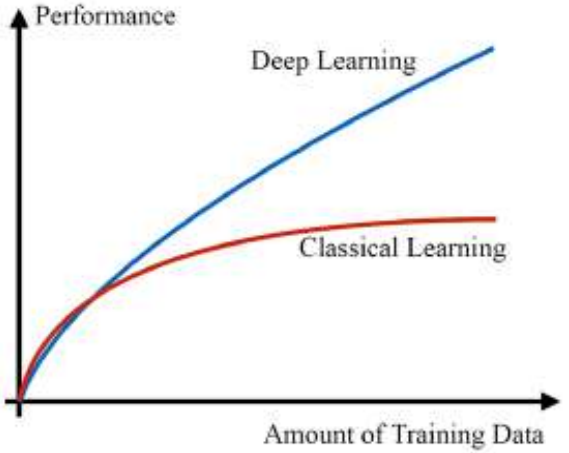
Deep Learning for Drug Discovery

What are the advantages of using a machine learning model?

- AtomNet® models use state of the art deep convolutional neural nets (DCNN)
- These models display best-in-class performance compared to other classical statistical approaches implemented in conventional scoring functions.

- In general, DCNNs are excellent at finding generalizable predictive patterns compared to classical statistical models.[1]
- Statistical models are good at drawing inferences from samples but are inferior to machine learning algorithms in prediction.[1]
- Data has shown that the performance of deep learning methods increases logarithmically based on volume of training data size compared to classical methods.[2,3]

- For further reading, please refer to:
  1. Nat Methods. 2018, 15, 233-234. DOI: 10.1038/nmeth.4642
  2. IEEE Int Conf Comput Vis., 2017, 843-852. DOI: 10.1109/ICCV.2017.97
  3. IEEE Trans Commun. 2019, 67, 7331-7376. DOI: 10.1109/TCOMM.2019.2924010

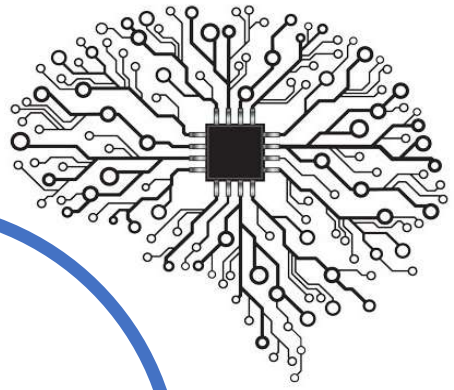


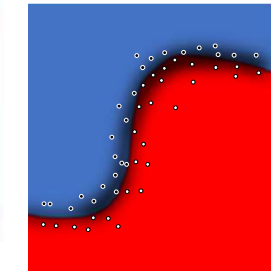
Classical and Deep learning vs. training set size [3]

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Leveraging AI to discover new medicines

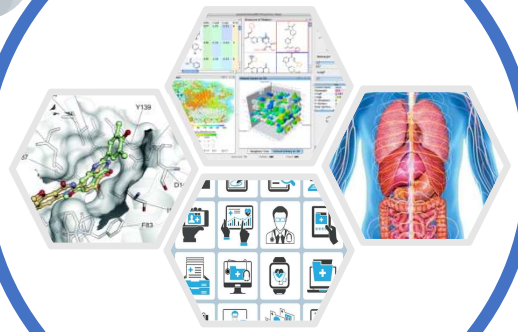
Doctors, Drug Hunters
& Data Scientists





Machine Learning

Data



amazon web services

Scalable Compute

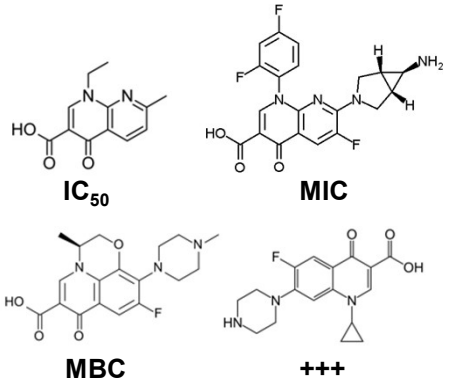
$$$

Capital

Laboratory

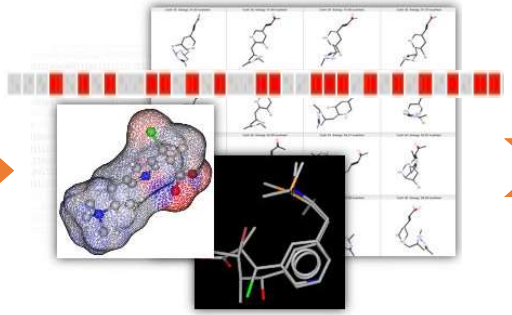
Advances in ligand-based predictive modeling

Training Data



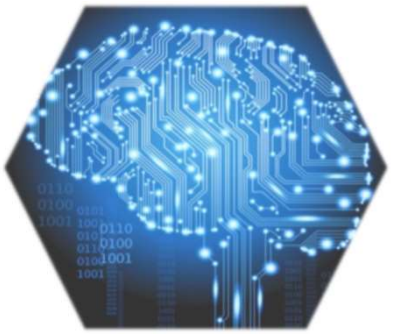
Featurize

Representation



ML

Statistical Model



Advantage

Generality

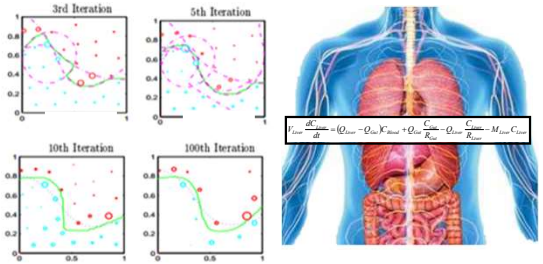
Challenges

Representational complexity

The data!

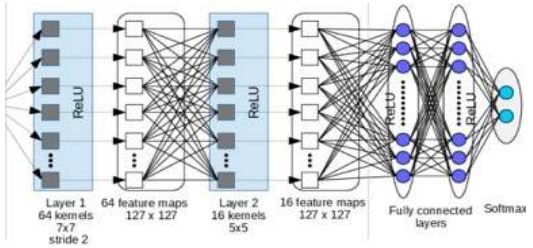
- Noisy
- Biased
- Small

Representations & Machine Learning: Matched to Problem & Data



High-D SAR

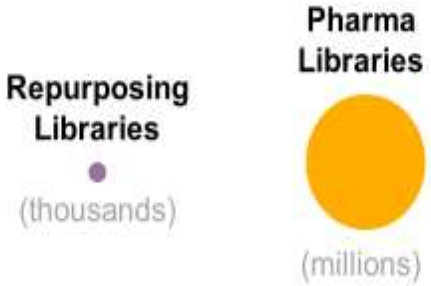
Mechanistic Multi-Task



Deep Neural Networks

Advantage of Scale

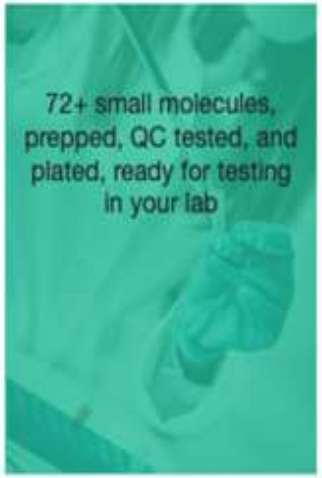
Future leads hidden in enormous chemical space



16+ Billion unique and orderable compounds

AIMS Awards Program

Artificial Intelligence Molecular Screen (AIMS) Awards provide valuable access to AI-based drug design to academic research labs.

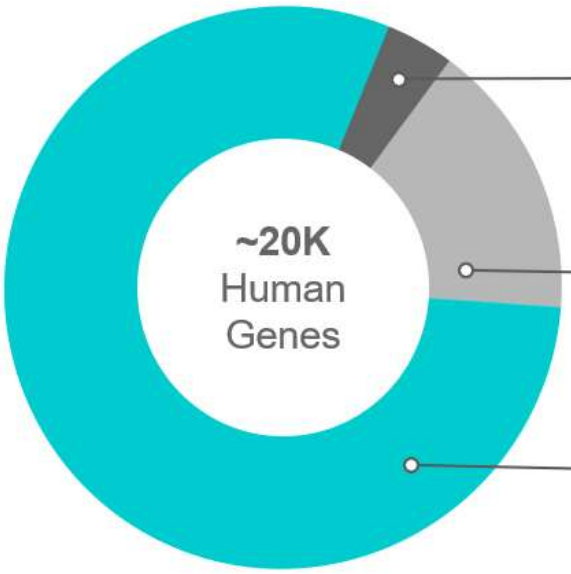


Atomwise

© 2020 Atomwise

Drugging the Undruggable

96% human genes represent the future of Drug Discovery



Low Hanging Fruit
Only 4% have FDA-approved drugs

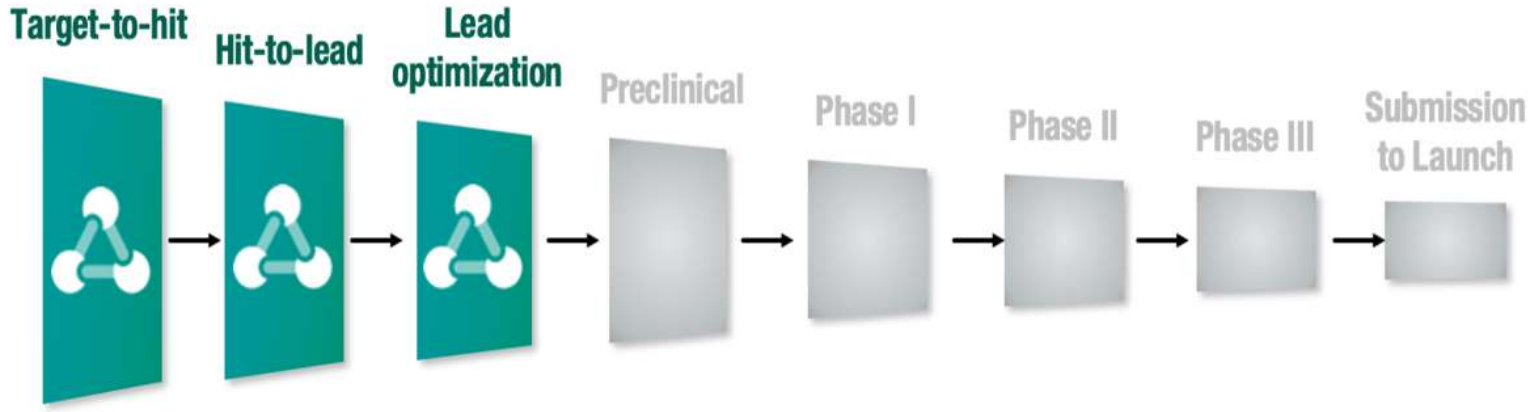
16% Targets with Disease Evidence

80% Remaining Human Targets

96% Undiscovered space
The most challenging targets, which AI can solve.

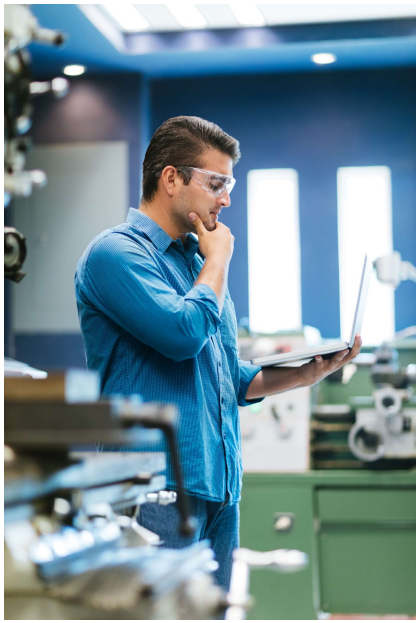
Leading AI Technology for Drug Discovery

Faster and cheaper small molecule drug design



CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Poll the Audience!



CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Have a question to ask our panel?
Open the ZOOM Q&A and type in your question at anytime !

We'll be sure to save time for them later!

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Today's Moderator and Panelists

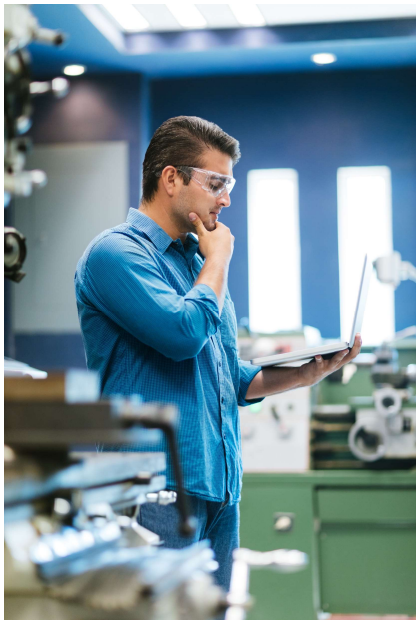
John Overington, Ph.D.
Chief Informatics Officer,
Medicines Discovery Catapult

John Griffin, Ph.D.
Vice President, Integral Health

Jeff Warrington, Ph.D.
Senior Scientist, Atomwise

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Poll the Audience!



CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Today's Moderator and Panelists

John Overington, Ph.D.
Chief Informatics Officer,
Medicines Discovery Catapult

John Griffin, Ph.D.
Vice President, Integral Health

Jeff Warrington, Ph.D.
Senior Scientist, Atomwise

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Some interesting developments in the field

Ranking-based learning

(12) **United States Patent**
Duffy

(10) **Patent No.: US 7,702,467 B2**
(45) **Date of Patent: Apr. 20, 2010**

(54) **MOLECULAR PROPERTY MODELING USING RANKING**

(75) Inventor: **Nigel P. Duffy**, San Francisco, CA (US)

(73) Assignee: **Numerate, Inc.**, San Bruno, CA (US)

Wang et al. "1-(5-Chloro-2-alkoxypheny)-3-(5-cyanopyrazi-2-yl)ureas as Potent and Selective Inhibitors of Chk1 Kinase: Synthesis, Preliminary SAR, and Biological Activities." *Journal of Medical Chemistry*. 2005 vol. 48(9): pp. 33118-3121.

Lerche et al. "A Comparison of Partial Order Technique with Three Methods of Multi-Criteria Analysis for Ranking of Chemical Sub-

Multi-purposing data

Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery

Jaak Simm,[1,8] Günter Klambauer,[2,8] Adam Arany,[1,8] Marvin Steijaert,[3] Jörg Kurt Wegner,[4] Emmanuel Gustin,[4] Vladimir Chupakhin,[4] Yolanda T. Chong,[4] Jorge Vialard,[4] Peter Buijnsters,[4] Ingrid Velter,[4] Alexander Vapirev,[5] Shantanu Singh,[6] Anne E. Carpenter,[6] Roel Wuyts,[7] Sepp Hochreiter,[2,9] Yves Moreau,[1,9] and Hugo Ceulemans[4,9,10,*]
[1]ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium
[2]Institute of Bioinformatics, Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria
[3]Open Analytics NV, Jupiterstraat 20, 2600 Antwerp, Belgium
[4]Janssen Pharmaceutica NV, Turnhoutseweg 30, 2340 Beerse, Belgium
[5]Facilities for Research, KU Leuven, Willem de Croylaan 52c, Box 5580, 3001 Leuven, Belgium
[6]Imaging Platform, Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, MA 02142, USA
[7]ExaScience Life Lab, IMEC, Kapeldreef 75, 3001 Leuven, Belgium
[8]These authors contributed equally
[9]Senior author
[10]Lead Contact
*Correspondence: hceulema@its.jnj.com
https://doi.org/10.1016/j.chembiol.2018.01.015

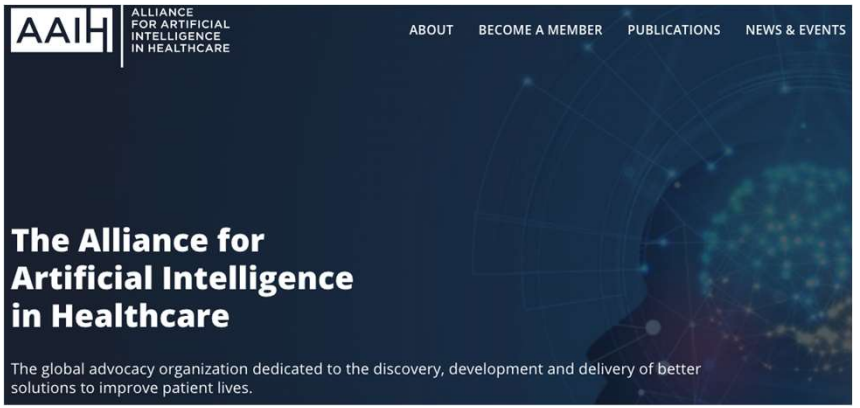
Integrating downstream

ARTICLE
https://doi.org/10.1038/s41467-019-11069-0 OPEN

Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings

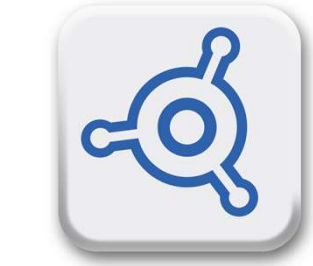
Charlotte A. Nelson [1], Atul J. Butte [2,3] & Sergio E. Baranzini [2,4]

Alliance for Artificial Intelligence in Healthcare (AAIH)



CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Questions?



CDD VAULT®
Complexity Simplified

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Today's Moderator and Panelists

John Overington, Ph.D.
Chief Informatics Officer,
Medicines Discovery Catapult

John Griffin, Ph.D.
Vice President, Integral Health

Jeff Warrington, Ph.D.
Senior Scientist, Atomwise

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

CDD Platforms - Enabling Modern Research Informatics

CDD VAULT®
Complexity Simplified™

Activity & Registration
Store and organize your data

ELN
Document all your research

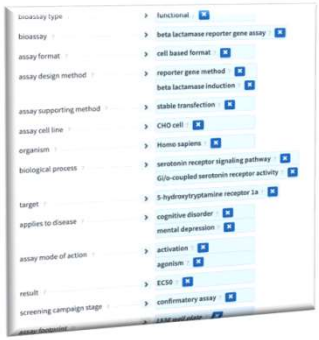
Inventory
Keep track of compounds

Visualization
Plot datasets and mine them

BioHarmony
DATA SIMPLIFIED

Content Store
is your centralized source for semantic drug data

Drug
Target
Disease
Clinical
Research
Metadata
Financial
Commercial
Timeline



Annotator
(formerly BioAssay Express)
automatically converts your assay metadata into semantic content

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

To find out more about CDD Vault and CDD BioHarmony and Bioharmony Annotator please contact us at:

info@collaborativedrug.com

CDD VAULT®
COLLABORATIVE DRUG DISCOVERY

Upcoming Events at CDD

Virtual Exhibition & Presentation — September 15-18
Discovery on Target 2020

Virtual Symposium — October 13
Informatics for Effective Drug Discovery
Presented by Optibrium StarDrop™ & CDD Vault

Webinar - BioIT World — November 5
Data Management & Analysis for Drug Discovery *[Tentative]*
Presented by Certara D360™ & CDD Vault

More Information:
collaborativedrug.com/blog/

CDD VAULT®
Complexity Simplified