Exploiting Bigger Data and Collaborative Tools for Predictive Drug Discovery

Sean Ekins, CSO CDD

This guy deserves a coffee

CDD website 2010-2013

Log in to CDD Vault Home Features News & Events About Us Sign Up

Streamline drug discovery with CDD's collaborative, web-based software.

CDD Vault is a web application for intelligent data management and secure collaboration. CDD makes drug discovery easier and more efficient for thousands of industry and academic scientists around the world.

Learn more Sign up now

Simple and secure data management

Securely store and easily mine experimental data, including bioassays and chemical structures, in a private data vault that CDD hosts for your group.



Perkin Elmer in Laboratory Informatics Guide 2014



Scientific breakthroughs don't happen every day. But with the TIBCO Spotfire[®] platform, they just might happen more often. TIBCO Spotfire[®] software visualizes your data in stunning, impactful analyses—giving you immediate insights and even answering questions you haven't posed yet.

With such a powerful tool, you can take complete control of your data:

- Combine data from multiple sources—chemical structures, text, numbers, images, chemical properties, biological assays, and more.
- Identify new relationships, isolate significant outliers, and easily spot trends and patterns.

TIBCO Spotfire® software connects to data sources with just a few clicks—and runs visualizations instantly. It's that easy.



Download your free trial and get started on your next breakthrough today. www.perkinelmer.com/spotfire-scientificcomputingworld





OPEN SCIENCE DATA CLOUD

Cloud Services for the Scientific Community

Science Clouds

Home Blog Clouds Ecosystem Appliances Contact



"by provisioning the right amount of storage and compute resources, cost can be significantly reduced with no significant impact on application performance"

CDD's Influence spreads beyond the cloud

2004 - present



 Home CDD Vault Community Blog About Us

CDD Vault®

Finally, a modern approach to drug research informatics

CDD Vault is a hosted biological and chemical database that securely manages your private and external data. Accomplish more with an intuitive solution designed by scientists.



A SYSTEM YOUR ENTIRE TEAM CAN ACTUALLY USE

Intuitive web interface
Economical cloud deployment
Biologists and chemists interacting



BEATS JUGGLING SPREADSHEETS OF SCREENING DATA Eliminates the risk of data loss
Unified data yields better results
Easier to find, analyze, and share data



2014

ScienceCloud by Accelrys Transforms Externalized Drug Discovery

New cloud solution for externalized life science research and development redefines collaboration and creativity

San Diego, CA, Feb. 6, 2014 / PRNewswire / — Accelrys, Inc. (NASDAQ: ACCL), a leading provider of scientific innovation lifecycle management software,



Transform Scientific Collaboration

ransform the way you run your externalized R&D project: multiple follaborators, single platform, one conversation.

Empower your virtual teams with a new generation of integrated, cloud-based applications built on a world-class scientific platform.



SOLUTIONS COSTOMENS SUPPORT ABOUT





Big Data, Life Sciences, and National Security

APR

New data science institute to help scholars harness 'big data'

By Robert Sanders, Media Relations | November 13, 2013

BERKELEY — In a world awash in data, UC Berkeley is meeting the flood head-on by establishing a new institute to support faculty, researchers and students in their efforts to mine this information in areas as diverse as astronomy and economics, genetics and demography.

LXR Small Data circa late 1990's

 $K_{\rm m(apparent)}$ for metabolism of substrates for expressed CYP2B6





Se 💷 🗴



How you dispense liquids may be important: insights from small data





Generated with Discovery Studio (Accelrys)

Cyan = hydrophobic

Green = hydrogen bond acceptor

Purple = hydrogen bond donor

Each model shows most potent molecule mapping



Future: sharing chemical relationships without structures

Private Data Shared Data Scaffold Groups Neighbors Molecule IDs M2 M2 M3 M3 M3 M1 M1 M7 M4 M7 M4 Active M8 M5 M8 M5 M8 M5 M9 Inactive M9 M6 M M6 \bigcirc M12 M12 M12 Structures Relationships Structures M11 M11 M11 M10 M10 M10 Relationships Activity Data Activity Data Activity Data compound set enrichment statistical corrections local hit rate ontology pattern identification marginal cost of discovery nearest neighbor clique-oriented prioritization workflow inference diversity-oriented prioritization Level 3 **R-Group Network** Scaffold Tree Scaffold Network M2 M2 M3 M3 M1 M1 M2 M5 R2 Level 2 M3 M6 R3 M7 M4 M7 M4 R4 M8 M8 M5 M5 MQ M9 R5 M6 M10 M6 M7 R6 M11 M8 M12 M12 M9 M12 M11 M11 M10 M10 Level 1 matched-pair analysis tree visualization network visualization Anonymized Candidate R-group analysis imputed trees imputed networks structure-activity relationships Scaffolds Network

Matlock and Swamidass J. Chem. Inf. Model. 2014, 54, 37–48

Drug Discovery Archeology

- Still a heavy emphasis on "testing" "doing " rather than 'learning'
- Mining data and historic data will increase in value
- Data becomes a repurposing opportunity
- How do we position databases for this?
- What about neglected diseases?



what loc skoute enow slout

Multi drug resistance in 4.3% of cases Extensively drug resistant increasing incidence one new drug (bedaquiline) in 40 yrs

PREVENT DISEASE



streptomycin (1943) *para-*aminosalicyclic acid (1949) isoniazid (1952) pyrazinamide (1954) cycloserine (1955) ethambutol (1962) rifampicin (1967)

TUBERCULOSIS

Tuberculosis kills 1.6-1.7m/yr (~1 every 8 seconds) 1/3rd of worlds population infected!!!!





Freundlich Laboratory Collaborations Rely on CDD for Data Tracking!



Drug Discovery Compound **Evolution** CDD VAULT Target **Chemical Probe** Identification & Validation **Evolution**

Three collaborations within Rutgers–NJMS

- Collaboration with Johns Hopkins, SRI, and CDD
- Collaboration with Johns Hopkins
- Collaboration with CDD

Supported by 7 Active NIH Grant



discoverv

Can we better understand the fundamental biology of M. tuberculosis to seed the discovery of novel



MM4TB



More Medicines for Tuberculosis

24 groups in this project use a single CDD VAULT





Godbole et al., Biochem Biophys Res Comm 2014, in press

More Medicines for Tuberculosis (MM4TB)

Fishing: Example of mimic strategy for bioB Rv1589

Biotin biosynthesis

Take substrate and generate 3D conformers and build a pharmacophore

Use the pharmacophore to search vendor libraries in 3D

Buy and test compounds









Searching Maybridge (57K) gives 72 molecules – many of them hydrophobic so they stand a chance of in vitro activity

Pharmacophore

Sarker et al., Pharm Res 2012, 29:2115-27

Over 5 years analyzed in vitro data and built models

High-throughput Mtb screening phenotypic Bayesian Machine Learning classification *Mtb* Model molecule *Mtb* screening database/s Descriptors + Bioactivity (+Cytotoxicity) $p(h \mid d) = \frac{P(d \mid h)P(h)}{P(d)}$ н Molecule Database (e.g. GSK malaria actives) virtually scored using Bayesian Models Top scoring molecules New bioactivity data assayed for may enhance models Ekins et al., Pharm Res 31: 414-435, 2014 Mtb growth inhibition Ekins, et al., Tuberculosis 94; 162-169, 2014 Ekins, et al., PLOSONE 8; e63240, 2013 Ekins, et al., Chem Biol 20: 370-378, 2013 Identify in vitro hits and test models Ekins, et al., JCIM, 53: 3054-3063, 2013 Ekins and Freundlich, Pharm Res, 28, 1859-1869, 2011 Ekins et al., Mol BioSyst, 6: 840-851, 2010 Ekins, et al., Mol. Biosyst. 6, 2316-2324, 2010, 8-10X enrich-

3 x published prospective tests >20% hit rate Multiple retrospective tests 3-10 fold enrichment

A summary of some of the numbers involved – filtering for hits. >250,000 molecules screened through Bayesian models ~750 molecules were tested in vitro **198 actives were identified** >20 % hit rate Identified several novel potent hit series with good cytotoxicity & selectivity Identified known human kinase inhibitors and FDA approved drugs as hits

Ekins et al., PLOSONE 2013 May 7;8(5):e63240; Ekins et al.,Chem Biol 20, 370–378, 2013 Ekins et al., Tuberculosis 94: 162-169

Taking a compound in vivo identifies issues



HN

compound18

MIC = 0.25 mg/mL vs. "wild type"

in vivo activity = +++ w/ INH @ ++++

• BAS00521003/ TCMDC-125802 reported to be a *P. falciparum* lactate dehydrogenase inhibitor

- Only one report of antitubercular activity from 1969
 - solid agar MIC = 1 μg/mL ("wild strain")
 - "no activity" in mouse model up to 400 mg/kg
 - however, activity was solely judged by extension of survival!



64X MIC affords 6 logs of kill

 Resistance and/or drug instability beyond 14 d

Vero cells : $CC_{50} = 4.0$ µg/mL

Selectivity Index SI = $CC_{50}/MIC_{Mtb} = 16 - 64$

In mouse no toxicity but also no efficacy in GKO model – probably metabolized.

Bruhin, H. et al., J. Pharm. Pharmac. 1969, 21, 423-433.

Ekins et al., Chem Biol 20, 370–378, 2013



Big Data: Screening for New Tuberculosis Treatments

gsk

National Institute of Allergy and Infectious Diseases Leading research to understand, treat, and prevent infectious, immunologic, and allergic diseases.

SOUTHERN RESEARCH

Tested >350,000 molecules >1500 active and non toxic Tested ~2M2MPublished 177100s

Others have likely screened another 500,000

U NOVARTIS

BRO

>300,000

800

How many will become a new drug? How do we learn from this big data?

Hunting High and Low for new molecules to test

We need to search

From the Oceans...

SO

To the ground To the trees To the air.

And do it virtually

Find new libraries

to screen virtually

and test

Take everything out of CDD public

- Run through TB Bayesian models
- Score

• Test



What is the next bottleneck?









Five-fold increase in the publication of TB mouse model studies from 1997 to 2009 Franco, *PLoS One* 7, e47723 (2012).

Billions of \$ of your money spent on TB but no database of mouse *in vivo* data !



Hunting for the *in vivo* data

It's out there.. be patient

Building the mouse TB database

Manually curated, structures sketched Mobile Molecular DataSheet (MMDS) iOS app or ChemDraw (Perkin Elmer)

Downloaded from www.chemspider.com

Combined with pertinent data fields 1 log₁₀ reduction in *Mtb* colony-forming units (CFUs) in the lungs

Publically available CDD TB database (In process)

30 years with little TB mouse in vivo data

6.12



Where are the New TB drugs to be found?

Low Shawno

Longitude West from London

85

2 25 Longit M. from Form





20

90

95

Machine Learning Models

Bayesian Support Vector Machine Recursive partitioning (single and multiple trees) Using Accelrys Discovery Studio and R.



JCIM In Press 2014

External Test set

11 additional active molecules obtained from 1953-2013

RP Forest	RP Single	SVM	Bayesian
1.4716	Tree	S.:	1 1
3 /11	4/11	7/11	8/11

778 Selected: 🚧 Plot - 🕞 Export - 🗞 Add to collection 🔯 Customize your report										
Select		Chemical Properties		TB mouse in vivo data from the literature						
all < none	Molecule 🗢	Molecular weight (g/mol) ≑	Mouse model 🌻	Treatment 🗢	Results ≑	In vivo activity 🗢	Activity Score 🗢	PubMed Id 🗢	Year 🗘	
	fiag outliers $ \begin{array}{c} & \mu_{\mu} \subset \begin{pmatrix} c_{\mu} \\ + \\ + \\ c_{\mu} \\ + \\ c_{\mu} \\ + \\ c_{\mu} \\ + \\ c_{\mu} \\ c_{\mu} \\ - \\ c_{\mu} $	416.449	GKO mouse model	22 days	Lung < 1 log CFU reduction, 25 mg/kg	Inactive	0	Phillips et al. 2012	2012	
	fire suffices	154.454	01/0	00 daug	Luna ed ha	la a alfred		Dialling of all	2042	
		451.454	GKU mouse model	22 days	Lung < 1 log CFU reduction, 25 mg/kg	inactive	0	Philips et al. 2012	2012	

A much higher ratio of compounds were tested in vivo to in vitro in the 1940s-1960s rather than now

The Clock is ticking

Infrastructure to provide a clear understanding of the position of compounds in the pipeline is essentially lacking

Shortage of new candidates suggest we may lack the commitment and resources we had 60 years ago

Use machine learning in vivo models to prioritize Mouse studies

.....

Ekins, Nuermberger & Freundlich Submitted

Abstract

Ce

Enhancing Hit Identification in Mycobacterium tuberculosis Drug Discovery Using Validated Dual-Event **Bayesian Models**

Sean Ekins^{1,2}*, Robert C. Reynolds³", Scott G. Franzblau⁵, Baojie Wan⁵, Joel S. Freundlich^{4,6}, Barry A. Bunin¹

1 Colaborative Drug Discovery, Burlingene, California, United States of America, 2 Colaborations in Overnistry, Fuquay-Vierina, Noth Carolina, United States of America, 3 Southern Research Institute, Nirmingham, Alabama, United States of America, 4 Department of Pharmacology & Physiology, UMDNU - New Jensey Medical School Newark, New Jenery, United States of America, 5 Institute for Tubercalosis Research, University of Illinois at Chicago, Olicago, Illinois, United States of America, 6Department of Medicine, Center for Emerging and Reemerging Pathogens, UMDNJ - New Jensey Medical School, Newale, New Jensey, Linited States of America



High-throughput screening (HTS) in whole cells is wild y pursued to find compounds active against Mycobacterium tuberculars (Mb) for further development towards now tuberculars (TB) drugs. Ht rates from these screens, usuall conducted at 10 to 25 µM constructions, typically range from less than 1% to the low single digits. New approaches to increase the efficiency of hit identification are urge ntly needed to learn from past scree



DADED

PLOS ONE

A collaborative database and computational models for tuberculosis

www.rsc.org/molecularbiosystems | Molecular Bio

Analysis and hit filtering of a very large library of compounds screened

st tuberculosis which annually claims

abit the growth of Mtb in vitro. We have collected data from

ntify potential leads is to screen in vitro small molecules

quently, it has been difficult to analyze molecular

intil recently there was no central repository to collect

ilosis

Jeremy Yang,8 Nicko Goncharoff,4 Moses M. Hohmana and Barry A. Bunin

publically available sources on over 300 000 small molecules deposited in the Collaborative Drug

Discovery TB Database. A cheminformatics analysis on these compounds indicates that inhibitors

of the growth of Mth have statistically higher mean logP, rule of 5 alerts, while also having lower

HBD count, atom count and lower PSA (ChemAxon descriptors), compared to compounds that

are classed as inactive. Additionally, Bayesian models for selecting Mtb active compounds were evaluated with over 100 000 compounds and, they demonstrated 10 fold enrichment over random

for the top rank ed 600 compounds. This represents a promising approach for finding compounds

active against Mtb in whole cells screened under the same in vitro conditions. Various sets of Mtb

industry to identify compounds with potentially reactive moieties. We found differences between

the number of compounds flagged by these rules in Mtb datasets, malaria hits, FDA approved

drugs and antibiotics. Combining these approaches may enable selection of compounds with

hit molecules were also examined by various filtering rules used widely in the pharmaceutical

drug discoverv[†]

Sean Ekins,*abcd Justin Bradford," Krishna Dole," Anna Spektor," Kellan Gregory," David Blondeau," Moses Hohman" and Barry A. Bunin"

Received 28th August 2009, Accepted 21st December 2009 First published as an Advance Article on the web 9th February 2016 DOI: 10.1039/b917766c

M co

DOI: 10.1039/c0mb00104j

cte

Received 15th July 2010, Accepted 13th August 2010

nds that inf

ag, in

Pina † Elec

suppl

The search for molecules with activity against Mycohacterium tuberculasis (Mtb) is employing many approaches in parallel including high throughput screening and computational metho We have developed a database (CDD TB) to capture public and private Mtb data while enabling data mining and collaborations with other researchers. We have used the public data along with several cheminformatics approaches to produce models that describe active and inactive pounds. We have compared these datasets to those for known FDA approved drugs and between Mtb active and inactive compounds. The distribution of polar surface area and pK_a of active compounds was found to be a statistically significant determinant of activity against Mtb. Hydrophobicity was not always statistically significant. Bayesian classification models for 220 463 molecules were generated and tested with external molecules, and enabled the

Pharm Res (2012) 29:2115-2127 DOI 10.1007/r11095-012-0741-5

RESEARCH PAPER

www.rsc.org/molecularbiosystems | Molecular BioSystems

Justin Bradford

Combining Cheminformatics Methods and Pathway Analysis to Identify Molecules with Whole-Cell Activity Against Mycobacterium Tuberculosis Malabita Saker - Carolyn Talcott - Reter Madrid - Sidharth Chopra - Bany A. Bunin - Gyanu Lamichhane - Joal S. Freundlich - Sean Búns Received: 31 August 2011 / Accepted: 16 March 2012 / Published online: 4 April 2012 C Springer Science+Bushes Media, 12.C 2012 STRACT the ninformatics a tubercul

combined with

their small mode.

Bayesian Models Leveraging Bioactivity and Cytotoxicity Information for Drug Discovery



*Department of Microbiology, Immundiogy and Pathology, Colorado State University, 200 Wast Lake Street, Fort Collins, CO 8 These authors contributed equally to this work

*Present address: Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, 1530 Third Av Birmingham, AL 35294-1240, USA

*Correspondence: ekinssean@yehoo.com (S.E.), freundjs@umdnj.edu (J.S.F.) http://dx.doi.org/10.1016/j.chembiol.2013.01.011

SUMMARY

Identification of unique leads represents a significant challenge in drug discovery. This hurdle is magnified in neglected diseases such as tuberculosis. We have leveraged public high-throughput screening (HTS)

data to exe	1	valid		unu:	er ag	
approach e	byi	law	1.0	els		
bioactivity i	nat	(sir	ev e	od		
as bioactiv	nd	oto	y i	nati	- N	1.5
event mode	/e	ally	991	a co	an.	1
library and	- Y	ally o	- fir	ad activ	w	

hit rates exceeding typical HTS results by one to two orders of magnitude. This initial dual-event Bayesian model identified compounds with antitubercular whole-cell activity and low mammalian cell



IN TRODUCTION

Modern drug discovery must be more time and cost efficient in discovering novel therapeutics. These challenges are felt even more significantly in the search for neglected disease treatments, where public-private partnerships coordinate drug existing HTS data, focusing on not just the few most promisi discovery with very limited resources. A prime example is tuber culosis (TB), caused by Mycobacterium tuberculosis (Mtb), and inactives. which infects approximately one-third of the world's population and results in 1.7-1.8 million deaths annually (Lienhardt et al., 2012a) New drugs active against Mtb are urgently needed to combat a pandemic heavily affected by resistance to available therapies and coinfection with HIV/ADS (Nuemberger et al., 2010). TB drug discovery is challenging, reflected in the lack of

in the dat als 840 set et al., 2012; Sacchettini et al., 2008). One response has be to screen very large compound libraries (Ananthan et al., 20) Maddry et al., 2009; Reynolds at al., 2012), hoping to dial



ortes. Fort pathogenic/ lorium smo noine hit that i clinical candidate bedaquiline (Andrias et al., 2005), where another resulted in the early-phase candidate SQ109 (L et al., 2003). Although SQ109 arose directly from a library congeners of the frontline drug athembutol, high-throughp

ina i ically does iver a clinica whatstvo t must or sidering pl afford clin following w notics, phar). The ren dates such current TB a an for purposing of scovering anti-conculars from decar ourposing of bacterials of (Lienhardt et al., 2012b). Despite these successful efforts, t expected failure of ~85% of clinical candidates (Lectord, 20

and the growth of TB drug resistance necessitate new ciril submissions, which ultimately require the discovery of no hits and leads. We assert that the TB field should further levera hits due to resource limitations but the entire data set of activ

We hypothesize that prior knowledge of Mtb actives a inactives, combined with machine learning models, can sign cantly focus compound selection and improve screening e clency (Ekins and Freundlich, 2011; Ekins et al., 2010c, 2010 2011), as practiced in the pharmaceutical industry (Prathip et al., 2008, to improve the performance of virtual screen a new TB focused therapeutic approved in over 40 years (Gros- (Schneider, 2010). These and other cheminformatics metho



sently

Drug Disc

94101 8

on.com: 1

tons In Chamistry, out Runnyn

*Ca

⁶Global Alliance for TB Drug Development, 40 Wall Street, 24th floor, New York, NY 10005, USA 10 Consulting Drive, Waterford, Connecticut, 06385, USA Division of Biocomputing, University of New Mexico, Albuquerque,

NM 87131 SureChem, The Macmillan Building, 4 Crinan Street, London UK NI 9XW

2316 | Mol. Bla5vst., 2010, 6, 2316-2324

Electronic supplementary information (ESI) available: Supplementary figures and methods. The Bayesian module crusted in Discovery Studio are available from the authors upon written request. See DOI: 10.1039/ dm100104i

carried out in various organizations before, until recently there has not been any central depository of the screening results. However, during the last two years there have been changes as

BCG at



Institute (SRI) on over 100 000 compounds purchased from commercial sources (Tuberculosis Antimicrobial Acquisition and Coordinating Facility, TAACF-NIAID-CB25). Our initial analysis of an earlier TAACF set, NIAID, GVKbio, MLSMR and Ballel datasets (Table 1) suggested that the mean value for various molecular descriptors is statistically different to that of FDA approved drugs,3 The mean value of polar surface area descriptors was frequently higher in active compounds, compared to the inactive molecules across four datasets. The molecular weight, logP, and Lipinski score were statistically significantly higher in the most active compounds in the MLSMR screening data, while the PSA is slightly lower compared to the inactive compounds. This analysis

Tuberculosis Computational Models with Public ening Aerobic Activity Datasets

INTRODUCTION

culos

appr

ted: 25 February 2011 / Published online: 10 March 2011 LLC 2011

notecules with activity against noteasingly uses high throughethods. Several public datasets

Higher levels of failures of SMARTS filters from different groups also correlate with the number of Lipinski violations. Conclusions These computational approaches may assist in finding desirable leads for Tuberculosis drug discovery.

gir/Mth), the causative agent of tuber-

ind of the world's population and kill

ach year (1). The survival of Mtb

lar functions carried outby metab-

stim, ed to maintain latent infection in

wd s esian models - Collaborative Drug dass fingerprints - mycobacterium ritative Structure Activity Relationship berculosis database

Novatic co SMARTS alerts to identify INTRODUCTION

pounds as a test set for the >4.0-fold enrichment over in the compuwas observed s in the drugs database. r tha known TB drugs.

Mycoladerium tuboratlosis (Mtb), the causative agent of tuberculosis (TB), infects approximately one third of the world's population, and 1.7-1.8 million people die from this disease annually (1). Agents that are active against Mtb are urgently needed to combat this global epidemic that is ds faile the Abbott SMARTS heavily influenced by resistance to the available regimen of drugs, lengthy treatment, and co-infection with HIV.

d The online version of this article tains supplementary material.

> Department of Pharmaceutical Sciences, University of Maryland Baltimore, Maryland, USA

S. Birst Department of Pharmacology, Robert Wood Johnson Medical School University of Medicine & Dentistry of New Jersey Pacataway, New Jersey 08854, USA

S. Frandich Dipartment of Biochemistry and Biophysics, Tiosa A&M University College Station, Texas 778/3, USA

overney een maami tuberculosis and annually 10 million new cases of tuberculosis occur and 1.7-1.8 million people die from it.1 Almost all of the currently used drugs against Mtb were discovered more than 50 years ago and because of the emergence of resistance and



- Invented for Pipeline Pilot: public method, proprietary details
- Often used with Bayesian models: many published papers
- Built a new implementation: open source, Java, CDK
 - stable: fingerprints don't change with each new toolkit release
 - well defined: easy to document precise steps
 - easy to port: already migrated to iOS (Objective-C) for TB Mobile app
- Provides core basis feature for CDD open source model service

Same datasets -Versus published data

Dataset

In vivo data (773 molecules) FCFP_6 fingerprints

Combined model (5304 molecules) FCFP_6 fingerprints

MLSMR dual event model (2273 molecules) and FCFP_6 fingerprints Leave one out Reference Leave one out ROC Published fingerprints

this study 0.75

0.71 *J Chem Inf* 0.77 *Model* 53:3054-3063. 0.86 *PLOSONE* 0.83 8:e63240

Clark et al., submitted 2014

Open fingerprints and bayesian method used in TB Mobile Vers.2



In vitro data

In vivo data



COLLABORATIVE DRUG DISCOVERY



COLLABORATIVE DRUG DISCOVERY





ADME/Tox data & Models

Data sources and tools we could integrate

Drug-like scaffold creation

TB Prediction Tools TB Publications

Future: How can we tackle more diseases?

We could do this again for a different disease Small data: Mouse *In vivo* model data

«Tuberculosis and mouse in vivo model » ~33 papers in PubMed «Malaria and mouse in vivo model » ~314 papers in PubMed





Chagas Disease

Reverse the mimic approach to predict targets of hits

Use pharmacophpores for targets e.g. CYP51

Use machine learning models to identify novel compounds













The new faces of personalized medicine: children with rare diseases

The Rare Disease Parent Odyssey

- Diagnosis of child
- Try to find out about disease papers behind paywall
 - Try to connect with scientists
- Form not-for-profit
- Raise funds
- Fund Scientific research on disease
- Advocate for support from NIH, FDA etc
- Start a company
- Try to find a cure before its too late

Could we create a rare disease community for scientists & foundations ?



Wood J, Drug Disc Today, 18: 1043–1051, 2013

The Solution Schematic



The Open Drug Discovery Teams (ODDT) iOS App



Future: how do we deliver data Rare diseases inspired an App that may be a new kind of database

upload molecules by tweeting them-1 tweet upload

Take our data with us anywhere

Bring data off the cloud into device

Advantages you get to analyze it in the Cloud on a plane



All at CDD and many others ...Funding: 1R41AI088893-01, 2R42AI088893-02, R43 LM011152-01, 9R44TR000942-02, 1R41AI108003-01, MM4TB, Software: Accelrys

