# Data relevance in pharmaceutical industry

Davide Branduardi
Applications Scientist
Schrödinger, Inc.
London, UK

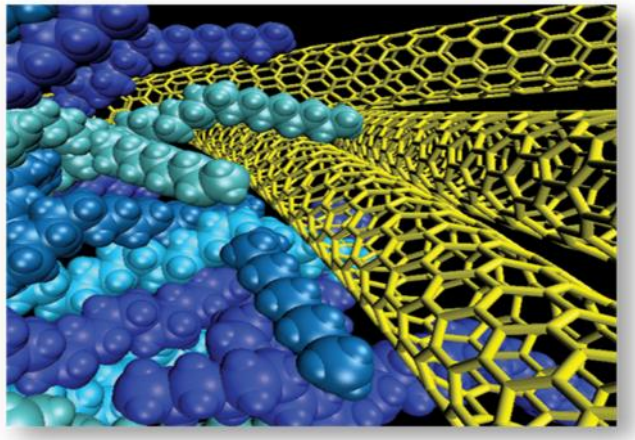
SCHRÖDINGER.

# What does Schrödinger do?

- Mission

  **Improving human health and quality of life through *advanced computational methods***

- Provides integrated software solutions and services to pharmaceutical/biotechnology and materials companies



SCHRÖDINGER.

# Who is Schrödinger?

- **Founders – Scientists from Academia**
  - Richard Friesner – Columbia University
    - Theoretical chemist focused on life sciences
  - Bill Goddard – Caltech
    - Theoretical Chemist focused on materials science

- **Investors – Patient; passionate about science**
  - David E. Shaw
    - Founder of D.E. Shaw Group, Hedge Fund
    - Chief Scientist – D. E. Shaw Research
    - Senior Research Fellow – Center for Computational Biology and Bioinformatics at Columbia University
  - Bill Gates
  - No institutional investors

SCHRÖDINGER.

# Schrödinger Offices and Business Partners


SCHRÖDINGER.

# Schrödinger contribution to structure-based drug discovery

Scientific advances in drug discovery; for example:

- **2004**: Glide – de facto standard in protein ligand docking
- **2005**: 1st reliable flexible-receptor ligand docking method (induced fit)
- **2009**: 1st rigorous treatment of protein desolvation ('hydrophobic effect')
- **2011**: Most accurate small-molecule force field
- **2014**: 1st benchmark method for accurate prediction of binding affinity

…together with a commitment in the open source visualization software Pymol.

SCHRÖDINGER.

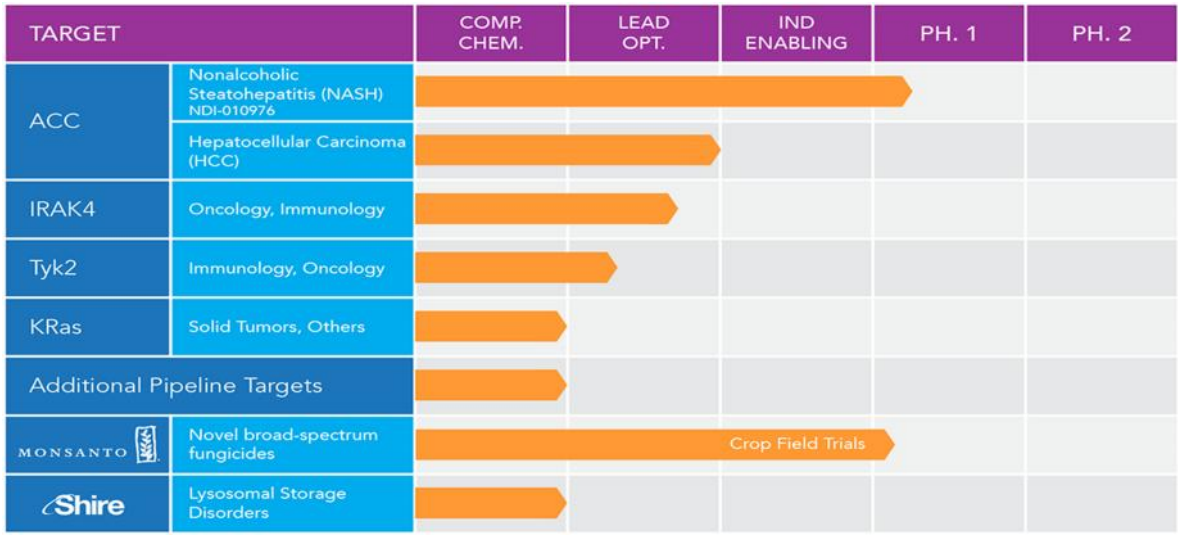
# Some Facts & Figures

- 24 Years of innovations in scientific research and product development
- ~350 employees, >55% Ph.D.
  - Scientists
  - Engineers
- Significant R&D effort and focus on customer support
  - R&D spending: ~50% of budget
  - Development: ~50% of employees
  - Internal Drug Discovery: ~10% of employees
  - Customer Support: ~15% of employees
- Revenue is reinvested in research and development
- Focus on discovery software & services for small molecules, biologics, and materials science
- Customers: 380 commercial (including all top 30 Pharma companies); 2100 academic; 130 government

SCHRÖDINGER.

# Nimbus Therapeutics

- Nimbus is pioneering a new computational technology-driven paradigm to rapidly advance a diverse pipeline into clinical development
- $72 Million from 7 Investors
  - Including Atlas Venture, Bill Gates, and Pfizer Ventures
- Schrödinger is a founding partner (please refer to www.nimbustx.com for the up –to-date information)




NIMBUS THERAPEUTICS

NIMBUS THERAPEUTICS

Home Contact Us Careers Search

Our Approach Our Pipeline Our Network News & Events Abo

Overview
Press Releases
Publications & Abstracts

# GILEAD SCIENCES ANNOUNCES ACQUISITION OF NIMBUS THERAPEUTICS' ACETYL-COA CARBOXYLAS (ACC) PROGRAM FOR NASH AND OTHER LIVER DISEASES

FOSTER CITY, Calif. & CAMBRIDGE, Mass., April 4, 2016—Gilead Sciences, Inc. (NASDAQ: GILD) and Nimbus Therapeutics, LLC today annour
that the companies have signed a definitive agreement under which Gilead will acquire Nimbus Apollo, Inc., a wholly-owned subsidiary of
Therapeutics will receive an upfront payment of $40
milestones over time.

SCHRÖDINGER.

Home Product Suites Support

Home › News & Events › News

## News

April 4th, 2016

**Schrödinger congratulates its collaboration partner, Nimbus Therapeutics, on the acquisition by Gilead of Nimbus's ACC program for up to $1.2 billion**

Gilead Sciences, Inc. (NASDAQ:GILD) and Nimbus Therapeutics, LLC, announced that the companies have signed a definitive agreement under which Gilead will acquire Nimbus Therapeutics' Acetyl-CoA Carboxylase (ACC) inhibitor program for up to $1.2 billion.

Schrödinger is a co-founder and equity holder of Nimbus and provided the computational platform and support for the ACC program. For more information, please read the **press release**.

Forbes / Pharma & Healthcare

Join Us For An Exclusive Webcast on Social Security

APR 4, 2016 @ 08:33 AM 5,191 VIEWS

# Atlas-Backed Nimbus Delivers Its Apollo Mission: A $1.2B Gilead Partnership

**Bruce Booth,** CONTRIBUTOR

*I write about emerging life-science startups and the industr*

FOLLOW ON FORBES (281)

Opinions expressed by Forbes Contributors are their own.

FULL BIO

Today Gilead announced the acquisition of Nimbus' l
program targeting NASH and related metabolic disor
$400M upfront and up to another $800M in develop
regulatory milestones (here). This transformative dea
culmination of over five years of work on the progran

Biotech

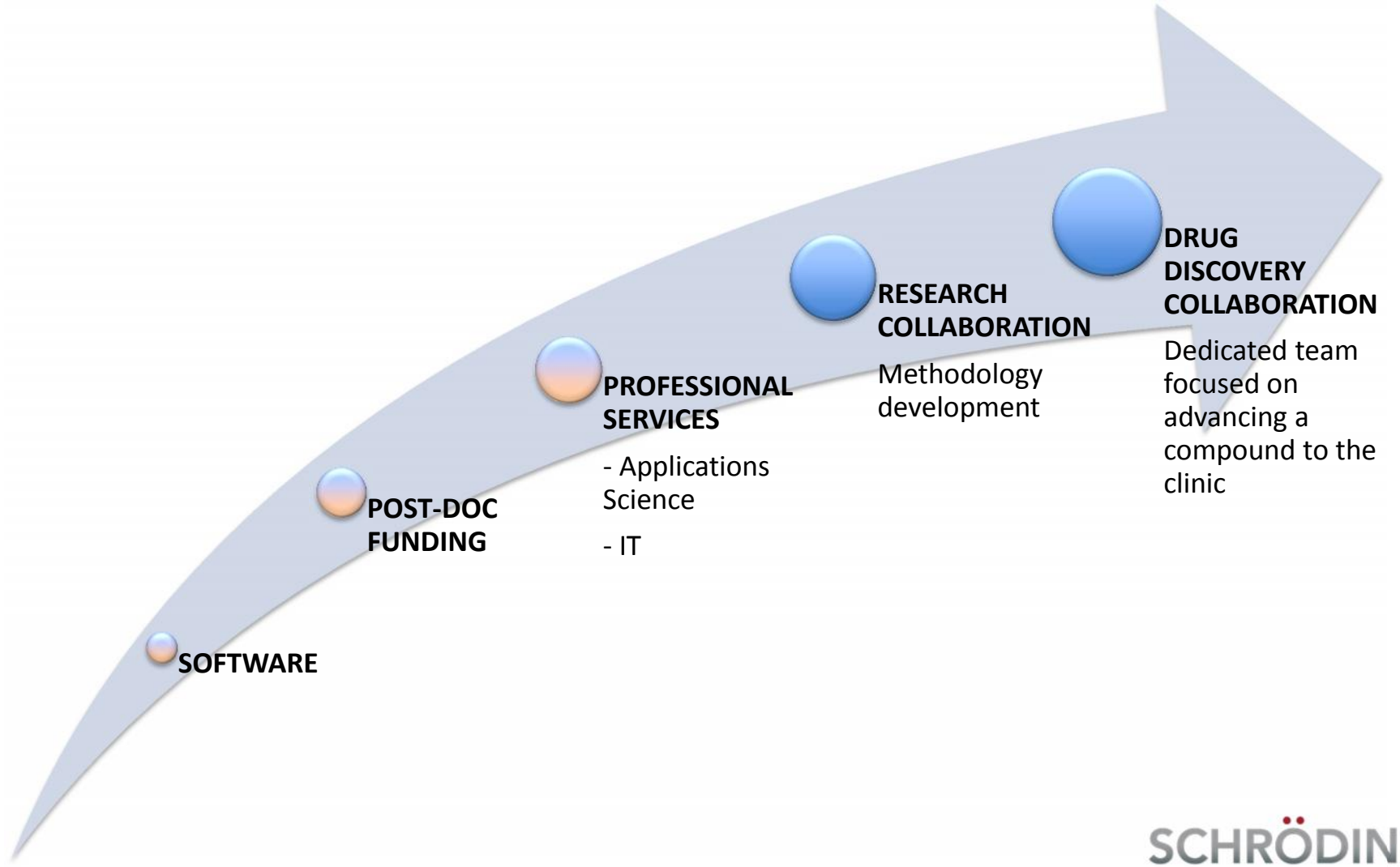
# Gilead pays Nimbus $200M after fast progress in NASH

by Nick Paul Taylor Nov 2, 2016 6:24am

Six months after Gilead paid Nimbus Therapeutics $400 million to buy a NASH drug, it has handed over another $200 million milestone payment.

SCHRÖDINGER.

# Schrödinger works operates in many ways




SCHRÖDINGER.

# Outline

- How a drug works and how is identified
- Pharma industry and data generation
  - What kind of data pharma industry generates
  - R&D issues: integration data challenge
  - Productivity
- Smart use of in-house data
- Smart use of external data
- A look to the future

SCHRÖDINGER.

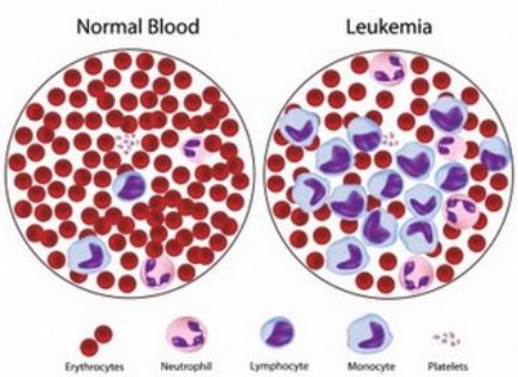
# Data in Pharma

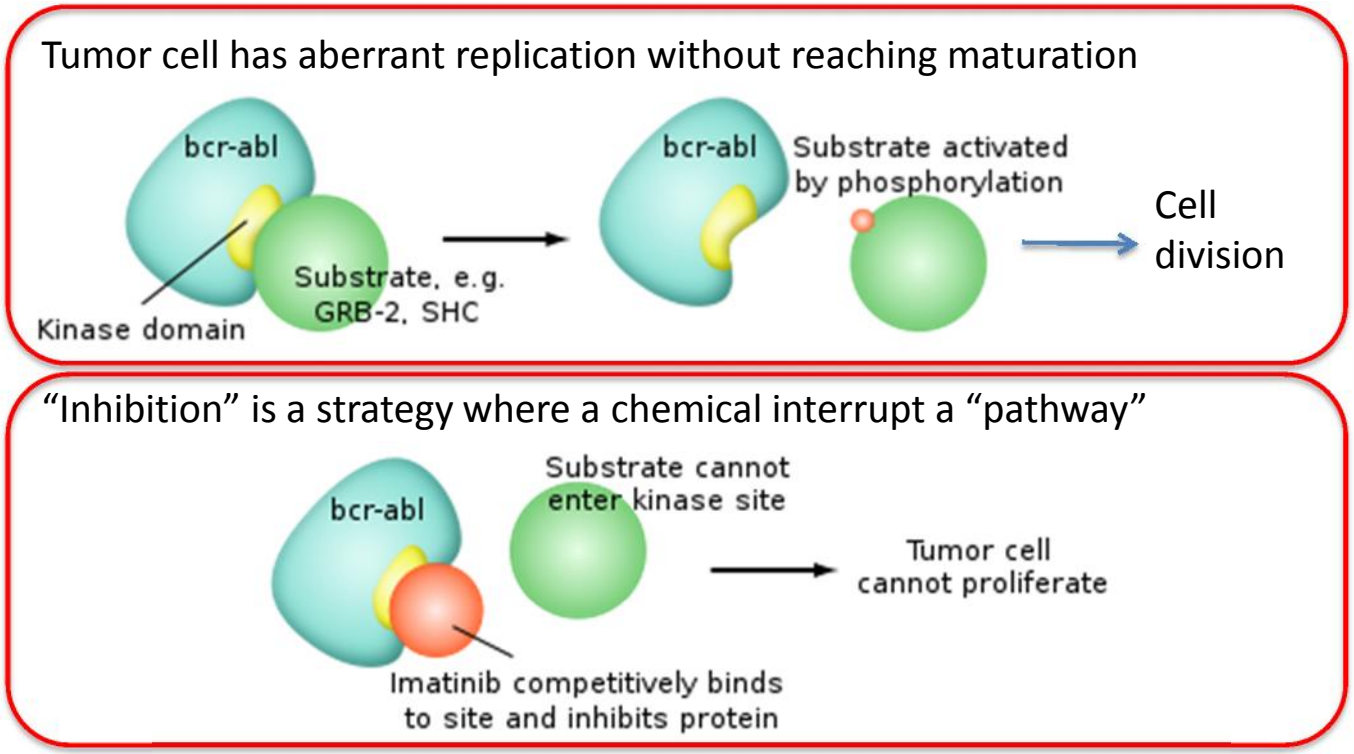
- Pharma is an interesting example of data science
  - **Research data** on drugs is very private. Attempts and failures are kept hidden for competition and driving stock price. But acceptance from the specialists happen on public
  - **Production data** is very open: regulatory agencies may want companies track batch numbers and difformities (e.g. In 2012-2013 flu pandemic, production was not effective for a change in production standard or Quinavaxem on hold in 2010)

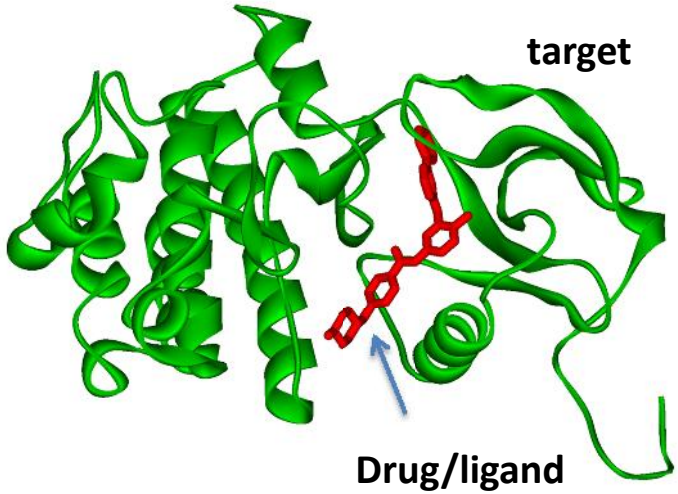
http://www.who.int/immunization_standards/vaccine_quality/outcome_quinvaxem_investigation_february_2011/en/

# How a drug works?

- An example: Chronic Myeloid Leukemia
- We haven't built our human body: finding mechanism (pathways) is hard



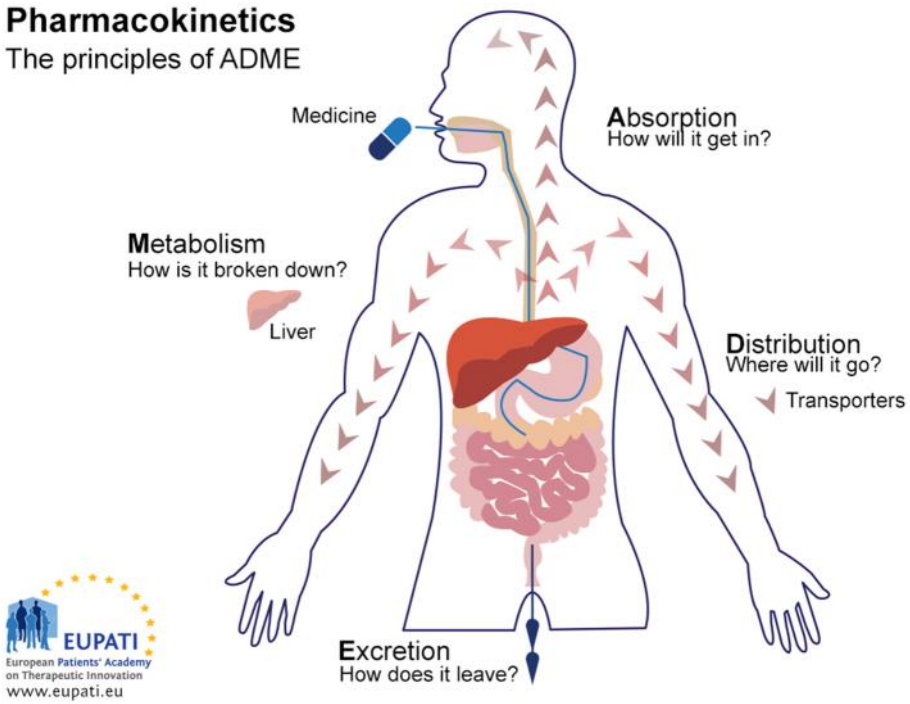




SCHRÖDINGER.

# How a drug works? From your mouth to the cell

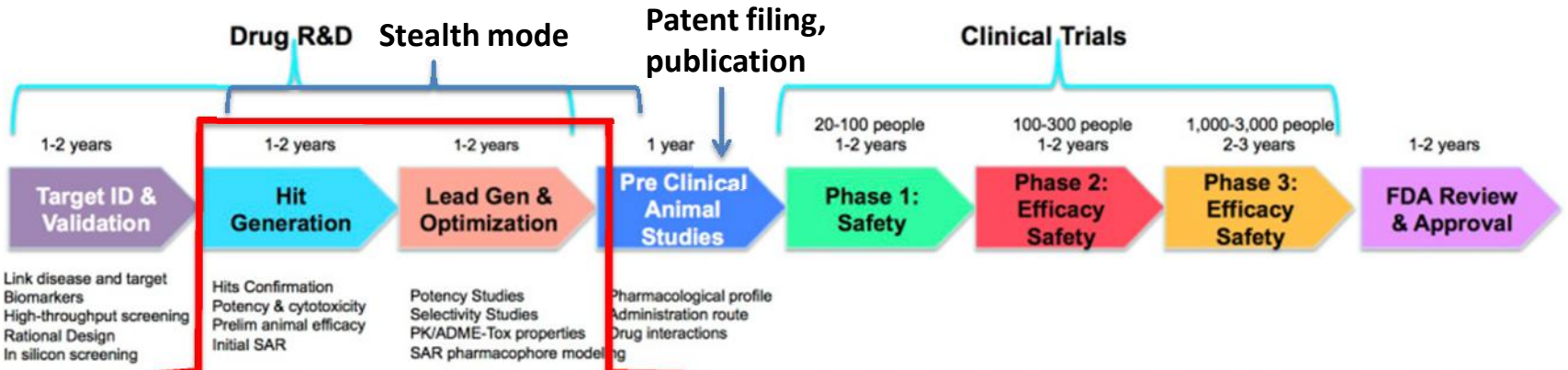
- Lots of things may happen from your mouth to the cell
  - May not penetrate the gut
  - May not be transported efficiently by blood
  - May be cleansed by liver very fast (not around for enough time to be effective)
  - May get more than anything to something which is not its targed (TOX!)
  - Metabolites can be cleared very fast

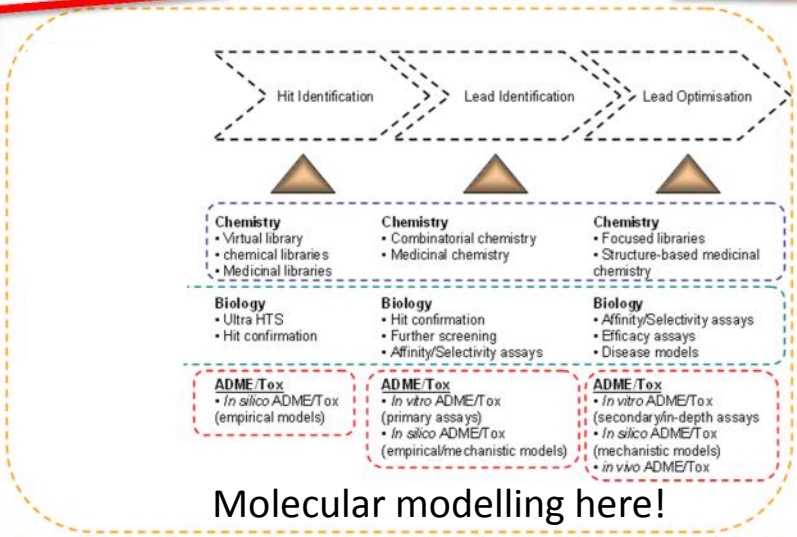



We do not know all about how our body works. We use animal studies to get as close as possible to the real scenario and also here it often does not work!

SCHRÖDINGER.

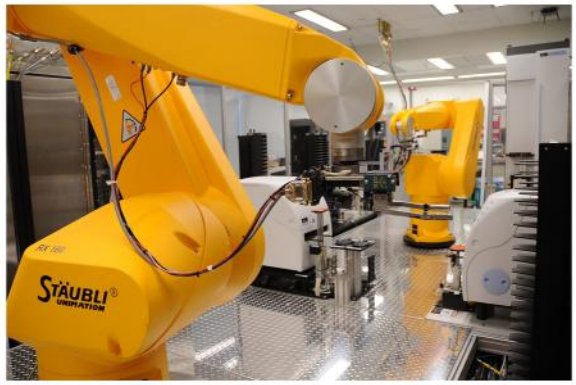
# How drugs are identified?




Molecular modelling here!

High Throughput screening robot.
$10^5$ compounds screened in weeks

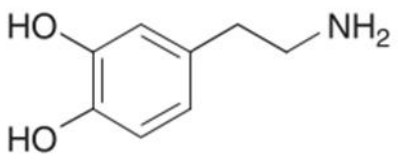


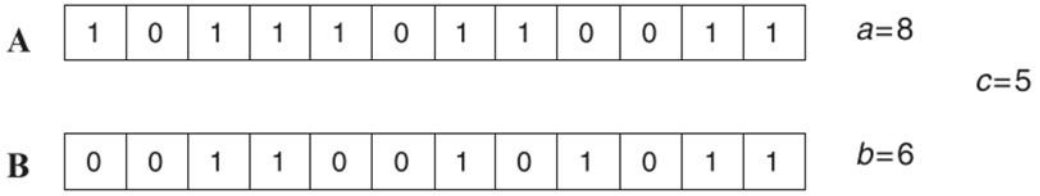

Libraries: millions of compounds

http://www.frost.com/prod/servlet/market-insight-print.pag?docid=135570876
www.brooks.com
https://newdrugapprovals.org/2014/02/

SCHRÖDINGER.

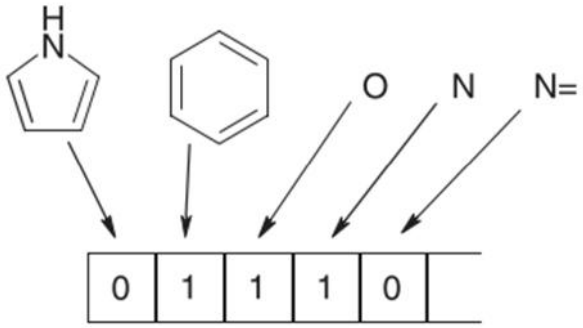
# Library design basic concepts




$$S_{AB} = \frac{5}{8+6-5} = 0.56$$

Tanimoto similarity




Bitstring (or fingerprints)

Clustering

Filtering

SCHRÖDINGER.

# Prioritize compounds with Molecular modeling: in-silico approaches

RCSB PDB PROTEIN DATA BANK An Information Portal to 124029 Biological Macromolecular Structures

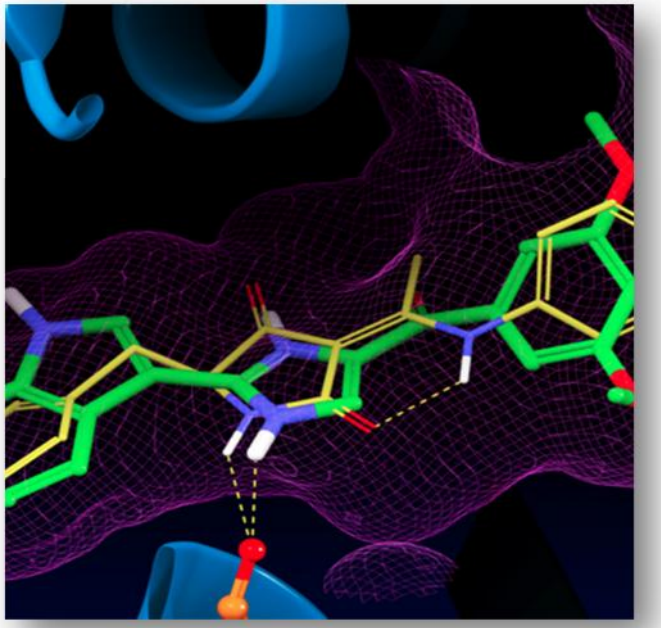
3D structures of targets

Virtual Database of compounds, filtered for the purpose

Ideas

Optimization of potency and physchem properties

Rigid docking/Free Energy Perturbation (via MD simulations)

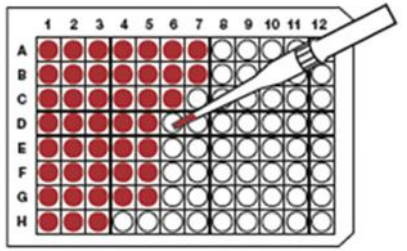


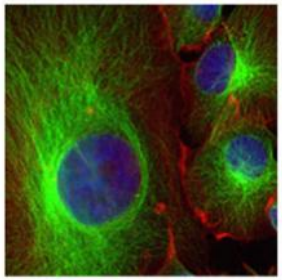
Ranking, Rationalize, Water network analysis

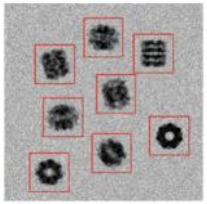
SCHRÖDINGER.

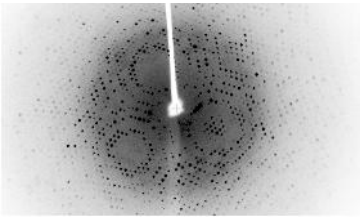
# Data in R&D is everywhere and very heterogeneous

- Managing compounds in stock (availabiltity, characterization, planning, production): 10^6
- Managing assay data
  - multiple experimental sources
    - Images
    - numbers
  - multiple reagents
  - multiple operators
- Managing structural data
  - Xray crystallography
  - Cryo-EM
  - NMR
  - Molecular modeling results
- All these data can be non integrated and redundant/outdated
- Data integration and analytics on all these
  - Spotfire (TIBCO)
  - D360 (Certara)
  - LiveDesign (Schrodinger)
- Managing Electronic lab nootebooks for intellectual property issues



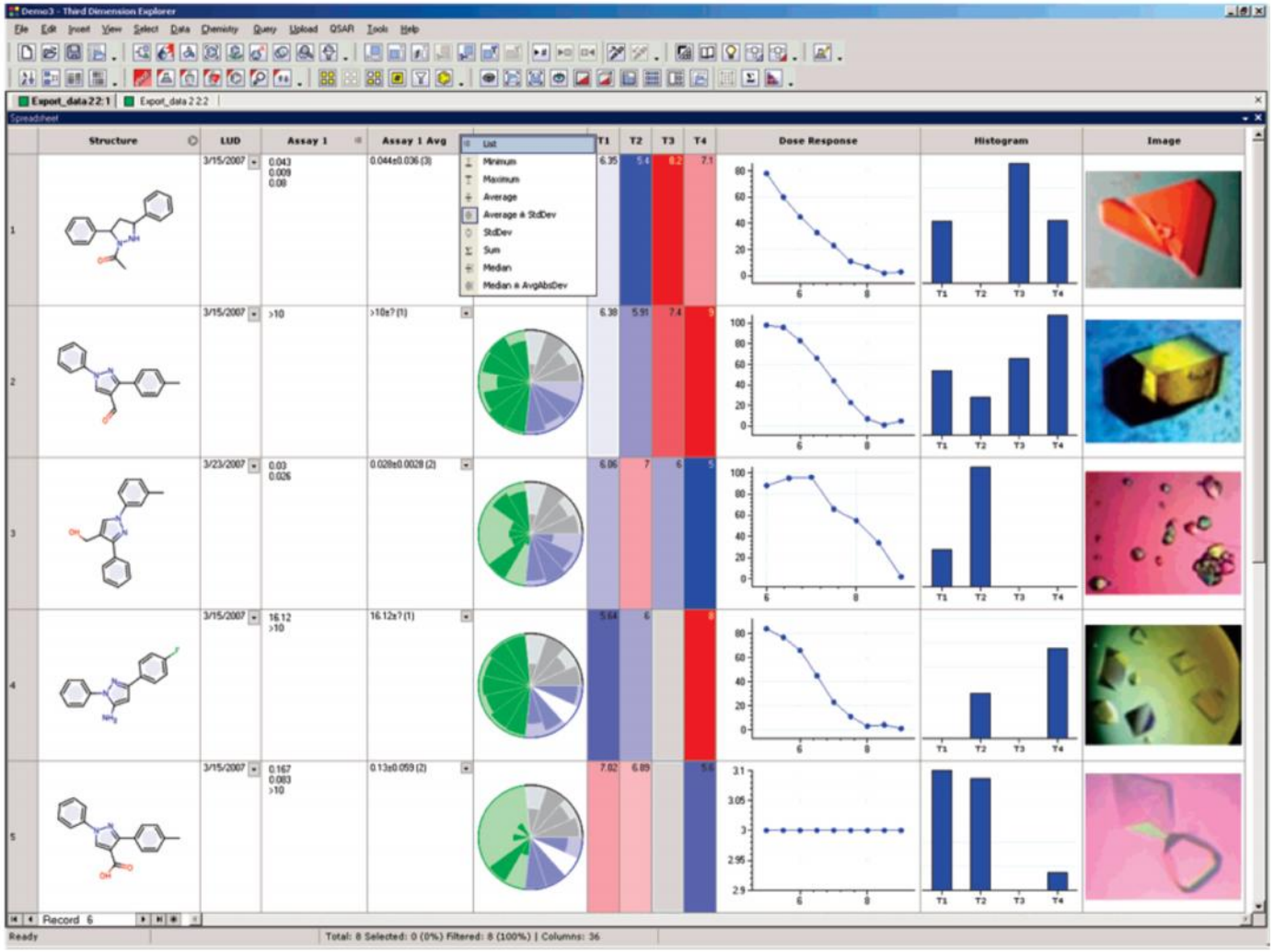







SCHRÖDINGER.

# Example of data integration: Janssen ABCD



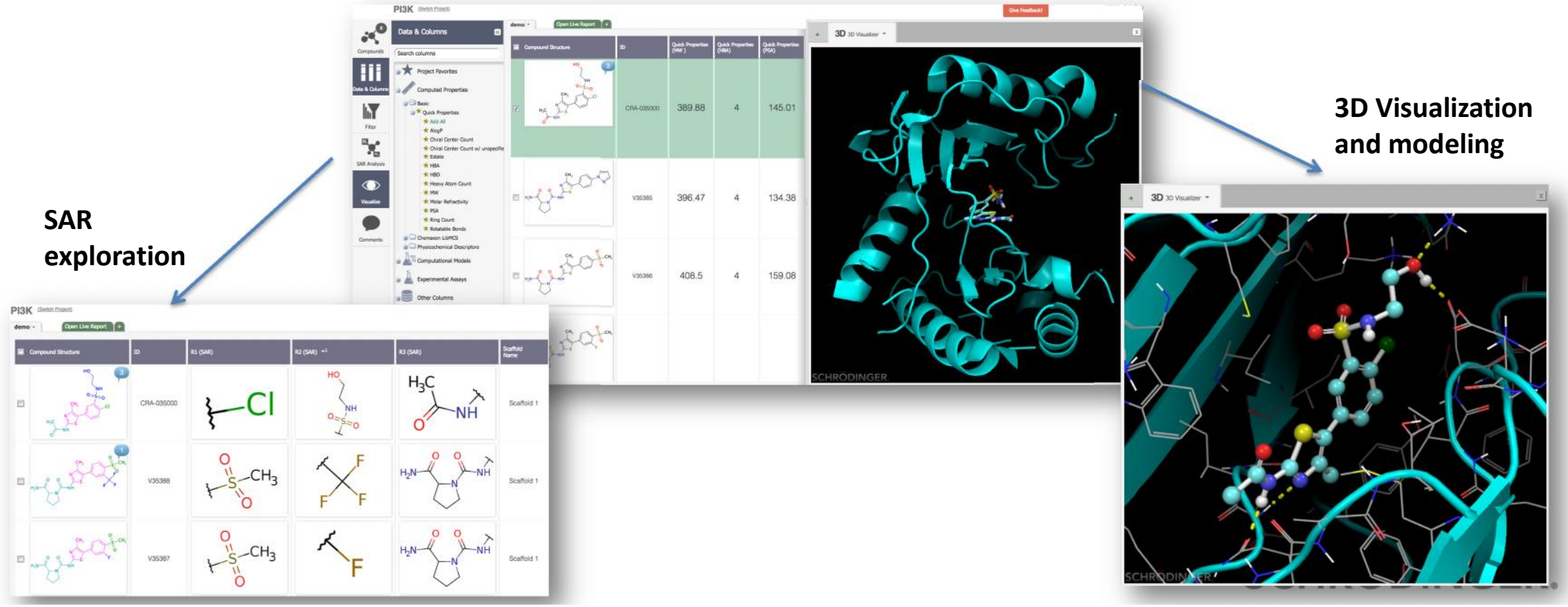

Agrafiotis et. al *J. Chem. Inf. Model., Vol. 47, No. 6, 2007* **2009**

SCHRÖDINGER.

# Integrating experiments and calculations: ideation engine

LiveDesign™

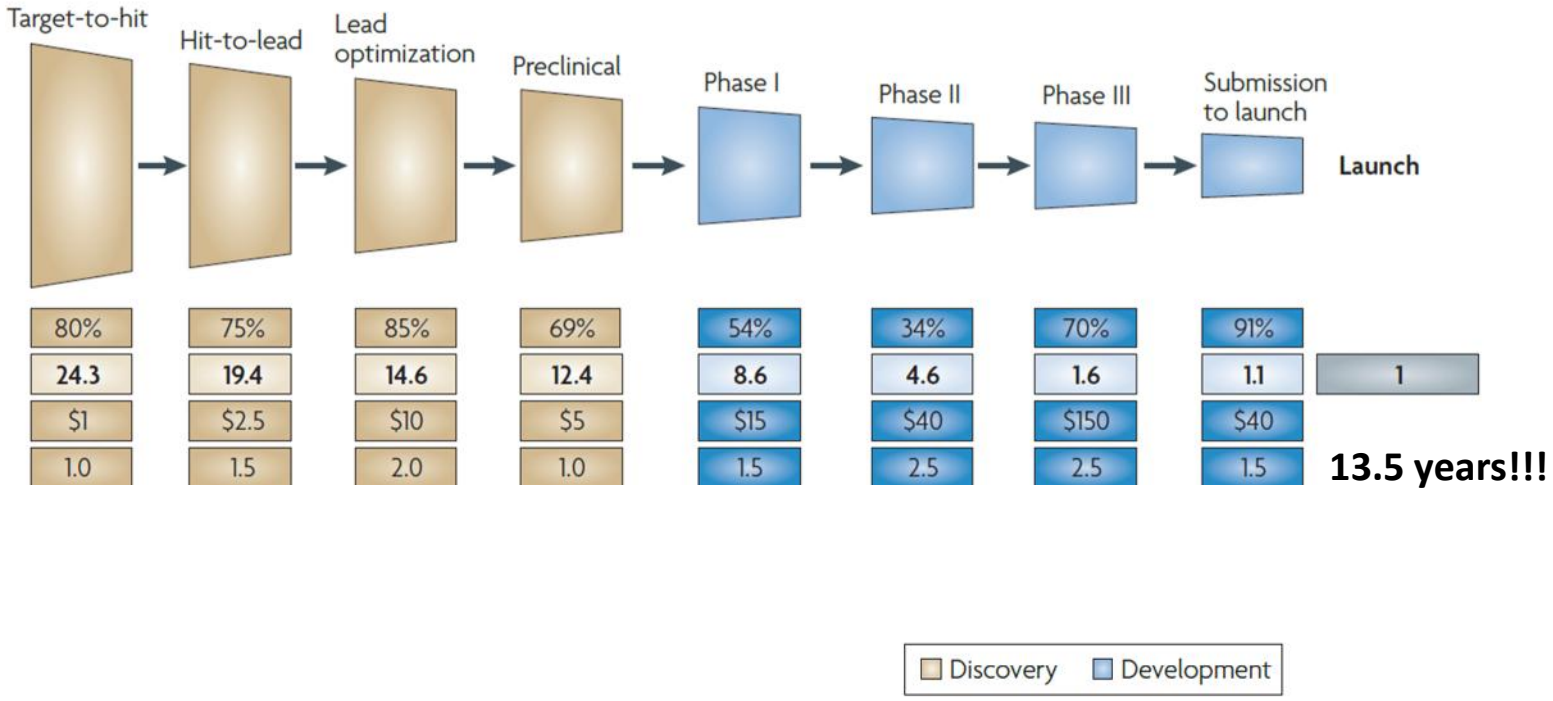
- LiveDesign™ is a browser-based enterprise platform
- Centralizes your small molecule data, ideas, and communication
- Designed to improve project efficiency



# From ideation to market: the path of a drug

$$P \alpha \frac{WIP \times p(\mathrm{TS}) \times V}{CT \times C}$$

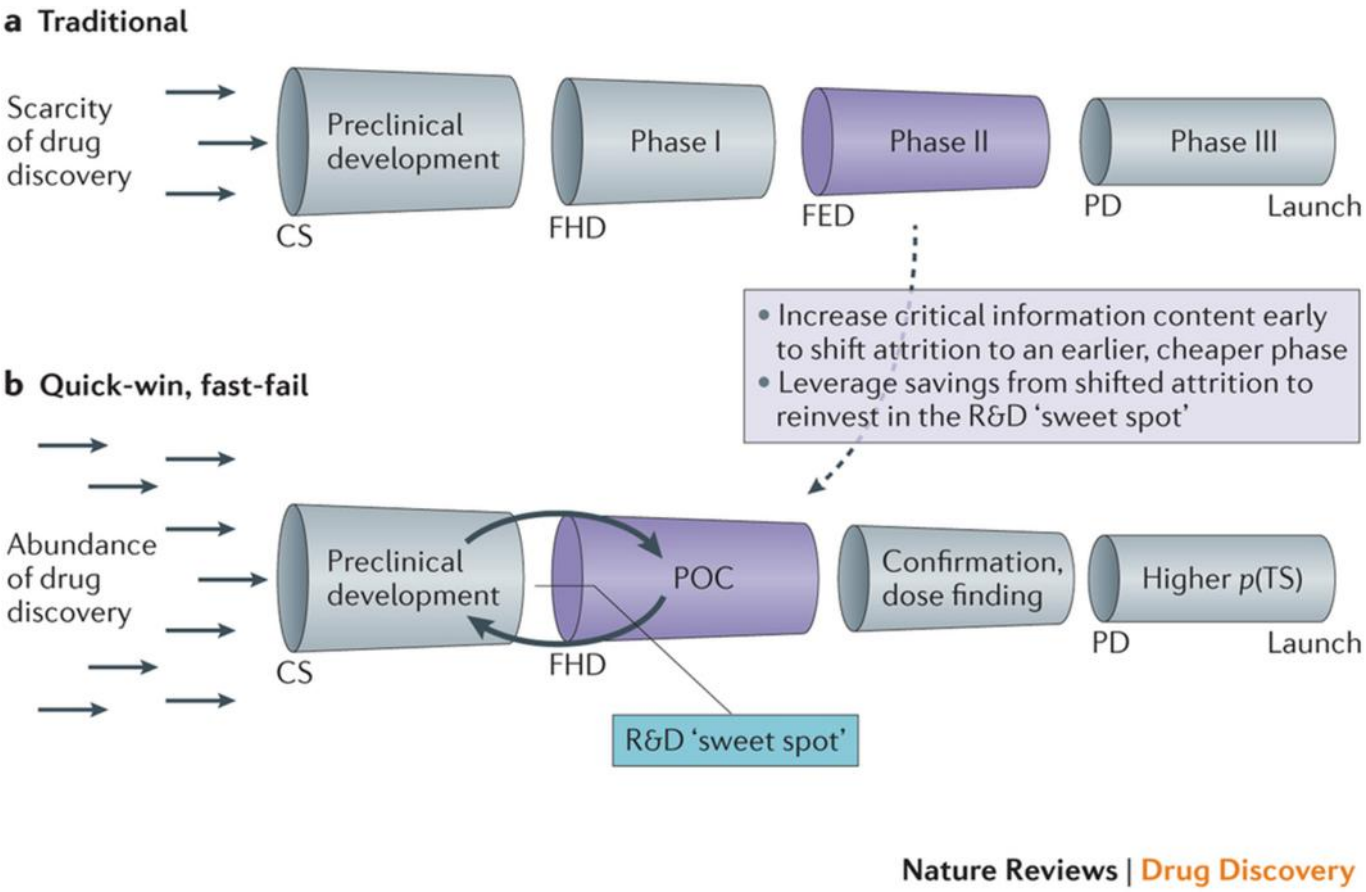
P= Productivity; WIP=Work in process; **p(TS)**=Probability of technical success; V=Value; CT=Cycle Time; C=Cost




Paul et al. Nat Rev Drug Disc (2010), 9, 2003

SCHRÖDINGER.

# Quick-win, fast fail

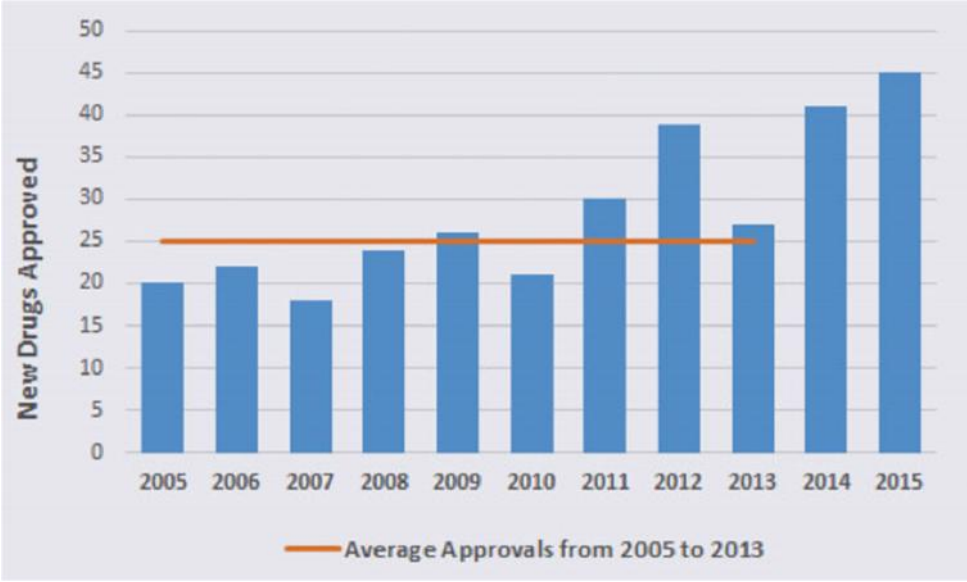



Owens et al, Nature Reviews Drug Discovery 14, 17–28 (2015)

SCHRÖDINGER.

# How many submission per year to FDA?

- Last year seem to see a new trend: finally out of “Ice Age” of pharma industry?




Maybe due to lots of first-in-class (get high chance of approval) and other FDA approved schemes **Fast Track, Breakthrough Therapy, Accelerated Approval, Priority Review (http://www.fda.gov/forpatients/approvals/fast/ucm20041766.htm)**

http://www.impactpharma.com/blog/record-numbers-of-fda-approved-drugs/

SCHRÖDINGER.

# Playing tricks

- Accelerating drugs through market

- (*The Wall Street Journal via NewsPoints Desk*)

## Drug Firms Buy $67.5 Million Voucher to Speed FDA Review

- (*The Wall Street Journal via NewsPoints Desk*)

*(Ref: The Wall Street Journal)*

July 31st, 2014

BioMarin got a voucher for contributing with a drug for a unmet medical need in paediatric area
Sanofi bought it for cholesterol reducing drug

## Sanofi, Regeneron's PCSK9 inhibitor Praluent gains FDA priority review

*(Ref: Yahoo!Finance, Morningstar, FinanzNachrichten, Forbes, PR Newswire, Bloomberg)*

January 26th, 2015

By: Matthew Dennis

Tags: Top Story evolocumab Praluent Amgen Regeneron Sanofi FDA Dyslipidemia Regulatory Affairs

https://www.firstwordpharma.com/node/1259857

SCHRÖDINGER.

# Where all this got us

## Cost and Time

- Cost of single drug is estimated to be around 1Billion $ !
- Time of getting a new drug is 13.5 years
- Lots of failures

## Data generated

- Compound libraries of millions of compounds, characterized, stocked, tested
- Combinatorial chemistry
- High throughput screening facilities
- Large databases, from chemical structure, to storage, to batchID, to experiment, to 3D structures, ADME/tox

# Are we using data in the right way?

- In-house data: are we looking to the data we already have in the right way?
- External data: are we accessing all the data which sits outside (institutions, companies) ?
- Data analytics offers now great opportunities: is it the case to teach a old dog (pharma) a new trick (data science)?

# Digging in-house data

Drug Discovery Today

Volume 20, Issue 6, June 2015, Pages 652–658

Feature

## Extending kinome coverage by analysis of kinase inhibitor broad profiling data

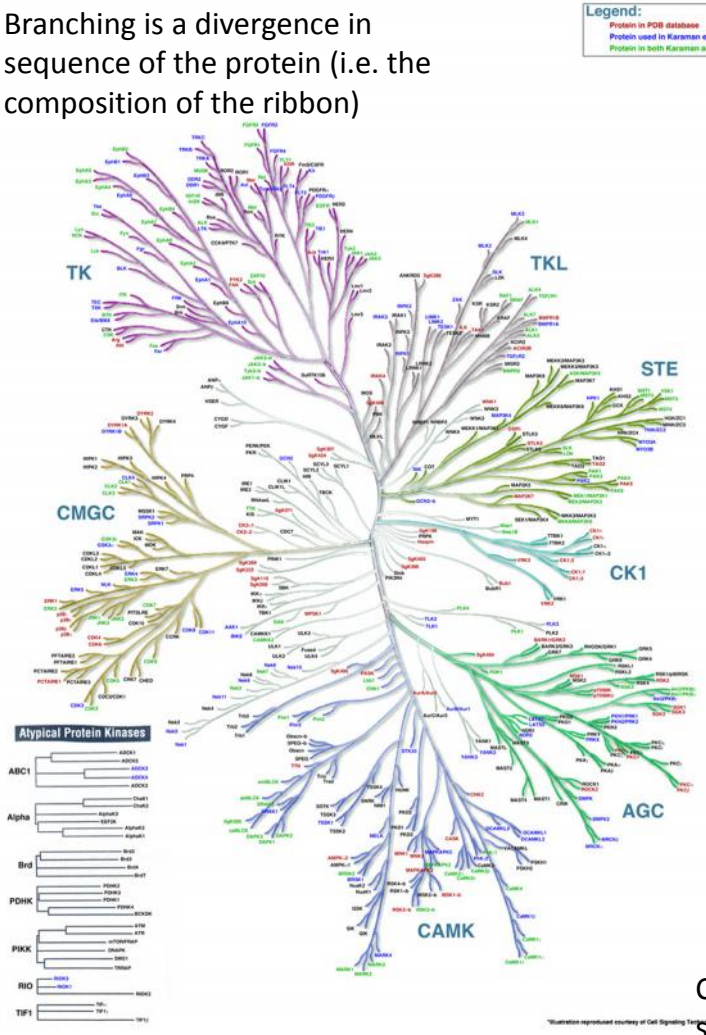
Edgar Jacoby[1], Gary Tresadern[1], Scott Bembenek[2], Berthold Wroblowski[1], Christophe Buyck[1], Jean-Marc Neefs[1], Dmitrii Rassokhin[3], Alain Poncelet[4], Jeremy Hunt[5], Herman van Vlijmen[1, 6]

- Janssen has an extensive compound library
- Over 40 project on Kinases in the years (~1.5billions$)
- 70K compounds synthesized to target kinases
- Can we capitalize on this gigantic effort to find new targets?

SCHRÖDINGER.

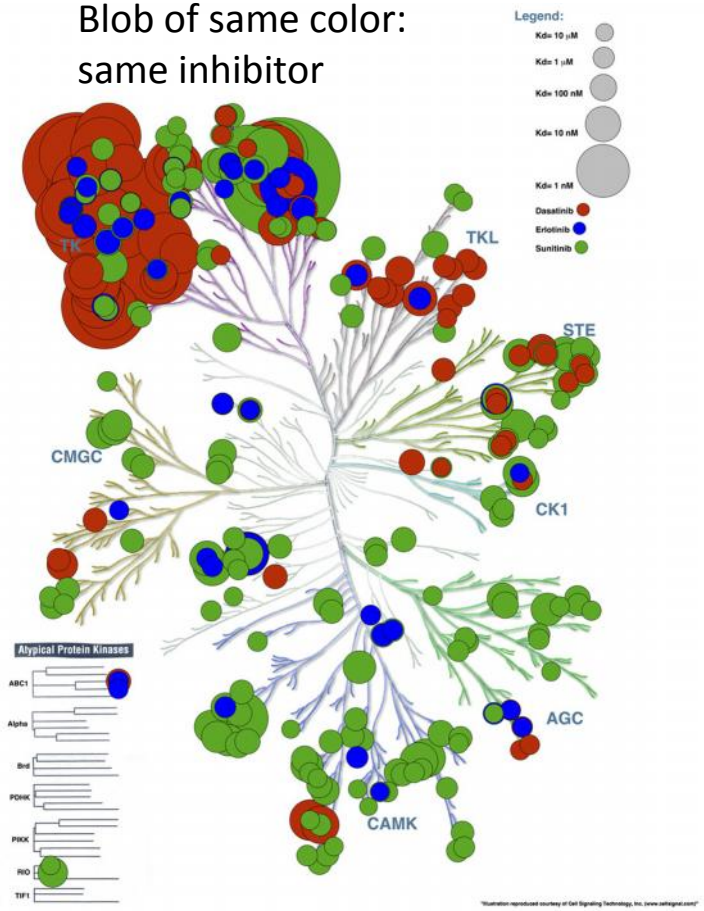
# The kinome: more than 500 similar proteins

Branching is a divergence in sequence of the protein (i.e. the composition of the ribbon)




- Specificity is important to limit the side effects
- For very similar proteins a limited degree of promiscuity is inevitable
- There are also a number of well documented classes of drugs
- Mostly linked to cancer therapies
- Finding new drugs with specificity of this kind would be already a success

Blob of same color: same inhibitor

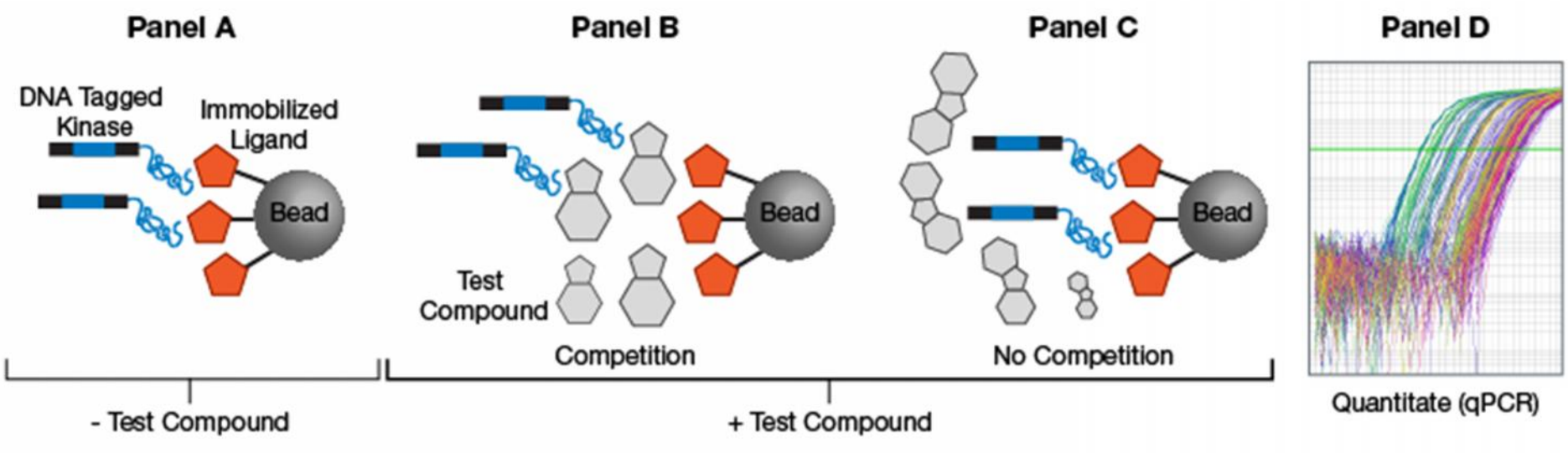



Chartier M, Chénard T, Barker J, Najmanovich R. (2013) Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. *PeerJ*1:e126

SCHRÖDINGER.

# DiscoverX Kinome Scan

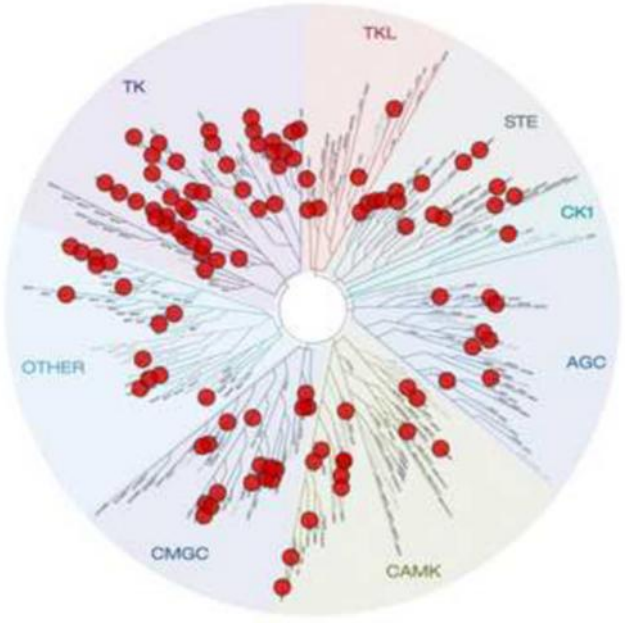
450 Kinases provided by DiscoverX* **3K compounds from Janssen** =13500K experiments? No!
-> Test each ligand with multiple kinases, then measure which kinases are attached on the bead. The ones which are not **attached have interacted with the ligand**




https://www.discoverx.com/technologies-platforms/competitive-binding-technology/kinomescan-technology-platform

SCHRÖDINGER.

# Results




KD/IC50 <= 10 nM S65 <= 0.05

129 kinases by 434 cpds

- New potent and selective compounds for many new kinases are found (55)
- New project were started as consequence of this effort
- Much of these compounds is already known beforehands since they've been amply characterized (cost/time cut)
- Good eye for new technologies provide new ways to benefit from material and data already present in house

SCHRÖDINGER.

# Share with care: pre-competitive agreements

According to Pistoia Alliance: "**aggregating, accessing, and sharing data that are essential to innovation, but provide little competitive advantage**".
Companies and Institutions put some data in a third party institution which act as a broker for the projects of each contributor **to protect everyone's intellectual property**

- Innovative Medicines Initiative
  - ETOX
  - K4DD
  - EMIF
  - OpenPhacts
  - European Lead Factory
  - Etc….
- MedChemica SALT


SCHRÖDINGER.

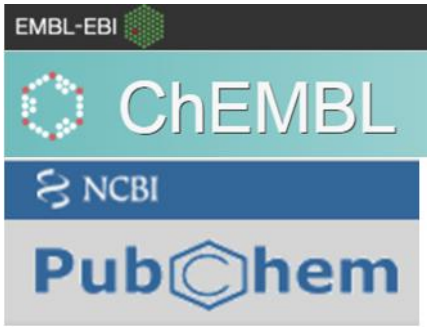
# European Lead Factory

- Idea: my competitor has compounds is not interested anymore. May I speed up my research by using them? After all, he is not much interested anymore in them!
- 30 Institutions and companies share proprietary compounds
- 500K compounds ready to be screened
- Facilities in Scotland (compound library) and the Netherlands for screening
- Scientists who contribute with novel compounds are rewarded
- Researchers and companies can ask to test a target against the compound collection
- Only the confirmed active (~50) will be shared and all the rest of the screened compounds remain unknown to protect the IP of those who shared the compounds
- Companies share lots of knowledge but they disclose very little at a time while having huge impact on productivity

SCHRÖDINGER.

# Digging in others' data: loads of info in the outside world

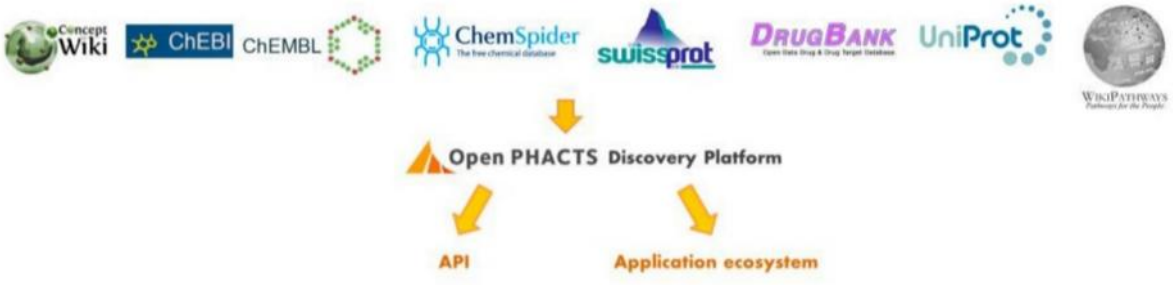
- ChEMBL: 1,686,695 cpds, annotated
- PubChem: 82 millions cpds
- SwissProt: protein database
- Uniprot: protein databse
- Genebank
- Literature




SCHRÖDINGER.

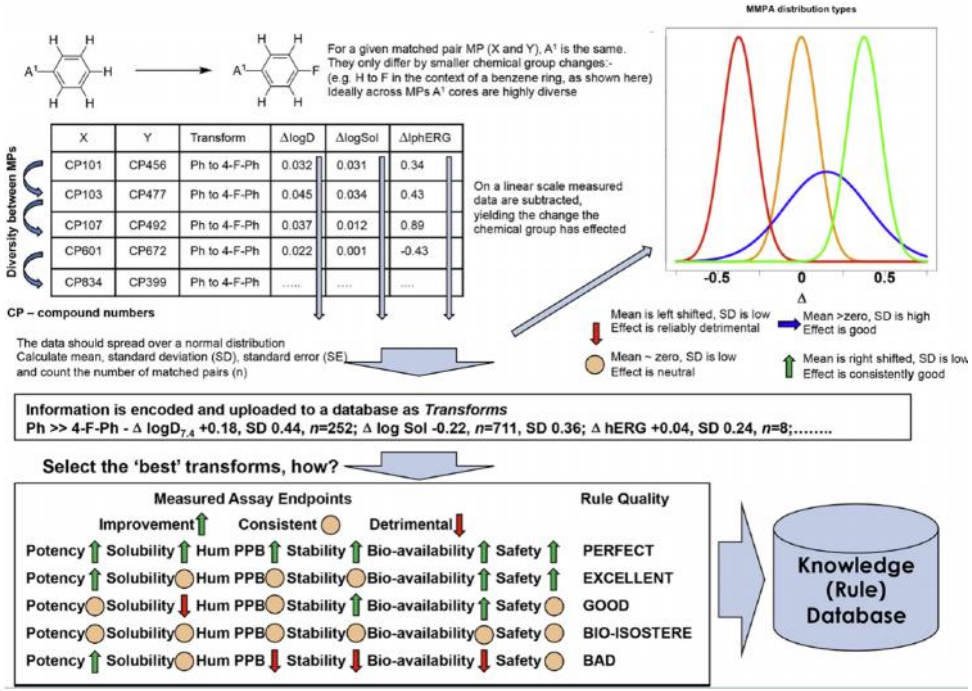
# OpenPhacts

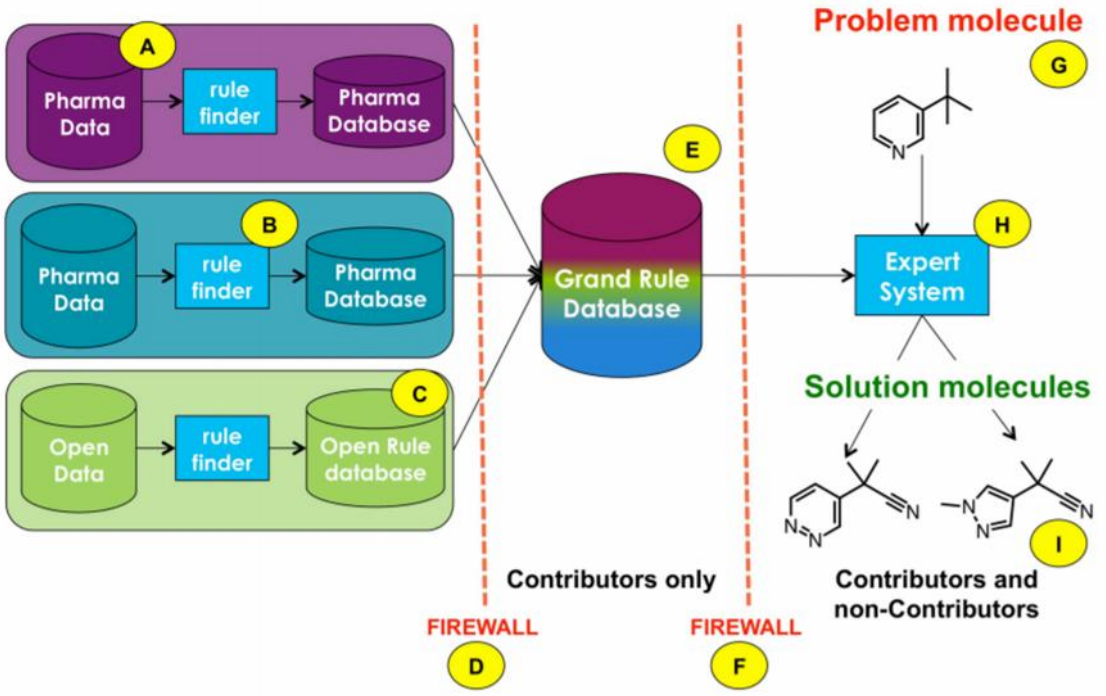

- Lots of databases: drugs, properties, proteins, genes
- Their information is somewhat connected but each single Pharma does the effort to integrate it: redundant efforts
- Task: create a “semantic integration hub”, a common standard and API where company can access all those data and integrate their own
- Integrate informations about  compound–target–pathway–disease/phenotype
- 31 academics, 9 pharma industries, 3 software SME




# MedChemica SALT: a privately owned

- Based on Matched Molecular Pairs



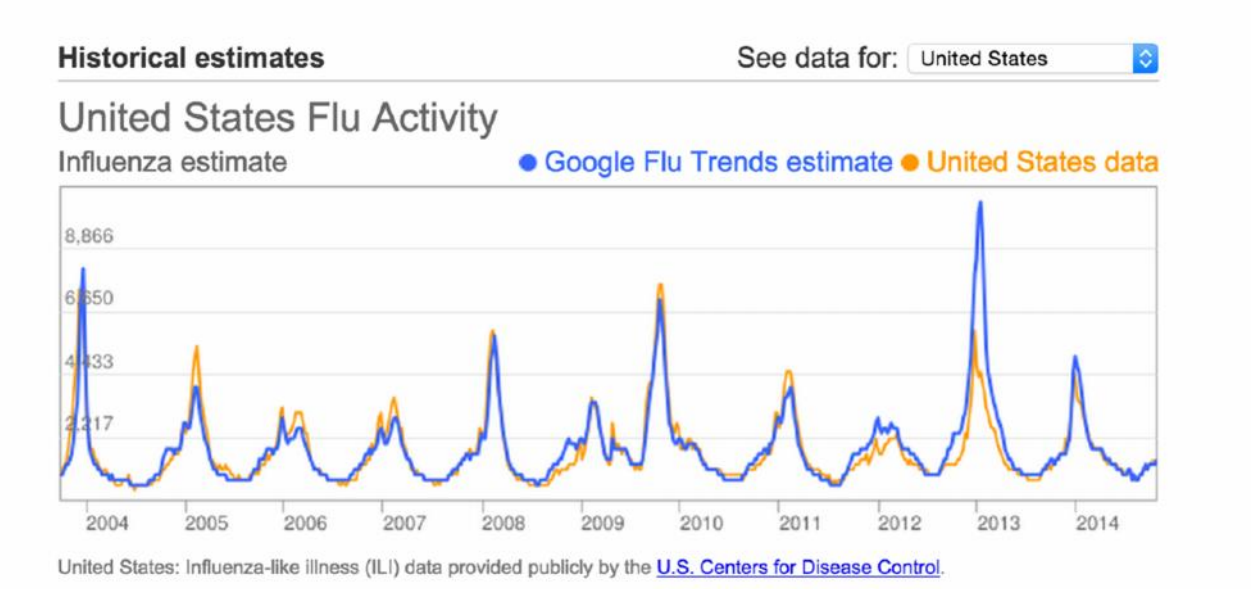




Dossetter et al Drug Discovery Today (2013) 18, 724–731

SCHRÖDINGER.

# Google FLU Trends

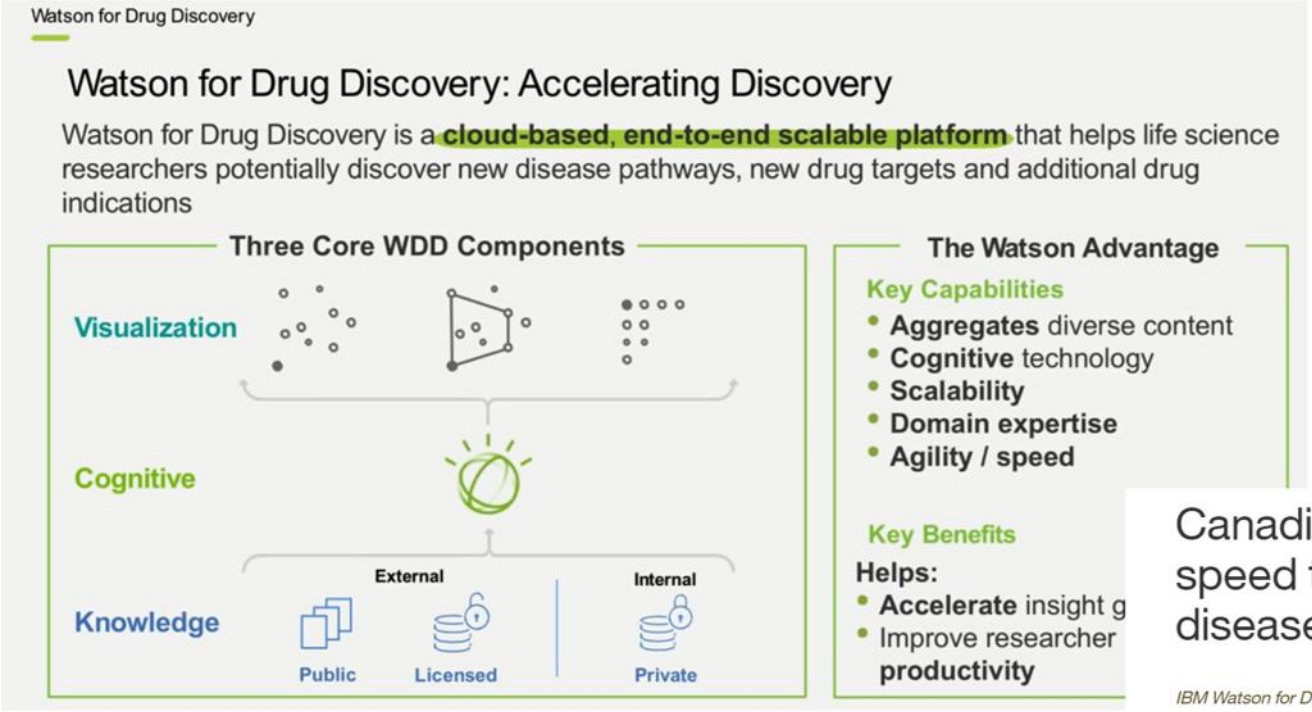
- Launched in 2008 now closed
- Based on google user searches for terms related to flu
- Could predict the spread of flu few weeks ahead respect to the Center for Disease Control and Prevention
- This can help adjusting medical support logistics

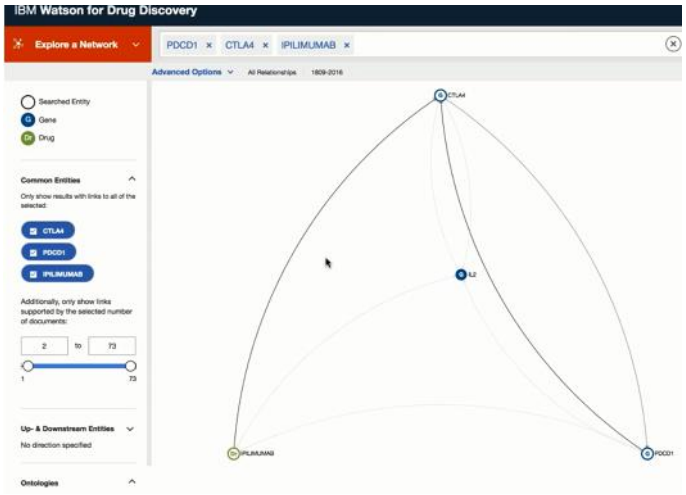



"Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."
Chris Anderson, Wired, 2008
https://www.wired.com/2008/06/pb-theory/

SCHRÖDINGER.

# And of course IBM Watson…




Canadian neuroscience leaders tap IBM Watson to speed time to discovering new drugs for Parkinson's disease

*IBM Watson for Drug Discovery chosen to help researchers more rapidly pinpoint promising drug targets*

Uses Natural Language Processing to dig out quickly unsuspected relations occurring in literature to highlight new targets, biomarkers etc.

Info Hub | **News releases** | Info Hub archives | News release archives

**TORONTO, ON – 12 October, 2016:** IBM (NYSE: IBM) Watson Health today announced that the Ontario Brain Institute (OBI) and the Movement Disorders Clinic (MDC) at Toronto's University Health Network (UHN) will embark on Canada's first ever Parkinson's disease research project using the recently launched IBM Watson for Drug Discovery.

# Genomics revolution

- Deciphering the human genome took 10 years and 3 Billion $
- Now a whole genome can be mapped in 24 hours and costs around 1000$

nature Vol 437|15 September 2005|doi:10.1038/nature03959

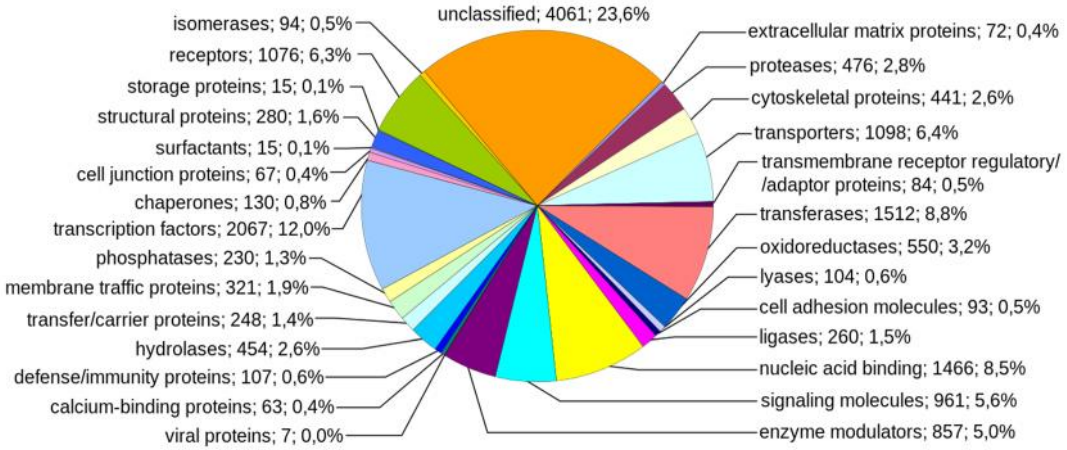
ARTICLES

## Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies[1]*, Michael Egholm[1]*, William E. Altman[1], Said Attiya[1], Joel S. Bader[1], Lisa A. Bemben[1], Jan Berka[1], Michael S. Braverman[1], Yi-Ju Chen[1], Zhoutao Chen[1], Scott B. Dewell[1], Lei Du[1], Joseph M. Fierro[1], Xavier V. Gomes[1], Brian C. Godwin[1], Wen He[1], Scott Helgesen[1], Chun He Ho[1], Gerard P. Irzyk[1], Szilveszter C. Jando[1], Maria L. I. Alenquer[1], Thomas P. Jarvie[1], Kshama B. Jirage[1], Jong-Bum Kim[1], James R. Knight[1], Janna R. Lanza[1], John H. Leamon[1], Steven M. Lefkowitz[1], Ming Lei[1], Jing Li[1], Kenton L. Lohman[1], Hong Lu[1], Vinod B. Makhijani[1], Keith E. McDade[1], Michael P. McKenna[1], Eugene W. Myers[2], Elizabeth Nickerson[1], John R. Nobile[1], Ramona Plant[1], Bernard P. Puc[1], Michael T. Ronan[1], George T. Roth[1], Gary J. Sarkis[1], Jan Fredrik Simons[1], John W. Simpson[1], Maithreyan Srinivasan[1], Karrie R. Tartaro[1], Alexander Tomasz[3], Kari A. Vogt[1], Greg A. Volkmer[1], Shally H. Wang[1], Yong Wang[1], Michael P. Weiner[4], Pengguang Yu[1], Richard F. Begley[1] & Jonathan M. Rothberg[1]

The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and *de novo* assembly of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy in one run of the machine.

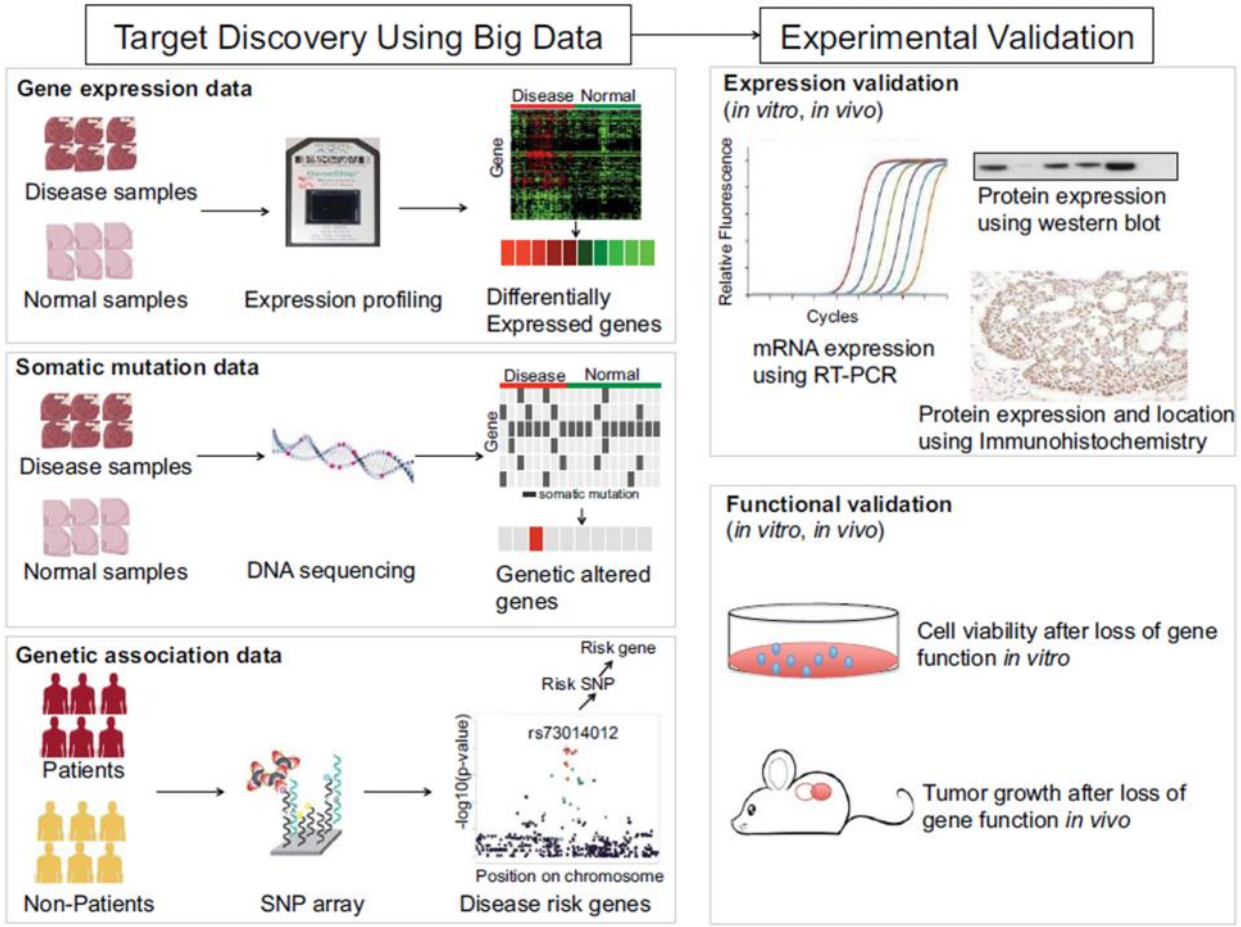
Human genome is 3 Billions of base pairs!




SCHRÖDINGER.

# Publicly available genome data: finding new targets

Detect a change in a known pattern (i.e. overexpression, mutation), consider high noise is expected (for full genome sequence) and druggability (i.e. has pockets for small mols in relevant regions) of the target must be considered




Validate the causal relation

May found that more than one drug is needed

Drug repurposing + personalized medicine!

Chen B and Butte AJ, **2016** *Clin Pharmacol Ther*

SCHRÖDINGER.

# Conclusions

- Technology innovation and integration is key in pharma industry
- Many branches of today's most exciting science under the same hood: chemistry, (molecular)biology, genomics, physical chemistry, molecular modeling
- Data integration plays a central role
- Standardization is ongoing
- New technologies appear, these give new opportunities to use existing data
- Using external data is becoming appealing too, in particular when using third parties that protect knowhow

SCHRÖDINGER.