Big Data Science in Drug Discovery and Development

CIKM 2016 Tutorial Part II

October 24th, 2016

Ping Zhang

Center for Computational Health
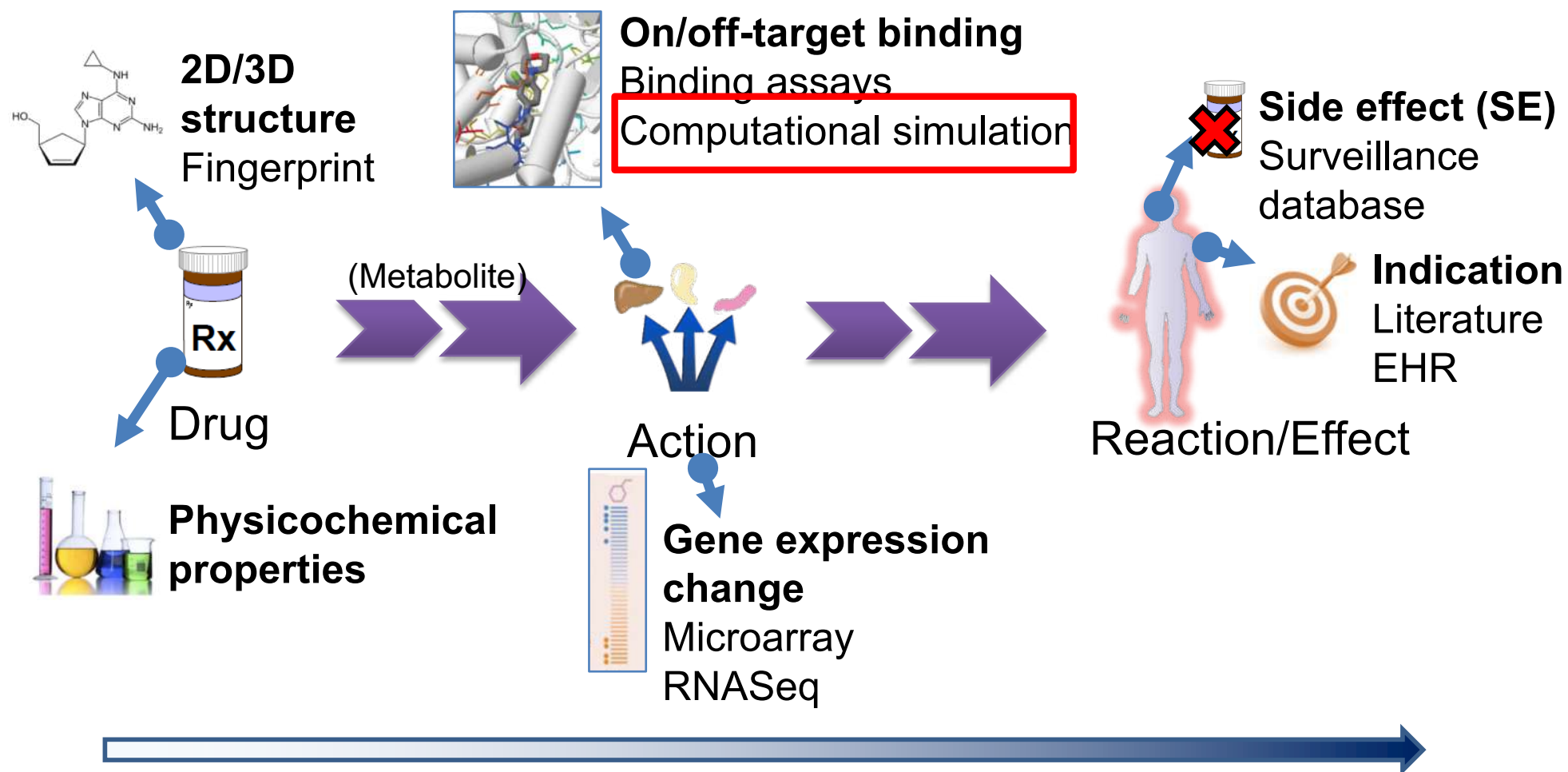
IBM T.J. Watson Research Center

pzhang@us.ibm.com

Outline

- Preclinical data analytics – chemical-protein interactome (CPI) as an example
    - Drug-drug interaction prediction
    - Drug repositioning
- Patient data analytics – real-world evidence (RWE) as an example
    - Drug safety signal detection from FAERS

Path from drug to effect

2D/3D structure
Fingerprint

Drug

Physicochemical properties

(Metabolite)

On/off-target binding
Binding assays
Computational simulation

Action

Gene expression change
Microarray
RNASeq

Side effect (SE)
Surveillance database

Indication
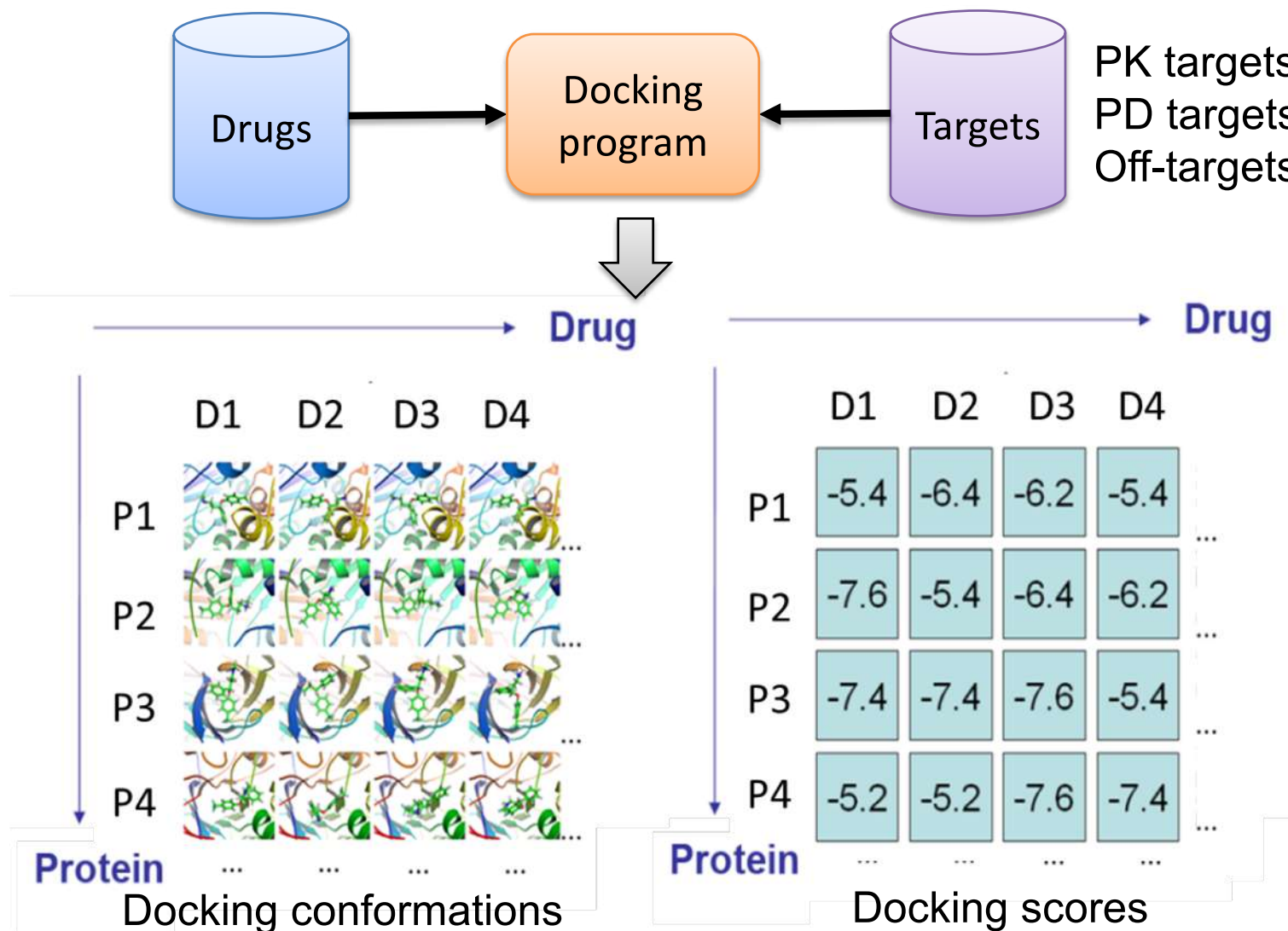Literature
EHR

Reaction/Effect

3

Molecular docking

- A docking program simulates the binding between a small molecule and a protein target.
  - Optimal binding position
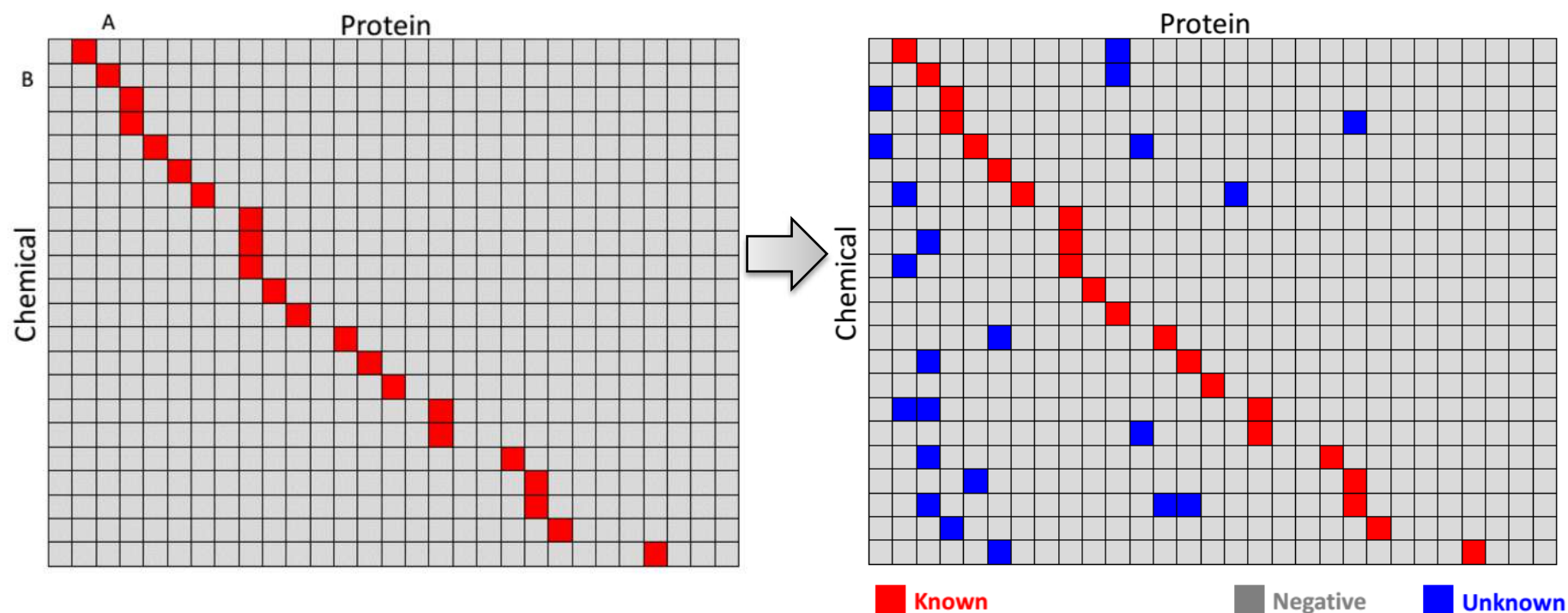 - Binding strength (docking score)
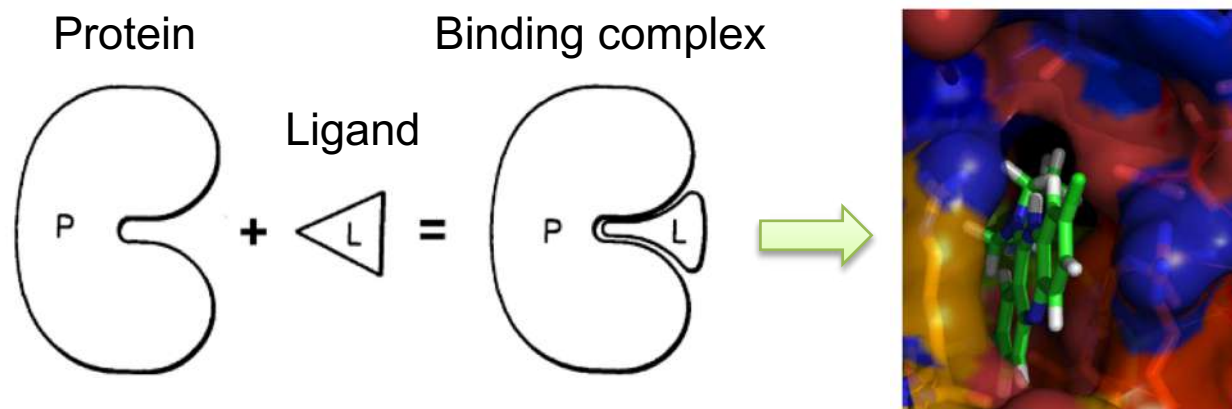
Drug binding inside the protein

Chemical-protein interactome (CPI)

Docking conformations

Docking scores

Simulation of a chemical-protein interactome (CPI)

Why chemical-protein interactome?

- Expand the existing knowledge
  - Identify potential off-target binding
- Fast – 1 minute for a drug-protein pair
- Cheap – compared to wet-lab experiments

Application 1: Drug-drug interactions

- Older patients usually take more than one drug
- 1/25 individuals have adverse reaction caused by drug–drug interactions (DDIs)

Types of DDIs
- Potentiation
- Interference
- Antagonism
- Displacement

Two drugs compete for protein binding sites - a *major* cause for drug-drug interactions

Workflow of DDI-CPI server

Model training

(A) 12,656 drug pairs
(DrugBank)

Drug pair	DDI?
A and B	Yes
B and C	Yes
A and D	No

(B) Docking scores (2,515 drugs against 611 targets)

Drug	T_1	T_2	...
A	-9.3	-9.8	
B	-8.4	-10.1	
C	-7.3	-9.1	

Combine

(C) Training set

The sum and the absolute difference of the docking scores as features

Drug pair	T_1		T_2			DDI?
	Sum	Dif.	Sum	Dif.		
A and B	-17.7	0.9	-19.9	0.3	...	Yes
B and C	-15.5	0.9	-17.6	0.6		Yes
A and D	-19.7	1.1	-20.3	0.7		No

Train →

(D) Logistic regression models

← Predict

Model prediction

(F) Docking score towards 611 targets

Drug	T_1	T_2	...
X	-6.0	-8.2	...

(E) Drug X

Submit ↓

Calculate ←

DDI-CPI server

Combine

(G) DDI predictions

Drug pair	T_1		T_2			DDI?	Probability
	Sum	Dif.	Sum	Dif.			
A and X	-15.3	3.3	-18.0	1.6		Yes	0.68
B and X	-14.4	2.4	-18.3	1.9	...	Yes	1.00
C and X	-13.3	1.3	-17.3	0.9		Yes	1.00

Luo[#], Zhang[#], et al. *Nucleic acids res.* (2014): gku433

Demo: DDI-CPI

Results

The ROC and precision-recall curve comparison for different DDI prediction methods based on independent validation

P-score: uses side-effect similarities to predict target sharing (Campillos, et al. Science (2008), 321, 263-266.)

S-score: uses drug-target network to predict DDIs (Huang, et al. PLoS Comput Biol (2013), 9, e1002998)

LR(S-score and P-score): integrates P-score and S-score by a Bayesian probabilistic model

DDI-CPI: predicts DDI using machine learning models via CPI

Case study - MAO-A inhibitors

Table 3 (adapted from reference 6,7)

Drugs to Avoid When Taking MAOIs

Amphetamines	Bupropion
Cyclobenzaprine	Dextromethorphan
Linezolid	Meperidine
Methadone	Mirtazapine
SSRIs/SNRIs	TCA's
Triptans	Tramadol
Vasoconstrictors (psuedoephedrine, phenylephrine, cocaine)	
Chlorpeniramine, brompheniramine	
St. John's Wort	General anesthesia

- SSRI with MAOI results in high extracellular serotonin (5-HT) concentration – serotonin syndrome.
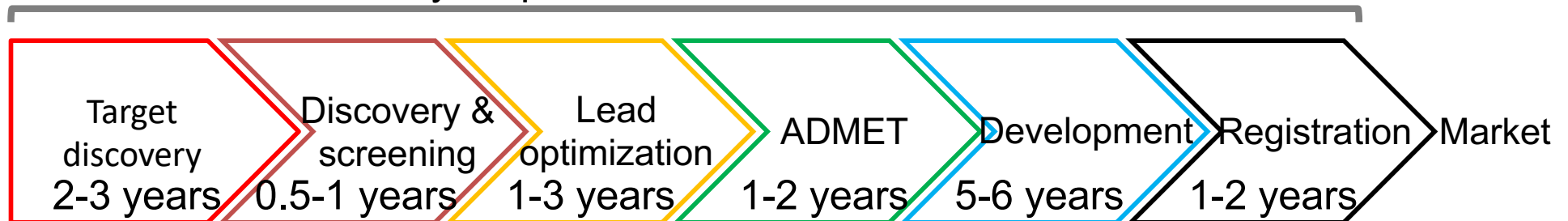

11

Case study - MAO-A inhibitors

Monoamine oxidase inhibitors (MAOIs)

Selective serotonin reuptake inhibitor (SSRI)

MAO-A targets

	2BXR	2Z5X	2Z5Y
Isocarboxazid	-9.4	-9.3	-9.2
Linezolid	-9.4	-10.2	-9.6
Naratriptan	-9.3	-8.8	-8.9
Sertraline	-8.8	-9.7	-9.7
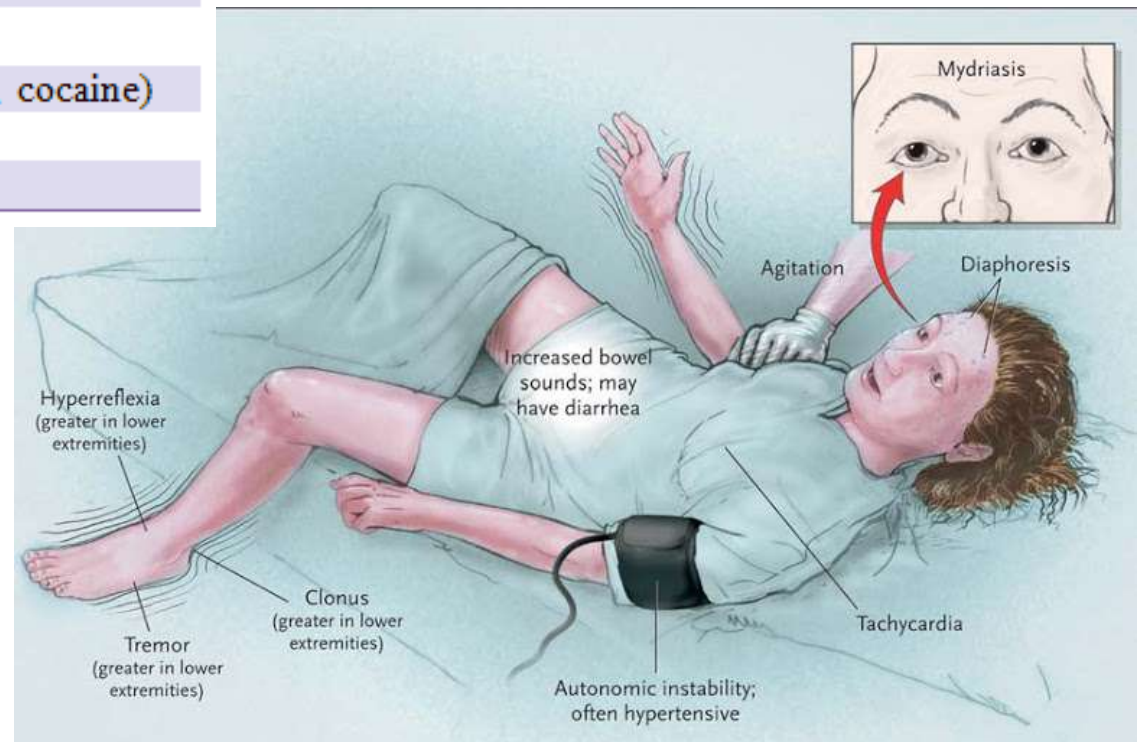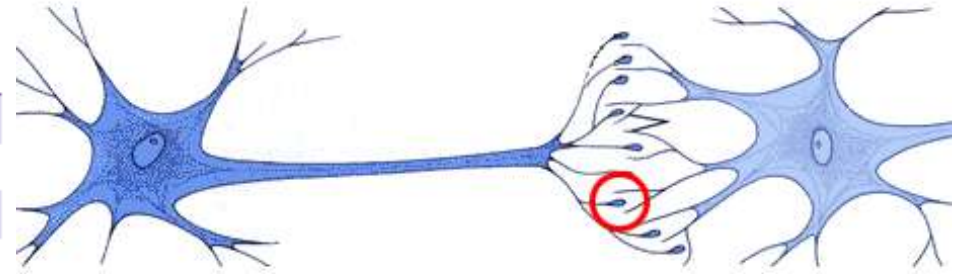
A subset CPI visualization between drugs and MAO-A targets
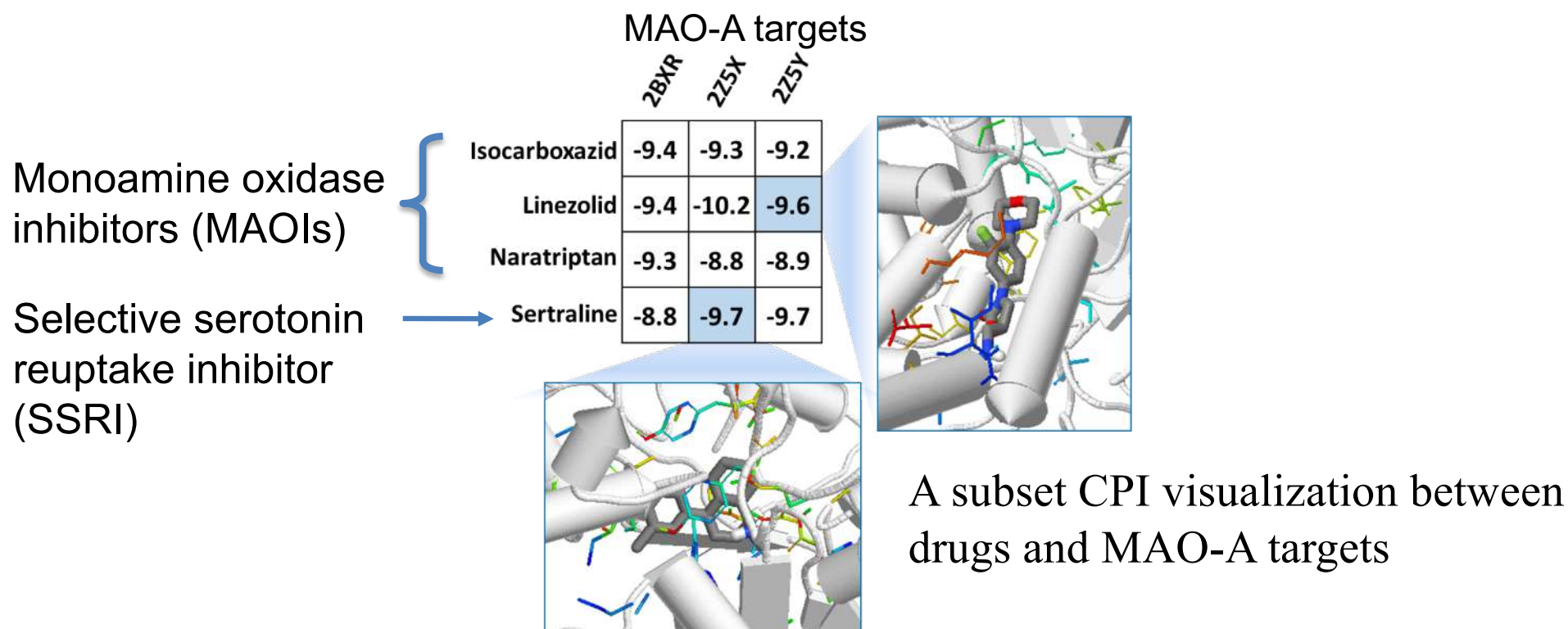
- The server predicts that sertraline may interact with isocarboxazid, linezolid, and naratriptan
- All of the predicted drugs can rank the MAO protein structures to the top 20% – possible mechanism suggested

Application 2: Drug repositioning

- Identify new indications for existing drugs.

De novo drug discovery and development
10-17 year process, <10% overall success rate

| Target discovery 2-3 years | Discovery & screening 0.5-1 years | Lead optimization 1-3 years | ADMET 1-2 years | Development 5-6 years | Registration 1-2 years | Market |

Drug repurposing
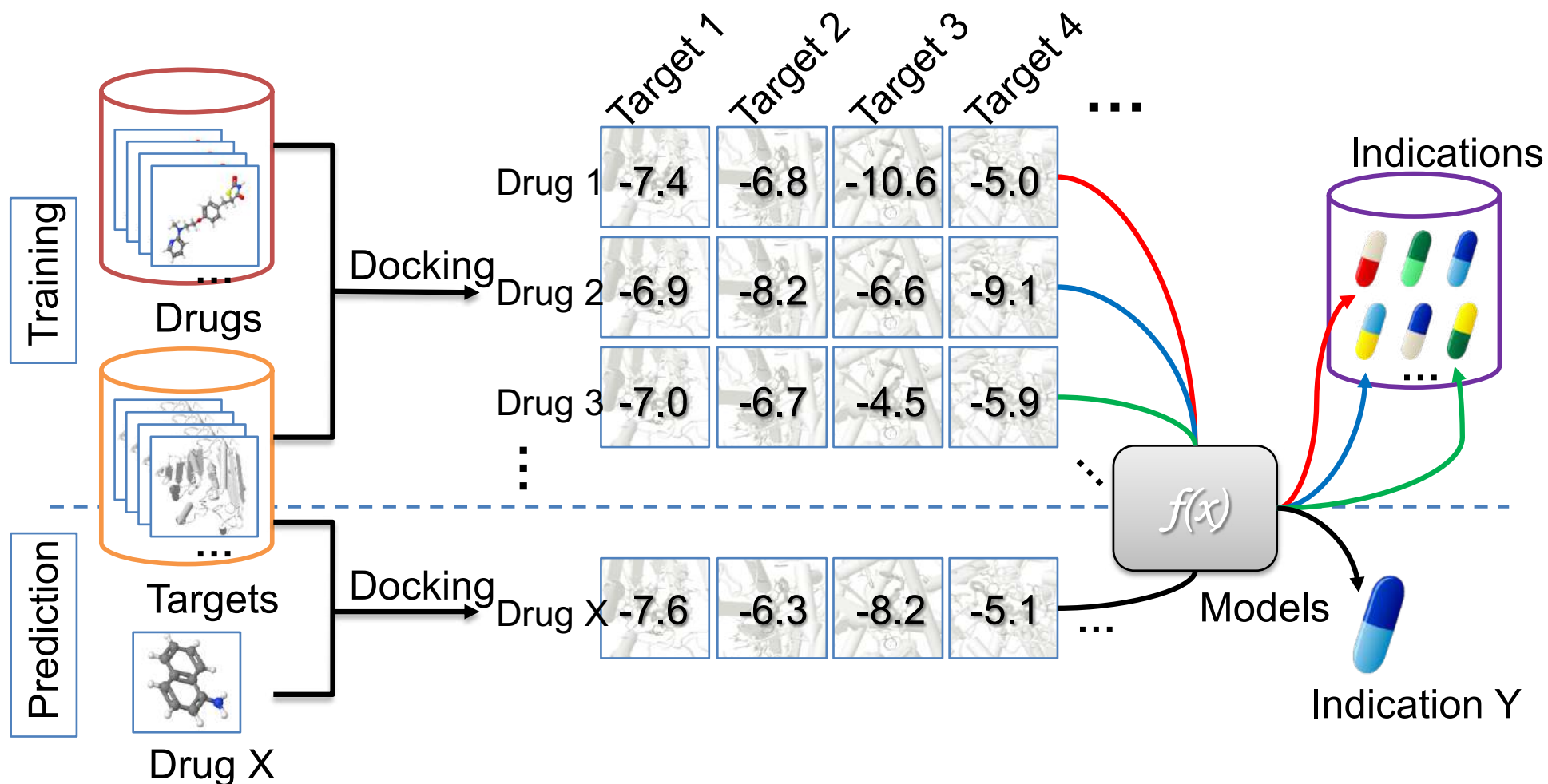3-12 year process, reduced safety and pharmacokinetic uncertainty

Less time
Lower cost

| Compound identification 1-2 years | Compound acquisition 0-2 years | Development 1-6 years | Registration 1-2 years | Market |

Ashburn, *Nature reviews Drug discovery* 3.8 (2004): 673-683.
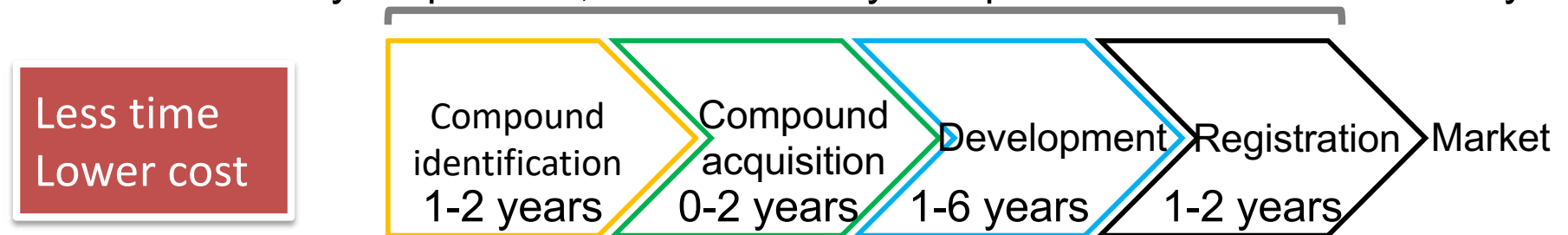
Revenue to pharmaceuticals

1. AstraZeneca Health Connects Us All
77% $$
54% Rx

2. Bristol-Myers Squibb
58% $$
15% Rx

3. Pfizer
50% $$
20% Rx

4. Lilly
44% $$
28% Rx

5. gsk GlaxoSmithKline
37% $$
24% Rx

6. NOVARTIS
29% $$
25% Rx

7. Abbott A Promise for Life
24% $$
17% Rx

8. MERCK Be well
20% $$
20% Rx

9. SANOFI
19% $$
21% Rx

10. Johnson&Johnson
13% $$
36% Rx

11. Bayer HealthCare
9% $$
47% Rx

12. Roche
6% $$
15% Rx

Contribution of the repositioned indications to the sales in 2011

Workflow of DPDR-CPI Server

Luo#, Zhang#, et al. Scientific Reports. (2016)

Demo: DPDR-CPI

State-of-the-arts: Various fingerprints

Positive set
Drugs known to treat disease Y

Descriptors

Descriptor	Treat Disease Y?
00101001001010101010101	Yes
11001010101010100001001	Yes
01010000101010101010101	No
00001001010010101001010	No

Treat Disease Y?

$f(x)$

Machine learning model for disease Y

PubChem 881 Fingerprints

Bit Position	Bit Substructure
0	>= 4 H
1	>= 8 H
4	>= 1 Li
5	>= 2 Li
9	>= 2 C
10	>= 4 C
...	
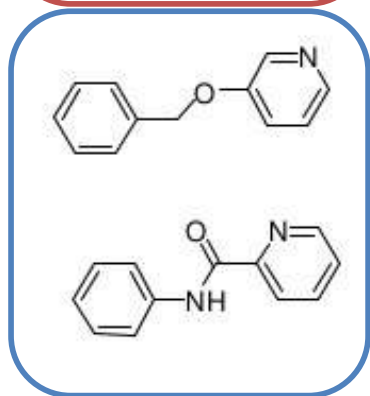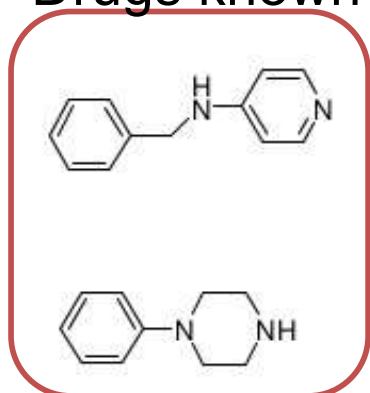
Negative set
Drugs unknown with disease Y

Performance comparison

Performance comparison between DPDR-CPI and chemical structure-based predictors based on independent validation set

Case Study - Rosiglitazone

User molecule → Submit → DPDR-CPI → Check results

Check CPI

Indication predictions

Target binding predictions
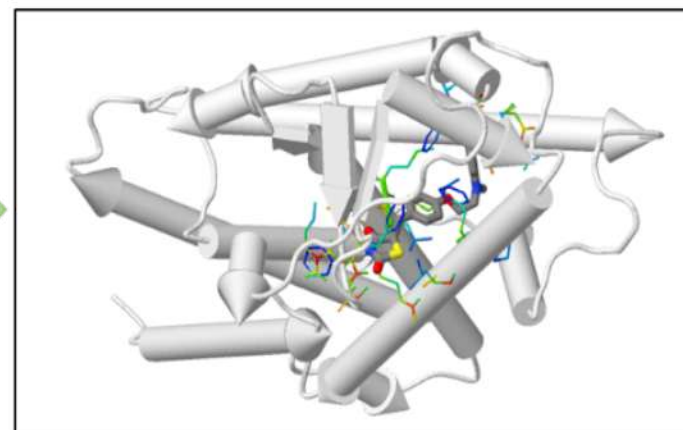
Visualize binding → **Binding conformation**
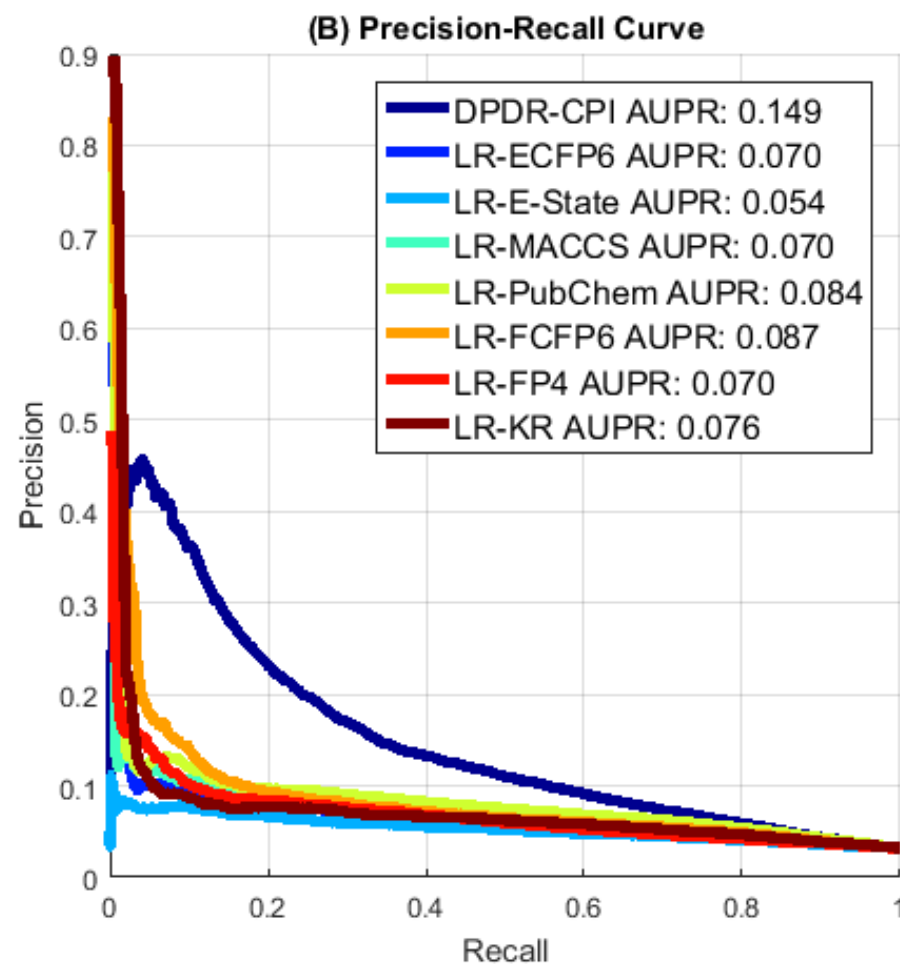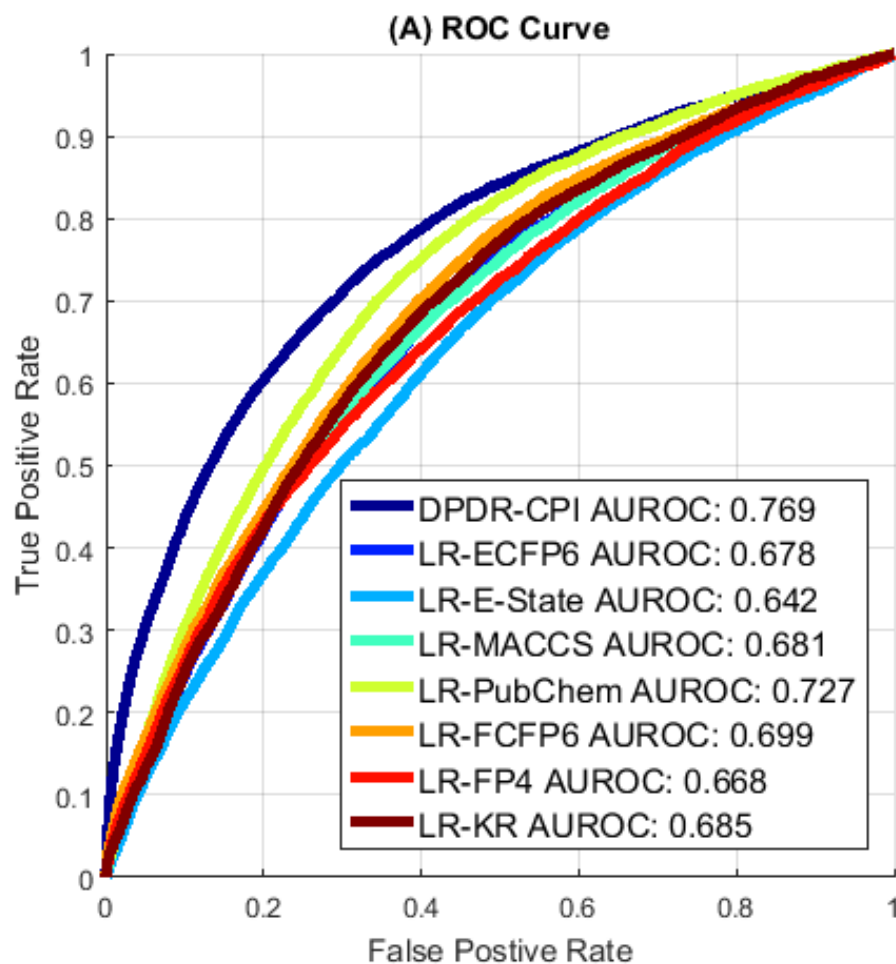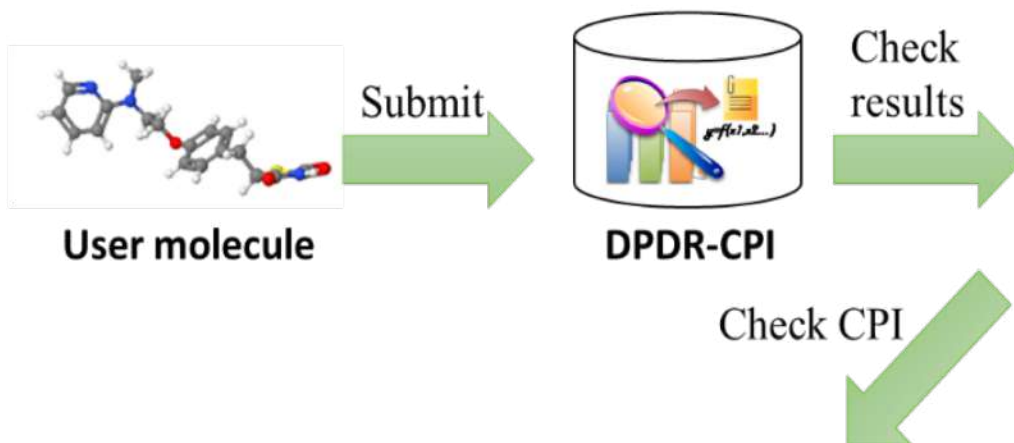
Free big data in the domain

Drug structures
2,198 approved drugs
5,022 experimental drugs
(DrugBank)

On/off-target binding
1,154,431 BioAssays (PubChem)
118,748 crystal structures (RSCB PDB)
551,193 reviewed protein sequences
62,148,086 not reviewed (UniProt)

Side effect (SE)
5,868 side effects
139,756 drug-SE pairs
(SIDER)
6,503,071 reports
(FAERS)

(Metabolite)

Drug

Action

Effect

Chemical structures
89,124,716 compounds
219,712,379 substances
(PubChem)

Gene expression change
3,775 human genomes (1000 genome)
15,819 sequencing platforms (GEO)
68,503 gene expression series (GEO)
1,801,592 gene expression samples (GEO)

Indication
57,805 drug-indication pairs (NDF-RT)
215,433 clinical trials (ClinicalTrials.gov)
22,000,000+ articles (PubMed)

20

Next: Multi-channel detailed computational hypothesis generation
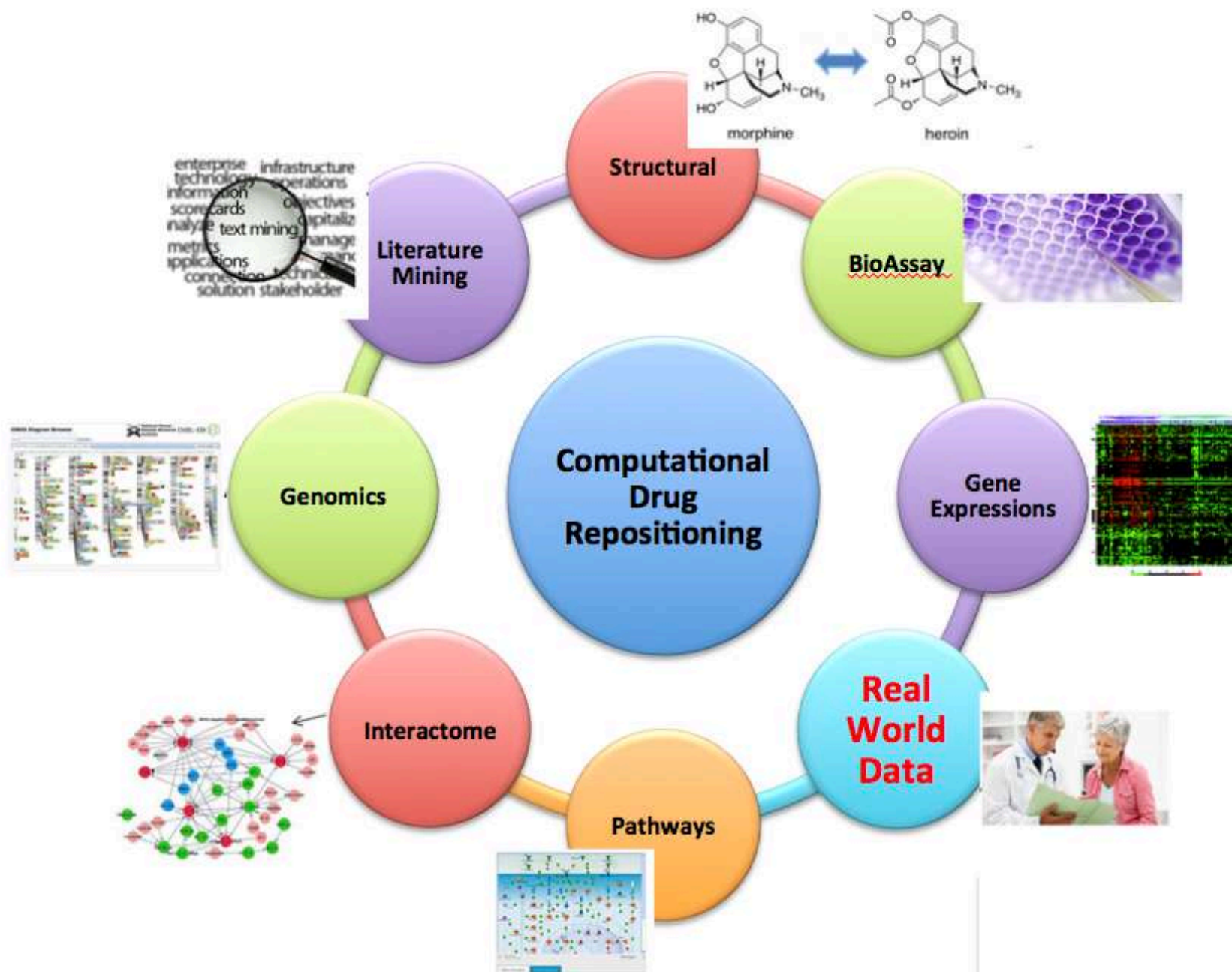
And even beyond the hypothesis generation…

biology | chemistry | dmpk | pharmacology | toxicology

Home » Pharmacology » Diabetes and Obesity » Obese Mice

ob/ob Diabetes Model – 16 Mice

$9,000.00 USD
per service

9 week
turn around time

Provided By

Service Description

Provider: is a US company with laboratories in Hangzhou, China. The laboratory has been offering exploratory (non-GLP) pharmacology services to US and Chinese biopharma since 2004.

Background: The obese mutant mouse model was first reported by Ingalls A *et al* from the Jackson Laboratory in 1951 (Obese, a New Mutation in the House Mouse [164 KB]). The obese mouse resulted from a spontaneous mutation in a gene that was named *ob* in the V stock. Mice homozygous for the obese spontaneous mutation, (Lep^ob^; commonly referred to as *ob* or *ob/ob*), are first recognizable at about 4 weeks of age. Homozygous mutant mice gain weight rapidly and may reach three times the weight of wild-type controls. In addition to obesity, mutant mice exhibit hyperphagia, a diabetes-like syndrome of hyperglycemia, glucose intolerance, elevated plasma insulin, subfertility, impaired wound healing, and an increase in hormone production from both pituitary and adrenal glands. Friedman J *et al* reported leptin in 1994, and demonstrated that leptin, the product of the *ob* gene, was produced in white adipose tissue and served as the peripheral signal to the central nervous system of nutritional status.

Service Details: This service offers a 28 day db/db mouse model of T2DM and obesity. Customer has various options that are conveyed to Links Biosciences using a Service Order Form. Customer assigns up to 16 mice to

Request Info

Add to Cart

SHARE

Be Brilliant™

Enter our online marketplace below to find, compare and purchase research services from hundreds of contract research organizations (CROs).

"Had I known that I can get chick embryo assays done for $2000 in four weeks, I would not have asked a postdoc to spend a year setting it up in our lab."

Holger Wesche, Principal Scientist, Large Pharma

Ask An Expert

Use our free service locator program to find the research services you need.

Register In Seconds

Get free access to detailed information on thousands of research services.

Best Price Guarantee

Purchase services with confidence that you are getting the lowest possible price.

click for more information ›

Big data researchers will have a higher impact in biomedicine ☺

Validation methods are increasingly commoditized

Outline

- Preclinical data analytics – chemical-protein interactome (CPI) as an example
  - Drug-drug interaction prediction
  - Drug repositioning
- Patient data analytics – real-world evidence (RWE) as an example
  - Drug safety signal detection from FAERS

What is "Real World Evidence" (RWE)

- RWE is clinical observations other than randomized clinical trials (RCT).
  - RWE are large-scale clinical observations from population
  - RCT are expensive and in far smaller scale

- RWE is observations on human in the clinical stage
  - Less of a translational issue
  - Other than "omics", numerous external factors (e.g., environment, diet and exercise) affect response to medication

- RWE is not only vast but also varied in type and source: electronic medical records (EMR), claims data, and even social media.
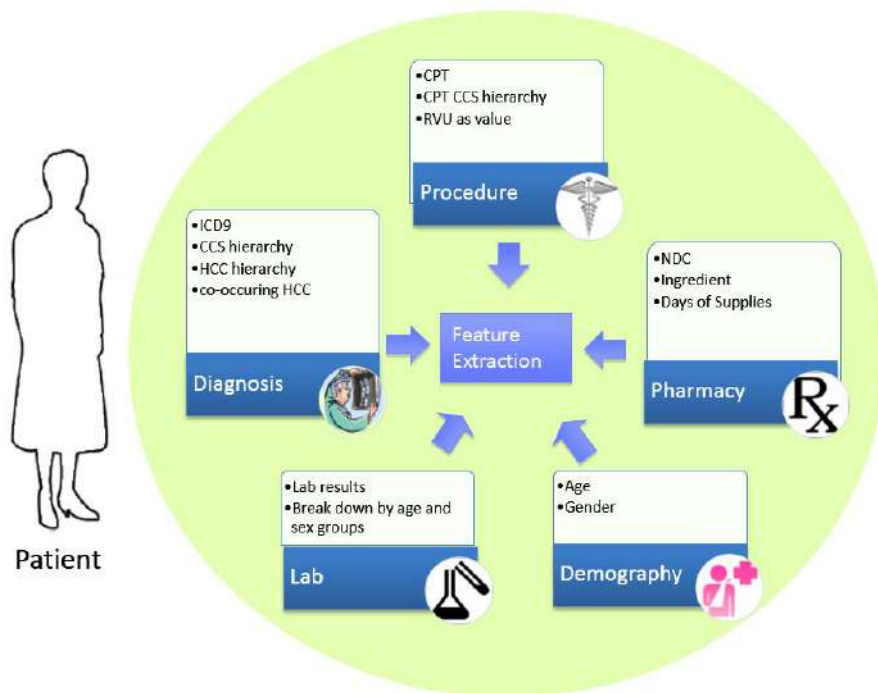
Application 3: Adverse drug reactions (ADRs)

- Post-approval ADRs remain a significant source of mortality and morbidity around the world
  - 2 million potentially preventable injuries, hospitalizations, and deaths each year in US alone
  - Associated cost estimated at $75 billion annually

The New York Times

F.D.A. Issues New Alerts About Cholesterol Drugs

By GARDINER HARRIS
Published: February 29, 2012

CORRECTION APPENDED

Federal health officials on Tuesday added new safety alerts to the prescribing information for statins, the cholesterol-reducing medications that are among the most widely prescribed drugs in the world, citing rare risks of memory loss, diabetes and muscle pain.

SIGN I MAIL

PRINT

SOUND OF IN THEATRES

Statins are considered some of the safest drugs

Merck Pulls Arthritis Drug Vioxx from Market

by RICHARD KNOX

September 30, 2004 12:00 AM ET

Pharmaceutical giant Merck & Co. is pulling its arthritis drug Vioxx from the market after a study confirmed earlier concerns that it raises the risk of heart problems, including heart attacks and stroke. Vioxx is currently used by 2 million people worldwide and has been used by more than 84 million people around the world, according to Merck.

- More than 140,000 cases of serious heart disease
- $4.85 billion for legal claims from US citizens

Data sources of drug safety information in post market stage
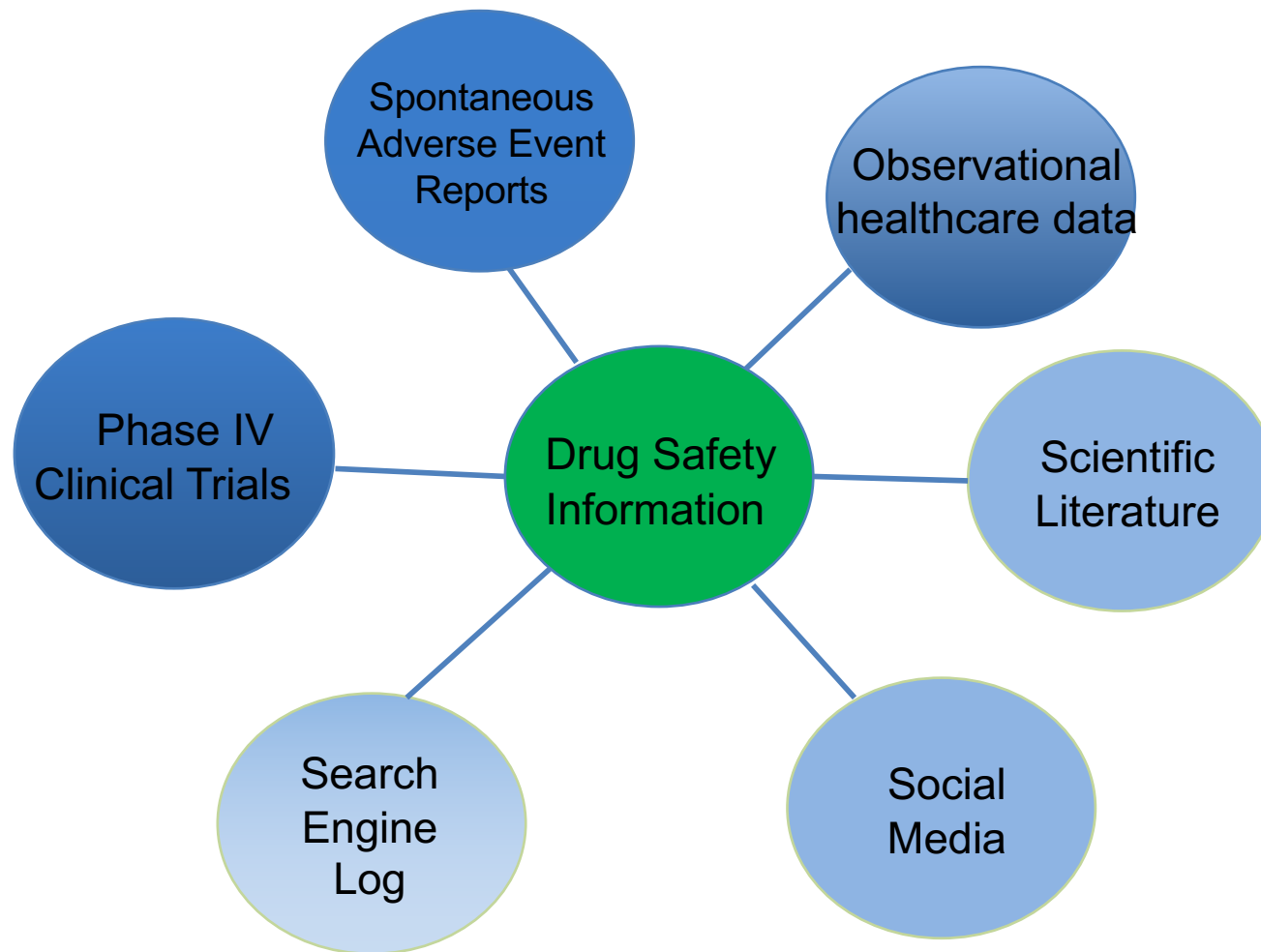
FDA Adverse Event Reporting System (FAERS)

- FDA Adverse Event Reporting System (FAERS)
  - FDA has maintained AERS since 1968
  - Spontaneous reports of suspected ADRs collected from healthcare professionals, consumers, and pharms
  - Data (from Jan 2004 to June 2016) is publicly available at FDA's website!
- Over 5 million reports collected so far:
  - patient: age, sex, weight, country   → Often sparsely collected
  - drugs they are taking
  - diseases they were being treated for
  - the adverse events that occurred to that patient

Source: fda.gov
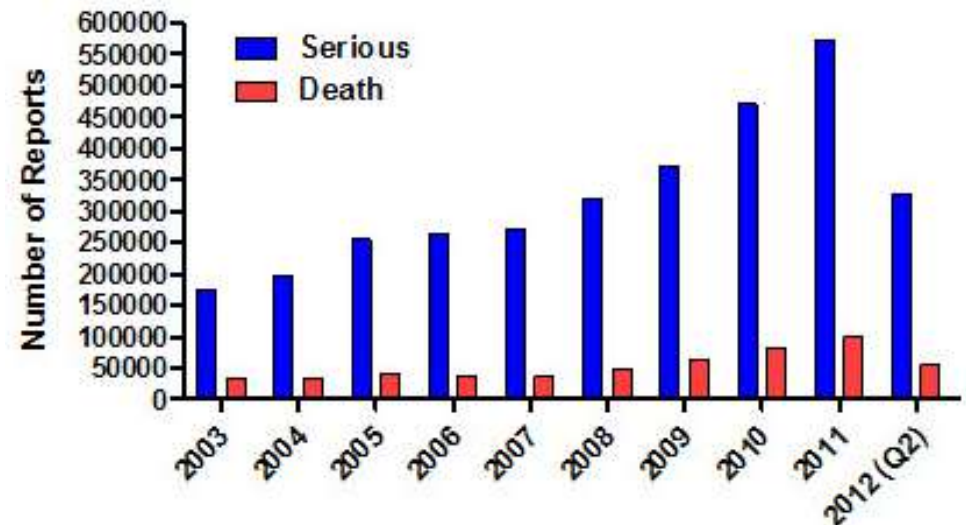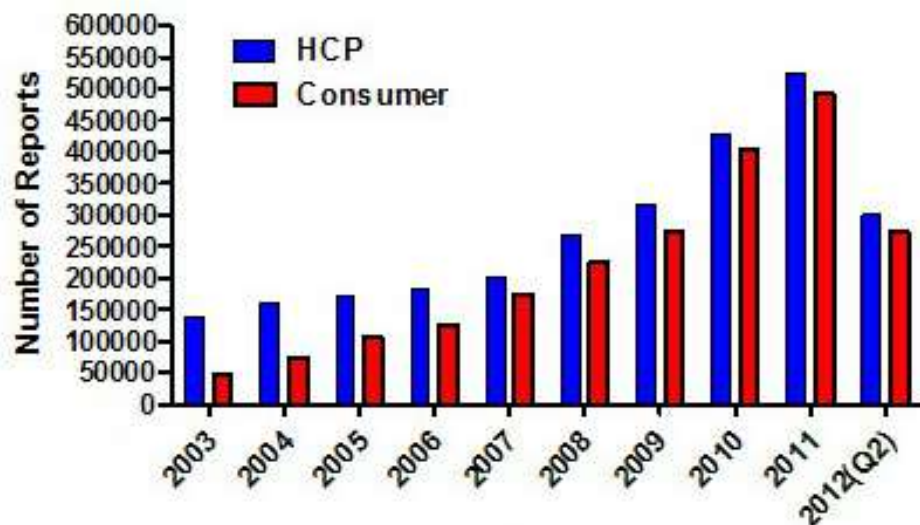
FAERS database structure

DRUG FILE
- *ISR*
- Drug sequence
- Role code
- DrugName
- Route of administration
- Dose
- Dechallenge
- Rechallenge

DEMOGRAPHIC FILE
- *ISR*
- CASE Number
- Event date
- Report date
- Age
- Gender
- Weight
- Occupation code
- Reporter country

THERAPY FILE
- *ISR*
- Drug sequence
- Start therapy
- End therapy
- Duration

OUTCOME FILE
- *ISR*
- Outcome code

RPSR FILE
- *ISR*
- Reporter's source code

INDICATION FILE
- *ISR*
- Drug sequence
- Indication

REACTION FILE
- *ISR*
- Preferred Terrm
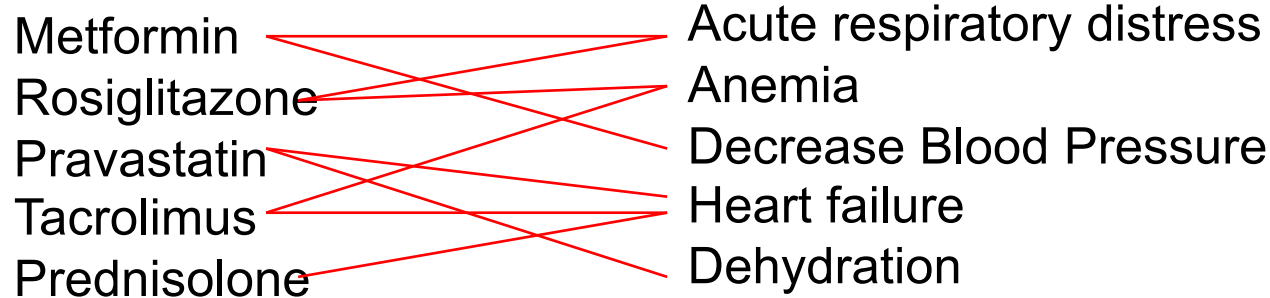
Poluzzi et al. DOI: 10.5772/50095

Interpreting those FAERS reports is hard

- Many drugs, many adverse events
  - what causes what?
  - Most of these red lines are false - which are true?

Drugs

Metformin
Rosiglitazone
Pravastatin
Tacrolimus
Prednisolone

Adverse Events

Acute respiratory distress
Anemia
Decrease Blood Pressure
Heart failure
Dehydration

- Signal detection algorithms for FAERS
  - Quantify "unexpectedness": to identify drugs that have a greater proportion of a particular event compared to the proportion seen for other drugs
  - Sampling variance
    - Underreporting
    - Over reporting
  - Selection biases
    - Causative covariates other than drug under analysis

Disproportionality analysis

	reports w ae	reports w/o ae	Total
reports w drug	a	b	a+b
reports w/o drug	c	d	c+d
Total	a+c	b+d	a+b+c+d

Measure of association	Formula	Probabilistic interpretation
Relative reporting (RR)[1]	$\dfrac{a(a+b+c+d)}{(a+c)(a+b)}$	$\dfrac{\Pr(ae \mid drug)}{\Pr(ae)}$
Proportional reporting rate ratio (PRR)	$\dfrac{a(c+d)}{c(a+b)}$	$\dfrac{\Pr(ae \mid drug)}{\Pr(ae \mid \sim drug)}$
Reporting odds ratio (ROR)	$\dfrac{ad}{cb}$	$\dfrac{\Pr(ae \mid drug)\Pr(\sim ae \mid \sim drug)}{\Pr(\sim ae \mid drug)\Pr(ae \mid \sim drug)}$
Information component (IC)[2]	$\log_2 \dfrac{a(a+b+c+d)}{(a+c)(a+d)}$	$\log_2 \dfrac{\Pr(ae \mid drug)}{\Pr(ae)}$

1. The RR, when implemented within an empirical Bayesian framework, is known as empirical Bayes geometric mean (EBGM); 2. The IC is a logarithmic RR metric that is implemented in a Bayesian framework.

- Modern signal detection algorithms (e.g., EBGM, IC) could address sampling variance
  - Estimate confidence intervals (CIs) for disproportionality statistics
  - Dampen drug-event signals that have little evidence to support them
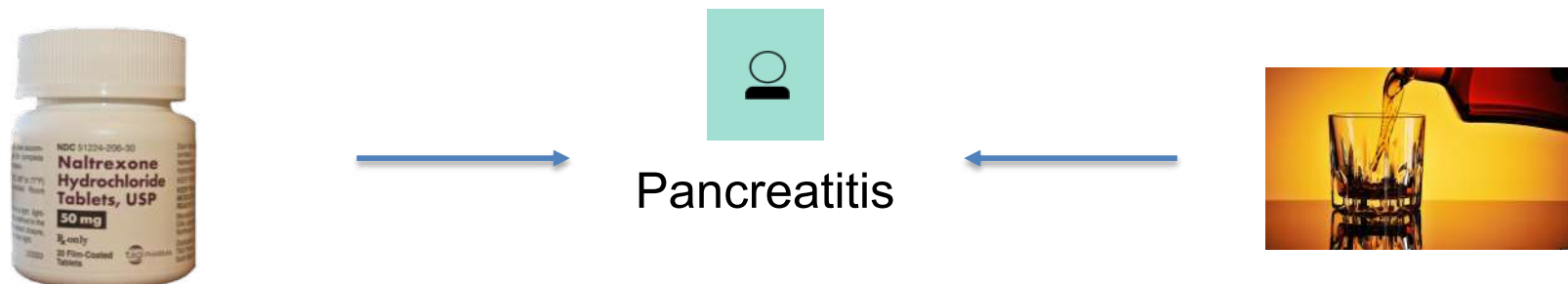- How to address selection biases?

The Confounding Effect poses many challenges for ADR detection of real world events

Co-Prescription Confounders

Mary has arthritis, and has to take painkillers everyday. She has been taking both Aspirin and Vioxx. Which drug caused her heart attack?
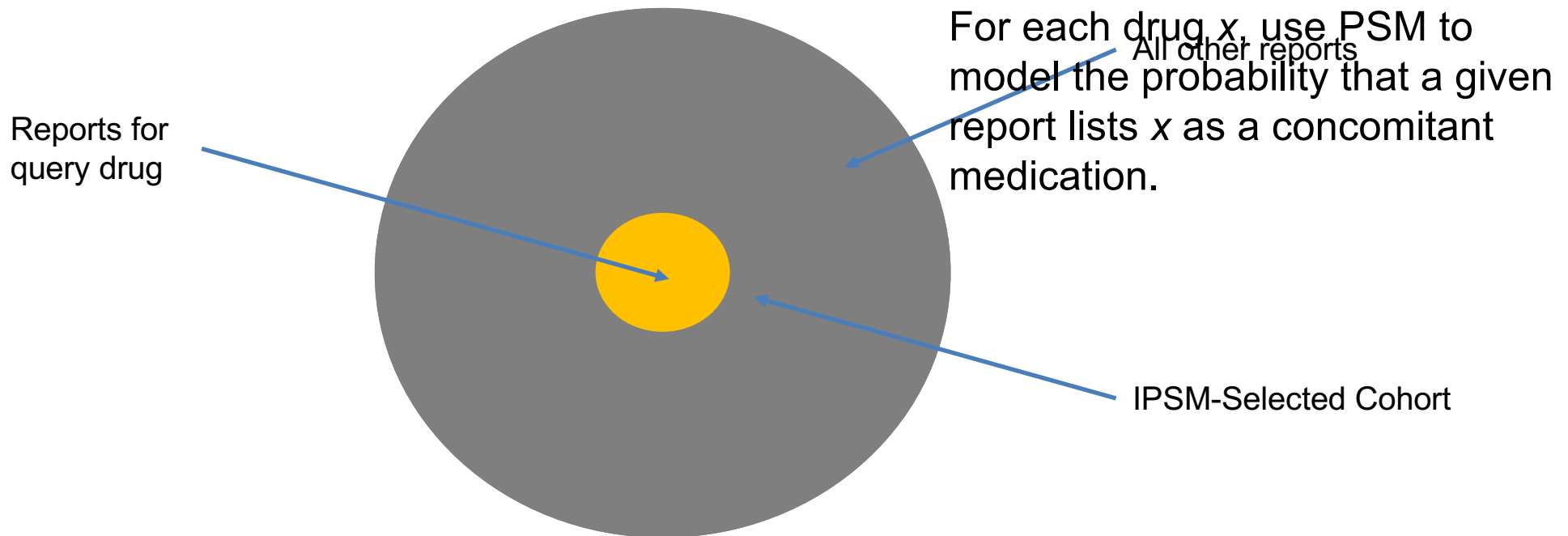
Drug Indicator Confounders

Pancreatitis

Joe is an alcoholic who develops Pancreatitis. He has been drinking daily and taking Naltrexone. What caused the Pancreatitis?

Selection biases in FAERS reports

- Selection biases introduce "synthetic associations"
  - (e.g.) from concomitant drug use (co-Rx effect)
    - drugs co-prescribed with Vioxx more likely to be associated with heart attacks
  - (e.g.) from indications (indication effect)
    - drugs given to diabetics more likely to be associated with hyperglycemia
  - (e.g.) co-Rx effect and indication effect extend to other covariates
    - Patients reported to be taking a cholesterol-lowering agent are more likely to be older, and this may cause these drugs to be synthetically associated with age-related effects, such as hypertension or myocardial infarction (age bias).

- Propensity score matching (PSM) corrects for bias of MEASURED covariates
  - Identify matched controls for the studied cases in observational clinical studies
  - Model the likelihood of a case being selected based on the covariates
    - PS = Estimated Pr(Exposed+| covariates)~ age + sex + weight + ...
  - Match each case with one or more controls with the same likelihood
  - However, PSM requires the covariates to be both known and measured; neither parameter is guaranteed to be present in FAERS

# Implicit Propensity Score Matching (IPSM)

- Invented by Tatonetti NP et al. *Sci Transl Med*. 2012;4(125):125ra31.

- Assumes combination of co-reported drugs and co-indications describes all patient covariates. Generate a probability of a patient receiving a drug given co-prescribed medications and comorbidities.

For each drug *x*, use PSM to model the probability that a given report lists *x* as a concomitant medication.

All other reports

Reports for query drug

IPSM-Selected Cohort

- First, reduce to only those reports that have co-prescribed prescriptions listed

- Second, reduce to only those reports that have correlated indications listed

Takes advantage of co-Rx and indication variables likely to co-vary with unmeasured covariates

IPSM produces better estimates of expected values
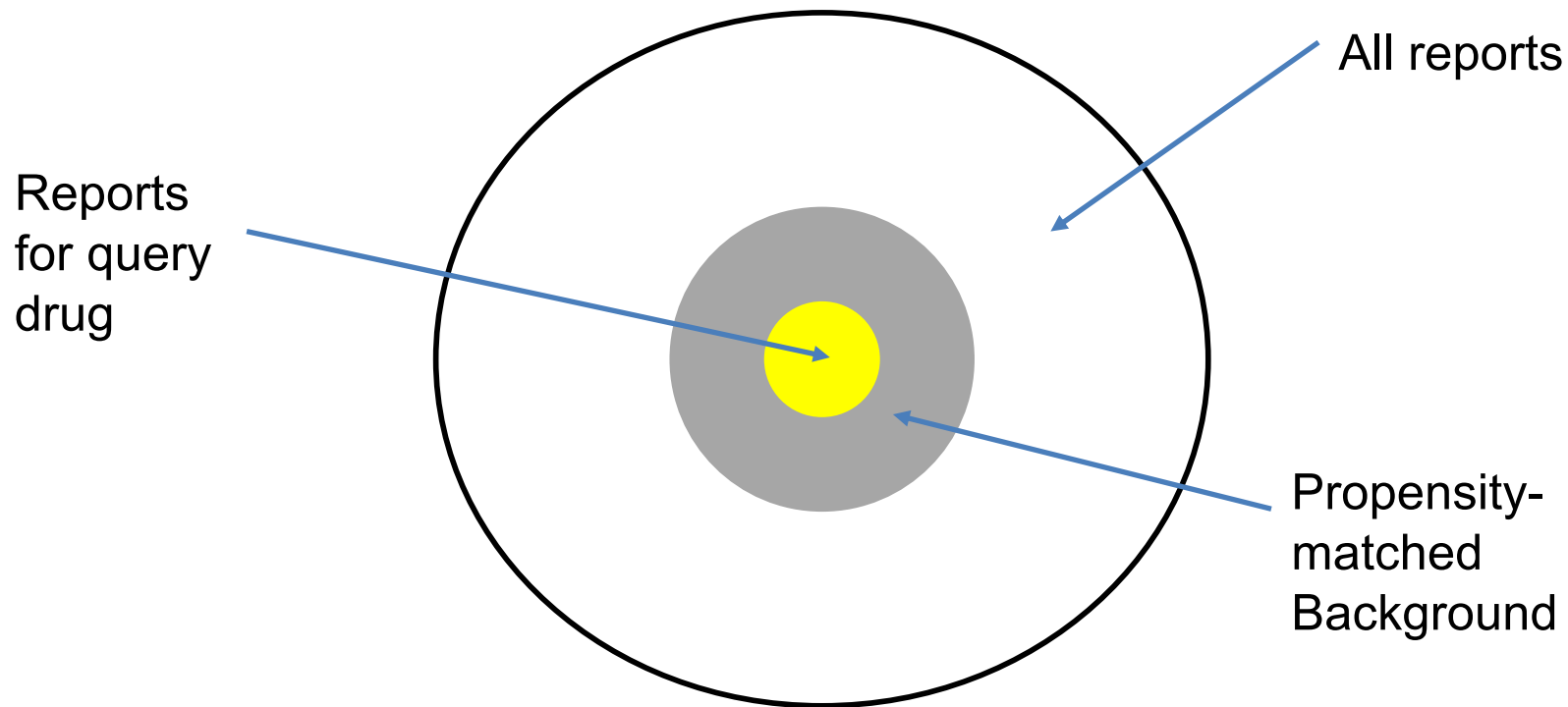
- Example: Reporting of hyperglycemia with diabetes drugs
- **Observed** reporting frequency: 17.7%
- **Expected** Estimates:
    - Entire database expected frequency: 1.5%
        - PRR = 17.7%/1.5% = 11.8!!!!!
    - IPSM-derived expected frequency: 17.6%
        - PRR = 17.7%/17.6% = 1.0 …

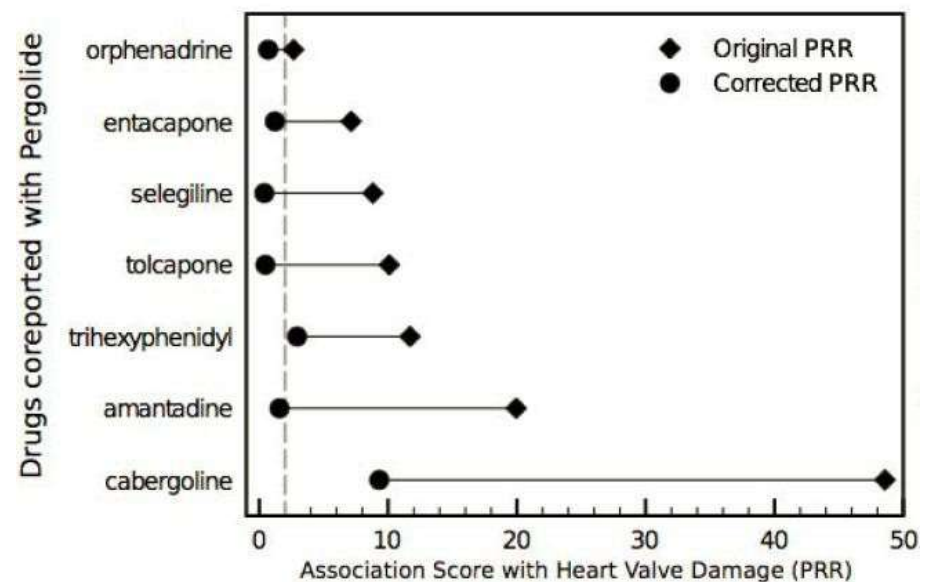# IPSM corrects for indication and co-Rx biases

Drugs given to Diabetics
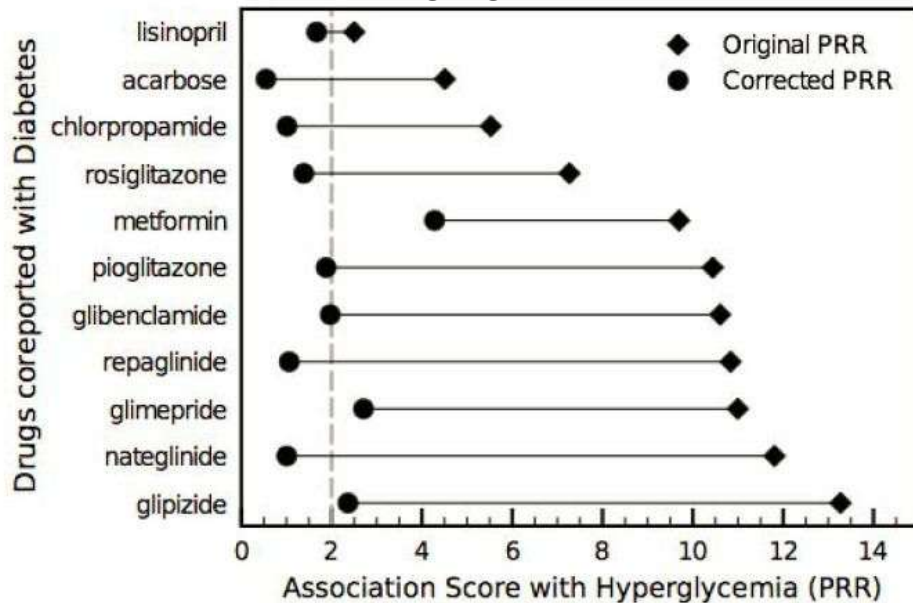
Anti-arrhythmics and Arrhythmia
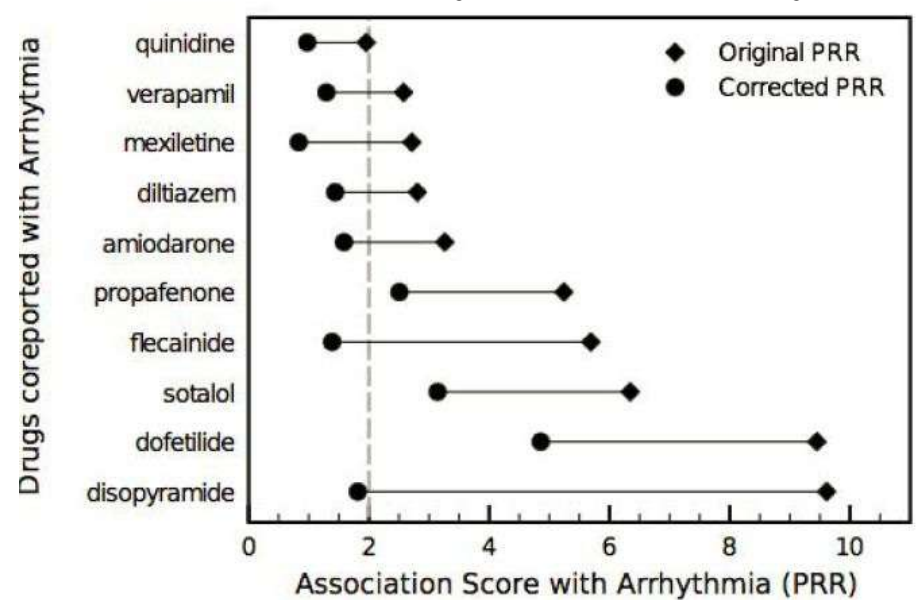
Drugs co-reported with rofecoxib (Vioxx)
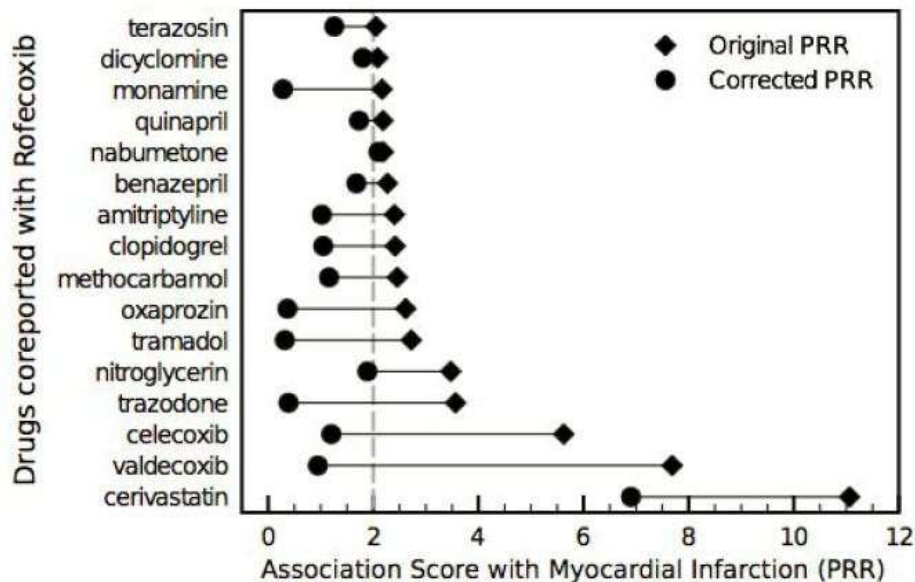
Drugs co-reported with pergolide

IPSM implicit correction for other biases

Drugs preferentially with males are more likely to be associated with 33 sex-related (male) effects
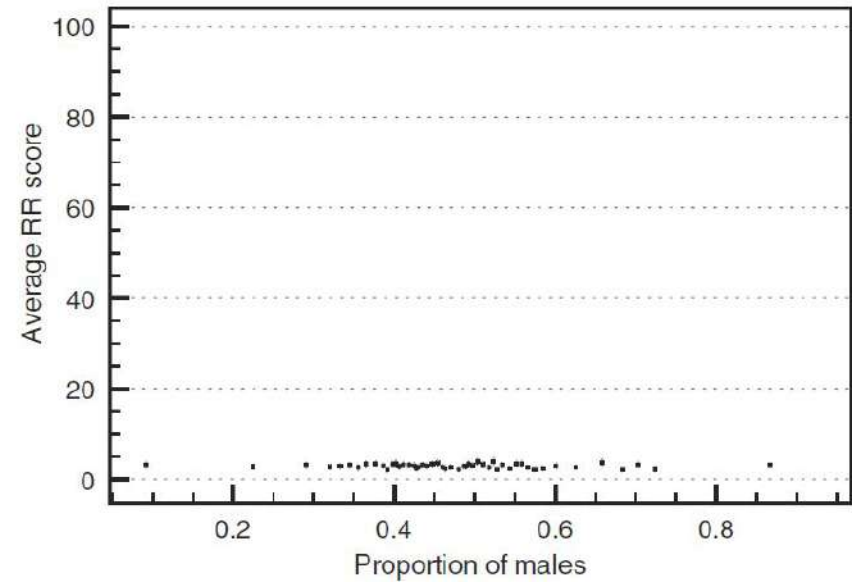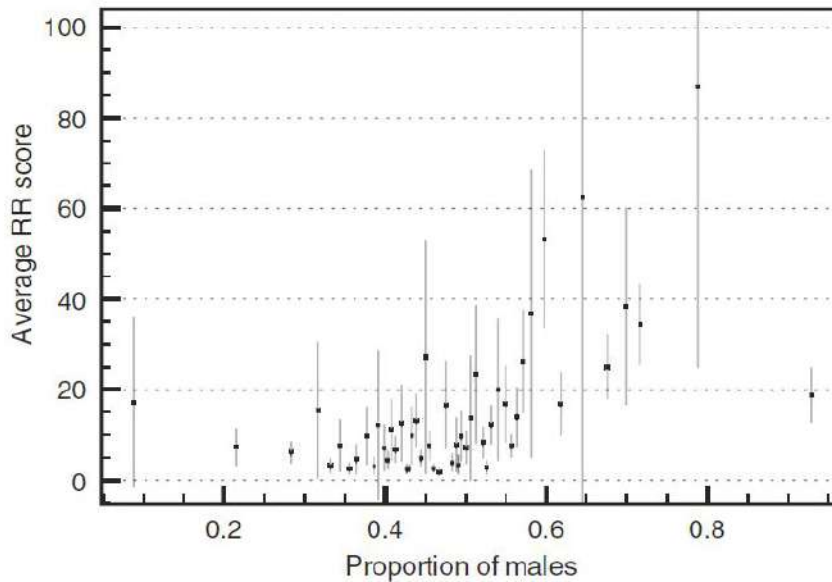
Drugs preferentially with young/old patients are more likely to be associated with 48 age-related effects
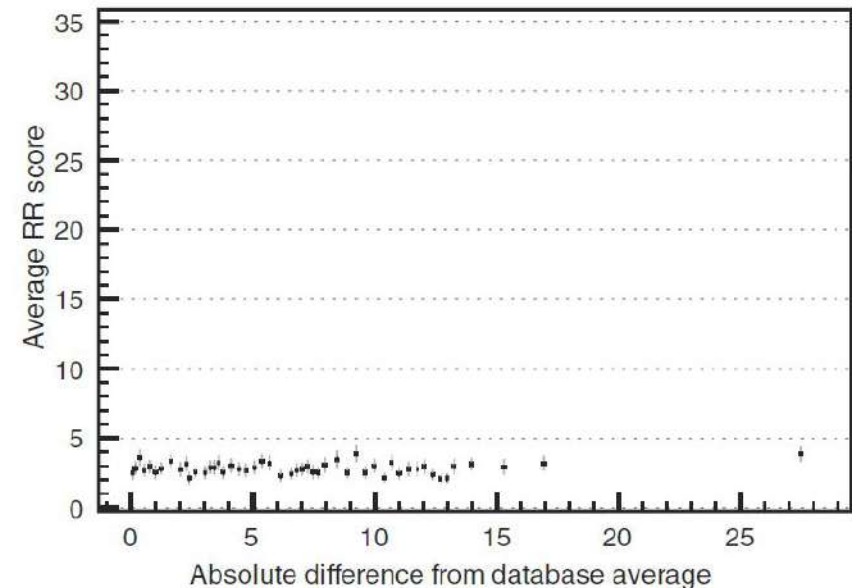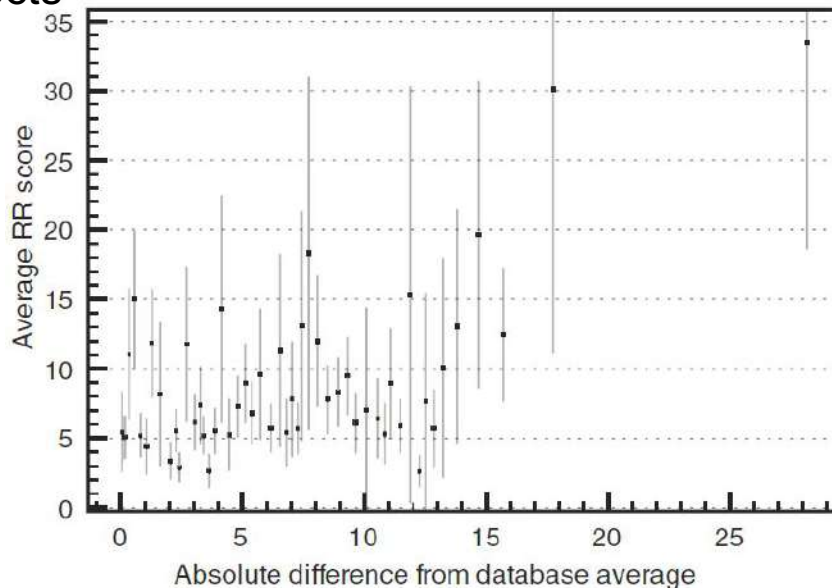
Performance of Pharmacovigilance Signal-Detection Algorithms for FAERS

- Data: FAERS data covered the period from 1968 through 2011 Q3, totaling 4,784,337 reports.

	Method name	Signal score computed
Disproportionality Analysis	Multi-item Gamma Poisson Shrinker (MGPS)	EBGM (empirical Bayes geometric mean): a centrality measure of the posterior distribution of the true observed-to-expected in the population EB05: lower 5th percentile of the posterior observed-to-expected distribution
	Proportional Reporting Ratio (PRR)	PRR: point estimate (mean) of the relative risk reporting ratio distribution PRR05: lower 5th percentile of the relative risk reporting ratio distribution
	Reporting Odds Ratio (ROR)	ROR: point estimate (mean) of the reporting odds ratio distribution ROR05: lower 5th percentile of the reporting odds ratio distribution
Multivariate Modeling	Logistic Regression (LR)	LR: point estimate of the odds ratio distribution obtained from logistic regression LR05: lower 5th percentile of the odds ratio distribution obtained from logistic regression
	Extended Logistic Regression (ELR)	ELR: point estimate of the odds ratio distribution obtained from extended logistic regression ELR05: lower 5th percentile of the odds ratio obtained from extended logistic regression

Harpaz, Rave, et al. "Performance of Pharmacovigilance Signal-Detection Algorithms for the FDA Adverse Event Reporting System." Clinical Pharmacology & Therapeutics 93.6 (2013): 539-546.

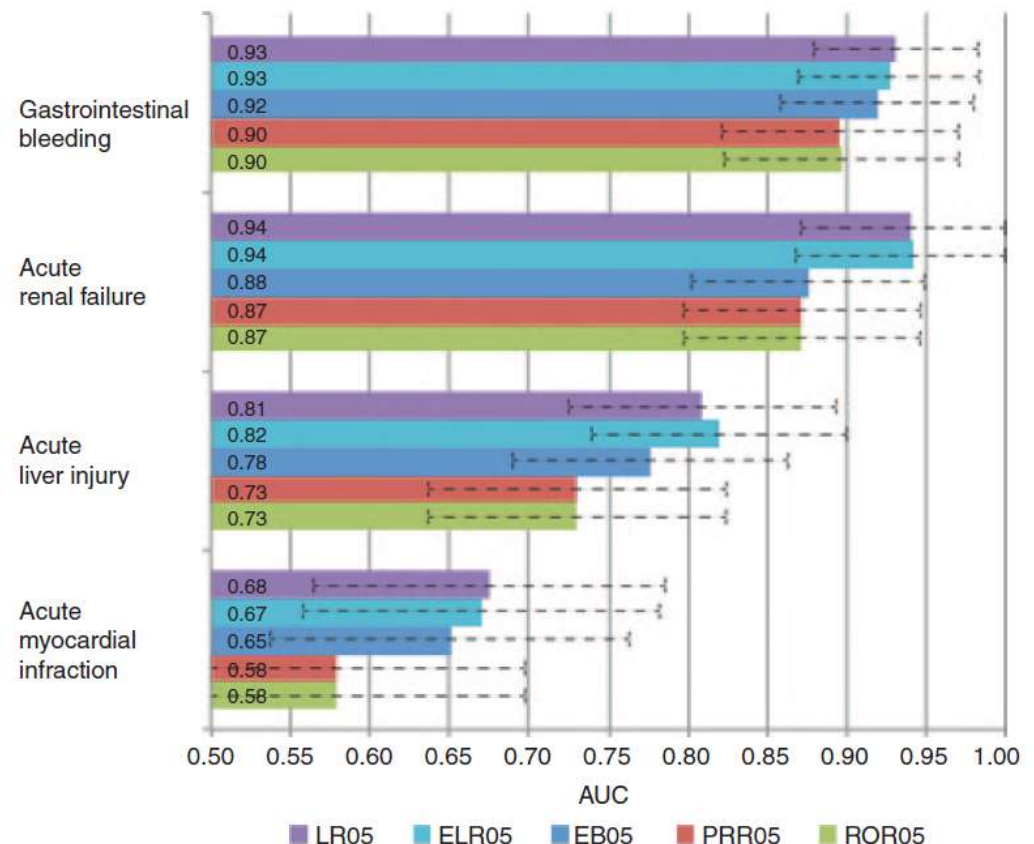# The application of biomedical gold standards

Positive Drug Set for an ADR:

- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"[35]
- Literature review identified no powered studies with refuting evidence of effect

Negative Set:

- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"[35]
- Literature review identified no powered studies with evidence of potential positive association

Event	Positive Cases	Negative Case	Total
Gastrointestinal Bleeding	24	67	91
Acute Liver Injury	80	37	117
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Total	164	234	398

Harpaz, R, et al. **Clinical Pharmacology and Therapeutics**. 2013

Summary - strengths and weaknesses of notable signal detection methods

Triaging to select signals and follow up

QUANTITATIVE "RULES"

- Apply fixed thresholds
  - EB05 ≥2; EBGM ≥2; EBGM ≥4;
  - PRR ≥2; a number of reports (N) ≥3; a Chi-square ≥4
  - Lower 95% CI of PRR ≥1
  - Lower 95% CI of ROR ≥1
  - IC025 > 0
- Apply flexible thresholds
  - Estimate the false discovery rate (FDR) to decide threshold on a signal-by-signal basis

QUALITATIVE "RULES"

- Novel
  - Not currently known and on drug label
  - New adverse event or new drug ("early warning")
- High potential relevance
  - Public health issue – e.g. important drug (serious indication, widely used), serious reaction, many cases
  - Change in merit/harm
- Strong evidence
  - Exposure-response relationship (site, time-to-onset, dose, reversibility in dechallenge/rechallenge)
  - Reasonable from a biological mechanism perspective
- Time trend
  - Surge in recent reporting, notable increase in reporting over time

Meyboom RH, et al. Drug safety. 2002 May 1;25(6):459-65.

Beyond ADR detection

Common drug combo increases diabetes risk

| Hypothesis generation based on FAERS | → | Signal validation based on EHR databases | → | Mice model validation |
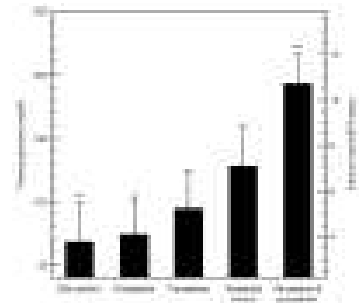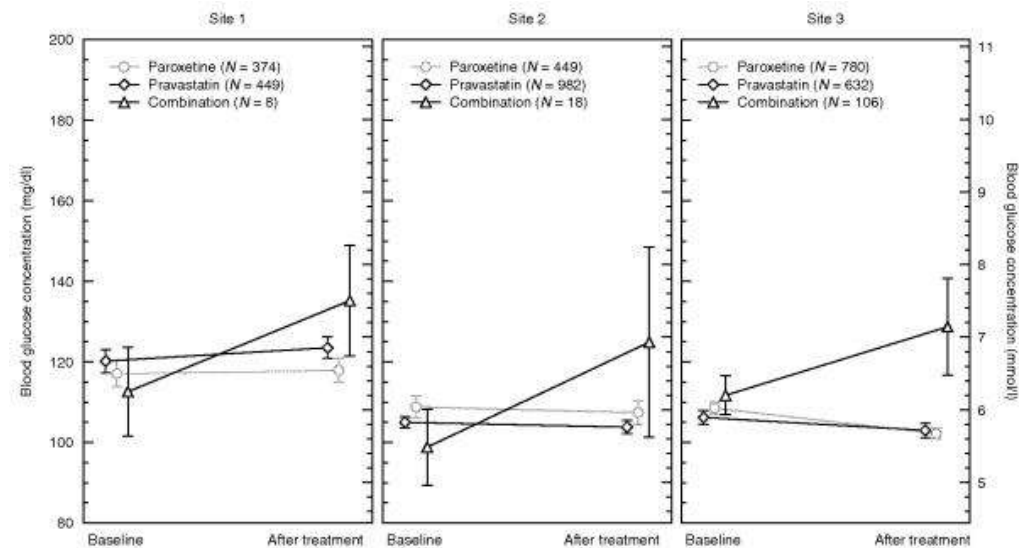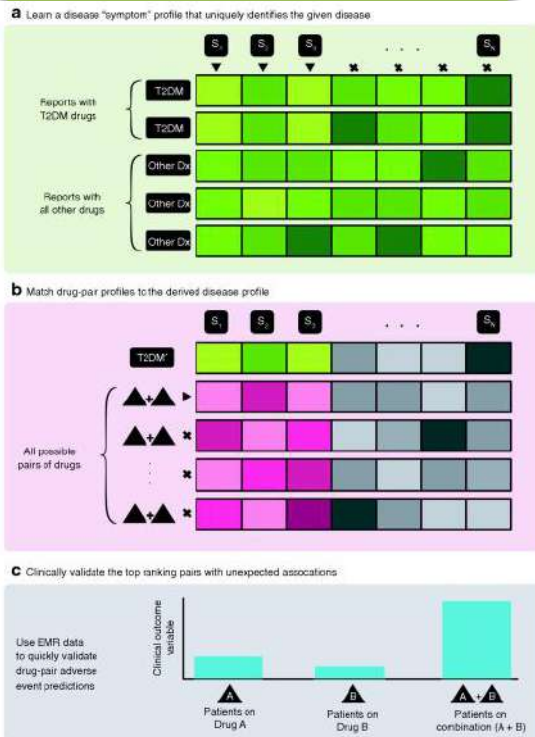
A combination of two common drugs – paroxetine (one an antidepressant), pravastatin (the other used to lower blood cholesterol) – that caused blood sugar to rise, may put people at risk of developing diabetes.

Beyond ADR detection

Common drug combo decreases ADRs

Data to knowledge generator

Structured and normalized spontaneous reporting systems

Uni-variate feature selection module

Propensity score computing module

Drug-ADR-Drug associations prediction Module

Outputs:
- Drugs that could reduce other drugs-induced ADRs
- Link drug combinations to their potential clinical effects
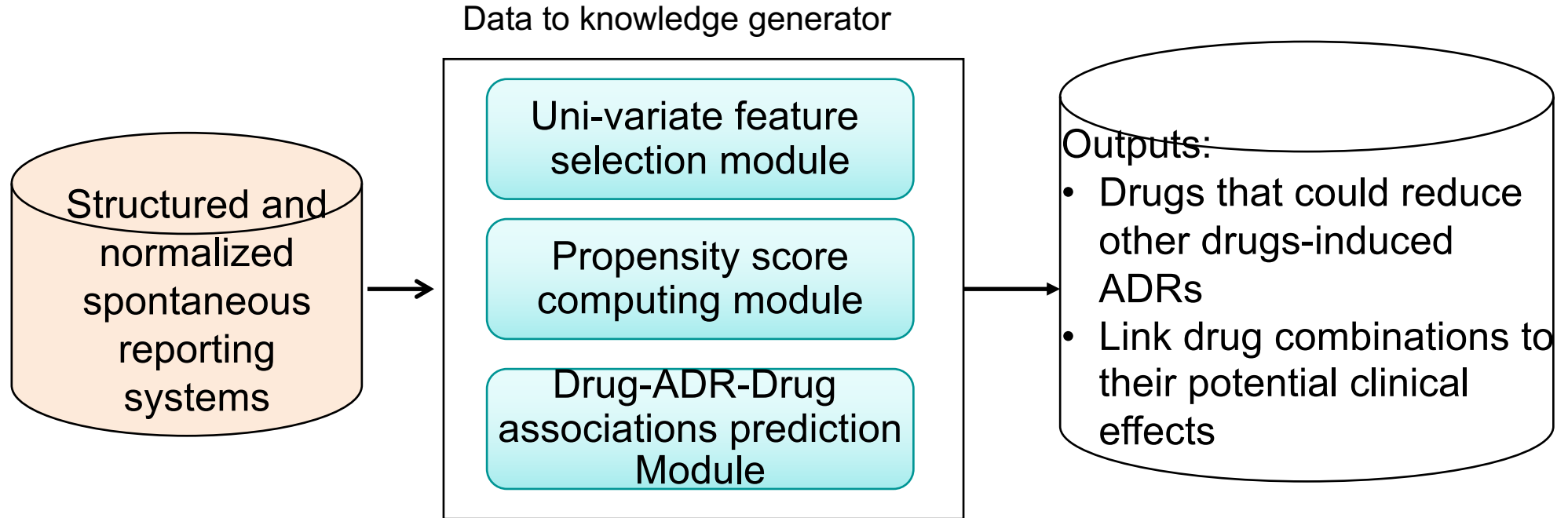
$$logit(P(ADR = 1)) = \beta_0 + \beta_1 DrugA + \beta_2 P_1 + \beta_3 DrugB + \beta_4 P_2 + \beta_5 DrugA * DrugB + \lambda|\beta|_1$$

The novel regularized logistic regression is able to reveal two different mechanism of drug combinations
- $(\beta_3 + \beta_5)$: the degree that a patient who is on Drug A could benefit or suffer from taking Drug B for the ADR of interest
- β_5: the degree that the interaction effect between Drug B and Drug A on the ADR

Li., Zhang., et al. AMIA 2016.

Clinical Validation

List of 15 predicted beneficial drug combinations and their ADR reduction

Pamidronate is used to treat high blood calcium levels

Drug A name	ADRs associated with drug A	Drug B name	Predicted beneficial score	Common ATC code	Evidence for combined use
benazepril	DIZZINESS	amlodipine besylate	-0.57	yes	F
atovaquone	PYREXIA	proguanil	-0.36	yes	F
rofecoxib	MYOCARDIAL INFARCTION	pamidronate	-0.33	yes	
rosiglitazone	MYOCARDIAL INFARCTION	exenatide	-0.32	yes	
progesterone	BREAST CANCER	adalimumab	-0.27	no	
trimethoprim	PYREXIA	sulfamethoxazole	-0.17	yes	F
exemestane	ARTHRALGIA	everolimus	-0.16	yes	III
amoxicillin	DIARRHOEA	clavulanic acid	-0.15	yes	IV
ampicillin	PYREXIA	sulbactam	-0.15	yes	F
desmopressin	HYPONATRAEMIA	somatropin	-0.15	yes	
sertraline	ANXIETY	nicotinic acids	-0.14	no	
sumatriptan	MIGRAINE	naproxen	-0.14	no	F
olanzapine	DIABETES MELLITUS	biperiden	-0.13	yes	
clindamycin	DIARRHOEA	benzoyl	-0.13	yes	F
fluticasone	DYSPNOEA	salmeterol	-0.13	yes	F

F: FDA approved drug combination; III: phase III clinical trial; IV: phase IV clinical trial

a NSAID. On September 30, 2004, Merck withdrew rofecoxib from the market because of concerns about increased risk of heart attack and stroke associated with long-term, high-dosage use.

Our commitment to Health – IBM Moonshot

"I'm telling you, our moonshot will be the impact we will have on Healthcare. It has already started. We will change and do our part to change the face of Healthcare. I am absolutely positive about it. And that, to me, while we do many other things, that will be one of the most important."

Ginni Rometty
IBM Chairman, President and CEO
April 16, 2015

IBM Life Sciences Solutions	Accelerated Product Innovation — Advance next generation discovery and development	Commercial Transformation — Act on insights to drive value	Analytics-Driven Care Management — Empower people to make better decisions to improve outcomes
IBM Watson Health	Data — Structured & Unstructured	Insights — Cognitive & Advanced Analytics	Solutions — IBM & Ecosystem Solutions
Key Acquisitions	explorys TRUVEN HEALTH ANALYTICS MERGE PHYTEL CÚRAM SOFTWARE The Weather Company		

Center for Computational Health @ IBM
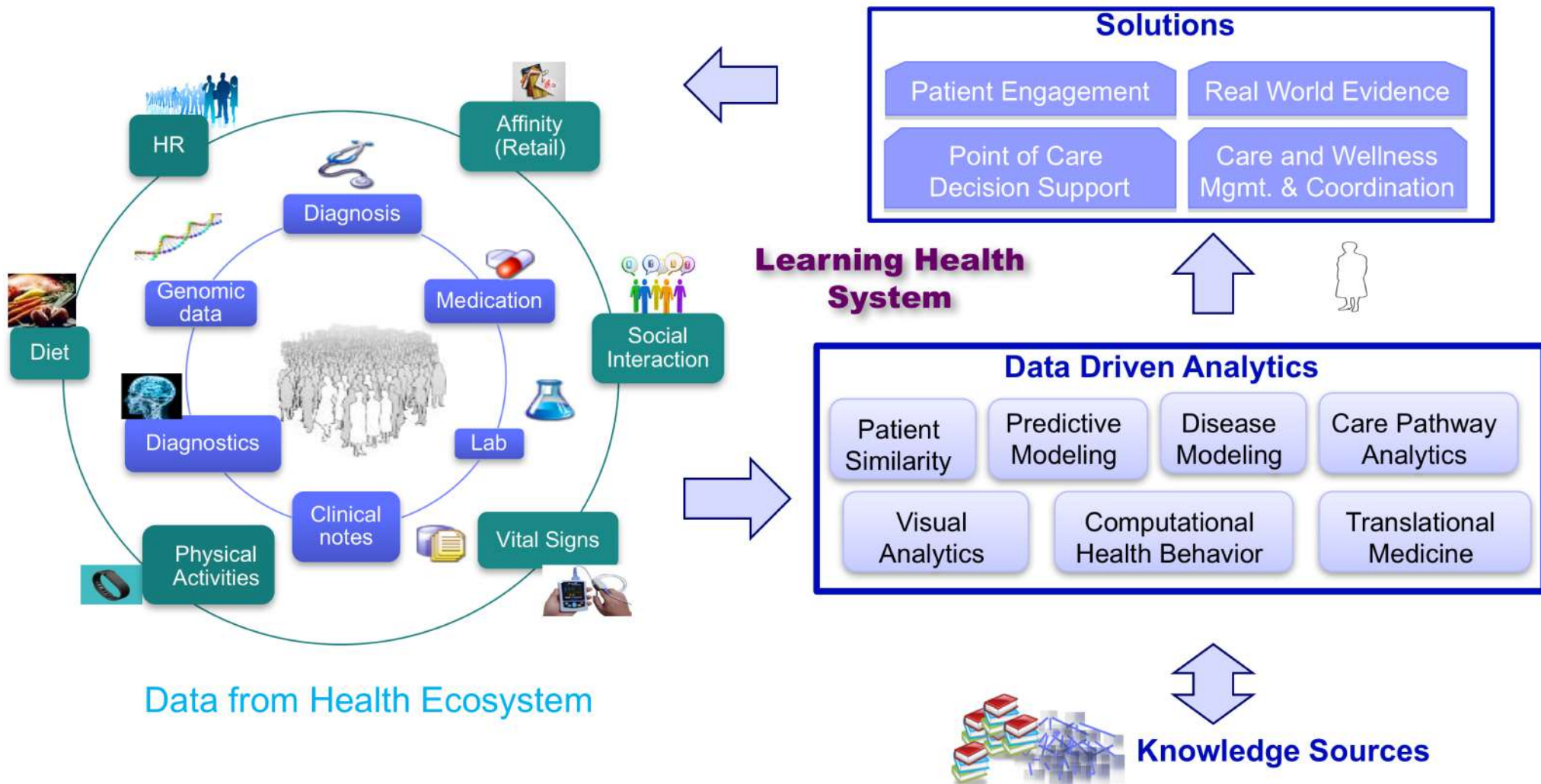
Multiple Positions Available:
- Interns
- Research Scientists
- Research Engineers

Contact:
pzhang@us.ibm.com