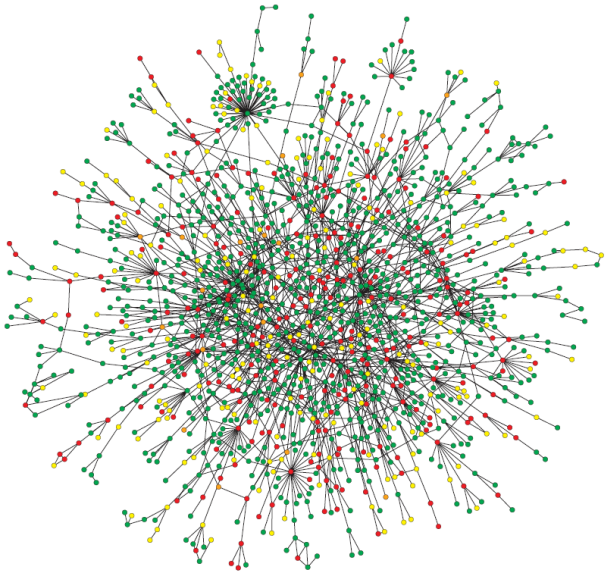
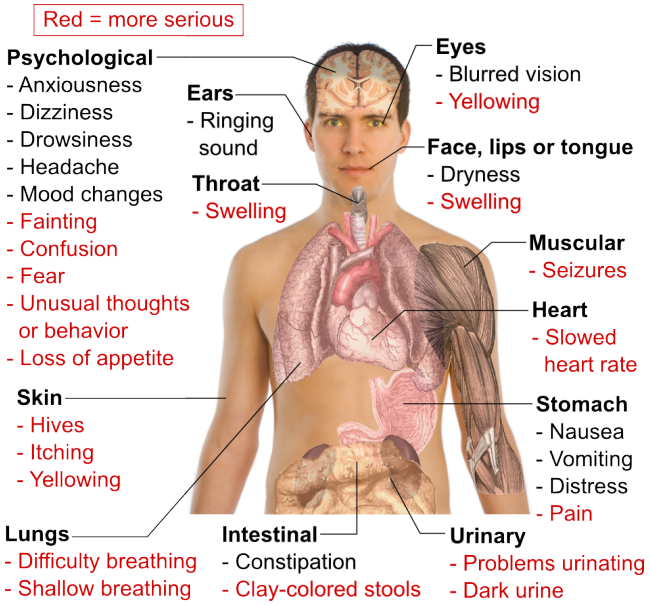
From Chemical to Systems Biology: How Chemoinformatics can contribute?

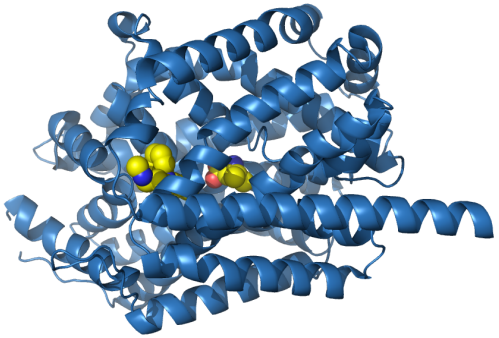




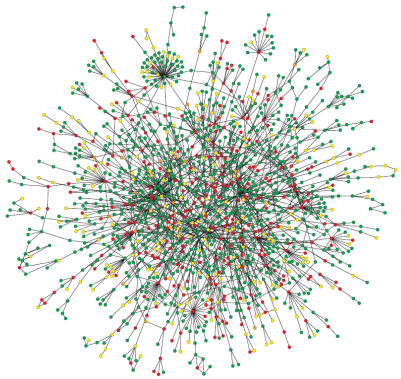
Computational Chemical Biology

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

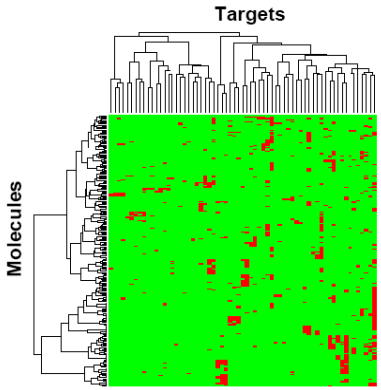
Objective: **Understand the relationship between chemical actions (environmental chemicals, drugs, natural products) and disease susceptibility genes.**



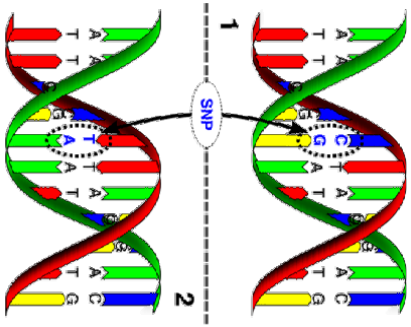
Chemoinformatics



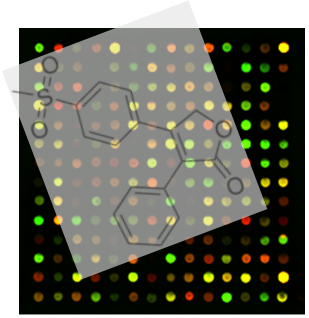
Biological networks



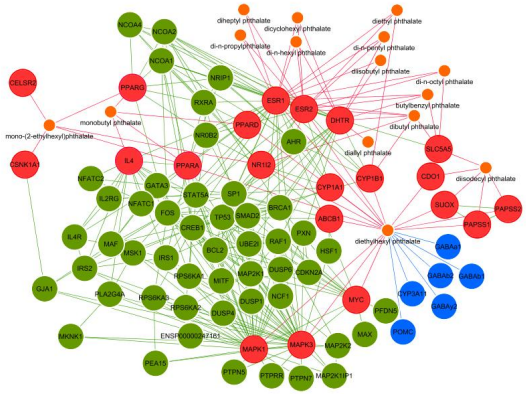
Gene expression data analysis



Functional human variation

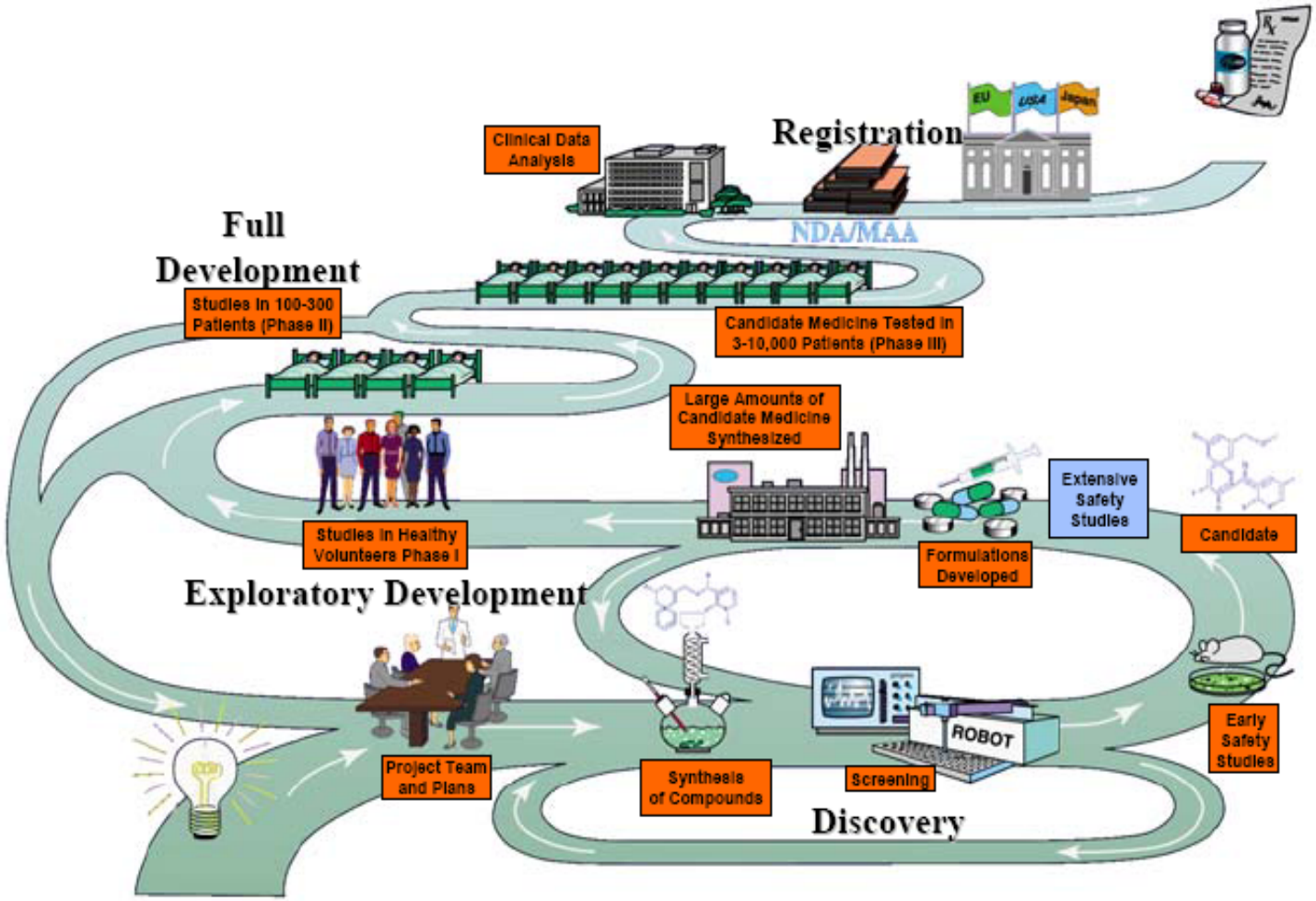


Toxicogenomics



Integrative Chemical Biology

The Long Road to a New Drug



From Chemical to Systems Biology

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Systems chemical biology

Tudor I Oprea, Alexander Tropsha, Jean-Loup Faulon & Mark D Rintoul

The increasing availability of data related to genes, proteins and their modulation by small molecules has provided a vast amount of biological information leading to the emergence of systems biology and the broad use of simulation tools for data analysis. However, there is a critical need to develop cheminformatics tools that can integrate chemical knowledge with these biological databases and simulation approaches, with the goal of creating systems chemical biology.

Small compounds

Structural information
Bioactivity information

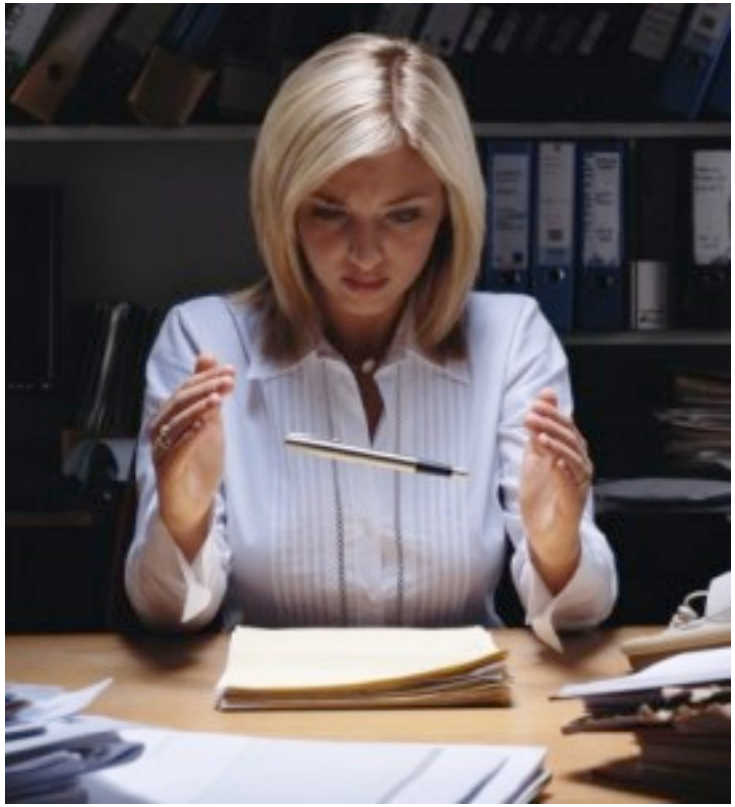
Human body

Biological pathways
Protein-protein interactions
Gene expression data
Disease phenotypes
Side effect data,
etc... etc...

POLYPHARMACOLOGY
CHEMOGENOMICS
NETWORK PHARMACOLOGY
SYSTEMS PHARMACOLOGY

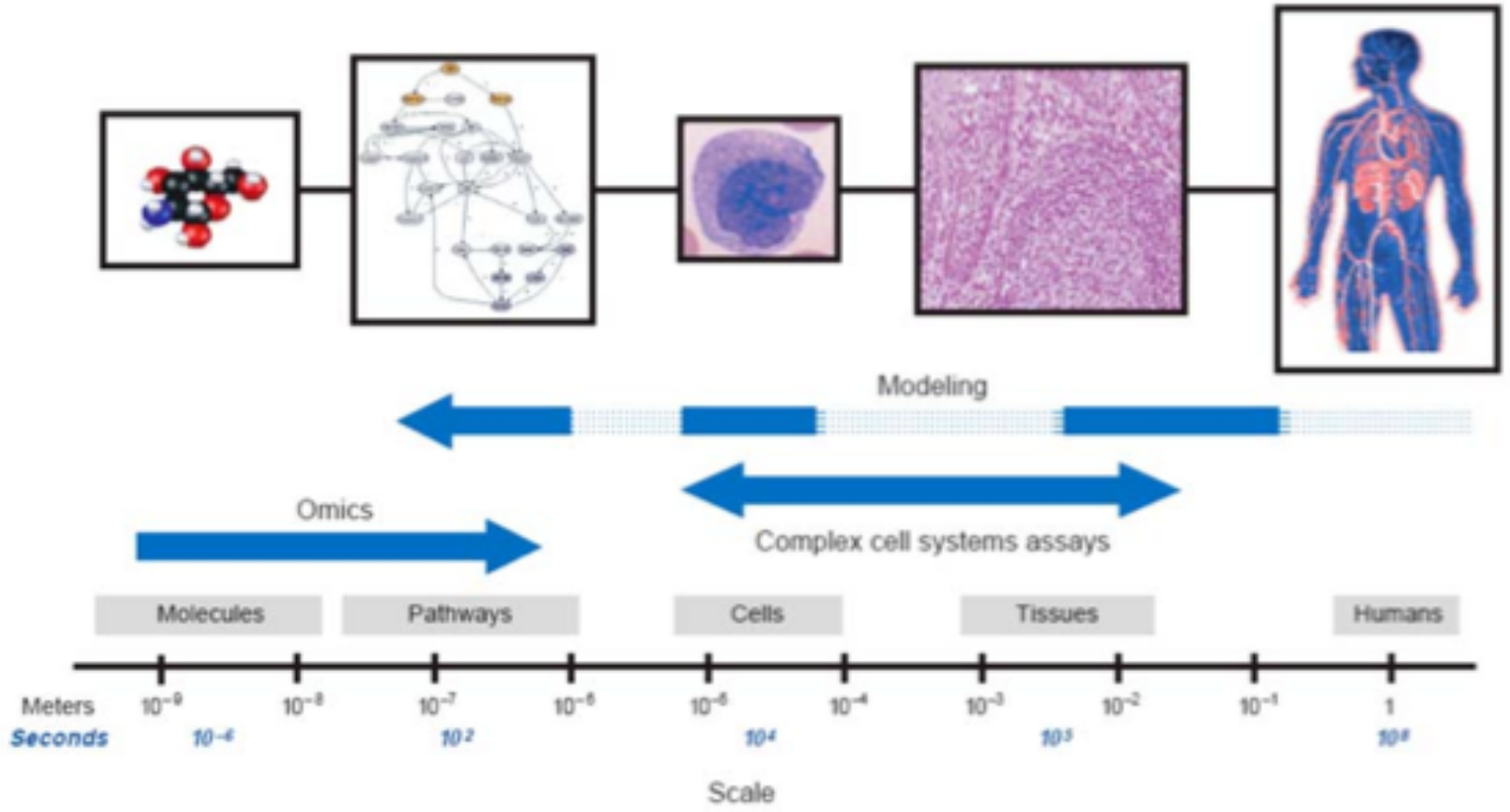
Oprea *et al. Nature Chem Biol* (2007) 3, 447-450

How can we do that?



Many possibilities…

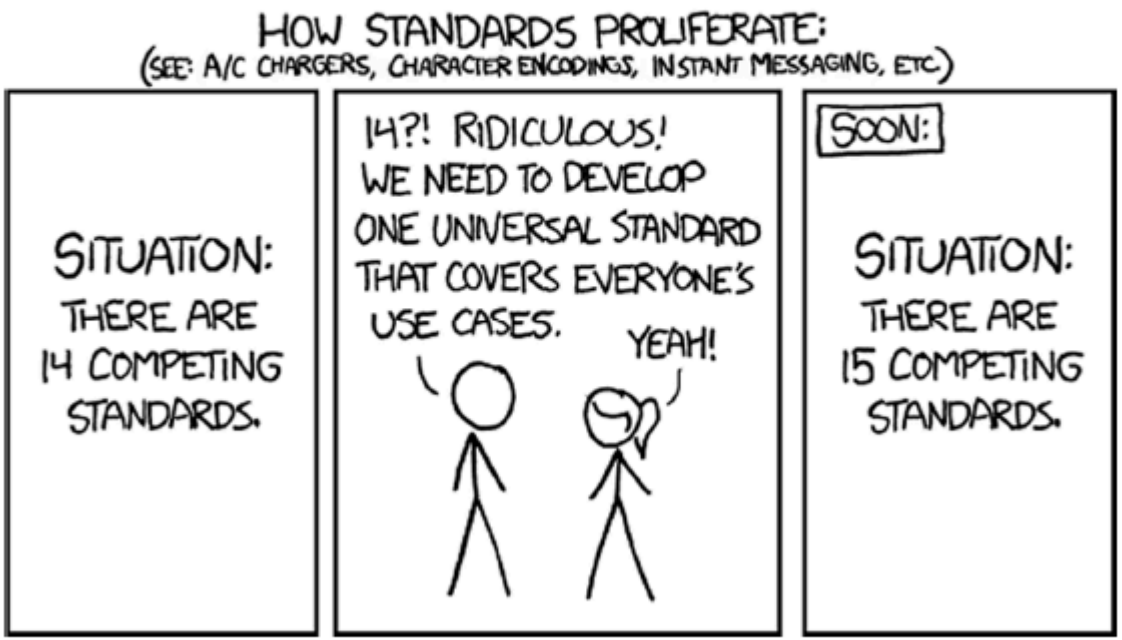
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Butcher EC, Berg EL, Kunkel EJ. *Systems biology in drug discovery. Nat Biotech 2004; 22: 1253-9.*

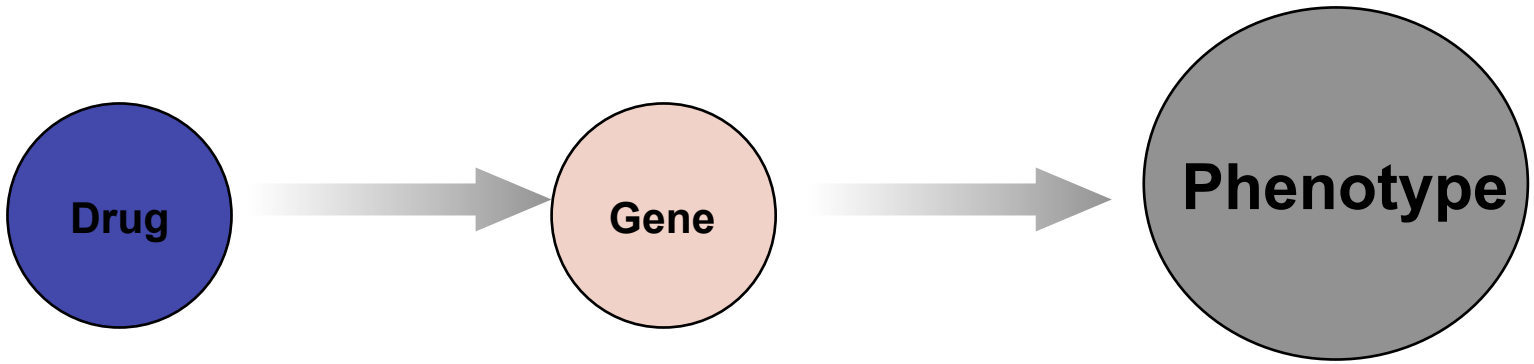
Where to start?

A Meta-Database?



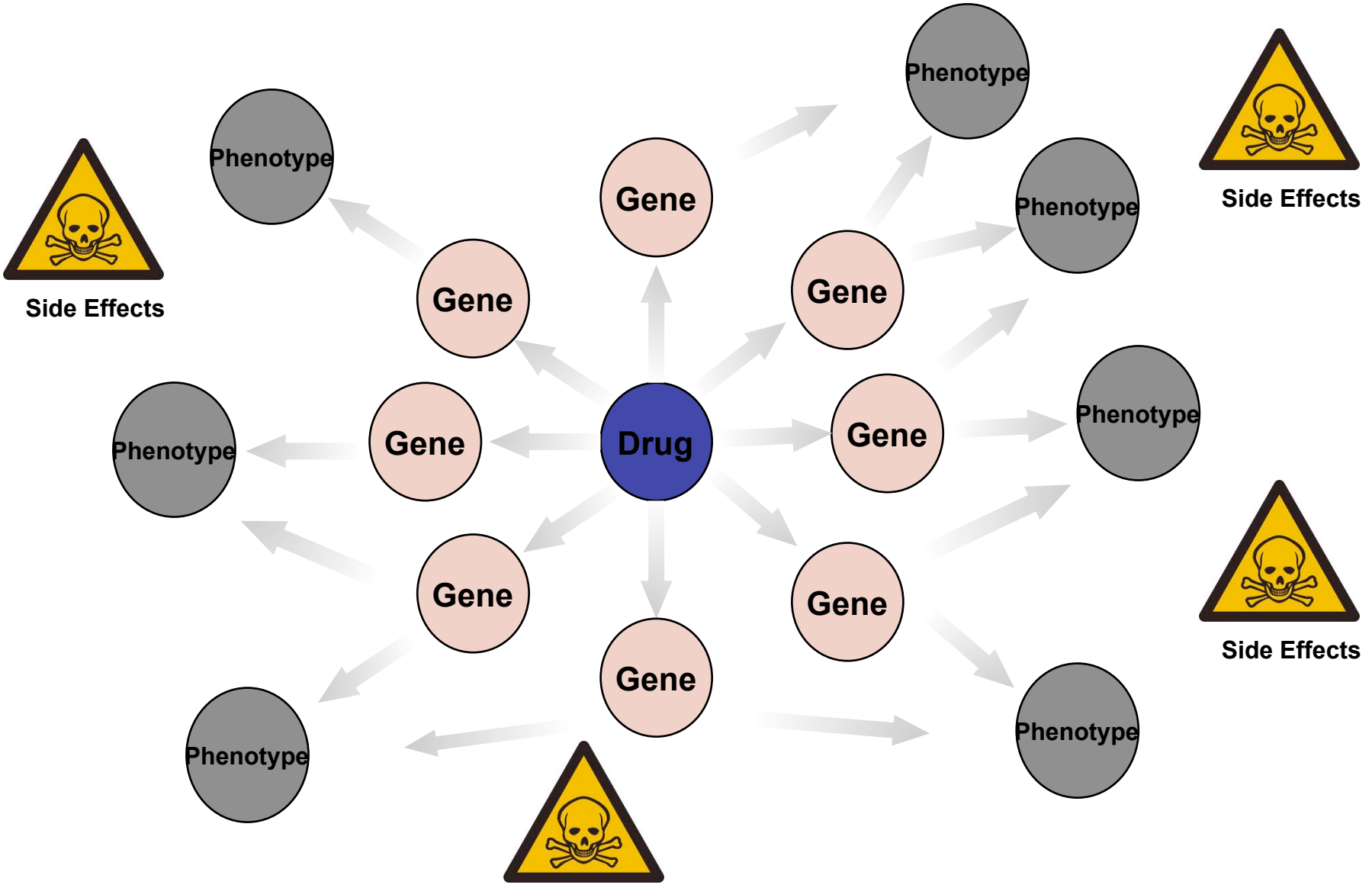
http://xkcd.com/927/

We hope for a simple concept…



But in reality it is not so simple

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



What is the number of targets for a drug?

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

DrugBank → 4400 drugs, 2.7 targets/drug in average

Wombat-PK → 1081 drugs, 5.69 targets/ drug in average

The topology of drug–target interaction networks: implicit dependence on drug properties and target families†‡

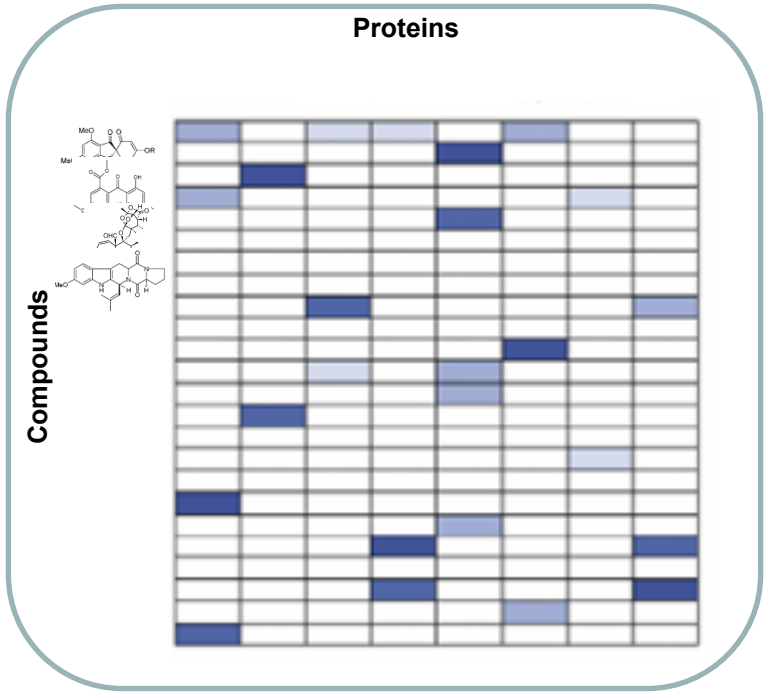
Jordi Mestres,*[a] Elisabet Gregori-Puigjané,[a] Sergi Valverde[bc] and Ricard V. Solé[bd]

Received 23rd March 2009, Accepted 26th May 2009
First published as an Advance Article on the web 8th July 2009
DOI: 10.1039/b905821b

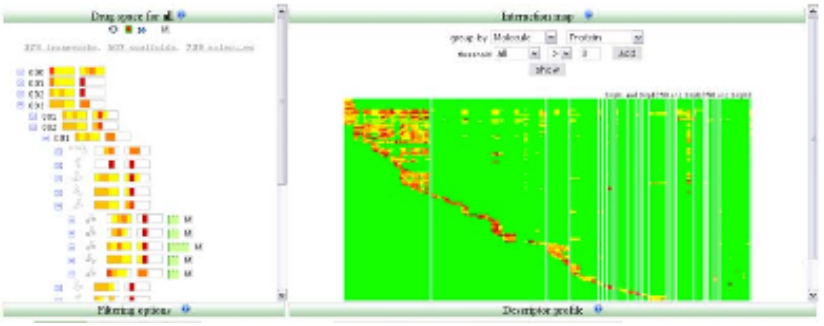
The availability of interaction data between small molecule drugs and protein targets has increased substantially in recent years. Using seven different databases, we were able to assemble a total of 4767 unique interactions between 802 drugs and 480 targets, which means that on average every drug is currently acknowledged to interact with 6 targets. The application of network theory to the analysis of these data reveals an unexpectedly complex picture of drug–target interactions. The results confirm that the topology of drug–target networks depends implicitly on data completeness, drug properties, and target families. The implications for drug discovery are discussed.

The pharmacology of a drug is still sparse

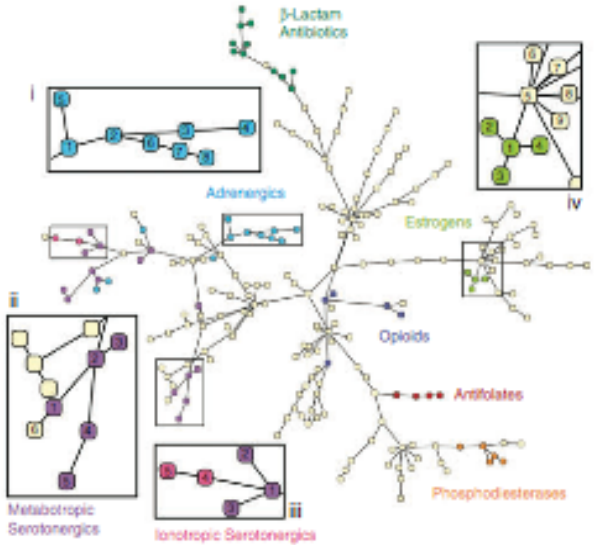
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Chemical similarity



Garcia-Serna R et al. Nat. Bioinformatics 2010



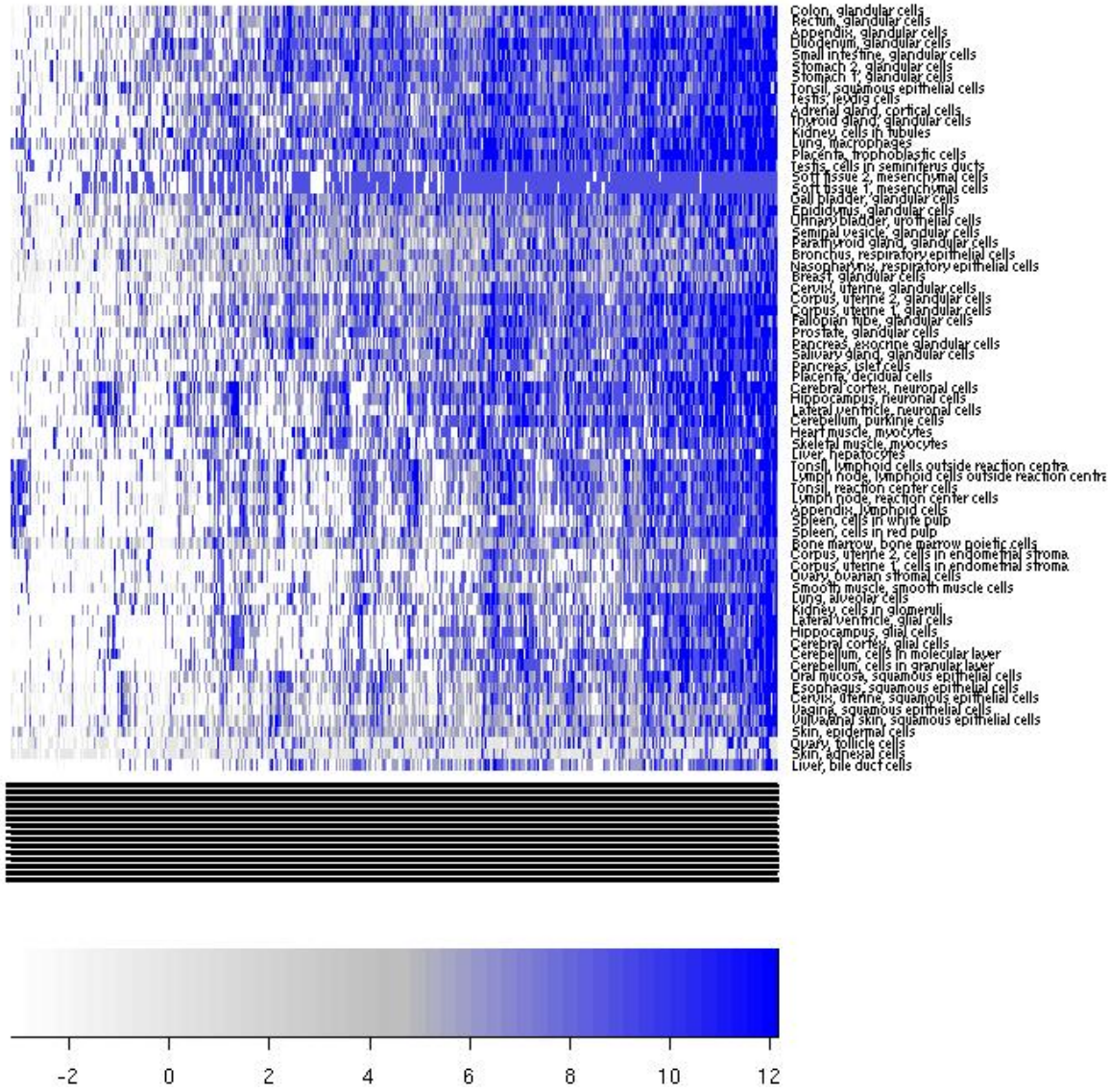
Keiser MJ et al. Nat. Biotech 2007

Drug-target network

Yildirim M et al. Nat. Biotech 2007

Genes-tissues specificity

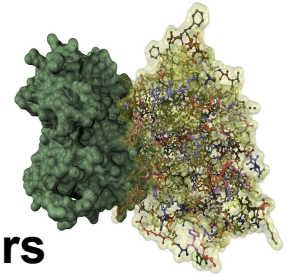
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

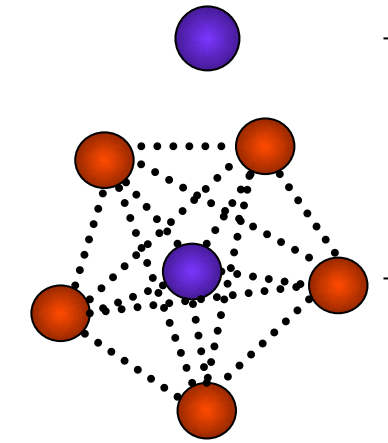


What about phenotypes?

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

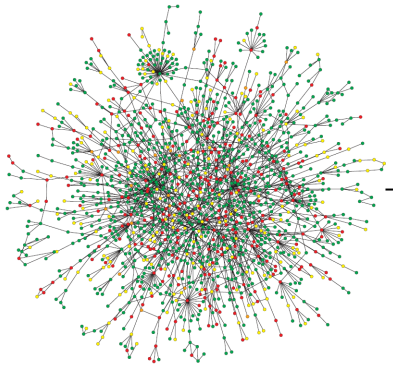
Protein-Protein interactions Network (PPI)





disease — **Especially genetic disorders (color blindness, Huntington's disease, Cystic fibrosis)**

disease — **Prader-Willi syndrome (7 genes)**



disease — **Cancer, diabetes, mental illness**

A quality-controlled human protein interaction network

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

MINT

BIND

DIP

GRID

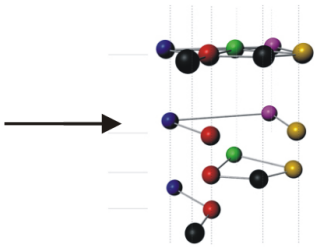
HPRD

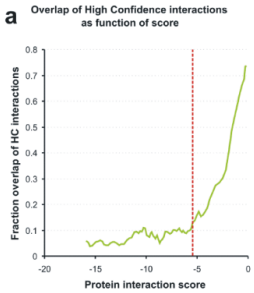
KEGG

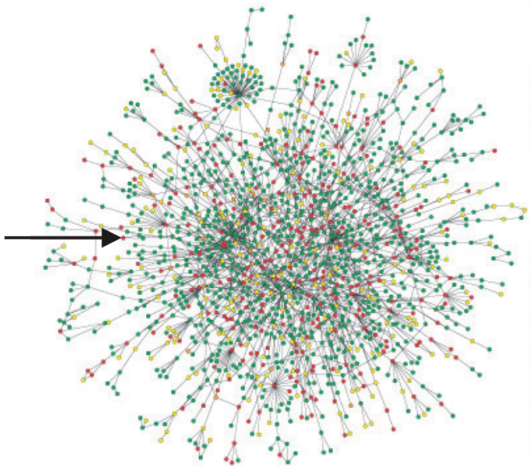
REACTOME

IntAct

Curated







Download and reformat PPI databases

Trans organism ppi transferral

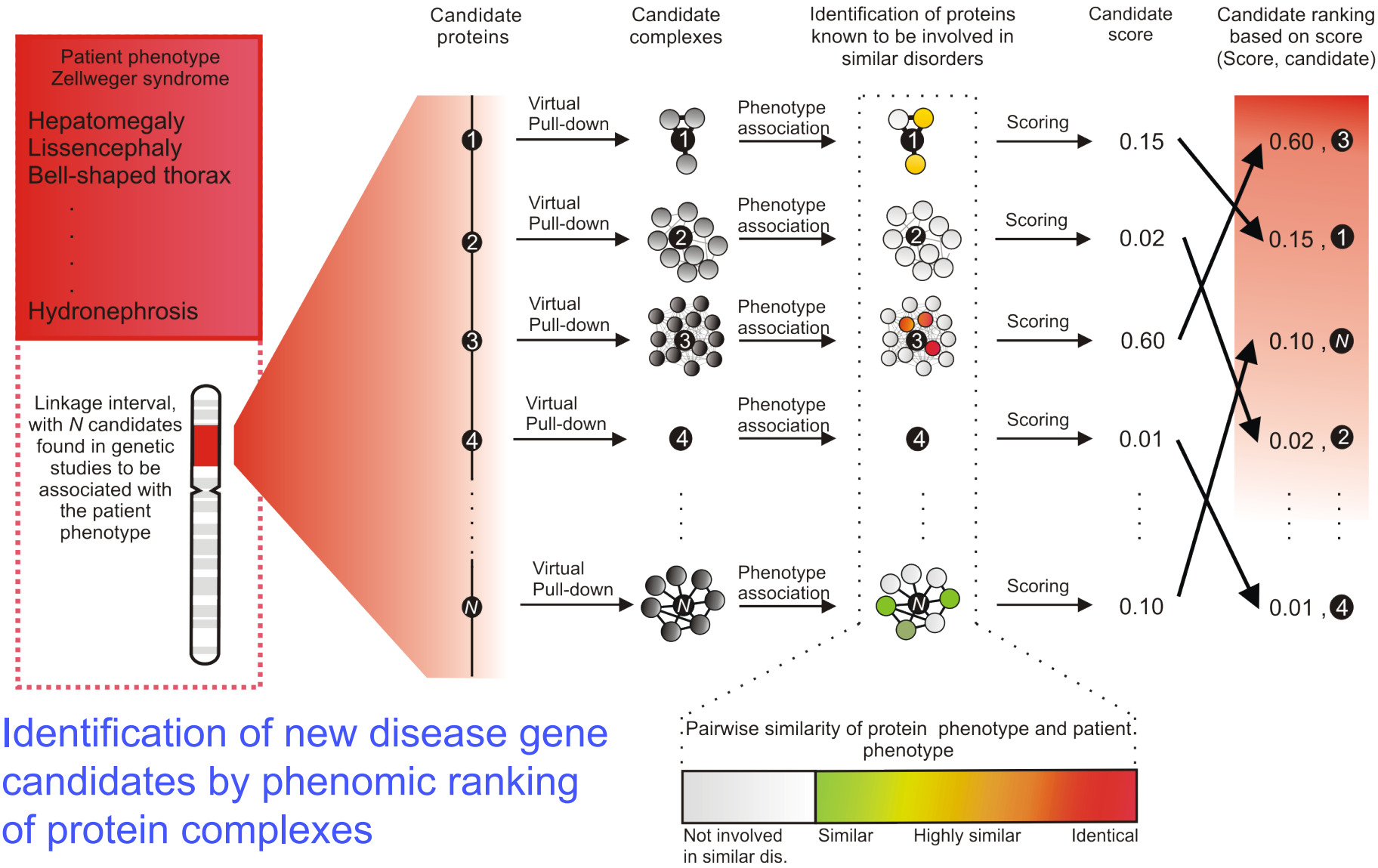
Automated scoring of all interactions

500 000 interactions between 10,300 human proteins

Lage et al. Nat. Biotech 2007

Ranking disease-protein complexes

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

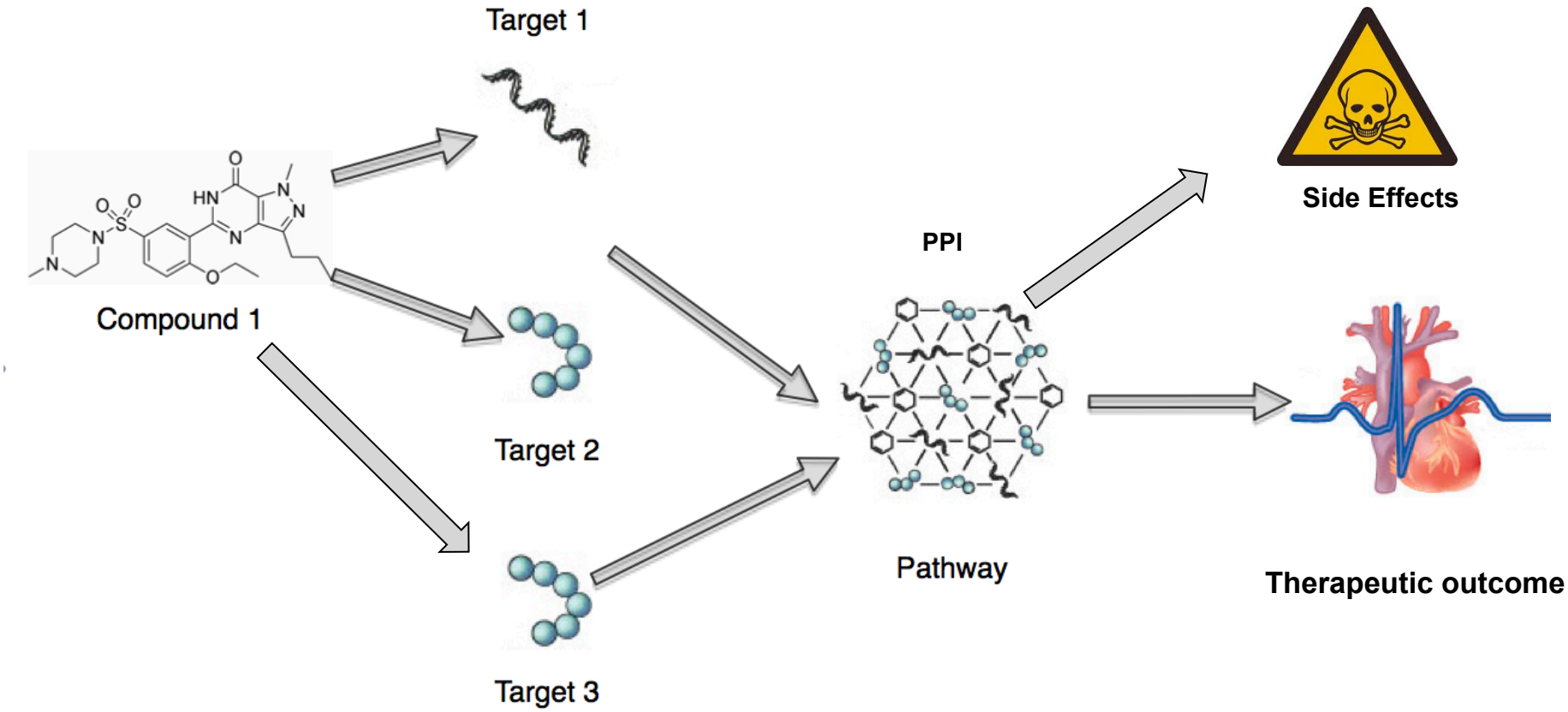


Identification of new disease gene candidates by phenomic ranking of protein complexes

Lage et al. Nat. Biotech 2007

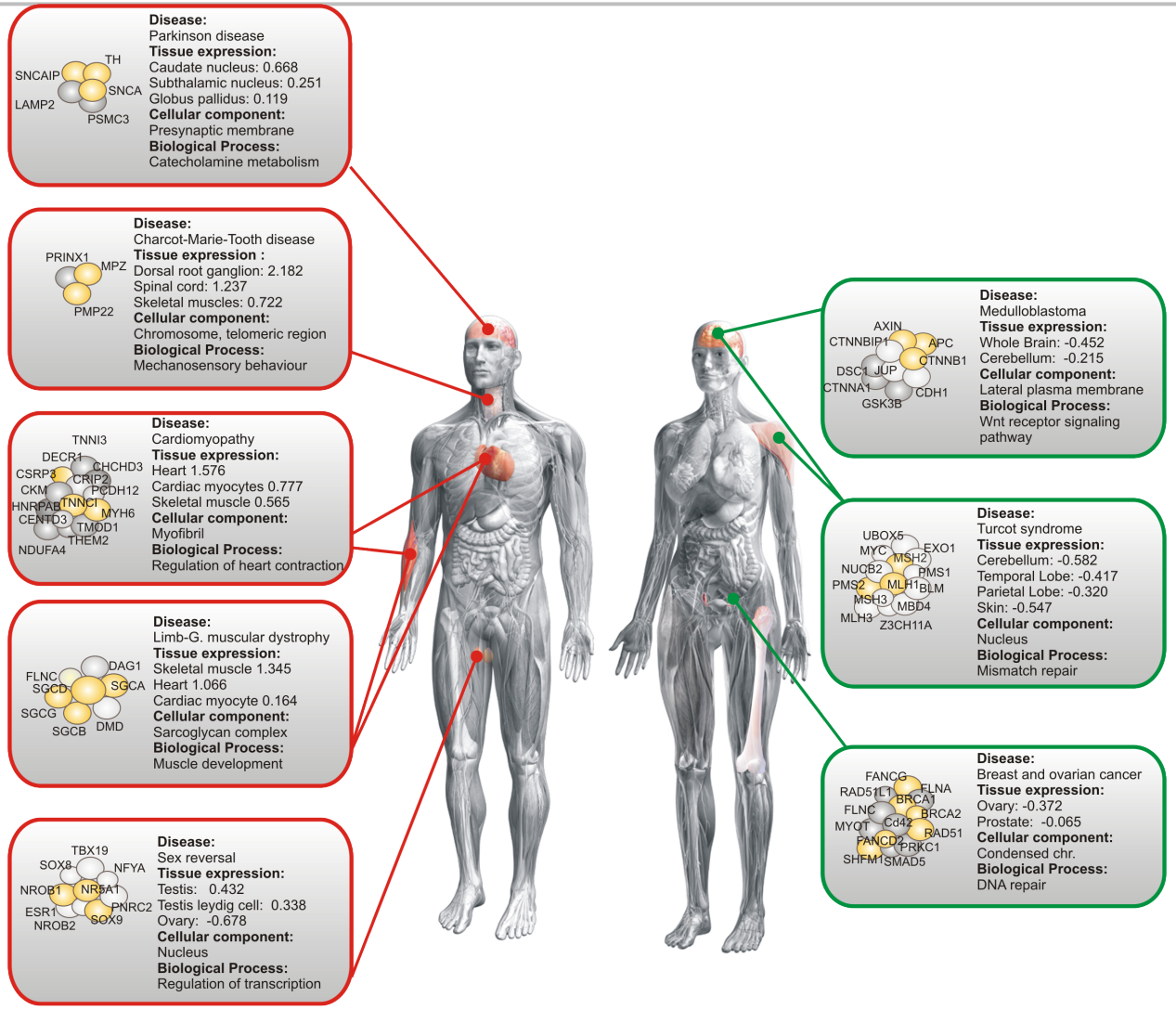
Not looking anymore at 1 protein at the time

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Tissue-specific pathology and gene expression of human disease genes and complexes

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS



Overexpressed non-cancer disease complexes

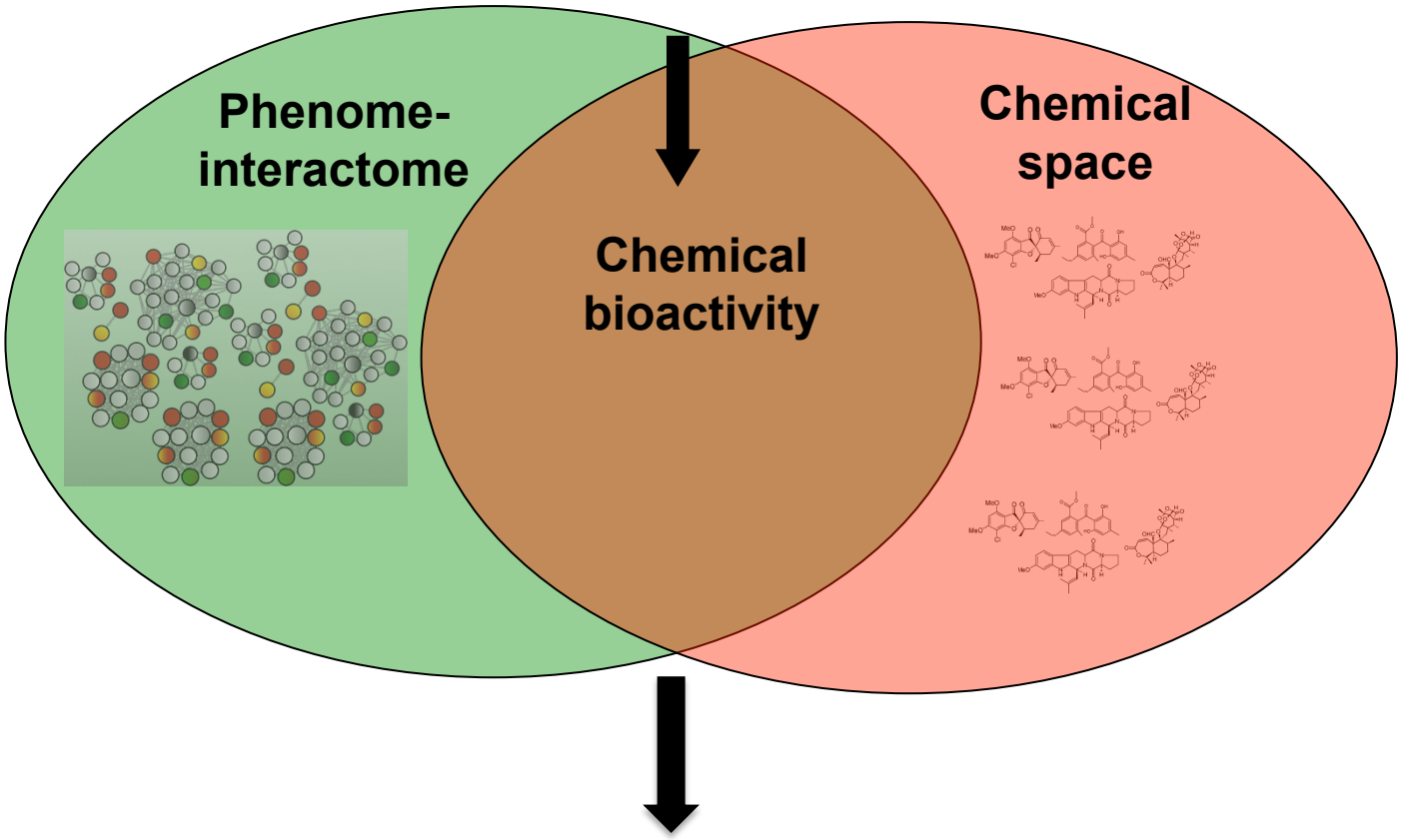
Underexpressed cancer complexe

Lage K, et al.,PNAS, 52, 20870-20875 (2008)

Integration of chemical space into biological space

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Multi-data integration to understand biological problems



Disease chemical biology network

Chemical collection with protein association

chEMBL

ctd™

STITCH

PDSP
K, Database

PubChem

DrugBank

BindingDB

ChemProt

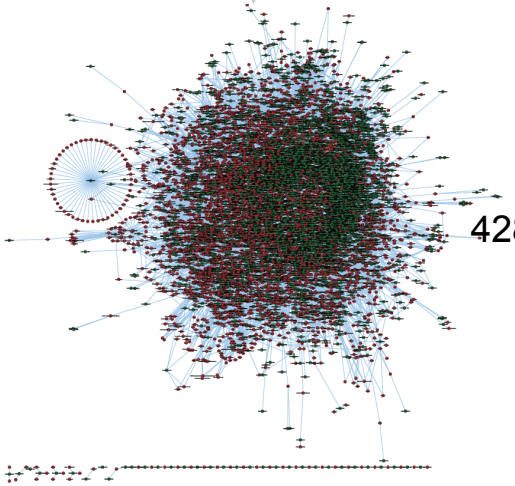
ChemProt: a disease chemical biology database

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

ChemProt

700.000 unique bioactive compounds for 30.000 proteins

Protein-Protein Interactions data integration**

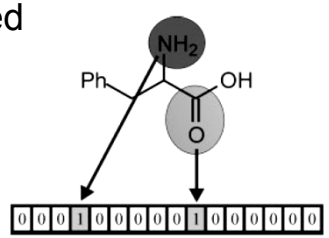


428000 PPIs

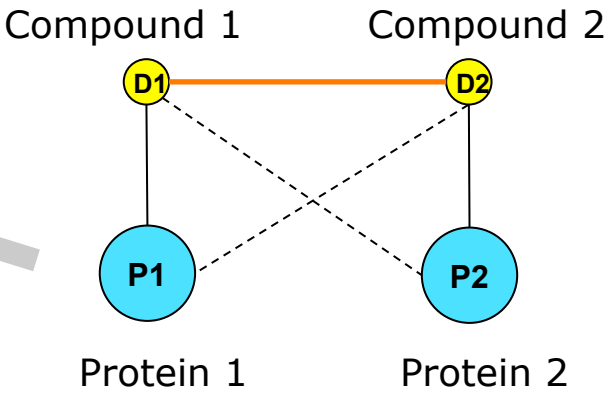
Disease-protein complexes

OMIM, BioAlma, GO, Tissue specificity

MACCs and pharmacophore fingerprints computed



Structural similarity search

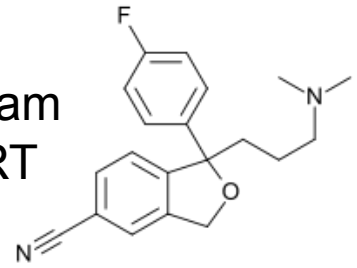


Taboureau O et al. 2010 NAR

An example with citalopram (antidepressant)

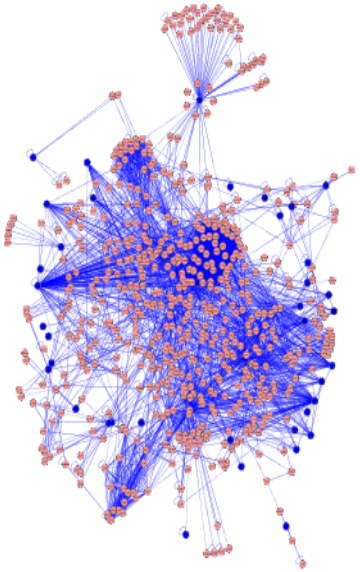
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

13 proteins associated to citalopram
Including the primary target SERT



ChemProt

PPIs



629 genes forming 4141
interactions

diseases

An example with citalopram: Genes enrichment

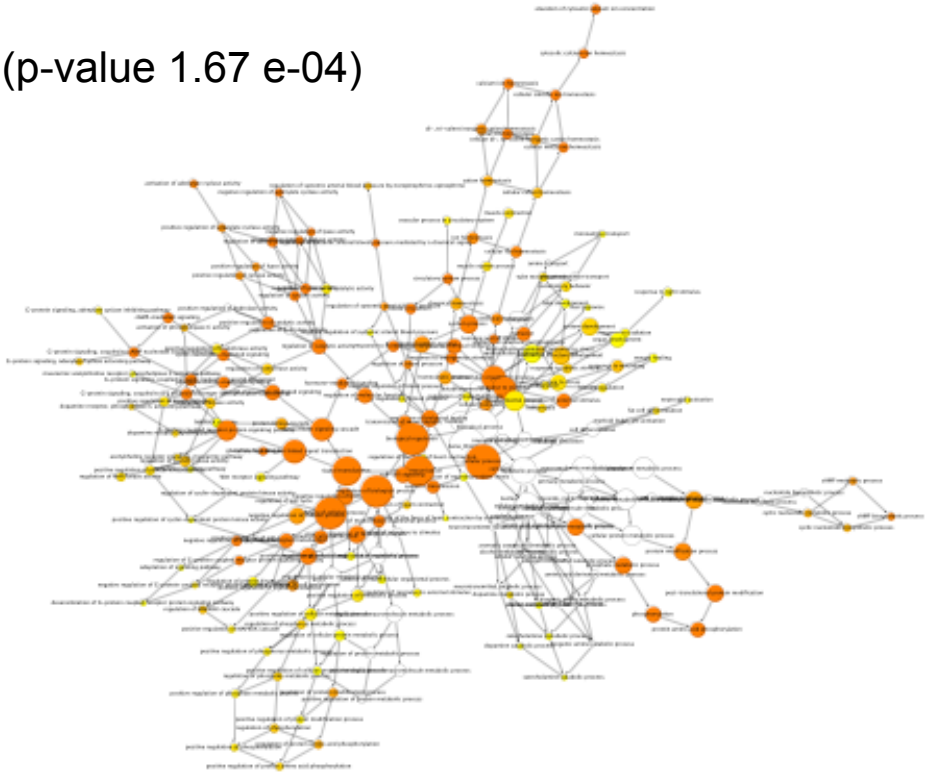
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

GO
- cell communication (p-value 1.32 e-86)
- signal transduction (p-value 4.07 e-81)

OMIM
- Major Depressive Disorder (p-value 3.77 e-06)
  TPH2;FKBP5;HTR2A

- Obsessive-Compulsive Disorder 1 (p-value 1.67 e-04)
  HTR2A;SLC6A4

BioAlma
- Schizophrenia 9.02 e-24
- Bipolar disorder 6.92 e-21
- Anorexia nervosa 6.61 e-10
- Bulimia nervosa 1.41 e-07
- Obesity 2.20 e-05



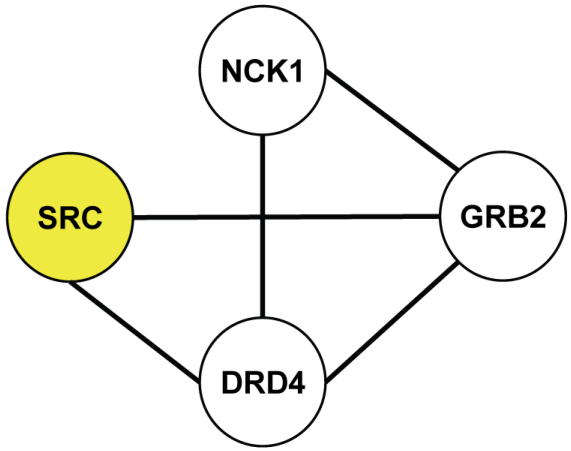
An example with citalopram: Genes enrichment for DRD4

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

BioAlma

GO biological process

HPA



OMIN

GO cellular component

mRNA expression

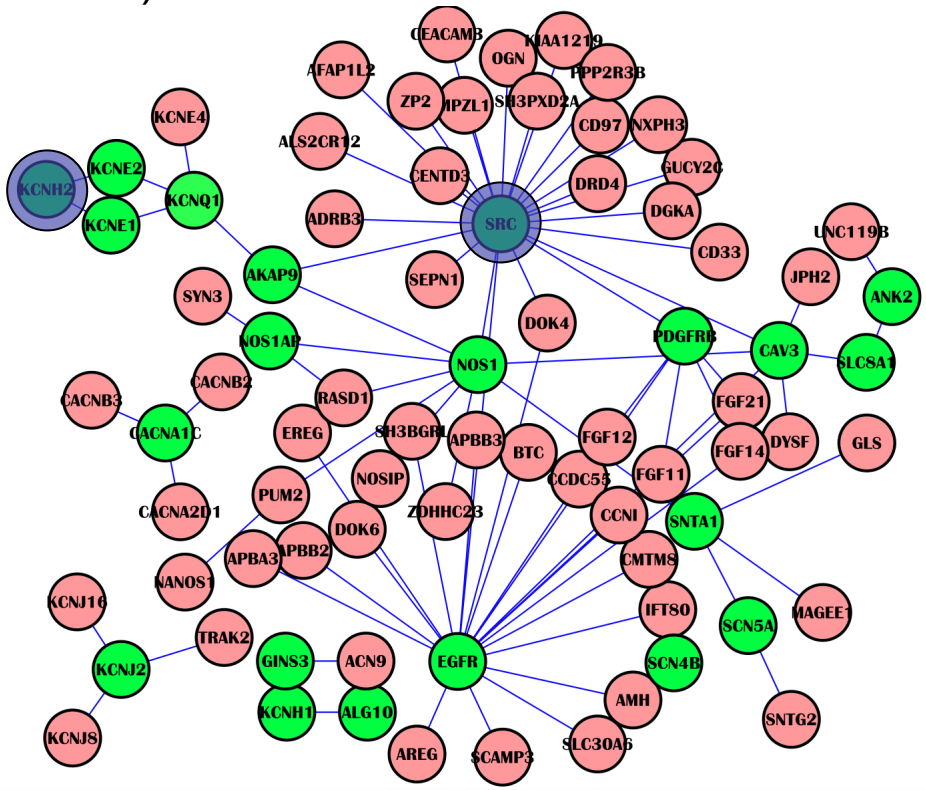
Leukemia

An example with citalopram: Src regulation of hERG?

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Citalopram inhibits hERG
(5.4 µM in ChEMBL)

Not associated to LQTS and arrhythmia



Cell Signal. 2008 Oct;20(10):1815-21. Epub 2008 Jun 19.

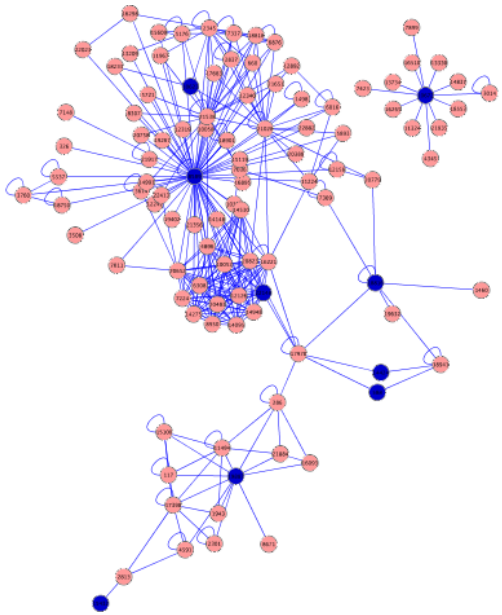
Both EGFR kinase and Src-related tyrosine kinases regulate human ether-à-go-go-related gene potassium channels.

Zhang DY, Wang Y, Lau CP, Tse HF, Li GR.

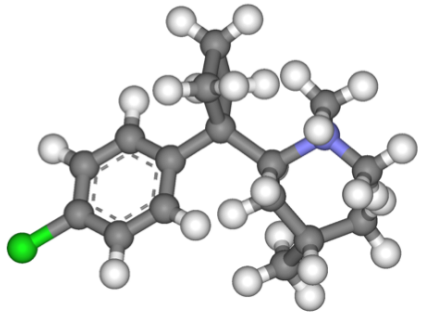
An example with sibutramine (anti-obesity)

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

7 proteins associated to sibutramine



105 genes forming 326 interactions !!!



GO - Monoamine transport 5.07 e-08

OMIM - Neuroticism anxiety 3.1 e-03

BioAlma
- Tourette syndrome 1.28 e-08
- Schizophrenia 1.22 e-07
- Obesity 3.0 e-07

86804 cohort in US who took an anti-obesity medication, 38% took antidepressants and 2.5% had schizophrenia as side effects (Bolen SD, Obesity, 2010)

Anatomical Therapeutic Chemical (ATC) classification

Drugs categorization based on the Anatomical Therapeutic Chemical (ATC) Classification.

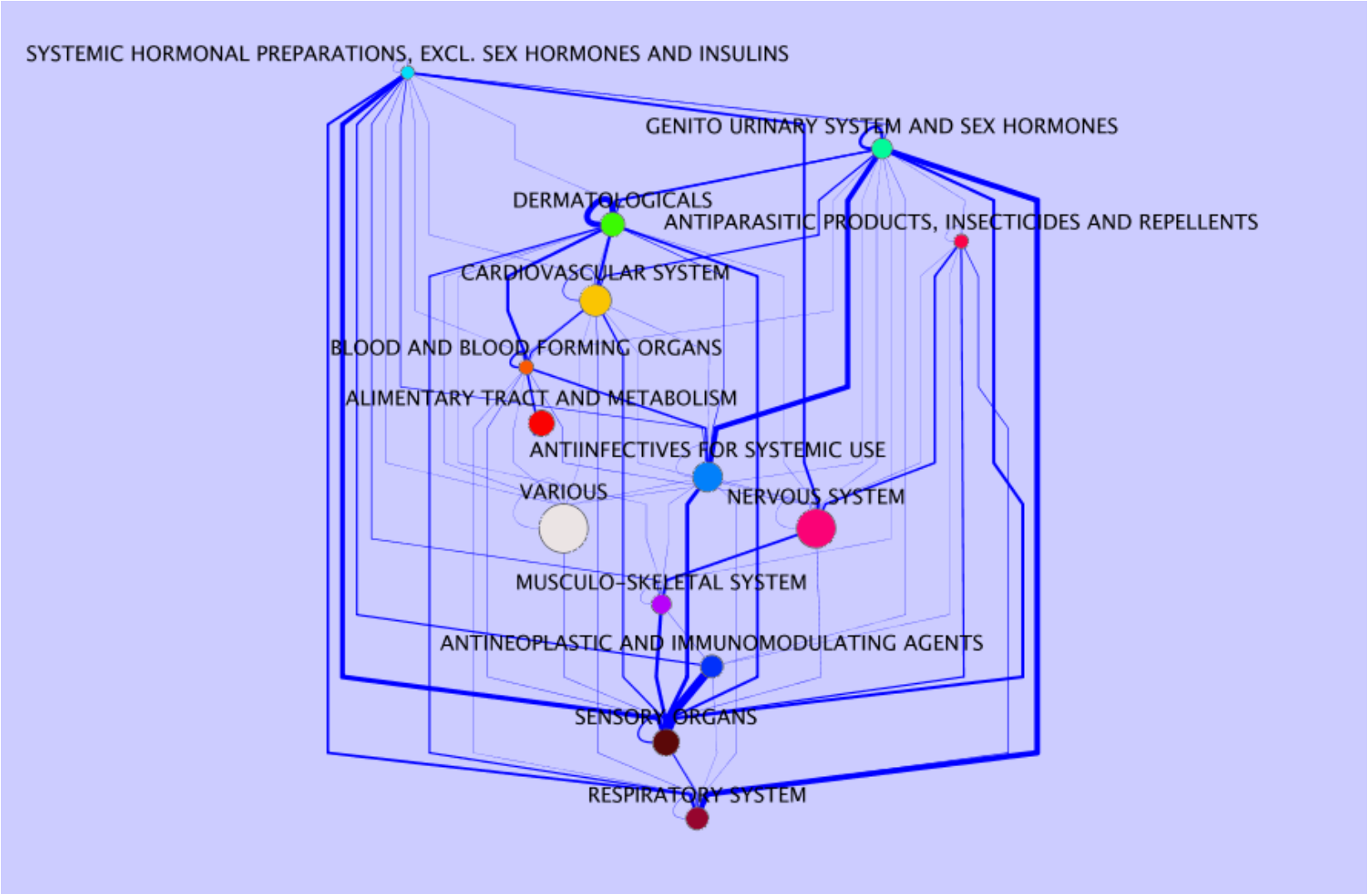
In ATC classification system, the active substances are divided into different group according to the organ or system on which they act and their therapeutic, pharmacologcal and chemical properties

A	Alimentary tract and metabolism (1st level, anatomical main group)
A10	Drugs used in diabetes (2nd level, therapeutic subgroup)
A10B	Blood glucose lowering drugs, excl. insulins (3rd level, pharmacological subgroup)
A10BA	Biguanides (4th level, chemical subgroup)
A10BA02	metformin (5th level, chemical substance)

Drug – Target – Disease Classification through ATC

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

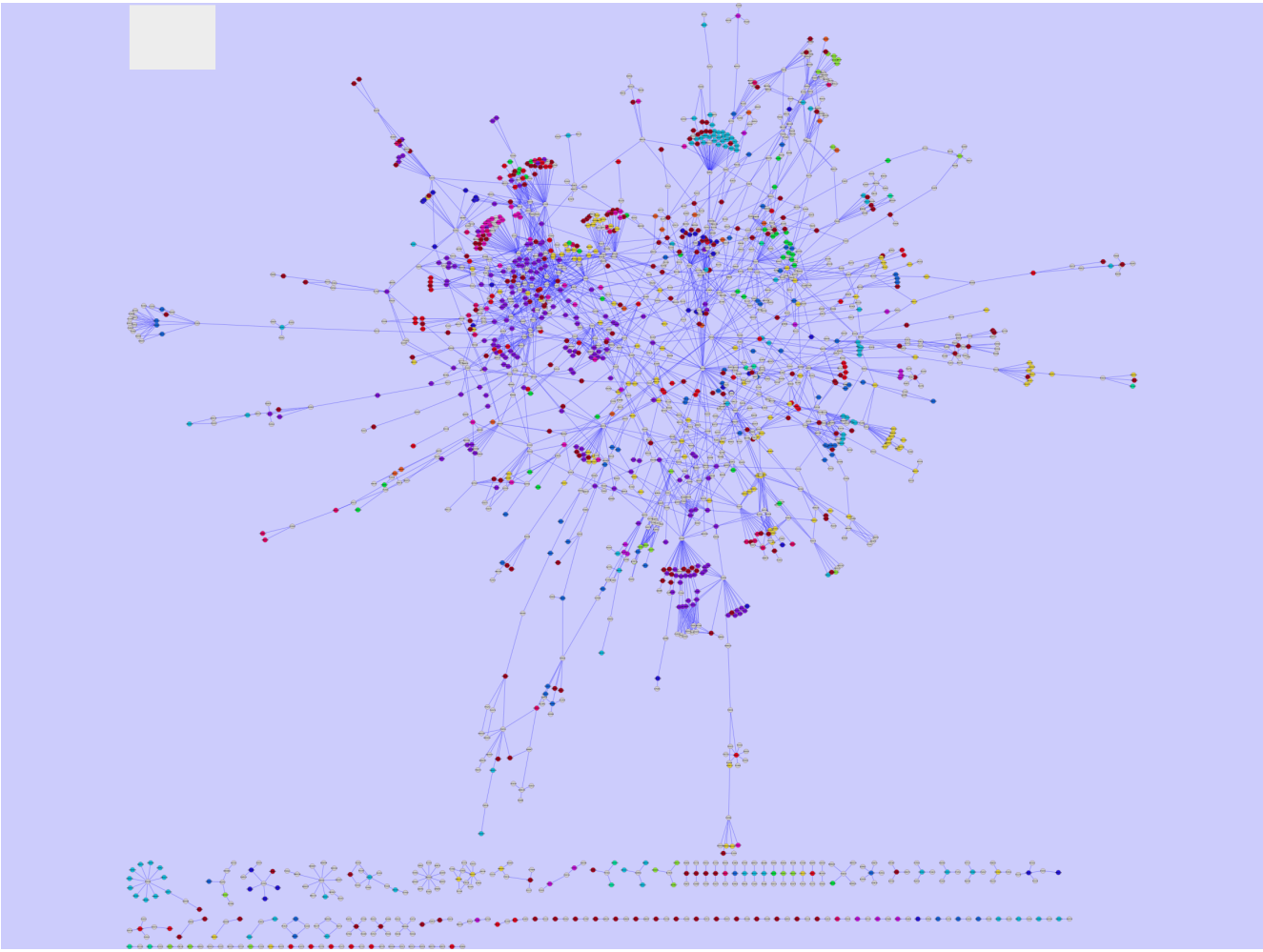
Drugs categorization based on the Anatomical Therapeutic Classification (ATC)



Drug – Target – Disease Classification through ATC

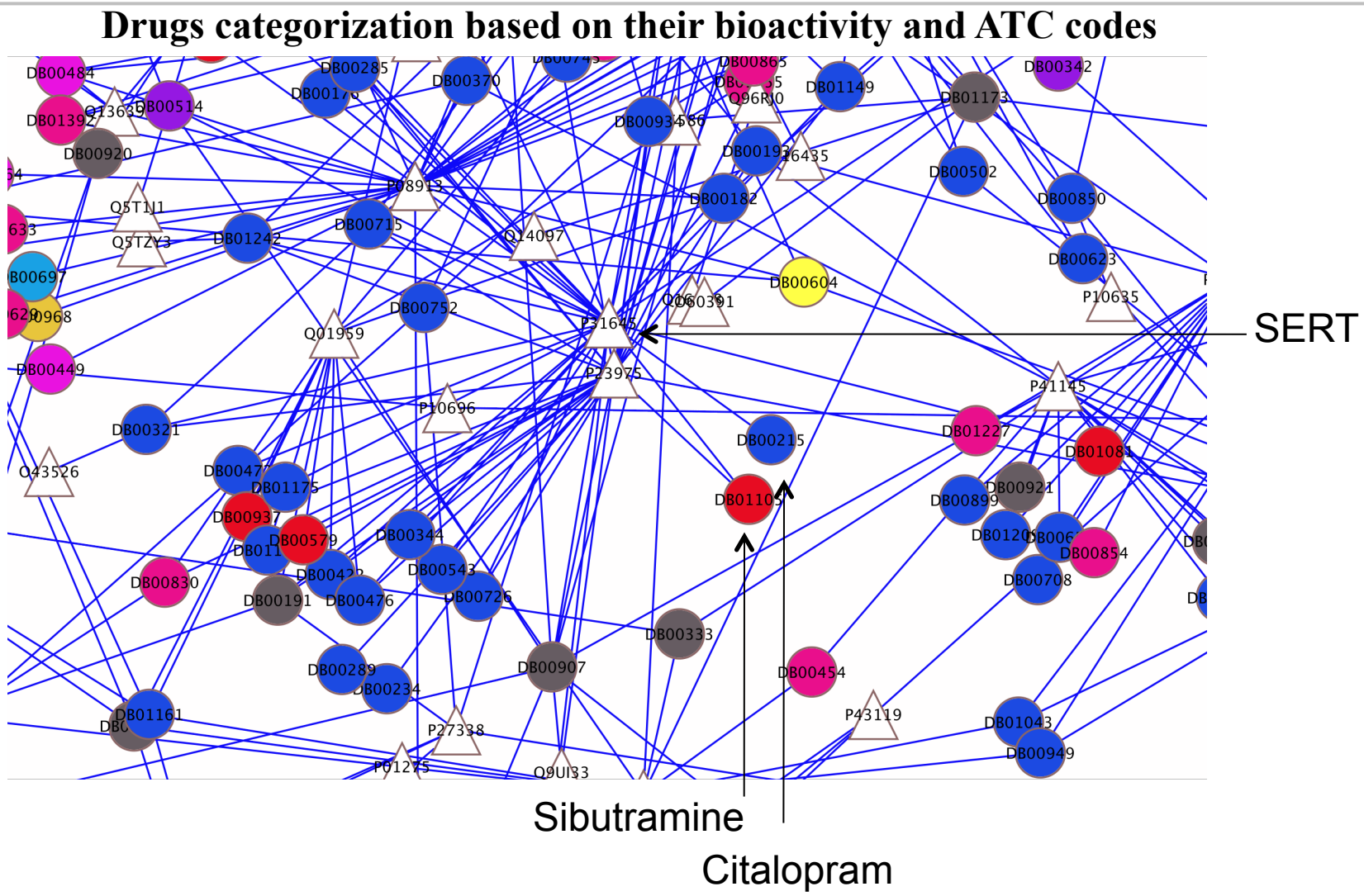
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Drugs categorization based on their bioactivity and ATC codes



Drug – Target – Disease Classification through ATC

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Drug – Target – Side Effects

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

- Over 2 million serious Side Effects (SEs) occur every year (in the world)
- Side effects potentially accounts for the 4th leading cause of death
- The underlying mechanisms of side effects are not understood (only partly)
- Greater effort on designing safe drugs

Side effect resource (SIDER database)

CENTERFO BIOLOGI ALSEQU NCEANA LYSIS CBS

Molecular Systems Biology 6; Article number 343; doi:10.1038/msb.2009.98
Citation: *Molecular Systems Biology* 6:343

www.molecularsystemsbiology.com

molecular systems biology

REPORT

A side effect resource to capture phenotypic effects of drugs

Michael Kuhn[1,4], Monica Campillos[1], Ivica Letunic[1], Lars Juhl Jensen[1,2] and Peer Bork[1,3,*]

[1] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, [2] Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark and [3] Max-Delbrück-Centre for Molecular Medicine, Berlin, Germany
[4] Present address: Biotechnology Center, TU Dresden, 01062 Dresden, Germany
* Corresponding author. Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg 69117, Germany. Tel.: +49 6221 387 8526; Fax: +49 6221 387 517; E-mail: bork@embl.de

Received 9.7.09; accepted 30.11.09

The molecular understanding of phenotypes caused by drugs in humans is essential for elucidating mechanisms of action and for developing personalized medicines. Side effects of drugs (also known as adverse drug reactions) are an important source of human phenotypic information, but so far research on this topic has been hampered by insufficient accessibility of data. Consequently, we have developed a public, computer-readable side effect resource (SIDER) that connects 888 drugs to 1450 side effect terms. It contains information on frequency in patients for one-third of the drug–side effect pairs. For 199 drugs, the side effect frequency of placebo administration could also be extracted. We illustrate the potential of SIDER with a number of analyses. The resource is freely

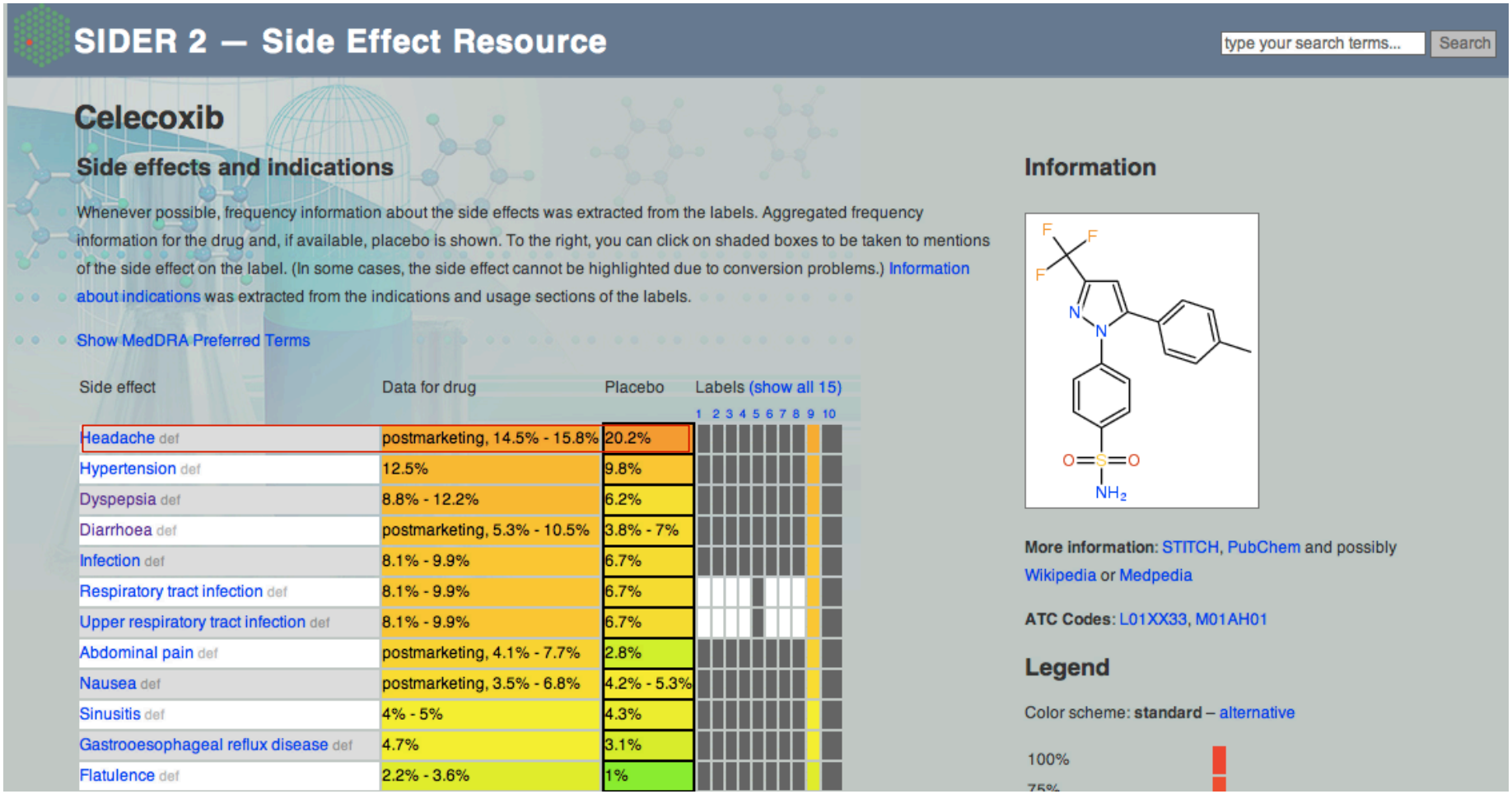
0; d :10.1038/msb.2009.98

type side effects

Update from 888 to 996 drugs - only ~ 200 drugs with placebo

Analyze carefully the information

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



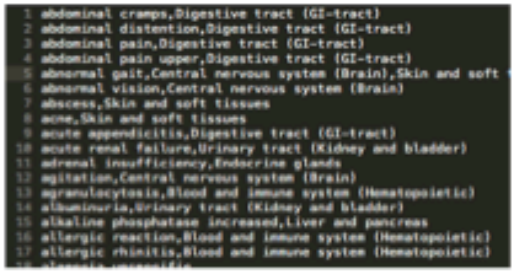
21% non significant pairs

Side Effect – Tissue integration

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

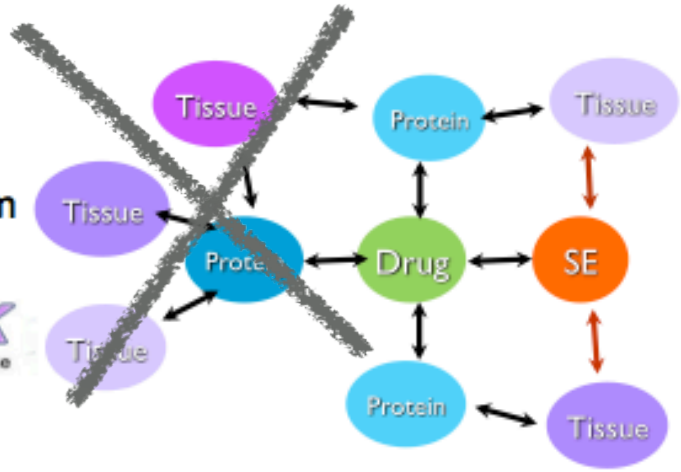
SE-drugs pairs

Mapped Tissue to SE
around 1000 terms



Add drug-targets from
DrugBank

DRUGBANK
Open Data Drug & Drug Target Database



12 Tissue categories

Central nervous system (Brain)
Blood and immune system
(Hematopoietic)
Skin and soft tissues
Endocrine glands
Liver and pancreas
GI tract
Lung
Cardiovascular system
Female Tissue
Placenta
Male Tissue
Urinary tract

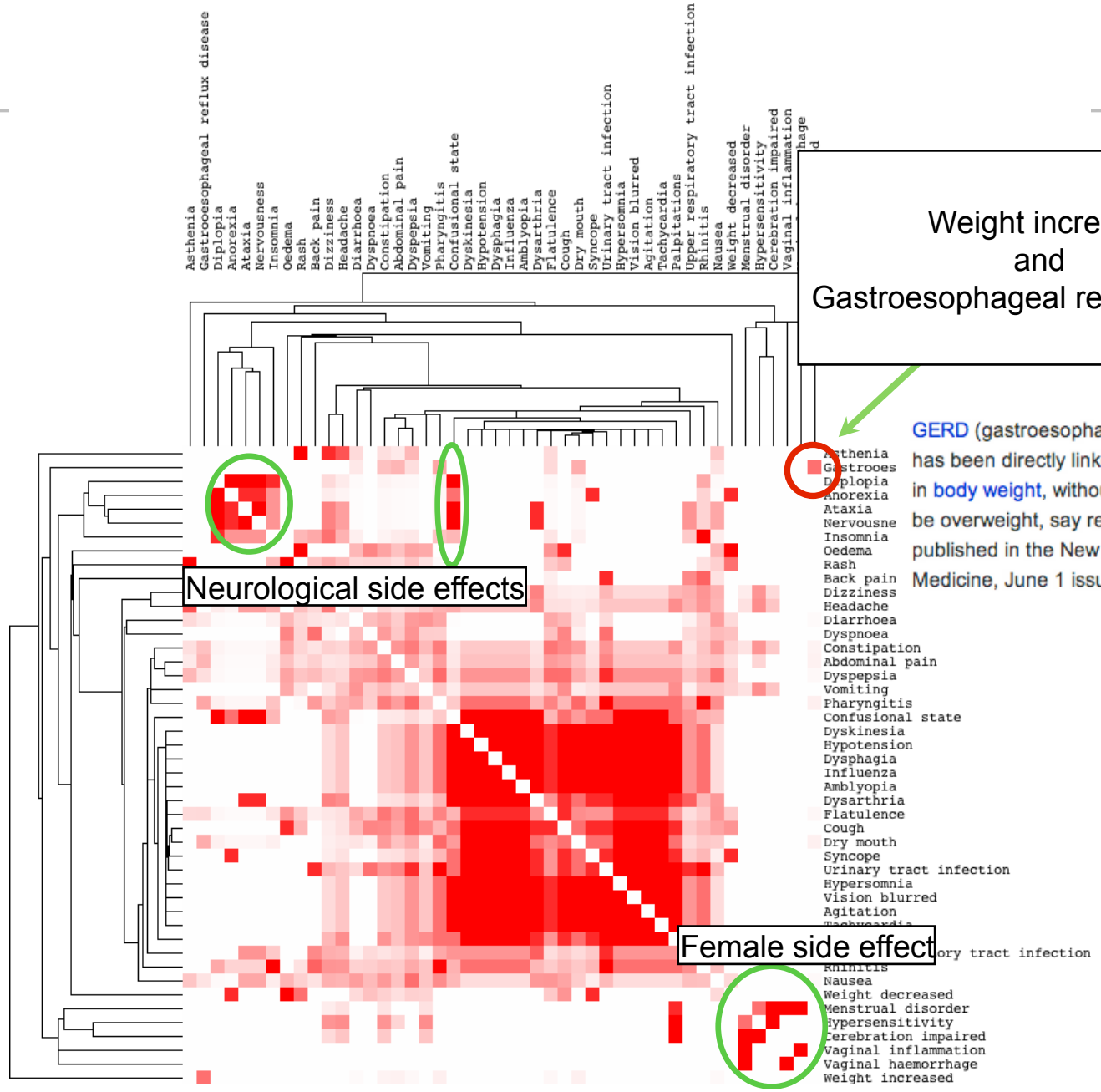
Add
secondary-targets from
ChEMBL
IC50 and Ki < 100 uM

ChEMBL

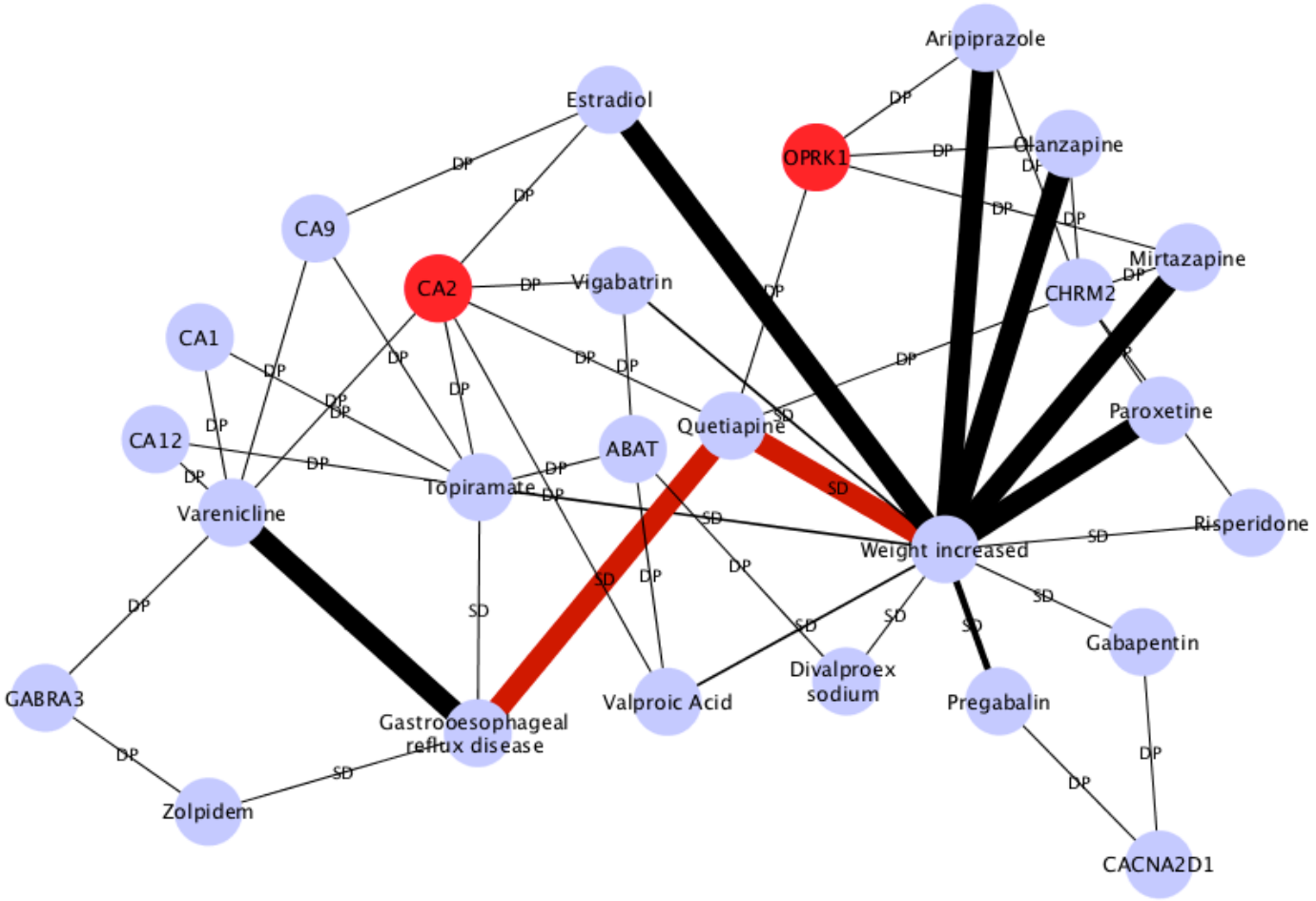
THE HUMAN PROTEIN ATLAS

Add
Tissue information to the
proteins from Protein Altas


Asthenia
Gastrooesophageal reflux disease
Diplopia
Anorexia
Ataxia
Nervousness
Insomnia
Oedema
Rash
Back pain
Dizziness
Headache
Diarrhoea
Dyspnoea
Constipation
Abdominal pain
Dyspepsia
Vomiting
Pharyngitis
Confusional state
Dyskinesia
Hypotension
Dysphagia
Influenza
Amblyopia
Dysarthria
Flatulence
Cough
Dry mouth
Syncope
Urinary tract infection
Hypersomnia
Vision blurred
Agitation
Tachycardia
Palpitations
Upper respiratory tract infection
Rhinitis
Nausea
Weight decreased
Menstrual disorder
Hypersensitivity
Cerebration impaired
Vaginal inflammation
Vaginal haemorrhage
Weight increased



GERD (gastroesophageal reflux disease) has been directly linked to even small gains in body weight, without the person having to be overweight, say researchers in a report published in the New England Journal of Medicine, June 1 issue.



Biol Res Nurs. 2007 Apr;8(4):294-9.

The effect of kappa opioid receptor antagonism on energy expenditure in the obese Zucker rat.

Jarosz PA.

Wayne State University, College of Nursing, Detroit, Michigan 48202, USA. ad9433@wayne.edu

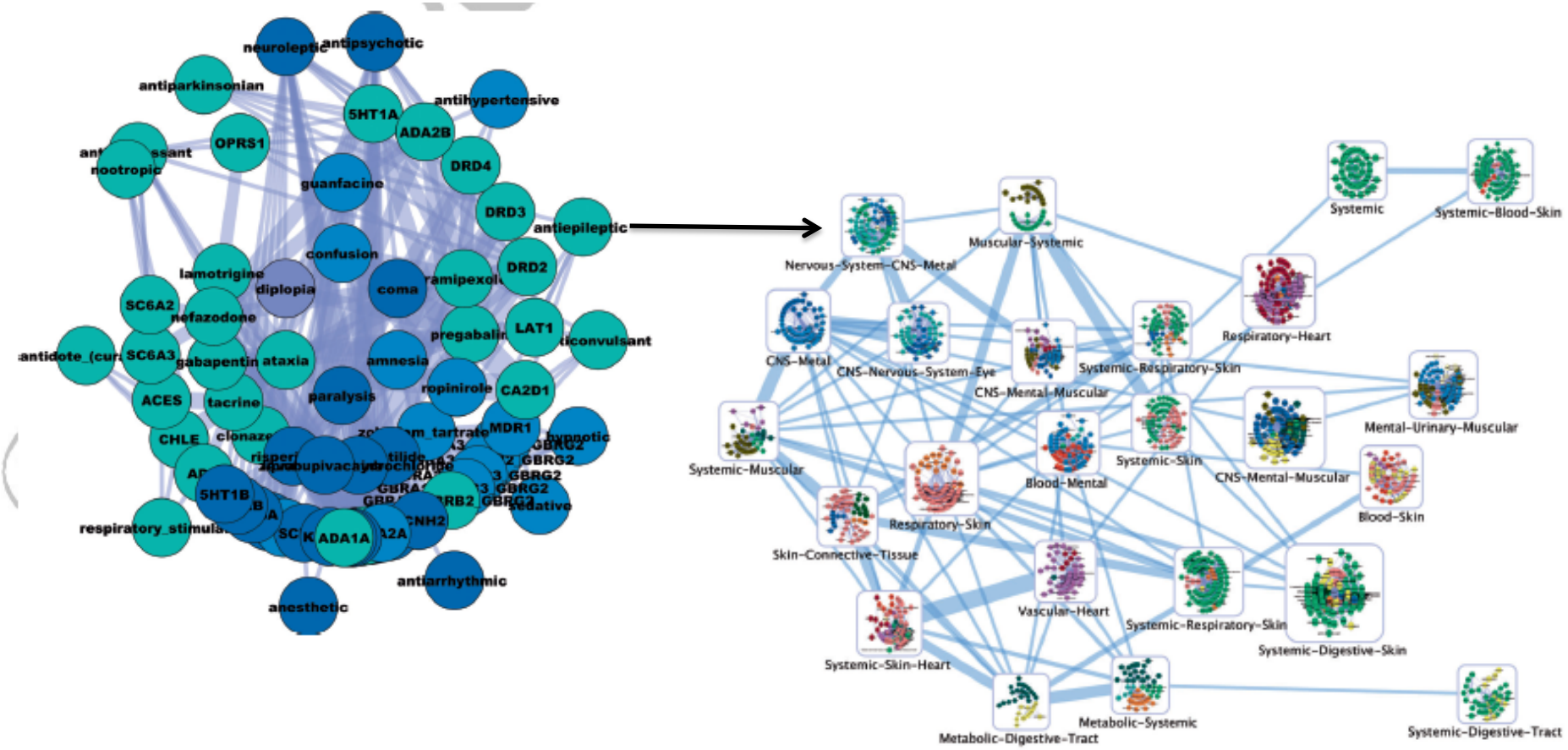
Abstract

Food intake and, subsequently, body weight are influenced by endogenous opioids acting in the central nervous system. Agonists for the opioid

Clinical outcomes

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Association drugs, targets and clinical outcomes



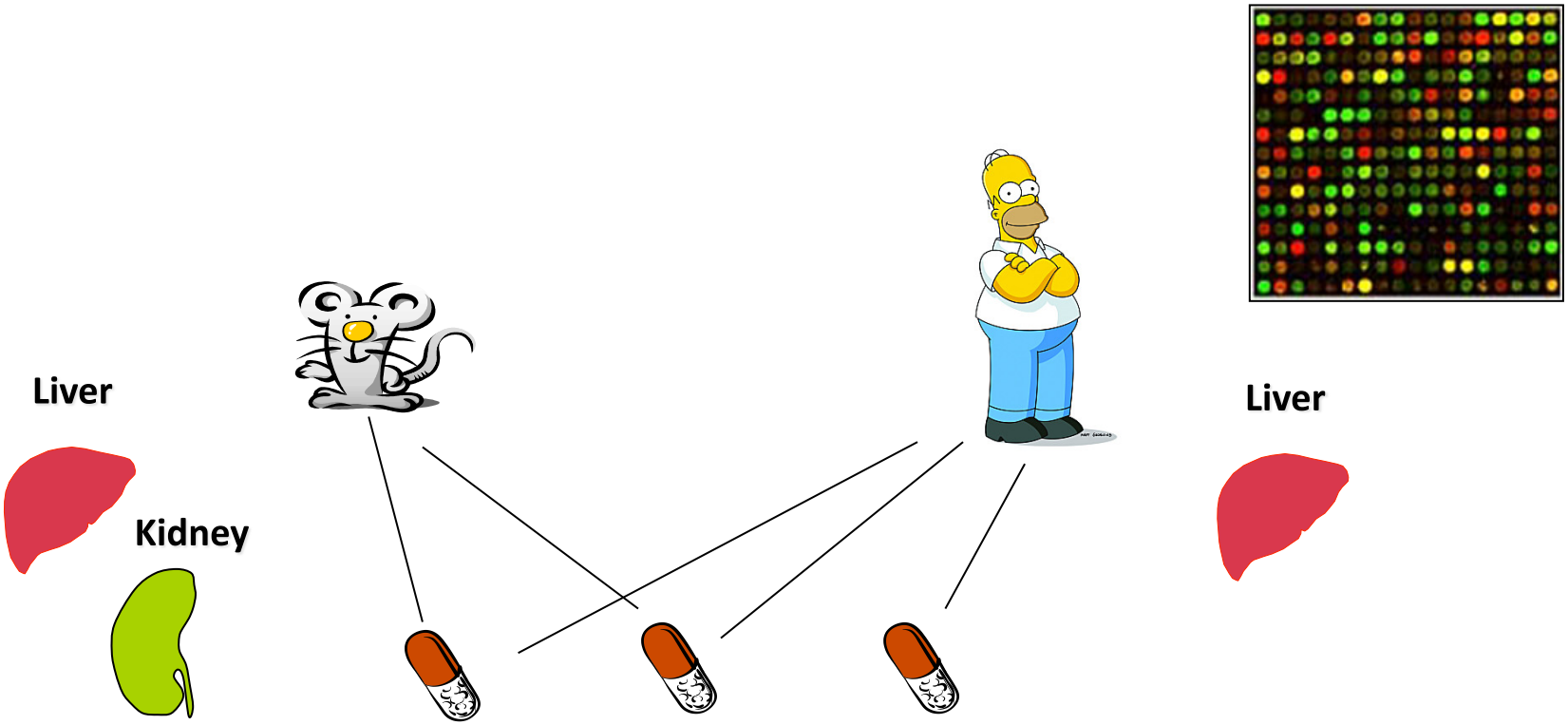
Oprea TI. et al. 2011, Mol. Inf.

Toxicogenomics

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Can we predict inter-species toxicological profiles.

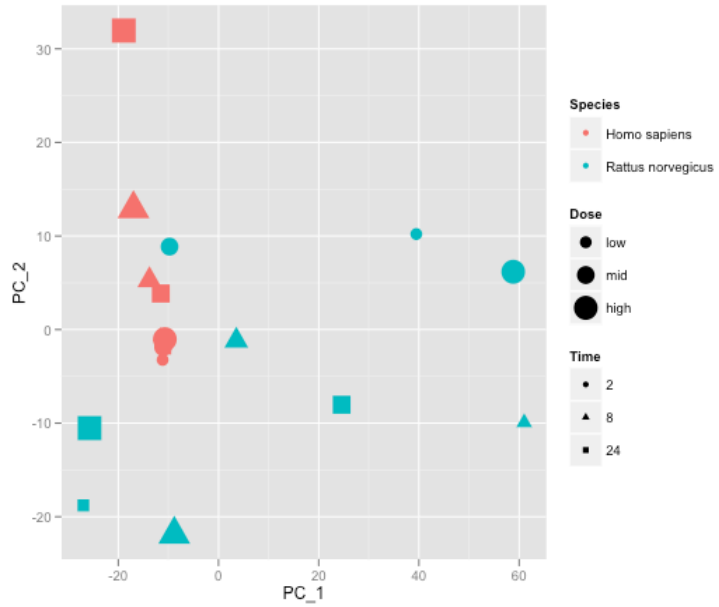
We are addressing this by comparing gene expression patterns in rat and *human, in vivo and in vitro,* after treatment with a set of 130 compounds from the TG-GATEs database

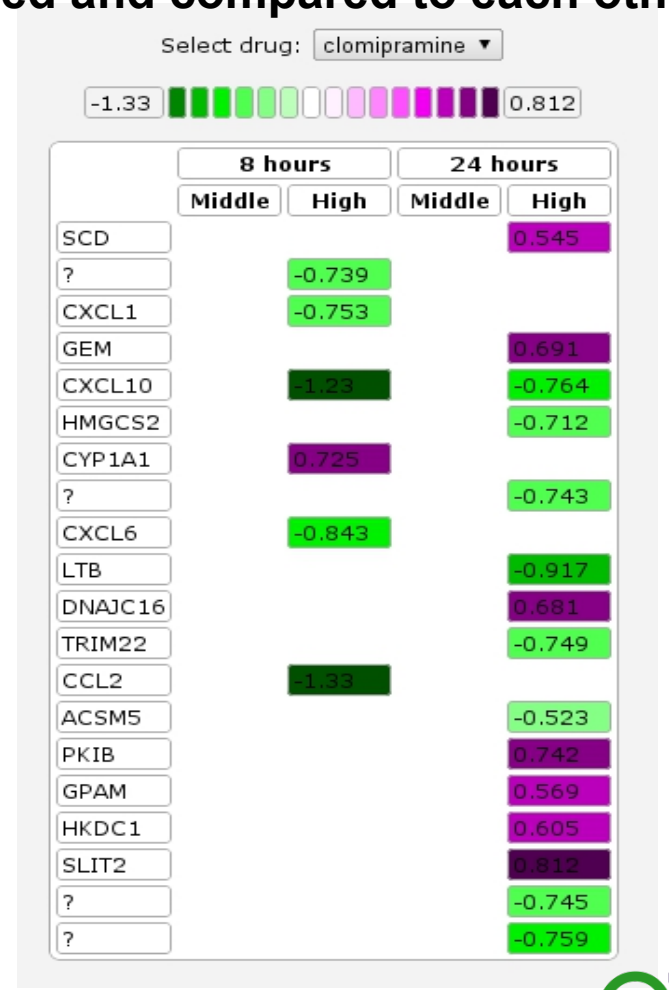


Toxicogenomics

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

The most perturbed genes, according to the conditions (species, organs, in vivo, in vitro, times and doses), can be visualized and compared to each other.



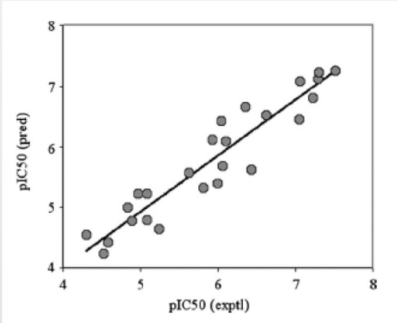


eTOX

Chemoinformatics and Bioinformatics link to Toxicogenomics

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

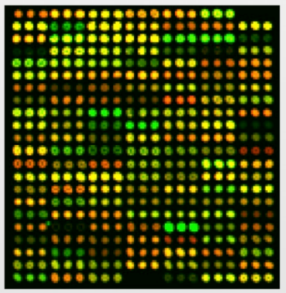
QSAR models based on chemical structure



Collection of compounds with known information for toxicity endpoints

-Phospholipidosis
-Hepatoxicity
-Nephrotoxicity
-Carcinogenicity
-...

Toxicogenomics models based on gene expression profiles



Hybrid models combining both data

Confusion matrix

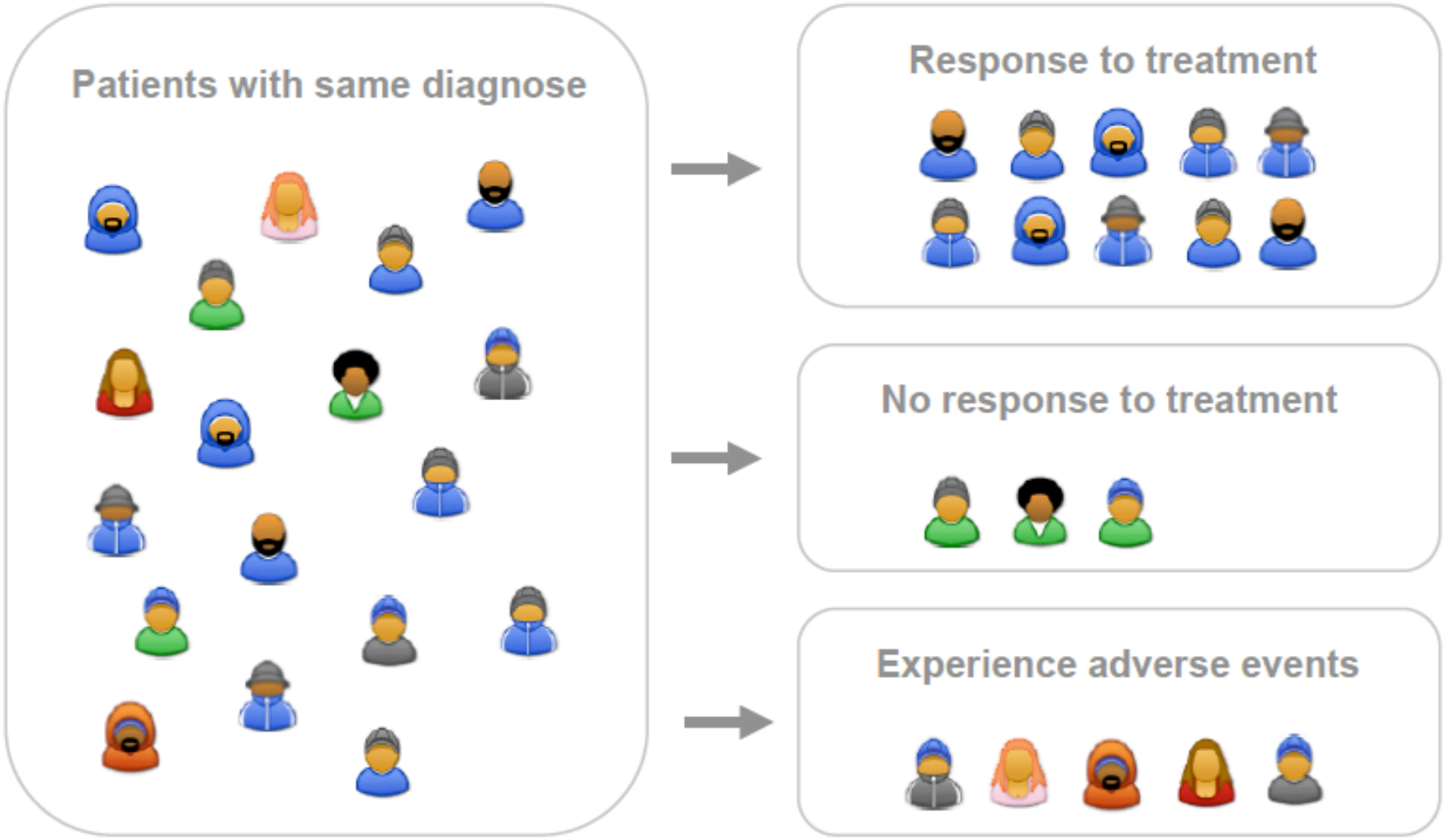
True Positive	**False Negative**
False Positive	**True Negative**

eTOX

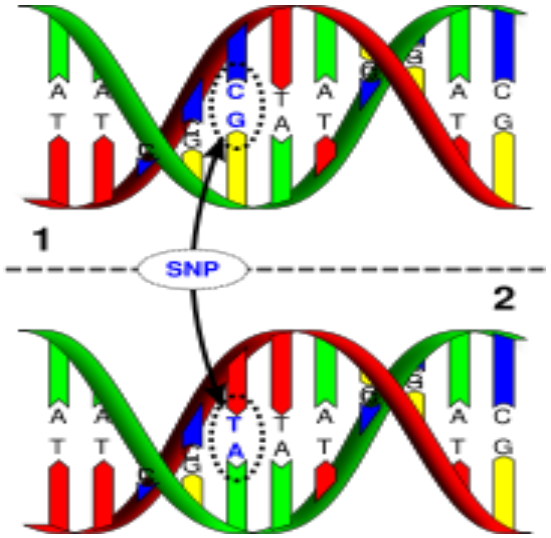
Low Y. et al. 2011, Chem. Res. Toxicol.

Pharmacogenetics

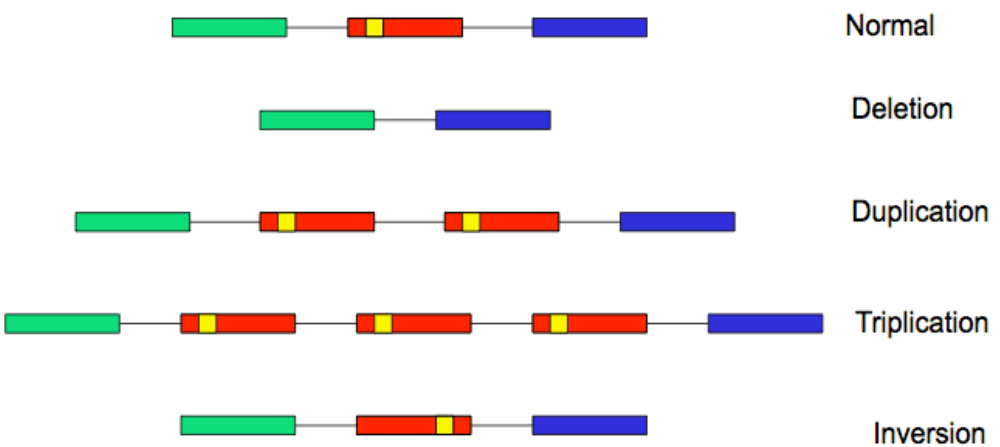
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Genetic variation

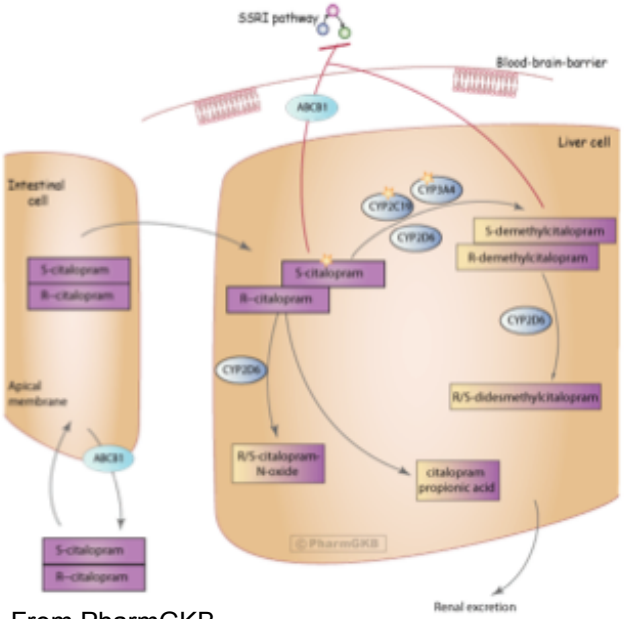


Single nucleotide polymorphism



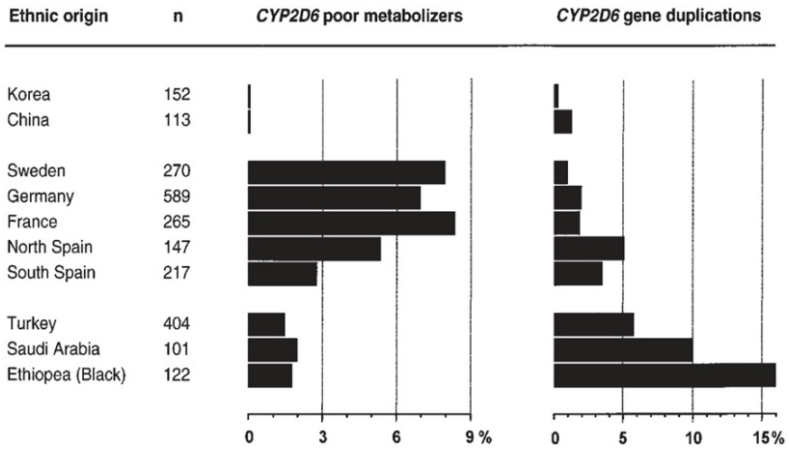
Copy number variation

Genetic variation: CNVs of CYP2D6



From PharmGKB

CYP2D6 genetic variation



Perry et al Nat Genet. 2007 39: 1256–1260.

Frequencies of common deletion and duplications alleles of ADME genes in three major populations determined by PCR-based techniques.

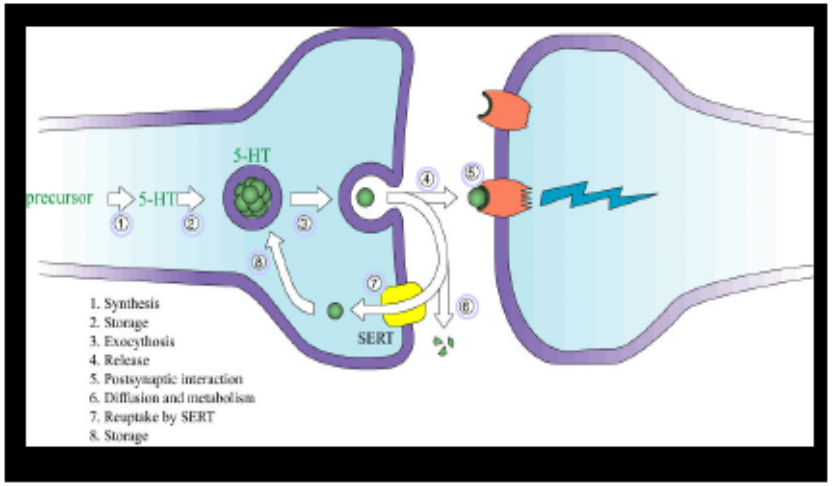
Gene	Type of variation	Decent (healthy subjects)			Enzyme substrates
		African	Asian	European	
CYP2D6	Duplication	0.016-0.136	0.000-0.010	0.011-0.070	Metabolism of variety of xenobiotics and environmental agents including antiarrhythmics, antipsychotics, adrenoceptor antagonists and tricyclic antidepressants.
	Deletion	0.006-0.061	0.045-0.062	0.016-0.073	

→ We are looking on a Danish and Icelandic cohort based on SNP array and qPCR for CYP2D6

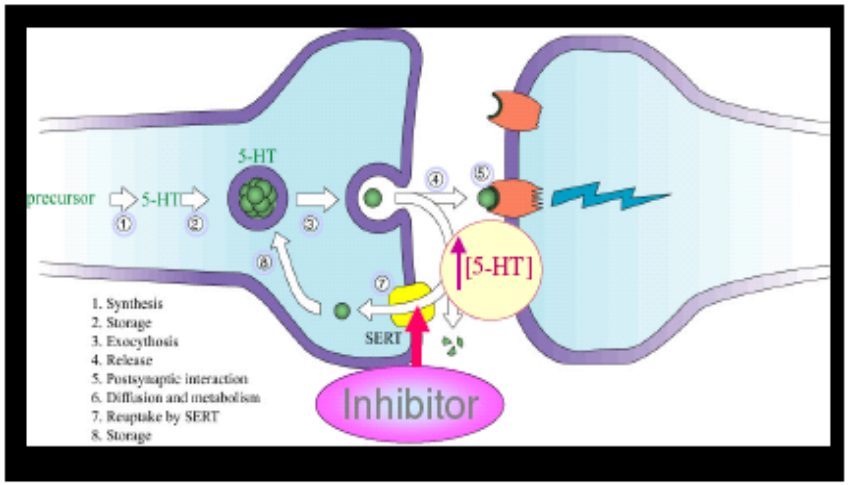
Genetic variation: Mutations study and SNPs on SERT

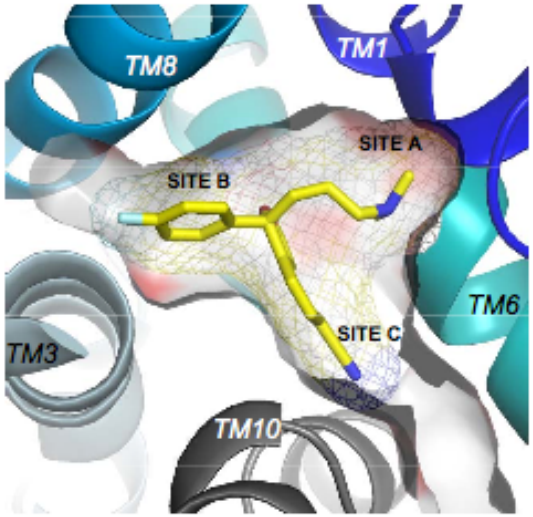
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

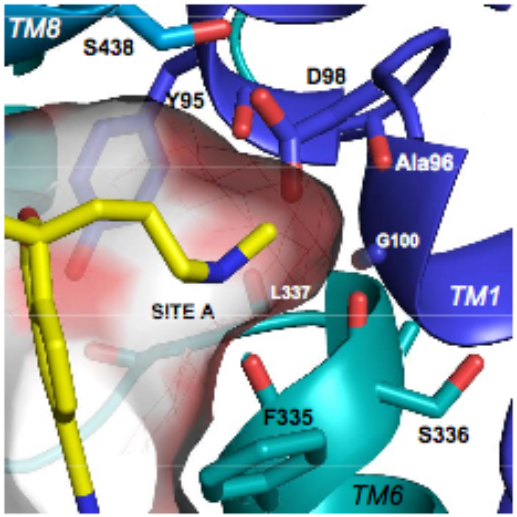
Modulation of neurotransmission

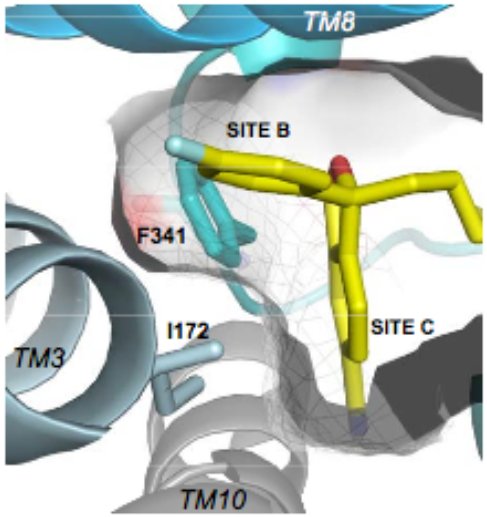


Modulation of neurotransmission





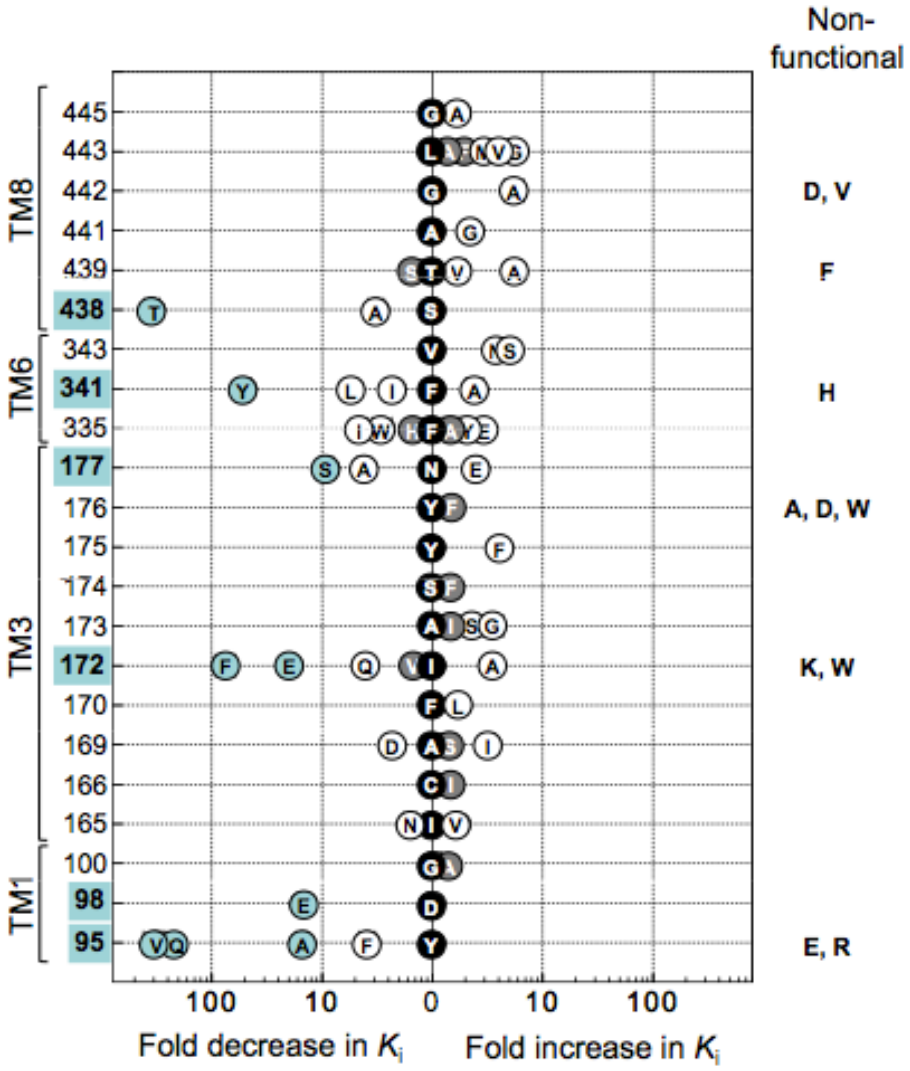


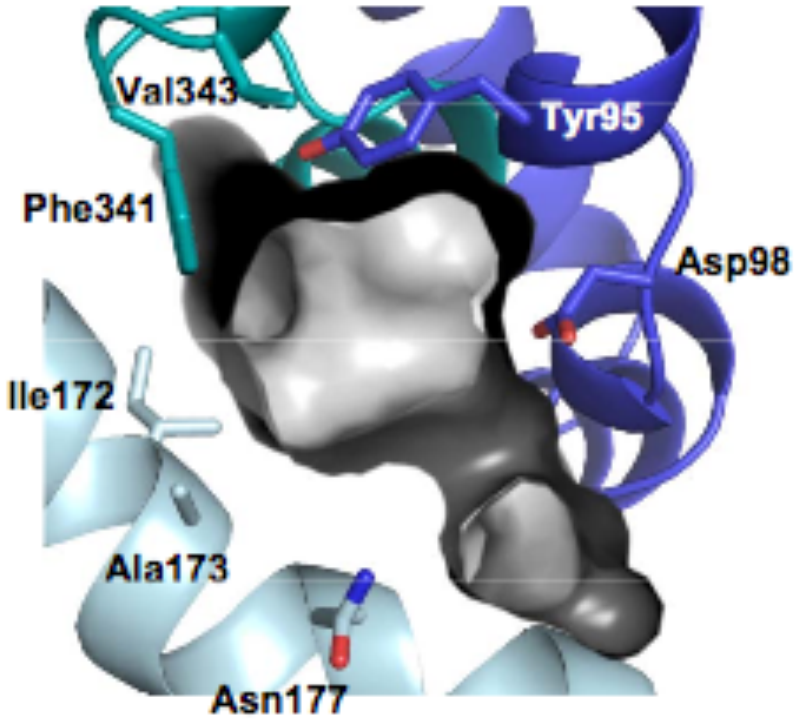


Andersen J et al., J. Biol. Chem. 284 (2009) 10276

Genetic variation: Serotonin transporter

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS





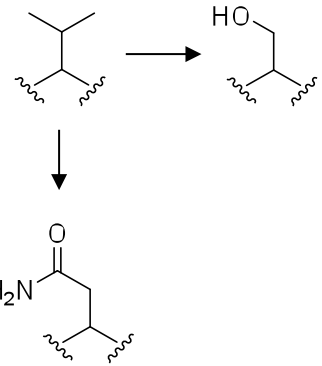
Andersen J et al., J. Biol. Chem. 285 (2010) 2051

What is the binding site for escitalopram?

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

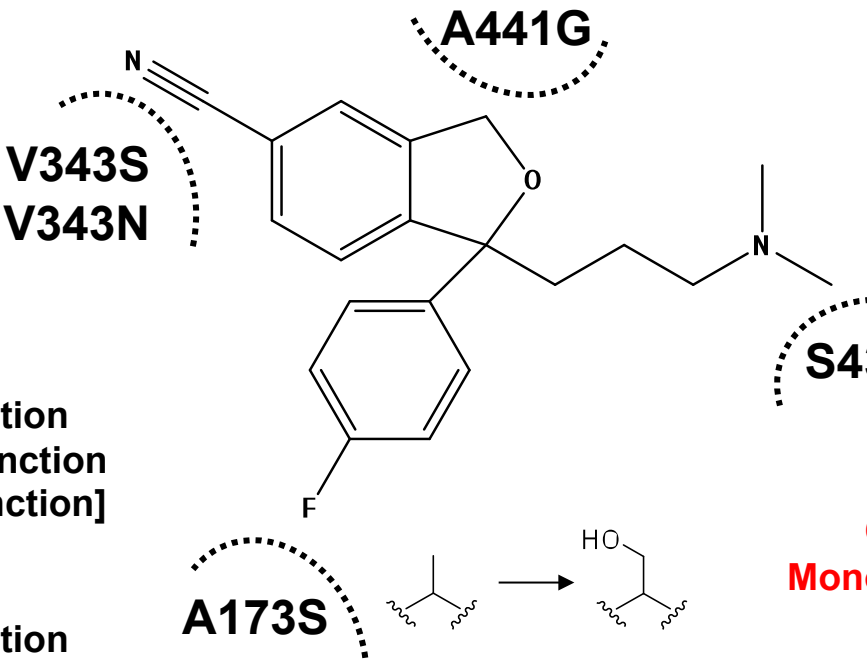
Mutagenesis study

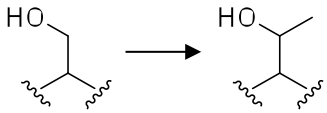
A441G
CIT: 2.5x *gain* of function



V343S
CIT: 4.5x *gain* of function
Des-CN: 4.9x *gain* of function
[Des-F: 7.4x gain of function]

V343N
CIT: 3.5x *gain* of function
Des-CN: 35.5x *gain* of function
[Des-F: 3.8x *gain* of function]



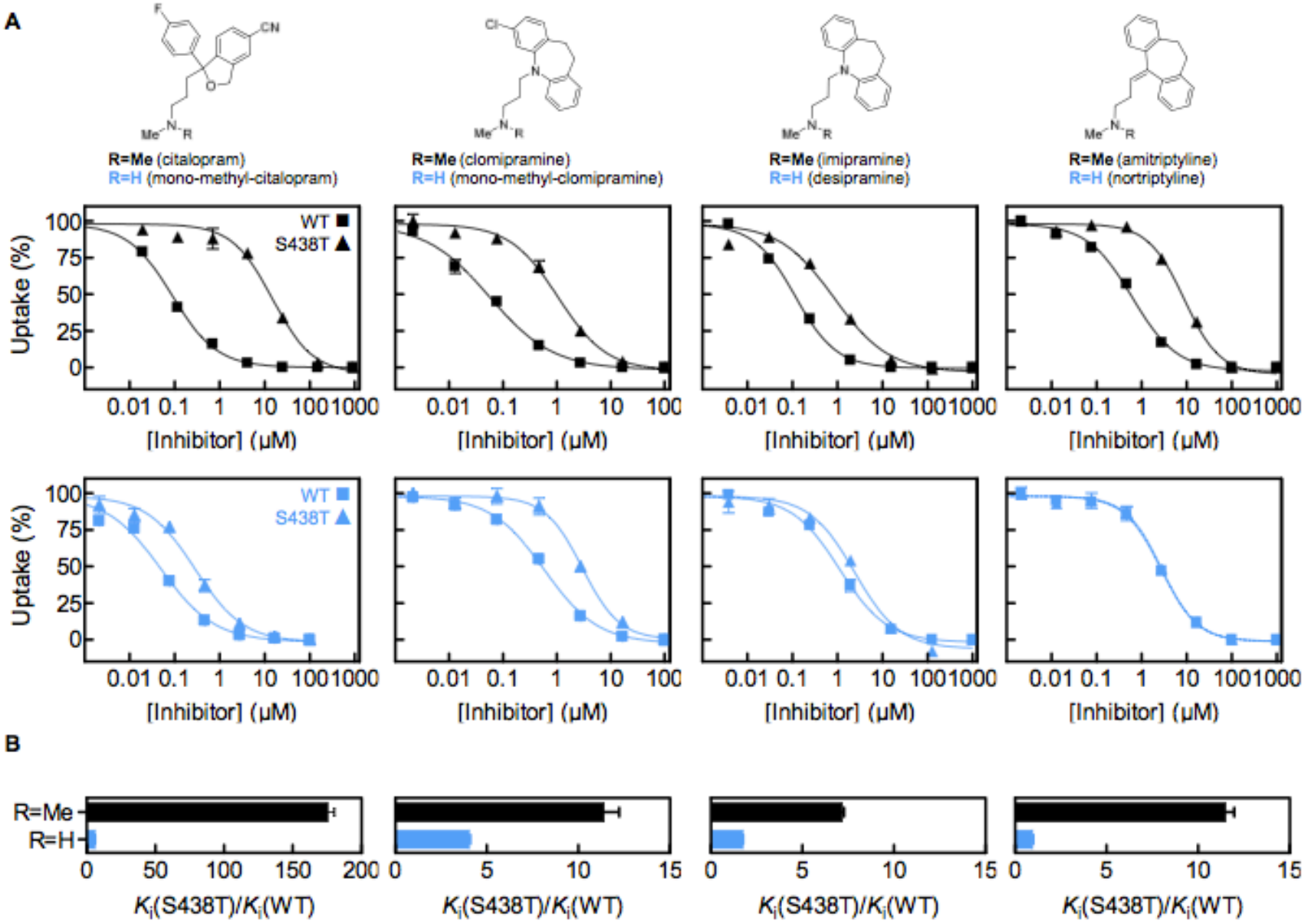


S438T
CIT: 175x *loss* of function
Monomethyl: 3.5x *loss* of function

A173S
CIT: 4.5x *gain* of function
Des-F: 16.5x *gain* of function

Doest that affect others antidepressants?

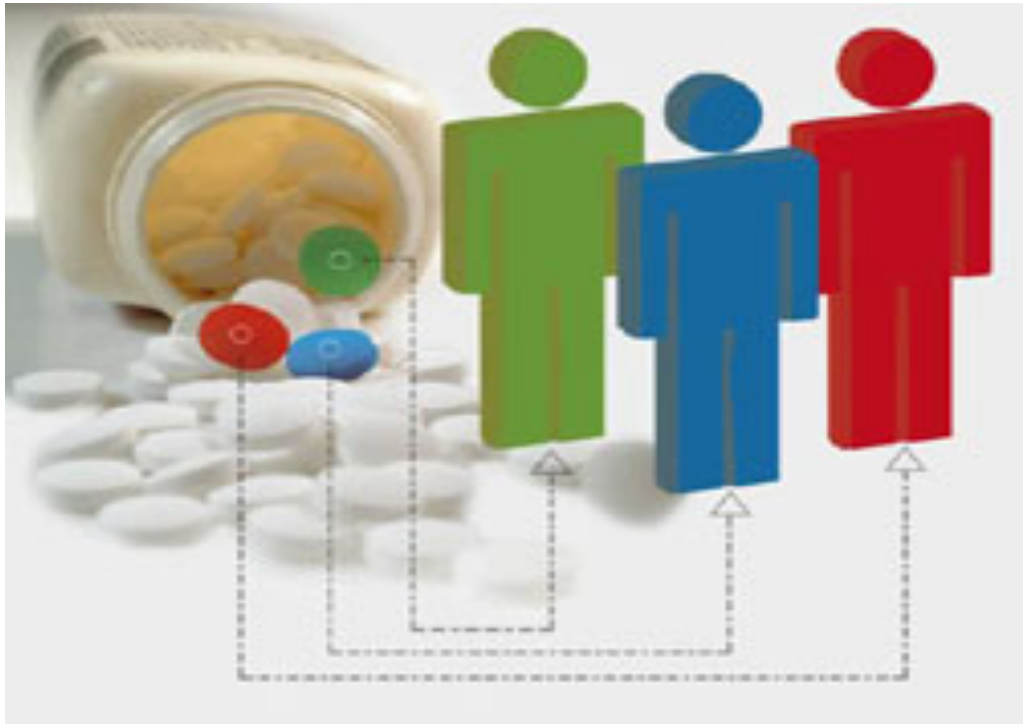
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS **CBS**



Pharmacogenetics and personalized medecine

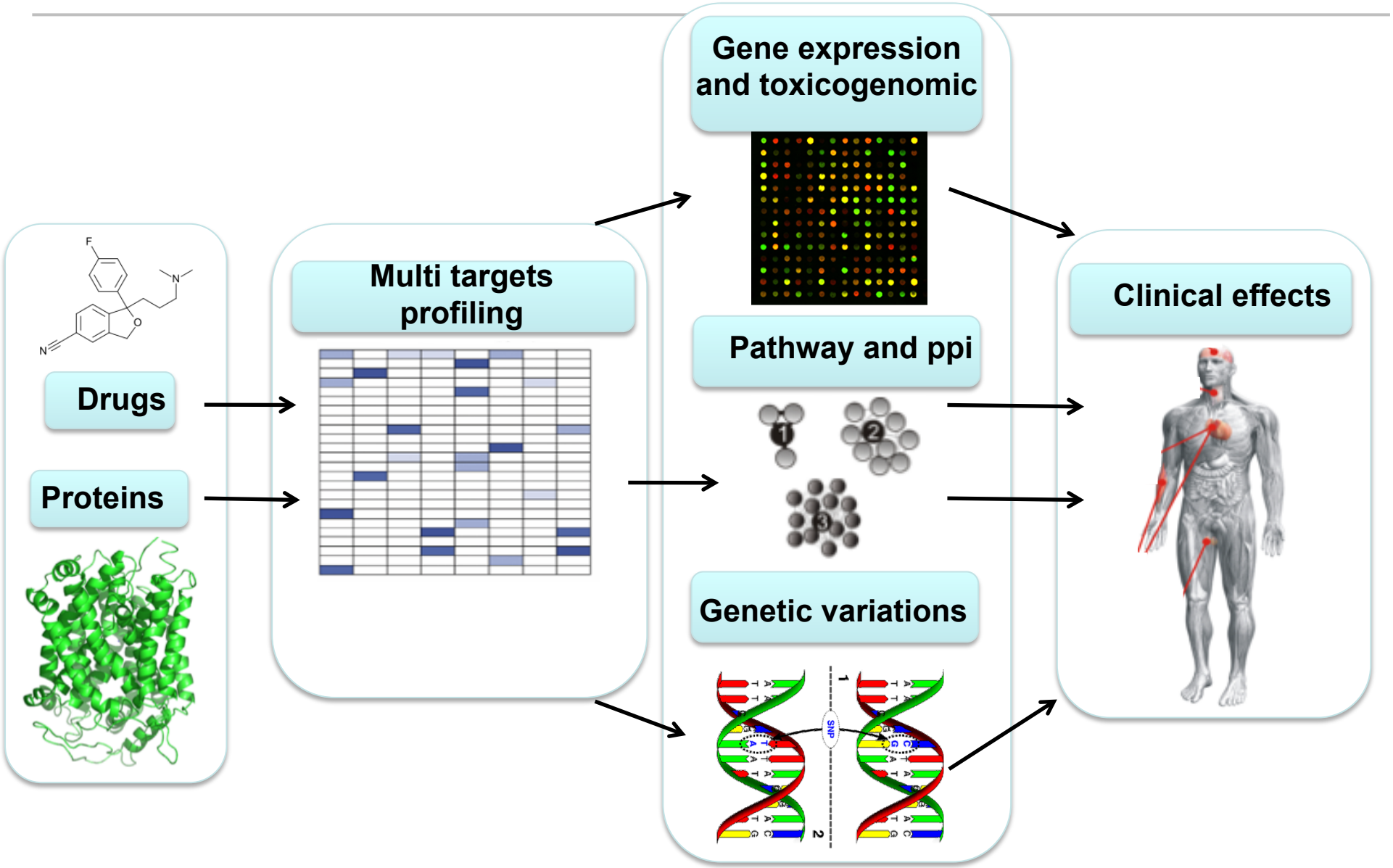
- **The study of how genetic variation influences the response and side-effects of a drug.**

- **Drug administration based upon genotyping of genes importance to drug response**



Conclusion

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



Acknowledgements

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

CBS

- Sonny Nielsen
- Karine Audouze
- Søren Brunak
- Tudor Oprea
- Integrative Systems Biology
- Computational Chemical Biology

Psychiatry Center Sct. Hans

- Henrik Rasmussen

EBI

- John Overington
- Anne Hersey

Copenhagen University

- Flemming Steen Jørgensen
- Kristian Strømgaard

DEER (EU)

ETOX project (EU-IMI)

OPENPHACTS (EU-IMI)

Villum Kann Rasmussen Foundation (Danish)

INDICES project (Danish)