Data Fusion by Matrix Completion for Exposome Target Interaction Prediction

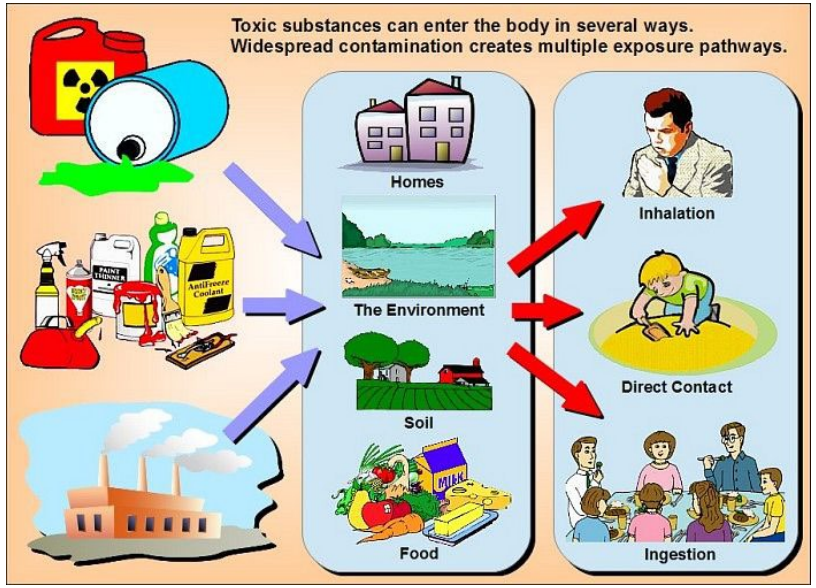
Kai Wang

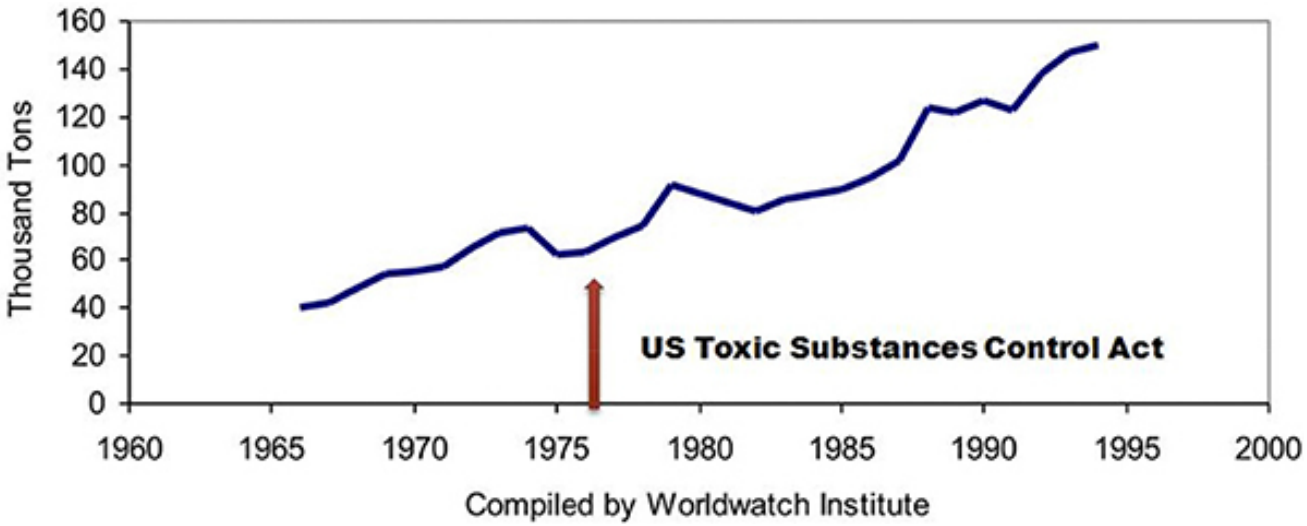
wangdaha@umich.edu

09/19/2023

Environmental Exposures

- Environmental exposures alone or in combination with genetics can affect human health and cause diseases.
- New chemical compounds are created and released into our environment every year. About 2,000 new chemicals are introduced into commerce annually in the U.S., at a rate of about seven new chemicals a day.
- Comprehensive toxicity assessment of all chemicals is an impossible task.

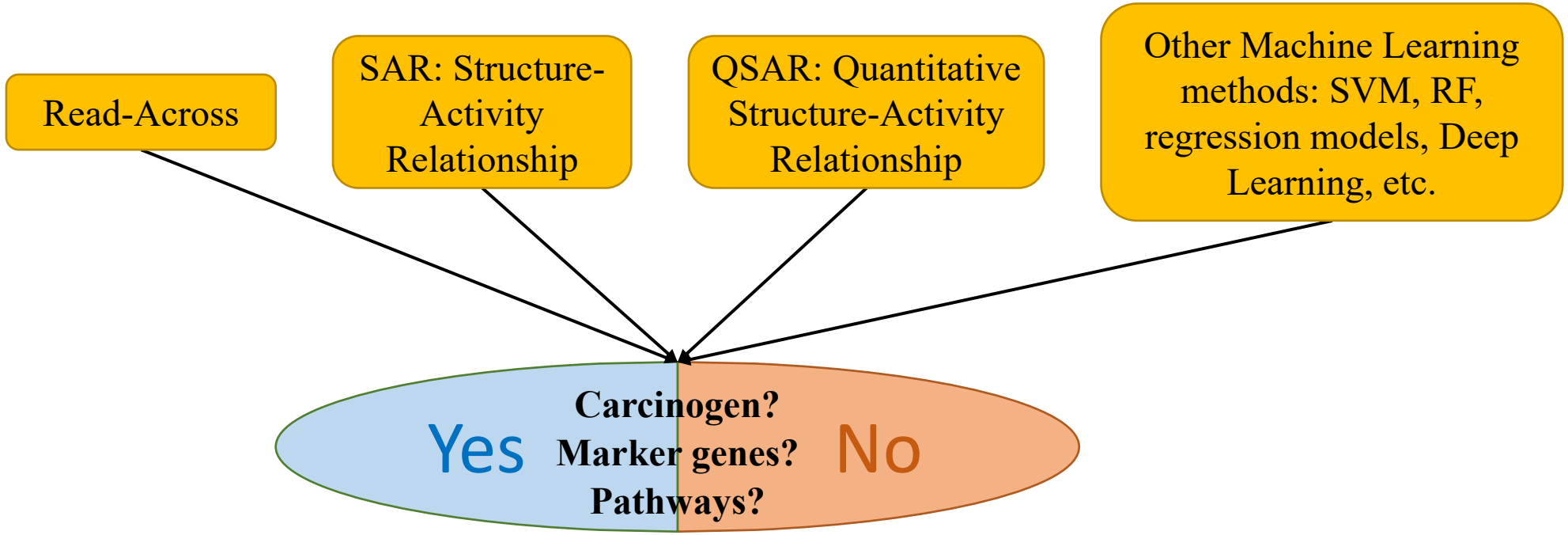




Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. Thorax. 2014 Sep 1;69(9):876-8.
Institute of Medicine R on EHS. The Challenge: Chemicals in Today's Society.

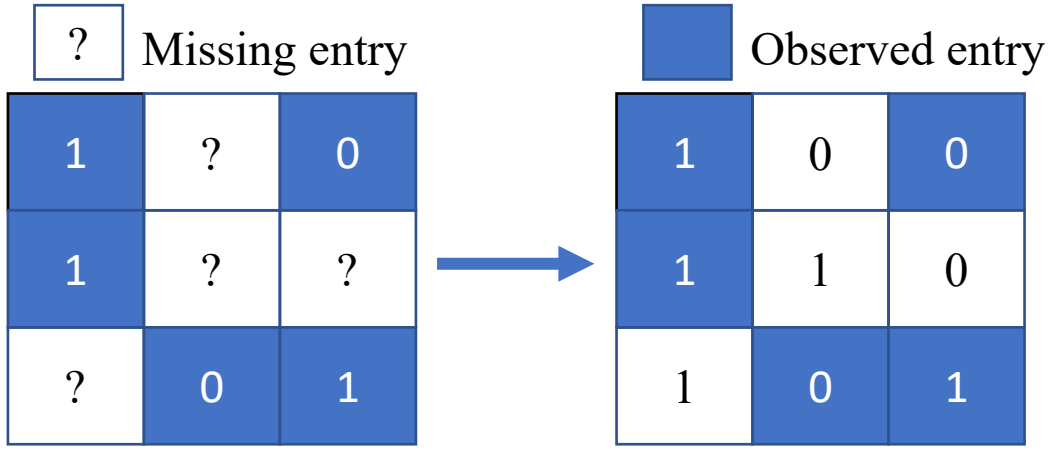
Computational Toxicology Methods

- Computational methods can be used to prioritize chemicals for *in vivo* or *in vitro* toxicity study.



Matrix Completion drug target prediction

Matrix completion is the task of filling in the missing entries of a partially observed matrix, which is equivalent to performing data imputation in statistics.



Drug target interaction: physical interaction
Environmental chemical: any types of interactions, transcription level, post translational modifications, etc.

Coupled Matrix-Matrix Completion

Input:

Main:

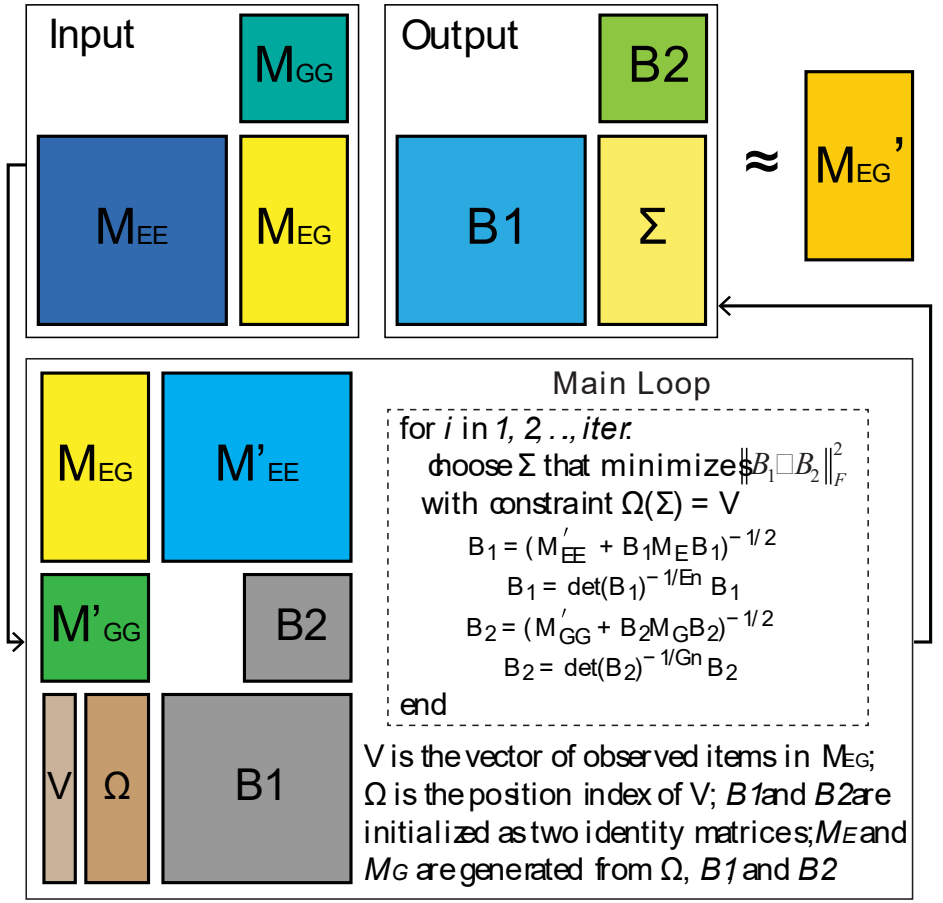
- Chemical-gene interaction matrix

Coupled:

- Chemical-chemical similarities
- Gene-gene similarities

Output:

- Chemical-gene interaction matrix with imputed values



Bagherian M, Kim RB, Jiang C, Sartor MA, Derksen H, Najarian K. Coupled matrix–matrix and coupled tensor–matrix completion methods for predicting drug–target interactions. Briefings in bioinformatics. 2021 Mar;22(2):2161-71.

The Comparative Toxicogenomics Database

- A curated database that promotes understanding about the effects of environmental chemicals on human health.
- 4,864 chemicals exposures
- 22,606 genes
- 201,375 known chemical-gene interaction

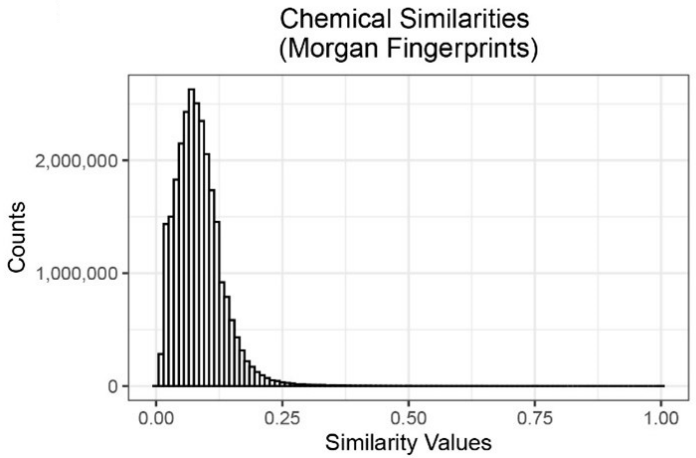
201,375/109,955,584 ≈ 0.183%

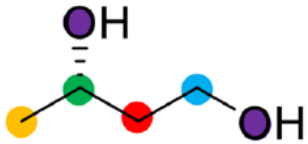
Chemical Similarities

- RDKit: Open-Source Cheminformatics Python package
- Two types of fingerprints were used to calculate chemical similarities.

Open-Source Cheminformatics and Machine Learning

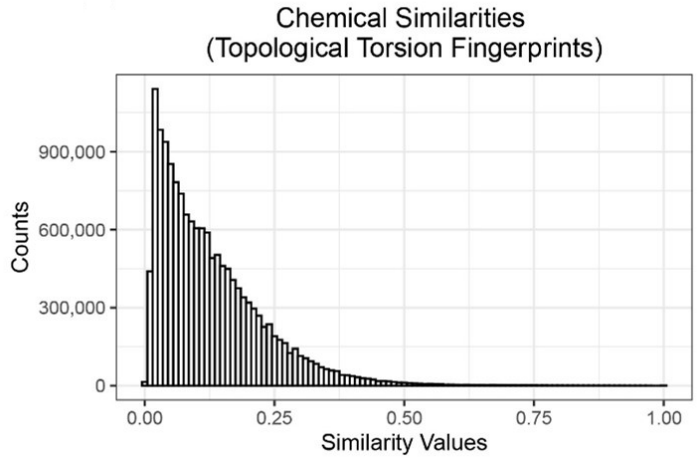
Morgan fingerprints





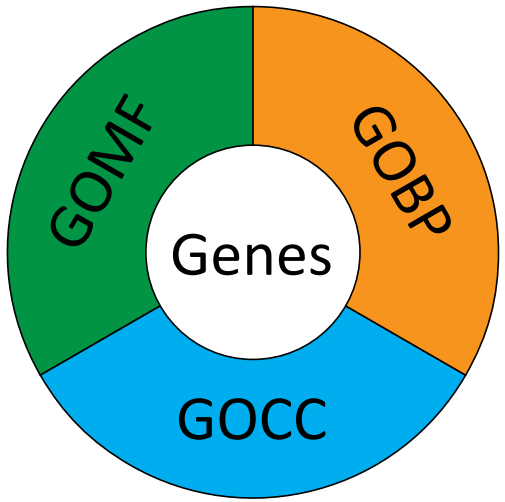
Molecular structure

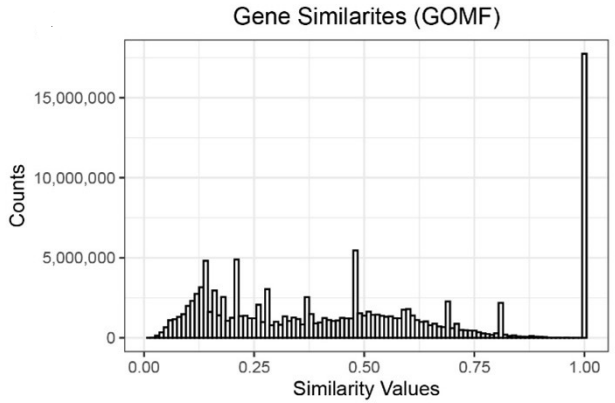
Topological torsion fingerprints

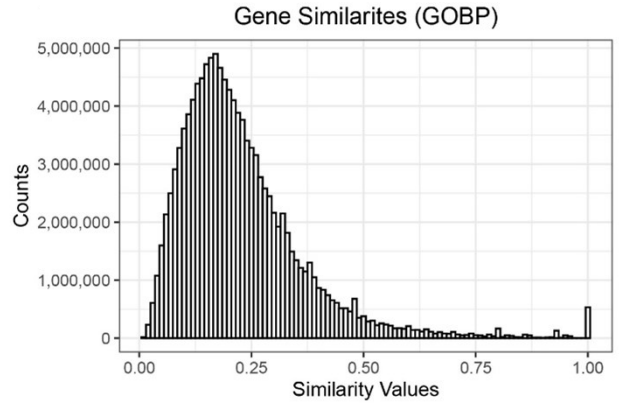


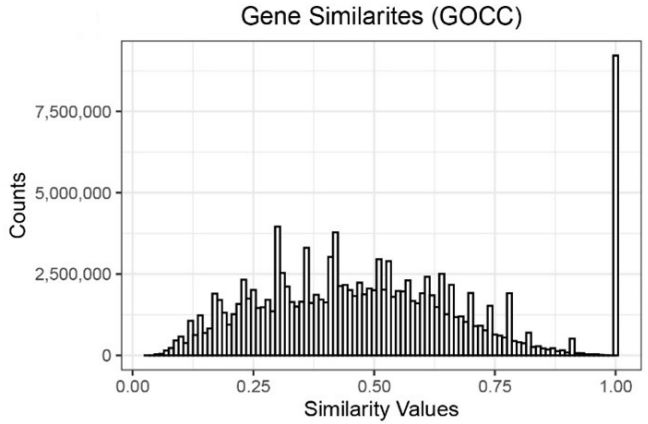
Gene Similarities

- GOSemSim: an R package for measuring semantic similarity among GO terms and gene products.





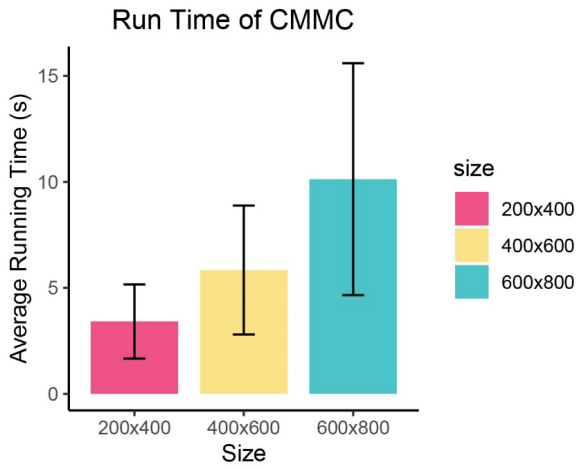


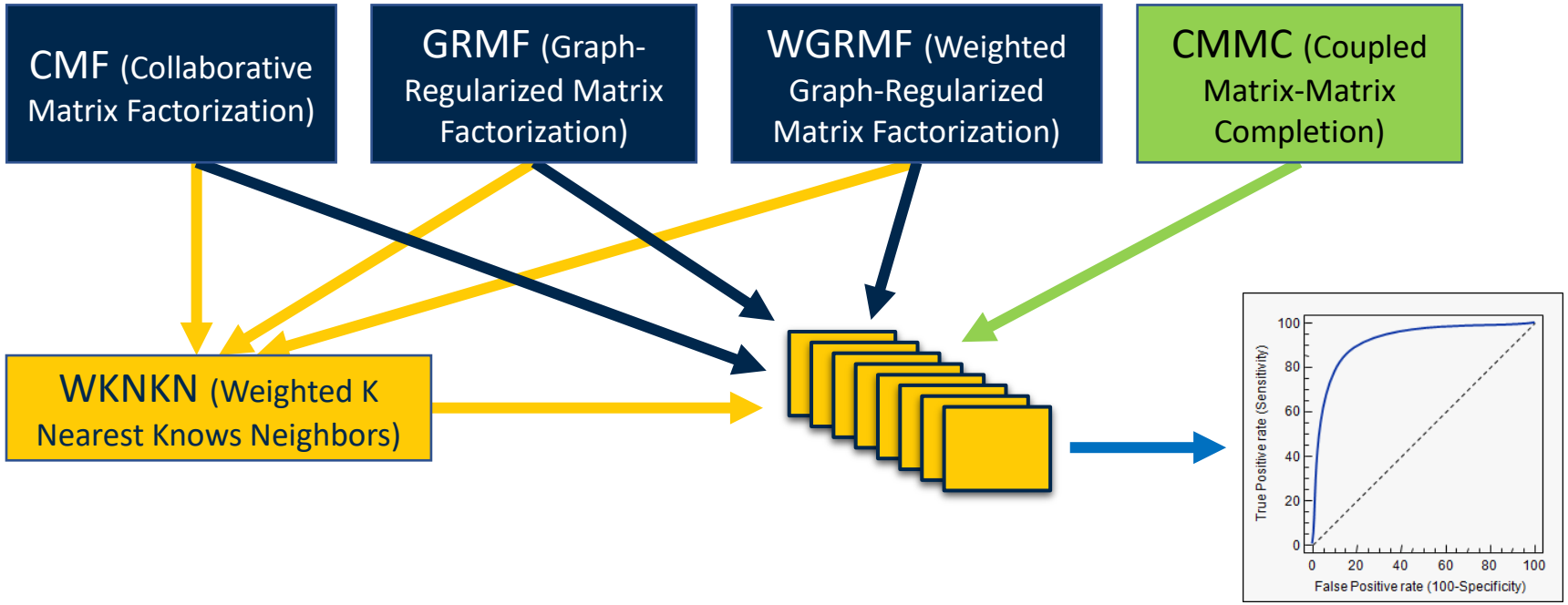


Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010 Apr 1;26(7):976-8.

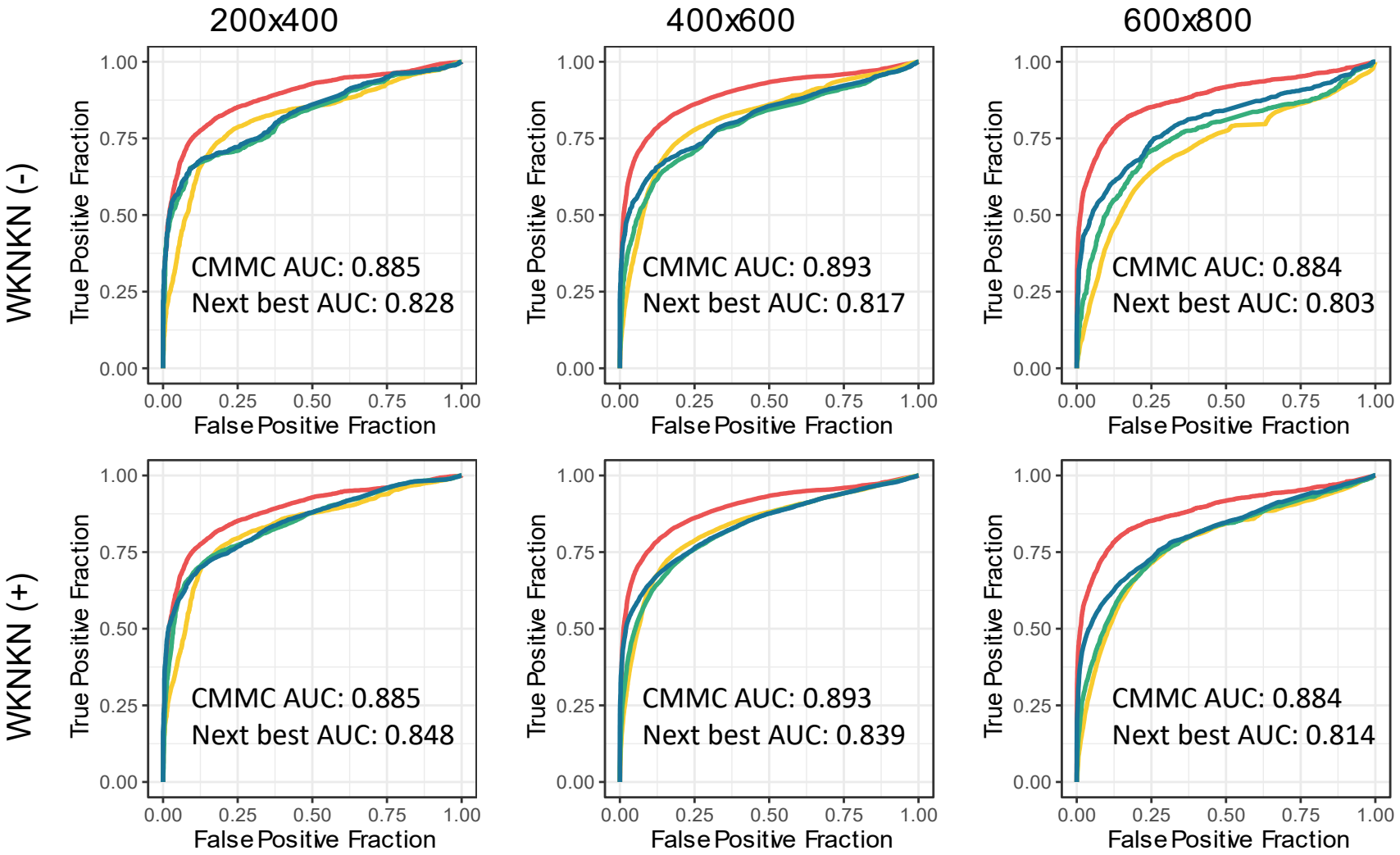
Performance Comparison

- MATLAB - CMF, GRMF, WGRMF
- CMMC - C++





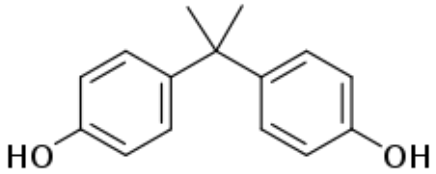
Performance Comparison



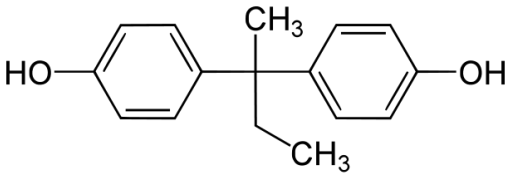
Case study - BPA with its analogs

Problem: how to get the target gene of novel chemicals without any biological activity data?

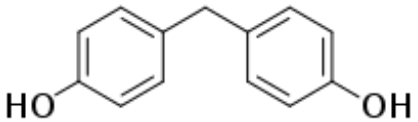
BPA



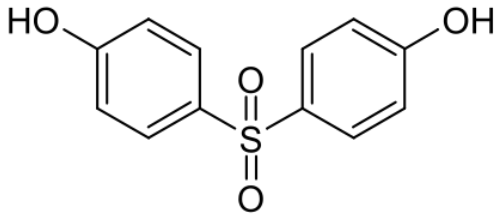
BPB

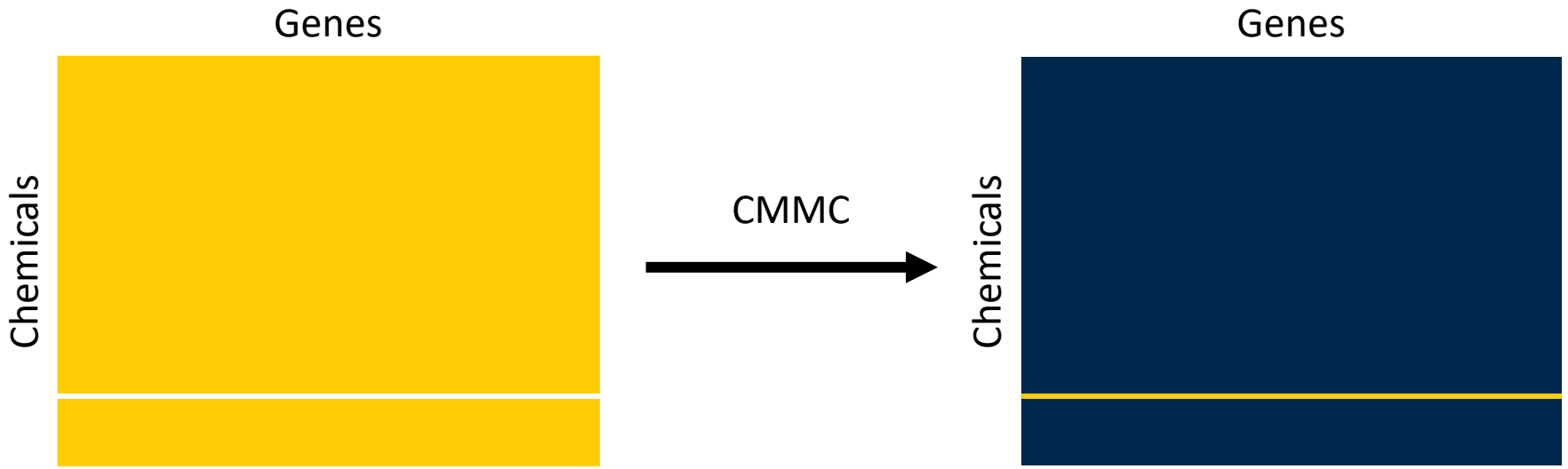


BPF



BPS

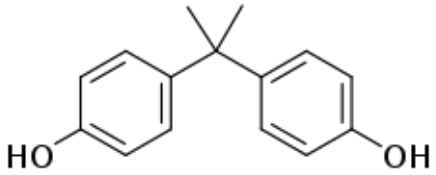




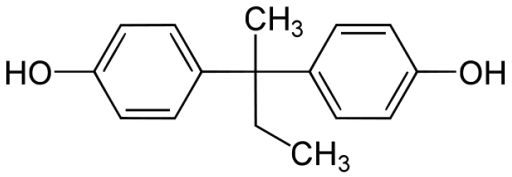
Case study - BPA with its analogs

Problem: how to get the target gene of novel chemicals without any biological activity data?

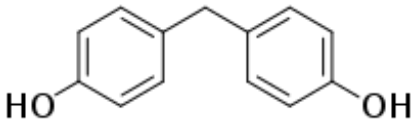
BPA



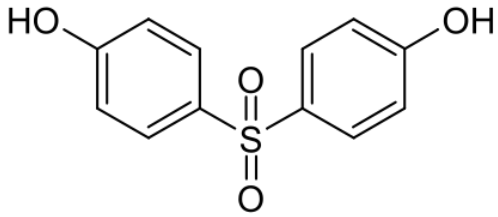
BPB

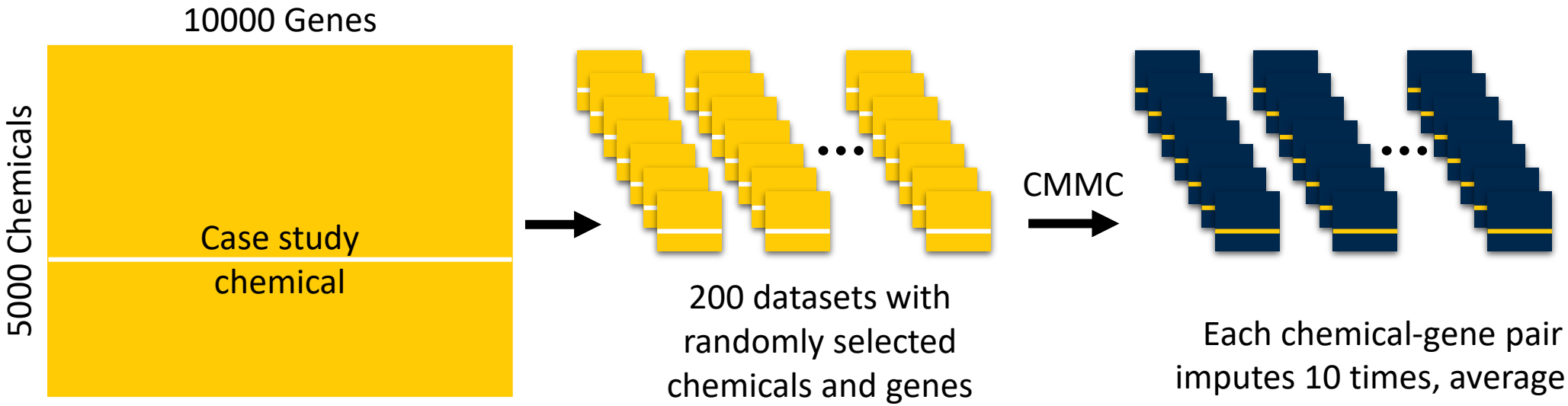


BPF



BPS

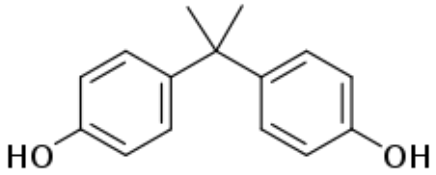




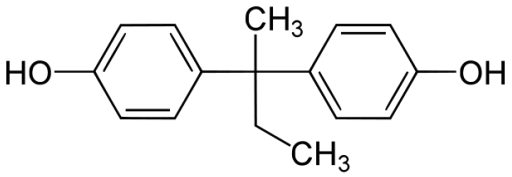
Case study - BPA with its analogs

Problem: how to get the target gene of novel chemicals without any biological activity data?

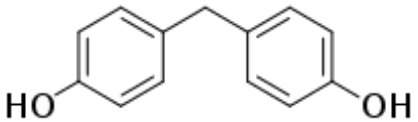
BPA



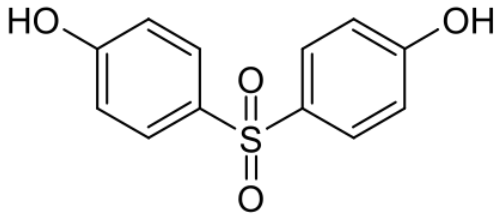
BPB

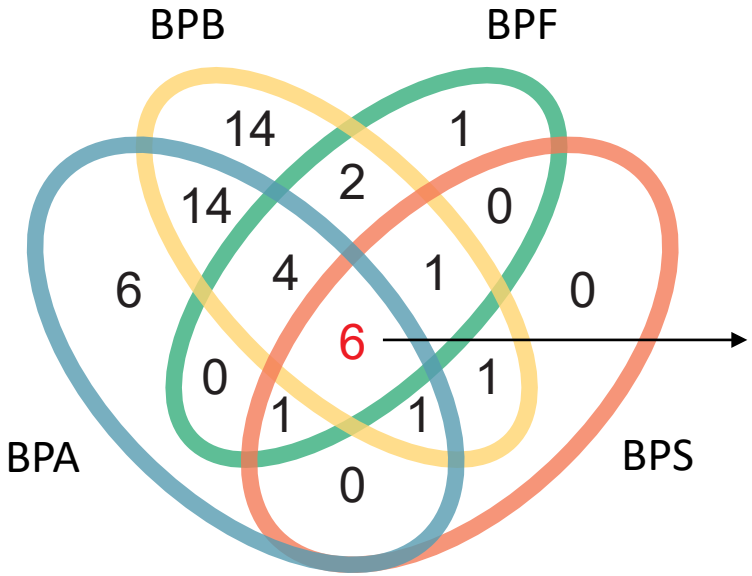


BPF

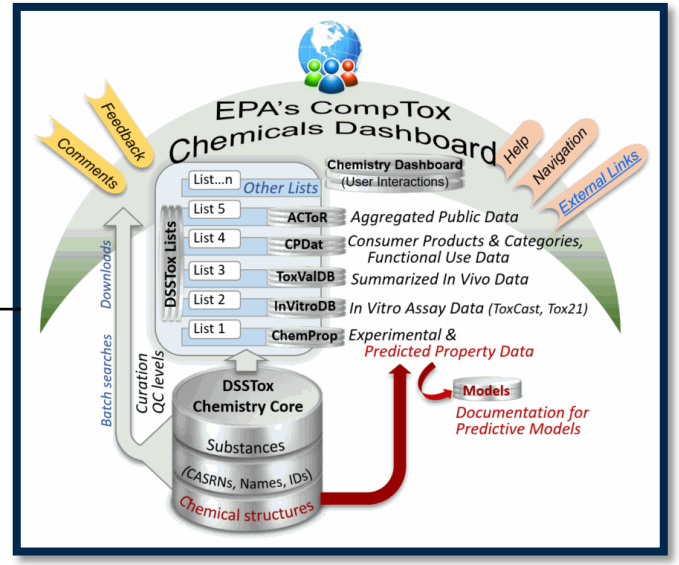


BPS





ESR1
ESR2
AR
NR1I2
PGR
PPARG
PPARD



GPER1: target gene of BPB

- G Protein-Coupled Estrogen Receptor 1: This gene encodes a multi-pass membrane protein that localizes to the endoplasmic reticulum and a member of the G-protein coupled receptor 1 family.
- CMMC predicted this gene as a top interactive gene only for BPB.

Bisphenol AF and Bisphenol B Exert Higher Estrogenic Effects than Bisphenol A via G Protein-Coupled Estrogen Receptor Pathway

Lin-Ying Cao,[†,‡] Xiao-Min Ren,[*,†] Chuan-Hai Li,[†,‡] Jing Zhang,[†] Wei-Ping Qin,[†,‡] Yu Yang,[†] Bin Wan,[†] and Liang-Hong Guo[*,†,‡]

[†]State Key Laboratory of Environmental Chemistry and Eco-toxicology, Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Beijing 100085, P. R. China

[‡]University of Chinese Academy of Sciences, Beijing 100039, P. R. China

…six BPA analogues bound to GPER directly, with bisphenol AF (BPAF) and bisphenol B (BPB) displaying much higher (~9-fold) binding affinity than BPA…

Future plan - CTMC and CTTC

CMMC

CTMC

CTTC

Acknowledgement

Dr. Maureen Sartor

Dr. Dana Dolinoy

Dr. Justin Colacino

Dr. Maryam Bagherian

Nicole Kim

Kai Li

Elysia Chou

Supported by NIH/NIEHS grants P30ES017885, R35ES031686, R01ES028802, and **K99ES034429**