

Fish Connectivity Mapping II: an Upgraded Library of Transcriptomic Profiles and its Toxicological Applications

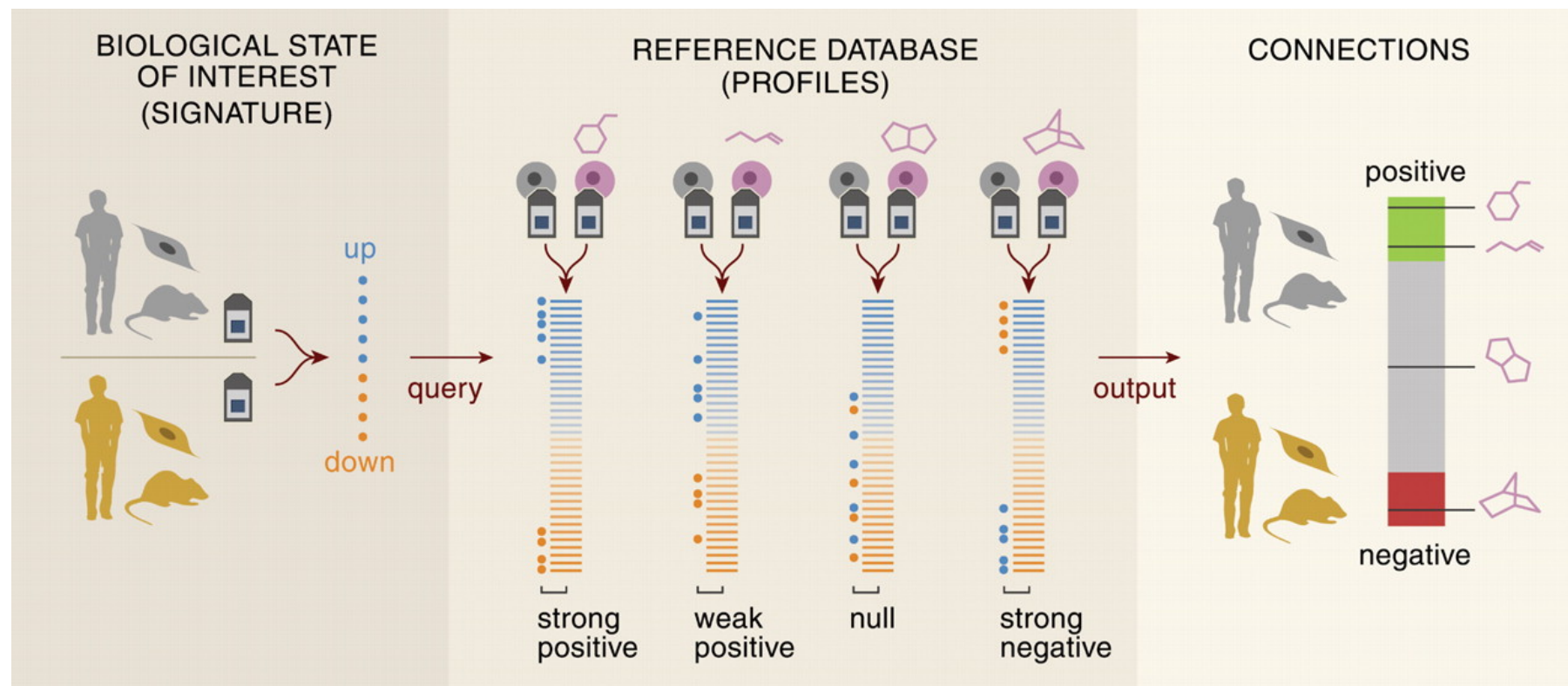
Rong-Lin Wang, David Bencic, Adam Biales, Robert Flick, Morgan Hu
Great Lakes Toxicology and Ecology Division/ORD/US EPA
Cincinnati, OH

Conclusions:

- A significant expansion of chemical/biological coverage in the upgraded Cmap library
- Improved cross-mapping methods for both signature queries and target transcriptomic profiles
- Utilities in chemical exposure diagnosis
- Utilities in chemical grouping/MOA discovery

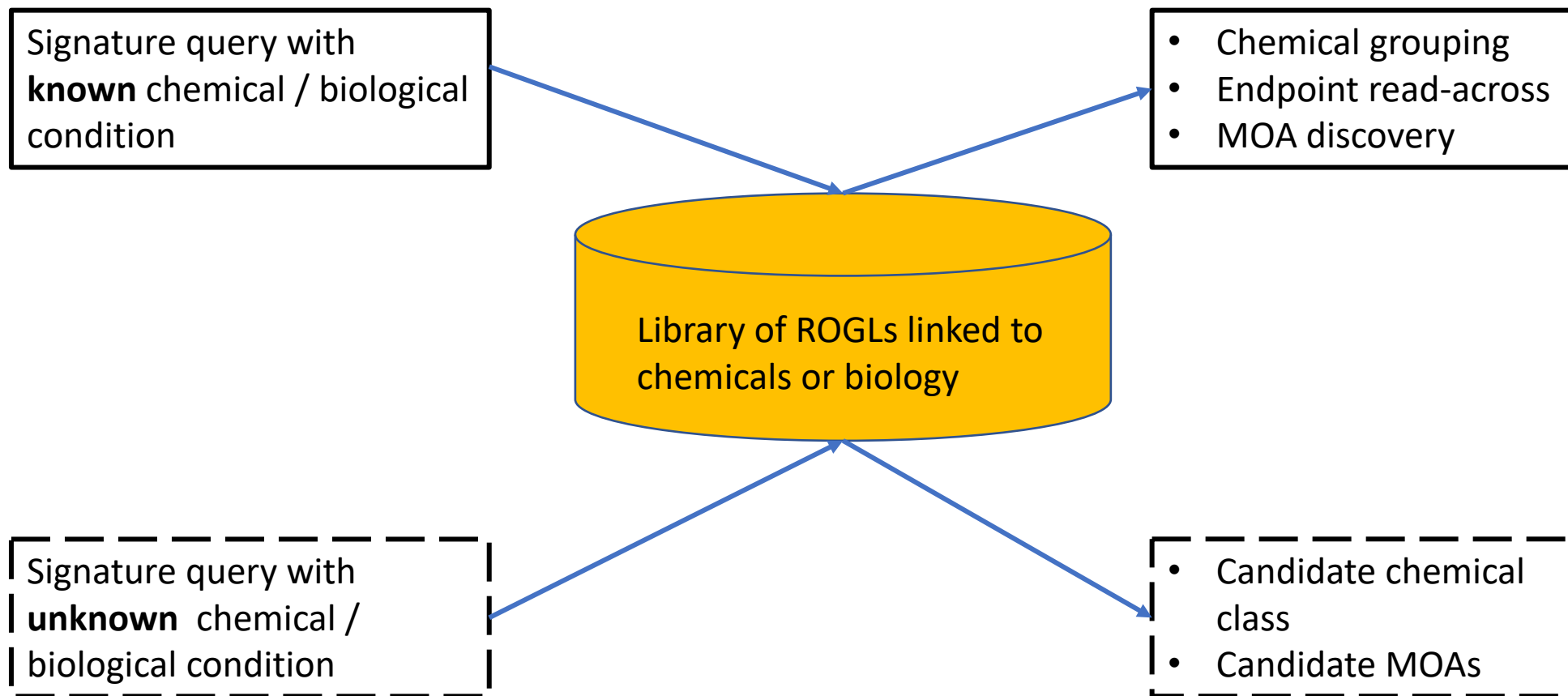
The Landmark Study

The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease

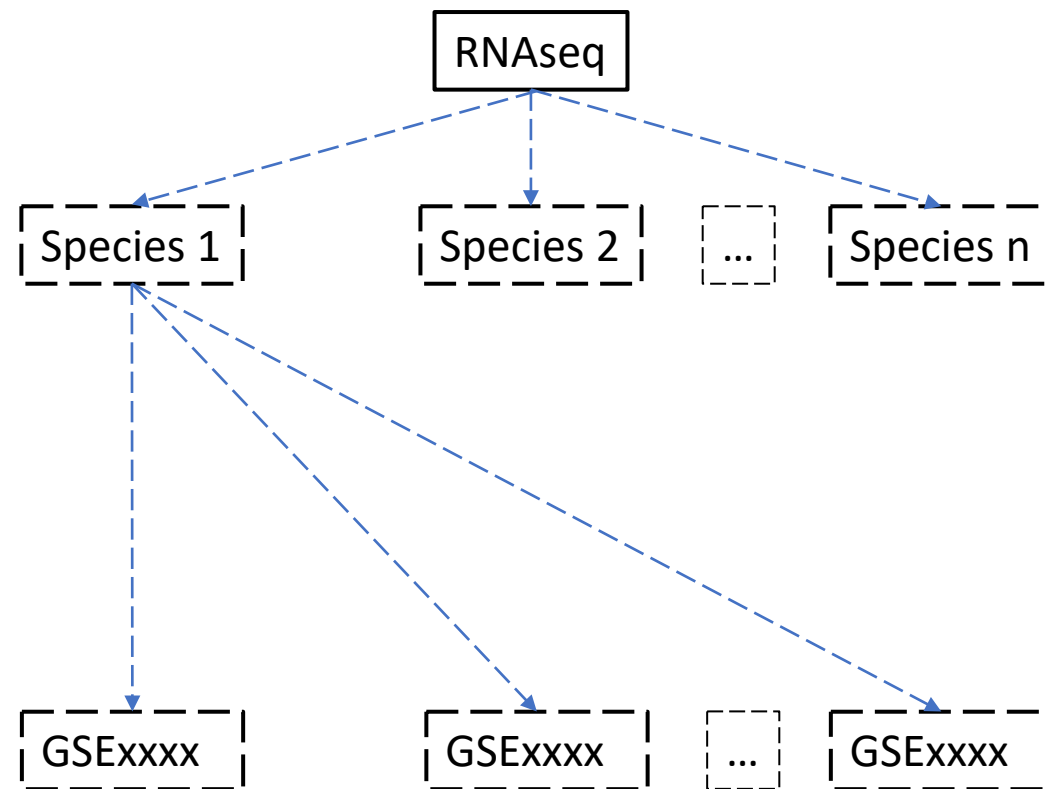
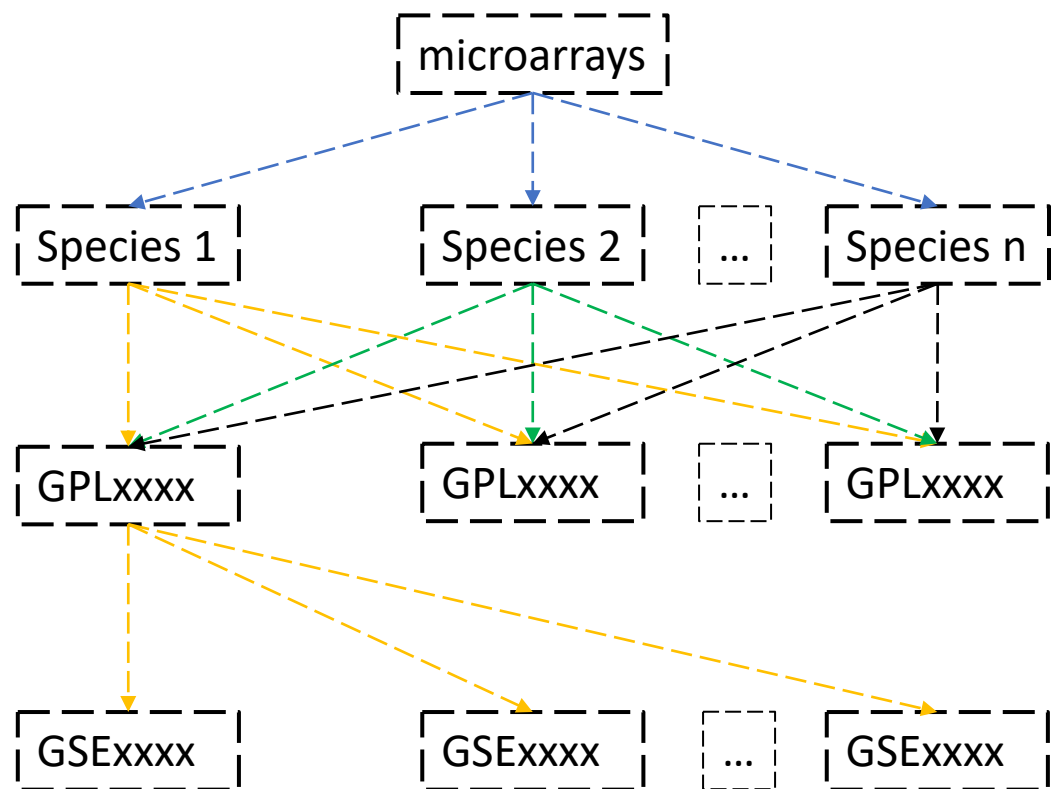


Justin Lamb et al. Science 2006;313:1929-1935

Cmap Toxicological Applications

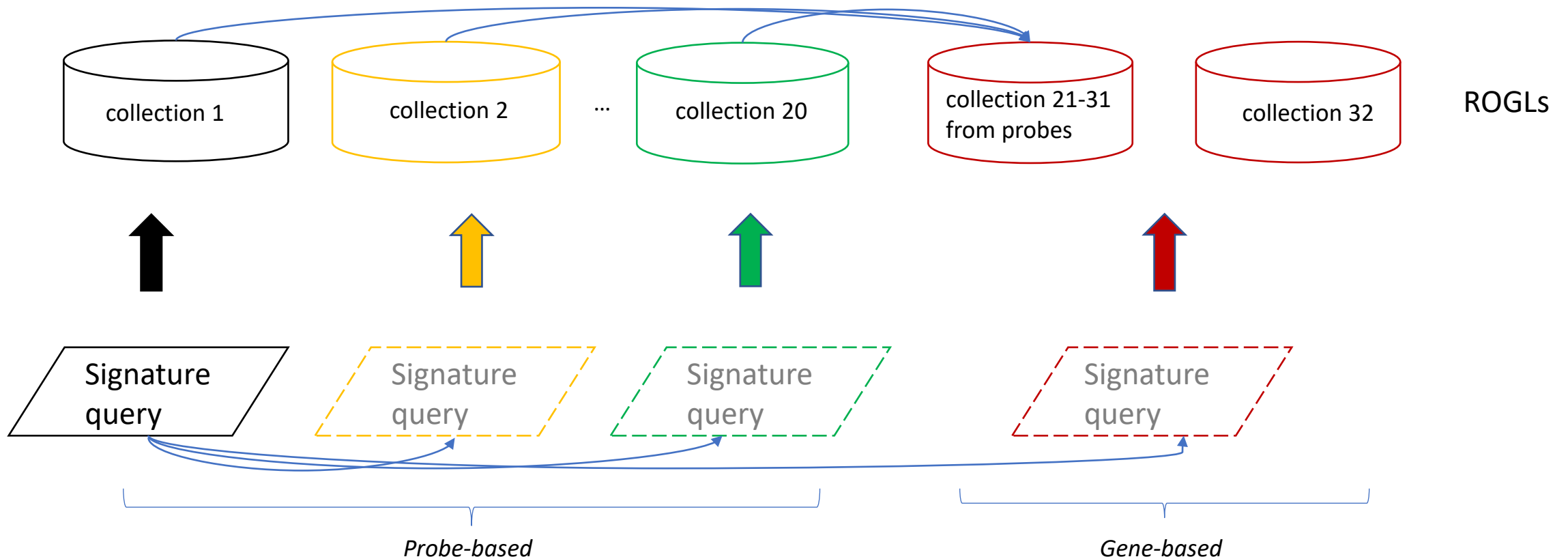


Cmap Challenges: Heterogeneity in Public Data



Virtual Integration of Cmap Library by Cross Mapping

Interspecific query enabled: 22 fish, human, mouse, rat, fruit fly



Cmap Workflow

RNAseq

Mapping reads to FHM
transcriptome / genome

Generate read counts

Generate query signatures
and ROGLs

Arrays

Compile a list of NCBI-GEO GSE

Parse and summarize GSE SOFT

Annotate metadata manually

Download all GSM raw data

Generate query signatures and
ROGLs

Cross map query
signatures and ROGLs

Set up and run Cmap

Enrichment analysis of
mapped ROGLs



Transcriptomic Datasets Included in This Study

Technology	NO. studies	NO. samples	Design platforms	Designs	Source
FHM array	68	4222	9	One color, two color common reference	NCBI-GEO
ZF Affymetrix array	151	2147	4	Affy GeneST, Affy Genome Array	NCBI-GEO
ZF Agilent array	214	4416	28	One color, two color common reference, two color A optimal loop, two color direct comparison	NCBI-GEO
RNA-seq	1	989	1		EPA
total	433	11774	42		

A Upgraded Cmap Library

Species / technologies	NO. ROGLs	NO. ROGL sets	NO. consolidated ROGL collections by probes	NO. consolidated ROGL collections by genes ²	NO. signature queries
FHM array	2818	448 ¹	4	3	232
ZF Affymetrix array	1307	349	3	2 ³	294
ZF Agilent array	3066	769 ¹	13	7 ³	604
RNA-seq	989	121	---	1	80
Total	8180	1683	20	12	1210

¹GSE27067 has four ROGL sets duplicated in both species.

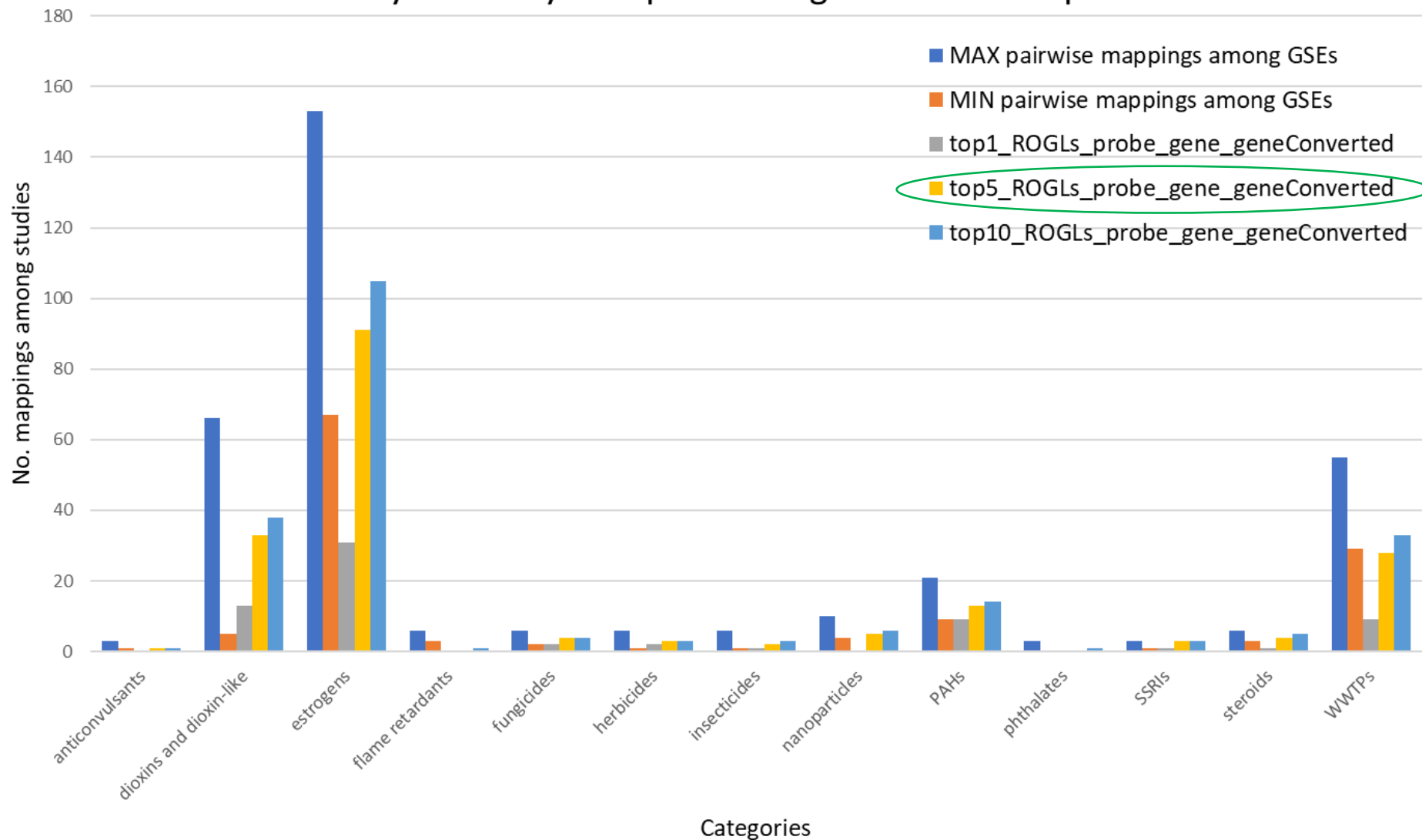
²ROGLs by gene IDs, both from the original FHM RNA-seq and those transformed from the probe - based ROGLs.

³One of the ROGL collections contains both ZF Affymetrix and ZF Agilent arrays.

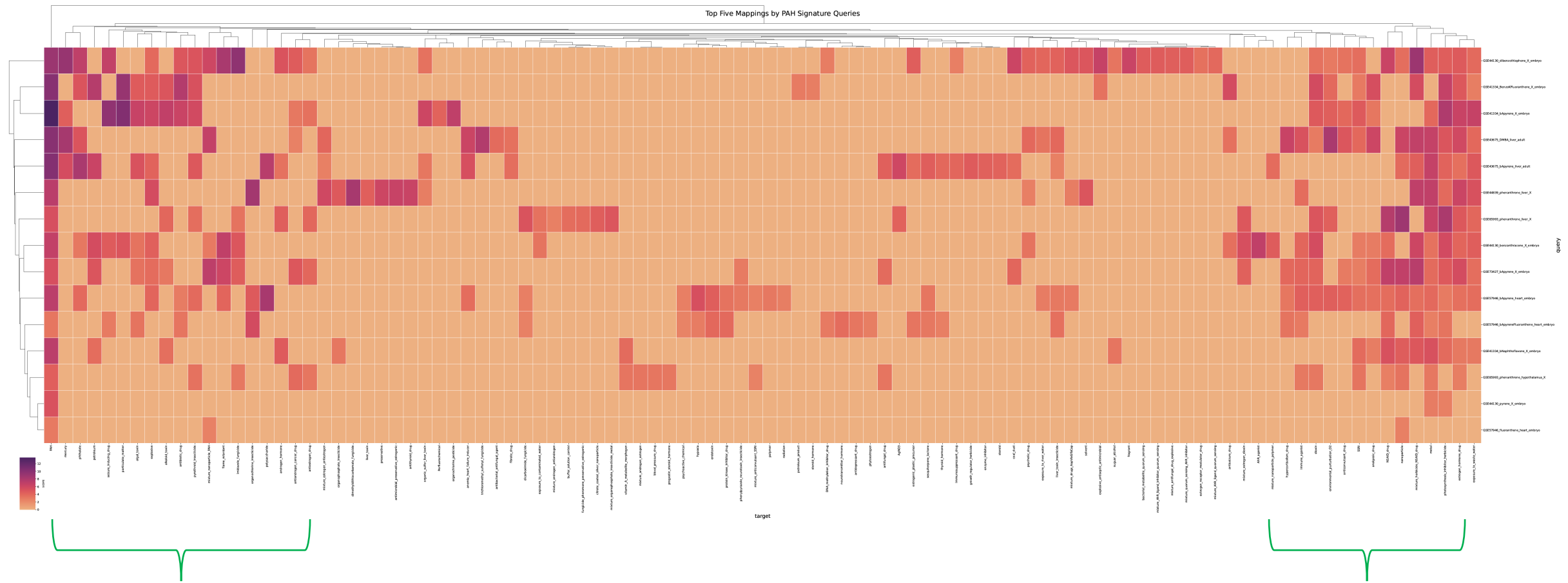
Performance Evaluation: 13 Test Class of Chemicals / Conditions

category	chemicals	No. GSEs
anticonvulsants	carbamazepine, diazepam	3
dioxin and dioxin-like	bApyrene, BenzoKFluoranthene, bNaphthoflavone, PCB126, TCDD	12
estrogens	BPA, BPAF, DES, E2, EE2, genistein, linuron	18
flame retardants	BDE209, MIPTphosphate, TBBPA, TDCIPP	4
fungicides	propiconazole, prochloraz, vinclozolin, Ziram	4
herbicides	atrazine, Diuron, linuron, Paraquat	4
insecticides	azinphosmethyl, cPyrifosCu, terbufos, Dichlorvos	4
nanoparticles	AgNP, PVP-AgNP, Ag+	5
PAHs	bApyrene, bApyrene + Fluoranthene, benzanthracene, BenzoKFluoranthene, bNaphthoflavone, dibenzothiophene, DMBA, Fluoranthene, phenanthrene, pyrene	7
phthalates	Diisobutyl phthalate, Butyl benzyl phthalate, Di-n-butyl phthalate, Di(2-ethylhexyl) phthalate	3
SSRIs	fluoxetine, paroxetine, sertraline, venlafaxine	3
steroids	beclomethasone, dexamethasone, prednisolone, triamcinolone	4
WWTPs	multiple sites in AZ, CA, MN, VA	11

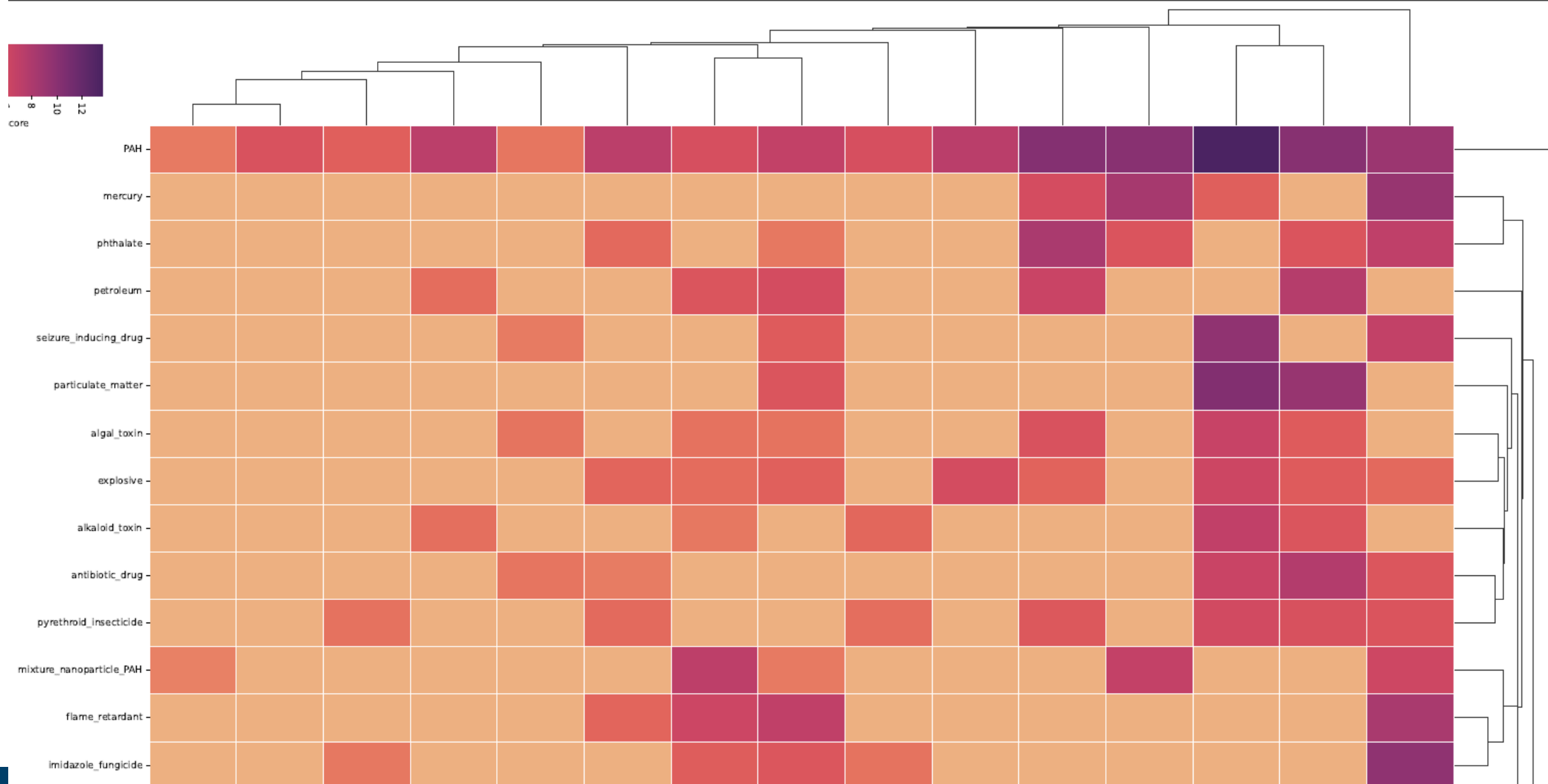
Performance by Recovery of Expected Targets across Independent Studies



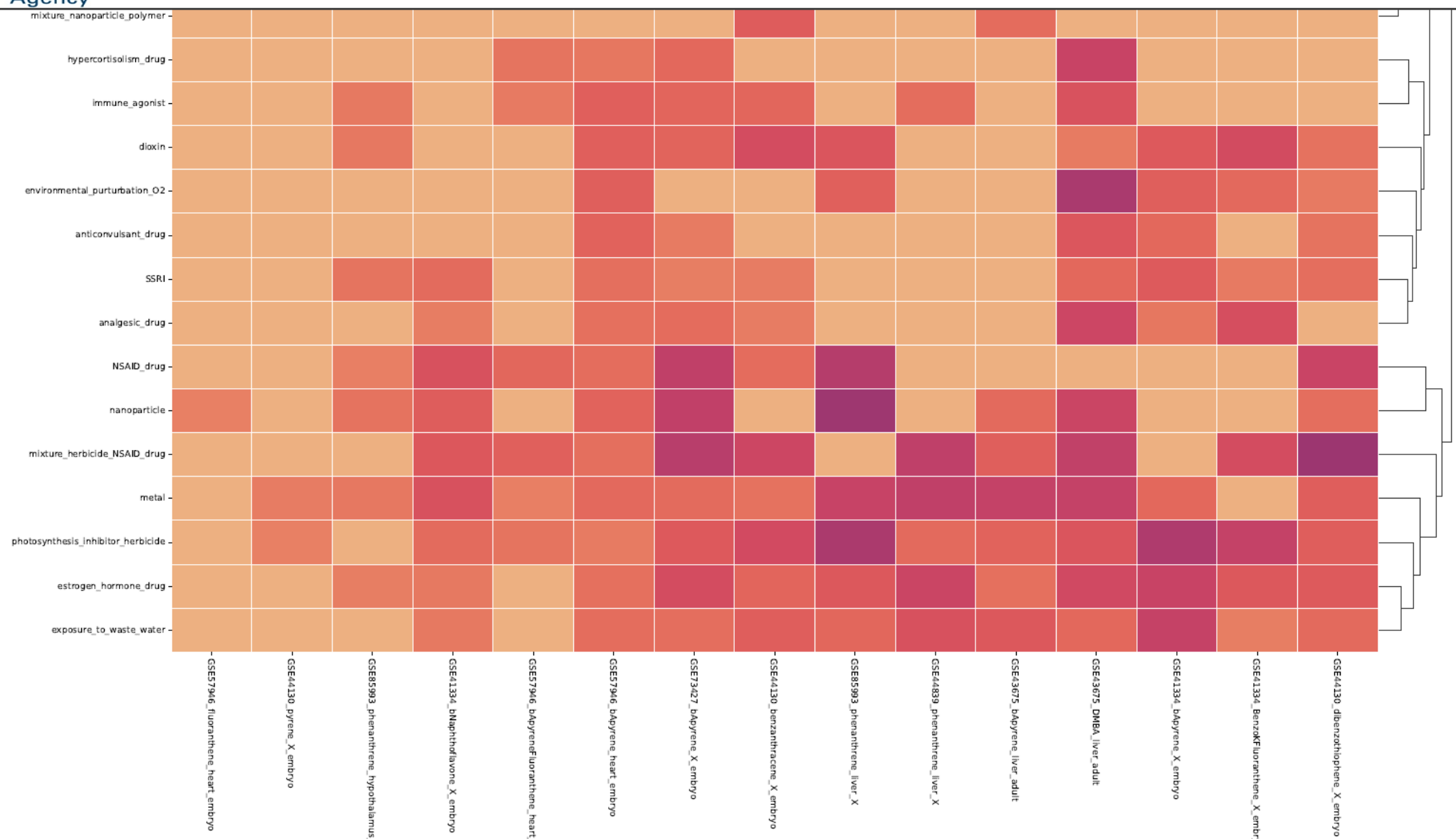
Top Five Hits by PAH Signature Queries



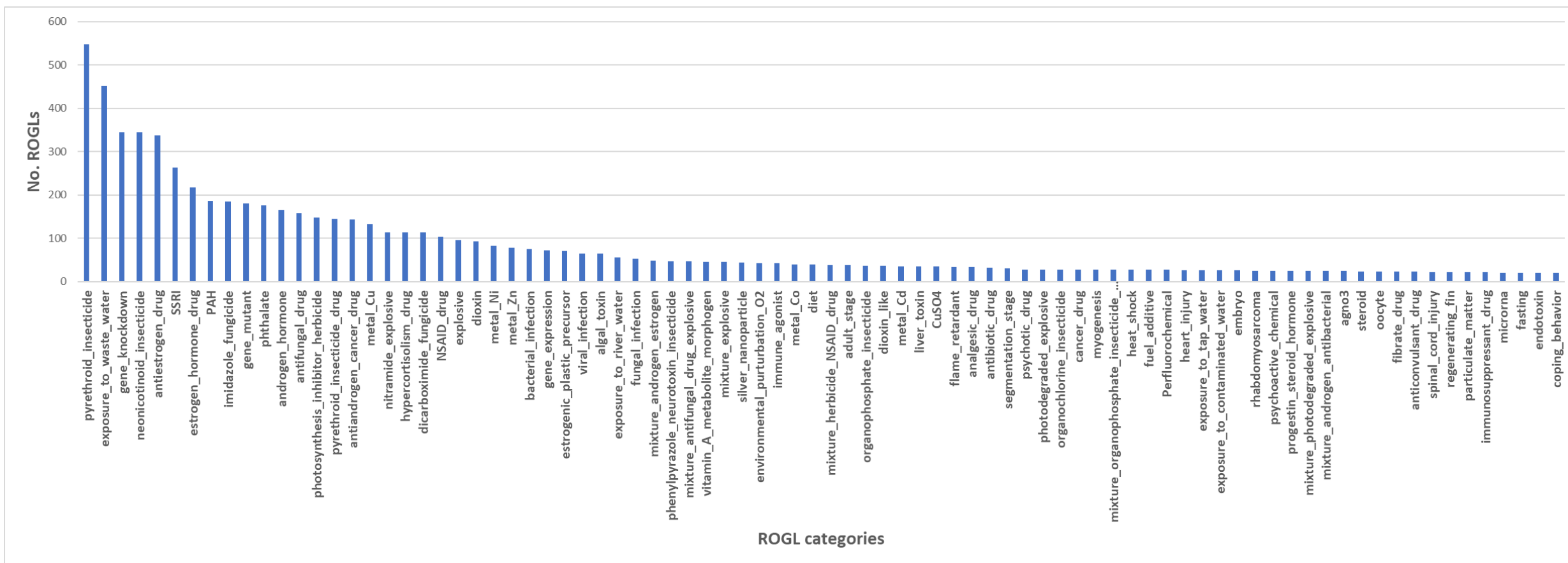
Top Five Hits by PAH Signature Queries---Potential Targets of Interest



Top Five Hits by PAH Signature Queries---Potential Targets of Interest



ROGLs Categorized by Applications/MOAs of Their Chemical/Biological Conditions



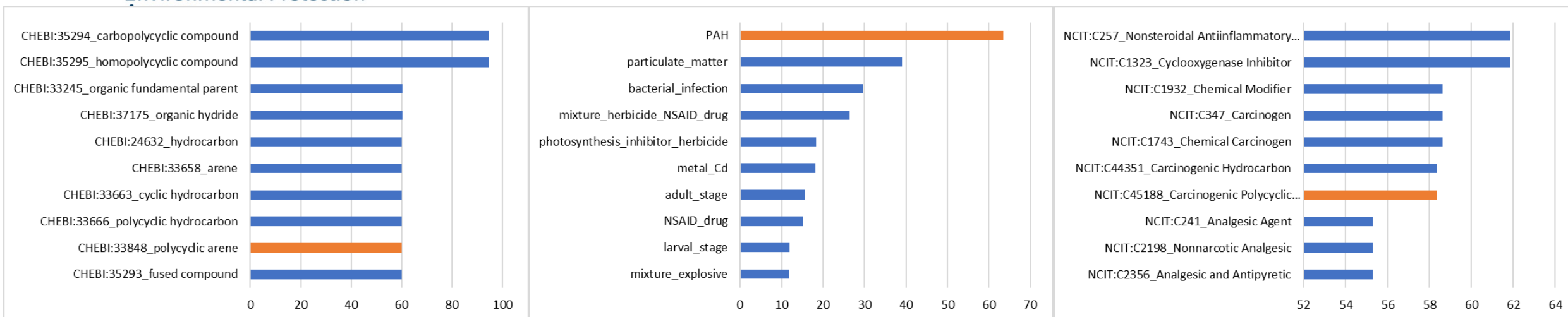
*among the 259 categories, only those >= 20 ROGLs are displayed

Ontologies Terms Mapped to Cmap Library

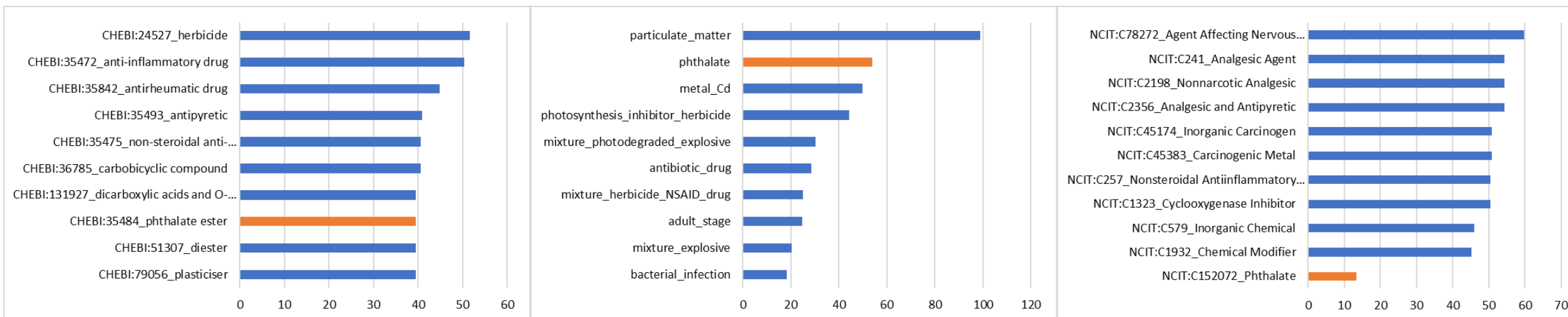
ontology	description	No. mapped terms	Ontology	Description	No. mapped terms
NCIT	NCI Thesaurus OBO Edition	235	ECTO	Environmental conditions, treatments and exposures ontology	9
CHEBI	Chemical Entities of Biological Interest	222	HP	Human phenotype ontology	8
UBERON	Uberon multi-species anatomy ontology	34	ZECO	Zebrafish Experimental Conditions Ontology	6
FMA	Foundational Model of Anatomy Ontology	23	MPATH	Mouse pathology ontology	6
ZFA	Zebrafish anatomy and development ontology	20	DOID	Human Disease Ontology	6
GO	Gene ontology	20	OBI	Ontology for Biomedical Investigations	5
XCO	Experimental condition ontology	19	TXPO	Toxic Process Ontology	4
MONDO	Mondo Disease Ontology	17	MP	Mammalian Phenotype Ontology	4
DRON	The drug ontology	17	ExO	Exposure ontology	3
NCBITaxon	NCBI organismal classification	11	SO	Sequence types and features ontology	2
ENVO	Environment Ontology	11	PCO	Population and Community Ontology	2
CL	Cell ontology	11	RBO	Radiation Biology Ontology	1
NBO	Neuro Behavior Ontology	9	CHMO	Chemical Methods Ontology	1

Enriched Ontology Terms and Categories in Mapped Targets

PAHs

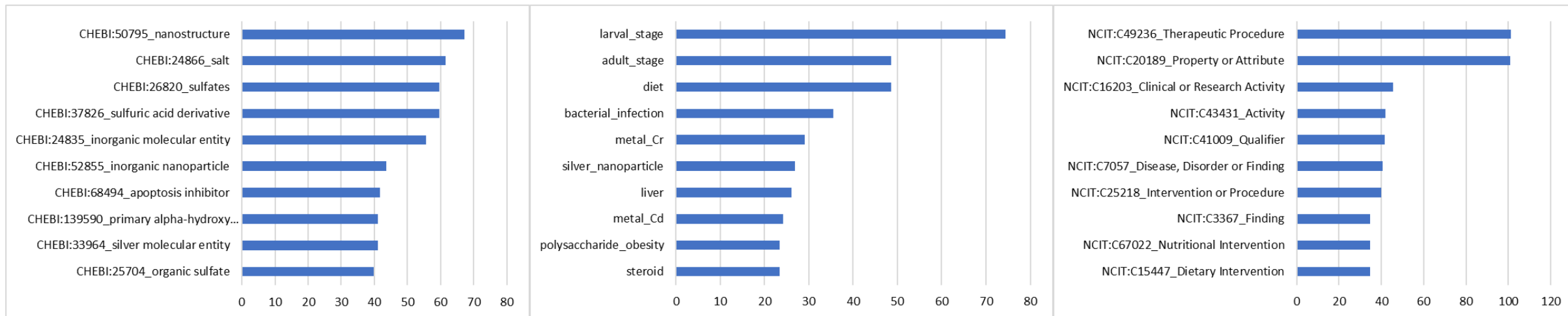


Phthalates



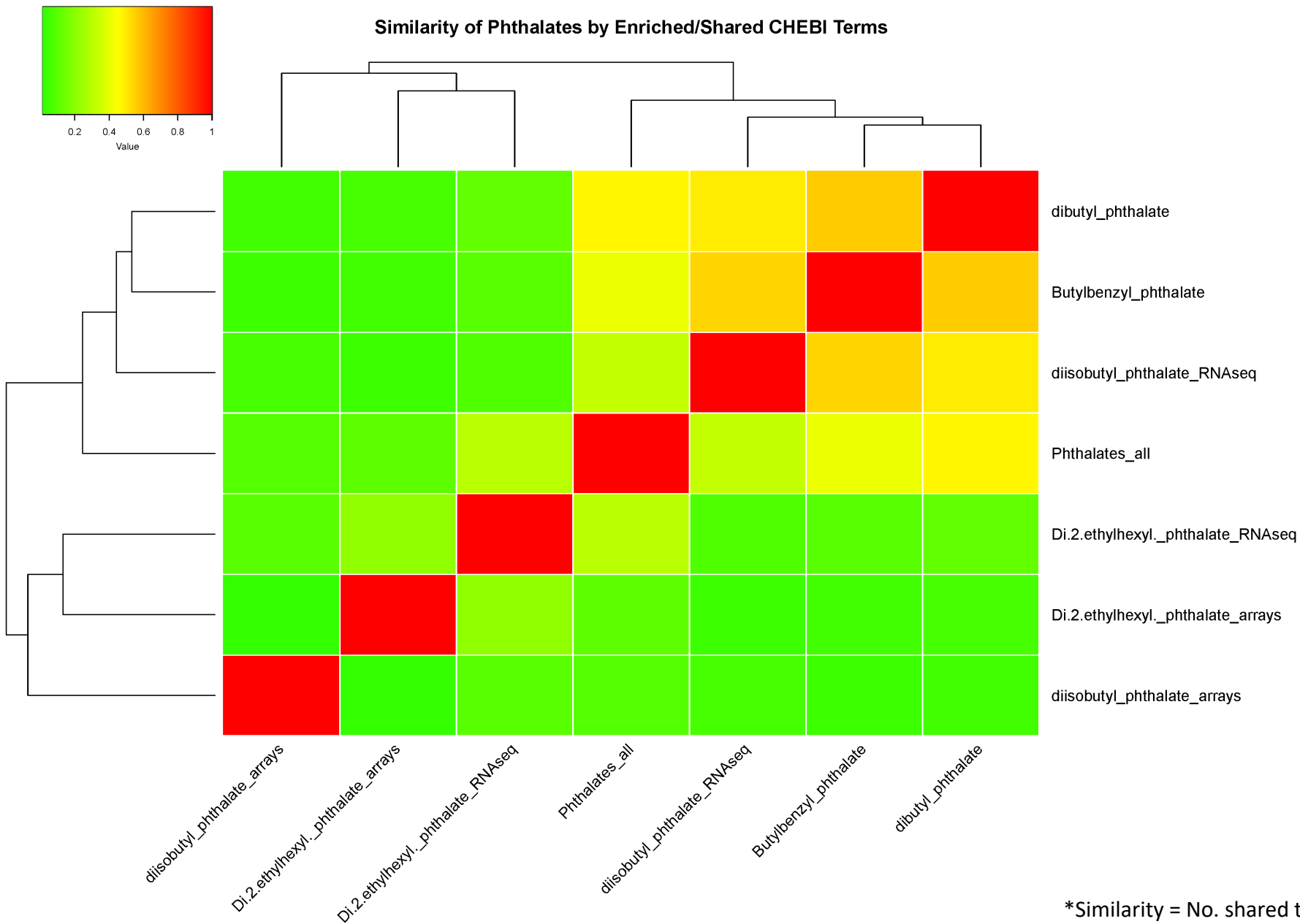
Enriched Ontology Terms and Categories in Mapped Targets

WWTPs



Top ten hits of Cmap on probe-, original gene-, and gene-ROGLs converted from probes





*Similarity = No. shared terms/total No. unique terms

Summary

- **Utilities in chemical exposure diagnosis**
 - Successful recovery from the Cmap library of diverse chemical classes same as signature queries as high-scored mappings
 - Enriched CHEBI terms encompassing the chemical classes of signature queries
- **Utilities in chemical grouping**
 - High-scored chemicals of a DIFFERENT class from a signature query
 - Enriched CHEBI terms more distantly related to a signature query
- **MOA discovery**
 - Enriched ontology terms and categories in mapped targets provide structured hypotheses grounded in most current biology of diverse knowledge domains
 - Well-studied chemicals from the library mapped to signature queries
 - Interrelationships of chemicals in the same class