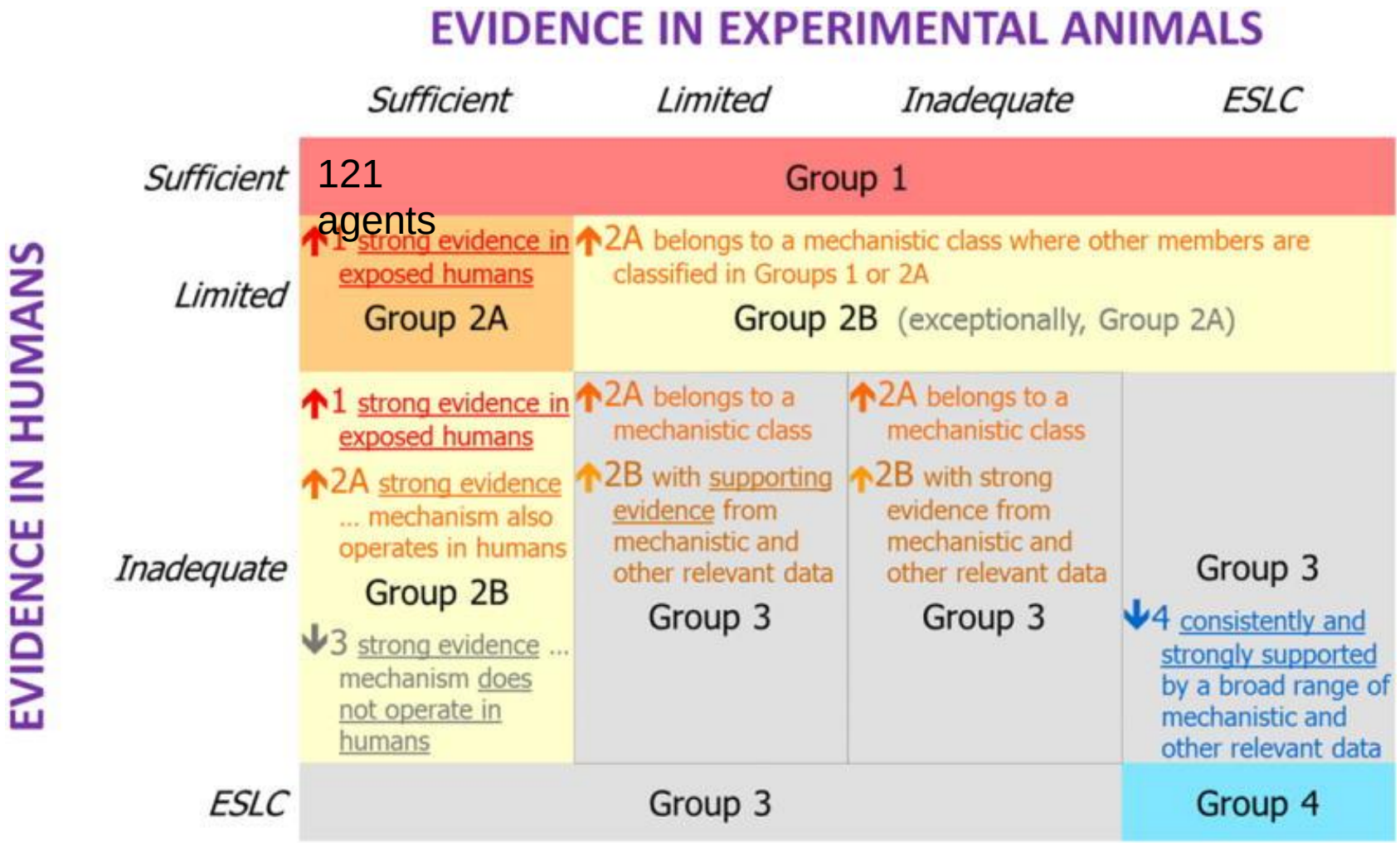
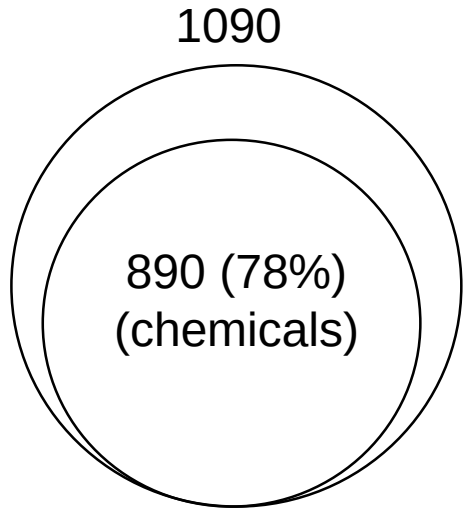
Prioritizing Chemical Agents for Cancer Hazard Identification

Dinesh Kumar Barupal, PhD
Assistant Professor
Integrated Data Science Laboratory for Metabolomics and Exposomics (IDSL.ME)
Department of Environmental Medicine and Public Health (EMPH)
Icahn School of Medicine at Mount Sinai (ISMMS)
New York, USA

Icahn
School of
Medicine at
Mount
Sinai

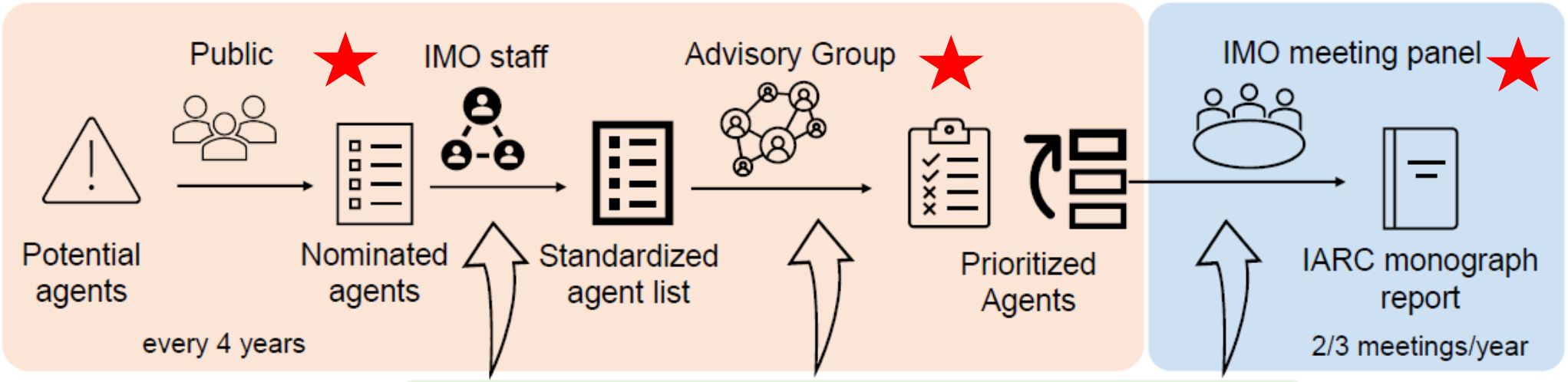
IARC Monographs – the encyclopedia of carcinogenic agents





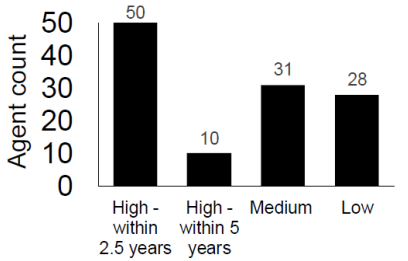
Source : PMC5783793

Cancer hazard identification process



33 millions papers (PubMed)

110 millions chemicals (PubChem)

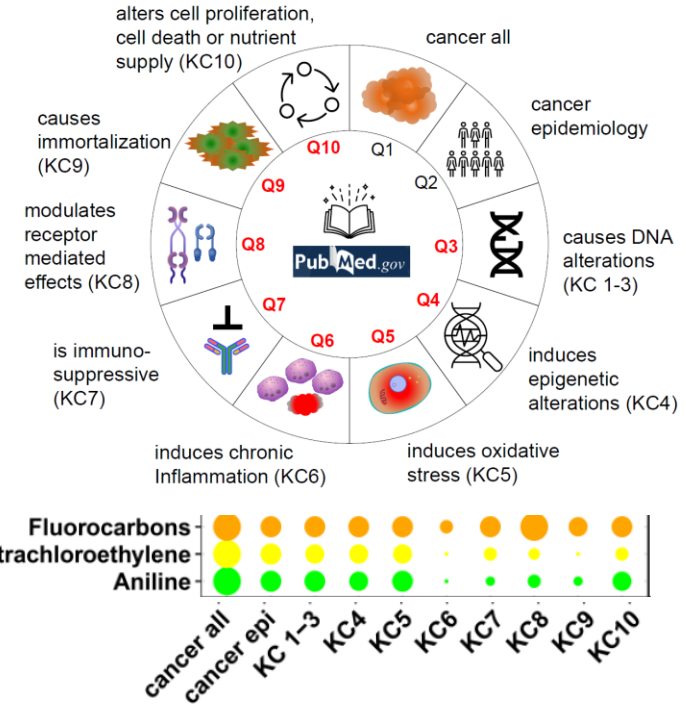


Need to develop new **computational approaches** to utilize **literature and chemical databases** for the cancer hazard identification process.

Which chemical agents should be nominated and evaluated ? → Chemical prioritization

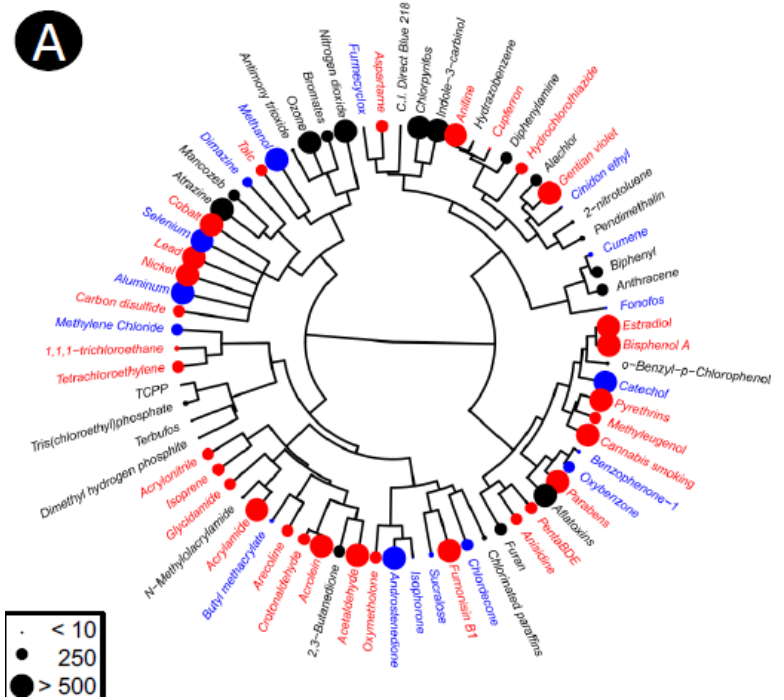
How to prioritize chemical agents ?

Chemical text mining



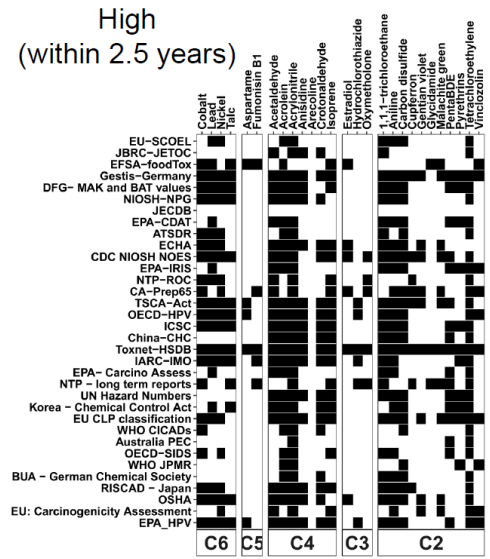
+

Chemoinformatics



+

Database fusion



1. Barupal, Dinesh Kumar, Mary K. Schubauer-Berigan, Michael Korenjak, Jiri Zavadil, and Kathryn Z. Guyton. "Prioritizing cancer hazard assessments for IARC Monographs using an integrated approach of database fusion and text mining." *Environment International* 156 (2021): 106624.
2. Guha, N., Guyton, K. Z., Loomis, D., & Barupal, D. K. (2016). Prioritizing chemicals for risk assessment using chemoinformatics: examples from the IARC monographs on pesticides. *Environmental health perspectives*, *124*(12), 1823-1829.
3. Barupal, Dinesh Kumar, and Oliver Fiehn. "Generating the blood exposome database using a comprehensive text mining and database fusion approach." *Environmental health perspectives* 127.9 (2019): 097008.

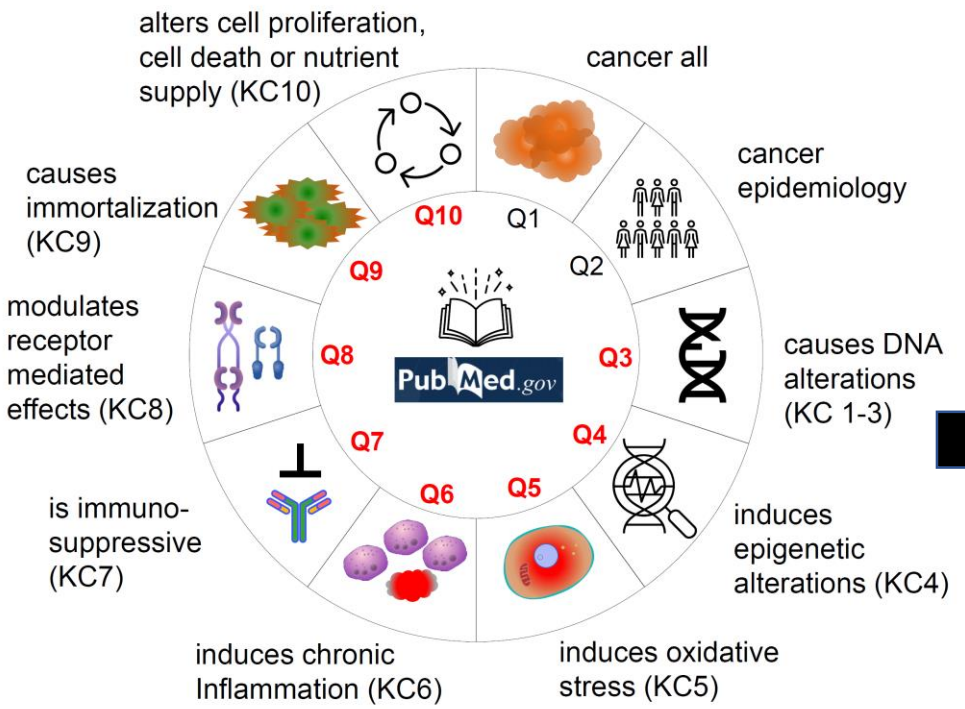
1) Chemical text mining

www.cancer.idsl.me

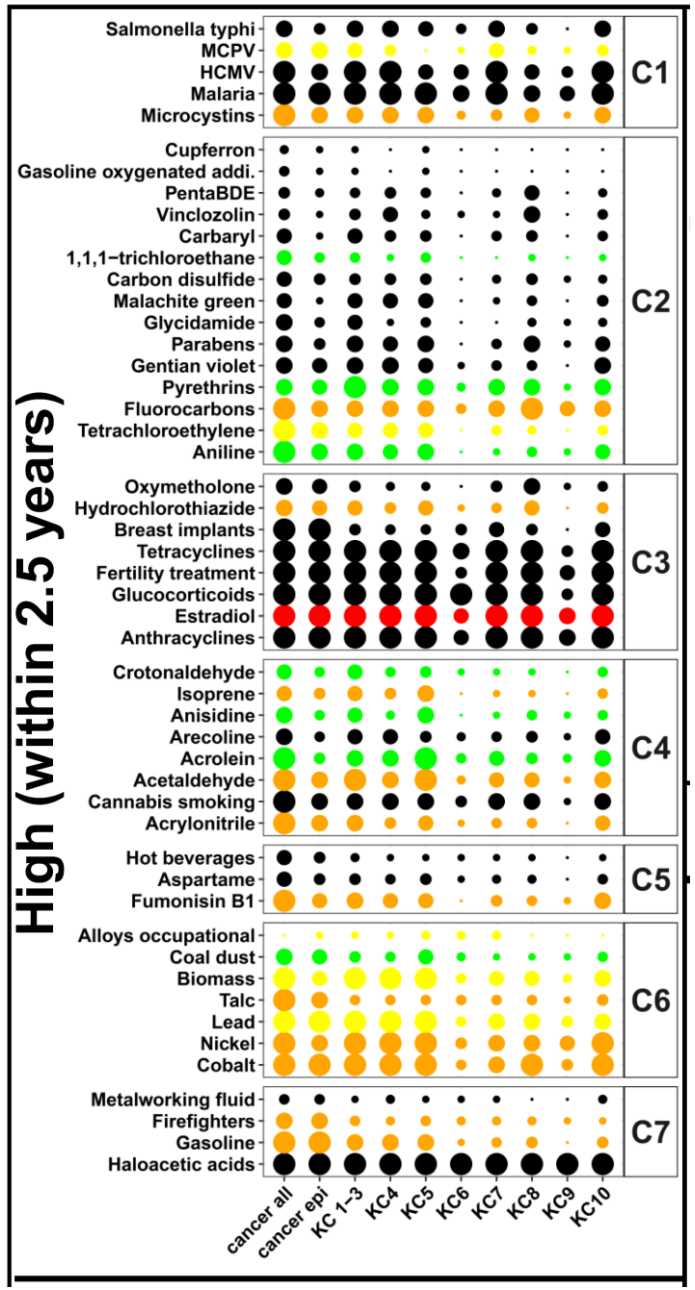
33 million abstracts

Queries

Apache Lucene Index for full-text searches



PubMed filters for key characteristics of a carcinogen (KCs)

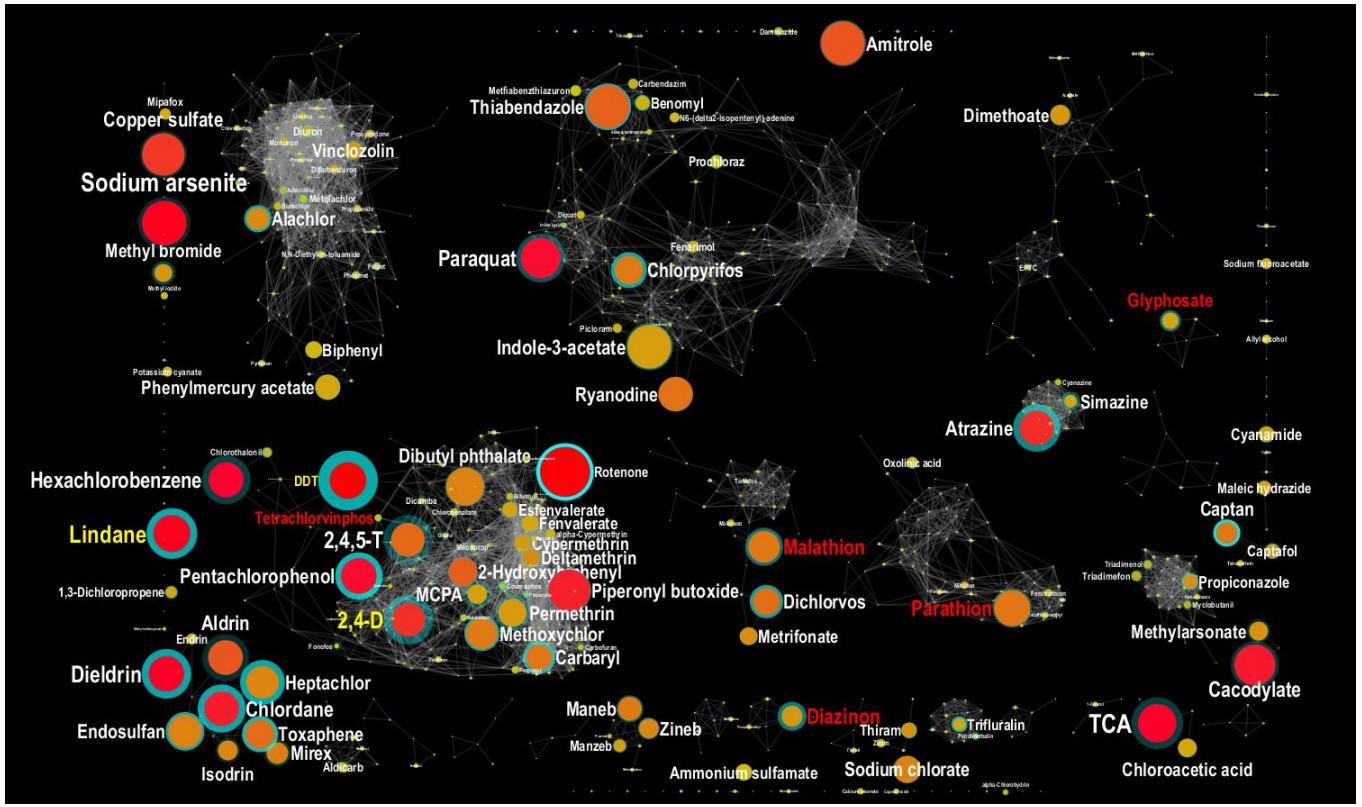


Literature count data

Provided an overview of available relevant literature data as well as highlighted the research gaps.

PMID:33984576

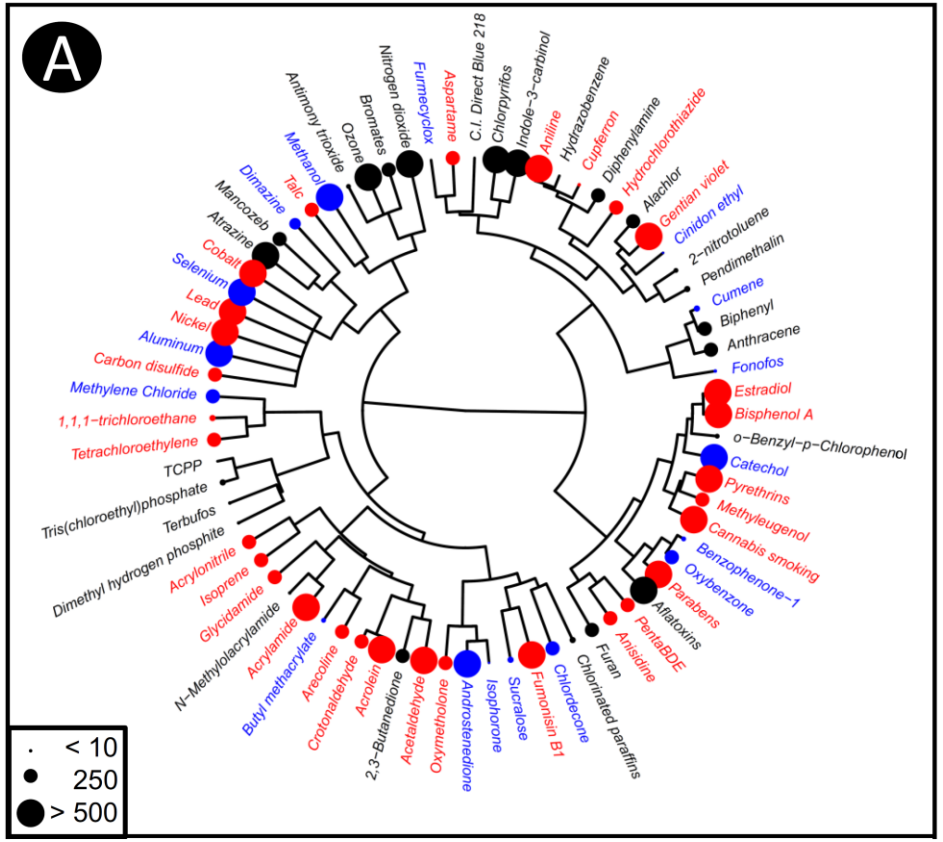
2) Chemoinformatics clustering of chemicals



Tanimoto chemical similarity coefficient

http://pesticide.barupal.org/

Chemical similarity map of ~1000 pesticides



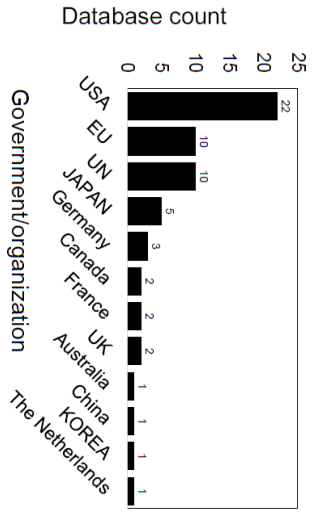
Chemical structure clustering of prioritized agents by the advisory group (2019) for IARC Monographs

Provided a logical grouping of chemically similar agents to be evaluated in a single IMO meeting.

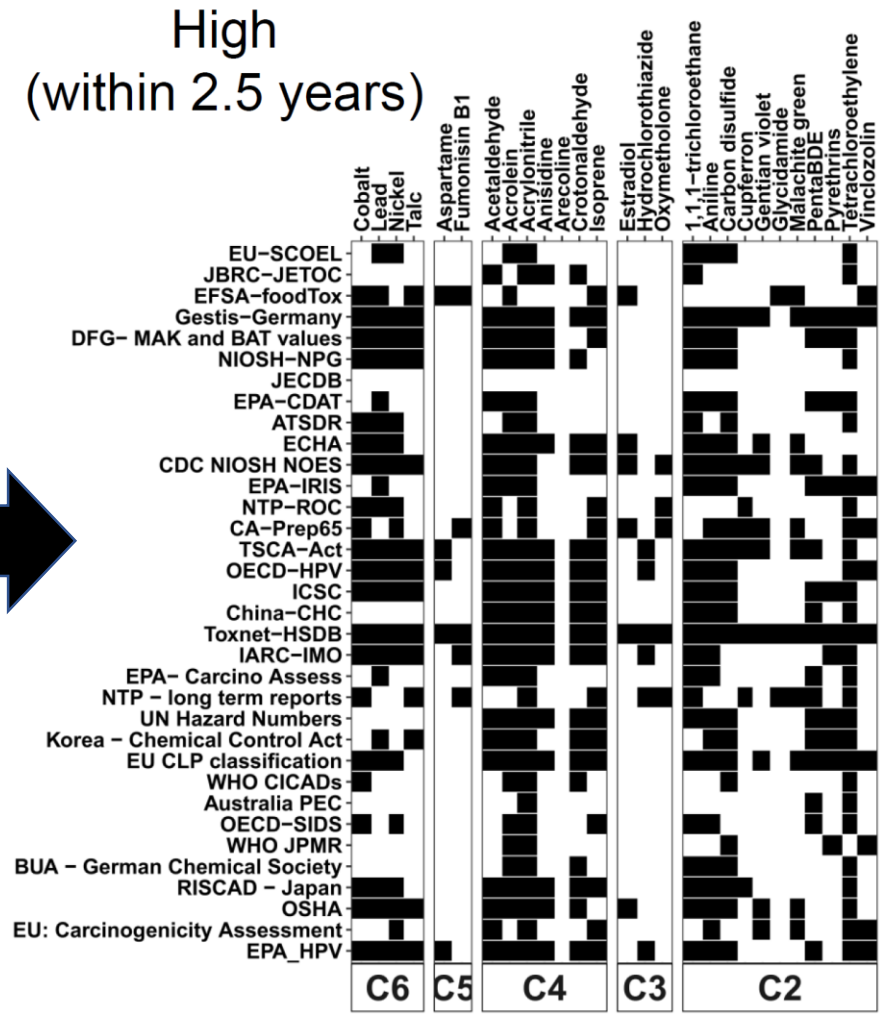
PMIDs: 33984576, 31005580, 27164621, 32593317

www.cancer.idsl.me

3) Database fusion



Database property			Database coverage				
Database	Database size	IMO coverage	All agents	High (2.5 years)	High (5 years)	Medium	Low
IARC-IMO	852	852	33	16	2	9	6
Toxnet-HSDB	5893	662	68	26	3	23	16
TSCA-Act	67270	489	52	20	3	17	12
Gestis-Germany	8014	479	62	20	3	22	17
CDC NIOSH NOES	9318	458	52	19	3	19	11
CA-Prep65	776	407	39	15	3	12	9
ICSC	1704	326	48	16	2	18	12
DFG- MAK and BAT values	1411	323	41	15	3	14	9
EU CLP classification	4007	300	50	18	3	15	14
NTP - long term reports	544	279	34	13	3	13	5
ECHA	14983	271	44	15	3	14	12
UN Hazard Numbers	2381	271	36	12	1	13	10
OECD-HPV	4636	247	43	17	2	16	8
RISCAD - Japan	4482	247	37	14	3	11	9
OSHA	901	247	40	16	2	13	9
China-CHC	2844	246	32	11	1	10	10
EPA_HPV	4308	246	43	18	2	15	8
EPA-CDAT	2626	238	31	10	2	11	8
NTP-ROC	227	198	19	9	2	4	4
EPA-IRIS	555	193	34	11	2	12	9
NIOSH-NPG	648	187	31	13	1	7	10
Korea - Chemical Control Act	2106	187	25	12	2	7	4
EU: Carcinogenicity Assessment	1243	186	24	9	1	7	7
EFSA-foodTox	5501	179	31	11	3	11	6
OECD-SIDS	1569	137	25	9	2	8	6
EPA- Carcino Assess	253	124	19	8	1	4	6
ATSDR	275	111	14	8	0	2	4
BUA - German Chemical Society	358	103	15	7	2	4	2
JBRC-JETOC	412	98	17	6	2	5	4
EU-SCOEL	269	81	14	8	2	2	2
WHO JPMR	325	60	14	5	0	8	1
WHO CICADs	346	49	9	6	0	2	1
Australia PEC	114	19	6	3	1	2	0
JECDB	445	16	1	0	0	0	1



~85,000 chemicals from 34 high priority databases

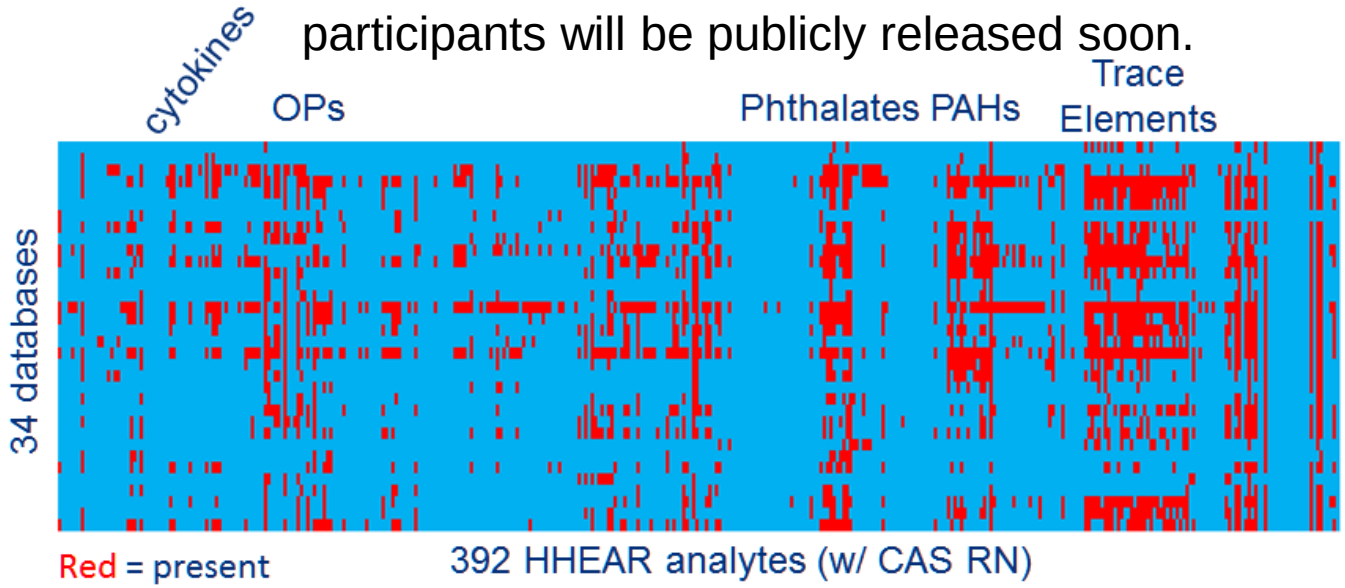
PMID: 33984576

Provided a useful list of candidate exposome chemicals and their coverage in high priority regulatory and toxicology databases.

https://chemcor.idsl.site/exposomechemicals/

Exposome and the cancer hazard identification

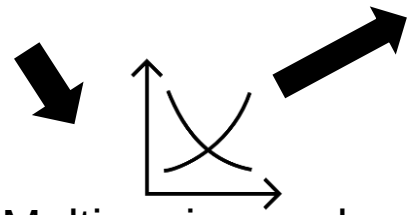
Data from 40 studies and ~28,000 participants will be publicly released soon.



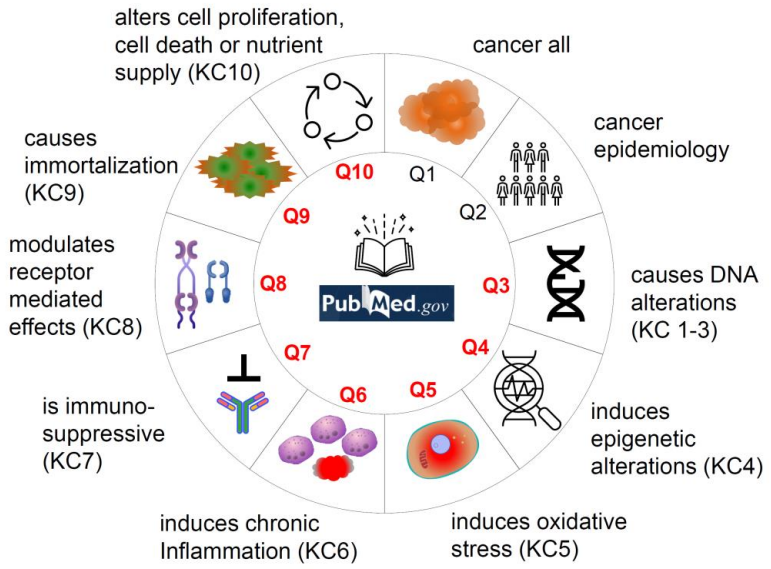
Liquid or gas chromatography
+ high resolution mass spectrometry
+ ipa.idsl.me

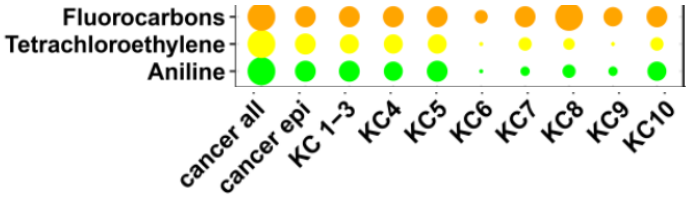
> 50,000 unique chemicals

Biomonitoring surveys using targeted and untargeted chemical analysis assays



Multi-omics analyses





HHEAR Human Health Exposure Analysis Resource

https://hhearprogram.org/

Exposome research have a great potential to fill gaps in the evidence-base for cancer hazard identification. HHEAR public data will be a first of its kind to support such investigations.

Conclusions

1. IARC Monographs uses a systematic and complex process to identify cancer hazards.
2. The process requires engagement among public, scientific experts, advisory board and IMO scientists.
3. The process can be benefitted by a data science approach of chemical text mining, chemoinformatics and database fusion to prioritize agents.
4. A computational approach has been developed to facilitate agent prioritization (cancer.idsl.me).
5. Exposome research has a great potential to fill research gaps in key characteristics of an agent to be classified as a carcinogen.

Acknowledgement

IDSL.ME team:

- Dr. Sadjad Fakouri Baygi (Post-Doc Fellow)

Mt Sinai collaborators :

- Lauren Petrick
- Douglas Walker
- Elena Colicino
- Dania Valvi
- Mary Wolf
- Susan Teitelbaum
- Chris Gennings
- Manish Arora
- Robert Wright
- Shelley Liu

Current and former external collaborators :

- Mary Schubauer-Berigan, Jiri Zavadil Michael Korenjak : **IARC**
- Kate Guyton, **NASEM**
- Neela Guha, **CA-OEHHA**
- Oliver Fiehn, Tobias Kind, Kent Pinkerton : **UC Davis**
- Steve Rappaport : **UC Berkeley**
- Ian Lipkin, Andrea Baccarelli : **Columbia University**
- Rima Kaddurah-daouk : **Duke University**
- Yashwant Kumar : **THSTI India**

Special thanks to NIH

NIH Common Fund Metabolomics Program

HHEAR Human Health Exposure Analysis Resource