



Prioritizing Chemicals for Risk Assessment Using Chemoinformatics

International Agency for Research on Cancer
Lyon, France

Neela Guha, PhD

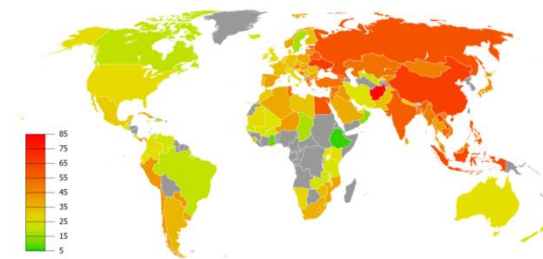
The IARC Monographs

- ✓ Identify carcinogens – the first step towards cancer prevention
 - Environmental, lifestyle, occupational, physical, chemical, biologic agents
 - Systematic reviews of epidemiological evidence, animal cancer bioassays, carcinogenic mechanisms
 - ~ 1000 agents evaluated since 1971
- ✓ Impact: Public health action to limit exposure to workers & general public
- ✓ Challenge: **Maximize public health benefit by prioritizing agents for IARC evaluation**

How are agents selected for IARC Monograph Review?

- ✓ Agents nominated by the public and scientists are prioritized by:
 - Public health importance worldwide
 - Global human exposure/Usage data
 - Evidence of carcinogenicity
 - Potential to update a previous IARC evaluation (new data)

- ✓ Chemoinformatics was used to select pesticides for IARC assessments (March & June 2015, Oct 2016)
 - Chemical structure + Bioinformatics



Pesticide Selection Using Chemoinformatics

Diverse "pesticides" recommended for IARC evaluation

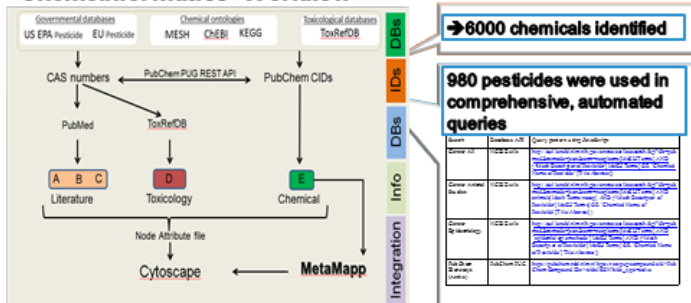
Data assembly, integration and visualization
(980 pesticides)

New or updated classifications

Key Highlights:

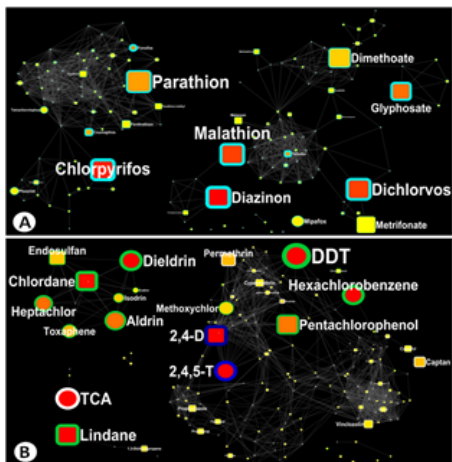
- Assembly of a comprehensive list of pesticides
- Data visualization for diverse pesticides by class
- Automated text mining of public databases
- Objective prioritization for IARC evaluation in Vol 112, 113, 117

Chemoinformatics Workflow



- A. PubMed cancer all
- B. PubMed cancer epidemiology
- C. PubMed animal cancer bioassays
- D. ToxRefDB carcinogenicity
- E. Chemical similarity scores

OP (A) and OC (B) Pesticides



- **Node:** Pesticide active
- **Size:** # of cancer pubs
- **Shape:** square = in ToxRefDB
- **Link:** if Tanimoto > 0.6 (60% chemical similarity)
- **Border width and node fill color:** # of epidemiology pubs (red > orange > yellow)
- **Border:** chemical class (green, OC; lt blue, OP; dk blue, phenoxy; white, other)

2015 IARC Monograph Pesticide Classifications

Name	Rank (within class)	Usage notes	PubMed Human cancer	PubMed Mechanisms	Prior IARC (year)	2015 evaluation (Human, Animal, Mechanisms)	Current IARC (2015)
Organophosphate pesticides							
Parathion	1	Restricted	8	578	3 (1987)	H: Inad. A: Sufficient	Group 2B
Malathion	2	High	12	370	3 (1987)	H: Limited A: Sufficient M: Strong	Group 2A
Diazinon	5	High	16	215	-	H: Limited A: Limited M: Strong	Group 2A
Glyphosate	7	High	9	204	-	H: Limited A: Sufficient M: Strong	Group 2A
TCVP	13	Active	1	40	3 (1987)	H: Inad. A: Sufficient	Group 2B
Organochlorine and chlorophenoxy pesticides							
DDT	1	Restricted POP	190	953	2B (1991)	H: Limited A: Sufficient M: Strong	Group 2A
Lindane	2	Active POP	51	545	2B (1987)	H: Sufficient A: Sufficient M: Strong	Group 1
2,4-D	1	High	84	420	2B (1991)	H: Inad. A: Limited	Group 2B

International Agency for Research on Cancer

<http://pesticide.barupal.org/>

Guha et al (2016). *Environ Health Perspect.* PMID: 27164621
 Guyton KZ et al (2015). *Lancet Oncol.* PMID: 2580172
 Loomis D et al (2015). *Lancet Oncol.* PMID: 26111929

Pesticides ranked for risk assessment, regulation, research

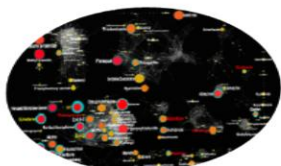
pesticide.barupal.org/dataTable.html

Pesticide name and Classification			Database Identifiers		Chemical Properties			Literature data			Toxicological data	
S.No.	Name	KEGG Classification	PubChemCID	MESH ID	MolecularFormula	Counts of chlorine atoms	Contain Phosphorus atom	Cancer All	Cancer EPI_Counts	Cancer Animal Studies	PubChem BioAssay (Active)	ToxRef DB Annotation
769	TCA	Others	6421	68014238	C2HCl3O2	3	No	731	146	286	4	N/A
798	DDT	Organochlorine	3036	68003634	C14H9Cl5	5	No	494	190	237	44	N/A
756	Rotenone	Others	6758	68012402	C23H22O6	N/A	No	420	15	228	234	No
371	Quinoxaline	Others	7045	N/A	C8H6N2	N/A	No	233	66	142	8	N/A
253	Sodium arsenite	Inorganic	443495	67017947	AsNaO2	N/A	No	239	15	128	0	N/A
707	Diethylhexyl phthalate (DEHP)	N/A	8343	68004051	C24H38O4	N/A	No	136	21	122	9	YES
383	Lindane	Organochlorine	727	68001556	C6H6Cl6	6	No	189	51	119	41	YES
706	Dieldrin	Organochlorine	969491	68004026	C12H8Cl6O	6	No	151	57	90	34	N/A
710	Diphenylamine	N/A	11487	68004159	C12H11N	N/A	No	147	20	86	4	No
			7855	68000181	C3H3N	N/A	No	129	68	65	5	N/A
							Yes	101	5	59	50	N/A
										58	0	N/A
										57	5	N/A
										47	74	N/A
										47	11	N/A
										44	4	YES
										43	37	YES
										42	21	YES
										39	23	No
										38	11	No
										29	0	N/A
										29	9	N/A
										27	16	YES
										54		YES

CarcinoMapper: Pesticides

Pesticides : Chemical Similarity Networks and Data Table

A list of 980 pesticides was queried against publicly available databases to retrieve published data and literature on chemical's carcinogenic potential in animal experiments and epidemiological settings. The complexity of the data was visualized using chemical similarity network graphs which allowed extension of existing classification schemes, such as the one obtained from KEGG, and provided an efficient way to visualize the retrieved rich information. To highlight the application for selecting candidates for IARC Monographs evaluation, pesticides containing chlorine and phosphorus atoms, including the two of the most commonly used pesticide classes of organochlorines and organophosphates, were visualized as focused network graphs. High-ranked pesticides were then further passed through additional pesticide specific filter to create a final list of selected agents. Additionally, an interactive data table is provided at the bottom of graphs with hyperlinks that redirect queries to the PubMed database to retrieve the most recent publications.



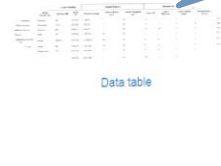
All Pesticides



Phosphorus containing



Chlorine Containing



Data table

Data Table

All Pesticides

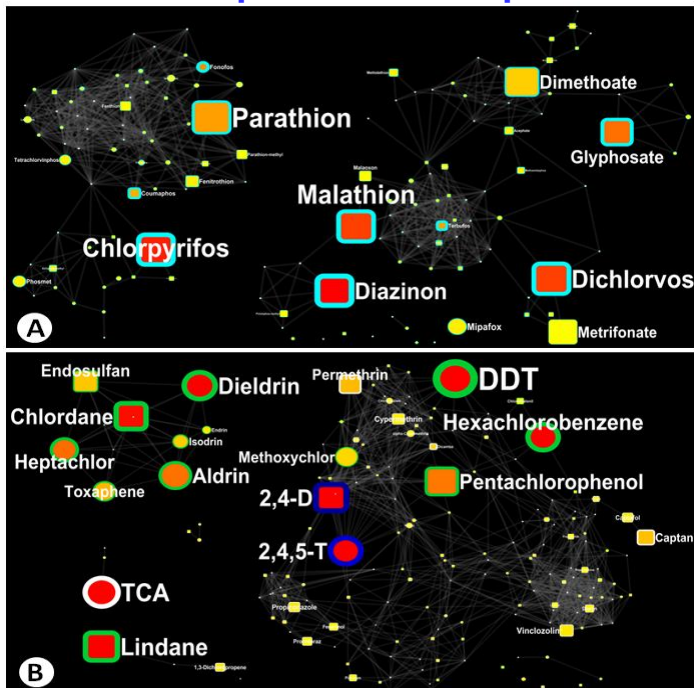
Phosphorus pesticides

Chlorine Pesticides

<http://pesticide.barupal.org/>

Pesticides selected translated to new IARC classifications

Network maps of OP & OC pesticides



Node: Pesticide active
Size: # of cancer pubs
Shape: square = in ToxRefDB
Link: if Tanimoto > 0.6 (60% chemical similarity)
Border width and node fill color: # of epidemiology pubs (red > orange > yellow)
Border: chemical class (green, OC; lt blue, OP; dk blue, phenoxy; white, other)

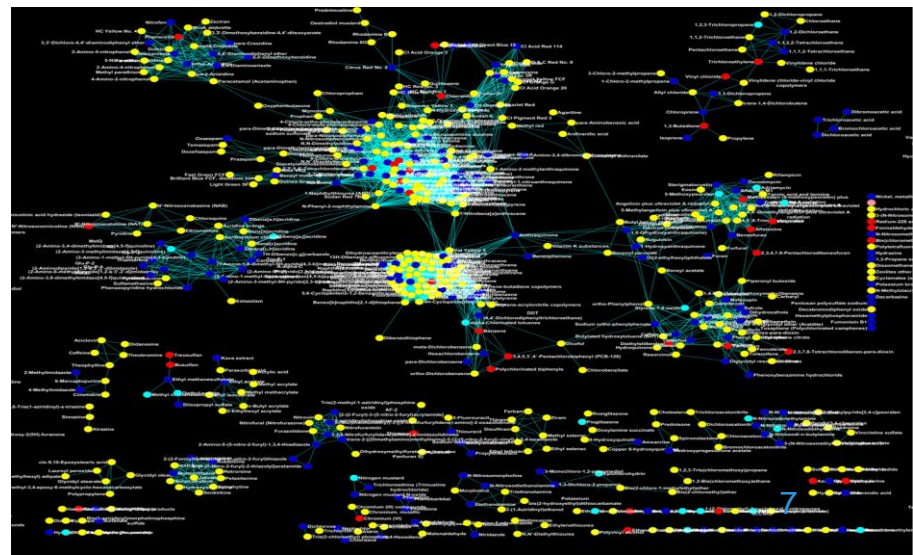
2015 IARC Monograph Pesticide Classifications

Name	Rank (within class)	Usage notes	PubMed Human cancer	PubMed Mechanisms	Prior IARC (year)	2015 evaluation (Human, Animal, Mechanisms)	Current IARC (2015)
Organophosphate pesticides							
Parathion	1	Restricted	6	578	3 (1987)	H: Inad. A: Sufficient	Group 2B
Malathion	2	High	12	370	3 (1987)	H: Limited A: Sufficient M: Strong	Group 2A
Diazinon	5	High	16	215	-	H: Limited A: Limited M: Strong	Group 2A
Glyphosate	7	High	9	204	-	H: Limited A: Sufficient M: Strong	Group 2A
TCVP	13	Active	1	40	3 (1987)	H: Inad. A: Sufficient	Group 2B
Organochlorine and chlorophenoxy pesticides							
DDT	1	Restricted POP	190	953	2B (1991)	H: Limited A: Sufficient M: Strong	Group 2A
Lindane	2	Active POP	51	545	2B (1987)	H: Sufficient A: Sufficient M: Strong	Group 1
2,4-D	1	High	84	420	2B (1991)	H: Inad. A: Limited	Group 2B

Next Steps – Extending the approach

- ✓ Extend the approach to occupational carcinogens, environmental pollutants, drugs, nutritional compounds
- ✓ Predict carcinogenicity of unclassified chemicals based on chemical structural similarity to classified carcinogens
- ✓ Overlay data on human exposure
- ✓ Prioritize pesticides to be studied in relation to cancer trends worldwide (collaboration with CSU)

Network Map of IARC Classifications



Conclusion

- ✓ 2015 evaluations gave new or higher IARC classifications for pesticides with worldwide exposures
 - Public health impact: worldwide regulation of glyphosate
- ✓ Diversity of chemical structures and data availability were visualized efficiently to prioritize pesticides for hazard assessment
- ✓ Data mining systematized the manual process for agent selection
 - Enhanced the traditional process
 - Automated literature mining of public databases (bioinformatics)
 - Displayed data systematically by volume, complexity, chemical similarity
 - Prioritized pesticides for IARC evaluations (March & June 2015, Oct 2016)

Acknowledgements

Dinesh Kumar Barupal (UC Davis)

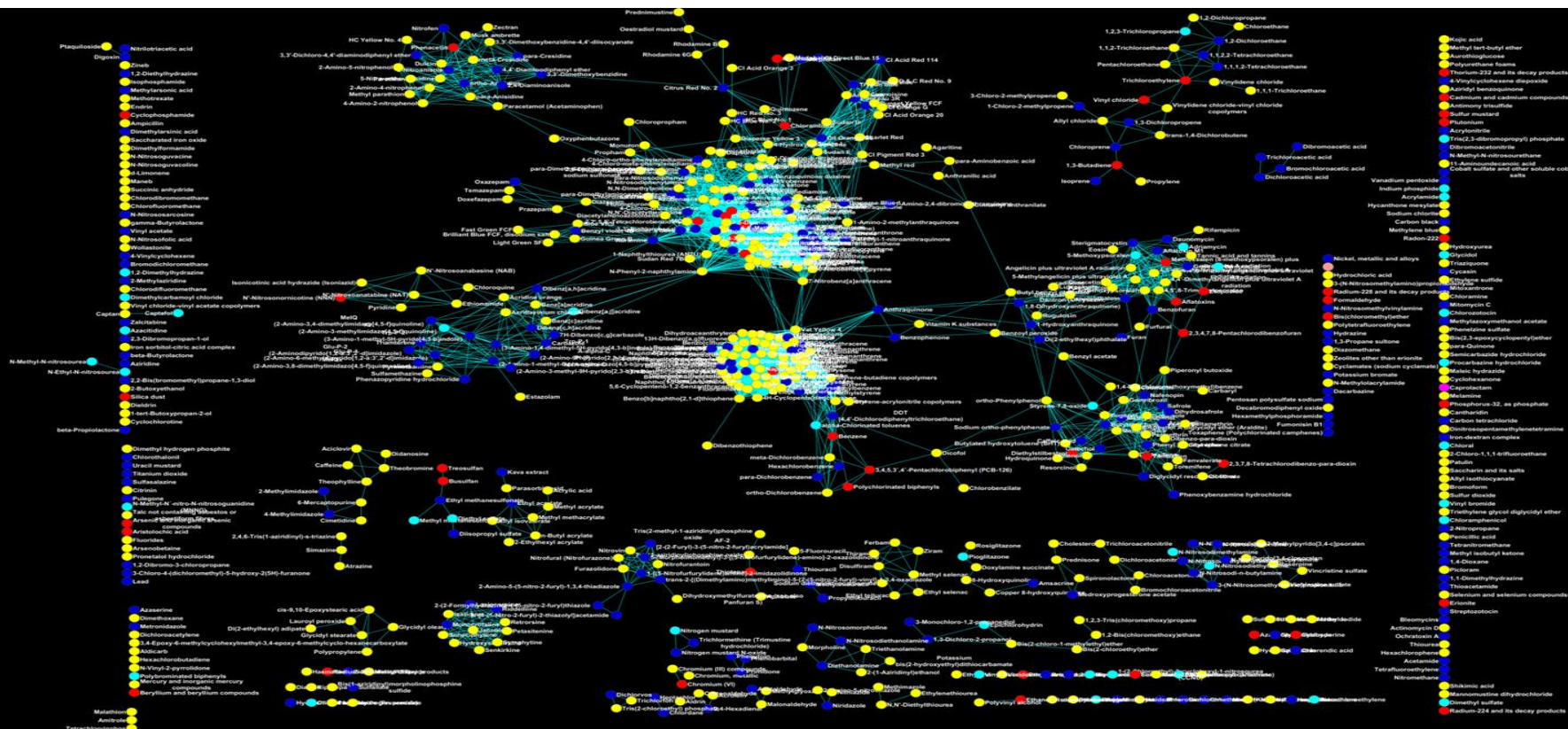
Kate Guyton (IARC)

Dana Loomis (IARC)

The IARC Monographs Programme gratefully acknowledges support from the US National Cancer Institute (U01CA033193), the National Institute of Environmental Health Sciences (LAC/IMO/2015/01) and the European Commission (VS/2015/0156).

Competing financial interests: NONE

Chemical Similarity Map of IARC Monograph Evaluations



What we did: Pesticides

- ✓ Compiled master list of pesticide compounds
 - Databases (DB): KEGG, ToxRef (Name, CASNr)
- ✓ Chemical structure obtained for DB linkage
 - PubChem: CASNr > CID
- ✓ Text mined public DB to assess scope of published literature (CID+CAS+search terms)
 - Cancer epidemiology (PubMed)
 - Animal Bioassays (PubMed)
 - PubChem BioAssay (ToxRefDB)
 - Chemical Classes (KEGG DB)
- ✓ Network Graphs created to visualize amount of literature, grouped by chemical similarity

Querying Databases: Pesticides & Cancer

Search	Database Used/ Web API	Search Terms/Query parameters
Cancer and Pesticides Animal Bioassays	NCBI Eutils	http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&rettype=count&retmode=json&term=cas numbers AND top three synonyms AND pesticide* AND cancer AND (animals[All Fields]) AND hasabstract[text] AND english[language]
Cancer and Pesticides Epidemiology	NCBI Eutils	http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&rettype=count&retmode=json&term=cas numbers AND top three synonyms AND pesticide* AND cancer AND risk*[All Fields] AND humans[All Fields] AND ("CI" OR "confidence interval" OR ratio*) AND hasabstract[text] AND english[language]
Active BioAssays counts	PubChem PUG	pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/XXX/aids/JSON?aids_type=active (XXX is PubChem Compounds identifier for the compound)
CAS to CID conversion	PubChem PUG	pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/YYYY/cids/JSON (YYY is a CAS number)
NCBI BioSystems	NCBI Eutils	http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=biosystems&rettype=count&retmode=json&term=XXX[PID] AND "conserved biosystem" [Filter] (XXX is a PubChem compound identifier)