Literature Mining Methods for Toxicology and Construction of Adverse Outcome Pathways

Nancy C. Baker, PhD –Lockheed Martin (contractor)
Thomas B. Knudsen, PhD -National Center for Computational Toxicology
US EPA, Research Triangle Park, NC 27711
baker.nancy@epa.gov
knudsen.thomas@epa.gov

1st Workshop/Scientific Expert Group meeting on the OECD Retinoid Project
April 26, 2016 –Brussels

Text mining – why do it?

Unstructured Text

Structured data

Chemical	Gene	Amount
Corticosterone	NR3C1	416
Tretinoin	RARA	397
Mifepristone	NR3C1	260
Fenofibrate	PPARA	189
Estradiol	EGFR	119
Tretinoin	RARG	111
Tamoxifen	EGFR	96
Tretinoin	EGFR	86
Triamcinolone	NR3C1	83
Lovastatin	LDLR	65
Estradiol	NR3C1	60
Estradiol	MMP9	54
Genistein	EGFR	50
Phenobarbital	NR1I3	48

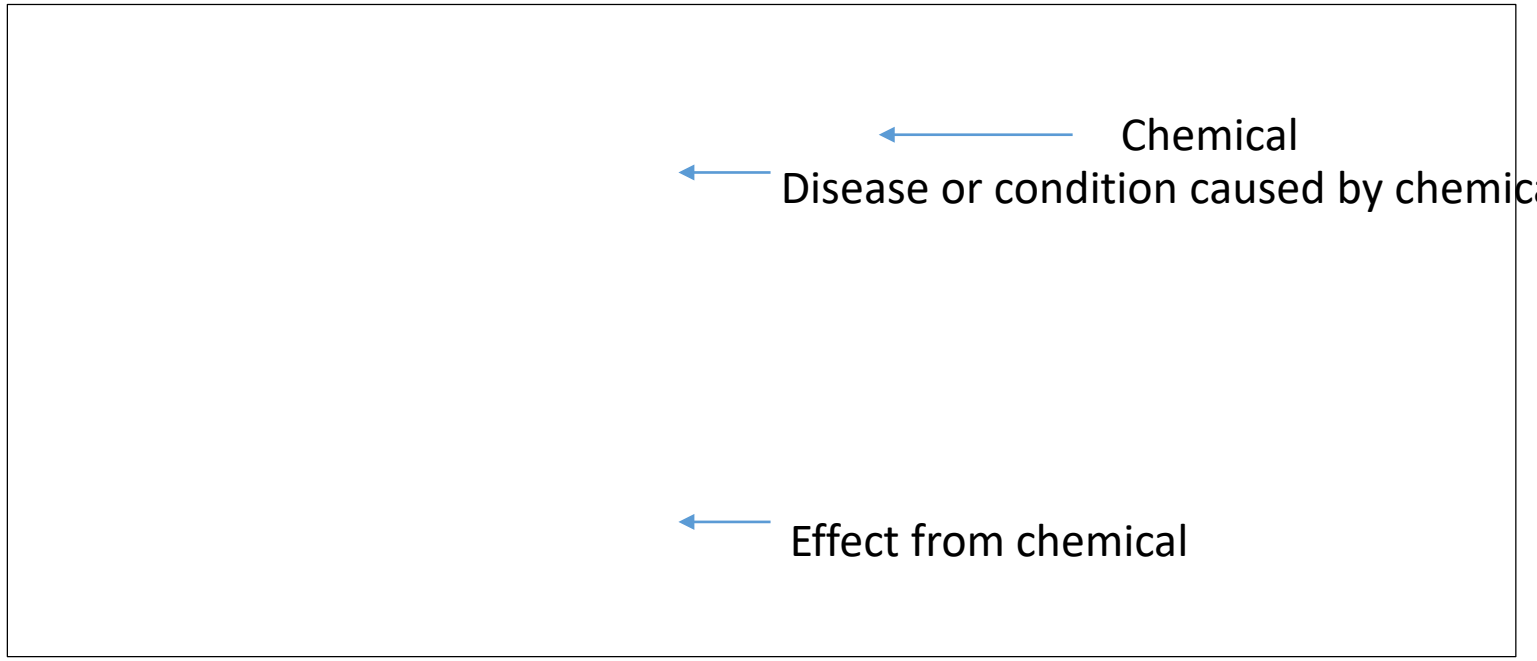
- Integrate
- Analyze
- Cluster
- Model
- Infer
- Visualize
- Discover
- **Deal with Error / noise / bias**

Anatomy of a PubMed record

MeSH Indexing annotations

MeSH

- **Me**dical **S**ubject **H**eadings
- Not designed to be used as data … but that's what many people do.
- But like data they are a controlled vocabulary



Format: Heading / subheading

https://www.nlm.nih.gov/mesh/MBrowser.html

Indexing terms → data

PubMed ID	MeSH heading	Qualifier / subheading	Majr
15907662	Chlorpyrifos	Administration & dosage	N
15907662	Chlorpyrifos	Toxicity	Y
15907662	Cleft Palate	Chemically induced	N
15907662	Fetal resorption	Chemically induced	N
15907662	Hernia, abdominal	Chemically induced	N
15907662	Neural tube defects	Chemically induced	N
15907662	Organogenesis	Drug effects	Y
15907662	Polydactyly	Chemically induced	N
15907662	Spine	Drug effects	N
15907662	Mice		

High-throughput text-mining: a few readouts per article, but it adds up …

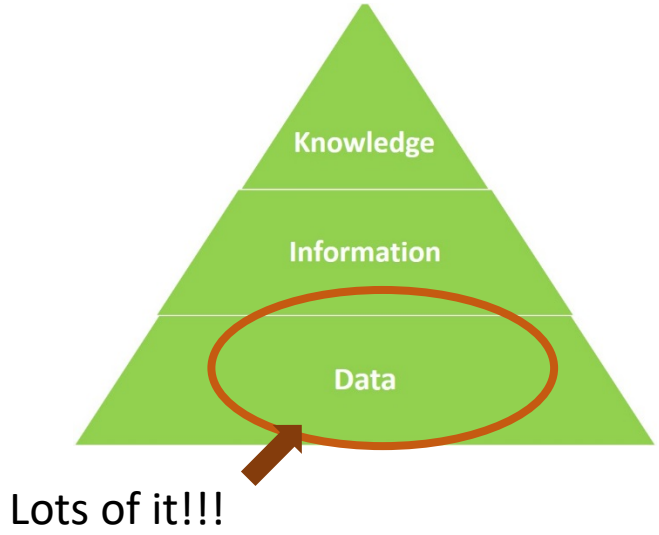
Chlorpyrifos – 2174 articles

Disease List		
Chemical	**Disease/condition**	**PMID Count**
Chlorpyrifos	Behavior, Animal	84
Chlorpyrifos	Motor Activity	60
Chlorpyrifos	Body Weight	42
Chlorpyrifos	Maze Learning	21
Chlorpyrifos	Cognition	14
Chlorpyrifos	Nervous System Diseases	13
Chlorpyrifos	Prenatal Exposure Delayed Effects	12
Chlorpyrifos	Memory	11
Chlorpyrifos	Reaction Time	10
Chlorpyrifos	Peripheral Nervous System Diseases	10
Chlorpyrifos	Psychomotor Performance	8
Chlorpyrifos	Occupational Diseases	7
Chlorpyrifos	Cognition Disorders	7
Chlorpyrifos	Agricultural Workers' Diseases	7
Chlorpyrifos	Hypothermia	7
Chlorpyrifos	Learning	7
Chlorpyrifos	Cat Diseases	6
Chlorpyrifos	Weight Gain	6
Chlorpyrifos	Reflex, Startle	5
Chlorpyrifos	Space Perception	5
Chlorpyrifos	Memory, Short-Term	5
Chlorpyrifos	Cattle Diseases	5
Chlorpyrifos	Fever	5
Chlorpyrifos	Feeding Behavior	5
Chlorpyrifos	Attention	4
Chlorpyrifos	Avoidance Learning	4
Chlorpyrifos	Memory Disorders	4
Chlorpyrifos	Impulsive Behavior	4
Chlorpyrifos	Visual Perception	4

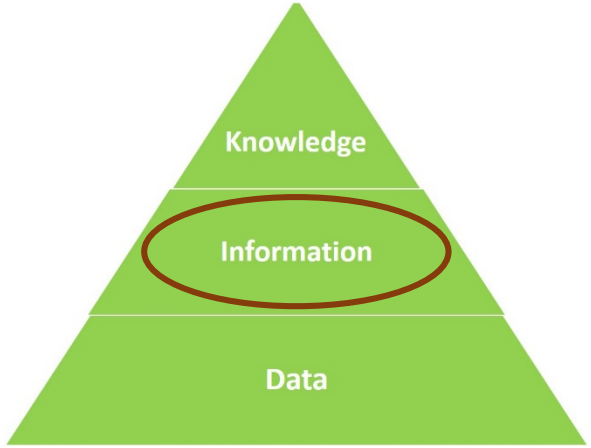
Protein / gene List	
Protein / gene	**PMID Count**
Acetylcholinesterase	488
Cholinesterases	465
Receptors, Muscarinic	57
Esterases	47
Aryldialkylphosphatase	47
Cytochrome P-450 Enzyme System	42
Glutathione Transferase	37
Carboxylesterase	35
Carboxylic Ester Hydrolases	35
Butyrylcholinesterase	33
Superoxide Dismutase	30
Catalase	29
Glutathione	24
Enzymes	23
Receptors, Nicotinic	19
Recombinant Proteins	19
Isoenzymes	17
Glutathione Peroxidase	16
Nerve Tissue Proteins	16
Choline O-Acetyltransferase	15
PON1 protein, human	15
Adenylate Cyclase	15
Enzymes, Immobilized	13
neurotoxic esterase	13
Oxidoreductases	13
Cytochrome P-450 CYP3A	12
L-Lactate Dehydrogenase	12
Glycine	12
glyphosate	12
Membrane Transport Proteins	10
CYP3A4 protein, human	10
Proteins	10
Receptors, Cholinergic	10

The numbers (approximately)

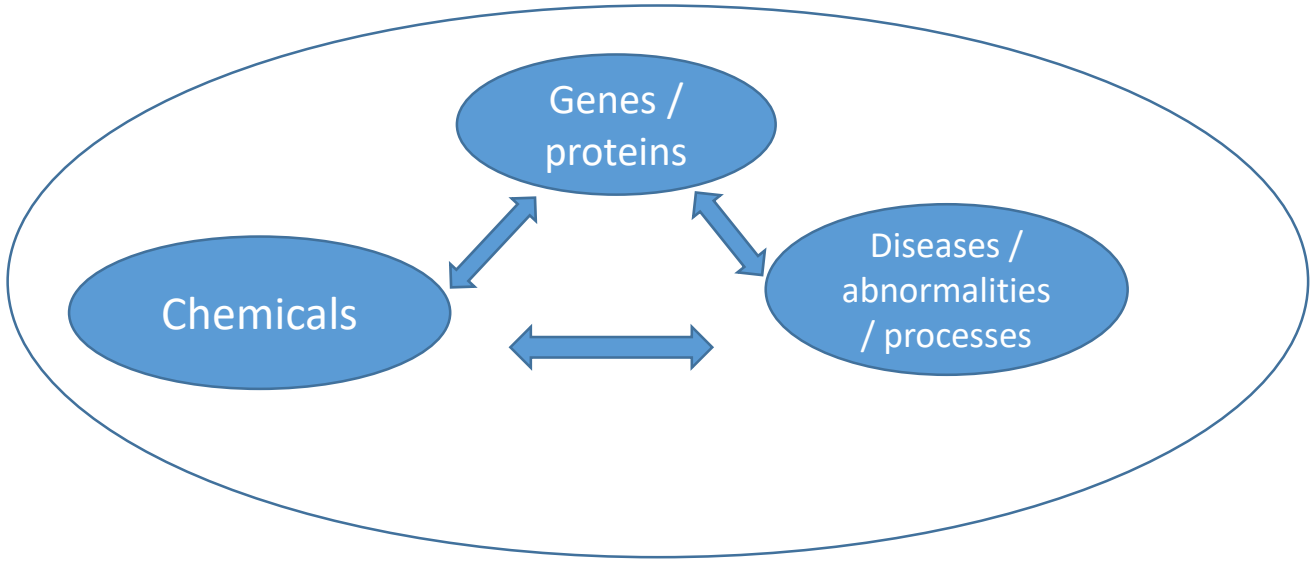
- 24 million articles in PubMed
- 12 million articles have chemical annotations
- 190 million MeSH annotations
- 12 million annotations are my additions
- Growth
  - 1 million annotations / month
  - More next year



So we have lots of data … what can we do with it?



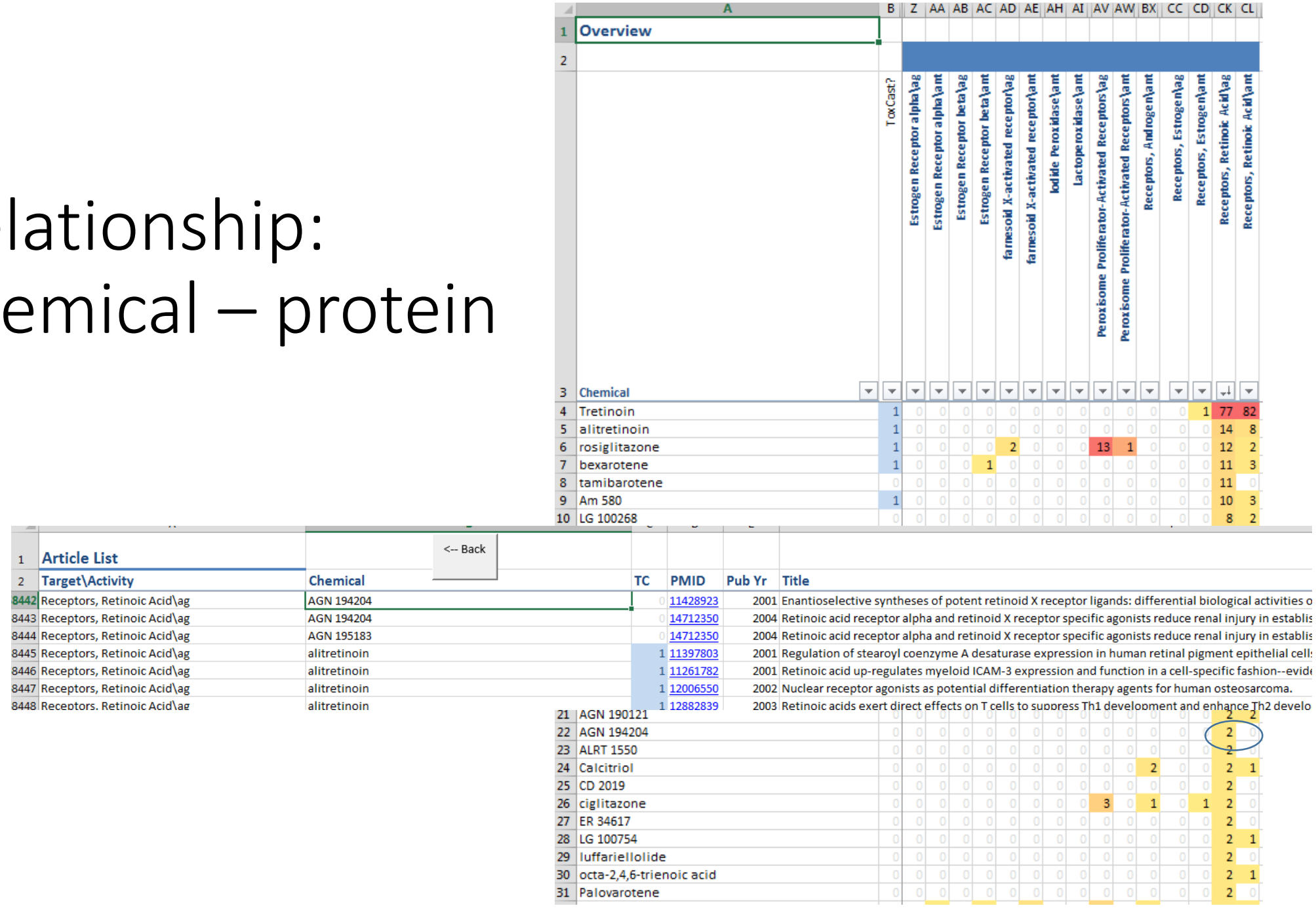
- Delivery forms
  - Dashboard – for basic information delivery
  - E-libraries – for very specific computationally intensive problems



And in what context:

- Species
- Life stage

Relationship: chemical – protein



Overview

	Chemical	ToxCast?	Estrogen Receptor alpha\ag	Estrogen Receptor alpha\ant	Estrogen Receptor beta\ag	Estrogen Receptor beta\ant	farnesoid X-activated receptor\ag	farnesoid X-activated receptor\ant	Iodide Peroxidase\ant	Lactoperoxidase\ant	Peroxisome Proliferator-Activated Receptors\ag	Peroxisome Proliferator-Activated Receptors\ant	Receptors, Androgen\ant	Receptors, Estrogen\ag	Receptors, Estrogen\ant	Receptors, Retinoic Acid\ag	Receptors, Retinoic Acid\ant
4	Tretinoin	1	0	0	0	0	0	0	0	0	0	0	0	0	1	77	82
5	alitretinoin	1	0	0	0	0	0	0	0	0	0	0	0	0	0	14	8
6	rosiglitazone	1	0	0	0	0	2	0	0	0	13	1	0	0	0	12	2
7	bexarotene	1	0	0	0	1	0	0	0	0	0	0	0	0	0	11	3
8	tamibarotene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0
9	Am 580	1	0	0	0	0	0	0	0	0	0	0	0	0	0	10	3
10	LG 100268	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	2
21	AGN 190121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
22	AGN 194204	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
23	ALRT 1550	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
24	Calcitriol	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	1
25	CD 2019	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
26	ciglitazone	0	0	0	0	0	0	0	0	0	3	0	1	0	1	2	0
27	ER 34617	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
28	LG 100754	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
29	luffariellolide	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
30	octa-2,4,6-trienoic acid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
31	Palovarotene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0

Article List

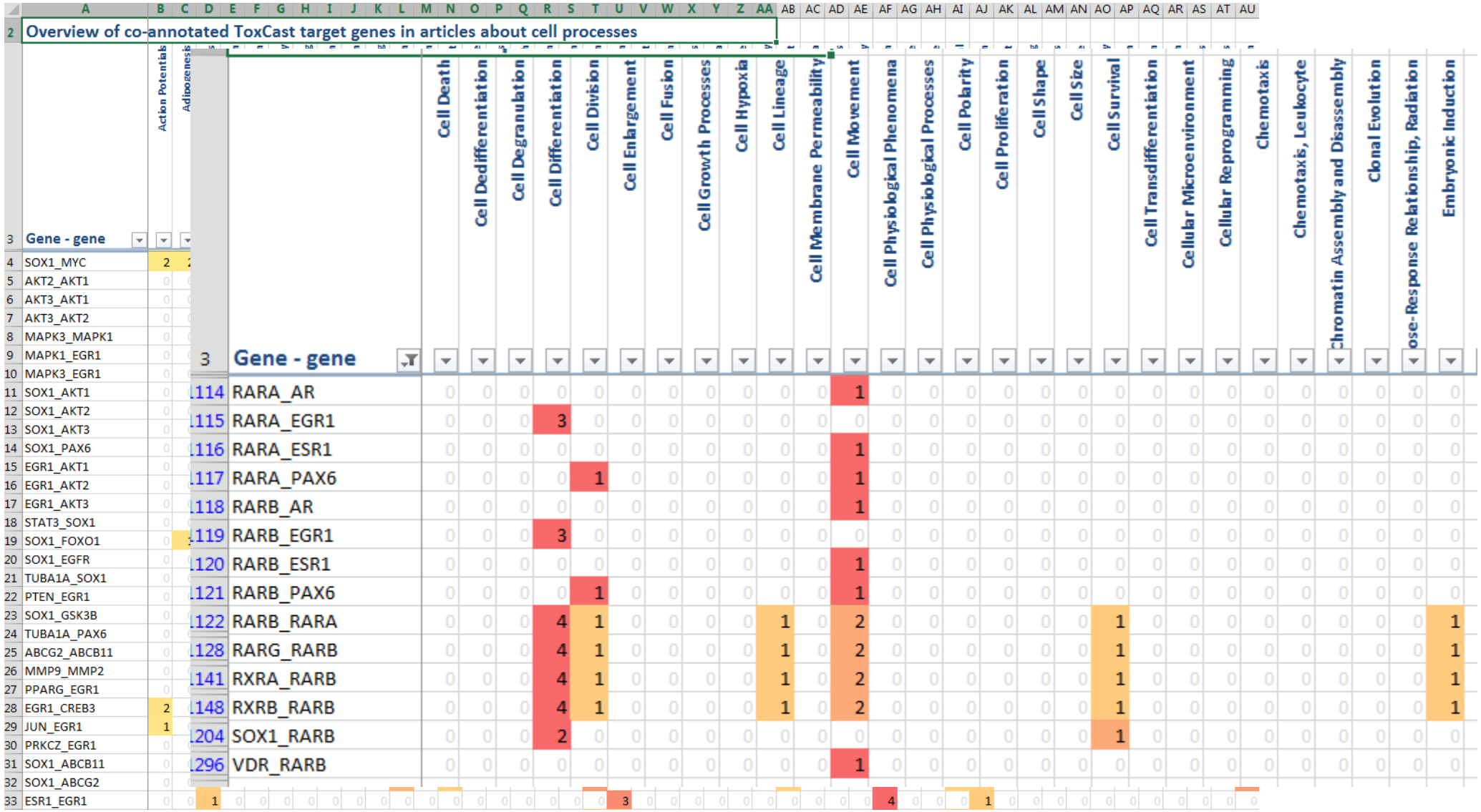
<-- Back

	Target\Activity	Chemical	TC	PMID	Pub Yr	Title
8442	Receptors, Retinoic Acid\ag	AGN 194204	0	11428923	2001	Enantioselective syntheses of potent retinoid X receptor ligands: differential biological activities o
8443	Receptors, Retinoic Acid\ag	AGN 194204	0	14712350	2004	Retinoic acid receptor alpha and retinoid X receptor specific agonists reduce renal injury in establis
8444	Receptors, Retinoic Acid\ag	AGN 195183	0	14712350	2004	Retinoic acid receptor alpha and retinoid X receptor specific agonists reduce renal injury in establis
8445	Receptors, Retinoic Acid\ag	alitretinoin	1	11397803	2001	Regulation of stearoyl coenzyme A desaturase expression in human retinal pigment epithelial cell
8446	Receptors, Retinoic Acid\ag	alitretinoin	1	11261782	2001	Retinoic acid up-regulates myeloid ICAM-3 expression and function in a cell-specific fashion--evid
8447	Receptors, Retinoic Acid\ag	alitretinoin	1	12006550	2002	Nuclear receptor agonists as potential differentiation therapy agents for human osteosarcoma.
8448	Receptors, Retinoic Acid\ag	alitretinoin	1	12882839	2003	Retinoic acids exert direct effects on T cells to suppress Th1 development and enhance Th2 develo

Relationship: chemical – disease or condition

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	Overview																													
2	Chemical Name	Hit group	Disorders of Sex Development	Fetal Death	Eye Abnormalities	Embryo Loss	Cleft Palate	Cleft Lip	Craniofacial Abnormalities	Fetal Resorption	Musculoskeletal Abnormalities	Abnormalities, Multiple	Prenatal Exposure Delayed Effects	Fetal Growth Retardation	Heart Defects, Congenital	Hydronephrosis	Microcephaly	Neural Tube Defects	Limb Deformities, Congenital	Hydrocephalus	Cheilitis	Ectromelia	Spina Bifida Occulta	Bone Diseases	Clubfoot	Jaw Abnormalities	Microphthalmos	Spinal Diseases	Spinal Dysraphism	Urogenital Abnormalities
3	Dieldrin	rar	1	2			1	1																						
4	Endrin	rar					2	1								1														
5	Endosulfan	rar		1						2		1					1													
6	Aldrin	rar					1	1																						
7	tributyltin	rxr	33	2	2	2	1		1	1	1																			
8	4-nonylphenol	rxr	3									2																		
9	triphenyltin hydroxide	rxr											1	1																
10	fenvalerate	rxr																												
11	tran-Retinoic acid	RA		12	5	2	35	1	12	3		23	2		14	1	4	13	6	5	4	4	4	2	2	2	2	2	2	2
12	Retinol	RA		6	2	2	29	2	1	5		8		1	4	2	2	7	2			3	1			3	1	1	5	
13	2,4,5-trichlorophenol			1			1																							

Examples



Overview of co-annotated ToxCast target genes in articles about cell processes

Gene - gene	Action Potentials	Adipogenesis	Cell Death	Cell Dedifferentiation	Cell Degranulation	Cell Differentiation	Cell Division	Cell Enlargement	Cell Fusion	Cell Growth Processes	Cell Hypoxia	Cell Lineage	Cell Membrane Permeability	Cell Movement	Cell Physiological Phenomena	Cell Physiological Processes	Cell Polarity	Cell Proliferation	Cell Shape	Cell Size	Cell Survival	Cell Transdifferentiation	Cellular Microenvironment	Cellular Reprogramming	Chemotaxis	Chemotaxis, Leukocyte	Chromatin Assembly and Disassembly	Clonal Evolution	Dose-Response Relationship, Radiation	Embryonic Induction
SOX1_MYC	2	2																												
AKT2_AKT1	0	0																												
AKT3_AKT1	0	0																												
AKT3_AKT2	0	0																												
MAPK3_MAPK1	0	0																												
MAPK1_EGR1	0	0																												
MAPK3_EGR1	0	0																												
SOX1_AKT1	0	0																												
SOX1_AKT2	0	0																												
SOX1_AKT3	0	0																												
SOX1_PAX6	0	0																												
EGR1_AKT1	0	0																												
EGR1_AKT2	0	0																												
EGR1_AKT3	0	0																												
STAT3_SOX1	0	0																												
SOX1_FOXO1	0	1																												
SOX1_EGFR	0	0																												
TUBA1A_SOX1	0	0																												
PTEN_EGR1	0	0																												
SOX1_GSK3B	0	0																												
TUBA1A_PAX6	0	0																												
ABCG2_ABCB11	0	0																												
MMP9_MMP2	0	0																												
PPARG_EGR1	0	0																												
EGR1_CREB3	2	2																												
JUN_EGR1	1	1																												
PRKCZ_EGR1	0	0																												
SOX1_ABCB11	0	0																												
SOX1_ABCG2	0	0																												
ESR1_EGR1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	4	0	0	0	1	0	0	0	0	0	0
RARA_AR			0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARA_EGR1			0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARA_ESR1			0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARA_PAX6			0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARB_AR			0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARB_EGR1			0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARB_ESR1			0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARB_PAX6			0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RARB_RARA			0	0	0	4	1	0	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
RARG_RARB			0	0	0	4	1	0	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
RXRA_RARB			0	0	0	4	1	0	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
RXRB_RARB			0	0	0	4	1	0	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
SOX1_RARB			0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
VDR_RARB			0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Disease and protein

	Proteins annotated with hypospadias			
2	**Protein**	**PMID**	**PubYr**	**Title**
3	17beta-hydroxysteroid dehydrogenase type 3	20059664	2010	Genetic polymorphisms of 17 Î²-hydroxysteroid dehydrogenase 3 and the risk of hypospadias.
4	17-Hydroxysteroid Dehydrogenases	6332300	1984	An improved method for evaluating testosterone biosynthetic defects.
5	17-Hydroxysteroid Dehydrogenases	6105737	1980	Pseudovaginal perineoscrotal hypospadias: genetic heterogeneity.
6	17-Hydroxysteroid Dehydrogenases	20059664	2010	Genetic polymorphisms of 17 Î²-hydroxysteroid dehydrogenase 3 and the risk of hypospadias.
7	17-Hydroxysteroid Dehydrogenases	9112555	1997	Defects of the testosterone biosynthetic pathway in boys with hypospadias.
8	20-Hydroxysteroid Dehydrogenases	3160950	1985	Elevated 17-hydroxyprogesterone and testosterone in a newborn with 3-beta-hydroxysteroid dehydrogenase deficiency.
9	3 (or 17)-beta-hydroxysteroid dehydrogenase	9112555	1997	Defects of the testosterone biosynthetic pathway in boys with hypospadias.
10	3-Hydroxysteroid Dehydrogenases	15181062	2004	Lack of defects in androgen production in children with hypospadias.
11	3-Hydroxysteroid Dehydrogenases	14764821	2004	Molecular study of the 3 beta-hydroxysteroid dehydrogenase gene type II in patients with hypospadias.
12	3-Hydroxysteroid Dehydrogenases	6603965	1983	Differences in inhibition by various steroids of rat testis and Pseudomonas testosteroni delta 5-3 beta-hydroxysteroid dehydrogenase.
13	3-Hydroxysteroid Dehydrogenases	3867211	1985	Male pseudohermaphroditism due to 3 beta-hydroxysteroid dehydrogenase-isomerase deficiency associated with atrial septal defect.
14	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	8723114	1996	Phenotypic classification of male pseudohermaphroditism due to steroid 5 alpha-reductase 2 deficiency.
15	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	1568634	1992	[A case of pseudo-vaginal, perineoscrotal hypospadia with 5-alpha reductase deficiency].
16	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	512549	1979	Reduction of androstenedione by skin in vitro and serum levels of gonadotrophins and androgens in men with hypospadias.
17	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	17609295	2008	Molecular characterization of 6 unrelated Italian patients with 5alpha-reductase type 2 deficiency.
18	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	8789759	1996	Molecular genetic analysis and human chorionic gonadotropin stimulation tests in the diagnosis of prepubertal patients with partial 5 alpha-reducti
19	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	3920857	1985	Endocrine and immunogenetic evaluation of an XX male infant with perineoscrotal hypospadias.
20	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	564935	1978	Metabolism of androstenedione in skin and serum levels of gonadotrophins and androgens in prepubertal boys with hypospadias.
21	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	22678668	2012	Environmental and genetic contributors to hypospadias: a review of the epidemiologic evidence.
22	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	8339743	1993	The androgen resistance syndromes: clinical and biochemical aspects.
23	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	18097518	2008	Molecular diagnosis of 5alpha-reductase-2 gene mutation in two Indian families with male pseudohermaphroditism.
24	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	2913055	1989	Intracellular and nuclear binding of [3H]dihydrotestosterone in cultured genital skin fibroblasts of patients with severe hypospadias.
25	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	6105737	1980	Pseudovaginal perineoscrotal hypospadias: genetic heterogeneity.
26	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	521711	1979	Further studies of testosterone 5 alpha-reductase deficiency in human fibroblasts [proceedings].
27	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	16736621	2006	[Mutation analysis of SRD5A2 gene in patients with hypospadias].
28	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	3263511	1988	Androgen receptor levels and 5 alpha-reductase activities in preputial skin and chordee tissue of boys with isolated hypospadias.
29	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	6699962	1984	Endocrine studies in patients with advanced hypospadias.
30	3-Oxo-5-alpha-Steroid 4-Dehydrogenase	6480803	1984	Partial androgen resistance associated with secondary 5 alpha-reductase deficiency: identification of a novel qualitative androgen receptor defe

Literature signal at a high level over several toxicity types

	A	B	C	J	K	L	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	**Overview**																											
2																												
3				DevTox			Obesity				ReproTox						Thyroid											
5	**Chemical**	**gsid**	**cas**	Abnormality	Embryonic Structures	Morphogenesis	Adipogenesis	Adipose Tissue	Clinical Conditions	Proteins and genes	Abnormality	Cancer	FemaleRepro	Infertility	MaleRepro	Processes	Binding Proteins	Hormones	Receptors	Synthesis	Hepatic Catabolism	TR Contolled Genes	Transporters	Body Temp Reg	Clinical Conditions	Cognition_IQ	Frog	Thyroid Gland
6	Diethylhexyl Phthalate	20607	117-81-7	79	55	51	5	8	97	11	9	11	79	20	268	407	1	18	2	2	19	1	0	0	0	7	0	5
7	Diethylstilbestrol	40770	6898-97-1	357	209	72	0	0	63	1	36	95	632	74	255	1216	8	70	0	24	8	0	0	3	0	3	0	4
8	Cyclophosphamide	208761	60030-72-0	390	185	101	0	1	113	1	0	45	150	88	234	513	0	36	0	10	6	10	1	4	7	26	0	6
9	bisphenol A	27480	2444-90-8	69	125	146	2	21	184	36	3	0	184	36	200	809	0	60	19	7	40	4	1	0	0	27	30	11
10	Dibutyl Phthalate	21781	84-74-2	111	33	45	0	0	32	3	21	3	22	12	194	212	0	5	6	1	2	0	0	0	3	1	4	0
11	Estradiol	20573	50-28-2	89	57	60	0	3	96	182	12	7	296	24	149	792	43	762	22	129	115	42	21	0	0	19	16	7
12	Ethinyl Estradiol	20576	57-63-6	103	47	87	0	3	149	9	0	4	232	8	107	799	14	66	0	6	19	0	3	1	0	7	26	6
13	Flutamide	32004	13311-84-7	63	1	13	0	2	18	5	15	0	12	4	91	77	0	3	0	0	4	3	1	0	0	0	0	1
14	methyl cellosolve	24182	109-86-4	85	31	27	0	0	34	0	0	0	15	4	87	141	0	0	0	1	1	0	0	2	0	0	0	0
15	mono-(2-ethylhexyl)phthalate	25680	4376-20-9	5	16	7	2	2	7	2	0	3	19	0	75	24	0	3	0	0	6	0	0	0	0	0	0	1
16	Atrazine	20112	1912-24-9	73	25	121	0	2	45	0	3	0	58	0	70	207	0	14	0	3	0	0	0	1	3	8	133	7
17	vinclozolin	22361	50471-44-8	32	9	36	0	0	15	0	12	0	10	12	69	156	0	5	0	0	0	1	0	2	0	0	0	1
18	3-dinitrobenzene	24065	99-65-0	0	0	0	0	0	5	0	0	0	2	8	66	12	0	0	0	0	0	0	0	0	0	0	0	0
19	tributyltin	40709	688-73-3	162	35	67	4	16	46	8	0	0	31	0	66	174	0	6	2	1	3	0	0	0	0	1	10	2
20	Testosterone	22371	58-22-0	38	7	11	0	1	19	52	2	0	41	8	63	96	9	301	6	10	28	2	0	2	2	0	0	4
21	carbendazim	58370	37574-18-8	9	14	8	0	0	12	0	0	0	7	4	59	56	0	5	0	1	0	0	0	0	0	0	0	1
22	Lindane	20686	58-89-9	24	30	28	0	16	50	0	0	0	31	12	58	192	0	20	0	5	2	0	0	7	5	8	0	7
23	Methotrexate	20822	59-05-2	293	35	44	2	1	91	5	0	27	45	20	55	427	1	7	0	5	13	12	33	3	2	154	0	3

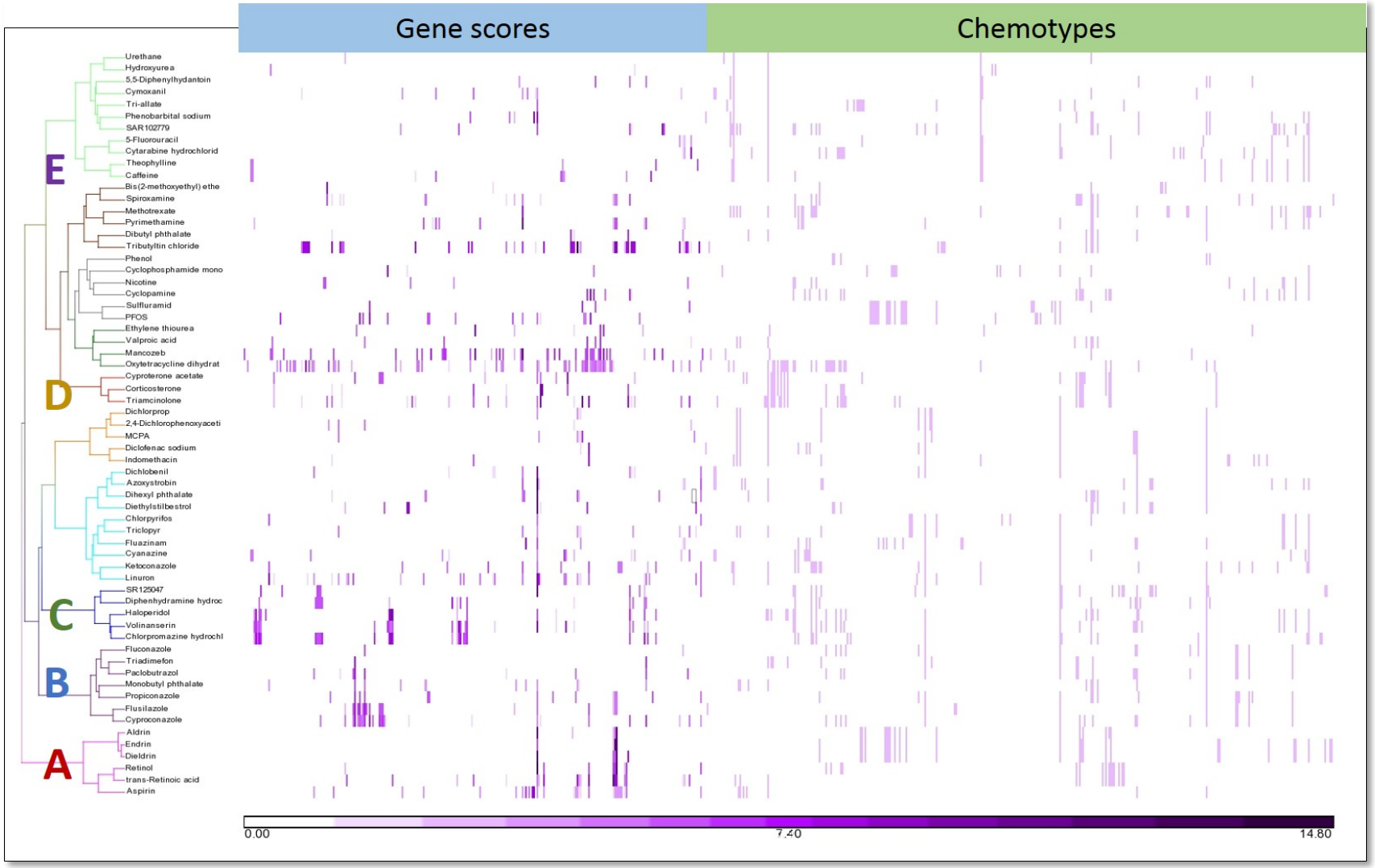
Integration: ToxCast and literature

	A	D	F	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	Overview																															
2				Scores (annotation count * arbitrary weights)																													
3		ATG_Ahr_CIS_up		DevTox			GeneTox				Obesity				ReproTox						Thyroid											
5	Chemical	AC50	Hit Call	Abnormality	Embryonic Structures	Morphogenesis	Mutagen	DNA Damage/Repair	Genetic Structures	Processes	Adipogenesis	Adipose Tissue	Clinical Conditions	Proteins and genes	Abnormality	Cancer	FemaleRepro	Infertility	MaleRepro	Processes	Binding Proteins	Hormones	Receptors	Synthesis	Hepatic Catabolism	TR Contolled Genes	Transporters	Body Temp Reg	Clinical Conditions	Cognition_IQ	Frog	Thyroid Gland
86	1-phenylazo-2-naphthol	-1.418118069	1	0	0	0	12	24	8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
97	Acetaldehyde	-1.252319017	1	1	4	2	24	12	9	0	0	0	2	0	0	0	1	0	4	5	0	2	0	0	0	0	0	1	0	0	0	0
102	1,2,5,6-dibenzanthracene	-1.21708267	1	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
138	indole-3-carbinol	-1.031070118	1	0	0	0	9	6	8	0	0	0	4	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	2
147	Apomorphine	-0.625806458	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	4	0	0	0	0	0	0	0
150	benz(a)anthracene	-0.390970114	1	0	0	4	15	18	21	0	0	0	1	0	0	3	3	0	1	4	0	3	0	0	1	0	0	0	0	0	0	0
287	4-chloro-1,2-diaminobenzene	-0.372897034	1	0	0	0	12	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
292	indeno(1,2,3-cd)pyrene	-0.289718708	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
293	Ketoconazole	-0.193907372	1	28	3	5	0	0	5	1	0	1	2	0	0	0	8	0	18	29	0	18	0	2	20	0	3	0	5	4	0	1
300	1,4-naphthoquinone	-0.192036977	1	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
304	beta-Naphthoflavone	-0.189962165	1	0	1	0	3	27	1	0	0	0	0	0	0	0	1	0	0	2	0	1	0	0	6	0	0	0	0	0	0	0
332	Cycloheximide	-0.124700601	1	4	4	1	30	24	183	2	0	0	0	0	0	0	1	0	0	8	0	43	7	12	0	0	0	0	0	4	2	0
353	3,3',5,5'-tetramethylbenzidine	-0.112828864	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
377	Methenamine	-0.064371953	1	0	1	0	12	3	2	0	0	0	2	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
430	dibenzo(1,4)dioxin	-0.058282855	1	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	1

Integration: ToxCast and chemical structure as chemotypes

In this figure: chemicals that cause cleft palate and have ToxCast results in the form of gene scores (summarized by gene and burst-adjusted)

The chemicals were clustered based on their ToxCast scores and structure. The hypothesis: chemicals that cluster together cause cleft palate through a similar mechanism.



Summary

- Literature mining can be used to investigate the retinoid system on several levels from several angles
- Integration possible
  - ToxCast and literature
  - ToxCast and chemical structure

Acknowledgements

- Tom Knudsen – NCCT
- Ann Richard – NCCT
- Chris Grulke – NCCT
- Nisha Sipes - NIEHS
- National Center for Computational Toxicology (NCCT), US EPA