Chemical and HTS Profiling of 63 Cleft Palate Teratogens from ToxCast

Nancy C. Baker[1], Nisha S. Sipes[2], Christopher M. Grulke[1], Maxwell C. K. Leung[3], Jill A. Franzosa[3], Barbara D. Abbott[4], Richard S. Judson[5], Thomas B. Knudsen[5]

[1]Lockheed Martin, [2]NIH/NIEHS/Division of the National Toxicology Program, [3]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, 37831, [4] National Health and Environmental Effects Research Laboratory, [5]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711

Nancy Baker | baker.nancy@epa.gov | 919-541-2680

Introduction

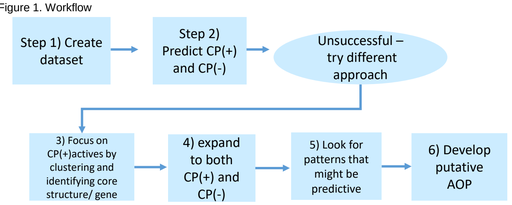
Cleft palate (CP) is a birth defect that has been linked to both genetic and environmental factors. To characterize the potential molecular targets and biological processes across mechanistically diverse teratogens that cause cleft palate in animal studies, we mined the ToxCast high-throughput screening (HTS) database for chemical-assay-structure relationships.

Objectives:

- Describe patterns and linkages in the data that identify molecular activity that may be linked to CP.
- Elucidate a plausible AOP based on high-throughput literature mining and analysis.

Methods

Figure 1. Workflow



Step 1) A dataset was created from 63 CP-active (CP(+)) chemicals (chemicals found either in ToxRefDB to cause CP or in articles found in the biomedical literature) and 437 CP-inactive (CP(-)) chemicals (chemicals that have been tested in developmental studies and that showed no CP). An array of data elements was constructed for each chemical. The arrays consisted of ToxCast gene scores [1] and chemotype information [2].

Step 2) Weka data-mining software was used to build models to predict CP activity [3]. The models showed consistently low performance, even when resampling techniques were used to address the class imbalance.

Step 3) We then focused on the CP(+) chemicals only. We clustered the CP(+) set using the same array of data elements used in the predictive modeling. Clustering was performed using Partek software and Pearson Dissimilarity / Ward's method. Several small but clear clusters were apparent. For each cluster, we identified the feature of the cluster (gene or chemotype) that was common to the chemicals and informative.

Step 4) We took each of the features identified in Step 3 and found all the chemicals (CP(+) AND CP(-)) that had that feature.

Step 5) We then looked for patterns, either through sorting or clustering that would help predict CP activity in this restricted set of chemicals and features [4].

Step 6) Starting with the genes found to be common and informative in each cluster we used the biomedical literature to build putative AOPs.

Figure 2. Illustration of modeling set construction

Chemical	CP?	ToxCast Gene Scores								Chemotypes (383 total)							
		Abl1	Ache	Acp1	Adcy5	Adora1	[...272 genes ...]	Vdr	Xbp1	O=P	H-S	O=C	C-N	N-C=S	[... 376 chemo types ...]	O-S	O=S=O
Chem 1	1	0	2.4	0	0	0		3.4	0.7	1	0	1	0	0	1	0	0
Chem 2	1	0	0	3.1	0.3	0		0.4	1.7	1	0	0	1	0	0	1	0
Chem 3	0	7.2	0	4.4	0	0.3		3.2	5.7	0	1	0	0	1	0	1	0
Chem 4	0	0	0	.09	0	0.5		0	0	1	0	1	1	0	1	0	1

1. Judson, R. S., et al. (2015). "Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity in the 694 ToxCast dataset (manuscript in preparation)."

2. Yang, C., et al. (2015). "New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling." Journal of Chemical Information and Modeling.

3. Hall, M., et al. (2009). "The WEKA Data Mining Software: An Update." SIGKDD Explorations 11(1).

4. Baker, N., et al. Manuscript in preparation.

Results

Figure 3. Clustering of CP(+) chemicals by gene scores and structural chemotypes. Scale represents range of gene scores (0-15). Chemotypes were scaled: 0 (not present) and 3 (present).

DISCLAIMER: does not reflect EPA policy

Results

Cluster A: Retinoic acid receptor gamma (RARG) active chemicals

CP(+) chemicals in the cluster: Dieldrin, Endrin, Aldrin, Aspirin, Trans-retinoic acid, Vitamin A

Common feature: RARG gene score

Chemical	Cleft palate? 0/1	Gene Score RARG
Retinol	1	13.5
Dieldrin	1	12.0
Endrin	1	11.9
Aldrin	1	11.6
Bromuconazole	0	10.8
Tributyltin chloride	1	8.6
trans-Retinoic acid	1	8.2
Rotenone	0	7.7
Fenpyroximate (Z,E)	0	7.6
Chlorothalonil	0	6.7
Triflumizole	0	6.5
SSR126768	0	6.4
Tebufenpyrad	0	6.3
... > 50 more chemicals		

Table 1. Expansion to CP(+) AND CP(-) with activity at RARG gene, sorted by the score. CP(+) highlighted in pink.

Chemicals with a gene score over 8.0 are enriched for CP(+) chemicals.

Cluster B: conazoles with activity at cytochrome P450s (Cyps)

CP(+) chemicals in the cluster: Paclobutazole, Triadimefon, Propiconazole, Monobutyl phthalate, Cyproconazole, Fluconazole, Flusilazole

Common feature: triazine ring

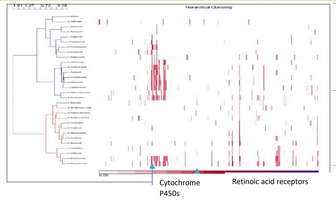


Figure 4. Clustering of CP(+) and CP(-) chemicals with triazine ring.

This cluster has both CP(+) and CP(-) chemicals.

This cluster has only CP(-) chemicals.

Cluster C: chemicals active at G-protein coupled receptors (GPCRs)

CP(+) chemicals in the cluster: Volinanserin, Chlorpromazine, Diphenhydramine, Haloperidol

Common feature: gene score for adrenergic alpha 2B

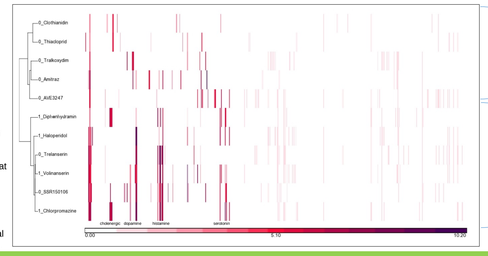


Figure 5. Clustering of CP(+) and CP(-) chemicals that have adrenergic alpha 2B activity by gene scores and structural chemotypes.

This cluster has only CP(-) chemicals.

This cluster has both CP(+) and CP(-) chemicals.

Cluster D: chemicals active at adenosine receptor and phosphodiesterase 10

CP(+) chemicals in the cluster: Caffeine, Theophylline

Common feature: Gene score at adenosine receptors and PDE10A.

Name	Cleft palate? 0/1	Gene scores ADORA1	Gene scores ADORA 2A
Prometryn	0	5.18	5.15
Caffeine	1	5.16	5.24
Cyanazine	1	5.12	5.08
Atrazine	0	5.08	5.16
Theophylline	1	4.86	5.18
Ametryn	0	4.80	4.86
Bromacil	0	4.72	0.00
Prometon	0	4.64	4.91
Desisopropylatrazine	0	3.34	4.96
3-Iodo-2-propynyl-N-butylcarbamate	0	3.05	0.00
Quercetin	0	3.01	0.00
Genistein	0	2.93	0.00

Sorted list of CP(+) and CP(-) chemicals that hit the adenosine receptors and PDE10A.

Name	Cleft palate? 0/1	Gene Score PDE10A
Caffeine	1	5.16
Theophylline	1	4.86

Chemicals that hit an adenosine receptor over 5.0 were likely to be CP(+), and chemicals that hit the phosphodiesterase receptor 10A were both CP(+).

Cluster E: corticosteroids active at the glucocorticoid receptor

CP(+) chemicals in the cluster: Triamcinolone, Cyproterone acetate, Corticosterone

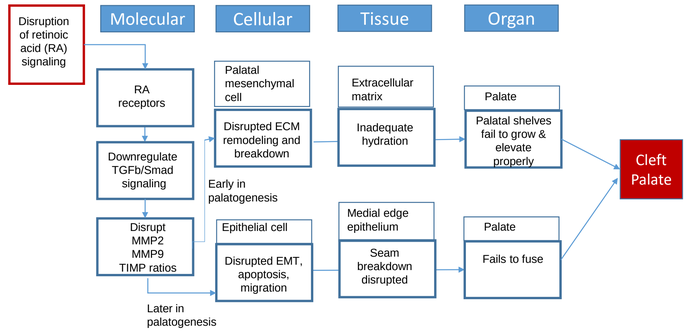
Common feature: gene score at glucocorticoid receptor (NR3C1)

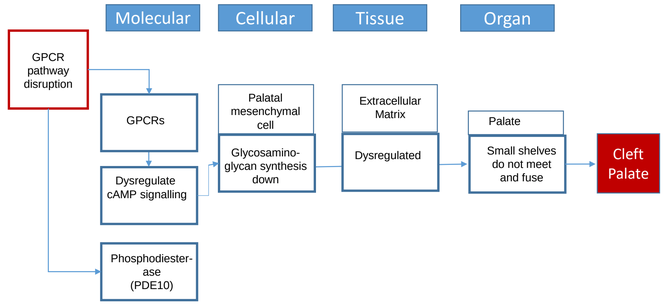
Chemical	Cleft palate 0/1	Gene Score GR (NR3C1)
Corticosterone	1	9.61
Triamcinolone	1	9.57
Mifepristone	0	5.79
Cyproterone acetate	1	4.74
Testosterone propionate	0	3.79
3,3',5,5'-Tetrabromobisphenol A	0	3.46
Triphenyltin hydroxide	0	2.48
Captan	0	2.39
Butachlor	0	2.30
Diethylstilbestrol	1	2.26
Chlorothalonil	0	2.25
Oryzalin	0	2.03
Hexaconazole	0	1.96
Bensulide	0	1.90
Dodecylbenzenesulfonic acid	0	1.88
PFOS	1	1.88
Spiroxamine	1	1.87
Difenoconazole	0	1.86

Sorted list of CP(+) and CP(-) chemicals that hit the glucocorticoid receptor (NR3C1).

Those chemicals that hit NR3C1 (GR receptor) with a potency over 4.00 have a high likelihood of being CP(+).

Putative Adverse Outcome Pathways (AOPs)





Conclusion

- Through a methodology of zooming in and out around key data points in our dataset constructed from ToxCast gene scores and chemotypes, we were able to find the level where ToxCast gene activity and structure data correlated significantly with CP activity.
- Using this ToxCast gene activity as a possible molecular initiating event and using the biomedical literature to support the relationships, we constructed putative adverse outcome pathways for CP.

LOCKHEED MARTIN

v-Embryo™