Building an Adverse Outcome Pathway Framework through HTS Data and Literature Mining Integration

Nancy C. Baker[1], Nisha S. Sipes[2], M. Shane Hutson[3], Thomas B. Knudsen[2]

[1]Lockheed Martin, RTP, NC, [2]U.S. EPA, ORD, National Center for Computational Toxicology, [3]Departments of Physics and Biological Sciences, Vanderbilt University, Nashville, TN

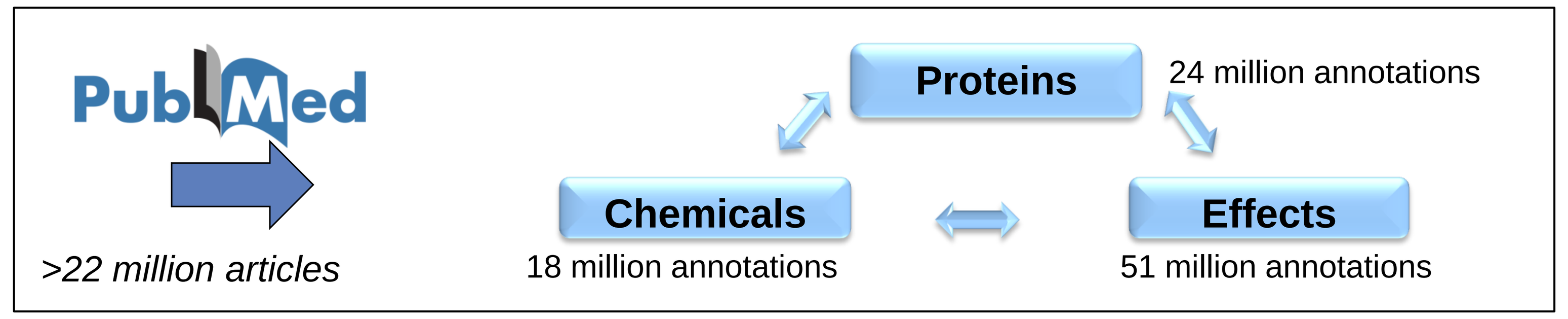
Nancy Baker I baker.nancy@epa.gov I 919-541-2680

Overview

Interpreting EPA's ToxCast in vitro assay data in the context of Adverse Outcome Pathway (AOP) development is a significant challenge. While chemical activation in these assays may shed light on the molecular initiating event, it can be difficult to identify the downstream effect of these activities at higher levels of biological organization (e.g., cellular, tissue, organ) that could potentially lead to a toxicity endpoint.

In this research, we explore applications of literature mining techniques that can be readily used to build and evaluate an AOP framework from in vitro data. We use cleft palate as a prototype and focus on the ToxCast assay targets that were found to have a significant univariate association with cleft palate / cleft lip (CLP) in ToxRefDB and the literature.

Literature Data



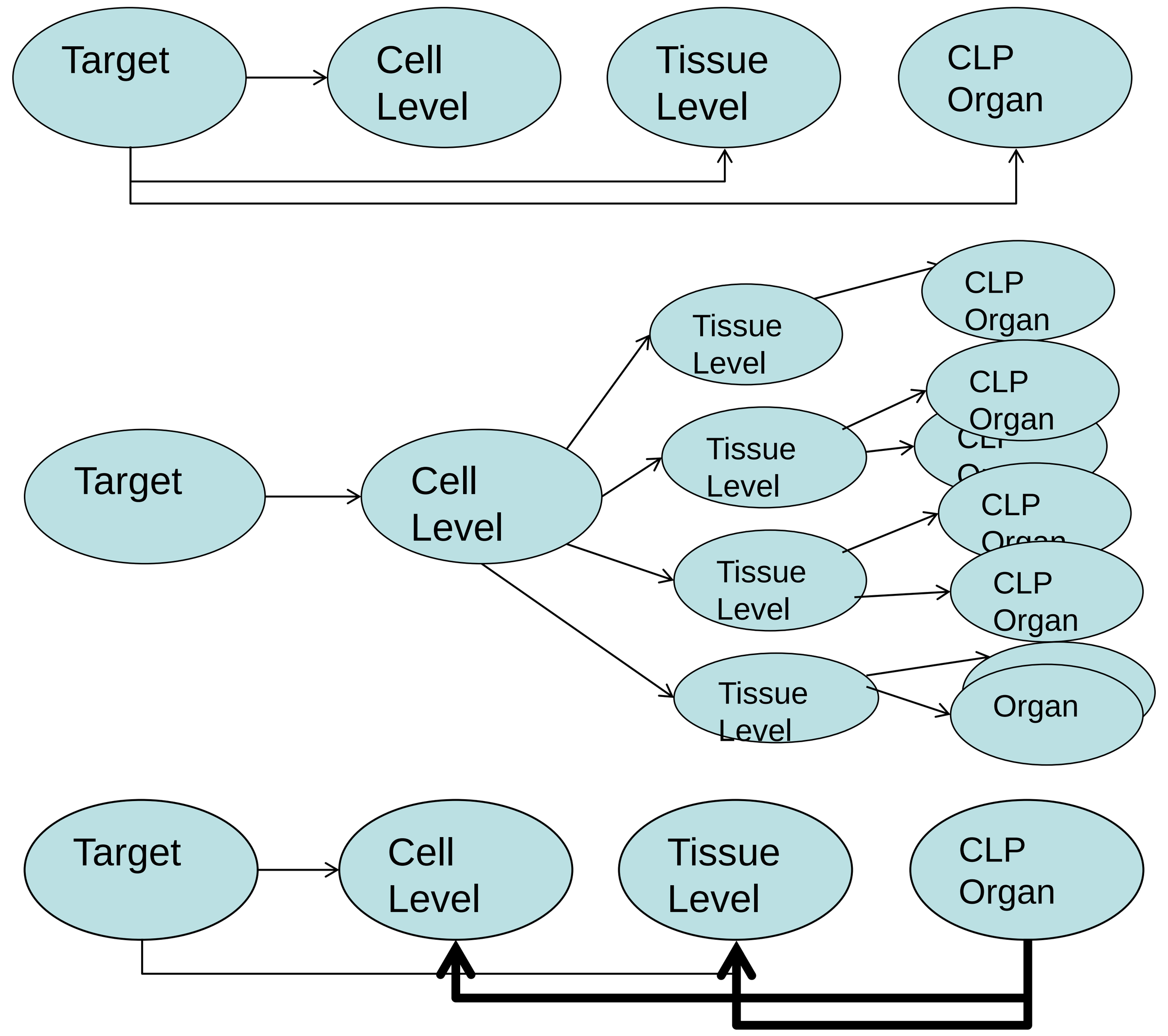
Strategy

Analyze the EPA's ToxCast HTS dataset containing > 1,000 unique chemicals tested across >500 in vitro assays to identify the assays with the most significant univariate association with chemicals that cause cleft palate or cleft lip. Find the literature connections between ToxCast targets and cell-level activity/entities and tissue-level activity and entities. Connections are defined as an article co-annotation between a target protein and the cell/tissue entity.

Option 1. Binary relationships

Option 2. Expand relationships

Option 3. Bidirectional relationships



Results

ToxCast: 56 chemicals in the ToxCast libraries were identified as CLP actives by reviewing data in ToxRefDB and the biomedical literature. A statistical analysis of the assay data identified 29 assays that correlated significantly with the CLP endpoints. (Student's T-test (p<=0.05)) A subset of those assays with the corresponding Medical Subject Heading (MeSH) term for the target protein is found in Table 1. These statistical correlations represent potential molecular targets that have been assayed in the ToxCast portfolio and serve as a potential entry point for AOPs leading to a cleft palate/cleft lip phenotype in pregnant rats and/or rabbits.

Literature: The search of EPA's biomedical literature database using the term Cleft Palate retrieved 36 unique tissue MeSH terms, 37 cell types, 34 cellular processes, and 54 non-cellular processes. Co-annotations of these tissue or cell types and any of the MeSH terms for the ToxCast targets (Table 1) were also retrieved. The articles were counted and the results output to the database and extracted (with hyperlinks) into the spreadsheets show below. Observations deduced from the subject categories serve as a starting point to fill in the biological space between the assays and the endpoint based on what is known in the literature.

AOP Framework:

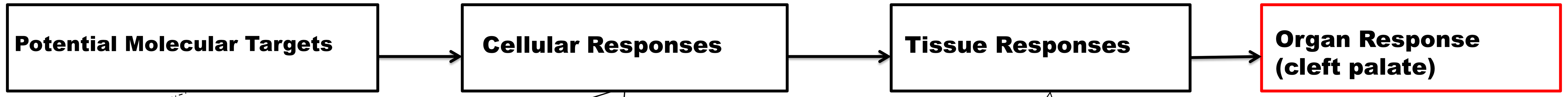
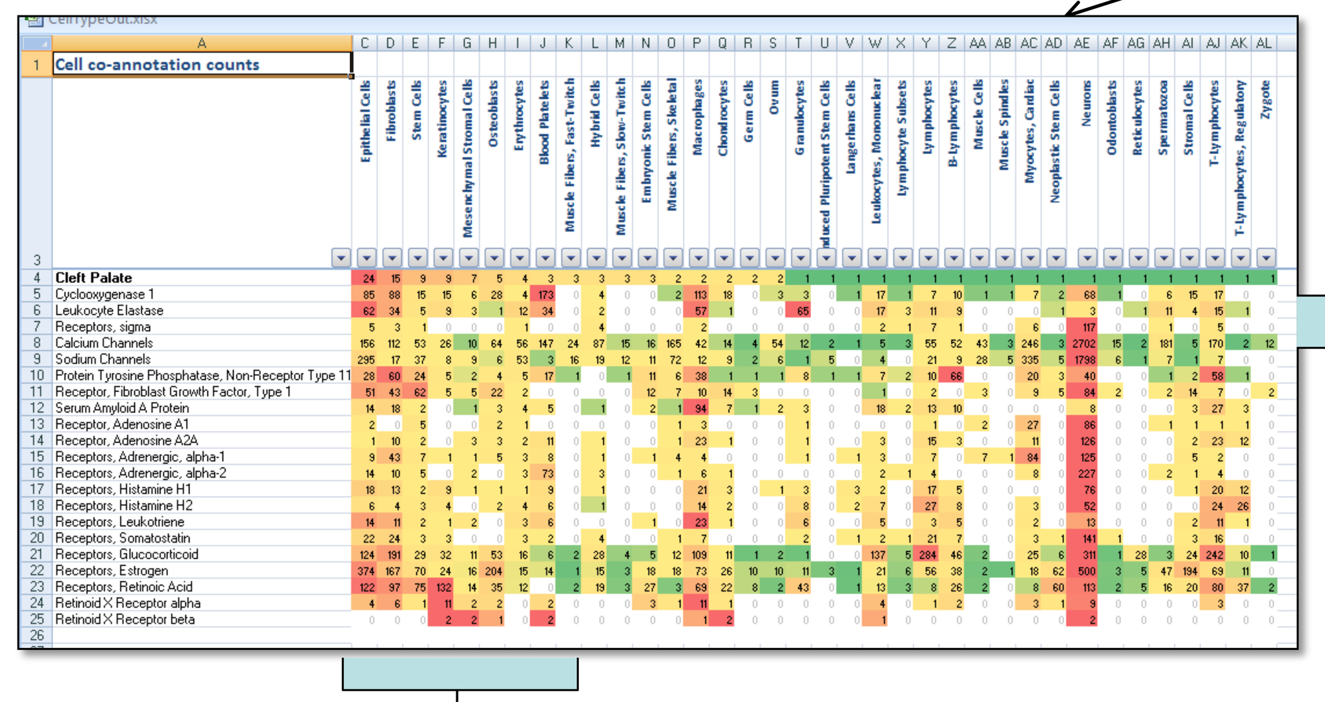
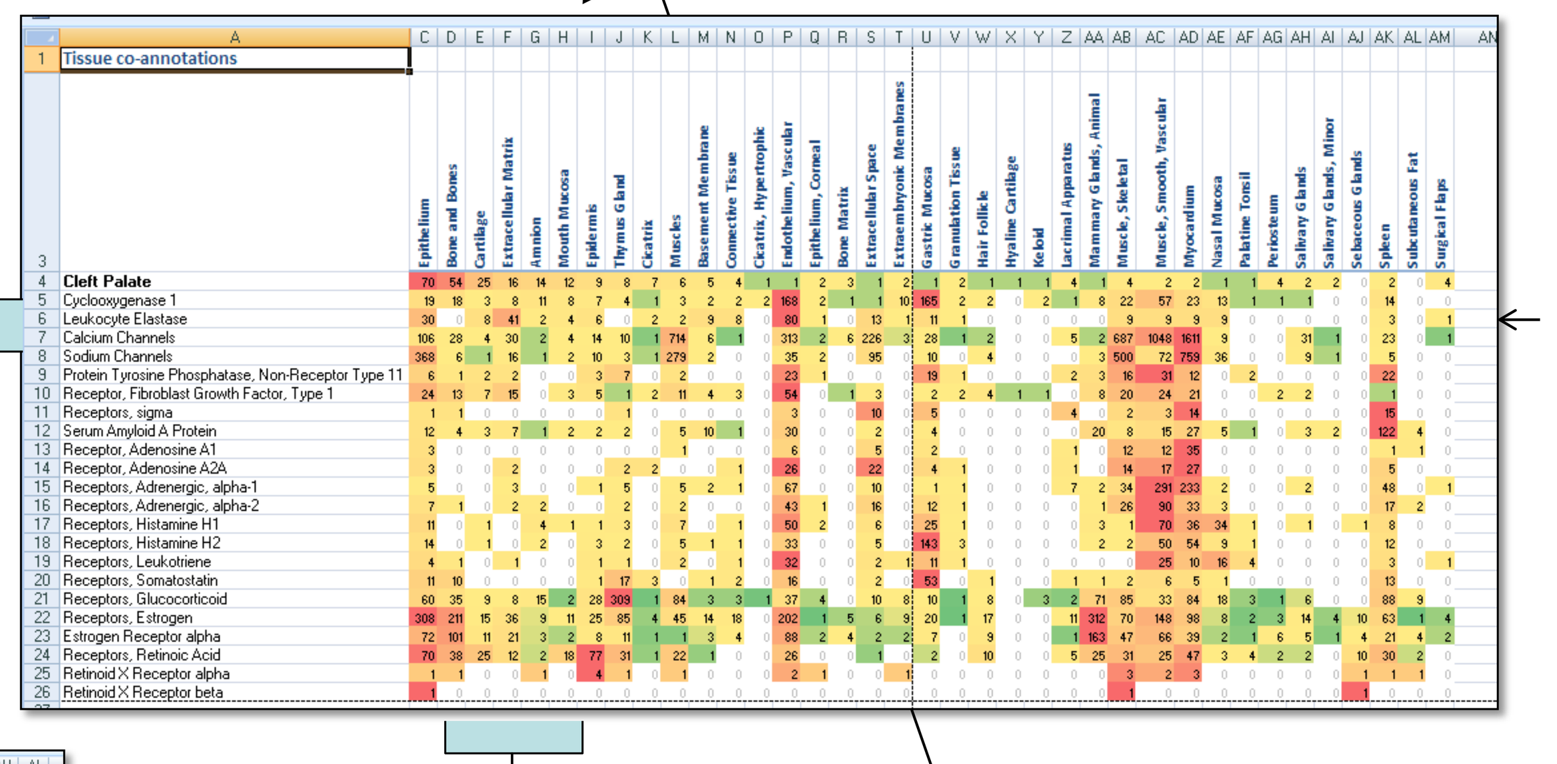


Table 1. Selected assays with significant univariate association with cleft lip / cleft palate; Blue: nuclear receptors; Gray: GPCRs; Green: other

ToxCast Assay	Target - selected MeSH term
NVS_ENZ_oCOX1	Cyclooxygenase 1
NVS_ENZ_hElastase	Leukocyte elastase
NVS_OR_gSIGMA_NonSelective	Receptors, Sigma
NVS_IC_rCaBTZCHL	Calcium Channels
NVS_IC_rNaCh_site2	Sodium Channels
NVS_ENZ_hPTPN11	Protein Tyrosine Phosphatase, Non-Receptor Type 11
NVS_ENZ_hFGFR1	Receptor, Fibroblast Growth Factor, Type 1
BSK_CASM3C_SAA_up	Serum Amyloid A Protein
NVS_GPCR_hAdoRA1	Receptor, Adenosine A1
NVS_GPCR_hAdoRA2a	Receptor, Adenosine A2A
NVS_GPCR_rAdra1A	Receptors, Adrenergic, alpha-1
NVS_GPCR_rAdra1B	Receptors, Adrenergic, alpha-2
NVS_GPCR_bH1	Receptors, Histamine H1
NVS_GPCR_gH2	Receptors, Histamine H2
NVS_GPCR_gLTD4	Receptors, Leukotriene
NVS_GPCR_rSST	Receptors, Somatostatin
ATG_GR_TRANS	Receptors, Glucocorticoid
ATG_ERE_CIS	Receptors, Estrogen
ATG_RARa_TRANS, ATG_RARb_TRANS , ATG_RARg_TRANS	Receptors, Retinoic Acid
ATG_RXRa_TRANS, ATG_RXRb_TRANS	Retinoid X Receptor alpha, Retinoid X Receptor beta

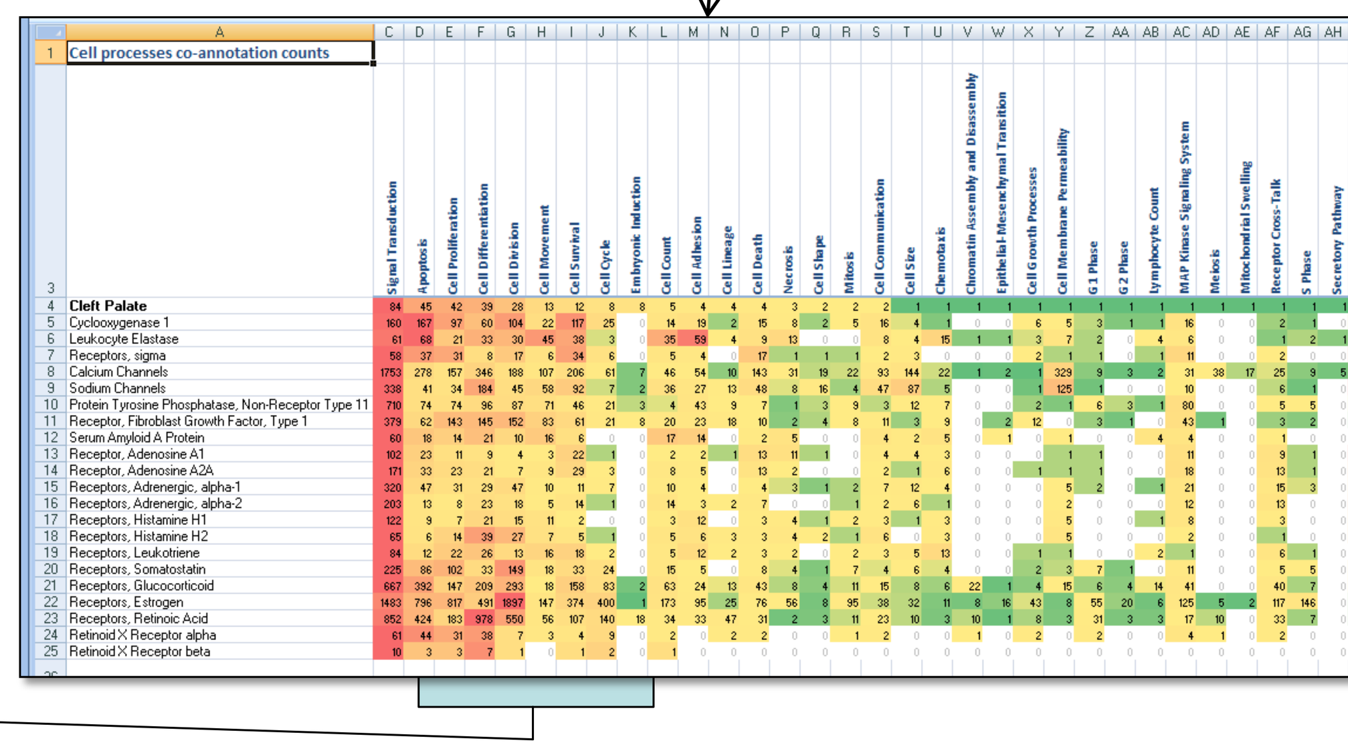


Observation: Leukocyte elastase is highly connected to epithelial cells, immune system cells, and cell adhesion and cell movement.

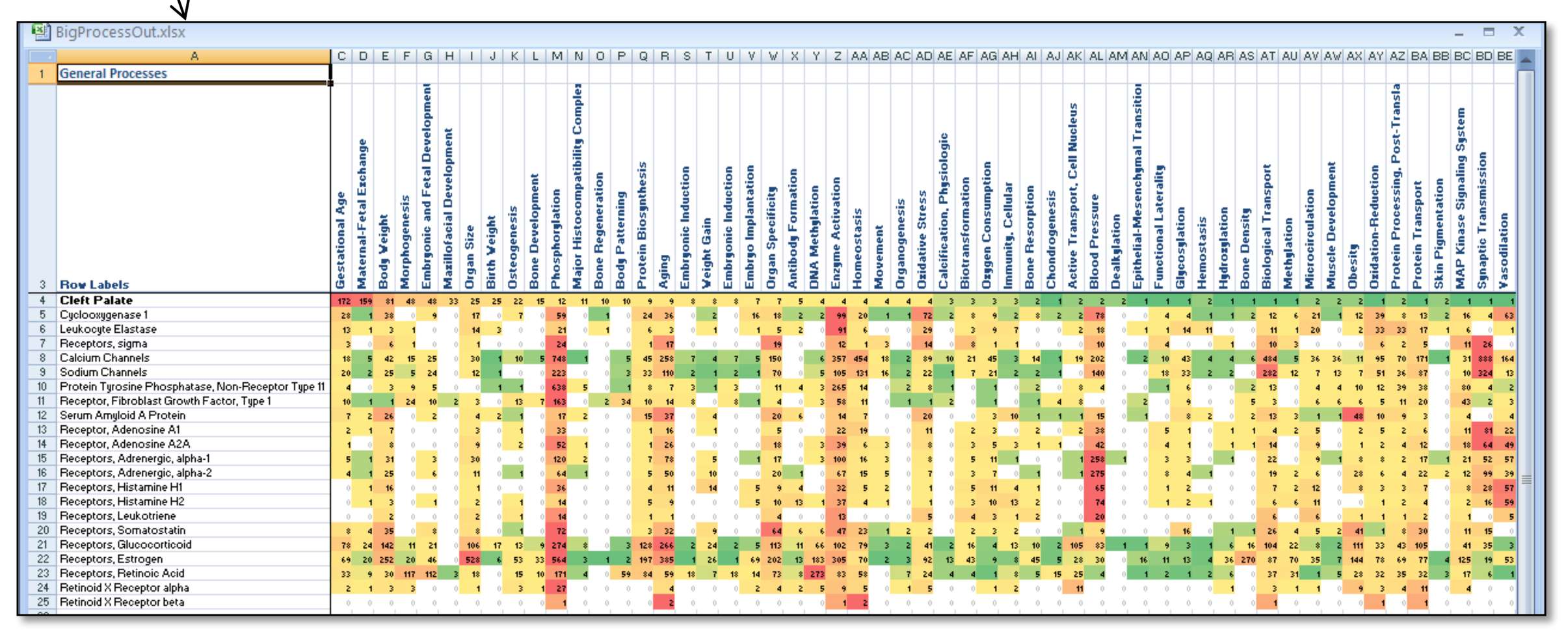


Observation: Leukocyte elastase is highly connected to the extracellular matrix; so is cleft palate.

Observation: Nuclear Receptors are more often associated with epithelial cells and keratinocytes than GPCRs are.



Observation: Nuclear receptors are more connected to Bone, Cartilage, and Extracellular Matrix than GPCRs are. GPCRs have more co-annotations with Vascular Endothelium, but Vascular Endothelium does not have many connections to cleft palate.



Observation: GPCRs are more often co-annotated with Synaptic Transmission than nuclear receptors are … but Synaptic Transmission is not connected strongly to cleft palate.

Observation: Nuclear Receptors are more often associated with cell differentiation than GPCRs are.

Looking for associations and connections in the literature can help direct the AOP development. The observation to the right are possible starting points for an AOP for CPL.

Summary:

•This approach to AOP construction utilizes Weight-of-Evidence from two diverse sources of information: (1) univariate associations between HTS assay (ToxCast) and apical endpoint (ToxRefDB), and (2) literature analysis for relevant biological knowledge processes linked to the candidate molecular targets.

•Benefits are demonstrated in this early AOP framework for cleft palate / cleft lip. CLP-actives could be classified by biological domain (e.g., nuclear receptors, GPCRs) for sorting the relevant literature (shown here) and future classification by chemotype (not shown).

•Coarse analysis of MeSH headings can identify and sort relevant literature based on levels of biological organization (molecular-cellular-tissue-organ). These tools allow for streamlined access to the articles for in-depth analysis that can help investigators deduce weight-of-evidence specific to the endpoint domain and extensible to knowledge from outside that domain.