

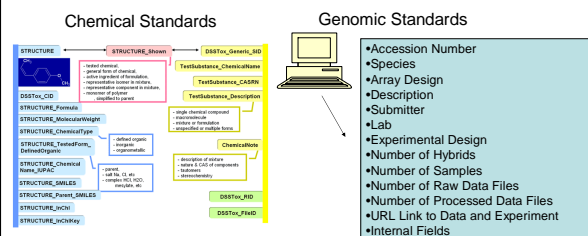
Exploration of the Chemical Space of Public Genomic Databases

ClarLynda Williams-DeVane¹, Maritja Wolf², Ann Richard³. ¹North Carolina State University/U.S. EPA COOP/ Bioinformatics, United States; ²Lockheed Martin (Contractor to U.S. EPA), United States; ³U.S. EPA/Office of Research and Development (ORD)/National Center for Computational Toxicology (NCCT)

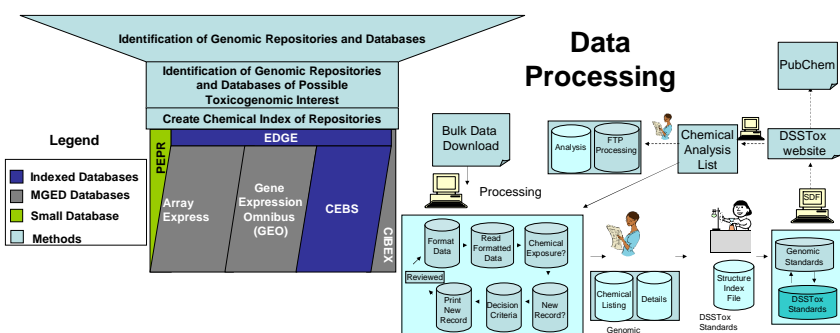
Project Goals

- ❖ Chemically Index the content of Public Gene Expression Databases and/or Repositories
- ❖ Provide Chemical Structure Index Files for each Gene Expression Database
- ❖ Integrate Public Gene Expression Data with Public Toxicology Data through Chemical Structure
- ❖ Develop Methodologies for use of public gene expression/toxicology datasets exploring data across:
 - ❖ Laboratories,
 - ❖ Chemical Space,
 - ❖ Platforms,
 - ❖ and Species

Standardization

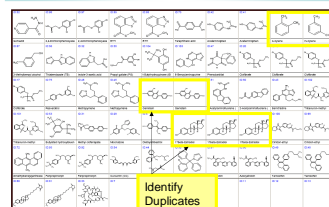


Methods/Approaches



Results

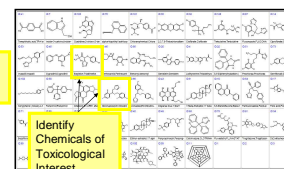
ArrayExpress Database



Statistics

- Initially 106 Chemical Exposure Records
- Currently 351 Chemical Exposure Records
- 106 Chemical Structures overlap 273 Toxicological Records

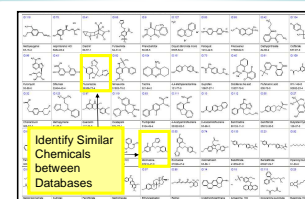
GEO Database



Statistics

- Initially 129 Chemical Exposure Records
- Currently Over 500 Chemical Exposure Records
- 129 Chemical Structures overlap 246 Toxicological Records

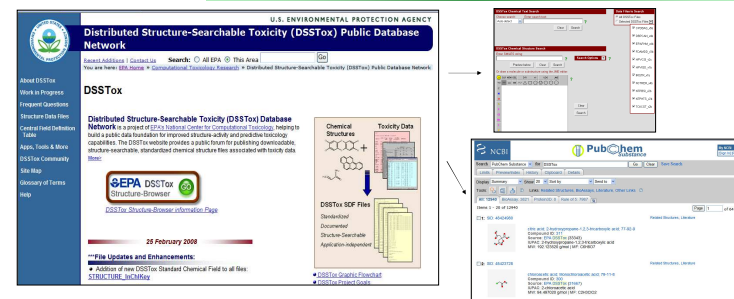
CEBS Database



Statistics

- Initially 135 Chemical Exposure Records
- Currently All Records contain Toxicology Information
- Chemical Structures overlap 538 Toxicological Records

Publication



Future Analysis

Hypothesis 1: Data from the same laboratory on the same species using the same chemical treatment completed at roughly the same time are consistent.

Hypothesis 2: Data from the same laboratory on the same platform using the same chemical treatment on different species yield comparable results.

Hypothesis 3: Data from different laboratories on the same species on the same platform using comparable chemical treatments use the comparable results.

Hypothesis 4: Data from the same laboratory on different platforms using the same chemical treatment on the same species yield comparable results.

Hypothesis 5: Experimental data deposited into two different genomic repositories, yield the same results.

References

1. ArrayExpress <http://www.ebi.ac.uk/ArrayExpress>
2. CEBS <http://cebs.niehs.nih.gov>
3. Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo>
4. Environment, Drug, and Gene Expression (EDGE) <http://edge.oncology.wisc.edu/>
5. Iconix B. Ganter et al, Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action, Journal of Biotechnology (2005) 119: 219-44.
6. PubChem <http://pubchem.ncbi.nlm.nih.gov>
7. DSSTox <http://www.epa.gov/ncct/dsstox/index.html>

This work was reviewed by U.S. EPA and approved for publication but does not necessarily reflect official Agency policy.