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)

Results **Project Goals** ArrayExpress Database **GEO** Database **CEBS** Database *Chemically Index the content of Public Gene Expression Databases and/or Repositories 3 concorror of of more & 200 00 -000 mm di vito *Provide Chemical Structure Index Files for each Gene Expression Database ŝ mall right right mar sho of a منه مند بع من من من مد 1000 *Integrate Public Gene Expression Data with Public Toxicology Data through to an Ex the war and and Romm-der word 3 Identify General Companyantine State State State Chemical Structure Annual Street 500 the as to at at the the Similar Chemical En the of the *Develop Methodologies for use of public gene expression/toxicology datasets The new she affer affer affer and and exploring data across: Identify at war ge Identify Similar 30 horse broke 38 27 3 mer 300 200 200 Chemicals of Chemicals Laboratories, NUMBER PORT PORT · # .** @ Toxicological between Interest of - and and all - the Databases Identify ad- Chemical Space, -189. Duplicates Statistics Platforms, Initially 129 Chemical Exposure Records Statistics Statistics Currently Over 500 Chemical Exposure Records Initially 106 Chemical Exposure Records Currently 351 Chemical Exposure Records Initially 135 Chemical Exposure Records 129 Chemical Structures overlap 246 Toxicological and Species Currently All Records contain Toxicology Information Records 106 Chemical Structures overlap 273 Chemical Structures overlap 538 Toxicological Records Standardization Toxicological Records Chemical Name Publication Genomic Standards Chemical Standards Level of Testing •Vehicle/Medium Accession Number STRUCTURE . •Vehicle of drug/compound Species ributed Structure-Searchable Toxicity (DSSTox) Public Database administration Array Design •MIAME score (0-5) Reference Used to mimic biological or TestSubstance CASR •Description •Array •Submitter Protocols environmental situations •Combination_Treatment or - single chemical compound macromotiecule STRUCTURE_Molece •Lab •Factors SSTox STRUCTURE_ChemicalType Experimental Design Combination_Reference - defined organic - inorganic - organometallic •Raw data STRUCTURE_TextedForm_ DefinedOrpanic •Number of Hybrids •Processed data Used in combination with other hable Toxicity (DSSToy) D STRUCTURE_Chemical Name_UPAC treatment or reference chemicals •Number of Samples - parent. - salt Na, CL etc - complex HCL H2O metylate, etc STRUCTURE_SWILES Unknown Pub@hem 85 W/W •Number of Raw Data Files 1 Br Level of testing not known STRUCTURE_Parent_SMI DSSToy BI •Number of Processed Data Files STRUCTURE INChI for chemical DSSTox_FileID •URL Link to Data and Experiment STRUCTURE_INCHINEY •Treatment SEPA DSSTo Internal Fields Primary Focus of Experiment **Methods/Approaches** \mathbb{R}^{k} $\gamma\gamma$ entification of Genomic Repositories and Databases Data PubChem Future Analysis References Identification of Genomic Repositories Processing and Databases of Possible ArrayExpress http://www.ebi.ac.uk/ArrayExpress **Toxicogenomic Interest** hypothesis 1: Data from the same laboratory of -lypothesis 4: Data from Create Chemical Index of Repositories 2 CEBS http://cebs.niehs.nih.gov the same species using the same chemical the same laboratory on Gene Expression Omnibus (GEO) http://www.ncbi.nlm.nih.gov/geo Bulk Data 3 DSSTox treatment completed at roughly the same time l egend EDGE lifferent platforms using Environment, Drug, and Gene Expression (EDGE) Download are consistent. Analysis website the same chemical http://edge.oncology.wisc.edu/ List treatment on the same Iconix B. Ganter et al, Development of a large-scale chemogenomics 5 Indexed Databases Hypothesis 2: Data from the same laboratory on species yield database to improve drug candidate selection and to understand the same platform using the same chemical MGED Databases Processing comparable results. mechanisms of chemical toxicity and action. Journal of Biotechnology ray CEBS reatment on different species yield comparable Small Database (2005) 119: 219-44. results. PubChem http://pubchem.ncbi.nlm.nih.gov 6 Methods Hypothesis 5: DSSTox http://www.epa.gov/ncct/dsstox/index.htm Experimental data 7 Hypothesis 3: Data from different laboratories deposited into two Structur Index File on the same species on the same platform different genomic This work was reviewed by U.S. EPA and approved for publication but does not using comparable chemical treatments use the epositories, yield the necessarily reflect official Agency policy. comparable results. same results **United States** Environmental Protection

NCSU Graduate Symposium