

The Comparative Toxicogenomics Database (CTD)

The CTD (<http://ctd.mdibl.org>) promotes understanding about the effects of environmental chemicals on human health by facilitating cross-species comparative studies of toxicologically important genes and proteins. CTD is now publicly available as a prototype. It provides annotated associations between genes, proteins, sequences, references and chemicals in vertebrates and invertebrates; integrates molecular and toxicology data; implements ontologies; and will describe gene-chemical interactions in diverse organisms. These data provide insight into the genetic basis of variable sensitivity to chemicals and complex interactions between the environment and human health.

Conclusions

Data produced by (eco)toxicogenomics investigations are growing in volume and complexity at a staggering rate. It is not trivial to define precise data content, presentation and exchange formats. However, there is a growing realization within the (eco)toxicogenomics community that, if we are to realize the opportunities offered by omics-based technologies, we will need to change our approach to data handling and work more collaboratively. The authors, also moderators of the RSBI working group, would like to emphasize the need for community participation in the integration of these standardization initiatives. It is hoped that highlighting these different initiatives will help to assess the commonality and optimize harmonization, thus minimizing duplication and incompatibility and achieving cost-effective results in a timely manner.

Acknowledgements

The authors would like to thank the following people for their assistance: Alvis Brazma, Ann Richard, Betty Cheng, Carole Foy, Carolyn Mattingly, Chris Taylor, Craig Zwickl, Dawn Field, Helen Parkinson, Henning Hermjakob, Jason Snape, Jessie Kennedy, John Lindon, Michael Waters, Nancy Doerr, Peter Lord, Raffaella Corvi, Robert Stevens, Cyril Petit, Thomas Papoian, Weida Tong and the SOFG organizers. Susanna-Assunta Sansone is supported by the ILSI-HESI Genomics Committee, Norman Morrison by the NERC Environmental Genomics programme,

Philippe Rocca-Serra by the European Commission NuGO project and Jennifer Fostel by the NIEHS NCT.

References

- ArrayExpress: <http://www.ebi.ac.uk/arrayexpress>
 ArrayTrack: <http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack>
 Ball CA, Sherlock G, Parkinson H, et al. 2002. An open letter to the scientific journals, published in: *Science* 298(5593): 539; *Bioinformatics* 18(11): 1409; *Lancet* 360: 1019.
 Brazma A, Parkinson H, Sarkans U, et al. 2003. ArrayExpress — a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31(1): 68–71.
 Brazma A, Hingamp P, Quackenbush J, et al. 2001. Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genet* 29(4): 365–371.
 BSC; <http://www.csbcn.org>
 CDISC; <http://www.cdisc.org>
 CEBS; <http://cebs.niehs.nih.gov>
 CTD; <http://ctd.mdibl.org>
 DSSTox Coordinating Public Effort project; <http://www.epa.gov/nheer/dsstox/CoordinatingPublicEfforts.html>
 DSSTox; <http://www.epa.gov/nheer/dsstox>
 EBI toxicogenomics; <http://www.ebi.ac.uk/microarray/Projects/tox-nutri>
 EBI; <http://www.ebi.ac.uk>
 ECVAM; <http://ecvam.jrc.cec.eu.int>
 Editorial. 2002. Microarray standards at last. *Nature* 419: 323.
 Editorial. 2003a. EU starts a chemical reaction. *Science* 300, 5618: 405.
 Editorial. 2003b. Europe whittles down plans for massive chemical testing program. *Science* 302, 5647: 969.
 Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1): 207–210.
 EPA DRAFT — Potential implications of genomics for regulatory and risk assessment applications at EPA; <http://www.epa.gov/osa/genomics.htm>
 ERCC; <http://www.cstl.nist.gov/biotech/workshops/ERCC2004>
 FDA draft guidance for industry pharmacogenomic data submissions; <http://www.fda.gov/cder/guidance/5900dft.doc>
 Hermjakob H, Montecchi-Palazzi L, Bader G, et al. 2004. The HUPO PSI molecular interaction format — a community standard for the representation of protein interaction data. *Nature Biotechnol* 22: 177–183.
 ICCVAM; <http://iccvam.niehs.nih.gov/home.htm>
 Ikeo K, Ishi-i J, Tamura T, et al. 2003. CIBEX: center for information biology gene expression database. *C R Biol* 326(10–11): 1079–1082.
 ILSI HESI; <http://hesi.ilsil.org/index.cfm?pubentityid=120>
 IPCS; <http://www.who.int/ipcs/en>
 Kepler; <http://kepler-project.org>
 KNB; <http://knb.ecoinformatics.org>
 LGC; <http://www.lgc.co.uk>
 MARG; <http://www.abrf.org/index.cfm/group.show/Microarray.30.htm>

- Mattes WB, Pettit SD, Sansone A, *et al.* 2004. Database development in toxicogenomics: issues and efforts. *Environ Health Perspect* **112**: 495–505.
- maxd; <http://bioinf.man.ac.uk/microarray/maxd>
- MfB; <http://www.mfbprog.org.uk>
- MGED RSBI Working Groups; <http://www.mged.org/Workgroups/rsbi>
- MGED; <http://www.mged.org>
- NAS Committee on Emerging Issues and Data on Environmental Contaminants; <http://dels.nas.edu/emergingissues/index.asp>
- NCTR-FDA; <http://www.fda.gov/nctr>
- NERC EGTDC; <http://envgen.nox.ac.uk>
- NICEATM; <http://iccvam.niehs.nih.gov/home.htm>
- NIEHS-NCT; <http://www.niehs.nih.gov/nct>
- OECD; <http://www.oecd.org>
- Pennie W, Pettit SD, Lord PG. 2004. Toxicogenomics in risk assessment: an overview of an HESI collaborative research program. *Environ Health Perspect* **112**: 417–419.
- PSI; <http://psidev.sourceforge.net>
- Quackenbush J. 2004. Data standards for 'omic' science. *Nature Biotechnol* **22**: 613–614.
- SEEK; <http://seek.ecoinformatics.org>
- SEND; <http://www.cdisc.org/models/send/v1.5>
- SMRS; <http://www.smrsgroup.org>
- SOFG; <http://www.sofg.org>
- Spellman PT, Miller M, Stewart J, *et al.* 2002. Design and implementation of microarray gene expression mark-up language (MAGE-ML). *Genome Biol* **3**(9): research0046.
- Stoeckert CJ, Parkinson H. 2003. The MGED ontology: a framework for describing functional genomics experiments. *Comp Funct Genom* **4**: 127–132.
- Stoeckert CJ, Causton HC, Ball CA. 2002. Microarray databases: standards and ontologies. *Nature Genet* **32**: 469–473.
- Tong W, Cao X, Harris S, *et al.* 2003. ArrayTrack — supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ Health Perspect* **111**: 1819–1826.
- Tox-MIAMExpress; <http://www.ebi.ac.uk/tox-miamexpress>
- Waters M, Boorman G, Bushel P, *et al.* 2003. Systems toxicology and the chemical effects in biological systems knowledge base. *Environ Health Perspect* **111**: 811–824.
- Waters MD, Fostel JM. 2004. Toxicogenomics and systems toxicology: aims and prospects. *Nature Rev Genet* **5**: 938–948.
- Xirasagar S, Gustafson S, Merrick AB, *et al.* 2004. CEBS object model for systems biology data, CEBS SysBio-OM. *Bioinformatics* **20**(13): 2004–2015.