Comments on November 2003 Draft Guidance for Industry on Pharmacogenomic Data Submissions [Docket No. 2003D-0497]

In the Federal Register, Vol. 68, on November 4, 2003, the FDA announced the availability of the Draft Guidance for Industry on Pharmacogenomic Data Submissions, Docket No. 2003D-0497.
The Draft Guidance includes a solicitation of comments from individuals and groups. The Authors, appreciate this opportunity to comment on the Draft Guidance.

1. Purpose

[Line 408] "*V. Format and Content of a VGDS*"

The Authors believe that further clarification would be helpful in this section, providing recommendations on the format and content of Voluntary Genomic Data Submissions (VGDSs) reports and data.

2. Recommendations

[Line 437] "*An evolving public standard for specific types of experiments, such as the Minimum Information About a Microarray Experiment (MIAME) standard for microarray expression data*."

The Authors recommend a clarification of the MIAME[1] concept. MIAME defines the minimum descriptors that must be reported about array based gene expression monitoring experiments, in order to ensure the interpretability of the results, as well as potential verification by third parties. By providing a structured framework for capturing information, MIAME is a data content standard, not a format standard.

The Authors recommend that MIAME is intended as a guide to Sponsors/Applicants and Agency's reviewers to determine whether the content of VGDS presents sufficient detail to interpret the results unambiguously.

[Line 433] *"Therefore, we offer the following examples of possible VGDS formats:…"*

The Authors bring to the attention of the Agency another possible data format the MicroArray Gene Expression (MAGE) standards[2], MAGE object model (MAGE-OM) and MAGE markup language (MAGE-ML). MAGE provides the formal standard that specifies the communication protocols, ensuring software interoperability and encoding all MIAME required information. MAGE is a stable, adopted specification of the Object Management Group (OMG)[3].

The Authors recommend the inclusion of the MAGE-ML as a possible format for VGDS.

3. Background and Rationale

MIAME and MAGE are the work of the Microarray Gene Expression Data (MGED) Society[4], an international organization of biologists, computer scientists, and data analysts from academia, government and industry across the globe. The current focus of the MGED Society is on establishing standards for microarray data annotation and exchange, facilitating the creation of microarray databases and related software implementing these standards, and promoting the sharing of high quality, well annotated data within the life sciences community. A long-term goal for the future is to extend the mission to other functional genomics and proteomics high throughput technologies.

MIAME and MAGE could assist Sponsors/Applicants to submit accurate information in a standard format and help the Agency to have reliable records on which to base scientific assessment and regulatory decisions.

[Line 429] *"We recommend only that, to achieve the goals of the VGDS process, the data submitted in a VGDS and the level of detail be sufficient for the Agency to interpret the information and independently analyze the data, verify results, and explore possible genotype-phenotype correlations across studies."*

The draft guidance does not indicate to the Sponsors/Applicants the minimal descriptors to be submitted to the Agency. MIAME would provide such minimal requirements to interpret an array based gene expression monitoring experiment and independently analyze the data, verify results. The microarray community response to MIAME has been very favourable and many instrument manufacturers, software developers, and international databases have moved to adapt their systems to capture and manage MIAME-compliant data. Furthermore, the MIAME guideline is used by editors and reviewers of several main scientific journals (including *Science*, the *Nature* group of journals, *Cell*, *The Lancet*, *EMBO*, *Toxicologic Pathology, Environmental Health Perspectives)* in their evaluation of whether or not a manuscript provides as much information as necessary for others to replicate and interpret the analysis presented. Some journals go a step further, requiring accession number from public MIAME-compliant databases (ArrayExpress[5] and GEO[6]) to be supplied at or before acceptance of publication. A third public MIAME-compliant database, CIBEX is currently under development at DDBJ, Japan.

[Line 427] *"Currently, consensus standards do not exist for presenting and exchanging genomic data, although such standards are evolving. Therefore, this guidance does not recommend a specific format for the VGDS."*

Currently the standard MAGE-ML format for data exchange is used by several databases, including: the Stanford Microarray Database (SMD)[7], The Institute for Genomic Research (TIGR) microarray database[8], ArrayExpress at the

European Bioinformatics Institute (EBI)[5], the German Genome Resource Centre (RZPD) and it is under construction at the NCI Microarray Database, at the NIH National Institute of Environmental Health Sciences (NIEHS) Microarray Center, at the National Center for Toxicogenomics Chemical Effects in Biological Systems (CEBS) Knowledge Base and at the FDA National Center for Toxicological Research (NCTR), Center for Toxicoinformatics. Furthermore, several microarray informatics tools are currently implementing MAGE-ML import/export, including Rosetta Resolver v4.0[9], GeneSpring v6.1[10] and BioConductor[11]. These developments simplify data submission and communication, illustrating the utility of MAGE-ML as a data exchange format.

Recently, as part of a collaborative undertaking, the ILSI Health and Environmental Sciences Institute's Committee on Genomics (HESI)[12], the NIH NIEHS National Center for Toxicogenomics (NIEHS-NCT)[13], the FDA NCTR, Center for Toxicoinformatics[14], and the EBI have initiated an harmonization process[15] for reporting and exchanging array-based toxicogenomics experiments. The MAGE-ML format will be used to capture both the information associated with the in-life component and the microarray part of toxicogenomic experiments. This collaboration has developed in the establishment of a Toxicogenomics Working Group (TWG) within the MGED society, providing a public forum for the creation of an internationally compatible informatics platform to exchange toxicogenomics and similarly pharmacogenomics data.

[Line 419] *"The purpose of the VGDS process is to provide the FDA access to emerging pharmacogenomic data so that a foundation can be built for developing scientifically sound regulatory policies. The Agency intends to gain experience and to develop an aggregate genomic knowledge database from multiple VGDSs that could be used to rationally facilitate the use of pharmacogenomics in drug development and to share what general knowledge is learned from the data repositories, where appropriate. The VGDS process will also provide a forum for scientific discussion of exploratory data within the FDA outside of the application review process."*

Critical components in developing such knowledge database would be the breadth, depth, and uniformity of the information provided in each VGDS by Sponsors/Applicants. It is critical to the utility of the knowledge database that the VGDS process is also the vehicle for promoting widely accepted structured framework for capturing information and international standards in data organization, respectively MIAME and MAGE as applied to array based gene expression monitoring experiments.

The Authors are confident that MIAME and MAGE can assist Sponsors/Applicants to submit accurate and information in a standard format and help the Agency to build the knowledge database on which to base scientific assessment and regulatory decisions.

The Authors thank the Agency in advance for the consideration of our comments.

February 2, 2004.

Susanna-Assunta Sansone (on behalf of the Authors)

The European Bioinformatics Institute (EBI),
EMBL Outstation,
Wellcome Trust Genome Campus,
Cambridge CB10 1SD, UK
direct: +44 (0)1223 494 691
fax: +44 (0)1223 494 468
email: sansone@ebi.ac.uk

"The Authors concur with this assessment of the FDA draft guidance document and endorse these comments. The Authors fully support the assessment and have agreed to sign."

The Authors in alphabetical order:

Alvis Brazma
(MGED Board member, Senior Scientist at European Molecular Biology Laboratory and EBI Microarray Informatics Group Leader, Cambridge, UK, email: brazma@ebi.ac.uk)

Charles Troup
(MGED Board member and Technical Staff Member, Agilent Technologies, Palo Alto CA, charles_troup@agilent.com)

Chris Stoeckert
(MGED Board member and Research Associate Professor at University of Pennsylvania, Philadelphia, PA, USA, email: stoeckrt@pcbi.upenn.edu)

Helen Causton
(MGED Board member and Head of CSC/IC Microarray Centre at Imperial College, London, UK, email: helen.causton@csc.mrc.ac.uk)

Helen Parkinson
(MGED Advisory Board member, Microarray Informatics Unit, European Bioinformatics Unit and Scientific Programme Manager, National Cancer Research Institute and ArrayExpress Curation Coordinator at EBI, Cambridge, UK, email: parkinson@ebi.ac.uk)

John Quackenbush
(MGED Board member and Investigator at TIGR, Rockville, MD, USA, email: johnq@tigr.org)

Michael D. Waters
(Assistant Director for Database Development at NIEHS-NCT, Research Triangle Park, NC, USA, email: waters2@niehs.nih.gov)

Michael Miller
(MGED Board member and Senior Application Developer at Rosetta Biosoftware, email: Michael_Miller@Rosettabio.com)

Pascal Hingamp
(MGED Board member and Lecturer at INSERM ERM 206, Marseille, France, email: hingamp@tagc.univ-mrs.fr)

Pierre Bushel
(Bioinformatics Manager of NIEHS Microarray Group, Research Triangle Park, NC, USA, email: bushel@niehs.nih.gov)

Richard S. Paules
(Director of NIEHS Microarray Group, Head of Growth Control and Cancer Group at NIEHS-NCT, Research Triangle Park, NC, USA, email: paules@niehs.nih.gov)

Ronald C. Taylor
(MGED Board member and Senior Research Scientist at Pacific Northwest National Laboratory, Richland, WA, USA, email: ronald.taylor@pnl.gov)

Susanna-Assunta Sansone
(MGED Toxicogenomics Working Group Leader and Nutri/Toxicogenomics Project Coordinator at EBI, Cambridge, UK, email: sansone@ebi.ac.uk)

Weida Tong
(Director of Center for Toxicoinformatics, FDA's NCTR, Jefferson, AR, USA, email: WTong@nctr.fda.gov)

4. References

[1] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C. *et al*. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29:365-71.

[2] Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S. *et al*. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol* 3(9), research0046.1-0046.9.

[3] OMG: http://www.omg.org

[4] MGED Society: http://www.mged.org

[5] Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen, A., Rocca-Serra, P. and Sansone, S.A. (2003) ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*, 31, 68-71.

[6] Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30, 207-10.

[7] Gollub, J., Ball, C.A., Binkley, G., Demeter, J., Finkelstein, D.B., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Kaloper, M., Matese, J.C., Schroeder, M., Brown, P.O., Botstein, D., Sherlock, G. (2003). The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res* 31(1): 94-96.

[8] Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbu (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34(2): 374-378.

[9] Rosetta Resolver, Rosetta Biosoftware:
http://www.rosettabio.com/products/resolver/default.htm

[10] GeneSpring, Silicon Genetics:http://www.silicongenetics.com/

[11] BioConductor: http://www.bioconductor.org/

[12] ILSI-HESI Genomics Committee:
http://hesi.ilsi.org/index.cfm?pubentityid=120

[13] Waters MD, G. Boorman, P. Bushel, M. Cunningham, R. Irwin, A. Merrick, K.Olden, R. Paules, J. Selkirk, S. Stasiewicz, B. Weis, B. Van Houten, N. Walker, and R. Tennant. (2003). Systems toxicology and the chemical effects in biological systems knowledge base. *Environ Health Perspect* Toxicogenomics: 111:811-824.

[14] Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, Harris A, Hong H, Xie Q, Perkins R, Shi L, Casciano D. (2003). ArrayTrack--supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ Health Perspect* Toxicogenomics: 111: 1819-26.

[15] William B. Mattes, Syril D. Pettit, Susanna-Assunta Sansone, Pierre R. Bushel, and Michael D. Waters. (2004). Database development in toxicogenomics: issues and efforts. *Environ Health Perspect* Toxicogenomics: doi:10.1289/txg.6697