

Validation of Toxicogenomics-Based Test Systems: ECVAM-ICCVAM/NICEATM Considerations for Regulatory Use

Raffaella Corvi, Hans-Jürgen Ahr, Silvio Albertini, David H. Blakey, Libero Clerici, Sandra Coecke, George R. Douglas, Laura Gribaldo, John P. Groten, Bernd Haase, Karen Hamernik, Thomas Hartung, Tohru Inoue, Ian Indans, Daniela Maurici, George Orphanides, Diana Rembges, Susanna-Assunta Sansone, Jason R. Snape, Eisaku Toda, Weida Tong, Joost H. van Delft, Brenda Weis, and Leonard M. Schechtman

doi:10.1289/ehp.8247 (available at <http://dx.doi.org/>)

Online 17 August 2005



Validation of Toxicogenomics-Based Test Systems:

ECVAM¹-ICCVAM²/NICEATM^{*} Considerations for Regulatory Use

Raffaella Corvi¹, Hans-Jürgen Ahr³, Silvio Albertini⁴, David H. Blakey⁵, Libero Clerici⁶, Sandra Coecke¹, George R. Douglas⁵, Laura Gribaldo¹, John P. Groten⁷, Bernd Haase¹⁸, Karen Hamernik¹⁹, Thomas Hartung¹, Tohru Inoue¹⁰, Ian Indans¹¹, Daniela Maurici¹, George Orphanides¹², Diana Rembges⁶, Susanna-Assunta Sansone¹³, Jason R. Snape¹⁴, Eisaku Toda¹⁵, Weida Tong¹⁶, Joost H. van Delft¹⁷, Brenda Weis¹⁸, Leonard M. Schechtman^{2,19}.

¹ECVAM – European Centre for the Validation of Alternative Methods, IHCP, JRC, Italy

²ICCVAM – US Interagency Coordinating Committee on the Validation of Alternative Methods

^{*}NICEATM – National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, NIEHS, Research Triangle Park, North Carolina, USA

³Bayer HealthCare AG, Germany; ⁴Hoffmann-La-Roche, Switzerland; ⁵Environmental Health Centre, Health Canada, Ottawa, Ontario, Canada; ⁶PCE - Physico-Chemical Exposure, IHCP, JRC, Italy;

⁷TNO, The Netherlands; ⁸Gene Logic, LTD; ⁹US Environmental Protection Agency, Washington, DC, USA; ¹⁰National Institute of Health Sciences, Tokyo, Japan; ¹¹Health Safety Executive, UK;

¹²Syngenta, UK; ¹³EMBL - The European Bioinformatics Institute, UK; ¹⁴AstraZeneca, UK; ¹⁵OECD, Paris, France; ¹⁶Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arkansas, USA; ¹⁷University of Maastricht, The Netherlands; ¹⁸National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA; ¹⁹Food and Drug Administration, National Center for Toxicological Research, Rockville, Maryland, USA.

Address correspondence to R. Corvi, European Centre for the Validation of Alternative Methods, IHCP, Joint Research Centre of the European Commission, Via E. Fermi 1, 21120 Ispra, Italy.

Telephone: +39 0332 785266. Fax: +39 0332 785845. E-mail: raffaella.corvi@jrc.it

Validation of Toxicogenomics-Based Tests

Key words

Acceptance, alternatives, biomarker, predictive test, regulatory use, standardization, toxicogenomics, toxicology, validation.

Abbreviations

European Centre for the Validation of Alternative Methods (ECVAM), European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL- EBI), Environment Protection Agency (EPA), Food and Drug Administration (U.S. FDA), Good Laboratory Practice (GLP), Good Manufacturing Practice (GMP), Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), International Life Sciences Institute Health and Environmental Sciences Institute (ILSI-HESI), MicroArray and Gene Expression (MAGE), Minimum Information About a Microarray Experiment (MIAME), Microarray Gene Expression Data Society (MGED), National Center for Toxicogenomics (NCT), NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Toxicology Program (NTP), Organization for Economic Co-operation and Development (OECD), Quality Assurance (QA), Quality Control (QC), Quantitative Structure-Activity Relationship (QSAR).

Acknowledgements

The authors gratefully acknowledge Dr. William S. Stokes, Director of the NICEATM, for his contributions to the planning and organization of the workshop as well as his insightful and constructive comments.

This document represents the consensus of the participants' views expressed as individual scientists, and does not necessarily represent the policies and procedures of their respective institutions. The authors declare they have no competing financial interests.

Descriptor: Toxicogenomics

Section Headers

Abstract

Introduction

Validation of Toxicogenomics: Focus on the Biological Systems

Strategy 1 - Use of Toxicogenomics Data to Define Mechanism and Identify Biomarkers

Strategy 2 - Use of Gene Expression Signatures to Predict Toxicity

Standardization/Validation of Toxicogenomics-Based Methods: Focus on the Technology

1. One-off Validation

2. Routine Validation

2.1. Biological Standards

2.2 Quality Assurance and Good Laboratory Practice

3. Validation as a Result of Procedural Changes

Data Management

Regulatory Acceptance of Validated Toxicogenomics-Based Methods

Potential Uses of Toxicogenomics Data in the Regulatory Area

Case for a Modular Approach to Validation

Standards for Comparison

Toxicogenomics Data from in vitro Systems and Data Relevance

Additional Regulatory Acceptance Issues

Conclusions

References

Tables

Figure Legends

Figures

Abstract

This is the report of the first workshop on the Validation of Toxicogenomics-based Test Systems held in Ispra, Italy on 11-12 December 2003. The workshop was hosted by the European Centre for the Validation of Alternative Methods (ECVAM) and organized jointly by ECVAM, the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The primary aim of the workshop was to discuss and define principles applicable to the validation of toxicogenomics platforms as well as validation of specific toxicological test methods that incorporate toxicogenomics technologies. The workshop was viewed as an opportunity for initiating a dialogue between technological experts, regulators, and the principal validation bodies, and for identifying those factors to which the validation process would be applicable. It was felt that to do so now, as the technology is evolving and associated challenges are identified, would serve as the basis for the future validation of the technology when it reached the appropriate stage. Due to the complexity of the issue, different aspects of the validation of toxicogenomics-based test methods were covered. The three focus areas included (a) biological validation of toxicogenomics-based test methods for regulatory decision-making, (b) technical and bioinformatics aspects related to validation, and (c) validation issues as they relate to regulatory acceptance and utilization of toxicogenomics-based test methods. This report summarizes the discussions and details the recommendations for future direction and priorities.

Introduction

Toxicogenomics, an emerging field in molecular toxicology, offers the promise of new approaches to identify and characterize such factors as the biological activity of new and existing chemicals and drugs and could play an important role in hazard assessment for human health. This revolutionary field can potentially affect many scientific and medical areas, including the development of a new generation of alternative predictive testing and screening methods that could lend themselves to the reduction, refinement and replacement of animals used for such purposes.

The European Centre for the Validation of Alternative Methods (ECVAM), the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) are currently investigating the specific considerations necessary for adequate validation of toxicogenomics-based test methods. The primary objective of ECVAM and ICCVAM-NICEATM is to facilitate development, validation, and regulatory acceptance of new, revised and alternative test methods that reduce, refine, and replace the use of animals (referred to as *the Three Rs*; Russell and Burch 1959) in testing while maintaining and promoting scientific quality and the protection of human health, animal health, and the environment. The efforts of such organizations as ICCVAM-NICEATM and ECVAM have helped foster the principles of the Three Rs and have contributed to progress in the use of alternative methods for regulatory, research, and educational purposes.

Experience in the validation of conventional alternative test methods has led to an understanding that new and innovative approaches will likely be necessary to standardize test methods based on toxicogenomics and to evaluate the scientific validity and regulatory applicability of such test methods. It is envisioned that the entire validation process will be more complex and challenging than that typically encountered thus far for other alternative test methods. This is because not only will the technology itself need to be standardized and validated, but also the methods that are based upon the technology and their predictive aspects will likewise need to undergo validation if they are to be employed in regulatory decision-making processes. In addition, the validation process will need to be able to accommodate the anticipated rapid changes in technology that could impact the performance of the test method and its reliability for a specific purpose.

Toxicogenomics-based methods, are being widely applied in toxicology and biomedical research. Since data are already being generated using these technologies, it is both timely and important to address the subject of validation now with the aim of establishing a foundation that will facilitate future regulatory acceptance of scientifically validated toxicogenomics-based test methods. By addressing the critical validation issues early on, and in parallel with the evolutionary and maturation phases of the technological development of toxicogenomics-based methods, it should be possible to pre-empt many potential pitfalls and data gaps encountered with retrospective method evaluations that could impede validation of this promising research and regulatory tool. Such a strategy will also facilitate early buy-in and confidence in the technologies by the regulatory arena in its quest for new, improved

and relevant methods by which to help ensure human health, protect the environment, and demonstrate responsiveness to animal welfare issues.

In consideration of all of these related issues, ECVAM and ICCVAM-NICEATM held the first of a planned series of workshops to address the validation principles that lend themselves to toxicogenomics-based test methods, e.g. gene expression technologies and associated bioinformatics. Given the complexity of the rapidly evolving toxicogenomics field, a variety of issues were addressed. These included, but were not limited to: (a) differences in and evolution of technology platforms including changes in genome coverage for model species, (b) quality assurance and Good Laboratory Practice (GLP) compliance, (c) technology standardization, transferability, and reproducibility, (d) relevance to *in vivo* biological responses, (e) yardsticks against which toxicogenomics responses should be measured, (f) data evaluation, statistical approaches and databases, (g) validation approaches, and (h) regulatory acceptability.

In order to begin to examine these complex issues, three Breakout Groups were formed. Each group concentrated on different aspects of the validation of toxicogenomics-based test methods, and the discussions were shared with the other participants in plenary sessions. The three focus areas included (a) biological validation of toxicogenomics-based test methods for regulatory decision-making, (b) technical and bioinformatics aspects related to validation, and (c) validation issues as they relate to regulatory acceptance and utilization of toxicogenomics-based test methods.

Validation of Toxicogenomics: Focus on the Biological Systems

The biological issues related to the validation of toxicogenomics-based test methods involved two strategies that were proposed for developing and validating such methods so that they can be employed to support regulatory decision-making. One strategy involves phenotypic anchoring of gene expression changes to identify molecular mechanisms and candidate biomarkers of toxicity (i.e. single genes, proteins or biological pathways). A second strategy involves the identification and validation of predictive gene expression signatures of toxicity. Validation considerations specific to data quality and cross-platform and inter-laboratory variability that are common to both strategies were identified. It is acknowledged that any new toxicogenomics-based methods will need to address established validation criteria for determination of reliability and relevance (Balls et al. 1995; ICCVAM 1997, 2003), as well as articulate the advantages and limitations of a given toxicogenomics-based test method. In addition, biological validation of such a test method, i.e. assessment of the concordance of gene changes with biological events, is essential but is contingent upon validation of the technology itself, which is addressed elsewhere in this paper.

Strategy 1- Use of Toxicogenomics Data to Define Mechanism and Identify

Biomarkers

Toxicogenomics offers the opportunity to enhance existing toxicity prediction strategies through elucidation of biological mechanisms around critical events. This sentiment is captured in the recent U.S. Environment Protection Agency (U.S. EPA)

and U.S. Food and Drug Administration (U.S. FDA) strategies regarding the inclusion of genomics data in submissions of regulated substances (U.S. EPA 2002; U.S. FDA 2003). While these agencies currently preclude basing regulatory decision-making on genomics data alone, they do encourage the voluntary submission of well-documented quality genomics data. Both agencies are considering the use of submitted data on a case-by-case basis for assessment purposes (e.g. to help elucidate mechanism of action or contribute to a weight-of-evidence approach) or for populating relevant comparative databases by encouraging parallel submissions of genomics data and traditional toxicological test results. This approach is appropriate given the state of scientific knowledge of toxicogenomics and the requisite need for a clear understanding of the toxicological relevance of the gene expression signals detected by this technology. There are a small but rapidly increasing number of published reports demonstrating a linkage between gene expression changes and adverse phenotypic changes (Huang et al. 2003; Orphanides 2003). These reports provide qualitative evidence of the power of genomics to link phenotype with gene expression, thereby contributing to an understanding of mechanism of action. Some such reports demonstrate the predictive power of these data to classify compounds. However, they fail to address adequately quantitative dose- and time-dependent (e.g. threshold) responses that are the hallmark of toxicological evaluation, making their immediate acceptance in regulatory arenas circumspect.

Nonetheless, toxicogenomics data may eventually have utility in hazard and risk assessment if data quality and validity can be adequately substantiated. Some regulators are finding that these data have the potential to add to the body of knowledge about compound mechanism of action. With appropriate dose- and time-

dependent measurements, gene and protein changes can be used to mark the molecular events that occur as an organism moves through the continuum from exposure to response. The obvious benefit is the identification of early markers of response, including responses that mark the point of departure from adaptation to toxicity. In addition, it may be possible to detect unforeseen effects at very low doses, or in unexpected tissues (Brown et al. 2002). This is important because changes in gene or protein expression alone are not sufficient to differentiate toxicity from biological adaptation following exposure to an exogenous compound. The challenge for predictive toxicology is to link changes in gene and protein expression to sequential changes in phenotype, both adaptive and adverse, in a manner that is consistent with the underlying biological mechanisms. For example, gene expression profiling has been used to classify hepatotoxins based on mechanism of action and to differentiate early, presumably adaptive, responses from later responses that are reflective of toxicity (Hamadeh et al. 2002a,b; Waring et al. 2001, Waring et al. 2003). The gene expression changes correlated well with changes in histopathology and clinical chemistry, supporting the liver as target organ for the test compounds.

While good technical progress has been made in recent years, additional proof-of-principle studies are needed for the regulatory community to become more accepting of the use of toxicogenomics data as part of the regulatory decision-making process. It would be important to demonstrate, for instance, that toxicogenomics can not only confirm what is already known about specific compounds and toxic endpoints (i.e. phenotypic anchoring), but also can accurately predict toxicity for unknown compounds. The task is to present regulatory scientists with new knowledge gained from toxicogenomics approaches in a familiar context. Ideally, at least in the short-

term, the focus will be the identification of single, or small sets of, genes or proteins that serve as biomarkers of response, as opposed to signatures of response that are the typical output of microarray experiments. Simple biomarkers of response are favored over complex expression signatures because they are familiar in toxicology assessment, are easy to maintain over time (e.g., are independent of the microarray platform) and can be readily validated. Validation strategies for toxicogenomics-based markers can be modeled after protocols for existing biomarkers. Thus, global gene expression technologies such as microarrays can be used to identify a specific gene marker, or a suite of markers, that can then be validated by conventional methods such as Northern analysis, *in situ* hybridization, and qPCR. This approach has advantages because regulatory agencies, such as U.S. FDA, have proposed procedures to address gene and protein biomarkers, and other organizations, such as the Organization for Economic Co-operation and Development (OECD; http://www.oecd.org/document/29/0,2340,en_2649_34373_34704669_1_1_1_1,00.html), are embarking on establishing similar guidance. (Supplemental Information, Section 1)

Proof-of-principle studies could be conducted concurrently with existing regulatory test methods using similar samples of test compounds. In such situations, it may be appropriate to use *in vivo* systems, which are widely accepted by the regulatory community. Parallel *in vitro* studies could be conducted in situations where an appropriate test system is available. It may be wise to focus initial efforts on defining relationships between gene expression changes and toxicity for individual compounds or compound classes with well-defined endpoints. The experimental design should address conventional aspects of dose and time (dose-response), species and strain

susceptibility, group size and sex, and selection of endpoints for study (e.g., histopathology, clinical chemistry). Numerous commercial microarray platforms offer genome-wide coverage for model systems such as rat, mouse, *C. elegans*, and humans. Commercial microarrays are also available for genes that are highly expressed in specific tissues (e.g., liver, breast) and during specific biological processes such as metabolism (e.g., P450 enzymes). Both genome-wide and dedicated arrays can be used with RNA samples from *in vivo* and *in vitro* (tissue and cell culture) systems, enabling parallel studies to be conducted with a single microarray platform. This is important because the results of microarray experiments can vary depending on the array design and the selection and performance of gene probes on the array. Encouraging results on cross-platform comparisons and between-laboratory reproducibility are now emerging (Bammler et al. 2005; Chu et al. 2005; Douglas et al 2004; Irizarry et al. 2005; Larkin et al. 2005). Toxicogenomics studies conducted in parallel and comparative systems can demonstrate the biological relevance of *in vitro* models as surrogates for *in vivo* models without the need to address cross-platform (technological) issues (Boess et al. 2003; Huang et al. 2003). While initial efforts should focus on defining simple gene and protein biomarkers for specific compound classes, endpoints, and model systems, the end goal is to establish a compendium of compound-specific knowledge that transcends technology platform. Ideally, the markers should be robust enough to withstand technological advances in toxicology that add to the existing knowledge about the compound. Once sufficient and adequately validated data are available, toxicogenomics can become part of a hierarchical approach to compound assessment.

The use of toxicogenomics to identify (screen) compounds with the potential to cause adverse effects may present opportunities to reduce the need for full animal tests, or perhaps refine animal use, and/or reduce the numbers of animals needed when *in vivo* tests are necessary. Of course, the statistical power of any test will influence the number of animals used in an *in vivo* test as well. Screening type assessments may be appropriate for priority setting, dose setting, chemical ranking, etc. The extent of validation required for screening tests may be different than that required for full replacement tests because negative compounds might still undergo a full animal testing. Establishing a compendium of compound-specific information will enable regulators and sponsors to access what is known about a compound across multiple test systems, species, and endpoints, thereby improving the biological relevance of regulatory decisions to safeguard human health and the environment.

Strategy 2 - Use of Gene Expression Signatures to Predict Toxicity

Toxicogenomics holds great promise for improving predictive toxicological assessments. Gene expression profiling has been used to classify compounds by chemical class and mechanism (Hughes et al. 2000; Scherf et al. 2000; Steiner et al 2004; Thomas et al. 2001), tumors by origin and type (Chung et al. 2002), and breast cancer patients for follow-up chemotherapy (van 't Veer et al. 2002). In all cases, classification was based on a set of discriminatory gene elements, between ten and several hundred, identified from a larger pool of genes on a microarray. The pattern of gene expression, not the measurement of a single or a small set of genes, was the basis for classification. A variety of gene expression analysis algorithms were used to discriminate samples based on gene expression signature. In all cases, the compound class or tumor status was known *a priori*, and gene expression signatures for known

samples were used to predict classification for other known but blinded samples (Blower et al. 2002; Brindle et al. 2002). Such models are currently being developed in the private sector (e.g., <http://www.genelogic.com>; <http://www.iconixpharm.com>) and are commercially available, but cannot, as yet, be exploited by regulators and the scientific community because the underlying data sets and algorithms have not been made available outside the private sector.

Predictive model development will require an extensive “training” set of gene expression measurements for classes of model compounds in a variety of test systems, both *in vivo* and *in vitro*, at multiple doses and time points. Initial studies can be conducted concurrently with conventional testing systems as a way to confirm model predictions. In the short term, it is unlikely that sufficient data will be available for gene expression signatures to replace conventional approaches. Until then, such data can be used as part of a hierarchical approach to toxicity testing in conjunction with accepted methods routinely used for regulatory purposes. In the long-term, sufficient data should accumulate from well-designed validation studies, such that gene expression signatures could be part of a battery of tests that reduce or replace animal procedures.

Model validation will necessitate multiple independent data sets and application of sophisticated statistical approaches. Acceptance of these models will require that research and regulatory communities have access to the data analysis tools used to build the models, and that they become familiar with the limitations and uncertainties of using these complex computational models. Confidence in and acceptance of these models will also require rigorous performance standards and appropriate controls to

ensure reproducibility and stability over time (see below), and adequate sensitivity and specificity to discriminate toxic from non-toxic responses. Initial model development could easily be accelerated through coordinated sector-spanning efforts. Coordinated efforts across academia, government and industry partnerships will accelerate progress in defining gene sets that are robust and discriminatory both within and across technology platforms. This is an ideal scenario given the rapidly advancing pace of technology development.

An important aspect of any toxicogenomics validation strategy is the need to measure the range of biological variability of gene responses for a given test system. Ideally, this should be accomplished by one species, tissue and endpoint at a time, in order to adequately assess cross-species differences that often hamper risk assessments. Measurements of biological variability under baseline and toxicant-challenged conditions will enable regulators to better discriminate biologically relevant responses from baseline homeostatic fluctuation. This is an important issue for toxicogenomics, as studies conducted on cell culture populations demonstrate a wide range of biological variability in gene expression measurements for individual cells under both baseline and challenged conditions (Kuang 2004). Therefore, it is necessary to define criteria to adequately address biological variability in a data submission and to establish whether the burden of maintaining these data is that of the regulator or sponsor.

The recommendations related to the biological validation of toxicogenomics-based test methods are listed in Table 1.

Standardization/Validation of Toxicogenomics-Based Methods: Focus on the Technology

Considerations given to validation of the technology encompassed the technical and bioinformatics issues related to the validation of toxicogenomics-based test methods. The starting premise adopted was that with the availability of bioinformatics expertise, biological data generated from toxicogenomics studies could be interpreted with a high degree of confidence. The ultimate aim was to identify a strategic approach that would enable credible biological observations and consequential judicious regulatory decisions, and that this approach would be independent of the toxicogenomic platform used. Moreover, standardization and validation of toxicogenomic platforms were seen as essential for identifying and reducing technological artifacts. Standardization would also be required to increase the certainty by which biological observations could be extrapolated across and between different microarray platforms. It is therefore important to build on the learning of previous and ongoing efforts in standardization of toxicogenomics (reviewed in Sansone et al. 2004).

Three distinct levels where validation is necessary were identified (see Figure 1 and discussion below). The first level of validation is the responsibility of the array manufacturer or provider and has to be performed only once. This can be seen as a “one-off validation” and it relates to both the microarray quality and the instrumentation. The second level of validation is the responsibility of both the experimental toxicologist and the array manufacturer or provider. This can be seen as “routine validation” or best practice to allow data comparability. It encompasses

quality control aspects of the critical experimental components and is a process that occurs on a regularly scheduled basis. The third level of validation, i.e. determination of reliability and relevance, is needed every time that a change is introduced into the test procedure. Performance standards developed based upon the original test method would serve as the criteria against which the revised method would be compared. Despite these multi-level validation needs, it was repeatedly emphasized that significant technological development and progress in microarray platforms is still underway and that efforts to validate and standardize these technological platforms must not be at the expense of innovation.

1. One-off Validation

The one-off validation is the responsibility of the array manufacturer or array provider. This is required to ensure that the array platform being used is robust and the inherent variability within the platform is transparent to the user and the regulator (Figure 1). The following were identified as being necessary for microarray-based toxicogenomics to be used in regulatory assessments:

- Microarrays should be fabricated in accordance with the principles of Good Manufacturing Practice (GMP).
- The specifications and performance criteria for all instrumentation and method components should be available.
- All quality assurance/quality control (QA/QC) procedures should be transparent, consistent, comparable, and reported.
- The array should have undergone sequence verification and the sequences should be publicly available.

- All data should be exportable in a MAGE (MicroArray and Gene Expression) compatible format.

2. Routine Validation

Routine validation is an on-going process that is the responsibility of the experimental toxicologist and the array manufacturer or provider (Rockett and Hellmann 2004).

Again, for microarray-based toxicogenomic assays to be used in regulatory decision-making the following important factors were identified (Figure 1):

- Oligos, cDNAs or clones that are arrayed should be randomly sequence-verified to ensure that no errors are introduced between batch syntheses. This verification process should be recorded and reported by the manufacturer.
- All reagent components should be identified. Reagents should be prepared according to GMP and/or GLP as appropriate. Data regarding batch variability should also be recorded and reported.
- Common reference RNA standards (house keeping genes) should be adopted to facilitate comparison between array platforms. This may be achieved in collaboration with the international Microarray Gene Expression Data (MGED) Society and other related efforts (see below).

2.1. Biological Standards

Performance standards, test component standards, and quality control measures are key components of any validation strategy for a toxicological test method.

Establishing standards is particularly important for gene expression technologies due to the inherent technological and biological “noise” in these systems. Commonly used

biological standards are reference RNAs that are competitively hybridized with the sample of interest in two-channel array formats, and *in vitro* RNA transcripts that are “spiked into” RNA samples of interest in either one-channel or two-channel array formats. Establishing accepted RNA standards will address concerns of regulatory reviewers about data quality and variability within and between laboratories and across different technology platforms. The standards will also provide a common benchmark for regulators to assess platform performance over time. To achieve this goal, any standards instituted must maintain a defined level of accuracy, sensitivity, specificity, and reproducibility across platforms.

Reference RNAs can be derived from tissue extracts, cell lines, or both and serve a variety of purposes. Workshops sponsored by governments and industry have focused on defining the specifications for reference RNAs for clinical and regulatory applications (<http://www.cstl.nist.gov/biotech/UniversalRNASTds/index.html>). There is a consensus that multiple RNA standards are needed to measure the accuracy, dynamic range, sensitivity and specificity of varied technology platforms under varied conditions. Important questions are whether regulatory agencies will define preferred sources of RNA standards, and, if so, who will generate and maintain baseline information about these standards. While the selection of a given RNA standard depends primarily on the purpose and application, all RNA standards should be tested for a clearly defined number of copies of a given sequence within an RNA preparation over some linear range (Cronin et al. 2004).

Some initiatives are raising awareness of the effects of variables that might hamper data comparability and are working towards developing best practice guidelines for

microarray-based measurements (www.mfbprog.org.uk). For example, recommendations for best practice in array normalization, together with performance characteristics in terms of sensitivity, accuracy and comparability of different array platforms (cDNA and oligo, spotted and in-situ synthesis) are beginning to emerge together with proposals for transparency and availability through publicly accessible databases (www.lgc.co.uk). Other initiatives are considering the use of quality metrics for standardizing and validating array-based toxicogenomics measurements. It remains to be seen as to the extent to which such efforts will be pursued and the impact they will have upon the standardization issues which are a necessary prerequisite to the validation exercises.

2.2 Quality Assurance and Good Laboratory Practice (GLP)

GLP is intended to promote proper documentation, quality and authenticity of toxicity test data and is required for data acceptance by regulatory agencies (e.g. U.S. FDA, U.S. EPA). At the international level, GLP has been promulgated under the OECD guidelines program (OECD 1998). As part of the progression towards regulatory acceptance, toxicogenomics experiments should ideally be conducted in accordance with GLP. However, at present, most large scale toxicogenomics efforts are not arising from GLP-compliant laboratories, and requiring compliance for data submission could greatly hamper the technical advancement of new technologies and retard their migration into the regulatory arena. In order to avoid discouraging technological progress while maintaining a level of GLP conformity, it could be argued that for research and technical development and improvement purposes it might be acceptable if array-based studies could at least measure up to the reporting standards required by GLP. However, with the adoption of the toxicogenomics-based

technologies into regulatory decision-making practices, GLP compliance undoubtedly will be expected. Procedural aspects of GLP compliance that are not currently captured in MIAME-Tox (Minimal Information About a Microarray Experiment – Toxicology) will need to be identified, but can be incorporated over time. Until then, it may be possible to allow for proof-of-principle and prevalidation studies to be conducted in accordance with the “intent” of GLP practices by requiring submitters to adequately document procedures and control measures and make experimental data open to regulatory review. “Best practices” for toxicogenomics can be established until formal procedures are adopted. This may be a more realistic solution that permits the advancement of science while addressing the need for quality assurance and control.

3. Validation as a Result of Procedural Changes

This third level of validation is necessary whenever a technical or methodological change is introduced into the test. Such changes might, on one hand, be restricted to the microarray technology (e.g. modification or addition of sequences to a microarray; changes in data analysis procedures, etc.). Alternatively, they could involve the experimental design (e.g. parameters such as dose, time, cell culture procedures, etc.). One consideration is that a distinction between minor and major procedural changes that might be incorporated into a test would help determine the extent of such validation necessary. Additionally, to facilitate the process, performance standards should be defined based upon the original validated test procedure. Minor changes would entail a demonstration of equivalence of results obtained with the modified test to that obtained from the validated test. Major changes would involve the need to define a new set of reference materials to be tested and a more extensive validation.

Guidance on the use of performance standards and the elements comprising them have been published (ICCVAM 2003) and have been employed for *in vitro* dermal corrosion assessment methods (ICCVAM 2004). Such guidance can likewise help facilitate the establishment performance standards for toxicogenomics-based test methods in which procedural modifications have been introduced after an initial validation exercise, thereby providing a basis for the comparison of reliability and accuracy of the modified method relative to the validated and accepted reference test method.

The concept of performance standards was originally developed to evaluate the acceptability (accuracy and reliability) of proposed test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect as an accepted (previously validated) test method. Since some regulatory authorities and international test guidelines programs (e.g. OECD) have restrictions regarding the use of proprietary test methods (methods that are copyrighted, trademarked or patented), performance standards also allow for the development and validation of comparable non-proprietary methods based on performance standards derived from the corresponding proprietary antecedent method. Under these circumstances, performance standards allow the characteristics and functional attributes of a proprietary method or technique to be described and offer a procedure for evaluating the performance of methods claimed to be substantially similar. A method that meets the established performance standards is considered sufficiently accurate and reliable for the specific testing purpose for which it is designed and is viewed as comparable to the original test method upon which it is based. If the correct performance standards have been developed, a method for which the results

have the same accuracy and reliability of the original one should by definition also be as relevant as the original method.

The conceptual framework and scope of performance standards could be expanded or adapted to include innovations or advancements in areas such as microarray or protein or metabolite separation and identification technology, where proposed improvements might or might not be generally or completely analogous to those in existing systems but would still enable similar applications. Performance standards could still provide a gauge for evaluating newer or revised technologies to ensure that their reliability and accuracy were at least comparable to that of existing acceptable techniques using similar chemicals even if essential test method components (i.e., structural, functional, and procedural elements of a validated test method to which a proposed, mechanistically and functionally similar test method should adhere) were not substantially similar.

This level of validation, which does not imply that a test needs to be completely re-validated, is of extreme importance for tests based on rapidly evolving technologies. It would be a mistake to immobilize these technologies by enforcement of a strict and inflexible validation approach, which would hamper progress and test improvement. Finally, a periodic reassessment of a test method's performance (accuracy and reliability) employing established performance standards would help to ensure adherence to essential test method components, and the reliability and accuracy of the modified test method relative to the validated antecedent method (Hartung et al. 2004). Such assurance could be best established and reported by international

validation bodies such as ECVAM and ICCVAM/NICEATM, which could track the history, performance and validation status of a given test.

Data Management

The lack of robust quality control procedures and capture of adequate metadata has caused problems with the analysis and reproducibility of array-based transcriptomics investigations. Consequently, the international MGED proposed standards for publication (Nature Editorial 2002) which were designed to clarify the MIAME guidelines (Brazma et al. 2001). As a result, a number of journals now require that articles containing microarray experiments must be compliant with the MIAME standard and some also require that the data integral to the paper's conclusions should be submitted to the ArrayExpress database at EBI (Brazma et al. 2003), GEO at NCBI (Edgar et al. 2002) and CIBEX at DDBJ (Ikeo et al. 2003), respectively the European, American and Japanese database counterparts.

There is a critical need for public toxicogenomics databases, due to the significant volume of data associated with these experiments, the complexity of comparing different gene annotations and splice variants across platforms, and the need for a resource for complex informatics analyses of the traditional toxicology and microarray data in parallel. However, to fully achieve the potential of this emerging interdisciplinary field, it is necessary to move toward the establishment of a common public infrastructure for exchanging toxicogenomics data (Mattes et al. 2004). The infrastructure should address (a) the technical problems involved in data upload, (b) the demand for standardizing data models and exchange formats, (c) the requirement for identifying minimal descriptors to represent the experiment, (d) the necessity of

defining parameters that assess and record data quality, and (e) the challenge of creating standardized nomenclature and ontologies to describe biological data. The goal is also to create an internationally compatible informatics platform integrating toxicology/pathology data with transcriptomics, providing the scientific community with easy access to integrated data in a structured standard format, facilitating data analysis and data comparison, and enhancing the impact of the individual datasets and the comprehension of the molecular basis of actions of drugs or toxicants. Ultimately, such a knowledge-base could be maintained (respecting confidentiality as appropriate) as a reference for regulatory organizations to evaluate toxicogenomics/pharmacogenomics data submitted by registrants to those organizations.

The potential exists for the international development of this public infrastructure. As part of the collaborative undertaking with the National Institute of Environmental Health Sciences (ILSI-HESI) Technical Committee on the Application of Genomics to Mechanism Based Risk Assessment, the EMBL-EBI (<http://www.ebi.ac.uk/microarray/Projects/tox-nutri/index.html>), the NIH-NIEHS NCT (<http://www.niehs.nih.gov/nct/>) and FDA NCTR (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/index.htm>) have worked closely together. The respective databases (Brazma et al. 2003; Tong et al. 2003; Waters et al. 2003), are based on the international standards developed by the MGED Society (<http://www.mged.org>). Following the very favorable response that the MIAME had received from the microarray community and key scientific journals (Ball et al. 2002, 2004; Nature Editorial 2002), the MIAME checklist was extended to describe array-based toxicogenomics experiments. The MIAME-Tox checklist

(<http://www.mged.org/Workgroups/rsbi>) is an attempt to define the *minimum* information required to interpret unambiguously and potentially reproduce and verify array-based toxicogenomics experiments. MIAME-Tox also supports a number of other objectives, for example, linking data from different experimental domains within a study, and linking several studies from one institution and exchanging toxicogenomics data sets among public databases. In fact, the major objective of MIAME-Tox is to guide the development of toxicogenomics databases and data management software. Without a sufficient depth of data in these resources, the scientific community's opportunity to develop consensus on analysis and application of these data for risk assessment or screening may be limited. The availability of this level of information regarding platform specification, appropriate common reference standards, and the toxicological study itself, will facilitate the predictive value of toxicogenomics across different array-based platforms. This, in turn, will result in a greater appreciation of and confidence in the value of toxicogenomics within a regulatory context, such that testing strategies can be optimized, predictive alternative models can be identified, and animal usage can be reduced (Supplemental Information, Section 2).

Moreover, the long-term provision of a MIAME-Tox compliant database with a MAGE-ML export is required for the long-term storage of toxicogenomics data. This would directly support the role of ECVAM, ICCVAM/NICEATM, and other validation bodies in the validation of toxicogenomics-based test methods.

The recommendations related to the technical and bioinformatics aspects of validation are listed in Table 2.

Regulatory Acceptance of Validated Toxicogenomics-Based Methods

Regulatory scientists are increasingly being called upon to consider incorporation of toxicogenomics data in regulatory assessment processes that involve evaluation of potential human health or environmental hazard and risk. Those scientists will need to be able to judge the level of confidence to place in both *in vivo* and *in vitro* toxicogenomics-based test methods and the resulting data that might be submitted in support of regulatory decision-making. Whether or not a method has been determined to be valid for a specific purpose will be an important factor for the consideration of its use for regulatory purposes. Furthermore, the level of confidence held by regulators will influence regulatory acceptance of methods and data, and will impact both the further pursuit of toxicogenomics technologies and technological improvements, and will affect the extent of industry application of these technologies.

Potential Uses of Toxicogenomics Data in the Regulatory Area

The potential of toxicogenomics-based methods in contributing to regulatory assessment processes is broad. Examples might involve, but would not be limited to, obtaining microarray data from individual *in vivo* bioassays or *in vitro* cell or tissue-based assays or from batteries of assays, using conventional or high-throughput approaches. In accordance with the current developing state of the science, realistic possibilities for initial uses of toxicogenomics data in regulatory settings might be first in the realm of hazard assessment, such as to support chemical mechanism of action arguments. Other early uses might include aiding individual chemical/chemical mixture screening or ranking exercises to set priorities for toxicity testing or to sort chemicals into batches. These types of applications might involve identification of

individual genes or gene patterns associated with particular toxic effects or pathways, adaptive responses, or metabolic pathways. However, global pattern recognition-type techniques are, as yet, not considered to be ready to fully replace traditional bio-analytical methods for predicting toxicity or elucidating information on mechanism of action or biochemical pathway component identification.

Using only human or animal *in vitro* or *in vivo* data derived from toxicogenomics technology to estimate such parameters as adverse/no adverse effect levels or to determine dose-response relationships for conducting risk assessments is regarded as a much longer-term goal. However, for hazard assessment purposes, the possibility of considering toxicogenomics data along with other types of toxicological information and data (e.g. from *in vivo* and *in vitro* studies, quantitative structure-activity relationship [QSAR] or SAR determinations, etc.) in a weight-of-evidence approach on a case-by-case basis was not discounted. Regulatory bodies have begun to craft preliminary proposals, policies and guidance for the submission and use of -omics-type data in regulatory deliberations and to provide encouragement for the use and further development of the technology (U.S. EPA 2002; U.S. FDA 2003).

Additionally, organizations such as the OECD are actively working with member countries on approaches that seek to harmonize the use of -omics-derived information for hazard assessment related to health and environmental effects.

Harmonization of toxicogenomics-based test methods will first necessitate the standardization and validation of the specific test protocol(s) developed for a specific purpose(s), as conducted by international validation bodies such as ECVAM and ICCVAM/NICEATM. It will then be important for such organizations to interface

with OECD in order to ensure the appropriate crafting of harmonized OECD toxicogenomics-based test guidelines that are based upon standardized, adequately validated procedures, and that are considered practical and permit consistent regulatory judgments.

Case for a Modular Approach to Validation

Because of the extraordinary rate at which toxicogenomics technologies are evolving, current validation processes might need to adapt so as to accommodate the rapidly developing changes and advancements while still observing the basic tried-and-true validation principles. To meet this anticipated need, a modular approach to validation (Hartung et al. 2004) was considered, not to abridge the process, but to allow for more flexibility in data collection and evaluation throughout the progressive changes that the technology will undergo. Typically, in the conventional validation procedures for an alternative test method, a sequential approach to the process is taken. The test protocol is first optimized and its transferability is determined. The resulting standardized method is then evaluated for within-lab and between-lab reproducibility and for its accuracy. Thus, an optimized, standardized protocol linked to specific test method elements and a prediction of outcome for given classes of chemicals, are evaluated together for performance characteristics and applicability. Such a linear validation model, although effectively employed for other test methods, might not be optimal for dynamic test methods in which changes are rapidly introduced that improve or alter the protocol or the technology incorporated in the protocol in any substantive way. The linear validation model might result in unnecessary delays in incorporating innovations into toxicogenomic-type test methods. In contrast, with a modular approach to validation, which capitalizes on the fundamental classical

concepts of validation as defined by ECVAM and ICCVAM (Balls et al. 1995, ICCVAM 1997, 2003), the different steps in the validation process are subdivided into independent modules, each of which can be assessed individually so that those components that have been completed need not undergo repeated validation. Further validation activities would, instead, be directed to only that part of the process flow where needed. The proposed model would accommodate validation of innovation affecting only a particular part of the sequence such that incorporation of advancements in a particular sector into testing strategies would less likely be impeded. At the same time, a modular approach to validation could efficiently handle information/data gaps that could be filled over time without derailing the validation stages already achieved. The modular approach complemented with the use of performance standards (discussed in section on Validation as a Result of Procedural Changes), are together, expected to facilitate and help expedite the validation of the toxicogenomics technology and test methods that are based on toxicogenomics.

The modular approach follows the fundamental classical concepts of validation as defined by ECVAM and ICCVAM. Validation is defined as the process by which the relevance and reliability of a test method for a specific purpose is determined (Balls et al. 1995; ICCVAM 1997, 2003). Adequate validation involves development of a standardized test method protocol, and assessment of the protocol's within- and between-laboratory variability, predictive capacity/accuracy, usefulness and limitations, and adherence to performance standards.

Standards for Comparison

As technological advancements are made and new, modified or revised toxicogenomics-type test methods are put forward for consideration, it will be necessary to have a means by which the performance of proposed methodologies can be compared with that of existing (traditional and non-traditional) methods, especially those that employ animals. The lack of an approach rooted firmly in high quality science could jeopardize attempts to seek or gain regulatory acceptance of toxicogenomics-based test methods and strategies. Evaluations of test method performance might be based on comparisons made between particular parameters, as dictated by the specific intent for which the assay was developed. Examples include:

- (a) *in vivo*–*in vivo* study comparisons so as to examine concordance of gene changes with such factors as onset, duration, severity, dose, age, possible temporal changes of effects, and species differences;
- (b) *in vitro*–*in vivo* study comparisons so as to explore gene changes associated with a critical event or end-point in an *in vitro* cell-based assay and an established *in vivo* biomarker of toxicity;
- (c) *in vitro*–*in vitro* study comparisons so as to analyze the responses of human and animal cell systems to xenobiotics;
- (d) technological comparisons so as to evaluate the effects of proposed technical improvements (e.g. comparing gene changes using different techniques of array/platform preparation).

Accordingly, in order to consider what type of validation activity and comparison are appropriate in a given situation, it is important that the specific purpose of the proposed methodology and a detailed description of all relevant procedures be clearly elaborated (Balls et al. 1995; Hartung et al. 2004; ICCVAM 1997, 2003).

Toxicogenomics Data from in vitro Systems and Data Relevance

At the present time, toxicogenomics data derived from *in vitro* systems have been considered to have limited utility in regulatory applications. However, a great deal of interest exists for the further development of *in vitro*-based toxicogenomics methods, for an examination of their potential applicability in the regulatory arena, and for an appraisal of their potential for contributing to improvements in animal welfare. It is anticipated that technological advancements will ultimately facilitate the utilization of *in vitro*-based methods as adjuncts to or surrogates for *in vivo*-based methods.

Possible areas where validated *in vitro*-based toxicogenomics test methods might play a future role include (a) preliminary assessments (pre-screens), (b) complementary testing that might assist in obtaining additional (e.g. mechanistic) information, and (c) surrogate tests that could help in the refinement, reduction and replacement of animals used for -omics-based or traditional testing methods. One exciting aspect of toxicogenomics technology is the prospect of being able to identify species differences and/or similarities in the response to a xenobiotic. Although this is not viewed as near-term prospect, it obviously has potential applications for hazard and risk assessment purposes and could also have an impact on previous regulatory decisions, when the technology becomes sufficiently advanced to permit such uses for it.

Additional Regulatory Acceptance Issues

In considering approaches to validation, achieving regulatory acceptance of toxicogenomic-based methods or acceptance of information/data derived from such methods is an important goal. Regulators will be asked to evaluate whether data

submitted using -omics technologies can be used in support of a particular or broader-based toxicological, pharmacological or physiological premise. For example, experiments using microarrays demonstrated increased expression of a cluster of related genes which was associated with enhanced activity and production of a microsomal enzyme important in the metabolic activation of a chemical to a toxic entity, which in turn was associated with a histopathological biomarker lesion in the liver with a known human cancer correlate. Each of the events in this example can be thought of as a sequence of separate critical steps or information levels (Figure 2) which progressively connect -omics data (from microarrays), to gene expression changes (increased expression), to a biochemical pathway (liver enzyme induction leading to toxic metabolite formation), to a toxicological effect *in vivo* (liver lesion) with human relevance (cancer). Moving between two levels involves a prediction of outcome linking both steps. At each of these prediction junctures, regulators would be looking for evidence to scientifically substantiate moving to the next step and whether or not the prediction linking the levels (e.g. in this example, prediction 1, 2, 3, or 4 in Figure 2) was adequately validated. Theoretically, with this type of system, validated links could be established between any two levels. Technological advancements or new information could be independently incorporated into a given level and considered/evaluated for the specific relevant prediction juncture. In this way, each of the prediction levels can be assessed independently and the validity of the links determined.

In the future, toxicogenomics-based test methods may be shown to have been adequately validated and technically suitable for certain specific purposes, but regulatory acceptability and implementation will depend, in part, on whether the

methods validated have utility to a given regulatory agency or program, i.e. are applicable to the products that fall within its regulatory purview. Some regulatory bodies may have internal peer review processes, particular regulatory mandates, and/or regulatory assessment procedures that also have a role in the determination of test method applicability in regulatory programs, even though a test method may have been appropriately validated.

The widespread use of –omics technologies will also bring about increasing demands on the regulatory community in terms of training of regulatory personnel in areas such as potential applications, data quality control, analysis and interpretation, statistical analysis, limitations of the technology and how the information might be incorporated into safety/hazard/risk assessment processes. To satisfy these needs, regulatory agencies have been engaging in developing and implementing training procedures, hiring scientists with the necessary technical knowledge and experience, establishing centers of excellence and dedicated laboratories focused specifically on the various –omics and related informatics areas (e.g. National Center for Toxicological Research [FDA]; National Center for Toxicogenomics [NIEHS], Minister of Health Labour and Welfare – National Institute of Health Sciences Project in Japan, Netherlands Genomics Initiative, EMBL-EBI where informatic scientists are working with experimental practitioners and MGED to ensure that transcriptomic experiments can be mapped on to regulatory toxicology studies). In addition, the regulatory arena has found that maintenance of open lines of communication with appropriate external scientists facilitates cooperation and the sharing of technical aspects, skills and practical experiences that help to broaden the collective knowledge-base. In any case, as the technology evolves further and finds wider application and acceptance, it will

be necessary to address such fundamental matters as (a) the generation, management, and interpretation of massive amounts of data, (b) the consequent complex questions that will undoubtedly arise (e.g. What constitutes an adverse effect as identified using the technology? How does a given gene pattern correlate with a particular toxic endpoint or relate to onset, duration, and severity of effects, age, dose and species?), and (c) the limitations to the technology. To address such issues most efficiently will warrant an ongoing dialogue between regulators and practitioners and a willingness to share relevant experiential and theoretical knowledge. Standard submission and presentation formats compatible with electronic data submission would likely need to be developed. Programs and staff would need to learn how information from the new technologies might be incorporated in regulatory practices and decision-making processes, and would also have to face dealing with possible incongruities between toxicogenomics-derived data and existing or future submissions of conventional toxicity data. A number of regulatory authorities have already begun to contemplate and make provisions for this enormous and challenging task, but others may not yet have committed the resources to do so.

The recommendations related to regulatory acceptance and utilization of toxicogenomics-based test methods are listed in Table 3.

Conclusions

This workshop was organized as a result of the rapid growth and technological advancements in the field of toxicogenomics, the promise it offers for numerous scientific arenas, especially human health and the environment, and the interest demonstrated by regulatory agencies as well as the industrial sector. As a result, it has become apparent that a considerable effort will need to be invested in the appropriate validation of both the technology itself and those test methods that incorporate the technology. The workshop provided a platform for technical experts in the field to become cognizant of the validation principles and regulatory issues to be encountered and for regulators and principal validation bodies to gain a better sense of those technological aspects that would lend themselves to standardization, harmonization, and validation. Thus, this workshop served as an important initiative that fostered an exchange of information fundamental to the ultimate adoption of toxicogenomics-based test methods for regulatory decision-making purposes. It is envisioned that the conclusions and recommendations that resulted will serve as a basis for future validation considerations for test method applications of toxicogenomics technologies in the regulatory arena and evaluating their potential utility for hazard/safety/risk assessments.

The following lists several aspects of the validation of toxicogenomics that were identified as needing further exploration in order to help facilitate regulatory acceptance of future toxicogenomics-based test methods.

- Conduct toxicogenomics-based tests and the associated conventional toxicological tests in parallel in order to: 1) generate comparative data supportive of the utilization of the former in place of the latter or, 2) provide relevant mechanistic data to help define the biological relevance of such responses within a toxicological context.
- Determine and understand the range of biological and technical variability between experiments and between laboratories and ways to bring about greater reproducibility.
- In the short-term favor defined biomarkers that are independent from technology platform, and therefore are easier to validate. But in the longer-term focus on pathway analysis (i.e. system biology approach), rather than just on individual genes.
- Harmonize reference materials, quality control measures, and data standards, and develop compatible databases and informatics platforms that are key components of any validation strategy for a toxicological method. This can only be achieved by promoting partnerships and collaborations among ongoing initiatives in toxicogenomics, standardization and validation.
- Determine performance standards for toxicogenomics-based test methods that will serve as the yardsticks for comparable test methods that are based on similar operational properties.
- Define further the modular validation scheme that would allow keeping up with methodological improvements and innovations without having to repeat the entire validation process, but would however, integrate ECVAM and ICCVAM principles of validation and acceptance.

References

Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, et al.

2002. Standards for microarray data. *Science* 298:539; *Bioinformatics*

18(11):1409; *The Lancet* 360:1019.

Ball C, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, et al.; MGED

Society. 2004. Standards for microarray data: an open letter. *Environ Health*

Perspect 112(12):A666-667.

Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, et al. 1995.

Practical aspects of validation of toxicity test procedures. *Alta* 23:129-147.

Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, et al.

2005. Standardizing global gene expression analysis between laboratories and

across platforms. *Nature Methods* 2:351-356.

Blower PE, Yang C, Fligner MA, Verducci JS, Yu L, Richman S, et al. 2002.

Pharmacogenomic analysis: correlating molecular substructure classes with

microarray gene expression data. *Pharmacogenomics J* 2(4): 259-271.

Boess F, Kamber M, Romers S, Gassei R, Muller D, Albertini S, et al. 2003. Gene

expression in two hepatic cell lines, cultured primary hepatocytes and liver

slices, compared to the in vivo liver gene expression in rats: possible

implications for toxicogenomics use in in vitro systems. *Toxicol Sci* 73(2):386-

402.

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(4):365-371.

- Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. 2003. ArrayExpress- a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31:68-71.
- Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, et al. 2002. Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using ¹H-NMR-based metabolomics. *Nature Medicine* 12:1439-1445.
- Brown HR, Ni H, Benavides G, Yoon L, Hyder K, Giridhar J, et al. 2002. Correlation of simultaneous differential gene expression in the blood and heart with known mechanisms of adriamycin-induced cardiomyopathy in the rat. *Toxicol Pathol* 30(4):452-469.
- Chu TM, Deng S, Wolfinger R, Paules RS, Hamadeh HK. 2004. Cross-site comparison of gene expression data reveals high similarity. *Environ Health Perspectives* 112:449-455.
- Chung CH, Bernard PS, Perou CM. 2002. Molecular portraits and the family tree of cancer. *Nature Genetics Suppl* 32:533-540.
- Cronin M, Ghosh K, Sistare F, Quackenbush J, Vilker V, O'Donnell C. 2004. Universal standards for gene expression. *Clinical Chemistry* 50(8):1464-1471.
- Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207-210.
- Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. 2002a. Gene expression reveals chemical-specific profiles. *Toxicol Sci* 67:219-231.
- Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. 2002b. Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* 67:232-240.

- Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, et al. 2004. A modular approach to the ECVAM principles on test validity. *ATLA* 32:467-472.
- Huang E, Ishida S, Pittmann J, Dressman H, Bild A, Kloos M, et al. 2003. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nature Genetics* 34(2):226-230.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. 2000. Functional Discovery via a compendium of expression profiles. *Cell* 102:109-126.
- ICCVAM. 1997. ICCVAM Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods, NIH Publication No: 97-3981. Available: <http://iccvam.niehs.nih.gov> [accessed July 2005].
- ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods, NIH Publication No. 03-4508. Available: <http://iccvam.niehs.nih.gov> [accessed July 2005].
- ICCVAM. 2004. ICCVAM Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion, NIH Publication No. 04-4509. Available: <http://iccvam.niehs.nih.gov> [accessed July 2005].
- Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y. 2003. CIBEX: center for information biology gene expression database. *C R Biol.* 2003 326 (10-11):1079-1082.
- Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, et al. 2005. Multiple-laboratory comparison of microarray platforms. *Nature Methods* 2:345-349.

- Kuang Y, Biran I, Walt DR. 2004. Simultaneously monitoring gene expression kinetics and genetic noise in single cells by optical well arrays. *Anal Chem* 76(21):6282-6286.
- Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. 2005. Independence and reproducibility across microarray platforms. *Nature Methods* 2:337-343.
- Mattes WB, Pettit SD, Sansone SA, Bushel PR, Waters MD. 2004. Database development in toxicogenomics: issues and efforts. *Environ Health Perspect Toxicogenomics*:doi:10.1289/txg.6697
- Nature Editorial. 2002. Microarray standards at last. *Nature* 419:323.
- OECD. 1998. OECD series on principles of good laboratory practice and compliance monitoring. Number1
http://www.oecd.org/document/63/0,2340,en_2649_34381_2346175_1_1_1_1.0.html [accessed July 2005].
- Orphanides G. 2003. Toxicogenomics: challenges and opportunities. *Toxicol Lett* 140-141:145-148.
- Rockett JC and Hellmann GM. 2004. Confirming microarray data – is it really necessary? *Genomics* 83:541-549.
- Russell WMS, Burch RL. 1959. *The Principles of Humane Experimental Technique*. London, UK: Methuen. Available:
http://altweb.jhsph.edu/publications/humane_exp/het-toc.htm.
- Sansone SA, Morrison N, Rocca-Serra P, Fostel J. 2004. Standardization initiatives in the (eco)toxicogenomics domain: a review. *Comp Funct Genom* 5:633-641; doi: 10.1002/cfg.447

- Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. 2000. A gene expression database for the molecular pharmacology of cancer. *Nature Gen* 24:236-244.
- Steiner G, Suter L, Boess F, Gasser R, de Vera MC, Albertini S, et al. 2004. Discriminating different classes of toxicants by transcript profiling. *Environ Health Perspect Toxicogenomics* 112(12):1236-1248.
- Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, et al. 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* 60(6):1189-94.
- Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, et al. 2003. ArrayTrack-- supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ Health Perspect Toxicogenomics* 111:1819-1826.
- U.S. EPA. 2002. Interim Genomics Policy, U.S. Environmental Protection Agency Science Policy Council. Available: <http://epa.gov/osa/spc/html/genomics.htm> [accessed July 2005].
- U.S. FDA. 2005. FDA Guidance for Industry-Pharmacogenomic [PG] Data Submissions (Final: March 2005) Available: www.fda.gov/cder/guidance/6400fnl.htm [accessed July 2005].
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530-536.
- Waring JF, Cavet G, Jolly RA, McDowell J, Dai H, Ciurlionis R, et al. 2003. Development of a DNA microarray for toxicology based on hepatotoxin-regulated sequences. *Environ Health Perspect Toxicogenomics* 111:53-60.

- Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, et al. 2001. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* 175(1):28-42.
- Waters MD, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, et al. 2003. Systems toxicology and the chemical effects in biological systems knowledge base. *Environ Health Perspect Toxicogenomics* 111:811-824.
- Yauk CL, Berndt ML, Williams A, Douglas GR. 2004 Comprehensive comparison of six microarray technologies. *Nucleic Acid Res.* 32(15):e124. doi: 10.1093/nar/gnh123

Table 1 Recommendations: Focus on Biological Systems

1	Encourage increased usage of toxicogenomics-based approaches to define the mechanistic context of toxic responses to exogenous compounds.
2	Promote greater understanding of the relationships between gene expression responses and altered phenotype, considering the biological pathways affected, dose-response and the point of departure from adaptive to toxic response.
3	Favor the identification of biomarkers that are independent of technology platform, but acknowledge the potential strengths of pathway analysis.
4	Characterize the range and extent of biological variability of responses for the test systems (e.g., diurnal effects, animal care and use, age-related context).
5	Encourage the immediate usage of toxicogenomics-based approaches in conjunction with conventional toxicity testing approaches.
6	Explore the extent to which toxicogenomics can address cross-species responses and specific disease states.
7	Promote the conduct of parallel and comparative <i>in vivo</i> and <i>in vitro</i> studies to identify <i>in vitro</i> systems that can serve as surrogates for <i>in vivo</i> systems.
8	Characterize predictive toxicology models with respect to parameters such as dose, time, study design, relevance; characterize the system to fulfill validation criteria.
9	Promote the identification of gene and protein biomarkers as early (prognostic) markers as a refinement to existing toxicity testing methods.
10	Establish a compendium of toxicant information based on gene expression responses for model compounds across multiple species, endpoints and test systems.
11	Foster the development of effective partnerships between academic, government and industry groups to promote collaborative efforts to validate toxicogenomics-based test methods and generate sufficient high quality data to support regulatory decision-making.

Table 2 Recommendations: Focus on Technology

<i>1</i>	Validation and QA/QC should be mandatory during the manufacturing of the arrays.
<i>2</i>	The array should undergo sequence verification and sequences should be available in the public domain.
<i>3</i>	MIAME guidelines should be adhered to.
<i>4</i>	Initially, develop “best practices” for toxicogenomics, including the interpretation of data and how to manage uncertainties and limitations.
<i>5</i>	Subsequent development of guidance for and adherence to GLPs for toxicogenomics experiments.
<i>6</i>	Common reference standards should be considered.
<i>7</i>	A workshop should be convened to address the development of standards for RNA sample preparation (and other biological aspects of microarray analyses).
<i>8</i>	Need for a “common” RNA standard including developing consensus about sources and maintenance of baseline data for regulatory and research purposes.
<i>9</i>	Studies should be MIAME-Tox compliant.
<i>10</i>	Performance Standards should be developed and implemented to evaluate reliability and accuracy of test methods incorporating procedural modifications.
<i>11</i>	An ongoing dialogue should be maintained between scientists in the various relevant disciplines including bioinformaticians; through meetings, published papers, and advisory/discussion panels (e.g. ILSI-HESI committee, NCT consortium, OECD panel).
<i>12</i>	Ensure that validation efforts and QA/QC criteria are not restrictive to the technology or its advancement.
<i>13</i>	Explore whether toxicogenomics measurements can define toxicological effects quantitatively.
<i>14</i>	Develop prediction models (e.g. algorithms) for toxicogenomics-based test methods.
<i>15</i>	Develop a data infrastructure for capturing, storing, and reporting toxicogenomics data.
<i>16</i>	Ensure continuation of financial support for long-term public database maintenance.

Table 3 Recommendations: Focus on Regulatory Acceptance of Toxicogenomics-Based Methods

1	Build on and/or learn from previous and ongoing efforts in toxicogenomics, standardization, validation, and harmonization efforts where possible (e.g. MIAME, ICCVAM, ECVAM, NCT, EMBL-EBI, ILSI/HESI, U.S. FDA, U.S. EPA, OECD).
2	Fund pilot programs to test possible validation strategies and processes.
3	Identify training needs and assist in developing training vehicles and ways of presenting the state-of-the-science to regulators and the regulated community (including electronic means).
4	Maintain transparency of validation processes.
5	Explore additions/amendments/revisions to ICCVAM and ECVAM validation guidance that would accommodate new and rapidly changing technologies.
6	Implement the modular approach to validation in order to accommodate existing knowledge and future technical developments.
7	Establish performance standards for toxicogenomics-based test methods and have them accommodate rapid technological advancements and procedural modifications.
8	Explore/develop/support sector-spanning world-wide harmonization entities.
9	Create confidence among regulators by involving them early on in discussions and various scientific fora that would facilitate application of the technology for regulatory purposes.
10	Encourage industry and other parties to share data, in part, to support validation comparisons.
11	Promote high quality science in supporting the use and development of the technology for regulatory purposes to further protection of human health and the environment.
12	Consider opportunities for synergy between QSAR, pharmacokinetic and pharmacodynamic (PKPD) modeling and other <i>in silico</i> efforts and the toxicogenomics communities.

Figure Legends

Figure 1. Scheme of the different steps in a toxicogenomics-based test. Three distinct levels were identified where validation is necessary. 1) “One-off validation”, which should be performed once and is mainly related with the quality of the microarray and the instrumentation (blue). 2) “Routine validation” and quality control representing the on-going requirements which are the responsibilities of the experimental toxicologist and the manufacturer (red). 3) The extent of validation necessary whenever a technical or methodological change is introduced in the test. A method should meet the pre-established performance standards in order to be considered reliable and relevant as the original test method (green).

Figure 2. Process flow showing different independent prediction levels considered important in assessing validity of a toxicogenomics-based test method.

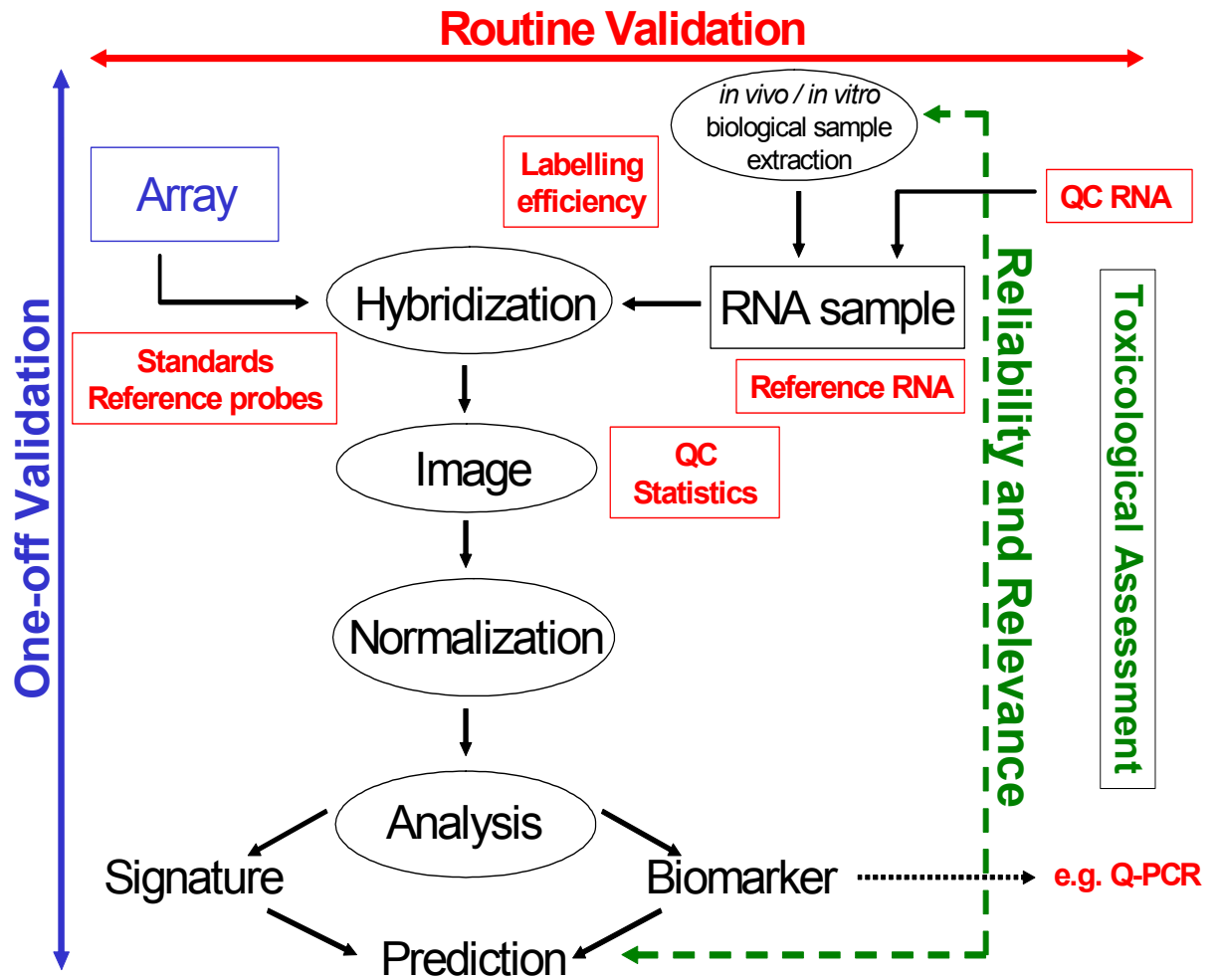


Figure 1

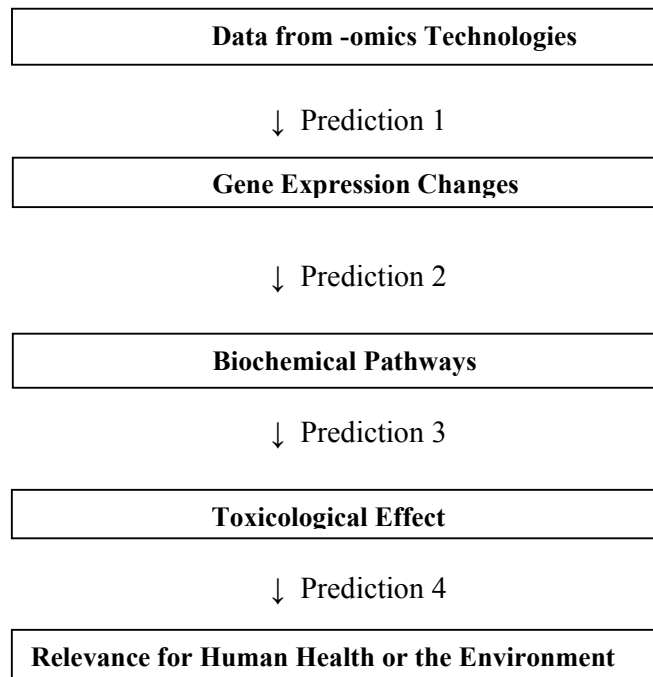


Figure 2