Integrating Omic Technologies into Aquatic Ecological Risk Assessment and Environmental Monitoring: Hurdles, Achievements and Future Outlook.

Graham Van Aggelen\textsuperscript{1}, Gerald T. Ankley\textsuperscript{2}, William S. Baldwin\textsuperscript{3}, Daniel W. Bearden\textsuperscript{4}, William H. Benson\textsuperscript{5}, J. Kevin Chipman\textsuperscript{6*}, Tim W. Collette\textsuperscript{7}, John A. Craft\textsuperscript{8}, Nancy D. Denslow\textsuperscript{9}, Michael R. Embry\textsuperscript{10}, Francesco Falciani\textsuperscript{6}, Stephen G. George\textsuperscript{11}, Caren C. Helbing\textsuperscript{12}, Paul F. Hoekstra\textsuperscript{13}, Taisen Iguchi\textsuperscript{14}, Yoshi Kagami\textsuperscript{15}, Ioanna Katsiadaki\textsuperscript{16}, Peter Kille\textsuperscript{17}, Li Liu\textsuperscript{9}, Peter G. Lord\textsuperscript{18}, Terry McIntyre\textsuperscript{1}, Anne O’Neill\textsuperscript{19}, Heather Osachoff\textsuperscript{3}, Ed J. Perkins\textsuperscript{20}, Eduarda M. Santos\textsuperscript{21}, Rachel C. Skirrow\textsuperscript{1}, Jason R. Snape\textsuperscript{22}, Charles R. Tyler\textsuperscript{21}, Don Versteeg\textsuperscript{23}, Mark R. Viant\textsuperscript{6}, David C. Volz\textsuperscript{24}, Tim D. Williams\textsuperscript{6} and Lorraine Yu\textsuperscript{1}

\textsuperscript{1} Environment Canada, Pacific Environmental Science Centre, Vancouver, BC, Canada
\textsuperscript{2} US Environmental Protection Agency, Duluth, MN, USA
\textsuperscript{3} Clemson University, Clemson, SC, USA
\textsuperscript{4} NIST, Hollings Marine Laboratory, Charleston, SC, USA
\textsuperscript{5} US Environmental Protection Agency, Gulf Ecology Division, Gulf Breeze, FL, USA
\textsuperscript{6} The University of Birmingham, Birmingham, West Midlands, UK
\textsuperscript{7} US Environmental Protection Agency, Athens, GA, USA
\textsuperscript{8} Glasgow Caledonian University, Glasgow, Scotland, UK
\textsuperscript{9} University of Florida, Gainesville, FL, USA
* Corresponding Author

J.K. Chipman

School of Biosciences, The University of Birmingham, Birmingham, B15 2TT, UK

Tel: +0044 121 414 5422

Fax: +0044 121 414 5925

Email: j.k.chipman@bham.ac.uk
Running Title
Omics, Pollutants and the Environment

Key Words
environment, environmental monitoring, fish, metabolomics, microarray, regulatory toxicology, transcriptomics

Acknowledgements
The authors thank staff at the Pacific Environmental Science Centre, Vancouver, Canada and Marilyn Cruz for hosting and co-organizing the workshop and the UK NERC for funding (Grant NE/D000793/1).

Competing Interests Declaration
A list of affiliations and potential financial interests has been submitted

List of Abbreviations
QA- quality assurance
NMR- nuclear magnetic resonance
PCR – polymerase chain reaction
SETAC - Society of Environmental Toxicology and Chemistry
US EPA – US Environmental Protection Agency
NERC - Natural Environment Research Council
GLP- good laboratory practice
Abstract

Background: In this commentary we present the findings from an international consortium on fish toxicogenomics sponsored by the UK Natural Environment Research Council (NERC) with
a remit of moving omic technologies into chemical risk assessment and environmental monitoring.

Objectives: The consortium from government agencies, academia and industry addressed three topics; progress in ecotoxicogenomics; regulatory perspectives on ‘roadblocks’ for practical implementation of toxicogenomics into risk assessment; and dealing with variability in datasets.

Discussion: Key points were: Examples of successful application of omic technologies have been identified. There remain critical studies to relate molecular changes to ecological adverse outcome. Recommendations were made for the management of technical and biological variation. The need for enhanced interdisciplinary training and communication. The need for considerable investment into the generation and curation of appropriate reference omic data.

Conclusions: The participants concluded that, while there are ‘hurdles’ to pass on the road to regulatory acceptance, omics technologies are already useful for elucidating modes of action of toxicants and can contribute to the risk assessment process as part of a ‘weight of evidence’ approach.
START OF MANUSCRIPT

The remit of this consortium was to assess current developments towards the incorporation of omic technologies into environmental risk assessment and environmental monitoring, particularly in relation to aquatic ecotoxicogenomics. Participants recognized that omic tools and associated endpoints are already significantly improving our understanding of how individual chemicals and mixtures affect organisms and could ultimately influence risk assessment and environmental management. While a significant amount of basic research and validation needs to be done before omic endpoints are incorporated as complementary data for routine assessments of environmental risk, it was generally agreed that there are no “roadblocks” for omics technology per se, but “hurdles” along the road of discovery, acceptance, and implementation of omic endpoints. Given the context of the workshop, the following quote is noteworthy: “the successful incorporation of toxicogenomics into regulatory frameworks may someday be regarded as the most important intellectual and practical contribution from this generation of ecotoxicologists” (Ankley et al. 2006).

Benefits and Successful Applications of Omics in Ecotoxicology and Ecological Risk Assessments

Historical Challenges and Recent Developments for Regulatory Implementation. Previous publications and workshops (e.g. Ankley et al. 2006; Boverhof and Zacharewski 2006) have discussed the potential application of omic technologies to risk assessment. The use of omic technology in toxicology (toxicogenomics) was initiated following the development of the first
high density techniques (microarrays). However, excitement surrounding this new technology generated “hype” which yielded unrealistic expectations of the timeline for incorporation into risk assessment. There is now a more realistic understanding of the potential contribution of omics to toxicology (NRC 2007). It is realized that a multi-level, systems biology approach to safety assessment, combining molecular- (including mRNA, protein and metabolites), cellular-, tissue-, individual- and population-level data represents a powerful new multidisciplinary approach that identifies biomarkers with much improved predictive capacity.

A number of initial concerns and difficulties have been overcome. The high cost of microarrays imposed severe restrictions on the number of doses, replicates and time points assessed following chemical administration to biological systems. As a consequence, reported omic responses frequently reflected pathological change with no evident predictive value. Methodology has now been improved, costs reduced, and microarrays are commercially available for a range of species. In the context of fish transcriptomics, a special issue of the Journal of Fish Biology (Miller and Maclean 2008) illustrates the progress that has been made with non-model organisms.

**Successful Prospective and Diagnostic Case Studies.** Transcriptomic experiments in aquatic toxicology have been diverse, encompassing different microarray platforms, test species and exposure routes, emphasising their use as “case studies” rather than standardized tools. This wide-ranging approach contributes to the elucidation of mechanisms of toxicity, including; dose-response relationships, differential species-sensitivity, and classification of chemical–specific biological responses. This approach also provides leads for identification of novel biomarkers of exposure and adverse effect.
Omic and bioinformatic tools offer substantial promise for discovery of gene, protein and/or metabolite alterations indicative of the mode of action (MOA) of chemicals and improved understanding of mechanisms in prospective studies (Ankley et al. 2006). Knowing MOA can reduce uncertainties in chemical risk assessments providing, for example, a basis for extrapolating effects across species (Benson and Di Giulio 2007). There is ongoing debate as to the appropriate role of biomarker data in ecological risk assessments (Forbes et al. 2006). Historically, most biomarker data employed in ecotoxicology was indicative of exposure, but had limited prediction of deleterious effects meaningful to risk assessment, namely; survival, growth and development, and reproduction (Forbes et al. 2006). This is largely due to a lack of mechanistic knowledge concerning linkages between molecular alterations and outcomes in the whole organism. Ideally, omics data would reflect both MOA and deleterious outcome(s). To achieve this, it is necessary to define the cascade of pathways associated with toxicity; from a molecular initiating event (e.g. receptor-binding) through subsequent biological alterations (reflected by omic and cellular changes) that culminate in a deleterious outcome (NRC 2007). However, as indicated below, there is also potential for a contribution to understanding ecological impacts. Furthermore, omic approaches can contribute to the reduction of animal usage and of severity of treatments, as more subtle changes can be identified and a more complete assessment of the health of individual animals or cell cultures can be achieved.

A recent example of how a toxicity pathway approach can be used to establish quantitative linkages across biological levels of organization was provided by Miller et al. (2007) and Ankley et al. (2008a), who investigated the consequences of molecular changes in the fish hypothalamic-
pituitary-gonadal (HPG) axis in terms of reproductive and population effects. Production of vitellogenin (VTG; an oocyte lipoprotein produced in the liver of oviparous female vertebrates) can be affected by a range of signaling events that alter steroid hormone production and activity. Analysis of an integrated dataset derived from fathead minnow (*Pimephales promelas*) reproduction studies, with five chemicals that decrease VTG and fecundity, but affect the HPG axis through different discrete mechanisms, demonstrated robust associations between steroid and VTG concentrations in female fish. This was predictive of egg production and, via modeling, could be used to tentatively forecast fathead minnow population status. Thus, through understanding the biological pathways leading to vitellogenesis, mechanistic molecular responses were successfully related to potential adverse outcomes meaningful to risk assessors.

Omics data can be used in diagnostic studies to determine the efficacy of pollution remediation as part of a weight-of-evidence approach (Roling et al. 2007). Furthermore, omic profiling can be used to identify chemical causation of effects induced by complex mixtures (Garcia-Reyero et al. 2008a). For example, in studies with the fathead minnow Filby et al. (2007) applied multiple quantitative PCR to identify diagnostic signatures from different chemicals that induce the same phenotypic effects. There were common features in the responses of fish exposed to estrogenic waste-water effluent and individual steroid estrogens. These data demonstrate that patterns of gene expression induced by estrogenic effluents, although complex, can be diagnostic for some of the estrogens they contain and could be used by regulators to determine the primary contaminant.

**The Utility of Omics in Studies on Environmental Sites**
The influence of the local environment on an organism’s transcriptome or metabolome can be exploited in environmental monitoring to characterise the effects of anthropogenic stressors such as pollution. In European flounder (*Platichthys flesus*) (Falciani et al. 2008), transcriptomics and genetic algorithm bioinformatic approaches were used to predict the site of origin of fish from the environment based on stress-responsive genes. Thus, although gene expression is affected by many environmental factors, a subset of genes with altered expression can inform on stress responses. The potential utility here is to improve biomarker identification and to identify patterns of gene expression associated with different types of pollution. Bundy et al. (2007) sampled an earthworm species (*Lumbricus rubellus*) from seven sites with different levels of metal contamination. Using NMR metabolomics it was shown that metabolic profiles of the earthworms could resolve individual sites. Despite the confounding influences of site parameters, specific metabolites were correlated with zinc, the major contaminant, across all seven sites. Another NMR metabolomics study, involving flounder sampled from industrialised and reference sites in the UK showed that water composition had a significant effect on the fish liver metabolome (Parsons et al. 2007). Uses of “omic” data for prognostic and diagnostic studies are summarized in Supplemental Material, Table 1. These important observations lend support to the implementation of omics as diagnostic tools in ecotoxicology.

**The Current Hurdles from a Regulatory Viewpoint**

**General Technical Hurdles in the Application of Toxicogenomics.** A proposed time-frame for realizing the utility of the omics in tiered testing was shown by Ankley et al. (2006). However, there are a number of factors limiting widespread acceptance for regulatory applications. As well
as complex relationships between omic responses and ecological outcome, there is a lack of standardized, validated exposure, assay and analysis procedures (Ankley et al. 2006). Omics technologies can be viewed as complementary testing procedures that can improve understanding of biological systems and can lead to development of simpler individual assays with defined endpoints. This approach, while valuable, does not exploit the full capabilities of omics, particularly their open nature, allowing discovery of unexpected changes. Intermediate technologies such as PCR-arrays have been proposed for use in clinical diagnosis (Bustin and Mueller 2005), but are not open systems. The validation required for uptake of multi-biomarker techniques for routine testing is time-consuming and complex. In the clinical setting some targeted microarray applications have now been approved by the FDA and EU for diagnostic use. These include the Amplichip CYP450® (Roche, Indiana, USA) for genotyping of human cytochromes P450 and Mammaprint® (Agendia, Amsterdam, Netherlands), a gene-expression microarray for breast cancer prognosis (Glas et al. 2006). Mammaprint validation required the testing of over 1,000 patient samples in 12,000 assays, highlighting the effort and investment necessary for such accreditation. Validation of an assay is a complex procedure encompassing the determination of its reliability and relevance. Hartung et al (2004) discussed a modular approach to validation of alternative tests as part of an initiative by the European Centre for Validation of Alternative Methods. This approach (Supplemental Material, Figure 1) is applicable in general terms to omics-based assays. Key to this procedure is defining a relevant endpoint. Biomarkers can be of exposure or of effect, and the choice between these, and endpoints they aim to predict, must be informed by the requirements of the regulators.
Corvi et al. (2006) suggested requirements for validation of transcriptomics in regulatory toxicology and it is unlikely that this will be rapid. However, a realistic application of omics techniques may be their use in pre-screening chemicals and mixtures for prioritization in further tests (Ankley et al. 2006). The ToxCast programme (Dix et al. 2007), employed a diverse selection of tests, including toxicogenomics, and showed the potential application of this approach. As discussed later, the interpretation of omic data is highly reliant on advanced computational and statistical methods which are still being developed. While QA procedures are paramount, at the pre-screening tier more flexibility is permitted. The US EPA (2002) currently accepts toxicogenomics data as part of a weight-of-evidence approach for establishing mechanisms of toxicity for regulated substances. While we discuss experimental variation below, it should be noted that data capture and archiving is an essential mechanism for highlighting and avoiding the pitfalls of inappropriate experimental design, such as the introduction of systematic variation during omics experiments. While transcriptomics databases are well-established, toxicology-specific omics databases are now emerging (Waters et al. 2008).

Environmental metabolomics recently benefited from the first inter-laboratory intercomparison exercise to evaluate the accuracy, precision and efficacy of $^1$H NMR metabolomics (Viant et al. 2009). Flounder liver extracts from contaminated and reference sites were analysed, and multivariate statistical analyses confirmed high reproducibility across all seven laboratories. Furthermore, the same metabolic biomarkers that discriminated fish from the two sites were discovered by all the laboratories (see Supplemental Material). For transcriptomics in fish, diversity of microarray platforms has precluded inter-laboratory comparisons, but inter-laboratory microarray comparison has been successful for mammalian species (e.g. Mattes 2008;
Shi et al. 2006). It is likely that in future, improvements in the technologies for assaying gene expression, such as high-throughput pyrosequencing and digital transcriptomics (Nielsen et al. 2006), will replace microarrays. Already pyrosequencing allows for fast construction of high quality oligonucleotide microarrays for non-model species (Garcia-Reyero et al. 2008b). The technology is constantly evolving, but the key question that must be addressed, whatever technology is in use, is: How do gene and protein expression and metabolite concentrations relate to ecological outcome? Progress has been made on this question and initial studies on population bases are now being published (Supplemental Material, Table 1). It should be recognized that, although biomarkers are valuable in regulatory and monitoring contexts, the meanings of such changes must be clarified, to allow efficient use in regulatory decision-making (Adelman 2005; Boverhof and Zacharewski 2006).

**Identification of Sources of Variation and Minimising their Effects.** Variability in omics data is an ongoing concern particularly in relation to multiple individual manipulations between biological sampling and data interpretation. The workshop identified sources of technical and biological variation and made recommendations on how these should be managed in terms of experimental design (Supplemental Material). Study design, inadequate sample numbers, methods of sample acquisition, preparation, storage, processing and analysis are key areas of possible technical artefacts. Methods of normalisation and statistical interpretation are major sources of variability. Careful study design is essential to minimise biological variability “intra-class” (e.g. stage of reproduction in a control group) thus maximising “inter-class” differences (e.g. control versus exposed groups). It is noted that inter-individual variability within a
population is essential for ecological health and therefore an impact on such variability from a stressor can be very important.

**Training and Communication.** To advance the application of omics technology into regulatory ecotoxicology and water quality policy, there is a need for effective scientific communication amongst academia, industry and regulators (Blunt et al. 2007). The benefits and limitations of these techniques need to be candidly discussed so that tools with the potentially greatest return-on-investment (both financial and knowledge-based) may be prioritized for utilization. Multidisciplinary workshops allow continued dialogue to inform all stakeholders of developments. These cross-functional meetings also provide researchers with an understanding of the priorities of regulatory authorities in order to discover practical ways of solving issues.

Advances in omics have significant implications for risk assessment practice and regulatory decision-making. The use of genomics technologies generates a large volume of data and the field of bioinformatics is evolving rapidly to meet data analysis needs. A Genomics White Paper identified areas likely to be influenced by omics (US EPA 2004). The Genomics Task Force recommended that the Agency develops training materials and modules to prepare risk assessors and decision makers who will be faced with the challenge of interpreting and applying omics information. Participants of this workshop also believed that training was critical for furthering the application of genomics technologies into monitoring and regulation, particularly as a means of interpreting and applying genomics data for risk assessment (Dearfield et al. 2008). Risk assessors must be able to communicate to managers and stakeholders both the underlying science and the interpretive tools and models used to develop the risk assessment. Likewise, it will be
important to provide training to risk managers regarding the benefits and limitations of genomics in risk assessments Haymes et al. 2008).

It was also recognized that there is a need to build capacity within academia, the private sector and government agencies to implement omic tools and to evaluate omics data, particularly with respect to biological and ecological significance. These institutions will require resources, support, and targeted training to bring scientists and decision makers within their organizations to a point where these tools can be effectively used in regulatory decision making, especially risk assessment (US EPA 2004). National and sub-national programs and agencies should apply strategic hiring practices to recruit individuals who possess omics skills. It would be useful to develop and initiate training in the near future to prepare risk assessors and risk managers to enable them to evaluate and incorporate omic data into environmental decision-making. Initial training could address basic omics concepts, technologies and potential applications and include the basic steps necessary to interpret and apply genomics data to risk assessment.

**Research Needs for Regulatory Implementation.** Research needs were reviewed in two successive SETAC Pellston workshops held in 2004 and 2005 (Ankley et al 2006; Ankley et al, 2008b; Benson and Di Giulio 2007). These efforts identified both short- and long-term needs which, due to resource constraints, have not yet been fully addressed.

The short-term needs identified were (1) formal standardization and validation of data collection, analysis, and presentation for standard test species, and (2) generation of libraries of gene expression, proteomic, or metabolite profiling data based on a set of reference chemicals with
well-defined, relevant MOAs. As explained above, there have been important advances in recent years in the context of both of these needs. The long-term needs identified were (1) generation of genome sequence data for ecologically relevant species, and (2) linkage of molecular and biochemical responses to adverse alterations in survival, growth and development, and reproduction.

Significant advances have also been made towards obtaining data for the development of reference gene expression profiling databases from species commonly used for regulatory assessments (Ankley et al. 2008b) although, much work remains. As toxicogenomics data will be most valuable for predictive toxicology and elucidating toxicologically-relevant MOAs for additional chemicals, a future database should be based on toxicity testing and monitoring protocols commonly used for regulatory purposes (eg. global pesticide registrations), as well as chemicals with well-known MOAs such as 17ß-estradiol and dioxin.

**Funding.** Chemical production is highest in the Organization for Economic Cooperation and Development (OECD) countries, particularly in specialty chemicals and the life science sectors. Moreover, innovation in new chemical development and manufacturing practices is extremely high due to advances in combinatorial chemistry, nanotechnology, and biotechnology. These changes question the sustainability of current approaches to prioritization, monitoring, and risk assessment. It may be difficult to allocate additional resources required to efficiently incorporate and understand omics data. It is important for programs and agencies to focus on these needs and to ensure that adequate funds and people are brought to bear on this need.
The OECD Environment, Health and Safety Programme has started cooperative work for the use of genomic information for risk assessment of chemicals. The scope of this activity is to explore and evaluate regulatory application of genomic methods in chemical hazard/risk assessment. To reduce redundancy and minimize the funding to develop these omic technologies, international cooperation and a common data-base are essential. Target and cross-species omic information and technologies should be developed to monitor animal species relevant to disparate countries.

There are many international examples of initiatives to enhance ecotoxicogenomics e.g. Ministry of Environment of Japan, Canadian Federal Government Interdepartmental Genomics initiative, the US EPA initiatives and the UK NERC Postgenomics and Proteomics research program.

To address the toxicogenomics needs, a considerable investment into the generation and curation of appropriate reference exposure data is required. For example, for most chemical registrations, chronic aquatic toxicity data are required for one freshwater fish species (early life-stage toxicity), one freshwater invertebrate species (full life-cycle toxicity), and, in most cases, one saltwater fish species (early life-stage toxicity) and one saltwater invertebrate species (full life-cycle toxicity). This leads to a large number of GLP-compliant studies and such a reference database would likely require ~$10 million over 3 to 5 years. While this is a substantial cost, a reference database is arguably the only means for successful and appropriate implementation of toxicogenomics data into the current ecological risk assessment paradigm. Without a database to compare chemicals with unknown MOAs, the risk assessor will not be able to interpret the significance of the gene expression responses within the context of characterizing ecological risk because many gene expression changes are not anchored to adverse effect and the risk-assessment requires knowledge of the MOA and the dose-response relationship. A good example
of the utility of a comprehensive database is (ToxRefDB; Martin et al. 2009), containing mammalian toxicity data.

The most feasible way forward would be funding through a multi-stakeholder consortium, such as has been achieved in the ACToR database (Judson et al. 2008). The final reference database would be open source and accessible through a website. The reference database could also be integrated into a tool similar to the U.S. Food and Drug Agency’s ArrayTrack to allow regulatory agencies to easily manage, analyze, and interpret omics data submitted by registrants or other government or academic laboratories using similar ecological species, testing protocols, and microarray, proteomic, or metabolomic platforms.

Conclusions

The omic technologies have advanced over recent years and continue to become more efficient, data-rich and economical in use. Proof of principle has been achieved in terms of potential application to environmental toxicology, and specifically the assessment of environmental pollution impacts in non-model organisms. The key needs to progress the utilization are in relation to the expansion of a reference database of ecotoxicogenomics and a better understanding of the relationships between specific responses and biomarkers to ecological adverse events. Through improved communication between the sectors, the aim of assisting in regulatory decisions can be expedited through the utilization of omic techniques.
Supplemental Material

Supplemental Material discusses variability in the omics. Supplemental Material, Table 1 summarizes key examples of the successful use of toxicogenomics data in ecotoxicology. Supplemental Material, Figure 1 illustrates the modular approach for applying the ECVAM principles on test validity.
References


Supplementary Text 1 – Variability in the omics

Sources of Technical Variability

Transcriptomics: The complex nature of microarray experiments, with many individual manipulations between biological sampling and data interpretation, make technical variability a serious concern. The process of interpretation is especially vulnerable to technical variability because of the simultaneous measurement of hundreds or thousands of endpoints and these are compounded by the high cost of the technique which has often led to the use of suboptimal sample sizes.

Technical variation in transcriptomics measurements can arise from many sources, including sampling processes, non-ideal RNA isolation and storage, variations in RNA labeling (including lability of probes), imperfect hybridization and subsequent data analysis methodologies. Transcriptomics has been the subject of quite extensive evaluations of technical variation (reviewed in detail in de Koning et al. 2007; Fuscoe, et al. 2007; Mattes 2008; Thompson and Hackett 2008; Walker and Hughes 2008). Several consortia have recently examined technical variability within laboratories, between laboratories and between different expression monitoring platforms (Bammler et al. 2005; Kuo et al. 2006; Shi et al. 2006; Beyer et al. 2007; Chen et al. 2007; Arikawa et al. 2008; Fielden et al. 2008; Kohlmann et al. 2008). These large scale studies have demonstrated that microarray analysis can be performed with good intra- and inter-laboratory reproducibility for genes that have copy numbers of at least 5 to 10-fold above...
their detection limits and this has been confirmed by alternative methods such as quantitative real time PCR (Canales et al. 2006).

Several key areas have been identified that can cause technical artefacts. These include study design, inadequate numbers of samples, and methods for sample acquisition, preparation, storage, processing and analysis. Once capture of the biological target has been achieved it is important that equivalent portions of tissue are dissected because of the likelihood of cellular and functional heterogeneity of the tissue and zonation of gene expression (Gebhardt 1992; Oinonen and Lindros 1998; Boedigheimer et al. 2008). It is important to recognize that batch effect as a source of error is likely and that biological samples are processed at random in each batch rather than each batch analyzing biological material from a single sampling event. Sample integrity is critical to successful analysis. Tissue thawing can lead to a rapid loss of RNA integrity thereby dramatically affecting array quality especially if probes are located more than 1000 bp from the transcriptional end of the mRNA (Thompson et al. 2007). Standardized protocols and external controls for quality control must be established to assure reliability. The quality of microarray analysis can be greatly improved by inclusion of adequate documentation that capture important aspects of experimental variables such as treatments, sample quality and methods, therefore helping to assess microarray quality issues due to protocol variation.

Design of microarrays can introduce technical variation especially if multiple platforms are used for analysis. Disagreement in results from different platforms can result from
probes hybridizing with different efficiencies to non-overlapping target sequences and in the case of cDNA probes, this may be compounded by cross hybridization with multigene family members. The format of the microarrays can affect variability with commercial products having better quality control and reproducibility with different sample labeling protocols contributing negligible variations in results (Kuo et al., 2006, Patterson et al., 2006). While different formats give similar results in terms of expression patterns, variation in signal intensity and resulting expression values can be obtained between different laboratories and microarray formats (Chen et al. 2007).

Perhaps the most important technical variable is in the normalization and statistical interpretation of array data. Preliminary steps should be employed to eliminate outliers (excessive chip-to-chip variation) that may be as high as 5-10% (Boedigheimer et al. 2008; Hershey et al. 2008). Normalization is essential to eliminate the effects of variable cDNA labeling but different normalization protocols and statistical analysis algorithms generate different lists of differentially expressed genes. For chip-based arrays, normalization by robust multichip average (RMA) is the usual approach and may have advantages over traditional normalization with housekeeping genes (Irizarry et al. 2003; Hershey et al. 2008), especially since in the present context toxicants are being studied and by definition these will adversely affect cellular architecture and basal metabolic processes. While some groups advocate simple cut off rules based on fold-change versus controls and significance tests (Shi et al. 2006), this can lead to high false positive errors if a no-treatment effect is not considered (Chen et al. 2007). Reproducibility can be maximized if analyses are based on biological functions such as defined by gene
ontology terms and toxicological function (Bammler et al. 2005; Fielden et al. 2008).
Analysis of species of ecological interest is complicated by lack of genome sequence and poor annotation.

Proteomics: Global analyses of the proteome are now becoming possible with new technological innovations, however, sample preparation and manipulations can be exceedingly complex and thus variability is of considerable concern. Recently there have been a few attempts to ascertain intra-individual variability but these have not, so far, been extended to inter-laboratory comparisons, possibly due to the lack of established standards for representing proteomic data.

Inter-experimental reproducibility is not good using 2-dimensional polyacrylamide gel electrophoresis approaches, however, an excellent differential display technique (DIGE) allows multiple samples to be compared on the same gel and is a powerful tool in biomarker discovery for laboratory exposure studies which is quantifiable. Potentially useful biomarker candidates must be carefully selected from proteins that are not subject to sexual, nutritional and naturally high variations in expression, moreover, they must be sufficiently abundant and resolvable from other proteins for reliable quantitation. Recent studies indicate that a sample size of some 7 replicates creates reliable data. There is obviously a pressing need for establishment of standardized selection procedures. Whilst proteins of genomically characterized animals can be identified relatively easily by mass spectrometry of excised proteins using MALDI-TOF instruments, protein identification in non-characterized organisms requires *de novo* sequencing methods, precluding routine use until sufficiently comprehensive databases have been assembled.
Increasingly, proteomic techniques use liquid-chromatography (LC) separations coupled with electrospray ionization (ESI) MS and tandem mass spectrometry (MS/MS) for the characterization of the separated peptides or proteins. Sample preparation procedures are potentially much less variable and theoretically they should be able to analyze a larger proportion of the proteome, however, extensive sequence databases of the study organism are required and methods for accurate inter-individual quantitation are lacking and still under development.

**Metabolomics.** The ultimate expression of an organisms’ phenotype is the profile of metabolites in its cells and bodily fluids. Often their existence is transient as many are rapidly biotransformed therefore sampling, sample preservation and extraction techniques are critical for high reproducibility. Standardisation of methods is therefore essential and progress is being made on this front.

Keun et al. (2002) assessed the analytical reproducibility of an NMR metabolomics experiment by analysing two identical sets of rat urine samples from an acute toxicity study. The analyses were performed at two sites and principal components analyses (PCA) revealed extremely similar descriptions of the metabolic responses to hydrazine.

In one study (Bertram et al. 2007), identical sets of human urine before and after dietary intervention were measured using 250, 400, 500 and 800 MHz NMR spectrometers. When analysed by partial least squares discriminant analysis (PLS-DA), the loadings were found to comprise of the same spectral regions implying that the same metabolites
were discriminating pre- and post-dietary intervention, independent of magnetic field strength. Most recently, an intercomparison exercise involving seven laboratories evaluated the accuracy, precision and efficacy of $^1$H NMR metabolomics for environmental research (Viant et al. 2009).

The study comprised the analysis of both synthetic metabolite mixtures as well as European flounder (*Platichthys flesus*) liver extracts from clean and contaminated sites. For both sample types, PCA revealed highly similar scores plots across all laboratories. Furthermore, the same metabolic biomarkers that discriminated fish from clean and contaminated sites were discovered by all the laboratories. Taken together, these studies clearly demonstrate that NMR-based metabolomics can generate data that are sufficiently reproducible between laboratories to support its use in regulatory studies. No such intercomparison exercises have yet to be reported for mass spectrometry based metabolomics studies.

**Sources of biological variability**

Biological variation can be conveniently sub-categorised into variation within the control group or population (i.e. intra-class variation that is often unrelated to the toxic stressor being studied) and variation between the control and exposed groups (i.e. inter-class variation). In general, experiments should be designed to minimise both technical and intra-class variation, thereby maximising inter-class differences that can be explored
using data mining techniques. Meaningful results depend on technical variability being less than biological variability.

As a prelude to this discussion, two distinct scenarios must be considered. In chemical testing (e.g. OECD) a limited number of model organisms are used worldwide for controlled laboratory studies. In environmental monitoring, locally relevant sentinel species, with little supporting genomics information, are typically used and sampled directly from the environment. These two scenarios will be associated with significantly different degrees of biological variation.

Variability between individuals within a given population is an essential component of population health and sustainability as it encodes for phenotypic flexibility and ability to acclimate to changing conditions and is the vehicle of evolution. Environmental stressors can impact this and affect the phenotypic variation between individuals. Reduced variability is potentially adverse to the sustainability of the population. Little is understood about the stressor-induced changes on these parameters in an environmentally relevant context. Some examples in the literature indicate that individuals within a population may have different degrees of susceptibility to estrogenic exposure, resulting in a large spread in the degree of response at low concentrations of estrogens and a more consistent response between individuals at high concentrations (Thorpe et al. 2001; Thorpe et al. 2003). The implications of the decreased variability caused by chemical exposure for the sustainability of wild fish populations are unknown. Potentially, if such an effect was also occurring in the wild, it could impact on the ability of the population to
cope with changing environmental conditions and compromise its sustainability. It is
clear that individual variation might be considered as an endpoint in toxicological studies
when population level effects are being addressed as chemicals may impact on this parameter and compromise the ability of populations to survive in environments constantly under pressure by changes in natural factors and anthropogenic stressors.

Inter-individual biological variation can be associated with two levels of organisation, generally defined as genotypic and phenotypic variation. Genetic variation is inherent in individuals within populations and accounts for the phenotypic plasticity allowing individuals to acclimate and populations to adapt to changing conditions. Genetic variation within a group of organisms is apparent at several levels, including polymorphisms, copy-number variants, alternative splicing, post-transcriptional and post-translational regulation and epigenetic modifications. Furthermore, organisms may vary in strain or clonal line. In general, genetic variability will be considerably higher (and less well characterised) for environmentally sampled organisms compared to model organisms unless out-bred colonies have been maintained. Furthermore, the identification of closely related species (eg *Mytilus spp.*) is sometimes difficult for environmentally sampled organisms, which can be a major source of error and a confounding issue in the interpretation of ecotoxicogenomics data. Phenotypic variation is strongly influenced by the environment and its interaction with the unique physiological conditions associated with individual organisms. This form of variation can arise from factors such as age, stage of life cycle and reproductive cycle, sex, nutritional status and general health. Phenotypic variation is also likely to be lower in studies of model organisms under
laboratory conditions where it is practical to use standardized experimental conditions. Many of these factors cannot be controlled for environmentally-sampled animals, but recommendations applied to current biological-effects monitoring regarding sampling and documentation (by EPA, ICES etc) should be adhered to. The NERC Environmental Bioinformatics Centre (NEBC) has recommended those parameters that should be recorded and reported for transcriptomic submissions (envgen.nox.ac.uk/posters/MIAME_Env.ppt) as has the Environmental Context working subgroup of the Metabolomics Standards Initiative for a metabolomics study (Morrison et al., 2007). While it might be anticipated that model organisms raised in a controlled laboratory exhibit less environment- and capture-induced variation than similar animals living in the wild, the situation is far less clear for wild animals housed within the laboratory for short periods. Hines et al. (2007) showed that direct sampling of *Mytilus galloprovincialis*, a marine mussel, from the environment resulted in less metabolic variation than for animals from the same cohort that were returned to the laboratory in an attempt to allow their metabolome to equilibrate in controlled conditions.

Mammalian toxicogenomics studies have provided valuable information on baseline fluctuations in gene expression due to study condition and/or endogenous factors. A consortial effort was recently undertaken by the HESI Genomics Technical Committee to examine microarray data from control animals from toxicogenomics studies of rat liver and kidney (Boedigheimer et al. 2008). Gender, organ section, fasting state, and strain emerged as study factors that contributed highly to variability in gene expression, whereas other factors, such as age, vehicle administration route, sacrifice method, and
dose frequency were not major contributors to baseline variance. Genes with high variability, identified in the study, include many of interest in toxicology, such as those involved in xenobiotic metabolism, androgen and estrogen metabolism, steroid biosynthesis, and antigen processing and presentation. Low variance genes were also identified, and included those involved in protein metabolism and immune response. Such low variance genes may prove valuable as study controls. This collaborative effort determined the impacts of key study factors on measured gene expression in a toxicogenomics study and illustrated the importance of defining the baseline gene expression against which stressor-induced gene expression changes are to be evaluated.

Others have considered expression level variation as a genetic Quantitative Trait that can be used to identify loci that regulate gene expression and thus explain the mechanisms behind the variation (Williams et al. 2007). While it is possible to establish the importance of genetic factors in dictating mRNA expression level, variation between independent studies frequently generates conflicting data which have sometimes been attributed to a technical failure of array technology from differential hybridization of array probes by polymorphic transcripts (Alberts et al. 2007). In the area of proteomics, Hu et al. (2005) investigated variability in expression of proteins of human cerebrospinal fluid and Zhang et al. (2006) reported variability in protein expression between 12 human liver samples. These, and other studies on fish (George et al. pers. comm.), identified a very large dynamic range of variation in protein expression.
Experimental design and biological variation

Here we highlight recent findings from studies of variability within both the chemical risk assessment and environmental monitoring scenarios. Hines et al. (pers. comm.) recently conducted an extensive investigation into sources of metabolic variation in marine mussels sampled from the environment. An initial analysis of the entire metabolic dataset showed large variability but the majority of this could be rationalised in terms of season (i.e. month when animal was sampled), sex, species and site effects. This data suggests that information about these parameters is necessary for interpretation of inter-individual biological variation and to potentially reveal pollutant effects.

Similarly, many studies in both fish and rodents have highlighted the strong influence that parameters such as sex and stage of the reproductive cycle have on an individual’s transcriptome and proteome, not only in the reproductive tissues but also in other commonly studied tissues such as the brain and liver. In many cases this can be attributed directly to interaction with sex hormone signalling pathways; however, more subtle effects can also be due to cross-talk between nuclear transcription factors. Therefore, it is essential to consider and document both the sex and the stage of the reproductive cycle in toxicological studies even when assessing chemicals which are not suspected of causing endocrine disruption. While differences in the gonad are to be expected, extensive sexually dimorphic gene expression is also found in somatic tissues in rodents (Boyle and Craft 2000; Waxman and O’Connor 2006; Yang et al. 2006), zebrafish (Robison et al. 2008; Santos et al. 2008; Sreenivasan et al. 2008) and M. edulis (Brown et al. 2006) and
is likely to occur across species and in particular in organisms developing as
gonochoristic males or females. Environmental factors such as temperature and hypoxia
have also been shown to strongly influence the transcriptomic and proteomic profiles in
poikilotherms. Indeed the expression of several enzymes and proteins which activate and
detoxify important chemical toxins, as well as those involved in RNA processing,
translation initiation, mitochondrial metabolism, proteasomal function, and essential fatty
acid synthesis in fish show a clear temperature dependence (Gracey et al. 2004). Thus the
response to chemical exposure may be significantly modulated by the ambient
temperature and this must be taken into account in field sampling. The general
recommendation of European legislative studies of chemical impacts is to standardise the
sampling season to minimise effects of temperature and the nutritional status and to
utilise animals (usually males) when they are either sexually immature or gonadally
quiescent.

Key components that require attention to minimise variability, as discussed above, are
sampling processes, sample isolation and storage, preparation of samples for analyses,
numbers of replicates, design of platforms and methods of normalisation. Particularly
important are study design and statistical analyses. Only with biological replicates is it
possible to apply statistical tests. Statistical justification and formula presented by
Kendziorski et al. (2003) and Peng et al. (2003) can be used to calculate the number of
biological samples and pools needed for an appropriately powered analysis. For omics
based studies the challenge in power analysis is how to determine the appropriate
variance since tens of thousands of elements (i.e. genes, proteins or metabolites) are
measured simultaneously and each one possesses its own variance. One option is to use a pooled variance, such as the value estimated from PCA or other error pooling algorithms. The second option is to calculate the variance for each element and take the n\textsuperscript{th} percentile as the value to use. Usually, detection of small changes for elements with large variances is unlikely. By adjusting the variance value and the effect size (i.e. fold change), power analysis can be tailored to choose the sample size that fits the purpose and significance of an experiment. Power analysis requires a pre-existing data set that resembles the proposed project as much as possible in order to derive an accurately estimated variance. The perfect data set comes from a pilot study for the proposed project. When that is impossible, we recommend matching the biological aspects (e.g. species and tissues) before matching the technical aspects (e.g. technologies and platforms), assuming the technical variance is smaller than the biological variance.

References


**Supplementary Table 1: Key examples of the successful use of toxicogenomics data in ecotoxicology.**

A. Prospective Studies

<table>
<thead>
<tr>
<th>APPLICATION</th>
<th>NATURE OF EXPOSURE</th>
<th>APPROACHES</th>
<th>SIGNIFICANCE OF RESULTS</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Signatures</td>
<td>Diet, static renewal, semi-static renewal or <em>ip</em> injection</td>
<td>Microarray</td>
<td>Demonstrates chemical class specific gene expression, and/or chemical specific signatures</td>
<td>Benninghoff and Williams 2008; Larkin et al. 2003; Hamadeh et al. 2002; McMillian et al. 2005; Moens et al. 2006; Hook et al. 2006; Ellinger-Ziegelbauer et al. 2008</td>
</tr>
<tr>
<td>Mixtures</td>
<td>Hepatocytes in vitro</td>
<td>Microarray</td>
<td>Demonstrates that mixtures show attenuated biomarker responses compared to those of the chemical signatures of classical environmental toxicants</td>
<td>Finne et al. 2007</td>
</tr>
<tr>
<td>Endocrine Disruption Pathways</td>
<td>Flow-through</td>
<td>Microarray</td>
<td>Links reduced fecundity/population decline to perturbations in endocrine pathways, vitellogenin, steroid concentrations, gonad weight, and specific perturbations in gene expression</td>
<td>Miller et al. 2007; Ankley et al. 2008; Villeneuve et al. 2007; Hoffmann et al. 2008</td>
</tr>
<tr>
<td>Pathways of Toxicity</td>
<td>Injection</td>
<td>Microarray</td>
<td>Demonstrates that TCDD perturbs fin regeneration by impacting the expression of genes involved in extracellular matrix composition and cellular differentiation</td>
<td>Andreasen et al. 2006</td>
</tr>
<tr>
<td>Pathways of Toxicity</td>
<td>Static exposure</td>
<td>Microarray</td>
<td>Demonstrates heart-specific mechanisms of AhR/TCDD-mediated toxicity</td>
<td>Handley-Goldstone et al.</td>
</tr>
<tr>
<td>Pathways of Toxicity/Biomarkers of Effect and Temporal Changes</td>
<td>Flow-through</td>
<td>Microarray</td>
<td>Used cDNA arrays and Q-PCR to identify potential indicators of thyroid axis and metamorphosis disruption in frogs</td>
<td>2005; Carney et al. 2006</td>
</tr>
<tr>
<td>Temporal and Adaptive Changes</td>
<td>Injection</td>
<td>Microarray</td>
<td>Demonstrates time-dependent adaptive changes prior to toxicity following TCDD treatment</td>
<td>Roling et al. 2006</td>
</tr>
<tr>
<td>Screening of Emerging Chemicals; Predicting Adverse Outcomes</td>
<td>Flow-through</td>
<td>Microarray</td>
<td>Uses microarrays to determine the potential mechanisms of PFOA toxicity</td>
<td>Wei et al. 2008</td>
</tr>
<tr>
<td>Emerging Chemicals</td>
<td>Static</td>
<td>Microarray</td>
<td>Uses microarrays to demonstrate differential effects of nanoparticles and their constituents</td>
<td>Griffitt et al. 2007</td>
</tr>
<tr>
<td>Inter-laboratory Comparisons</td>
<td>Flow-through, other</td>
<td>Microarray</td>
<td>Meta-analysis of data from environmental estrogen exposure that produced new, sensitive biomarkers of exposure</td>
<td>Gunnarsson et al. 2007</td>
</tr>
<tr>
<td>Computational</td>
<td>Flow-through</td>
<td>Microarray</td>
<td>Assesses sources of variation in fish microarray experiments. Chemical class prediction using bioinformatic classification software such as Support vector machines</td>
<td>Wang et al. 2008a, Wang et al. 2008b</td>
</tr>
<tr>
<td>Chemical Signatures/Adaptive Response</td>
<td>Flow-through</td>
<td>Metabolomics</td>
<td>Demonstrates compensatory mechanisms and adaptive recovery from an estrogen. Also demonstrates the potential of metabolomics in</td>
<td>Ekman et al. 2007</td>
</tr>
<tr>
<td>Temporal Changes</td>
<td>Flow-through</td>
<td>Proteomics</td>
<td>Provides early response indicators to thyroid hormones in frogs</td>
<td>Domanski and Helbing 2007</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>-----------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Chemical signatures</td>
<td>laboratory</td>
<td>Metabolomics</td>
<td>A mechanism of action from combination of NMR metabolite profiling and neural network classification</td>
<td>Ott et al 2003</td>
</tr>
</tbody>
</table>
### B. Diagnostic Studies

<table>
<thead>
<tr>
<th>APPLICATION</th>
<th>NATURE OF EXPOSURE</th>
<th>APPROACHES</th>
<th>SIGNIFICANCE OF RESULTS</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Genetics</td>
<td>None/unknown field</td>
<td>Microarray</td>
<td>Demonstrate variation in gene expression and potential difficulties of using particular genes as biomarkers</td>
<td>Oleksiai et al. 2002; Oleksiai et al. 2005</td>
</tr>
<tr>
<td>Sample Monitoring</td>
<td>Wastewater treatment</td>
<td>Microarray</td>
<td>Effect of Mixtures; endocrine disruptors and no-endocrine disruptors</td>
<td>Filby et al. 2007b; Garcia-Reyero et al. 2008; Filby et al. 2007c</td>
</tr>
<tr>
<td>Sample Monitoring</td>
<td>Field</td>
<td>qRT-PCR</td>
<td>Demonstrates that seal thyroid hormones are sensitive to disruption by pollutant stress</td>
<td>Tabuchi et al. 2006</td>
</tr>
<tr>
<td>Population Genetics</td>
<td>Field</td>
<td>Microarray</td>
<td>Demonstrates differential adaptation to distinctly different field sites</td>
<td>Larsen et al. 2007</td>
</tr>
<tr>
<td>Chemical Remediation</td>
<td>Field</td>
<td>Microarray</td>
<td>Demonstrates the utility of microarrays as an additional weight-of-evidence approach to monitor remediation at a polluted site</td>
<td>Roling et al. 2007</td>
</tr>
<tr>
<td>computational</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site Monitoring</td>
<td>Field</td>
<td>Subtractive hybridization, Differential display</td>
<td>Demonstrates differential expression at polluted and references site. Links to lab work to field sampling</td>
<td>Maples and Bain 2004; Roling et al. 2004</td>
</tr>
<tr>
<td>Computational</td>
<td>Field</td>
<td>Metabolomics</td>
<td>Improved discrimination between sample</td>
<td>Parsons et al. 2007</td>
</tr>
</tbody>
</table>
classes compared to unscaled, autoscaled or Pareto scaled data. Leads to better site classification

<table>
<thead>
<tr>
<th>Method</th>
<th>Type</th>
<th>Tools</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site monitoring</td>
<td>Field</td>
<td>Metabolomics</td>
<td>Demonstrates differential metabolic fingerprints in earthworms (<em>Lumbricus rubellus</em>) from sites with differing metal contamination.</td>
<td>Bundy et al. 2007</td>
</tr>
<tr>
<td>Site monitoring</td>
<td>Field</td>
<td>Metabolomics</td>
<td>Demonstrates differential metabolic fingerprints in marine mussels (<em>Mytilus edulis</em>) at polluted and reference sites.</td>
<td>Viant, unpublished</td>
</tr>
</tbody>
</table>
REFERENCES


Supplementary Figure 1.

**Timeframe for Realizing Omic Efficiencies in Tiered Testing**

Reproduced from Ankley et al (2006), with permission.
Supplementary Figure 2.

The modular approach for applying the ECVAM principles on test validity

A “yes” indicates that the appropriate information for the module is adequate for entrance into the peer-review process. All seven modules have to be satisfactorily completed, as judged by the Validation Management Group, before a method can enter the peer-review process.

ECVAM = European Centre for the Validation of Alternative Methods.

Reproduced from Hartung et al (2004), with permission.