The National Toxicology Program Roadmap; a Decade Out

John R. Bucher[1], Mary S. Wolfe[1], Linda S. Birnbaum[1] Christopher J. Portier[2]

[1] Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, 27709

[2]

Abstract

In 2004 the NTP released *"A National Toxicology Program for the 21st Century, a Roadmap for the Future*. The *Roadmap* was organized around three broad initiatives; (1) refining traditional toxicology assays, (2) developing rapid, mechanism-based predictive screens for environmentally induced diseases, and (3) improving the overall utility of NTP products for public health decisions. In the last decade considerable progress has been achieved on all three fronts. This communication reviews initiatives and actions by the NTP in response to the *Roadmap*, discusses how they are changing the science of toxicology, and identifies technological and procedural advances that should result in improvements in human health protection from exposures to agents in our environment.

Text

The US National Toxicology Program (NTP) was established in 1978 as an interagency program, headquartered at the National Institute of Environmental Health Sciences (NIEHS), and charged with coordination of toxicology research and testing across the Department of Health and Human Services. Over its 35-year history, the NTP has performed toxicology and carcinogenesis studies on thousands of agents of environmental concern. The program also carries out important literature analysis activities including issuance of the Report on Carcinogens and reports and program opinions on non-cancer health outcomes generated through the Office of Health Assessment and Translation. In addition, the NTP provides support for the activities of the Interagency Coordinating Committee on the Validation of Alternative Methods, charged with facilitating the acceptance of toxicology testing methods that reduce, refine, or replace animals in current Federal regulatory agency testing frameworks.

> *During 2003, the NTP developed a vision for how it saw toxicology evolving during the 21st century. As stated, the main goal of the Vision was "To move toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon a broad inclusion of target-specific, mechanism-based, biological observations"(NTP 2004a; Bucher and Portier 2004).*

In 2004, following a year-long development and consultation process, the NTP released a strategic plan for the new century, *"A National Toxicology Program for the 21st Century, a Roadmap for the Future* (NTP 2004b). The *Roadmap* was organized around three broad initiatives; (1) refining traditional toxicology assays, (2) developing rapid, mechanism-based predictive screens for environmentally induced diseases, and (3) improving the overall utility of NTP products for public health decisions. As stated:

> *"The challenge for the NTP and health regulatory and research agencies is to examine the impacts of the growing list of anthropogenic exposures as well as those of natural origin encountered in our daily lives, while also providing the best scientific data possible on the toxic potential of each exposure. Unfortunately, several factors complicate this challenge: the large number of chemical substances in commerce, the complexity of environmental exposures, and the uncertainties about how genetic variability and/or lifestage might interact with environmental exposures to affect risk" (NTP 2004b).*

The purpose of this communication is to take stock of the progress the NTP has made towards achieving these goals, and to illustrate how these accomplishments are changing the way toxicology is contributing to the improvement of public health and the prevention of disease.

Refining Traditional Toxicology Assays

> *"The initial phase of implementation of the NTP Roadmap will involve a formal review of the designs of all NTP assays to critically analyze their predictive power and determine whether the protocols for these studies should be altered" (NTP 2004b).*

Beginning in 2005, NTP sponsored public workshops to gather ideas on changes that could be made to enhance the utility of existing standard toxicity study designs (Chhabra et al. 2003). The NTP rodent cancer bioassay was a major focus of attention. The long used inbred Fischer 344 rat and the hybrid B6C3F1 mouse were compared with potential alternative inbred and outbred strains, and the use of multiple strains in the cancer bioassay was debated (King-Herbert and Thayer 2006; King-Herbert et al. 2010). Although the rat strain ultimately was changed to the Harlan Sprague Dawley, the B6C3F1 hybrid mouse was retained despite it's recognized propensity to develop high rates of liver tumors. The desire to increase the genetic "coverage" of the rodent bioassay was recognized as a major deficiency and a multiple strain approach (Festing 1995) was heatedly debated from both a statistical and logistical standpoint, but in the end, was considered impractical to implement.

Another workshop explored the general area of hormonally mediated cancers with respect to cross species predictivity (Thayer and Foster 2007). While most known human carcinogens are mutagens or show other evidence of genotoxicity, hormonally mediated cancers also occur, and the question examined was whether the tumor profiles in hormonally sensitive organs in rodents were sufficiently similar to those in humans that rodents could be considered adequate surrogates. The short answer was no, and the workshop formulated several recommendations to potentially compensate for the deficiencies of the current models and study designs. These included focusing on selected non-neoplastic changes as biomarkers of carcinogenic activity in rodents, utilizing mammary gland longitudinal "whole mounts" to allow a more complete assessment of mammary ductal structures, incorporating chemical exposure during the perinatal period to potentially address windows of enhanced susceptibility, and placing greater reliance on other approaches to discover chemical-hormone receptor interactions utilizing *in vitro* assays.

The program decided to accept several of the workshop recommendations, including to adopt exposures during the perinatal period beginning with rats (NTP 2013a), and to employ mammary gland whole mounts for studies of hormonally active substances (Rudel et al. 2011). Adoption of the Harlan Sprague Dawley rat was critical to this change, as neither the F344 rat, nor the B6C3F1 mouse, are sufficiently robust breeders to be able to reliably and economically generate the required numbers of pups for study designs that both avoid litter effects and allow the study to commence within a reasonably constrained time period. The Harlan Sprague Dawley is an outbred animal, with excellent breeding characteristics and is commonly used in toxicology studies. Rats of the Sprague Dawley lineage also have a long history of use in studies of mammary carcinogenesis, with altered terminal end bud development considered an early biomarker of cancer (Russo and Russo 1996). In contrast, the B6C3F1 mouse was retained in part because, as will be discussed later, the program wished to develop tools to evaluate chemically induced epigenetic changes using new sequencing technologies, and use of a hybrid mouse from consistent crosses of the parental inbred C57BL/6 and C3H/HeN strains would allow this both in future studies, and potentially in examinations of tissues from archived studies.

The adoption of routine perinatal exposures as part of NTP standard studies (NTP 2013a) prompted much thought about the best way to efficiently incorporate what is in essence a reproduction and developmental toxicity study as the front end of studies evaluating cancer, immune function, neurobehavior, etc. The "workhorse" NTP reproduction and development study design was devised in the late 90s, termed Reproductive Assessment by Continuous Breeding (Chapin and Sloane 1997), and involved multiple matings across three generations. Recently, the NTP has been testing a modified one-generation reproduction and development protocol (NTP 2013b). This design has advantages in that it reduces animal use, provides dose setting information for perinatal cancer studies, and allows for the addition of "cassettes" for evaluating other toxicity endpoints such as immunotoxicity, while

still retaining the capacity and power to detect and characterize chemical effects on reproduction, development and/or teratology. This design is consistent with and responsive to the historical NIEHS emphasis on the developmental origins of health and disease, and should also result in a reduction in the numbers of animals currently required.

Over the past decade many other procedural changes have taken place in NTP studies, ranging from incorporation of improved statistical methods for cancer bioassays (Peddada and Kissling 2006), to enhanced pathology assessments of the nervous system (Rao et al. 2011), to improved methods for scoring micronuclei (Witt et al. 2008) and incorporation of the comet assay (Recio et al. 2102) in our genetic toxicity battery. Toxicokinetic studies have been and remain a frequently performed adjunct to provide important parameters for physiologically based pharmacokinetic models to assist in risk assessments (Collins et al. 2000). Many chemically induced rodent tumors are now routinely examined for molecular changes in oncogenes or tumor suppressor genes, and mutation spectra in these genes are often compared with the changes seen in comparable human tumors (Hoenerhoff et al. 2009). In addition, gene expression studies are now a routine part of characterizing the toxic response to chemical treatments (Boorman et al. 2005a; 2005b, Auerbach et al. 2010).

The NTP maintains an extensive archive of materials from studies performed over its 35-year history. The archive currently contains over 5 million tissues as formalin fixed paraffin embedded (FFPE) blocks. These materials have been available and used for studies of chemically induced changes in DNA sequence (SNP analyses and mutation spectra studies), but until recently gene expression studies could only be performed on a much smaller set of saved frozen tissues. Recently we have developed methods to recover usable RNAs from FFPE blocks for both microarray analyses (Merrick et al. 2012) and for RNA seq (Merrick et al. 2013), providing powerful tools for future use in addressing questions about the molecular pathobiology associated with chemical exposures for which phenotypic outcomes are known.

The ability to mine additional knowledge from NTP archival specimens was greatly enhanced by the recent acquisition of a commercial database of Sprague Dawley rat toxigenomic profiles for 638 different compounds. This gene expression database is called DrugMatrix (NTP 2013c). The NTP has made this, one of the world's largest molecular toxicology reference databases, freely publically available, allowing researchers to compare their toxicogenomic findings against genomic signatures representing approximately 140 molecular pathways of relevance to toxicology and disease, and almost 50 pathological phenotypic outcomes.

These new tools are providing opportunities to examine chemical–biological interactions in toxicology studies at an unprecedented level of detail. But the *Roadmap* also pointed out the critical need to develop ways to approach the extent

and underlying mechanisms of variability in population response to toxicants. Of great concern is the assumption that toxicity results from studies in one or a few strains of rats and mice form an adequate assessment of human hazards and risks.

In 2003/2004, the NTP and NIEHS applied DNA sequencing technologies to "resequence" the genomes of 15 mouse strains, including 11 commonly used domesticated inbred lines, plus 4 wild-derived lines, to gain a better understanding of the phylogenetic relationships between the strains (Frazer et al. 2007). Over 8.27 million single nucleotide polymorphisms (SNPs) were identified, and their density and locations were used to establish genome-wide haplotype maps, which were then used to track the ancestral contributions of the Mus musculus subspecies. The results showed that large segments of the mouse genome in the commonly used inbred strains were similar or identical, suggesting a much more restricted genetic diversity than was originally thought.

To expand the diversity, a programmed cross breeding effort has created hundreds to thousands of genetically distinct mouse strains through a consortium termed the Collaborative Cross (Churchill et al. 2004). A different approach to the generation of genetically heterogeneous mice is the Diversity Outbred (DO) mouse population created at the Jackson Laboratory (Churchill et al. 2012). These new research resources are providing an opportunity to study toxicant responses in populations of mice that are claimed to exceed the range of genetic variation in the human population.

NTP has used the DO mouse in several studies to gain an appreciation for the limitations reliance on one or a few strains of rodents really represents in our attempts to provide toxicology information for the protection of public health. One study involved inhalation exposure of DO mice to the known leukemogen benzene, with measurement of micronuclei frequencies in blood and bone marrow at the conclusion of the 28-day study. The variation in micronuclei response spanned over 200 fold among identically exposed individual mice, and genotyping of these mice allowed identification of a sulfotransferase variant as a candidate gene primarily responsible (French et al. 2014). Further studies of this type will help to guide regulatory science practices in the coming years.

The decreasing cost of whole genome sequencing is also providing an opportunity to begin to examine epigenetic effects of chemical exposures in stunning detail. A current NTP and NIEHS project is deep sequencing the DNA from the livers of B6C3F1 mouse hybrid offspring and their parental C57/BL 6 and C3H/HeN strains when each strain provides either the maternal or paternal alleles. The DNA sequence information is being aligned with RNA expression data generated through RNAseq. This will hopefully allow a better understanding of the inheritance of the methylome and other epigenetic marks and how these influence gene expression. The goal of these studies is to develop a toolbox of rapid screening approaches to apply to the tissue archives to expand even more, opportunities for in depth study of chemical target tissue interactions.

Developing Rapid, Mechanism-based Predictive Screens for Environmentally Induced Diseases

> *"The NTP envisions that over the next decade utilization of our rapidly expanding knowledge of the physiological, biochemical, and molecular bases of disease will lead to the development of, and a gradual transition to, vastly improved and higher-throughput methods for predicting the toxicological impacts of environmental agents. The NTP recognizes that this transition in methods from predominantly mammalian screens toward more in vitro systems and non-mammalian models must be carefully planned and systematically evaluated to assure scientific and regulatory utility. The NTP is confident that through sustained leadership in creating and applying these mechanistic toxicology tools, we will generate the scientific information and understanding necessary for public health decision-makers to use these new tools to reduce the burden of environmental disease" (NTP 2004b).*

By 2001 it was becoming apparent that advances in chip technologies, both computational and experimental, were likely to have a great impact on regulatory toxicology (Portier 2001; Simmons and Portier 2003). The ability to obtain massive amounts of data on the interaction of thousands of chemicals with hundreds of biological "targets" in automated robotic *in vitro* assays combined with the computational power to organize and analyze these data created an opportunity to change the practice of toxicology. When these ideas were proposed at NTP workshops charged with helping to formulate the NTP Roadmap, they were met with much skepticism. Nonetheless, the Roadmap was clear in setting out an ambitious goal for the NTP and for the science of toxicology.

The NTP supported a number of workshops discussing the use of single endpoint high-throughput screening (HTS) as a means to identify chemicals that alter critical targets related to toxicity. The choice of targets, the designs of these assays and the types of analyses needed to evaluate the findings were discussed in great detail. Early findings (Xia et al. 2008) from the initial studies were promising and led to development of a larger effort that is still in its formative stages (Xia et al. 2009; Parham et al. 2009; Martin et al. 2011; Sedyck et al. 2011).

These concepts were debated, adopted and ultimately actively promoted by a National Research Council (NRC) Committee that was charged with formulating a long-range vision for toxicity testing and a strategic plan for implementing that vision (NRC 2007; Anderson et al. 2010). This committee drew heavily on concepts articulated in the NTP Vision and Roadmap in an interim report (NRC 2006). The NRC Committee report was very effective in raising the consciousness of the toxicology community and as evidenced by the prominent role these ideas played in

keynote addresses, symposia and talks at subsequent Society of Toxicology Meetings.

The NRC Report (NRC 2007) advanced the concept of "toxicity pathways" as an organizing principal around which to build a new science. Toxicity pathways were described as biological processes reflecting perturbations to a cell or organism of sufficient magnitude that the host was unable to adapt because of underlying nutritional, genetic, disease, or life-stage status. Then biologic function would be compromised leading to toxicity and disease. We had earlier proposed that mechanism-based biological observations might be more health protective than the results of standard cancer bioassays for establishing carcinogenic risk (Bucher and Portier 2004). Gohlke et al. (2009) advanced the concept that perturbations in signaling and/or metabolic pathways, whether induced through genetic or chemical interference, were critical to the identification of a mechanistic basis for disease.

In 2008 we announced a cross agency initiative to shift toxicology assessments from primarily *in vivo* animal studies to *in vitro* assays, *in vivo* assays with non-mammalian organisms, and computational modeling (Collins et al. 2008). In collaboration with the EPA National Center for Computational Toxicology, the National Human Genome Research Institute NIH Chemical Genomics Center, and in 2010, the Food and Drug Administration (FDA), the NIEHS/NTP joined to form what is now known as Tox21. This collaboration is the federal government's response to the call for creation of an agency dedicated to these purposes proposed by the NRC Committee (NRC 2007). The results of this effort have been published in over 100 reports as of December 2013, and a critical assessment of the challenges and accomplishments of the collaboration up until the time the FDA joined in 2010, was recently communicated by Tice et al. (2013). The initial results of this exceedingly complex undertaking are promising (e.g. Thayer et al. 2012; Thomas et al. 2012), and we believe will ultimately be successful in providing an alternative data stream on which to select only the highest priority chemicals for further in-depth study, to guide efforts to develop new chemicals for new purposes that have a lower probability of risks to human health and the environment, and ultimately use as the basis for assessments of human risk.

Improving the Utility of NTP Products for Public Health Decisions

> *"The ultimate responsibility for making decisions that impact public health rests predominantly with international, national, and state regulatory authorities. But, the private sector must also understand the basis for regulatory decisions that affect each and every person. The NTP must work with all its stakeholders to ensure that the context and nature of data*

generated by the NTP are clearly articulated and support risk assessment decisions based on scientific information of the highest quality" (NTP 2004b).

Communicating the results of NTP studies and explaining the basis for programmatic public health decisions has been an important goal for NTP since its establishment in 1978. The NTP rodent cancer assay technical reports have been a model for the complete and thorough reporting of these complex studies for decades (NTP 2013d). The program published comprehensive pathology atlases of neoplastic lesions in the Fischer rat and the mouse in the 1990s (Boorman et al. 1990; Maronpot et al. 1999), and we are about to begin release of non-neoplastic lesion atlases in a web-based format (Cesta et al. 2014). This atlas is an offshoot of a years long effort to "digitize" pathology images from NTP studies. Because of this effort, it's now possible to carry out remote pathology consultations or full reviews using computer images, providing greater flexibility in achieving consensus diagnoses and international harmonization of lesion terminology.

In 2007 the NIEHS rolled out the CEBS Database (Chemical Effects in Biological Systems) as a public repository for toxicogenomics data, including study design information, and clinical chemistry and histopathology associated with microarray and proteomics data (Waters et al. 2008). In 2009 the NTP adopted the CEBS database framework as the primary repository of all electronic data from NTP studies and efforts were begun to incorporate historical data from prior studies, along with all newly generated data into this searchable integrated database. Although still under development, CEBS is the primary portal for public access to NTP data (NTP 2013c). The intent is to create a database capable of supporting a variety of inquires across study types, from traditional *in vivo* toxicology studies to genomic and HTS data, revealing associations between findings that may not be otherwise apparent.

Technical reports from the rodent cancer assays have for many years included detailed interpretive language for "levels of evidence for carcinogenic activity". No comparable standard language has been available for other study types. In 2008 we convened working groups to evaluate and road test proposed language for a consistent assessment of immunotoxicity, reproduction and development studies. Guidance documents are available on our web site (NTP 2013e,f,g) that if adopted widely should improve the consistency of interpreting and reporting these studies across the toxicology community.

The NTP also produces influential public health documents that incorporate information produced by the program, as well as surveys of the existing literature on specific topics. The Report on Carcinogens is a congressionally mandated activity requiring the NTP to produce a biennial report that lists agents either "known" or "reasonably anticipated to be human carcinogens". Although the specific criteria for listing substances in these categories were last reviewed and approved in 1995, we've made several modifications to the process by which the background information on which listing decisions are made is collected, analyzed and

summarized (NTP 2013h) in an effort to improve the transparency for how the listing decisions are reached.

From 1998 through 2010 the NTP Center for the Evaluation of Risks to Human Reproduction (CERHR) issued a number of reports evaluating the evidence for potential adverse effects on reproduction and development of current levels of human exposures to a variety of substances. These reports were compiled and the supporting literature interpreted largely by expert panels convened by the program according to criteria that were established by each panel. In 2011, the mandate for CERHR was expanded to include health endpoints other than reproduction and development and the name changed to the Office of Health Assessment and Translation (OHAT)(Bucher et al. 2011). Part of the charge to this new office was to develop innovative approaches for information management and data integration across human, experimental animal and mechanistic or mode of action data streams, to reach consistent and clearly articulated public health decisions. Working from processes developed to systematically evaluate and integrate data from randomized clinical trials, OHAT staff have been actively developing the tools and processes to enable these methods to be applied to the range of human epidemiology, animal toxicology and mechanism of action studies that form the literature base in environmental health sciences (Birnbaum et al. 2013). Efforts are underway to develop these new literature assessment tools in concert with EPA, ATSDR and other government agencies that are faced with similar challenges of sifting through ever growing literature of diverse types and quality to reach conclusions in a systematic and transparent manner.

The application of systematic review to literature-based health assessments is particularly challenging when the literature is large, diverse and conflicting. This situation is best showcased by studies of the potential for adverse health effects from exposures to environmentally relevant levels of bisphenol A (BPA). In 2008 the NTP CERHR reviewed the already voluminous literature on the reproductive and developmental effects of BPA and was the first Federal organization to declare "some concern" that current levels of exposure to human fetuses, infants and children may result in developmental changes in the prostate gland and brain, and diminish sexually dimorphic behaviors (NTP CERHR 2008). These findings remain controversial (Bucher 2009). Although literally hundreds of studies have been published, many by academic investigators funded by NIEHS (2013), that demonstrate a growing list of health effects associated with environmentally relevant exposures to this endocrine disruptor, the standard toxicology studies done according to prevailing regulatory guidelines provided essentially no findings that would prompt concerns about exposures even many fold higher than is currently the case.

Regulatory agencies around the globe are struggling to devise appropriate approaches to deal with a variety of agents capable of disrupting endocrine system actions at environmentally relevant levels of exposure (EPA 2013; EFSA 2013). Members of the Endocrine Society have offered recommendations for incorporating

basic endocrine principles into safety assessment studies for endocrine active agents (Zoeller et al. 2012), prompting a heated disagreement between some members of the toxicology and endocrine communities (Dietrich et al. 2013; Gore et al. 2013).

Although extraordinary NIEHS and NTP attention is being given to studies of BPA to try to clarify its true risks to humans (Birnbaum et al. 2012), this is an answer we must find. If exposures to low doses of a wide variety of hormonally active substances during critical windows of development are affecting human health, then our science is failing its basic public health mission. The question of whether guideline compliant toxicology studies of current designs are up to the task of identifying agents such as BPA is being put to the test in an unusual consortium of government and academic investigators (Schug et al. 2013). The NIEHS and NTP have arranged and are funding a very large new perinatal exposure chronic rat study underway at the FDA's National Center for Toxicological Research. The study is employing a broad range of BPA doses, an expanded set of endpoints, and is being performed with extraordinary attention to potential for contamination by BPA from common laboratory equipment and supplies. The truly unique aspect is that a competitive grants program was made an integral part of the study. The NIEHS Division of Extramural Research and Training has funded a dozen academic investigators, many of whom have published findings implicating BPA in low dose health effects, to be full partners in the study and to carry out a wide range of investigations on tissues and animals from the NCTR study. The results of this study will hopefully address some of the fundamental issues that currently divide the endocrinology and toxicology communities.

The principal NTP agencies (NIEHS, NCTR/FDA and NIOSH/CDC) have worked closely to assure that NTP studies are addressing current needs of the regulatory agencies. NIEHS has an established, 20-year interagency agreement with NCTR focusing on toxicity studies of chemicals in foods, herbal supplements, nanoscale materials and endocrine active agents in addition to BPA. A somewhat similar arrangement is in place with NIOSH that focuses on complex occupational exposures, and has recently been used to provide exposure measurements for a wide variety of substances including manufactured nanomaterials, molds, and popcorn butter flavoring agents. These interactions provide an important grounding in NTP study designs such that they are optimally useful for regulatory purposes. Conversely, many of the initiatives mentioned earlier are providing information that may ultimately influence regulatory agency practices and testing requirements.

The BPA consortium study is one of the most important ongoing efforts to improve the utility of not only NTP studies, but ultimately the current practice of toxicology for the protection of public health. As we move into the future and make choices about resource investments in traditional toxicology studies and the newer Tox21 approaches we will need to continue to develop an appreciation for the strengths and weaknesses of all of these approaches. An OHAT workshop carried out in January 2011 provided an excellent example of how these deliberations could be

informed (Thayer et al. 2012). Focused on examining the growing evidence for environmental agent contributions to the emerging epidemics of diabetes and obesity, we assembled scientists with a broad range of expertise including toxicology, medicine, epidemiology, endocrinology, statistics, high throughput screening (HTS), etc. While reviewing the more traditional lines of evidence linking various agents with these outcomes in animal toxicology and human epidemiology studies, we also provided them with some preliminary findings from the EPA ToxCast portion of the Tox21 consortium related to HTS chemical "hits" for biological targets that were considered relevant for these toxicity/disease pathways. The discussions produced a number of testable hypotheses and overall the HTS data evaluated in this type of meeting format was considered a useful approach.

What Does the Future Hold?

The NTP Roadmap was an important tool in developing a long-term strategy to support change in how public health decisions are informed by science. The Roadmap provided a means through which change could be accomplished based upon a careful evaluation of the scientific merit of a multitude of approaches that had and could in the future provide insight into the processes by which environmental agents might impact public health. The essential paradigm change evolving from the Roadmap is to use a series of inexpensive, rapid, informative assays to identify compounds that act in similar fashion and have evidence suggesting toxicity. The results of these assays could directly inform public health decisions if the information is sufficient, or guide more resource-intensive assays if data gaps need to be filled before the information can be used. Some of the initial scientific challenges of this paradigm have been addressed while others remain. The effort to address these will continue.

Many scientific and technological advances have allowed the NTP to address the goals of the Roadmap over the past 10 years, and some advances have been significantly stimulated by the Roadmap. Systems on a chip, inexpensive gene expression technologies, whole-genome sequencing methods, epigenetic assessments, HTS screens, genetically modified and diverse rodent models and new ways to computationally and systematically review and evaluate vast amounts of information have created a far different science than could be envisioned 10 years ago. The NTP and the field of toxicology must continue to evolve to take best advantage of these new tools and must not hesitate to replace old tools that are not as informative or more limited in scope. The NTP, as a national resource in toxicology remains dedicated to improving the science of toxicology in the service of public health. The 2004 Roadmap provided a powerful stimulus for change. The challenge is to intelligently utilize these changes to transform public health protection.

Acknowledgments

The authors are deeply indebted to all the NIEHS, FDA and CDC staff that have served the program through the years.

References

Andersen ME, Al-Zoughool M, Croteau M, Westphal M, Krewski D. (2010). The future of toxicity testing. J Toxicol Environ Health B Crit Rev. 13:163-96.

Auerbach SS, Shah RR, Mav D, Smith CS, Walker NJ, Vallant MK, Boorman GA, Irwin RD. (2010). Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. Toxicol Appl Pharmacol. 243:300-14.

Birnbaum LS, Bucher JR, Collman GW, Zeldin DC, Johnson AF, Schug TT, Heindel JJ. (2012). Consortium-based science: The NIEHS's multipronged, collaborative approach to assessing the health effects of bisphenol A. Environ Health Perspect. 120:1640-1644.

Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS. (2013). Implementing systematic review at the National Toxicology Program: Status and next steps. Environ Health Perspect. 121:a108-a109.

Boorman GA, Eustis SL, Elwell MR, Montgomery CA Jr, MacKenzie WF. Eds. (1990). *Pathology of the Fischer Rat.* Academic Press, Inc., San Diego.

Boorman GA, Blackshear PE, Parker JS, Lobenhofer EK, Malarkey DE, Vallant MK, Gerken DK, Irwin RD. (2005a). Hepatic gene expression changes throughout the day in the Fischer rat: implications for toxicogenomic experiments. Toxicol Sci. 86:185-93.

Boorman GA, Irwin RD, Vallant MK, Gerken DK, Lobenhofer EK, Hejtmancik MR, Hurban P, Brys AM, Travlos GS, Parker JS, Portier CJ. (2005b). Variation in the hepatic gene expression in individual male Fischer rats. Toxicol Pathol. 33:102-10.

Bucher JR. (2009). Guest Editorial, Bisphenol A, Where to Now? Environ Health Perspect. 117:A96-A97.

Bucher JR, Portier C. (2004). Human carcinogenic risk evaluation, Part V: The National Toxicology Program vision for assessing the human carcinogenic hazard of chemicals. Toxicol Sci. 82:363-366.

Bucher JR, Thayer K, Birnbaum LS. (2011). Guest Editorial, The Office of Health Assessment and Translation: A problem-solving resource for the National Toxicology Program. Environ Health Perspect. 119:A196-197.

Cesta MF, Malarkey DE, Herbert R, Brix A, Hamlin M, Singletary E, Sills RC, Bucher JR, Birnbaum LS. (2014). The National Toxicology Program Web-based Nonneoplastic Lesion Atlas: A Global Toxicology and Pathology Resource. Toxicol. Pathol. (in press).

Chapin RE, Sloane RA. (1997). Reproductive assessment by continuous breeding: evolving study design and summaries of ninety studies. Environ Health Perspect. 105 Suppl 1:199-205.

Chhabra RS, Bucher JR, Wolfe M, Portier C. (2003). Toxicity characterization of environmental chemicals by the US National Toxicology Program: An overview. Int J Hyg Environ Health. 206: 437-445.

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD et al. (2004). Complex Trait Consortium. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet. 36:1133–1137.

Churchill GA, Gatti DM, Munger SC, Svenson KL. (2012). The Diversity Outbred mouse population. Mamm Genome. 23:713-8.

Collins BJ, Grizzle TB, Dunnick JK. (2000). Toxicokinetics of phenolphthalein in male and female rats and mice. Toxicol Sci. 56:271-81.

Collins FS, Gray GM, Bucher JR. (2008). Transforming environmental health protection. Science. 319:906-907.

Dietrich DR, Aulock SV, Marquardt H, et al. (2013). Scientifically unfounded precaution drives European Commission's recommendations on EDC regulation, while defying common sense, well-established science and risk assessment principles. Chem Biol Interact. 205:A1–A5.

EFSA. (2013). Committee, Scientific opinion on the hazard assessment of endocrine disruptors: scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment, EFSA J. 11: 3132–3216.

EPA. (2013). http://epa.gov/ncct/download_files/edr/NMDR.pdf

Festing MF. (1995). Use of a multistrain assay could improve the NTP carcinogenesis bioassay. Environ Health Perspect. 103:44–52.

Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, Pethiyagoda CL, Stuve LL, Johnson FM, Daly MJ, Wade CM, Cox DR. (2007). A sequence-based variation map of 8.27 million

SNPs in inbred mouse strains. Nature. 448:1050-3.

Gohlke JM, Thomas R, Zhang Y, Rosenstein MC, Davis AP, Murphy C, Becker KG, Mattingly CJ, Portier CJ. (2009). Genetic and environmental pathways to complex diseases. BMC Syst Biol. 3:46.

Gore AC, Balthazart J, Bikle D. et al. (2013). Policy decisions on endocrine disruptors should be based on science across disciplines: A response to Dietrich et al. Endocrinol. 154:3957-3960.

Hoenerhoff MJ, Hong HH, Ton TV, Lahousse SA, Sills RC (2009) A review of the molecular mechanisms of chemically induced neoplasia in rat and mouse models in National Toxicology Program bioassays and their relevance to human cancer. Toxicol Pathol. 37:835-48.

King-Herbert AP, Sills RC, Bucher JR. (2010). Commentary: update on animal models for NTP studies. Toxicol Pathol. 38:180-1.

King-Herbert A, Thayer K. (2006). NTP workshop: animal models for the NTP rodent cancer bioassay: stocks and strains--should we switch? Toxicol Pathol. 34:802-5.

Maronpot RR. (Ed.) (1999). *Pathology of the Mouse.* Cache River Press, Vienna, IL.

Martin MT, Knudsen TB, Reif DM, Houck KA, Judson RS, Kavlock RJ, Dix DJ. (2011). Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol Reprod. **8**5: 327-339.

Merrick BA, Auerbach SS, Stockton PS, Foley JF, Malarkey DE, Sills RC, Irwin RD, Tice RR. (2012). Testing an aflatoxin B1 gene signature in rat archival tissues. Chem Res Toxicol. 25:1132-44.

Merrick BA, Phadke DP, Auerbach SS, Mav D, Stiegelmeyer SM, Shah RR, Tice RR. (2013). RNA-Seq profiling reveals novel hepatic gene expression pattern in aflatoxin B1 treated rats. PLoS One. 8(4):e61768.

NIEHS (2013) http://www.niehs.nih.gov/research/resources/bpa-related/index.cfm

NRC (National Research Council). (2006). Toxicity Testing for Assessment of Environmental Agents: Interim Report. Washington, DC. The National Academies Press.

NRC (National Research Council). (2007). Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. Washington, DC. The National Academies Press.

NTP. (2013a). http://ntp.niehs.nih.gov/?objectid=6B4371F1-F1F6-975E-798578C7A860D385

NTP. (2013b). http://ntp.niehs.nih.gov/index.cfm?objectid=B6813A81-92F9-4809-9FB4C5F76EFF5FA8

NTP. (2013c). http://ntp.niehs.nih.gov/?objectid=72016020-BDB7-CEBA-F3E5A7965617C1C1

NTP. (2013d). http://ntp.niehs.nih.gov/?objectid=7DA86165-BDB5-82F8-F7E4FB36737253D5

NTP. (2013 e). http://ntp.niehs.nih.gov/ntp/htdocs/levels/09-3566%20NTP-ITOX-R1.pdf

NTP. (2013f). http://ntp.niehs.nih.gov/ntp/htdocs/levels/09_3566_NTP_ReproTOX_R1.pdf

NTP. (2013g). http://ntp.niehs.nih.gov/ntp/Test_info/NTP_DevTox20090507.pdf

NTP. (2013h). http://ntp.niehs.nih.gov/?objectid=03C9B512-ACF8-C1F3-ADBA53CAE848F635

NTP CERHR. (2008). http://ntp.niehs.nih.gov/ntp/ohat/Bisphenol/Bisphenol.pdf

NTP. (2004a). Toxicology in the 21st Century: The Role of the National Toxicology Program. http://ntp-server.niehs.nih.gov/ntp/main_pages/NTPVision.pdf

NTP. (2004b). A National Toxicology Program for the 21st Century. http://ntp-server.niehs.nih.gov/NTP/About_NTP/NTPVision/NTProadmap_508.pdf

Parham F, Austin C, Southall N, Huang R, Tice R, Portier C. (2009). Dose-response modeling of high-throughput screening data. J Biomol Screen. **14**(10): 1216-1227.

Peddada SD, Kissling GE .(2006). A survival-adjusted quantal-response test for analysis of tumor incidence rates in animal carcinogenicity studies. Environ Health Perspect. 114:537-41

Portier C. (2001). Chipping away at environmental health risk assessment. Risk Policy Report. 8, 37-38.

Rao DB, Little PB, Malarkey DE, Herbert RA, Sills RC. (2011). Histopathological evaluation of the nervous system in National Toxicology Program rodent studies: a modified approach. Toxicol Pathol. 39:463-70.

Recio L, Kissling GE, Hobbs CA, Witt KL. (2012). Comparison of Comet assay dose-

response for ethyl methanesulfonate using freshly prepared versus cryopreserved tissues. Environ Mol Mutagen. 53:101-13.

Rudel RA, Fenton SE, Ackerman JM, Euling SY, Makris SL. (2011). Environmental exposures and mammary gland development: state of the science, public health implications, and research recommendations. Environ Health Perspect. 119:1053–1061.

Russo J, Russo IH. (1996). Experimentally induced mammary tumors in rats. Breast Can Res Treat. 39:7-20.

Schug TT, Heindel, JJ, Camacho L, Delclos KB, Howard P, Johnson, AF, Aungst J, Keefe KD, Newbold R, Walker NJ, Zoeller RT, Bucher JR. (2013). A new approach to synergize academic and guideline-compliant research: The CLARITY-BPA research program. Reprod Toxicol. 40:35-40.

Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A. (2011). Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect **.119**: 364-370.

Simmons PT, Portier CJ. (2002). Toxicogenomics: the new frontier in risk analysis. *Carcinogen.* 23, 903-5.

Thayer KA, Foster PM. (2007). Workgroup report: National Toxicology Program workshop on Hormonally Induced Reproductive Tumors - Relevance of Rodent Bioassays. Environ Health Perspect. 115:1351-6.

Thayer KA, Heindel JJ, Bucher JR, Gallo MA. (2012). Role of Environmental Chemicals in Diabetes and Obesity: A National Toxicology Program Workshop Report. Environ Health Perspect. 120:779-789.

Thomas RS, Black MB, Li L, Healy E, Chu TM, Bao, W, Andersen ME, Wolfinger WE. (2012). A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. Toxicol Sci**. 128**: 398-417.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. (2013). Improving the human hazard characterization of chemicals: A Tox21 update. Environ Health Perspect. 121:756-765.

Waters M, Stasiewicz S, Merrick BA, Tomer K, Bushel P, Paules R, Stegman N, Nehls G, Yost KJ, Johnson CH, Gustafson SF, Xirasagar S, Xiao N, Huang CC, Boyer P, Chan DD, Pan Q, Gong H, Taylor J, Choi D, Rashid A, Ahmed A, Howle R, Selkirk J, Tennant R, Fostel J. (2008). CEBS--Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res. 36: D892-900.

Witt KL, Livanos E, Kissling GE, Torous DK, Caspary W, Tice RR, Recio L. (2008). Comparison of flow cytometry- and microscopy-based methods for measuring micronucleated reticulocyte frequencies in rodents treated with nongenotoxic and genotoxic chemicals. Mutat Res. 649:101-13.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, Jadhav A, Smith CS, Inglese J, Portier CJ, Tice RR, Austin CP. (2008). Compound cytotoxicity profiling using quantitative high-throughput screening. Environ. Health Perspect. 116 (3): 284-291.

Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt KL, Inglese J, Tice RR, Austin CP. (2009). Identification of chemical compounds that induce HIF-1alpha activity. Toxicol Sci. 112(1): 153-163.

Zoeller RT, Brown TR, Doan LL, Gore AC, Skakkebeck NE, Soto AM, Woodruff TJ, Vom Saal FS. (2012). Endocrine-disrupting chemicals and public health protection: A statement of principles from The Endocrine Society. Endocrinol. 152: 4097-4110.