Response to the NCTR SAB Subcommittee Review of the Division of Bioinformatics and Biostatistics

Weida Tong, Ph.D
Director, Division of Bioinformatics and Biostatistics

We appreciate the thorough examination and review of the Division of Bioinformatics and Biostatistics (DBB) research by the Science Advisory Board (SAB) Subcommittee chaired by Dr. Katrina Waters and Dr. Pamela Lein, Chair and co-Chair, respectively. Special thanks to Drs. Cynthia Afshari, Ying Lu, John Quackenbush, and Kenneth Ramos who were the subject matter experts on the subcommittee and spent a tremendous amount of time to provide constructive comments during and after the division review. Both Drs. Waters and Lein are members of NCTR SAB, and Dr. Afshari is member of FDA's Science Board. We also thank the representatives from FDA product centers that participated in the division review.

We appreciate the overall positive comments and we take seriously the constructive recommendations in the report; the thought and effort that went into the preparation of the report by the Subcommittee members are obvious. The comments and recommendations will inform us of possible changes as well as new initiatives and strategies to strengthen the division's role in supporting both the NCTR and FDA's mission. Our reply follows the original structure with "point-by-point response" style, where the text from the SAB Subcommittee is in *Italics* while our response is in plain text.

Theme 1: Precision Medicine

- **Comment (general)**: *The Subcommittee recommends that future plans emphasize issues of toxicological significance, such as individual differences in susceptibility to toxicity and disease, response to pharmacological treatment, and adverse drug reactions.*
    - **Response:** We appreciate very much the comment and agree totally with the Subcommittee. We have two main precision medicine projects in the division, (1) MAQC/SEQC consortium to assess the reliable use of genomics technologies in regulatory decision-making and (2) repurposing marketed drugs for the treatment of rare diseases. Both projects are taking toxicological significance into consideration. For the rare disease project, we pay specific attention to drug safety. For example, there are safety concerns of using oncologic drugs for the treatment of rare diseases, a new concept we have developed and recently published (see #1 in Support Materials). We are developing methodologies to understand how the dosage and route should be modified to change safety profiles of cancer drugs in this type of application. Re the MicroArray Quality Control (MAQC) consortium projects, since we started MAQC in 2005, toxicological significance has always been one of the essential considerations in all the MAQC projects. For example, MAQC 1 applied the MAQC main findings to

a toxicogenomics study design involving both liver and kidney carcinogenicity, MAQC2 assessed microarray-based predictive models for liver carcinogenicity, and MAQC3 compared RNA-seq with microarrays in toxicogenomics (see #2-4 in Support Materials). The toxicological significance has continually be a part of focus in the new MAQC project (MAQC4), with aims at being able to detect the genetic bases for disease and adverse events.

- **Comment (Rare Diseases)**: *A lack of a clear definition of anticipated impacts and relevance to the Agency mission.*
    - **Response:** Out of ~7000 known rare diseases, only around 500 orphan drugs are available. In the past 5 years, 64 orphan drugs were approved by the FDA, representing 35% of all 182 drugs approved during this period with 47% in 2015 and 41% in 2014. Falling under different provisions of the Orphan Drug Act, these approvals should alleviate suffering among some people with terrible diseases; our research aims at addressing a very important question of how to improve therapeutic options for rare disease through the bioinformatics approaches. The private sector has limited financial incentive to invest in this area, and thus it becomes an important responsibility under the broad FDA's mandate to protect and promote public health. To make our research relevant to regulatory applications, we are actively engaging the scientists/reviewers from the regulatory centers to seek input.

- **Comment (SEQC2)**: (1) *The Subcommittee recommends sharpening of focus and closer alignment to the Agency's mission to facilitate continued maturation of the Division. (2) Various FDA centers have encountered NGS data in regulatory applications, closer alignment between FDA needs and the Division activities is highly desirable.*
    - **Response:** Multiple FDA regulatory centers have encountered next-generation sequencing (NGS) data in regulatory applications. We have been working closely with many reviewers and scientists from all FDA centers on refining the SEQC2 study design and objectives accordingly. The project now consists of three specific aims: (1) to develop quality metrics for reproducible NGS results from both whole genome sequencing (WGS) and targeted gene sequencing (TGS), (2) to benchmark bioinformatics methods for WGS and TGS towards the development of standard data analysis protocols, and (3) to assess the joint effects of key parameters affecting NGS results for enhancing understanding of fit-for-purpose applications with NGS. Realizing the rapid development in this field, the initial emphasis of the project will be placed on evaluating the challenges that are outlined by the FDA's discussion papers entitled, "Developing Analytical Standards for NGS Testing" and "Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper" along with other FDA internal discussion documents to fulfill the immediate needs in the FDA for review of NGS data. We believe these efforts align with the Subcommittee's recommendation of continual sharping of focus and closer alignment to FDA mission and needs. Actually, such alignment has

been integral with earlier MAQC projects; for example, the findings and recommendations from the first MAQC project were incorporated into the draft companion Guidance for Industry on Pharmacogenomics Data Submissions. This draft Guidance is being updated and incorporated into the parent Guidance as the final Guidance to the industry. We are participating in this revision effort so that findings and recommendations from SEQC/MAQC3 about RNA-seq can be incorporated.

- **Comment (SEQC2)**: **(1)** *Caution should be exercised in making investments to define best practices and standards in a fast evolving field. The team should not lose sight of the fact that the ultimate goal is to systematically evaluate quality metrics and standard practices using comprehensive and diverse datasets. (2) In addition to establishing the performance characteristics of different techniques, it will be critically important to develop relative ranks and amount of variations introduced by different sources.*
    - **Response:** We appreciate the Subcommittee's recognition of the fast evolving nature of these emerging technologies and the related challenges of assessing them. As suggested, we will focus on systematically evaluating quality metrics and standard practices using comprehensive and diverse datasets. Understanding the sources of variations and their impact has been an emphasis throughout earlier MAQC projects. Specifically, we plan to evaluate these sources of variations such as sample preprocessing, gene capture panels, library preparation, and sequencing instruments. We are interested in studying these parameters with various biological samples. We will maintain this focus in SEQC2. While not a goal per se, the process we engaged in, and continue to do so, inherently provides insights into what practices may enhance or deleterious affect validity and reliability.

- **Comment (SEQC2 Collaboration)**: (1) *The Biostatistics Branch has the expertise in experimental design and should be more tightly integrated at the planning stage to design more efficient experiments and to facilitate development of innovative statistical designs. (2) Clear evidence was provided of the collaborative nature of the program, but the rationale for selection and identification of partners was not provided. Plans to promote the diversity in composition of the proposed working teams were judged to be a useful mechanism to ensure that the regulatory science goals are incorporated into the design of the study. Also promising was the intent to solidify the details of the experimental design based on community-wide input. Stakeholder involvement and workshops will be used to evolve ideas, embrace needs of the community and be more responsive. This approach will help advance efforts toward regulation without stifling innovation, a difficult balance to achieve.*
    - **Response:** Thank you for the positive comments. Collaboration is central to the success of SEQC2, as it has been the three preceding MAQC/SEQC phases. In the past we utilized many venues such as a federal registry notice, scientific conferences and workshops, MAQC consortia, FDA webpages, and social media to engage stakeholders and built consortia with diverse partners from academia,

industry and government agencies across the world. A sign of the success in a diverse, inclusive collaboration is the intensity of debates that have accrued, but which eventually evolved into broad agreement. We promote open participation as a common practice for the government-funded project to ensure transparent assessments of technologies. We have been involving the Biostatistics Branch for study design, data collection, and data analysis. The biostatisticians have been working with SEQC2 scientists and participating in the SEQC2 working group.

Theme 2: Predictive Toxicology

- **Comment (general):** *Predictive toxicology is one of the main research areas within the Division. The primary objective of the program is to improve drug safety by understanding and predicting adverse events through computational modeling and biomarker development. The Predictive Toxicology program has a diverse research program, ranging from curation of data to assembling knowledge bases relevant to interpreting toxicology assays to developing and testing biomarkers. The Division has established itself as a leader in toxicogenomics research and created resources. Research in predictive toxicology has been a major focus of the Division and general, good examples of what the Division could do as many address questions relevant to the FDA's regulatory and health assurance mission.*
    - **Response:** Thank you for general positive comments regarding our predictive toxicology efforts.

- **Comment (LTKB-RO2)**: **(1)** *The RO2 rule for assessing the liver toxicity of orally administered drugs represents a place where the Division could have done more, looking to leverage what was learned in this effort to other drug administration routes, to studying other compounds that cause liver damage, or to incorporating other data and information (such as chemical structure) that is recorded in the LTKB. (2) It was not clear that the Division had worked closely with the other FDA divisions to assess using the RO2 as part of the regulatory process.*
    - **Response:** Thank you for the affirming comments regarding the potential of RO2. We tried to show the Subcommittee the utility of RO2 in the FDA during the review. We co-chair the Liver Toxicity Interest Group at the FDA, which consists of over 30 FDA scientists and reviewers with most of them from CDER. We have communicated the LTKB results with the interest group and they have utilized our results in the review process. More specifically, as of now, we have been asked to evaluate liver toxicity in 16 submissions using RO2. The submissions are across IND, Phase I, Phase II, Phase III, and NDA submissions. This Interest Group has succeeded in bringing together reviewers and researchers. As an outcome, we have co-authored with a CDER reviewer a report on the success of the RO2 rule in predicting hepatotoxicity potential of direct-acting antivirals for treatment of chronic hepatitis C. In additional, we have also developed a bioinformatics tool to assist the reviewers in applying the RO2 rule and more generally the Liver Toxicity Knowledge Base (LTKB) in their regulatory review.

- **Comment (MicroRNAs as biomarkers for hepatotoxicity):** *The analysis allowed the group to identify a candidate "best practices" data analysis pipeline, but the small sample size and the overall experimental design was not sufficient to find robust candidates. This project could serve as a springboard for larger more directed studies or for the analysis of miRNA seq data from other projects across the FDA, and eventually this might lead to the development of a predictive set of circulating miRNAs that might help predict liver damage.*

- **Response:** We do apologize for not making the objectives of this project clearly. The primary objective of the miRNA project is to understand whether expression profiles of miRNAs in the rat liver tissue can be used as mechanistic biomarkers of human DILI (drug-induced liver injury) predictive value. We chose next generation sequencing (NGS) for miRNA expression profiling for its potential in discovering novel miRNAs. However, NGS has not been widely used for miRNA profiling, such that a standard analysis practice has yet to be established, particularly in application to the toxicology studies which involve multiple doses and time points and thus complicated dose response curves and time-dependent patterns. That is why our first step is to determine which bioinformatics pipeline would best meet our needs. At the time of this review, this work was in the first step in which we had carried out a moderately comprehensive assessment of various data analysis pipelines. However, this is not the main objective of the project. As outlined in our presentation slide at the review (see #5 in Support Materials), this first phase will be followed with Phase II to discover miRNA biomarkers for liver carcinogenicity, Phase III to discover miRNA biomarkers for drug induced liver injury, and Phase IV to confirm these biomarkers. The findings are expected to beneficially inform our larger and more inclusive efforts in advancing DILI prediction accuracy. Recently we have written two manuscripts (one submitted to Arc Tox, see #5 in Support Materials) from this project. The first manuscript reports our discovery of miRNA biomarkers for liver carcinogenicity and the second for DILI biomarker identification. We are aware that many studies have been focused on ***circulating*** miRNAs as non-invasive diagnostic biomarkers for organ injury. However, this is not the focus of our project. We choose to study rat liver ***tissue*** miRNA expression as mechanistic biomarkers that may convey better translational potentials from rats for predicting DILI in humans.

- **Comment (EDKB)**: *The EDKB project represents an important contribution to the overall FDA mission. The information gathered is essential for the development of predictive models of response to endocrine disruptors, and additional work to use this resource to develop robust, predictive, quantitative models has the potential to be of broad interest and use across the FDA and beyond.*
  - **Response:** Thank you for the positive comments. Over a 17 year history, a huge diversity of models have been built and published, many with a large number of citations. These range from three-dimension quantitative, Comparative Molecular Field Analysis, to chemometric classification models, ensemble models, and docking models. Currently, the EPA has a large collaboration underway that includes predictive toxicology in which we are participating.

- **Comment (Toxicogenomics): (1)** *The group has done a great deal of work to assemble a large toxicogenomics dataset comprising more than 15,000 samples and profiling with 131 drugs using three different assay protocols (in vitro assays, in vivo repeated dose studies, and in vivo single dose studies). (2) It was difficult to extrapolate the in vitro*

results to the in vivo response, but that careful experimental design might help address this problem.

- **Response:** Thank you for recognizing the importance of extrapolating in vitro results to in vivo. We understand the challenge. We are working with Drs. Jurgen Borlak (Hannover University, Germany) and Ruth Roberts (Apconix, UK) in this field by taking into consideration the specific study design of the data set we are using to address the extrapolation problem. As a matter of fact, our first manuscript to describe this effort is about to be submitted (see #7 IN Support Materials).

- **Comment (Overall)**: (1) *The Predictive Toxicology program has made important contributions, the most significant of which are the knowledge bases that they have assembled. (2) The Division should consider working more closely with the regulatory Centers within the Agency to explore how they could use the information captured in these knowledge bases to develop applications that could help inform the regulatory process and advance the broader mission of the FDA.*
  - **Response:** Thanks for the positive comments in recognizing the significance of our knowledge bases. Again we apologize for not making their regulatory application clear in our presentation due to limited time. We strive to work closely with the regulatory Centers by forming interest groups, e.g., FDA liver toxicity interest group. This is a mechanism we use to disseminate our research outcomes to be considered for application in the regulatory settings. This particular mechanism has been performing very well. As mentioned above, we have co-authored with a CDER reviewer a report the success of the RO2 rule in predicting hepatotoxicity potential of direct-acting antivirals for treatment of chronic hepatitis C. In addition, we have worked closely with Center for Tobacco products to build a database containing 8700 tobacco product chemicals, which has powerful search and annotation capability.

- **Comment (biomarker): (1)** *The work on biomarkers is also interesting, but the sample size and experimental design are too limited to draw firm conclusions. (2) This exploratory work might represent a space where the Division could work closely with programs such as the National Toxicology Program to collect relevant data from animal testing being performed by that Program.*
  - **Response:** Thank you for the great suggestion. We always strive to build more collaborative relations with related research programs. We have approached the National Toxicology Program on the study of DILI biomarkers. In fact, we are developing a biomarker project that has been communicated with the National Toxicology Program (NTP) with potential for funding. This project was included in the review as a poster presentation. The project aims to evaluate the utility of in vitro predictive toxicology biomarkers for predicting the DILI potential of drugs and chemicals.

- **Comment (general):** *The focus and closer integration with the other programs within the FDA would only serve to make the predictive toxicology program stronger and even more relevant to the FDA's and NCTR's missions.*
    - **Response:** We completely agree with the Subcommittee and we have been striving for that. For that reason, we developed the R2R (Research to Review and Return) program to enhance the interaction between our research and FDA review activities. The importance of the R2R program has also been recognized by the FDA. The R2R program has just been awarded the highly prized Commissioner's Special Citation for its innovative cross-Center bioinformatics projects benefitting regulatory business processes.

Theme 3: Biostatistical Approaches and Applications

- **Comment**: *A total of 6 FTE, or 67% of Branch's effort is devoted to supporting of other NCTR Divisions with a wide range of study types. We consider the support efforts to be critical and appropriate to the mission of NCTR. We would recommend that this support effort be recognized as a key performance measure for the Branch and those significant contributions to statistical support of collaborative and team research efforts be considered an important metric for promotion consideration.*
  - **Response:** Statistical support and collaboration have been an essential responsibility for the Biostatistics Branch. Four members of the branch members are supporting the National Toxicology Program and there are established performance metrics to assess the performance of their support effort. For other members in the branch, their support role is mainly recognized through co-authoring the work they support. Both mechanisms are essential components in promotion consideration.

- **Comment**: *The Biostatistics Branch has four research areas: (1) risk factor identification and characterization; (2) statistics and data mining for large-scale data inference; (3) foodborne pathogens genomics knowledgebase; and (4) health risk assessment methodology. Many of the projects represent collaborations within NCTR and within FDA. These are important research areas and critical to the mission of the FDA. The group has expertise and significant achievements in all these areas.*
  - **Response**: We appreciate the generally positive comments that our research areas are important and critical to the mission of the FDA.

- **Comment**: *It was not clear how the Biostatistics Branch sets research priorities for Branch scientists and statisticians.*
  - **Response:** The Biostatistics Branch scientists set research priorities based on the following criteria: FDA impact; intra-Division and inter-Division collaboration; the scientist's research interests and analytical strengths; and research opportunities established through the Office of Chief Scientist (OCS), which foster inter-Center collaboration, and additional outside funding.

- **Comment**: *It seems many of these are initiated in an ad hoc manner by investigators based on their own research interests and/or historical collaborations.*
  - **Response:** Research scientists in the Biostatistics Branch vary in their research interests and bring unique talents to the division. This diversity is what allows our small team the ability to meet the various analytical needs of NCTR. Indeed, many of the research projects were initiated as a result of work with other FDA centers in order to meet a specific research need since all research must demonstrate FDA significance. In recent years, several projects were developed based on OCS proposals. The Branch has received two FDA Office Women's Health research awards (2011-2013 and 2014-2016) and one CDER funded collaborative project.

- **Comment**: *There was negligible discussion on work in research area 4 (Health risk assessment methodology), which is the traditional strength of this group and very much related to the core mission of the NCTR.*
  - **Response:** The NCTR Biostatistics program has been recognized as experts in quantitative health risk assessment methodology before 2007. Since then, two colleagues with extensive risk assessment expertise have left NCTR (one retired in 2005 and the other took a faculty position in 2007). Currently, contemporary statistical research favors methods for analyzing molecular data. As a result, the Biostatistics Branch has shifted its research focus to precision medicine and machine learning methodologies for high dimensional and big data analysis. Regarding research in health risk assessment methodology, the Director of this group has collaborated with former colleagues, as a co-investigator, to develop a dose response model for estimating infectious doses for microbial pathogens (and a statistical model for calculating benchmark dose for categorical response data, both of which led to publications. Since 2000, he has written several review articles regarding risk assessment.

- **Comment**: *Also the link between next generation sequencing (NGS) and the 4 research areas was not obvious. [four research areas are (1) risk factor identification and characterization; (2) statistics and data mining for large-scale data inference; (3) foodborne pathogens genomics knowledgebase; and (4) health risk assessment methodology.]*
  - **Response:** Several FDA centers have encountered NGS data for various FDA regulatory applications. In recent research, NGS data analysis is connected closely with the research areas (1), (2) and (3). The Biostatistics Branch has developed two projects involving NGS applications and data analysis. The first project is a protocol to develop statistical methods for RNA-seq analysis including gene set analysis, sample size determination, and outlier sample identification, which links to research areas (1) and (2). The second project develops biostatistics and bioinformatics pipelines to *Salmonella* NGS data analysis and data mining methods for foodborne pathogen detection and source tracking, links to research areas (2) and (3).

- **Comment**: *Strategic planning within the Branch to identify research directions and priorities is strongly recommended.*
  - **Response:** The field of biostatistics proceeds in tandem with advances in biotechnology. Biostatisticians constantly learn fundamental concepts of new biotechnology and develop new/improve methods for data analysis. Traditional statistical test procedures were developed for risk factor identification. However, today, statistical models and test procedures are developed to identify biomarkers for high dimensional molecular experiments. Thus, research for the Biostatistics Branch will focus on two main areas: (1) statistical and machine learning for high dimensional data and big data and (2) statistical methods for precision medicine. New projects are subjected to Division-level and Center-level reviews. The Center-

level review process encompasses the solicitation of comments from the Center Director from individuals at FDA product centers in order to ensure that projects meet the needs of the Agency.

- **Comment: (1)** *Promotion of Research Biostatisticians based on methodological publications and as Co-PIs on collaborative projects, which is a good practice. (2) It is important to protect independent research time for statisticians working in collaborative team research efforts so that they can push the envelope beyond conventional approaches. Many efforts in collaborative research are opportunities to develop new methods.*
    - **Response:** We agree with these comments. Statistical consulting and collaboration have been an important mechanism for statisticians to learn new technology and generate new research ideas. For example, a research statistician published a paper on outlier detection for RNA-Seq data, which resulted from a collaborative project with a PI in the Division of Neurotoxicology. Research statisticians in the Branch are mandatorily evaluated at five year periods of cyclical review under the NCTR Research Peer Review System. They may request to be reviewed early for promotion. My experience is that the peer review evaluation considers both primary author and co-author publications. Two branch research statisticians were promoted recently (2012 and 2015).

- **Comment:** *It is important to have statistical contributions to science be counted as a key performance measure for promotion. The contribution to science by statisticians is more significant than the methodology itself.*
    - **Response:** We fully agree with the comment. A statistical method used in the analysis may be simple, but the contribution to the science in the study may be still significant. Unfortunately, this may not be recognized in the evaluation.

- **Comment (*Big data research):*** *The committee strongly encourages to actively engaging in big data. We noticed that the Branch has been involved in big data research activities such as the study of the AE reporting system, topic modelling as an unsupervised leaner, text mining, etc. As mentioned by the CBER representative, the FDA has a program to access public health care databases (including data from the Department of Veterans Affairs), which can complement the FDA AE reporting system. The information of toxicity data in NCTR and pre-marketing trial data in FDA will also be great data sources. Integration of these data will definitely be a big data research project and can provide insights on drug safety, including multiple drug interactions. The Biostatistics Branch is encouraged to collaborate with the System Biology Division in the NCTR and other statisticians in the CDER to engage in big data effort.*
    - **Response:** As noted by the Subcommittee, we have been involved in big data research activities such as the study of post-market adverse event reporting data, the application of topic modeling, etc. Following the committee's suggestion and FDA's research strategy, we are turning our research priority to big data analysis. We have just developed three proposals and submitted them

to the FDA Office of Chief Scientist for internal funding consideration. Big data analysis is the key component of addressing different endpoints in these proposals. In addition, we totally agree with the suggestions of the committee, and we are including the VA hospital systems as one of the EHR (Electronic Health Records) data sources and collaborating with CDER, CVM, and CDRH colleagues in an effort to use big data to promote regulatory science.

- **Comment (Additional research opportunities):** *NCTR has state-of-the-art imaging equipment and experimental data. The Subcommittee considers "radiomics" and/or imaging related research a potential missed opportunity for the Biostatistics Branch to develop novel statistical approaches while supporting both the NCTR and Agency missions.*
    - **Response:** We appreciate the recommendations of the Subcommittee about research in medical imaging data. We are currently developing a protocol on quantitative techniques for analyzing bioimaging data. We have discussed our ideas with the NCTR expert in MRI imaging in the Division of Neurotoxicity. In addition, NCTR is setting up a bioimaging analysis core. We will participate in the bioimaging core and provide statistical support on bioimaging data analysis once the core is formed.

- **Comment:** *Expand research collaborations with the Bioinformatics Branch within the Division on genomic sequencing, with the System Biology Division within the NCTR on safety prediction and drug interactions, and biostatisticians in the CBER/CDER/CDRH on precision trials.*
    - **Response:** We have been involved in LTKB and NGS projects. We also plan to utilize research statisticians and their expertise in the preplanning phase of analysis (e.g. experimental design) and in the analytical phase to provide support for future projects in the Bioinformatics Branch. The Biostatistics Branch has provided statistical support for several projects and as co-investigators from the Division of System Biology (DSB). One Branch protocol involves a PI from the DSB to develop data mining techniques for prediction of drug-induced cardiovascular adverse events.

- **Comment (Succession plan, recruitment and retention):** There were *expressed concerns about recruitment and retention of talented statisticians. It is important to develop a succession plan for the Biostatistics Branch, including the Branch Director as well as statisticians at all levels. Statisticians are in demand nationwide and extra efforts have to be made to attract talent to the NCTR. We recommend the NCTR senior leaders work with the Branch to develop a comprehensive plan with necessary resources to attract both senior and junior statisticians, to address performance review processes for retention, and to enhance diversity in the Branch.*
    - **Response:** The field of Ph.D. statisticians currently has a supply problem, with many more job openings than candidates to fill them. The FDA as a whole and our Biostatistics Branch has continually struggled to recruit and retain highly

qualified mathematical statisticians at all levels. We are very grateful that NCTR senior leaders have been very supportive in recruiting FTE statisticians. Thanks for the suggestions and we will work with the NCTR senior leaders to develop a plan for hiring and retention of talent individuals.

Theme 4: R2R Framework & Activities

- **Comment: (1)** *The goal of this program is to provide an informatics and analytics framework that enables the FDA product reviewers to become more efficient in their reviews. (2) In this end, this should provide efficiency to the review process, more effective flagging of issues in sponsor data, and enable more rapid decision-making with respect to products.*
    - **Response:** We agree that this is one of the important goals of the R2R projects. We appreciate the Subcommittee's recognition of this goal and the importance of the R2R Framework in general. We plan to manage and execute R2R projects to better align with this goal.

- **Comment (Strengths): (1)** *This is a major project with a goal that is wholly integrated with supporting the FDA initiative of evolving FDA's regulatory science. Its goal to provide an informatics structure to sponsor data, FDA reviewer documents and drug labels will facilitate future research and data mining for multiple purposes. Significant progress has been made in a very short time (approx. 2 years). (2) This project is highly integrated and collaborative with other product centers within FDA. (3) The reiterative, recursive, nature of this project that is designing with the reviewer in mind is poised to quickly deliver important tools to the FDA reviewers and also to the broader FDA research community. The flow of tool creation and feedback will ensure that the tools will be impactful and useful and will not become obsolete shortly after they roll out. (4) The focus on "end user" and "user friendly" tools is another advantage for a highly impactful product. (5) The development of application of text mining tools is a unique aspect of this project and will provide an important foundational tool for the R2R workflows.*
    - **Response:** We appreciate the detailed, insightful and positive comments of the SAB subcommittee towards the R2R program.

- **Comment (Areas for Development):** *Since the project is so focused on the product reviewers, it might be helpful for some of the fellows within the NCTR Biocomputing division to spend some time doing rotations working from the White Oaks campus.*
    - **Response:** Thank you for the fantastic idea. As a matter of fact, this suggestion has also been brought up independently by some product reviewers that we have been working with for the R2R projects. We very much agree with the comment and are exploring options. However, we do need to work within some budget constraints, particularly the travel budget.

- **Comment:** *This project may benefit from a business analyst/project manager that could provide expertise for planning and prioritizing projects with the mindset of "return on investment".*
    - **Response:** We agree with the Subcommittee. We recognize the benefit to have a coordinated effort for planning and managing R2R projects to better realize the return on investment. We have been discussing with the NCTR management about possible reorganization to better address this need.

- **Comment:** *NCTR scientists should look for solutions to the needs of the project from sectors in industry and academia beyond the internal capabilities within NCTR.*
    - **Response:** We agree with the suggestion. DBB is a member of Arkansas Bioinformatics Consortium (AR-BIC) which consists of the major universities in Arkansas. We will leverage this activity to explore new resources to support our R2R effort.

- **Comment:** *It is recognized that there are some challenges with this, especially the lack of "nimbleness" in government contracting, but NCTR management is encouraged to try to find a way to ensure this can be pursued.*
    - **Response:** Thank you for the suggestion.

- **Comment:** *This is an important project that will develop tools for the FDA reviewer and impact the FDA goal to enhance regulatory science. As such it will be important to consider the right metrics for tracking impact of the output.*
    - **Response:** Thank you the positive comments. Indeed the importance of the R2R program has been recognized by the FDA. Several projects have won FDA Honor Awards. For example, the R2R program has just been awarded the highly prized Commissioner's Special Citation for its innovative cross-Center bioinformatics projects benefitting regulatory business processes. The FDALabel team (the project under R2R) won a 2016 Scientific Achievement Award from the Office of Chief Scientist for outstanding inter-center scientific collaboration. The Patient Narrative Analysis team was recognized for developing innovative tools to manage and analyze unstructured patients narratives which are an important part of clinical submission. We have developed some metrics for tracking the impact. The focal point of the metrics is to what extend the tool or service has been adopted in the regulatory review process.

- **Comment:** *It is important to note that this work may not always result in peer reviewed publications and therefore, metrics appropriate to inform NCTR career promotion decisions is important to consider for those staff working on this project.*
    - **Response:** We totally agree with the comments. As mentioned above, we are in discussion with NCTR senior leaders for reorganization to better manage the R2R projects and those staff working on R2R project. The reorganization has several benefits: 1) the support function will be better coordinated and supervised, improving efficiency and productivity; 2) job responsibility and career development path will be better clarified; and 3) the oversight of current projects and identifying future needs will be improved.

Theme 5: Service & Support Functions

- **Comment:** *Scientific Computing and the Biostatistics Branches, encompassing approximately 60% of the total DBB activities. The semantics of what qualifies as "support" and what qualifies as "research" was confusing throughout all of the presentations. The inclusion of these activities into the calculation of "support" skews the perception of balance between support and research and should be reevaluated to be consistent with other divisions within NCTR to ensure that all Division activities serve the mission of the Agency.*
  - **Response**: We apologize for not making it clear about the distinction between "support" and "research" in the division and glad to have this opportunity to clarify this issue. Re "support", the major "support" activity is in Scientific Computing Branch and Biostatistics Branch. The Scientific Computing Branch (18 full time positions) provides database administration, management and support of legacy applications, software development, IT investment and portfolio management, IT quality control and IT liaison functions which are all 100% for the support of the NCTR community and research mission. Thus, this branch is 100% in "support". Within the Biostatistics Branch, 4 FTEs supports the National Toxicology Program and 100% of their activities are considered as "support". In addition, the research scientists in both Biostatistics and Bioinformatics Branches provide both biostatistical and bioinformatics supports, including setting up Galaxy platform for analysis of next-generation sequencing, job scheduling for conducting data analysis in high performance computing environment, developing customized code and software for specialized purpose, and conducting training of bioinformatics tools in addition to routine data analysis support when requested by the scientists at NCTR and beyond (i.e., other centers at FDA). This activity consumes many FTEs, which is considered as "support". Altogether, around 60% of human resource engages "support" to the center-wide research and the agency-wide research/regulatory activities. Re "research", any activities towards working with approved research protocols is considered as "research". The NCTR has a standard protocol approval mechanism based on which the approved protocols become a part of NCTR budgeting process. With that said, in some situations, the line between "support" and "research" is blurring. For example, we had one project funded by FDA's Center of Tobacco Products to develop a knowledgebase for tobacco constituents. This project involves both conventional IT skills (i.e., database administration) and chemoinformatics research. Thus, we consider this activity covering both "research" and "support".

- **Comment (Scientific Computing Branch):** *It would be worth exploring how Reach could be opened up to the rest of the FDA and provide a cornerstone of communication strategy for the entire agency.*
  - **Response**: Current plans call for this to be discontinued in favor of a web solution promoted by FDA. The Agency solution will enable customized pages for each

laboratory and will be accessible by all FDA staff. This has the added benefit of being managed, documented and funded at the Agency level.

- **Comment (Scientific Computing Branch)** *There was an apparent lack of infrastructure support for the Bioinformatics Branch, other than hardware support.*
    - **Response**: The Scientific Computing Brach together with on-site Office of Information Management and Technology staffs provide various IT support to the Bioinformatics Branch such as setting up high-performance computing for data intense study, maintaining database for ArrayTrack and other software developed by the Bioinformatics Branch. With that said, Scientific Computing Branch and on-site OIMT is a part of basic investment of NCTR to support the entire center. Specifically, both groups provide IT support such as database management and programming support for all NCTR Divisions and Offices including both sibling branches in the Division of Bioinformatics and Biostatistics. This arrangement has a benefit for enhanced IT security which has become a major consideration in the government operation. In addition, it also provides a platform for better use of IT resource at NCTR.

- **Comment (Scientific Computing Branch):** *The value of customized software should be balanced with the amount of effort that is spent supporting legacy applications, some of which are under-utilized.*
    - **Response**: HHS and FDA have adopted a government or commercial off-the-shelf (GOTS or COTS) first approach to eliminate redundant, outdated and underutilized software. With the governance of the Center IT Board, NCTR will follow this lead. The SCB has performed an inventory of the applications and databases hosted on the NCTR servers. As the portfolio of legacy software is rationalized and requests for new applications are received, solutions available elsewhere in FDA, open source products and COTS alternatives will be evaluated. Buy versus make has and will prove to be challenging as many of the legacy applications are linked via shared data structures/databases and a change to one may require a simultaneous change to another to ensure interoperability and access to historic data.

- **Comment (Scientific Computing Branch):** *The Division should explore if any of their current software tools could be replaced with publicly or commercially available alternatives that are cheaper to license than the investment required to maintain or fix custom software over the long run.*
    - **Response:** This comment is well received and timely. As part of the Enterprise Performance Lifecycle (EPLC) and Capital Planning and Investment Control (CPIC) policies, HHS and FDA have begun requiring alternatives analysis for all system development and software purchases. Under the governance of the NCTR Center IT Board, existing applications and new requests will be reviewed and an alternatives analysis performed. While the Scientific Computing Branch and other IT staff will propose alternatives, acceptance by the research and

management staff (i.e., customers) will be essential. Experience has taught us that COTS or GOTS solutions normally require some amount of customization, which can greatly increase their cost and extend their implementation timeline.

- **Comment:** *In the emerging era of Precision Medicine, integrating the needs for privacy and HIPA compliance into the NCTR computing plan is essential. Given that it takes multiple years for strategic planning and capital investment, it is recommended that NCTR stay on the forefront of technology, including migration to cloud computing platforms, enable high performance computing capabilities and increased bandwidth with infrastructure modernization through 10 gigabit/second or 100 gigabit/second connections, with strategic planning that looks beyond the 2-3 year time frame.*
    - **Response:** NCTR continues to work with FDA's Office of Information Management and Technology as well as the Chief Technology Officer to pursue increased bandwidth and access to cloud platforms such as AWS and Salesforce. Many of the security concerns have been addressed however funding considerations persist. Scientific Computing Branch staff represents the Center on multiple workgroups and subcommittees to ensure the needs of NCTR are considered when assessing cloud access and network infrastructure improvements.

- **Comment (Biostatistics Branch - related to "support"):** *The Biostatistics Branch provides statistical support to NCTR/FDA scientists through 5.5 FTEs of effort. Despite the fact that this accounts for approximately 50% of the Biostatistics Branch's total budget, there is minimal support (approximately 1.5 FTE) available for NCTR internal research divisions.*
    - **Response:** Half of the Biostatistics Branch staff involves providing a supporting role to the National Toxicology Program. The other half is engaging research, either as PI or in collaboration with other divisions. The Branch has been successfully utilizing ORISE Post-Docs for both research and collaborated-research activities. With that said, due to a shortage of Ph.D. level statisticians, our statistical support activities in collaboration within the division branches and with other divisions are seriously limited. We have discussed with the NCTR senior leaders to improve this important function of the division.

- **Comment (Biostatistics Branch - related to "support"):** *A full 4 FTEs are devoted 100% to support data analysis for National Toxicology Program studies, estimated to number from 4-8 studies per year. There appears to be a missed opportunity to leverage the expertise of these individuals in a shared capacity across both support and research activities, to develop advanced statistical methods and push the engagement of biostatistics into the entire lifecycle of drug and food safety regulation.*
    - **Response:** Four FTE statisticians are funded through an IAG with the National Toxicology Program (NTP). The NTP IAG cites specific responsibilities for these statisticians which are primarily of a support nature. There are areas in which advancement of statistical methodology might fall within the purview of the NTP

IAG, such as analysis methods for mutigenerational studies or for vaginal cytology data. However, NTP-funded statisticians tend to be staffed sufficiently for statistical support of NTP-funded studies, but not sufficiently for additional research.

Overall Subcommittee Conclusions and Suggestions:

- **Comment:** *Overall, the Division staff collectively has a strong research record, which predates its current incarnation as a Division. The quality of science conducted by the Division is overall very strong to outstanding, and in many instances is contributing to significance advances in the respective discipline. … One notable exception is the work on microRNA, which is lagging behind the state of the science in the field.*
  - **Response:** We appreciate the positive overall comments from the Subcommittee. This young division consists of three branches that have been around for many years (way before the division was established in 2012). Although the distinct history behind each branch presents the challenge of developing a coherent strategy, it does provide a diverse staff with multidisciplinary expertise that is important to the future of the division. Thus, in the next five years, we will focus to building synergy within the Division and collaboration with other NCTR divisions. We aim to strengthen and focus our "research" efforts in the three themes presented during the review: Precision Medicine, Predictive Toxicology, and Biostatistical Approaches & Application. Aided by the specific comments and suggestions from the Subcommittee, we are confident of continually making significance advances in these research areas.
  - We do not understand the comment on microRNA. As we mentioned earlier in our response that the primary objective of the miRNA project is to understand whether expression profiles of miRNAs in the rat liver tissue can be used as mechanistic biomarkers of human DILI predictive value. We chose next generation sequencing for miRNA expression profiling for its potential in discovering novel miRNAs. In order to address the biology question, we first determine which bioinformatics pipeline would best meet our need. In the first phase of the project where we were at the time of Subcommittee review, we carried out a moderately comprehensive assessment of various data analysis pipelines. However, this is not the main objective of the project. As outlined in the presentation, this first phase will be followed with Phase II to discover miRNA biomarkers for liver carcinogenicity, Phase III to discover miRNA biomarkers for drug induced liver injury, and Phase IV to validate these biomarkers. We have made tremendous progress since then. Two manuscripts have been prepared. The first reports some miRNA biomarkers for liver carcinogenicity and is currently under review. The second reports some biomarkers for drug induced liver injury and will be submitted soon.

- **Comment:** *The Division has also contributed significantly to the mission of the Agency by creating resources such as ArrayTrack, the Endocrine Disruptor Knowledge Base (EDKB), and the Liver Toxicity Knowledge Base (LTKB) that have been very useful across multiple NCTR Divisions and FDA Product Centers. Similarly, the Division has done a good job of supporting the bioinformatics needs of other NCTR Divisions and FDA Product Centers, representatives of the latter who participated in the site visit provided strong and enthusiastic support for the Division's service and support activities, suggesting a justification for more resources to expand the Division's support function. Indeed, given*

that biology in general, and toxicology itself, are increasingly data driven, and because the FDA is increasingly seeing submissions that include large-scale genomic and other data sets, the expertise represented in the Division is essential to the future success of the FDA. While the work of the Division is overall commendable, its dual role might be the source of some of its weaknesses. It was not at all clear in some instances where the boundary was drawn between research and support, or which people were contributing to which aims of the Division. There also appears to be a lack of clarity as to the overall mission of the Division, with some evidence of "mission creep" with research projects investigating areas that are not clearly linked to the NCTR or the FDA missions.

- **Response:** We appreciate the Subcommittee's recognizing the importance of the Division's "service" and "support" activities and recommendation of more resources to expand such support function. To better manage the balance between research and support is very critical to the healthy functioning and growth of the Division. To better address this need, I would like to clarify three terminologies that describe the division activities. These are "Conventional Service", "Data Analysis Support", and "Research". (1) **Conventional Service**: it includes activities conducted by Scientific Computing Branch and statisticians in Biostatistics Branch to support National Toxicology Program, which accounts for 22 FTEs (18 from SCB and 4 from Biostatistics Branch). These activities have been established a long before the division was established. (2) **Data Analysis Support**: As more and more research at NCTR involves data science, effective application of the data analysis methods requires a broad-scale of bioinformatics/biostatistical capabilities and effective communication with the biologists for data interpretation. These new requirements demand not only the conventional IT support (e.g., setting up the HPC, storage and backup) but, most importantly, expertise beyond IT (such as domain specific knowledge, developing customized algorithms, setting up specialized programming environment). They are also different from the conventional services described above which are mainly handled by both SCB and Biostatistical Branches; these two branches have been around for many years with established support mechanisms and staffs to traditional requirements/needs in their respective fields at NCTR (e.g., animal care, NTP support and etc.). However, the fast growing demand of support for the new data stream is beyond the current expertise resided in these two branches. This "Data Analysis Support" is to address this gap in this fast growing data science field. (3) **Research:** It includes research activities towards working with approved research protocols and externally funded research projects. During the Subcommittee review, both "Conventional Service" and "Data Analysis Support" were described as "Support", which may have caused some confusion in respect to the boundary between "Support" and "Research". Clearly, many new technologies adopted at NCTR generate massive data that demand enormous bioinformatics/biostatistics support. This paradigm shift also blurs the boundary between "research" and "support"; the former frequently applies bioinformatics tools and methodologies. The "blurring boundary" is also due to the fact that supports in these areas are challenging since they require

research and development with scientific (domain specific) knowledge through collaboration. For that reason, we divide "support" to two components, "Conventional Service" and "Data Analysis Support". Specifically, we have developed a proposal to enhance "Data Analysis Support" in the division. We have discussed this proposal extensively with the NCTR management. One goal of the reorganization is to maintain this culture despite ever increasing demands that are the consequence of IT security, and new data streams.

Recommendations:

- **Comment:** (1) *The balance between primary research and service to support the Agency mission was a recurring issue across the Branches and Programs within the Division and the Subcommittee came away from the site visit with concerns about "mission creep". (2) The Subcommittee strongly recommends that the Division conduct an internal review to evaluate each research program and clarify its alignment with the Agency mission. (3) Does each program fit within a broader, more coherent research mission for the Division? (4) Do the research programs within the Division address not only bioinformatics and biostatistics support within the FDA, but also research necessary to support the Agency's regulatory mission, including research relevant to analyzing the new data types that the Agency is seeing or anticipates seeing in submissions? (4) As part of an internal review, the Division might also reconsider its current mission statement to determine whether it is reflective of ongoing Division activities.*
    - **Response:** We appreciate the Subcommittee's concern of "mission creep" and suggestion for an internal assessment to ensure the divisions' projects are in line with the overarching mission in FDA in general and at NCTR in specific. While every on-going and future project at the division is approved by the NCTR for their relevance to the FDA mission, the primary research does compete with service due to the division's dual responsibilities. With that said, I don't feel both are necessary competing for resources and investments, and it can turn to beneficence as long as we plan carefully and strategically. To better integrate them and balance between them, we are considering the establishment of a fourth branch to better coordinate "support" and "research" activities in the division. As for the research programs, over the years the NCTR has developed a thorough and rigorous vetting process to ensure each FDA funded project aligns well with the Agency mission. This process includes internal division review and endorsements from other FDA centers. Usually we need to have co-investigators from other FDA centers to be part of the proposals. The projects we presented in this review have all gone through this process. A standard and important part of each research proposal is its relevance and projected impacts to FDA regulatory mission. With that said, we do understand the "mission creep" concern. That is why we group our research programs into three themes (i.e., Precision Medicine, Predictive Toxicology for Drug Safety, and Biostatistical Approaches & Application) to concentrate our research activities that fits better with both our skill and the agency's mission.

- **Comment:** *Resources such as ArrayTrack, the Endocrine Disruptor Knowledge Base (EDKB), and the Liver Toxicity Knowledge base have been very useful within the FDA, but there are other projects such as the analysis of breast cancer to search for biomarkers that, while good work, might have been better if applied to a problem more directly relevant to either the NCTR's or the FDA's mission.*
    - **Response:** We appreciate the very affirming comments on ArrayTrack, EDKB, and LTKB. We have been striving to develop resources that have broad application in the FDA and beyond. Work along this direction will continue to be a focus for the Division. As for the specific project on breast cancer, we developed a new bioinformatics methodology and tested it with some breast cancer datasets to see how it performed in search of biomarkers for diagnosis and prognosis. We chose breast cancer since it is a primary concern to the Office of Women's Health (OWH). Through these analyses, we intend to develop a research proposal and submit it to OWH for grant application.

- **Comment:** *It was the Subcommittee's impression that there was a disconnect between projects, between Branches within the Division, and between the NCTR and other Product Centers with regards to scientific and research overlap with the result that potential synergies are not being exploited. A notable example is the apparent lack of substantive interaction between the Division of Bioinformatics and Biostatistics and the Division of Systems Biology. Another possible "missed opportunity" is the application of the expertise within the Division to the analysis of imaging data available within the NCTR, which is anticipated to become more common in submissions in the Agency. It is also not clear where the modeling expertise resides within NCTR, and whether that expertise is integrated into the Division's mission and activities.*
    - **Response:** We do apologize for not conveying this message clearly during the review. Our division is very collaborative. The collaboration comes naturally for this division because we seldom produce our own data. A good example is that every project under R2R involves multiple divisions and centers. For example, FDALabel (a R2R project) involves personnel from the Division of Systems Biology, NCTR's Office of Scientific Coordination (Project Leader) and CDER's Office of New Drug and Office of Computational Science. Since the collaboration is a natural thing for this division, our presentation and preparation of the written materials are more focused on division initiated research projects. With this emphasis, it does not mean that our connection and collaboration with other NCTR Divisions and Product Centers are not extensive. We apologize for not making that clear. While we are preparing this response, NCTR formed a bioimaging data analysis group and our division has a strong presence in this group.

- **Comment:** *The R2R Program has great potential to improve and enhance the Agency mission. However, this program is not likely to produce significant numbers of publications, therefore, the Division is urged to work with the NCTR Director and the other Product Centers to define and collect metrics for assessing the impact of the R2R program on the Agency mission. It is also recommended that the R2R Program be integrated into all Division activities.*
    - **Response:** Thank you for recognizing the great potential of the R2R program. We agree with the Subcommittee about the importance to define and collect metrics for assessing the impact and success of the program.

- **Comment:** *Professional reward systems are needed within the Division that extend beyond the publication of high quality papers. Other metrics of professional success should be implemented, not the least of which is impact to FDA operations and facilitation of stakeholder engagement in the regulatory process. It is recommended to work with the Centers to define and capture metrics for impact, including outreach to the global regulatory community.*
    - **Response:** We agree with the Subcommittee. Properly designed and implemented reward and career advancement systems for the support professionals will be extremely important to retain high caliber support staff. We will work with the Center Management to define and implement non-publication centered metrics for professional success.

- **Comment: (1)** *The Subcommittee shares concerns raised by at least one presentation during the site visit regarding leadership training and a succession plan for the Division. (2) It is not clear that there is a succession plan in place. Who are the future leaders in the Division? (3) Do the current Division leaders have a plan for growing internal talent or recruiting new talent? (4) What is the vision of the Division regarding diversity within the Division leadership and staff? The Subcommittee strongly recommends that the Division leadership work with NCTR leadership to develop concrete plans to address these issues.*
    - **Response:** The Division is relatively new. We appreciate the Subcommittee for raising these important questions to guide our future growth. We will incorporate the comments into our strategical plan for the next five years. As a move along that direction, we are proposing a fourth branch to focus on R2R projects and support new data stream, e.g., the bioimaging data. This plan will also strength our succession plan.

Support Materials:

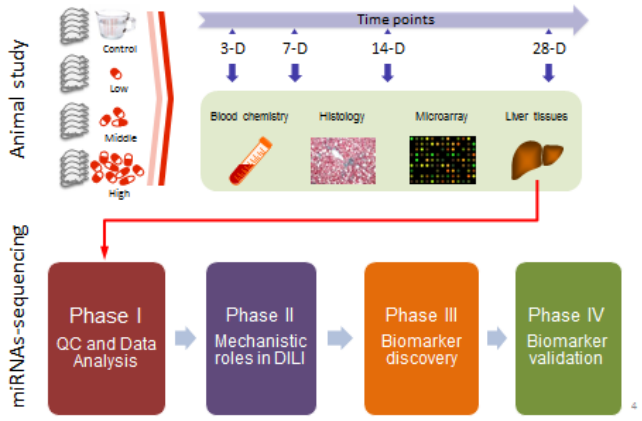
1. Zhichao Liu, Hong Fang, William Slikker, Weida Tong, Potential Reuse of Oncologic Drugs for the Treatment of Rare Diseases, Trends in Pharmacological Sciences (IF=10.15), (37)10, 843–857, 2016

2. Rat toxicogenomic study reveals analytical consistency across microarray platforms, Nat Biotech, 2006 (Cited 379 times)

3. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models, Nat Biotech, 2010 (Cited 439times)

4. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance, Nat Biotech, 2014 (Cited 111times)

5. Study design and milestone of the microRNA project in the division



6. Harsh Dweep, Yuji Morikawa, Binsheng Gong, Jian Yan, Zhichao Liu, Tao Chen, Halil Bisgin, Wen Zou, Huixiao Hong, Tieliu Shi, Ping Gong, Christina Castro, Takeki Uehara, Yuping Wang, Weida Tong, **Mechanistic Roles of microRNAs in Hepatocarcinogenesis: A study of Thioacetamide with multiple doses and time-points of rats**, Arc Tox, 2016 (submitted)

7. Abstract:

   ***In Vitro* to *In vivo* Extrapolation (IVIVE): A Genome-Wide Analysis of Drug-Induced Liver Injury Using a Drug-Pair Ranking (DPRank) Method,**
   Zhichao Liu, Hong Fang, Jürgen Borlak, Ruth Roberts, Weida Tong

   There has been considerable concern over the fact that preclinical animal toxicity studies may not accurately predict human toxicity. In light of this, *in vitro* systems have been developed that have the potential to supplement or even replace animal uses. We examined *in vitro* to *in vivo* extrapolation (IVIVE) of gene expression data retrieved form Open TG-GATEs for 131 compounds given to rats for 28 days, and to human or rat hepatocytes for 24 hours. Notably, a Drug-Pair Ranking (DPRank) method was developed

among the three testing systems and the IVIVE potential was quantified by the DPRank score based on the preservation of the order of similarity rankings of drug-drug pairs between the platforms using a receiver operating characteristic (ROC) curve analysis to measure Area Under the Curve (AUC). A high IVIVE potential was noted for rat primary hepatocytes (DPRank score = 0.71) whereas the IVIVE potential for human primary hepatocytes was lower (DPRank score =0.58), indicating species difference playing a critical role in IVIVE. When limiting the analysis to only those drugs causing drug-induced liver injury, the IVIVE potential was slightly improved for both rats (from 0.71 to 0.76) and humans (from 0.58 to 0.62). Similarly, DPRank scores were improved when the analyses focused on specific hepatotoxic endpoints such as hepatocellular injury, cholestatic injury or 'mixed'. In conclusion, the toxicogenomics data generated *in vitro* enables the ranking of drugs for their potential to cause toxicity prior to *in vivo* investigations and the proposed DPRank methodology could be applied to other cell-based assay systems where insight on IVIVE is required.