

Division of Bioinformatics and Biostatistics (DBB)

Joshua Xu, Ph.D.

On behalf of Weida Tong, Ph.D.

Views expressed in this presentation are those of the presenter and not necessarily those of the U.S. Food and Drug Administration

Division Staff

- Three branches — Full Time Employees (FTEs)
 - Bioinformatics Branch: Research-centric (17 FTEs)
 - Biostatistics Branch: Research + Service (9 FTEs)
 - Scientific Computing branch: Service-centric (17 FTEs)
- Immediate office: 2 administrators + one senior advisor
- ~10 Postdoctoral fellows + 1 graduate student
- Division activities:
 - 40% in Research and 60% in Service



Division Mission (Vision)

- **Research** - To conduct bioinformatics and biostatistics research to support FDA's mission of improving the safety and efficacy of FDA-regulated products
- **Service** - To provide research and regulatory support to NCTR and FDA scientists in bioinformatics, biostatistics, and scientific computing
- **Focused on FDA Relevance:**
 - Diligently seek to have direct impact on the review process
 - Seek and strengthen linkages with product centers, and evolve our capabilities to meet current and future FDA needs

Division Mission (Strategies)

- Developed the R2R framework (i.e., Research-to-Review and Return)
 - Research-to-Review via “knowledge uptake”
 - Review-to-Research via “data liberation”
 - Partnership between OCS/CDER and NCTR
- Example projects:
 - Collaborating with CDER/OTS on the DASH system (Data Analysis Host System)
 - Tracks progression from INDs to NDAs or BLAs and the approval of NDAs and BLAs
 - Began by upgrading the technology
 - Progressed to providing means for text mining and analytics of documents
 - FDALabel: a modern web-based database for FDA-approved drug labels
 - Began as a research need and has progressed to its being integrated in the review process
 - Collaborating with ORA to develop an intelligent recognition system for food-contaminating bugs



Five Themes Reviewed by the Subcommittee

- Theme 1: Precision Medicine
 - Focused on assessing enabling technologies for precision medicine
- Theme 2: Predictive Toxicology
 - Focused on drug safety
- Theme 3: Biostatistical Approaches and Applications
 - Big data methodologies
- Theme 4: R2R Framework & Activities
 - Regulatory application
- Theme 5: Service & Support Functions



Structure of Subcommittee Report

- Overview of the Subcommittee review process and the Division
- Review comments for each theme
- Overall Subcommittee Conclusions and Suggestions
 - i.e., overall comments

Structure of Response Presentation

- Appreciate thoughtful comments
- Respond to overall comments
- Respond to theme-specific comments

Overall Comments: Bioinformatics Resources and Service (page# 20)

- 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.
- Notes: Page numbers are referring to our response document. Comments are shown without quotation marks.

Overall Comments: Boundary Between Research and Service (page #21)

- *"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.*
- *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."*

Response (page# 21-22)

- Three different functions:
 - “Conventional Service”
 - Very much legacy support
 - Mainly in Scientific Computing Branch and Biostatistics Branch
 - Well ingrained and working mechanism
 - “Data Analysis Support”
 - Involve research and method development
 - During the Subcommittee review, both were described as “Support”, which may have confused the boundary between “Support” and “Research.”

Response (cont.)

- Many new technologies generate large and complex data
 - Demand enormous bioinformatics/biostatistics support
 - Blurs the boundary between “research” and “support”

- Analytics can be difficult
 - Often falls into methods R&D
 - Often requires close collaboration for domain-specific knowledge

- Proposal put forth to enhance Division's “Data Analysis Support”
 - Discussed extensively with NCTR leadership
 - A reorganization maintaining the interdisciplinary teamwork culture (despite ever-increasing demands stemming from IT security, and new data streams).

Other Overall Comments

(1, page #22)



- *The Subcommittee strongly recommends that the Division conduct an **internal review** to evaluate each research program and clarify its alignment with the Agency mission. Does each program fit within a broader, more coherent research mission for the Division? 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?*

Response to Other Overall Comments (1, page #22)



- The NCTR has developed a thorough and rigorous vetting process to ensure NCTR projects align well with the Agency mission.
 - Internal division review and endorsements from other FDA centers
 - Co-investigators from other FDA centers to be part of the proposals
 - An integral part of each research proposal to state relevance and anticipated impacts to FDA regulatory mission.
 - The above is why we grouped our research programs into three themes (i.e., Precision Medicine, Predictive Toxicology for Drug Safety, and Biostatistical Approaches & Application) to illustrate how research activities align between our skills and the agency's mission.

Other Overall Comments

(2, page #23)

- 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.*

Response to Other Overall Comments (2, page #23)

- We do apologize for not conveying this message clearly during the review.
 - Our presentation and preparation of the written materials are more focused on division-initiated research projects.
- Our division is very collaborative.
 - Collaboration comes naturally as bioinformatics is inherently multidisciplinary and we seldom produce our own data.
 - FDALabel, SEQC2, R2R, CTP projects are all great examples
- NCTR has formed a bio-imaging data analysis group where our Division has a prominent presence.

Response to Other Overall Comments (2, cont.)



- Cross-center collaboration
 - Identifying risk factors for safety assessment of drug-induced QT prolongation in cardiotoxicity (CDER collaboration)
 - R2R projects
 - CTP projects
- Regional activities:
 - MidSouth Computational Biology and Bioinformatics Society (MCBIOS)
 - One staff member is the President for 2016 and another serves as a board member.
 - Arkansas Bioinformatics Consortium (AR-BIC):
 - Established in 2014, comprising the major Arkansas universities plus NCTR.
 - The division formulated the concept and facilitated its establishment and development.

Response to Other Overall Comments (3, page #24)

- 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 value and potential of the R2R program. We agree that defining metrics and collecting data will foster success and impact.

Response to Other Overall Comments (4, page #24)

- **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 emphasis. Properly designed and implemented reward and career advancement protocols for support professionals will increase retention likelihood of high caliber staff. We will work with the Center leadership toward this goal.

Responses to Theme-Specific Comments



Theme 1: Precision Medicine (1)

- **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:** Page #1-2
 - SEQC2: assess the reliable use of genomics technologies in regulatory decision-making
 - Toxicological significance has been an essential element in all the MAQC projects.
 - Rare diseases: repurpose marketed drugs for the treatment of rare diseases.
 - A new concept: using oncology drugs to treat rare disease
 - Developing methodologies to understand how the dosage and route should be modified



Theme 1: Precision Medicine (2)

- **Comment (SEQC2):** *The Subcommittee recommends **sharpening of focus and closer alignment** to the Agency’s mission to facilitate continued maturation of the Division. Various FDA centers have encountered NGS data in regulatory applications, closer alignment between FDA needs and the Division activities is highly desirable.*
- **Response:** Page #2-3
 - Working closely with many reviewers and scientists from all FDA centers
 - Three specific aims:
 1. Develop quality metrics for reproducible NGS results
 2. Benchmark bioinformatics methods towards the development of standard data analysis protocols
 3. Assess the joint effects of key parameters affecting NGS results for enhancing understanding of fit-for-purpose applications with NGS.
 - Initial emphasis on evaluating the challenges outlined by the FDA’s discussion papers
 - *Developing Analytical Standards for NGS Testing*
 - *Optimizing FDA’s Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper*
 - Such alignment has been maintained throughout the MAQC project series

Theme 1: Precision Medicine (3)

- **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: Page #3**
 - Understanding the sources of variations and their impact has been an emphasis throughout MAQC projects.
 - Evaluate these sources of variations such as sample preprocessing, gene capture panels, library preparation, and sequencing instruments.
 - Studying these parameters with various biological samples.
 - To provide insights into what practices may enhance or deleteriously affect validity and reliability.
 - Focus on quality metrics and standard practices

Theme 2: Predictive Toxicology (1)



- **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:** Page #5
 - Co-chair the Liver Toxicity Interest Group at the FDA
 - Used to evaluate liver toxicity in 16 submissions using RO2.
 - The submissions are across IND, Phase I, Phase II, Phase III, and NDA submissions.
 - 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.
 - Bioinformatics tool to assist the reviewers in applying the RO2 rule and LTKB in their regulatory review.
 - DILIScore incorporated formation of reactive metabolites to predict the severity of DILI risk

Theme 2: Predictive Toxicology (2)

- **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:** Page #4-5
 - The primary objective: 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.
 - NGS profiling miRNA expression to discover novel miRNAs
 - Phase I: determine which bioinformatics pipeline would best meet our needs.
 - Phase II: discover miRNA biomarkers for liver carcinogenicity
 - Phase III: discover miRNA biomarkers for drug induced liver injury
 - Phase IV: confirm these biomarkers.
 - Progress updates: two manuscripts
 - Not focused on **circulating** miRNAs

Theme 2: Predictive Toxicology (3)

- **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: Page #6**
 - Published a diverse set of models
 - three-dimension quantitative using Comparative Molecular Field Analysis (CoMFA)
 - chemometric classification models
 - ensemble models
 - docking models
 - Participating in a large predictive toxicology collaboration organized by the EPA

Theme 2: Predictive Toxicology (4)



- **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: Page #7**
 - LTKB: initiate the interest group to communicate our results
 - Use R2R to facilitate the translation of LTKB for use in review process
 - Tobacco Constituents Knowledge Base (TCKB) another example with chemical structure and biological effects data for > 9600 chemicals

Theme 3: Biostatistical Approaches and Applications (1)

- **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:** Page #9
 - We appreciate the generally positive comments that our research areas are important and critical to the mission of the FDA.

Theme 3: Biostatistical Approaches and Applications (2)

- **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 learner, 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.*
- **Response:** Page #11-12
 - Transitioning the research priority to big data analysis is well underway
 - Three proposals submitted to the FDA Office of Chief Scientist for internal funding consideration
 - EHR (Electronic Health Record) data will be combined with the FDA AE database
 - Collaboration with CDER, CVM, and CDRH colleagues in an effort to use big data to promote regulatory science

Theme 3: Biostatistical Approaches and Applications (3)

- **Comment:** *Strategic planning* within the Branch to identify research directions and priorities is strongly recommended.
- **Response:** Page #10
 - The field of biostatistics proceeds in tandem with advances in biotechnology
 - Traditional statistical test procedures for risk-factor identification.
 - Today, statistical models and test procedures are developed to identify biomarkers for high-dimensional molecular experiments.
 - Focus #1: statistical and machine learning for high-dimensional big data
 - Focus #2: statistical methods for precision medicine.
 - New projects are subjected to Division-level and Center-level reviews.
 - Seek guidance from FDA product center colleagues in order to ensure that projects meet the needs of the Agency.

Theme 4: R2R Framework & Activities (1)

- 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:** Page #14

 - We appreciate the detailed, insightful, and positive comments of the SAB subcommittee towards the R2R program.

Theme 4: R2R Framework & Activities (2)

- **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:** Page #15
 - Importance of the R2R program has been recognized by the FDA
 - R2R program won Commissioner’s Special Citation for its innovative cross-Center bioinformatics projects benefitting regulatory business processes
 - FDALabel won a Scientific Achievement Award from the Office of Chief Scientist for outstanding inter-center scientific collaboration
 - Developed some metrics for tracking the impact
 - Logging of users and usage
 - Collection of use cases and feedback

Theme 5: Service & Support Functions (1)

- **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:** Page #18
 - Pursuing increased bandwidth and access to cloud platforms such as AWS and Salesforce
 - Addressed many security concerns
 - Persistent funding constraints
 - Represents the Center on multiple workgroups and subcommittees to ensure the needs of NCTR are considered when assessing cloud access and network infrastructure improvements.

Theme 5: Service & Support Functions (2)

- **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 underutilized.*
- **Response: Page #17**
 - HHS and FDA have adopted a government or commercial off-the-shelf (GOTS or COTS) first approach to eliminate redundant, outdated and underutilized software.
 - NCTR will follow this lead
 - Performed an inventory of the applications and databases hosted on the NCTR servers.
 - Evaluating solutions available elsewhere in FDA, open source products and COTS alternatives
 - Buy versus make has and will prove to be challenging
 - Many legacy applications are linked via shared data structures/databases
 - Simultaneous changes are required to ensure interoperability and access to historic data.



Thank You!