NCTR Division of Bioinformatics and Biostatistics (DBB) Review
Subcommittee Site Visit Review, November 5-6, 2015
Submitted May 10, 2016

Overview

Review Subcommittee and Expertise

The names and affiliation of the site visit reviewers are provided in Appendix 1. Two members (Waters and Lein) of the NCTR Scientific Science Advisory Board (SAM) participated in the review. Waters and Lein served as subcommittee chair and co-chair, respectively. Afshari, who is a member of the FDA Science Advisory Board, participated as a temporary member consultant on the subcommittee. All other reviewers participated as subject matter experts in content areas of interest to the Division of Bioinformatics and Biostatistics (DBB, referred to hereafter as "the Division"). The subcommittee members received a written overview of their charge in a memorandum dated July 13, 2015 from Daniel Acosta, Jr., Ph.D., NCTR Deputy Director for Research and Donna L. Mendrick, Ph.D., Associate Director for Regulatory Activities. The charge memo is attached in Appendix 2. Site visit reviewers were provided with project overviews divided into thematic areas on October 20, 2015 and a slightly revised version of these materials on October 22, 2015.

Primary and secondary reviewers were assigned to each topical theme area in advance of the meeting. Reviewer assignments are shown in the table below. The reviews of each theme area have been written primarily by the assigned experts.

Theme	Primary Reviewer	Secondary Reviewer
Precision Medicine	Kenneth Ramos	Ying Lu
Predictive Toxicology	John Quackenbush	Cynthia Afshari Pamela Lein
Biostatistical Approach & Applications	Ying Lu	John Quackenbush
R2R Framework & Activities	Cynthia Afshari	Kenneth Ramos
Service & Support Functions	Katrina Waters	Pamela Lein

Agenda, Reviewed Materials and Process

The agenda for the two-day site visit is shown in Appendix 3. Electronic files of materials were provided to reviewers prior to the site visit and included: (1) Executive summary of the Division written by the Division Director, Dr. Weida Tong; (2) Written summaries of the major research programs of the Division; (3) current CVs of the Principle Investigators in the Division; and (4) The strategic plans of the FDA and the NCTR.

Subcommittee members were instructed to review these materials prior to the site visit at the NCTR campus. When the reviewers arrived for the site visit, a hardcopy of the materials listed above was also provided in a tabbed notebook, along with printouts of PowerPoint slide

presentations, a written overview of the Division, copies of posters and a list of Division publications from 2014 up until the November 2015 site visit.

Division of Bioinformatics and Biostatistics Overview

The Division is a relatively new division of the NCTR, having been established in May of 2012. The Division's mission is to apply bioinformatics and biostatistical methods to conduct research and support diverse needs within NCTR and the FDA. Thus, the Division plays a dual research and support role within the FDA, with approximately 40% of its efforts focused on research and the remaining 60% directed to support services in Information Technology (IT), software development/support and bioinformatics and biostatistics support for NCTR experiments and the National Toxicology Program. The Division has significant interactions with other Divisions within the NCTR, with other Product Centers within the FDA and with the regional scientific community (e.g., the MidSouth Computational Biology and Bioinformatics Society and the Arkansas Bioinformatics Consortium).

Administratively, the Division, which is directed by Dr. Weida Tong, is divided into three Branches: (1) the Bioinformatics Branch, led by Dr. Tong: (2) the Biostatistics Branch led by Dr. Chen; and (3) the Scientific Computing Branch led by Mr. Edward "Ted" Bearden. At the time of the site visit, the Division included about 50 staff members, including visiting scientists, 10 postdoctoral fellows and 4 graduate students.

The Division's research and service programs are designed to align with FDA Product Center needs, both current and prospective. The Division's activities are grouped into 5 themes: (1) Precision Medicine – assessing emerging technologies and their application for rare diseases and clinical application; (2) Predictive Toxicology – improving drug safety with predictive modeling and biomarker development: (3) Biostatistics and Bioinformatics Methodologies – developing and apply new statistical methods; (4) R2R or Review-to-Research and Return Program – enhancing the interaction between review and research; and (5) Service and Support – primarily supporting NCTR regulatory science research.

The charge to the Subcommittee was to:

- Review the quality of the science conducted by the Division. Is it state-of-the-art?
- Evaluate the impact of the Division's activities on the mission of the FDA. Are the Division activities critical to the Agency's mission? Are they forward-looking and anticipatory of Agency needs?
- Provide feedback to the Division on how it should position itself for the future.

Theme 1: Precision Medicine

The stated objective of the Division of Bioinformatics and Biostatistics at the National Center for Toxicological Research (NCTR) of the Food and Drug Administration (FDA) is to assess emerging precision medicine technologies (primarily in the genomics space) and their application in pharmacogenomics biomarker development and the treatment of rare diseases.

In a relatively short period of time, the Division has played a major role in advancing the adoption of emerging technologies within the FDA review process, increased understanding of these technologies within the Agency's workforce, and advanced the future application of genomic technologies. Their efforts have clearly been paradigm shifting for the Agency, as evidenced by a large number of high quality publications, successful collaborations with prominent scientists around the world, and agency-wide adoption of several key recommendations made by the Division.

While overall efforts within the Division are laudable, and significant progress has been made in several areas, the primary objective of the Division was not tightly woven into present and future activities. As such, deliberate efforts should be made to more closely align the work proposed by scientists in the Division with the stated mission of the NCTR and the Division's objectives. Specifically, 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. On a related note, it was stated during the site visit and in writing that a focus on rare diseases is warranted given that one third of the new drugs approved by the FDA in the past ten years are for rare diseases. However, no information was provided about how the proposed efforts help to advance the mission of the NCTR. In general, the presentations on site were well-conceived and delivered with clarity, enthusiasm and understanding of the subject matter. However, most of these lacked a clear definition of anticipated impacts and relevance to the Agency mission. Specific recommendations were made by Subcommittee members for improvement of the scientific quality of individual projects.

Dr. Weida Tong is a talented and highly energetic scientist and his team clearly views him as an effective leader. The quality of the work produced by the Division is evidence of his strong leadership for a relatively young division. He has established a culture of high productivity and his efforts have enabled significant advances for the Division and the Agency. Under his leadership, the team proposes to continue to build on efforts to support the development of FDA companion guidance for industry on pharmacogenomics data submissions, further advance SEQ2 with the development of quality control metrics, assessment of reproducibility and continuity, and benchmarking of bioinformatics approaches for clinical use of WGS and targeted sequencing data from NGS. While these plans are well within the scope of expertise of the Division, and likely to generate valuable data and resources for the Agency, as noted above, the Subcommittee recommends sharpening of focus and closer alignment to the Agency's mission to facilitate continued maturation of the Division. It should be noted that various FDA centers have encountered NGS data in regulatory applications, including FDA oversight of NGS-based assays for diagnosis and prognosis, applying NGS in pathogen detection and outbreak detection, reviewing NGS data for drug efficacy and safety for both clinical and preclinical assessments and NGS as an improved tool for studying immunogenicity of vaccines. Therefore, closer alignment between FDA needs and the Division activities is highly desirable. Furthermore, 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.

The MAQC model enables community consensus on the fit-for-purpose use of NGS, facilitating transition of NGS data toward science-based regulation in the FDA. However, 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. In view of the rapid evolution of technology, 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.

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. Stakeholders involved 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.

Theme 2: Predictive Toxicology

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 group has been developing and applying a broad range of methods, including toxicogenomics, systems toxicology/pharmacology, computational toxicology methods (such as quantitative structure-activity relationships [QSARs]), database and knowledge base development, data standards and ontology, FDA resources such as the FDA's Adverse Events Reporting System [FAERS], drug labeling, and data from the large government initiated programs such as Tox21 and ToxCast.

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 such as the Liver Toxicity Knowledge base (LTKB) that have been very important in advancing understanding in the field. Research in predictive toxicology has been a major focus of the Division and the individual research projects that were described in the annual report and presented were, in general, good examples of what the Division could do as many address questions relevant to the FDA's regulatory and health assurance mission.

The review panel was presented with four research projects in the Division's report, two of which were presented in detail during the review panel meeting. The first was a project designed to identify predictors of drug-induced liver injury (DILI). The approach involved the use of the resources assembled into the LTKB and an extensive analysis of the factors involved using a variety of methods applied to a large number of different drugs and compounds. The most robust and predictive method that emerged from this analysis was a relatively simple "rule of two" (RO2) for assessing the liver toxicity of orally administered drugs: The RO2 was based on analysis of 164 FDA-approved oral medications and showed association of high daily doses (≥ 100 mg/day) and lipophilicity (partition coefficient, logP ≥3) with significant risk for DILI. The RO2 was further verified in an independent test set of 179 oral medications, drug pairs with similar chemical structures and molecular targets but different DILI potential, and in clinical case studies with complex co-medication regimes. While an important advance, this work 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. Further 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.

The second project presented to the Subcommittee during the site visit was a study of microRNAs as biomarkers for hepatotoxicity. The project involved the analysis of miRNA-seq data in individuals with miRNA biomarkers for Non-alcoholic Steatohepatitis (NASH) and required the development of analytical protocols for miRNA-seq data and its analysis to identify differentially expressed miRNAs that might be predictive of liver damage. They investigated 384 representative data analysis pipelines that consist of four tools (mirDeep2, mirExpress, miRNAkey, sRNAbench), 24 profiling choices, and four normalization options on miRNA-seq data generated from rat liver samples in four time points and three dose levels. 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. However, the 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.
Two other projects were presented in the written materials provided to the Subcommittee. The first was the Endocrine Disruptor Knowledge Base (EDKB), which like the Liver Toxicity Knowledge Base (LTKI), represents an important contribution to the overall FDA mission. The information gathered in the EDKB 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. The second project described in the written materials was the Division's overall involvement in toxicogenomics. 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). They used "topic modeling" to analyze the collected data and set of candidate genes for toxic response. However, they found that it was difficult to extrapolate the *in vitro* results to the *in vivo* response, but that careful experimental design might help address this problem.

Overall, the Predictive Toxicology program has made important contributions, the most significant of which are the knowledge bases that they have assembled. 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. The work on biomarkers is also interesting, but the sample size and experimental design are too limited to draw firm conclusions. However, 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. This 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.

Theme 3: Biostatistical Approaches and Applications

Division accomplishments

The Biostatistics Branch in the Division consists of 9 FTE members. The Branch Director, Dr. JJ Chen, is a seasoned statistician and has been with NCTR for 32 years. Dr. Chen is an internationally recognized leader in statistical methods for toxicology, biomarker identification, and statistical modeling for quantitative risk assessment. In addition to Dr. Chen, the branch has the following principal investigators: Drs. CW Chang, YC Chen, NI George, and W Zou.

While the division reported that they spent 50% efforts to conduct peer-review research in development of statistical methods to support FDA's missions with the remaining 50% effort focused on support of other studies conducted at NCTR, the actual effort for support seems to be higher. According to Slide 4 of the Biostatistics Branch Presentation, 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 these 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 that significant contributions to statistical support of collaborative and team research efforts be considered an important metric for promotion consideration.

The Biostatistics Branch has successfully competed for 6 FDA investigator-initiated grants. Three research projects were included in the report and site visit presentation: (1) The predictive biomarker identification and subgroup selection for treatment optimization; (2) Safety signal detection and analysis in the FDA's adverse event reporting systems; and (3) Statistical analysis and challenges in next-generation sequencing. All three projects were intimately tied to FDA's mission. A total of 39 peer-reviewed publications from three funded studies were listed. Considering the size of the Branch that supports NCTR studies, we congratulate the Branch for their impressive research accomplishments.

Quality of the Science

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. These are important research areas and critical to the mission of the FDA. The group has expertise and significant achievements in all these areas.

The three research presentations provided during the site visit demonstrated state-of-the-art research accomplishments. For example, Dr. Chen used a model of three factors for biomarker identification, which includes a latent subgroup variable. This is different from conventional approaches in which subgroup variable is the result of splitting of biomarkers, rather than being considered a factor by itself. His approach allows a proper evaluation of the utility of biomarkers in predicting the latent subgroup, if they indeed exist. Other work in safety signal detection, topic modeling, and NGS are all important with great applications. Many of the projects represent collaborations within NCTR and within FDA.

Throughout the site visit presentations and discussions, it was not clear how the Biostatistics Branch sets research priorities for Branch scientists and statisticians. It seems many of these are initiated in an *ad hoc* manner by investigators based on their own research interests and/or historical collaborations. For example, 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. Also the link between next generation sequencing (NGS) and the 4 research areas (as identified above) was not obvious. Strategic planning within the Branch to identify research directions and priorities is strongly recommended.

Positioning for the Future

Dr. Chen asked for Subcommittee feedback on the following points.

(1) Promotion of Research Biostatisticians based on methodological publications. The challenge of appropriately evaluating statisticians' contributions in team science is not unique to NCTR. The Branch has identified biostatisticians as Co-PIs on collaborative projects, which is a good practice. 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. On the other hand, it is important to have statistical contributions to science be counted as a key performance measure for promotion. Sometimes an innovation may not be sufficient for a methodological publication but is so critical to the science such that without this innovation the project might not succeed. Other times, the contribution to science by statisticians is more significant than the methodology itself. In academia, such contribution can be accounted via their co-first author or co-senior author publications (based on significant contributions) or a letter of support by collaborators.

(2) Whether and how the Biostatistics Branch should be involved in Big Data. The committee strongly encourages the Branch to actively engage in big data research. 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.

(3) Subcommittee suggestions for additional research opportunities that the Biostatistics Branch might consider.

(3.1) 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.

(3.2) 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.

Succession plan, recruitment and retention

During the visit, Dr. Chen 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.

Theme 4: R2R Framework & Activities

The Research-to-Review and Return (R2R) program represents an important collaboration between NCTR and the FDA Centers, especially the Center for Drugs (CDER). 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. It will provide a structure to sponsor data that is easily analyzed and compared to other existing data. In the 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.

This collaborative project is delivering important tools for the FDA reviewers. This includes templates for the IND and NDAs that will facilitate uniformity of these documents to ease future data mining efforts across products. The impact of being able to compare data across products, as opposed to handling each project one by one, is significant. This will enable consistency across reviewers and ensure sponsors are held to similar standards. This should create a benefit to both FDA and patients. In addition to tools to standardize regulatory review documents, labels will be captured as well. A drug review dashboard will promote cohesiveness to the workflows. In addition to providing a platform for more effective review, a major goal of R2R is to provide data for regulatory research that helps basic scientists and computational biologists to create new analytic models back to the reviewers. An example of a model that FDA/NCTR aspires to create is a model related to drug-induced liver injury (DILI).

A major principle about how the R2R framework is structured is the use of a recursive programming approach that allows a seamless interface between the developers and end users. This ensures that tools are rapidly deployed for "real-world" needs that enables iterative improvement and adjustment to refine.

Strengths:

- 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).
- This project is highly integrated and collaborative with other product centers within FDA. At the review meeting, several scientists from other product centers spoke very passionately and positively on behalf of the interactions and dependency that they have for the scientists within NCTR's Bioinformatics and Biocomputing Division to deliver the implementation tools for this project.
- 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.
- The focus on "end user" and "user friendly" tools is another advantage for a highly impactful product.
- 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.

Areas for Development:

- The strength of the collaboration between the FDA product centers is apparent within this project. However, 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. In doing so, they could sit side by side with the reviewers interacting with them and observing them conducting their review work. This ability to "walk in the shoes" of the reviewer should help to facilitate R2R tools.

- 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". Maintaining focus will be key to realizing impact for the amount of resource going into the project(s).
- NCTR scientists should look for solutions to the needs of the project from sectors in industry and academia beyond the internal capabilities within NCTR. It is possible that some projects could move more quickly by leveraging other applications developed already externally. 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.
- 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. 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.

Theme 5: Service & Support Functions

The Service and Support activities provide research and regulatory support to NCTR and FDA Center scientists in all areas of bioinformatics, biostatistics and scientific computing. Since more and more submissions include bioimaging or next-generation sequencing data, this function of the division has become an essential part of the regulatory process. While all three branches of the DBB are engaged in some support function, they are largely represented by the Scientific Computing and the Biostatistics Branches, encompassing approximately 60% of the total DBB activities. However, 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.

Scientific Computing Branch:

The Scientific Computing Branch (SCB) is responsible for the development, customization and maintenance of software applications and databases supporting both research and management functions. This includes quality control, Good Laboratory Practices compliance, document tracking, and the management of all of NCTR's IT investments. Ted Bearden is an enthusiastic and competent leader for the Scientific Computing Branch, representing NCTR's IT needs on several Agency committees and serving as a liaison between NCTR and the Office of Information Management and Technology.

The SCB fills the needs of the research groups through custom software solutions or customized features of commercial software. Examples include the Radial Arm Maze Control software, Electron Microscopy Electronic Notebook, Protocol Data Collection System and Multispecies Behavioral System, which support several research branches within NCTR. Another impressive example was the development of a custom communications hub similar to Facebook called NCTR Reach. 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. A contrast to these examples was the apparent lack of infrastructure support for the Bioinformatics Branch, other than hardware support. 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. 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.

The SCB management of all IT infrastructure at NCTR is a huge lift given the dynamic nature of technology advancements and cybersecurity requirements for data security. 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 the 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.

Biostatistics Branch:

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. 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.

Overall Subcommittee Conclusions and Suggestions:

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. 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 that have been very useful across multiple NCTR Divisions and FDA Product Centers. Similarly, the Division seems to have done a good job of supporting the bioinformatics needs of other NCTR Divisions and FDA Product Centers and 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. This is perhaps not surprising given that the Division is relatively new as an independent Division and is still growing into its own, but it was driven home by the fact that few of the presentations during the site visit indicated how Division activities were advancing or impacting FDA regulatory activities.

Recommendations:

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". 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? 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.

2. 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.

3. 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.

4. 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.

5. 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. While publishing is important for external validation of research, a balance between publishing and tangible impact to food or drug safety should be a priority. It is recommended to work with the Centers to define and capture metrics for impact, including outreach to the global regulatory community.

6. The Subcommittee shares concerns raised by at least one presentation during the site visit regarding leadership training and a succession plan for the Division. While the Division has extremely competent leaders in Dr. Tong, Dr. Chen and Mr. Bearden, it is not clear that there is a succession plan in place. Who are the future leaders in the Division? Do the current Division leaders have a plan for growing internal talent or recruiting new talent? 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.

Appendix 1. Division of Bioinformatics and Biostatics Site Visit Review Subcommittee Members

Cynthia Afshari, PhD, DABT, FDA SAB Member, Temporary subcommittee consultant
Scientific Executive Director
Discovery Toxicology, Comparative Biology and Safety Sciences
Amgen, Inc., Thousand Oaks, CA 91320
Expertise: Molecular Toxicology, Systems Biology, Organ Injury and Genomics

Pamela Lein, PhD, Site Visit Co-Chairperson and NCTR SAB Member
Professor and Vice-Chair, Molecular Biosciences, University of California, Davis
Director, UC Davis CounterACT Center of Excellence
School of Veterinary Medicine, UC Davis, Davis, CA
Expertise: Molecular Toxicology, Neurotoxicology

Ying Lu, PhD
Professor of Biostatistics, Stanford School of Medicine
Director of Palo Alto Cooperative Studies Program Center
Palo Alto VA Healthcare System, Palo Alto, CA
Expertise: Biostatistics and Meta-Analyses of Clinical Data

John Quakenbush, PhD
Professor of Biostatistics, Harvard T.H. Chan School of Public Health
Professor of Computational Biology and Bioinformatics, Dana-Farber Cancer Institute
Director, Center for Cancer Computational Biology, Dana-Farber Cancer Institute
Boston, Massachusetts 02215
Expertise: Bioinformatics and Big Data Analyses

Kenneth Ramos, MD, PhD, PharmB
Associate Vice President of Precision Health Sciences
Executive Director, Center for Applied Genetics and Genomic Medicine
University of Arizona Health Sciences, Tucson, AZ
Expertise: Personalized (aka precision) Medicine and Genomics

Katrina Waters, PhD, Site Visit Chairperson and NCTR SAB Member
Scientist, Earth & Biological Sciences
Deputy Division Director for Biological Sciences
Pacific Northwest National Laboratory, Richland, WA
Expertise: Bioinformatics and Computational Toxicology

Appendix 2. Charge Memo, July 13, 2015

Dear SAB Subcommittee Review Members:

Donna and I welcome your participation in the review of the Division of Bioinformatics and Biostatistics (DBB) at the National Center for Toxicological Research (NCTR). The NCTR conducts external reviews of its research programs by eminent scientists in their fields as a means to provide independent scientific guidance, technical advice, and recommendations on strategic direction and mission relevance to the NCTR leadership and program staff. The Science Advisory Board (SAB) Subcommittee reviews of the research programs have been a cornerstone for research management in assessing the research conducted at the NCTR for quality, performance, scientific soundness, and pertinence in addressing public health issues. To accomplish this purpose, the SAB Subcommittee review is designed to focus on the relevance and quality of the science performed within the Division's research program and evaluates the overall progress as related to the strategic goals of the NCTR/FDA in advancing regulatory science. On November 5-6, 2015, the DBB will be undergoing a SAB subcommittee site visit meeting as part of the cyclic review of research programs at the NCTR.

It is anticipated that the SAB Subcommittee review of the DBB portfolio will provide objective advice to the NCTR Director, the Division Director, and researchers on strengths and perceived weaknesses of each aspect of the research program.

- Research projects proposed at NCTR are in response to requests from the regulatory centers and/or are PI initiated and conducted by senior scientists with the assistance of staff fellows and postdoctoral fellows. Thus, these projects are critical to the success of NCTR's mission and goals, and the quality of the science must be state-of-the-art and worthy of publication in peer-reviewed journals. What is SAB's evaluation of the research being conducted in DBB?
- Evaluate how NCTR and its Divisions may improve horizon-scanning for emerging sciences and comprehensive safety assessment approaches

In addition, the Subcommittee should:

- Identify and discuss critical emerging regulatory, research, scientific issues, trends, and needs in relation to the research capabilities of the NCTR/FDA
- Assist in identifying areas that are less relevant to the FDA's public health mission

Two weeks prior to the meeting, you will be receiving:

- Individual written summaries on the major research programs of the DBB. Included will be language on any areas that are being done at the behest of a FDA regulatory center for which guidance is not needed
- The PowerPoint presentations of the Director, his research scientists, and current CVs of the Principal Investigators in the Division
- The most recent strategic plans of the FDA and NCTR

After the meeting

- The committee will generate a report of their findings within 4 months.
- The Director of the Division under review will provide his/her feedback within 3 months.

We look forward to your visit and your review of the DBB.

Sincerely,

Donna and Dan

Appendix 3. Division of Bioinformatics and Biostatistics Site Visit Agenda

Thursday, November 5, 2015	
7:00 am	Transportation depart Hotel
8:00 – 8:30 am	Continental Breakfast – Bldg. 12 Auditorium

Thursday Morning

8:30 – 9:00 am	**Welcome, Purpose of Review, Opening Remarks**
	William Slikker, Jr., Ph.D., Director, NCTR Katrina Waters, Ph.D., Chairwoman, SAB Site Visit Team Pamela Lein, Ph.D., Co-Chairwoman, SAB Site Visit Team
9:00 – 10:10 am	**Division Overview**
9:00 – 9:10	Weida Tong (Division Director): Division Overview
9:10 – 9:30	Weida Tong: Branch Overview - Bioinformatics
9:30 – 9:50	James Chen: Branch Overview - Biostatistics
9:50 – 10:10	Edward (Ted) Bearden: Branch Overview - Scientific Computing
10:10 – 12:00 am	**Theme 1: Precision Medicine**
10:10 – 10:30	Weida Tong: Sequencing Quality Control: Phase II (SEQC2) – A community wide consortium effort (15 min+5 min Q&A)
10:30 – 11:00	BREAK
11:00 – 11:20	Huixiao Hong: Developing the best practice for variants detection with whole genome sequencing (WGS) (15 min+5 min Q&A)
11:20 – 11:40	Wenming Xiao: Benchmarking personal genome assembly with whole genome sequencing towards precision medicine (15 min+5 min Q&A)
11:40 – 12:00	Joshua Xu: Assessing detection power of deep-sequencing for rare mutants (15 min+5 min Q&A)

12:00 – 1:00 pm LUNCH and Poster Viewing – SAB Conference Room, Bldg. 12

Thursday Afternoon

1:00 – 2:30 pm	**Theme 2: Predictive Toxicology**
1:00 – 1:20	Weida Tong: Introduction
1:20 – 2:00	Minjun Chen: Liver Toxicity Knowledge Base (LTKB) (25 min+15 min Q&A)
2:00 – 2:30	Yuping Wang: MicroRNAs as preclinical safety biomarkers for hepatotoxicity (20 min+10 min Q&A)

2:30 – 4:00 pm	**Poster and software demo**
4:00 pm	Transport to Little Rock
6:00 pm	Dinner: TBD

Friday, November 6, 2015	
7:00 am	Transportation depart Hotel
8:00 – 8:30 am	Continental Breakfast – Bldg. 12 Auditorium

8:30 – 9:50 am	**Theme 3: Biostatistical Approaches and Applications**
8:30 – 9:10	James Chen: Biostatistical Research in Predictive Medicine and Machine Learning (25 min+15 min Q&A)
9:10 – 9:30	Wen Zou: Topic Model in Text Mining (15 min+5 min Q&A)
9:30 – 9:50	Ching-Wei Chang: Statistical Analysis and Challenges in Next-Generation Sequencing (15 min+5 min Q&A)
9:50 – 10:20	BREAK
10:20 – 11:40 am	**Theme 4: R2R (Research-to-review and return) – a cross-center collaboration framework and activities**
10:20 – 10:40	Roger Perkins (Senior Advisor): Introduction
10:40 – 11:00	Zhichao Liu: FDALabel as a means for an enhanced regulatory science (15 min+5 min Q&A)
11:00 – 11:20	Joshua Xu: Food contamination detection with image analysis (15 min+5 min Q&A)
11:20 – 11:40	Joseph Meehan: DASH to support FDA review and decision-making (15 min+5 min Q&A)
11:40 – 12:00 pm	**Final comments and questions from SAB**

12:00 – 2:00 pm	Closed door Subcommittee session with working lunch SAB Conference Room, Bldg. 12
2:00 pm	Transport to airport