FDA U.S. FOOD & DRUG ADMINISTRATION

Division of Bioinformatics and Biostatistics

Weida Tong, Ph.D.

National Center For Toxicological Research

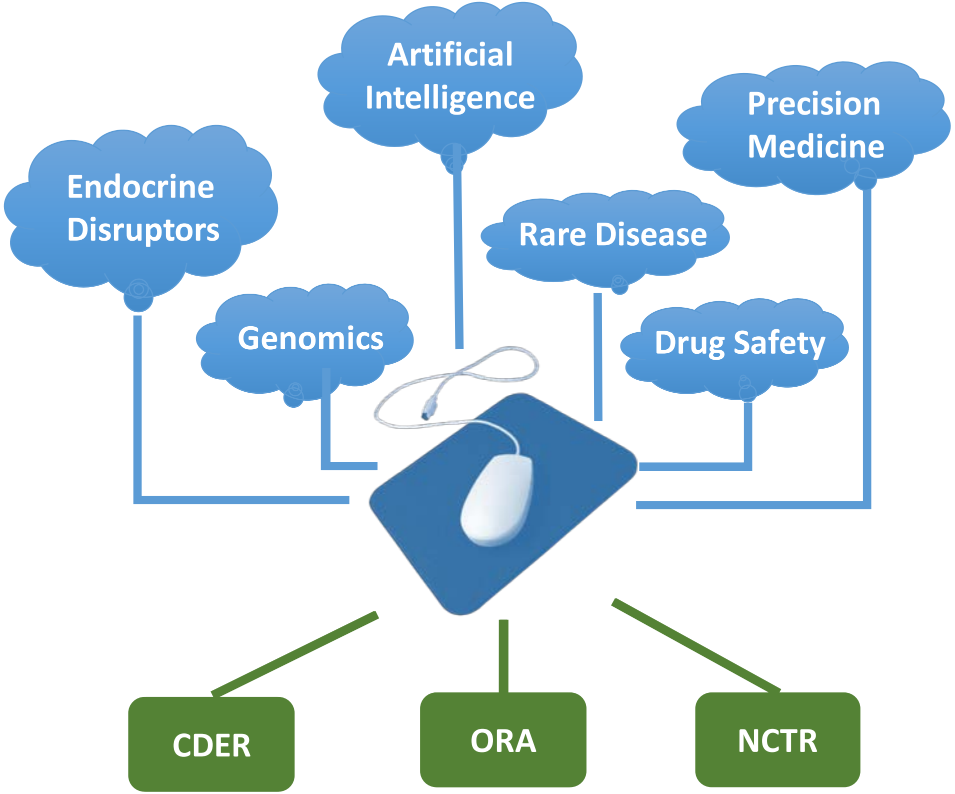
U.S. Food and Drug Administration

Division Staff (>50)

- Four branches: Bioinformatics, Biostatistics, R2R and Scientific Computing
- Government Positions (# FTE = 40 plus 9 vacancies)
  - Research Scientists, Staff Fellows & Visiting Scientists : 14
  - Support Scientists : 22
  - Administrative : 4
- ORISE Post Docs and Graduate Students: 11

- Division at-a-glance
  - Multidisciplinary
  - 40% in research and 60% in support/service

Division Overview and Missions



Research

- To conduct integrative bioinformatics and biostatistics research to support FDA's mission of improving the safety and efficacy of FDA-regulated products.

Support

- To ensure that the division's activities relate to FDA's review process, our linkages with product centers continue to be strengthened, and our capabilities evolve to meet the current and future needs of FDA.

Support/Service at NCTR

- Legacy activities: a part of center-wide infrastructure and investment
  - Working with on-site OMIT staff (12 FTEs) to take care of IT infrastructure and related support: Computer Center (135 servers, PB of data storage, HPC cluster)

- Bioinformatics specific support: Establish data analysis environment, manage commercial and in-house software tools, and conduct training courses
  - Manage HPC for big data analytics
  - Next-generation sequencing (NGS) data: implement Galaxy platform and manage CLC Genomics Workbench
  - Offer annual hands-on training to use these tools (collaborated with OSC)

Collaborations within NCTR

- Selected cross-division collaborations:
  - NeuroTox: 2 projects on sequencing data analysis
  - DGMT: 1 project on sequencing data analysis
  - DBT: 3 projects, two on text mining to support monograph review and 1 for genomics data analysis

- Develop FDALabel to manage the FDA drug labeling data to support drug review and regulatory application
  - Led by NCTR/OSC, developed by DBB, and expert consultant from DSB
  - Partner with CDER (Led by CDER/OTS/OCS)
    - OCS (co-operating with NCTR) for collecting requirements from reviewers, provide training and user support http://inside.fda.gov:9003/CDER/OfficeofTranslationalSciences/OfficeofComputationalScience/ucm449277.htm
    - OND (Drug Labeling Expert): Recommended as a drug review tool http://inside.fda.gov:9003/CDER/OfficeofNewDrugs/ImmediateOffice/LabelingDevelopmentTeam/ucm025576.htm
    - Communicating with OGD and OPQ of CDER as well as CBER and CVM
  - FDA resource for public
    - https://www.fda.gov/science-research/bioinformatics-tools/fdalabel-full-text-search-drug-labeling
    - https://www.fda.gov/drugs/development-resources/labeling-information-drug-products

Collaboration with CDER

- Completed Projects in 2019:
  - Breakthrough Therapy Designation (BTD) system (CDER/OND)
  - Text mining study of OND regulatory documents (Meeting Minutes) (CDER/OND)

- On-Going Projects:
  - Support DASH (Data Analysis Search Host) Tool (CDER/OTS)
  - Develop *IND Smart Template* to standardize the IND data submission and management (CDER/OCS)
  - Risk Evaluation and Mitigation Strategy (REMS) (CDER Office of Communication)

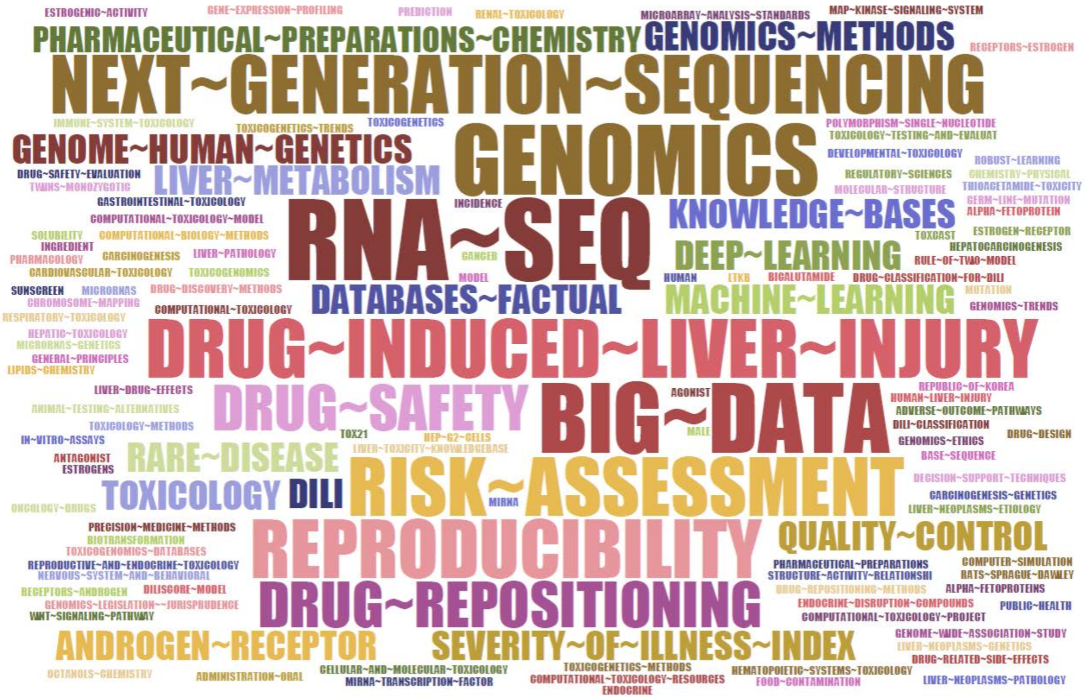
- New Project: Develop Safety Policy Research Team (SPRT) system (CDER/OND)
  - Establish a post-market safety database that will facilitate systematic analyses of post-market safety actions, policies, and outcomes
  - Leverage information from existing databases/spreadsheets that have been developed to evaluate post-market drug safety issues
  - Safety information in SPRT will be linked to associated data in DASH
  - Natural Language Processing of regulatory documents will be used to help populate SPRT

Collaboration with ORA

- Prototyping Automated Laboratory Information System (ALIS)
  - Near completion: the module for salmonella testing
    - On-site demo to the ORA labs here and at New York
    - Remote demo to other ORA sites
    - Discussion for production
  - Pesticide detecting module: on-going and estimated to complete by the end of FY20

- Two artificial intelligence (AI) centric projects:
  - Deep learning for image analysis of identification of storage pests fragments contaminating food product
  - Machine learning of mass spectrometer data for identification of persistent organic pollutants (Chief Scientists Challenge Grant project)

DBB Research Priorities and Accomplishments



Accomplishment #1: Genomics

- MicroArray and Sequencing Quality Control (**MAQC/SEQC**): An FDA-led consortium effort to assess technical performance and application of emerging technologies for safety evaluation and clinical application
  - Started in 2005 and completed 3 projects by 2014 with ~30 publications

- On-going (Fourth project): Sequencing Quality Control Phase 2 (SEQC2)
  - Area #1: Cancer genomics using whole genome sequencing
    - One paper is tentatively accepted, conditional on the acceptance of the other one which is under review, both are with Nat Biotechnol
    - Additional 2 papers will be submitted by FY20
  - Area #2: Cancer genomics using target gene sequencing - 4 papers are scheduled for submission in FY20
  - Area #3: Reproducibility of whole genome sequencing - 2 papers are scheduled for submission in FY20
  - Area #4: Epigenomics – 2 papers in queue for submission

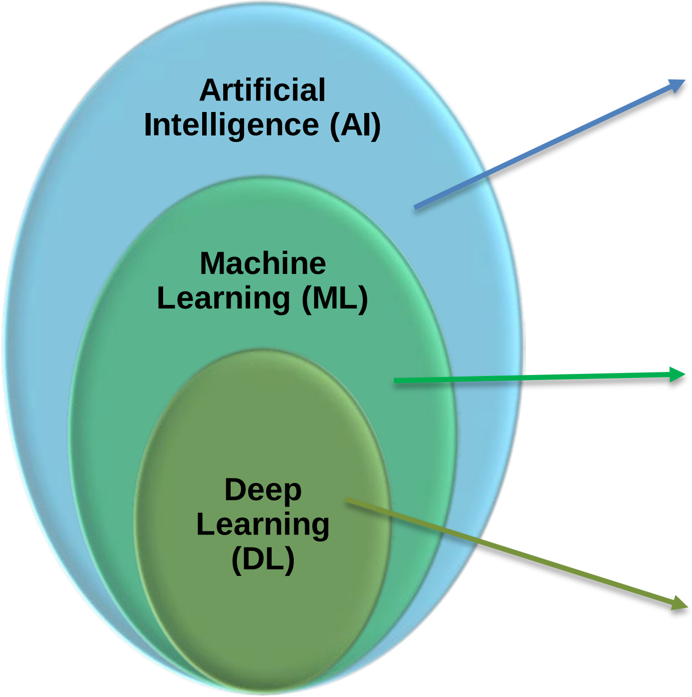
Accomplishment #2: Drug-Induced Liver Injury (DILI)

- Background:
  - About 50% drugs fail in clinical settings due to DILI not being detected by preclinical models
  - Alternative methodologies, particularly these high-throughput methods such as in vitro and in silico approaches can play a role in detecting human DILI
  - Most of these methods rely on a large list of drugs with known human DILI
  - One of the key components of Liver Toxicity Knowledge Base (LTKB) is to produce such a list to evaluate alternative methodologies

- Drug lists with known human DILI from LTKB:
  - 2011: Benchmark DILI dataset – around 280 drugs were classified based on FDA drug labeling documents (published in DDT)
  - 2016: DILIrank – >700 drugs were annotated with causality assessment (Published in DDT)
  - 2019: DILIst – around 1300 drugs were classified (in press, DDT)

Big Data Analytics and Artificial Intelligence (AI)

- AI Research Force (AIRForce)

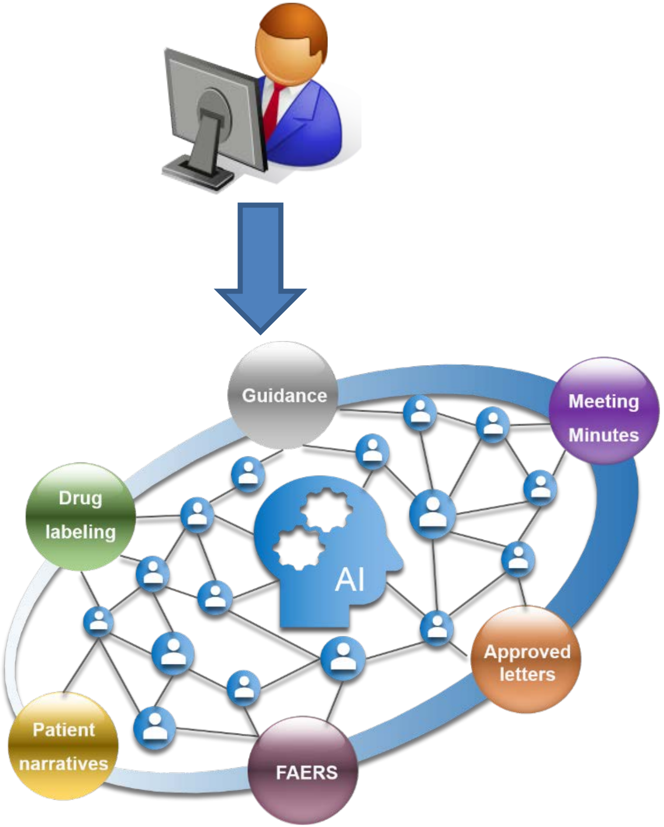


- DeepReviewer: An AI framework to support regulatory review process (on-going)
- Genomics Biomarkers for DILI: A crowdsourcing project with CAMDA (Accomplishment #3)
- AI Challenge project: collaborating with PrecisionFDA (on-going)

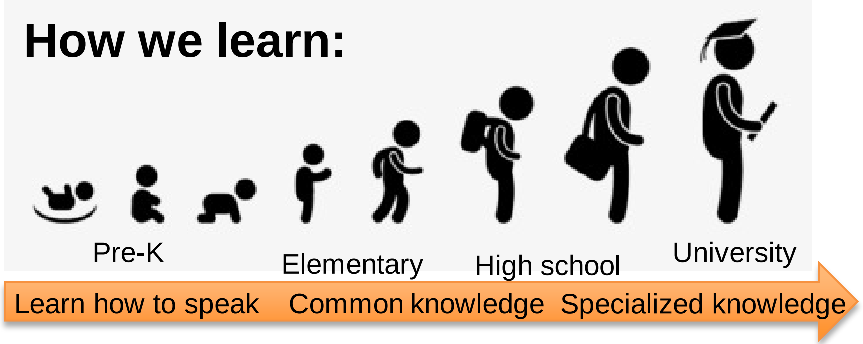
DeepReviewer

- Challenges in review process
  - Difficulty in accessing historical knowledge due to turn-over
  - Rapid access to relevant knowledge (internal and external)
  - Maintaining the institutional memory

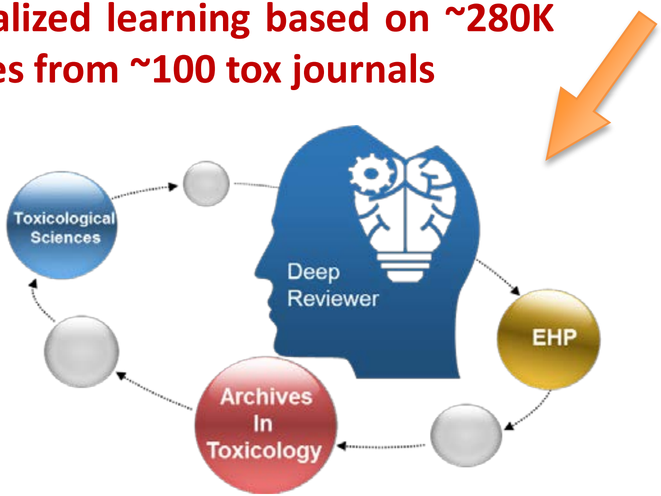
- Development of an AI framework to assist review process
  - Help access both public documents and regulatory-related documents at FDA



Proof-of-Concept: Safety Assessment



Specialized learning based on ~280K articles from ~100 tox journals



Information Retrieval: the activity of obtaining information system resources that are relevant to an information need from a collection of those resources

Text Summarization: Text summarization refers to the technique of shortening long pieces of text to create coherent and fluent main points outlined in the document.

Questioning & Answering: Given a question and a set of candidate answers, answer selection is the task of identifying which candidate answers the question correctly.

Sentiment Analysis: refers to the use of NLP to systematically identify, extract, quantify, and study affective states and subjective information

Example 1: Common vs. Specialized Knowledge

Query “liver”

Learn from Google web	Learn from Tox journals
kidney (0.739)	**hepatic (0.716)**
pancreas (0.723)	kidney (0.683)
kidneys (0.717)	pancreas (0.553)
livers (0.656)	lung (0.516)
lung (0.639)	tissue (0.509)
bone_marrow (0.621)	**hepatocellular (0.503)**
internal_organs (0.617)	**hepatocytes (0.503)**
intestine (0.607)	spleen (0.499)
liver_kidneys (0.603)	testis (0.482)
liver_disease (0.599)	intestine (0.478)

Query “acetaminophen”

Learn from Google Web	Learn from Tox journals
Ibuprofen (0.658)	**APAP** (0.773)
Acetaminophen (0.649)	paracetamol (0.709)
NSAID (0.642)	AAP (0.661)
Decongestants (0.640)	bromobenzene (0.603)
pain_relievers (0.630)	hydroxyacetanilide (0.594)
Paracetamol (0.624)	**overdosed** (0.571)
NSAIDs (0.623)	galactosamine (0.570)
Dextromethorphan (0.622)	**ALF (0.563)**
Acetominophen (0.620)	amap (0.553)
Tylenol_acetaminophen (0.617)	diclofenac (0.535)

Example 2: Basic vs. Advanced Algorithms

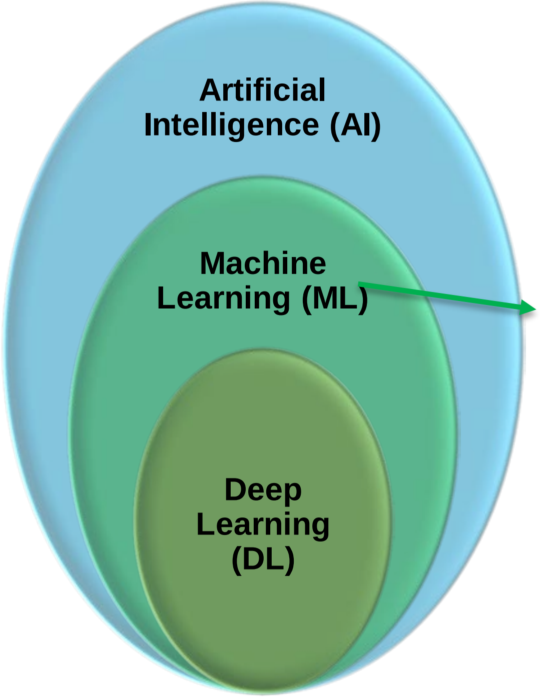
Query "liver"

Learn from Google web	Learn from Tox journals
kidney (0.739)	**hepatic (0.716)**
pancreas (0.723)	kidney (0.683)
kidneys (0.717)	pancreas (0.553)
livers (0.656)	lung (0.516)
lung (0.639)	tissue (0.509)
bone_marrow (0.621)	**hepatocellular (0.503)**
internal_organs (0.617)	**hepatocytes (0.503)**
intestine (0.607)	spleen (0.499)
liver_kidneys (0.603)	testis (0.482)
liver_disease (0.599)	intestine (0.478)

Query "acetaminophen"

Word2vec	FastText
APAP (0.773)	aminophenazone (0.770)
paracetamol (0.709)	paracetamol (0.760)
AAP (0.661)	propacetamol (0.730)
bromobenzene (0.603)	acetamidophenol (0.729)
hydroxyacetanilide (0.594)	acetamide (0.720)
overdosed (0.571)	aminophenazine (0.717)
galactosamine (0.570)	bisacetamide (0.717)
ALF (0.563)	thioacetamide (0.711)
amap (0.553)	acetamido (0.709)
diclofenac (0.535)	aminophenol (0.707)

MAQC Consortium Projects (2005 – 2014)



- Evaluated machine learning for gene expression based predictive models and biomarkers
- Unresolved questions
  - Data size
    - More samples and a better model, but how much is enough
    - Do more features lead to a better model?
  - Do sophisticated methods offer opportunities, e.g., deep learning?

MAQC-II, Nat Biotechnol (2010)
Wang et al. Nat Biotechnol (2014)
Su et al. Genome Biology (2014)
Zhang et al. Genome Biology (2015)

Accomplishment #3: cMAP Drug Safety Challenge

- 2018: Led a CAMDA Challenge for AI/ML to predict DILI with genomics data
  - CAMDA = Critical Assessment of Massive Data Analysis (established in 2000), is a platform to evaluate big data analytics using a crowdsourcing challenge mechanism
  - Participants: 11 teams from nine countries
  - Observations: Deep Learning outperforms conventional machine learning methods, however, we need to
    - Have large datasets to confirm this finding
    - Examine its generalization, particularly for clinical application

Collaborating with PrecisionFDA on AI Challenge

- Assessing AL/DL for biomarker development (planning stage)

- Office of Health Informatics of Office of Chief Scientist (OCS) established PrecisionFDA in 2105 with a focus on
  - Development of standards and tools of Next Generation Sequencing and omics technologies
  - Helping advance regulatory science by running scientific Challenges with cross-Center participation

- OCS/OHI has conducted several PrecisionFDA challenges:
  - The CFSAN Pathogen Detection Challenge
  - The CDRH Biothreat Challenge
  - The NCI-CPTAC Multi-omics Enabled Sample Mislabeling Correction Challenge

- Has 4000+ registered users

Future Directions

- Research:
  - To continually develop big data analytics, particularly in the area of AI for FDA data (e.g., DeepReviewer and DeepLabel)
  - To study computational reproducibility
  - To investigate real-world data as real-world evidence to support FDA missions such as electronic health records (EHRs) data
  - To evaluate alternative methodologies for predicting drug safety such as DILI
- Support:
  - Increase data analysis support such as imaging and NGS data
  - Further improve collaborations with other Centers

Trends in Pharmacological Sciences

CellPress REVIEWS

Feature Review

Lessons Learned from Two Decades of Anticancer Drugs

Zhichao Liu,[1,*] Brian Delavan,[1,2] Ruth Roberts,[3,4] and
Weida Tong[1,*]

Tremendous efforts have been made to elucidate the basis of cancer biology
with the aim of promoting anticancer drug
past 20 years, anticancer drug development
cytotoxic agents to target-based and im
quently, more than 200 anticancer drugs a
ever, anticancer drug development still suf
phases of clinical development and is cons
therapeutic category within the drug devel
performance of investigational anticancer
some shortcomings in the translation of pre
to humans, and that heterogeneity in the pat
cant challenge. Here, we summarize both su
anticancer development during the past 20
current paradigm may be suboptimal. We
improvement.

Trends

Current Progress in Anticancer Drug Develo

Cancer, which is characterized by the uncontrolled
most difficult and complex diseases to treat [1–3]. F
rates, which range from 1.1% for prostate cancer to 9
of cancer diagnosis. Therefore, anticancer drug rese
lenging and daunting activity, and the likelihood of fail
compounds developed reach the market [5]. Furtherm
per attempt that is approximately greater than on
therapeutic categories, such as cardiovascular dis
are many difficulties and barriers in anticancer drug
still pursuing opportunities for anticancer drug candi
[7,8]. For example, oncology is ranked in the top t
amounted to US$78.94 billion in 2015.

Approved Anticancer Drugs

The ultimate task for an anticancer drug is to ki
proliferation to prolong patient survival and impro
are many different mechanisms by which this can be
mechanism of anticancer action, an anticancer drug
biosynthesis blocker; (ii) structure and function of D
and RNA synthesis blocker; (iv) protein synthesis and
stasis influencer; or (vi) immune system modulator.
produced four major groups of anticancer drugs:

852 Trends in Pharmacological Sciences, October 2017, Vol. 38, No. 10
Published by Elsevier Ltd.

Trends in Pharmacological Sciences

CellPress REVIEWS

Opinion

Toxicogenomics: A 2020 Vision

Zhichao Liu,[1,*] Ruili Huang,[2] Ruth Roberts,[3,4] and Weida Tong[1,*]

Toxicogenomics (TGx) has contributed significantly to toxicology and now has great potential to support moves towards animal-free approaches in regulatory decision making. Here, we discuss *in vitro* TGx systems and their potential impact on risk assessment. We raise awareness of the rapid advancement of genomics technologies, which generates novel genomics features essential for enhanced risk assessment. We specifically emphasize the importance of reproducibility in utilizing TGx in the regulatory setting. We also highlight the role of machine learning (particularly deep learning) in developing TGx-based predictive models. Lastly, we touch on the topics of how TGx approaches could facilitate adverse outcome pathways (AOP) development and enhance read-across strategies to further regulatory application. Finally, we summarize current efforts to develop TGx for risk assessment and set out remaining challenges.

Toxicogenomics in Regulatory Application: Challenges and Opportunities

Animal models are used to assess and avoid risk to humans from exposure to potential hazards, but their use is under constant review, especially in the light of some reports of poor extrapolation for complex endpoints, such as hepatotoxicity and carcinogenicity. Consequently, 21st century toxicology emphasizes alternative means of risk assessment and the promotion of the 3Rs (replacement, reduction, and refinement of animals in toxicology testing) [1]. In Europe, great efforts have been made to advance the 3Rs with the aim of developing animal-free risk assessment methodologies. To this end, several high-profile programs are underway, such as the Framework Programme 7 (FP7), Horizon 2020, and some public–private partnerships, including Safety Evaluation Ultimately Replacing Animal Testing (SEURAT-1) and the Innovative Medicines Initiative (IMI). Furthermore, a series of EU Legislative directives have been developed and implemented over the past three decades, with an emphasis on moving away from animal testing; since 2013, animal models have been prohibited for testing cosmetics or household products in the EU, as well as in Israel and India [2]. In the US, government-initiated efforts comprise advanced regulatory sciences proposed by the US FDA [3] and Tox21 [4] [which involves four government agencies, including the Environmental Protection Agency (EPA), National Center for Advancing Translational Sciences, National Institute of Environmental Health Sciences, and the FDA] and ToxCast [5] (by the EPA). These ongoing efforts actively advocate and promote *in silico* and *in vitro* approaches, including **toxicogenomics (TGx)** (see Glossary), for prioritization and also for a potential application in risk assessment.

TGx, as a subdiscipline of toxicology, has been successfully implemented to address critical issues and questions in a broad spectrum of toxicology. The rapid advancement of next-generation sequencing (NGS) technologies has gained traction in clinical application, particularly in personalized cancer diagnosis and prognosis, offering great opportunities for precision medicine. Meanwhile, the entire toxicology field has also been impacted by the rapid development with these advanced sequencing technologies. Equipped with innovative technologies in

Highlights

Together with the promotion of non-animal testing, *in vitro* toxicogenomics (TGx) may play a vital role in the next-generation risk assessment paradigm.

A strategic shift in risk assessment provides an unprecedented opportunity for repositioning TGx in the regulatory setting.

As the emerging technique continues to impact the TGx field, novel genomic features such as miRNAs, ncRNAs, and circular RNAs may provide more resolution towards better understanding of the underlying mechanisms of toxicological processes.

Advances in machine learning and artificial intelligence are gaining ground for their applicability in biomedical fields. In the near future, these advances may be further applied in the TGx field to improve predictive power.

[1] National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas 72079, USA
[2] National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, USA
[3] ApconiX, BioHub at Alderley Park, Alderley Edge, SK10 4TG, UK
[4] University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

*Correspondence:
zhichao.liu@fda.hhs.gov (Z. Liu) and
weida.tong@fda.hhs.gov (W. Tong).

92 Trends in Pharmacological Sciences, February 2019, Vol. 40, No. 2 https://doi.org/10.1016/j.tips.2018.12.001
© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

CrossMark

Please cite this article in press as: Liu et al., Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We?, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.08.006

Trends in Genetics

CellPress REVIEWS

Review

Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We?

Zhichao Liu,[1,*] Liyuan Zhu,[1] Ruth Roberts,[2,3] and Weida Tong[1,*]

Next-generation sequencing (NGS) technologies have changed the landscape of genetic testing in rare diseases. However, the rapid evolution of NGS technologies has outpaced its clinical adoption. Here, we re-evaluate the critical steps in the clinical application of NGS-based genetic testing from an informatics perspective. We suggest a 'fit-for-purpose' triage of current NGS technologies. We also point out potential shortcomings in the clinical management of genetic variants and offer ideas for potential improvement. We specifically emphasize the importance of ensuring the accuracy and reproducibility of NGS-based genetic testing in the context of rare disease diagnosis. We highlight the role of artificial intelligence (AI) in enhancing understanding and prioritization of variance in the clinical setting and propose deep learning frameworks for further investigation.

Introduction to Rare Diseases

Approximately 7000 rare diseases have been recognized, a substantial number of which are life-threatening or chronically debilitating [1]. Around 80% of rare diseases are genetic in origin. A single rare disease affects a small number of the population (defined as < 1/15 000 in the US and < 1/2000 in Europe) but on aggregate, an estimated 350 million people globally suffer from rare diseases. Most rare disease patients (50%–75%) show onset at birth or in childhood. As many as 30% of rare diseases patients die before the age of 5 years. Furthermore, each rare disease patient has been estimated to cost a total of 5 million dollars throughout their life span.

Incomplete knowledge of **natural history** (see Glossary) and lack of awareness confounds rare disease diagnosis. The average length of accurate diagnosis of a rare disease is 4.8 years and involves more than seven physicians or specialists who may be geographically distributed. An often-protracted path to the diagnosis of rare diseases poses an immense burden and psychological distress to patients and their family and a strong challenge to the current healthcare system [2]. The rare disease patients and family may benefit from genetic diagnosis. The genetic diagnosis may not be directly associated with any treatment options, and physicians will continue to treat symptoms, albeit in a more informed way based on likely prognosis of the case. Therefore, genetic diagnosis could be of benefit beyond treatment management as it can offer information to families, many of who just want to know what is wrong with their family member, and can also inform fertility decisions.

Next-Generation Sequencing-Based Genetic Diagnosis: Challenge and Opportunities

Emerging genomics technologies, such as next-generation sequencing (NGS), have been intensively applied in a research setting but also offer great opportunities in the clinical setting [3–5]. Despite the remarkable progress of NGS-based genetic testing[i] for improving the discovery of genetic variants in rare disease, the translational gap between NGS-based genetic testing and clinical implementation remains. Many factors contribute to the suboptimal translation of NGS technology into a rare disease diagnosis. The acceptability and uptake of NGS-based genetic testing depends upon a clear demonstration of patient benefit driven by providing physicians with the tools for enhanced decision making. In this context, real-world evidence in support of NGS-based genetic testing is often limited.

There are several challenges to overcome before NGS can deliver its potential for patients, clinicians, and society. The key debate is whether NGS-based genetic testing can produce accurate and reproducible results that would support clinical decision making for rare disease diagnosis. To address this,

Highlights

NGS-based genetic testing in the diagnosis of rare diseases holds great promise to serve as a first-tier genetic testing tool in the near future.

Advancement of NGS technologies provides many options for diagnosing rare d
with different
iants. Factors m
'fit-for-purpose

The accuracy a
NGS should be
ical setting to
netic testing
clinical stages

Artificial intelli
central role in
diagnosis info
enhanced diag
diseases.

[1] National Center
Research, U.S. Fo
Administration, Je
[2] ApconiX, Alderle
SK10 4TG, UK
[3] University of Bir
Birmingham, B15

*Correspondence
zhichao.liu@fda.h
weida.tong@fda.h

Trends in Genetics, ■■, Vol. ■■, No. ■■ https://doi.org/10.1016/j.t
Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/

Trends in Pharmacological Sciences

CellPress

Opinion

Potential Reuse of Oncology Drugs in the Treatment of Rare Diseases

Zhichao Liu,[1,*] Hong Fang,[1] William Slikker,[1] and Weida Tong[1,*]

Cancer research has made remarkable progress with the help of advancing genomics techniques, resulting in more precise clinical application and many new anticancer drugs on the market. By contrast, very few treatment options are available for rare diseases that are often progressive, severe, and life-threatening. In this opinion we elaborate on the possible association between cancers and rare diseases across three different levels including clinical observation, crosstalk between germline mutation and somatic mutation, and shared biological pathways. Consequently, by utilizing systematic drug-repositioning approaches, and taking safety issues into consideration, we suggest that oncology drugs have great potential for reuse in the treatment of rare diseases.

Significance of Rare Diseases in Public Health

Rare diseases are usually chronic, serious, and even life-threatening, which creates a burden on society and public health systems [1]. Each rare disease individually affects only small number in the population (<200 000 persons in the USA and <50 000 persons in Europe). For instance, LEOPARD syndrome (LS) is an extremely rare genetic disorder characterized by significant cardiac and skin abnormalities. Only ~200 patients with LS had been reported worldwide by the year 2008 [2]. However, as a group, 'rare diseases' are far from rare, and collectively they affect approximately 30 million Americans [3]. About 7000 rare diseases have been identified, and most have a genetic origin, affecting patients from birth. The molecular etiology of rare diseases is often poorly understood; thus far research has only identified the genetic mutations that underlie one-third of rare diseases [4].

Around 80% of rare diseases are caused by germline mutation, and the majority of these mutations are single-nucleotide polymorphisms (SNPs) [4]. For example, Tay–Sachs disease is a rare fatal genetic disorder caused by mutations in *HEXA*, the gene encoding β-hexosaminidase A, an enzyme that plays a crucial role in the brain and spinal cord. SNP mutations in *HEXA* result in severe damage to the central nervous system (CNS) and ultimately death [5]. Rare diseases are considered to be an issue of 'precision medicine' because accurate diagnosis is generally not based on symptoms but instead on genetics and genomic laboratory tests [6]. The identification of causative genes for rare diseases has made great progress with the advance of new genomic techniques such as next-generation sequencing (NGS) [7]. Specifically, whole-exome sequencing (WES) that aims to detect genetic variants located in the protein-coding region is a powerful tool for the identification of rare disease genetic variants [8,9]. In the past 5 years alone, researchers from the Centers for Mendelian Genetics using genome-sequencing techniques have identified more than 700 genes linked to Mendelian diseases (http://mendelian.org/).

The lack or limited knowledge of the natural history of rare diseases is the biggest obstacle for improving their treatment [10]. It is difficult to collect a sufficient number of patients to conduct

Trends

Evidence suggests that germline mutations associated with rare diseases may predispose patients to develop somatic mutations that trigger tumor development.

In some cases, cancers and rare diseases may perturb the same biological pathways, suggesting the possibility of shared therapeutic targets.

Anticancer drugs may therefore have potential to be repurposed to treat some rare genetic diseases.

[1] National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA

*Correspondence:
Zhichao.liu@fda.hhs.gov (Z. Liu) and
weida.tong@fda.hhs.gov (W. Tong).

CrossMark

Trends in Pharmacological Sciences, October 2016, Vol. 37, No. 10 http://dx.doi.org/10.1016/j.tips.2016.08.010 843
Published by Elsevier Ltd.

Feedback Requested

- Recruiting/filling vacancies:
  - On-going effort: bringing students from the local universities via ORISE, and converting them if qualified
  - Using social media such as LinkedIn
  - Other mechanisms?

- Working with Electronic Health Records (EHRs):
  - We are working with VA EHRs, many challenges and costly
  - We are looking into MIMIC and NHANES
  - Other datasets?

FDA U.S. FOOD & DRUG ADMINISTRATION