TRANSLATIONAL BIOINFORMATICS 101

JESSICA D. TENENBAUM
Department of Bioinformatics and Biostatistics, Duke University
Durham, NC 27715 USA
Jessie.Tenenbaum@duke.edu

SUBHA MADHAVAN
Innovation Center for Biomedical Informatics, Georgetown University
Washington, DC 20007 USA
Subha.Madhavan@georgetown.edu

ROBERT R. FREIMUTH
Department of Health Sciences Research, Mayo Clinic
Rochester, MN 55905 USA
Freimuth.Robert@mayo.edu

JOSHUA C. DENNY
Department of Biomedical Informatics, Vanderbilt University
Nashville, TN 37203 USA
josh.denny@Vanderbilt.Edu

LEWIS FREY
Public Health Sciences, Medical University of South Carolina
Charleston, SC 29425 USA
frey@musc.edu

1. Workshop Focus

This Workshop will give an overview of key topics in the young field of Translational Bioinformatics (TBI). TBI has been defined as the “development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data, and genomic data, into proactive, predictive, preventive, and participatory health.”[1] With PSB’s stated focus on research in databases, algorithms, interfaces, natural language processing, and modeling, the bioinformatics aspect of TBI is a natural fit for this audience. Further, the US government’s recently announced Precision Medicine Initiative makes it particularly timely for researchers to learn about and explore the translational side of our field. Specifically, this workshop provides context for how the various bioinformatics methods may be applied toward the enhancement of human health, enabling healthcare providers to deliver the right intervention for the right person at the right time.

This workshop covers major themes within the field of TBI, as put forth by a recent review in IMIA’s (International Medical Informatics Association) Yearbook of Medical Informatics[2], and supplements those topics with a section on relevant data standards from the clinical and translational domain. Each presenter is a nationally or internationally recognized expert in their respective areas.

2. Workshop Agenda

Title	Speaker
Introduction	J Tenenbaum
Clinical "big data" I The use of EHR data for genomic discovery: the eMERGE network	J Denny
Clinical "big data" II Clinical3PO: Deep Phenotyping to Precision Medicine	L Frey
Omics for drug discovery and repurposing	J Tenenbaum
BREAK	
Intro to the clinic I Standards for Translational Bioinformatics	R Freimuth
Intro to the clinic II G-DOC *Plus*: A TBI platform for novel hypothesis generation in precision medicine research	S Madhavan
Personal genomic testing and related ethical, legal, and social issues	J Tenenbaum

3. Workshop Contributions

The use of EHR data for genomic discovery: the eMERGE network - J. Denny

Precision medicine offers the promise of improved diagnosis and more effective, patient-specific therapies. Typically, such studies have been pursued using research cohorts. Across the Electronic Medical Records and Genomics (eMERGE) Network, we have explored use of the electronic health records (EHRs) linked to DNA biobanks to do genomic and pharmacogenomic discovery. This combination allows study of the genomic basis of disease and drug response using real-world clinical data. Finding phenotypes in the EHR can be challenging, but the combination of billing data, laboratory data, medication exposures, and natural language processing has enabled efficient study of genomic and pharmacogenomic phenotypes. These studies have replicated many known associations as well as posited new genetic findings for diseases not yet studied via other methods. A particular advance may be for drug response traits, for which the EHR has proven cost efficient and effective. The EHR also enables the inverse experiment – starting with a genotype and discovering all the phenotypes with which it is associated – a phenome-wide association study (PheWAS). PheWAS requires a densely phenotyped population such as is found in the EHR. We have used PheWAS to replicate >300 genotype-phenotype associations, characterize pleiotropy, and discover new associations. We have also used PheWAS to identify characteristics with disease subtypes.

Collectively, EHR-linked biobanks across the U.S. alone are approaching 1 million people, portending a future in which these will play an increasingly important role. Indeed, the recently-announced presidential Precision Medicine Initiative highlights the role of EHR-based molecular study as an efficient and powerful longitudinal health data discovery platform. This national research cohort will enroll over 1 million individuals who are re-contactable and share biospecimens and health data. Many of these participants will likely share sensor and mobile technology data as well.

Clinical3PO: deep phenotyping to precision medicine - L. Frey

We will demonstrate components of the open source big data Clinical Personalized Pragmatic Predictions of Outcomes (Clinical3PO) platform, developed for the U.S. Department of Veterans Affairs (VA) along with its extension to typical healthcare environments. Its data representation is the Observational Medical Outcome Partnership (OMOP) common data model, which has been encoded to lower the barrier for cross-institutional data analysis. Using OMOP we will demonstrate the feature extraction module of Clinical3PO for deep phenotyping cohorts and feeding machine learning predictive analytic pipelines. The ability to support big data analytics for deep phenotyping using Clinical3PO applied to medical data will be described.[3] We will show how the pipeline can be used to create machine learning models focused on the care patterns of individuals. The path forward using the system to advance precision medicine will be discussed.

Omics for drug discovery and repurposing - J. Tenenbaum

Much has been written about the notoriously lengthy and expensive processes of drug discovery and FDA approval. From target identification to FDA approval, it is not uncommon for the process to take well over a decade and $1 billion. One major contribution of translational bioinformatics has been to apply a data-driven approach both to drug discovery, and drug repurposing: identifying existing FDA approved drugs that may help treat conditions for which they were not initially intended. High throughput omics technology can be used to identify promising pathways and molecular targets for a given disease, and also to identify molecular signatures of drug administration. These drug signatures can then be compared to signatures that characterize different diseases. Drugs that tend to have opposing effects on disease-related genes and pathways may be good candidates for treatment of those conditions. By effectively bypassing lead identification and Phase 1 trials, this approach can save both time and cost in FDA approval for new indications of existing drugs.

Standards for translational bioinformatics - R. Freimuth

The rapid growth of the biomedical domain has presented researchers and clinicians with more opportunities to share data and knowledge than ever before, but the diversity of data types, analysis methods, and contexts can pose significant challenges to the meaningful exchange, integration, and use of information. Standards can reduce barriers to semantic and syntactic interoperability. This presentation will review examples of existing and emerging standards within the translational bioinformatics community, including data, terminology, and message standards.

The generation, annotation, interpretation, and clinical reporting of genetic test results will serve as a use case throughout the presentation. Specific challenges in this process, such as variability in the representation of genetic data (e.g., nomenclature systems), and the impact that those challenges have on patient care and translational research will be discussed. Existing efforts by both international standards organizations and national consortia to develop normalized systems for the exchange of clinical genetic test results will be reviewed. Finally, methods for sharing knowledge, including that expressed in clinical genomic guidelines and decision support rules, will be summarized.

G-DOC *Plus*: A TBI platform for novel hypothesis generation in precision medicine research - S. Madhavan

G-DOC is a feature-rich shareable translational research infrastructure that allows physician scientists and translational researchers to mine and analyze a variety of "omics" data in the context of consistently defined clinical outcomes data for cancer patients.[4]

Scientists today are using not only a combination of clinical, NGS and omics data for analysis, but also medical and digital images for validation of analysis results. Currently, numerous tools and software exist that specialize in handling and processing of one or two "omics" data types, or only NGS data types, most of which need a bioinformatician to help with analysis. To drive hypothesis generation and validation of molecular markers for biologists and researchers, it would be convenient to have a "one–stop" system that can handle all these data types, including NGS and medical images, in one location without having to switch to other tools or resources for analysis.

With the goal of improving overall health outcomes through genomics research, we present G-DOC *Plus*, a web-based bioinformatics platform that enables the integrative analysis of multiple data types to understand mechanisms of cancer and non-cancer diseases at a systems level for systematic conduct of research in precision medicine. It currently holds data from over 10,000 patients selected from private and public resources including Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA) and the recently added datasets from REpository for Molecular BRAin Neoplasia DaTa (REMBRANDT), caArray studies of lung and colon cancer and the 1000 genomes data sets. G-DOC *Plus* allows researchers to explore clinical-omic data one sample at a time, as a cohort of samples; or at the level of population, providing the user with a comprehensive view of the data.

Three case studies in Pharmacogenomics, cancer variant search, and gene network analysis will be demonstrated to educate workshop attendees on features of G-DOC Plus for novel hypothesis generation to advance precision medicine research.

Direct to consumer genetic testing, and related ethical, legal, and social issues- J. Tenenbaum

Direct to consumer (DTC) genomic services enable individuals to obtain their own genetic data without a healthcare provider acting as intermediary either to order the test or interpret the results. For several years after these services were introduced, it was unclear whether or how they should be regulated by the government. In 2013, the FDA strongly asserted that these tests do indeed fall within their purview, and each health-related association must be separately validated. Their rationale for stepping in was that people might make healthcare decisions, drastic ones even, based on information obtained through these services.

In this portion of the workshop, we describe a pregnancy management case in which a treatment plan was modified based on a DTC result. A woman with no personal or family history of blood clotting-related complications learned through DTC testing about a heterozygous prothrombin (factor 2) gene mutation. Twice daily injections of enoxaparin were recommended throughout pregnancy for this patient based on this genetic information combined with other risk factors including advanced maternal age and pregnancy with twins. Genetically based medical guidelines are a moving target, however, and treatment of thrombophilic conditions in asymptomatic patients is controversial, with guidelines continuing to evolve.

We will also discuss ethical, legal, social, and economic issues raised by this case and its impact on the patient's subsequent efforts to obtain life insurance, which unlike health insurance, is not covered under the Genetic Information Nondiscrimination Act of 2008.

4. References

1. AMIA. *Translational Bioinformatics | AMIA*. 10/5/15]; Available from: http://www.amia.org/applications-informatics/translational-bioinformatics.
2. Denny, J.C., *Surveying Recent Themes in Translational Bioinformatics: Big Data in EHRs, Omics for Drugs, and Personal Genomics.* Yearb Med Inform, 2014. **9**: p. 199-205.
3. Frey, L.J., L. Lenert, and G. Lopez-Campos, *EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group.* Yearb Med Inform, 2014. **9**: p. 206-11.
4. Madhavan, S., et al., *G-DOC: a systems medicine platform for personalized oncology.* Neoplasia, 2011. **13**(9): p. 771-83.