

# Making it personal: translational bioinformatics

doi:10.1136/amiajnl-2013-002028

Atul J Butte, Lucila Ohno-Machado, *Editor-in-chief*

One of the most exciting research areas in Translational Bioinformatics<sup>1,2</sup> is related to the redefinition of fundamental notions of what constitutes a 'disease.' Nosology, the systematic classification of diseases, dates back to Carl Linnaeus, with the *Genera Morborum*<sup>3</sup>. Today, the improvement in our abilities to make molecular measurements related to health and disease has largely driven the revolution towards personalized medicine. For example, in diseases like non-small cell lung cancer or breast cancer, standard-of-care is now including sequencing of genes such as EGFR or quantitating panels of RNA such as those included in Oncotype DX, respectively, to drive therapeutic decisions for new subtypes of patients. While experts, including those at the National Research Council, are seeing the potential of scaling beyond these early case examples towards redefining our entire nosology,<sup>4</sup> it is in the field of cancer where personalized or precision medicine has had best traction. It is no coincidence that many contributions to this special issue of JAMIA focus on cancer. Personalized medicine, also known as precision medicine, has often been equated with the use of molecular measurements to characterize disease. The special feature in this issue of JAMIA challenges this limited view.

Personalized medicine starts even before a disease is manifested in an individual, many times at a point when the disease or condition is preventable. Researchers use data from different sources to develop preventive models. For example, smoking is still the strongest preventable risk factor for many cancers, most notably lung cancer, yet it is hard to extract this information from the electronic medical record, since it is often contained in narrative portions of the record. Bush (*see page 652*) shows that many clinicians use ICD-9 tobacco-use codes and that these codes can indeed be accurately used to identify current smokers, for behavioral or even pharmacological interventions. Once clinicians get this information, they still need to effectively implement prevention strategies. Wagholikar (*see page 749*) shows that automated methods to identify patients who need cervical cancer screening can miss many crucial details, but these details can be 'filled in' by a community of experts. Evidently, genetics is also involved in predicting disease

susceptibility. Urbanowicz (*see page 603*) shows how a learning classifier system can be used to find combinations of alleles and environmental factors (eg, a difference at a single base pair in DNA along with a multi-pack-year history of smoking) that can predict those at higher risk of bladder cancer. These articles describe different ways to get to personalized risk prediction and prevention strategies that focus on different types of cancer. Other successful informatics approaches to study cancer are presented by Kim (*see page 637*), Chen (*see page 659*), and Warner (*see page 696*). Lussier (*see page 619*) and Moore (*see page 630*) also describe generalizable models that use data from genome-wide association study (GWAS) to study gene interactions, and Alexov (*see page 643*) proposes a novel method to understand disease-causing mutations. Additionally, this issue also presents translational bioinformatics articles related to prognostication.

Once a patient has a disease or condition, the stage and severity of the condition often provides the personalized context for that disease. Taking advantage of the novel public resource of cancer molecular measurements and images contained with The Cancer Genome Atlas (TCGA), Huang (*see page 680*) shows how histopathological characteristics can be learned and then used to estimate or impute genetic changes within cells. Some of these characteristics can predict survival with triple negative breast cancer, a particular form of cancer for which there are currently not many targeted therapeutics. Shin (*see page 613*) applies semi-supervised learning on data from the National Cancer Institute Surveillance, Epidemiology and End Results (SEER) Program to learn clinical features that can help identify breast cancer patients who are likely to have a favorable outcome. It is illuminating to see how these two groups were able to use different public data resources to implement their prognostic solutions.

After diagnosis, pharmacological therapy is often the next step, and this step has an urgent need of personalized solutions. For the earliest stages of drug development, Sarkar (*see page 668*) analyzes data from ClinicalTrials.gov to suggest that, since many drugs in trials are actually plant-based derivatives, newer drugs could also be

found from natural products. To match drugs with specific therapeutic conditions, Haibe-Kains (*see page 597*) demonstrates an interesting use of machine learning using drug treatment data from the Cancer Cell Line Encyclopedia and the Cancer Genome Project to discover gene-specific therapeutics that can be prescribed based on the particular molecular pattern of a patient's cancer. Instead of molecular-based predictors, Mani (*see page 688*) shows how MRI images can be used to drive decision support for neoadjuvant chemotherapies in breast cancer.

The informatics community has been studying new ways to provide Electronic Health Records (EHRs)-based decision support for a long time, and this issue also focuses on human factors that impact the implementation and utilization of these systems, a key next step in personalized medicine. Embi (*see page 718*) describes how clinicians and administrators utilize these systems for documentation, Singh (*see page 727*) focuses on how primary care clinicians manage EHR-based test results, and Holup (*see page 787*) studies EHR utilization in residential care facilities. Deutsch (*see page 700*) describes best practices in documenting transgender status in EHRs, and Lee (*see page 778*) proposes a method to predict complications in interventional cardiology using electronic data. Three articles address the issue of EHR costs: Lussier (*see page 708*) presents a critical appraisal of the costs involved in transitioning to ICD-10-CM, which is currently underway in a large number of institutions in the USA and highly motivated by government-specified EHR meaningful use criteria. Bassi (*see page 792*) reviews the literature on the economic impact of EHR systems, and Driessen (*see page 743*) models return on investment for an EHR system implemented in Malawi. Bass (*see page 736*) describes how information technology fills information needs of house staff, and Wu (*see page 766*) discusses unintended consequences of communication systems in five teaching hospitals. Informatics systems are also taking important steps towards direct use by patients: Veinot (*see page 758*) and Li (*see page 704*) present user-requirements for an informatics intervention addressing sexual health and alert patients about privacy policies of health social network sites, respectively.

This issue displays a board array of informatics research and applications. We present high-level experiences in implementing EHR and decision support systems all the way to personalized models that use molecular information. As the articles in this issue exemplify, extending the scope of informatics and developing new ways to conceptualize health and disease requires exploration of deeper levels of disease taxonomies. This is similar to studying trees at a new level. While many can recognize the beauty in trees, whether tall or short, there is an entirely different beauty that is revealed when one starts studying the leaves. Imagine we were just recognizing the beauty and variance in leaves for the first time, after being so

familiar with trees. We would learn that the study of leaves can help us identify much of what the tree should look like, and the many different ways in which leaves differ from the tree trunk. Perhaps it is this same sort of beauty we should be searching for in personalized medicine. ICD-9, ICD-10, and other taxonomies will continue to form a strong structure for the organization of health and disease concepts. But only by seeing and learning from the variance in patients and across their conditions and afflictions we will be able to design and implement new personalized systems to keep these patients healthy. This new perspective will enable us to see, understand, appreciate, and treat each patient in his or her own individual way.

**Competing interests** None.

**Provenance and peer review** Commissioned; internally peer reviewed.

### REFERENCES

- 1 Butte AJ. Translational bioinformatics: coming of age. *J Am Med Inform Assoc* 2008;15:709–14.
- 2 Sarkar IN, Butte AJ, Lussier YA, *et al.* Translational bioinformatics: linking knowledge across biological and clinical realms. *J Am Med Inf Assoc* 2011;18:354–7.
- 3 Linnæe Cv, Schröder J. *Genera morborum*. Upsaliæ: Apud C. E. Steinert, 1763.
- 4 National Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease. *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press, 2011.