Yuri Quintana, PhD, delivered a webinar in AllerGen's Webinars for Research Success *series on February 27, 2018, discussing different approaches to biomedical informatics and innovations in big-data platforms for biomedical research. His main messages and a hyperlinked index to his presentation follow.*

WHAT IS BIOINFORMATICS?

Bioinformatics is an interdisciplinary field that develops analytical methods and software tools for understanding clinical and biological data. It combines elements from many fields, including basic sciences, biology, computer science, mathematics and engineering, among others.

WHY DO WE NEED BIOINFORMATICS?

Chronic diseases are rapidly expanding all over the world, and associated healthcare costs are increasing at an astronomical rate. We need to develop personalized treatments tailored to the genetics of increasingly diverse patient populations, to clinical and family histories, and to environmental factors. This requires collecting vast amounts of data, integrating it, and making it accessible and usable.

CHALLENGES IN BIOINFORMATICS

Data collection, coordination and archiving: Many sources of biomedical data—hospitals, research centres and universities—do not have complete data for any particular disease, due in part to patient numbers, but also to the difficulties of data collection, inter-operability and sharing. Many centres do not collect environmental or family histories. Linking clinical data with genomic data is challenging; doing so across institutions and countries is even harder, and there is no convenient place to archive—and make available for reuse—data from large clinical studies.

Data classification and codification: There are too many ways to classify and organize data. There are too many standards, and within any given standard, there are different ways of implementing the codification of data.

Data management and integration: A major impetus for the further development of bioinformatics has been the advancement of technologies to analyze the human genome. These technologies have evolved to the point that we can now analyze not only at the DNA level, but down to the level of proteins. DNA sequencers, DNA microarrays, and mass spectrometers are generating tremendous volumes of data, and each year brings a new generation of tools producing more micro-granular data. The challenge is not only to manage this data, but to integrate data that has been analyzed to different degrees of complexity using different generations of technology.

User interfaces: Many data management products, commercial and institutional, never reach their potential because they do not dedicate sufficient resources to ensuring the tool's usability: users can not easily navigate the interfaces. Such systems may be very powerful in terms of functionality, but difficult to use. Thus, developing effective user interfaces is a niche area where work is needed.

BUILDING A BIOINFORMATICS INITIATIVE

Convene the right team: The single most important factor in building a successful bioinformatics initiative is bringing together the right combination of people. Many complex skills are needed, so you need very highly trained, multidisciplinary personnel, including: chief scientists; a research informatics director who understands how the information must be organized; someone clinical that understands data integration from the clinical side; a biomedical platforms architect that can bring these two sides—the life sciences and the clinical side—together; programmers and engineers, including people with a deep understanding of taxonomies and classification; a team for assuring data quality; experts in cybersecurity and cloud computing; data scientists with machine learning expertise to create good user interfaces and analytical tools; and good ethics and privacy directors: people who communicate internally but also with external stakeholders to build consensus and trust around how the data are to be used, archived and secured.

Institute good governance: Strong leadership by keen individuals who understand the global landscape and challenges involved is required for creating a cohesive environment in which all the different disciplines involved are respected.

Involve patients: It is vital to involve patients as true partners from the beginning. Their insights and concerns should inform research design and help generate research questions. It is crucial to ensure that they feel respected and are comfortable with the plans for how data is to be collected and used.

Agree on standards: To enable the merging of data, agreement is needed among researchers and research centres on data formats, standards and coding—including the taxonomies and classifications to be used.

Ensure transparency: Internal and external transparency is fundamental, regarding data use and how benefits are to be shared. If you plan to partner with pharma or other groups, this must be accounted for from the outset, in everything from the design of consents to plans for data storage and retrieval.

Start focused, then expand: Bioinformatics initiatives that endeavour to do too much too quickly often end up “feature rich” but “function poor.” It is best for such initiatives to have an initial focus that is feasible and that allows for a certain level of short-term success, but that can later be scaled up, in terms of both data and research questions.

RESOURCES

Tools to advance the sharing of genomic data:

- **i2b2:** Informatics for Integrating Biology & the Bedside
- **tranSMART:** An Open Source Knowledge Management & High Content Data Analytics Platform

Large datasets available for sharing and advancing research:

- **NCI Data Commons:** A unified repository enabling data sharing across cancer genomic studies to support precision medicine
- **NIH Big Data to Knowledge** repository

INDEX OF WEBINAR CONTENT

Available for this webinar: slides (in PDF) | video recording

***Yuri Quintana**, PhD, is Director for Global Health Informatics in the Division of Clinical Informatics, Beth Israel Deaconess Medical Center, and Assistant Professor of Medicine, Harvard Medical School, Harvard University. His research is focused on developing innovative technologies that empower patients, families and health professionals.*