Translational Bioinformatics and Clinical Research (Biomedical) Informatics

S. Joseph Sirintrapun, MD[a,*], Ahmet Zehir, PhD[b], Aijazuddin Syed, MS[b], JianJiong Gao, PhD[b], Nikolaus Schultz, PhD[b], Donavan T. Cheng, PhD[b]

KEYWORDS

• Translational informatics • Bioinformatics • Clinical research informatics • Biomedical informatics • The Cancer Genome Atlas • TCGA • cBioPortal • Cancer genomics

ABSTRACT

Translational bioinformatics and clinical research (biomedical) informatics are the primary domains related to informatics activities that support translational research. Translational bioinformatics focuses on computational techniques in genetics, molecular biology, and systems biology. Clinical research (biomedical) informatics involves the use of informatics in discovery and management of new knowledge relating to health and disease. This article details 3 projects that are hybrid applications of translational bioinformatics and clinical research (biomedical) informatics: The Cancer Genome Atlas, the cBioPortal for Cancer Genomics, and the Memorial Sloan Kettering Cancer Center clinical variants and results database, all designed to facilitate insights into cancer biology and clinical/therapeutic correlations.

OVERVIEW OF TRANSLATIONAL BIOINFORMATICS AND CLINICAL RESEARCH (BIOMEDICAL) INFORMATICS

Translational bioinformatics and clinical research (biomedical) informatics are the primary domains related to informatics activities that support translational research. Although arguably distinct, clinical research (biomedical) informatics and translational bioinformatics are often used interchangeably. Translational bioinformatics focuses more specifically on the computational techniques in the areas of genetics, molecular biology, and systems biology.[1,2] By contrast, clinical research (biomedical) informatics involves the use of informatics in the discovery and management of new knowledge relating to health and disease.

Clinical research (biomedical) informatics uses computational techniques related to secondary research use of clinical information for understanding disease processes. These computational techniques span a wide set of interdisciplinary fields and encompass resources, devices, and methods that optimize the acquisition, storage, retrieval, transformation, and communication of clinical information.[1,2]

Driving both translational bioinformatics and clinical research (biomedical) informatics is the management and refinement of data: how data are captured, transmitted, processed, and conveyed into information in order to generate meaningful knowledge. How data are captured for translational bioinformatics begins after tissue acquisition and tissue processing, and uses advanced molecular techniques for data generation. How data are captured for clinical research (biomedical) informatics starts with data compiled from health information systems (discussed in an article elsewhere in this issue).

Disclosures: None.

[a] Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA; [b] Memorial Sloan Kettering Cancer Center, 417 East 68th Street, New York, NY 10065, USA
* Corresponding author.
E-mail address: sirintrs@mskcc.org

Surgical Pathology 8 (2015) 269–288
http://dx.doi.org/10.1016/j.path.2015.02.015
1875-9181/15/$ – see front matter

One application of clinical research (biomedical) informatics is managing information related to clinical trials. Another application is linking large-scale DNA data banks with electronic medical record systems for discovery of genotype-phenotype associations.[3] Informatics of biospecimens and biorepositories also falls under the scope of clinical research (biomedical) informatics and is discussed briefly.

With biospecimens and biorepositories, there are immense infrastructural needs from information technology. Biospecimens and biorepositories must have associated quality clinical and pathology information with the specimens, which means efforts to determine which data elements to capture and easy mechanisms to associate and annotate samples. Optimal information systems can update whether studies have institutional review board approval using samples and associated clinical data elements. Moreover, there should be security maintenance and processes in place for de-identification of protected health information. Tools for de-identification could include an honest broker system, which maintains linkages between samples and clinical data elements through a third-party mediator.

Information systems for biospecimens and biorepositories should encompass operational logistics, such as inventory tracking, sample processing, storage, and distribution management. Sophisticated information systems have barcoding systems to facilitate such operational logistics. Crucial are functionalities to document how specimens are acquired and collected. Other functionalities include refrigeration and location, specimen distribution, and usage and control user accessibility. Biospecimens and biorepositories are costly investments and there are pressures for such information systems to enable cost recovery measures.[4]

Creating an optimal information systems infrastructure for biospecimens and biorepositories has proved daunting. The cancer Biomedical Informatics Grid (caBIG) initiative began in 2004 to create an interoperable academic/commercial biomedical information system, built on community-driven, precompetitive open source standards for data exchange and interoperability in the cancer research enterprise. This initiative held hopes for widespread dissemination throughout the cancer community. The guiding principles of caBIG of open access, open development, and open source were appealing. The ideal vision for caBIG was to make large and diverse cancer research data sets sustainably available for analysis, integration, and mining. In doing so, caBIG would become the platform by which cancer researchers would access data and biospecimens across institutions to perform genomic analysis and to find and analyze clinical data. The caBIG initiative never achieved its ideal vision for multiple lengthy reasons which will not be discussed and, sadly, the caBIG initiative was retired.[5]

ILLUSTRATIVE EXAMPLES OF TRANSLATIONAL BIOINFORMATICS AND CLINICAL RESEARCH (BIOMEDICAL) INFORMATICS

This article details 3 projects that are hybrid applications of translational bioinformatics and clinical research (biomedical) informatics. The first is TCGA, the second is the cBioPortal for Cancer Genomics, the third is the MSKCC CVR system database; all were designed to facilitate insights into cancer biology and clinical/therapeutic correlations.

PART 1. THE CANCER GENOME ATLAS

TCGA is a comprehensive and coordinated multi-institutional effort to create a detailed catalog, or atlas, of genetic mutations in cancer using advanced genome sequencing and translational bioinformatics associated with specific types of tumors to improve the prevention, diagnosis, and treatment of cancer. Its mission was to accelerate the understanding of the molecular basis of cancer through the application of genome analysis and characterization technologies.

TCGA began in 2006 as a pilot project funded by the National Cancer Institute (NCI) and National Human Genome Research Institute, both parts of the National Institutes of Health. Initially, TCGA focused on characterization of only 3 types of cancers but since has grown to at least 30 tumor types and many more subtypes.[6,7] The cancers were selected by TCGA because of their poor prognosis and overall public health impact. The power of the project is the quality of tissue acquisition. TCGA samples are consistent in their processing with extensive quality controls (QCs) in place. TCGA research network encompasses centers for genome characterization, protein characterization, and genomic data analysis centers, which enable the process for genomic discovery. TCGA network comprises scientists, bioinformaticians, bioethicists, doctors, nurses, and cancer advocates. The data generated includes gene expression, protein expression, DNA copy number alterations (CNAs), epigenomics (noninherited DNA modifications), and microRNAs (miRNAs), which are short RNAs that control gene expression. Genome sequencing centers perform exome (coding gene region) sequencing on all cases, with some cases