The Role of Information Systems in Clinical and Translational Research (Frontiers: The NIH Clinical and Translational Science Award Driving Information Systems)

Paul F. Terranova[1], Richard J. Barohn[2], Lauren S. Aaronson[3], Andrew K. Godwin[4], Peter Smith[5], and L. Russ Waitman[6], University of Kansas Medical Center

Elias Zerhouni, former NIH Director, charged the NIH with building a biomedical research enterprise that enhanced the translation of scientific discoveries into improved health care for our nation (1). One component of that charge was to establish the NIH Clinical and Translational Science Award (CTSA) program, which currently is located in the National Center for Advancing Translational Sciences (http://www.ncats.nih.gov/research/cts/ctsa/ctsa.html. The goals of the CTSA program are 1) to improve the manner by which medical research is performed in the US, 2) to move discoveries in the lab more quickly into the clinic, 3) to integrate clinical research into communities, and 4) to train the next generation of medical researchers in the art of clinical and translational research. Approximately 60 institutions in 30 states including the District of Columbia and Kansas are part of the CTSA consortium working toward these goals.

Headquartered at the University of Kansas Medical Center, the NIH-CTSA-supported Frontiers program, more formally called The Heartland Institute for Clinical and Translational Research, is a network of scientists from institutions in Kansas and the Kansas City region (www.FrontiersResearch.org). The Frontiers program is developing more efficient use and integration of biorepositories, genomic information and biomedical informatics which are important components for attaining the CTSA goals. Specific goals of Frontiers are to: 1) create a new academic home with innovative education and training programs for clinical and translational investigators that will transform the type, and increase the number of faculty and investigators needed to bring discoveries and research findings more rapidly to the point of care; 2) provide an enhanced coordinated translational research infrastructure that will speed the process from discovery to community research in Kansas and the Kansas City region; and 3) actively engage the community in developing, testing, and disseminating translational research through existing and new networks in Kansas and the Kansas City region.

[1] Vice Chancellor for Research; [2] Director, Frontiers; [3] Deputy Director, Frontiers; [4] Professor & Director, Molecular Oncology; [5] Director, Institute for Neurological Discoveries, Co-Director, KIDDRC; [6] Director, Medical Informatics.

The infrastructure needed (Frontiers goal #2) to speed the process of discovery, includes, in part, fundamental components such as bio-repository, genomics, and biomedical informatics. Each of these promotes and enhances discovery and each is foundational for moving discoveries from the bench to the bedside and becomes the foundation for this presentation. Each of these is described in order to provide an overview of how each is a component of translational research.

Frontiers: The Heartland Institute for Clinical and Translational Research

Many partners are involved with accomplishing the goals of our Frontiers program. These partners include academic institutions (University of Kansas Medical Center in Kansas City, University of Kansas School of Medicine in Wichita, University of Kansas at Lawrence, University of Missouri at Kansas City, and the Kansas City University of Medicine and Biosciences) and health care institutions and centers in the region (University of Kansas Hospital, Wesley Medical Center in Wichita, Via Christi Health System in Wichita, Kansas City, VA Medical Center, Children's Mercy Hospitals and Clinics, St. Luke's Health System, Truman Medical Center, Swope Health Services, Center for Behavioral Medicine, and Center for Practical Bioethics. The basic infrastructure of Frontiers includes several components described below.

1. Clinical Research Development Office. This is the literal and virtual (via a web portal) front door where investigators go to access information about services and resources. Information on the investigators research interests are logged so they may be contacted by Frontiers faculty regarding special initiatives and for collaboration. Thus, this office brings together the scientific expertise and other resources to support faculty projects, ensuring appropriate multidisciplinary contribution and inclusion of relevant components of the CTSA such as biostatistics, biomedical informatics, community engagement and ethics programs.
2. Clinical and Translational Research Education Center. This is the site of educational and career development programs that use teaching and learning technologies and faculty/staff to educate students and faculty at many levels of development. This Center provides innovative programs for mentoring skills and development, formal training in the entrepreneurial skills necessary to bring new discoveries to the market, and a program for minority recruitment and retention. This Center also includes a NIH TL1 pre-doctoral program and a NIH KL2 post-doctoral program. In addition to formal training in research methods, scholars and trainees take advantage of a mentorship program and become integrated into multidisciplinary research teams.
3. Biomedical Informatics. This program provides information resources, services, and communication technologies in support of translational research. Russ Waitman, PhD, associate professor and director of Medical Informatics in the Depart-

ment of Biostatistics leads this program and is providing a chapter on this topic in this Merrill conference (2012). The program has several goals: a) to provide a portal for investigators to access clinical and translational research resources, track usage and outcomes, and provide informatics consultation services; b) to create a repository to integrate clinical and biomedical data for translational research aligned with national standards; c) to advance medical innovation by linking biological tissues to clinical phenotype and research laboratory results for phase I and II clinical trials; and d) to leverage an active, engaged statewide telemedicine and Health Information Exchange to enable community based translational research.

4. Biostatistics. This program, offered through the Department of Biostatistics, provides the statistical and data management support for Frontiers studies.
5. Clinical and Translational Science Unit (CTSU). This unit is a new state-of-the-art facility with capability to conduct first-in-human proof of concept trials as well as traditional clinical research activities within the unit and in clinics and the community (using our "CTSU without Walls"). This program also developed a clinical research participant identification and registration system that our Biomedical Informatics program is further developing to provide easy cohort identification and contact information on patients willing to consider participating in research and to incorporate national developments in this area.
6. The Institute for Advancing Medical Innovations. This program provides infrastructure for drug and device development that lead to clinical applications.
7. Translational Technologies Resource Center. This program provides access to specific technologies through four cores: In vivo Imaging, Cell and Tissue Imaging, Molecular Biomarkers, and a Biological Tissue Repository. The overall goal of this program is to eliminate barriers to incorporating new technologies and research approaches into research studies for both new and seasoned investigators by providing access to equipment and expert faculty collaborators in these technologies.
8. Pharmacokinetics/Pharmacodynamics (PK/PD). This program provides PK/PD support to investigators who prefer to focus on other aspects of their investigator-initiated trials. The program also supports clinical and translational research by training investigators who wish to develop such skills to become competent in designing, analyzing and interpreting PK/PD data obtained from their investigator-initiated trials, and by providing a training ground for graduate and post-doctoral students on PK/PD.
9. Personalized Medicine and Outcomes Center. This program focuses on T2 research innovations and support and has five cores: Analysis of Large Databases; Quality Assessment/Quality Improvement Analysis;

Economics, Health Status & Decision Analysis; Survey Design and Qualitative Research Support; and Outcomes Education.

10. Pilot and Collaborative Studies Funding. This program serves as the centralized Frontiers resource and catalyst for soliciting, reviewing, awarding and tracking clinical and translational research pilot and feasibility studies.
11. Community Partnership for Health. This program supports and assists investigators with accessing diverse populations for participation in clinical and community based research projects—historically one of the larger barriers for many clinical investigators. It ensures investigators are skilled and knowledgeable about working in and with diverse communities through a novel training and certification program on community engagement for research. This program is provided online and will be coordinated with all other annual investigator certifications (e.g., on Human Subjects and Biosafety) required for maintaining Institutional Review Board approvals.
12. Regulatory Knowledge and Support. This program supports investigators to ensure their clinical and translational research is in compliance with ethical and regulatory standards.
13. Ethics. This program provides individual ethics consultations to Frontiers investigators and formal research ethics education and training programs.

Bio-repository

The bio-repository includes numerous components requiring integration of multiple sources of information flow with a goal of enhancing discovery to improve health. The bio-sample repository has several major functions including recruitment, sample acquisition, processing, storage, distribution and analysis. The repository also includes tissues, bodily fluids and cell lines. The process usually begins with recruitment and informed consent of the patient in order to collect bodily tissue and/or fluids (e.g., serum or plasma). Informed consent may occur during pre-admission testing, an outpatient visit or upon admission to the hospital. Sample collection may occur in the hospital, operating room and/or outpatient clinic. Samples are proactively acquired for future testing and may eventually be sent to multiple recipients. Information on personal and family histories, clinical intervention, and lifestyle factors is collected for each sample as well as: demographic, family history, medical history, epidemiologic risk factors, and clinical history. Annotated, high quality samples are collected and stored. Specific procedures are followed for the tissue banking scheme. For example, surgical specimens are collected by the pathologist with the aid of their physician assistant. Select tissue is identified for sampling and this information is entered into a database. A sample is then snap frozen and kept in liquid nitrogen or prepared for frozen sectioning. Samples frozen in liquid nitrogen may be stored and/or prepared for DNA and RNA extraction. Another sample of the same tissue is placed in fixative for sub-

sequent histopathological analysis by storing in paraffin blocks for subsequent preparation of histological sections including routing staining, immunostaining, or tissue microarrays. Blood is separated into serum or plasma, red blood cells, leukocytes, DNA and stored at -80C. Blood samples are banked within 4 hours of draw after being barcoded. Sample quality is critical to the success of the bio-repository. Quality measures taken include compliance with best practices, qualification for membership in the International Society for Biological and Environmental Repositories, insisting on standard operating procedures, and certification by the College of American Pathologists (and others). NIH supported projects and others will/may require appropriate mechanisms for data sharing. Thus, patient consent, sample processing and inventory, questionnaire data, medical records and tumor registry data (when appropriate) must all be stored in databases, managed, and appropriate samples identified when requests are made for procurement. Bio-repository management is complex with many components requiring oversight such as equipment, supplies, dedicated resources, liability, data management, standard policies and procedures, quality control, and compliance. Access to bio-samples is regulated by an advisory committee which considers requests from within and outside of the medical center.

Our goal is to enter genome and other types of 'omic' data such as metabolome, transcriptome and proteome into the system. Thus, sequencing data, genotype, methylation, expression, single nucleotide polymorphism data, chromatin remodeling, etc. are collected and entered. The samples are annotated and the data placed in the Healthcare Enterprise Repository for Ontological Narration (HERON—see Waitman presentation in this Merrill series, 2012). HERON also integrates with billing records, medical records, and the KU Hospital Tumor Registry along with all of the 'omic' and other types of data. The Bio-sample Repository supports several programs including, Frontiers, Early Detection Research Network, Alzheimer's Disease Center, Clinical Trials and Protocol Support Lab, the Consortium of Investigators of Modifiers of BRCA1/2, Pharmaceutical and Biotech collaborations, the Triple Negative Breast Cancer Network, faculty collaborations including investigator initiated studies, and the Gynecological Oncology working group.

Compliance is important in the bio-repository and includes federal and local components such as Health Insurance Portability and Accountability Act and Institutional Review Board of protocols.

Bioinformatics Services

Bioinformatics services (http://www2.kumc.edu/siddrc/bioinformatics) at KUMC are provided largely by the Smith Intellectual and Developmental Disabilities Research Center (SIDDRC; http://www2.kumc.edu/siddrc/) which is a collaborative effort with KU at Lawrence (http://kiddrc.kumc.edu/). The Center is supported by a NIH grant (HD002528), which has been held for 47 consecutive years and the grant will continue at least into its 50th year. Additional support for bioinformatics comes from NIH including the Center for Biomedical

Research Excellence in Cell and Developmental Biology (RR024214) and the Kansas Idea Network of Biomedical Research Excellence (GM103418, RR016475). Bioinformatics studies and utilizes methods for storing, retrieving and analyzing biological data, such as DNA, RNA, proteins and their sequence as well as their structure, function, pathways and interactions. The overall mission of the core facility is to advance the understanding of integrative functions in biological systems, including human, through the application of computational models and data analysis with focus on Next Generation Sequencing Data Analysis and Microarray Data Analysis. The Bioinformatics core provides consulting, data analysis solutions and analysis of large-scale biological datasets produced by high-throughput genomics experiments. Thus, the core identifies opportunities and implements solutions for managing, visualizing, analyzing, and interpreting genomic data, including studies of gene expression (RNA-seq and microarrays), pathway analysis, protein-DNA binding (e.g. ChIP-seq), DNA methylation, and DNA variation, using high-throughput platforms in both human and model organisms.

A variety of services in bioinformatics and computational biology are provided by the Bioinformatics Core. For example:

High-throughput sequencing

- **RNA-Seq:** provides an unbiased deep coverage and base level resolution of the whole transcriptome.
- **Chip-Seq:** combines chromatin immunoprecipitation with high-throughput sequencing to provide an unbiased whole genome mapping of the binding sites of DNA-associated proteins.
- **Whole Genome Sequencing:** sequences the whole DNA sequence of an organism's genome.
- **De novo Sequencing:** provides the primary genetic sequence of an organism.
- **Metagenomic Sequencing:** sequencing and identifying the genomes of whole microbial communities.
- **Methyl-Seq:** analysis of methylation patterns on a genome wide scale.

Microarray analysis

- **Affymetrix 3′ Expression Arrays:** target the 3′ end of genes.
- **Affymetrix Exon Arrays:** provides expression levels for every known exon in the genome.
- **Affymetrix miRMA Arrays:** provides measurements of small non-coding RNA transcripts involved in gene regulation.
- **Affymetrix Genome-Wide Human SNP Array:** copy number analysis
- **Affymetrix GeneChip Tiling Arrays:** gene regulation analysis

Biological Functional and Pathway Analysis: software from Ingenuity Systems analyzes expression data to ascertain the top biological functions and pathways associated with them. **Biological Literature Survey:** software from Acumenta (Literature Lab) that performs data mining tasks on experimentally derived gene lists. **miRNA Target Prediction:** in house software and open source software such as TargetScan, miRanda for detecting genomic targets for miRNAs.

Transcription Factor Binding Site Prediction: in house software and open source software such as MEME, Homer, PGMotifScan to identify protein DNA interaction sites.

Medical Informatics

The KUMC Medical informatics program is discussed in more detail in another chapter of this Merrill document (2012). The work of the medical informatics division (within the Department of Biostatistics in the School of Medicine) includes the following:

- HERON (http://informatics.kumc.edu/work/wiki/HERON) healthcare enterprise repository
- CRIS (http://biostatistics.kumc.edu/bio_cris.shtml) Comprehensive Research Information System powered by Velos
- REDCap (http://informatics.kumc.edu/work/search?q=redcap) for REDCap at KUMC and (http://project-redcap.org/) for REDCap at the CTSA consortium.
- RequestTracking (http://informatics.kumc.edu/work/wiki/RequestTracking) place for notes on using REDCap for request tracking (CTSA, ADC, HERON)
- EnterpriseAnalytics (http://www2.kumc.edu/aa/ir/)
  - TEAL data warehouse (http://informatics.kumc.edu/work/wiki/TEAL)
- NetworkInnovation (http://informatics.kumc.edu/work/wiki/NetworkInnovation)
- Other informatics efforts at the University of Kansas not mentioned in the article above include:
  - The Center for Health Informatics (http://www2.kumc.edu/healthinformatics/)
  - The Center for Bioinformatics (http://www.bioinformatics.ku.edu)
  - The K-INBRE Bioinformatics Core (http://www.kumc.edu/kinbre/bioinformatics.html)
  - Bioinformatics Computational Biology Laboratory (http://www.ittc.ku.edu/bioinformatics/)
  - Biodiversity Informatics (https://journals.ku.edu/index.php/jbi)

Reference:

Zerhouni EA Translational and clinical science–time for a new vision. NEJM 2005, 353:1621-23.