

Signature Center for Computational Diagnostics

Translating Experimental Technologies into Clinical Research

Susanne Ragg, MD, PhD^{abc} Gunther Schadow, MD, PhD^{bcd},

^aDepartment of Pediatrics, ^b Indiana University School of Medicine, ^c Regenstrief Institute, ^d Indiana University School of Informatics

Abstract

The goal of the Center for Computational Diagnostics [<http://www.iupui.edu/~compdiag>] is to serve as a “reactor” of innovative research for the integration of diverse high throughput technology into clinical trials to allow clinical researchers to obtain a more comprehensive view of the disease states. It has become clear that to better understand diseases we cannot continue to only focus on single genes, proteins or metabolites operating in single linear ordered pathways. Large scale high throughput technologies such as applied in genomics, proteomics and metabolomics allow for a more comprehensive view of the complex interactions occurring within body fluids or tissues at any one time. The Center operates by addressing the three different areas that are required to successfully integrate high throughput methodologies into translational research:

1) High throughput biospecimen banking; 2) Generation of high quality datasets and 3) Workflow development for data storage and analysis.

To support *high throughput bio-specimen banking* we have co-developed caTissue Suite under the national caBIG effort. The key developments of the Center have been:

1. CaTrack, an intelligent barcode-based automatic data capture system
2. protocol-driven Study Calendar
3. xCaCORE, an innovative XML-based data import and export software program
4. Scalable, globally unique specimen identification utilizing an ISO Object Identifiers encoding scheme
5. Barcode generator and label printing

We have implemented the Pediatric Biospecimen Repository to be able to develop and test these informatics tools. In addition to functioning as the development and test site for informatics research, the repository also develops research protocols and stores biospecimens for Pediatrics, Ophthalmology and Obstetrics. We also support protocol and data management related to biospecimens for the CTSI and the Fairbanks Institute. We currently have 92 active protocols and data on 119,319 specimens in our production instance of caTissue Suite. This includes specimens from 1200 well-defined healthy control subjects across all age groups including 600 children. To develop the computational and statistical workflows for data storage and analysis, we have generated large well-designed datasets for coronary artery disease (LC-MS/MS, NMR, protein antibody arrays); cancer (osteosarcoma and Wilms tumor; LC-MS/MS, NMR, protein antibody arrays) and ophthalmologic diseases (glaucoma; protein antibody arrays). Our main focus is to develop analytical workflows that translate the large datasets into relevant information for clinical researchers, focusing on the biological interpretation of the results. In this context we developed statistical models for protein quantification for LC-MS/MS and protein antibody arrays. These workflows were implemented in the open source statistical software R and published under the R-based project Bioconductor.