

Informatics for Integrating Biology and the Bedside (i2b2): A Translational Engine at the National Scale

RESEARCH IMPACT: Since its inception in 2004 i2b2 has been designed to provide the instrumentation for *using the informational byproducts of health care and the biological materials accumulated through the delivery of health care to* – and as a complement to prospective cohort studies and trials - *conduct discovery research and to study the healthcare system in vivo*. The utility of this approach is demonstrated by the grass-roots adoption of the i2b2 platform by over 84 academic health centers (AHCs) internationally, each implementation of which represents a major, local institutional commitment. **IMPACT EXAMPLE 1. Genomic Disease Studies.** As presented in our recent Nature Genetics Review (1), the field of Electronic Health Record (EHR) Driven Genomic Research (EDGR) has come into its own. We have made significant contributions with our validations of findings made in other studies in a broad array of phenotypes (e.g. RA, MDD, Asthma, IBD (2-6)). In all studies the directionality of the odds ratios of SNPs reproduced with magnitudes within 95% confidence limits, all at least 1-2 orders of magnitude *faster and cheaper*. We were furthermore able to measure the effect size of SNPs in minority populations due to overrepresentation in AHC EHRs. **IMPACT EXAMPLE 2. Pharmacovigilance (as a public health application).** Our team has successively used EHR mining to confirm the association of increased MI mortality with Vioxx use (7), the elevated MI risk with Avandia usage (contributing to the FDA “black box” warning) (8), and in collaboration with SIMBIOS and Vanderbilt to rapidly confirm an FDA alert regarding increases in blood glucose in patients taking both paroxetine and pravastatin (9). **Publications.** The core i2b2 team has produced 185 peer reviewed papers exclusive of 121 publications directly resulting from our Natural Language Processing (NLP) Challenges and including over a dozen in journals with impact factors of 20 and higher. Several of them are the first of their kind in demonstrating direct utility of EHR data.

COMMUNITY RESOURCES/SOFTWARE/COLLABORATIONS: i2b2’s software platform (“i2b2”), designed to enable discovery research from existing healthcare information, has provided a major leveraging factor for population-based studies, quality care and outcomes initiatives, adverse event monitoring, and novel hypothesis-driven investigations. The freely available i2b2 toolbox is now used and is being extended by an Academic Users’ Group (AUG) now numbering over 300 members. Attendance at our Annual Users’ Group Conference, which now exceeds 125 members representing all key constituencies (CTSAs, AHCs, HMOs, Industry, Disease Networks), affirms the value of this product and the community collaboration that has developed to push it forward. i2b2 has led with the Harvard CTSA the development and deployment of a web-based network, SHRINE, that enables data sharing across i2b2 (or other) platforms (10, 11), as exemplified by the University of California’s recent deployment of SHRINE to allow the analysis of 11 million patients across their AHC’s (12). To advance the essential NLP tools necessary to crisply define phenotypes derived from clinical data, we have developed and hosted 6 International NLP Challenges based on annotated, de-id’d patient data sets that have resulted in participation by 138 international teams from 11-45 organizations, a Research Data Set available from our website that has 3,117 unique downloads from a user base of 230 academic researchers, graduate students, industry, and course developers, and 121 publications.

TRAINING & DISSEMINATION: i2b2’s Ed/Dissemination Core has prioritized the recruitment of talented undergraduate students into graduate study in the area of healthcare informatics by establishing a Summer Institute in Bioinformatics and Integrative Genomics that has graduated 94 students (8 programs), including 35 URMs. Of the 64 who have graduated college, 42 are now in MD, PhD or MD/PhD programs, including 16 URMs. We have in addition participated in the training of 30 graduate students and 33 postdocs with a stable core faculty size of 15. 20 of the postdocs now hold faculty appointments, 10 are still in training. Support to our user communities is provided by an active Community Wiki, AUG listserv, twice yearly software workshops, and annual AUG Conferences, NLP Workshops, NLP Challenges, and SHRINE National Conferences. www.i2b2.org.