Patient-Centered Information Commons

Harvard Medical School

PI: Isaac S. Kohane

Grant Number: 1-U54HG007963-01

We propose to create a massively scalable toolkit to enable large, multi-center Patient-centered Information Commons (PIC) at local, regional and, national scale, where the focus is the alignment of all available biomedical data per individual. Such a Commons is a prerequisite for conducting the large-N, Big Data, longitudinal studies essential for understanding causation in the Precision Medicine (1) framework while simultaneously addressing key complexities of Patient Centric Outcome Research studies required under ACA (Affordable Care Act). Our proposal is solidly grounded in our experience over the last 25 in harnessing clinical care data to the research enterprise. In creating PIC we propose to focus on: 1. Enable the identification and retrieval of all data that pertain to individual health by creating a data sharing architecture that is capacious enough for all relevant data types and that enables patient and institutional autonomy to be respected. 2. Test fully-scaled implementations of the proposed architecture early in the development process, with the active involvement of a committed user community that seeks to use allowed us to refine our designs to facilitate subsequent robust dissemination and adoption. 3. Provide commodity workflows that can be used to 'clean' and complete the often noisy and sparse data gathered in the course of observational studies. 4. Embrace decentralization while enabling the construction of a nationally or regionally-scaled patient-centered information commons. 5. Encourage the selection of standards through the tools that enable the construction of patient-centered information commons. 6. Employ diagnostic classification and prognostication as figures of merit to measure how well a patient-centered information commons adds the understanding of patient populations. In addition to the research and development agenda we have also taken on the development of educational opportunities for end user community to become more familiar with the methods and challenges of data science. PUBLIC HEALTH RELEVANCE: Large populations of individuals characterized by many different and complementary types of data, for example genetic, environmental, imaging, behavioral and clinical findings will allow significant progress in our ability to accurate classify individuals as to their disease or disease risk and provide more precise predictions of their disease course. The proposed toolkit enables such characterization at the local and national scale.