Additional Priority Idea

Precision Medicine

How can we exploit clinical and genomic “big data” to predict and individually improve each person’s health trajectory?

Medicine is entering a new era in which the acquisition and interpretation of vast quantities of data from human populations is not only beginning to enable individually-tailored clinical care, but is also opening previously inaccessible avenues for understanding the biology of health and disease. Electronic medical records give access to high quality phenotypic data, including medical history, family history, environmental exposures, and medication use. DNA sequence data give access to the underlying genetic information that encodes and influences phenotypes. This integration of clinical and genetic information promises to transform our understanding of human biology, and to boost discovery for diagnosis, individualized treatment, predictive medicine, and disease prevention.

Historically, personalized medicine began with blood transfusion, involving the assignment and matching of blood groups between donors and recipient patients. Now, the most visible recent development is individualized cancer therapy. For example, if a non-small cell lung cancer tumor is positive for the PD-L1 protein, it can be treatable with specific antibodies, known as checkpoint inhibitors, which enable the immune system to eliminate the cancer. Underscoring the broader significance of this approach to medicine, in 2016 the NIH announced initial funding “to inform efforts to accelerate the understanding of individual differences that play a role in health, with the goal of informing better prevention and treatment strategies tailored for each person.” Although there has been a remarkable pace of medical progress along these lines, there remains vast unexplored territory in understanding human biology and disease. For example, of the ~20,000 genes of the human genome, genetic variants in only 57 genes are currently considered to be “medically actionable,” that is, potentially targetable by therapeutics. Therefore, the grand challenge ahead is to comprehensively identify and functionally explain all individual genetic variants that correlate with human health and disease. These discoveries will have major scientific, medical, and economic impact.

There is a need for scalable, automated, and predictive algorithms for Precision Medicine. A top priority toward linking genetic data with human biology and disease is the need for generating algorithms that will derive novel insights from the vast amounts of multidimensional data entailed in the Precision Medicine enterprise. This will require intellectual expertise and innovation in data science and full participation of experts in this rapidly evolving field. Clearly, assistance will come from Machine Learning, Artificial Intelligence, and mathematical models, and the contributions of leading investigators in those areas will be essential (See Data Science above). At the same time, patient privacy and data security will remain paramount. Thus, a growing need for collaborative teams of big data scientists will confront biologists and clinicians engaging in the Precision Medicine enterprise.

Personalized Medicine at Yale

In 2009, Yale investigators developed exome sequencing, the small subset of the genome that encodes for protein function, selectively capturing and sequencing the exomes of expressed genes at high efficiency and low cost. The utility of this technology was demonstrated by performing the first genetic whole exome diagnosis of a chronically ill infant, identifying a homozygous defect in a bicarbonate/chloride exchanger of the kidney. In other studies, family pedigrees with the highest and lowest values of blood pressure were studied by exome sequencing, identifying a variety of kidney channels and transporters implicated in renal salt processing, thus establishing salt handling as a primary element responsible for blood pressure variation.
A major investment in 2010 established the Yale Center for Genome Analysis (YCGA), one of the first CLIA/CAP-certified facilities for diagnostic DNA testing of human samples, providing state-of-the-art low-cost sequencing and data processing, and leadership in developing new technologies for genomic analysis. YCGA is a University-wide resource used by over 400 Yale investigators, and has the capacity to annually sequence 10,000 human genomes at the exome level in high-priority areas including prenatal medicine, newborn disorders (e.g. congenital heart disease), and cancer. The enormous accumulation of exome, and recently, whole genome data, places Yale scientists in an ideal position to make informative and often unanticipated observations about the genetic differences between individuals. Such valuable information provides the basis for creative biological hypotheses, allowing our investigators to ask crucial scientific questions about human health and disease.

A major asset is Yale’s close partnership with the Yale New Haven Health System, which, as part of its role in healthcare provision and clinical trials, collects a comprehensive array of electronic health records. These data are maintained in a standardized, research-compatible form via Epic, and include content that is particularly valuable for biomedical discovery. Moreover, the Health System serves a diverse population, reflecting the varied demography of Connecticut and the nation as a whole, providing insight into population-level heterogeneity in disease risk and treatment outcomes. In many cases, medical information has been collected repeatedly from the same individuals over time, providing longitudinal data. Thus, the combination of YCGA sequencing capability with the strong local medical record system, and the demography of the New Haven area, place Yale in a uniquely advantageous position to push genomic discovery in medicine, nursing and public health to new limits.

The USSC endorses the ongoing efforts in this area, and sees increasing investments in Data Science and improvements in clinical trial support, both described elsewhere in this report, as critical factors in developing our leadership in this field.