Genomic data on chronic lung disease made readily available on new website

March 22 2012

The constant focus on customer needs that drives the design of everything from automobiles to personal computers has now been applied to a field traditionally immune to such concepts: the scientific study of disease.

On a new website, the Lung Genomics Research Consortium (LGRC) -- an alliance of scientists at five U.S. institutions -- makes a broad range of genomic data on chronic lung disease available in a format specifically tailored to investigators' needs. In contrast to other genomic research websites, which typically upload massive amounts of information without much regard for how the data will be used, the new site, www.lung-genomics.org, was designed entirely with end users in mind.

"We built this portal by asking who the users will be and how they're likely to interact with the data," says John Quackenbush, PhD, director of the Center for Cancer Computational Biology at Dana-Farber Cancer Institute, which serves as the LGRC's data-coordinating center. "We considered the type of data they're likely to need and the questions they're likely to ask."

The LGRC includes researchers at Dana-Farber, the University of Pittsburgh, Boston University, National Jewish Health, and the University of Colorado. The principal investigators are David Schwartz, MD, and Mark Geraci, MD, of the University of Colorado, Quackenbush of Dana-Farber; Avrum Spira, MD, MSc, of Boston

University; and Naftali Kaminski, MD, and Frank Sciurba, MD, of the University of Pittsburgh.

The site provides access to data from a comprehensive genomic analysis of lung tissue samples from 400 patients with chronic obstructive pulmonary disease (COPD) or interstitial lung disease (ILD), as well as detailed clinical information about those patients. For each sample, the site includes genome-wide data on microarray DNA expression, microRNA expression, RNA sequencing expression, small RNA expression, SNP variation, DNA methylation, and a mix of whole-exome and whole-genome data.

The site is designed for use by two groups of researchers: quantitative scientists, on the one hand, and basic and clinical investigators on the other.

For quantitative scientists, the site makes it possible to aggregate genomic data into useful subsets. "Even high-volume users will want to 'shop' for the data most relevant to their research questions," says Kaminski.

For basic scientists, the site offers a new tool for single-gene investigations. "Basic researchers can specify which genes or gene pathways they're interested in, and we'll report back to them with detailed information across diagnostic categories or disease subtypes," says Kaminski.

For translational scientists, the site provides access the LGRC's expertise in clinical data analysis. "We're able to use clinical data to identify cohorts of patients based on diagnosis, gender, ethnicity, smoking history, and other factors," explains Schwartz. "These findings can then be used to understand the genetic factors that help influence disease and develop tools to implement personalized medicine approaches in chronic

lung diseases."

"We've created a system where people across the spectrum of <u>lung disease</u> research can gain access to raw data but do not need to be data-analysis experts themselves to make valuable discoveries," says Spira. Mick Correll of Dana-Farber adds that the project's overall goal "is to allow our users to become 'editors' of genomic content."

The LGRC also geared a major portion of the website to a third constituency who, although not scientists themselves, are the ultimate supporters and beneficiaries of lung cancer research: the general public. The new website includes a public-outreach section that describes the mission, goals, and strategy of the LGRC in reader-friendly terms.

"The new LGRC web portal promises to allow researchers everywhere to quickly browse the genomics datasets to analyze for answers to their scientific questions," says James Kiley, PhD, director of the National Heart, Lung and Blood Institute's Division of Lung Diseases. "This data-sharing initiative represents a great example of a way to improve access to the data available on the genomics of lung diseases for the broader scientific community."

Chronic lung diseases are rapidly emerging as a leading cause of morbidity and mortality in the U.S. and represent a significant burden on health care resources. The project was established in 2009 with a two-year, $11 million Grand Opportunity Grant from the National Heart, Lung and Blood Institute.

Provided by Dana-Farber Cancer Institute