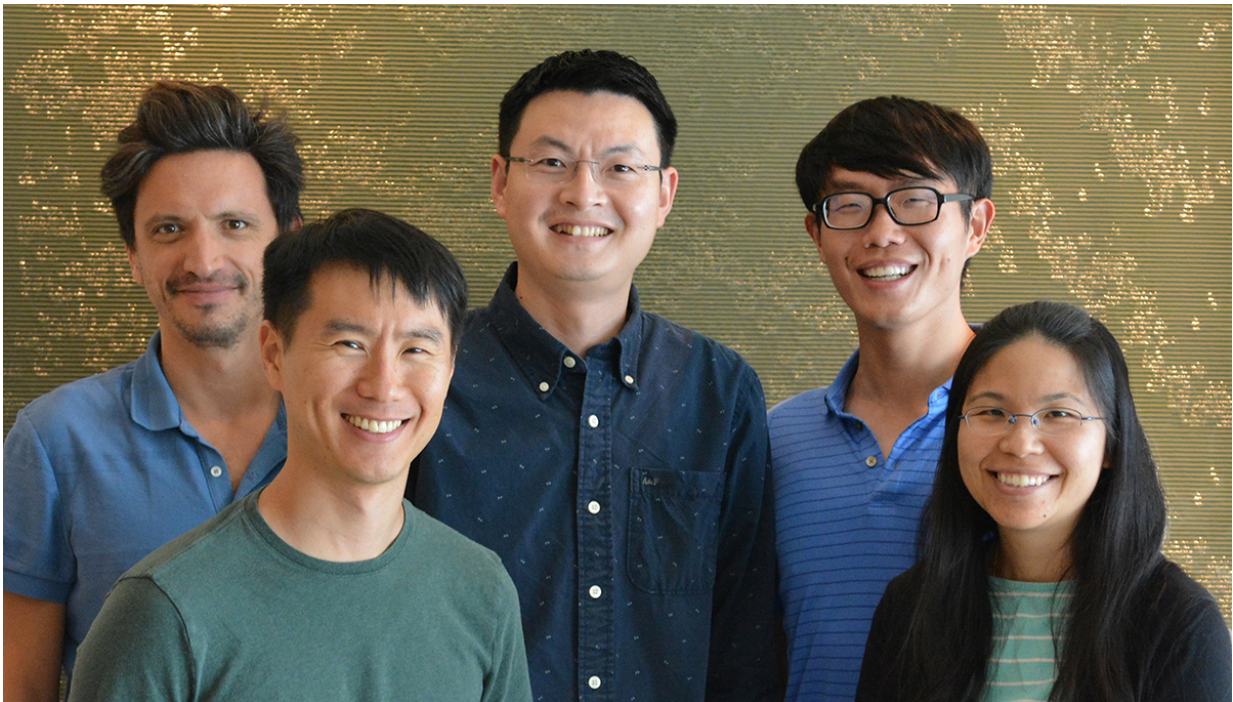


Team streamlines biomedical research by making genetic data easier to search

May 10 2016



Members of the Scripps Research Institute team include (left to right) Lelong, Andrew Su, Chunlei Wu, Jiwen Xin and Ginger Tsueng. Credit: Photo courtesy of the Scripps Research Institute.

Call them professional "data wranglers." A team of scientists at The Scripps Research Institute (TSRI) is expanding web services to make biomedical research more efficient. With their free, public projects, MyGene.info and MyVariant.info, researchers around the world have a

faster way to spot new connections between genes and disease.

"This is about how to deliver information quickly to biologists," said Chunlei Wu, associate professor of molecular medicine at TSRI.

Wu and TSRI Associate Professor Andrew Su co-led a new study published in the journal *Genome Biology* reporting on progress in setting up these services and the positive response from users so far.

Good News, Bad News

Here's the good news: Genetic sequencing is faster and more affordable these days, giving scientists a better understanding of mechanisms behind many diseases. The bad news? This flood of [genetic data](#) means scientists have to wade through multiple databases and PDF files to gather useful information.

Wu said he has spent hours downloading and parsing data, often running into problems when he discovers that the original data creators didn't annotate information in a standard way.

With support from the National Institutes of Health's (NIH) "Big Data to Knowledge" (BD2K) initiative, Wu, Su and their colleagues have begun to tame this problem by creating a data-harvesting platform to automatically import and update data from a variety of public databases. The data they aggregate are then structured and delivered via two high-performance web search services, MyGene.info and MyVariant.info, powered by the latest cloud-computation technology.

"Now researchers can focus on their own work instead of going through the data-wrangling effort," said Wu.

MyGene.info and MyVariant.info are also powerful because of their

ability to scale up as the user base and datasets grow.

MyGene.info holds information on more than 13 million genes from about 15,000 species. The service receives four to five million user "queries" each month, and the researchers are prepared to accommodate even more by expanding their use of Amazon cloud servers.

MyVariant.info currently covers more than 316 million unique variants gathered from 14 community data sources.

The services have received positive feedback from the research community so far, said Ginger Tsueng, scientific outreach project manager in the Su lab and co-author of the new study. In just this year, MyVariant.info has received more than four million hits, while MyGene.info has handled more than 17 million.

A Foundation for Future Applications

The researchers have made these services open source to encourage others to use the data and develop their own applications.

For example, researchers at the University of Washington have built an interface that retrieves data from MyGene.info and contributes additional information to run MyGene2.org, a site that aims to connect patients who share rare genetic diseases. MyGene.info also provides the backbone for BioGPS, a resource for learning about gene and protein function, run by Su, Wu and TSRI programmer Max Nanis.

Another project in the pipeline is an app built on the MyVariant.info platform that displays variants when a user scans a gene name—from a poster at a scientific conference, for example.

"Bioinformatics tools and analyses are highly dependent on having solid foundations of other tools on which to build," said Su. "MyGene.info

and MyVariant.info are key pieces of infrastructure that many bioinformaticians are using every day."

More information: Jiwen Xin et al, High-performance web services for querying gene and variant annotation, *Genome Biology* (2016). [DOI: 10.1186/s13059-016-0953-9](https://doi.org/10.1186/s13059-016-0953-9)

Provided by The Scripps Research Institute

Citation: Team streamlines biomedical research by making genetic data easier to search (2016, May 10) retrieved 26 April 2024 from <https://medicalxpress.com/news/2016-05-team-biomedical-genetic-easier.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.