

April 29, 2016 | By Daniel Kane

Bioengineers Create First Online Search Engine for Functional Genomics Data

University of California San Diego bioengineers have created what they believe to be the first online search engine for functional genomics data. This work from the [Sheng Zhong bioengineering lab](#) at UC San Diego was just published online by the journal Nucleic Acids Research. This new search engine, called [GeNemo](#), is free for public use at: genemo.org.

GeNemo addresses a pressing challenge: effectively searching functional genomic data from online data repositories. (The name GeNemo is a combination of “Ge” from the word gene and Nemo from the movie “Finding Nemo.”)

The functions of an organism’s genome, captured in functional genomic data, are directly relevant to health and disease. Functional genomics data record the diverse activities of every piece of an organism’s genome. The new search system may lead researchers to uncover the functional aspects in specific parts of genomes that are associated with normal physiology or disease of specific organs and tissues.

[GeNemo](#) queries user-input data against online functional genomic datasets, including the entire collection of ENCODE and mouse ENCODE datasets. Unlike text-based search engines, GeNemo’s searches are based on pattern matching of functional genomic regions.

Instead of just “searching by text,” the new tool allows researchers to search inside the functional data. Searching for binding patterns that are similar to that of a novel transcription factor is just one example.

“If you think of functional genomic data files as video files, then the ‘text search’ is like searching by keywords in the title or the description of a video file. The ‘inside data search’ is like searching for a video clip by pattern matching within the video itself,” explained Zhong.

“Functional genomic assays are producing massive amounts of data, in challenging data types. We have developed an online tool that empowers users to input any complete or partial functional genomic dataset, for example, a binding intensity file like bigWig, or a peak file,”

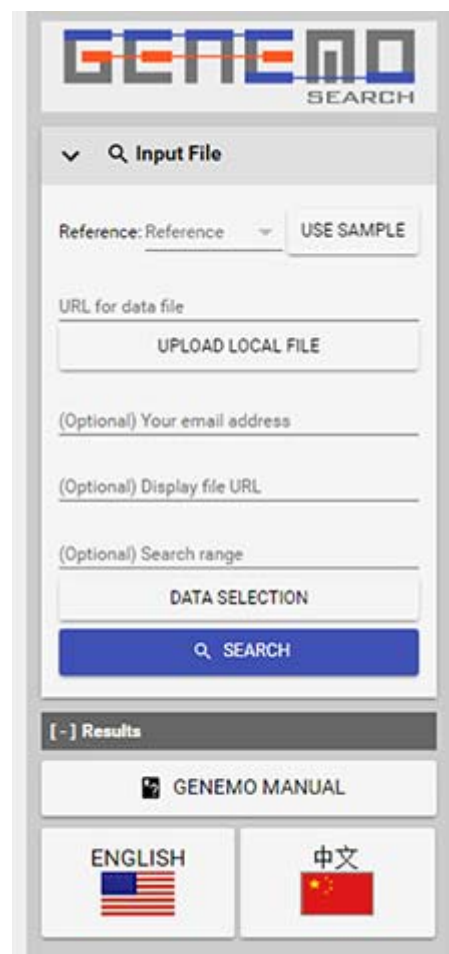
explained UC San Diego bioengineering scientist Xiaoyi Cao, a joint first author on the paper. “GeNemo reports any genomic regions, ranging from 100 bases to 100,000 bases, from any of the online ENCODE datasets that share similar functional patterns such as binding, modification and accessibility.”

Functional genomic assay data opportunities

Leveraging DNA sequencing such as a high-throughput readout, functional genomic assays can interrogate genome-wide distributions of transcription factor binding (ChIP-seq), epigenetic modifications (ChIP-seq), regulatory regions (DNase-seq, FAIRE-seq) and other functional outcomes. The results are typically stored as genome-wide intensities (WIG/bigWig files) or functional genomic regions (peak/BED files). These data types present new challenges to big data science.

According to the researchers, this is the first software to be released for executing functional genomic data searches online.

“I am excited to see how different research teams from around the world use this powerful new tool to make better use of the massive amounts of functional genomic data that is being generated every day,” said Zhong.



Screenshot from GeNemo online search engine for functional genomics data.

MEDIA CONTACT

Daniel Kane, 858-534-3262, dbkane@ucsd.edu

UC San Diego’s [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.