scienceWATCH.com
TRACKING TRENDS & PERFORMANCE IN BASIC RESEARCH
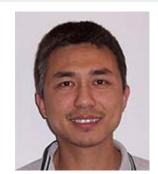
Interviews   Analyses   Data & Rankings

2010 : April 2010 - Fast Breaking Papers : Da-Wei Huang Talks About The DAVID Knowledgebase

fast breaking papers - 2010

April 2010

📄 PDF

Da-Wei Huang talks with *ScienceWatch.com* and answers a few questions about this month's Fast Breaking Paper Paper in the field of Biology & Biochemistry.

Article Title: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources
Authors: **Huang, DW**;Sherman, BT;Lempicki, RA
Journal: NAT PROTOC, Volume: 4, Issue: 1, Page: 44-57, Year: 2009
* NCI, Lab Immunopathogenesis & Bioinformat, Clin Serv Program, SAIC Frederick Inc, Frederick, MD 21702 USA.
* NCI, Lab Immunopathogenesis & Bioinformat, Clin Serv Program, SAIC Frederick Inc, Frederick, MD 21702 USA.

SW: **Why do you think your paper is highly cited?**

The DAVID Bioinformatics Resources (DAVID) was one of the earliest bioinformatics tools for the emerging needs of the high-throughput gene functional annotation analysis of large gene lists, which are usually derived from genome-wide biological studies (such as from microarray or proteomics studies).

After continuous development and improvement for several years, DAVID's quality and comprehensiveness place it in a leading position among other similar tools, as evidenced by the fact that DAVID has been used effectively in approximately 1,600 genomic publications according to Google Scholar in January 2010.

This *Nature Protocols* paper summarizes the latest functions of DAVID for users to more efficiently and more smoothly conduct the data analysis tasks.

SW: **Does it describe a new discovery, methodology, or synthesis of knowledge?**

The higher quality and comprehensiveness of DAVID come from several novel methods and ideas that other tools have not been able to extensively address.

1) With DAVID, we developed a novel single-linkage-based agglomeration method to construct one of the largest bio-knowledge databases of this kind, the DAVID Knowledgebase. This work integrated more

than 40 well-known functional annotation categories from dozens of public databases, thereby providing a solid foundation for DAVID data-mining algorithms.

2) We developed a set of novel fuzzy-logic algorithms to address the redundant relationships among many-genes-to-many-terms. This work can make the annotation results more focused and cleaner by reducing the redundancy and repeats.

3) We created a unique memory-based data IO method to enhance calculating speed more than 10 times faster than the method based on database query.

SW: **Would you summarize the significance of your paper in layman's terms?**

"...to the best of my knowledge, the paper provides the clearest principles and guidelines for genomics researchers to conduct data analysis of this kind."

The modern high-throughput technologies, such as gene chip, allow investigators to simultaneously measure the changes of genome-wide genes under certain diseases (e.g., cancer vs. normal).

After hundreds or even thousands of genes with changes are able to be identified, biological interpretation of them is an important and necessary downstream step for investigators to understand the biological mechanisms of the diseases they are studying (e.g., cancer).

The *Nature Protocols* paper describes both underlying analytical principles and step-by-step procedures to interpret biological mechanisms from the results derived from the high-throughput technologies.

SW: **How did you become involved in this research, and were there any problems along the way?**

I had obtained extensive hands-on bioinformatics experience, particularly in the areas of high-throughput DNA sequencing and microarray, through a couple of jobs in biotech companies since 2000. I had almost decided my next move would be to a pharmaceutical company, Eli Lilly, in the summer of 2004.

In the end, I decided to move back to an academic R&D environment to lead the DAVID Bioinformatics Lab at NIH/SAIC-Frederick, the operations and technical support contractor for NCI. After joining and leading the DAVID Bioinformatics lab, I realized that I had made the right decision because I completely fell in love with the DAVID project as proposed and focused by the group.

The DAVID project allows talented team members to fully combine their skills and knowledge of molecular biology, statistics, and computer science. The team efforts eventually turned out nicely, with the current version of DAVID becoming quite popular in the genomic community.

During the R&D course, we had many scientific challenges, which I actually enjoyed very much rather than thinking of them as problems. I found that it is indeed a challenging and difficult task to coordinate activities, goals, and projects across inter-disciplinary lines, with molecular biologists, computer programmers, and statisticians all as members of the same team. Since bioinformatics represents a joining of these areas, it seems that I will have to face such challenges during day-to-day project management for the rest of my professional life.

SW: **Where do you see your research leading in the future?**

The next generation of high-throughput technologies is moving toward measurements of finer biological elements from gene intensity level, such as ChIP-on-chip to measure promoter activities, SNP microarray to measure genetic land markers, and exon microarray to measure alternative splicing events.

We will expand DAVID functions to align with the analytical needs of the newer technologies, which needs are still unmet in the genomic community.

SW: **Do you foresee any social or political implications for your research?**

Although my work does not have political impact, some aspects of social implications are of immediate application. For example, to the best of my knowledge, the paper provides the clearest principles and guidelines for genomics researchers to conduct data analysis of this kind.

Da-Wei Huang, M.D.

Supervisor, DAVID Bioinformatics Lab

Laboratory of Immunopathogenesis and Bioinformatics

Clinical Services Program

SAIC-Frederick, Inc.

National Cancer Institute (NCI) at Frederick

Frederick, MD, USA

KEYWORDS: FUNCTIONAL-ANALYSIS; ONTOLOGY; TOOL; DAVID bioinformatics resources.

PDF

back to top