UT SOUTHWESTERN
MEDICAL CENTER
NEWS

Media Contact: Amanda Siegfried
214-648-3404
amanda.siegfried@utsouthwestern.edu

EMBARGOED UNTIL NOON CST, WEDNESDAY, JAN. 23, 2008

Computer-based tool aids research,
helps thwart questionable publication practices

DALLAS – Jan. 23, 2008 – A new computer-based text-searching tool developed by researchers at UT Southwestern Medical Center automatically – and quickly – compares multiple documents in a database for similarities, providing a more efficient method to carry out literature searches, as well as offering scientific journal editors a new tool to thwart questionable publication practices.

The eTBLAST computer program is efficient at flagging publications that are highly similar, said Dr. Harold "Skip" Garner, a professor of biochemistry and internal medicine at UT Southwestern who developed the computer code along with his colleagues. Not only does the code identify duplication of key words, but it also compares word proximity and order, among other variables.

The tool is especially useful for investigators who wish to analyze an unpublished abstract or project idea in order to find previous publications on the topic or identify possible collaborators working in the same field.

Another application of eTBLAST is to aid journal editors in detecting potentially plagiarized or duplicate articles submitted for publication. Dr. Garner and his colleagues explored that application in two recent articles: in a scientific paper in the Jan. 15 issue of *Bioinformatics* and in a commentary in the Jan. 24 issue of *Nature*.

In the first phase of the study, published in *Bioinformatics*, researchers used eTBLAST to analyze more than 62,000 abstracts from the past 12 years, randomly selected from Medline, one of the largest databases of biomedical research articles. They found that 0.04 percent of papers with no shared authors were highly similar and cases representing potential plagiarism. The small percentage found in the sample may appear insignificant, but when extrapolated to the 17 million scientific papers currently cited in the database, the number of potential plagiarism cases grows to nearly 7,000.

The researchers also found that 1.35 percent of papers with shared authors were sufficiently similar to be considered duplicate publications of the same data, another questionable practice.

In the second phase of the study, outlined in the *Nature* commentary, Dr. Garner and Dr. Mounir Errami, an instructor in internal medicine, refined their electronic search process so that is

(MORE)

THE UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER AT DALLAS

UT Southwestern Medical School • UT Southwestern Graduate School of Biomedical Sciences • UT Southwestern Allied Health Sciences School
UT Southwestern University Hospitals & Clinics

Office of News and Publications • 5323 Harry Hines Blvd., Dallas, TX 75390-9060 • Telephone 214-648-3404 • Fax 214-648-9119
www.utsouthwestern.edu

was thousands of times faster. An analysis of more than seven million Medline abstracts turned up nearly 70,000 highly similar papers.

Plagiarism may be the most extreme and nefarious form of unethical publication, Dr. Garner said, but simultaneously submitting the same research results to multiple journals or repeated publication of the same data may also be considered unacceptable in many circumstances.

When it comes to duplicate or repeated publications, however, there are some forms that are not only completely ethical, but also valuable to the scientific community. For example, long-term studies such as clinical trial updates and longitudinal surveys require annual or bi-annual publication of progress, and these updates often contain verbatim reproductions of much of the original text.

"We can identify near-duplicate publications using our search engine," said Dr. Garner, who is a faculty member in the Eugene McDermott Center for Human Growth and Development at UT Southwestern. "But neither the computer nor we can make judgment calls as to whether an article is plagiarized or otherwise unethical. That task must be left to human reviewers, such as university ethics committees and journal editors, the groups ultimately responsible for determining legitimacy."

Dr. Garner said eTBLAST not only detects the prevalence of duplicate publications, but also offers a possible solution to help prevent future unethical behavior.

"Our objective in this research is to make a significant impact on how scientific publications may be handled in the future," Dr. Garner said. "As it becomes more widely known that there are tools such as eTBLAST available, and that journal editors and others can use it to look at papers during the submission process, we hope to see the numbers of potentially unethical duplications diminish considerably."

Other UT Southwestern researchers in the McDermott Center who were involved in the research are computer programmer Justin Hicks, postdoctoral researcher Dr. Wayne Fisher, network analyst David Trusty and staff member Tara Long. Dr. Jonathan Wren at the Oklahoma Medical Research Foundation also participated.

The research was funded by the Hudson Foundation and the National Institutes of Health.

###