

KILPATRICK TOWNSEND & STOCKTON LLP

Joseph Petersen (JP 9071)
Robert Potter (RP 5757)
1114 Avenue of the Americas
New York, NY 10036
Telephone: (212) 775-8700
Facsimile: (212) 775-8800
Email: jpetersen@kilpatricktownsend.com

Joseph M. Beck (admitted *pro hac vice*)
W. Andrew Pequignot (admitted *pro hac vice*)
Allison Scott Roach (admitted *pro hac vice*)
1100 Peachtree Street, Suite 2800
Atlanta, Georgia 30309-4530
Telephone: (404) 815-6500
Facsimile: (404) 815-6555
Email: jbeck@kilpatricktownsend.com

Attorneys for Defendants

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

**DECLARATION OF NEIL R. SMALHEISER IN SUPPORT OF
DEFENDANTS' MOTION FOR SUMMARY JUDGMENT**

I, Neil R. Smalheiser, pursuant to 28 U.S.C. § 1746, hereby declare as follows:

1. Since August, 1996, I have been a faculty member in the Department of Psychiatry, University of Illinois at Chicago, in which I teach courses and conduct research on neuroscience and information science. Currently I am Associate Professor with Tenure. I submit this declaration in support of the defendant libraries' (the "Libraries") motion for summary judgment. Unless otherwise noted, I make this declaration based upon my own personal knowledge.
2. I received a Bachelor of Arts degree in Mathematics from the University of Iowa in 1974 and received my MD-PhD in Medicine and Neuroscience from the Albert Einstein College of Medicine in 1982.

3. I have worked in the field of text mining since 1991. “Text mining” is the use of technology to identify and extract new pieces of information from the enormous amount of knowledge available in large bodies of text. While text generally is written for people to read, text mining does not involve reading the text; instead, it uses text in digital form as data to be analyzed and processed through algorithms, which are sets of instructions or rules applied—usually by a computer—to compute a result.

4. Text mining can be applied to many different types of uses, such as retrieving and classifying documents; identifying new, interesting or particularly controversial findings; or identifying new emerging trends. In different contexts, the techniques of text-mining can be put to a variety of uses, including identifying influential experts (thought leaders) in a particular subject, predicting civil unrest in third world countries, or tracking the emergence of infectious disease outbreaks or terrorist cells.

5. A simple example of these many uses of text mining is as follows: Assume a historian discovers an unpublished manuscript of a play written in absurdist style—he suspects that it may have been written by Edward Albee or Harold Pinter. A text mining approach to this question might be tackled by collecting all of the known works of Edward Albee digitally and tabulating all of the words and phrases and punctuation marks used therein. Besides counting their individual frequencies, they can also be classified in different aggregate ways—e.g., counting the frequencies of proper names, active verbs, mentions of geographical locations, or calculating the average difficulty of the text in terms of the grade level required to understand it. This creates an overall profile of Edward Albee, and the same can be done for the known works of Harold Pinter. The profile of the unpublished manuscript is compared to the profiles of Edward Albee and Harold Pinter—if it is very similar to Albee and not to Pinter, this would

provide evidence that Albee is the likely author. If not very similar to either, this would suggest that some other author entirely may be responsible for writing it.

6. In fact, I understand that a professor at Vassar College, Donald Wayne Foster, used a form of text mining to identify Joe Klein as the writer of “Primary Colors,” a thinly veiled exposé of President Clinton’s 1992 run to the presidency which was originally published anonymously.

7. As I will discuss in more detail below, my personal experience in text mining has mostly been in the biomedical field. However, text mining processes and methods could be employed to conduct research over digital textual material of virtually any subject matter to discover new relationships, trends, correlations, and other information that may not be recognized through manually reading the texts, or that may only become apparent upon analysis of such a vast dataset that it would be virtually impossible to realize through reading.

8. I have published more than 90 peer-reviewed publications, of which more than 20 concern text mining. I have received five research grants for text mining from the National Institutes of Health (NIH) and private foundations. I have been a member of the program committee of many international conferences on medical informatics, am a member of eight journal editorial boards, and have been in leadership roles in prominent professional societies including the American Medical Informatics Association, Association for Computing Machinery, American Society of Information Science and Technology, and Society for Neuroscience. I have served on numerous grant review panels for NIH and the National Science Foundation (NSF). Attached as Exhibit A is a true and correct copy of my most recent curriculum vitae.

9. I have been asked by Kilpatrick Townsend & Stockton LLP to describe certain of the types of research that can be performed using a digital repository of works such as the repository of works offered by the Libraries through the HathiTrust Digital Library (“HDL”).

10. In working on this assignment, to date, I have read and/or referred to the HathiTrust website at hathitrust.org.

The Emerging Field of Text Mining

11. The studies of one of my mentors, Dr. Don Swanson, during the period 1986 to 1993 were an early impetus for the development of automated text mining research processes and their application in the biomedical field. Dr. Swanson developed the technique of combining separate statements, found in separate works, together to form new statements that represent new scientific hypotheses.

12. For example, suppose the statement “A affects B” appears in one work, and the statement “B affects C” appears in another work. These two works may have been published in different years by different authors, in different medical sub-fields, and no one person may have even read both of them. However, juxtaposing and viewing both statements together, one may well infer the possibility that “A affects C,” and that statement might be novel and potentially represent an important scientific discovery.

13. Dr. Swanson used this type of procedure to propose several significant medical hypotheses that were subsequently tested and confirmed clinically. For example, he proposed that fish oil supplementation would ameliorate Raynaud’s syndrome¹ and that magnesium

¹ Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med.* 1986 Autumn;30(1):7-18. Raynaud syndrome is a disorder, believed to be the result of decreases in the blood supply to parts of the body, that causes pain to and discoloration of the fingers, toes, and other areas. In some cases, the effects can be more significant, including necrosis and gangrene.

supplementation would ameliorate migraine headaches.²

14. Dr. Swanson's early studies employing this technique were carried out by hand, reading numerous articles and identifying patterns. While a researcher might be able to identify a few "A – B – C" correlations of this type manually by reading articles or other texts, Dr. Swanson and I quickly realized that through computers it is possible to search through thousands of articles to identify a large number of potentially new scientific hypotheses. Such automated search processes carry the hope of discovering correlations that individuals could not discover without computers.

15. Dr. Swanson and I created one such computer program together, called Arrowsmith,³ which was designed to consider data in the bibliographic records for biomedical articles in medical databases (e.g. the PubMed database⁴), and which given a topic A, would identify topics C that were likely to be related to it, on the basis that both topic A and topic C have some relationship to common topic B. Arrowsmith used article bibliographic records to identify these "A – B – C" correlations where no articles explicitly mentioned A and C together.

16. Arrowsmith operated by first running searches for a topic A (e.g., Huntington's Disease) and retrieving the bibliographic records for all articles that discuss that topic. Next, it created a list of all of the terms included in the titles of those articles, and these terms were treated as the B items that had a relationship to topic A and might serve as a link to identifying

² Swanson DR. Migraine and magnesium: eleven neglected connections. *Perspect Biol Med*. 1988 Summer;31(4):526-57.

³ Swanson DR, Smalheiser NR. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence* 1997; 91: 183-203.

⁴ The PubMed database consists of bibliographic data concerning ~20 million biomedical articles (including author names, title, abstract, affiliation, Medical Subject Headings, etc.). (No full-text articles are contained within the PubMed database.) Public users can query the PubMed database freely at <http://pubmed.gov>, or can apply for a relatively unrestricted license to download the entire database and manipulate the data locally on their own computers.

new topics C that had not previously been identified as related to topic A. The program ran searches for these B items and retrieved the bibliographic records for all the articles that discussed each one, creating a number of B article sets. Arrowsmith then created lists of all of the terms in the titles of each B article set, and the terms in these lists became the C items. To exclude from the results any A – C connections that may have been mentioned within the articles themselves, the program deleted from the lists of C items any terms that also appeared in the titles of the articles retrieved with the searches for topic A. Then the program ranked the remaining C items by potential relevance, according to the number of different B article sets in which they appeared (the more different B items that resulted in identifying a particular C item, the higher the possibility that the C item shared a relevant connection with the initial topic A). As a result, Arrowsmith provided a ranked list of items that may have been related to a topic but that were not identified in the existing medical literature as being related to that topic.

17. Using such a procedure, I identified a particular class of molecule called “microRNAs” as particularly likely to be involved in Huntington’s Disease, and this prediction was confirmed by subsequent research in this field.

18. In the years since we first designed and implemented the “Arrowsmith” technology we have improved upon it and made modifications to it that have enabled new discoveries.

19. For example, during the time period 2008, I was engaged in writing a review article on microRNA regulation⁵ and became interested in assessing whether “phosphorylation,” a common modification of proteins that regulates their function, might be involved in regulating the formation of microRNAs. At the time of my analysis, many proteins had been reported to

⁵ Smalheiser NR. Regulation of mammalian microRNA processing and function by cellular signaling and subcellular localization. *Biochim Biophys Acta*. 2008 Nov; 1779(11): 678-681.

interact with microRNAs, and in separate studies many proteins were known to be phosphorylated, but no one had investigated directly whether phosphorylation was responsible for regulating microRNAs.

20. I hypothesized that microRNAs (topic A) were meaningfully linked to phosphorylation (topic C), and using a modified version of the Arrowsmith program, I sought to make a list of proteins (the B items) that were candidates to mediate this connection. I used the Arrowsmith system to carry out two searches of the PubMed database (one on microRNAs and one on phosphorylation), to collect all of the titles in each set of articles, and to identify all of the words and phrases that were shared in common in both sets. The Arrowsmith system then filtered the list of words and phrases to identify the names of proteins, and then ranked the proteins according to their likely relevance (using an algorithm that we developed). The result was a shortlist of proteins that represented good candidates for further study of their possible action in regulating microRNAs by virtue of their phosphorylation.

21. The analyses described above could not reasonably be carried out manually. Not only is it necessary to use computers in order to conduct the searches of thousands of articles identified in each set (A and C), but we needed to carry out statistical modeling based on many searches in order to create a quantitative model that could predict which B items are most likely to be relevant.

22. Automated text mining continues to evolve at a remarkable pace. As more full-text becomes accessible and technology advances, increasingly these techniques focus on the full text of books and other texts, both in the general domain of digitized books (as illustrated by the example of assessing authorship of a manuscript in Paragraph 5, above) and in the biomedical domain.

The HathiTrust Digital Library and HathiTrust Research Center

23. As described in the examples above, because of the scale on which it is conducted and the complexity of the algorithms applied, a great deal of valuable text mining research cannot be carried out manually, but requires large databases of digital textual material that can be processed by computers.

24. I understand that the HDL is a shared database of over ten million digitized volumes, many of which had not previously existed in digital form, from the library collections of major research universities.

25. I believe that the HDL, as a large database of widely varied digital textual material, presents an opportunity for valuable educational and scholarly text-mining research to be conducted in a broad range of subjects and disciplines. Indeed, the same text mining techniques described above could be used to identify previously unknown trends, correlations, and relationships from information contained in the different books in the HDL.

26. I understand that the HathiTrust, through the HathiTrust Research Center, is exploring ways of enabling research similar to the text mining research conducted by myself and others as described above.

27. In my opinion, the HDL corpus is amenable to many of the same types of text mining analyses set out above. For example, scientists have developed algorithms and visualization tools designed to analyze digital text and detect “bursts,” which are sudden increases in data, and in the context of text mining, refer to sudden increases in appearance or usage of a word or topic. These tools have been used by researchers in the science community to identify major research topics and to trace research topic trends.⁶ Similar algorithms and tools

⁶ Mane KK, Börner K. Mapping topics and topic bursts in PNAS. Proc Natl Acad Sci U S A. 2004 Apr 6;101 Suppl 1:5287-90.

could be applied to the corpus of the HDL, or perhaps to a topic- or time-period-based subset of that corpus that is particularly relevant to a researcher's goals, and used to identify "bursts" or other trends in usage of specified words over time (or between different categories of published works). This type of information could provide new insight valuable to the scholarly work done in variety of subject areas, including history, political science, linguistics, literature, anthropology, sociology, philosophy, and economics.

28. For instance, I understand based on my review of information available through the HathiTrust website that Ronnie Lipschutz, a professor in the Politics Department at the University of California, Santa Cruz, is currently utilizing software text analysis techniques to document the usage of terms and concepts related to human rights in Jane Austen's novels. A resource like the HDL could allow other researchers to conduct similar research with respect to 20th century literature or other types of works not believed to be in the public domain.

29. In summary, I believe that the range of text mining analyses that can be performed on the in-copyright material in the HDL corpus are substantial, and can be beneficial for the public good.

30. I have not provided testimony as an expert in any cases in the last four years.

31. My consulting rate to review material, participate in conference calls, and to prepare this declaration is \$200/hour.

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct.

Date: June 26, 2012



Neil R. Smalheiser, MD, PhD

EXHIBIT A

CURRICULUM VITAE

NAME: Neil R. Smalheiser, MD, PhD

POSITION: Associate Professor (with tenure), Department of Psychiatry, as of 8/15/08; Adjunct Associate Professor, Department of Anatomy & Cell Biology; Member, Psychiatric Institute, University of Illinois at Chicago (9/96 - present).

ADDRESS:

Department of Psychiatry, UIC Psychiatric Institute M/C 912
1601 W. Taylor Street, room 525
Chicago, IL 60612
Phone: 312-413-4581; fax 312-413-4569; neils@uic.edu.

EDUCATION

University of Iowa, Iowa City, IA, (major: mathematics)	B. A. with Honors	1974
Albert Einstein College of Medicine, New York, NY (PhD in Neuroscience)	MD-PhD	1982

PREVIOUS EMPLOYMENT

University of Chicago, Chicago, IL, Department of Pediatrics: Intern, Postdoctoral Fellow, Instructor, and Assistant Professor 1982-1996.

University of Illinois at Chicago, Chicago, IL. Department of Psychiatry, Research Assistant Professor and Assistant Professor 1996-2008.

LICENSURE

Licensed physician, State of Illinois 1983 – present.

MEMBERSHIPS IN PROFESSIONAL ORGANIZATIONS

American Association for the Advancement of Science; American Medical Informatics Association; American Society for Information Science and Technology; Associate, Behavioral and Brain Sciences; Association for Computing Machinery; International Brain Research Organization; International Society for Neurochemistry; The RNA Society; Society for Neuroscience.

ACADEMIC HONORS AND FELLOWSHIPS

Ford Future Scientists of America Regional Award, 1968.

National Merit Finalist, 1971.

B. P. O. Elks Scholarship, 1971.

Honors Scholarships, University of Iowa, 1971-1973.

Phi Beta Kappa, 1972.

Graduation with Honors and with High Distinction, 1974.

NIH Medical Scientist Training Program Fellowship, 1974-1981.

NIH NRSA individual postdoctoral training award, 1984-1985.

Scheppe Foundation career development award, 1987-1990.

Andrew W. Mellon Foundation Fellow, 1988-1989.

1. TEACHING ACTIVITIES

- Instructor, Anatomy/Cell Biology 523, Biology of microRNAs and other Small RNAs, graduate seminar series, 2006, 2009, 2011. (created this course and taught solo).
- Instructor, Honors College core course 134, The Process of Scientific Discovery, 2010 (created this course and taught solo).
- Laboratory supervisor in Medical Neuroanatomy course for 1st year medical students 1997-2005.
- Lecturer in Neuroscience seminar series for psychiatry residents (have lectured on developmental neurobiology) 1999-present.
- Lecturer in Introduction to Biological Psychiatry course for PGY-1 psychiatry residents 2006-2010.
- Lecturer in Biological Sciences 582, graduate course on Experimental Methods in Modern Neuroscience. (have lectured on antibody methods, RNA interference, microRNAs and informatics) 2000-2004; 2008; 2010, 2011.
- Lecturer in Anatomy/Cell Biology 520, graduate course on Synaptic Structure and Function (2000, 2001).
- Lecturer in Biological Sciences 286, Biology of Brain (lectured on neurobiology of schizophrenia) (2001, 2002).
- Lecturer in GCLS 502, Molecular Biology, core course for UIC graduate students, lectured on microRNAs (2007-present).
- Lecturer in CS 582 - Information Retrieval, graduate course for UIC computer science students, January 2012.
- Lecturer in Graduate course at UIUC, Graduate School of Library and Information Science, "Literature Based Discovery", October 2008.
- Organizer and lecturer in 3 day workshop at UIC, "Informatics Tools for Discovery and Collaboration," 9/03, 9/04.

- Supervisor of undergraduate students in Biological Sciences 299 and 399 and volunteer research rotations. Logan Grewal, 1998. Mauli Verma, 1999. Rima Patel, 2002 (now a graduate student at UIC School of Public Health). Cristina Floreani, 2003 (now a MD-PhD student at UIC in Anatomy/Cell Biology). Atena Lodhi, 2004.
- Sponsor of high school students, Illinois Math and Science Academy, Student Inquiry and Research Program: Kinga Wilewska, 2004-2005. Kyle Schirmann, 2006-2007. Matthew Liu, 2007-2008.
- Mentor of Honors College undergraduate students, 2009-present.
- Sponsor of postdoctoral fellows:
 - Marc Weeber, PhD, 2001-2002, now working in industry (Knewco, Inc.).
- Supervisor of graduate research assistants:
 - Wei Zhou, 2002-2008. Wei obtained the best results (out of 30 entries nationwide) in the 2006 Genomics TREC competition. Now working at Ingenuity Systems, Inc.
 - Wei Zhang, 2002-2006. Now working at Microsoft.
 - Giovanni Lugli, PhD, 2001-present, Research Specialist in Health Sciences in my laboratory. With my support and encouragement, he is now enrolled in the Neuroscience Training Program as a PhD candidate at UIC, while continuing to work full-time in my laboratory. His thesis project concerns localization and processing of microRNA precursors within mature forebrain neurons; successfully defended his thesis on 5/19/11.
- Member of PhD thesis examination committee:
 - Wei Zhou, 2008.
 - James Gocel, 2009.
 - Sachin Moonat, 2009.
- Professional mentoring:
 - Vetle Torvik, PhD, 2001-2008, was Research Assistant Professor in my laboratory. He is developing his own line of research concerned with analyses of collaboration behavior of MEDLINE authors, and was recipient of a Summer Faculty fellowship at the National Center for Supercomputing Applications, working under Noshir Contractor. Vetle is now Visiting Assistant Professor at UIUC. Using the Author-ity author name disambiguation dataset developed at UIC, he successfully wrote a NSF grant proposal to merge Author-ity with a disambiguated US Patent database (with Lee Fleming, Harvard Business School, dual PI), beginning in 2010.
 - Carole L. Palmer, PhD. Dr. Palmer is Associate Professor at UIUC. I invited her to undertake the study of information-seeking behavior in the Arrowsmith field testers, which has developed into a NSF-funded 3 year grant that she directed.
 - Ramin Homayouni, PhD. Dr. Homayouni is Associate Professor at University of Memphis, where he now chairs the Bioinformatics Program. I assisted his informatics efforts during the period when he was a subcontract PI on my Arrowsmith grant.

- Hong Yu, PhD. Dr. Yu is Assistant Professor at University of Wisconsin-Milwaukee. I have been assisting her in writing R01 grants (am listed as a subcontract PI on an upcoming grant of hers submitted in March 2007) and in finding biologists to collaborate with in the development of biology-oriented information retrieval systems.
- Larissa Nonn, PhD. Dr. Nonn is Assistant Professor at UIC who studies the involvement of microRNAs in prostate cancer. I contributed a letter of support for her successful NCI Transition Career Development Award (K22).
- Department of Anatomy & Cell Biology, member of PhD thesis advisory committee for Paul Kim.
- Faculty Medical advisor for William Ruzicka, Anita Seibold.
- Participant in Medical and MD-PhD admissions interviews.
- Member, MD/PhD Program training faculty, Neuroscience PhD program and Biomedical Neuroscience training program, and the Graduate College.
- Fellow, UIC Honors College, 2009-present.

CURRICULUM DESIGN ACTIVITIES

Advisory Committee Member for The Scientific Communications Initiative, 2006-2009. This is a NSF-funded curriculum grant in bioinformatics centered at the Graduate School of Library and Information Science at University of Illinois Urbana-Champaign. PIs are Carole Palmer and P. Bryan Heidorn. The Scientific Communications Initiative is developing a biological informatics masters degree program for Scientific Communication Specialists (SCS). Unlike most existing educational programs in bioinformatics, the SCS program takes a broad view of biology and informatics to train professionals to bridge arenas of information technology development in the biological sciences. Other advisory committee members are chosen nationally from a variety of institutions including the American Museum of Natural History, the Smithsonian Institution, the Missouri Botanical Garden, the Peabody Museum at Yale, and the Biomedical Informatics Research Network.

INVITED PRESENTATIONS

Invited Presentations at International Conferences since 1996:

- Lecturer, Green College Thematic Lecture Series on Creativity, University of British Columbia, Vancouver, Canada, January 2002. This is a University-wide event inviting distinguished visitors from around the world, and the lectures are collected and published in book form by University of Toronto Press.
- Organizer, workshop on Informatics, Intl. Congress for Schizophrenia Research, Colorado Springs, March 2003.
- Organizer, workshop on "Informatics for Neurochemists," Intl. Soc. Neurochemistry meeting, Hong Kong, August 2003. (Meeting cancelled because of SARS epidemic.)

- Organizer, technology panel on MicroRNAs and RNA Interference in the Nervous System, Asian-Pacific Society for Neurochemistry Biennial Meeting, Hong Kong, February 2004.
- Speaker, panel on “Mining the Literature to Promote Biomedical Discoveries” at Medinfo [International Medical Informatics Association triennial meeting], San Francisco, September 2004.
- Plenary speaker and session chair, 8th International Conference on Discovery Science, Singapore, October 2005.
- Discussant, First Monday FM10 Openness Conference, Chicago, May 2006.
- Speaker, Workshop on Scholarly Databases & Data Integration, Bloomington, IN, August 2006.
- Discussant, Pacific Symposium on Biocomputing, Maui, HI, January 2007.
- Speaker, T-FaNT 07 (Tokyo Forum on Advanced NLP and Text Mining), Tokyo, Japan, March 2007.
- Co-organizer, workshop on Fragile X protein/microRNA pathways in neurons, International Society for Neurochemistry biennial meeting, Cancun, August 2007 (meeting canceled due to Hurricane Dean).
- Chair and speaker, symposium on Non-coding RNAs and Synaptic Plasticity, International Society for Neurochemistry biennial meeting, Athens, Greece, August 2011.
- Speaker, International Congress of Human Genetics, Oct. 11 - 15, 2011, Montreal, session on "Functional genomics of long non-coding RNA in mammalian systems."

Invited Presentations at National Conferences since 1996:

- Speaker, Society for Neuroscience Satellite Meeting on the Human Brain Project, November 2002.
- Organizer, panel session on Literature-Based Discovery, Am. Soc. For Information Science and Technology, Washington, DC, October 2003.
- Speaker, Short Course on Bioinformatics, Society for Neuroscience meeting, New Orleans, LA, November 2003.
- Speaker, symposium on RNA interference at the Am. Soc. Neurochemistry annual meeting, NYC, August 2004.
- Speaker, Cambridge Healthtech Institute conference on RNA Interference, San Francisco, June 2005.
- Speaker, panel on “"Enabling Biomedical Research with Literature Access and Mining: Progress and Challenges," American Medical Informatics Association annual meeting, Washington, DC, October 2005.
- Speaker, panel on “Literature-based Discovery,” American Medical Informatics Association annual Spring Congress, Phoenix, AZ, May 2006.
- Panelist, NIH Knowledge Environments for Biomedical Research (KEBR) Conference, Bethesda, Maryland, December 2006.

- Speaker, meeting on Unique Identifiers for Authors/Contributors sponsored by CrossRef, Washington, DC, February 2007.
- Speaker, Cambridge Healthtech Institute conference on microRNA in Human Disease & Development, Boston, MA, March 2007.
- Speaker, PubMed Plus conference, sponsored by the Society for Neuroscience, St. Louis, MO, June 2007.
- Participant, NSF Biomedical Informatics workshop, Portland, OR, December 2007.
- Speaker, Symposium on Computational Approaches to Creativity in Science, Stanford, CA, March 2008.
- Participant, IARPA M² Conference on Technical Discovery, Extraction and Organization, Northbrook, IL, October 2008.
- Speaker, Cambridge Healthtech Institute conference on microRNA in Human Disease & Development, Boston, MA, March 2009.
- Speaker, panel: Beyond (simple) Reading: Strategies, Discoveries, and Collaborations, Am. Soc. For Information Science and Technology, Vancouver, BC, November 2009.
- Participant, “Integrating, Representing, and Reasoning over Human Knowledge: A Computational Grand Challenge for the 21st Century,” August 7-14, 2010, at the Snowbird Ski and Summer Resort Conference Center, hosted by the Institute for Computing in Science (ICiS).

Invited Presentations within UIC since 1996:

- Dept. of Anatomy & Cell Biology, 1996.
- College of Medicine, MD-PhD Training Program, March 2005.
- Honors 201 Seminar, “Networks in Life Sciences,” March 2006.
- Autism Study Group, February 2009.
- Panel on Open Access journals, Daley Library, October 2009.
- Frontiers of GI Research Conference, February 2012.

Invited Presentations at other Universities since 1996:

- Northwestern Univ. Medical School, 1996.
- Univ. Florida at Gainesville Dept. of Pharmacology, 1996.
- Chicago Institute for Neurosurgery and Neuroresearch, 1996.
- Second Intl. Oxidative Stress and Brain Damage Symposium, 1997.
- UIUC, Graduate Library and Information Sciences School, 2001.
- UIUC, Beckman Institute, 2002.
- Stanford Univ., Division of Child and Adolescent Psychiatry, November 2002.
- Tennessee Bioinformatics Consortium, March 2004.
- Michigan State Univ., Dept. of Pharmacology and Toxicology, September 2004.
- RIKEN Biological Resource Center, Tsukuba, Japan, October 2005.

- University of Wisconsin-Milwaukee, Medical Informatics program, February 2007.
- Chicago Biomedical Consortium, RNA Symposium, June 2007.
- Chicagoland RNA Club, February 2008.
- Merck Serono (Research Knowledge Management), Geneva, Switzerland and Darmstadt, Germany, June 2008.
- Harvard Business School, Science-Based Business Initiative Seminar, February 2009.

2. RESEARCH ACTIVITIES

RESEARCH GRANTS

(active grants are indicated in bold)

- NIH NRSA individual postdoctoral training award, National Eye Institute, 1984-1985. Smalheiser, N. R., PI.
- Block Fund grant (University of Chicago), 1986. Smalheiser, N. R., PI.
- Brain Research Foundation grants, 1984-1987, 1993. Smalheiser, N. R., PI.
- Dysautonomia Foundation grants, 1986-1988. Smalheiser, N. R., PI.
- March of Dimes Basil O'Connor Starter Scholar award, 1987-1989. Smalheiser, N. R., PI.
- March of Dimes, "Laminin as a molecular and genetic probe of neurites," 1990-1992. Smalheiser, N. R., PI.
- NIH FIRST award, "Molecular and cellular basis of cranin's action on neural cells," 1988-1992. Smalheiser, N. R., PI.
- Scottish Rite Schizophrenia Research Program, "Heat shock protein 60 serum antibodies in schizophrenia," 1993-1994. Smalheiser, N. R., PI.
- NIH Program Project, "Biological basis of mental retardation," National Institute for Child Health and Human Development, 1992-1995. Schwartz, N. B., PI (I was Project P.I. of Project #2).
- Office of Naval Research, "ARROWSMITH Analysis of Biomedical Innovation and Discovery," 1999-2000 (\$50,000 direct costs). We were specifically invited to write this application by the ONR. Smalheiser, N. R., PI.
- NIH R03, "Circulating Reelin and Psychosis Vulnerability," National Institute of Mental Health; 9/00-8/02. (\$50,000 direct costs per year for 2 years). Smalheiser, N. R., PI.
- National Alliance for Autism Research, "Circulating Reelin and Autism Spectrum Disorder," 7/01-6/03 (\$45,000 direct costs per year for 2 years). Smalheiser, N. R., PI.
- NIH R01, "Arrowsmith Data Mining Techniques in Neuro-Informatics," 6/01-5/07. Human Brain Project grant, co-funded by NLM and NIMH. Funded on the first submission. (This is a large grant representing a multi-institutional consortium of six

sites, of which UIC is the home site. The overall budget is \$500,000 direct costs per year for five years.) Smalheiser, N. R., PI.

- NIH R21, “RNAi-Mediated Gene Suppression in the Adult Mammalian CNS,” National Institute of Drug Abuse; 9/30/02-9/30/05 (\$100,000 direct costs per year for 2 years, currently on no-cost extension). This is a CEBRA grant funded by NIDA for “cutting-edge” innovative high-risk, high-payoff investigations. Funded on the first submission. Smalheiser, N. R., PI.
- NIH R21, “Author Name Disambiguation in Medline,” National Library of Medicine; 1/15/05 – 6/30/08. \$125,000 direct costs per year. Funded on the first submission. This is an effort to disambiguate authors (many different people may have the same last name, first initial). We will assign all articles in Medline in clusters according to the individuals who wrote them. Smalheiser, N. R., PI.
- NIH R01, “Function of FMRP in the mouse olfactory system,” National Institute of Deafness and Other Communications Disorders; 07/01/03 – 06/30/08 Larson J., PI (N. Smalheiser, co-I, 10% effort). \$175,000 direct costs per year for five years. This is a grant to study the role of the fragile X mental retardation protein in olfactory perception and memory.
- High Q Foundation, “Literature-Based Discovery Techniques to Identify Novel Huntington Disease Modifiers, Treatments or Targets”, 8/15/07 – 2/14/08, Smalheiser, N. R., PI., \$24,000 direct costs.
- NIH R21, “Validating microRNA Analysis in Human Postmortem Brain” (Y. Dwivedi, N. Smalheiser, dual PIs). National Institute of Mental Health, 7/1/07 – 6/30/09, \$125,000 direct costs per year for 2 years requested. Funded on the first submission.
- Stanley Medical Research Institute proposal, “Prefrontal Cortex microRNAs in the Stanley Neuropathology Consortium,” Smalheiser, N. R., PI, \$75,000 per year for 2 years. 8/1/08-7/31/11.
- **NIH R01, LM010817-01, “Text Mining Pipeline to Accelerate Systematic Reviews in Evidence-Based Medicine,” Smalheiser, N. R. and Cohen, A.M., dual PIs. This is a multi-institutional consortium encompassing 4 sites, of which UIC is home site. About \$442,000 direct costs per year for 4 years. 9/30/2010 – 9/29/14. Funded on the first submission.**
- **Alzheimer’s Association, IIRG-11-202853, “Plasma microRNAs as biomarkers for Alzheimer disease,” Smalheiser, N. R., PI. 11/1/11 – 10/30/14. total \$200,000 direct costs.**
- **Dept. of the Army – USAMRAA, “Cellular Basis for Learning Impairment in Fragile X Syndrome,” Larson, J. R., PI. 04/01/2012 - 03/31/2015. \$750,000 direct costs per year for 3 years. My role is co-Investigator.**
- **University of Illinois at Chicago CCTS-0512-03, “ Plasma Small RNAs as Biomarkers for Pediatric Bipolar Disorder”, Dwivedi, Y., PI. 5/1/12 – 4/30/14. \$30,000 direct costs per year for two years. My role is co-PI.**

Pending Proposal:

NIH/NIA P01, Innovation in an Aging Society, Bruce Weinberg, PI.

Title: Innovation in an Aging Society

Agency: National Institute on Aging

Total Direct Cost Year 1: \$998,013; Total Cost Year 1: 1,419,245; Total Direct Cost for 5 Years: 5,318,371; Total Cost for 5 Years: \$7,686,358. Dates: 12/1/12 – 11/30/17

My role is co-Investigator.

About half-a-dozen proposals planned in the coming year:

NIH, Brain Research Foundation, Simons Foundation – grants on depression, autism, small RNAs, plasma microRNA biomarkers.

INVENTIONS AND COMMERCIALIZATION

Developer of two monoclonal antibodies against cranin (dystroglycan) that were licensed commercially by Chemicon.

Co-developer, with Don R. Swanson (Univ. of Chicago), of ARROWSMITH, a computer-assisted strategy for information retrieval.

Co-developer, with Vetle Torvik, of Author-ity, which utilizes a new monotone Boolean method of data mining. The Author-ity database is a resource that disambiguates author names for papers in MEDLINE. Licensed to NIH (NCBI) in 2009. Licensed to LnxResearch in 2009. Other licenses pending.

Co-developer, with Vetle Torvik, of ADAM, a database of abbreviations in Medline that includes both acronyms and non-acronyms.

Developer of WETLAB, an open source electronic notebook programmed in JAVA.

Co-developer, with Vetle Torvik, of Anne O'Tate, which facilitates summarization, drill-down and browsing of PubMed search results.

Co-developer, with Vetle Torvik, of a novel quantitative model to measure the type and amount of implicit information linking two sets of articles. Licensed to Merck Serono in 2008.

Press Coverage:

Profiled in *The Scientist* 12: 12-13, 1998.

Profiled in *Science* magazine 310: 1401, 2005.

Mentioned in an editorial in *Nature* magazine 440: 1090, 2006.

Genetic Engineering & Biotechnology News (<http://www.genengnews.com/>) rated the Arrowsmith Project website “Excellent” in their Best of the Web: Reference” list, December 2007.

Profiled/interviewed in *Biomedical Computation Review* 4: 16-27, 2008.

Mentioned in a news feature in *Nature* magazine 463: 416-418, 2010.

In addition, I have been interviewed as an expert source to comment on my own or others' work for various online news stories (e.g. *Nature*, Medicine Online, The Discovery Channel, *The Scientist*, *Biomedical Computation Review*, MyScienceWork, etc.)

PEER REVIEWED PUBLICATIONS (name is in bold if senior author)

A note on journals:

The publications span numerous specialties both within biomedical research and information sciences, and recording impact factor is misleading because different fields vary significantly in the impact factor of their leading journals. However, *Journal of Biological Chemistry* is the most important journal in the field of biochemistry; PNAS is one of the top 5 general-interest scientific journals; *Artificial Intelligence* is the leading journal in its field; *Archives of General Psychiatry* is the #2 journal in psychiatry; *Trends in Neurosciences* has the highest impact factor in neuroscience; *Journal of the American Society for Information Science and Technology* is the most prestigious journal in information science; JAMIA has the highest impact factor in medical informatics; *The New England Journal of Medicine* is the leading general-interest journal in medicine; PLOS Biology is the leading general-interest open access journal in biology; and *Trends in Genetics* is one of the top journals in genetics. *Annual Review of Information Science and Technology* is the most prestigious review journal in its field. Finally, note that the lab generally presents 2-4 abstracts at meetings each year; however, they are not listed in this curriculum vitae because they are not mature publications.

A note on author order:

We follow the convention of many biomedical laboratories, in which the person who acquires the primary data in a study and prepares the figures and tables is listed as first author. Often, but not always, this person is also the one who wrote the first draft of the paper. Other authors are listed in order of their relative contributions, except the PI who is generally listed last. This does not imply that the PI has a relatively minor role or is listed as a courtesy.

A note on open access:

Since the launching of PubMed Central, BioMed Central and Public Library of Science, my policy has been to publish articles in open access journals whenever possible.

1. **Smalheiser**, N. R. and Crain, S. M. (1978) Formation of functional retinotectal connections in co-cultures of fetal mouse explants. *Brain Res.* 148: 484-492.
2. **Smalheiser**, N. R., Crain, S. M., and Bornstein, M. B. (1981) Development of ganglion cells and their axons in organized cultures of fetal mouse retinal explants. *Brain Res.* 204: 159-178.
3. **Smalheiser**, N. R., Peterson, E. R., and Crain, S. M. (1981) Neurites from mouse retina and dorsal root ganglion explants show specific behavior within co-cultured tectum or spinal cord. *Brain Res.* 208: 499-505.
4. **Smalheiser**, N. R., Peterson, E. R., and Crain, S. M. (1981) Specific neurite pathways and arborizations formed by fetal mouse dorsal root ganglion cells within organized spinal cord explants in culture: a peroxidase labeling study. *Dev. Brain Res.* 2: 383-396.

5. **Smalheiser, N. R.** (1982) Positional specificity tests in co-cultures of retinal and tectal explants. *Brain Res.* 213: 493-499.
6. **Smalheiser, N. R., Crain, S. M., and Reid, L. M.** (1984) Laminin as a substrate for retinal axons in vitro. *Dev. Brain Res.* 12: 136-140.
7. **Smalheiser, N. R. and Crain, S. M.** (1984) Radiosensitivity and differentiation of retinal ganglion cells within fetal mouse explants in vitro. *Dev. Brain Res.* 13: 159-163.
8. **Smalheiser, N. R. and Crain, S. M.** (1984) The possible role of “sibling neurite bias” in the coordination of neurite elongation, branching, and survival. *J. Neurobiol.* 15: 517-529.
9. **Smalheiser, N. R. and Schwartz, N. B.** (1987) Cranin: a laminin binding protein of cell membranes. *Proc. Natl. Acad. Sci. USA* 84: 6457-6461.
10. **Smalheiser, N. R. and Schwartz, N. B.** (1987) Kinetic analysis of ‘rapid onset’ neurite formation in NG108-15 cells reveals a dual role for substratum-bound laminin. *Dev. Brain Res.* 34: 111-121.
11. **Schwartz, N. B. and Smalheiser, N. R.** (1989) Biosynthesis of glycosaminoglycans and proteoglycans. In: *Neurobiology of Glycoconjugates*, ed. R.U. and R.K. Margolis, Plenum Press, NY, pp. 151-186.
12. **Smalheiser, N. R.** (1989) Morphologic plasticity of rapid-onset neurites in NG108-15 cells stimulated by substratum-bound laminin. *Dev. Brain Res.* 45: 39-47.
13. **Smalheiser, N. R.** (1989) Analysis of slow-onset neurite formation in NG108-15 cells: implications for a unified model of neurite elongation. *Dev. Brain Res.* 45: 49-57.
14. **Smalheiser, N. R.** (1989) Altered cell shapes in mouse 3T3 fibroblasts treated with 5'-deoxy, 5'-methyl thioadenosine: relation to morphogenesis of neural cells. *Dev. Brain Res.* 45: 59-67.
15. **Smalheiser, N. R.** (1990) Neuronal growth cones: an extended view. *Neuroscience* 38: 1-11.
16. **Smalheiser, N. R.** (1990) Cell attachment and neurite stability in NG108-15 cells: effects of 5'-deoxy, 5'-methyl thioadenosine (MTA) compared with laminin, kinase inhibitor H-7, and Mn^{2+} ions. *Dev. Brain Res.* 51: 153-160.
17. **Smalheiser, N. R.** (1990) Cell attachment and neurite stability in NG108-15 cells: What is the role of microtubules? *Dev. Brain Res.* 58: 271-282.

18. **Smalheiser, N. R.** (1991) Role of laminin in stimulating rapid-onset neurites in NG108-15 cells: relative contribution of attachment and motility responses. *Dev. Brain Res.* 62: 81-89.
19. Pomeranz, H. D., Sherman, D. L., Smalheiser, N. R. and Gershon, M. D. (1991) Expression of the immunoreactivity of a neurally related cell surface laminin binding protein by neural crest-derived cells migrating to and within the gut: relationship to the formation of enteric ganglia. *J. Comp. Neurol.* 313: 625-642.
20. **Smalheiser, N. R.** and Collins, B. J. (1992) Characterization of a novel set of membrane antigens associated with axonal growth. I: Biochemical and functional studies. *Dev. Brain Res.* 69: 215-223.
21. **Smalheiser, N. R.** and Collins, B. J. (1992) Characterization of a novel set of membrane antigens associated with axonal growth. II: Expression in the chick central nervous system. *Dev. Brain Res.* 69: 225-231.
22. **Smalheiser, N. R.,** Collins, B. J., and Sharma, S. C. (1992) Characterization of a novel set of membrane antigens associated with axonal growth. III: Expression in the regenerating goldfish optic nerve and tectum. *Dev. Brain Res.* 69: 277-282.
23. **Smalheiser, N. R.** and Rossulek, M. (1992) Morphometric and time lapse analyses of rapid-onset neurites stimulated by cycloheximide in NG108-15 cells. *Int. J. Dev. Neurosci.* 10: 467-472.
24. Landis, C. A., Collins, B. J., Cribbs, L. L., Sukhatme, V., Bergmann, B., Rechtschaffen, A., and **Smalheiser, N. R.** (1993) Expression of EGR-1 in the brain of sleep-deprived rats. *Molec. Brain Res.* 17: 300-306.
25. **Smalheiser, N. R.** (1993) Monensin-sensitive cellular events modulate neurite extension on laminin: an example of higher order regulation of cell motility. *Cell Motil. Cytoskel.* 24: 256-263.
26. **Smalheiser, N. R.** (1993) Acute neurite retraction elicited by diverse agents is prevented by genistein, a tyrosine kinase inhibitor. *J. Neurochem.* 61: 340-343.
27. **Smalheiser, N. R.** (1993) Cranin interacts specifically with the sulfatide-binding domain of laminin. *J. Neurosci. Res.* 36: 528-538.
28. **Smalheiser, N. R.** and Swanson, D. R. (1994) Assessing a gap in the biomedical literature: magnesium deficiency and neurologic disease. *Neurosci. Res. Commun.* 15: 1-9.
29. **Smalheiser, N. R.** and Ali, J. Y. (1994) Acute neurite retraction triggered by lysophosphatidic acid: timing of the inhibitory effects of genistein. *Brain Res.* 660: 309-318.

30. **Smalheiser**, N. R. (1994) Three good things about “bad” science. *Perspect. Biol. Med.* 38: 58-60.
31. **Smalheiser**, N. R., Dissanayake, S. and Kapil, A. (1995) Regulation of neurite outgrowth and retraction by phospholipase A₂-derived arachidonic acid and its metabolites. *Brain Res.* 721: 39-48, 1996.
32. **Smalheiser**, N. R. and Kim, E. (1995) Purification of cranin, a laminin binding protein. Identity to dystroglycan and reassessment of its carbohydrate moieties. *J. Biol. Chem.* 270: 15425-15433.
33. **Smalheiser**, N. R. (1996) Proteins in unexpected locations. *Molec. Biol. Cell* 7: 1003-1014.
34. Belkin, A. M. and **Smalheiser**, N. R. (1996) Localization of cranin (dystroglycan) at sites of cell-matrix and cell-cell contact: recruitment to focal adhesions is dependent upon extracellular ligands. *Cell Adhes. Commun.* 4: 281-296.
35. **Smalheiser**, N. R. (1996) The importance of parametric approaches in the analysis of cell behavior. *Perspect. Biol. Med.* 40: 60-65.
36. **Smalheiser**, N. R. and Swanson, D. R. (1996) Indomethacin and Alzheimer's disease. *Neurology* 46: 583.
37. **Smalheiser**, N. R. and Swanson, D. R. (1996) Linking estrogen to Alzheimer's disease: an informatics approach. *Neurology* 47: 809-810.
38. Swanson, D. R. and Smalheiser, N. R. (1997) An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artif. Intell.* 91: 183-203.
39. Peng, H. B., Ali, A. A., Daggett, D. F., Rauvala, H., Hassell, J. R., and **Smalheiser**, N. R. (1998) The relationship between perlecan and dystroglycan and its implication in the formation of the neuromuscular junction. *Cell Adhes. Commun.* 5: 475-489.
40. **Smalheiser**, N. R. and Swanson, D. R. (1998) Calcium-independent phospholipase A₂ and schizophrenia. *Arch. Gen. Psychiat.* 55: 752-753.
41. **Smalheiser**, N. R. and Swanson, D. R. (1998) Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses. *Computer Meth. Prog. Biomed.* 57: 149-153.
42. **Smalheiser**, N. R., Haslam, S. M., Sutton-Smith, M., Morris, H. R., and Dell, A. (1998) Structural analysis of sequences O-linked to mannose reveals a novel Lewis X structure in cranin (dystroglycan) purified from sheep brain. *J. Biol. Chem.* 273: 23698-23703.

43. Impagnatiello, F., Guidotti, A., Pesold, C., Dwivedi, Y., Caruncho, H., Pisu, M.G., Uzunov, D.P., Smalheiser, N.R., Davis, J.M., Pandey, G.N., Pappas, G.D., Tueting, P., Sharma, R.P. and Costa, E. (1998) A decrease in reelin expression as a putative vulnerability factor in schizophrenia. *Proc. Natl. Acad. Sci. USA* 95: 15718-15723.
44. **Smalheiser**, N. R. (1998) Conserved amphipathic helices near the N-terminus and C-terminus of the alpha subunit of cranin (dystroglycan). *Cell Adhes. Commun.* 6: 401-404.
45. Swanson, D. R. and Smalheiser, N. R. (1999) Implicit text linkages between Medline records: using Arrowsmith as an aid to scientific discovery. *Library Trends* 48: 48-59.
46. **Smalheiser**, N. R., Costa, E., Guidotti, A., Impagnatiello, F., Auta, J., Lacor, P., Kriho, V. and Pappas, G. (2000) Expression of reelin in adult mammalian blood, liver, pituitary pars intermedia and adrenal chromaffin cells. *Proc. Natl. Acad. Sci. USA* 97: 1281-1286.
47. **Smalheiser**, N. R. (2000) Walter Pitts. *Perspect. Biol. Med.* 43: 217-226.
48. **Smalheiser**, N. R. and Collins, B. J. (2000) Coordinate enrichment of cranin (dystroglycan) subunits in synaptic membranes of sheep brain. *Brain Res.* 887: 469-471.
49. Manev, H., Uz, T., Smalheiser, N. R. and Manev, R. (2001) Antidepressants alter cell proliferation in the adult brain in vivo and in neural cultures in vitro. *Eur. J. Pharmacol.* 411: 67-70.
50. **Smalheiser**, N. R., Manev, H. and Costa, E. (2001) RNAi and Memory: Was McConnell on the right track after all? *Trends in Neurosci.* 24: 216-218.
51. **Smalheiser**, N. R. (2001) Predicting emerging technologies with the aid of text-based data mining: the micro approach. *Technovation* 21: 689-693.
52. Swanson, D. R., Smalheiser, N. R. and Bookstein, A. (2001) Information discovery from complementary literatures: categorizing viruses as potential weapons. *J. Am. Soc. Information Sci. Technol.* 52: 797-812.
53. Kim, H.M., Qu, T., Kriho, V., Lacor, P., Smalheiser, N., Pappas, G. D., Guidotti, A., Costa, E. and Sugaya, K. (2002) Reelin function in neural stem cell biology. *Proc. Natl. Acad. Sci. USA* 99: 4020-4025.
54. Das, A., Smalheiser, N. R., Markaryan, A. and Kaplan, A. (2002) Evidence for binding of the ectodomain of amyloid precursor protein 695 and activated high molecular weight kininogen. *Biochimica et Biophysica Acta (General Subjects)* 1571: 225-238.

55. **Smalheiser**, N. R. (2002) Informatics and hypothesis-driven research. *EMBO Reports* 3: 702.
56. **Smalheiser**, N. R. (2003) Linking investigators: A centralised linking facility for data sharing and coordination of samples in banks. *EMBO Reports* 4: 108–110.
57. Dong, E., Caruncho, H., Liu, W.-S., Smalheiser, N. R., Grayson, D. R., Costa, E. and Guidotti, A. (2003) A reelin-integrin receptor interaction regulates Arc mRNA translation in synaptoneuroosomes. *Proc. Natl. Acad. Sci. USA* 100: 5479-5484.
58. **Smalheiser**, N. R. (2003) EST analyses predict the existence of a population of chimeric microRNA precursor – mRNA transcripts expressed in normal mouse and human tissue. *Genome Biol.* 4: 403. <http://genomebiology.com/2003/4/7/403>
59. Lugli, G., Krueger, J. M., Davis, J.M. Persico, A. M., Keller, F. and **Smalheiser**, N. R. (2003) Methodological factors influencing measurement and processing of plasma reelin in humans. *BMC Biochemistry* 4: 9. <http://www.biomedcentral.com/1471-2091/4/9>
60. Gardner D, Toga AW, Ascoli GA, Beatty JT, Brinkley JF, Dale AM, Fox PT, Gardner EP, George JS, Goddard N, Harris KM, Herskovits EH, Hines ML, Jacobs GA, Jacobs RE, Jones EG, Kennedy DN, Kimberg DY, Mazziotta JC, Miller PL, Mori S, Mountain DC, Reiss AL, Rosen GD, Rottenberg DA, Shepherd GM, Smalheiser NR, Smith KP, Strachan T, Van Essen DC, Williams RW, Wong ST. (2003) Towards effective and rewarding data sharing. *Neuroinformatics.* 1: 289-295.
61. **Smalheiser**, N. R. (2003) Bath toys: a source of gastrointestinal infection. *New Engl J Med.* 350: 521.
62. **Smalheiser**, N. R. and Torvik, V. I. (2004) A population-based statistical approach identifies parameters characteristic of human microRNA-mRNA interactions. *BMC Bioinformatics* 5:139.
63. Torvik, V. I., Weeber, M., Swanson, D. R. and **Smalheiser**, N. R. (2005) A probabilistic similarity metric for Medline records: a model for author name disambiguation. *J. Am. Soc. Information Sci. Technol.* 56: 140-158.
64. **Smalheiser**, N. R. and Torvik, V. I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends in Genetics* 21: 322-326.
65. Lugli, G., Larson, J., Martone, M.E., Jones Y. and **Smalheiser**, N. R. (2005) Dicer and eIF2c are enriched at postsynaptic densities in adult mouse brain and are modified by neuronal activity in a calpain-dependent manner. *J. Neurochem.* 94: 896-905.
66. **Smalheiser**, N. R., Perkins, G. A. and Jones, S. (2005) Guidelines for negotiating scientific collaborations. Endorsed by the Am. Medical Informatics Assn. Working Group on Ethical, Legal and Social Issues. *PLOS Biology* 3: e217.

67. Zhang, W., Yu, C., Smalheiser, N. R. and Torvik, V. I. (2005) Segmentation of Publication Records of Authors from the Web. (poster paper) In the Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE'06). Atlanta, GA, April 2006. (this conference was peer-reviewed and had overall 31% acceptance rate)
68. **Smalheiser, N. R.** and Torvik, V. I. (2006) Alu elements within human mRNAs are probable microRNA targets. *Trends in Genetics* 22(10), 532-536.
69. Zhou, W., **Smalheiser, N. R.** and Yu, C. (2006) A tutorial on information retrieval: basic terms and concepts. *J. Biomed. Discovery Collaboration* 1: 2.
70. **Smalheiser, N. R.**, Torvik, V. I., Bischoff-Grethe, A., Burhans, L. B., Michael Gabriel, M., Homayouni, R., Kashef, A., Martone, M. E., Perkins, G. A., Price, D. L., Talk, A. C. and West, R. (2006) Collaborative development of the Arrowsmith two node search interface designed for laboratory investigators. *J. Biomed. Discovery Collaboration* 1: 8.
71. Swanson, D. R., Smalheiser, N. R. and Torvik, V. I. (2006) Ranking indirect connections in literature-based discovery: The role of Medical Subject Headings (MeSH). *J. Am. Soc. Information Sci. Technol.* 57: 1427-1439.
72. Zhou, W., Torvik, V. I. and **Smalheiser, N. R.** (2006) ADAM: Another database of abbreviations in MEDLINE. *Bioinformatics* 22: 2813-2818.
73. Zhou, W., Yu, C., Smalheiser, N., Torvik, V. and Hong, J. (2007) Knowledge-intensive Conceptual Retrieval and Passage Extraction of Biomedical Literature. *Proc. 30th Ann. Intl. ACM SIGIR Conf. on Research & Development on Information Retrieval (SIGIR'07)*, pp. 655-662, 2007, Amsterdam, Netherlands (this conference was peer-reviewed and had overall 18% acceptance rate).
74. Torvik, V. I. and **Smalheiser, N. R.** (2007) A quantitative model for linking two disparate literatures in MEDLINE. *Bioinformatics* 23(13): 1658-1665.
75. **Smalheiser, N. R.** and Torvik, V. I. (2008) Author name disambiguation. *Annual Review of Information Science and Technology* 43: 287-313.
76. **Smalheiser, N. R.** (2007) Exosomal transfer of proteins and RNAs at synapses in the nervous system. *Biology Direct* 2:35.
77. **Smalheiser, N. R.**, Zhou, W. and Torvik, V. I. (2008) Anne O'Tate: A tool to support user-driven summarization, drill-down and browsing of PubMed search results. *J. Biomed. Discovery Collab.* 3:2.
78. **Smalheiser, N. R.** (2008) Regulation of microRNA processing and function by cellular signaling and subcellular localization. *Biochim. Biophys. Acta Gene Regulatory Mechanisms* 1779:678-681.

79. Lugli, G., Torvik, V.I., Larson, J.R. and **Smalheiser, N. R.** (2008) Expression of microRNAs and their precursors in synaptic fractions of adult mouse forebrain. *J. Neurochem* 106: 650-661.
80. **Smalheiser, N. R.** (2008) Synaptic enrichment of microRNAs is related to structural features of their precursors. *Biology Direct* 3: 44.
81. **Smalheiser, N.R.**, Lugli, G., Torvik, V.I., Mise, N., Ikeda, R. and Abe, K. (2008) Natural antisense transcripts are co-expressed with sense mRNAs in synaptoneurosomes of adult mouse forebrain. *Neurosci. Res.* 62: 236-239.
82. **Smalheiser, N. R.**, Torvik, V.I. and Zhou, W. (2009) Arrowsmith two-node search interface: a tutorial on finding meaningful links between two disparate sets of articles in MEDLINE. *Comput. Meth. Programs Biomed.* 94: 190-197.
83. Torvik, V. I. and Smalheiser, N. R. (2009) Author name disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data* 3(3):11.
84. **Smalheiser, N. R.** and Lugli, G. (2009) microRNA regulation of synaptic plasticity. *NeuroMolecular Medicine* 11: 133-140.
85. **Smalheiser, N. R.** (2009) Do Neural Cells Communicate with Endothelial Cells via Secretory Exosomes and Microvesicles? *Cardiovascular Psychiatry and Neurology*, 2009: 383086.
86. **Smalheiser, N. R.**, Lugli, G., Lenon, A. L. Davis, J. M., Torvik, V. I. and Larson, J. R. (2010) Olfactory discrimination training up-regulates and reorganizes expression of microRNAs in adult mouse hippocampus. *ASN Neuro* 2(1):art:e00028.
87. Cohen, A.M., Adams, C.E., Davis, J.M., Yu, C., Yu, P.S., Meng, W., Duggan, L., McDonagh, M., and Smalheiser, N.R. (2010). Evidence-based medicine, the changing landscape of the medical knowledge base, and the need for automated text mining tools. *ACM 1st Intl. Conference on Health Informatics* 1:376-380.
88. **Smalheiser, N. R.**, Lugli, G., Rizavi, H., Torvik, V. I., Turecki, G. and Dwivedi, Y.(2012) MicroRNA Expression is Down-Regulated and Reorganized in Prefrontal Cortex of Depressed Suicide. *PLoS ONE* 7: e33201.
89. **Smalheiser, N. R.**, Lugli, G., Thimmapuram, J., Cook, E. H. and Larson, J. (2011) Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. *RNA* 17: 166-181.
90. **Smalheiser, N.R.**, Lugli G., Zhang, H., Rizavi, H. S., Torvik, V.I., Pandey, G.N., Davis, J. M. and Dwivedi, Y. (2010) microRNA expression in rat brain exposed to repeated inescapable shock: differential alterations in learned helplessness vs. non-learned helplessness. *Int. J. Neuropsychopharmacol.* 14: 1315-1325.

91. **Smalheiser, N.R.**, Zhou, W. and Torvik, V.I. (2011) Distribution of “characteristic” terms in MEDLINE literatures. *Information*, 2(2), 266-276.
92. **Smalheiser, N.R.** (2011). Sometimes non-IRB approved research deserves a second look. *J. Clinical Research and Bioethics* 2:104.
93. Piriyaongsa, J., Jordan, I.K., Conley, A. B., Tom Ronan and **Smalheiser, N.R.** (2010) Transcription factor binding sites are highly enriched within microRNA precursor sequences. *Biology Direct* 6: 61.
94. **Smalheiser, N. R.** (2011) Literature-based discovery: beyond the ABCs. *J. Am. Information Sci. Technol.* 63: 218-224.
95. **Smalheiser, N. R.**, Lugli, G., Thimmapuram, J., Cook, E. H. and Larson, J. (2011) Mitochondrial small RNAs that are up-regulated during olfactory discrimination training in mice. *Mitochondrion* 11: 994-995. [doi:10.1016/j.mito.2011.08.014](https://doi.org/10.1016/j.mito.2011.08.014)
96. **Smalheiser, N. R.** (2012). The search for endogenous siRNAs in the mammalian brain. *Exp. Neurol* 235: 455-463.
97. Lugli, G., Larson, J., Demars, M.P. and **Smalheiser, N. R.** (2012) Primary microRNA precursor transcripts are localized at postsynaptic densities in adult mouse forebrain. *J. Neurochem.*, submitted.
98. Shu, L., Lin, C., Meng, W., Han, Y., Yu, C. T., Smalheiser, N. R. (2012) A framework for entity resolution with efficient blocking. 13th Intl. Conference on Information Reuse and Integration (IRI), in press.

Manuscripts in preparation:

Smalheiser, N. R. and Manev, H. (2011) A case of opportunistic discovery: analysis and an aesthetic principle.

Smalheiser, N. R., Larson, J. and Dwivedi, Y. (2011). Global shifts in microRNA expression in mammalian brain: methodology, mechanisms and biology.

Smalheiser, N. R. (2011). From genome browser to text browser: a public platform to support multi-scale text annotation, corpus sharing, information retrieval and knowledge discovery.

INVITED BOOK CHAPTERS

Smalheiser, N. R. (2005) The Arrowsmith project: 2005 status report. *Discovery Science 2005. Lecture Notes in Artificial Intelligence* vol. 3735, eds. A. Hoffmann, H. Motoda, and T. Scheffer, pp. 26-43, Springer-Verlag Press, Berlin. (Invited lecture at the

8th International Conference on Discovery Science / 16th International Conference on Algorithmic Learning Theory (Singapore, October 2005), published as a book chapter.)

Smalheiser, N.R. and Torvik, V. I. (2006) Complications in mammalian microRNA target prediction. In "MicroRNA: Protocols", ed. S.-Y. Ying, in the series "Methods in Molecular Biology", published by Humana Press, pp. 115-127.

Smalheiser, N. R. and Torvik, V. I. (2008) Models of microRNA-target coordination. In "microRNAs: From Basic Science to Disease Biology", ed. K. Appasani, Cambridge University Press, pp. 221-226.

Smalheiser, N. R. and Torvik, V. I. (2008). The place of literature based discovery in contemporary scientific practice. In "Literature-Based Discovery", ed. P. Bruza and M. Weeber, Springer Press, pp. 13-22.

Lugli, G. and **Smalheiser, N. R.** (2011). Preparing Synaptonemes from Adult Mouse Forebrain. In MicroRNA: Protocols, part of the series Methods in Molecular Biology published by Humana Press. Submitted.

BOOKS AND JOURNAL SPECIAL ISSUES EDITED OR CO-EDITED

Tiffany C. Veinot, Ümit V. Çatalyürek, Gang Luo, Henrique Andrade, Neil R. Smalheiser (Eds.): ACM International Health Informatics Symposium, IHI 2010, Arlington, VA, USA, November 11 - 12, 2010, Proceedings. ACM 2010, ISBN 978-1-4503-0030-8.

Andrade, H. and Smalheiser, Neil R. (eds.): Journal of Medical Systems special issue, 2011.

SCIENTIFIC CORRESPONDENCE, EDITORIALS AND BOOK REVIEWS

Smalheiser, N. and Philipson, L. (1984) Alternative medicine. *New Engl J Med* 310: 791.

Smalheiser, N. (1984) More on the Medical College Admission Test. *New Engl J Med* 311(12): 803.

Smalheiser, N. (1988) Means to immortalize neural cells. *Trends in Neurosci.* 11: 307.

Smalheiser, N. R. (1990) Young scientists and the future. *Science* 249: 1486-1487.

Smalheiser, N. R. (1992) Teaching the Human Genome Project as a case study. *J. College Science Teaching.* 22: 7.

Smalheiser, N. R. (1994) review of *Evolution without Selection: Form and Function by Autoevolution.* *Perspect. Biol. Med.* 37: 312-313.

Smalheiser, N. R., De Groote, S. L. and Case, M. M. (2009) Open-access publishing: a new path. *J. Biomed. Discovery Collaboration* 4: 6.

JOURNAL COVERS

Cell Adhesion & Communication 5: (6), 1998.

Cerebral Cortex 9: (8), 1999.

PUBLIC WEB-DEPOSITED DATABASES

Smalheiser, N.R. and Torvik, V.I. (2004) A statistical approach predicts human microRNA targets. *Genome Biol.* 5: P4. <http://genomebiology.com/2004/5/2/P4>.

Zhou, W., Torvik, V. I. and **Smalheiser, N. R.** (2007) A database of terms in MEDLINE abstracts that co-occur frequently and share the same semantic category. Deposited on the Arrowsmith website.

PROJECT-RELATED PUBLICATIONS (supervised but was not a co-author)

Zhou, W. and Yu, C. (2005) Experiment report of TREC 2005 Genomics track ad hoc retrieval task. *The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings*, Baltimore, MD. Technical report, <http://ir.ohsu.edu/genomics/>.

Swanson, D. R. (2006) Atrial fibrillation in athletes: Implicit literature-based connections suggest that overtraining and subsequent inflammation may be a contributory mechanism. *Med. Hypotheses* 66: 1085-1092.

Swanson, D. R. (2008) Running, esophageal acid reflux, and atrial fibrillation: a chain of events linked by evidence from separate medical literatures. *Med. Hypotheses* 71: 178-185.

Swanson, D. R. (2011) Literature-based resurrection of neglected medical discoveries. *J. Biomed. Discovery Collab.*, in press.

TECHICAL REPORTS (not peer-reviewed)

Zhou, W., Yu, C., Torvik, V. I. and Smalheiser, N. R. (2006) A concept-based framework for passage retrieval in Genomics. *Fifteenth Text REtrieval Conference (TREC 2006) Proceedings*, Baltimore, WA.

Torvik, V. I., Smalheiser, N. R. and Weeber, M. (2007) A simple Perl tokenizer and stemmer for biomedical text. Posted on the Arrowsmith website to accompany the Biomedical Stemmer and Tokenizer tool.

FORMAL RESEARCH COLLABORATORS SINCE 1996 (shared active grants, were co-authors on published papers, or submitted research grant applications together)

Hong Kong University of Science and Technology, Department of Biology, Hong Kong
Benjamin Peng

Imperial College London, Department of Biological Sciences, London, UK
Anne Dell

Maryland Psychiatric Research Center, Baltimore, MD
Robert McMahon, William T. Carpenter

McGill University, Montreal, Canada
Gustavo Turecki

Ohio State University
Bruce Weinberg (plus multi-institutional collaborators on his program project)

Oregon Health and Science University, Portland, OR
Aaron Cohen, Marian McDonagh

Stanford University, Division of Child and Adolescent Psychiatry
Allan Reiss

State University of New York – Binghamton
Weiyi Meng

Univ. California-San Diego, National Center for Microscopy and Imaging Research
Maryann Martone, Guy Perkins, Diana Price

University "Campus Bio-Medico", Laboratory of Molecular Psychiatry, Rome, Italy
Antonio Persico, Flavio Keller

University of Chicago
Don Swanson, Abraham Bookstein, Yves Lussier, Andrey Rzhetsky

UIC, Department of Anatomy and Cell Biology
Orly Lazarov

UIC, Department of Biological Sciences
Arnold Kaplan, Thom Park

UIC, Department of Communication
Steve Jones

UIC, Department of Computer Science
Clement Yu, Bing Liu, Philip S. Yu

UIC, Department of Medicine
Larissa Nonn

UIC, Department of Pharmacy Administration
Bruce Lambert

UIC, Department of Psychiatry
Erminio Costa, John Davis, Yogesh Dwivedi, Robert Gibbons, Dennis Grayson,
Alessandro Guidotti, John Larson, Hari Manev, Rudmila Manev, George Pappas,
Kiminobu Sugaya, John Sweeney, Vetle Torvik, Tolga Uz.

UIC, Department of Psychology
Michael Ragozzino

Univ. IL-Urbana Champaign, Beckman Institute
Michael Gabriel

Univ. IL-Urbana Champaign, Graduate School of Library and Information Sciences
Chip Bruce, Carole Palmer, P. Bryan Heidorn

University of Indiana at Bloomington, School of Library & Information Science
Katy Borner, Ying Ding

University of Nottingham, UK
Clive Adams

University of Tennessee at Memphis, Center for Genomics and Neurobiology
Elissa Chesler (now at Oak Ridge Natl. Labs), Ramin Homayouni (now at U of
Memphis), Rob W. Williams

University of Wisconsin at Milwaukee, Department of Health Sciences
Hong Yu

3. SERVICE ACTIVITIES

ADMINISTRATIVE ACTIVITIES

Reviewing for NIH Study Sections: (including neuroscience, drug abuse, bio-computing and informatics programs)

- BISTI National Centers for Excellence in Bio-Computing Special Emphasis Panels, 4/01, 9/01, 3/02.
- Neuroinformatics Special Emphasis Panel (Human Brain Project), 9/01, 12/04.

- National Library of Medicine Special Emphasis Panels 3/03, 4/04.
- Molecular, Cellular, and Developmental Neuroscience Integrated Review Group 7/04.
- NIDA CEBRA Award review 9/04; R21/33 review 5/09.
- Challenge grants 2009.
- NCRR Centers (COBRE and RCMI), 2009; P41, 2011.
- National Library of Medicine Technology Review Panel (ARRA contracts), 8/04.
- National Center for Complementary and Alternative Medicine (NCCAM), 2/12.
- NSF Smart Health & Well Being Type 1 EXP Panel in the Information & Intelligent Systems Division (IIS), 6/12.

Reviewing for other funding agencies:

- National Science Foundation (programs on Developmental & Cellular Neuroscience and Genes & Genome Systems).
- US Army Medical Research and Materiel Command.
- Department of Health, U. K.
- US-Israel Binational Science Foundation.
- Israel Science Foundation; Basic Science Foundation (Israel Academy of Sciences and Humanities).
- University of Liège, Belgium.
- Alzheimer's Association.
- Autism Speaks.
- Research Grants Council (RGC) of Hong Kong.
- Kentucky Commercialization Fund.
- Netherlands Genomics Initiative (Horizon programme).
- Research Fund "Medizinische Forschungsförderung Innsbruck" of Innsbruck Medical University.
- Parkinson's Disease Society (UK).
- Prinses Beatrix Fonds, The Netherlands.
- India Alliance (Wellcome).
- Medical Research Council (MRC), UK.
- Netherlands Organisation for Scientific Research (NWO).

Leadership positions in National Organizations:

American Medical Informatics Association:

Ethical, Legal & Social Issues Working Group Chair-Elect/Chair/Past Chair 2003-2007.
 Knowledge Discovery and Data Mining Working Group Chair-Elect, will proceed as
 Elect/Chair/Past Chair 2008-2011.

Scientific Program Committee, 2012.

Society for Neuroscience:

Neuroinformatics Committee, member, 2009-2010.

Association for Computing Machinery (ACM):

Special Interest Group on Health Informatics (SIGHIT), Vice Chair, 2011-2013.

Member, ACM Health Informatics Task Force, 2011- present.

American Society for Information Science and Technology (ASIST):

Committee on Communications and Publications, Co-Chair, 2011-present.

Member of Program Committee for International Conferences:

- The 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, September 18-22, 2006, Berlin, Germany.
- BioCreAtIvE - Critical Assessment for Information Extraction in Biology Conference, April 23-25, 2007, October 7-9, 2009; Madrid, Spain. 2011, TBA.
- Pacific Symposium for Biocomputing, Hawaii, HI, January 4-8, 2008.
- IDAMAP: Intelligent Data Analysis in bioMedicine And Pharmacology, Verona, Italy, 2009; Washington, DC, 2010; Pisa, Italy, 2012.
- Intelligent Systems for Molecular Biology Conference, Boston, July 9-12, 2010.
- ACM 1st International Conference on Health Informatics, Washington, DC, November 11-12, 2010. Program Committee co-Chair for Medicine.
- EFMI (European Federation for Medical Informatics) Special Topic Conference, Lasko, Slovenia, April 14-15, 2011.
- 7th Conference of the Austrian Computer Society (OCG) Workgroup: Human-Computer Interaction & Usability Engineering (HCI&UE), Graz, Austria. November 25-26, 2011.
- 1st International Conference on Health Information Science, Beijing, China, April 8-10, 2012.
- Medical Informatics Europe (MIE) Conference, Pisa, Italy, August 26-29, 2012.
- HI-BI-BI, International Symposium on Network Enabled Health Informatics, Bio-Medicine and Bioinformatics, Istanbul, Turkey, 27-28 August, 2012.
- Program co-Chair, The First International Workshop on the role of Semantic Web in Literature-Based Discovery, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, October 4-7, 2012.

Membership on Editorial Boards and Advisory Boards:

- Founding Editor-in-Chief, *Journal of Biomedical Discovery and Collaboration*. Published by BioMed Central, 2005-2008; hosted by University of Illinois, 2009-present. This peer reviewed, open access journal has the unique goal of bringing together three different groups of researchers in a common forum for the first time: namely, laboratory investigators, informatics researchers who make tools to enhance

discovery and collaboration, and social scientists who study scientific practice. The Editorial Board includes internationally known leaders in each of these 3 disciplinary areas, including deans, department chairmen, named professors, program/center directors, and a Nobel laureate.

- *Biology Direct*. Open access, BioMed Central. Editorial board member, 2005-present.
- *PLOS ONE*. Open access, Public Library of Science. Editorial board member, 2011-present.
- *Frontiers in Neuroinformatics*, Frontiers Research Foundation. Open access. Editorial board member, 2007- present.
- *Biomedical Informatics Insights*, Libertas Academica. Open access. 2007-present.
- *Health Information Science and Systems (HISS)*. Biomed Central, open access. 2011-present.
- *Network Modeling and Analysis in Health Informatics and Bioinformatics*. Springer, 2012-present.
- *Health Systems*, Palgrave Macmillan, 2011-present.
- *Transactions of the IL State Academy of Science*. Editorial Board member and Chair, Science, Mathematics and Technology Education Division, 1994-1996.
- Member, Technical Advisory Board for “VIVO, Enabling National Networking of Scientists,” 2009-present. This is a NIH-funded multi-institutional consortium (Mike Conlon, Univ. of Florida, PI) that will use Semantic Web-enabled technologies to facilitate querying and collaboration across disciplines and institutions.

Ad Hoc Reviewer:

- *Neuroscience and Psychiatry Journals:*

Behavioral and Brain Sciences; Brain Research; Cardiovascular Psychiatry and Neurology; Cellular and Molecular Neurobiology; The Cerebellum; Journal of Cerebral Blood Flow and Metabolism; Journal of Neurochemistry; Journal of Neuroscience; Journal of Neuroscience and Behavioral Health; Journal of Neuroscience Research; Molecular Psychiatry; Nature Reviews Neuroscience; Neuropharmacology; Neuroreport; Neuroscience; Neuroscience Research; Restorative Neurology & Neuroscience; Trends in Neurosciences.

- *Other Biomedical Journals:*

Acta Histochemica; Biochemical Journal; Biochemical Pharmacology; Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms; BMC Developmental Biology; BMC Genomics; BMC Systems Biology, Briefings in Functional Genomics and Proteomics; Cell Research; Cellular & Molecular Biology Letters; Experimental Cell Research; International Journal of Biochemistry & Cell Biology; IUBMB Life; Journal of Biological Chemistry; Journal of Cell Biology; Journal of Clinical Investigation; Journal of Heredity; Life Sciences; Mechanisms of Aging and Development; Mobile Genetic Elements; Molecular Biology and Evolution; Nature Communications; Nature Structural and Molecular Biology; Nucleic Acids Research; Oncogene; PLOS Computational Biology; PLOS One; Proceedings of the National Academy of Sciences

USA; Proceedings of the Society of Experimental Biology and Medicine; RNA; Trends in Genetics; Wiley Interdisciplinary Reviews: RNA.

▪ *Informatics Journals:*

Annual Review of Information Science and Technology; Bioinformatics; BMC Bioinformatics; BMC Medical Informatics and Decision Making; Frontiers in Neuroinformatics; IEEE/ACM Transactions on Computational Biology and Bioinformatics; Information Processing & Management; Journal of the American Society of Information Science & Technology; Journal of Biomedical Informatics; Journal of Medical Internet Research; Neuroinformatics.

▪ *Multi-Disciplinary and Humanities Journals:*

Isis; Issues in Integrative Studies; Perspectives in Biology and Medicine; Synthese.

▪ *Conferences and Books:*

American Medical Informatics Association; Medinfo (International Medical Informatics Association); MIE (European Federation for Medical Informatics, EFMI); American Society for Information Science and Technology (ASIST). Blackwell Press (for a book on scientific discovery and one on exosome biology); EFMI Special Topic Conference.

Service for NIH Office of Neuroinformatics

Leader of Human Brain Project Working Group on Data Mining, 2005-present.

University of Illinois at Chicago Service Involvement:

- UIC Faculty Senate Academic Freedom and Tenure Committee, 2013.
- Ad hoc reviewer, Campus Research Board.
- Reader, Phi Beta Kappa nominations.
- Coordinator, multi-college UIC-UIUC Visiting Speaker Program, sponsored by the UIC Humanities Laboratory 2001-2002.
- Member, Dept. of Communication faculty search committee, 2002.
- Director, Corner for Collaborative Informatics, 2002 – present.
- Member, Chancellor's Committee on LGBT Issues, 2004-2005.
- Member, UIC Health Informatics Task Force, 2002- 2006. This is an inter-college committee that reported to Dean Tate.
- Member, Clinical and Translational Science Award (CTSA) Informatics Working Group, 2006-present. UIC received a CTSA planning grant in September 2006, and this multi-college working group was charged with planning and implementing informatics activities to support a CTSA grant application in January 2008 (which received funding).
- Affiliated member, Project Biocultures.
- Department of Psychiatry Review Committee for research involving human subjects, 2008, 2009.

Service for Industry

- Consultant to System Biosciences (SBI), 1616 North Shoreline Blvd., Mountain View, CA.

- Consultant to Acidophil, LLC, 2330 West Joppa Road, Suite 330, Lutherville, MD 21093.

COMMUNITY ACTIVITIES

Rider, Twin Cities-Chicago AIDS Ride, 1998.

Member, Lincoln Elementary School PTO Technology Committee (Oak Park, IL) 2000-2001.

Finisher, Chicago Marathon, 2004, 2006.

Invited speaker, Seminar for Scholars, Niles West High School, Niles, IL, March 2009.