



COLLEGE OF NATURAL SCIENCES
THE UNIVERSITY OF TEXAS AT AUSTIN

Department of Chemistry and Biochemistry • Austin, Texas 78712-1167 • FAX (512) 471-8696

November 28, 2005

Dr. Yves Brun
Systems Biology / Microbiology Faculty Search
Department of Biology
Indiana University
Jordan Hall 142, 1001 E 3rd St
Bloomington IN 47405-7005

Dear Dr. Brun,

It is my deep pleasure to write this letter of recommendation in the strongest possible terms to support Dr. Insuk Lee's application for a tenure-track faculty position. Insuk is a remarkable computational biologist, an excellent scientist, and one of the most productive and deep-thinking members of my group. He has an exceptional level of expertise in this very cross-disciplinary research area and I have high, and I think reasonable, expectations for his success as an independent scientist.

I have known Insuk since the fall of 2001, as a Ph.D. student in a "Bioinformatics" course I taught. He already showed a high level of scientific ability in his class projects, tackling an extremely ambitious class project. He applied the bioinformatics skills he learned to his Ph.D. dissertation research, along with a number of new ideas he developed, drawing on his self-motivated attendance of several years of computer science classes, and published his work. This research, which started from his project in my class, was published in the journal *Nucleic Acids Research*. After he finished his Ph.D. dissertation in the area of molecular genetics under the supervision of Dr. Rasika Harshey, Insuk joined my laboratory as a post-doctoral research fellow. This switch in careers from "wet" bench-top experiments to "dry" computational data mining in biology was rather unusual, but Insuk now has a high level of expertise in both molecular genetics and computational analyses.

Insuk's major post-doctoral accomplishment has been to devise a method to interpret the flood of functional genomics and proteomics data being generated by biologists. In Insuk's approach, the literally millions of experimental data points from diverse, heterogeneous functional genomics data sets can be interpreted and compared under a single, unified criterion, essentially measuring the "functional coupling" between genes. Insuk has applied this method, combined with extensive Bayesian statistical analyses, to calculate the probabilities that genes work together (are "functionally linked") within a yeast cell. The result is a network of thousands of relationships among yeast genes, defining the systems in a yeast cell as seen from the perspective of functional genomics. One important point to note about this work is that Insuk developed some extremely effective approaches for extracting pathway information from DNA microarray data. In particular, the approaches he developed were provably more accurate than many high-throughput experimental approaches (such as high-throughput yeast two-hybrid maps).

Insuk made a number of other substantial advances in the discovery of gene networks: he improved upon classical naïve Bayesian approaches for information integration by developing a simple, but quite effective, heuristic approach that allowed him to account in a parameter-sparse way for the relative independence of the different data sets—this turned out to be very useful for data integration. He also developed a simple way of finding additional linkages between genes by "bootstrapping" from the existing linkages, employing genes' network contexts in order to find new links. He also developed a simple but extremely effective statistical framework for assign confidence scores to each interaction measured in high-throughput protein interaction assays (something that the original researchers measuring the interactions frequently do not do.) Insuk's resulting gene network for yeast is the current "world champion"—an extremely accurate and extensive gene network for the genes of yeast, published in *Science* in 2004.

Several of Insuk's other research projects bear mentioning. He has since refined the computational approaches for extracting functional links from functional genomics data, first described in his *Science* paper, and presented by him at last's ISMB computational biology conference. He has already extended the methods from yeast to higher eukaryotes, including *C. elegans* and human genes. His *C. elegans* gene network is the subject of a collaborative project between my group and that of Andrew Fraser at the Sanger Genome Center in Hinxton, UK. For this project, Insuk has figured out a very

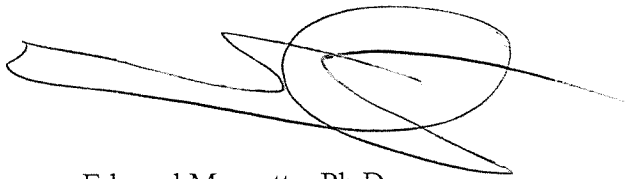
creative approach for extracting pathway information from *C. elegans* RNAi phenotypic data. His first draft *C. elegans* network is in the process of being experimentally validated by Dr. Fraser's group. Insuk's initial human gene network represents work in progress, but he has already established that the techniques he has developed are extremely effective for discovering pathways from human functional genomics data. Finally, Insuk has also developed a gene network for *Mycobacterium tuberculosis*, which serves as a framework for interpreting mycobacterial proteomics data we have collected. This work will be submitted for publication early next year. I expect Insuk's *C. elegans* network paper to be a significant one and be published in a high profile journal.

Insuk has also further refined the global yeast gene network. The improved network appears to be of high accuracy (& more accurate than the version published in *Science*), judged not only by Insuk's computational benchmarks but by experimental validation—a graduate student in the lab has tested the involvement of uncharacterized yeast genes in ribosomal biogenesis, and 8 out of 8 raw predictions have been confirmed so far. Remarkably, the topological properties of yeast genes in Insuk's latest network model turn out to predict biological phenotypes quite accurately. In particular, the essentiality of a gene correlates quite well with its connectivity in the network, with network topology able to account for ~80% of the variance of yeast gene essentiality. Thus, the network topology, which differs from physical protein interaction networks, appears to capture the biology of yeast cells in some predictive fashion, and Insuk is currently investigating this, as well as writing up this work for publication.

In his time as a post-doc in my group, Insuk has been extremely prolific, consistently producing work of high quality and creativity. He is interactive and brings a considerable intellectual maturity to my group, and the entire group has benefited from this. Insuk is a careful worker and his work is of the highest caliber; he takes the necessary pains to ensure that every step is correct, refining and extending his ideas. He may in fact be a bit too much of a perfectionist, and I think the number of his publications reflects this—however, this is balanced by the high quality and essential correctness of the work he produces, as well as by the fact that he has material for three first author papers still

unpublished or in preparation. His written English could also stand some improvement, but the ideas behind the sentences are excellent, and I do not think this will be a barrier to his competing effectively for grants or publications. He is extremely creative in the approaches that he takes, and I very much enjoy interacting with him. He has the scientific ability and maturity to effectively manage a research group, and I believe that he will be a successful independent investigator. He would be an excellent choice for a tenure-track faculty position, and I recommend him highly. Please feel free to contact me by phone (512-471-5435) or e-mail (marcotte@icmb.utexas.edu) if you should desire more details.

Sincerely,

A handwritten signature in black ink, consisting of a large, stylized 'E' followed by a long horizontal stroke that loops back under the 'E'.

Edward Marcotte, Ph.D.

Director, Center for Systems and Synthetic Biology

Associate Professor, Department of Chemistry and Biochemistry

Institute for Cellular and Molecular Biology

University of Texas at Austin