Claremont
GRADUATE UNIVERSITY

August 28, 2005

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

School of Mathematical Sciences

RE: Recommendation for Dr. David Wild, Faculty Position in Systems Biology/Microbiology

Dear Professor Brun,

I am delighted to write in support of Professor David Wild, whom I have known and worked closely with since the summer of 2000, when we agreed to jointly supervise the research of CGU Ph.D. student Dr. Claudia Rangel. That research (funded under Professor Wild's own sponsored research), titled "Modeling Biological Responses Using Gene Profiling and Linear Dynamical Statistical Models," required us to work across several disciplinary boundaries: molecular biology, mathematics, computational science, probability, and statistics. This collaboration also provided the foundation of an important bridge between Keck Graduate Institute (his home institution) and CGU, which I expect to eventually lead to a joint degree program in computational and systems biology.

My association with Professor Wild on this research was a resounding success. I found him to be a first-rate researcher, extremely knowledgeable and creative in the disciplines needed to carry out the research, and exceptionally resourceful in pulling together expertise from a variety of prominent scientists. Professor Wild has a most extraordinary capacity for intellectual work, but most importantly from my viewpoint, he was an excellent mentor for Dr. Rangel, ultimately providing her the advice and necessary connection to receive a post doctoral appointment with the prestigious computational molecular biology unit of the University of Southern California under the supervision of Professor Simon Tavare, a world renowned mathematician now working in computational molecular biology. In fact, Dr. Rangel is arguably one of the most successful Ph.D. graduates from CGU in the past 15 years, and I attribute collaboration with Professor Wild to be largely responsible for her growth and success.

Dr. Rangel's Ph.D. was granted in Spring of 2003, but the collaboration and hard work did not end there, nor did Professor Wild's enthusiasm for research in the area. Professor Wild continued to organize and motivate high quality publications related to the research. These included a book chapter, and at least three other papers in refereed journals. In particular, our collaborative work in this area was recently published in the 2004(20) issue of the journal *Bioinformatics*.

In summary, I am delighted and fortunate to have had the privilege of working with Professor Wild, and would welcome and look forward to future collaborations with him. He has my highest recommendations as a researcher and colleague.

Yours truly,

John E. Angus
Dean and Professor of Mathematics
School of Mathematical Sciences
Claremont Graduate University
John.angus@cgu.edu, 909-607-3376

SIMON TAVARÉ
George and Louise Kawamoto Professor of Biological Sciences
PROGRAM IN MOLECULAR and COMPUTATIONAL BIOLOGY
DEPARTMENT OF BIOLOGICAL SCIENCES
UNIVERSITY OF SOUTHERN CALIFORNIA
1050 CHILDS WAY, Rm 201B
LOS ANGELES, CA 90089-2910
Fax: (213) 740-8631, Phone: (213) 740-8766
e-mail: stavare@usc.edu, URL: http://www.cmb.usc.edu

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

September 5, 2005

Dear Dr Brun:

This letter is written in support of the application of **Dr. David Wild** for the systems biology position in the Department of Biology.

I have known David since he joined the Keck Graduate Institute of Applied Life Sciences, where I am a Professor-at-Large. Although we have not co-authored any papers, we have interacted many times. We were part of the organizing committee for the 2004 proteomics program at the Institute for Pure and Applied Mathematics at UCLA, and I was on the committee of one of David's PhD students who now works with me as a postdoctoral asso-ciate.

David is a bioinformatics expert whose recent research has focused on the application of advanced machine learning methods to problems in functional genomics, systems biology and structural proteomics, in collaboration with both experimentalists and theoreticians. He has considerable experience in the analysis of microarray experiments, much of this in the context modeling T-cell activation and identifying biomarkers from gene and protein expression data. He is also working on protein fold prediction, remote homolog detection, and computer simulation of structure/function relationships in drug resistant mutant proteins. Among the other-related aspects of his research is a new method for classification of microarray cancer expression data using Gaussian processes.

David's publication record is typical of someone who has spent many years in industrial or

consulting positions (such as Glaxo or EMBL). Since joining KGI in 1999 he has published many papers on a broad range of topics in bioinformatics, and his record now looks more like the conventional academic one. He has had several postdocs and PhD students, and he has a very active research program.

In addition to his interdisciplinary research program in bioinformatics and structural biology, David has several other important strengths. He will serve as a strong link between the informatics and bioscience communities, to which he also brings a statistical emphasis. This is important, because many current quantitative problems in genomics have a large statistical component as well as interesting computer science aspects. David also has a serious interest in eScience—one of his current research projects involves molecular dynamics simulations via grid-based computing. He has had a lot of experience in developing and teaching in a two-year bioscience Masters program at KGI. His experience in this program will be very useful in the development of bioinformatics training in your department.

David is a very interactive person, and is easy to get on with. He has experience in collaborative research proposals (for example, he has an active NIH grant of this type focussing on structural genomics).

In conclusion, I recommend David in the strongest terms for a position. He has a very broad perspective on computational biology, he has an active and expanding research program, and he has a lot of experience in industrial, institute and university biology. He will make a strong bioinformatics appointment.

Yours sincerely,

Simon Tavaré
Professor of Biological Sciences,
Mathematics and Preventive
Medicine

GATSBY COMPUTATIONAL NEUROSCIENCE UNIT

University College London

Alexandra House ● 17 Queen Square ● London WC1N 3AR

Tel: +44 (0)20 7679 1176
Fax: +44 (0)20 7679 1173
Email: zoubin@gatsby.ucl.ac.uk
http://www.gatsby.ucl.ac.uk

September 8, 2005

Letter of Recommendation for Prof David Wild

I have known Prof David Wild since 1997 when we started collaborating on an application of statistical machine learning methods to protein secondary structure prediction. David had read a paper of mine (on factorial hidden Markov models) and had the insight that the ideas could be adapted to this important problem in Bioinformatics and Computational Systems Biology. Our collaboration started there and has continued and grown over the last 7 years.
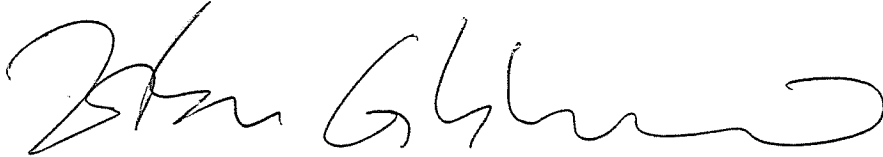
The interface between statistics, machine learning, and bioinformatics is growing to be one of the most important research areas in the life sciences. These fields fuel important advances in understanding and analyzing large data sets of genomic and proteomic data. David is a leading figure at this interface, having extensive knowledge and experience in the underlying biology, the advanced statistical modelling tools, and the latest research in the field of machine learning. David has already made many significant contributions and I'm sure will continue to do so. I will mention a few of David's contributions which I am most familiar with:

- David has developed new methods for inferring gene regulatory networks using dynamic Bayesian networks with hidden variables. These hidden variables can capture effects which cannot be directly measured in a gene expression profiling experiment, for example: genes that have not been included in the microarray, levels of regulatory proteins, the effects of mRNA and protein degradation, etc. Sophisticated Bayesian methods were required to infer these models from data.

- David, a postdoc (Chu Wei) and I have been working on predicting protein secondary structure and contact maps using hidden semi-Markov models. Unlike other machine learning methods, this approach has attempted to incorporate a large number of biological constraints into the statistical model. By combining approximate and exact inference methods, state-of-the-art prediction results can be obtained in an interpretable model.

- David has pioneered using non-parametric Bayesian mixture models for clustering of gene and protein expression data. This is an important advance on some of the ad hoc methods previously used for clustering as it allows an accurate estimate of the uncertainty in the clustering and inferences on the number of clusters in the data. David has also been collaborating more recently on testing a novel Bayesian hierarchical clustering method on biological data.

- David has worked with my group on biological applications of Gaussian process models—kernel methods similar to but more statistically motivated than support vector machines (SVMs). These have been applied to Gleason score prediction from microarrays and can be used for a variety of classification, regression and ordinal regression tasks.

David is one of the leading figures in Bioinformatics and Computational Systems Biology, and his work is sure to continue to make significant impact in this growing area, especially as it is one of the most sophisticated from the point of view of statistics and machine learning. David is very well at ease in a leadership role, knowing how to direct research and obtain the best performance out of his students and postdocs while creating a positive and nurturing environment for research. I have had the fortune of hearing David give tutorials and found that he has excellent teaching skills. He is also a very interactive personality, maintaining a wide network of international collaborators.

David is a highly productive and ambitious researcher. I am constantly impressed by his drive to explore new ideas, and by his ability to develop these into solid publications. David is also highly adept at seeking sources of funding and writing high quality proposals. I think all these skills make David Wild ideally suited for a senior faculty position in Bioinformatics and Computational Systems Biology. To summarize, I give David my *highest* recommendation.

Sincerely

Zoubin Ghahramani

Reader in Machine Learning
Gatsby Computational Neuroscience Unit
17 Queen Square, Room 403
University College London
London WC1N 3AR
United Kingdom

Associate Research Professor
Center for Automated Learning and Discovery
School of Computer Science
Carnegie Mellon University
Pittsburgh 15213
USA

Tel   +44 (0)20 7679 1199
Fax   +44 (0)20 7679 1173
Email zoubin@gatsby.ucl.ac.uk
WWW   http://www.gatsby.ucl.ac.uk/~zoubin

From Jan 15, 2006:

Professor of Information Engineering
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge CB2 1PZ
United Kingdom