

Department of Genetics

January 13, 2004

Professor Rob de Ruyter van Steveninck
Biocomplexity Faculty Search Committee
Biocomplexity Institute
Indiana University
Swain Hall West 117
Bloomington, IN 47405-7105

Dear Search Committee,

This letter is to recommend Dr. Eric Snyder for the bioinformatics position in your department. I have known Eric since 1989 when he was an undergraduate and took my course on Biological Sequence Analysis. The course was an even mix of undergraduates and graduate students, from biology as well as computer science. Eric was far and away the best student in that class. The following year he was accepted into our graduate program in Molecular, Cellular and Developmental Biology, and joined my group at the end of his first year. By that time he already had quite a bit of experience in experimental biology and wanted to focus on a computational thesis. In many respects Eric was the best student I ever had. Certainly in terms of productivity he was able to accomplish more faster than anyone else. He completed his Ph.D. degree in four years, which I think is still a record for that department. And it was not a trivial thesis, in fact I still consider it one of the best things to come from my lab. Recall that in 1992 the state of the art in human gene prediction was GRAIL, which had been published the previous year. It was a significant improvement over previous methods, but did not really predict mRNA structure, it just indicated genomic regions that were likely to contain exons. It was still up to the user to somehow piece them together, finding likely splice junctions that could be used to join exons in-frame. The Gene-ID program came out in 1992, which attempted to do that, but used an exhaustive search method that was very inefficient and not practical on long sequences. Eric's thesis project was the program GeneParser, which was the first method that took various kinds of information, such as codon usage, splice site patterns, etc, and employed dynamic programming to obtain optimal (and also easily returns ranked suboptimal) predictions. He used a neural network method to learn the appropriate weights for the different features, based on training examples. In many respects this is the forerunner of HMM methods, such as GeneScan. In fact, after I presented a talk about GeneParser at a conference, David Haussler showed me how to

modify the approach and turn it into a fully probabilistic model. We then wrote a paper together describing that approach (Stormo, G.D. and Haussler, D. (1994) Optimally Parsing a Sequence into Different Classes Based on Multiple Types of Evidence. In: Proceedings of the Second International Conference on Intelligent Systems in Molecular Biology , pp. 369-375.) which is generally considered the first paper on generalized HMMs for gene prediction. It is a bit ironic that it came out before the second, and final, paper on GeneParser, even though it was based on a modification of that method, but that is due to the delay in publishing in a journal such as JMB. So Eric's thesis was far from ordinary, in fact it was revolutionary. Now current gene prediction methods are much improved over GeneParser, and also much more efficient, but it set off on a new course of combining various kinds of information and producing guaranteed optimal (i.e. highest scoring) predictions via dynamic programming that is the basis of current methods. And he was incredibly efficient at producing that work. He would often come to me with some test results and we discussed possible improvements that should be tried. I expected it would take a week or two to implement them and then evaluate their performance, but he was often back in a couple of days with not only those results but also some other ideas that he had tried. He was not only an efficient worker, but incredibly motivated and diligent about what he was doing. His first GeneParser paper, in NAR in 1993, showed the potential of the approach in a pilot study, which he completed in the first year in my group. He had completed the much more thorough training and evaluation, as reported in the 1995 JMB paper, by early 1994 when he completed his Ph.D. thesis.

Since then I have kept up with his career to some extent, by email and when I run into him occasionally. He went immediately from graduate school to head of the bioinformatics (or Genome Analysis) group at Sequana Therapeutics. After a few years, I think when Sequana was bought, he moved back to Boulder to head up the Computational Biology section of Genomica. When that was bought, and he was left without a job just as the biotech bubble was losing air, he decided to look for an academic job. That was not easy to find at the time; there weren't too many available then and he hadn't published much during those years in the biotech industry. He found the job he currently has, at the Pennington Biomedical Research Center, part of Louisiana State University, but it hasn't turned out to be what he expected. For the first two years he was supported by a contract with a biotech company interested in finding obesity genes. As he tells it, he did all of the computational work and they kept all of the data; small bits were published in some minor medical journals, but the bioinformatics tools he had to develop to obtain the data were not given much credit. Since then he has been trying to get his own research program going, applied for an NIH grant but was turned down. He is

planning to send in a revised version. But one of the critiques, which he well appreciates, was that he is in an isolated environment without others around doing anything similar and without really a lot of support provided. Hence his interest in moving to a new environment where he will have colleagues with similar interests. I have encouraged him to do that as I think he has a lot of potential in this field, but it won't be realized until he finds a supportive environment.

In summary, Eric is a very bright and highly motivated individual with extensive experience and skills in bioinformatics, much more than are evident from his publication record. I think he only needs a better location, with interested and interactive colleagues, in order to flourish. He is very personable and always got along well with others in my group, and was always willing to help other people with their projects if he could. I can't say much about his teaching ability because he didn't do that as a graduate student, but his talks were always clear and he was good at helping new students in the lab get started. In short, I think you would find him a true asset to your group and I recommend him highly and without reservation. If you any further questions that I might be able to answer, feel free to contact me at the email address or phone number given below.

Sincerely,

A handwritten signature in black ink, appearing to read "Gary D. Stormo". The signature is fluid and cursive, with a long horizontal stroke extending to the right.

Gary D. Stormo
Professor
Director, Computational Biology
Program
stormo@genetics.wustl.edu
(314)747-5534