November 14, 2005

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

To Whom It May Concern:

This letter is in support of Guo-Cheng Yuan. Guo-Cheng was a postdoc in my lab at Harvard from January 2004 until June of 2005, when our lab left for UTSW. At the time when Guo-Cheng joined our lab, he was in the early stages of transitioning from the mathematical sciences to biology. We proposed that he join a maturing collaboration started with Ollie Rando, helping to refine and develop analytical techniques for interpreting tiled microarray data to predict nucleosome positions. This collaboration was an opportune venue in which Guo-Cheng could fruitfully apply his strong data analysis background, while learning both about chromatin biochemistry and, more broadly, about systems biology. During the time of this project, I witnessed Guo-Cheng develop from a curious outsider in biology to a mature contributor in his chosen area of interest: chromatin bioinformatics.

Guo-Cheng made three substantial contributions to our lab's effort to map nucleosome positions in yeast chromosome III. First, he helped set the normalization of the data on a sound statistical footing. The linear normalization scheme he finally settled upon was extremely appropriate to the problem, being neither too sophisticated nor too naïve. He was able to defend his approach with compelling and systematic studies at group meetings.

Second, he continued development work on a Hidden Markov Model (HMM) to detect nucleosome positioning, started by Jong Liu in our lab for his senior thesis. Guo-Cheng's modification to the hidden-state topology allowed for the detection of "fuzzy," or poorly positioned, nucleosomes. This was a key step in our ability to discern signal from noise in this data set. This task involved both understanding the mathematical underpinnings of our model, as well diving into the existing implementation of the model, built on top of Murphy's Bayes net package for Matlab. The importance of being able to understand and modify existing code is a skill that is often underappreciated in academics, and is all the more important now that the development of distributed, multidisciplinary software platforms is more common among those working in bioinformatics.

Finally, and scientifically most importantly to the collaboration, Guo-Cheng spearheaded our effort for finding a signal in the data to show that transcription factor binding sites tend to be nucleosome depleted at the genomic-scale, whereas unbound DNA sequences containing transcription factor motif sites do not have this property. This last effort really required patience, clear thinking, and scientific maturity. Comparing our data with existing transcription factor data required careful alignment of data sets and genomic features at different spatial resolutions. Finding and carefully comparing different methods for obtaining a depleted nucleosome signal in the transcriptional start site region required him to explore new analytical techniques and to develop visualization tools. As with all of Guo-Chen's work, I was impressed that his approaches were convincingly the simplest, yet most effective methods to solving our problems.

At the time when we started this collaboration with the Rando lab, the use of microarrays to map high-resolution nucleosome positions lacked precedent. As a further challenge, little validation data existed to provide the feedback that typically allows refinement of algorithm development. Hence, the project demanded good sense, a good eye, and tenacity, to develop a new platform with solid statistical underpinnings. Guo-Cheng showed all of these qualities, and has shown dedication and willingness to apply his analytical skills more generally to the rapidly growing area of chromatin bioinformatics. During the time he was in our lab, he showed a keen interest in talking directly to the experimentalists in the collaboration. Instead of simply taking the data as they came, he acquired first-hand knowledge of the potential sources of noise and bias in the experiments and was eager to develop methods to reduce the impact of these problems.

Guo-Cheng is a very careful analyst, always checking and re-checking his work. He is fun to brainstorm with and works long hours implementing many different approaches to solving problems. Guo-Cheng is very easy to work with and got along with everyone in the lab. Guo-Cheng has a lot of talent and I strongly recommend him for a faculty position in your institute. Do not hesitate to contact me if you need further information.

Best regards,

Steve Altschuler
Green Center for Systems Biology
Pharmacology Department
UT Southwestern Medical Center at Dallas
6001 Forest Park Road – ND9.202
Dallas, Texas   75390-9041