

## Bayesian approach to transforming public gene expression repositories into disease diagnosis databases

**Jasmine Zhou**  
University of Southern California, USA

The rapid accumulation of gene expression data has offered unprecedented opportunities to study human diseases. The National Center for Biotechnology Information Gene Expression Omnibus is currently the largest database that systematically documents the genome-wide molecular basis of diseases. However, thus far, this resource has been far from fully utilized. This paper describes the first study to transform public gene expression repositories into an automated disease diagnosis database. Particularly, we have developed a systematic framework, including a two-stage Bayesian learning approach, to achieve the diagnosis of one or multiple diseases for a query expression profile along a hierarchical disease taxonomy. Our approach, including standardizing cross-platform gene expression data and heterogeneous disease annotations, allows analyzing both sources of information in a unified probabilistic system. A high level of overall diagnostic accuracy was shown by cross validation. It was also demonstrated that the power of our method can increase significantly with the continued growth of public gene expression repositories. Finally, we showed how our disease diagnosis system can be used to characterize complex phenotypes and to construct a disease-drug connectivity map.

### Biography

Jasmine Zhou has completed her Ph.D. in Bioinformatics at the Swiss Federal Institute of Technology (Zurich), and conducted her post-doc training at Harvard University. She is currently a professor of biology and computer science at University of Southern California. She is also the PI of the NIH center for disease knowledge base within the MAP Gen consortium. She is a recipient of an Alfred Sloan fellowship and a NSF Career award.

[xjzhou@usc.edu](mailto:xjzhou@usc.edu)