THE AUTHOR FILE

Xiang Zhou

A love of math and biology comes together with a SPARK.

Xiang Zhou is happy to be multitudes: he's a neurobiologist and statistical genomicist, algorithm developer and coder. He joined the faculty of the University of Michigan School of Public Health in 2014.

Zhou helps experimental biologists with statistics challenges in genome-wide association studies or single-cell sequencing and keeps his eye on tool needs for emerging technologies. The new Spatial Pattern Recognition via Kernels (SPARK) method from his lab is designed to help people assess spatially resolved transcriptomic data. Biologists need to find the subset of genes with marked spatial patterns of gene expression.

Analysis methods exist, but some tools “can take hours, even days,” says Zhou. The same analysis takes SPARK minutes on a laptop or desktop. “SPARK is very computationally efficient,” he says, and comes with a user manual and a dataset for users. “Once they see that, then they can feel more comfortable to run SPARK on their own datasets,” he says.

To develop SPARK, Zhou put on his multitude of hats. First was the translation of the biological problem into math. Next came the choice of model: here, a generalized linear spatial model, which is classically used for predictions. “We essentially adapt the model to do hypothesis testing,” he says. The team needed an approach for parameter estimation such that the statistical analysis leads to calibrated *P*-values.

To make inferences within this model, the team adapted the penalized quasi-likelihood (PQL) algorithm. It fits the model with variables and weights to make calculating spatial differentially expressed genes more computationally tractable.

The algorithm delivers parameter estimates, each of which needs a *P*-value to check statistical significance of a given gene expression change. Zhou's team combined *P*-value calculation methods and mixed χ^2 distributions with the Cauchy combination rule so that SPARK delivers a *P*-value that is calibrated for false positives, the so-called type I error that is so important to biologists. “You can't just give them a *P*-value without a type I error guarantee,” says Zhou. For now, SPARK analyzes one tissue section at a time. Soon, experimental biologists will have large-scale spatial transcriptomic

Xiang Zhou

datasets, and SPARK is set up to be scalable. Confounding factors and batch effects await when multiple labs look at the same tissue under dissimilar experimental conditions with different technologies. But Zhou thinks there is way to integrate results into a more comprehensive and precise tissue description. “We see a lot of opportunities,” he says. “There's so many different extensions we can do.” And other statisticians or computational biologists might make their own SPARK extensions.

People in the Zhou lab have various backgrounds. Those who know more computing get to work on algorithm-intensive projects, while biology students take on analysis-focused ones. Both learn other aspects along the way. To prepare students and postdocs for their next challenges, Zhou makes sure their solid statistics training gives them a deep understanding of models and inference. They need to know how stable different algorithms are for different models and “what type of tricks you need to know in order to make your model work.” Implementing a model well takes good coding skills, and biology knowledge helps one appreciate a collaborator's data.

After undergraduate studies in biology at Peking University, Zhou moved to Duke University, where he completed his PhD in neurobiology and took statistics classes on the side. He completed a master's degree in statistics and even considered a second PhD in statistics but chose a statistics-oriented postdoctoral fellowship instead.

At the University of Chicago, he worked in statistical genomics, developing methods for genome-wide association studies in Matthew Stephens's lab. He worked on genomics analysis methods for Yoav Gilad's team and taught in the statistics department.

“When I was in high school, I really loved mathematics,” says Zhou. After placing first on the national college entrance exam in his native Zhejiang province, he had freedom of choice. He picked Peking University and biology as his major to follow what other high-scorers chose. “Certainly I loved biology, but I always felt I was missing something, I'm missing the mathematics,” he says. The course load was too heavy for additional courses, so he made up for that later at Duke.

“When I was in high school, I really loved mathematics.”

When Zhou has time, he hikes. He starts his day with run of a mile or two. “I do that to replace my coffee, so I don't need to drink coffee anymore,” he says. Beyond that, he plays with his two small children outside or watches cartoons with them.

“Xiang is one of those rare people who, it seems, can not only turn his hand to almost anything, but also excel at almost anything,” says Matthew Stephens. He hesitated to take on a biologist, who might lack the statistical and mathematical skills, but was impressed during the interview. “I remember really well the time when I realized I had nothing to worry about,” he says. Two weeks into the fellowship, Zhou walked into Stephens's office with a stack of linear algebra derivations like he had never seen before — or since. “While I would not say I am usually overly impressed by fancy algebra,” says Stephens, “I have to admit that this was the moment that I realized I was working with someone really special.” □

Vivien Marx

Published online: 27 January 2020
<https://doi.org/10.1038/s41592-020-0738-7>

Reference

Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* <https://doi.org/10.1038/s41592-019-0701-7> (2020).