



photo courtesy of Junhyong Kim

Junhyong Kim, Ph.D.

*Junhyong Kim is the Edmund J. and Louise W. Kahn Term-Endowed Professor in the Department of Biology, School of Arts and Sciences, with a joint appointment in the Computer and Information Science Department at Penn. He received his Ph.D in Theoretical Ecology at SUNY Stony Brook with Lev Ginzburg in 1992 and completed a postdoctoral fellowship with Margaret Kidwell in *Drosophila* genetics at the University of Arizona. Currently, he is the Co-Director of the Penn Genomics Institute, a university-wide institute for genomics research, and he is a member of the Penn Center for Bioinformatics. His research focuses on evolutionary genomics, comparative neuroscience, and comparative development using genome-wide gene expression data in yeast.*

How did you get interested in science?

I have to admit there was some brain-washing involved (just joking). My parents were humanities scholars and—presumably, from the “grass is greener on the other side” syndrome, [they] always encouraged all of us (four siblings) to be interested in science. So when I was young, for no deep reasons I always said I wanted to be a scientist—then I spent my college years rebelling and partying. Somehow, I meandered through and ended up in grad school where I finally took serious interest.

Well, now I love my job and what I do. The best part is that after all these years there is still so much I want to do [in science]. Everyday there are problems I want to work on. It also feels good that what I work on may at some point have some deeper significance, maybe even make a difference. It doesn’t get much better than that.

To those who do not know much about science, how would you describe what genomics is about?

The genome, loosely speaking, is the DNA sequence in every organism that contains instructions for generating the organism (from single cells) and “running” the organism. In particular, since the work of [Jim] Watson and [Francis] Crick, we know that the “sequence” of the genome, the order in which the four different letters of DNA are strung together, is how the instruction is specified.

Genomics started in the late 80’s to mid 90’s. As DNA sequencing technology was improving, scientists started getting the idea that instead of each lab studying pieces of the genome at a time, that

we should systematically collect all the genomic information up front as a single effort. There was a huge resistance to this idea at first. The resistance was based on the thought that science should be hypothesis driven—that is, we should first generate some hypothesis about biological workings and then try to study only the part of the genome that contains the information for that particular function. Instead, the early genomicists proposed that we should just gather all the DNA information first, hopefully leveraging economies of scale, and then figure what each part is doing and what it is useful for later. This kind of approach was derided

“we found that the workings of an organism are far more complex and interconnected than we ever thought”

as non-scientific and wasteful. Fortunately, the genome projects were funded.

Two major unexpected things came out of the genome projects. The first is that the whole strategy of doing things at a large systematic scale worked spectacularly well. The technology for sequencing genomes improved at an exponential scale such that a genome that took several years to sequence ten years ago can now be sequenced in less than a week. The second, and the more important lesson, is that we found that the workings of an organism are far more complex and interconnected than we ever thought. Before we had exhaustive enumeration of parts of a genome, people thought that a given gene might interact with a few, at

most tens, of partners. Now we know it may interact with hundreds if not thousands of other parts. It is as if we used to read books one sentence at a time from random pages and suddenly somebody read the whole book from front to end and you finally learned the story of Romeo and Juliet. Genomics now encompasses not only sequencing projects but a whole new scientific paradigm where we first systematically measure as much as we can [of DNA sequences, RNAs, proteins] and then try to develop computational, mathematical, and conceptual models of how they work together.

What is your research about and how did you get involved with it?

My lab works on topics related to evolution and theoretical biology. Two areas of interest right now are the evolution of gene expression (how DNA is transcribed into RNA) and the evolution of neurons in mammals. While we do some direct bench experiments, most of what we do is theoretical and computational biology. So the major tools that we use are mathematics, statistics, and computational analysis. Our questions always have a theoretical flavor. For example, we are interested in not so much the detailed mechanisms of how genes control each other’s expression but how the architecture of control systems is put together and how they evolve.

Being a theoretical biologist means we work on lots of other problems too, usually in collaboration with other researchers. We are engaged in many other projects ranging from fly olfaction, genes involved in autism, evolutionary history reconstruction, to biotechnology development. The latter work, technology development, is fun and I feel criti-

cal. While I personally try to work on theoretical problems, I feel it is critical to have new kinds of data. Once new data is available, new theories will arise—somewhere, sometime. Theory is autocratic, requiring the right person with the right thought. Data is democratic. This is why we at the Penn Genomics Institute have the goal of making Penn the premier place for acquisition of novel genomic data.

What are some future applications of Genomics?

These days genomics permeates all of biology. The basic thing is that we've now all realized that all these molecular parts interact in a large system doing all kinds of complicated interesting things such that we cannot really understand biological phenomena without taking into account genome-level complexities. So one quick answer is genomics applies to all biology. But for more specific examples, several things jump to mind. Cancer is really a disease of genomic instability; we are just beginning to understand this and I expect that we will finally make rapid progress after nearly 40 years of "war on cancer."

“we cannot really understand biological phenomena without taking into account genome-level complexities... genomics applies to all biology.”

We are also realizing that drug efficacy and toxicity has subtle and not-so-subtle dependency on individual genomes. Thus the future of pharmacology will be in using a patient's genomic information to tailor the therapy appropriately, what we are calling "individualized medicine."

Genome-scale genetic mapping studies are revealing the genetic risk factors underlying complex diseases such as autism. In evolutionary biology, we are using genomic information to reconstruct evolutionary histories and to decipher the relationship between the diversity of genomes and the diversity of life. And most importantly, genomic technologies will bring in new kinds of biological data, for example, real-time measurements of molecules, which will lead us to things we have not begun to think about.

What has been the greatest challenge you have had to face in your research?

Every researcher will tell you that the greatest difficulty in research is finding the funding to keep it going, especially if you have unorthodox ideas. But besides from this banal reality, there is a critical challenge that arises from the fact that my research and much of genomics in general requires really broad interdisciplinary approaches. A good project might require one to use knowledge and techniques from molecular biology, evolutionary biology, chemistry, math, statistics, mechanical engineering, algorithms, etc.

Unfortunately, we usually put premium value on specialization. For example, I am sure all the students heard of phrases like "Professor X is the world's foremost expert on Y." There aren't a lot of personal rewards for being very broad or for collaborating in an interdisciplinary team. Thus the main challenge for all of us is first, to get over the culture of "expertism" and second, to personally put our own egos to the side. Fortunately, I find that researchers at Penn are incredibly willing to do exactly that.

What kind of contributions have undergraduates made to your research?

I've always had one or two undergrads in my lab, but because of the nature of our work, we usually do not have many undergrads. Genomics and computational work has a bit of a stiff learning curve so it takes a while for students to settle in. I also try to encourage undergrads to work on semi-independent projects rather than just routine lab work. Those that have passed through have been wonderful assets to the lab and have gone off to continue research training—graduate school, MD/PHD programs, and the like.

One word on undergraduate research experience: It is often very difficult not because of technical difficulty but because research takes a lot of concentrated time and persistence. This is really difficult for an undergraduate with courses to take and other obligations. Also, the fact that something doesn't work the first time (or the tenth time) can be quite frustrating. But, if you do put in the time and are persistent, undergraduates can achieve significant work as long as they start early enough, say in your second year, and stay with it. Again, it is important to stay with a topic rather than move from one to another. There is nothing like having a first au-

thored paper in a serious research journal to get you to the next step, whether that is grad school, med school, or the industry.

What are your interests outside of academia?

I have three beautiful young children who are at maximally interesting ages and, a beautiful wife with whom I share scientific interest as well - she is a cognitive psychologist. Besides that, I used to be total gym rat so I like doing anything with a ball, although I don't get to do as much as I would like. I wish I had some eccentric hobby but can't say that is true—does being a fan of Sci-Fi and the TV show *Battlestar Galactica* count?

— Interviewed by Ningkun (Nancy) Li