



October 25, 2012

## Q&A: Mount Sinai's Andrew Kasarskis on Teaching Students how to Analyze their Own Genomes



**Name:** Andrew Kasarskis

**Age:** 39

**Position:** Vice chair and associate professor, Department of Genetics and Genomic Sciences; co-director, Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, since 2011

**Experience and Education:**

Senior director, multiscale biology, Pacific Biosciences, 2010-2011; Head of strategic initiatives, Sage Bionetworks, and affiliate investigator, Fred Hutchinson Cancer Research Center, 2009-2010; Scientific director of genetics and other positions, Rosetta Inpharmatics (owned by Merck), 2002-2009; Senior scientist and other positions, DoubleTwist, 2000-2002; Scientific curator, department of genetics, Stanford University School of Medicine, 1998-2000; PhD in molecular and cell biology, University of California, Berkeley, 1998; BS in biology and BA in chemistry, University of Kentucky

**EARLIER THIS MONTH**, Mount Sinai School of Medicine announced that it is running a course this fall in which students will sequence, analyze, and interpret a human genome — either their own or an anonymous reference.



www.genomeweb.com  
The GenomeWeb Intelligence Network

© 2012 GenomeWeb, LLC. All rights reserved. Copying, photocopying or duplicating this publication in any form other than as permitted by agreement with GenomeWeb, LLC is prohibited and may constitute copyright infringement subject to liability up to \$100,000 per infringement. For photocopy permission, back issues, bulk distribution or site licenses, please contact +1-212-651-5632 or reprints@genomeweb.com.

The elective course, called “Practical Analysis of Your Personal Genome,” is offered through the Genetics and Genomic Sciences training area within Mount Sinai's Graduate School of Biological Sciences, and 20 students, including MD and PhD students, medical residents, genetic counseling students, and junior faculty members, were selected to participate.

Last week, *Clinical Sequencing News* met with Andrew Kasarskis, vice chair of the Department of Genetics and Genomic Sciences, to talk about the design of the course and what he and his colleagues hope students will take away from it. Below is an edited version of the conversation.

### How did you select students for this course?

We had a lot more interest than we had slots — at least a factor of five, if you include all the faculty interest — so we had to select. We put in place a few prerequisites, and we had a lot of instructor input as to who we were going to select.

We wanted to get a student body that had a number of different perspectives to facilitate the training, and also our teaching. You can imagine that if you are a genetic counseling master student, your world view is very different than if you are a PhD student working in a laboratory doing bioinformatics or molecular biology, or a medical genetics resident.

We tried to select people who had a strong background in genetics, and we wanted to have that diversity because we thought that would help address the full spectrum of things you need for medical whole-genome sequencing. Also, we wanted to bring home the point that no one actually would do a general analysis of a whole human genome on their own; people can't do that. These things, functionally, in a clinical context, are done by teams of people, people who have expertise in the sequencing technology, the bioinformatics, in the annotation databases, in [the functional interpretation and clinical decision-making].

We have specifically turned down vast numbers of post-docs, faculty members, administration, all kinds of people

*continued on next page*

October 25, 2012

who wanted to take the class. It's not like a lecture where people can sit back, learn a little bit and not disrupt things. One of the things we wanted to get across is that the Unix command line is not a terrifying thing. We do have people in there who have never touched one before and who are now relatively proficient. That in itself is a bit of a triumph.

We are going to need to have a bridge if we are going to have effective teams of people doing this sort of work down the road. One of the reasons we set this course up was, we did not have an incredible mass of people who could think this way and do this at Mount Sinai, and we figured it would be good to increase that. So it's good to train people for wherever their jobs take them, and we will be studying that over time, but more importantly, just in a completely selfish way, I need folks who can do this kind of work clinically here and now.

### **What did students have to do prior to entering the course?**

We held a 26-hour summer course to prepare students for the fall course. I came to the conclusion that the only way you could properly educate people about what you would learn in a lab course looking at a whole human genome that might even be your own was to have them do an intensive sort of boot camp type course, a little like those cooking shows they have on TV, where you mix all this stuff, and then they take something already baked out of the oven, and you show it to the audience and say, 'This is how it looks when it's done.'

We kind of did that with a reference genome over the summer. They learned the operations, and then we discussed the results and what the implications of the results might be. We hunted for variants, talked about what they might mean, getting people in the frame of mind of what they would learn and what they would not learn.

Importantly, we discussed the limits of the resolution and knowledge of these things. Even if you are a geneticist, few people will appreciate all the things that diminish the usefulness of genome sequence data. The annotation databases are relatively incomplete and poor, for instance. If you are a bioinformatics jock used to doing research studies in mice with RNA-seq, you might not appreciate the degree to which the human databases are lousy. If you are a medical geneticist, you might think that the accuracy or the coverage of the sequencing is quite a bit better than it actually is. Just realizing that most variants are going to be unknown, most variants are not going to be high penetrance, having that discussed with people with different

perspectives was a vital part of that summer course.

### **Were students made aware of potential risks of sequencing their genomes?**

It was 26 hours of, essentially, group genetic counseling over the summer. We discussed that at great length. We had some discussions about [the Genetic Information Nondiscrimination Act]; we spent a lot of time just on the limits of knowledge.

The thing that I think many people have been very concerned about with regard to genome [sequencing], especially people who are in the business of providing definitive genetic tests, is that you are going to end up with all these variants and no one will know what they are. So if you are prone to concern, you might be very worried by that.

One of the lectures in the fall course I gave was on bogus results in medical genetics, just to bring home the limits of the literature as a useful tool for interpreting a personal genome. We are talking about a single genome; it's not like you've got safety in numbers. You don't have any family structure to help you out; you are looking at two alleles and trying to see what they might mean. There is a limit to what you can do with that, and we wanted to make sure that was brought home.

### **How and where are you going to sequence course participants' genomes?**

Right here in our Mount Sinai [Genomics Core] CLIA facility, using the Illumina HiSeq 2000, shooting for 30x coverage. It will be a little over two terabases of DNA sequence for the entire class. We have actually done the sequencing at this point. We went from blood draws to complete sequence in a little under a month.

### **Why did you choose to do this with whole-genome sequencing and not exome sequencing, where most of the interpretable variants seem to lie today?**

I personally have a strong belief that whole-exome sequencing will become a pretty irrelevant technology relatively soon because you have to do a capture, and the capture is never going to be 100 percent accurate. It's just another step that can get mucked up; it's just another step that costs money. And as throughput increases, you are probably just going to want to do the whole thing anyway.

And if you have got the computing power, which we do, to analyze the whole thing, why not? Also, if you are really trying to educate people about analyzing a human genome, you

*continued on next page*

October 25, 2012

might as well analyze a human genome.

If you know the exome, you sort of punt on all the structural variants and things like that because you know you are not going to really get them, and if you are doing the whole genome, at least you can talk about the limits of the technology for doing structural analysis. A lot of what people see clinically is not allelic variants that are SNPs but rather structural rearrangements or inherited structural variants. If you want the whole picture, you do the whole genome.

### **What kind of data do the students get back, and how do they analyze it?**

They are sitting at a computer and they are getting back probably a little over a billion FASTQ reads. They will run it through a fairly standard pipeline, based on the 1000 Genomes pipeline.

We talked a little bit in the lectures here about how you might want to annotate that, so they will have access to the same sorts of databases we use for our clinical annotations here, like PharmGKB for pharmacogenomics or HGMD for human mutations. Some of them will probably develop an interest in certain areas and build out databases, just as you would do in a clinical context for things that interest you.

We are going to reference the 1000 Genomes data for allele frequencies. If it's a really rare allele you might view it differently than if it's a common allele, or not, depending on what your question is.

From here until the end of the semester, people will be spending their time getting their genome sequence through there.

### **Do students have the option to exempt certain parts of their genome from the analysis?**

Absolutely. Part of the summer course was just to make sure that everyone had, based on discussions in a group setting, thought through all the issues about knowing your own personal genome sequence. This is a self-selected group of people who thought the idea was a good idea, but you want to make sure you don't have people just blithely saying, 'Hey, let's go get a tattoo,' and then regretting it for a little while. We wanted to make sure people had an opportunity to think that through.

One of the early exercises in the fall course was to compile lists of genes that would fit in various categories — for example, adult-onset neurological disorders like Huntington's or early-onset Alzheimer's, or cancer predisposition. So we divvied them up, and people came up with lists of variants,

we converted those to intervals, and you can simply exclude reads in those intervals from the analysis if you want to, and you can choose whatever intervals you want to exclude.

### **Will students have to produce some kind of report on their genomes?**

They have to report an analysis of some aspect of their personal genome. It can be whatever they want — it can be ancestry, it can be pharmacogenomics, it can be disease susceptibility, it can be an odd structural thing that they think they detected.

Given that it's a new course, and we don't know what to expect, we are really grading on participation. We do test their knowledge with a questionnaire as part of the research study we are conducting, and there is an overall oral presentation, but we are not giving letter grades on the course.

### **If a student comes up with something in their genome that seems clinically relevant, like a BRCA mutation, how can they follow up on that?**

If people saw something that they thought was medically interesting — a good chunk of the class is medically inclined — they would have access to genetic counseling and [would be referred to] medical genetics for that. And that could be done in a way that the course instructors know nothing about it, of course.

### **Is any information from the course going to go into their medical records?**

Absolutely not. There is nothing clinical about what's being done here, this is entirely an exercise for young scientists to learn how to apply certain skills to a human genome sequence that happens to be their own and therefore potentially a bit more interesting than just a random, unknown sample.

A central hypothesis of the research study is that students who have sequenced their own genomes will report more engagement than those that did not. Whether we will actually be able to test that hypothesis will be a function of how many people choose to analyze their own genome.

### **Did anyone request not to sequence their own genome?**

I have no idea. They have made that decision, but the course instructors don't know and never will. We'll know how many [decided to sequence their own genome], but probably only at the end of the course.

*continued on next page*

October 25, 2012

**What are you looking at in the research study, and did students sign an informed consent for that?**

We have a questionnaire-based research study. There are questions about what students learn, how they go about making decisions, do they experience what's called in the lingo 'decisional conflict,' and how is their psychological well-being.

Interestingly, there is very little evidence that people who have done direct-to-consumer testing, or participated in other courses that have looked at personal genomic and genetic information, have experienced psychological distress.

[For the research study,] a formal informed consent document was not required by our IRB as the questionnaire involved minimal risk, making the study exempt.

For the course itself, it was not a research study, so a research informed consent document was not appropriate. Students in the fall course had finished the 26-hour summer course, and had risks and benefits clearly explained verbally and in an extensive information sheet comparable to an informed consent document.

**How is the course financed, and how much does it cost?**

It's basically floated via department funds; it's not that much money when you think about it. The standard costs of putting together a course are what they are, and the sequencing itself we run in our internal core facility. The department is willing to eat the cost of labor and instrument depreciation, and the reagent cost comes out to a few thousand bucks per genome. That's money, and it does mean that you can't scale this to the entire medical school second year class at Mount

Sinai or something like that, but it's not an ungodly sum of money. Next to many of the courses that get offered at lots of places in a medical center, this is pretty light.

**Is there anything else you'd like to mention?**

There is a lot of concern about genetic information. There is genetic exceptionalism, genetic determinism in general, nasty histories of eugenics and other sorts of discrimination, potentially for life insurance even now. There are the limits of knowledge of genetic information, and how it might be misapplied, and the fact that it sticks with you forever.

Similarly, of course, whenever you are teaching students, there is a concern about coercion. They are considered by most IRBs to be a vulnerable population. It's not people coming off the street by their own free will; they are part of a program. So we went to very great lengths to make sure that we were non-coercive. We provided opportunities to get the full experience with your genome and without it in ways that the instructors would not know, so that the students would experience complete freedom of choice on these sorts of things.

And also, we wanted to make sure they were very well educated about what they were getting themselves into, because even though this is obviously a self-selected group of people who thought the idea would be interesting, we thought it was important to probe that a little bit, make sure you really understand what you might learn.

I think it's a good course. I think it's doing exactly what we need to do.