# Q&A: Kevin Shianna on Ramping up Sequencing for the New York Genome Center



**Name:** Kevin Shianna
**Age:** 39
**Position:** Senior vice president, sequencing operations, New York Genome Center, since July 2012



**Experience and Education:**
Director, Genomic Analysis Facility (2005-2012), director of operations, Center for Human Genome Variation (2008-2012), and assistant professor (2008-2012), Duke University School of Medicine
Postdoctoral fellow, Duke University School of Medicine, 2001-2005
PhD in microbiology and genetics, North Carolina State University, 2001
MS in microbiology and genetics, University of North Dakota School of Medicine and Health Sciences, 1997
BS in microbiology and genetics, University of Minnesota, 1995

**Kevin Shianna** recently joined the New York Genome Center as senior vice president of sequencing operations after heading the genomic analysis facility at Duke University School of Medicine for seven years.

Earlier this month, *In Sequence* visited Shianna at the NYGC's pilot laboratory at Rockefeller University, where he is building up the center's sequencing operations prior to its move to the

permanent location in lower Manhattan next year (*IS 7/24/2012*). Below is an edited version of the conversation.

---

**What attracted you to come to New York to head sequencing operations at the New York Genome Center?**

It started up by [NYGC Executive Director] Nancy [Kelley] and the NYGC team visiting Duke last year. I spent a day with them and we sort of hit it off. Then I started consulting in October, and I really liked the feel of NYGC and what was being built there. It had a startup mentality, and because of that, there was an excitement; I could feel the excitement in the city, and the interest from the institutional founding members and companies that are doing genomics work. It just seemed like something to become part of.

**How long has the pilot lab here at Rockefeller been operational? How are you equipped with instrumentation?**

We started at Rockefeller roughly in May. We have two HiSeq 2000s now, a third one will be added next week, and then we will have a fourth one coming in around mid-October − that will be an early-access HiSeq 2500. All our machines will be upgraded to HiSeq 2500s when the commercial upgrade is available, which gives us flexibility to run them in fast mode or slow mode. In addition, we have one MiSeq, we have two Caliper systems for automation of sample prep, and we have all the other miscellaneous devices that you need in place.

Right now, we partner with Illumina through the Illumina Genome Network to do [whole-]genome sequencing. But at some point, we will bring those genomes in house.

**How is the NYGC staffed right now?**

We are about 35 people, roughly half on the operational and half on the business side. As we grow, the lab and bioinformatics side will greatly expand.

The members of our group come from great labs with a lot of sequencing operations. One of us oversaw a lab at Roche, another individual came from Washington University, and then of course there is me from Duke. We all have a lot of experience running decent-size labs, but what we are trying to build here is much higher throughput. The way we are approaching this is to try to evaluate for every step of the way what's going to make us most efficient.

For example, for fragment analysis, often times we use the [Agilent] BioAnalyzer, but there are multiple other things that are a little higher throughput, so we are going step by step testing those. Because if you are going to prep, say, 3,000 samples, if you can cut off 10 minutes per sample, that's going to be a big deal.

On the bioinformatics side, Dirk Evers has been brought in as senior vice president of informatics. He used to work for Illumina in the UK. We have created a team of five bioinformaticians who are working on setting up basic high-throughput analysis pipelines for genome sequencing, exome sequencing, and RNA-seq. For genome and exome, it is going from alignment to a reference genome to calling and annotating variants, and for RNA-seq, it will be quantifying transcripts and looking for splice junctions.

Beyond that, we'll ramp up the team to look in specific areas, like setting up a program for Mendelian disease or trio analysis. We want to have units within bioinformatics where we'll build up expertise. This will be driven by either internal pilot projects or because we know we are going to be doing cancer genomics bioinformatics, for example. And then as projects come in, we will have some base expertise. There are a lot of groups we will deal with that don't have bioinformatics expertise, so this approach helps both parties.

**How are you planning to expand?**

We will start ramping up early next year by hiring five to 10 technicians. On the bioinformatics side, our high-throughput pipelines should be set up and fairly automated by then, so if we throw in 20 machines and start doing genomes or exomes, to get to annotation, it likely won't take that many more bioinformaticians. But the sample prep is always a big deal. If we transition from doing exomes to more genomes, then of course it's taking more machine capacity, and you don't need to prep as many samples. The balance [between exomes and genomes] will dictate how many techs to hire.

**How much informatics support do you offer to users?**

For genome and exome sequencing, you get the sequencing and basic bioinformatics, all the way to the point of annotation. For RNA-seq, it will be basic transcriptome alignment and quantification. In addition to that, we throw in two years of data storage and access to our computational servers. If an individual said, 'This analysis you've done is great, but we want to take it a step further; we want to do some other alignment program,' they can actually do that on the NYGC servers.

Beyond that, we go into customized projects, and that's done based on an hourly rate or another agreement we have with the group.

**How are you equipped with computational hardware?**

At Rockefeller, we have a small data storage area that is enough to get us by if the connection to our co-location facility breaks − about a month's worth of storage. In addition, we'll have some computational power here that also allows us, if there is a disruption, to look at and analyze the data. But the main activity will be happening at our co-location facility, which is the [Sabey Intergate.Manhattan facility in downtown Manhattan].

When we move to our permanent location, we will have a decent-size server room that will be large enough to hold six months' worth of data, but the plan still is to have that as a staging area, and once the data is analyzed and we know what we want to keep, it gets pushed over to the co-location facility, which then is backed up someplace in Washington. In addition to that, we will have some decent computational power, but the majority of our computational clusters will be at the co-location facility.

We will also have access to a cloud-based solution, so if we have a huge project, we can temporarily expand data storage and computation as needed. I think right now, it's still cheaper, or equal price, to set up what we've set up versus the cloud, but once that changes, it's fairly obvious what we should do. There are issues as far as the CLIA lab we're setting up, but I would guess that over time, those issues would be solved.

**To what extent will your projects determine what technologies you will bring in?**

You have to let science drive your decisions, but I would hope that the Innovation Center is thinking a step ahead of that. We are a genome center; we want to look at all technologies that are coming out independent of what the scientific question is. So then, when the scientific question comes in, we say, 'We have already tested these, what would work best for that is X.'

With the Ion Proton, for example, we'll say 'let's sequence an exome that we have sequenced on the HiSeq on the Proton I chip, and see what's the difference.' There is no scientific question behind that; it's just testing the technology. It's not even just equipment; it's also different ways to prep samples — FFPE for RNA-seq and for exomes, for example — as new kits come out and even as we develop homebrew kits.

**Have you found vendors to be cooperative in granting you access to new technologies early?**

At Duke, we had people interested in us because we are doing great science, but at NYGC, everyone wants to be associated and work with us. So yes, they have been very willing to work with us on trying early-access stuff.

**What are the greatest challenges of building a genome center from scratch?**

Putting all the pieces together, I think, is the biggest thing. You can buy or lease a building, you can buy machines, you can hire people, but you don't just unlock the doors and turn the lights on to have it work. And that, from my perspective, has been one of the hardest things, to try to get groups to interact. For example, lab folks and bioinformatics folks, they have to work together when you are doing sequencing. That process takes time. A lot of groups can do it successfully, but remember, we are going for ultimate efficiency here.

We will have some internal pilot studies set up within the sequencing facility to help drive that, actually testing our lab technicians and the bioinformatics. For example, sequencing one of the

[HapMap] CEPH trios and then doing a complete analysis of that. By doing that, you can evaluate how the sequencing process went, how the bioinformatics went, what the QC was like. So it's sort of a dry run on what a project will look like, and it's very simplistic, but you can actually track through and determine what's working and what's not working.

**What types of projects will you be taking on?**

A lot of them will be driven by the interest of the scientific director [who is currently being recruited], as well as by the institutional founding members. Right now, we have had probably 60 or 70 discussions about different projects, anything from looking at sequencing flies to sharks. Obviously, there is going to be a focus on cancer because cancer is so big in this area.

**Can you talk about ongoing projects?**

For the Alzheimer's disease project with the Feinstein Institute for Medical Research, a fair number of samples are in the process of being sequenced. In addition, we have roughly five projects that are ongoing, working with various institutional founding members, and there are probably another 15 that are likely to happen within the next three months or so. These vary in size, and it's a mix of RNA-seq, exomes, and, to a lesser extent, genomes. A lot of them go beyond the typical analysis, where they'll need additional bioinformatics support.

**How are you planning to help bring high-throughput sequencing to the clinic?**

The new facility has a side section for CLIA sequencing, the size of which is expandable. We are in talks with the individual who set up Illumina's sequencing CLIA lab, who is consulting with us to evaluate what needs to be done to get that off the ground. When we move into the new facility, the plan is that we have a CLIA lab open shortly after that.

For validation testing, we will likely start with a cancer panel and a Mendelian disease panel, but that's just to start. We are also working with the state for the ability to get an exome and a genome CLIA-certified in New York. That would be our long-term goal.

**How is the NYGC pilot lab interacting with the Innovation Center at Memorial Sloan-Kettering Cancer Center that is receiving four Ion Proton sequencers?**

We will have a working group to discuss how projects from the institutional founding members will get on the Ion Proton to test the machine. We will actually likely be pulling one of the machines to Rockefeller here, so that Sloan-Kettering will run three and we will run one.

But the NYGC will also have internal projects, not so much to answer scientific questions but to test the technology, whereas at Sloan, the runs will likely be tied to specific scientific projects.

**How will the NYGC pilot lab interact with the sequencing core facilities of the institutional founding members?**

The idea is a hub-and-spokes model. NYGC is in the middle, and the core facilities are out, and we are trying to collaborate closely and not necessarily directly compete with them. For example, if someone contacts us with a small project to sequence 10 exomes, the first thing we ask them is, 'Have you spoken with your core facility about this?' If they come with a project to us saying, 'We want 2,000 exomes,' oftentimes, the cores are not equipped for that throughput. There are some decent-sized core facilities here that are doing wonderful things, and we see that there is a good complement between NYGC and institutional founding members' core facilities, and it's not a direct competition.

**How does the NY Genome Center plan to distinguish itself from other large genome centers in the country?**

The all-in-one package I described, starting from sequencing all the way to computational analysis, data storage, and bioinformatics, and the ability to use our computational power, I think that model is somewhat unique. And I think the critical mass of bioinformaticians that we are going to build is the other piece. Definitely, in New York, the core mass of this bioinformatics know-how will be unique. There are other institutes [elsewhere] that have that, too, but likely not to the level that we will eventually build up to, potentially hiring 100 to 150 bioinformaticians.

**With the availability of commercial whole-genome sequencing services, is there still a need for large academic sequencing centers?**

We deal with researchers who don't always want to ship their samples off to some faraway location. The proximity, I think, is one important thing. Obviously, people are interested in cost, and we are not necessarily going to compete on cost because we feel the packages that we are putting together bring added value that others will not bring. I also think if you can go out there and produce the best quality, that actually is different than just sending off samples to someone [with whom] you potentially have less interaction. Any of the institutional founding members can come to the facility, and if they wanted to, they could see their run on the machine and walk through the QC on their samples. It's a little harder to do that with other service providers.

I think the other thing is, you can't just look at NYGC as a bunch of sequencing machines. You are getting your sequencing, you are getting data storage, you are getting your analysis, you are getting the ability to come in and use the computational power, you are getting the exposure to training courses, and you are getting the exposure to new technology: that's what NYGC is.