



Q&A: UCSD's Olivier Harismendy Compares Enrichment Methods for Targeted Sequencing

December 01, 2009

Newsletter: [In Sequence](#)
[In Sequence - December 1, 2009](#)



Name: Olivier Harismendy

Age: 33

Position: Project scientist, Moores UCSD Cancer Center and Department of Pediatrics (part of Kelly Frazer's group, Division of Genome Information Science), University of California San Diego, since 2009

Experience and Education:

Staff scientist, Scripps Translational Science Institute, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 2007-2009

Postdoctoral research associate, department of neurobiology, The Scripps Research Institute, 2005-2007

PhD in microbiology, Paris 7 University – Denis Diderot, 2004

MS in microbiology, Pasteur Institute and Paris 7 University, 2001

MS in process engineering, ENSTA-ParisTech, 2000

As a project scientist at the Moores UCSD Cancer Center and the Department of Pediatrics at the University of California, San Diego, and a former staff scientist at the Scripps Translational Science Institute, Olivier Harismendy has been exploring second-generation sequencing platforms for targeted sequencing applications ([see *In Sequence* 4/7/2009](#)).

Recently, he and his colleagues explored two methods for targeted enrichment of DNA — Agilent's HybSelect in-solution hybridization probes and RainDance Technologies' microdroplet PCR — and published their results last month in [Genome Biology](#) and [Nature Biotechnology](#), respectively.

In Sequence spoke with Harismendy two weeks ago to discuss the pros and cons of each approach, and studies to which he plans to apply them. Below is an edited version of the conversation.

How did you decide to test the RainDance and Agilent SureSelect technologies? Why did you choose these two?

At the time we started the study, people were starting to develop these targeted sequencing technologies. There were some microarray experiments described in the literature, but the specificity of the capture was not really great; it was around 50 percent or even less, meaning that half of what you elute from those microarrays is not what you want but wasted sequencing. Another technology that was published was padlock probes, or the molecular inversion probes, but they were pretty much in development and not optimized at the time.

We saw an opportunity because in our program, we always want to look at many samples, so we were really interested in increasing the throughput and, of course, the quality of the targeting.

We started to talk with Agilent, which was already developing, with Andreas Gnirke [at the Broad Institute], what they now call SureSelect, a hybrid selection method. We decided with Agilent that we would help them optimize the product, so we beta-tested their SureSelect technology.

RainDance was a very innovative approach to targeted sequencing. We have longstanding experience with PCR, and anybody aiming at targeted sequencing will start thinking about a PCR solution, but multiplexing PCR has always been a challenge. The number of primer pairs you can put in one tube is limited, and the design of these primer pairs becomes very complicated if you want to pool them. So having each primer pair encapsulated in a microdroplet was really a very interesting approach, and we wanted to test this with RainDance. We do not have their instrument in house — we helped them design the primers for the targets we had in mind, they did the merger of the microdroplets with genomic DNA and the amplification, and we did the sequencing and the analysis. In parallel, we did a manual, traditional PCR with exactly the same primer pairs in house, so in the end, we were really able to compare the performance of a traditional PCR that is done in microtiter plates and the performance of the microdroplet PCR.

From your experience, what are the strengths and limitations of the Agilent and RainDance approaches?

I think they are not for the same purpose, really. The experimental design — what you want to do with your science — is the key factor to decide which technology you want to use.

The Agilent approach is a very scalable method. It can process multiple samples simultaneously, and that's one of the great advantages over microarrays or even over the RainDance instrument. However, SureSelect, despite all the improvements that they did and the maturation of this technology, still has 50 to 60 percent specificity — it might be a bit higher in the commercial product today. But anyway, there is still some wasted sequencing that is going to be produced using SureSelect.

On the other hand, the RainDance approach is pretty much a PCR approach. Everybody knows PCR — the bias of PCR but also the advantages. And the greatest advantage of PCR is this phenomenal specificity. You pretty much have more than 90 percent specificity over the target, so everything that you want to capture is amplified.

Another difference is the success of design for the type of region you want to target. Since SureSelect is a hybridization method, it cannot target the repetitive fraction of the genome. If you are interested only in exons, there are no repetitive elements in exons, so it's not that much of a problem, and the probe design success is over 95 percent in those regions. However, if you want to target a contiguous interval of the genome to resequence — maybe a few hundred kilobases or something like that — or you want to target a 3' UTR of a gene, those elements have repeats in them, and with the strategy that SureSelect uses, which is the same as microarrays, you can't target those repetitive elements; they will be excluded from the design.

With the RainDance technology, on the other hand, because it is PCR, you have two primers, so you can always anchor one of the primers in a unique region and go through the repetitive region and still amplify that, so the targeting success will be higher with the RainDance approach, especially for large intervals.

How does the cost of the two technologies compare?

Obviously, for these two papers, it was a collaboration between the companies and our laboratory, so we were not affected by the cost.

I would assume that the cost of SureSelect is probably similar to microarrays, maybe a bit more expensive. It decreases as you order more — the ballpark is, if you order only for five capture experiments, it's probably going to be around \$1,000 per capture. However, if you order for 500, then it decreases and you end up with a couple hundred dollars per sample. For the SureSelect, you buy those libraries and the kit that goes with it, and there is no capital equipment involved.

For the RainDance instrument, it's a bit different, because you actually need to either go through a service provider who has the instrument, or you will have to have access to or purchase an instrument [which, as of February, had a list price of more than \$200,000, see *In Sequence 2/24/2009* — ed.]. The RainDance solution makes sense in a core facility, for example, where an instrument would be shared for several projects.

On top of the direct cost of a technology, you have to include all indirect costs, which is mostly the labor. The advantage of the RainDance instrument is that it's fully automated. You put a plate with your primer on one side, the plate with your sample on the other side, and it is fully automated to generate those samples. It takes probably 10 minutes hands-on per sample, plus some PCR.

On the other hand, with the SureSelect, it's a lot of pipetting and washes and hybridization, so there is a bit more hands-on time. There is the possibility of automating, obviously — implementing the SureSelect on a liquid-handling robot, and I know the Broad is developing something like that.

Another difference [in cost arises from] the specificity — there is very high specificity for RainDance, and the specificity is lower for SureSelect. In the end, it's going to impact the cost of sequencing, because for the same amount of region you target, you are going to have to sequence, probably, twice as much using the SureSelect than the RainDance.

So RainDance has more upfront costs, but maybe less indirect costs. However, SureSelect has less upfront costs but probably more indirect costs. In the end, it's quite difficult to compare them.

So what is your recommendation for someone who wants to do targeted sequencing?

If someone comes in and says he wants to do targeted sequencing, I would say, 'Show me the region that you want to target, what type of region is it, what type of DNA is it, what type of coverage do you need? Are you aiming to just get an overview of the region, or do you want to go very deep and find some very rare mutations? How many samples do you have — is it just some kind of proof-of-concept study where you look at five samples, or do you intend to do a sequenced-based association study where you look at several hundred samples?' It really depends on scale of the study and on the type of region you want to target.

What about the number of targets you can tackle with each technology?

That's also a difference. We tested, with RainDance, 4,000 amplicons — that's what they are selling for now. They intend to go to 20,000 amplicons next year. So it's a bit lower than what you can find with the SureSelect, which I believe now is offering a full exome, about 27 or 30 megabases of sequence. There is definitely a difference in scale here.

You used the Illumina GA in one study, and both the Illumina and the 454 FLX in the other — in how far does the downstream sequencing platform influence what target selection method you choose?

For now, it's not really a criterion. For deep sequencing, we really need to use a short-read instrument because it can generate way more sequences.

How are you equipped with sequencers?

UCSD has a core facility that is equipped with three Solexa [Illumina GA] sequencers, and one of them has been acquired by the cancer center and is reserved for cancer center projects. In the department of pediatrics as well, Gabriel Haddad has an additional Illumina sequencer to which we have access. And we are currently analyzing some data that has been generated on SOLiD instruments via a collaboration with Life Technologies. Since our team is new at UCSD, we don't have our own sequencer, but there is sequencing bandwidth that is available for our projects. As we gear up, we might acquire our own instrument; for now, we are just using what's available on the campus through collaborations.

Are you using both methods in projects now? Can you talk about any of these?

Yes, absolutely. We did all this work, and these two papers helped to prove to ourselves that it was doable and that we were obtaining good quality. But now we want to use these methods, of course, to address some interesting biological questions.

Kelly Frazer's group here at UCSD — which I am part of — belongs to two entities, one is the department of pediatrics at the School of Medicine, the other is the Moores Cancer Center, which is one of National Cancer Institute's comprehensive cancer centers. For

pediatric diseases, we just saw recently [a paper](#) from the University of Washington, by Jay Shendure and Debbie Nickerson, resolving one of those rare Mendelian diseases, Miller syndrome. In the department of pediatrics, that's one of the very first things we want to do. We are collaborating with a geneticist, Ken Lyons Jones at Rady Children's Hospital here in San Diego, who has a lot of experience with those Mendelian diseases. We are going to select some kindreds of children who are affected by those diseases and try to solve one of those thousands of syndromes that are still unexplained. That would be whole-exome sequencing on a few samples and trying to find coding mutations that are causing those syndromes.

In the cancer center, we are interested in two different aspects. One is maybe more diagnostic, although it can be used in research — targeting a subset of all known mutations in cancer. We would like to resequence them, and we want to go very deep. We want to be able, in one sample, to detect mutations that could be present in one to five percent of the tumor cells. Cancers are heterogeneous, especially solid tumors. From some of the studies out there, we know that, for example, cancer reoccurring in a patient can be derived from a sub-clone, a mutation that was there but that was not targeted by the therapy. We really want to interrogate those tumors at a very deep level. In order to do deep sequencing, we need to do targeted sequencing to be cost-competitive. We are going to target maybe 100 or 200 genes — we are still compiling the list of genes that we want to target — genes highly mutated in cancer.

There is another aspect in cancer. We want to target specific pathways, so it's more like a discovery approach. One pathway we are really interested in is the DNA repair pathway, where some mutations in DNA repair genes might increase the sensitivity to some drugs, for example. So if we monitor the DNA repair genes in these cancers, we could probably customize the treatment of these cancers a little more.

When will the cost of whole-genome sequencing decrease so much that it's no longer cheaper to do targeted sequencing?

That's a very good question. There will be a break-even price, and there might be several, depending on the application. If your application is just to sequence, not so deeply, the whole exome, there is going to be one day where the whole genome is going to be just cheaper. However, if you want to go very deeply, 1,000x, this time will probably come later. It really, again, depends on what you want to look at, how you want to use the technology.

Are you considering whole-genome sequencing for future studies, or is that too far off?

We are looking at some whole-genome sequencing in the cancer, but it's really more of a discovery phase, some very early projects. The question we address there will be only on a few samples now. We don't have the bandwidth to generate data for dozens and dozens of samples at the whole genome level for now.

Targeted sequencing is also targeted questions on more samples. Usually, if we use targeted sequencing, the questions are more defined, you have a precise hypothesis you are going to test, and you know targeted sequencing is the best way to test that. Whole-genome is more like a big question mark, 'Let's do everything,' like rearrangements, all mutations, coding, non-coding. But what you realize is that people who are sequencing

whole genomes for now are still focusing their analysis on the coding mutations, because there is very little we know about the non-coding mutations.

In terms of translating the sequencing into the clinic, I think it will be sooner with the targeted sequencing approach like the one I described than the whole-genome approach where it's very difficult to interpret.

Genomeweb system

These settings are generally managed by the web site so you rarely need to consider them.

Issue Order: 4

