

<http://www.genomeweb.com/sequencing/qa-mls-alex-kohlmann-targeted-diagnostic-sequencing-leukemias>

Q&A: MLL's Alex Kohlmann on Targeted Diagnostic Sequencing of Leukemias

February 15, 2012



Name: Alexander Kohlmann

Age: 36

Position: Head of next-generation sequencing and microarray department, MLL Münchner Leukämielabor, since 2008

Experience and Education:

Research alliance manager, Roche Molecular Systems, Pleasanton, Calif.; Roche Diagnostics, Penzberg; and Roche Diagnostics, Rotkreuz, Switzerland, 2005-2008

PhD in leukemia research, Ludwig-Maximilian-University, Munich, 2004
Undergraduate degree in biology, University of Würzburg, Germany, 2001

Based near the university hospital of the Ludwig-Maximilians-University Munich, MLL Münchner Leukämielabor is a leukemia diagnostic reference laboratory that serves doctors and patients across Germany. Last year, the lab, which focuses on leukemias, lymphomas, myelodysplastic syndromes, and myeloproliferative diseases, performed diagnostic tests on 45,000 samples.

Assays performed by MLL include cytomorphology, cytogenetics and FISH, immunophenotyping, and a battery of molecular-type tests, such as real-time expression analysis of fusion genes, mutational analysis, and quantitative assays. Since mid-2010, MLL has been providing routine diagnostic sequencing on the 454 platform. The firm is currently equipped with three 454 GS FLX instruments, four 454 GS Juniors, and two Illumina MiSeq platforms and recently participated in Illumina's global early-access program for MiSeq.

Clinical Sequencing News visited Alex Kohlmann, MLL's department head of next-generation sequencing and microarrays, last week to find out more about diagnostic next-gen sequencing performed and developed at the laboratory. Below is an edited version of the conversation.

Since when have you been exploring next-generation sequencing as a diagnostic method? What are the advantages over other assays?

The scouting happened early in 2009. At that time, we were investigating the short-read platforms of SOLiD and Illumina, but also the long-read platform of 454.

There are many genes in the hematological space with insertions or complicated insertions and deletions, and we saw a particular strength in the long-read technology of 454 to provide us this type of information.

Initially, we designed two proof-of-principle experiments with 454. One was a capture-based assay for the identification of unknown fusion genes, the other focused on amplicon deep sequencing. At that time, we were already thinking of placing conventional Sanger-type tests onto the 454, based on the short turnaround time of this assay, and we were also seeing the advantage of greater sensitivity of the 454 and, if the facility is equipped in the appropriate way, its high throughput.

Since when have you been applying 454 sequencing routinely in diagnostics?

After our initial proof-of-principle studies, which were published in the [Journal of Clinical Oncology](#) and in [Leukemia](#), we decided to prepare ourselves for ISO certification and accreditation. We have been accredited since June 2010, and since then, all the 454 methods for certain candidate target genes have been running in daily routine operations.

There are about 10 genes, in small sets of gene panels, that are being routinely applied to patients, both at diagnosis but also to detect minimal residual disease, which is a really powerful utility of this assay in following up patient-specific individual mutations during the course of disease, or checking the success of a transplantation.

At the moment, we perform several FLX runs a week that are complemented from time to time by Junior runs. There are certain assays that we perform on the Junior, such as the *CEBPA* gene. Because of the high GC content, as [published](#) by my colleague, this requires a different master mix in the emulsion PCR, so this is being set up in a separate solution on the Junior.

We always have a mixture of routine patients complemented by novel genes that are currently being developed that are spiked in as research samples.

About how many of the 45,000 samples your lab analyzed last year underwent next-gen sequencing?

In a typical month, we currently provide next-gen sequencing data both for diagnosis and for research purposes for about 1,000 cases. Last year, we sequenced more than 100,000 amplicons at 500-fold coverage, which makes us one of the largest amplicon deep-sequencing 454 labs worldwide.

Are your diagnostic next-gen sequencing assays reimbursed by health insurance in Germany?

In cases where we substitute standard direct Sanger sequencing, the next-gen sequencing approach is reimbursed accordingly, but in other cases, this is not possible today.

What have been the greatest challenges in implementing next-gen sequencing in routine diagnostics?

First of all, we had to build up the facilities from scratch, which take up about 350 square meters [3,767 square feet] of dedicated lab space. We have strict separation of the pre- and post-PCR areas and have different rooms for the post-PCR according to function, like breaking of emulsions, enrichment of DNA-carrying beads, and sequencing instrumentation, which requires air conditioning and temperature control.

Secondly, of course, we had to find the right staff, to train the staff, and have an efficient workflow, also in terms of the data analysis pipelines. Since we established our in-house bioinformatics group, we tried to automate as much as possible. We are not facing the challenges others might face when sequencing whole exomes or genomes because we just sequence 10 to 20 genes, and these genes are sequenced over and over again, week after week, so we can really focus on a robust and highly sophisticated pipeline to provide all the information that's needed to interpret those reads from a diagnostic perspective. Last year, we published a [toolbox paper](#) on that.

This all went along with a great learning experience in terms of different sample prep options. We were looking into providing extra levels of accuracy and robustness by using pre-pipetted plates and are working with Roche Applied Science and 454 to develop customized plates that contain primers with barcodes in a lyophilized state. This is part of the so-called IRON [Interlaboratory RObustness of Next-generation sequencing] study. We have also been looking into the Fluidigm AccessArray system very early on, and have been collaborating with Fluidigm to establish the use of that system in a routine environment.

We spent a lot of time and effort upfront in getting the sample prep robust and at least semi-automated, working with companies like Beckman Coulter to develop robots and platforms for PCR

product purification, and with the REM module from Roche/454 for bead-enrichment.

On the back end, data analysis-wise, we worked with JSI Medical Systems to find the right software for our needs to interpret the results from a functional relevance perspective to create meaningful diagnostic reports. None of the software out on the market — neither from Roche/454 nor from other vendors — really has a complete package that includes everything from interpreting the raw data, generating quality reports, monitoring the quality and the efficacy of an experiment, but also, at the end, to allow a diagnostic interpretation. We follow the principle of different people looking at the same result, and at the moment, also different software programs interpreting the same result, so there is an extra layer of security, accuracy, and robustness of our data.

Can you briefly talk about the IRON study?

It came out of a discussion at a meeting we held in May 2010. The first idea was to look into the robustness of next-gen sequencing, because at that time, a lot of labs showed interest, the Junior was just about to be launched, and the question was how 454 could find its place in daily routine operations in clinical labs.

We headed the study, which was conducted in the summer of 2010. We decided to conduct replicate analysis in 10 labs from three continents across the world. We shipped out 18 anonymized and blinded specimens, and the labs, just trained by a web protocol, then performed one sequencing run using the specimens they got from us. The results were [recently published](#) in *Leukemia*, one of the earliest studies to demonstrate the robustness of 454 sequencing in a clinical environment. The selected study participants had a strong background in molecular technologies and clinical technologies. They all see patients at diagnosis and included labs that are national reference laboratories for leukemia study groups.

Since this was a very successful project, we are moving now into a second phase, where the study will be open again to additional sites. We are gathering more than 25 labs from 14 countries at the moment, and we are creating five distinct working groups across the different myeloid and chronic lymphatic disease types that will be looking at specific scientific questions. Roche, again, will develop content together with us and the IRON study participants, so there will be an “assay-on-demand”, a customized 96-well plate according to our input where certain genes, or panels of genes, are being prepared for sequencing by distinct working group members. This has technically started now; we had a kick-off a few weeks ago at the European LeukemiaNet conference, which is one of the important networking conferences in our field. At that meeting, we decided to disclose the plate information to the participants, and we have now opened the study for sequencing patient samples.

What other next-gen sequencing platforms, besides the 454, are you looking into?

We are quite aware of how rapidly the field is evolving. It's really a tremendous time to be part of at the moment, especially in next-gen sequencing. Since we scouted the technologies back in 2009, the development has been marvelous. Seeing sequencing costs for the whole genome coming down to about \$1,000 later this year is astonishing, and I'm confident it will soon find its place in clinical applications.

But for the mid term, and we are talking now for the next two to three years maybe, we still see a situation where sequencing of panels, a limited number of genes, will be the most promising approach. But we cannot just add more and more genes into the system. We need to have a rationale for why a certain gene is sequenced. So we are always scouting both on the sample prep front and on the sequencing back end. That's why we have, for example, a limited testing phase with RainDance to look into creating single-plex PCR-based droplets. We are looking into a 31-gene panel at the moment on the RainDance platform. We were also part of Illumina's early-access

program for the MiSeq and the custom TruSeq assay that Illumina has developed.

So we are always trying to develop the field and be ahead of the field when it comes to developing diagnostic utility.

Any feedback on your initial experience with the MiSeq?

That's something we cannot disclose at the moment, but it looks promising.

When do you think there will be a switch from targeted amplicon sequencing to whole-exome or maybe even whole-genome sequencing for diagnostic purposes?

We are certainly aware of the price drops for enriching whole exomes on either the SOLiD or Illumina platforms, and maybe the [Ion Torrent] Proton soon. That's really incredible how prices have come down. However, in enrichment data we have seen so far, even in high-ranked publications, the coverage is not 100 percent, and from the diagnostics perspective, if I'm a patient, and I'm asking a doctor or a diagnostic lab to sequence a certain number of genes, I want to have 100 percent accuracy — that 100 percent of a certain target coverage has been achieved, so that a really meaningful diagnostic report can be made at the end.

The problem we see at the moment with all these enrichment technologies is that they are far away from giving us the same highly reliable amount of information that amplicons do at the moment. Of course this is all undergoing constant change, and it all might look different a year from now, but for the moment, we think the intermediate will be larger gene panels. Then, at some point, it might make sense just to limit the analysis of whole-exome sequencing to a few candidate genes that are relevant, and the rest is already being prepared and can be interpreted when clinical follow-up data becomes available.

For this, turnaround time will also be critical. Our current 454 workflow provides meaningful results within four working days; this cannot be achieved with whole-exome sequencing today on any platform because we need to enrich first, we need to sequence a large amount of bases, and we need to go through this large amount of data, and the pipelines are getting even more complicated. From the bioinformatics perspective as well, handling amplicon data, since it's the same genes and the same few amplicons over and over again, can be automated very nicely, but automating whole-exome data can be very challenging. Algorithms are constantly being refined for looking for insertions or deletions. What also makes whole-exome sequencing very challenging is just the sheer amount of SNPs or copy number variants that are just intrinsic in any patient that is sequenced. This requires sequencing the non-leukemic material, also.

Can you talk about MLL's expansion plans, and how that will affect your next-gen sequencing facility?

The team right now encompasses 17 people who are directly involved with generating sequencing data, and an additional three or more biologists generate the diagnostic reports at the end. This number will certainly increase as of this summer when we open a new sequencing facility that will double the lab space. In this, we want to combine CGH arrays with the sophisticated next-gen workflow. We are currently evaluating both RainDance's and Illumina's technologies, and we will certainly look at even more platforms, like nanopore sequencing, as they become commercially available.

Is there anything else you'd like to mention?

I think what makes it so interesting to work with this technology today is that we screen more and more genes per patient because patients and doctors want to know, and we provide not only information for classifying a disease, but we can really provide prognostic information. Since we screen more and more genes, we will find more and more patient-specific mutations. And this will give us, for the first time, additional knowledge that we can offer to a patient to follow and monitor his personal disease in an individual manner over the course of treatment.

This is something that has not been there in the past. For sure, there were specific mutations in genes like *NPM1* or *FLT3*, or the fusion genes that happen very often in leukemias, but we need, for many of these patients, additional information that, for the first time, can now be generated in a quantitative manner by using the deep sequencing approach.

So we are really entering a new era, where 454 long-read deep sequencing is something that's adding to the diagnostics. It's not replacing anything but it's adding and gives new and unprecedented information to patients and doctors.