Unmasking the Viral Etiology of Cancer and Immune Disease

Karolinska Institutet researchers use Illumina sequencers to identify novel HPV types associated with non-melanoma skin cancers.

Introduction

Despite decades of research, the causes of cancer seem to many to be as mysterious and unresolved as ever. Yet, there are certain types of cancer where the drivers of malignancy have emerged into stark focus. Scientists know that human papillomavirus (HPV) causes 99% of cervical cancer cases[1]. HPV is a large DNA virus that establishes infections in keratinocytes of the mucous membranes, as well as the skin where it causes squamous cell papilloma or warts. If HPV is responsible for most cervical cancer cases, could it be responsible for certain skin cancers?

The strong link between HPV and cervical cancer prompted Emilie Hultin, PhD, to wonder if there was a connection between HPV and nonmelanoma skin cancer. A geneticist at the Karolinska Institutet in Sweden, Dr. Hultin's HPV research has spanned the transition from microarrays to next-generation sequencing (NGS), revealing novel HPV types present in nonmelanoma skin cancer. Relying on Illumina HiSeq®, MiSeq®, and NextSeq® Systems, her research underscores the ability of NGS to unravel some of medicine's biggest mysteries.

Unlocking HPV Secrets

Dr. Hultin began her science career studying chemical engineering at the Royal Institute of Technology (KTH) in Stockholm. During her PhD studies at KTH, she developed new technologies and methods for array-based SNP genotyping and HPV typing. After a 2-year sojourn working in industry, Dr. Hultin moved to the Karolinska Institutet in 2011. She is currently a research coordinator in the lab of Joakim Dillner, MD, PhD, where she runs the group's deep sequencing lab.

Drs. Hultin and Dillner are interested in identifying viruses associated with different human diseases, especially cancer. Their work has focused mainly on detecting causal links between HPV and cervical cancer. In the genital tract, certain types of high-risk HPV viruses are much more likely to turn cervical cells cancerous, whereas low-risk HPV viruses are limited to causing genital warts. Because HPV can also infect skin, they decided to investigate whether the HPV-cancer link might also include nonmelanoma skin cancers. They decided to exclude melanoma from their studies because of its strong association with sun exposure.

"We knew that there were many HPV strains on the skin, but weren't sure whether they could cause nonmelanoma cancers," Dr. Hultin said. "Establishing the presence of a link required identifying small amounts of HPV DNA in skin samples. That was not an easy task."

Because they didn't know what they were looking for, they used NGS for their studies. "Microarray technology is directed toward known DNA sequences, so we couldn't use it to identify novel viruses," Dr. Hultin stated. "NGS enabled us to sequence all the DNA in skin samples, making it possible to detect novel HPV types."

Dr. Hultin's arrival in Dr. Dillner's lab coincided with their first purchase of an NGS sequencer, the Roche 454 System. The team later used a core facility's Thermo Fisher Scientific Ion Torrent sequencing system. The team soon realized that neither system provided the sequencing depth to find novel viruses. "The viral signals were being swamped out by the overwhelming amount of skin cell DNA in the samples," Dr. Hultin said. "Only about 0.2% to 0.5% of all the sequencing reads were actually viral DNA—the rest was human. We also wanted a sequencing system that could multiplex, enabling us to sequence cost-effectively. It's too expensive to run one sample per sequencing run."

MiSeq System Uncovers HPV Diversity in Skin

The MiSeq System offered the perfect alternative, meeting their deep sequencing requirements, and capable of multiplexing 96 samples per run. They conducted a head-to-head comparison of the MiSeq platform, 454, and Ion Torrent systems to see how each system performed. The MiSeq System delivered, providing more reads and a 1,000 fold more sequencing depth than the other systems. Yet, the study was much more than a sequencer showdown. Dr. Hultin and her team used the systems to conduct a set of experiments looking for novel viruses in 142 different skin samples from immunosuppressed individuals with actinic keratosis, basal, and squamous cell carcinomas.[2]

Dr. Emilie Hultin is a Research Coordinator at the Karolinska Institutet in Sweden.

Study samples consisted of whole genome amplified and PCR amplified sample pools. “There’s an advantage in sequencing unbiased and biased sample pools,” Dr. Hultin said. “When we sequence unbiased whole genome amplified material we are able to detect more HPV types, but we also get a lot of human reads. PCR amplified samples enable us to obtain more HPV-only reads. However, we’ll miss some HPV types that are not complimentary to the primers we are using.”

The 2 sample pools required the use of 2 different Illumina library prep kits. “We used the Nextera® DNA Library Preparation Kit for the 82-sample whole genome amplified pool and were able to prepare libraries from all these samples in just 1-2 days,” Dr. Hultin said. “It was convenient to analyze a full plate at the same time as we prepared the samples.”

> “The MiSeq System identified HPV types that were not found using the 454 and Ion Torrent systems.”

“We also had samples that were amplified with specific PCR primers, creating amplicons of about 450 bp,” Dr. Hultin added. “That was the perfect size, so we used the TruSeq® Nano DNA Library Preparation Kit for these samples. We eliminated the fragmentation, end-repair, and size selection steps so we could ligate the sequencing adapters to the ends of the PCR products immediately. The Illumina library kit protocols and MiSeq workflow reduced our hands-on time in the lab dramatically.”

The study found an unexpected level of diversity, with no less than 396 HPV types, containing 229 putative virus types that had never before been identified. In contrast, a sequencing study performed earlier with the 454 and Ion Torrent systems identified 273 different HPV types, and only 47 unknown viruses.

“We thought that we were going to find a few novel HPV types, because there hadn’t been many studies of skin HPV,” Dr. Hultin said. “We didn’t expect to find such extreme diversity of cutaneous HPV types. The MiSeq System identified HPV types that were not found using the 454 and Ion Torrent systems.”

After 2 years using the 454 and Ion Torrent systems, Dr. Hultin made the switch to the MiSeq System, enabling her to make greater strides in her research. “It wasn’t until we switched to the MiSeq System that we had the sequencing depth that we needed, yielding data that enabled us to identify novel HPV types, Dr. Hultin said. “The results of that study marked a fundamental change in our understanding of HPV diversity.”

Short-Read Sequencing Identifies Novel HPV Type

The team continued their research, sequencing serum, fresh frozen (FF) biopsies, and formalin-fixed paraffin embedded (FFPE) skin samples. Because the DNA in FFPE samples is highly degraded, they thought that longer reads would be better in identifying novel HPV types. But as they worked, they found that shorter paired-end sequencing (151 bp) actually gave them the best results for FFPE samples.

Using the HiSeq and MiSeq Systems, Dr. Hultin led a follow-up study that looked for HPV in 91 skin samples that were a combination of FF biopsies and FFPE samples from keratoacanthomas, actinic keratosis, and basal cell carcinomas.[3] Multiplex, paired-end sequencing on the HiSeq System using dual indexes enabled them to sequence the samples using a non-biased approach. They found 4 previously unknown putative HPV types, including HPV197, which was highly present in the different cancer samples. HPV197 is only 75 percent identical to the most closely related type of HPV. Its primer regions possessed many nucleotide differences that would have prohibited PCR amplification.

“HPV197 would never have been discovered without using an unbiased method,” Dr. Hultin said. “This virus is abundant among these nonmelanoma skin cancer samples, so it seems likely that it is a very important skin HPV type.”

A Shift to the NextSeq 500 System

In June 2014, Dr. Hultin added a NextSeq 500 System to the lab and has been thrilled with its performance. Regardless of whether they’re doing PCR amplification or not, Dr. Hultin’s team has found that the NextSeq 500 System is perfect for their studies. “The NextSeq 500 System enables us to sequence FFPE samples effectively, and assemble novel viruses from FF biopsy samples,” Dr. Hultin said. “It offers greater sequencing depth and an ability to sequence pools of 96 samples. We’ve shifted primarily to the NextSeq 500 System for our studies, using the MiSeq System for confirmation or pilot studies.”

The team’s bioinformatics pipeline is optimized for paired-end sequencing data. Given that viruses like HPV can integrate themselves into the host genome, paired-end sequencing with the NextSeq 500 System enables Dr. Hultin to identify locations where the virus has integrated. “We can look at the pairs and see if one pair is aligned to the virus and the other pair is aligned to the human genome,” Dr. Hultin said.

> “HPV197 would never have been discovered without using an unbiased method.”

Dr. Hultin is thrilled with the performance of the NextSeq 500 System and how easy it is to use. “The MiSeq System training was easy and we couldn’t believe that sequencing could get any easier,” Dr. Hultin said. “Then we were trained on the NextSeq 500 System and realized that Illumina had indeed made sequencing a bit easier. We almost didn’t need any training at all.”

Viral Etiology in Cancer and Immune System Diseases

NGS has opened up a window into the number of HPV viral types that exist. “A couple of years ago, there were about 100 different HPV types known,” Dr. Hultin stated. “Now more than 200 types have been identified.”

Their success with HPV has enabled the researchers to look more broadly at the role of viruses both in cancers, and in diseases and disorders that impact the immune system negatively. Nobel

laureate Harald zur Hausen, MD, who first discovered the role of HPV, has proposed that immunosuppressed individuals are more likely to get cancers that are caused by viruses, because their own immune systems are no longer able to keep the infections in check. An example is AIDS, where patients are at much higher risk of nonmelanoma skin cancers such as Kaposi's sarcoma.

> "The NextSeq 500 System enables us to sequence FFPE samples effectively, and assemble novel viruses from FF biopsy samples."

Dr. Hultin and her team have been looking for viral signatures in cancers associated with immunosuppression in the hopes of finding a potential viral link. They are also conducting a study on multiple sclerosis (MS) to see if there's evidence of a viral etiology. Dr. Hultin has obtained serum samples from MS patients taken several years before their diagnosis to identify viruses that could be linked to MS. The first NextSeq 500 System sequencing runs have been completed and the group is working on the confirmation phase.

Although the role of HPV in cervical cancer has become clear, scientists are just beginning to understand the role of HPV and other viruses in the development of cancer. Dr. Hultin believes that many more types of cancer might have a viral etiology, even if viral infection doesn't automatically cause cancer down the line. With the help of NGS, she believes she will ultimately find these signatures.

References

1. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol.* 1999;189:12–19.
2. Bzhalava D, Mühr LS, Lagheden C, et al. Deep sequencing extends the diversity of human papillomaviruses in human skin. *Sci Rep*. 2014; 4: 5807.
3. Arroyo Mühr LS, Hultin E, Bzhalava, et al. Human papillomavirus type 197 is commonly present in skin tumors. *Int J Cancer,* 2015; 136(11): 2546–2555.

Learn more about the Illumina products and systems mentioned in this article:

- MiSeq System, www.illumina.com/systems/miseq.html
- HiSeq System, www.illumina.com/systems/hiseq_2500_1500.html
- NextSeq System, www.illumina.com/systems/nextseq-sequencer.html
- Nextera DNA Library Preparation Kit, www.illumina.com/products/nextera_dna_library_prep_kit.html
- TruSeq Nano DNA Library Preparation Kit, www.illumina.com/products/truseq-nano-dna-library-prep-kit.html

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

 Pub. No. 770-2015-023 Current as of 15 October 2015

illumina®