Deciphering the Role of Long Non-Coding RNA in Cancer

Researchers are using RNA-Seq to reveal how lncRNAs could be used to identify, measure, and treat cancer.

Introduction

Professor Jo Vandesompele, PhD has a particular affinity for RNA. It began with his PhD thesis research where he used reverse transcriptase quantitative PCR (RT-qPCR) to study the role of RNA in neuroblastoma. He's been hooked on RNA ever since. It's the focus of his functional cancer genomics group at Ghent University and of the company he founded in 2007, Biogazelle, which offers RNA analysis and experimental design services.

"I think RNA is a fascinating molecule," Professor Vandesompele said. He became laser focused on a specific type of RNA, long non-coding RNA (lncRNA) after hearing a presentation about it at a conference. lncRNAs are thought to be responsible for the large number of specialized cells in the human body and orchestrate how a cell is organized transcriptionally. Certain cancer cells appear to be addicted to the presence of certain lncRNAs, and die when the transcript is silenced. "We immediately recognized the power of lncRNAs and began performing studies using different library prep methods to study these transcripts and their role in various cancers," Professor Vandesompele added. "Our teams at Ghent University and Biogazelle are recognized as world experts in lncRNAs."

iCommunity spoke with Professor Vandesompele about the value of lncRNAs in understanding cancer versus focusing on DNA markers alone, and how they could be used as targets for developing new cancer diagnostics and therapeutics. We also discussed how they are using the NextSeq™ 500 System and TruSeq™ RNA Exome to identify RNA markers in highly fragmented samples, such as formalin-fixed, paraffin-embedded (FFPE) tissue samples and liquid biopsies.

Q: What triggered your interest in RNA and, specifically, lncRNA?
Jo Vandesompele (JV): I became interested in RNA when I conducted neuroblastoma studies in grad school. I wanted to access the cells of origin and they were rare cells. Northern blots and microarrays required too much RNA, so I used RT-qPCR, which was a novel method at the time.

I was at a Keystone meeting in 2009, when I first learned about lncRNAs. At that point, only about 1700 lncRNA sequences had been identified. We recognized the power of lncRNAs as biomarkers and designed qPCR assays to quantify them. Through our efforts and those of researchers worldwide, the lncRNA database grew to 4000, and later to 18,000 sequences.

Q: What technologies have you used to study lncRNAs?
JV: As the lncRNA database grew, it became clear that qPCR wasn't an efficient assay to use anymore. We designed a microarray and then transitioned to total RNA-Seq using different library prep methods. Not many researchers were performing total RNA-Seq at the time and that's still true today. Most researchers rely on polyA RNA-Seq, even though it only captures RNA with polyA tails. We estimate that half of lncRNAs do not have such tails.

Q: What is the role of lncRNA in the human body?
JV: We are just scratching the surface in understanding the role of lncRNA in the human body. As we perform experiments, one of the themes that is emerging is the extreme tissue and cell type specificity of lncRNA. It seems that lncRNAs might explain the complexity of human life. Although flies, tomatoes, and humans are quite different, they have almost the same number of protein coding genes. Yet, humans are far more complex with many specialized cell types throughout the body. Researchers once thought this was due to alternative splicing or posttranslational protein modifications. That might be true to some extent. However, it has become clear that lncRNAs play an instrumental role in orchestrating how a cell is organized from a transcriptional regulation perspective. We think that lncRNAs act as glue, either bringing proteins to the DNA, or attaching proteins to RNA, or glueing different RNA molecules together into a scaffold. Another possibility is that they titrate important molecules away from where they're not supposed to be or bring them to a location where they need to be.

Jo Vandesompele, PhD, is a Professor at Ghent University, and Co-Founder and Chief Scientific Officer at Biogazelle.

Q: What could lncRNAs tell us about the status of a cell?
JV: The expression specificity of lncRNAs can be exploited to uncover a cell's status. There are about 50,000 lncRNAs identified to date and that number is growing. lncRNAs are very specific in time and location of expression, with only a tiny fraction expressed in any given cell type. Once we figure out the code of when and where a specific lncRNA is expressed, we can identify the cell type and the status of the cell, such as whether it's healthy or diseased. We can learn a lot by studying these cell type specificities.

> "The expression specificity of lncRNAs can be exploited to uncover a cell's status."

Q: How could lncRNAs be used as biomarkers for diagnostic and therapeutic development?
JV: If a lncRNA is specifically expressed in a certain cell or tissue, it's logical to assume that it must have an important function in that cell. It turns out that some cells, particularly certain cancer cells, are addicted to lncRNAs being present. One of the first lncRNAs that we identified was SAMMSON, which is only expressed in skin melanoma. Together with collaborators, we found that silencing SAMMSON disrupts certain mitochondrial functions and kills the cancer cells.[1]

SAMMSON is one of hundreds of cancer cell type–specific lncRNAs, where if we remove it, the cancer cell is unhappy and either no longer proliferates or dies. We can perform rational drug design against lncRNAs to develop therapeutics and at the same time have the companion diagnostic. If we know the lncRNA sequence and can design assays to measure it, we can identify the disease or cancer cells that are addicted to these lncRNAs and shut them down using silencing technologies. There are CRISPR- and siRNA-based approaches in development. Antisense nucleotides are the most advanced and provide a promising approach to developing lncRNA-targeted therapies.

lncRNAs offer promise in the fields of oncology, neurology, and other disease types. There are a handful of antisense drugs on the market, such as inotersen, nusinersen, eteplirsen, and mipomersen, and more than a hundred clinical trials evaluating antisense oligonucleotides. Such drugs are easy to administer and act by reducing RNA or modifying splicing.[2] We're at the beginning of a completely new type of medicine.

Q: What type of lncRNA studies are you performing at Ghent University?
JV: While we initially focused on a childhood cancer, we realized that our technologies could be applied to all cancer types. We're now focused on about a dozen cancers, including prostate, melanoma, ovarian, esophageal, lung, and neuroblastoma. We're trying to be as broad as possible and are working with expert collaborators and key opinion leaders to advance these studies.

Q: What is the makeup of the Biogazelle customer base and what services do you provide?
JV: We work mainly with pharmaceutical, biotechnology, and clinical researchers. We provide an A to Z breadth of services, from RNA biomarker discoveries through customized RNA quantification studies, to performing high-throughput antisense oligonucleotide screenings, identifying, measuring, and silencing lncRNA targets, and assisting customers with lead drug discovery and preclinical studies.

We don't just deliver data to our customers. We assist them in thinking through projects, providing insights and results that are actionable so that they can proceed. We have PhD-level project managers that are involved in all project discussions and assist with experimental design. We apply our methods in a quality-controlled environment, are ISO 17025 accredited, and can work under Good Clinical Laboratory Practice (GCLP) standards, which is very important for our pharmaceutical customers.

We also have a research department where we interface a lot with Illumina. The team is focused on developing new RNA quantification methods that enable our customers to do things that were not possible before, especially in the liquid biopsy space.

> "Most researchers rely on polyA RNA-Seq, even though it only captures RNA with polyA tails. We estimate that half of lncRNAs do not have such tails."

Q: When did you begin using Illumina NGS systems?
JV: Our first NGS system at Ghent University was a Roche 454 System. We only used it for DNA analysis. People complained that it was difficult and expensive to use. Around the same time, we purchased a Genome Analyzer™*II* System. In 2014, the diagnostic laboratory in our center got a MiSeq™ System. They were very happy with it and thought it was easy to use and very affordable. At about the same time, the Biogazelle team purchased a NextSeq 500 System to perform RNA-Seq. Since then, we've acquired a second NextSeq 500 System.

Q: What types of samples are you analyzing?
JV: Most of the samples that we analyze are highly fragmented cancer samples, such as FFPE tissue samples and liquid biopsies.[3] TruSeq RNA Exome is a great solution to address FFPE and liquid biopsy samples that combine the challenges of ultra-low input and highly fragmented RNA. It converts total RNA into template molecules of known strand origin, followed by sequence-capture of coding RNA. Other total RNA prep methods consume many reads on introns that are not always meaningful and generate lower quality data.

Q: After you process the sample, how do you decide whether you have sufficient data?
JV: At every step, we perform quality control (QC). At sample reception, we review the volume to make sure it is what the customer states on the label. For FFPE samples, we perform a quantification in terms of purity and integrity. Unfortunately, for liquid biopsies we can't access integrity or purity initially because the levels are way too low. The first time that we can verify these for liquid biopsies is when we have precapture cDNA to look at the prepurified libraries on a fragment analyzer. This can be assessed with any microfluidic device and more than 90% of the time it works. Sometimes we have to repeat the analysis, so we try to make sure that we have sufficient RNA.

After enrichment and library purification, we perform another QC in terms of concentration to make sure that we have equimolar pooling. After those QC steps are passed, we perform RNA-Seq on the NextSeq 500 System to obtain high-quality data.

It's very important to generate high-quality data. What many service providers don't realize is that batch effects occur in sequencing. To avoid them, we've learned to scramble and randomize the samples. When the customer sends us samples, there's always a specific order. So, we randomize the samples at sample reception, during extraction, and during library prep, to eliminate batch effects. We also make sure that we have quality- and batch-controlled reagents so we that we don't introduce batch effects. From a sequencing standpoint, we try to expand on the number of indexes to have as many samples as possible. For almost all our libraries, we now have up to 96 indexes and for some applications (eg, 3'-end sequencing) even higher.

> "If we know the lncRNA sequence and can design assays to measure it, we can identify the disease or cancer cells that are addicted to these lncRNAs and shut them down using silencing technologies."

Q: Are you using RNA-Seq to identify mutations present in the sample?
JV: We would love to identify the mutations present in a cancer specimen, as well as splice variants and gene fusions, but many people are conservative and want us to perform only gene-level quantification.

It's a bit unfortunate, because they're not tapping into the richness of RNA-Seq data. For example, alternative splicing is easy to detect. Also, there's now a novel type of splicing called circular RNA (circRNA) that originates through back-splicing. Three recent papers document the potential of circRNA as therapeutic targets or novel biomarkers.[4-6] CircRNA cannot be identified with classic polyA sequencing because they don't have a polyA tail and that's probably why people haven't paid attention to them.

We are pushing the technology to see if we can also identify other mutations in RNA-Seq data. Customers understand the value at the DNA level. It's a different ball game at the RNA level. For example, RNA editing is a confounding, but interesting, phenomenon. I think we are just beginning to understand the importance of RNA editing in disease biology. However, the dynamic range of RNA is a problem as the coverage depends on the expression levels of each gene. If we want to detect low frequency mutations with a decent level of sensitivity, we need 30× coverage for the low-abundant RNAs. That eats up all your sequencing reads with the top abundant genes.

> "RNA editing could be valuable in assessing tumor mutation burden (TMB)."

Q: What are the possible applications of RNA editing?
JV: RNA editing could be valuable in assessing tumor mutation burden (TMB). TMB measures all mutations carried by cancer cells and is important for predicting the response or lack of response to immune-oncology therapies.

Researchers have been studying TMB at the DNA level, probably because it's so easy. While it's a challenge with RNA, it's doable and valuable. Tumor cells with high TMB have more neoantigens, which can be recognized by the immune system to elicit an antitumor response. Nonsynonymous mutations that are expressed, and hence identified in RNA-Seq data, might result in neoantigens, with RNA editing as an additional source of neoantigens.

We have preliminary data that shows that there is strong correlation between what we refer to as expressed TMB (eTMB) and the DNA-based TMB score. We're working with partners to perform large-scale studies with clinical data to show that if we combine expressed variants of the neoantigens resulting from RNA editing, we might have a better predictor for response to checkpoint inhibitors. Assay sensitivity won't be as important because the TMB score is the number of nonsynonymous mutations per sequenced megabase of protein coding genes. The current TMB liquid biopsy methods look at only a few hundred genes. With RNA-Seq, we could look at many more genes, resulting in richer data. We see significant potential in exploiting RNA for this type of analysis.

Q: Does Biogazelle also assist companies with preclinical studies?
JV: While we apply RNA sequencing in our discovery programs, we switch to qPCR for the development of diagnostic assays for use in preclinical studies to triage subjects and to act as a companion assay. In one of the development phases, we use the

same samples that we used for our RNA-Seq discovery studies to make sure that we can replicate the findings using qPCR as an orthogonal validation method. Most of our customers and partners want the clinical grade assay and often qPCR is good enough. That may change in the future, but today there aren't too many clinical assays for RNA using sequencing. Using qPCR, we can create a quality-controlled, clinical-grade diagnostic kit.

> "There are significant growth opportunities for RNA analysis in the liquid biopsy space and outside of the field of oncology."

Q: What types of data analysis do you perform?
JV: We first perform differential gene expression analysis comparing samples or sample groups. We then move beyond this classic way of looking at RNA to assess gene signatures using two advanced bioinformatic methods. The first is gene set enrichment analysis using the Broad Institute's gene set enrichment analysis algorithm. It can rank the genes by fold-change differences or P value. It is a robust and sensitive computational method to discern subtle patterns. It can reveal gene set or pathway-level information that enables us to understand what's going on in the sample group.

The second method is deconvolution analysis to estimate the contribution of the different cell or tissue types that are present in the RNA sample. A tumor is mostly made up of cancer, stromal, and immune cells. Using deconvolution methods, we can tell which cell types are in a tumor and at what proportion. We are also applying this method to fluids, although success depends on the quality of the gene signatures that are characteristic of cell and tissue types. However, the results of deconvolution in fluids have been revealing. We see various cell type signatures in plasma, urine, and cerebrospinal fluid (CSF) that are meaningful and show changes over time. These RNA signatures could be used to demonstrate whether a cancer is present, has returned, or is in remission. I expect that some important discoveries will be made using these advanced analysis methods.

Q: What data do you provide to customers?
JV: The data that we provide to customers depends on the question that they need answered. Typically, they want gene-level quantification data. However, requests for data on splice variants, mutation analysis, RNA editing, and circular RNA data are increasing. Customers receive raw and processed data, together with a results report.

Q: What is the value of RNA-Seq versus DNA sequencing data for disease assessment?
JV: I believe that RNA-Seq data provides the same, if not more, informative data as DNA sequencing. In addition to gene expression analyses, we can determine copy number variations[7, 8], mutations[7], and RNA editing, and perform T-cell receptor profiling[9] and immune cell profiling.[10] The biomarker potential is also significantly greater if you look beyond protein-coding regions into lncRNA, with 50,000 potential genes in lncRNA compared with 20,000 protein-coding genes. Ultimately, the value might be in looking at RNA and DNA signatures together to assess a cancer or disease sample.

Q: How do you think liquid biopsy testing will evolve in the future?
JV: Liquid biopsy testing based on DNA analysis is growing rapidly for oncology applications. I just spoke at the 2019 Precision Liquid Biopsy Conference and hope I was successful in communicating the value of RNA analysis for liquid biopsies. DNA alone is not dynamic or responsive enough. At best, ctDNA says something about the cancer cells and the mutations present. It says nothing about how the host is responding to the tumor or how the immune system or stromal cells are interacting with it. We can see these signals with RNA analysis, including what's occurring with vesicles and platelets. There is a new field focused on tumor-educated platelets and new therapeutic approaches are being developed around the use of purified vesicles for treating various types of human diseases such as cancer or heart disease. There are significant growth opportunities for RNA analysis in the liquid biopsy space and outside of the field of oncology.

Q: What are the next steps in Biogazelle's growth?
JV: We want to grow the Biogazelle customer base beyond Europe and continue to provide unique services. We are more than just an RNA-Seq service provider. We design experiments and conduct extensive bioinformatic analyses that provide our customers with insight on the data and support their next steps in developing valuable diagnostics and therapeutics. We just need to spread the word.

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

NextSeq 550 System, www.illumina.com/systems/sequencing-platforms/nextseq.html

TruSeq RNA Exome, www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-rna-access.html

Read more about RNA-Seq workflows at Biogazelle:

Biogazelle Case Study, www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/vandesompele-biogazelle-case-study-nextseq-lncrna.html

References

1. Leucci E, Vendramin R, Spinazzi M, et al. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature*. 2016;531:518-522.
2. Blokhin I, Khorkova O, Hsiao J, et al. Developments in lncRNA drug discovery: where are we heading? *Expert Opin Drug Discov*. 2018;13:837-849.

3. Biogazelle. Expertise/Liquid biopsies. www.biogazelle.com/expertise/liquid-biopsies. Accessed June 17, 2019.
4. Vo JN, Cieslik M, Zhang Y, et al. The Landscape of Circular RNA in Cancer. *Cell*. 2019;176:869-881.
5. Chen S, Huang V, Xu X, et al. Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell*. 2019;176:831-843.
6. Smid M, Wilting SM, Uhr K, et al. The circular RNome of primary breast cancer. *Genome Res*. 2019; 29:356-366.
7. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet*. 2013;93:641-651.
8. Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352:189-196.
9. Li B, Li T, Pignon JC, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet*. 2016;48:725-732.
10. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453-457.

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