The Emergence of RNA Diagnostics

Emergence of RNA Diagnostics

In any industry, there is an inherent tension between the cost of a product, the speed at which it can be delivered, and its quality. Sometimes known as the Triple Constraint or the Iron Triangle, the principle suggests that only two of those attributes can be selected at any given time. This raises an interesting predicament in the healthcare industry, particularly in the diagnostic sector, where rapid, accurate and accessible tools are needed to diagnose and respond to human disease. More specifically, is it possible to rapidly diagnose a condition with high accuracy and at low cost? Recognizing, of course, that both speed and cost are somewhat relative measures.

A recent example of this problem can be found in the news with the rapid spread of the Zika virus. Private companies and government health agencies worked frantically to provide diagnostic tools to those potentially infected. Fortunately, the technology for detecting the RNA genomes of viruses such as Zika is not only well established, it is becoming faster and cheaper while providing high levels of accuracy. A team of synthetic biologists from Harvard recently unveiled a test that could check for a specific strain of Zika in three hours.

RNA provides a remarkably powerful avenue for detecting not only viral infection, but many other health conditions as well. Thanks to advances in genomic technology over the past decade, the field of RNA diagnostics is reaching a point where it arguably fits the "better, faster, cheaper" ideal.

At IQuity, our interest is in giving healthcare providers information found in blood cells so that they can efficiently diagnose autoimmune disease.

Overview of Genetic Material

It's worth stepping back for a moment and reviewing where RNA is situated in the biochemistry of cellular processes. The Central Dogma of molecular biology presents a three-step process where DNA makes RNA makes protein. While findings over the past decade have added many caveats to this workflow, the essential points of this model hold true. DNA is the permanent code that contains all the information necessary for an organism to develop, grow and function. Depending on a cell's needs and the relevant signals, stretches of DNA are selectively transcribed into the intermediate molecule, RNA. Transcription is typically thought of in the context of genes that ultimately give rise to proteins.

Emergence of RNA Diagnostics

However, there are many "non-coding" regions of DNA that are now known to be transcribed as well. While these non-coding regions do not result in the production of a functional protein, these molecules often function in gene regulation influencing other RNAs and even proteins. In the case of canonical protein-coding genes, once an RNA molecule has been transcribed, it is transported out of the cell's nucleus, where it is then translated into protein. These proteins go on to carry out their programmed function inside or outside of the cell.

For decades, particularly in the '90s and '00s, a special focus within the scientific community was on DNA and unlocking humanity's genetic code. With the completion of the first working draft of the human genome in 2001, the scientific community believed that it had the information needed to understand both health and disease with the discovery of ~20,000 protein-coding genes encoded on the 23 human chromosomes. It quickly became clear that this was not the case. For one thing, long stretches of DNA, which were thought to be junk, turned out to encode many non-coding RNA molecules mentioned above. Additionally, studies in the field of genetics found that small mutations in protein-coding genes didn't always provide answers as to the etiology or progression of disease.

In the end, it all came down to how the whole process was regulated. Understanding this took more than a list of 3 billion human DNA base pairs, as originally presented. While DNA provided the information, a follow-up area of investigation sought to understand what a cell did with this complex DNA code. To figure that out, scientists actively sought answers one step down the line at the level of RNA. As it happens, the amount of DNA is static, but the non-static levels of a given RNA molecule can provide deep insights into how a cell is functioning. Too much or too little of a particular RNA molecule can indicate deviations from homeostasis. In some cases, mutations in DNA can also influence individual RNA levels in cells. Put another way, mutations like single nucleotide polymorphisms (SNPs, where only one base pair sequence deviates from the norm) can be associated with disease or disease risk. However, presence of these DNA changes does not guarantee that the disease will manifest. Thus, researchers have looked to investigations of RNA to figure out what is actually happening in the cells and tissues of the body, rather than what might happen if an abnormality in the DNA is detected.

Using RNA to detect the underlying pathogenesis of disease could have numerous advantages. One of the most significant benefits to using RNA is that it can show what is happening in real time. Because RNA is transcribed from DNA, increased levels of a DNA signature may only suggest an increase in the number of cells with that signature. However, measuring changes in RNA levels provides an immediate indicator of specific cellular activities. With DNA, presence of a particular mutation may confer a level of risk for developing disease or predict responses to a drug regimen. RNA, by contrast, has the potential to explain exactly what is happening in the sample. Thus, RNA presents significant diagnostic advantages and can serve as a tool for understanding the molecular portrait of disease processes.

Using RNA to detect the underlying pathogenesis of disease could have numerous advantages. One of the most significant benefits to using RNA is that it can show what is happening in real time.

The History of RNA Detection Methods

FISH

Fluorescence has been a lynchpin of RNA detection throughout the history of the field. In the 1980s, fluorescence in situ hybridization (FISH) was developed, a breakthrough that simplified earlier methods using radioactive RNA or DNA probes. In this technique, fluorescent probes are synthesized that target specific RNA (or DNA) sequences. This allows for detection and localization of RNA transcripts for which the sequence is known. Thus, a known splice variant or a sequence associated with a disease can be illuminated in a sample. Depending on the target, presence or absence, quantity, and/or the location of the transcript within a tissue sample can be relevant for understanding cellular processes. Similarly, localization of transcripts can provide insights into the function of that particular RNA. More recently, additional refinements to probe-based technologies have made it easier to localize RNA at high resolution and lower copy number, providing greater insight into processes such as proliferation.

Microarray

Detection of RNA transcripts can be carried out outside of the cell, as well. Microarrays were first developed in the mid-1990s to look at gene expression in a plant model organism. The original investigators attached a set of cDNA sequences to an assay plate, after which they added purified test samples. If a particular sequence was present in the test sample, it resulted in a detectable signal on the plate. Thus, the researchers were able to determine both the presence and quantity of a particular sequence. Today, this same basic technique is used to elucidate molecular signatures in diseases, including certain cancers.

However, a critical limitation of both FISH and microarray is that they are dependent on knowing sequences *a priori*. One can only find genes or transcripts they are intentionally looking for and were included in the assay. Thus, while both FISH and microarray have been powerful tools for many years, they do not allow for unbiased discovery of molecular signatures. In some cases, this is a strength. For the Zika virus test mentioned above, targeting a very limited set of molecules is exactly what allows healthcare workers to have an answer within three hours. They know what they are looking for and have a test to find it, with no distractions. However, for the broader research community trying to make new discoveries and when dealing with more complex issues in the clinical setting (such as pharmacogenomics), this limited scope is indeed a deficit.

A critical limitation of both FISH and microarray is that they are dependent on knowing sequences a priori. One can only find genes or transcripts they are intentionally looking for and included in the assay.

RNA-Sequencing

The inherent bias of hybridization techniques provided an opening – and need - for something better. The most recent development in the field of RNA biology is RNA sequencing [RNA-seq], developed in the late 2000s using rapidly advancing technology from the sequencing field.

RNA sequencing has historically relied on Sanger sequencing of cDNA reverse transcribed from an RNA library. Reverse transcription is still used, but dramatically faster and cheaper reads are now possible using high-throughput techniques that emerged in the mid-200s with a new machine developed by Ilumina.

Emergence of RNA Diagnostics

Another significant advantage to RNA-seq is that it does not require specific sequence-based probes the way hybridization techniques do. This is particularly relevant for diagnostics. If a specific disease-related variant is known, hybridization may be sufficient because a probe for that sequence can be used to test a sample. Discovering new variants or the relevance of previously unknown differences in transcript levels requires the unbiased sequencing derived from RNA-seq. Thus, RNA-seq across both research and clinical applications present a very powerful platform for discovery and quantification of RNA targets that was previously unattainable.

RNA-seq starts by reverse transcribing RNA. Essentially this reverses the normal biological process by creating complementary DNA (cDNA) sequences out of RNA molecules. Then, adaptors – short DNA sequences that are not part of the original gene – are added to the ends of the cDNA fragments. Using the adaptors as starting points, the fragments are sequenced using high-throughput techniques. These often include adding in one deoxynucleotide (dNTP) at a time and then detecting a signal, often a tiny flash of light, produced by incorporation (or lack thereof) of that nucleotide. The process is repeated sequentially for each nucleotide (A, T, C, and G) for about 100-150 cycles to give reads of up to several hundred bases in length. Software then scans these individual fragment reads and assembles them.

Analyzing the resulting sequencing data at extremely high resolution and specificity is critical. This is another area where RNA-seq has significant advantages over older array-based hybridization techniques. RNA-seq can detect transcript levels along a highly dynamic range and detect low abundance transcripts more effectively than hybridization. Lastly, RNA-seq results in lower background than microarray and eliminates cross-hybridization (i.e., non-specific binding between probe and sample that leads to false-positives), cleaning up the analysis. Having clean, low-background and high dynamic range results is particularly important. Sometimes the existence of a mutation in a genetic sequence is indicative of a problem or disease process. In many other cases, though, it's the relative level of a particular RNA transcript that points to abnormalities in normal cellular processes that can lead to disease even in the presence or absence of a change in the DNA code.

Though still a relatively young technology, these features demonstrate the potential power of RNA-seq to help pinpoint specific biomarkers of disease. Importantly, the technologies described are not only extremely effective at characterizing molecular signatures and therefore facilitating the diagnosis of disease by providers, but they can also do this in a relatively short period of time and at a cost that continues to become more affordable. In coming articles, we will explore some of the successful uses of RNA diagnostics.

Though still a relatively young technology, these features demonstrate the potential power of RNA-seq to help pinpoint specific biomarkers of disease.