

Computational Genomics Through the Lens of Dr. Miler Lee

Whenever scientists have access to unmanageable quantities of data, computational innovation always follows. Next Generation Sequencing and recent improvement in genomic data production have thus highlighted the importance of computation in extracting biological insight. Naturally computational genomics, the marriage of computational biology techniques and large genomic data sets, has started to gain traction within the last few decades. According to Nature, Computational Biology is “an interdisciplinary field that develops and applies computational methods to analyze large collections of biological data, such as genetic sequences, cell populations or protein samples, to make new predictions or discover new biology.” Such methods are used frequently by researchers here at the University of Pittsburgh to investigate the mysteries of embryonic development.

I recently had the opportunity to sit down with an expert computational biologist that works with Next Generation Sequencing genomic data. Pitt’s very own Dr. Miler T. Lee, an assistant professor in the Department of Biological Sciences, works extensively with zebrafish in tandem with computational analysis to “study gene regulation during animal development.” His computational analysis of developing pluripotent zebrafish embryos has helped characterize transcription factors crucial in early stage development.

But, what exactly is pluripotency? “Pluripotency is the capacity for a cell to differentiate into any adult cell type,” says Dr. Lee. “This is a special property that most cells lack, and is the reason that, say, heart cells normally can’t become brain cells.” An example of pluripotency is the use of stem cells in bone marrow transplant, which has been practiced for the last six decades. This procedure replenishes the ability of blood cell precursors inside bone marrow to differentiate and create new cell types, such as those in blood. The importance of pluripotent stem cells cannot be understated. “Imagine if we could make a new organ for a patient by taking some of their skin cells, inducing them to pluripotency in a petri dish, then growing them up as a new tissue type,” ponders Dr. Lee. If this were possible soon, the issue of organ donor shortages

may vanish. As such, it is imperative to better comprehend the mechanisms of cell differentiation.

The Lee lab uses a common model organism, zebrafish, to study the underpinnings of pluripotency. “We investigate the molecular mechanisms that induce eggs to become pluripotent embryonic stem cells, focusing on the model organism zebrafish (*Danio rerio*),” states Dr. Lee. “My lab relies heavily on high-throughput experimental methods and computational genomics. Being able to draw from a large toolbox of different bioinformatics techniques gives us a lot of power to move the field forward.”

Zebrafish have become an exceptional model organism for a variety of reasons. Their genome sequence is well documented, verified, and deposited on multiple online servers around the world. This allows for greater collaboration across laboratories and increased data accessibility. Their embryonic development is rapid, more easily visualized, and occurs outside their mother (in eggs). They are a prominent resource for developmental biologists. Not only are they well suited to research, zebrafish also have direct implications for understanding human health.

Interesting parallels between zebrafish and humans include the fact that 70% of genes in humans have a zebrafish counterpart and 84% of genes associated with human disease have a zebrafish counterpart. “Although I happen to work on something that has direct parallels to human embryonic development, it’s important to realize that almost all biological inquiry has the potential to impact human health,” says Dr. Lee. “Many important discoveries were made by scientists who did not originally set out to cure diseases.” One such example could be CRISPR/Cas9 that was discovered by studying how bacteria defend against foreign biological matter. But how did we come to realize the medical potential of these findings? Computational genomics helps provide the understanding necessary to fully harness basic-science innovations like CRISPR/Cas9 into medically relevant treatments. Techniques like RNA-Seq and ChIP-Seq help researchers like Dr. Lee understand the genomic foundations of pluripotency.

Dr. Lee uses RNA-Seq and ChIP-Seq to quantify how gene usage changes over time and map protein interactions with a given genome. Through computation, these

techniques have yielded interesting discoveries into the development of animal embryos such as how RNA expression levels, produce specific embryo phenotypes. Using such tools, Dr. Lee has discovered novel transcription factors important for the proper development of endothelial and blood cell precursors in humans and zebrafish.

These tools are representative of how computers have guided the field of biology into a new era of data analysis. “I don’t like to think of Computational Biology as a field, as opposed to just a way to address biological questions with a different set of tools, computational tools. A lot of it may seem like really specialized skillsets right now, but I would not be surprised if a decade from now, a lot of what we do today will be automated and contained in a piece of benchtop equipment,” explains Dr. Lee. “I don’t mean to belittle what computational biologists do – it’s precisely their innovation that is propelling the field forward.” One great example is BLAST, the most widely used computational biology tool for quantifying similarities between distinct DNA or protein sequences. Before its development, such analysis required meticulous database mining that is now available at the push of a button.

Considering the breadth of computational biology, it can be challenging to figure out where to begin as an aspiring scientist and in which aspect of the field to partake. “There are a lot of resources online, including courses and tutorials that will give you a taste for what computational biologists do. Start there, then pick something that looks interesting and do some further investigation,” Dr. Lee advises. “There is a ton of biological data freely available, thanks to resources like NCBI’s Sequence Read Archive, and a great way to get experience is to play with some of this data. For example, you could find a journal article that reports some finding, look up the raw data it was based on, and attempt to replicate it.” The possibilities are endless. His research into pluripotency is a clear proof of concept for how computational techniques improve and accelerate biological inquiry.