New Single-Cell Sequencing Methods Aim to Move Beyond Gene Expression

Aug 31, 2018 | Monica Heger

SAN FRANCISCO (GenomeWeb) – Two single-cell sequencing methods described this week in *Science* aim to push the field beyond looking primarily at gene expression.

One method, developed by researchers from the University of Washington, combines RNA sequencing with chromatin accessibility, while the other, developed by researchers from Harvard University, focuses on analyzing genome structures.

Both groups said that analyzing other features at the single-cell level in combination with gene expression would help in gaining a more complete understanding of various cell types and their functions.

With sequencing costs continuing to decline and a range of technologies and methods for analyzing single cells now available, a lot of single-cell sequencing data is being generated, but the vast majority of that data is focused on gene expression, Jay Shendure, a professor of genome sciences at the University of Washington, said. "That's leaving a lot on the table," he said. "Gene expression is certainly very interesting, but there are other things going on inside the cell."

Shendure's lab collaborated with Cole Trapnell's lab at UW and researchers at Illumina to develop a single-cell method that combines RNA sequencing with a sequencing method to analyze chromatin accessibility known as ATAC-seq.

The method builds off previous research the group has published based on combinatorial indexing, which the UW team has harnessed as a way of analyzing thousands of single cells without having to rely on microfluidics or other equipment. Combinatorial indexing involves a series of barcoding and dilution steps that results in single cells being uniquely tagged without having to physically separate the cells.

The researchers originally published their ATAC-seq protocol in 2015, and have since described single-cell combinatorial indexing methods for RNA-seq, whole-genome and Hi-C, haplotyping, and methylome profiling.

The recent *Science* study is the first demonstration of two of the methods being combined in one assay. "The ability to get multi-omics, different aspects of gene regulation, at the single-cell level might create the opportunity to get closer to causality," Shendure said.

Shendure noted that combining the methods took "quite a bit of optimization," primarily by lead author Junyue Cao, who exploited the combinatorial nature of the protocol to test various optimizations. In

general, the method works by incorporating the first index barcode during reverse transcription within the nuclei followed by in situ tagmentation for the ATAC portion that also contains a well-specific barcode. The nuclei are then pooled and redistributed.

To demonstrate the method, the researchers first applied it to cells in culture that had been subjected to treatment at different time points with dexamethasone, which mimics cortisol and alters gene expression. They obtained both RNA-seq and ATAC-seq data from 4,825 cells, showing how gene expression and chromatin accessibility changed over time.

Next, they tested the protocol on mouse kidney cells, generating both RNA-seq and ATAC-seq data for 11,296 cells and identifying nearly 9,000 genes that were differentially expressed across 14 cell types.

Shendure said the combination of gene expression and chromatin accessibility data is helpful for a number of reasons. Single-cell ATAC-seq data can be quite sparse, since the data are spread out across the entire genome, as opposed to single-cell RNA-seq, which focuses only on the transcripts. "It's hard to cluster cells into cell type based on ATAC-seq data alone, so you can identify cells with the RNA-seq data, but then also have chromatin accessibility data for those cells."

In the study, the researchers used the ATAC-seq data after clustering cells by cell type with the RNA-seq data to analyze how chromatin accessibility differed across the cell types.

Going forward, Shendure said that his team plans to continue to improve on the method and also to apply it. For instance, he is interesting in looking at different time points in development to see whether changes in chromatin accessibility foreshadow changes in gene expression.

In addition, he said, such co-assays could be beneficial for large-scale projects like the Human Cell Atlas initiative, which is now generating gene expression data across thousands of single-cells. "Ideally, there will be other information layered onto such an atlas," Shendure said, including chromatin accessibility and spatial information, that are jointly measured in the same cells.

Dip-C

Similarly, Sunney Xie, who is now director of the Beijing Advanced Innovation Center for Genomics at Peking University, said that a main goal of his team's work, also published in *Science* this week, was to develop a single-cell sequencing method that could deliver more than just sequence information. His group developed a method that builds off advances from the Hi-C method to capture genome structure information, which is thought to play a role in gene regulation and cell function. The method, which the researchers have dubbed Dip-C, combines a transposon-based whole-genome amplification method with an algorithm for imputing chromosome haplotypes.

Other groups have previously adapted Hi-C sequencing for single cells in order to study genome structure, including Shendure's team as well as a group from the Babraham Institute in the UK. Xie noted that the method differs from previous Hi-C methods in that it can detect an order of magnitude more chromatin contacts per cell, giving it a resolution of 20 kilobases, and also incorporates a haplotyping algorithm to phase the genome.

The method enables researchers to see how individual chromosomes come together in a diploid cell, Xie said. In the study, the researchers performed the method on cells from a lymphoblastoid cell line as well as whole blood. They isolated cells using flow cytometry and used a slightly modified version of the Hi-C chromatin conformation method, skipping the biotin pulldown step. Then, the team performed a multiplex transposon-based whole-genome amplification method for each cell followed by sequencing.

The team sequenced 17 cells from the lymphoblastoid cell line and 18 whole-blood cells. The researchers used the structural information to type cells, showing for instance that the lymphoblastoid cells clustered together, and that cells from the blood samples clustered together that seemed to correspond to T lymphocytes, B lymphocytes, and monocytes/neutrophils.

In addition, by phasing the data, the researchers were also able to look at structural differences between maternal and paternal alleles. For instance, in the lymphoblastoid cell line, the researchers showed that the two X chromosomes had very different genomic structures, correlating with one of the chromosomes being inactivated. Xie said that specific cell line was already known to have X chromosome inactivated due to methylation, but that inactivation hadn't previously been linked to its structure. The inactive chromosome is more folded, he said, while the active one is more extended.

Ivan Liachko, CEO of Phase Genomics, which sells Hi-C kits, said that the Xie team made a number of "improvements on both the laboratory and computational fronts that add up to solid advance in the field" over standard Hi-C. He noted, however, that the single-cell sequencing and transposon-based library prep method "add complexity to the protocol, which while not uncommon in the field, does make it less likely to be widely adopted in the short term."

Xie noted the researchers plan to continue to refine the Dip-C method, including developing it into a kit to make it easier for others to use. He also said he is considering commercializing the technology. The researchers have filed a patent application on the method. Cost is also still an issue, with the protocol costing around 10 times more than single-cell RNA-seq methods.

Xie also echoed Shendure's opinion that projects such as the Human Cell Atlas should incorporate more than gene expression data, adding that structural information would also be valuable. He said one planned project is a way to "simultaneously sequence the transcriptome and get the structure at the single-cell level." He added that structural information is relevant for basic developmental biology as well as for certain diseases, like cancer.

Filed Under Sequencing Informatics Editor's Pick epigenetics genome mapping NGS RNA-seq single-cell sequencing