

## Sifting Through Cells

Advances in sequencing technology like the NextSeq<sup>®</sup> 500 system are giving scientists a peek into the processes that guide cellular development.

### Introduction

A single cell in the human body boasts about 6 billion base pairs of DNA and 600 million bases of RNA. Understanding how this deluge of information works to construct healthy organs or how it breaks down to provoke cancer, is tricky. Biologists must meticulously comb through these sequences to grasp thoroughly how these molecular instructions make cells tick.

At Stanford University Medical Center, DNA Sequencing Core Director Norma Neff, Ph.D., manages support for a team that is sequencing RNA with the NextSeq 500 system to parse the thousands of genes that assemble vital organs such as the brain and pancreas. The team is led by Stephen Quake, D. Phil., a genomics innovator and Professor of Bioengineering at Stanford. Instead of exploring the genetic makeup of cell populations, the team uses a technique called single-cell sequencing, which reads the gene expression of one individual cell at a time. In this way, the group can catch the unique genetic composition of a cell of interest, rather than relying on the consensus in a group of cells. This effort is part of a grant from the California Institute for Regenerative Medicine (CIRM).

Before Dr. Neff was recruited to be director of the Sequencing Core at Stanford by Dr. Quake in 2008, her work in yeast genetics at Memorial Sloane Kettering Cancer Center in New York helped discover a yeast protein with an intein that is capable of splicing itself. She also worked on the rare inherited disorder called Bloom syndrome, which makes a person more cancer-prone, at the New York Blood Center.

iCommunity spoke with Dr. Neff to understand how singlecell sequencing is transforming our understanding of the developmental programs that drive cells toward health and disease.

Q: When did you start using Illumina sequencing systems? Norma Neff (NN): I started working with Illumina sequencing systems with the ENCODE (Encyclopedia of DNA Elements) project in 2007 with Richard Myers, Chairman of Genetics at Stanford. We were looking at binding sites for different transcription factors. When I started working with Steve Quake, the Stem Cell Institute had moved off campus to a building on Arastradero in Palo Alto. I was recruited to help set up the sequencing core for the Stem Cell Institute in 2008.



Photo courtesy of Jun Seita

The Stanford DNA Sequencing Core is located in the Lokey Stem Cell building, where a Dale Chilhuly glass chandelier graces the lobby.

# Q: The team at Stanford has a grant from CIRM to perform single-cell sequencing on the brain and the pancreas. What are the goals of your studies?

NN: We're really digging into these organs to understand their development and characterize the different cell types in a precise way. We're trying to develop gene sets that will identify specific cell types in these various organs. The idea is to see what we can learn from the most abundant and moderately expressed genes in the cell. With single-cell RNA-Seq (mRNA Sequencing), you can look specifically at gene expression patterns within individual cells.

#### Q: Does this work have applications in cancer?

NN: Prospectively, down the road we'll work on cancers, but right now the basic focus is on early development. Recent studies suggest that cancers arise in a stem or early progenitor cell, so a deeper understanding of the normal molecular developmental program will help us understand how cells become tumorigenic.

#### Q: Why is this work so challenging?

NN: Single, isolated cells are fragile and contain very small amounts of mRNA. Contamination and RNA degradation are constant problems.

#### Q: What observations have you made?

NN: We've learned that there aren't as many genes expressed in single cells as originally thought. When looking at the ENCODE project, there are 15,000 to 20,000 genes expressed in the cell populations. With single cells, we really get somewhere between 6,000 and 8,000 genes being expressed in individual cells. So you don't have to sequence as deeply as you do for other types of RNA-Seq. Also there are disparities between mRNA levels and protein expression. Single cells that are sorted using antibodies to abundant surface markers don't always have abundant mRNA for that protein.

#### Q: Why did you choose the NextSeq 500 system?

NN: We chose it for the high number of reads you get and the fast turnaround. The NextSeq 500 System has really been the backbone of this pipeline, enabling us to pool 100–300 individual single-cell RNA-Seq libraries together and sequence them. We've been using the new 384-index system for the Nextera<sup>®</sup> libraries and that's helped out a lot. If we find cells that look more interesting and want to sequence more deeply using the large index set, the probability that you can go back and sequence that one cell more deeply in another pool is much higher with more indexes.

#### Q: How has the NextSeq 500 system performed?

NN: We've been running it quite successfully six days a week; we give it one day off. Primarily, we perform 2 x 75 bp runs for the RNA-Seq libraries. That's a perfectly good read length when you're just aligning it to the reference genome. It keeps the error rate low, everyone is very happy with the data quality, and the cost per Gb is nearly identical to HiSeq<sup>®</sup> system runs.

### "The data quality that we get off the NextSeq 500 system at 2 × 75 bp is exactly what we're looking for in our studies."

### Q: How does the NextSeq 500 system compare to the $\rm MiSeq^{\tiny (0)}$ system you have in your lab?

**NN:** The data quality that we get off the NextSeq 500 system at 2 × 75 bp is exactly what we're looking for in our studies. The flow cells are shipped dry so there is no cleaning. After sequencing, the system automatically washes itself using a bleach solution added into the reagent cassette, so there is no delay in starting the next run. From a speed standpoint, it actually makes the MiSeq<sup>®</sup> system look slow.

### Q: What library prep kit are you using?

NN: We're using the Illumina Nextera XT® kit very successfully. We're really glad to have it, primarily because we make hundreds of libraries at a time. The 96 capillary AATI Fragment Analyzer is a critical instrument in our single-cell sequencing pipeline. Once you quantitate the cDNA (complementary DNA) with capillary electrophoresis, the Nextera libraries are so reproducible that you just mix 2 microliters of your libraries and clean up a single pool and submit them for sequencing. There's no need to quantitate the individual library preps because they're coming out to be so uniform from plate to plate. We're very happy with this product.

### "It's both the throughput and the turnaround time of the NextSeq system that is helping to advance this research field."

### Q: What types of sequencing studies are you performing with the MiSeq system?

NN: We now use the MiSeq system for experimental library construction and for metagenomics primarily. We're doing a lot of metagenomics in collaboration with JGI (Joint Genome Institute). The long-term goal of the project is to develop methods for sequencing and assembling single cell bacterial genomes from environmental samples. We also used the MiSeq system to sequence single nucleotide polymorphisms (SNPs) for cell lineage studies from various types of cancer cells.

### Q: Why is now the time for single-cell sequencing?

NN: The technology to isolate cells and amplify RNA has gotten much better. We rely on the Fluidigm C1 system to capture single cells and synthesize cDNA in nanoliter microfluidic chambers. There's an opportunity for people to make an unbiased assessment of the cells in their population with regard to mRNA expression.

Illumina sequencing has improved to the point where the turnaround time is much faster and the throughput is high enough to be able to get 1–2 million reads for several hundred single cell libraries in a 30-hour run. It's quick enough to support multiple single-cell projects with timely data production. It's both the throughput and the turnaround time of the NextSeq system that is helping to advance this research field.

### Q: What improvements must be made to make single-cell sequencing more useful or universally available?

NN: We are focused on improving protocols for making single cell suspensions and investigating novel ways of displaying the data. Making single suspensions out of solid tissue is a particular challenge, primarily because you have this inherent feeling that the gene expression is changing at the time you're doing your prep. So the sooner the cells can be captured and made into cDNA, the better off you'll be. Collecting, analyzing, and displaying the data in useful ways are all part of an evolving process. Integrating and sharing this new data in an efficient and friendly format among different institutions is a long-term goal of the CIRM project.

### Q: What are the next trends in this field?

NN: The Stem Cell Institute is interested in defining what are the true stem cells, both for various types of developmental programs as well as cancer stem cells. We're trying to isolate and transplant these cells to see what their potential is and learn something about their gene expression program.

We collaborate with researchers in the Stem Cell and Cancer Institutes to try to find new cell surface markers and new transcription factors by single-cell RNA-Seq to help develop better tools to isolate these cells in a purer form, especially with regard to stem cells and cell-based therapy. You want to give people the right cells.

For cancer therapy to focus on cancer stem cells, you want to be able to target the correct cell population without causing damage to other cells. That's not really a new trend, but the technology has gotten better. By looking at cells on an individual basis we have a new window to try to understand what these cells are, what they're doing, and how to target them more effectively.

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

FOR RESEARCH USE ONLY.

© 2014 Illumina, Inc. All rights reserved. Illumina, HiSeq, MiSeq, Nextera, Nextera XT, NextSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 770-2014-015 Current as of 30 October 2014

# illumina