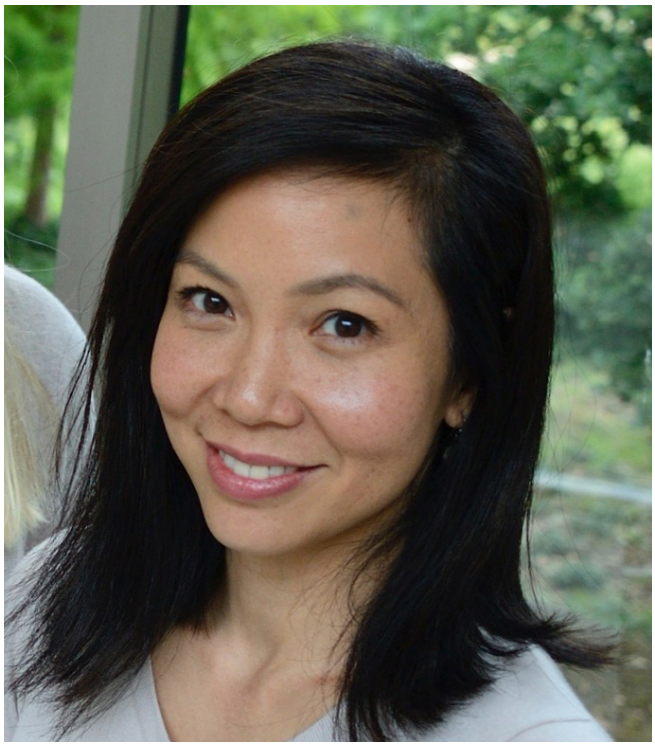
Whole Genome Sequencing Gives Scientists a Detailed Map of Cancer Mutations

Dr. Serena Nik-Zainal, a clinical geneticist, and her team at the University of Cambridge are studying how cancer genome profiling could help clinicians better fight the malignancies



Dr. Serena Nik-Zainal, a Cancer Research UK Advanced Clinician Scientist and Honorary Consultant in Clinical Genetics.

Introduction

Cancer starts silently in our genes—not just with a few mutations, but with thousands of them. "If you can sequence the whole genome, you can get an entire map of the genetic defects in a cancer," says Dr. Serena Nik-Zainal, a clinical geneticist at the University of Cambridge. That's exactly what Dr. Nik-Zainal and her team are doing in an effort to understand cancer's genetic roots and improve treatment.

When scientists finished sequencing the first human genome in 2003[1], it had cost three hundred million dollars and taken thirteen years to complete. Today, a human genome can be sequenced in two days for under a thousand dollars[2]. Armed with this powerful technology, Dr. Nik-Zainal and her colleagues are using whole genome sequencing to identify and interpret cancer's myriad mutations. "Now that we can see the whole human genome and all those thousands of mutations, we've realized that they are not random. They seem random when you're looking at the three billion bases—it looks like noise," explains Nik-Zainal.

> "But actually, you can make sense out of the mutations. They occur in patterns. Understanding these patterns could one day give researchers and clinicians critical insight into the causes of a patient's cancer and the best strategies to treat it"

iCommunity spoke with Dr. Nik-Zainal about why she decided to pursue research in cancer genomics, how whole genome profiling can benefit cancer patients, what mutational signatures could reveal to researchers and clinicians about the disease, and what she believes are the significant trends in the future of her field.

Q: What has been your journey so far in the field of genomics and cancer research?

A: I'm a medical doctor by training and I specialize in the area of clinical genetics, which is for diagnosing and managing people with genetic disorders. They may have their genetic disorder from birth, and it may give them, for example, an increased risk of developing cancer or greater risk of having some kind of neurological illness. In my field, we try to diagnose these people and manage them to the best of our ability.

In the 2000s all these new technologies emerged, where people were able to read the genome a lot quicker. I found that the technology would sometimes reach me as a trainee specialist, and I wasn't able to understand it. I would try to explain to a family, for example, why their child has a genetic disease and I would look at a result but wasn't sure I fully understood it. So, I decided to do a Ph.D. in technology and informatics at the Wellcome Sanger Institute.

That was about 2009—only a few years after the end of the Human Genome Project. We now had the ability to look at the whole genome. So, I thought: Why don't I get involved in that? At the time, though, whole genome sequencing was still an enormous task. It took us months to make what we call "libraries" in order to read even one cancer genome. But today, people can produce the sequence of a whole genome in 24 to 48 hours. Even in the decade that I've been a scientist, the speed and scale of sequencing has increased dramatically.

Now I can read the whole human genome for diseases like cancer. Cancer is a disease of the genome, where the genome that you're born with has changed dramatically. It's mutated—and we're not talking five or ten mutations, we're talking about thousands of mutations.

> "Those mutations can tell you something about why that person developed cancer. If you can sequence the whole genome, you can get an entire map of the genetic defects in a cancer."

So I went into cancer genomics to study how we can read all these mutations in cancer genomes and importantly, how to interpret them and how to use that information for a patient's benefit. We're still in this journey, we're still learning—but it's been a really exciting time.

Q: Can you talk more about your experience using genetic sequencing technology in the early days of your research, compared to today?

A: It was incredibly difficult just generating the data. In those days, it took us weeks to even get the data. In the first study I published when I was at the Wellcome Sanger Institute, it was only twenty-one samples of cancer from women around the world. It doesn't sound like much at all, but in those days it was a big deal.

Now there are other versions of sequencing, such as exome sequencing, where you just look at the coding sequences. There are also targeted sequences where you only look at certain genes. People are doing a lot of all of these smaller amounts of sequencing, and importantly, they are sharing that data. It's wonderful because when you share, you learn from each other's data. There is a communal benefit of new knowledge and insights.

That's one of the things I've noticed in the past 10 years. It was very difficult at the start and then it started to get easier as people started sharing more data. The process of generating data has become easier. But it's still very exciting and new, with all these discoveries being made.

Q: How exactly can whole genome profiling help cancer patients and clinicians?

A: Among the thousands of mutations in a cancer patient's genome, there are four main types of mutations: single building block changes, insertions and deletions, rearrangements, and sometimes the number of copies of our chromosomes can be increased or decreased.

Whole genome profiling basically pulls all the information together. You can tell from looking at it whether it has abnormalities, such as a defect in a DNA repair pathway. When you put all this information from whole genome profiling together, you get these images—almost like a picture of the human genome. And then we have computational algorithms that make calculations and report to us the probability of, for instance, a tumor being mismatch repair deficient or something else that can be targeted by a certain drug. That's how we use whole genome profiling.

Q: What exactly are mutational signatures and why are they important in cancer genetics?

A: In the last four decades, people have focused on very few mutations that we call driver mutations—those are mutations that occur in genes. But genes make up only one and a half to two percent of the genome, which means that everyone has focused on a very small amount of the genome. Of course, it was very important to identify driver mutations because people created drugs for those specific mutations, and those drugs work very effectively.

> "But now that we can see the whole human genome and all those other thousands of mutations, we've realized that they are not random."

They seem random when you're looking at the three billion bases—it looks like noise. But actually, you can make sense out of the mutations. They occur in patterns.

The mutation pattern is what we call mutational signatures. They are a signature of something that has happened. Here's an analogy: if you look at a beach that has a lot of footprints, they may seem random. But if you study the footprints, you would be able to tell if they belonged to a human or animal, if they're from a child or adult, the direction the person was walking in, and more. In a similar sense, we're studying the footprints of these mutations to see what we can learn from them. Some of those mutational signatures tell us, for example, about abnormalities that could be targeted by certain drugs. They may even tell us whether someone has a tumor that's very bad and they have a very poor prognosis. That's why we're looking at patterns—these mutational signatures—in cancer genomes.

Q: What is tumor mutational burden, and can it be used as a new type of metric to evaluate cancer?

A: Tumor mutational burden is the total number of mutations that are present in a cancer. There has been work by Luis Diaz *et al*.[3] on mismatch repair deficient cancers—mismatch repair is a kind of DNA repair pathway which normally mitigates damage. When you have a deficiency in this pathway, you cannot fix damage anymore and there is a huge increase in mutations. Tumor mutational burden is being used as a proxy for mismatch repair deficiency. There is a cutoff, and if your tumor mutational burden is above this cutoff, then potentially your mismatch repair deficiency could be sensitive to specific drugs—in particular, checkpoint inhibitors. That was why it was designed.

Tumor mutational burden works quite well as a biomarker in things like colorectal cancer and uterine cancers, where mismatch repair deficiency is most common. But mismatch repair deficiency occurs in many different tumor types. In those other tumor types, the cell cycle doesn't turn over as quickly as in the colon or the uterus. Gene panels large enough to evaluate tumor mutation burden sometimes misses these mismatch repair deficient tumors.

Tumor mutational burden has a problem with specificity as well. Mismatch repair deficiency is not the only reason why someone might have a high mutational burden—it can occur for a number of different reasons. So, if we're just using a total number as a cutoff, there's a risk that we will identify a cancer as mismatch repair deficient when it's not. But we've got to start

somewhere, large gene panels are actually effective assays for detecting tumors with mismatch repair deficiency for inhibition therapy.

Q: Where is your research headed now?

A: My team is now trying to optimize the work we're doing. A lot of our work is algorithmic and computational, which means we need to invest a lot of energy to make sure it works well. Whole genome sequencing is no longer the biggest hurdle.

> "People used to worry that whole genome sequencing was expensive and difficult, but it's no longer that way. Illumina has shown us that it can produce large numbers of whole genome sequences fantastically as it has improved its technology enormously."

Now we can do a whole genome sequence for almost less than the cost of a complex CT-scan. The cost of sequencing is no longer a limitation. The real limitation now is analysis and interpretation.

My team is focused now on trying to make analysis and interpretation as quick as possible so that we can help clinicians use that information to the best of their ability. They shouldn't be asking, "What do I do with this data?" They should say, "Okay, I've got the results, now how will I treat this patient?" That's what we need to do, and that's where we're headed.

Q: Have you noticed any trends in cancer genomics research that you expect will be significant in the future?

A: There are two trends here: from the clinical perspective and from the research perspective. From the clinical perspective, genomics is still in its early stages. A lot of genomics work up to this point has consisted of people looking for one mutation or one genetic defect. For example, with an important gene like *PIK3CA* in breast cancer, people look for the presence of one mutation and they use that one mutation to decide what treatment to give a patient, or whether to give a patient treatment or not. It's very binary. But the human genome is huge. A patient might have a *PIK3CA* mutation, but they might have a whole lot else wrong with their cancer. We need to use all of that information to best manage patients. That requires a shift in mindset. It's still not quite happening, but it will come. I think people are starting to see that.

From the research perspective, a lot of academic work so far has focused on targeted and exome sequencing, I think that will change towards a whole genome approach, as well as how people use genome sequencing in conjunction with other modalities such as adding on transcriptomics or proteomics. Whether that becomes something that's ever useful clinically, I don't know. Certainly, it's an exciting area from a research perspective.

The world does not revolve around the genome. All my work revolves around the genome, because we can get a lot of information from it. But the genome is one part of a very busy ecosystem. There are proteins, there's this soup of cellular material. The genome interacts with the immune system, with the environment. There is a lot more for us to learn about how a cell goes from being normal cells to a cancer cell.

References

1. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003 Apr 11;300(5617):286-90.

2. Illumina, Inc. NGS for Beginners: Cost of Next-Generation Sequencing. https://sapac.illumina.com/science/technology/next-generation-sequencing /beginners/NGS-cost.html

3. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017 Jul 28;357(6349):409-413.

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