Targeted Resequencing Enables High-Throughput Detection of Somatic Mosaicism

Dr. Saumya Jamuar uses targeted resequencing on the MiSeq™ System in the first systematic, high-throughput approach to identify rare mosaic mutations in brain malformations.

Introduction

Somatic mutations, which lead to two or more populations of cells with distinct genotypes, have been characterized in cancer. However, while the role of somatic mutations in genetic disorders has gained recognition, their prevalence in neurodevelopmental diseases has only recently been evaluated systematically.

Traditionally, identification of genetic variants has been accomplished with capillary electrophoresis (CE)/Sanger sequencing. However, this iterative process is time-consuming, labor-intensive, and has a limit of detection that prevents identification of rare variants.[1] New technologies, such as next-generation sequencing (NGS), enable scientists to increase the scope of their studies and shorten the path to discovery.

Saumya Jamuar, MD, Clinical Geneticist at KK Women's and Children's Hospital, Singapore, and cofounder of Global Gene Corp, an innovative genomics research and development company, created the first systematic, high-throughput approach to identify somatic mutations associated with cerebral cortical malformations.[2] He performed this study using NGS with the MiSeq System and TruSeq™ Custom Amplicon assay.

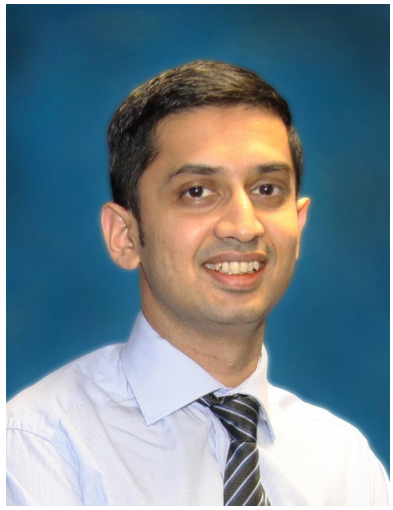
iCommunity spoke with Dr. Jamuar about the development of his targeted resequencing panel and how he is using NGS to improve human health.

Q: Why did you become a physician?
Saumya Jamuar (SJ): Medicine is a profession that allows one to transform the lives of other people in a very direct manner and at a point when they are most vulnerable. As a pediatrician, I always find it challenging to help children, but it's very satisfying to hear "thank you" at the end of a consult, especially when they do get better. That gratification and the job satisfaction drives my whole practice.

Q: What prompted you to pursue scientific research in addition to practicing medicine?
SJ: I studied and completed my pediatric residency in Singapore, followed by a government-sponsored genetics fellowship at Harvard Medical School. After completing my fellowship, I stayed on at Harvard and performed postdoctoral work in the lab of Dr. Christopher Walsh. That's where I made the switch from being a clinician to becoming a clinician scientist. I realized that when I am in the clinic, I can only help one patient at a time. However, in the lab I can help many more patients by conducting research that can impact human health.

I realized that there was a major gap in translational research. Translational research wasn't happening at the same pace as basic science research or clinical trials, and application of basic science research to patient care was lacking. My background in clinical medicine and the basic research capabilities of the Walsh lab provided me with the perfect opportunity to practice translational medicine. That's when I also became familiar with Illumina.

Q: What kind of research did you perform in the Walsh lab?
SJ: I've always been interested in neurological disorders, especially brain development and how errors in that process lead to brain malformations. Professor Walsh is a neurologist and leader in the field of genetics of brain malformation and neuropsychiatric diseases, so the Walsh lab was the perfect place for me. My project was aimed at developing a systematic approach to investigating somatic mutations in brain malformations.



Saumya Jamuar, MD, is a clinical genetecist at KK Women's and Children's Hospital in Singapore, and the cofounder of Gene Corp.

Q: Were Illumina systems and products being used in the Walsh lab?
SJ: The lab hadn't invested in an Illumina system. Instead their sequencing was performed at core or service labs. They would send sequencing samples to different providers, some of whom used Illumina technology. As part of the background work for my project, I read information about the different products that were available, which in 2012 included Illumina TruSeq Custom Amplicon, Agilent HaloPlex, and Ion Torrent amplicon sequencing, among others. Those were the three that I considered, and, after much deliberation, I decided to go with TruSeq Custom Amplicon.

> "I realized that when I am in the clinic, I can only help one patient at a time…I can help many more patients by conducting research that can impact human health."

Q: What were the reasons that you chose TruSeq Custom Amplicon?
SJ: One reason was the relative ease of designing the experiment and customizing the panel. When I tested TruSeq Custom Amplicon using DesignStudio™ Software, the process was fairly simple. I input gene names and specified how much flanking I wanted at the exon-intron boundary. Two days later, I had a design that met most of my requirements. Then it was a matter of a few additional steps to optimize the panel, which was also a simple process.

The second reason was the reputation of Illumina technology. In speaking with people who were familiar with the different technologies, I was advised to use Illumina technology because the data quality was robust. In the end, it worked out well for us.

Q: In the days before NGS, you would have used Sanger sequencing to perform this study. At the time, what were the strengths of Sanger sequencing to identify mutations in variant detection?
SJ: Actually, there weren't many advantages to using Sanger sequencing for variant detection. It was a lengthy process because Sanger sequencing proceeds gene by gene. At times, the lab would sequence many genes and not get any results. With NGS, we put all the genes of interest together into one panel, one reaction, and waited for the results.

Also, Sanger sequencing has a limit of detection of about 15–20% for mosaic variants. When trying to detect mosaicism, we showed that the proportion of these variants can be as low as 5% and Sanger sequencing would not be able to detect these variants.[2] Using TruSeq Custom Amplicon, we could easily detect mosaicism down to a theoretical limit of detection at 1% based on our calibration samples.[2] Compared to Sanger sequencing, it was a relatively straightforward process. Researchers have been trying to detect mosaicism systematically for over a decade. With the TruSeq Custom Amplicon panel and MiSeq Reporter Software, I was able to identify those mosaic variants within a week. NGS turned mosaic variant detection into a highly efficient process.

However, the validation of those mosaic variants took longer. We had to demonstrate that the variants were real and not due to a sequencing error. We used Sanger sequencing of subcloned, individual colonies to prove that, and it took an additional 6 months to accomplish. Validation of low frequency variants is still an issue, and different labs are looking at alternative methods of validation of variants that are present below 30%.

Q: Why did you look specifically at somatic mosaicism and brain malformations?
SJ: The Walsh lab found evidence of mosaicism in the doublecortin-encoding *DCX* gene that results in double cortex (DC) syndrome and X-linked lissencephaly (XLIS), literally a "smooth brain" malformation.[3] Children with lissencephaly are typically neurologically devastated; they can suffer from epilepsy, developmental delay, feeding issues, and aspiration pneumonia.

Brain malformations are easy to identify with structural imaging such as MRI. The lack of sulci (furrows) and gyri (ridges) in a lissencephalic brain can be visualized. But DC is a spectrum disorder; some individuals have epilepsy but do fairly well. They aren't developmentally delayed. MRI would show some changes that resembled lissencephaly, but not completely. This is due to the cell autonomous nature of neurons. Neurons lacking the mutation will grow and migrate as expected, forming a normal cortex. Neurons bearing a *DCX* mutation will display migration defects resulting in a heterotopic band of neurons that underlie the normal cortex, hence the name "double cortex". Mosaicism in the brain was evident from the radiological findings, but the question remained. How significant was it? How many of these families with evidence of brain malformations on imaging actually had mosaicism that Sanger sequencing had missed?

> "I was advised to use Illumina technology because the data quality was so robust. In the end, it worked out well for us."

Q: By using TruSeq Custom Amplicon, how were you able to clarify some of the results that didn't make sense with Sanger sequencing?
SJ: Some mosaic variants were missed with Sanger because they were below the level of detection. One of the patients had undergone whole-exome sequencing and again the variant was missed because the sequencing coverage was at a very low level. TruSeq Custom Amplicon enabled deep sequencing of at least 200–500× of a select region of the genome, targeting genes that have been associated with brain malformations

Q: Have targeted applications been used to look at other neurodevelopmental disorders?
SJ: The Walsh lab and others have used targeted sequencing to detect somatic mutations in other neurodevelopmental disorders such as autism.[4]

> "Using TruSeq Custom Amplicon, we could easily detect mosaicism down to a theoretical limit of detection at 1%…NGS turned mosaic variant detection into a highly efficient process."

Q: Are other labs using NGS to detect somatic mutations in different neurological conditions?
SJ: There was a recent study that used whole-genome sequencing to detect variants causing severe intellectual disability.[5] There have also been papers that have shown somatic mutations in epilepsy using NGS.

Q: So, your research opened up a new door for people to investigate the role somatic mutations are playing in these diseases?
SJ: We definitively showed that NGS was a better alternative to Sanger sequencing for detecting somatic mutations. We found somatic mutations in eight individuals, and remarkably almost two-thirds of them were missed with traditional Sanger sequencing. One of these individuals went on to whole-exome sequencing looking for the answer. But the truth was that one had to perform deep sequencing to find the mosaic mutation.

I must emphasize that we were not the first to use NGS to detect somatic mutations. There had been individual case reports where researchers performed whole-exome sequencing and found somatic mutations. However, we developed the first systematic, high-throughput approach to identify somatic mutations. And now, some clinical labs have started using deep targeted sequencing to look for somatic mutations as well.

Q: Why did you choose the MiSeq System?
SJ: It came down to the MiSeq and HiSeq™ Systems, and the HiSeq 2000 System had a turnaround time of 14 days. I wanted to know quickly whether there was an issue with my design, so I chose the MiSeq System because the throughput gave me results within 40 hours. I would load the libraries on the system, and the next day I had the results.

Given that now the HiSeq 4000 System has a turnaround time within 1–3 days, it's an option that offers much higher throughput. It can enable researchers to run more samples.

Q: How did MiSeq Reporter Software help in your study?
SJ: When I began working with NGS, I knew Illumina produced great sequencing chemistry and made great panels, but I was pessimistic about the bioinformatics capability. That changed after I used MiSeq Reporter. I had tried to perform my own bioinformatics using different algorithms and pipelines, and I wasn't getting the answers I wanted. I decided to try MiSeq Reporter and found it was a quick process, and I was ultimately successful.

Having a DNA-to-report solution definitely helps. If you use a sequencing panel, you run it on a sequencer, but that's only the first step. The bioinformatics is important, but it can also be challenging. The trick with using MiSeq Reporter was that it performed a localized alignment to the target region. I tried aligning FASTQ files myself, but I was aligning them to the entire genome. For a large gene such as *FLNA*, which encodes the Filamin-A protein, I was only getting 50% alignment because many of the reads were mapping incorrectly and being discarded. When I used MiSeq Reporter software and ran the same data, my alignment increased from 50% to 90%. MiSeq Reporter enabled me to detect mutations that otherwise would have gone undetected, not because the data was bad, but that I was not using the right tools to align.

Q: Now that you're back in Singapore at the KK Women's and Children's Hospital (KKH), are you using Illumina NGS systems for your current studies?
SJ: After returning to Singapore, I became the clinical lead for the undiagnosed disease program at KKH. We are using the HiSeq 4000 System to perform whole-exome sequencing for these subjects. For syndromic individuals, we are using the TruSight™ Inherited Disease Sequencing Panel on the MiSeq System as a first pass to identify mutations in known genes. If those are negative, then we move on to either whole-exome or whole-genome sequencing, the latter we are doing on the HiSeq X™ Ten System.

> "MiSeq Reporter enabled me to detect mutations that otherwise would have gone undetected, not because the data was bad, but that I was not using the right tools to align."

Q: Were those Illumina systems at KKH before you got there, or did you influence their acquisition?
SJ: We do not own any HiSeq Systems, but use those owned by our collaborators. However, I encouraged our research lab to purchase the MiSeq System because of my previous positive experience with it.

Q: What is Global Gene Corp?
SJ: While in Boston, I cofounded Global Gene Corp with one of my Harvard colleagues, Professor Jonathan Picker. Our goal is to understand the underexplored populations of the world using genomics and provide affordable and effective genomic solutions in Asian markets such as India, China, and the Middle East to

disrupt traditional health care and enable delivery of precision medicine in these economies.

> "NGS has changed how I practice medicine. It helps me explain to parents that their child has a set of problems because of a genetic condition, and I can identify the mutation and give them a more precise answer.

Q: What problems are you solving at Global Gene Corp?
SJ: To benefit from the advances of precision medicine and genomics, one needs to link genomics data and the clinical context, and then build applications to benefit the end patient. There are many experts in precision medicine in the US and in Europe, but South Asia, Southeast Asia, and the Middle East are lagging behind the rest of the world. The result is more than 60% of the data in the scientific literature comes from less than 1% of the global population. For instance, India has a population of 1.3 billion people and there are less than 100 publicly available whole-genome sequences. While the 1000 Genomes Project (1KGP) has some Indian representation, they come from specific subpopulations within India, including the Gujarati, Punjabi, and the Sikhs, and those are just four or five distinct subgroups. It has been demonstrated that Indians are different from Europeans and Caucasians, but within India individuals from different parts of the country tend to segregate based on principal component analysis.

Global Gene Corp, along with our collaborators, is creating the foundation of precision medicine in Asia, starting with India, and then bringing genomics to life in these markets through customized applications.

Q: This is a pioneering effort; how are you accomplishing this?
SJ: Global Gene Corp is focused on using genomics to impact every individual's life in a positive manner. This means creating relevant and customized applications for specific populations that solve for both the emotional and financial cost of disease.

For example, India has almost 1 million new cancer patients every year with 700,000 deaths estimated in 2015. In certain types of cancer, such as breast and ovarian cancer, 150,000 new cases are detected every year. The 5-year survivability in India is 52% compared to 89% in the United States with research demonstrating low effectiveness of oncology drugs as a class. We are launching personalized chemotherapy in India in collaboration with our partner in the UK – which combines phenotype-genotype and drug interaction information to identify the most effective treatment plan for every individual. We also support pharmaceutical companies in creating better, innovative, personalized drugs relevant for their target populations with our expertise in and access to genomics insights and cohorts in these populations.

Q: Why is affordability a key focus area?
SJ: I was brought up in India, and I am intimately familiar with the challenges and opportunities in delivering products in an affordable manner. In the markets of focus for Global Gene Corp, while there are significant numbers of patients who can afford any price point, the pyramid of affordability naturally means that the higher the affordability of treatment, the greater the number of people who can benefit from this exciting technology. Hence, one of the key questions that is core to our ethos is, how can we work through the delivery chain to make it affordable? Ultimately, the goal is to innovate on the application of the science to create personalized products and also on efficiencies of cost-structures to deliver it and benefit as many people as possible.

Q: During your time as a physician, what has NGS meant to medicine?
SJ: NGS has transformed how I practice medicine. It helps me explain to parents that their child has a set of problems because of a genetic condition, and I can identify the mutation and give them a more precise answer. More importantly, I can tell them whether they need to worry about future recurrence risk. Even if there is cause for worry, there are ways for us to guide them during pregnancy because we know the mutation. Before NGS, we couldn't do this. There is a term called 'reproductive stoppage' where families decide not to have more children after having a child with a disorder. It increases if it's a genetic disorder with an undiagnosed molecular etiology. Knowing the molecular diagnosis changes everything because now I can tell parents that we can help them.

> "Actually, there weren't many advantages to using Sanger sequencing for variant detection. It was a lengthy process…With NGS, we put all the genes of interest together into one panel, one reaction, and waited for the results."

Q: What do you see of the future of using NGS? How will NGS impact medicine in the future?
SJ: I believe that when we look back, the greatest achievements of this century will be related to looking inwards into the "human being" and understanding how we work. I believe that genomics is the disruptive "technology" of our lifetime that will enable us this privilege of looking inwards and putting the pieces of the puzzle together.

Previously, clinical genetics was about identification of a potential syndrome and managing complications in an anticipatory fashion. With NGS, we can go deeper to identify causative mutations and in the future, possibly correct them so that people can live healthier lives.

> "Our goal is to understand the underexplored populations of the world using genomics and provide affordable and effective genomic solutions in Asian markets..."

What I foresee is the increasing value and utility of tools like CRISPR and other related genome editing technologies. Thalassemia is a common genetic disorder in Southeast Asia, 3% of the population is a carrier for a mutation in the globin gene. I believe we will be using genome editing tools in the not-so-distant future to edit the genome and actually cure such monogenic diseases. Obviously, this is in the future and the use of CRISPR is still developing. If researchers can optimize CRISPR or some other genome editing technology, it has the potential to change medicine. That is how NGS will fundamentally change the practice of clinical genetics, and that's what excites me.

Q: What are the next steps in your work?
SJ: At Global Gene Corp, our plan is to see how we can build these genomic cohorts and identify potential novel pathways, novel targets for drug development, and therapy. Long term, we are working to personalize medicine at the individual level rather than what we have now, which is one solution for everyone. We hope that we can customize solutions for each individual.

> "Researchers have been trying to detect mosaicism systematically for over a decade. With the TruSeq Custom Amplicon panel and MiSeq Reporter Software, I was able to identify those mosaic variants within a week."

Q: How does it feel to help many people versus just one person?
SJ: I used to wonder what I could do as a scientist when I was already helping patients as a clinician. However, science allows me to help individuals on a much bigger, possibly even global, scale. Some of the credit goes to Illumina for making NGS systems that make it easy for us to do this type of research. Science is always more fun when you get results, and more satisfying when you make a positive difference to your patient.

References

1. Gajecka M. Unrevealed mosaicism in the next-generation sequencing era. *Mol Genet Genomics*. 2016;29(2):513-530.
2. Jamuar S, Lam AN, Kircher M, et al. Somatic Mutations in Cerebral Cortical Malformations. *N Engl J Med*. 2014;371(21):2038.
3. Gleeson JG, Minnerath S, Kuzniecky RI, et al. Somatic and Germline Mosaic Mutations in the doublecortin Gene are Associated with Variable Phenotypes. *Am J Hum Genet*. 2000;67:574–581.
4. D'Gama AM, Pochareddy S, Li M, et al. Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. *Neuron*. 2015;88(5):910–917.
5. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344–347.

For more information about the products, systems, and applications mentioned in this article:

TruSeq Custom Amplicon Panel, www.illumina.com/products/truseq_custom_amplicon.html

MiSeq System, www.illumina.com/systems/miseq.html

Targeted Resequencing, www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing.html

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

illumina®

For Research Use Only. Not for use in diagnostic procedures.