Targeted Multigene Sequencing Panels Enable Fast, Cost-Effective Variant Discovery

Researchers at the Carlo Besta Neurological Institute perform targeted sequencing with the MiSeq® System to uncover novel variants associated with metabolic and neurological disorders.

Introduction

Franco Taroni, MD, heads the Unit of Genetics of Neurodegenerative and Metabolic Diseases at the Carlo Besta Neurological Institute (INCB) in Milan, where his team studies metabolic and neurological disorders. For 25 years, they've used Sanger sequencing for research studies to define the molecular basis of rare neurodegenerative disorders, such as cerebellar ataxias, spastic paraplegias, and hereditary neuropathies. Three years ago, his team switched from Sanger to targeted next-generation sequencing (NGS) on the MiSeq System, developing a unique workflow using different targeted multigene panels to identify potential genetic variants associated with these neurodegenerative diseases.

NGS enables Dr. Taroni's group to identify variants in a fraction of the time and at a significantly lower cost than Sanger sequencing. In 2016, the team discovered novel mutations that cause Charcot-Marie-Tooth disease, a hereditary neurological disorder affecting peripheral nerves.[1-3] They were also able to screen many subjects with cerebellar ataxia in order to identify mutations in the giant gene *SYNE1*.[4] Dr. Taroni's team has several papers in review that focus on the discovery of novel variants associated with leukodystrophy and peripheral neuropathy.

iCommunity spoke with Dr. Taroni about his use of targeted multigene sequencing panels in his research studies, and how the MiSeq System is expanding and accelerating his research efforts.

Q: What first interested you about neurological and metabolic diseases?
Franco Taroni (FT): I trained as a neurologist in medical school. When I started working in the laboratory in the early 1980s, I focused on the biochemical and metabolic aspects of neurological and metabolic diseases. Solving the difficult problems that these diseases present is an exciting challenge.

Q: What is the mission of the Carlo Besta Neurological Institute?
FT: Research at the Institute is focused on neurological and metabolic disorders in children and adults. We see examples of rare genetic diseases that we try to decipher using sequencing to identify novel variants.

Q: What's the research focus of your team?
FT: Our team has a long-standing interest and expertise in spinocerebellar ataxias, which are disorders of the cerebellum and the afferent and efferent pathways that lead to movement disorders. These diseases are different from Parkinson's disease

because, in addition to tremors, there's also a lack of coordination of movement.

We also focus on spastic paraplegias, which is a disorder of the corticospinal motor pathway, and peripheral nerve and myelin disorders that impact myelin formation in the central nervous system. Neurological system disorders are characterized by heterogeneous genetic etiology, often with more than 50 genes responsible for each condition. For example, for a subject presenting with a cerebellar dysfunction who we suspect of having a genetic disorder, there might be more than 80 different genetic causes responsible for the condition.
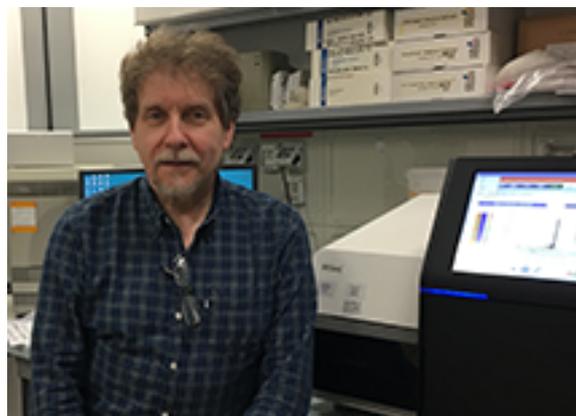
Q: How long did your lab use Sanger sequencing?
FT: We started using Sanger sequencing in 1989 and transitioned to using NGS three years ago when we acquired the MiSeq System.

Q: What prompted you to make the transition from Sanger sequencing to NGS?
FT: The heterogeneity of neurological and metabolic diseases caused us to consider using NGS in our research. Up to 90–100 genes can contribute to the phenotypes of the diseases we're studying. With Sanger sequencing, we could test for only 7–12 genes at a time. Using Sanger sequencing to analyze 80 genes is not affordable.

The NGS approach enabled us to analyze many genes simultaneously and cost-effectively, helping us to understand the complexity of these diseases. We now use Sanger sequencing to validate variants that we identify using NGS.

Franco Taroni, MD, heads the Unit of Genetics of Neurodegenerative and Metabolic Diseases at the Carlo Besta Neurological Institute (INCB) in Milan, Italy.

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

Q: Why did you choose to perform targeted multigene panel testing rather than whole-genome or exome studies?

FT: As a first approach, whole-genome sequencing (WGS) studies are expensive for us and bioinformatically complex. The additional expense of whole-exome sequencing (WES) is not warranted because our priority is to have a panel with virtually 100% coverage of the genes of interest. We chose to create five targeted sequencing panels for different diseases, including ataxia and spastic paraplegia, neuropathy, leukodystrophy (central myelin disorder), amyotrophic lateral sclerosis, and epilepsy. In addition to being less expensive, targeted multigene panel testing eliminates the complex issue of incidental findings.

> "With NGS, it takes approximately two days to analyze one subject sample for 200 genes. If we used Sanger sequencing, it would probably take months to complete the same analysis."

Q: What is the NGS-based targeted sequencing workflow that you use to identify variants associated with neurology and metabolic diseases?

FT: Our NGS workflow is based on the MiSeq System, which can process 12 to 24 subject samples per run. We start with an inexpensive, small prescreening panel and shift to larger, targeted multigene panels when necessary. Depending on the number of genes we are targeting, we use Nextera® XT DNA or Nextera Rapid Capture Custom Enrichment Library Prep Kits to prepare samples for the multigene panels.

First, we screen samples with several genes that are most commonly mutated for the disorder the subject is presenting. We capture 25–35% of pathogenic variants with this targeted screen. The gene set differs depending on the disease or phenotype that we're interested in analyzing. For instance, ataxia subjects typically have 4–6 commonly found genes, and those are the ones that we screen for first. For these narrow targeted multigene panels, we use the Nextera XT DNA Library Prep Kit to prepare the subject samples and create amplicon libraries.

If the screen is negative for these genes, we cast a wider net, performing a second round of targeted sequencing with a multigene panel. Each panel consists of virtually all the genes associated with a specific disease (70–200 genes per panel). For example, a panel for ataxia includes all the genes that are known to cause ataxia as a main symptom. The leukodystrophy panel includes all the genes that cause demyelination in the central nervous system, and so on. For these broad targeted multigene panels, we use the Nextera Rapid Capture Custom Enrichment Kit. This kit is useful, particularly when the subject has a continuity of phenotypes with many genes involved. For example, someone with ataxia might also have spastic paraplegia features and vice versa. It's an advantage to have a single panel, rather than two separate panels to test for various disease genes. With

Nextera Rapid Capture Custom Enrichment, we can also analyze copy number variation, enabling us to pinpoint deletions or duplications that might cause the disease.

If we don't identify the presence of disease variants using the disease panel, we turn to exome sequencing to cast an even wider net to identify variants of interest.

Q: How do you incorporate new genes into your panels?

FT: Our panels are regularly redesigned. Each time we reorder them, we have the opportunity to add new genes that we, and others, might have identified in our research. Although the Nextera Rapid Capture Custom Enrichment Kit allows for many genes, we sometimes have to take out some genes to make room for new ones.

Q: Why did you choose the MiSeq System to perform your targeted resequencing studies?

FT: We chose the MiSeq System based on our budget and its reliability in terms of data quality, particularly in sequencing through long stretches of repeat bases. The speed and versatility of the system also factored into our decision.

> "Depending on the number of genes we are targeting, we use Nextera XT DNA or Nextera Rapid Capture Custom Enrichment Library Prep Kits to prepare samples for the multigene panels."

Q: How do you analyze the data?

FT: Data analysis begins with VariantStudio Software. The data then enters a pipeline that includes some publically available commercial software, such as the CLC Genomics Workbench and Database. We use Database to filter the variants, and several other tools, including EVS, HGMB, ExAC, and 1000 Genome.

We've built our own local database to study variants found in the Italian population. Most publically available databases are based on North American or Anglo-Saxon genotypes. Our database addresses the Italian variability of polymorphic distribution effectively, which is important for bioinformatic analysis.

Variants are also analyzed with bioinformatic tools for their predicted functional effects, eg, Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen) for amino acid substitutions, NNSplice and Alternative Splice Site Predictor (ASSP) for RNA splicing. In vitro functional evaluation of protein and RNA levels is often required to validate the in silico prediction. Although we try to validate the variants using wet lab techniques, this is not always possible. There often isn't a functional assay for the particular gene we are assessing.

Q: How does the output and turnaround time of NGS-based panels on the MiSeq System compare to using Sanger sequencing to identify variants?

FT: NGS is superior to Sanger sequencing in terms of the number of genes analyzed and the complexity of the output of the report and analysis. In the Sanger era, we were able to sequence only one gene at a time and produced a report for the particular sequencing section of that specific gene. The turnaround time was quite long and we could sequence only a limited number of genes.

With NGS, it takes approximately two days to analyze one subject sample for 200 genes. If we used Sanger sequencing, it would probably take months to complete the same analysis.

> "We have identified novel genes and phenotypes quickly using a targeted sequencing process with the MiSeq System."

Q: What genes have you uncovered using this process?

FT: We have identified novel genes and phenotypes quickly using a targeted sequencing process with the MiSeq System. We designed our panels to be as inclusive as possible, which means that they even include uncommon genes that cause a particular phenotype of interest. Sometimes we come across unexpected results, such as a mutation in a gene that usually is attributed to a phenotype that is different than how the mutation is presenting in the subject sample. In those cases, we have found a new phenotype in already known genes.

In two cases, we identified novel genes with exome sequencing. They were samples from subjects with myelin disorder that had tested negative for the initial screen and panel sequencing. The mutations are relatively rare. We're completing studies of the two genes and the work is as yet unpublished. It's always interesting to find novel genes associated with a phenotype.

We're also working on two papers that report the results of interesting cases where we identified many mutations with our leukodystrophy and peripheral neuropathy panels. For subjects with rare diseases, it is very frustrating not knowing the cause of the disorder. Thus, increasing our potential analysis rate is very important.

Q: How has your lab perceived the transition from Sanger to NGS?

FT: It has been a dramatic change switching from Sanger to NGS. NGS is more complex and not everyone is using this new technology in a completely autonomous way. Relatively few members of our team have prepared a Nextera Rapid Capture library or analyzed the results, but many have created Nextera XT libraries and performed sequencing runs. It's changed the culture of our lab and the way people look at the possibility of this system.

Q: What are the next steps in your research?

FT: We'd like to work with clinicians to acquire samples from patients, family members, and other relatives and create a database of the results. It would make a significant difference in our research. It will allow us to have a detailed database of normal and pathological variants in the Italian population, thus enhancing the filtering power in sequencing projects. It isn't something that we can do ourselves, because we do not deal directly with patients.

We just joined a European project called Beyond the Exome. The scope of the project is to integrate different research approaches (RNA sequencing/transcriptomics, proteomics, and WGS) to unsolved cases that have already undergone exome sequencing. We'll be extending our sequencing studies to single cells and performing functional analyses. We're looking forward to performing RNA-Seq on the NextSeq®500 System.

Learn more about the Illumina systems and products mentioned in this article:

MiSeq System, www.illumina.com/systems/sequencing-platforms/miseq.html

NextSeq 500 System, www.illumina.com/systems/sequencing-platforms/nextseq.html

Nextera XT DNA Library Prep Kit, www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html

Nextera Rapid Capture Custom Enrichment Kit, www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-rapid-capture-custom-enrichment.html

VariantStudio Software, www.illumina.com/informatics/research/biological-data-interpretation/variantstudio.html

References

1. Corrado L, Magri S, Bagarotti A, et al. A novel synonymous mutation in the MPZ gene causing an aberrant splicing pattern and Charcot Marie-Tooth disease type 1b. *Neuromuscular Disord.* 2016; 26(8):516–520.

2. Piscosquito G, Saveri P, Magri S, et al. Screening for SH3TC2 gene mutations in a series of demyelinating recessive Charcot-Marie-Tooth disease (CMT4). *J Peripher Nerv Syst .* 2016; 21:142–149.

3. Piscosquito G, Magri S, Saveri P, et al. A novel NDRG1 mutation in a non-Romani patient with CMT4D/HMSN-Lom. *J Peripher Nerv Syst.* 2016; Dec 16. doi: 10.1111/jns.12201. [Epub ahead of print]

4. Synofzik M, Smets K, Mallaret M, et al. *SYNE1* ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study. Brain. 2016;139:1378–1393.

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

illumına®