

# The Time is Now for Microbiome Studies

Whole-genome shotgun sequencing and transcriptomics provide researchers and pharmaceutical companies with data to refine drug discovery and development.

## Introduction

Joseph Petrosino, PhD has found himself in the right place at the right time not just one time, but three times since he arrived at Baylor College of Medicine in 1993. Each time taking a step towards becoming an integral part of the initiation and later expansion of microbiome studies from academia to the commercial arena.

First, he was at the Human Genome Sequencing Center at Baylor when the Common Fund Human Microbiome Project (HMP) was announced. Because it was a National Human Genome Research Institute (NHGRI) initiative, genome centers were spearheading the jumpstart phase. Richard Gibbs, the director of the Human Genome Sequencing Center, asked Dr. Petrosino to help the Center build the initiative at Baylor. With his background in bacterial genomics, Dr. Petrosino “was very intrigued to leverage next-generation genomics platforms to provide a census and functional capacity of microorganisms from any type of sample, without having to cultivate them.”

When the HMP efforts wound down in 2010, Dr. Petrosino approached the college with the idea of building a translational microbiome research center. The goal was to move past genomics and discovery-based pipelines and into the dissection of host-microbial relationships, particularly relating the functions in microbial communities back to disease. The Center for Metagenomics and Microbiome Research (CMMR) was founded a year later with Dr. Petrosino as its Director.

Soon the CMMR was receiving requests from pharmaceutical companies to assist with microbiome projects. To handle the tighter deadlines required for commercial projects, Baylor created a microbiome services company, called Diversigen, in 2015. Dr. Petrosino was the company’s Chief Scientific Officer, spending about 10% of his time acting as a resource for Diversigen projects and the other 90% leading CMMR.

The CMMR “benefits from Diversigen activities,” according to Dr. Petrosino. “As a commercial entity, Diversigen enables the CMMR team to relate its knowledge to people who are on the frontline of developing new therapies and diagnostics. As a result, it’s accelerated the translation of CMMR academic discoveries into beneficial commercial products.” Recently, Diversigen was acquired by Orasure Technologies, as they compile an end-to-end service offering in microbiome analytics.

iCommunity spoke with Dr. Petrosino about the value of microbiome research, how study design strategies and the use of NGS has evolved over the last decade, and what the future holds.

## Q: What was the original vision of the Human Microbiome Project?

**Joseph Petrosino (JP):** There were many investigators around the world who were studying commensal organisms and who had begun implementing next-generation sequencing (NGS) technologies for microbiome studies. The NIH recognized that the microbiome could impact many disease areas and that would impact numerous NIH institutes. They also realized that there was a need for developing methodologies, strategies, and best practices concerning which tools were best suited for microbiome studies and for what purposes. The HMP effort closely paralleled the Human Genome Project in terms of gathering data, understanding best practices and methods, and developing a reference data set and protocols that investigators could use to build their own research programs.

## Q: What did you learn in setting up the HMP protocols?

**JP:** Our HMP effort was focused on building a cohort of 300 subjects from the Baylor College of Medicine and Washington University in St. Louis, two core clinical centers. Much time was spent on the clinical side in developing the recruitment protocol. The 300 subjects would be sampled at up to 18 body sites, up to three times in a 12-month period.

From a methodology standpoint, we identified the best extraction methods that could work with the many different sample types that were collected. Understanding how to benchmark and validate those protocols was a more refined and rigorous effort than we initially thought it would be. The first observation that we made from those 300 subjects was that even though organism diversity was different from person to person, functional components were conserved at each body site.



Joseph Petrosino, PhD is the director of the CMMR and Chief Scientific Officer of Diversigen.

**Q: Why was Diversigen founded?**

**JP:** The Center for Metagenomics and Microbiome Research was created to perform translational academic research in a medical center environment. However, we had several pharma commercial teams come to us with microbiome projects. While they were supportive of academic research, they weren't comfortable with, for example, a first-year graduate student working on a project for them. They wanted projects completed on tighter timelines with a seasoned individual who could troubleshoot and analyze data sets. Diversigen was launched in 2015 to address the needs of commercial entities for microbiome studies.

When I launched the Center, I had pointed out the potential for building a microbiome services company to give us a window into how pharma was looking at the microbiome. As academics, we like to think we know best in our own field, and where the field should grow to best serve public health. However, pharma may have a much different viewpoint from their vantage. Diversigen gives us a lens into that space and has changed how we think about the microbiome. It's been a lucrative operation from the get go.

**Q: What is unique about Diversigen?**

**JP:** Diversigen prides itself in being a Rolls Royce microbiome service provider to commercial customers, particularly pharmaceutical and biotech companies. The academic lab can be thought of almost as part of the Diversigen research and development department, enabling the company to offer state-of-the-art frontline applications and mature strategies to deliver data that compare favorably to what we produce academically. Diversigen has since built an extensive internal analysis team that has developed its own best practices that are ideally suited for commercial studies.

The Diversigen team engages with customers from the concept stage through results and beyond. We collaborate on study design, provide advice on how to collect, store, and ship samples, and recommend which sequencing methodology will best provide the answers they need.

"The ability to multiplex on the NovaSeq 6000 System has helped us quite a bit. We'll continue to use the NovaSeq for more applications."

There's been significant growth in microbiome companies, with many founded by microbiome experts straight from academia. Many of these companies, especially in the pharmaceutical arena, don't have a cadre of microbiome analytics experts on hand. Having worked on over 400 academic projects, with over 200 collaborators around the world, we have significant insights on various diseases and the potential role the microbiome plays in

them. These companies rely on the Diversigen team to give them feedback, advise, and consult with them on next steps and potential downstream applications.

"We have turnaround times of days rather than weeks. Illumina NGS systems enable us to schedule projects ahead of time so that the sequencers are constantly running."

**Q: What is the average sample number for Diversigen projects?**

**JP:** We used to perform studies with 5-20 samples. Currently, most of our projects have hundreds of samples, with some in the thousands. We've established best practices on how to manage projects with large cohorts that are sampled longitudinally. This includes how to collect, store, and batch these samples for maximum efficiency while reducing confounding noise in the data.

**Q: Are you working on clinical trials with pharmaceutical customers?**

**JP:** When we first started working with pharmaceutical companies, we began developing Clinical Laboratory Improvement Amendments (CLIA) and College of American Pathologists (CAP)-certified pipelines. We wanted our pharma customers to feel comfortable that the data we produced were of high quality, reproducible, validated, and ideal for regulatory submissions.

Several of our pharmaceutical company partners are entering the clinical trial stage and we're performing sequencing and analytics for some of these trials. These are often multiyear projects where additional best practices are being employed so that the most information possible can be extracted from the resulting data.

**Q: What are the sample types and sizes that you receive typically?**

**JP:** The microbiome field's favorite sample type is still stool. The advent of stabilization reagents means that it's no longer necessary to ship samples on ice to maintain sample stability, and that's made things easier.

Of course, we have many projects where we receive subideal, low-biomass samples. Because of our expertise in building the HMP protocols, we have an established team that develops and validates protocols for difficult sample types such as these. Our goal is to obtain meaningful data from whatever type of sample we receive, no matter how small.

**Q: What types of sequencing do you use in Diversigen projects?**

**JP:** We perform 16S sequencing on multiple variable regions within the 16S gene, depending on the community that's being analyzed. Certain variable regions provide better resolution for specific types of communities, such as oral vs. skin vs. stool.

We perform internal transcribed spacer 2 (ITS2) sequencing in fungal and eukaryotic communities and 18S sequencing for microeukaryotes. We use whole-genome shotgun sequencing for

metagenomic analyses, microbial community strain, and species- and functional-level determinations. We also perform long-read sequencing. Depending on the question and the applications, we sometimes use long-read and short-read sequencing together to provide better sequence quality for single-strain sequencing, just like in the old days, but now it's faster and cheaper.

We have a transcriptomics or metatranscriptomics pipeline that we are continuing to refine from a methodology and analytical standpoint. Viral metagenomics are also a significant part of what we do. We have spent a significant amount of time enhancing our ability to analyze viral metagenomic samples from a clinical perspective. Some of these samples are vanishingly small and spinning down the viruses and purifying them in a sedimentation gradient, the way we would in a traditional virology laboratory, is not possible or scalable. We take a 'less is more' approach in terms of sample handling, using analytics to pull out as much viral information as possible from these difficult samples.

"Given the size of a typical microbiome project, the capacity of the NovaSeq 6000 System will keep us in business for a long time."

**Q: How have study design strategies and the use of NGS evolved over the last five years?**

**JP:** In early microbiome studies, during the HMP for example, people were often leveraging samples collected for other purposes to perform analyses that were cross-sectional, rather than longitudinal, for a particular cohort. Samples may not have been collected or stored properly for microbiome analyses, and as we have learned since, single timepoint analyses don't always provide the full picture of what is transpiring in a given microbial community. Now, an increasing number of projects involve well-matched cases and controls for a particular disease and use preclinical models to dissect the functional mechanisms underlying microbial associations with health and disease.

Also, as companies realized the limitations of 16S sequencing and have increased their budgets for microbiome-based studies, we've seen a shift to whole-genome shotgun sequencing to obtain functional gene pathway and strain-based information to better study the microbial communities that are present. Likewise, we are also seeing an increase in transcriptomics, which enables the assessment of which genes are turned on at the particular time when a sample is collected.

**Q: What can metagenomic and the transcriptomic data tell you about a disease?**

**JP:** Metagenomics and metatranscriptomics let us determine what microbes and functions are present in a healthy individual or associated with a particular disease, and how they relate to that disease. With longitudinal samples, the data can be associated with the onset or exacerbation of disease. With transcriptomics,

specifically, one can assess if gene expression is associated with a particular disease or exacerbation state.

**Q: Does the data from metagenomic and transcriptomic studies provide targets for therapeutic or diagnostic development?**

**JP:** Yes, it's important to note that 'omics-based approaches enable target discovery. We realize that we need to go back and perform mechanistic and functional studies, and create or leverage preclinical models to understand why 'omics-based differences exist and how they impact the disease itself. Data from these studies inform decisions to move forward with a potential therapeutic or diagnostic candidate and provide information about whether a drug candidate is impacting the microbiome.

**Q: How much has your customer base grown since 2015?**

**JP:** Our customer base has grown exponentially and there's no sign of it slowing down. There's an explosion in the number of companies that are creating and advancing microbiome-based therapeutics. Pharma companies are also embracing pharmacomicrobiomics or how drugs are metabolized by the microbiome. So, even developers of chemical-based drugs are interested in whether they are modified by the microbiome.

**Q: What Illumina NGS systems are you using?**

**JP:** We have an iSeq™ 100 System, a couple of MiSeq™ Systems, a NextSeq™ 500 System, a HiSeq™ 3000 System, and a NovaSeq™ 6000 System. From our laboratory standpoint, we love them and they provide us with a lot of flexibility.

A project's read length, data, and multiplexing requirements dictate which pipeline we use. We use the iSeq 100 System for quality control and validation of new protocols. It also provides us with a cost-effective tool to perform phage or small-genome sequencing projects with a fast turnaround. We use the MiSeq Systems for amplicon-based sequencing. The NextSeq 500, HiSeq 3000, and the NovaSeq 6000 Systems are used to perform RNA-Seq and metagenomic sequencing. We have also validated various robotic handlers and an Echo instrument to help us build sequencing libraries.

**Q: How has the NovaSeq 6000 System performed in your studies?**

**JP:** The NovaSeq 6000 System has been a recent addition and its different flow cells enable us to be flexible with the types of sequencing we use for projects. It's running well, with new barcodes being added as we speak. The ability to multiplex on the NovaSeq 6000 System has helped us quite a bit. We'll continue to use the NovaSeq for more applications. It's been exciting to add it to our sequencer fleet.

**Q: Is sequencing efficiency important?**

**JP:** Efficiency is absolutely essential to the success of Diversigen. We have a small, experienced team of 12-17 people working in sample intake all the way through primary analysis. We have turnaround times of days rather than weeks. Illumina NGS systems enable us to schedule projects ahead of time so that the sequencers are constantly running.

"We use the iSeq 100 System for quality control and validation of new protocols. It also provides us with a cost-effective tool to perform phage or small-genome sequencing projects with a fast turnaround."

**Q: What are your goals for the next five years at CMMR and Diversigen?**

**JP:** We want to continue to grow and are interested in developing intellectual property (IP) for some of our CMMR projects. At CMMR, we've been involved in cancer immunotherapy and social anxiety projects, among others, and have conducted several clinical studies. We'd like to identify ways in which we can apply these initial observations into building cohorts that can help us reveal additional candidates to take into diagnostic and therapeutic development through a Baylor/Diversigen partnership. Diversigen also wants to become more IP driven with the technologies and analytics that it's developed.

**Q: How do you see NGS and Illumina sequencing systems impacting the future of microbiome studies at Diversigen?**

**JP:** In the short-term, amplicon sequencing is a great way to profile communities. The shift away from 16S sequencing to whole-genome shotgun sequencing will continue as the cost of sequencing decreases and we become better at multiplexing and building analytics, and discovering biomarkers to profile microbial communities effectively.

Given the size of a typical microbiome project, the capacity of the NovaSeq 6000 System will keep us in business for a long time. As the commercial business grows, we might need a second one to enable us to keep multiplexing projects at a rate that supports optimal turnaround times.

### **Learn more about the systems mentioned in this article:**

NovaSeq 6000 System, [www.illumina.com/systems/sequencing-platforms/novaseq.html](http://www.illumina.com/systems/sequencing-platforms/novaseq.html)

iSeq 100 System, [www.illumina.com/systems/sequencing-platforms/iseq.html](http://www.illumina.com/systems/sequencing-platforms/iseq.html)

NextSeq 500 System,\* [www.illumina.com/systems/sequencing-platforms/nextseq.html](http://www.illumina.com/systems/sequencing-platforms/nextseq.html)

MiSeq System, [www.illumina.com/systems/sequencing-platforms/miseq.html](http://www.illumina.com/systems/sequencing-platforms/miseq.html)

\*The NextSeq 500 System has been updated to the NextSeq 550 System.

**Illumina, Inc.** • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • [techsupport@illumina.com](mailto:techsupport@illumina.com) • [www.illumina.com](http://www.illumina.com)

© 2020 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html). Pub. No. 770-2019-025-A QB9280

**illumina**<sup>®</sup>