

0 5 . 0 9 . 1 7

# WorldQuant Perspectives

F E A T U R E D I N T E R V I E W



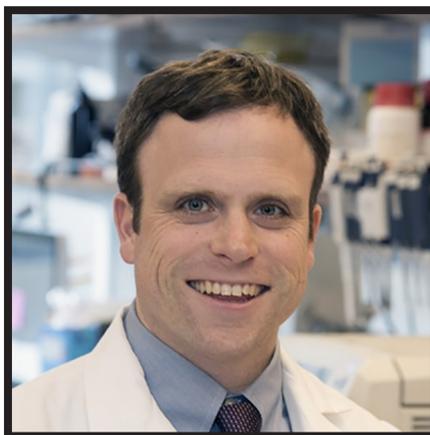
## USING BIG DATA TO MAKE BIG STRIDES IN MEDICINE

Weill Cornell Medicine's Christopher Mason explains how a new initiative funded by WorldQuant founder Igor Tulchinsky will help medical researchers make better predictions.

**WORLDQUANT**<sup>®</sup>

WorldQuant, LLC  
1700 East Putnam Ave.  
Third Floor  
Old Greenwich, CT 06870  
[www.weareworldquant.com](http://www.weareworldquant.com)

IN THE SPRING OF 2014, WEILL CORNELL Medicine associate professor Christopher Mason received an email from the school's then dean asking if he would have lunch with a potential donor who was interested in big data. Mason, who has a Ph.D. in genetics from Yale University, oversees a lab focused on developing novel techniques for DNA sequencing and algorithms to study the human genome and tumor evolution. As they dined on grilled fish at Weill Cornell's Griffis Faculty Club, Mason explained to the potential donor, WorldQuant founder Igor Tulchinsky, that his lab was busy swabbing hundreds of New York subway stations for DNA samples to create a microbial map of the city. He also talked about the work they were doing with NASA, including the Twins Study of astronauts Scott and Mark Kelly, to investigate "how microgravity and other space-related environmental factors influence changes in RNA and DNA." After lunch Mason took Tulchinsky on tours of his team's wet lab, showing him the tiny nanopore DNA sequencer that NASA is using in space, and the dry lab, where the computational analysis takes place. "I thanked him for visiting, offered an open-ended invitation to the lab, and that was kind of it," says the Wisconsin native, who learned a few months later that Tulchinsky had donated \$1 million to Weill Cornell Medicine to fund an annual \$50,000 WorldQuant Foundation research scholarship.



CHRISTOPHER MASON  
WEILL CORNELL MEDICINE

Since then Mason has received the scholarship three times and has become good friends with Tulchinsky as the WorldQuant founder's interest in applying data science outside finance has grown. "It's been really great," says Mason, sitting in his crowded office, whose walls and door are covered with newspaper articles, posters and clippings from scientific journals chronicling his obsession with data. "What started out as a hyper-technical lunch conversation about everything from how you measure antibodies or immune response to microbiomes and space has turned into an incredible collaboration."

This spring Tulchinsky took the next step in that collaboration by donating \$5 million to establish the WorldQuant Initiative for Quantitative Prediction at Weill Cornell Medicine. Mason and co-director Olivier Elemento — a Weill Cornell associate professor who specializes in applying big data to understanding

and treating cancer — will oversee the research initiative housed in the HRH Prince Alwaleed Bin Talal Bin Abdulaziz Al-Saud Institute for Computational Biomedicine at Weill Cornell Medicine, which will also collaborate with scientists and investigators at the institution's Caryl and Israel Englander Institute for Precision Medicine and the Sandra and Edward Meyer Cancer Center. Mason recently met with WorldQuant Global Head of Content Michael Peltz to discuss the genesis of the initiative, the goals it hopes to achieve and the role that WorldQuant researchers may play in reaching them.

WHEN DID YOU AND IGOR  
START TALKING ABOUT THE  
INITIATIVE FOR QUANTITATIVE PREDICTION?

**Christopher Mason:** We have been discussing it for a few years, but in the past six to nine months the seed really sprouted. If we had a paper come out, I'd send Igor all the cool highlights from the lab and think about how we could combine them with financial models. We started having dinners as well with Lew Cantley, who is the Meyer Director of the Meyer Cancer Center, and Olivier Elemento, the co-director of the initiative, and then Igor said: "What's the next big thing? What is the biggest challenge?" I told him, "Well, the next big thing is us being on Mars, but that's a bit further away." The next proximal big thing is a result of the fact that we're actually in a time of euphoria in terms of genetics discoveries and applications. We can monitor and sequence and profile individual cells and the communication between cells. We can also examine every mutation in any cancer cell. We can see all this new information, but the two biggest challenges are, one, technologically the ways you can actually capture and correctly profile individual cells or individual molecules from patients and then, two, the data analytics — the way that we model changes in cancer. In other words, when you sequence a piece of DNA, how much can it tell you forensically about where it comes from in the world and what potential risk or signal it can be for medicine?

Igor and I have talked a lot about those two big challenges and about what the Elemento lab is doing, just down the hall from our lab. Both of our labs do a lot of big data and modeling, and Olivier has been trying to predict why drugs fail in clinical trials.

So, from our perspective, we want to look at a sequence of DNA that comes from a particular cell in the body and understand everything we possibly can about it. What does it mean for your risk for disease based on the mutation state of it? What does it indicate about health? What drugs will work? To answer those questions, you'll need big data and better technology.

#### HOW DID THE IDEA FOR THE INITIATIVE EVOLVE?

We started talking about the need at the Meyer Cancer Center to jump-start some of the more cutting-edge work to model disease. And that became a discussion of what that would entail. Would it take cash? Cash is needed for everything, of course.

Then it became a discussion about whether the initiative could lead to creating new models. At WorldQuant they build models all the time — what they call alphas. They're building thousands and thousands of models that last for days. We're building one or two models that last for years. It's like this inverted modeling paradigm in which we want to see if we can actually learn from each other.

So it became this idea: There are many Ph.D.s, computer scientists, programmers and IT professionals in both places; they are similar types of people, using the same programming languages, but they're just using their brains on different kinds of data sets. What we could learn from each other became some of the discussion. Could we make a large experiment that blends quantitative finance and quantitative genomics?

#### WHAT WILL THE \$5 MILLION BE USED FOR?

A mixture of both people and hardware. We have some of the expertise here, but we have to bring in more people. The gift will allow us to hire five more people over the next few years —

*“ IT'S GOING TO BE THE FIRST INITIATIVE THAT BLENDS BIG DATA IN CANCER BIOLOGY, THE MICROBIOME, AND BIG DATA IN WHAT'S CALLED METAGENOMICS, OR ALL THE GENOMES. ”*

*“ WE'VE CREATED AN INFRASTRUCTURE THAT CAN MODEL PATIENT RISKS AND PREDICT WAYS TO DEFEND AGAINST THEM. IT SOUNDS A BIT GRANDIOSE AND A LITTLE PIE-IN-THE-SKY, EVEN TO SKIES ON MARS, BUT IT'S TRUE. ”*

extraordinarily nimble and smart programmers — to do some of the modeling. We also need to purchase more hardware to be able to look at individual cells or do a complex analysis of tumor cells. One of the instruments we're looking at, an imaging mass cytometer, costs \$1 million just for the hardware. It lets you take a complex tumor sample, for example, and map everything that's there — DNA, RNA, protein — in three dimensions. It's basically what pathologists do. The goal is essentially digital pathology — imaging what's inside cells and getting actual quantified data instead of a qualitative assessment of the data.

So you take a tumor sample — and for the Meyer Cancer Center that's the main focus — and you use dozens of markers instead of one or two markers of what's present in a cell mixture. The reason that is important is you don't just want to say, “Well, do I have cancer, yes or no?” or “Is it aggressive, yes or no?” You also want to be able to say whether the immune system has begun to respond: Can you see infiltrating macrophages, for example?

We'll also be purchasing more sequencing hardware — DNA sequencers and small molecule sequencers to look at what's in the cells. We'll use some of the funding to improve how fast we can sequence something; you can buy hardware for that, too. The high-throughput sequencers we're buying, optimizing and building are very rapid, so within ten minutes you could diagnose all the DNA molecule samples.

*AND AS PART OF THE INITIATIVE, WORLDQUANT RESEARCHERS WILL WORK IN YOUR LAB.*

They would come as visiting fellows, yes. The idea is that they will come here for six to 12 months. We will embed the WorldQuant researchers as data scientists, immersed in medical data; I want

them to come with new ideas. We could also send grad students or postgrads in the other direction to go spend a summer at WorldQuant. We'll make it a two-way transfer. The students would love it. They would be blown away. Quantitative finance has risen in appeal over the years because of some of the new and exciting approaches to big data.

#### WHAT DO YOU HOPE TO ACHIEVE?

When I first spoke to Igor, I said, "I want to be able to predict everything about anything," which is a bit too broad. But the goal is embedded within every project in the lab. What I mean is that for medicine, for most of biology, we often don't know the most informative places to look. Where and what kind of molecular signatures are the most indicative for health or disease are still being discovered. So when I say I want to predict everything about anything, it means I want to construct frameworks that can leverage as much as possible to build a better view of cancer evolution, of how we understand infectious disease. It's going to be the first initiative that blends big data in cancer biology, the microbiome, and big data in what's called metagenomics, or all the genomes.

At the very end I want to feel like we have a better model that can actually help. We're already helping to some degree with NASA, which is why I put that poster on my door. Think about it: We are literally helping not only patient health and wellness here on Earth, but also the very exceptional astronauts that are going to be traveling to Mars. I feel like we've created an infrastructure that can model patient risks and predict ways to defend against them. That's it in a nutshell. It sounds a bit grandiose and a little pie-in-the-sky, even to skies on Mars, but it's true.

#### WHAT ARE THE DIFFERENCES WHEN IT COMES TO ANALYZING BIG DATA IN MARKETS VERSUS MOLECULES?

“GENERALLY, IN MEDICINE YOU HAVE TO BE RIGHT AT LEAST 80 TO 85 PERCENT OF THE TIME OR ELSE PEOPLE DON'T TAKE YOU SERIOUSLY.”

“THE TIME FRAMES FOR MARKETS AND MEDICINE ARE VERY DIFFERENT, BUT THE ALGORITHMS THAT ARE USED IN ANALYZING DATA IN SOME CASES ARE THE SAME.”

There is a kind of inverted relationship. Financial researchers have arguably more data; they have more data coming at a faster rate, but its utility is very short-lived. They modify or replace their models very quickly because the market changes. They have to readjust.

In some sense, we want to do the opposite. We want to have a model that helps you plan for 30 to 40 years. So we've had these funny discussions with WorldQuant about whether they could ever imagine a financial model that would be good and last for 30 years, which of course is almost impossible to do. But it's what's essential for predictive medicine in that you need to say, "Okay, given your genome at age 15 or 18, here is whether or not you should have a prophylactic bilateral mastectomy." If you have a BRCA1 mutation [which significantly increases the odds that a woman will develop breast or ovarian cancer], you have to decide whether you would literally remove parts of your body to save the rest of your body. And that's a decision based on a lifelong time frame.

#### WHAT ARE THE SIMILARITIES BETWEEN MARKETS AND MEDICINE?

The time frames for markets and medicine are very different, but the algorithms that are used in analyzing data in some cases are the same. For example, machine learning methods, such as recurrent neural networks, deep learning methods and support vector machines, are used in both fields. The importance of metadata is critical for both medicine and markets; you need to be able not just to measure something but also to have the context for it.

Another similarity is the belief that when you incorporate as many data types as possible, you often can get more power. That's what we're both trying to do. There's always this sense of "If I could just tease out one key feature that no one else has seen before, I will

have the best alpha, or model.” Or in medicine, “If I can tease out this new molecular signature, I’ll have a great paper, get a patent or I’ll see something new that no one has discovered before and help people avoid disease altogether.”

Also, I think there’s this reverence for discovery that pervades both fields, and the gratification is also really wonderful for both of us. Ideally, as we both succeed we can help each other’s clients. For example, if you have a really great alpha and you make more money for the firm, that could then help pension funds, which could help everyone retire earlier. And if we have predictive models that ensure people avoid disease and monitor health, they can stay healthy for much longer. If we both do well, people will be able to retire earlier and then have a longer time to enjoy that retirement. That is our dream.

**TO BE SUCCESSFUL, A QUANTITATIVE INVESTMENT FIRM’S PREDICTIVE MODELS NEED TO BE RIGHT A LITTLE OVER**

**HALF THE TIME. DO YOU NEED A MUCH HIGHER ACCURACY RATE IN MEDICINE?**

That’s a great question. It’s a good differentiator, actually. If you said, “I’ll correctly guess whether your cancer is aggressive or not about 51 percent of the time,” that would be awful. Generally, in medicine you have to be right at least 80 to 85 percent of the time or else people don’t take you seriously. You wouldn’t want to tell someone, “Well, you know, you might die or you might not; I’ll just go flip a coin.”

I would say that the accuracy of models is one case where markets and medicine diverge. I mean, what if we could be financially correct 85 percent or 90 percent of the time? That would obviously be great for investors, but I don’t know whether applying techniques from quantitative genomics and medicine will make that happen. It’s a really big difference. The implications are more stark. If you lose money, you can maybe make it back later. If you lose your life, you can’t — currently, anyway — get it back later. ◀

---

*Thought Leadership articles are prepared by and are the property of WorldQuant, LLC, and are circulated for informational and educational purposes only. This article is not intended to relate specifically to any investment strategy or product that WorldQuant offers, nor does this article constitute investment advice or convey an offer to sell, or the solicitation of an offer to buy, any securities or other financial products. In addition, the above information is not intended to provide, and should not be relied upon for, investment, accounting, legal or tax advice. Past performance should not be considered indicative of future performance. WorldQuant makes no representations, express or implied, regarding the accuracy or adequacy of this information, and you accept all risks in relying on the above information for any purposes whatsoever. The views and opinions expressed herein are those solely of the author, as of the date of this article, and are subject to change without notice, and do not necessarily reflect the views of WorldQuant, its affiliates or its employees. No assurances can be given that any aims, assumptions, expectations and/or goals described in this article will be realized or that the activities described in the article did or will continue at all or in the same manner as they were conducted during the period covered by this article. Neither WorldQuant nor the author undertakes to advise you of any changes in the views expressed herein. WorldQuant may have a significant financial interest in one or more of any positions and/or securities or derivatives discussed.*