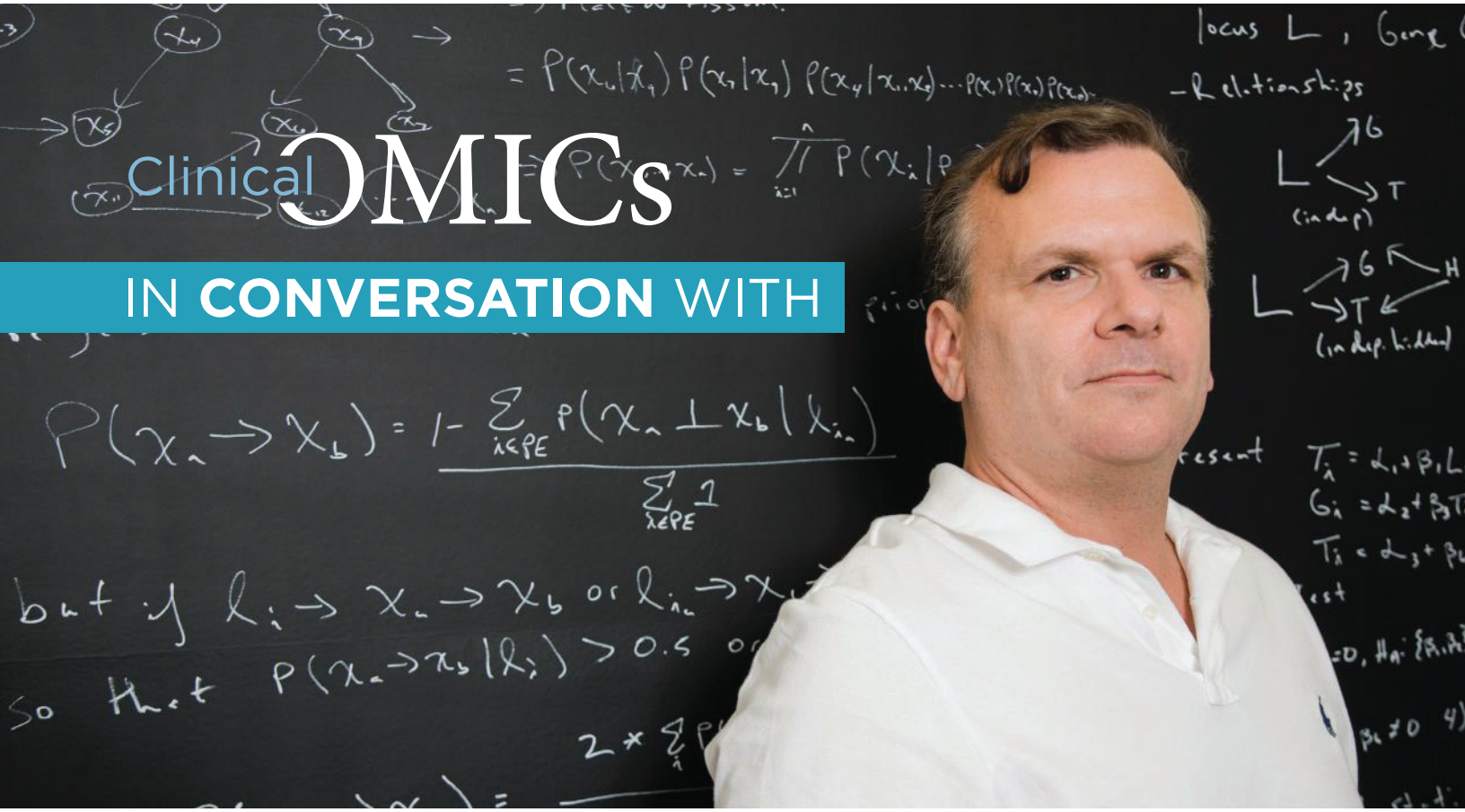


Clinical OMICs

IN **CONVERSATION** WITH

# Eric Schadt

## CEO, Sema4

Eric Schadt, Ph.D., founder and CEO of Sema4, is a mathematician and data scientist on a mission to collect as much patient data as he can with the ultimate goal of improving patient health. At Sema4, a tenet of the company is to put the patient in control of their own health data while providing a data-driven approach to fuel deeper analysis, and increased engagement to improve the diagnosis, treatment, and prevention of disease. Schadt serves as the Dean for Precision Medicine and Mount Sinai professor in predictive health and computational biology at the Icahn School of Medicine at Mount Sinai, where, immediately prior to founding Sema4, he was the founding director of the Icahn Institute for Genomics and Multiscale Biology, and professor and chair of the Department of Genetics and Genomic Sciences. He is an expert on constructing predictive models of disease that link molecular biology to physiology, a skill he has employed over the past 20 years, at pharmaceutical and life science companies including Merck, Rosetta, Sage Bionetworks, and Pacific Biosciences. Schadt took a few moments earlier this year to talk with *Clinical OMICs* Editor in Chief Chris Anderson about the founding of Sema4 and how it plans to help deliver actionable information for precision medicine.

***Clinical OMICs: Sema4 is a spinout of Mt. Sinai. Talk about the work that was being done while you were there that was the impetus for you to spin out a new company.***

**SCHADT:** I came to Mt. Sinai in late 2011, early 2012 with a charter to help the Mt. Sinai School of Medicine, now the Icahn School of Medicine, enter the modern era with respect to genomics, next-generation sequencing, and big data analytics. We built out the first CLIA-approved next-generation sequencing core in the state of New York and then hired an army of bioinformaticians to help understand how to look at the data off those instruments and how to integrate it with the digital universe of data.

I was lucky enough to get people like Jeff Hammerbacher, the founding chief data scientist at Facebook. He was classmates with Zuckerberg and helped bring Facebook to the big time. He then founded Cloudera which was created to bring Hadoop-style computing to the masses. We got him at a phase in his life where he wanted to get into the life sciences. He came for four years and helped build out web-scalable infrastructure and hired lots of statistical geneticists. His investment was in the AI-driven, probabilistic, causal reasoning areas.

We built out that engine and then hooked it up to clinical scientists by recruiting high-end investigators who had one foot in their disease focus and one foot in my department—people like Allison Goate out of Washington University for Alzheimer's and Judy Cho out of Yale for IBD, and on and on.

It all went really well. We helped put Sinai on the map for genomic and data sciences. I took them from nearly a bottom-of-the-barrel ranking in terms of NIH funding in genomics, to now they are number three.

What also was interesting was they had this clinical arm of the department. I very naively felt, as a mathematician, if I can own the clinical arm, then the doctors will just have to do what I tell them to do. I learned very quickly that it doesn't work that way.

But I learned that we better translate all this AI/machine learning stuff into the clinic. Being in that ecosystem helped me understand how medicine works. How do you work with physicians? How do they make decisions? How do you communicate complicated stuff to them to enable them to make more informed decisions?

That is where the genomics testing came in. Because of the information platform I built at Sinai, one of the first places we focused it was at next-generation sequencing diagnostics. That was driven by the clinical genetics arm of Sinai. It was all going well and we grew that internally, and it was making good cash to fuel the research.

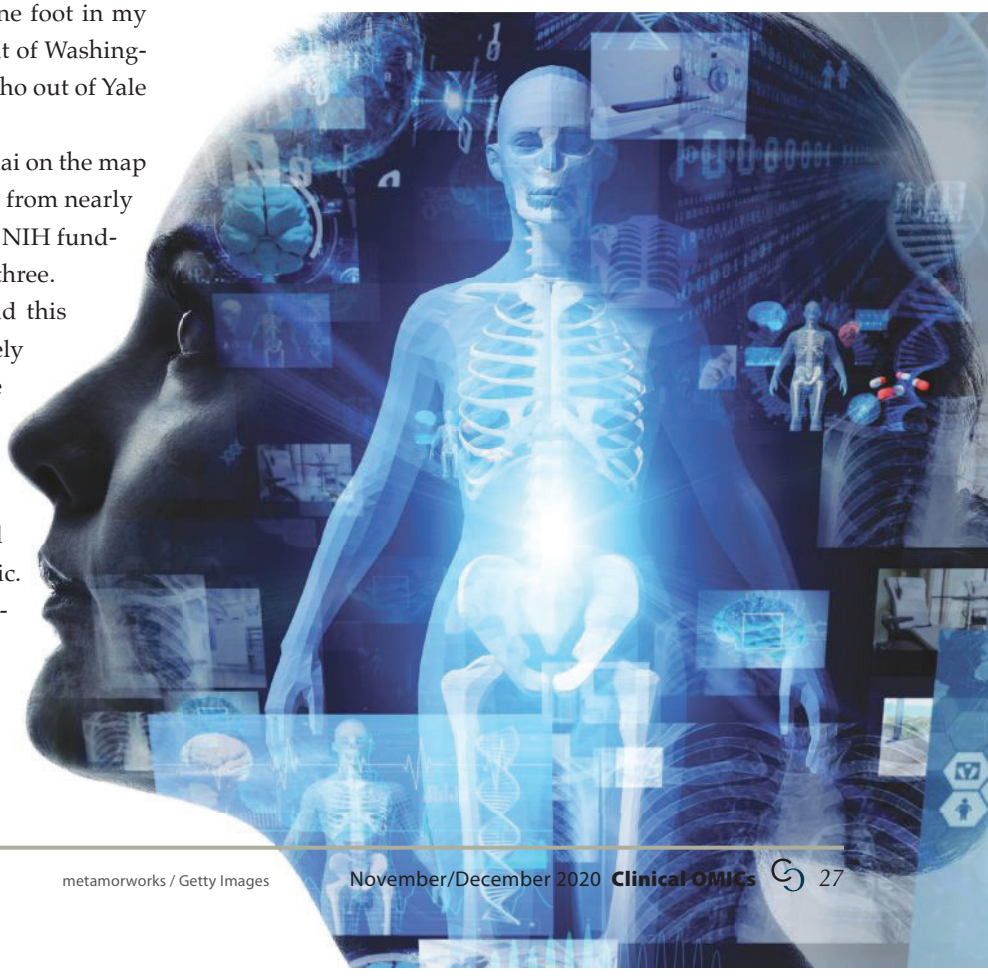
***If it was all going well, why did you decide spinout Sema4?***

**SCHADT:** It was two things. One was, "Gee, it sure was hard to convince clinicians to take all of these advanced analytics-derived insights and put them into practice." It was way harder than the success I achieved in pharma of informing on drug discovery programs. This was another level of hardness. When I took a step back I asked: "Why is that? Why don't they just want to change their standard of care?"

It was because the models themselves, while good enough to drive drug discovery, and good enough to get cool publications, they were not good enough to meet the clinical threshold of going into practice. That was a much higher bar.

But I wondered why weren't the models good enough? It turns out it wasn't algorithms, it was scale of

*(continued on next page)*



metamorworks / Getty Images

(continued from previous page)

data. Even with a system like Mt. Sinai's, and millions of patients of data, it simply wasn't dense enough and longitudinal enough to build the most accurate models for clinical decision making. If you wanted to do that, you would need to go beyond the walls of Mt. Sinai.

So Sema4 was driven by how can we leverage what we are doing in testing, how do you engage physicians and patients through standard of care and have them as partners? Because if you are operating as standard of care you can use that as a growth hack engine to go national, to get to the number of patients and data needed to play the machine learning/AI game. That was the drive, and of course the amount of capital you need to scale that and continue to grow—and then the data science and the software engineering you need to better engage physicians and patients digitally—that is real money. Forming Sema4 was to help generate the fuel to do that.

> **"If I could see what is going on in every single cell, in every part of the body across the entire human population, across all the dimensions of molecular profiling, I would do it."**

***Has that focus changed at all since the company was founded?***

**SCHADT:** It definitely has not changed. The chase for the bigger scales of data is still on. What has changed is that while I had viewed patients as one of the growth hack engines to the data, I hadn't appreciated as strongly that doing precision medicine-based deals with health systems was an efficient way to get access to scales of data too.

The data is not as good, because it is what you are pulling out of their EMR. You are not tracking the patient every day, but it is still useful. So I would say we have layered on additional ways to growth hack data.

***Is this similar to what Flatiron Health is doing, where they leverage data derived from their customer's EMR to improve their clinical support engine?***

**SCHADT:** Well it is different, but it also has some overlap. Certainly, the EMR-based data we are getting through partnerships with health systems is similar because we are taking the EMR data. In many of those collaborations, we have access to de-identifiable data where we help abstract information knowledge from the unstructured data like physician notes. We then structure those data and link them back to the structured EMR data to provide ways for those systems to leverage that information to have a better understanding of their patient population. In that sense, it is similar.

But we have a lot of other components in play that you need to wire together in order to get to a precision medicine solution for a health system, that a Flatiron doesn't have. In women's health, newborn screening, drug safety, heritable cancers, somatic cancers, we are the information-driven diagnostic testing partner for those institutions.

We directly engage the patients, since we are doing the genetic counseling, and we are delivering the test results. When you do that, you have the ability to form a partnership with the patient. Not only are we generating a lot more molecular data around the patient, we are also engaging and consenting the patient for longitudinal access to that patient—outside the health system—to acquire more data. We can help them manage their data and be able to re-contact them, to enroll in clinical trials, and so on. So, it is a more holistic play than Flatiron takes.

***How are you consenting the patients?***

**SCHADT:** We come in with a holistic service. The physicians read the script to the patients of what tests are available, they choose our test, and then we do everything from that point out. The physicians don't need to do anything. That includes scheduling appointments for the samples, pre-test genetic counseling, getting the results, post-test counseling—all of that is us engaging the patient. And we do that digitally.

When a patient signs on to our platform to set an appointment to do the genetic counseling, or whatever it is, then they are also provided the opportunity to consent in an IRB-approved way. This is a high bar consent in terms of how it informs the patient of what we are requesting. But it is really a consent to say: "We can serve on your behalf as data partner. We can get data from any and all health systems you've been a part of, manage it for you, provide it for you in ways that are portable, and we can set it for recontacting. Over 80% of patients who flow through our platform consent in that way and it's because we are a trusted clinical partner.

***This seems like an opportunity to talk about patient data and how it is handled. I've heard you speak passionately about protecting patient data and empow-***

***ering patients via their own health data. Why is that important to both you and the company?***

**SCHADT:** On the safety and security side, our most sensitive information is encapsulated in our medical record. From the diseases you are exposed to, to the diseases you have, to HIV status, there are a whole range of things that could be very devastating and could lead to discrimination. Then there is also HIPAA and others who require protection of that data. We view protection of the data as one of the top priorities of the company. You can't handle information like that digitally unless you employ the most state-of-the-art measures to protect it.

On the enabling the patient side—or why do this for the patient—my view is data that is being generated on you, that is your data. It is being paid for by you, whether it is your insurance or however, but that should be your data. It is enabling patients to have some control over that information, how that information is used, and the ability to share that information. What do we see in these cancer patient journeys or disease diagnostic odyssey cases? These patients don't see one doctor, they see ten doctors across ten different systems and why should they not be able to have a portable version of their data that they can share with anyone who they want to help in their fight?

It is also about how do you return some agency to the patient? Today, in the U.S., medicine is built around you give up your agency as a patient and trade it for the physician fixing whatever is wrong. That is the contract.

My view is doctors should be more like a co-pilot. They should help a patient navigate whatever it is to help them get better or to keep them on a well trajectory. It is all about how we empower patients to achieve that. Empowering patients to take more control over their life course, and health course, leads to better outcomes.

***You were leveraging genomic data and EMR data at Mt. Sinai. What other kinds of data are you working with at Sema4?***

**SCHADT:** I'm at heart a data scientist. To me, more is always better. If I could see what is going on in every single cell, in every part of the body across the entire human population, across all the dimensions of molecular profiling, I would do it. That is where I want to be.

We try to incorporate all the different data dimensions. The only limitation is, what can you afford and what are the medical establishment and payers willing to do. Is it informative enough, and cheap enough, to do at scale?

The way we think about the modeling is, any data we can get our hands on around the patient, we will collect and just view it as additional features that we may find ways to be predictive about their health.

***So, when a patient consents you will collect other data from them including other tests they may have had in the past, even though you may not yet have a broader use for that information?***

**SCHADT:** Right. Think of it as a big table where each row is a feature and the columns are people. It is about how many features you can collect. Because what I've learned over and over again in my career is that the things that are most predictive, and even causal, are the last things you would expect. The idea that we know that map we are trying to look at is just false. It is still very much a knowledge, discovery, understanding process. It is taking a very data-driven objective approach and we don't know a priori what all the right features are to look at. So, the more we collect, the faster we are going to learn.