**SEASON TWO**Dr. John Quackenbush: Precision Medicine Beyond Simple Mutations

January 9, 2020

Karan Cushman: Welcome to season two of the Precision Medicine Podcast sponsored by Trapelo. This is the podcast where experts come to discuss the problems oncologists, reference labs, and payers face as precision medicine grows and consider solutions for advancing the quality of patient centered cancer care. Be sure to subscribe at precisionmedicinepodcast.com to get the latest episodes delivered straight to your inbox.

Jerome Madison: Welcome to another episode of the Precision Medicine Podcast. I'm Jerome Madison, Vice President at Trapelo and today we have John Quackenbush, Ph.D., chair of the department of biostatistics at Harvard university's T.H. Chan School of Public Health. John, it's a privilege to have you on the podcast.

J. Quackenbush: Good afternoon, Jerome. It's a privilege to be here with you, so thank you very much for inviting me.

Jerome Madison: For sure. I mentioned this because I heard you speak not too long ago, and I've heard you speak over the years, and you had kind of have that distinction in my mind, and I'm sure others, of being that person who can speak about the human genome in a way that kind of helps us visualize what in the heck we're looking at and talking about and represent the genome in a way that we can begin to learn and understand the scope of what actually we're dealing with, because you know, what does the genome look like? And I think I figured out why.

Jerome Madison: So, you were one of the first folks that were hired to work on the Human Genome Project, which predates the Cancer Genome Atlas Project. So, there are people who say that where we are now are the early days of kind of genomic and genetic evaluation and using that information. So, when you started that had to literally be like the Flintstones. So-

J. Quackenbush: Yeah, we had little stone cars with big stone wheels and dinosaurs coming to help us cut the genome up into smaller pieces [inaudible 00:02:08] sequence it. It was exciting. So, what was it like? Well, it was a very interesting time, and thank you for the kind words. I always sort of attribute it to the fact that I had to learn biology and genomics kind of from the ground up.

J. Quackenbush: My Ph.D. was actually in theoretical physics, and as you mentioned, I got to a position working on the Human Genome Project very early on back in 1992, and it was a really exciting time. In the 1940s, people had started to decode the structure of DNA, and I think Watson and Crick's famous paper was published in 1953.

J. Quackenbush: We knew that DNA was the genetic material, we knew it was made up of these individual subunits we call basis A, C, G, and T, adenine cytosine, guanine, and thymine. We knew it formed this double-helix structure where A on one strand pairs with T on the other, C on one strand with G on the other. But we didn't have a good way of understanding what the genes were, where they were encoded within this genome. We didn't really have any technology to read off that string of A, Cs, Gs, and Ts that we now think of as a genetic code.

J. Quackenbush: It was really in the 1970s, I think 1977, when Fred Sanger invented what we now refer to as Sanger sequencing. And it was the first technology we could think about using to read off the DNA sequence. But to scale that from a few bases at a time to reading off the three billion bases of the human genome was a real challenge. And just to give the listeners a perspective, three billion is a number that it's sort of hard to conceive of-

Jerome Madison: Right.

J. Quackenbush: Right? It's a little bit bigger than our salaries I would assume. It's a tiny fraction of the national debt, but it's…three billion is the number of seconds in 95 years. Right? So, if we could read a base a second and lived a very long life, we would've gotten to the end of one copy of our genome. And we actually carry two: one from our mothers, one from our fathers.

J. Quackenbush: So, it was a time when…in the early 1990s, scientists sort of realized that we could take this technology that Sanger developed and scale it up and scale it in such a way that we could automate a lot of the processes and then read off the sequence. Then once we had that sequence, we had to figure out what to do with it, which is kind of where I came into this. Somebody with a quantitative background in physics, one of the questions was, well how do we take all of this data and make sense of it? So, that was really my entree into the Human Genome Project back in the early 1990s.

Jerome Madison: And to fast forward, I guess with everything that you've done in helping create, being the chair of the biostatistics program, you also teach computational biology and bioinformatics. So, I guess for our listeners also, can you explain to them what those terms mean and how these disciplines are being used to enable precision medicine?

J. Quackenbush: Sure. So, as I was getting ready for this podcast, I was actually making up a homework assignment for my students. So, if the listeners want to drop me a note, I'll send them the assignment, we'll see what we can do. But I've actually been here at Harvard since March of 2005, and I've been involved in the department of biostatistics, which at the time I came didn't quite match up with what people were doing in computational biology and bioinformatics.

J. Quackenbush: But our department now really has kind of a broad scope and that, we realized that biostatistics is part of that cornerstone of what we think of as quantitative health science. There are other important pieces now, things like machine learning and data science and computational biology and bioinformatics.

J. Quackenbush: So, I got my start in the early 1990s, as we talked about earlier, working in the Human Genome Project, and one of those questions that we were all asking was, what do we do with these three billion bases once we have them? And it really was a whole series of questions, all of which inevitably boiled down to, well, we have to develop computer programs and software that's going to allow us to make sense of this massive data. And that quest to build these programs and come up with new methods and handle these massive datasets is what we started to refer to as either bioinformatics or computational biology.

J. Quackenbush: In the early days, some of the questions where, well, can we just take these pieces of DNA we can read off? And at the time there were a few hundred base pairs in length. So how do we take all of these little pieces and put them all together to reassemble a representation of our genomes? And then there were other questions we started asking like, well, what are the genes and where are they within that genome?

J. Quackenbush: So, a lot of the early questions were really around the mechanics of taking what was coming out of a DNA sequence or a string of A, Cs, Gs and Ts, putting them together, and asking what was there. But then once we started to find genes, we started asking questions like what do they do? How are they put together? Can we find similar genes in a mouse or a plant like Arabidopsis thaliana or a yeast like Saccharomyces cerevisiae where we could do experiments to understand what these genes were actually doing?

J. Quackenbush: So, as the genome project sort of grew and evolved, this whole discipline of computational biology started to evolve really around how do we work at the interface between computation and biology to ask and answer meaningful questions?

Jerome Madison: It's amazing. I think we're still asking that question, what do we do with this information? But from the perspective of your team, the people that you've worked with internationally on this project, we have programs, we have software, we have platforms to be able to measure this in. In the world of knowledge, how much do you think we actually have a good understanding of, in order to do something with it? If you could put it in a percentage?

J. Quackenbush: Yeah, I don't know. It's hard to say where we are. I can tell you were a lot farther along than we were in 1992 when I started working on the genome project. We're much farther along than we were when we had the first reference genome completed in 2000 or published in 2001.

J. Quackenbush: What's really exciting for somebody like me today is that technologies to generate DNA sequence data have just grown tremendously. The cost has fallen. The time it takes to sequence a genome has fallen dramatically. So, now we have access to literally thousands of genomes that we can start to interpret and there's a cost curve that I often show when I talk about sequencing the genome. When the first gene was sequenced in 2000, 2001, the estimate was that sequencing the next genome was going to cost about a hundred million dollars. That's a price most of us can't even imagine.

J. Quackenbush: By 2009, because of these new technologies, it had dropped to about a hundred thousand dollars, and even then we knew enough about some diseases like cancer and the genetic variants that are associated with the disease that I would say very honestly, if my wife or son had a rare tumor, I would mortgage our house and sequence their genomes. Today, the cost of sequencing a genome is on the order of $1,000, and people can quote you different numbers depending on what you're going to do and how much sequence you're going to generate. But $1,000 is a sort of good working number.

J. Quackenbush: When I think about that, we're moving from this domain where you say, "Well, I have to mortgage my house to sequence my wife or son's cancer genome," now to being able to say, "Well I could pay for that with a credit card." And that really sort of transforms the way we think about generating data, but also using it. And for somebody like me who likes to deal with ever-larger quantities of data, it gives us the opportunity to ask and hopefully answer much more meaningful, complex questions.

Jerome Madison: Yeah. Precision medicine as it relates to genomic testing for oncology has been in the mindset of one gene: one drug, can have a one-to-one relationship, but your research recognizes that it's not individual genes that influence disease development or response to drugs, but it's more of a complex network of interacting genes that determine how, I guess, cells behave, and you call those gene networks. So how can we begin to understand what a gene network is? Because a lot of professionals are still trying to wrap their mind around just one target. And, and how does this approach help patients clinically?

J. Quackenbush: So, there are a couple of really good questions there. So the first question is: why do we worry about networks? And the analogy I always like to give is that if we think about the cells in our body, I just like to think of them as little machines made of proteins. And one of the acting definitions we have for a gene—and there are a variety of different definitions people use—but a pretty common definition, and sort of at the core of most of what we talk about is that a gene is a stretch of DNA that includes the blueprints for making a protein. Okay?

J. Quackenbush: So, the genome gave us a list of genes, that gave us a list of proteins, that make up the cell. All right? So we can think about this in a variety of different ways. The first way, I would say, encourages us to think beyond individual genes, is the idea that if I were to look at a machine like your car, all right, if you were to tell me about a single piece, right? That single piece, if it breaks, might prevent your car from working. So, if the fuel pump dies, that car might stop. And if you could figure out that it was the fuel pump that broke, then you would know that if you replaced the fuel pump, you can fix the car, right? That single element, that single protein in a cell that corresponds to the fuel pump could be something you focus in and really try to understand.

J. Quackenbush: On the other hand, what might happen in your car is the fuel pump might still be working but not completely, and the fuel filter might be a little clogged, and some of the fuel lines might be a little cracked and leaky. So if you start to think about reasons your car might not run, it could be one element or it could be a system, a group of elements that interact together to keep your car functioning and if they break down, even a little bit, might prevent your car from starting. It's a much harder problem to solve if it's not one thing you can replace or one thing you can target…if you've got a larger set of elements interacting together.

J. Quackenbush: That's kind of the way we have started to think about human cells. That any individual protein or any individual gene that corresponds to that protein…any individual protein by itself doesn't really give us the whole picture, that it's part of a larger set of interactions that govern the way in which the cells work. Okay? So, when we start to think about biological complexity, we recognize that individual genes by themselves only give us part of that more complete picture.

J. Quackenbush: The second thing we start to recognize is that while we talk about a human cell in the human genome, your brain cells and your liver cells and your kidney cells all have the same genome, the same DNA. They encode the same set of genes. So, what we now understand is that in a brain cell and a liver cell, a lot of the same genes are turned on because the cell has to run, it has to consume oxygen, it has to produce carbon dioxide. It consumes sugar to drive its energy metabolism, and we have core functions that a cell just has to do.

J. Quackenbush: But then, there are other properties or other functions that are carried out by the cell that are unique, right? Your brain cells are making neurotransmitters, your liver cells are making digestive enzymes. So, as we start to think about this bigger picture, what we realize is that that genome, which encodes all these proteins that make up all the parts is actually activating different elements in different cell types. Again, a lot of the way we think about that is that those genes which are being activated are regulator controlled in different ways.

J. Quackenbush: So, we can talk about networks of proteins to make yourselves function. We can also talk about regulatory networks of genes interacting together to activate different processes in different cells. And just like there are different processes or networks of interacting elements that turn on certain genes in your brain and not in liver or vice versa. There are a set of interactions between these genes that turn on certain processes in a healthy cell and turn them off in a tumor cell. Or turn them on in a tumor cell, which is much more likely, and keep them off in a healthy cell.

J. Quackenbush: So, the reason we start to thinking about networks is we realized that, just like all the parts of your car don't tell you how the car runs, a parts list for a cell doesn't tell you how the cell runs, that we have to think about how these parts go together and we have to start thinking about them in context to understand what makes one type of cell different from another.

Jerome Madison: So, for our listeners, I'll remind you that I did say that you are one of the most gifted speakers that that helps us understand how what we're dealing with because that car analogy was teaching gold man, everyone's had a problem with a car, and it's not just a fuel pump. They get you on a diagnostic tool and work the networking, you go in for a fuel pump and walk out with a brand-new engine or something, you know? That's just gold.

Jerome Madison: So, it would seem to me that gene networks would be even more personalized and precise because what's a normal gene function for one person…in the expression for one patient… is completely different from another. Then there is kind of the environmental factors that might influence that. So, what kind of opportunities come out of knowing this about the way our genome works?

J. Quackenbush: Well, it's given us a lot of insight into different aspects of what makes cancer cancer, and it has given us insight into ways that we can think about treating cancer. The first step to all of this is really understanding. So, my colleagues and I have developed a whole series of methods around the idea that we can look at differences in gene regulation.

J. Quackenbush: So, what's this process? What are the things in the cell that are turning on certain functions in one cell type and not in another? So, in a healthy cell or tumor cell. And as we started to look at this, what we've come to recognize is that when we look at different groups, we can start to see differences in that regulatory process. So, one of the nicest examples of this is a study that Camilla Lopez-Ramos and Dawn DeMeo and Kimberly Glass, and, I think, Marieke Kuijjer and I published in October of 2018, and it was a study of sexual dimorphism and colon cancer.

J. Quackenbush: So, sexual dimorphism is this observation that, in many diseases, we actually see differences in the clinical manifestation of that disease between men and women. So, if we look at colon cancer, there are differences in disease risk, there are differences in disease severity once disease starts to develop, and there are differences in response to chemotherapy that we see between men and women. One of the questions that's been a long-standing open question is, “What drives these differences between males and females?”.

J. Quackenbush: We started to ask that question by looking at the individual genes and their mutations and didn't see anything that was informative. We then looked at what genes were simply turned on and off, in men and women. When we asked that question about what's being turned on and off, the data we get is often just a snapshot. Right? And I often liken this to trying to learn the rules of football by looking at the pictures of a football game that might appear in a newspaper, right? You have a dozen shots, right? How do you learn the rules, right? What makes males different from females?

J. Quackenbush: And when we started to look at this, we didn't see any clear pattern emerging. But we stood back and used some of the methods we have for estimating networks, for learning networks in male tumors and female tumors. And when we asked how do the networks between men and women differ, what we saw was that there were differences in the pathways that turn on or turn off drug metabolism in men and women.

J. Quackenbush: So, that gives us a really interesting insight, because the regulation of these pathways is controlled differently in males and females. Our snapshots were really tumors before treatment had started. But our models told us that we should expect differences in how male and female disease respond.

J. Quackenbush: So, from that first step, we haven't yet moved to treating men and women differently. But we've developed some hypotheses for how we might want to do that and those are things that we can test in the laboratory. We can test in cell lines and animal models, and we can really start to understand whether or not we should use some of these. And, as a first step to better precision medicine, just ask the question, should we be treating males and females with colon cancer differently?

J. Quackenbush: And this, a phenomenon of sexual dimorphism occurs not just in colon cancer. We see it in other cancers, head and neck cancer, lung cancer. We see it in diseases like chronic obstructive pulmonary disease and emphysema. We see it in Alzheimer's. So, as we start to tease apart these regulatory networks, we can start to ask questions that have alluded us before, like do men and women differ? And there are a whole host of other things we can do to using these kinds of methods.

Karan Cushman: You've been listening to part one of our conversation with professor John Quackenbush, chair of the department of biostatistics at the Harvard T.H. Chan School of Public Health. Be sure to look out for part two where we'll continue our discussion on precision medicine beyond simple mutations. We hope you'll tune in.

Karan Cushman: You've been listening to the precision medicine podcast sponsored by Trapelo. Trapelo is the first clinical decision-support tool to align the interests of oncologist, labs, and payers to give patients the best chance at beating cancer. To learn more, visit gettrapelo.com. To subscribe to the podcast or download transcripts of any episode, visit precisionmedicinepodcast.com.

Karan Cushman: We invite you to join the conversation on social media. You can find us on Twitter at PMPbyTrapelo or on LinkedIn at the Intervention Insights company page. If someone who would enjoy the Precision Medicine Podcast, please share it. They'll thank you, and so will we. We hope you'll tune in for the next episode.

A person smiling for the camera

Description automatically generated

**About Our Guest: John Quackenbush, Ph.D.**

John Quackenbush is Professor and Chair in the Department of Biostatistics at the Harvard TH Chan School of Public Health. John’s Ph.D. was in Theoretical Physics, and he received a Human Genome Project fellowship in 1992. This led him through the Salk Institute, Stanford University, The Institute for Genomic Research (TIGR), and then to Harvard in 2005. John uses massive data to probe how many small effects combine to influence health and risk of disease. His work has been cited more than 70,000 times. Among his honors is recognition as a White House Open Science Champion of Change, which he received in 2013.