MIT CSAIL Alliances | Podcast_ManolisKellis_04/29/2020.mp3

[MUSIC PLAYING]

Welcome to our podcast series from the Computer Science and Artificial Intelligence Lab at MIT. I'm Lori Glover. And today, I'm here with Professor Manolis Kellis, who heads the computational biology group at MIT CSAIL Manolis, thank you so much for joining me this morning.

Lori, it's a pleasure to be here. Thank you so much for doing this

Absolutely. So can you talk a little bit about your area of research and some of the most compelling challenges that you're working on?

It's not a small goal. What we're trying to do is basically understand the basis of human disease, understand the function of every nucleotide in the genome in every one of your cell types, how that function is encoded, and what causes that function to change in human disease. And that basically means we're trying to understand the language of DNA, which is kind of cool for someone who grew up in Greece, in France who speaks three languages with my kids at home. But I feel like I speak even more languages with my collaborators because we have to speak biology, we have to speak computer science, we have to wear many hats, think like an engineer, think like a scientist, like a doctor, think like a patient, think like a biology, like a computer scientist.

So our work is extremely interdisciplinary. And again, that's part of the hats. All of the other hats is blending that with a teaching environment where we're working closely with students, we're mentoring them, we're seeing them grow scientifically, we're seeing them mature as scientists, we're seeing them flourish as human beings, we're seeing them grow and learn how to share, how to collaborate, how to understand each other's need.

So not unlike how I'm watching my kids grow up and learn how to share and to collaborate at home. And that again, is just one aspect. And again, working with you, we have to blend that with entrepreneurship, with startups, with competent collaborators, with industry, with hospitals. And taking our research to the world.

And that's why I'm so excited to be here and to be part of this podcast And at the same time we're working with doctors, with hospitals, with clinics, with health centers trying to really have a practical impact to understand and address the needs that they have. And of course, blend all that with understanding patients and empowering them with knowledge about their own genome, prognosis,

diagnosis, biomarkers, wearables. So it's a very complete picture that spans the scientific, the educational, and of course the entrepreneurial, and the interpersonal.

Wow, that's rather amazing. I like that analogy to it as a language as well. And you are a computer scientist. And how is the biology really blended in with computer science?

I'd like to answer the question exactly the opposite way. How could we possibly do biology without computer science?

OK.

Namely, DNA is a programming language. It's not just about, "Oh, there's lots of data." At the foundation of computation biology lies the fact that there's a programming code in there that all of our cells are running in order to understand how to interact with the environment, in order to translate our genetic makeup into biological functions.

So what we're trying to do with computational biology is extremely natural. It's basically understand the language of DNA, the programming language of DNA, and understand how mutations in this language. If I take out a go-to statement or a function call or a variable, how is that going to affect the overall system?

We're trying to understand the circuitry of every cell type. How every single control element is connected to every single gene? How the genes are turned on and off? How that varies from cell type to cell type?

There's no way biology would function without computer science. It's not just that it's an aberration. It's the most natural way to think of doing things.

And part of that is understanding how to number 1, of course, integrate vast, vast data sets that are just impossible with humans alone. And number 2, even understanding and thinking like a computer scientist. What type of data should I even gather? What is the next data set that I should go and get?

So despite being a computer science department and a computer science lab, a big chunk of my team is actually experimental. We go out and gather massive amounts of data. We are now gathering perhaps the largest collection of single-cell data on the human brain in the world across dozens of disorders. We're basically looking at cognitive, psychiatric, neurodegenerative disorders. And we have more than 1,000 different brain samples that we have now profiled at a single-cell level.

So yes, it is a computer science lab, yes computer science is central to all this. But also, we have a vast, vast effort on the experimental side to both gather data sets for integration. But also, generate validation for our predictions. In my view, it's impossible to do it any other way than to look at disease in a holistic kind of fashion. And in the same way that DNA unified biology, I think, the DNA will unify disease--

Interesting.

--because being able to understand the commonalities in that molecular function. We're kind of solving the same problem all over again in every single disease department right now. And I think this makes no sense. It makes more sense to say we're going to have a department that seeks to understand the human epigenome, a department that seeks to understand the regulatory circuitry. And now, that department will be cross cutting and applied to many different disorders.

And if you look at how hospitals are organized and how pharmaceutical industry is organized. There's a huge separation between diseases. And these departments are not talking to each other when, in fact, we have so much to learn by looking at many disorders. And I think that's one of the strengths of our group.

The fact that we have so many collaborators across metabolic, immune, psychiatric, and regenerative cancer. There's just so much to learn that we are learning in one area. And that we're then applying in a completely different area. I think, we're still very naive about what really truly underlies all these different conditions.

Just a few years ago, we basically were looking into genetic variants underlying Alzheimer's disease. Everyone was expecting to see them light up in brain. No, they light up in the immune system instead. So there's a set of immune cells that make up a small fraction of our brain. And that's where the vast majority of the genetic predisposition for Alzheimer's disease appears to be, I think.

Oh, wow!

So that changes our view. And suddenly, you need immunologist to try to understand brain disorders. And the same thing with cancer. And the same thing with metabolic disease. Every single one of those we're learning just how much we were ignoring about the true molecular basis of the pathophysiology of the disease.

Well, that is a really interesting approach. And certainly, when you look at all the diseases and kind of looking at that commonality, that really could be game-changing. The goal of your research, you can

talk a little bit about your vision and where you like your research to go.

So the goal of our field-- and it's a very lofty goal-- is to really, number 1 understand the molecular basis of disease, understand the human genome and dramatically transform the therapeutic landscape. So basically, truly enable, for the first time, a rational approach to drug development. And ultimately, transform human health, improve the human condition, understand our body better. And perhaps even more importantly, understand our brain better, human psychology, motivation, depression, aging, neurodegeneration, psychiatric disorders.

We have not even begun to understand how our brain can affect our body. If you look at the placebo effect, this is the most dramatic evidence that if the brain thinks that it's getting better, it actually makes the body better. So you take a little sugar pill and suddenly you're beating your cancer, you're beating your immune disorder et cetera. And that's because of all these interconnections between our brain and its ability to actually control our body.

And I think that, yes, we've come a long way, and it's an extremely complex world. But despite our still rudimentary understanding of the human brain and its function, we've accomplished so much. If we now understand this brain better. If we can feed it what it really needs to succeed and what it needs to keep our body healthy and to even keep our mind at ease, I think, that we can have such a dramatically transformed, not just health care system but, society, by better understanding both the human body and the human brain and their interconnection.

Yeah. That mind-body connection. That's great. So in your lab what are some of your most recent discoveries?

So we work in many areas we work in immune, cancer, metabolic, brain. So we've had a series of papers that are basically using this holistic approach that I'm mentioning of large scale data integration and profiling samples across different data modalities, understanding the genetic differences, understanding the epigenomic differences, understanding how gene expression changes, understanding at the single-cell level, how complex tissues are, in fact, made up of a diversity of small parts and the variation in those. And that has led to insights in many different areas. The most recent one is a paper that we just published on the single-cell profiling of the human retina.

So for the first time, we actually were able to look at thousands of cells, individual cells from a human retina and understand what are the genes that are active in each of those cells. Then, we were then able to use this information to go back at the genetics of age-related macular degeneration, which is

actually one of the disorders that I have a genetic predisposition for. And then, look at where are those genetic variants acting, what genes are they affecting, what cell types within the human rhythm are they affecting.

Just a few months ago, we had a paper in *Nature* looking at the first single-cell dissection of Alzheimer's disease, looking at 80,000 different sales across 48 individuals to basically understand how are men different from women before Alzheimer's. Are men different from women after Alzheimer's? How is each one of our cell types responding to Alzheimer's? And which of those changes are likely to be causal and underlying Alzheimer's?

One of the dramatic surprises there was that oligodendrocytes, which are the cells that coat the neurons to make electric transductions more immune to perturbation that myelinate these neurons. These cells turn out to be dramatically different between men and women after Alzheimer's disease. So the men are able to upregulate these myelination processes and protect their neurons, whereas women are not able to do that.

Wow!

So our prediction from the single-cell data was that if we went back to the pathology information, and we look at the white matter loss between men and women that it would be much more dramatic in women. So using data that was decades old, we saw that there was a dramatic difference between the two that just had never been noted in the literature.

Interesting.

So that basically showcases these very naive understanding that we have right now of all of disorders and how these massive amount of data can allow us to really take a rational approach and understand those. In metabolism, just a few years ago, we basically showed how there's a switch within our fat cells. That switch is between storing fat and burning fat. And that was just simply not recognized before.

There's two master regulators of metabolism that were just previously completely unknown. And that, again, changes the way that we see the equation of nutrition. It's not just energy in, energy out. It's not just what you eat and whether you exercise or not, what you burn. It's really also what is the energy lost.

In any kind of thermodynamic system, there's three parts-- energy in, energy out, energy lost. I do lose energy as heat. It turns out that within our fat cells, there's a switch that controls thermogenesis,

the generation of heat. And that's actually a switch that our cells can turn on to burn heat. And that, in fact, underlies the strongest genetic association with obesity. Understanding that opens up new therapeutic, new approaches for now manipulating that pathway.

How interesting?

In cancer, in a paper that we're just submitting this week, we were able to show that the are circulating blood contains these small vesicles called exosomes, that within them carry RNA from diverse cells across the body. By looking within the blood, we are now able to predict not only how is the cancer doing in a tissue that is inaccessible, but also which patient will actually respond to immunotherapy and which ones will not. And that can have dramatic implications for the ability to combat disease. So again, across all these areas it's always the same principle-- large-scale data integration, collaboration with the best experimental scientists, validation, and then dissemination of the results for the whole community.

Can you walk me through an example of the impact that gene variation has on disease? And how your work can impact our understanding and treatment of that disease?

So as I mentioned earlier, for these metabolic switch underlying obesity, in 2007, the community was surprised and shocked to find that genetic variants were discovered that underlie obesity. Because up until then, basically, people were just, I don't know, blaming what you're eating and blaming whether you exercise or not. But what the community found was that there was one locus in FTO, a gene that was then renamed fat and obesity-associated, where a dramatic association with obesity was discovered. So suddenly it meant that there's a genetic basis. That it's not just your fault and your choices, but there's something in your body that changes.

And the whole community focused on this one gene-- FTO-- that was sitting square on top of the association. And for more than a decade, there was just this dramatic focus on understanding that function of that gene. And a huge amount was learned about that gene in all the processes it's involved in.

But what we showed in 2015 is that these gene, in fact, probably has nothing to do with obesity. That the true targets are sitting 1.2 million nucleotides away from that gene. That what these genetic variants that are associated with obesity are doing are not changing the function of that gene. That gene is unchanged, its expression is unchanged. And its protein sequence and protein function is unchanged with those genetic variants.

What's happening is that is that those genetic differences that are sitting inside the FTO gene are controlling IRX3 and IRX5, two genes that are sitting 1.2 million nucleotides away and 600,000 nucleotides away.

Oh, wow!

And it turns out that those genes are the real culprits, which are then controlling this process of thermogenesis. And those genes have just never been associated with either obesity or thermogenesis before. And now, they're at the core of a whole new program for targeting these genes to manipulate human body composition for where and how do you use your calories.

Do you burn them? Do you store them? And that decision is made by those genes.

So what we showed is that you can trace these large region of an association that contains 89 common genetic variants across 10,000 nucleotides into a single nucleotide. So we could find that a single letter change can actually switch between lean and obese phenotypes. And we use CRISPR-Cas9 genome editing to go in and change that one nucleotide from a C to a T and from a T to a C.

And what we found is that if you take primary fat cells from a risk individual who's unable to burn calories in their fat cells and you fix that one nucleotide, their cells are now able to completely burn those calories as if they were completely healthy. So these therapeutic implications are huge--

Huge, huge.

--because it basically means that we can have three interventions. Of course, everybody should eat less. Of course, everybody should exercise more. But at the same time, we can control that last switch, that last knob of how many calories do you actually burn inside your fat cells.

And that will be incredibly helpful.

I could certainly use some of that.

Not you. But I mean, for the world. I mean, think of the surgeries and things that people go through and the other health effects also that are associated with obesity to be able to have a healthy or healthier approach to that. That is fantastic thing.

Our society is plagued by obesity right now. I mean, the US, but also the whole world is-- you mean, obesity rates are just climbing everywhere. And it's not just affecting your body. It's affecting your mind. It's affecting your self-esteem. It's affecting so many different other diseases get so much

worse with more weight.

Oh, absolutely.

And if we're able to help these people, not only exercise and eat more healthy, but also therapeutically help their own metabolism, It would be just transformative with societal impact.

Yes, absolutely. That's fantastic. So talking about kind of the biggest challenges that you're facing. And one of the things that as you're talking is just occurring to me is all this data and the privacy laws and getting access to what you actually need. I'm sure, that's got to be a challenge too, right?

So for more than 20 years, I've struggled with this. We've basically reached out to clinics that want us to analyze their data. And we want to analyze their data, but there's months and months and sometimes years of paperwork to be done.

And one of the things that we're doing right now is actually trying to overcome this by a new startup that we started. Secure AA Labs sale. That basically is trying to make data sharing so much more seamless and so much easier. Every hospital is constrained to not bring the data of the patients outside, which basically makes it very hard to analyze data in silos because every one of those hospitals has only so few patients that we're just statistically underpowered to really discover these associations.

So what we're trying to do in Secure AA Labs is enable the movement of computation rather than the movement of data. So what we're basically doing is enabling with both a hardware and a software solution. The complete security and privacy of all these data sets.

And then, computation, which is written by third parties to then be trusted and checked at every point of the way about exactly what types of operations you can actually do on those data. And then, go and gather snippets of analysis without the patient data ever leaving the hospital, you can have bits of, integrative knowledge that spans across many patients from that hospital, which can then be leveraged and utilized and integrated across many, many different hospitals.

So that's one of the challenges, you mentioned in data sharing. But the discovery challenge that we have is not just about data. It's every part of the way. So basically profiling, analyzing, discovering, validating-- every single one of those aspects has dramatic challenges. And I think, that's what our everyday job is about, how do we basically make every one of those components easier.

That's great. So most people are familiar with computer science and biologists separate disciplines.

We talked about computational biology. And we hear a lot today about personal medicine and precision medicine. Can you talk about how computational biology applies to those areas?

Computational biology applies absolutely both. And thanks for mentioning both areas because people very often confuse the two. So in my view, precision medicine is really what I've been describing.

Basically, being able to take these 3.2 billion nucleotides in the human genome and say which one do I have to change to fix that process. And it doesn't get more precise than that. Basically, being able to know what cell type to act in, what nucleotide to change, and exactly what to measure.

Basically, what are the upstream control regulators, what are the control regions, what are the downstream target genes, and what are the processes that are affected by that. So in my view, that's precision medicine instead of just throwing a bazooka at the disease, you're going in with a gorilla team that says, "This is the nucleotide that we're going to change and this is the one gene that we're going to affect and this is the one cell type and the one tissue that we're going to manipulate." So that's precision medicine.

Precision medicine is being able to, instead of saying, "Take an aspirin," say, "Let's change this one thing." And a lot of medicine has been affecting these dramatic processes. And that's why there's all these side effects with so many different pills and especially as individuals age.

You basically have dozens of pills and their interactions are completely detrimental to the liver, to the whole body of those individuals. If instead, we were able to have a huge catalog of drugs where you can mix and match exactly which ones you need to target those three genes in that pathway with only a minute amount, I think, the side effects can be decreased dramatically. So that's what precision medicine is about.

Now, personalized medicine means a lot more. It means not just being able to identify exactly what we should change for every disorder, but understand how every person should be treated differently. And again, the economics of personalized medicine are right now extremely daunting because everybody says, "So which company is going to invest a billion dollars developing a drug for me."

But I think the dynamics should be thought of very differently. Namely, yes, there are 6, 7 billion people in the planet, but there's only 20,000 genes. And there's only so many cell types in the human body. And there's only so many ways that we can modulate those.

So if instead, we start thinking of a library of compounds which can individually be tested for safety.

And then, if the FDA is thinking of the clinical trial at the level of the algorithm of how we're going to be combining these compounds and how we're going to be rationalizing the identification of exactly the treatment that every person needs, then you can basically mix and match from these catalogs to create a treatment at almost no extra cost for each person. And that's what personalized medicine will be about.

And of course, genomic medicine means that some of that personalization will come from your own genome. So in my view, it's not just a genome. It's of course, the entire environmental information and the metainformation and the exposome if you wish, of that person which has to be combined with the genome to truly understand the personalization of the treatment.

Namely, every patient would not only have access to their whole genome, but possibly have a lot of information from wearables, a lot of information about their sleeping patterns or walking patterns, exercise patterns, movement, eye movement, body movement, and so on and so forth, as well as possibly a blood biopsy that can tell you about metabolite, that can tell you about blood biomarkers of a lot of inaccessible tissues and organs that you will now have more information about.

And then, to make a decision about number 1, diagnosis, but also number 2, prognosis-- what does the genome tell you about what you might have tomorrow and what is your current blood draw tell you about a disease that might be developing that might not manifest yet. So instead of treating the symptoms, we can start treating the causes. Instead of responding, we can start anticipating.

And again, everybody is wondering about the cost in the health care system of the society, but we have to embrace the fact that we're all in this together. And if we can decrease health care costs for 1 out of 1,000 people by, I don't know, a million dollars, this is worth $1,000. So in other words, even if these treatments are costing us $1,000 for every person. And even if only 1 out of 1,000 people gets that benefit, the math still works out.

But my goal is that these interventions that these measurements will cost so much less than $1,000 that is going to be $10 or $100. And then, the moment you spread that out across the whole health care system, everyone gets better. And a few people benefit dramatically more. And therefore, society as a whole benefits dramatically more.

And I'm just thinking as you're talking about the so-called designer drugs in cancer that really attack a particular cancer and like a chemotherapy drug that doesn't go after your hair and things like that where they've had great advances in it. But it's still on a personal level can be incredibly effective with one patient and not effective at all on the other. So really it's kind of taking that precision and

that personal and putting it together for the best health benefit of all.

I completely agree. And in my view it's really a combination of understanding the genetic makeup of each person, understanding the genetic makeup of that tumor, and understanding the current status of your immune system. The most effective therapies are really combination therapies because you're not just going after one thing, which then allows the cancer to find ways around it and evolve out of that one treatment, but it's using the combination. And the most effective way in the last few years has truly been to use the immune system as your friend and as your partner to basically help the immune system of those individuals fight the tumor.

One of the hallmarks of cancer that has really been recognized recently is the fact that the tumor is actually manipulating your immune system to escape evasion. It's basically telling your immune system, "Oh, everything is fine. Don't worry. Nothing's going on here." It's hiding itself from the immune system and so on and so forth.

The most effective treatment, which we are currently will be putting a big focus in our lab is really the ability to train the immune cells of the person to make the cancer visible to those immune cells, to stimulate the response in a very precise way. And there are new companies out there that are basically trying to build vaccines that are personalized for each individual's tumor. They basically say, "What are the recognizable features of that tumor. And then, can we use these features to now help the immune system."

So I it's going to be a multipronged approach. So on one hand, traditional chemo. On the other hand, traditional drugs. And on the third hand, how do we help the immune system fight the tumor.

Great. So you are very well known for completing the roadmap epigenetics project to create a map of the human epigenome. Can you talk about that project? And why it's so important?

In the same way that the human genome is a foundation for all of genetic studies, in my view, the human epigenome is the foundation for all functional studies of how are the cells functioning, how are they using this genome in every single cell type differently. So if you look at your body, there are hundreds of different cell types. There's so many different tissues and every one of these tissues has a complete and complex mix of dozens of different cell types.

How is that possible? Every one of your cells has exactly the same genome. It's an automaton that is running exactly the same code. How can the exact same code lead to a neuron and to a hair cell into a retinal cell and to your heart muscle and your blood circulating in those veins. The reason this is

possible is because a different subset of the human genome book is being used by every one of our professional cell types.

To build an analogy, it's as if every profession on the planet received the same encyclopedia of 23 volumes, which contains all of human knowledge regardless of your profession. And some people will need the plumbing books or the plumbing chapters. And some people will need the nuclear physics chapters. And some people will need the ballet chapters, and so on and so forth.

And what the epigenome does is mark up those chapters for every one of those professions so that the liver cell has bookmarks for every chapter that's related to liver function. And the neuronal cell has bookmarks for all of the chapter related to neuronal functions. So that's what the human epigenome is about.

So what we did in the Roadmap Epigenomics project is trying to systematically understand the human epigenome across hundreds of different cell types. And we've just posted a paper online a few months ago, giving the next generation of that map. So in the Roadmap Epigenomics project, we had mapped 117 different cell types, 111 different cell types, 127 total by combining with the ENCODE project at the time. And we now have 830 different human cell types that we have now mapped and released to the community.

So we're getting a lot of feedback from folks who are already using this map to now start understanding hundreds of different disorders because it's only by understanding what is the circuitry of every one of the cells from these epigenome mapping that we can then understand how are the genetic variants that are associated with all these diseases manifesting in the human body. What are the cell types that they're affecting? I gave you the ultimate the example of Alzheimer's disease and all those genetic variants localizing in immune cells. How did we find that out? Through the Roadmap Epigenomics project.

Because we had mapped all these 111 cell types, we can now start asking which one of those contains the genetic variants that underlie Alzheimer's disease. We could do the same thing with Alzheimer's with psychiatric disorders. And then, we're finding brain and neurons where we can do the same thing with heart disease and we're finding heart muscle. And we can do the same thing with lung function and liver function and immune disorders, and so on and so forth.

And we're finding in every single case what are the cell types that appear to be underlying these disorders. So these maps are basically both allowing us to map human disease, genetic variants. But also allowing us to paint the circuitry of how are those genetic variants connected to their target

genes.

Good. And thank you for bringing up ENCODE because one of my questions is going to be the relationship between the genome and epigenome. But I think you really addressed that. Tell me a bit more about the ENCODE project.

So the ENCODE project basically stands for encyclopedia of DNA elements. So think of it and I like to go back to this joke that Eric Lander made when he was speaking at the Nobel Prize ceremony where he had to describe the human genome in seven words. And he basically used those words. He said, "Genome-- bought the book, hard to read."

So what the ENCODE project is about is this hard to read part. So how do we now systematically understand every nucleotide in the human genome. And what the ENCODE project is trying to do is carry out in a very organized fashion, in a very systematic fashion, a very coordinated and common fashion a series of experiments that we now know are extremely informative for mapping the function of the human genome in all of the different cell types.

And that's what the ENCODE project is about. It's basically trying to build an encyclopedia of what are all the functional elements in these 3 billion bases. Only a small fraction of that genome is in fact useful. Only about 20% or so appears to be containing these very precise gene regulatory elements.

How do you now map all of that? Well, you have to map every nucleotide. You have to do these hundreds of experiments. And instead of every lab doing a slightly different version of that experiment, which would make integration very difficult, the NIH basically said, "Let's do all these experiments systematically with a small number of assays and a small number of teams that will coordinate with each other to make sure that the data is as clean and as reproducible as possible."

Yes, it's a huge endeavor. , Yes it's very difficult. But if we do it in a haphazard kind of way, the end result to the community will not be as clean.

It's not like genome sequencing, where in the end, there's a letter there or not. It's very difficult to make sure that all these experiments are in fact compatible with each other and comparable with each other and easily integrated. And that's what the ENCODE project is about.

That's fascinating. So this relationship between the human genome and the epigenome, how is this important in your research?

I like to think of the epigenome as the great integrator. There's a very, very large gap between

genetic variation and disease. For every complex trait, there are thousands of genetic differences scattered across these 3 billion bases that are contributing minutely to this disorder.

If you look at Alzheimer's disease yes there are less than 100 strong effects, but there are thousands of statistically reproducible weak effects. And if you now ask how are those genetic variants mapping to disease, this is a very, very large gap. What the human epigenome allows us to do and what all of these molecular profiles allow us to do is basically bridge that gap.

And ask, how our genetics first affecting the epigenome and how is the epigenome affecting gene expression patterns and how are the gene expression patterns affecting endophenotypes, the intermediate phenotypes associated with disease, for example, your lipid levels and how are those then affecting disease. So the epigenome gathers information from the genome. It is influenced by the genome. But it is also influenced by the environment.

So it integrates information from both genetics and environment. It is also influenced by disease. It can reflect the disease status. It can also reflect the expression pattern of different genes and the exposure to different environmental stimuli and the immune system functionality.

So this great integrator is basically integrating all this information across all of the different processes telling us both about genetics and environment and disease status. So by systematically mapping the epigenome, we can now understand how to bridge this gap and how to integrate both environmental and genetic hints to enable us to predict disease. And to understand disease status and to understand disease progression and to even understand treatment progression as the patients are moving through a particular course of treatment.

Interesting. So you actually can predict disease?

To some degree, yes. And that's where I want to go back to common variants and also rare variants. Common variants are common in the population. They are maybe at 5% frequency, sometimes 20%, 40% frequency. The genetic variant that I mentioned earlier-- rs1421085-- which is associated with FTO and obesity, that genetic variant is at 40% frequency in Europe, 44% frequency in Southeast Asia.

It sounds like a disease variant, but no. It actually used to be a beneficial variant when food was scarce, it actually made sense to store all the calories rather than to burn them away. That genetic variant only became detrimental after World War II. If you look at longitudinal profiles of individuals who are carrying these genetic variant before and after the war, you will see that before the war,

there was like no association with obesity. It's only after the war when everybody started having office jobs and walking less and eating at McDonald's that suddenly FTO became a major plague for our society.

How interesting?

And it's always a combination of genetics and environment so for these complex traits there are thousands of genetic variants in each of them is contributing only a small amount so by knowing your top five genetic variants associated with obesity. I really can't say much about your risk, but if instead I look at thousands at a time. And that's what polygenic risk scores are doing or PRS. I can then start predicting, to some degree, your risk for the disease.

And when I state to some degree, well it depends on the disease. For most diseases, it can be less than 5%. For some disease, it can be a lot more. So that's polygenic risk score for common variants. Then, there's a whole other aspect, which plays a much bigger role for personalized medicine of the rare variants.

So rare variants tend to have much stronger effect or to actually put it a little differently in causal terms, variants that have strong effects tend to be kept at low frequency. Why is that? Because if something contributes dramatically to schizophrenia, the impact this will have on the fitness of a person, on the number of offspring that that person can successfully put into the world and the number of offspring that these kids will have as they grow up is dramatically altered.

If you live in a family with dramatic mental disorders, chances are there's going to be fewer of you down the generations than in other families that don't have these features. So over generations, genetic variants that have strong effect tend to be kept at low frequency by evolution. Genetic variants that have only modest effect, can rise to high frequency just by drift or sometimes by selection as I mentioned for the case of FTO.

So being able to predict disease means being able to combine both common and rare variants to then say something about what's going to happen in the lifetime of an individual when you first see their genome at conception or at birth or at any point in their life. If you now combine that with biomarkers, that basically tell you about how the blood is circulating or what your heart rate is and all kinds of other things. If you combine that with family history, if you combine that with, I don't know, reading through the email of the person or understanding their Google searches or looking at their Facebook profile, and so forth, then, oh yeah, you can predict disease.

You can predict disease very well. And it's by integrating all of that information that we will then be able to not just carry out better diagnosis, but also carry out better prognosis.

Interesting. So like these kits like 23andMe and those-- I mean, they are saying they can provide insights into disease based on your genetics for DNA that you submit. What is-- I mean, are they accurate? Or what are they looking at in?

So I mentioned this complete picture of your medical history, your family history, your biomarkers, your common variants, and your rare variants. 23andMe is only looking at the common variant. And that's a small part of this body of information that you would need in order to be able to predict disease. So for most disorders, They're all going to be very predictive, and that's OK.

What's great about 23andMe is that it really raises people's awareness of the genome, of genetic information and the importance of the genome for understanding your ancestry understanding your overall health and so on and so forth. And yes, it's kind of fun to be able to do that and it's kind of fun to put it in the hands of consumers. But at the same time, we should be very cautious because some people might read this information and they misinterpret it.

They might say, "Oh, you have a 1.5 fold increased risk for this." But in fact, this only has a 0.3% risk to start with or you have now 0.05% higher risk for this. Should you worry about that? Well, it depends. Most likely, no.

But I think people have a hard time with probabilities, understanding that something can happen at 1% frequency. Some people interpret this as, "Oh, it could happen? It could just not happen. There's two choices, 50% each." So you're off by a factor of 50.

So yeah, I like the way that most people understand statistics is actually quite rudimentary. And that's just a flaw of human cognition. Basically, we tend to overemphasize the risk of rare events. Every time you go swimming, you're afraid of a shark eating you, but sometimes you forget to put mosquito repellent on when you go to the forest. Well, mosquito bites kill millions more people than sharks every year.

That's true.

And yet in our risk assessment, we're so focused on these extremely rare events. And that's just flaws in human cognition. So the problem with these flaws is that when a company then goes out and says, "Hey, figure out your risk for x and y and z." And you see that these probabilities-- and they're being very honest about those probabilities. Many people will misinterpret this. That's the first problem.

The second problem is the science is not quite there yet. Even if you take all of the common genetic variants, we just haven't done all the studies to be able to really predict how all of them combined will lead to your disease predisposition. And the third, as I mentioned earlier, is that it's a very partial test. It's one component. So it's still at the early stages, but it's definitely on the correct path to using more information, to better inform our decisions.

Great. Why are genetics so important to the pharmaceutical industry? What genetics gives you is very unique. So we're all talking about predicting disease. And yes, genetics can, in some cases, predict disease but to only a modest amount compared to all these other information that I mentioned earlier. But what genetics gives you that this other information does not give you and what makes genetics so unique is causality.

What does causality mean? Those individuals who drink coffee also tend to live longer. Is the coffee causing them to live longer? Or is it that those individuals who drink more coffee have better jobs?

They can afford coffee. They can afford and also better health care. Maybe, it's the fact that they have a better socioeconomic status. That makes them live longer, not the coffee. Maybe the coffee is just an indication of the fact that they have a better socioeconomic status. And therefore, an indication that they're going to live longer.

So that's the difference between correlation and causation. So every time those studies come out. That's "a coffee makes you live longer" and you read, it's actually just correlation. It turns out that if you drink coffee already, you're probably going to live longer anyway.

But what genetic information gives you is much, much richer than that. And the reason for that is that you inherit those genetic variants before you're born. And those genetic variants can tell you about disease causation. So what these genetic variants are allowing you to do is not just say, "Hey, you have a higher risk for that." But it's actually understanding the mechanism of the disease.

Why is that important? Because understanding mechanism, understanding that these genetic variant even if it only causes a 1% change in the disease, but if you know that genetic variant has a causal path to the disease. Then, you know that if you intervene in whatever that genetic variant is affecting, then you will have a causal implication of the disease.

And causality means that the drug that you're going to build now is much more likely to succeed. So there have been studies that basically say which drugs succeed more or less. And what they have

found is that one of the best predictors of whether a drug will succeed is if there was genetic support for that drug.

So the reason why the pharmaceutical industry cares about genetics so much is because they can save billions and billions of dollars by increasing the success rate of their drugs ever so slightly. So by going after the right target in the first place, namely those that are supported by genetic information, that are supported by this mechanistic models that we're discovering through genetics, then we can be much more confident that the drug will succeed.

And that's why we're collaborating with so many different pharmaceutical companies right now because they want to use our models to understand genetics. Genetics is wonderful because it tells you about causality. It tells you that these genetic variant will have some causal impact on the disease.

And what makes it powerful is that it doesn't matter how the genetic variant acts. You can predict its impact will be there. So that's the power of it. Regardless of the mechanism, genetics works. You can look at genetic variation how it correlates with disease and get causal relationships back.

The downside is exactly that. That, yes, after you figured out the association, you have no idea how it works. And that's what our group is trying to do.

What we're trying to do by these massive data profiling, this massive data integration and the systematic validation is it predicts the mechanism through which those genetic variants are acting so that when the pharmaceutical industry wants to target obesity, they're not going to say, "Let's go after the brain." They're not going to say, "Let's go after the muscles." They're going to say, "Let's go after the adipocytes, after the fat. Let's go after this pathway, this gene, the cell type, that developmental stage," and so on and so forth.

And that's what genetics gives you. Genetics gives you insight into the mechanism of disease. And it gives you much higher confidence that what you manipulate is causal and to look at the correlation versus causation question.

If you see that immune cells are going up for Alzheimer's patients, I don't know, should you bring them down? Are they causing the disease? Are they just correlated? Do they have new role? Or maybe are they fighting the disease? Are they going up because they're fighting the disease?

So if you see that firemen are always correlating with fire, are they the arsonists or are they your best help? If you're looking from space and your like, "Whoa, they're firemen every time there's fire." Let's

get them out of there." Fire might just get worse. So that's what correlation versus causation gives you by knowing exactly where to intervene and how in order to causally affect the disease, you can increase the chance of therapeutic success dramatically.

That's great. So you mentioned you were working with some pharmaceutical companies. Are you working with just pharmaceuticals or other companies? And how does your group work with the industry?

We have several collaboration with the industry. And they usually take the form of joint research projects. Basically, the goal is to build resources that they need.

And there's no better way to have an impact than to truly understand what does the pharmaceutical industry need for that disorder. Because if you just stay among academics, you might be misled into thinking that these are just the only important things that we should be focusing on. It turns out that the pharmaceutical industry might not care about them at all.

And the reason why I'm so excited to work with your group and reach out to industry and to pharma is because understanding their needs means that we can have a much higher impact on patients. By solving the problems, by understanding the needs that they have, we can basically have a much more relevant research program. Of course, you have to have both basic research and applied research. But if we can drive the needs of the basic research, the needs of the basic research is addressing based on the pressing problems in the field right now, we will have a much better impact.

So we are going to generate these resources anyway, but understanding which resources are the most useful to them, allows us to prioritize them. And what is the advantage to them? The advantage to them is that they can be part of that creative process. They can have early answers as we're building these resources, understanding that we may need to change them if we find mistakes, before we release them to the rest of the world.

But they are very open to the fact that, yes, even their competitors will have access to the same resources because everything we do with will eventually be published. But by being part of the creative process by being part of the discovery process, they can figure out also how internally they can repeat that process for many other resources that they're building internally.

And I think, there's a growing pushing the industry right now to bring more resources together to build these tools together to basically say instead of you building your, I don't know, own master immune map of how cancer interacts with the immune system and me building the same map with slightly

different data and that other person and that other company building the same map with slightly different data, let's all just collaborate in a precompetitive kind of way. And then, after we've built the best possible map, we will all benefit.

So society will benefit. Academia and industry will benefit. Hospitals will benefit. And then, each of us can then go off and compete in who builds a better drug, but let's start with common biological knowledge. And I think that has created an environment and an ecosystem where industry, academia, and hospitals and government, like the NIH, can all come together and say, "Listen, guys, we're all in this together. Let's just pool our resources and build better infrastructure for all of us."

And that's what a lot of these collaborative projects are about. We want to benefit the whole world yet giving them early access, having weekly meetings, learning from each other, collaborating closely. And we have much to gain, and they have much to gain. And very often collaborating with academia means that they will generate the same resources at a tiny fraction of the cost because this is what we do for a living, because we can leverage all these other tools and all these other things that we have built to basically make the next resource even better.

And for us, it means that our data will be better our tools will be better because somebody will immediately be using it rather than just having to wait until somebody contacts us a year later saying, "Actually, this is not very helpful. Here's what we would have liked instead." We can do that from the onset. And so the more you can build your alliances program to include these companies and the more we can really help them recognize the huge value of industry and academia working side by side, I think, the better it will be for society and for medicine and for the whole world.

That's great. That's fantastic. Now, you and I are both here nursing a hot cup of tea in this January flu season. So I have to ask about the evolution of diseases and variants and similar like I heard the flu vaccine was developed thinking one strain was going to go, but it's actually another. Does your work address any of that?

This is such an important topic. So it's extremely important to understand that disease. It's not just about genetics. It's the interplay between genetics and environment. And if the environment changes, the genetic impact might completely disappear or it might actually be reversed. I mentioned earlier how these FTO genetic blockers had no association with obesity 100 years ago.

It's only 15 years ago that it actually started having an association. That's because the environment changed in the same way. Our body is not isolated from hundreds of other organisms that are living within us.

So within our gut, there's a human microbiome with so many different bacteria that are sometimes beneficial, sometimes it's detrimental. It's a whole ecosystem that we're carrying around. There are more bacterial cells in your body than there are human cells.

As you walk around, you're mostly a carrier of bacteria. And along for the ride come a few human cells that make up you. So we have to constantly embrace this fact that we are co-inhabiting the space around us with thousands of different types of environmental exposures including viruses, including bacteria, including so many different microorganisms.

And I think, that what we need to understand is not just the human genome, but also the genome of all those other species and the interplay between not just our cellular circuits, but also their cellular circuits. And the viral circuits very often go through the human host. They go through the vector, the carrier that brings them to us, whether that's air or whether that's a mosquito and so on and so forth.

We have to build not just self-contained gene regulatory networks, but also trans-organism and trans-kingdom regulatory networks that incorporate the viruses and the bacteria and the full set of metabolites and chemicals and exposures that we have and so on and so forth. So that's very much where the world is heading. But it's only very early steps in that direction.

OK. Will your work help to develop better treatments or even cures for diseases, such as diabetes and Alzheimer's?

Well, that's exactly what we're trying to achieve. And I'm hopeful that this will come a lot faster than the field is currently anticipating. Basically, we're currently working on this many, many years horizon for when our therapies will actually when our new set of circuitry will actually lead to new therapies. But I'm very optimistic basically seeing the rate at which research is being translated.

And as soon as we write a paper, my friends forward me emails basically saying, "Hey, your paper has been circulating within a pharmaceutical company. You know, the CEO is sending an email saying, hey, pay attention to this just came out."

And when our FTO paper came out in the *New England Journal of Medicine,* we had a meeting with someone from pharma just like five days later. And we're starting to describe in general terms what we were doing. And he's like, "Wow, that sounds like a paper that I was reading last week." I was like, "Which paper is that?" He's like so and so. I'm like, "Oh, that's our paper."

He had received this paper from five different email threads, basically saying pay attention to this.

This is very important for us-- from five different sources. And this is so exciting. It basically says that academia is having a huge impact in industry even when we don't know it.

If I hadn't had that meeting with him, I would have no idea that this was happening. And he didn't even know it was me. He thought it was just like some unrelated paper. He's like, "Oh, yeah, are you familiar with this group?" "That's us."

So I think that both the methodologies that we're using and also the insights that we're making and that we're gaining will dramatically transform therapies, dramatically transform our approach even to new therapies. So the best way we can do that right now is to disseminate the results as widely as possible. But also, work with a few select partners that can help us stay honest that can basically help us understand whether we're really achieving this goal of enabling them or whether we're missing the mark.

And we're generating resources that, yeah, not everybody will use. So I think that we have to have both. We have to have the wide dissemination as broad as possible to everything we're doing in ways that make it much more usable, much more understandable, just so easy to just build on.

But at the same time, work with partners that will constantly tell us, "I don't understand this. What do you actually mean?" Or that will tell us, "I understand it this way. Is this the right way?" And we can say, "Oops, we were not super clear. And let's change that for now in the whole world."

I think, that's the impact of truly transforming not just an industry right now. But also the whole field of therapeutics and ultimately our societal impact of really approaching disease and approaching our understanding of medicine in a completely new way and much more rational way much more data-driven way.

Wow, your work is amazing and sure has so much impact for our world. We've talked a lot about the things that could happen or might happen. On a closing note, what else would you think the future would hold in this?

So thank you so much for the comments. I think, what I want to emphasize that really this is a huge team effort and a team effort not just here at MIT, but across many, many different institutions. That basically, the number of players there are contributing to every paper we write is huge. And this is through building on the work of giants, basically building on papers that are previously published before us, learning so much which is constantly disseminating, building on resources that many, many other groups are or are building.

But also collaborating so closely with doctors, the generosity of patients for making their samples available, the generosity of families for making postmortem samples available from their loved ones to really help science, the incredible creativity of our students who are really the drivers behind all of these discoveries. And it's really just, it takes a village. It takes really a huge effort with people playing very different roles to really create all these "children"-- all the papers that we're writing, all the discoveries that we're making. It really takes a village to make them grow and make them have an impact.

So what does the future hold? I mean, my view is that, number 1, a lot more collaboration, a lot more interdisciplinary collaboration. I think this is where the field is really heading. And I think, this makes my day so much fun, working with people who know things that are just so different from what I know makes me feel like a kid I'm constantly learning. I'm constantly like bright-eyed, soaking in information about how the world works, about knowledge that they have that I couldn't even dream of finding the textbook.

And a better understanding of disease freely fundamentally, sometimes throwing away the textbook view of some diseases and just starting from data, starting from a completely renewed understanding, much more modern understanding of these data sets. In some cases, reinterpreting old data in a very kind of new way and saying, "Well, this is actually what was causing it." I mean, going back to the example of FTO. Somebody had knocked out the gene, had deleted the genetic segment and said, "Oh, great."

This is clearly how that function, but it turns out what they had done is deleted the circuitry that was unaffecting those other genes. And that perfectly explains their experiments. So even though we had an experiment that was showing that ablation of that gene leads to that phenotype, that gene is not involved at all because something else was there.

So again, being able to embrace the continuum of science, the fact that we can publish as much information as possible so that others reading our papers, down the road, maybe 10 years down the road can basically say, "Aha, they misinterpreted their data. It's actually that other pathway," and so on and so forth. And of course, building all those experiments that can allow us to validate that what our computational causal models were showing are indeed truly causal.

So basically, number 1, a lot more collaboration, number 2, a lot deeper understanding of the disease, and number 3, a lot more integration of the genome in all these components. Better integration of truly DNA-based knowledge on every aspect of science. And then, the last thing is

really something that I was saying earlier, understanding that we're all in this together.

That if we do a better job to improve the health of everyone, then everyone will benefit. Not just, "Hey, you know who cares about this person is not going to have to pay a million dollars." No, our health care system is paying for those million dollars, not just some random person. It's your own tax dollars that are going to treating much more severe cases and if we're doing a better job at prognosis, a better job at diagnosis for everyone.

So we have to embrace the fact that these are overall health care costs for all of us. And if we improve the health of every person in our society, we are all better off because diseases won't spread as widely because the knowledge gained will help me and will help my family, will help my loved ones when we get to that stage because you don't know what the future might hold for you and your health and your family. By supporting a much more inclusive health care system where we are really building these genomic resources for everyone and distributing the cost across all of society, we will save so much money in end.

And so in the end, not only are we better human beings, but we're actually were better off economically because we are not waiting until the symptoms are huge to address them. And the health care cost is huge on all of society. But instead, making a new future where we can better understand the molecular base of disease, where we're not going be giving the wrong drug to the wrong people, where we're not going to be completely misdiagnosing conditions until years, years down the road.

And I think that future is really a bright one. It's a future of more knowledge. It's a future of better health. It's a future of a more inclusive society. And it's a future where, I think, our children can grow up, understanding exactly what's causing their illnesses rather than taking wild guesses and throwing bazookas at it. It's really making it personalized, making it much more precise, and making it much more effective.

That's great. Well, thank you for all the work that you do. And thank you so much for joining me this morning, Manolis.

Thank you so much, Lori. And thank you for all the work that you're doing to really help industry and academia come together and really create a better society for all of us.

Thank you.