

Webinar Genetics Gallery - November 1, 2023

Opening Remarks 1 - Vicky Whittemore	2
Opening Remarks 2 - Oved Amitay	4
Lived Experience Perspective - Hayla Sluss	5
Session 1: Large data-sets and Genome Wide Association Studies (GWAS) Talk 1: Genetic risk factors of ME/CFS: a critical review (GWAS) - Chris Ponting	6
Talk 2: Combinatorial analysis of Genetic Risk Factors for ME/CFS, UK Biobank - Stephen Gardner	15
Talk 3: Genetic susceptibility in Long Covid and ME/CFS: Long COVID Host Genetics Initiative (international network) - Hanna Ollila - Anniina Maria Tervi	23
Session 1 Q&A	29
Session 2: Enriched cohorts and Epigenetics: Talk 4: Characterizing the Genetic Basis of ME/CFS through Case-Control and Family Studies - Fereshteh Jahaniani - Varuna Chander	35
Talk 5: Contribution of Epigenomics to ME/CFS Pathogenesis: Past, Present and Future - Alain Moreau	42
Closing Remarks - Kristina Allen-Brady	49
Meeting Adjourned - Vicky Whittemore	53

**ME/CFS Research Roadmap Webinar Series – Genomics and Genetic Susceptibilities**  
**Open Session**  
**Wednesday, November 1, 2023**

**Vicky Whittemore:** Okay, I think we'll get started. So welcome everyone to the fourth in a series of eight webinars that we're presenting as part of the ME/CFS research roadmap webinar series. I'm Vicky Whittemore. I'm a program director at the National Institutes of Neurological Diseases and Stroke at the National Institutes of Health, where I oversee grants on ME/CFS and epilepsy, as well as work with the Trans-NIH ME/CFS Working Group. So what we've planned and have been working with the NINDS ME/CFS Research Roadmap Working Group of Council is this series of webinars to understand the state of the art, what we know, what we don't know, and what we need to know to really move and accelerate research forward on ME/CFS. So I would just like to acknowledge all of the individuals who are participating as members of the Research Roadmap Working Group of Council. So it's representing individuals from other federal agencies, other NIH institutes, investigators, clinicians, as well as individuals who are individuals with lived experience, either themselves have ME/CFS, or are leaders of advocacy organizations, as well as other individuals who are participating in this effort.

I would like to, for this particular webinar, acknowledge Oved Amitay from Solve ME/CFS, as well as a group of individuals who are part of the Genomics Genetic Susceptibility webinar planning group who have planned and organized this particular webinar and also the team I'm working with at NIH across NINDS and also acknowledge our contractors at RLA, Holly Riley and Damon Cain in particular, who have really been instrumental, all of these individuals in making this webinar series possible. So just, we're at the halfway point with these webinars. There are four additional webinars coming up and you can see the dates here from November 30th through January 11th for the remaining four and you can go to this link. If you go to NINDS and about NINDS and then to the NINDS, who we are in the Advisory Council, you can get to this link and see the information for all of the upcoming webinars. So just some quick guidance for webinar participants.

So the goal of this webinar, as I said, is to identify research priorities for research on the genomics and genetic susceptibility to ME/CFS. And we're really very excited about this webinar because I think we're going to be hearing information that we have not seen presented before about genomics and genetic susceptibility. And so what the goal of each of these webinars is to really identify the research priorities that will all come together to form a report that will go to the NINDS Advisory Council and NINDS leadership at their May meeting in 2024. So we ask that the questions that you put in the Q&A for each presentation, and we'll allow time after each speaker for some Q&A, be really focused on research and research priorities. And we ask that you don't ask questions related to your individual health issues or

related to your own potential genetic findings because we're really not able, in this format, in this forum, to answer those kinds of questions. So you could ask clarifying questions or questions about research priorities as they've been presented by the speakers.

So, for additional feedback, you can send emails to this email address, [mmecfsresearchroadmap@ninds.nih.gov](mailto:mmecfsresearchroadmap@ninds.nih.gov), and the best way to receive announcements and updates from NIH about this webinar series and about all events and activities related to ME/CFS is to go to this website, [www.nih.gov/mecfs](http://www.nih.gov/mecfs) and sign up for the email listserv. And you can watch for announcements there to participate in discussions on the research priorities on our crowdsourcing platform IdeaScale. So we've just we're just about finalizing the research priorities from the very first webinar on the nervous system that will be posted and so watch for that announcement coming very soon.

So with that, I will turn it over to the chair of this webinar planning group, Oved Amitay, to introduce our first speaker. Thank you.

**Oved Amitay:** Thank you very much, Vicky. I'm Oved Amitay from Solve ME, a national patient advocacy organization that was established in 1987 and serves as a catalyst for critical research into diagnostics, treatments for ME/CFS and associated diseases. We work at this intersection of science, policy, and patient empowerment. I'm a pharmacologist by training, and prior to joining Solve, my professional career was dedicated to developing therapies for people with rare genetic diseases. So, my work over 25 years has led to eight treatments approved by the FDA that are now used by many people around the world. I mention this because this is actually a high number for an industry that has notoriously a very high failure rate. In fact, only around 10 percent of drug development projects make it all the way from phase one to approval. However, programs that use genetically validated targets, like those that were used in the studies that I'll be involved in, are twice as likely to be successful. That's a tremendous difference. So, although ME/CFS is not likely to be fully explained by a change in a single gene, I believe that genetics can provide incredibly important clues to understand the disease, and more importantly, to successfully develop solutions.

Unfortunately, the research into the genetics/genomics of ME/CFS has been limited, so as Vicky said, our webinar today is a bit different from the other webinars in this series. There is not that much to review from existing published information. Instead, we try to really bring us, today, an update on where research is at this moment with some information that will be shared today for the first time.

Unfortunately, we had some last minute changes to the program. Dr. Petrovsky from AstraZeneca had an unexpected committee and he cannot join us today and Dr. Ashley Beckham from 23andMe could not present today, but we do hope that she will be joining us later today so that her contributions to the discussion will inform us as well. But that gives us a little bit more time for questions and perhaps even a slightly longer break.

So, to get us started, it really is my pleasure to introduce Dr. Hayla Sluss, a researcher and a caregiver for a person with ME to get us started. And Hayla is sharing with us through prerecorded comments.

**Hayla Sluss:** Hello, my name is Hayla Sluss. I have been a caregiver to someone with ME/CFS. I wanted to thank the NIH and the participants for putting together these webinars. It is an honor to be asked to speak about my lived experience with ME/CFS. Like many people, I had a family member that was quite ill with many symptoms, including fatigue, sleep issues, abdominal pain that would often lead to the ER, and leg pains that would ache so much they couldn't walk. After exertion, they would often have to be in a completely quiet place and would almost appear to have paralysis. There were other symptoms like hypermobility, and this was on top of a diagnosis of severe asthma. It was many trips to the ER and many doctors. It was a tenacious disease that really didn't want to let go. I am also a researcher, so I applied my skills to figuring out what it could be. Immune dysfunction, IBS, metabolic disease, adrenal insufficiency, primary sleep dysfunction, and on and on. I suspected ME/CFS. It is, after all, a little bit of everything.

After a diagnosis, this did not lead to any clear path to treatment or relief or acceptance of my family member's disease. One thing that is not discussed enough is how ME/CFS is a family disease having effects on all parts of the family, extended family, and friendships, and are feeling connected to the world. The course I took to belong was to begin studying the comorbidities and commonalities of associated disease with ME/CFS in a study called Reclaim, which I began while still being a primary caregiver. With this debilitating family disease, it is important to have something that can connect you. The multisystem aspects of ME/CFS are a clue to the disease that ME/CFS is a multisystem dynamic disease. There's likely a genetic component and an environmental component that contributes to it. This genetic study section of this research roadmap is so important. Genes linked to ME/CFS have been described, and these studies here are even with greater number of patients that are sure to provide statistical resolution that even the most doubtful cannot deny the disease. Not only will these studies help identify risk for ME/CFS, but it will also be a launching point for personalized medicine. The studies in this genomics section are continuing the amazing work, which will be critical to understand the complex risk of ME/CFS. Thank you, and "obrigada".

**Oved Amitay:** Thank you very much for those comments, Hayla, and sharing your personal experience and describing so eloquently why today's focus on genetics is so important. So the study of genetics and genomics of ME/CFS could come from two different directions. One is from looking at large populations, large data sets, which will be the first part of our webinar today. And then also, on the other hand, from looking at smaller cohorts, families, which we'll do in the second part. So to get us started, our first speaker is Dr. Chris Ponting from the University of Edinburgh in Scotland. And I actually can't think of a better speaker to open the session and give us an overview of this research approach, and it's used in ME/CFS so far, but mostly because Dr. Ponting is now currently leading the DecodeME study, the largest study of this kind in ME/CFS. So we're very excited to hear about that. Chris, please.

**Chris Ponting:** Thank you very much, Oved. It's a pleasure to be here, and thank you very much for the invitation to contribute. So I don't have lived experience of ME, and I need to say this upfront. I do know people with ME, and having heard so much from people with ME over the years, it fortifies me to doing as much as possible. And what we're doing at the moment is the DecodeME study. And this is funded by the UK Medical Research Council and the National Institutes of Health and Care Research. It didn't come out of nowhere. It was a long time coming, it was being cooked up from about 2015 onwards and a whole variety of people were advocating for it inside and outside of the scientific community. So in the end this is a project that I hold close to my heart but also I know is a project that means a lot to people in the United Kingdom and beyond its borders.

So what do we know? We have been asked to address these questions in the presentation, so I just thought I'd go straight ahead and give you what I think of as being my answer. Yes, we just heard from Hayla. Yes, ME is in part genetic. It is clearly not in full genetic. It is in part genetic. There will be and definitely are environmental contributions. Now, what don't we know and what do we need to know if I put those together? I think we need to know the genetic risk factors. That isn't merely because I'm a human geneticist. It is because once we know the associations, we know the genetic risk factors, they will point out the organ systems, the cells, the genes that cause ME/CFS disease and its symptoms. It is a multi-organ system disease. We've heard again from Hayla about that. It is going to be a complex disease and we will need to tease it apart and ensure that if there are differences among people with ME as to why they've got the disease, then we need to know those differences as well as their commonalities.

So what are the research priorities for this area, which is genetic and genomic susceptibility? Well, I think, and I would say this, wouldn't I? But I'll try and explain why, that in the US you will need to have a genetic cohort. It needs to be large for reasons that I'll explain and it needs to be representative and it will provide DNA discoveries and be the launch pad for future clinical studies because it's not just, as I'll explain, that people can be recruited into a genetic study. But they can be recruited into further studies down the line, according to their genetics, according to their symptoms and their comorbidities. How much would this cost? Roughly \$5 million, roughly. So why are we even talking about genetics here, given that the trigger for two-thirds of people is an infection, it would appear? Well, it is that there can be and often is multiple members of the same family that are diagnosed with ME/CFS. We've known this for some time. And one very cogent explanation of that is that it's in the genes.

Now, if you look at individual genes and this is a study, you don't need to know necessarily what the figure means, but this is a study of every gene in the human genome. And you ask whether there are rare variants that break these genes. You ask whether any one of those genes increases ME/CFS risk. Then in this cohort, which is a couple of thousands of people with ME, there is no such rare variant that can explain their ME. So we do not think that there's a single variant, and

at the moment we don't know that there's a single gene that increases ME/CFS risk. We think there will be many variants and many genes. We don't know that, but if we read the literature of many other diseases, then that turns out to be what happens in those diseases. So let's assume that and let's follow what the methods are to try and find what those genes are and what those variants that predispose people to ME are.

So the relative risk for relatives is high. So what does this mean? This means that if you have ME, then you are twice as likely to have a first or second or third degree relative also with ME. And this is interesting because it's not just the first and perhaps second degree relatives who might share the same environment, but also the third, which immediately gives credence to the idea that it's not a shared environment that predisposes people to ME, but rather shared genetics. So why ME genetics again? Well, Oved's already said this, and here's the first paper that said this, there's been at least one since, that if we want to look forward to, as we do, clinical trials that are successful for treatments of ME, then we know from looking back at genetic studies that having evidence for the gene doubles the success rate of those clinical trials. So that's really important. It's important not just to convince pharmaceutical companies, but also, and we heard this from Hayla, to convince everyone in society that it's not all in your head, that there is a part to play of genetics.

The next thing to say about genetics is that it's comprehensive. It looks across all of the chromosomes, all of the DNA, all of the 3 billion letters that make up our genomes. It comes to the problem unbiased. We don't have a view as to what we're going to find. We look in an unbiased fashion across every part of the genome that is commonly variable in the human population and ask, is this variation different between people with ME and others in the general population? So we're not rooted in necessarily known biology. We don't have our favorite gene. We don't have a, you might say, a prejudice as to what we're going to find. And I think that's important too, because then we'll be able to find new things that hadn't been previously suspected. But more importantly, if we find that genetic factors are common to a particular bodily system, then that pinpoints that bodily system as being the seat of the disease, what has caused it. We don't know this yet, whether it's in the muscles or central nervous system. It may be a neuroendocrine problem or in the immune system or in the autonomic system, or indeed in combinations of them.

So this is what we do in a large genetic study, such as DecodeME. We look at the DNA of individuals and look particularly at positions where people commonly have differences in their letters, which are the Gs and Cs and As and Ts here. And we count, in the people with ME, whether they have one of those letters, C, for example, versus the other, T in this example. And if there are more Cs in people with ME, then would be expected from the general population, and that is significant, I'll come on to that, then we have a genetic association. We have a link between a genetic variant and the likelihood of someone being diagnosed with ME. And as I

said, we do this across every single chromosome and every single one of our 19 to 20,000 protein coding genes. So what will we find? To date, we don't know. So we have to use statistics. And the common statistic in the field, in fact, in science, is that our probability,  $p$ , is less than 5 percent. And this is that the probability of a null hypothesis is true. However, if you do this a million times, you'll get 50,000 false discoveries. So we can't use the common 5 percent threshold. Rather, we reset that threshold as that 5 percent and we divide through by the million. And that comes to a new significance threshold of 5 times 10 to the minus 8. So we have to have statistical significance to that degree of certainty. And for that, we need large numbers of people. So, this is why in all of these studies, and in particular here for DecodeME, we need not tens or hundreds of people, but thousands and tens of thousands, and in the future, I'm going to argue 100,000 people.

So, I'm going to give you a couple of examples where genetics has made a difference. The first is a recent one of acute COVID-19. This was a tremendous effort involving so many people in ICUs and in university departments during the COVID pandemic. And what they found was a link to a part of the genome that had a gene called TYK2. And that immediately implied to the researchers that an inhibitor of TYK2 might be useful in treatment of people with acute COVID. And when that inhibitor, baricitinib, was used, it saved lives. So genetics here has been used very quickly in a very short order to identify a drug that was already available, it was on the shelf to be used, that saved lives.

Secondly, Crohn's disease. So, this is a much older study, and I told you previously about a threshold of 5 times 10 to the minus 8, and this is the dotted line here. And so, any of these dots that lie above that line represent a signal for statistical significance of association with diseased individuals rather than in cases, therefore, rather than controls. And this gave a signal at the gene, which was called the interleukin 23 receptor. And this was a long time ago and was heralded as a breakthrough. And indeed, this drug has since been used for refractory Crohn's disease and it is effective. Again, genetics has been used to help to treat individuals. This is called repurposing. And at my most optimistic times in my research career, I hope that what is found in ME genetics, by us or by others, are associations which immediately allow the repurposing of a pre-approved drug, approved for other purposes but can be used for treatment of ME. We know that it is safe for use in people.

So let's think about that association that was made 17 years ago and ask, what's happened in that intervening time? Well, I'll tell you one thing that has happened. One thing that hasn't happened, there hasn't been an ME study of this size that was required here in 2006 to find a genetic association. So here's the last 15 years of genetics discovery, GWAS discovery, in a review article from this year. There are three things that are being highlighted here in different colors. First of all, in brownish color at the bottom is the average sample size between roughly 5,000 and 140,000. So you can see that over 15 years, the sample sizes have gone up



enormously. The second thing, in green, are the number of places in our chromosomes that have been indicated as being associated with these diseases. And that number started off at a few, up to tens, and for some diseases now, it's 30 to 40 or hundreds or even thousands for some. But not for ME/CFS. We are at least 10 years behind other diseases. And so you have to ask the question, well, why is that? Why are we so far behind? And even if we don't want to answer that question, what are we going to do about it now?

Well, our study in the United Kingdom aims for 20,000 samples. We have 26,000 people who have answered the questionnaire and 21,000 people who have signed up for offering their DNA, which is absolutely tremendous. And it's absolutely fantastic that so many people have placed their trust in us. But, it's also a devastating thing because they've wanted to contribute to the study because there hasn't been a previous one of this type or size. And we hear often from them, they say they're desperate to be involved in studies like this, but they have not been able to for 10 years and more. At 20,000 samples, according to this graph and according to our grant application, there is an expectation of five associations that we will find. Now you'll see that if you go from 20,000 and you increase the sample size more to like 100,000, then you'll go to 30. And 30 discoveries will tell you with a great degree of certainty what is going wrong in ME. Five will tell you something, but what it won't tell you is with certainty what is going wrong across the whole spectrum of people with the common and the diverse symptoms of ME. So what I'm going to say is that I think that the world needs to come together and bring genetic cohorts together to perform what is called a meta-analysis of all the data so that we can find this number of associations and work out what is going wrong.

So where are we at the moment before DecodeME? Well, the largest one thus far is round about 2,000 people. There are no risk loci at this number of individuals. This came from UK Biobank. Interestingly, it's not just numbers. We also need ancestrally more diverse cohorts. In the UK, and I'll say this in a minute, our cohort is not ancestrally diverse, and we need to study more people and more diverse people.

So, this is a slide from a couple of days ago. So we are almost at 21,000 and we are close to closing our recruitment in the United Kingdom. We are recruiting people of all severities, adults of 16 years or older. Anyone can participate from home. If they wish, they can complete a paper questionnaire, but most people complete an online questionnaire in their own time. They can go away and come back later on a different day and again and again until the questionnaire is completed. In one go, it often takes about quarter of an hour. We are pledged for sharing the data to bona fide researchers who will come to us with ideas about how to use the data and are involving people with lived experience in their projects. And I will tell you something which I had never appreciated before that doing this as a co-production with people with lived experience was essential for delivering on this critical question.

So, it all can be done from people's homes via the mail system. And so, that overcome some limitations of clinic-based recruitment, but I recognize that it is the case that some would say that the phenotyping, in other words, the collection of the symptomology, may not be as well-defined. When we took receipt of the first 17,000 questionnaires, we were able to look at what was common in their symptoms of all of those people. And what we're showing here is the symptoms, how often a symptom is reported by people between zero and 17,000, zero at the middle of these circles and 17,000 at the edge of the largest circle. So, you'll see the post-exertional malaise is reported by the most, as expected, with fatigue and refreshing sleep, brain fog, concentration issues, gut symptoms, muscle pain, noise sensitivity, flu-like feelings, all of which are not a surprise, I think, to anyone who's been working on ME for a long time. But what it does do is tell a story of over 100,000 years of lived experience.

So, nothing about ME without me. This was a co-production and it was really important for us to co-produce this because of the different ways in which value was given to the project by people with lived experience. And we used the UK national standards, which were incredibly important, to deliver our project and to ensure that we had the trust of the people. We think our project is representative. It has recruited from across the United Kingdom and across all of the different ages. We have recapitulated what people believe as being this strong female bias, almost 5.5 to 1, we see. Now we're able to compare that with data from NHS England to say, is that fully representative? Well, 3.9 to 1 female to male ratio for people in GP practices in England. And we think that the slight difference between 4 and 5.5 to 1 is due to the fact that we're sampling more females at the younger age because the female to male ratio is higher at younger ages.

What we are not good at in representation is the non-white population. There's a profound bias in DecodeME. Actually, it's a profound bias that occurs among people in England who are diagnosed with ME. I won't have time to go through what this means, but essentially, there are some communities in England where the diagnosis rate is 10 times fewer than among the self-reported white British. So, we're not representative in with respect to severity. We would recognize that if one believes that 25 percent are housebound and bedbound. We're not seeing those numbers in the severe and very severe categories. And the reason is, we think, that we're using the National Institutes of Health and Care Excellence definitions, which are overly stringent with respect to severity.

So I think that I will recommend to the USA, a large 50 to 100,000 strong cohort. To look at common DNA variants, it's not that expensive. To compare females versus males and their genetic predisposition, infection versus not at onset, and then to move on to rare variants, whole genome sequencing, which is more costly, but if you start with the most severely affected, then I think you will begin to see genetic signals faster. Then, with all of that data, you can compare against other genes, other diseases in a genetic way, not just ME with long COVID, but any other

condition. And then importantly, that cohort is then available for downstream trials by any criteria, by severity, by sex, by onset, by geography, whatever. And then lastly, my recommendation to you is gain from PPI, share the data that has already been gathered and all the data that has yet to be gathered and the expertise globally, because it is a global problem. I want to thank everyone in DecodeME and the funders, and I'm very happy to take questions.

**Vicky Whittemore:** Thank you so much, Chris, for that excellent presentation and overview. We have several questions that have come in, so I'll start down the list here. Literature shows that some people with ME/CFS have human leukocyte antigen HLA associations. Are HLAs the same thing as genes, or how are they different? And are there HLAs that make it difficult for some people to filter out toxins from their bodies?

**Chris Ponting:** Thank you for that question. HLA genes are genes and associations have been made to them. They do not reach the statistical threshold that I was trying to explain. So I do not think of them today as being shown to be statistically significant. So we don't know that HLA genes are truly associated with ME. Will we know? Well, I hope that we'll know through DecodeME or through the US cohort if there is one, because we will definitely be looking there as we will be looking everywhere for these signals.

**Vicky Whittemore:** Thank you. So the next question, I'll ask the question, but we may want to defer this because I know Alain Moreau's going to be talking about this. Is it possible that ME is an epigenetic phenomenon on top of a prior epigenetic phenomenon? And if so, how would that mechanism be discovered? I don't know if you want to comment, Chris, or we can punt that until after Alain's presentation.

**Chris Ponting:** I'll say something quickly and just say that everything in epigenetics is indeed on top of the genetics. And so often you can use the genetics to see whether something can be described as epigenetic. That sounds like an oxymoron, but it is not one or the other because you can look at the consequences of genetic variation through the lens of epigenetics.

**Vicky Whittemore:** So the next question is very specific. So to what degree do MTHFR, so methylene tetrahydrofolate reductase enzyme, genetic defects contribute to ME/CFS? Do we know at this point?

**Chris Ponting:** So there are not any common variants, DNA variants that are very common in our population in this gene that are more common in people than ME than the healthy population at a statistical level of significance. And that is true for every single gene in the human genome at the moment, unfortunately.

**Vicky Whittemore:** So I'll ask this question and actually direct it to Oved about, and you can maybe answer briefly, Oved, and we'll certainly talk about this more later. The question is, has NIH considered integrating ME/CFS genetics research component with the pre-existing All of Us effort? So I'll give you a chance to say something briefly here, Oved, and then we can come back to that in the discussion later.

**Oved Amitay:** Sure. So for those who are not aware of the study, All of Us is a U.S.-based study with the goal of enrolling 1 million people, getting their genetic information actually to the level of whole exome or even whole genome sequencing. And on top of that, get their clinical manifestations. The first tranche of the data from about 300,000 people was released earlier this year. And it's already made available for investigation. We have solved and looked at the data and it looks like there already are people with diagnosis of ME/CFS in this cohort. And now the goal would be to get to the level of really looking at the genetic information. That's going to require more expertise. So definitely we will need some effort from NIH to enable to do that.

But I will just say that even if, you know, we get to a million people. And if, you know, the epidemiology suggested it's about maybe one, two percent of the population has ME, you will only get us to that 20,000 people that, Chris, you already have. So that's not going to be sufficient. We're going to need to go beyond that to really create a cohort that's much more well-defined and has the right number of people to get us there. So in the short run, definitely we should use that because this is already happening and it's paid for by the by the taxpayers' money. So we should do that, but it's not going to be sufficient.

**Vicky Whittemore:** Thanks. Next question for you, Chris. There are many families with several cases of ME/CFS or unexplained illness that overlaps significantly. So I know the question really is also saying that there's been little effort to recruit. I know that the Stanford group who's speaking later will discuss that. And I guess specifically for you, Chris, do you know if there are individuals with ME/CFS from the same family represented in DecodeME?

**Chris Ponting:** Anecdotally, because they've told us, yes, definitely. But we also will analytically know that. We can tell from the data how people are related and so it goes beyond the immediate family. The types of analysis that we can do go beyond immediate family because we're able to look far beyond to people that have never met one another who are still relatives who might then have common variants that we can find as being associated. If I can just add to the All of Us discussion and say many of those people I would imagine were recruited in clinics, and the most severely affected perhaps would not be represented well in all of us. And to do this in a representative way would therefore require more spit and post, we call it in the UK, approaches rather than clinic-based recruitment.

**Vicky Whittemore:** Actually, I just signed up at my local pharmacy. So it's not everyone is, as a control. I don't have ME/CFS, but to participate in All of Us. So it is possible for people in many different ways to participate easily. Oved, did you want to comment?

**Oved Amitay:** No, I think that's exactly right. And it also speaks to why we need to have a special effort. But one of the biggest advantages of All of Us is that it is actually very, very diverse. So the study coordinators have done a phenomenal job of really getting people from across different ethnic groups, among different parts of the US. So it's truly a representative study and I think in that sense it will give us a lot of information that we do not really have at all.

**Vicky Whittemore:** So a question for you for DecodeME. Are you considering an omni-genetic approach, rare plus common variants? In other highly polygenic disorders, it's been shown severity at a younger age tend to have both several rare variants and highly polygenic background.

**Chris Ponting:** I hope that in time we'll be able to do that, but at the moment we're only measuring common variants, so we can't look at the rare variants and their combination with the common variants, but I will say that we have split the DNA sample into two and stored one to allow the kinds of sequencing-based approaches to identify the rare variants. We don't have the funds to do this, but we would be delighted for anyone who wishes to sequence large numbers of people of stored DecodeME DNA samples to come forward because it's taken a long time to get this repository and it's available for use.

**Vicky Whittemore:** You touched briefly on the ME severity. If you're accepting all severities, would that not cloud your data as to what really is genetically behind ME as it might not be all the same or similar?

**Chris Ponting:** Right. So we can split our participants by their symptoms. And we indeed are going to do this with respect to infection and non-infection at onset, but we could indeed go further and split by severity. The numbers of people severe and very severe are quite low, below the 2,000 or so, which is recommended for this type of analysis. We need more, I think, to do that kind of analysis.

**Vicky Whittemore:** The question about what types of control patients are planned to be used in your GWAS study.

**Chris Ponting:** And we're spending all our money on people with ME. And not on the control individuals. We're fortunate to have a half million strong population cohort called UK Biobank. And we're emulating what they did for half a million people across the UK in DecodeME so that

we absolutely match our cases and controls. And I would recommend you do that in the US too, because of your population cohorts.

**Vicky Whittemore:** So there's a comment, question here about being impressed with the study that's underway in the UK, but taking away from the presentation that 20,000 really is not enough people. And so the question is specific to you. Are there plans to increase the size for DecodeME study specifically and to actively recruit more diverse patients? And I'll let you answer and then I have a comment about that.

**Chris Ponting:** Yeah, so we costed up how to increase the size, we costed up how to increase diversity. But as I said, there is a great representation of people being diagnosed in certain ethnicities in the UK. Those costs were too great to fit within our budget. And we would wish to do so much more, but at the moment, we are focused on delivering on this one aspect.

**Vicky Whittemore:** Yeah, so I think look forward to what you're going to hear in additional presentations today. We had hoped to have a presentation on data that's been collected by 23andMe, which they weren't able to present here. But I think that they have hundreds of thousands of individuals. And so I think one of the things we'll come back to in the discussion and as we wrap up today is how can we bring all this data together? And how to move this forward in a way that really brings all of this GWAS and genomics and genetic susceptibility data together. There are several questions we didn't get to. Hopefully we may have time later in the discussion to get to them, but thanks very much, Chris, and I'll turn it back over to you, Oved.

**Oved Amitay:** Thank you very much, Chris. This was fascinating. Thank you so much. And congratulations for this extraordinary study. And we can't wait to get to the results. But as you helped us to understand, this is just the first step to bring us to where we needed to be 10 years ago. So there's more for us to do for sure. Our next speaker is Dr. Steve Gardner from PrecisionLife, a private company based in the UK. And Dr. Gardner's work is taking GWAS one step further, looking at those interaction of multiple genes as a risk factor. So we're delighted to hear from you, Steve, please.

**Stephen Gardner:** Well, thank you very much Oved and Holly et al. for inviting me to speak. I'm going to talk both about our work in ME/CFS primarily, but also a little bit of breaking news that we have from long COVID analysis, which I'll cover very briefly towards the end of this presentation. So everything that I'm going to talk about today is covered in four papers, the references for which are all here, and they cover ME/CFS analysis, then long COVID and the commonalities with ME/CFS. And there will be some additional reference to work that we did previously in COVID-19. As Oved said, we're a commercial organization. We are a precision medicine analytics business, but we work very closely with patient charities, with research consortia and other groups, both to understand what the patient communities would like out of research and also to benefit from the deep disease insights that are enabled by the key opinion leaders in the space.

So aside from a collaboration with Chris, and Chris has done a fantastic job of setting the scene in terms of where traditional GWAS analysis can get us, we're also working with Nick Lemoine, who is very active in the COVID world in the UK and also Paul Elliott, who's led the REACT studies in the UK, including the collection of genetic evidence on around 150,000 patients. We also work with an institute out of Salt Lake City called Metrodora, who specialize in diseases of the neuroimmune axis, which covers ME/CFS, long COVID, and a variety of other conditions like POTS and EDS. So we, well, actually I've been involved with the Human Genome Project. I should tell you, I'm a computational biologist by training. I got involved with the Human Genome Project back in 1994 when Jim Watson was on one of our scientific advisory boards. And we've seen the transformational impact that that project has had on oncology and idiopathic rare diseases, rare genetic disorders.

We've been very frustrated that the same transformation has not been possible in more complex diseases. So as well as ME/CFS and long COVID, I would include neurodegenerative diseases like Alzheimer's, as well as ALS, schizophrenia, cardiovascular disease, and a range of other disorders where genomic medicine has had less of an impact. So really the purpose of PrecisionLife, the mission was to be able to take large patient data sets including not just genomics, but also their clinical histories and electronic health records, transcriptomic data, and those epidemiological and environmental impacts on disease, and use them to understand what is driving disease for individual patients, and herein lies one of the first differences.

We explicitly recognized that complex diseases may have multiple causes. In other words, it's possible to have the same symptoms showing up in a doctor surgery, but for those symptoms to be caused by different genes, different pathways, different mechanisms in the biology. And this is a very good example. This is early onset Alzheimer's. The colors represent different subgroups of patients who have a different genetic etiology of their disease. And why is that important? Well, number one, you want to be able to tell which group is which, and number two, they're likely to respond to different medications. So being able to tell those apart is really

important. We can do this because we have access to around about 35 sets of data. So, as well as the UK Biobank that Chris referred to earlier, we have worked with Optum and UnitedHealthcare in the US. We've worked with Sano Genetics to collect long COVID data. And I'm very happy to say that we're also working with University of Edinburgh and DecodeME and Action for ME on the ME side.

So what did we want to do with this data? So those insights into disease can be pretty powerful. Many of you will know that recently there have been approvals for drugs in Alzheimer's for the first time in decades of study. And the challenge there, the reason there's been 130 failed clinical trials is because the mechanism that has been chased, the amyloid and tau hypothesis, is the most important disease driver, but actually only in about a third of patients. And existing methods don't connect the disease biology to the individual patients in a particularly precise way. That's what we aim to do. So in other words, we can use these, the insights and the genetic associations we have here as biomarkers to identify the mechanism that's at fault in individual patients. And that allows us to select those patients for clinical trials, which means the trials can be smaller, faster to read out, which is important, but much more likely to be successful.

It also means that we can use these kind of tools in the clinic, not only to assess the risk for an individual, but also to choose the medicine that they're most likely to respond to. And once we've actually looked across multiple diseases, we also get to do that repurposing work that Chris described in a very principled way. In other words, if we've seen medicines targeting a gene that is encoded for in this green population, and we know that medicine is safe, we can choose that medicine and go and test it against this green group of patients. So that's basically what PrecisionLife does.

The starting point for this has to be being able to find quite a lot of signal out of existing data sets. And so we work a little bit differently from the way that Chris described. Instead of looking for single mutations that are associated with a patient population instead of a control population, we realize that chronic diseases are multifactorial. They arise because of the interaction between multiple genes. And there are often multiple populations, as I said, with different causes of disease. And actually a lot of the factors driving disease may not even be in the data set. They may be environmental and epidemiological. The only way to capture all of that signal is to look for combinations of features that together are associated with the disease. And this captures the non-linear, the unpredictable effects of interactions between those features. So many genes will interact through metabolic networks and there will be feedback loops in the biology that either inhibit each other or amplify each other's effects. You can't reconstruct that signal from a traditional GWAS results set, you have to go looking for the combinations in the first place, and that's really hard to do.



So this is a real example. This 6-SNP signature occurs in 150 women who've got pathogenic BRCA2 variants, but they don't get breast cancer. And, you know, aside from the obvious clinical risks, the utility of that in assessing risk, it's also very useful to start to pull together a testable hypothesis of what might be going on. And here we think that the insulin receptor is being blockaded and that prevents tumors being formed, the activation of an oncogene, which means that your genome surveillance apparatus being broken is less of a problem to those patients. But this testable hypothesis is one of about 3,000 signatures that we get out of a single GWAS dataset. Now we've run this analysis across around about 50 different datasets. We have a focus in precision neuroscience, but we have also done a lot of work in pathogen-mediated diseases. And I would put all of the COVID, long COVID, ME/CFS into that kind of category.

In terms of the ME/CFS study, we actually started with UK Biobank and we had two populations. The first was defined by a pain questionnaire, and this is really ME/CFS cases, round about 2,400 of those. And we match those by gender and ethnicity, sorry, sex and ancestry, that should say, to healthy controls, we then had a completely separate data set. And this is primarily a CFS diagnosis in about 1,300 cases. So these are separate and different patients from the first data set. And we used a similar number of controls for this analysis. So what we're trying to do here is not only find interesting signal, but then also be able to demonstrate that some of those signals replicate in a slightly different cohort. And I'm using the word slightly different because these all come from UK Biobank and all have the same fundamental biases in terms of ancestry that Chris was describing.

In terms of the GWAS analysis on these data sets, as Chris has eloquently said, they're too small in order to give signal above the 5 times 10 to the minus 8 threshold. But when we ran them, we were able to identify 84 combinations of features. We call these risk signatures or disease signatures. There were 25 SNPs that occurred in all of those risk signatures, and they map to about 14 genes. And in total, when you add up the cases represented, the percentage of cases that have one or more of these disease signatures, it was about 90 percent. Interestingly, the risk signatures were between three to five SNPs in combination. So even if GWAS had somehow been able to find this, they wouldn't have found this particular set of signal because it's looking for single snips. The p-values range from 10 to the minus 10 to 10 to the minus 72. And the odds ratio of the average odds ratios of the combinations is about 3.7, which is a pretty healthy odds ratio.

Interestingly enough, about 95 percent of the SNPs that we identified were in non-coding regions. That's to say they wouldn't have shown up in a whole exome data. You need either genotyping coverage or you need whole genome sequence to find them. When we look at the genes involved, we see a range of genes turning up and they're associated with many of the cellular processes that you would expect to see. So some autoimmune components, some metabolic components, some circadian rhythm issues, some neurotransmitter and mood disorder

kind of correlated processes being shown up. This is the genetic architecture that we identified. So here you can see a particular subgroup of patients. In this particular case, it represents about 27 percent of the cases and they share risk signatures associated with these 29 SNPs. The p-value is very low, which is good. The odds ratio is pretty high, which is also good. And when we look at the genes that these SNPs map to, they're primarily associated with mitochondrial respiration or the replenishment of cellular energy stores post-exercise, primarily around a gene called AKAP1.

So the obvious hypothesis is that the patients in this subgroup ought to be those who are significantly enriched for a fatigue phenotype. And in fact, that is what we found when we looked at the phenotypes of those patients, they exhibited mainly the fatigue phenotype. As opposed to the group next to them in blue, this group, the defects were in a series of genes associated with neurotransmitter, precursor metabolism and transport. And these patients you would expect and in fact did present with mild cognitive impairment, aberrant stress responses, sleep disturbance, and a higher level of phenylalanine, which is a neurotransmitter precursor. We also saw overlap in this cohort with the verbal interview, the more of the CFS cohort. And then, just to highlight another group here, we found the CLOCK gene. CLOCK and the insulin receptor actually show up in our long COVID study, as I will show you in a second. But this maps very clearly to the management of circadian rhythms within the cells and is a very strong signal within this. So we're able not only to identify signal but also to map that to individual patients and as I said earlier that's really important when you come to thinking about how to evaluate what therapies might work for that individual patient.

So here you can see that those risk signatures, in total almost 200 SNPs in 84 different combinations. What we were able then to do was to go back and count how many times we saw those risk signatures in cases versus healthy controls. And as you'll see here, the distributions of these are very different from one another. And in fact, if you take a standard analysis of the top 25th percentile of data, you'll see the odds ratio of predicting a case accurately from a control is about nine. And just to put that in context, BRCA, which is a well-known monogenic risk factor. The odds ratio is between 5 and 7.5 at predicting breast cancer, depending on the population. So this gets really interesting. I should say for all of the geneticists in the audience, this does replicate in disjoint populations. I'll show you that evidence in a second. But what it opens up is the potential that we can put those 200 SNPs on a low density genotyping array. This is a test that can be delivered in bulk for probably about \$50 per patient, but it can be run off a saliva sample. So as Chris said, even a spit and post kind of analysis. This is capturing not just the fact that this patient is likely to have ME/CFS, but also the mechanistic defects that are likely underpinning their particular form of the disease.

So this opens up some really interesting opportunities if it is proven and it needs to be replicated. So here we can see the original pain questionnaire data. When we look at running the same type

of analysis in the verbal interview, we get a slightly lower odds ratio, but still extremely good. As we go down, we looked at post-viral fatigue syndrome as well, similar kind of results. And then we even tried seeding. In other words, taking the hits from one of these populations and using it to direct the search and we got a much higher odds ratio. All of those lead us to a superset of predictive signatures. So what we did for the three populations I've just described, we added all of those risk signatures together and then ran them again against each of the patient populations. And we still see, even in this replicated analysis, we are getting very healthy odds ratios, certainly well beyond what you would expect given the lack of genetic evidence for associations in the disease.

This opens up the possibility, not just of analyzing this for ME/CFS, but also doing a differential triage against multiple other diseases which share similar symptoms, and so perhaps being able to differentiate between those. We have seen overlap between ME/CFS, post-viral syndrome, fibromyalgia, long COVID, multiple sclerosis. Those are described in the papers. Most recently, and this should be out any day, we've also extended this analysis into long COVID. This is really carrying on analysis that we did in the very, very early days. So back in May 2020, we published this analysis. We found genes underpinning most of the major symptoms associated with severe COVID. What we have also been able to do is to turn that repurposing opportunity that Chris described into an engine. So basically looking for any given drug where you know its target, analyzing whether that target is implicated in disease, any of the 50 diseases that we've studied, and then designing a new clinical trial with a patient stratification biomarker to support its analysis.

Out of this, we found Dutasteride. This was published in May 2020, which is a medicine used for benign prostatic hypertrophy. So a very different indication, but it's read out in double-blind randomized clinical trials and been shown to reduce the severity of severe COVID by almost half and the need for ICU by about 40 percent. So a very significant impact on disease for a particular high testosterone patient population. So we've extended that into long COVID working with Sano. I'm not going to go through the details of these, but just to say, we analyzed two data sets, one which is very severe long COVID, one of which is particularly associated with fatigue. I've put the slides in here primarily so that you can refer to them. But the bottom line is we have found 73 genes associated. We found five that, sorry, we found 53 that are associated with both the fatigue and severe cohorts in long COVID. These are the top five most significant. We've also identified differences between those two disease presentations. Some of the most notable genes, I'd point you at Toll-like receptor 4, which we think is a very strong repurposing opportunity, certainly for long COVID, and that may actually translate into ME/CFS, but there are a variety of others described in the paper.

When we look at the overlap of those long COVID results with ME/CFS, we get 13 SNPs that are identified in the long COVID cohorts, and they map to these genes. So as I said, CLOCK

and insulin receptor, very strong signals that show up in these studies, which opens up a lot of potential opportunities. All of this comes with exactly the caveats that Chris described. These are small data sets, largely self-reported diagnoses. There may be a high rate of misdiagnosis. There's limited diversity. It's UK Biobank, so there is an elderly population recruited here, which is mismatched to the presentation of the disease. And we had very limited information on disease onset relapses and recovery through the process.

What I'm going to say in terms of the most important research priorities, number one has got to be share the data. We know that there are some very large cohorts of particularly of long COVID information out there that have been sat on for years and not made available to the community to analyze. I would call on everybody who has that data to make them available. Innovative analytic tools will find different results which we can then go and analyze and replicate and validate. But what else? The community needs better diagnostics. We need mechanism-based therapy selection, particularly if that comes from drug repurposing. We need to understand novel therapeutic approaches, targeting key disease mechanisms and organs and cell types. And we also need to be able to identify any disease biology, sorry, any biology which is actively protecting against disease because I believe that there are several mechanisms in the background which are preventing some of these patients from actually getting these diseases. So I'll stop there. There was deliberately more slides on there so that you at least have information to refer to and I would encourage you if you're interested to go back and read the papers that we put in the beginning.

**Vicky Whittemore:** Thank you so much, Steve. That was absolutely fabulous. So to pick up on almost your very last point, in your long COVID studies, are you looking at individuals who had acute COVID and did not develop long COVID to look at for protective genes?

**Stephen Gardner:** Yeah, absolutely. So there's actually a number of different things that we can do. That is clearly one approach to finding protective effects. The other that we are very keen to pursue is to find a population that has all of the, that has a very high risk score for all of the factors that should drive them to ME/CFS, and then to identify those who haven't had the disease. Now, in a pathogen-mediated disease, you know, some of them are not going to be exposed to the triggers, but there should be enough signal to enable us to identify the similarities within those patients. And that may point us at actively protective biology. And I think that is, you know, potentially those are, that is, it's both a very powerful and also a very low impact way of affecting the biology of some of these complex diseases.

**Vicky Whittemore:** So here's a question from Varuna. Since most of the significant SNPs are non-coding variants, do you know their impact on downstream gene expression? Has the genetic data been integrated with transcriptomic information?

**Stephen Gardner:** Yeah, we're trying to. The challenge with transcriptomic information is, of course, we don't know the cell types that are involved. And so getting single cell data is really quite complex. What we can do, so Chris stressed genetic associations. One of the things that we can do is tease apart the contributions of each of the components of our risk signatures. And it allows us to evaluate the causality of those features. So we can look at the consistency of the directionality of effect that they're causing. Are they causing protective or risk increase? And the degree to which we expect them to impact. So what it allows us to do is to look through the pathways and identify genes that may be up or downstream of individual targets and, with some certainty, decide that they should also be included in any of these, you know, the repurposing or novel target discovery approaches.

**Vicky Whitemore:** Yeah, so this question is about trying to understand all of this genetic information regarding different cellular processes, metabolic, autoimmune, sleep disturbances. So how do you know when a gene's responsible for ME/CFS versus another condition? So might some genes be responsible for or expressed in different pathologies? And I think maybe we showed that in some of the overlap with fibromyalgia and long COVID, but if you'd address that.

**Stephen Gardner:** Yeah, we see this routinely. I think there are many genes which are drivers of pathophysiology, which gets modulated by another gene or an external factor, or perhaps presents itself in one tissue versus another tissue. And they end up being slightly different diagnoses. So when, by the time you're looking -- you know, we diagnose in these diseases by observation in the clinic. You know, if you have the same process going wrong in the lung versus in the vasculature versus in neural tissue, you're going to get different phenotypes coming out, but there may be a common genetic driver behind a number of those issues. So yes, we explicitly look for commonalities across, well, all 50 different diseases that we've done. In the long COVID paper, we compared all of the genes that we found against 170 conditions. So a very, very large scale cross-disease analysis.

**Vicky Whitemore:** Yeah, thank you, Steve. So, you know, a message I think that's coming through loud and clear is finding a way to share all this data across the studies that have been done or are underway. And so the last question I'll ask you here is what further evidence would you need to be able to put these SNPs on a panel to be able to report back to a given individual their high polygenic burden or risk?

**Stephen Gardner:** I think what we primarily need is replication. So we are, we're collaborating with Chris. We have a grant from Innovate UK to use the DecodeME as a replication cohort, also probably a discovery cohort as well, but certainly a replication cohort. Again, you know, we'll still be in a Caucasian population, so we would like to have a much broader diversity of ancestries to work from. But that said, you know, as far as evidence goes, the amount of evidence that we've got so far in the crossover between multiple disease populations is actually

pretty compelling. We've also done the same in ALS, we've done it in endometriosis, diabetes complications and a range of other diseases. So we have a degree of confidence in that. But it's one of the other reasons for collaborating with Metrodora because they have patient populations in which we could test such a diagnostic.

**Vicky Whittemore:** So I'll ask one last question before we move on. Have you looked at your data and separated it by male and female? Because we know coming from a lot of the recent research coming out of the ME/CFS centers in the U.S. and other research that there are significant differences between what people are seeing in proteomics and other metabolomic studies. Have you looked at that to see if, with ME/CFS, the population separated out even different separately?

**Stephen Gardner:** Yeah, we did not have enough data to run that study and find results that we were confident of in both categories. So we're very hopeful that we will be able to do that with the DecodeME data set. I think it's one of the major questions we'd like to answer.

**Vicky Whittemore:** Excellent. All right, thanks so much Steve. Excellent presentation and we're answering questions and I'll go back to you Oved.

**Oved Amitay:**

Thank you, Steve. This was fascinating. And Steve already started to build some of the connection between findings in long COVID and ME/CFS. Obviously, for many of us, this was something that we're concerned about at the very beginning of the COVID-19 pandemic, that there will be some people who will not recover from the acute disease and will continue to exhibit something that looks like ME/CFS and, of course, that's what we've seen. And the COVID-19 pandemic not only was a global tragedy, but in some cases really led to an international collaboration that we've never seen before.

So we really have the privilege of having Dr. Hanna Ollila and her colleague Vilma Lammi and Anniina Tervi today are from the University of Helsinki in Finland and they were really at the front of getting this kind of cross-border collaboration, getting data from different countries, to understand long COVID and ME/CFS. So it really begins to speak to the point of how can we leverage those discrete cohorts in data that we have in different places. So Hanna, we're really looking forward to your presentation as learning some of the findings and also to get a bit of insights from you on how do you actually get those different data sets to work together. So Hanna, please.

**Hanna Ollila:** Thank you so much for the invitation and we're super excited to show the results and we try to make also as much as possible like data available, so all of the summary statistics from the analysis that we have run are also available for the research community. So, when COVID-19 hit, there was a global initiative to understand susceptibility and severity and this was really because people had such a different response to the initial infection relatively unpredictable as well. And this was coordinated from the University of Helsinki from FIMM by Andrea Ganna and Mark Daly and it was a massive effort. It was very successful and a bunch of different cohorts very much globally pitched in and did analysis locally and then meta-analyzed and produced these results. And what we are doing under the umbrella of the COVID-19 host genetics initiative is to understand the genetic risk factors for long COVID.

And we really are doing genetics because it provides an overview across the whole genome that was like beautifully described earlier and then that can really help understand underlying biology. So this, the variants that we're currently discovering, do not represent really like individual risk, but they help to understand by underlying biology that contributes then to development of long COVID. And one of the key symptoms that comes up across many of our cohorts is fatigue. And this really connects together with the ME/CFS field as well. And then we have a large variety of other symptoms that come up in different cohorts. And I just want to highlight the diversity of the data that we're putting together. Some of the cohorts are based on electronic health records and some are questions of symptoms after COVID-19 infection or suspected COVID-19 infection. So we do not limit the data sets based on whether they have ICD-10 codes or whether they have questionnaire data.

To understand how we do the analysis, I want to show you briefly that there is the population that is infected with COVID-19 or is suspected to have been infected with COVID-19. So we're comparing the first analysis we're doing is comparing long COVID to those individuals who were infected with COVID-19 but did not develop long-term symptoms. And then, the other comparison, a separate analysis that we're doing, is to compare long COVID cases to everyone else in the whole population, whether they had a COVID-19 infection or not.

And in the Data Freeze 4, that is, where we have results currently publicly available, we have 24 studies from 16 different countries, over 6,000 long COVID positive study participants, and over 40,000 people who had positive COVID-19 infection, and then 1 million people from the population controls. We do data freezes every four months and currently we are on Data Freeze 6. And you can see that really the effort is growing and I invite everyone to collaborate with us so that we've increased the sample size so that there is now 19 countries, 14,000 long COVID patients, 200,000 COVID-19 positive controls, and then 1.5 million population controls. Most of the results are pre-printed in Medrxiv, but I'm also showing unpublished results and Anniina will continue on our ME/CFS work in a bit.

So, one goal that we really wanted to have in this effort is to maximize different ethnic groups and diversity across populations. So we have majority of samples are still from the European populations, but we have also samples from Asian and then other ethnic groups. And this will be quite important when I show the results in a bit. And Chris already nicely outlined -- so essentially we're looking the whole genome, common variation, comparing those with the long COVID cases to either COVID or then to population controls. And currently, we have one genome-wide significant hit. This is a locus at FOXP4, and it's the strongest signal we're currently seeing after test verified SARS-CoV-2 infection versus population control. And this is partially because we have, with the population control, we have the largest power to also detect these associations.

And this is a new part, so we have now replicated the signal also in independent cohorts from six different cohorts. And then, there are two points to take home from this slide. So first of all, the risk is pretty systematic in all of these cohorts and then when you meta-analyze the replication you have a significant replication in the cohorts. And the other point is that then we have them systematically all on the positive side from each of the individual contributing studies in the replication.

And here is the reason why I think it's really important to have diverse ethnic groups, is that in European populations, the risk variance is actually relatively rare, so we have one to 2 percent allele frequency, whereas especially in the Asian cohorts, also in the Finnish cohort, we have higher allele frequency. And this means that we have more power to detect genetic associations, simply because we have more people with the risk variant in these populations.

And so, what is FOXP4? So FOXP4 is a transcription factor. This means that it activates other genes and regulates their expression levels in different tissues. There are three major tissues where FOXP4 expression has been implicated. There is the lung, there is the hypothalamus, and then there are immune cells. So, first, individuals with the risk allele have higher expression of FOXP4 in the lung. And if we compare how the genetic variant affects FOXP4 expression in the lung, we can see that the signals co-localize with each other. So it's probably the same variants that regulate FOXP4 expression in the lung that regulate also that affect long COVID susceptibility. If we go more into detail, like what type of cell types participate in this effect, we have used previously published single cell sequencing data from healthy individuals across different lung cell types and we can see that there are these alveolar cells, type two, that have the highest expression, but there are also immune cells, like granulocytes, that express FOXP4 in the lung. So this might point us to the function of the of the gene and the variant how it contributes to long-term consequences after COVID-19.

So, FOXP4, signal itself, it has been implicated before also in COVID-19 research. So FOXP4, the same variants are also risk factors for severe COVID-19 and COVID-19 hospitalization.



And it's not only the COVID-19 phenotypes that come up, if you look kind of like horizontally across many different phenotypes, what kind of other traits associated with this variant. But, we're looking here in Japan biobank, just because the allele frequency is higher in the Asian population. And we're seeing that in addition to severe COVID, we have these other lung phenotypes that associate with FOXP4 variants that predispose to long COVID and particularly lung cancer is one of those phenotypes that comes up here.

And then, there are earlier studies that have suggested that one of the risk factors for long COVID might be COVID-19 severity. It's not the only risk factor, but probably like one of the risk factors. So we also tested using Mendelian randomization, which is a genetic technique to test causality between two traits. So we tested, took all the variants that contributed to COVID-19 hospitalization and then predicted whether those variants or the COVID-19 hospitalization is a risk factor for long COVID. And we discovered that COVID-19 hospitalization indeed was a risk factor for long COVID.

And someone might say that, okay, so we're just picking up COVID-19 severity variants in the scan, but that is actually not the case. So, FOXP4, if we take all of these severity variants and look their effect sizes in long COVID, FOXP4 stands out as an outlier, really having a different level of association with long COVID than any of the other severity variants. So, with that, I really want to acknowledge everyone who participated in this effort. This is a massive collaboration across many different cohorts, especially Samuel Jones, Vilma Lammi, and Tomoko Nakanishi have contributed a lot as lead analysts for the project, which many, many other cohorts running the GWAS separately first at their cohorts and then providing it for them, for the meta-analysis. So, I want to actually stop here and then let Anniina show our results from ME/CFS.

**Anniina Tervi:** Yes, hello. So my name is Anniina Tervi, I'm a doctoral researcher at Hanna Ollila's group and I'm going to briefly talk about genetic correlations with ME/CFS, which relates to basically getting the summary statistics from these genome-wide associates and studies of what can we do with them and actually what was mentioned before by corresponding compare with other diseases. So, as we know, there are many known comorbidities with ME/CFS. I've listed here a few, for example, Raynaud's disease, POTS, HEDS and fibromyalgia. But then one question we can ask from this is, do these comorbidities show genetic correlation with ME/CFS? So we wanted to look into this question by using these summary level statistics from different European ancestry population cohorts, so mainly FinnGen from here, Finland, and then using the UK Biobank as well to conduct this analyses. And what these analyses basically are is a statistical method, what we can use to measure the extent to which two complex traits possibly share a genetic similarity.

So I have here a genetic correlation plot. I'll walk you through it. So in the horizontal lines, I have three different definitions of ME/CFS, so diagnosis-based, so ICD, and in this case, mainly the post-viral fatigue, and then also self-reported, which is from the UK Biobank, meaning the verbal interview that was mentioned also in the last talk. And then in the vertical lines here, we have different diseases or syndromes, which we tested than with these different definitions of ME/CFS. Now the coloring of the plot, so the deeper the blue we have, we have more sort of a positive correlation, and so the deeper red we have, we then have a negative correlation. And then here, I've highlighted the ones with the stars, so the stars indicate the statistical significance, A.K.A, have the actual or most likely have the correlation with ME/CFS and as you can see I've highlighted for example Raynaud's disease, hEDS, insomnia, asthma and Sjogren disease for example.

Now these preliminary results that we have indicate that there might be a shared genetic component with several diseases on traits with ME/CFS, for example now the Raynaud's syndrome, hEDS, or asthma. And this is just to show that these results are just a great example how this kind of population level data can be used to study not only the underlying biological mechanisms of ME/CFS but also the possible shared mechanisms with its comorbidities. And from here I want to sort of wrap up both of our talks and have some recommended areas of research focus. So these align nicely with the previous talks somewhat. So this allows us a hypothesis-free screen for variants associated to either COVID or ME/CFS or both to possibly discover predisposing or then protecting genetic factors. Then, not only we can look at these variants and possibly, if we're lucky, find some variants, but then go further from there to conduct function analysis, whether it is with the data level that we have, or then go to next level with cell models and possibly also animal models to study the mechanisms and disease pathways. But then also given how the heterogeneity between these diseases to possibly finding these patients subpopulations and are there different mechanisms between this subpopulation and then using that information to have optimized treatments for these patients. So with that, I want to thank you for your attention and we are happy to take questions.

**Vicky Whittemore:** Thank you both very much for an excellent presentation. So I don't see any questions that have come in from the Q&A, but I'll ask if any of the panelists would want to raise their hand and ask questions from the speakers. Oved.

**Oved Amitay:** So, Hanna, thank you very much for your presentation. It was really fascinating. And I guess my first question really has to do more with how you did that, to really be able to get so many cohorts from different countries, particularly with genetic data, which is normally the most stringent in terms of sharing with external investigators. Can you tell us a bit more about how did you actually do that? What were the barriers and really how you were able to overcome that? And hopefully if we did that for COVID-19, presumably we could do this for other diseases. So what have you learned from that?

**Hanna Ollila:** Yeah, absolutely. I think, so my background is in genetic epidemiology of post-infectious diseases and particularly sleep disorders. And what we learned earlier from those studies is that it's very rare for like one group to have sufficient sample size to do the analysis alone. So instead, also the psychiatric genetics field and the sleep disorders have started to share data with each other to be able to grow the sample size and have the numbers that are needed for genetic discovery. The way we do that is that we don't share any of the genetic data itself. So each group is analyzing their data in-house. So usually we have a lead analyst and a senior person from each cohort that performs the analysis, and then sends the genome-wide summary statistics to the group that is coordinating the study. In that way, there is no personal information or genetic information that is leaving the initial analysis team. There is only p-values and standard deviations and effect estimates for the DNA-wise stats.

So that is the easiest way to do that and at the end of the day we anyway share those data as part of the manuscript so that has made it feasible to do it and practically barriers and challenges of that is that you need to have a really good analysis plan how everyone is analyzing the data so that everyone does exactly the same analysis and then you need to have a designated person who's willing to troubleshoot with the teams. Because even though you share the same code, there is sometimes these mysteries like that programs works tiny bit different and you need to troubleshoot something. So spending the time and figuring out like what is going wrong in analysis and is there alternative way of doing that? So that is really, I would say the most important and practical aspect of the part. But once you have the results run by each team, then putting those together, that is relatively, then straightforward and is still a massive amount of work and all kind of like things can practically come up as with any data analysis challenge. So those QC steps and being able to coordinate with the local teams, I think that's one of them and having a good analysis plan.

**Oved Amitay:** Are there any challenges with having different areas or different, where, you know, where the sequencing itself, perhaps is done on different platforms? Was that a challenge from a technical perspective?

**Hanna Ollila:** A little bit, yes, but that hasn't been like the main thing that has come up and a lot of those things have been already solved as part of like, almost like a standard QC imputation. So those usually come up before we go into the actual analysis testing. So if the groups are working with genetic data already, then they've already kind of like troubleshooted the practical analytical part that comes from the genotyping or the array side.

**Vicky Whittemore:** So if I can comment, so there's an epilepsy group that was supported by NIH, by NINDS, to do whole exome sequencing of 4,000 individuals with epilepsy that very quickly grew organically into an international collaboration and is now that was called Epi4k it's

now was Epi25k and is really being coordinated through the Broad Institute to do sequencing of individuals with epilepsy from around the world um I think at last count they maybe are now at something like 35,000 individuals that have been sequenced. So it's possible and I think the genetics community is, I think, really quite eager to come together in these kinds of huge consortium collaborations. So I think that's really, really excellent. Chris, I see you turned your camera on. Did you have a comment?

**Chris Ponting:** Yeah, I just wanted to know whether you're fielding questions, as we are, about rare variants and what your feelings are about trying to have a rare variant meta-analysis community?

**Hanna Ollila:** Absolutely. So we haven't worked too much, like a little bit with rare variants. I haven't seen like a meta-analysis of rare variants put together yet. I think that's, I would love to continue that discussion like maybe also later and think about how to do that.

**Oved Amitay:** Thank you for bringing this up. I was actually, at the very least, was going to bring it up later in our discussion session. I think, you know, this has come up as a key point of deficiency for our community. We don't really have a place where we curate all those rare variants. I'm personally aware of, you know, people shared with me. You know, they're all genome sequencing data, so I'm aware of rare variants that have come up. All that I've seen all made sense. There weren't pathways that you would expect and Steve alluded to some of that before. So I think at the end of this webinar today, maybe we will have a volunteer or Vicky, maybe that's a call for NIH to designate someone and enable them with some financial support to be the curators of those rare variants. They're not going to get published, so the common practice of how we collect this information is just not going to happen, because right now people do that in sort of in a personal setting. So we're not going to be able to collect that with the usual means, and we need to have a concerted effort to do that. And actually to just one more comment, the All of Us study that we talked about before does report rare variants to the individual, but again, it's all hinged on the individual and what we do with that. So if we can get a place to a common place where we all share this information, I think that's going to be a huge step forward for our community.

**Vicky Whittemore:** And I absolutely agree with that. And I think that's the goal of this webinar and of this actual, of this whole process is to really think strategically about how to move this research forward.

So at this point, I'd like to invite all of the speakers from the session to turn your cameras on for a more general discussion. So I think one, just one question that did come up that I can address to all of you is, how do you account for individuals when they're self-reporting a diagnosis, or in other words, lack of an actual clinically diagnosed chronic illness? So how does that factor into your genetic analyses?

**Christ Ponting:** Okay, I'll start. Steve, you're putting your hand up. So we don't ask people to self-diagnose. We ask them whether they have a diagnosis of ME or ME/CFS. And what we're then doing, obviously, is trusting people to tell the truth. And I think there's ample evidence that people do tell the truth. And evidence from UK Biobank in concordance between various lines of evidence that they have, GP records, hospital records, et cetera. And I also want to say that there is genetic evidence that often people's recollections of their diagnosis are even better than hospital records are, because those genetic signals are stronger for self-reported observations than are sometimes hospital records. So I know and I hear you when you say this, but I also think that it isn't as great a problem as some might think.

**Vicky Whittemore:** Steve, did you want to comment?

**Stephen Gardner:** Yeah, thank you. I have a couple of observations. I mean, the first one is in the long COVID study. We did use a lot of survey data to validate that they had had persistent changes in symptomology and we separated that into a severe cohort and a fatigue-dominant cohort based around those symptoms. There was a, you know, there were a subset of patients who ended up in that study who proclaimed that they actually hadn't had COVID and yet they ended up in a long COVID study. And there was a little bit more bouncing around with some of the data than we were comfortable with. But the second piece is that quite often we deal with complex diseases with multiple etiologies and those are very easy to misdiagnose. Or to confuse for one another. And it is one of the reasons why we choose to do the analysis the way that we do associating signal with specific patient populations, because you can then go back and see whether the phenotypes of those patients that you have grouped together actually match a mechanism of action hypothesis and whether they represent a specific community in and of themselves. And I think in the analysis that I presented, we saw differences between fatigue-related symptomology and those symptoms that would be related more to classical ME-type symptoms. And I think in the fullness of time, as we get more data, we may be able to clarify some of these diagnoses a little bit better.

**Hanna Ollila:** I think it's really important to understand like how clinicians use the diagnostic codes, especially the new diagnostic codes for long COVID. And partially also like, because of that reason, the way that we have done, currently the ME/CFS work is to use three different phenotypes. Use clinician diagnosed phenotype, self-reported, and then self-reported plus the clinician diagnosed phenotype. And there are great examples from other post-infectious or triggered diseases. And for example, in narcolepsy, the diagnostic delay is often 10 years or can be extremely long. And in that case, you might gain benefit from understanding the symptomatology and what the patients and people are reporting in the questionnaires, in addition to the diagnostic codes that you can retrieve from the datasets.

**Vicky Whittemore:** Hayla, you had a comment or a question? You're muted.

**Hayla Sluss:** That would help, right? Thank you, those were lovely presentations, great work. I think this all reflects, especially for the United States, that there is a difficulty in getting a diagnosis. So what I'd love to see is that along with all the work to do the genetics there is also an effort for like physician and provider education, because there's some people that have it that like I've explained to them, no, I think you have ME, you know, I think, right? I'm not a physician. So I just want to, I wanted to just mention that. So thank you.

**Vicky Whittemore:** So there were several questions about, throughout the talks and I'll come back to it, about mitochondrial involvement. And I know Steve, you presented some SNPs with, I think it was you, mitochondrial respiration.

**Stephen Gardner:** Yeah, it was.

**Vicky Whittemore:** So the connection between those kinds of things, other known mitochondrial diseases, or what you're seeing, maybe just summarize again, what you're seeing in that realm.

**Stephen Gardner:** Yeah, so we were seeing fairly strong signals associated with mitochondrial dysfunction. So a lot of that is AMPK-mediated kind of processes through genes like AKAP1, for example. I think what's interesting about that from our side is that we actually see many of those same signals occurring in other disorders which are not traditionally, you know, post viral or post infection syndromes. We see them associated with neurodegenerative processes. We see a lot in ALS and Parkinson's and Alzheimer's as well. So I think that, you know, back to that question earlier, are the other common bits of pathophysiology, which in general may end up being stressors on a cellular system, which exacerbate other areas as well as being primary drivers of specific symptoms? I think the answer is probably yes. Our initial evidence would definitely say so. I think there's a lot of value in exploring that sort of mitochondrial access for many of these disorders.

**Vicky Whittemore:** So then I'll turn to you and see if there are key discussion points you'd like to raise with the group or also, again, ask any of the panelists to raise their hands and ask questions or comments. I see Maureen turned her camera on. Maureen, would you like to?

**Maureen Hanson:** Yeah, I'd like to ask whether the mitochondrial genome was analyzed in these many studies or whether it was just the nuclear genes including mitochondrial protein.

**Chris Ponting:** There are mitochondrial variants that are being surveyed in the genotyping that we're doing, but without sequencing, we're not able to find it or associate to any other variants.

**Stephen Gardner:** Yeah, our studies were all autosomal.

**Maureen Hanson:** Well, I see that maybe Hanna has a comment there.

**Hanna Ollila:** I wanted to take a comment for the Q&A, so I'll do that after then.

**Maureen Hanson:** Oh, okay. I'm just thinking that it would be useful because if you're doing whole genome sequencing, you should be able to get the mitochondrial sequences too.

**Stephen Gardner:** Yeah, I would totally agree with that. And we certainly see, we have other disease data sets as I was alluding to, where we see important signal in the mitochondrial DNA.

**Vicky Whittemore:** I see, Fereshteh, you have your hand up. Is it related to this discussion or a new question?

**Fereshteh Jahaniani:** Yes, actually regarding mitochondria because I highly agree with these questions and also studying mitochondria. One issue we have with mitochondria is the hemoplasma in mitochondrial mutation and knowing that some particular tissues might have more of these mutations being accumulated in one tissue, for example, muscle, rather than in your cells. I'm just questioning how we really can get to the bottom of this. It's hard to have access muscle tissues, but I know some people have muscle tissues. So I think that could really be a nice way to start somewhere.

**Vicky Whittemore:** Anyone have a comment on that or?

**Chris Posting:** No, I mean, it's a really excellent study that needs to be done. And for that there needs to be a cohort and I just would propose again that you develop a cohort in the US that you can draw upon in specific ways such as this for particular purposes. And if people wanted to draw upon the DNA captured as a resource also, then they could make use of that. So this is a

generic answer to a very specific question, but I see a whole gamut of different opportunities that could arise from developing a cohort for genetics and beyond.

**Stephen Gardner:** Yeah, if I may just add to that if I can draw analogies to the ALS community. The ALS community has been very forward looking in collecting data above and beyond the genetic information of patients. Even before those data could be effectively analyzed and some of those code, so they're collecting things like telomere length and epigenetics and a variety of other factors. And it is now paying absolute dividends in being able to correlate risk factors associated with that rather horrible disease.

**Vicky Whittemore:** Thanks. So back to you, Hanna. Did you have a comment about something in the Q&A?

**Hanna Ollila:** I wanted to comment on the FOXP4 associations with the clinical subtypes and the symptomatology. So one approach we're taking is to understand like if FOXP4 association is a general association across all the different fields types, or if it's like related to one specific subtype.

**Vicky Whittemore:** Any other questions or comments?

**Oved Amitay:** Just maybe a quick comment on the mitochondrial findings. And I think many of you are aware of the recent publication of a rare variant that was described by Dr. Wang. And I believe he will present in one of our upcoming webinars. And these findings are for rare variant that really has a postulated role in the mitochondrial dysfunction. And clearly, yeah, can explain at least some of the symptoms. So I think that we need to work both ways in terms of looking at individuals at the exhibit graph variants that could lead to an explanation of a pathway. And of course, the other way around, which was really our discussion today. So we need to make sure we do it both ways.

**Vicky Whittemore:** Well, Steve, I see you have your hand up.

**Stephen Gardner:** If I may just comment on that, I mean, I think one of the potential utilities of some of the analysis that we've done is identifying a cohort for whom mitochondrial dysfunction is most likely an underpinning factor of their disease and then enriching the cohort in that respect because it will reduce the number of samples that are required in order to perform effective analyses and it may yield, you know, a more tractable patient population for further analysis. And again, all of the data that we have generated, all of the risk signatures and everything else is in the papers. But if anybody does want to engage around those, we put all of that stuff in the public domain and we'd be very happy to help in using that in, you know, selecting those patients.



**Vicky Whittemore:** Fereshteh, I see you have your hand up again.

**Fereshteh Jahaniani:** Yes, I actually love the suggestion Steve just mentioned regarding mitochondria, because sometimes you see that within a family study, it seems that the disease is coming from the mother side of the family, which is mostly indicating that there is a possibility of mitochondrial elements into this. So enriching for that would be very helpful. I, if possible, I would love to ask at this comment, someone was talking about the epigenetic in the question this morning. I would love to say that knowing how important is that integrating all the -omics data together to really create a bigger picture for those who have genetic material saved in a minus 80, I would love to see that further with the budget to do combining the genetic study with the epigenetic study and making sure that this is really genetic driving the phenotype or epigenetic driving and changing the genetic around it. So that will be an area that we definitely need more funding and more studies around.

**Vicky Whittemore:** Yeah, yeah, excellent comment. So thank you, everyone. That was an excellent session. Really, thank you all very much. And at this point, we're going to take a break and we'll come back at, I don't have the time, in 25 minutes, I believe, correct?

**Oved Amitay:** 20 minutes. We're going to start at 1:25 Eastern Time or 10:25 on the Pacific Coast and in Europe you are in different time zones. So 20 minutes from now.

**Vicky Whittemore:** Great, thank you.

**Vicky Whittemore:** Yes, I'd like to welcome you all back from the break. And then once again thank all the speakers from the first session. That was absolutely fantastic. And I'll actually turn it over to you to introduce the session two.

**Oved Amitay:** Thanks. Thank you. So, in the first part we looked at how investigating large populations in terms of genetics can really help us to understand ME/CFS and find therapeutic options. Now we'll shift to the other side of the spectrum and really look at what can we learn from looking at studying families and more enriched cohorts. And then we'll talk about the epigenetic aspects of that. But for the first part, I'd like to invite Dr. Fereshteh Jahaniani and her colleague, Dr. Varuna Chander from Stanford University to share the research using case control and family studies. So Fereshteh, please.

**Fereshteh Jahaniani:** Good morning or good afternoon, everyone. I hope everybody's enjoying this wonderful talk as I'm enjoying it and learning so much. I would like to thank the organizer for the opportunity for letting us to share some of the work you have been doing on ME/CFS pathophysiology. Our talk today is going to be devoted to ME/CFS genetic predisposition. By now, we all know that ME/CFS is developed by the combination of genetic and environmental risk factors. We also know that ME/CFS can run in families and affects multiple generations within the same families. We also know that there have been multiple attempts, at least a few attempts, to identify genetic risk factors associated with ME/CFS through multiple case control analyses at different levels, including running a GWAS study on UK Biobank. While the results are very interesting and interesting candidates have been found, unfortunately the data is very inconsistent across the studies. There is no single gene or locus that has been replicated in two studies. There could be multiple reasons behind these inconsistencies in research findings, such as variation in a patient's population, varied definition of case and controls across the studies, sample size, different methods that has been used, and as well as the complex polygenic nature of the disease itself. As you heard in the morning, we all agree that ME/CFS might not stem from one single gene mutation, but instead it emerged from the combination of the complex interplay of multiple genetic and environmental risk factors. How to understand this intricate web of interaction between all these diverse genetic and environmental factors requires a combinatorial approach.

To this end, I personally believe that employing case control studies with family-based studies along with AI tools can provide us the necessary power we need to better understand ME/CFS underlying causes. To this end, we have collected a substantial cohort of samples through collaboration with multiple labs. One portion of our cohort came from Maureen Hansen's lab at Cornell, the other one from UK Biobank, and in collaboration with Michael Schneider, Ron Davis, and Michael Zinnis at Stanford, I could collect samples from over 150 families with ME/CFS. Unfortunately, due to funding issues and limitations, we could only sequence about 364 of these samples, including 22 families, nine twin pairs. Our samples are well balanced, fairly balanced in terms of male and female, and case and controls. We also included samples from 20 very severely ill ME/CFS patients who were bed-bound and included them into the study. I had to travel to their house and collect samples in their own place.

For today's talk, I'll delve into the genetic and multiple mix analysis I have done on one of this family and my colleague Varuna is going to talk about the GWAS analysis she has done in detail. Let's begin with the family pedigree for this particular family. Our patient is a male Caucasian in his late 30s. He has ME/CFS. He also presents hypermobility spectrum disorder or HSD phenotype. His sister has EDS type 3 or Ehlers-Danlos syndrome, and his father also presents hypermobility spectrum disorder. As we heard from talks earlier today, research also shows that over 70 percent of ME/CFS patients have comorbidities with connective tissue disorder, including EDS and hypermobility spectrum disorder. In our cohort, I also noticed that

over 50 percent of our patients, they have either HEDS, hypermobility type, or HSD. On the right, what you see is patients' long time of ME/CFS. Patients' symptoms began in 2004. And integrating clinical data with health data revealed that a patient had two episodes of leukocytosis, which is elevated of white blood cells count, prior to the onset of his condition. The first episode was caused due to a mononucleosis in junior high school and due to a Epstein-Barr virus infection. Patient symptoms gradually worsened from mild to moderate and to severe and from severe to very severe and finally to extremely severe in 2014, which rendered patient fully bed-bound and dependent on caregiver for all aspects of his life. After a decade looking for an answer, patient finally got diagnosed with ME/CFS in 2011 or 2012.

So, for this family, we could generate whole genome sequencing for the patient and also whole exome sequencing for the entire families. And we have used a hydrogen clinical insight platform to analyze the data and the variants. On the left, this figure on the left shows the number of the variants identified across all these five samples. We started over 5 million variants to begin with, which were subsequently reduced to 4 million variants after filtering out and excluding variants with low confidence. For this particular analysis, I'm just only focusing on variants with lower allele frequency in the population, less than 5 percent, and excluded common variants unless these common variants have deleterious effect. Finally, we refined this list to only 28 variants, based on deleteriousness score, considering HMGD, human genome, genes database, and as well as ClinVar and CADD score values.

On the right, you see these 28 variants I was just talking about. Patient's whole exome sequencing and whole genome sequencing have been included on the probing column, and then his father, mother, and sister's whole exome sequencing data have been included in the control column. We observed that some of the variants only have been found in identified in both whole genome sequencing and whole exome sequencing platform. However, some of the variants only have been identified in one flat form and absence in the other one. For example, we noticed that SMPD1's mutation on the variants on the top only have been identified in whole genome sequencing and is absent in whole exome sequencing in the patient. So, there is a possibility, because this variance actually is exonic variant. So we expected to see this variant in our whole exome sequencing data as well. So the fact that this is missing could raise the possibility that it's not a real variant and could be most possibly a platform specific artifact variants or false positive because of the platform specific artifact. We will do central sequencing to validate these findings.

As we just heard this morning, everybody has been repeating this wonderful point that we don't think ME/CFS stems from one single gene mutation, rather is the combined effect of multiple genetic variation at single nucleotide polymorphism or even maybe copy number variation. So, for this, here on the left, you see the list of 30 variants identified in these patients. These are all variants with lower allele frequency in the population, less than 5 percent, and have been

identified both from our whole-exome sequencing and whole-genome sequencing, and identified based on their deleteriousness score, as I mentioned. To better understand what these 30 variants, what pathways these 30 variants might be targeting, we have done a gene set enrichment analysis. The figure on the right shows the top dysregulated pathways that have been affected by this variance. Pathway analysis shows that complements and coagulation cascade as well as ATP binding cassette transporters are among top dysregulated pathways that are being affected by this variance. ABC transporters are ATP dependent pumps that they are important for carrying in or out the reagents or nutrients to the cells and taking them outside of the cells or any other type of substrates such as cations, anions, amino acids, and some specific phospholipids. Having dysregulation in complement and coagulation cascade could actually cause coagulopathy and affects patient's blood coagulations system as well as affecting their endothelial integrity and causing endothelial dysfunction and leaky blood vessels that have been long speculated to be one of the underlying causes in ME/CFS.

Further, genetic analysis also indicated a predisposition, or suggested a possibility for a predisposition to post-virus sequelae, including herpes virus-associated encephalitis in probiotics. A patient inherited a mutation, a variant in PKP2 genes from his father, and also another variance on TLR3 from his mother. PKP2 is a gene associated to cardiac function, and this mutation in these genes is linked to arrhythmogenic right ventricular dysplasia. However, this gene is also important for endothelial barrier function, specifically in the oral cavity and also for fighting against viral infection. TLR3 is a gene that we actually heard about another analog of this gene TLR4 in long COVID. TLR3, or two like receptors three, is part of TLR families that this particular TLR is important for recognizing viral associated molecular pattern and is very important to help host to fight viral infection associated with viral infection. Another very interesting thing about TLR3 is that an agonist of TLR3, ampullogen, has been subjected to clinical trial for treating ME/CFS patients. As mentioned earlier, integrating clinical data with the health data and -omics data also showed us that patients did indeed have two episodes of infection, and one of them was viral infection prior to the onset of the disease, which could be actually part of the disease trigger. Further, studying this clinical data has revealed that patients had elevated antibody against herpes virus 6, HSV-6 virus infection, as well as elevated EBNA-1 IgG against EBV. This data together further supports that genetic data and combining genetic data with clinical data and family history could actually help us to further understand the underlying causes in ME/CFS in individual patients and maybe combining this in many more patients.

To further understand how ME/CFS affects different biological pathways in this patient, we have done plasma cytokine and also PBMC transcriptomics for filing this patient. I apologize for this busy slide, but in summary, on the left, the figure on the left shows plasma cytokines level for over 68 cytokines that you can see their name here, including different classes of interleukins to interferon alpha, gamma, and the rest of the other cytokines. Patient's data is marked in red.

And I hope that you all agree that just following the red line, we can compare to the other lines, which is in orange, representing sisters, and the lines in gray and blue, which represents the parents. We can see that the patient actually has much higher elevated cytokines in plasma for many cytokines compared to his sister and also his elderly parents, which we expect in elderly parents, we see actually a higher level of inflammation while we see it actually a profound inflammation in the patient due to similar phenomena we see in COVID patients we named cytokine storm.

PBMC transcriptomics in patients and comparison to its family members also validates our cytokine score filing data. We can see an increased activation in interferon signalings and many other system associated to immune system. And we also can see that PBMC landscape in this patient mimics to what we see on the right figure, what we see in patients with relapsing and remitting multiple sclerosis, as well as inflammatory demyelinating disease. This data together further proves that ME/CFS is not in the patient's head or is not a functional disorder. It simply says that ME/CFS emerged from multiple, from a significant dysregulation in multiple biological system in patient's body and can point to the diverse and multisystemic nature of this condition.

I hope by now I could convince you that taking combinatorial approach and including and adding the family-based study to case control study can actually help us to better understand what are the underlying causes of ME/CFS at individual level and hopefully across a subset of the patients. It could enable our identification of rare potentially causative genetic variants. It can help us to identify the noble mutations that could be actually linked to ME/CFS development. It can offer potential target for precision medicine as we notice for the TLR3 variants and having a clinical trial actually address to target this particular gene. It can pave paths for personalized treatment approach for ME/CFS, and it can actually enhance our case-controlled GWAS studies to better understand ME/CFS root cause and underlying pathophysiology. With this, I really appreciate your attention. I would like to give the stage to my colleague, Varuna.

**Varuna Chander:** Thank you, Fereshteh, for this nice introduction. Okay, great. Thank you everybody for this opportunity to talk about ME/CFS disease and the kind of work we are doing at Stanford. So, really quick introduction about myself, I'm currently a postdoc at the Snyder Lab at Stanford University and I finished my PhD at Baylor College of Medicine in Mendelian Disease Genetics. I'm really excited to study ME/CFS disease architecture and get right into the details right now.

So, real quick, before I talk about the results that we have generated so far, I just want to give a quick introduction about the genetic architecture of Mendelian and common diseases. And this is a very nice graph from one of the papers, which clearly shows that there are two ends of the spectrum. We have these rare monogenic variants in Mendelian disease. At one end would have

high penetrance and the other end of the spectrum, we have this polygenic model with common variants across multiple genes together giving additive power, which can be identified by GWAS. And so, in my opinion, whatever we know today of the contribution of genetic loci towards common complex disease was due to the foundational knowledge of our understanding of, you know, from Mendelian disease genetics.

Considering that how heterogeneous ME/CFS is, just like how previous speakers had spoken about, and Chris Ponting also had mentioned in his earlier talk, that it's such a complex heterogeneous condition with multiple factors contributing to the disease etiology. So, it is a combination of not just genetic, but also environmental factors that could be driving the pathogenicity of this disease. And for that, we need an ensemble approach, you know, not just looking at one part of the spectrum, but kind of integrating both. Like looking into both the rare variants and the common variants and how they contribute to disease, and kind of bring into different methodologies, like how Fereshteh spoke about the family studies that we're doing here, but integrate those findings with the GWAS, and also conducting thorough systematic case control studies and using computational methods and statistical approaches to investigate the findings which are unbiased.

So, like I said, there is a need for case control studies through sequencing patient codes that are enriched and also very well phenotyped, and keeping that in mind, we have conducted a study here at Stanford where we've gotten samples, so enriched patient codes from collaborators, Maureen Hanson's lab from Cornell, and Fereshteh has collected her samples like she spoke about at Stanford here. And we also got samples, cases, and controls from UK Biobank. So essentially, we have three or ME/CFS disease cohorts, for which the samples have been sequenced, whole genome sequencing, and well phenotyped. And using these, we conducted a genome analysis workflow for GWAS, which is pretty much going through the standard pipeline, nothing new here, but essentially, taking the FASTQ files and generating the alignment files, the BAM files, and performing the variant calling. All of this was done at large scale, using Sentieon pipeline on the AWS platform, and they have all been joint-genotyped. And the downstream approaches include, you know, thorough stringent QC control metrics and filtering and annotating these variants and conducting the common variant analysis using GWAS and the rare variant analysis using burden tests, which I'll show on the next few slides. And the last approach, which we're also looking into right now, is using unbiased nonlinear methods like machine learning approaches to look into the cumulative effect of both the rare variants and the common variants like I spoke about. And each of these steps have to be done rigorously and systematically. Make sure that the sample level QC is conducted and filtering for the variance quality and to make sure that we have the final confident, high confidence variance set to conduct these systematic analysis.

And the implementation of the GWAS QC pipeline is as seen in the slide. I know this is kind of a little busy slide, but just to talk about the high level importance of certain metrics is, we conducted a thorough systematic analysis of the QC at the sample level, like I spoke about, and also the variant, at the variant level. To ensure that the variants are of high quality, we took into account the genotype missingness, deviation from Hardy-Weinberg equilibrium test, and also looked into the heterozygosity rates within the samples and factored all of that into account in our QC. And also, for the GWAS specifically, we retained one sample from each family member to account for kinship reference. And last but not the least, we also did the population stratification to remove any, confounding biases as a result of that. At the end we had 947 samples, which went into the GWAS analysis.

As you can see, as expected, it's such a limited sample size, so we really did not see any significant hits in GWAS, which is not surprising because we spoke about the fact that we need such a large sample size to actually see a significant signal here. So this was not surprising, but I threw this plot here to show that we did this analysis, and GWAS did not give us any significant hits due to the limited sample size. But also, another important point to consider here is that it could also be that there are rare variants playing a role in driving the disease ideology, which takes me to the next slide, which is what we did.

We performed rare variant association analysis for these patient cohorts. Each of them had their cases and controls, and we also have controls within Stanford that are well sequenced and well phenotyped using the IPOP samples, which is a longitudinal study here at Stanford, of supposedly healthy individuals that we are tracking over a considerable long period of time, for which we have high confidence quality sequencing data. So using these cases and controls, I performed the rare variant association analysis where we looked into those rare variants that are less than one person frequency and perform the burden tests. And the burden tests are pretty standard tests in the field. And we looked into the CMC test and the scatter test, and also adjusted for covariates like age, sex, and ethnicity, which is very important, and also corrected for multiple testing burden for these two burden tests.

And the findings that I'm showing here is preliminary results, we just got these results like last week, and we're still looking, into the details of it, but it's very interesting to see that the CMC test gave two genes with, high statistical significance, after correcting for one for any correction, and the p value is here. And these two genes play very important roles, for example, the ATP2C2 encodes for an ATPase that plays an important role in mitochondrial respiration. And the agent has been well studied in inflammatory conditions like Alzheimer's. And the SCATO results also gave a little more significance compared to the CMC wall, which is as expected. Because we are looking at, you know, a study that takes into account, you know, effects of the variance in both, you know, trade increasing and trade decreasing, and also had a little bit more power in terms of the burden test. And the first candidate is very interesting, it's a

multifunctional mitochondrial amino transfer, it's the AGXT2. And other candidates also play significant roles in mitochondrial function, T cell receptor activity, and inflammation. Like I said, we're still looking into these candidates. We got to go in deep into those variants and look much closer and replicate these findings before we can come into any conclusion.

Before I wrap up my presentation, I also want to talk about this approach that we are taking, which is the HEAL method. It's actually a machine learning framework that we have implemented in the past, and it has been published in Cell in 2018, where the HEAL method has identified abnormal aortic aneurysm disease genes, 60 of them, and also identified the underlying biological pathways that have been ablated. The way this method works is that it's constructed using, you know, understanding the mutational burden between cases and controls for rare pathogenic variants in a very nonlinear and unbiased fashion. What I mean by that is very agnostic, so you don't need any, you know, presumption on existing knowledge, neither do you require a large number of samples. And it is an hierarchical estimation from agnostic learning. So the model essentially analyzes the consequences of the mutations, you know, and predicts the pathogenicity and also looks into the disease ideology from the network perspective. And also, identifies those pathways that have been implemented in this disease. So using this approach, HEAL has identified 60 abdominal aortic aneurysm disease chains and pathways. And also --

**Vicky Whittemore:** Varuna, we need you to wrap up in one minute.

**Varuna Chander:** Okay. Not just identify the genes, but also accurately predicted the disease with clinical utility, as you can see the AUC, you know, specs here. And this is comparable to the existing tests that are present currently in the clinic today. So, we have just applied HEAL for our ME/CFS cohort and we have a very interesting preliminary findings and we have identified the top genes with pathogenic mutations and also its ability to predict the disease status as you can see here, but it's just for the genome. So we got to integrate this model with clinical features from the HR and other predictors to boost the predictive power. So you're going to go hear more from us next time.

And in summary, like I said the GWAS results did not yield any significant hits as expected, but the rare variant analysis did identify top candidate genes regulating playing important role in mitochondrial function cell signaling and inflammation and T cell receptor finding. We are really excited about the HEAL model and its potential to identify novel disease genes, and not just identification but also the prediction of disease status and clinical outcomes in the future.

And the top research priorities real quick, Fereshteh had mentioned already a few of them, which is we need to integrate the family and population studies to better understand the underlying disease ideology for ME/CFS. Considering how complex and heterogeneous this condition is like, you know, but we also need to clearly distinguish causal contributing and course



aggregating variants from artifacts that could be in the analysis, and put it all together, like integrate the different layers of a multi multi-omics layers of approach to better understand the downstream, you know, pathways implicated. And last but not the least, also bring in unbiased nonlinear computational methods, like machine learning frameworks, and use the insights from them to build predictive models so we can better predict the disease for clinical outcomes. And all of these insights we're hoping will lead to better stratification of our patients and identify those patients subgroups in biomarkers, which will eventually hopefully lead to therapeutic strategies in the future.

So this is truly a collaborative effort. I'd like to thank everybody, Mike Snyder and Ron Davis for, you know, helping us do our studies and funding for this research, and also other collaborators so have any graciously given their samples and expertise to drive this research, and my colleagues here at Stanford. Thank you.

**Vicky Whittemore:** Thank you to both of you for really excellent presentations. We need to move on, so I would ask you to answer any of the questions that are in the Q&A. So over to you, Oved, for the next speaker.

**Oved Amitay:** Thank you. I'm sure there are questions, so we'll try to answer them later. So finally, I'd like to welcome Dr. Alain Moreau, who's obviously very familiar to all of us in the ME/CFS community from Université de Montréal in Quebec, in Canada. And Alain, you will be presenting on the contribution of epigenomics to ME/CFS pathogenesis. So we're looking forward to hearing from you about the past, present, but mostly about the future.

**Alain Moreau:** Thank you very much, Oved, for the opportunity to present. We will change a bit the gear today because our excellent speakers so far in session one and session two present a lot about genetics and genomics. And I would like to connect the dots in my 25 minutes presentation about how the epigenetics and epigenomics in particular are connecting with genetics. So this is what we know about, yes, there are some familial predisposition, and this is a true family where you can see on the left open part, the pedigree. And this is not only a family with several affected individuals, but you can see that every generation we increase the number to the third generation, all the offspring are affected by ME/CFS. So this is a very, very special family. And for Chris, this is a UK family that we were able to collect, and we have the methylome as well as the whole genome sequencing data from them. So yes, there are some maybe familial predisposition of ME/CFS and it's worth to collect these special families to further understand. Moreover, I would like you to add that most of the rare variants identifying this affected family members also overlap with the severely ill ME/CFS participation from the Stanford cohort. So this is unpublished results, but we already tested that. So again, these rare familial form or rare families can reveal a lot more than looking at the unrelated case, even if you increase the numbers.

Unfortunately, for my friends in genomics and genetics, I would say that epigenetics probably may have a larger or more important contribution, because you know that there are several triggers like virus as well as bacterial infection, including also the mold toxins and other chemicals or heavy metals that can be pretty good triggers that lead individuals to develop over time ME/CFS. So we think that even though those triggers are relatively well known, it remains unclear if those triggers will introduce epigenetics alterations or there will be also some genetic susceptibility to response to those different triggers. So it's still a debate, although we all agree that environment factor are really contributing to the development of the disease onset with few cases that are going in remission. And I think it's also worked to study them to understand what we mentioned about this morning about protecting biology that how we explain that people are either less inclined to develop ME/CFS or eventually can get in full remissions. And also to further add to the complexity of studying ME/CFS is the fact that having a complex chronic disease that is developing an aging population, you will have several probabilities that will add to the noise to further understand what is happening, and these comorbidities can also be part of the aggravation of some symptoms.

So what do we know about the epigenetics? There are three main epigenetic mechanisms. One involves non-coding RNA, the other one, DNA methylation alterations, and the third one is histone modification. For the sake of my presentation and for your mental health, I will limit to the first two. So non-coding RNA, we have three main non-coding RNA. Those are RNA that are not producing and messenger RNA that would be the code to produce protein. So non-coding can be a small microRNA called microRNA that we can see in this diagram. They are produced as immature pri-microRNA that will go through a maturation process. And eventually

this microRNA has the property to recognize some region in the messenger RNA and by attaching to this messenger RNA or specific messenger RNA, they will form a risk complex with proteins and this will result in the blockade of the translation so we will have a dramatic reduction of the projection of the protein, normally encoded by this messenger RNA, and also often there will be a degradation of the transcript itself.

The other form of non-coding RNA, like long non-coding RNA, that can also act as a natural antisense that can block directly messenger RNA, and also they can act as sponges to sponge different type of microRNA. And there are the third form called circular microRNA that also can participate in the regulation of genes. DNA methylation is well known. So as you can see on this diagram, you may have some methyl group that will be added to some specific position in the DNA, often if it's in the promoter regions that will prevent the activation or the transcription of a gene. So that will be a big stop, especially if there are increase in number. And also if the gene is hypo insulated, we expect to see an overexpression of that gene. However, the same methylation may occur in the body part of the genes. And in that case, that can also lead to over increase of the expression. So it's not something automatic that if you see hypermethylation, the gene would be off. And finally, the third modification is histone modification, where histone will lead to acetylation or deacetylation of the chromatin, which will bring the chromatin, so the chromosome or the part of the genome that will be tightly packed, which preventing the presence and the activation by some transcriptional complex. And when it's acetylated, okay, it's losing and allowing the access to different active complex to increase the transcription of the genes. So very, very important mechanism that you don't need to have mutations, but even without mutation, those alteration may have a dramatic impact, not only on the messenger RNA, but also on the production of synthesis of protein.

So what do we know about the role of microRNA in ME/CFS pathogenesis? So again, just to give you the right context, we have about approximately 20,000 genes. This is the DNA, our database in every single cells, but not every cells are producing the same messenger RNA. So that leads us to an overall of 140,000 transcript. And depending which cell type you are, okay, you will activate some of these transcript in messenger RNA that will be decoded by what we call the ribosomal machinery to produce the functional product, the protein. We have roughly 100,000 protein, but that number increased to one million because those protein can be truncated, can be modified. There are some splice form in different isoforms. So you see that the complexity from DNA to protein increase.

One of the great advantages of the small microRNA, although they are very powerful epigenetics regulator, there's about 2,600 mature microRNAs. So you can see that we can easily reduce the complexity by targeting microRNA as opposed to other molecular aspects of ME/CFS. But those microRNA can also regulate the expression and synthesis of protein. So one microRNA can target up to 200 different messenger RNA and one messenger can be targeted by more than

one microRNA. So this is where we have an additional complexity. So here is, this is a table obtained from a very recent papers. So the good news is there'll be several microRNA associated with ME/CFS. But you can see that many studies, including from my home group, okay, are converging. So meaning that we identify more or less the same microRNA by different methods in different cohorts, which is a good sign. So there is a converging effort, although there are still some microRNA that await to be replicated in additional cohort as well.

Also, not only we can look about microRNA in different biofluid in circulation, like the plasma, serum, or even urine, or also cerebrospinal fluid, but we can also look at the level of cells. So this is an example, a publication from Petty et al about some, for microRNA in particular, been identified in NK cells and they demonstrate by transfecting two specific microRNA, the miR99b and 230-3p, they were able to diminish the complexity and affecting the NK cell function. So this is one beauty of working with microRNA is that you can also do functional assay and to further confirm their role in that case, the activation of the NK cells function. So I think this paper was also very important because we all know that NK cells are also impaired in many ME/CFS patients.

What we don't know is about the other type of non-cutting RNA, like the long non-cutting. So this is a very first paper from the group of Carmen Scheibenbogen at Berlin Charité Hospital in Germany, where she report at 10 different non-cutting, long non-cutting RNA. They have a special name as you can see in the bottom panel, and I did ask my team to do an engineering pathway analysis. This is a colorful diagram you can see on the right, where to see if some of these long coding RNA are targeting either genes known to be involved in chronic fatigue syndrome, as well as regulating some of the microRNA that I previously summarized in the previous table. And you can see that there are some interaction with very well-known microRNAs. So I think there's a lot of works to be done about the contribution along non-coding, because at the end of the day, they can act, as I said before, as sponges, and they can reduce to zero probably, at the action of some micro and even if you can detect them, they're increased. So the theory is a bit more complex. So I think there are more works to be done by looking at the contribution of long non-coding in biofluid as well as in circulation.

What we don't know also is exactly the origin of circulating microRNA how they are transported and why do you have at this cell target, so how some microRNA can be produced, for instance, in the brain but may act on the liver or some produced by the liver it naturally can act at the brain level. So I don't have time to go into detail on all the details about the biogenesis of the microRNA but suffice to say that they are produced often in a small vesicle or exosomes that leads to their protection. But at the surface of those vesicle or exosome, there are also additional proteins that can ease the targeting. So for instance, sometimes you may find fibronexin protein at the surface that will ease the targeting of bone cells because there are plenty of alpha-5, beta-1 integrin that will act as a receptor. But even without specific protein at the surface, they can also

easily be are uptake by different cell type. And we're not very good about to know, to understand when you measure globally in the plasma or in the serum of patients to really understand where they are coming from. So we can start to recognize by proteomic assay, those exosomes of physical surface marker that can ease that recognition, whether they are coming from brain or even some specific part of the brain or different cell type. And I think this is something that remains important on the point of view of eventually developing some therapeutic treatment, manipulating those microRNAs.

So again, the impact of polypharmacy. This is a big, big challenge in the development of biomarker and in that case of microRNA. So microRNA are also very sensitive to the diet, to different supplements and drugs. And this is a recent paper published in 2019 by the group of Elisa Ultra in Spain. And she report, you can see some molecules of prescribed drug that cannot use for patients. And you can see that these microRNA are affected differently locally during the blood cells, PBMCs or plasma and it can increase or decrease. So we need to be aware about that because that can impact the way you look at your patient group or your cohorts. And if those patients are always on the treatments, and sometimes it's more than one drug as you may know, so that may change the value of your biomarkers and their clinical utility. So this is something that we are not that good, so, and we need to further explore that. On the other end, we can use this knowledge to manipulate the expression level of those microRNA as well. So it's a catch-22 situation, but we can also engineer this knowledge to do better for the sake of patients.

So this is exactly what I mean that what do we need to know to how can we develop pharmacological therapies that can modulate the expression of specific microRNA for ME/CFS. So there are direct restoration. So some drugs can increase or decrease some microRNA and sometimes we don't need to totally increase or totally decrease. So it's not a do or die thing. It's just some time to adjust the right pathology for a specific patient and also we can also work with this other type of approach to have an indirect restoration so there are I think a lot of efforts to be done in that direction but there are, this is well known in the field for other disease than ME/CFS and should be further explored in the near future. A good example from my homework is about, can we use also the microRNA to predict a therapeutic response for some subset of patients? The answer is yes. So this is an example about how can we select a patient that can respond to amplitgens because amplitgens are interacting with a total receptor called TLR3. And unfortunately, 75 percent of patients will have a very, very high level of miR-6819-3p, which is targeting TLR3. So if the target is disappearing, it's very hard to allow a compound to interact with. So, which explained that so far the clinical data about the effect of amplitgene, and I'm not here to support the use or own use of amplitgene. I'm just stating the fact that the own data from amplitgene group is reporting that about 25 percent of patients seems to have positive effect. And this is what we are forecasting by measuring this microRNA in ME/CFS patients. So I think

there's a great clinical utility of using microRNA to predict a therapeutic response. And this is certainly a game changer for clinician and patients as well.

So what do we need to know a bit better, I would say, is to be or not to be provocative. That is the question for ME. And as you may know, there is the famous CPET test. This is two cardiopulmonary exercise tests, which are separated by 24 hours for recovery, and that's allowed to not only assess the debilitating consequence of ME/CFS, but also can be used to study using different -omics approaches to further characterize those individuals. The big problem with the CPET is the fact that most of the patients must be able to walk in in a clinic, which means that you are dealing with mild to moderate symptom patients, as opposed to the tests that we developed using an inflatable cuff that applied a passive exercise during 90 minutes and that allowed us to visit patients at home so we can test mainly housebound patients who are really participating in clinical trial. And this is also a game changer so that the test is more flexible, can use all the type of patient, but certainly we are able to test for the very first time housebound patients. So this is another game changer in terms of methods that we can reveal what is going on.

Also, the things about this provocation maneuver or standardized provocation maneuver is it neutralized the effect of body pharmacy. So we don't really care if a patient is taking 7, 9, up to 12 different medications because each participant is being in his own control. So we have a T0 baseline value, and in our case, we measure 90 minutes after the stimulation. So it's long enough to create or to induce post-aggrandizement class, and the change is mainly linked to PEM induction and not to the effect of the drug. So we can see change affecting brain oximetry oxygen level in the brain as well as the cognitive impairment and also more global whole body changes that we can measure by combining proteomic and metabolomic assessment. At the same time we are doing that we are also using connected tools like the exoskin vest where allowed us to capture chronotropic insufficiency in some cases as well as hyperventilation in some participants.

So this test revealed that we developed the first molecular testing with a panel of 11 microRNA that allowed to satisfy patients along four clusters that are associated with specific symptoms that's been published in 2020 in scientific reports. You can see on this diagram in light blue the 11 microRNA and the green their target And you see that they are associated with that constellation of a symptom associated with ME/CFS, and we apply that more recently to fibromyalgia patient. And what we observe in fibromyalgia patients is all the 11 microRNA are downregulated severely, where they are upregulated in ME/CFS. And when the patient is ME plus FM, having FM as a co-morbidity, so we have a intermediary score, as you can see in these two examples with two of the 11 microRNA. This is also a game changer because the open diagrams see that 92 patients came to us with a clinical diagnosis of ME by physicians, but in fact 52 of them were truly ME. 31 were ME-FM and 9 were FM. So this is very important for

the clinician as well as the patients to establish a molecular diagnostic and an evacuee that can assist a researcher to clean up their data whatever the -omics approach you would like to do. So this is certainly a new way to do research and to clean data and would be very important for any clinical trial.

So we make the connection with the same approach with long COVID. So the 11 microRNA that you can see here are connected with the key symptoms that are shared between long COVID and ME/CFS. And we apply this to long COVID, so the SCOPIMED cohort, and we were able to stratify the long COVID in six subgroups and further clusterize them in specific groups. So again, this is the upcoming papers that will be submitted soon, but we are probably the first team that is able to do the molecular certification of long COVID in different groups. Half of them will end up as ME or FM, and other will be something else. Respiratory illnesses, neurological illnesses, and one group that we call for now severe allergy that have very different signatures.

So what we need to do also is the influence of ethnicity on microRNA expression. We don't know much because like Chris Ponting and others that present before me, we are working with European ancestry population, white, Caucasian mainly, but this is an example on this slide, as far as I know, diabetes, okay, with women, and you can see in different population, you have different set of microRNA, even though some population are more or less relatively related, and you see the big difference with Asian and European ancestry population. So, we need to do better works to define whether some macronutrient can be conserved or some would be different and that would be maybe a European or Asian or even African panel for diagnostic purposes. That is maybe a possibility, but I think by applying a provocation maneuver, we may avoid that situation because we will respond more or less the same way to this type of provocation.

Same with about the sensitivity of the test. This is moving, how can we move microRNA diagnostic tests from discovery to market? So this is about colorectal cancer example, where you can see the sensitivity and specificity. So you don't need to know to reach 100 percent. So, so far our tests and tests prepared by others reach over 90 percent so we aren't there yet, so it's working in terms of sensitivity and specificity but again we need to be aware that they might have a difference in different population and that remain to be validated. There are commercial tests so this is two examples that I pick up from the literature so one about metastatic colorectal cancer that can also use as a tyrannistic test to select patients for EGFR treatment or anti-EGFR treatment and the other one for deteriorated cancer. So those tests are commercial. So what I'm telling you is the future is near for me before ME/CFS. I'm not talking about science fiction. So there are great hopes that we can maybe in the next maybe two, three years that we may work with the right partners to have the conversation at the very first test for using microRNA.

Methylation is a bit of a bigger problem because it's a lot of variation. So we are just measuring a very, very small part of the human genome. So right now it's less than 900,000 CPG that we

can measure using chips. There are over 30 million CPGs, so we are covering just a small portion of the genomes and has greater variation in according to the mythology and things. So in other words, there is no convergence yet on the mid-Silon avenue and that will vary depending the type of the cells that you take. So a lot of variation plus if you add to the phenomenon that you may compare FM patient with ME patient or maybe something else that further create more trouble in the way that we try to use methylation.

But nevertheless, we developed this approach for long COVID. We prevent the selection of patient using drugs for the long COVID that modify the DNA methylome, the methylation itself, which is also another obstacle that can increase the noise in different methylation studies. And what we did is we did the provocation studies that allowed us to reveal to, first of all, we can see that we can separate long COVID to a short COVID. So those that are having a full recovery, we have a nice signature panel here. And what we can do is we further stratify, thanks to this provocation essay, short COVID with severe outcome versus long COVID with a better outcome that will over time recover. So this is again a game changer in terms of prevention, so early detection can allow us to break the vicious circle to allow long COVID to develop more permanent long-term sequelae. So that will be soon, I hope, be the published papers.

So to wrap up my study, what we don't know yet is about a change about the good day, bad day. So there's a more dynamic aspect of the DNA methylations. And this is a work from the group of Warren Tate where they can see that at different time point for the very same patient. So the green and the brown is two patients and they measure the DNA methylome on six specific sites where you see an increase or decrease of methylation on those sites, depending if it's a good day or bad day. So we don't know much here. So this is something that we should be aware, especially in the context of ME/CFS or long COVID, as well as to study discordant nine identical twins for ME/CFS. So we collect twins that are discordant, nine for the disease as well as count down one because it's a very powerful method to study the effect of environmental versus genetic, as well, looking at the family.

So this is the very first, the same family that I showed you before with a strong genetic component. What we demonstrate with that family, there is a transgenerational epigenetic inheritance, which is still a debated subject, but in that family, we can prove it that there are all affected members share a strong hyper-methylation or hypo-methylation compared to the unaffected individual. So this is something that must deserve additional studies. So there are already tests available using a panel for methylation on different diseases. So again, this is not a science fiction, but will need further work. And the research priority is again, to further refine the association and identification of circulating microRNA with more specific symptoms like PEM, this has been done, but also more about dysautonomia, orthostatic intolerance, brain fog and other cognitive dysfunction, as well as immune cell dysfunction, and maybe also those that are contributing to produce autoantibodies. And you can see the additional challenges that we



have to do. So finally, I would like to thank my team, the support of Open Medicine, and my collaborators involved in this different aspect of my program on the CFS and non-COVID. Thank you.

**Vicky Whittemore:** Thank you very much, Alain. I don't see any specific questions that have come in for you, so I think we're going to move on to Kristina. Oved, would you like to introduce Kristina for summary?

**Oved Amitay:** Absolutely. Thank you very much. And really to help us put it all together, I'd like to invite Dr. Kristina Allen-Brady from the University of Utah to share with us what are the key takeaways. I really appreciate your doing this with us today.

**Kristina Allen-Brady:** I thank the meeting organizers for the opportunity to summarize this meeting, and also many thanks to the speakers for their excellent presentations. We heard today from Hayla Sluss, who is a caregiver for a family member with ME/CFS. Hayla told about the multiple physical challenges of her affected family member. While ME/CFS symptoms can differ from patient to patient, it's important for us all to remember that ME/CFS affects relationships. It affects relationships with families, extended families, friendships, and feeling connected to the world. We heard from Dr. Chris Ponting, from the University of Edinburgh and principal investigator for the DecodeME study in the UK. He reminded us that ME/CFS is, in part, due to genetic factors. The evidence for this is the increased risk of ME/CFS observed in first degree relatives, second degree relatives, and third degree relatives of probands affected with ME/CFS. As there is an increased risk for ME/CFS in distant relatives who likely don't share the same household, this provides evidence that shared inherited factors also contribute to ME/CFS. Dr. Ponting also pointed out that because ME/CFS does not follow a predictable Mendelian pattern of inheritance, the genetics of ME/CFS are likely to be complex with multiple genes involved.

It is important to understand the underlying genetic contributions to ME/CFS because doing so gives the research community a better chance. It doubles the success rate at developing new drugs or repurposing existing drugs to successfully treat ME/CFS that target key disease pathways and mechanisms. Genetic studies are also unbiased and they search the entire genome. Dr. Ponting suggested that genetic studies of ME/CFS are 10 years behind other disorders. Genome-wide association studies for ME/CFS have been done, but sample sizes are around 2000 participants. Significant genetic findings have not been identified and or replicated. Larger and more genetically diverse cohorts are needed. Dr. Ponting, as we know, is leading the DecodeME study in their sample of approximately 21,000 participants with DNA. They expect to find only five genes or variants that contribute to ME/CFS. This is a good start, but certainly, there will be many more genes variants that contribute to ME/CFS that need to be identified.

With regards to what ME/CFS research priorities are needed, Dr. Ponting recommended a large genome-wide association study to be conducted in the United States with 50,000 plus participants. He has suggested comparing the underlying genetic basis in females to males and comparing those with infection at the onset of disease to those without an infection. While identification of rare variants is more expensive because it requires whole genome or whole exome sequencing, he suggested that rare variant sequencing studies be conducted in individuals who are most severely affected. Furthermore, as ME/CFS overlaps with many other disorders such as long COVID, other chronic overlapping pain conditions, and other autoimmune

disorders, comparisons can be made between identified genes for ME/CFS with these other conditions. Shared genetic factors among these disorders might suggest new pathways for treating these conditions. Dr. Ponting emphasized, and I fully agree, that patient and public involvement in the ME/CFS research process should continue. Data and expertise should be shared globally.

We heard from Dr. Stephen Gardner, the CEO of Precision Life, who shared the results of a combinatorial analysis of genetic risk factors for ME/CFS and long COVID using UK Biobank data. The goal of their method is to capture genetic and molecular interactions that are significantly associated with specific groups of cases selected from a larger case controls design. Their data mining methods look for features such as specific genotypes that are overrepresented in a specific group. And the method continues to add new features until no additional features can be added that improve the score. They use a random permutation process to determine the significance of specific features. Dr. Gardner and colleagues applied their combinatorial method to ME/CFS cases in UK Biobank data. They selected approximately 24 ME/CFS cases, 24,000, 2,400, excuse me, ME/CFS cases, who ever had ME/CFS on a pain questionnaire for the primary analysis, and a second set of approximately 1,300 ME/CFS cases, based on verbal interview as having chronic fatigue syndrome for validation, and controls were selected as having no evidence for chronic fatigue or similar disorders. They identified no genome-wide significant variants in their over 500,000 markers tested. However, after running their combinatorial analysis, they identified 84 groups of SNPs, or risk signatures as they call them, each with three to five SNPs per signature, and 199 total SNPs of interest. Dr. Gardner and colleagues identified 25 of these disease, 25 disease SNPs that were part of multiple disease signatures, and these were in 14 different genes. These genes are involved in host response, metabolic function, sleep disturbance, and autoimmune function.

Dr. Gardner showed the advantage of identifying ME/CFS risk variance as low-density genotype arrays can be developed, and patients can be tested to determine their risk of ME/CFS for a relatively low price. Dr. Gardner showed results for a recently accepted manuscript where they applied their method to the study of long COVID. They identified 13 SNPs and 6 genes that were within 10 kilobases up or downstream from the genes they identified in their first ME/CFS combinatorial genetic study. With regards to what ME/CFS research priorities are needed, Dr. Gardner noted that the available ME/CFS datasets are small, and many with self-reported diagnoses. The datasets contain limited diversity, including ancestry, sex, age, social determinants of health, comorbidities, and psychosocial factors. The available data sets also contain limited longitudinal clinical history, including disease onset, relapse, and recovery. Replication studies are challenging. He stressed that data should be shared that is collected. Better genotype diagnostic tools could be helpful for triaging patients and for evaluating therapy. We need to identify any disease biology that is protective of ME/CFS.

We heard from Dr. Vilma Lammi from the University of Helsinki about a genome-wide association study for long COVID and the challenges of designing such a study. Dr. Lammi and colleagues ultimately included up to 6,000 long COVID cases and over a million population controls from 24 studies across 16 countries using the COVID-19 host genetics initiative. The team identified a genome-wide significant variant in the FOXP4 locus. The FOXP4 locus has been previously associated with COVID-19 severity, lung function, and lung cancer. They have replicated their results in six independent cohorts.

We also heard from Dr. Anniina Tervi from the University of Helsinki, who shared results of a genetic correlation analysis for ME/CFS and some common ME/CFS comorbidities. The goal of their study was to look at the genetic similarity of these pairs of diseases. They found significant genetic similarity between ME/CFS and Raynaud's syndrome, asthma, irritable bowel syndrome, Sjogren's syndrome, and other disorders, suggesting that these disorders might share the same underlying genetic component with ME/CFS. With regards to what ME/CFS research priorities are needed, doctors Lammi and Tervi recommended the use of hypothesis-free genome-wide screens to identify genes variants associated with long COVID and ME/CFS. These genome-wide screens can be used to identify predisposing and protective factors. They also suggested that finding disease subpopulations that may have different underlying disease mechanisms may lead to more optimized treatments.

We heard from Dr. Jahaniani from Stanford University. She highlighted the dilemma of ME/CFS genetics research. Multiple association studies have been done, some even using the same UK Biobank dataset, and yet, the findings have been inconsistent. Part of the problem is that the number of included participants has been small. Family-based studies involving families with multiple cases of ME/CFS are complementary to case control association studies, and they are better able to identify rare variants that contribute to ME/CFS. Dr. Jahaniani shared their genetic results obtained from a nuclear family where the father was affected with hypermobile spectrum disorder, a daughter was affected with hypermobile Ehlers-Danlos syndrome, and an affected son had both hypermobile spectrum disorder and chronic fatigue syndrome. These individuals were either whole genome sequenced or whole exome sequenced, and shared common variants were identified among the family members. They identified rare deleterious variants that were part of pathways implicated in dysregulation in energy metabolism, ABC transporters, coagulation cascades, and the immune system. They included a multi-omics approach and identified immune system dysregulation in the proband.

Dr. Chander shared GWAS and rare variant burden tests run in a relatively small sample of sequenced cases and controls. They included about 400 ME/CFS cases and about 500 controls obtained from Cornell, Stanford, and the UK Biobank. They did not see any significant genome-wide hits. Based on their burden tests, they identified several genes involved in mitochondrial function, cell signaling, and immune function. With regards to what ME/CFS research priorities

are needed, doctors Jahaniani and Chander thought it important to integrate family and population studies to better understand ME/CFS disease architecture. They stress the importance of applying a multi-omics and machine learning approaches to understanding ME/CFS etiology.

Dr. Moreau from the University of Montreal spoke about the contribution of epigenetics to ME/CFS pathogenesis, or the modifications to DNA from environmental factors that affect gene expression. We know that genetic factors predispose to ME/CFS disease onset and progression. Epigenetic alterations acting by environmental influence can also contribute to ME/CFS onset and disease progression. Examples of epigenetic mechanisms include non-coding RNA, DNA methylation, and histone modification. A number of microRNAs have been found recently to be differentially expressed in ME/CFS cases, with the most significant abnormalities seen in natural killer cells, suggesting that defective natural killer cell function could contribute to ME/CFS pathology. Emerging roles in ME/CFS pathogenesis are being identified for long, non-coding RNAs. However, there's still much to learn about the role of other non-coding RNAs and the origin of circulating microRNAs.

It has recently been realized that taking multiple drugs, or polypharmacy, to treat ME/CFS may result in drug-induced alterations in microRNA expression and may cause drug-disease interactions. As clinical therapy can modify microRNAs, Dr. Moreau suggested that microRNAs could be used to predict therapeutic response among different patients' subgroups. His group has already shown differential gene expression of microRNAs associated with post-exertional malaise. In a small sample, Dr. Moreau's team was able to show that 11 microRNAs were upregulated in ME/CFS and can be used to differentiate ME/CFS from fibromyalgia. MicroRNAs have been suggested as potential biomarkers that could be used for diagnostic purposes, although differences exist by race and ethnicity. DNA methylation alterations occur in ME/CFS, similar to micro RNA signatures, methylation patterns may also be useful as biomarkers for clinical testing. There appears to be a shift in the immune inflammatory response from transient to chronic in ME/CFS as reflected in DNA methylome changes. DNA methylation changes can be transmitted from an infected parent to an offspring.

With regards to what ME/CFS research priorities are needed, Dr. Moreau suggested that it is important to identify circulating microRNAs associated with ME/CFS and microRNAs associated with specific symptoms, including post-exertional malaise, dysautonomia, and immune cell dysfunction. Dr. Moreau suggested that it will be important to identify the genes targeted by pathogenic microRNAs and what the origins of these pathogenic microRNAs are. It will be important to screen for drugs and other compounds that reverse the effects of pathogenic microRNAs. And lastly, it will be important to understand post-infection syndromes, including long COVID, DNA methylation alterations, and ME/CFS onset.

In summary, there is much happening globally in genetic studies of ME/CFS, as we have heard here today, but there are many gaps that still need to be solved, including the need to identify validated genetic variants and epigenetic alterations that can contribute to ME/CFS disease onset and progression. I believe we need a mix of multiple study designs to answer these questions. Different study designs have different strengths and different weaknesses. The field of ME/CFS will greatly benefit from larger genome-wide association studies that capture more of the diversity of ME/CFS, including ancestry, gender, age, disease severity, and virus present or absent at onset. Larger genome-wide association studies will require global collaborative efforts. Family studies are also important as they are useful for understanding rare variant contributions to ME/CFS. Rare variants usually have higher penetrance. Genetic studies of ME/CFS and associated comorbid conditions might suggest new shared pathways for treating these often co-occurring conditions. Identification of biomarkers, both genetic and or epigenetic, are needed to facilitate diagnosis. Functional studies are needed to understand how the mutations or epigenetic alterations affect gene function and downstream proteins. Ultimately, identified genetic variants will be useful to improve diagnosis to identify key targets for developing new and repurposing existing drugs for treatment and to promote prevention of ME/CFS in high-risk individuals. Thank you.

**Vicky Whitemore:** Thank you so much, Kristina, for that excellent summary. With that, we will need to bring this webinar to a close. I would like to thank all of the speakers, as well as all of the participants, all of you for your excellent questions and participation today. And I will turn it over to you, Oved, for the last word.

**Oved Amitay:** Thank you so much, Vicky and I think we really have heard from so many wonderful presentations today about where we are. And I think that there's definitely some clear takeaways for what we want to propose in terms of priorities. So really just to thank everyone for their participation today. I look forward to continue this discussion in the coming weeks. Thank you.