Tokyo Workshop on Genomic Medi...eutics, and Health 20240412-06

Fri, Apr 12, 2024 1:17PM • 52:58

**SUMMARY KEYWORDS**

data, variants, individuals, work, sequencing, participants, called, structure, analysis, ancestry, genetic, disease, population, cluster, frequency, mutation, interesting, include, quebec, structural variants

00:08

So I also been to

00:15

maybe more Mondrian, Mount St. Augustine, working both sides as much as the other side is the university.

00:28

So this is correct. So, the first speaker from the University of Leuven, but indeed there are two main University ones row one on each side of the mountain. So let me just start by thanking Mark Latham for inviting me and also all the votes for this three to eight. Meeting. This is the first meeting I go since depending with this format of loading up with meeting is 12 minutes past and I really, really enjoyed the presentation and the discussion afterward. For some obscure reason I've been given 45 minutes so I'm very grateful. It hasn't been time for discussion and question. So I gave them an ambitious talk. My goal is really to introduce chapter nine. As the chair makes it, I want to dyslexic direct, and maybe hopefully convince you of its value and maybe think about project. So this is my talk and I'll start with the obvious so I know you're aware Canada is a very big country. That's the festival ocean depend will be way far down the Atlantic. Montreal will be somewhere here. Quebec is the largest province. We like to brag it's quite big. It's three times the size of France or the five times the size of Japan. But, but most of the people don't live across the whole problem. We all live in the south, and that's going to become important for what I'm going to tell you about. So this is there's a zoomin Boston will be somewhere here and get entrance to the big river. So we have a lot of fresh water. About 4% of the world's freshwater is incorrect. So and that play a really truly important role in the history, including genetic history of our population is mostly of your immune system. They're making 6 million living right now. 7.3 million they're doing this deeper today. All what we call friends. With that mean is that as people move from Europe to what is called Canada back then, but moved to North America. A lot of those people that moved to Quebec are aware of French ancestry and we can trace back the history of her friends getting these into about 8000 8500 settlers that left friends incidents in the 18th century. I'm also going to mention that Canada is a young country, but there have been people living in Canada for 1000s of years. First Nation in which I was talking about this, a lot of them today, as we heard for us about the case and this is also a sensitive issue in Canada. So the focus of my talk is not to say that it's not important to also include isn't involved in the recording but the focus of my talk will be on non Indigenous. And just to give you a reference, there's less than one set of I mean, frankly, new genome, which is of Indigenous ancestry, the bulk of it is a French ancestry and we can we have a sense of where in France because we have the record. So we've got all these settlers that moved to what's called now back. We know that they're mostly from the north west part of France. And that has the impact on the structure of people's lives. describe exactly what that means. The reason why it has an impact is because the genetic diversity that we find now in modern back does not reflect the genetic diversity in the forefront. It reflects the diversity of these individuals that move like them. And because it's mostly from the northern part of France, that's the genetic target factor benefactor. In distant genetics, as you all know, it's called the sound effect. So in Tibet, we have a strong stronger effect. That combined with the fact that families were very large, we have a very large Thunderbolt connection. So for the geneticists in the room, you're probably aware of gymnasts under second Iceland Miko also in Finland, we have some other very similar Andre, in fact, and that has both advantages and disadvantages. And so I will be referring to this Franconia founder population several times. As I mentioned, that is not only a French Canadian province, it is a cosmopolitan province. It is a descendant of a basin defying, famous European ancestry. Minorities and we look at more of these individuals. It's very diverse and mostly recent immigration, order of minority or first nation but then we have individual Trump is really very beans or French or French colonies, the former French colonies from Africa for less than and then you can see me why am I showing you this is because as we think about bringing it precision medicine strategy to Quebec, we cannot focus only on the majority we are talking about these other group and you'll see that this is a big focus. So, what is that design sort of these days, comorbidities of genetic and demography. After that, you have there the mission and the mandate over the years, we're looking at 43,000 participants from six centers. And this is really a population based cohort with a focus on chronic disease. So individual that we're going to work between 40 and 6969 years old at the time. The timeline looks like this. So we started recruiting though this was before me. So everything is 1000 paid for pizza, the first round of recruitment 20,000 People in 2010, then the second round, called Phase II, and that was completed in 2014 with a total of 40,000 participants. And at the top here, we can see the different follow up. We've been following our people, our participants, most recently, through a study with the Government of Canada on the effect of radon and toxic gas on health. And then obviously many many, many incidents studies related to the COVID. Are we on Martin our participant with questionnaire and the recollection. So this this describes where we have but what I wanted wanted to emphasize it's still a very active billboard where we're still collecting information about and we're trying to engage your participants. What type of data do we have? We have ill health listen here. I don't think this is any different than what people here in Japan and elsewhere have been doing so successfully. quite extensive information and health and lifestyle habits formed through that 1000 participants. We have quite extensive diet information on and the reason being that there were PI's installing this new beginning that we're very much interested in the impact of health and nutritional health. So we have a lot of nutrition data, physical measuring things that are used. In clinic, we have linked or participant we can learn things about air pollution. Then we have more wealth, more details in a type on a subset including MRI and also the COVID biospecimen this is all store and available in state of the art biobank. So we have all these biological samples you see on the left. For 30,000 participants. We have the most measuring profile, the blood is 3000 participants, and then what's the focus of mine today? We have separate 30,000 participants and we just finished the sequencing, high high coverage sequencing of 2200 I'll be telling you a lot more about this. So this is why we like to think that despite the fact that in the world, they are very, very large studies and we've heard about some of them today or during the week because it has some unique features. We have universal health care in Tibet. So we have any labor participants in data that offers many advantages. This is not to say that realizing this data is simple. The data obviously stored in the government is not meant to be analyzed. So we presented challenges, but it is possible to use fancy billing code to infer phenotypes and in particularly confirm who has developed the means to recall that so it is possible to define or derive warnings in cases that has many, many fields in terms of thinking about developing prediction strategies with a very broad consent with collaboration both in academia and industry. And we also have bugs in the contact report business and then we just marking the first project within vulnerable contact based on genetic information. So we're excited to see that that's inside of the scope of all participant. Training unique about that wasn't is ineligible construction. So I mentioned earlier that most of the individuals that are called friends and community come from 1000 settlers. So it is possible to theory to build the ancestral tree and actually this was done and you'll see soon by the bow.

10:30

So Balzac is is really, truly unique to Quebec. The idea of marriage certificates have been recorded ever since the beginning of the call and he was 16. So it is possible to retrieve these these certificates to know who got married with who and more to parents. So you can imagine being able to construct from the very beginning the tree of different Canadian back until now we have this type of information about the million French Canadian and our thanks to AI and the retention patterns. The database is being built such that potentially more than 8 million people will be linked to that do you know which is which would be truly unique to get in the world. There's nothing like this. The most famous of theology is probably the one in Iceland but you have 300,000 back now it was very, very, very large as I said 4 million people bring you every one being free. And that present with advantages and I'll show you the exercise promising for human genetics and humans. Alright, so now moving on, moving on to the structure and I can spend a lot of time but you all know what the TCA new maps are. So I'm gonna jump in somewhat describe how we do this analysis but some insight that we get and looking into engineer structure. Let's get that right. So it looks like this. I know you've seen PCA before, but this is a very multi ancestry cosmopolitan PCA. The axes are very similar between European ancestry and African ancestry here, and then we have the other axis of European ancestry with Asian ancestry. And you have here all the different clusters that can take all these PCs to a new map which is shown right, and then all the numbers and this is a very clever way to represent these individuals by Alex resistant small. So all these different clusters here, refer to these populations on the bottom. So instead of assigning a slight ancestry or self declare incident to these cluster, what Alex did is look at the country where the parents of these individual comments and you see for instance, that the cluster is zero here almost the individual whose rent as far from at other cluster, which means in we're very careful with things that definitely because obviously everybody in that cohort are giving positive feedback. They're Canadian, Asians, or so that's one of my students who presented the thing can also infer quite nicely, the diversity. I'm going to spend quite a bit of time to tell you about meaning and structure. But before that, I wanted to show you what we can do, even with smaller groups within the corporate setting. And this is an analysis done by Jeff Lemire again in small habits lab, where it's look at the structure of electro catalytic parties participant with four rural grandparents. So this is an unsupervised clustering analysis is a mixture in selenium. Basically, you tell the software, try to cluster your participant in X number of clusters. So when you take a while for your testing, well foster them within four. There's four corners here. And so for instance, I can show you friends here, you're gonna be most green. You have the African country here, but that's mostly blue. And if you look at the Moroccan country, Morocco, you see here that there are two groups there's like purple and the red. So it really looks like there are two type of ancestry within the Moroccan actually, if you go and ask now doing the same analysis, but with five cluster, you can see that this this group here Moroccan, poor and middle and Moroccan grandparents become orange. So we've created reason the new clusters, which is very specific. It's interesting, because just except the EVO Jr, which is not so far, and should be very, very similar. So what did we learn when we do this and we query the Kepler data set, then we're going to ask them what's the difference between these two? And when we look at self important ancestry, individuals that are orange, some declare themselves as Jewish as the red cells of evidence of an era. So that's interesting. Now there's a fair number of individual European ancestry, that in fact, you find it in some Jewish and then if you look at when these people move, emigrate, as you see a huge difference. So the individual that Moroccan individual define themselves moved into that much earlier. And that follows the independence moral, and they'll end the Six Day War in this harsh world, whereas invisible, again, American living today, we still move into 2000. So it makes for a nice story illustrates nicely how we can better understand the history in Quebec by looking at the data. It also has an implication in terms of precision medicine. Clearly not individuals themselves that are identified as neurochem are the same from genetic sample. And it would be wrong to assume that polygenic risk scores they calibrated for Moroka would work as well for individuals that self declare themselves. This is a little bit unfair themselves. So this is just a very small short analysis of information that I believe is truly truly interesting in terms of thinking about how we use precision ism in the cosmopolitan province. Now moving on to French Canadian because we know a lot about this the US air shows exactly what I showed you earlier but focus on enjoying ourselves determine size trends. And you can see that he's not a genius. So there's a minister of the analysis that we saw earlier for Japan where you're clearly have clusters in the result which are genetically different although all these initials are. So Paul. What is interesting is that we can link these individual with geography in that word and these images will map on the map of that and what we can see and then some some of you are very familiar with this, is that we have structure which maps to the geography so for instance, into that we have a very strong founder of bottleneck effect in the second in that same region. And if you look these individuals here, are very, very different. We also have individuals from Bas and ohana, which would be here along the river, and then in Joe Tombos, which will be here now. So they're structuring Quebec and again, we're gonna think about categories and polygenic risk for buildings because the President mentioned strategies, you need to know about that structure and design which is diverse with the participant because it allows you to do that. Similarly, make it make it this week, but works that way. Actually went one step further and tried to infer the history. So wet by explain these pattern of variation. And you show very, very nicely the deeper reaches there is very nicely all important in the riverbeds or in the history of it. So essentially, you can imagine that 1600 When people from Europe looking back, and they were forests, and rivers were no roads. So basically, people were using rivers to move the province. And it's very, very clear that the main axes of variation are parallel to the big rivers. And so you can have the St. Lawrence River you have to segment downstream so the river with both so very striking. So not only are we learning about diversity, not only are we learning about the fact that we need to be careful being about visitors, but also learning about the history and the demography. So lots of things that we can do when we combine the genetic data with with other parts. Now moving on to health and what we can learn from from from taxpayers money as an advice and help health care and disease prevention. Treatment. I decided to divide this in two parts. I'll focus mostly on rare disease first, and then I'll tell you more about common disease. And I'll start by one example. So there is a disease called primary ciliary dyskinesia. Probably there are people in the room that know better than disease. It's a very rare autosomal recessive disorder remanence of one. So it's very rare, very rare. And we have their description. It's characterized by a impair lung function, essentially. So one of our colleagues at the hospital Neil Shapiro, identify many friends and families with a decent, decent gene called it so interesting, but what can we learn from this? So you contacted as hunger user, population geneticists working with small fidelity Genome Center, and ask, is this a founder, founder mutation, it is awkward that this mutation occurred again and again. So what is the identify who in fact doesn't care? Just remember it's a recessive disorder from the art carriers. And what I humbly did is build the ancestral graph. So think about it as a core lesson. Now this origin firm, we're just looking at doing a haplotype analysis trying to see are the appetites are more similar among the characters than the areas and what was very, very striking. Set 22 of the 31 character, aquatics within the last 10 generation. What that means is that all these individuals all come from the same ancestor, less than 10 generation in a way that's the binding permission of father. So this is exactly what happened, and then went down within this took this one step further. And asked now, what's in the Janell is it's a member we have the analogy on most hard on more than 10,000.

20:31

So what can we learn about the generic so I didn't want to make this will include fine two couples from a small village called married in 16. So the mutation the standard mutation found found into that comes from two couples that come from a small village in that path and make friends marriages. This is interesting and mixlr good stories. It's also interesting from a clinical standpoint, because now we know that this mutation is a funder mutation. And we also know wary, mostly frequent internet, not Adam Shapiro, who see these these patient with his rare disease. Systematic know that he makes a patient come from this part of the bank, they most likely carry. This doesn't have an implication for the treatment yet. But in terms of screening, isn't that easily measured, that becomes really easy to identify, what are you seeing or decide whether you should sequence this and we have multiple stories like this, I forgot to pull the slide but we also found that and this was when I started my lab years ago, mutation design callback, which is responsible for all pretty much all dilated cardiomyopathy cases that comes from another reason is typical gets busy. So we now know at the hospital that if patients have ancestry from espeasy and have dieted cardiomyopathy they're most likely have mutations. So where are we are helping streaming clinical screening as well, using this. So we realized that rare, rare variants would be interesting in the context of the Fondaco patients. So I mentioned we have whole genome sequences how we did all the PCA analysis. We are also now sequencing a whole part and we just finished the sequencing of first trenching homing flow from 200 individuals. We select the 1700 and we also selected the middle from two groups and the middle with four patient records cast and really for grandparents why these two routes? That's because when we decided to set up the project, we looked in the world and we thought who should receive funds that have not been resequenced elsewhere and these are the two groups that are important important in Quebec. They're important minority in Quebec, for whom we don't have a look at what we decided to select these individuals. You have there at the bottom, the number of variants and numbers that are important. And you see we have 16% of variance, our normal project, the bulk of these mobile variants were not found in any of your foundation and again, highlighting the fact that by sequencing under diverse reporting, we're gonna incite more thinking. And so we're truly excited. Okay, so now one question that I get all the time. And then I end you could ask any clinician, any geneticist in Montreal that we get all the time just get with a call from a clinician based in Tibet, but he's an attorney on the province, okay, are based in New England, in the north of us say I have this patient, I found this variant. Is it a sponsor? If so, is it is it an invitation or a recipient or it's a variant in your population? So are you working with them to build the browser and the frequency browser for building this? I mean, with no mod, it's exactly exactly the same thing. So now we have two days off, so it's nothing it's possible, you know, frequency and variance. That will identify is this interesting as important? It is, and it's already being asked the hospital where I work, people are already using it. I showed you there what it looks like. And the idea isn't until there was one carrier the sequencing project, obviously, as we sequenced more in visual, they estimate that it was better. But for me, we work on a disease called sickle cell disease. It was interesting to see that by sequencing individuals, Asian and Hispanic ancestry, we also retrieve the sickle cell mutation. And now we have in the first line, a really good insight on the frequency of that sickle zone patient in electing them. And that as invitation for testing. So the browser is there. If you are interested. I know there are some meal tools to do maybe working in this location. So the browser is available. On this part, we will look at the frequency of variants found in French Canadians have a different getting on the y axis and frequency. We know that non Finnish European, this is law of scale, but security shows that there's a good alignment. So there's frequencies that are not that different. And I plotted they're only high impact variants. So these are variants that were changed in the last two pages and stuff pulled off chance to central splice sites, most very divergent systems and the frequency between no matter what you see all the little blue diamond. So these are all known. Friends getting into many of the known mutations that cause rare diseases. But I think what is more exciting is all these other little gray dots that are enrich in fragmented for for whom we don't know whether they're, they're clinically important. So we're providing to our community or to those analysis now much better insight, much bigger information. In terms of witchcraft is a place that was mentioned but black bar environment is a real positive control. So the other thing that we did and I thought that was never going to work is look at structural variants. The sequencing that we did was classic low, high pass short reads DNA sequencing. And it's definitely known that it's hard to call structural variants using this data is very inaccurate. When I was reading student in my lab, told me she wanted to do it. I said, Fine, please you're gonna learn how to run all these algorithm. I don't know what we're gonna get. So she was very intimidated. She ran different algorithm, integrated all the data, defined a method using a new algorithm to calibrate the quality and we ended up with a group of almost 5000 variants. We have validated all three of these using that method, but what got me the most excited was the slug here because I never thought that was going to work is the frequency of the structural variants on the on the x axis frequency industry. On the y axis and you see a pretty good alignment. I am sure that there are many false positives. And yet, we calibrate quite well. With no not suggesting that many of the structural variants that we identify are active and we are currently trying to assign phenotypes through association, the structural. So this is true from the more common structure and what about the rare structural, getting anything from this type of analysis? And again, I never thought that was going to work but there is a very famous flexural deletion in front team this was identified by fluview that Nolan has Michael Brown and Joe Goldstein, Lauria and Nomura, who found this deletion of LDLR separate for bad cholesterol first example. So this deletion has been known is very, very famous. And so can we find that in our population? And the answer is yes, this is the coverage at the top here on the bottom here, that was the reference allele and then you can see that coverage in the interest is the coverage shown by roughly half so this is a clearly nice example of how we can use more genome sequencing even short reads to identify that the gene has structural variants. And sure enough, we confirm the breakpoint is the classic 15 deletion. For the first time we have an idea of the correctness of this dedication and it is very strongly associated with LDL cholesterol. This is also the first example of us returning information. So we usually don't do that. But in this case, it was an ancillary project that we set up, made sure that these artists were aware of their high cholesterol and RFMD treatment. We knew that it wasn't a bad surprise all these carriers were already taking massive steps. Others virtual very variants, the answer is yes. And you have here three disease that I know very little isn't the last one. But these are structural variants that will reduce that and find meaning known to cause these diseases. And now we have a better idea of frequency. Of these degenerate structural variants. As we move forward and thinking about the future, obviously we are fully aware that short reads of the way those structural variants and we are talking with industry and funders to also include long term sequencing into our so the last part is more common GWAS in this world, are we can we learn something by studying common genetic variation? 30,000 participants once we have access to vitamins, or the first thing that we know that we're trying to do is to improve the patient. It was already mentioned a very nice talk this morning that depends on when you use it by making a Reference panel in the implementation and this is a slide that we published in 2016. We had another low pass

30:26

in already as in 16, with low pass data, we knew that we could improve in patient for rare repair. So knowing that now that we have better sequencing height that we are very much interested in building this and this is working for us. Daniel Italian faculty at the Yale is very interested in buying this. Unfortunately, panel is not ready. But we're working hard on building it and again that the quality of your thunder will depend on the quality of the spacing. And so what Daniel has done is very extensive calibration and phasing. Just to orient you here, what we're looking at the switch error ratio, what is the quality of the phasing data and then by allele frequency, and here what he did is he phase independent data using just the same reference title, or 14 plus 1000s. And in a nutshell, it shows that for most population, adding the 1000 genomes data else the quality of the phasing for for anything does not so we if we use just the cap data in our genome sequence we face in living the Franconia Do you know the quality is as good with or without. And this is true for common variant as well as for so we think that using this data, and using Danielle's expertise, we're going to have a very, very robust for genotyping tradition very quickly, and we hope to make this panel also available. But what can we do in the meantime, while we can do what people are doing that is your top net appetite to include genotyping and that's what we did. So we invested 100 million variants in the very positive and important discipline. We ran Rajini, which is a very convenient Miss Miss, Miss model for Association testing. And we were in this morning 700 phenotypes and numbers there. 300 or so binary phrases disease, and then this is very, very diverse. It is somewhat similar to all the big fi Wi Fi was announced that have been done in the moment with with phenotype that people are all sometime wondering why, but he wasn't nearly so we ran. We found more than 30,000 genome wide significant signal, avatars and beyond Mendelssohn, the new discoveries and but there wasn't planning to walk us through all of the known and novel discoveries here. I think that we are well aware that GWAS works, we are well aware of it seeing the data is available online. So we also have I was happy to see that by Wednesday 21 this morning. So Catherine now has a few ads and this is based on Sunderland and Daniel tell you Islam so that you can go look at the Favorites unified. You'll have wrapping up last time I had the frequency, you know, which was it. And then if I can bring your attention on the circle here, you can also download straight from the website to summarize. You don't have to go over the last catalog and go through all very slow healthcare pages. And just see when when you see the type, click the button. So hopefully, we can convince you. But there's a question of the agility of doing GWAS genetic study of common disease common traits in cattlemen. Again in all the existing data so what can we do with this lesson we learned so I think there are at least three opportunities of using them already. And it's really me trying to convince you to look into these data. The first one that's got that is pretty recent. It's pretty new so that it hasn't been included in existing large means analysis. We often forget now in 2024, but it remains that replicating sociation result is the gold standard expensive and that's true positive. Well, everyone can offer this awesome thing, because it's probably not in the in the amygdala, so GWAS. That you have. Number two is something that we work a lot in my lab, and I'm going to focus on today that is driving in any organization instrument specific. And then the last point and they'll have three example is calibrating existing closing principle to our population experience. And so this is work done by Florian Zuma, who was a postdoc in time and is looking at very well known who's responsible for heart disease and see how they perform in me. Just for you, for those of you that are less familiar has been focused here. You have the distinction of the sport control ring, and in many cases in orange, you see that as a subversion was a perfect that's his classic sport. And we look at incident in cases of coronary disease, you can see that this Court didn't perform very, very well given the limitation of these positions. So this is as good as it gets given the performance of the sport. And there are many people including us here, I was going to be presenting tomorrow are working on developing methods to improve this in Quebec Telecaster. My lab also work on the genetic of lung disorders, and then parking a non malignant and we have papers looking at the genetics of blood cell traits. And what we we rent Iwasa inside their anus and when we tested those scores together with our colleagues in interval in the UK, we also did that as you can see here on the right, the performance of music rescore with deposition data, as well as interval so it shows that even the data is smaller than if you're more into biomed. It remains useful to calibrate and validate and I'm sure that's the case for most smaller study, which have not been included in some of them we've heard this week in the my last example, is actually a nice surprise. So as I was preparing this talk, I look at what's published about gathering because obviously I don't see all the papers. So I stumbled upon this paper and I knew nothing about it. This is a group in Australia, West access to get that data to validate your opposing response for divert or disease of emphasis. If you don't know exactly what that diseases but it was a very nice surprise to this group in Australia that had access to compression data to validate are very exciting to see that now the user is extending outside of it. So this is the classic GWAS where we would normally can can partially petition Milan lab has been asking themselves, how can we make new discoveries and it's hard because he wasn't been very, very successful. There's one thing wondering, are they still discovering me? And so we decided to take advantage of the founder thing my definition for the founders that means is that they were very young to have much higher the frequency in today than elsewhere. And so we selected variants that are four times more frequent in our opinion than other locations. And then we divided this into three groups based on functional occasions we have this group of high impact, thinking about it as nonsense in the supply side, we have found that very conservative reasoning and also that is reactive which means that most cell type most species, if not to, common sense, reason that that probably drive to we subsidize these and then our fee was data to identify are they these variants that probably wouldn't be missed in other professions because much rarer that analysis here we have two examples to show you. I apologize if you've got the ears and get small. This is for artists to making normal levels. And in blue this is as usual, all the band 100 Me and Burnett were found and then you have the three different groups that are mentioned. And you see that there's a black dot there. That's that's outside of the diagonal. So we're curious to know what is very good, it isn't missense variants. The variants about five times more frequent remains rare, but it's five times more frequent. In French Canadian than most other populations. This variance will not be picked up in the study like they used to dieting because it is so rare or I should say we're not reaching a wide significance. We were in this quarter by using standard analysis. For the most part, as looked into this and read the data described in another chord, which we've heard a little bit about this morning, and other things important that has spread as well. So you see this variant and interestingly, these are mapped to an amino acid in the TSH. receptor. So lots of files that mentions here where you have a rare variant in the last year receptor associated with the level of normal.

39:27

Another story is this season with HDL cholesterol, but at the same logic here in this solution, there has to do with HDL cholesterol. And then, if I can draw your attention, there's a little yellow dots that I should have zoom in because it doesn't get justice in the data. But this apparently is more significant than what we would expect in maps to a very conserved region. What is this variant? Again, same story. It's a variant that is much more frequent in French and Indian than other population what would not have been picked up in motion or in the last four divatress has not been picked up. And again, the same story, almost in the West. And if there's any significant that we reduce the number of variants, testing is lower, and we were getting in so safe, and this is just to show where the variant is highly highly conserved region. We're interested because there are no genes. There's no gene A 600 KV window. The gene that would make more sense is ABCA ones you would tell you the disease, it's 3.1 megabase way could still regulate the expression of ABC one but 3.1 megaways starting to do. So, in conclusion, I hope I've convinced you that there's extreme value in these populations specific or I feel like I don't need to do a big, big lot of effort here, given what we've heard about all the different biobanking and I do think that there's many authors and we need to work in different position as an initiative in Quebec in Canada and abroad. I do think it can have us that small can have a global impact on medicine. And in my last point, which might be more important is that government is open for business so you can access the gap that India is very straightforward. There are three steps. The first one is you think about your project and you email the divine man, they will do a feasibility analysis. So they also you know, if someone wants to work on something and we want to defend it, but you will know right away, if it's possible, if not. If it's possible, then you are invited to submit that form. ation, process is really fast. And the beauty is that that means that you can still download the data. So you don't need individual level data and you don't like work yet. You can still access the raw data on a server and you have to get the QR code otherwise it's very easy to find those there's a lot of great people working on this project and these are wonderful picture of wonderful people that I work with. I mentioned there are PI's here and I tried to name all of them but also graduate student and postdoc and and then put them in any particular order actually put them in the order I received their picture. But all these people are key to make a really great project. And there are lots of fun to work with. And then finally, this is our design team. There's only eight people working. So that's already something. We're three scientific, dark, random. I wanted to end by thinking and when does that thank you again for your thanks

42:25

any questions?

42:33

Thank you for the very interesting I would like to know the differences between this cohort and

42:41

socio ecological study of aging. It's a pan Canadian study. So we've had recruited people across the whole country by Skype that really is only based in but the two studies are are quite similar and I think there are around six or 7000 individual NCAA participants CLSA that are in connect so that enables

43:00

these. So they are also included.

43:03

They're independent.

43:08

So for the other ethnicities, which are no offense, French Canadian, what would you do with those individuals?

43:18

The easy answer is I would like to sequence everybody. And I think that me see the shift now with with all of us in the US. We are learning a lot by sequencing in a small number. of individuals. The example I gave with the Moroccan quoi living in the back of Morocco, is quite striking when we have two different groups with grandparents living that came from the same country and if you don't know this, you can apply you can think about applying prediction model music was for knowing this information, and then you realize, well, he breaks obviously first because there are differences. But I want to see I think everybody is interesting and important. And I think that now we've passed the era of GWAS in the sense that we don't need a million people to do something really cool with genetics and even the smaller group are becoming really important.

44:10

And would you do other omics studies, other than just Yeah,

44:15

I would like to do everything by free so obviously, my mind is shifting toward long read sequencing, but it would be very nice. I mean, we have blood. So you know, we just anything that we have, I didn't mention about we have RNA sequence above 1000 people. We don't have single cell data. You know, there's many things that we can do proteomics at some point. It also means about the funding. That's why I would ask her about you know, teaming up with the industry. How can we interest thank you

44:49

for the question, CLSA as we were discussing a lot about also like trying to combine data across cohorts, and this might be some individuals actually that are the same maybe in the two cohorts, and how would you be able to discover that there are some issues?

45:08

So the answer is easy. I know they're not because we looked at the genetic data. We made sure that they were but But you're right, in theory someone could be invited by Katherine NCLs, eight and participated in both, because we can still download the data. It's quite easy. Very cool. So in the lab, we have data from both datasets. So it's quite easy to know. And he said that we have obviously related individuals. So something that that so we don't have the same role in both, but we have related individuals. So it's not the same. Some extent it's the same problem but

45:49

thanks for the talk. I really enjoyed so. So you started your talk with the genealogy and I realized that story about the founder attacks or you find the grandparents as they are related because of the gene. So and then I think that sort of leading towards that polygenic risk score, and wondering how can we use that genealogy information to improve the QRS? So that seems to me the sort of the long term, ultimate, you know, a lot of the ultimate goal I used

46:20

in terms of design better terms of improving

46:23

PRs, finding genes that maybe high rare variants, but you know, through the founder in fact, you sort of even the Bayesian approach where you can actually, you know, have a higher posterior for the effect sizes. It's

46:37

interesting for I think, for the geologists personally useful for rarely, so I don't think so, and we tend not to include those in posing risks, so they might be a way to improve the ball by including rare returns that are maybe more frequent in this population. I'm not sure that the journals per se will help. The existing carriers are mostly based on common variants. So but including the rare that's something that I liked very much so modeling the genealogy into Karis. Right now, we know ultimately, since you're

47:16

showing us the strength versus monster I think this this database has maybe a specific strengths because of the very distinct population kind of just mostly French Canadian, but also some authority. So I'm wondering if you have some disease or phenotypes that you don't see in French Canadians and you don't see the minorities but maybe mixed visuals, so something that arises. So there's into

47:52

three most many of those, like for instance, in a visual, his father would be aged from 18 and multiple different so we don't have a lot of these very recently addonics. It's a very interesting public doesn't do magic compound apps of variant that are come from various different populations. So it's a very interesting, I think, so that that's the major focus of all of us, all of us will have a lot of these individuals. And this calls for new methods to think about PRs or even finding cause of disease. So it's a really active area right now. Like how do you do a PRs on someone whose father was African ancestry? The mother was East Asian six.

48:44

My question is about the reason type. You said the reason structure Barians such as vision relates to some disease, diseases. Use the introduction of the structure Tantek structure, genetic structure as diversity. So the genetic structures diversity in Canadian population, the structural diversity of stocks and barriers, relates to sudden disease.

49:33

So if I understand the question, right, you're asking whether we can use a structural variant to see the structure in the population and the answer is yes. So you can run a PCA say with structural variants and you will see the same representation, then using separate structural events and capture the diversity that a structure

50:03

can kind of have wisdom to you. You said one of the possible uses of this Bible is to use it in the romanization these kind of studies we need to make sure that the link is disciplinary structure. You said also seven is copepods, has some some unique structures and how we think we need to somehow adjust where you fit.

50:39

So it's a good question for you. In terms of common variants, French Canadians are very different. And but where I think it pays off is with these rare variants that are more common ish in printing. It's an instrument that will be very specific in franking that would not be picked up like this. The example I gave you see, these will be instrument that will never come up from a study done in omega n or or U Matic, and now we have a very powerful instrument in our book nation which we have never discovered. We have known the sequence Okay, thank you.

51:29

The University of Tokyo, so he is what is now the latest technique based sequencing approaches. Using those techniques use characterize is called in to obtain the a single cell level access.

52:03

So, just just give me one more set activate the most customize that okay. Is the point or is is thank you for earning some time to me. So, this is a finite Ready, set. My name is Taka, somebody from the University of Tokyo. I'm really honored to be invited to this cutting edge meeting. So but before starting anything else, let me give you several authorities to apologies. The first one is my initial title single cell I'm not at the biobank level is too much exaggeration. By enemies said we have the number of the individuals we have analyzed sulfur is just