AI-generated content may be incorrect

Meshkin, Seyed Hamed * started transcription

LZ **Li, Zhihua** 0:03

Then.

OK screen.

Can everyone see my presentation?

KH **Howard, Kristina** 0:24

Yes.

LZ **Li, Zhihua** 0:26

Yeah. So basically we wanted to.

Make some make some proof of concept investigation into is, if it's even possible to do some computational modeling in this space and this all starts with with this road map.

Earlier this year.

To phase out animal testing for monoclonal antibodies.

And other drugs.

The road map mentioned a bunch of new approach methods and cynical and computational models are one of them, and specifically one AAM model paper published by Wang in 2024 was mentioned in that road map.

KH **Howard, Kristina** 1:10

Mm-hmm.

LZ **Li, Zhihua** 1:14

The road map even.

Frame provided a framework on how to use this kind of models highlighted below.

This time recently developed machine learning models analyzed amino acid sequences of an antibody's variable region to predict whether the antibody is likely to have high or low immunogenicity.

Reference 11 is basically that paper.

It says this kind of tools can flag problematic sequences early, early guiding engineering to derisk the product before.

It ever enters an animal.

Human I guess this is the idea.

How this can can be used to lower the number of animal studies, so we wanted to take a look at this paper since it's referenced by the road map.

This paper uses large language models in the space of antibody sequences.

Large language models.

They are the hottest AM models right now because of Chatgbt and other stuff. They all have two steps.

KH **Howard, Kristina** 2:17

Mm-hmm.

For pre training in this step you collect you collected large amount of amount of data. Of course for chatbot it's natural text but for this kind of bottle you collect it.

Sequences from available regional many different antibodies and then you send into this model and you you randomly randomly marked one of the words or in in our case mark one of the M assets.

And then you ask the model to predict the the hidden one and then you will learn from this error.

So from many, many rounds of learning, the model can can learn the the quote, UN quote, grammar of the language, or the asset sequence in antibodies.

KH **Howard, Kristina** 3:09

Uh-huh.

LZ **Li, Zhihua** 3:09

After that, the model is ready to for inference or predict to do prediction.

Now if you have.

If you have a new antibody, you take the heavy chain light chain of the M nitric acid, and then you first you do.

A pre process called tokenization basically converts it into words and then send them into the model. Then the model will give you something after post processing. Then you can classify them to anybody to be either homogenic or not.

So that's the basic idea of this paper and this kind of model.

KH **Howard, Kristina** 3:41

Mm-hmm.

LZ **Li, Zhihua** 3:45

So while we were going through the paper.

We found two potential issues in the paper.

Why is that?

The authors claim that because the endpoint is kind of a percentage of patients developing ANTIDRAUGANIBODIES or ADA in clinical studies.

But they claim the numbers.

Were extracted from FDA website.

However, most of the antibodies they use in the paper were not approved, which means we have the data, but they shouldn't have.

I don't know how they were able to get the data, but that's the claim.

KH **Howard, Kristina** 4:18

Right.

LZ **Li, Zhihua** 4:21

That's one issue.

Another issue is that every antibody drug in the label.

This is a warning, we said that.

The percentage of antibody ADA is highly dependent on many factors, including the assays being used.

So 10% in this label, if you change the method you it may become 2030%.

So the authors, yeah, or 2%? Yeah. The authors try to use a cutoff.

KH **Howard, Kristina** 4:46

Or 2%.

LZ **Li, Zhihua** 4:52

To to define the anybody's into immorality or nonogenic, I guess their their idea was that now we are not predicting a number, we're not predating 10% or 20%, we're predicting positive or negative. So that make that partially helps with this ambiguity in the percentage, but.

Why 2% is kind of arbitrary?

No concrete.

KH **Howard, Kristina** 5:17

It's extremely, yeah, that's extremely arbitrary.

And it's way too low.

Yeah, yeah, yeah.

So we wanted to see if there were other, better ways to do it.

Solution to the issue one is easy.

We just do our own data collection, goes going through dogs and dog breeds, of course time consuming.

KH **Howard, Kristina** 5:36

Mm-hmm.

LZ **Li, Zhihua** 5:39

But it's it's doable out of the 199, anybody's used 87 or FDA approved, meaning they have the label which provides provides the information.

KH **Howard, Kristina** 5:40

Huh.

LZ **Li, Zhihua** 5:51

For the rest were able to find 83 or 83.

With emergency information from FDA internal databases, others we cannot find, but also there were a lot more unapproved antibody drugs in the database but were not in this 199.

So if necessary, we could expand it to cover more unapproved antibody drugs.

So solution number solution two. We hypothesize that there is a group of antibodies that are non immunogenic.

Our assumption that if since it's negative nonhelogenic, then no matter what kind of assay is being used, the percentage of subjects developing ADA is close to 0 in theory 0. But in reality it's not 0 due to.

False positive or et cetera.

But it should be below a certain threshold.

But for the other group for this. So if if this group is non homogenic but for for the rest which are homogenic the 88 percentage can vary according to the essay. I think we we we did see in the in one of the drugs in the lab.

The per percentage at first was like 10%. Then after switching to.

A drug colorable assay. It became.

30% so we are we assume that this kind of entity drugs, their presence can vary but generally will be above a certain threshold.

So there is a threshold to to divide them into negative or positive with negative, always staying close to 0. But that's our only our hypothesis.

And we don't.

We don't know if that's true.

And also we we don't know if it's possible to to distinguish.

Between these two kinds of antibodies, by looking at the only the sequences, because that's how that model was built.

So we'll ask question, do these two groups of antibodies look different in terms of their amino acid sequences in a variable region?

If so, what kind of threshold can define non versus immunogenic?

So what we did was that we tried very different cut off thresholds, as you can see on the X axis .5% order to 10% for each cutoff threshold.

We just we apply this cut of a threshold to this, about 170 anybody drugs and then divide them into two groups. Below is noninogenic. Above is immunogenic.

Then we ask what?

How different?

These two groups, in terms of sequence similarity.

So the Y axis is the degree of sequence dissimilarity between the two groups, immunogenic versus nonogenic. Of course, with every cut off we have two.

KH **Howard, Kristina** 8:54

Mm-hmm.

LZ **Li, Zhihua** 8:57

Well, two different groups.

So we can always compute the distance.

We we saw that if one cutoff is good or optimal then the distance will be maximized.

Mean the YS will be a lot higher.

There were different ways to to calculate the dissimilarity, but for this particular school metric we we chose method. We chose 1% shows the highest.

Dissimilarity between the two groups.

So in other words, if we use 1% as a cut off.

Anybody's above the threshold and those below form 2 distinct clusters.

Based on the variable region sequences. So in this exercise we wanted to see if we can use 1% as a cutoff, but again this with different methods we may find different thresholds, but this is our first try.

KH **Howard, Kristina** 9:58

Mm-hmm.

LZ **Li, Zhihua** 9:59

Another way to look at it is that we we asked we we grouped the 170 antibodies into four groups based on their lab region sequence similarity.

It's done here represents one anybody.

So if the dots are in the same group or box or cluster, then they have a close sequence similarity. If they are in different groups or different clusters, then the the sequence are different or yeah more different.

KH **Howard, Kristina** 10:18

Mm-hmm.

LZ **Li, Zhihua** 10:33

The this is unsupervised learning, meaning when the model tries to pull them into the boxes or clusters, the model did not know about.

Ada percentage, etcetera.

It this was solely based on sequence similarity.

Then we'll color the dots according to their ADA percentage red dots being above 1%.

Blue dots being below 1% with if if this sequence based clustering has some correlation with the with the ADA percentage then we should see some pattern.

And we did see some pattern. The we have 4 clusters cluster 234.

They seem to be red dominated, meaning most of the antibodies in the cloud in the three clusters are immunogenic or ADA, presented above 1%.

With for the first after, it is half and half again the the clustering is is based on sequence only. So that means most of the immunogenic immunogenic antibodies.

We have some sequence features that are distinct from the first group, so they were put into the 234.

KH **Howard, Kristina** 11:50

Mm-hmm.

Li, Zhihua 14:14

Tell me which one is positive.

So that's our current finding.

So in summary, we improved upon the AM model referenced by the road map.

We use our own knowledge gathering.

And our assumption that.

For non immunogenic antibodies, the ADA percentage is likely to be below certain threshold.

No matter what assay was used and we well, our initial assessment suggests that that threshold could be 1%.

And we also found that immunogenic and non immunogenic antibodies may have different sequence features.

That's why they were put into different clusters and our model has really high accuracy in identifying positive ones.

And it has lower accuracy in identifying.

Negative or non immunogenic ones?

So next step for this current model, there are not some things to do. Clean up the data.

Because first of all, some heavy light chains of antibodies, we got it from the.

We got them from the paper, but they may need to be updated and also the endpoint 88%.

Kind of mixed right now. Some of them are treatments induced or enhanced. Others include background rate because when anybody was approved in the label and the label reports the 88 percent, sometimes they just give you.

A single number covering both background and induced.

We can try to separate them.

And also strengthen the model and also like I said, there were many more unapproved antibodies in the internal databases that we did not touch right now. And also for endpoints, a lot of people say that 88%, they don't care what they care is usually antibodies or.

Ada that affect PK PDE, so we could.

We could just reprocess the data, get the number for this.

Board a clinically important ones and also last of course other models and the points where we we want you to hear your ideas about.

So that's pretty much it.

KH **Howard, Kristina** 16:38

OK.

Thank you.

1st I'm gonna say that considering that you have absolutely 0 immunology experience, I'm impressed that you got this far.

Mm-hmm.

KH **Howard, Kristina** 17:25

The the problem.

I have not read the Wang paper, so I do have to read it.

And I'm. I'm just gonna qualify what I'm saying by saying I have not read that paper. So I I don't know what their experience and or background is in how they did their model.

But basically speaking, the concept so from an immunogenicity standpoint, the unfortunate thing.

Is that when you're looking at the label?

And you look at the percent immunogenicity.

It's not really what I would call a real number, and the reason I say that is because it's based upon the assay they did at the time that the original that, that product, the originator product was approved.

LZ **Li, Zhihua** 18:03

Yes.

KH **Howard, Kristina** 18:14

So for example, you can have products that were approved in the late 90s, early 2000s that are monoclonal.

Specifically, I'm thinking about infliximab.

And adalimab.

And there are others.

And those products, the assays we had available and our understanding of binding and neutralizing antibodies.

Was much less.

Defined as compared to the assays we have available now and what we know.

And.

We don't necessarily care and and I'm saying this we generically this is going to be clinical people, right, the the clinicians, right and they don't necessarily care what the percentage immunogenicity is on the label because.

That's just what they saw in the clinical trial.

There are actually many trials that are done.

Post approval and many trials that are not even associated with the approval.

And that's where we get a much better idea, because you have many more patients that are enrolled. So for example, if in the clinical trial you were allowed to use healthy volunteers, you might get one rate of immunogenicity. But the actual patient population is cancer patients and it.

Going to be significantly lower in them because they're immune suppressed because they're undergoing cancer treatment.

Conversely.

We have products.

That were originally approved for cancer patients, so they are listed as having very low immunogenicity, but now they want to use them in people who have rheumatoid arthritis who have hyperactive immune systems. And you're going to have much higher rates of immunogenicity.

So and immunogenicity as we're referring to it is just whether or not we see anti drug antibodies, we care predominantly about neutralizing rather than just binding.

Because obviously the neutralizing can impact whether or not the drug actually works.

Or you can detect PK.

But the other thing too is that you are also concerned about them.

Because when you make antibodies against a drug, it's also possible they can class switch and then you can end up with antibodies that you know 'cause anaphylaxis. They can switch to IG E.

So most things that you're looking at in the clinical trial.

That you looked at meaning in our databases internally with the approval, that's just IG G.

So it's gonna be very narrow and I don't think there's anything wrong with that, OK.

But when we're trying to say, well, what's a relevant measure of whether or not it's really immunogenic or not?

So for example, you can have a product like. I'm going to say rituximab, right when it was tested originally.

The immunogenicity is less than 5%, which is basically anything under 10 or 15%.

88 percentage, certainly less than 10%.

Neutralizing antibodies would be considered low immunogenicity in the field, so clinicians would not think it's high, right?

Antibodies that would be considered to have high levels of anti drug antibodies, in particular neutralizing those are things where you're seeing typically 30 to 60% so much higher levels in patients.

And those usually end up requiring the patient to stop the medication and change to something else.

And so I think in the Wang paper why they chose 2% beats the living heck out of me because 2% is way too low.

The clinician would never go for that if if you follow my meaning, it's saying sure.

Li, Zhihua 22:17

Yeah. Can I can I?

Can I just? I think that they they they took 2%.

I went with 1%. I think one reason that it's easier to to predict.

I think that's the that's.

KH **Howard, Kristina** 22:29

Oh, of course, of course.

Now let me explain to you what makes something actually immunogenic.

LZ **Li, Zhihua** 22:32

Yes.

KH **Howard, Kristina** 22:35

And I'm gonna just verbally explain it, but I can do a short talk like in two weeks when I'm back from leave.

So antigens.

Any particular thing that is picked up and so these are monoclonal antibodies and a monoclonal antibody is going to be an IgG molecule that is going to have specific.

For the drug, right?

And when it has specificity for the drug, it will bind that and either prevent you from measuring PK or neutralize it.

It may or may not, it just depends.

So you can get and let me just share my screen for a second.

I'm gonna see if I can't.

Find something that would be very helpful in explaining this. I hate saying it, antigen processing in presentation is probably one of the most complicated parts.

To immunology, it's really complicated, which is why it's really, really hard to model this stuff.

All right. We don't want class one. We want class two. There we go.

All right, I'm going to share my screen 'cause. It's gonna be easier for me to explain, and I apologize in advance that this is gonna be really kind of crazy looking.

OK.

Let me share my screen. OK so.

There are two different pathways for antigen processing.

One is class one and the other is Class 2. The one that we're talking about here is predominantly Class 2.

And what that means is.

MHC class one and Class 2 refer to the HLA type OK, so this is.

The easiest way to say explain this.

Is there are three alleles for class one and there are way more than that for class two. Class 1 identifies you as you, meaning you're going to have class one on your cells and that tells

cells like NK cells and cdat cells.

Don't kill me. I'm me.

Please don't kill me.

I'm me, right?

It sets out the I me signal right the self.

Class 2 is used for anything that is processed outside of that so.

Class two was only on antigen presenting cells like.

Dendritic cells, monocytes, etc.

And Class 2 is what sits on the surface of those cells and then talks to T cells and So what typically happens in an antigen presenting cell is you're going to have something come into the antigen presenting cell. It comes in through endocytosis, most monoclonal antibodies.

This is the good. This is the reason we use them rather than basically getting chopped up and processed OK.

Instead what happens is.

So if you have a normal antibody, it gets in the endosome, it binds FCRN and it gets recycled up to the surface of the cell and it doesn't get chopped up. But a lot of times these antibodies will get chopped up because over time they get beat up.

And when they get chopped up, they go to the endosome.

They get chopped into pieces.

And those peptides go into pieces that are typically, classically speaking, a peptide that is presented is going to be somewhere between 16 and 22 amino acids in length. And so exactly how they're going to chop them up.

And there are specific amino acid residues that they like to chop at.

So you can predict what the actual presented antigen will be.

If you know the sequence of the peptide and these are then chopped up, they get loaded onto Class 2, which is what these things are.

These then get attached into a transport vesicle and they get brought up to the surface and they are then presented on the surface.

And if you have a T cell come along and it sees this and it sees the sequence that matches its own.

Antigen recognition receptor.

Then you start making an immune response and this is what dictates whether or not it's immunogenic.

So you have two kinds of T cells.

You have the kind that are going to actually move forward, an immune response and you have regulatory cells which say no, no, no.

This is the same sequence as a self protein.

We're not making a response to that and that that regulatory T cell will come here and say, Nope, this is self, we're not going to make a response to that.

And it tells it to.

It tells the APC to shut it off, but this one it'll say Oh no, this is a totally different peptide.

We don't recognize this.

We need to make a response to this.

And so typically what you're going to have.

In what it is that's going to be recognized is going to be 16 to 22 amino acids that fill in the cleft, and that's what's going to get presented and it's not going to be a human peptide sequence.

Because your regular tour T cells will prevent that from happening.

Least they should. If you get autoimmune disease, it means your regulatory T cells aren't working so hot because they should have told the other T cells not to make a response, but.

So some of the variability that you're seeing in your categories is because you're looking at the amino acid sequence without considering.

Could this be presented and is it a self peptide sequence?

Now there are databases out there. I mean, they've been mapping whether or not a peptide is immunogenic forever.

I mean, this is something they've been doing since 2005 and believe it or not.

Even 20 years later, they still cannot reliably predict.

Is this definitely going to be immunogenic?

There, there are a couple of companies that have and I'm not saying all this to say pour cold water on this.

On the contrary, I I'm trying to point out there actually is.

A. An opening for this because we really do need have need to have a better method of prediction and what has typically been used. I'm gonna pull up a presentation by epivax.

They have a a tool that they use to do to predict.

And so give me just a second and I'll show that to you.

Give me one second.

I should have already had it open, but I was just.

I've been in meetings the entire day to.

I'm a little bit behind the 8 ball.

That's fine.

I remember I think.

This this kind of tool is this from adivax.

KH **Howard, Kristina** 29:37

Yep, epivax correct. Yep.

LZ **Li, Zhihua** 29:38

Yeah, I I think that the the tool may have been mentioned by that one paper.

So I think I I did not read.

KH **Howard, Kristina** 29:43

Yeah, I'm sure.

Yeah, I'm sure it was mentioned.

LZ **Li, Zhihua** 29:47

Yeah.

KH **Howard, Kristina** 29:47

Let me just see here. If yes, here they have it, OK.

So I'm gonna.

Pull this up and they have.

Sequences. OK, so this.

Ogd has a contract currently with Epivax.

And they've done a lot of peptide mapping and things like that.

So let me just show you a little bit of this to try to explain what they do and what they have, what they're doing with their model, all right.

And let me do that.

So this is based on salmon calcitonin, which, by the way is highly immunogenic because it only has 60% homology with human sequence of calcitonin and it's Salman calcitonin because it literally is coming from the fish salmon.

It's named salmon calcitonin because it comes from fish.

So so you can imagine it does not have a great amount of.

Overlapping.

Human and that doesn't always matter.

So in this one, these are peptide sequences.

They're going through the length and they're identifying, you know, what things could be, whether they're hydrophobic or not, whether or not they have any hits. They call their things indicating immunogenicity epibar. And I'll show it here.

So and this is looking at the HLA types just for DRB 1.

So HLA Class 2, you have DRB 1.

This is the primary antigen presentation model in in in humans there are actually many others, but this is the primary.

And each of these represents a field, so 01/01 is an HLA type.

There's actually more digits that can be behind that, but you have a 10103010401701080111011301 and 1501.

And each of these because of the structure of their Class 2 is gonna have a different binding cleft than the next one will.

So if you look at this peptide here that they highlighted.

For these two particular HLA types, we have very high likelihood of binding, meaning if this peptide is seen in these two different.

Class 2 HLA.

Types you're very likely to have immunogenicity whereas you have a lower likelihood as the numbers go down, there's less likelihood and here of course, like this is 0 point, you know minus 0.4, it's not immunogenic.

Minus zero 1.27. It's not immunogenic, so it gives you an idea of, OK, when we look at this for almost every single person, this peptide will probably be immunogenic.

But only for some of these people like this peptide is only immunogenic to this one.

HLA class, right.

Let me see if I can find another one.

This is looking at different sequences.

They're not all the same.

They have different reports here that they've done, but this right now is just all salmon calcitonin.

And what they've done is they they changed their model to do different things. And I'm not saying that your model has to do all this, but this is what leads to the variability and the difficulty in prediction because not every person is going to see a peptide as.

Being immunogenic.

In in other words, just because it's immunogenic to one person doesn't mean it's immunogenic to another.

I I probably should, you know, just stop there just because.

This is probably like throwing a huge amount at you guys all at one time.

LZ **Li, Zhihua** 33:50

Yes, very helpful.

KH **Howard, Kristina** 33:50

But.

But The thing is, is that recently I attended a seminar in which a big pharma company had done some modeling to try to predict immunogenicity.

And they had also used ADA rates. And they said, well, the problem is, is ADA rates don't actually predict what they see in the clinic because the ADA rate on the label is just from a couple of small clinical trials.

I realized, yeah, it's probably 2000 people, but relative to, you know, hundreds of thousands potentially taking the drug, you're going to see a different actual frequency in clinic when many people take it, who actually have the disease that's being treated.

LZ **Li, Zhihua** 34:17

Mm-hmm.

Which could be lower or higher.

And they found that and as you said, if you raise the threshold, it's much harder to predict.

I don't think that they got over 60% predictive ability in their model.

And another issue that they pointed out was that when looking at these models, it's not looking at the likelihood that that particular antigen is even going to be processed.

And turned into something that's presented, which is, you know, the whole front end stuff. It's like, OK.

You may know the sequence, but what's the likelihood? It's even going to be taken up and processed and presented, let alone then be seen as not self.

And so it's a complicated process and I think if and I appreciate the viewpoint that you guys took because.

It's a non immunologic approach and sometimes I think if you want to predict.

Something you do have to think about different ways to look at the same problem when you have people like me looking at it, we look at and say oh, but we have this variable and that variable and that variable and that variable and you have to take all.

Them into consideration and you still can't predict it.

And I I think that perhaps.

Refining what you have done to take some of these other factors into consideration might improve the predictive ability.

And maybe allow a better division of what is immunogenic and what's not. In other words, how do we define it?

What do we define as being predictive?

And it may just be, well, clinically, what is it that we actually see with this product?

If we look over a wide range of papers, So what Jenny did in her in her analysis is she pulled up every paper there was for a whole lot of different biologics to say, well, what's the immunogenicity we see with these products?

And I think if we were to use that rather than what's in the actual label or the clinical trials that support the label, that might actually be a better.

Predictor of what is considered to be immunogenic by clinicians because that's what we really care about is what is a clinician think is immunogenic rather than what they put on the label. Because remember, in some populations that you give a product to, I, I almost think you have.

To divide it up into who's getting the drug.

So if it's an oncology application, if you manage to get over 10 percent, 20% immunogenicity in an ecology patient, you've got something horrifically immun.

But if you're in a rheumatology or an ulcerative colitis patient and it's 10%, you're like, we're set. We're good to go. This is great.

In other words, it's the context, and so I think that if there is a little bit more context in the model.

As far as how you, in other words, putting all monoclonals into the model together at one time may not necessarily be the right way to try to predict them.

Because some of the variables go beyond the peptide sequence and and just seeing what is it that's immunogenic because it's also a question of who's getting it. You know you're going

to have such a different response if it's a cancer patient.

I mean, honestly, with cancer patients, we we don't get excited until we see something that's at least 10 or 15% because they can't make an immune response.

On the other hand, we have a rheumatology patient.

We kind of assume.

Well, 20% might be normal for them because they're so they're so hyper responsive, right?

LZ **Li, Zhihua** 38:21

Mm-hmm.

KH **Howard, Kristina** 38:21

And so where we would put that, I would almost think about it from an AI perspective of, you know, having and I know you need training data.

So I get that.

I understand that you got to have training data.

But it's almost like you want to have a little bit of a gauntlet that you run it through. So if if the patient is this, then your cutoff is gonna be in this range.

If your patient is of this type, it's gonna be in that range. You know, you you wouldn't wanna have your model try to predict all monoclonals all at the same time, because I think that's part of the reason that many of these models fail is because they're trying.

To be so.

They're either so detailed and specific, and they have every variable that you can't, that you then don't have any training data because everything's an end of one, right?

LZ **Li, Zhihua** 39:10

Yeah.

KH **Howard, Kristina** 39:10

Or on the other side you have everything in the same bucket and you can't tell the difference.

Between them, because they're all used for different things. There are different patient populations and you don't have like a pre screening.

So I'm almost wondering if we were to even just put in something simple like a pre screening.

Of, well, what kind of product is this?

What's what's the patient status? And and just dividing them that way, would this model change the output? Meaning, would the buckets that you see change and how they divide out change if you simply change to the patient?

Population was.

Because I and then you could change the percentages and it wouldn't be a big deal if if the percentage was over, you know, one or two percent because by not having things that are really different from each other that are going in very different patients.

I think you would end up getting a more predictive model.

I I don't want to do all the talking, but does does any of that make sense?

Yes, sure, sure.

KH **Howard, Kristina** 40:25
Right, right.

LZ **Li, Zhihua** 40:28
Information we're not using the the patient population sub population is obviously 11 big factor as you mentioned and also of course the label says the kind of assay you used, all those things will impact what you get in terms of 88 percentage. Right now we're using.

KH **Howard, Kristina** 40:36
Yeah.

LZ **Li, Zhihua** 40:46
Only one kind of input because it is a proof concept, but we we.

KH **Howard, Kristina** 40:49
Right, right.
Well, it's right.
It's it's an basically, you know it's a pilot study, that's what it is.

LZ **Li, Zhihua** 40:55
Yes, yes.
Yeah. So we you know we we are we we did think about after this prior study expands AM model at least for the puts we can expand it to cover other factors like patient factors for outputs we can do more refinement like neutral antibody etcetera but that does.

KH **Howard, Kristina** 41:15
Right, of course.

LZ **Li, Zhihua** 41:21
Indeed, what we we have in mind, yeah.

KH **Howard, Kristina** 41:24
Yeah.
No, I I mean I definitely think it is.
It is worth pursuing now. One thing I wanted to mention, and I I literally went front, went to this meeting from another meeting.
So I'm involved in in some Hesse ITC groups.
And you're familiar with ITC Hessi, right?
No. OK, so hessie ITC.
Hessie is basically organizations that are.

A collaborate.

Collaborative space, where people who are in sponsors and government and academia. You know CR OS, they can all work together on problems that are common throughout the industry.

So in other words, what is the problem that we're having and can we put our heads together and can we come up with a problem, you know, can we solve the problem together rather than all working individually?

And maybe if we pull our resources.

And pull our thinking we might, you know, get something better so and they do many different projects and the one that the methods group is right now just beginning to start scoping is doing basically immunoinformatics and trying to do in silico modeling of immunogenicity and tying it into.

Do we get real results with in vitro assays, you know IE?

Can we get a mock?

Can we make a model that works and what was funny is?

Our group.

Is primarily based in the US, one of the people who was there from Europe said.

Oh yeah, there's an EU group.

They're also just starting this, and one of the things and I brought up the fact that this is something that internally at FDA, we're also looking at and one of the comments was you know this is potentially something where economies of scale and trying to figure out, OK.

Maybe if we all try to work on this together, there are pieces and parts that can be done by different, you know.

Groups or individuals and we can actually get a better economy of scale and we can what?

#2, we'd also be more likely to get internal funding if we're working with industry. Just saying.

OK, you always got.

I always think about the money 'cause if you can't pay people, then how are you gonna do the research, right?

But this is the kind of thing that I think if they were to see an overall effort between industry, government particularly of course FDA right working with industry to try to develop a model that is gonna be more predictive using these in silico tools. And I mean lots.

Of companies have insilical tools, but everyone has their own innate.

Experience and biases.

You know, based on what they've done before and there are lots of great modeling groups you guys have obviously done fabulous modeling of different things.

So you know, I I would never say a negative word about the modeling you do, OK, I'm not a modeler.

I don't do it, silico.

I can just tell you the biology.

I do the other part the you know the all the little detail, nitpicky things that are very hard to figure out, but I think that this is something, yes, we should probably.

Scope this internally, but I think it would also be worthwhile to consider you know if the group if the external hessey group. If this is something that we could actually work on as part of a larger project, we might actually be able to move the ball further than.

You know, just your group alone working on it, but also having, you know, industry work on it.

Because remember, they're a little bit more constrained as far as what they're allowed to work on and how much time they can spend, even though they have research and they do conduct modelling as you are aware.

I mean, industry does lots of modelling but their modeling is generally more specific to a particular product or a particular disease.

Process. It's not to try to solve big problems and they don't have lots of spare people.

Oh well, I can just devote.

These people to this and it'll be great.

But I think that this is something that would really be a good selling point of potentially trying to connect this up because your group has resources that industry might not have to do some of the modeling work.

LZ **Li, Zhihua** 45:50

So I think we I'm not familiar with this particular activity from here see but other activities from here see the CPA that have been involved and that's a that's calibration between government, academia and industry. Yeah, so.

So we look at, you know, all kinds of methods in vivo and vitro and silico.

It is the methods group that I said I'm on two different hessi groups. One is methods.

But this group is specifically interested in this, and I know that there's a group in Europe that's just starting to scope this out, and I'm sure it's all because everybody is looking at the road map at the exact same thing that you looked at.

LZ **Li, Zhihua** 46:33

Yes.

KH **Howard, Kristina** 46:35

And we're all saying, OK, how can we address this? Because the problem is, is what epivax is doing. Yeah. Great.

LZ **Li, Zhihua** 46:36

Yeah.

KH **Howard, Kristina** 46:43

That tells you what could be immunogenic, right?

I mean it's, you know, there are models out there that can spit out for you.

Yes, this peptide on this HLA type is going to be immunogenic.

The problem is is that that is at such a refined level, and there's so much detail.

You're never going to have a training, you know you're not going to have training material to train AI to figure it out.

Well, it's too detailed.

Do you know what I mean?

And so we're we're kind of looking at something that is not too far in the weeds, but not flying at 50,000 feet.

LZ **Li, Zhihua** 47:10

Yeah, yeah.

KH **Howard, Kristina** 47:16

We want to.

Kind of be in the middle and if we can get to a model that's in the middle that generically can say, OK, this does or doesn't look human, that part we would want to have is knowing whether or not it would think it was human, because that's an.

Important part to whether it's going to be immunogenic.

But we have that and on the other side of it, we have the actual AI and the running through, you know, simulation after simulation and figuring out, OK, how can we tweak this model to get something that's be more predictive because it would be really helpful if we.

Had an insilico model that was much more predictive for this and the way it's been gone after has been at the micro level and the Super macro level. And we know these methods don't work.

And I don't want to say they don't work. They don't give us the answers we're looking for.

It's not that they don't work, it's just we want a much more specific answer.

Does that kind of help at all?

LZ **Li, Zhihua** 48:14

Yes, sure. Yeah. Yeah, yeah.

So it's for example, if it would expand our model.

Now we say this model considers the level region sequences and also the patient sub population and some other factors.

KH **Howard, Kristina** 48:22

Right.

Right.

LZ **Li, Zhihua** 48:28

And then you predict in these patients we have this.

KH **Howard, Kristina** 48:31

Exactly right.

LZ **Li, Zhihua** 48:33

Yeah, yeah, yeah.

KH **Howard, Kristina** 48:34

Right. In other words, we want some amount of detail, but not the huge minutia, because you're never going to be able to train the model on that because there's no data to train it on.

I I mean, you know, you have to be realistic, but I did want to ask, did anyone else on the team have any questions that you're just dying to ask immunologically while you have me?

LZ **Li, Zhihua** 48:43

Yes.

MJ **Mann, John** 48:54

Well, I I did.

I did have a question when you commented on like the the older essays being like ineffective.

KH **Howard, Kristina** 49:01

Yep.

MJ **Mann, John** 49:01

Is there a like a general cutoff that you think that over time the essays are like near current standard? Like if we were to look back at our data and exclude older essays, is there a time where it gets good enough?

Fig. 5: Trafficking pathways for MHC class II molecules.

From: A guide to antigen processing and presentation