I was Neilatha saying co-founder and chief scientific officer at Biomark IQ where we are developing Barci lung, a clinical inference engine to detect lung cancer at its earlier stages.

So, what you see on the on your screen is a very unfortunate problem.

So, currently more than 90 percent of the lung cancer cases get detected at stage 3 and 4.

Above table is the data from Aims tell me that shows that almost 96 percent cases are coming in late stages when the survival of chances is less.

So, how do we increase the chances of survival of these patients?

It is by detecting the cancers early and what we need is a solution that is non-invasive radiation free and can be available across the length and breadth of the country.

What we are developing is Barci lung which is a screening tool that is an ELISA based blood test that measures different biomarkers in the patient blood and puts it into a computational algorithm which then generates a report and clinical insight based on which confirmatory diagnostic tests can be prescribed.

So, what is our value proposition is that we predict a sensitivity and specificity of more than 90 percent and in AUC of greater than 0.8 and since our test principle is based on ELISA it can be adapted by any diagnostic lab which has a basic ELISA machine and therefore no upgrade of infrastructure is required.

What we offer is an ELISA kit with all the antibodies and solutions that you need to measure different biomarker levels.

A user interface where you have to input these biomarker levels into the system and a computational algorithm will run on the cloud and generate a reference report as you can see on your screen.

So, how did we select our biomarkers?

So, this biomarker selection was an extensive process involving machine learning as well as scientific expertise.

Over the last 1.5 years we carefully went through more than three lakh articles and had our scientific matrices made on which we selected these biomarkers and finally we reached 12 biomarkers which are all protein based and these are predicted to give an early state sensitivity of more than 92 percent and specificity of more than 96 percent.

We ruled out CFDNA and CTC's based on their low sensitivity in the early stages as well as technical feasibility in adoption.

So, now when we came to our 12 biomarker panel that we found out through our extensive data analysis what we went ahead is with an inpatient validation study where we are currently doing a study in 100 subjects 50 lung cancer patients and 50 healthy volunteers and these are our inclusion and exclusion criteria where you know we have classified lung cancer patients as well as healthy and our reference standard is tissue biopsy and index test is VARCE score.

So, I would like to show some of the data that we have obtained till now we have done these in 40 subjects 20 lung cancer patients and 20 healthy controls and what you can see on your screen are 3 different machine learning models that we have run and with each different iteration of the model we do see an increased performance of the model and ROC curve.
So, on the extreme left what you can see is different biomarkers and how much correlation they show with the lung cancer.
From this this study is still under progress as you can see this is just 40 patient and we are currently doing 100 after which we would be able to plot our confusion matrix and get our sensitivity and specificity data, but these subjects we already have.
So, we plan to finish this study by January 2025 and the next thing that we are planning to do is a multi centric clinical trial to validate our pilot data and for this we want to collaborate with Tata Memorial and for recruitment of these patients and conducting this study.
So, the societal impact of our product is multifold as timely detection of lung cancer would lead to improved treatment options, reduced cost of healthcare and increased survival rates.
In the last almost 2 years of our incorporation we have received significant traction, multiple environment grants and fellowship and this is our team being led by Malasik Harsnes who is PhD from Isarpune myself I am a PhD from NCCS and postdoc from University of Pennsylvania.
Atharv is leading our technology team and he is an expert in machine learning and AI models.
We are being advised by Dr. Manojkumar Ghat distinguished professor and former director NCCS and Dr. Sasmitha Panda who is an on call just at Nana Vati hospital.
Thank you and I would love to take your questions.
Yes.
So, currently we have taken AIDS smoking history gender as well as occupational hazard all these factors into play.
So, from the 40th subject study that we see right now we do not see much of association like significant data is not obtained but we are looking into that and we are looking forward to the complete study and then maybe we would be able to tell if it is good.
No, so biomarker selection when we did we did take into account age and so these biomarker levels are not dependent on smoking history is what we see.
So, we have taken into account race age gender smoking history occupational hazard alcohol consumption previous history so these things so and these are not dependent from our in silico data we can say.
Thanks for just wondering about these things about this algorithm methodology algorithm

you are pursuing on that.
So, methodical algorithm in person to have these biomarker to be selected as you mentioned
that AOC is good enough.
So, what is that outcome you have chosen to look on that over there.
So, that may be more on the methodological aspect you are looking.
Because machine learning gives you so many algorithms and if you put something definitely
you are going to get something.
Correct, correct.
So, we are looking at a balance between precision and recall values and for that we are using
basically 6 machine learning algorithms are what we are utilizing currently.
So, based on that we will be like precision and recall values is what we are selecting
the biomarkers on.
And in terms of the validations still you are looking for 1000 new patients to be valid
it not with these 40 patients you do not have any validations go.
Correct, so with these 100 we would be able to get our you know sensitivity and specificity
values that we in lab.
But to confirm that and to see that whether if it stands true in a larger data set we need
more data.
So, I think previous speaker also mentioned that you know with machine learning algorithms
and AI you need even more data set.
But yes 1000 would be something where we will plait you with our sensitivity and we could
be confident that this is what holds true.
Thanks.
I have similar question I mean your sample size is 40 days of now.
Sir, I was just wondering with the 40 samples how did you manage to train your email?
How did we?
How did you manage to train your muscle level?
Yeah, so currently what we are doing is we are these 40 we are amplifying 2000 data set
and then predicting but we are yeah but we are not able to plot confusion matrix because
of that.
These 40 cases are what?
This is a far reason.
No sir.
So 40 is basically 20 healthy 20 lung cancer and in this we have I think three stage one
we have rest of them are stage two to stage four.
So, stage two to stage four will not be hardly lung cancer.
Correct, that is true sir.
Your objective will be to screen.
Yeah.
Screen will be very early lung cancers.
True.
So, whatever model you develop or stage two to four will not apply to very early lung
cancers.
Yes.
So, are you identified the biomarker to immunocaptoromics or part board?

What has been your tool to reach these 12 biomarkers as your final targets?
Sir, I will answer your question in two parts.
The first part was I forgot the first part.
So, okay I will go with the second one.
So, the first part that it would not be.
It should be all one.
So, sir, as I shown in the first slide itself the problem is that early lung cancer patients
are less.
So, what can be done in this scenario is if you take the later patients and do a logistic
regression and then you have stage one few patient and then you overlap both the data.
This is mathematically how we can make sense with even less stage one cancer patients and
more of stage later.
So, this is first part.
The second part how we identified these biomarkers right.
So, for the identification of these biomarkers what we have done is that we have gone through
the scientific databases and through those 3.5 lakh articles that are published in the
lung cancer and diagnostic domains we have utilized natural language processing and from
there and scientific expertise we have curated our scores and divided biomarkers into different
personas and then selected with a human in the loop system.
And then that we are validating in the lab last 12.
So, these are all serum markers and yes, so, cell free DNA are comparatively less in earlier stages.
So, we are taking into account 39 all the 39 subtypes like even under adenocarcinoma's
forma cell all the ones.
But our clinical data if I talk about 20 we have from adenocarcinoma and swamas is majorly
biased.
So, we are saying about the 17 biomarker here.
So, saying about the 17 biomarker are you making the consolidated scoring system by 17 biomarker
or making any thresholding on that?
Yes.
So, these 12 biomarkers in different concentrations.
So, even for example, some of the biomarkers are present in TB cases as well in pneumonia
cases as well.
So, these are the differential concentrations and the algorithm which is running in the
back end that is differentiating between the cancer and the healthy.
And by healthy I mean not cancer it could be any other lung.
So, when you say it is specific and it is comparing with some other lenses also whether
it is turning up false in that case or not.
So, that is right but in our pilot study we are currently just focusing on lung cancer
and once we solve this problem then we would be correct that is.
So, thank you Dr. Snailata.
Thank you so much.
Thank you very much.
Thank you.

Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.
Thank you.