tactiq.io free youtube transcript

HITS colloquium: Julio Saez-Rodriguez on models for personalized medicine

https://www.youtube.com/watch/3JKGE0K-Gqw

00:00:05.359 yeah so we'll talk broadly about

00:00:06.480 computational models that use this large

00:00:08.080 omics data in the context of human

00:00:10.080 disease of personalized medicine and

00:00:11.679 rebecca already say that we are in the

00:00:12.960 medical faculty

00:00:14.400 and i should say for

00:00:16.560 uh openness that we also get funding

00:00:18.960 from

00:00:19.760 g-scan sanofi and i had fees from two

00:00:21.840 small companies called traveronestex

00:00:25.199 but that out of the way

00:00:26.800 as we were saying so

00:00:28.720 broadly speaking in our group we are

00:00:31.039 interested in how we can use these

00:00:32.399 different types of omics data in the

00:00:34.000 context of human disease and biomix it

00:00:36.399 can be

00:00:37.280 measuring many proteins proteomics many

00:00:40.000 metabolites metabolomics

00:00:42.079 many

00:00:43.120 rna molecules transcript transcriptomics

00:00:45.520 and so forth

00:00:46.800 and these technologies allow us to

00:00:48.640 measure

00:00:50.079 in the case of transatomic transcripts

00:00:52.640 of all genes around 20 000 in the sense

00:00:55.280 of proteomics maybe half of that and and

00:00:58.079 so forth so not necessarily a complete

00:01:01.039 picture of the different molecules in

00:01:02.800 the cell but large numbers and then we

00:01:05.280 can measure these across samples

00:01:07.280 these samples can be from the laboratory

00:01:09.520 from

00:01:11.200 cell lines or so forth or from patients

00:01:14.560 and we can also measure this not only

00:01:16.799 looking at

00:01:18.320 the the human cells of the patients but

00:01:20.080 also

00:01:20.960 in in the look at the microbiome our

00:01:23.920 guest cells in our organisms

00:01:26.640 furthermore this type of data

00:01:29.439 historically we could measure it what's

00:01:31.119 called in bulk so if i get a sample from

00:01:34.000 a patient let's say the lung of a

00:01:35.520 patient or a tumor i measure the average

00:01:38.320 proteome or the average transcriptome

00:01:40.640 but now we can also do this at the

00:01:42.000 single cell level namely from that

00:01:43.840 sample we can

00:01:45.759 split up the cells and look at each

00:01:47.200 individual cell what's inside and i

00:01:49.360 think you can imagine that then the

00:01:51.040 numbers explode

00:01:52.880 and then we need computational methods

00:01:55.200 to make sense of this

00:01:56.960 in our case with the ultimate goal to

00:01:58.880 use this to understand what's going on

00:02:00.799 in disease and how we can find the right

00:02:02.240 therapies to treat patients

00:02:05.520 and of course we need computational

00:02:07.520 methods and in particular

00:02:09.840 many people use statistics and machine

00:02:11.599 learning to make sense out of this data

00:02:14.080 right to find patterns in the data that

00:02:17.040 maybe uh we can just understand our

00:02:19.280 disease or build more predictive models

00:02:21.840 to for example

00:02:23.520 try to estimate whether a therapy is

00:02:25.120 going to work on the patient or not

00:02:27.280 is that clear so far

00:02:30.000 okay so as an example something we work

00:02:32.080 in the past a lot

00:02:33.680 is

00:02:34.480 using cell lines so cell lines growth in

00:02:37.360 the in the laboratory this is with

00:02:39.040 matthew garnett and the

00:02:40.840 scientifi in cambridge so you take one

00:02:43.440 thousand differential lines these are

00:02:45.040 samples coming from different tumors

00:02:47.360 from lung cancer breast cancer brain

00:02:49.120 cancer and so forth

00:02:50.720 um maybe 50 or so of each of these tumor

00:02:53.120 types each of these cell lines is

00:02:54.800 different so you characterize with these

00:02:56.239 different omnics like gene expression

00:02:58.080 transcriptomics

00:02:59.519 mutational data and others now there is

00:03:01.920 also proteome and so forth and then also

00:03:04.640 the cell lines you treat them with drugs

00:03:07.120 so these are drugs

00:03:09.040 either drugs used in the clinics like

00:03:10.879 chemotherapy or particular inhibitors

00:03:13.599 but also more experimental drugs

00:03:16.239 this is now 400 drugs and also now

00:03:18.000 there's a lot of data combining drugs

00:03:20.400 so you know

00:03:21.599 for each cell line how much drug you

00:03:23.760 need to kill

00:03:25.760 half of the cells like the ic50 by

00:03:28.400 putting different amount of drugs and

00:03:29.760 then looking at the survival of the

00:03:31.599 cancer cells and what you want to know

00:03:33.840 is why

00:03:34.879 some

00:03:35.680 cell lines are have a high acid 50 so

00:03:38.000 they are resistant to the treatment so

00:03:39.920 this would be

00:03:41.280 reflecting when a drug doesn't work on a

00:03:43.599 particular patient and why in other

00:03:45.360 cases the ic50 is very small with a

00:03:47.840 small amount of drug you can kill the

00:03:49.599 cell line meaning

00:03:51.200 you can stop the growth of a tumor

00:03:54.319 and so here you can then apply

00:03:56.959 any machine learning method

00:03:58.799 and we are many others try different

00:04:00.640 things from very simple linear models to

00:04:03.439 more advanced biasing multitask learning

00:04:05.200 there are people have some deep learning

00:04:06.959 and when you when you just use these

00:04:09.360 methods and i will just not get into the

00:04:12.000 specific results but what happens is

00:04:14.319 that

00:04:15.200 these models are able to predict

00:04:16.720 efficacy of drugs only with limited

00:04:19.358 capacity perhaps in some contexts like

00:04:22.000 some drugs it works a bit better than

00:04:23.520 others but in general it's not that good

00:04:25.759 and also very importantly it's very hard

00:04:27.759 to understand why so even when the model

00:04:30.720 predicts from transcriptomic data

00:04:32.560 whether a drug is going to kill

00:04:34.560 a cancer cell or not

00:04:36.720 we don't know why mechanistically and

00:04:38.240 this is very important if we we want to

00:04:40.800 have trust on these approaches i'm going

00:04:42.240 to take them further let's say

00:04:44.320 uh in the context of of of taking this

00:04:46.720 insight into treating actual patients

00:04:50.240 so for this reason uh the emphasis of

00:04:52.240 our group

00:04:53.280 is how we can

00:04:55.199 help these pure computational black box

00:04:57.520 methods with biological knowledge so

00:04:59.600 there is a lot of things we know about

00:05:01.360 these processes that happens in the cell

00:05:04.479 like we know different pathways

00:05:06.400 different networks and how they work so

00:05:08.639 how we can use it

00:05:10.400 to help

00:05:11.600 the machine learning

00:05:13.120 and the idea as i will show you is out

00:05:15.520 of this large

00:05:16.880 thousands of genes we can extract a

00:05:19.520 smaller number of features

00:05:21.680 and these features because they are

00:05:23.520 fewer

00:05:25.280 increase the statistical power of our

00:05:26.720 methods so less input features more

00:05:28.720 power

00:05:29.919 and second they are rooted in well

00:05:31.919 understood biochemical processes or

00:05:34.880 fairly understood which means that they

00:05:37.120 are more interpretable but they're also

00:05:38.639 more meaningful for follow-up analysis

00:05:40.960 and so forth

00:05:42.479 and all the tools and that i will show

00:05:44.800 you that we do in the lab are free and

00:05:46.479 they are most of them are packages or

00:05:48.160 python packages and we're always happy

00:05:50.320 when people can use them

00:05:53.680 so the first thing is where do we get so

00:05:55.919 where do we get this biological

00:05:57.039 knowledge

00:05:58.400 and

00:05:59.440 so there are many uh good databases

00:06:02.880 there are also many good resources that

00:06:04.880 combine different

00:06:08.319 also databases like you may know also

00:06:10.960 here from embls from board group string

00:06:14.880 or stitch or others

00:06:16.560 so what we did is to focus on a second

00:06:20.000 type of biological knowledge that is the

00:06:22.400 one that is really highly curated but

00:06:26.400 maybe not so large so the things that we

00:06:28.720 know really well to bring them together

00:06:30.720 under one portal and for this we

00:06:32.240 developed this resource which is called

00:06:33.520 omnipath

00:06:34.880 where we include information

00:06:36.800 created information

00:06:38.560 about which protein can activate which

00:06:40.639 protein how this happens complexes

00:06:43.520 annotations

00:06:45.120 and also about localization and really

00:06:47.600 the emphasis in is in keeping all the

00:06:49.440 underlying annotations

00:06:51.360 uh which is now over 2 million coming

00:06:53.280 from over 100 different databases and

00:06:55.120 this is a resource

00:06:56.479 that is driven by our own research

00:06:58.160 questions but of course we make public

00:06:59.680 available for

00:07:01.039 scientists to use

00:07:03.199 and

00:07:04.080 and the idea behind is that

00:07:06.639 there are many databases many places

00:07:09.039 where people have put together knowledge

00:07:10.479 that we can use

00:07:11.919 with different emphasis

00:07:13.599 maybe some people

00:07:15.039 were focused on a particular type of

00:07:16.720 biology like the immune system or cancer

00:07:19.840 other people had a different emphasis

00:07:22.319 and so through omnipot you can pluck and

00:07:24.240 play different resources

00:07:26.160 and the same way that if you wear

00:07:27.360 glasses when you go to try your new

00:07:30.080 glass you try different lens

00:07:32.000 and then you find the right one that

00:07:33.520 allows you to see sharply so in the same

00:07:35.919 way with only path you can

00:07:38.479 mix and match or pick the right resource

00:07:40.240 for your analysis

00:07:42.319 and one thing that uh i wanted also to

00:07:45.520 mention is

00:07:46.639 that uh uh

00:07:48.879 we are now um

00:07:51.199 uh trying to develop uh kind of uh

00:07:54.400 a language that allow us not only to

00:07:56.960 look at our resources that we have

00:07:58.560 through omnipot

00:07:59.919 but to share

00:08:01.440 a common language with other domains of

00:08:04.960 biological knowledge

00:08:07.039 and this is part of i mean the technical

00:08:10.000 and

00:08:11.280 refactoring of

00:08:12.639 of nif path using neo4j but what we hope

00:08:16.160 is it will allow us is that for example

00:08:18.560 there are great meta resources on the

00:08:20.560 chemo informatics world or

00:08:22.720 or on the

00:08:24.000 role of mutation of variance in in

00:08:26.400 disease

00:08:27.520 or indirect resource that does actually

00:08:29.680 text mining so idea is that

00:08:32.159 we can not only use the things that we

00:08:33.679 have access to only but combined with

00:08:35.440 other type of knowledges and if maybe if

00:08:37.760 some of you are developing or are aware

00:08:40.000 or are using

00:08:41.360 some domain specific database that could

00:08:44.399 be interesting to combine

00:08:46.560 let me know because we are now trying to

00:08:48.080 expand this yeah

00:08:50.959 okay so that's a bit how we get the

00:08:52.399 knowledge from

00:08:53.760 and now i will tell you a bit how we use

00:08:55.600 this biological knowledge to extract the

00:08:57.600 signatures from molecular data

00:09:01.040 and one thing that happens very often in

00:09:03.680 molecular networks is that

00:09:06.640 what we want to measure is what we want

00:09:08.880 to know about is hard to measure

00:09:11.040 for example it's very hard to measure at

00:09:13.200 large scale the activity of kinases or

00:09:15.760 of transcription factors of key

00:09:17.120 molecular players but this omics

00:09:20.240 well they are telling us is what is the

00:09:22.640 effect of these activities

00:09:24.880 so if if i'm um

00:09:27.600 if i know how my

00:09:29.519 interest feature my process of interest

00:09:31.279 affects those downstream omics like the

00:09:33.200 footprint and this is what i get from

00:09:35.120 the biological knowledge these causal

00:09:36.560 links i can estimate the former from the

00:09:39.279 latter now by looking at changes in the

00:09:41.440 downstream effect i can see what's the

00:09:43.519 activity of a particular process

00:09:46.640 so to be a bit more concrete um if i

00:09:49.200 have transcriptomics or changes in

00:09:51.040 expression of rna molecules

00:09:53.519 i can use it to estimate the activity of

00:09:56.160 signal impacts not just so this is

00:09:57.920 showing a cell the receptors the

00:09:59.760 different pathways

00:10:01.279 and downstream the control by this would

00:10:03.440 be transcription factors of gene

00:10:05.200 expression

00:10:06.480 and there are tons of methods to look at

00:10:08.720 pathways from transcriptomics but most

00:10:11.120 of them they would look at what is the

00:10:13.279 change on the expression

00:10:15.200 on on these blue components here but

00:10:18.320 these blue components are proteins the

00:10:20.160 signaling cascades are primarily built

00:10:22.079 by proteins so that there is more rna

00:10:24.560 doesn't mean that there is more protein

00:10:25.920 and even if there is more protein

00:10:27.040 doesn't mean that it's active because to

00:10:28.399 be active

00:10:29.519 it maybe needs to go somewhere or needs

00:10:31.680 to be phosphorylated and so on so for

00:10:33.839 this reason we and also others

00:10:35.920 instead look at the footprints so which

00:10:37.519 are the genes that we know change when a

00:10:39.600 pathway is active and by doing this we

00:10:41.600 get a much more accurate

00:10:43.519 estimation of its activity so the same

00:10:46.800 thing i can do to look at transcription

00:10:48.240 factors

00:10:49.360 and it's very analogous so again whether

00:10:52.480 the rna of a transcription factor is

00:10:54.640 higher or lower only roughly correlates

00:10:56.880 with its activity but we know

00:10:59.680 for many description factors with

00:11:01.760 reasonable quality what are the target

00:11:04.160 genes

00:11:05.279 so then this is called regular so we can

00:11:08.000 use this to estimate the description

00:11:09.519 factors

00:11:11.519 the same can go to other type of formics

00:11:14.079 so if i have phosphoproteomic

00:11:16.800 the phosphorylation is driven by the

00:11:18.399 activity of kinases so if i know which

00:11:21.760 phosphorylation sites are controlled by

00:11:23.360 which kinases i can use

00:11:25.279 phosphoproteomic to estimate activity of

00:11:26.880 kinesis that will be here in green

00:11:29.360 and for metabolomics i can do something

00:11:31.440 similar so metabo

00:11:33.680 lights are controlled by metabolic

00:11:35.440 enzymes it's a bit more complicated but

00:11:37.360 in essence it's the same idea that by

00:11:38.880 looking

00:11:39.760 at the changes in metabolites i can

00:11:42.880 estimate

00:11:44.959 the activity of metabolic enzymes

00:11:48.800 anyone is still with me still okay

00:11:51.920 okay

00:11:53.440 good

00:11:56.160 so once we have these features then what

00:11:58.560 we try to do is to integrate them in the

00:12:00.800 network and also here we depicted a

00:12:03.440 network of how different proteins

00:12:05.519 interact so what we try is then to

00:12:07.760 connect them and for this we develop a

00:12:09.680 framework called

00:12:11.040 cosmos which is tries to find causal

00:12:12.880 links connecting different processes

00:12:15.920 molecularly

00:12:17.760 and i will tell you a bit about this

00:12:19.680 using a study we did recently

00:12:22.320 so in this case we're looking at kidney

00:12:24.399 cancer so with colleagues in in rafael

00:12:27.279 kramen christian fritz in cambridge

00:12:29.360 who's actually non-colon and jesper

00:12:30.800 olsen in copenhagen

00:12:32.480 so inaccurate they got these kidney

00:12:34.880 samples

00:12:35.920 and then

00:12:37.279 our colleagues is the transcriptomic the

00:12:38.880 phosphoproteomic and the metabolomic

00:12:41.519 using mass spec and rna-seq

00:12:44.000 then as a first step just what i told

00:12:45.920 you until now from each of these omics

00:12:48.000 we can estimate the activity of

00:12:49.920 particular

00:12:51.680 key players transcription factors

00:12:53.519 kinases and metabolic enzymes

00:12:55.600 and then

00:12:56.720 with this method cosmos what it does is

00:12:59.360 to map this on a large network i will

00:13:01.040 show you in a moment

00:13:03.120 and then trying to find which causal

00:13:04.800 links in these networks can explain

00:13:07.600 what i see so can i find a pathway that

00:13:10.399 explains why when a kinasis is going up

00:13:12.399 here that transcription factor here is

00:13:14.320 going down and maybe a metabolic enzyme

00:13:16.160 here is going up and this example is a

00:13:18.480 schematic so it's trivial we can do it

00:13:20.240 by hand but to hit the large networks we

00:13:22.720 we develop a method based actually in

00:13:25.040 integer linear programming to find an

00:13:27.200 optimal network

00:13:28.959 because indeed the networks can get

00:13:30.480 quite complicated so this is just a

00:13:32.560 herbal to illustrate how large it is the

00:13:35.200 the starting network that we use that

00:13:37.040 comes from omnipot also

00:13:39.839 stitch

00:13:40.720 from mbl to look at

00:13:42.959 protein metabolites binding and recon 3d

00:13:45.120 large metabolic network

00:13:46.880 so you have this network it's not much

00:13:48.560 you can learn from it now you can map on

00:13:50.959 it this kinases transcription factors

00:13:52.959 and metabolic levels and find the key

00:13:55.760 paths that connect them so that you get

00:13:57.839 the network the network on the right

00:14:00.399 still is

00:14:01.519 a bit complicated but certainly less so

00:14:03.440 than the one on the left

00:14:05.120 and then here we can really go deeper

00:14:07.440 and find

00:14:08.639 um

00:14:10.560 specific paths that connect

00:14:13.279 important players and as here you know

00:14:16.000 why a particular kinase how based on

00:14:18.959 what we know of the network how this

00:14:20.240 could be driving changes in a particular

00:14:22.000 metabolite and this in turn affect the

00:14:24.639 transcription factor and and these are

00:14:27.199 some of these findings we when we look

00:14:29.360 further we saw that they were known

00:14:31.839 things in the literature some others are

00:14:34.240 new hypotheses and one thing that i want

00:14:36.720 to emphasize is all these methods are

00:14:39.279 really hypothesis degeneration tools

00:14:41.360 that you should follow up so

00:14:43.680 in making these analysis and these

00:14:44.959 models we are also doing assumptions we

00:14:47.040 know the data is not perfect we know the

00:14:49.120 networks are not perfect

00:14:50.880 so

00:14:51.600 the way to see them is a way to distill

00:14:53.680 from large large data sets specific

00:14:57.600 is that for example it can be validated

00:14:59.920 in a follow-up experiment

00:15:04.160 but just to summarize this part so

00:15:07.519 uh

00:15:08.800 i try to explain you how we use the

00:15:10.399 omics data as an indirect measurement as

00:15:13.519 a footprint of key molecular processes

00:15:17.440 and then later how we can bring them

00:15:19.440 together into large networks using these

00:15:21.120 customers methods to find

00:15:23.120 mechanisms and potential causal paths

00:15:26.560 connecting processes across signaling

00:15:29.120 gene regulation and metabolism

00:15:32.560 okay

00:15:33.759 so far so good

00:15:35.519 any question

00:15:37.759 yes so

00:15:38.959 you mentioned this problem that is just

00:15:40.800 you know there's basically just

00:15:41.839 predictions so how do you give

00:15:43.279 information about the reliability or

00:15:46.399 confidence

00:15:48.079 of the for example of interactions or of

00:15:50.320 the data yeah you can wait them so in

00:15:52.639 the optimization you can wait uh every

00:15:55.279 component uh basically give a different

00:15:57.440 penalty like

00:15:58.880 i choose this

00:16:00.639 one one number score or they see a

00:16:03.839 decomposed sort of score or what how do

00:16:06.480 you tell them so you you would at the

00:16:09.600 end

00:16:11.120 summarize into one score that you try to

00:16:12.959 optimize to find the model and in this

00:16:15.040 score

00:16:16.079 uh you include how well you explain the

00:16:18.399 different data but also some

00:16:21.199 occam's razor some parsimony to have

00:16:23.360 simpler models

00:16:24.720 and and then that's what you use but

00:16:26.480 what also happens is that

00:16:28.639 there is typically many solutions that

00:16:31.120 can be equally good according to your

00:16:33.360 metric and then of course you should

00:16:34.959 report them all and say okay it could go

00:16:36.880 this way or that way and in both cases i

00:16:38.880 could explain the data

00:16:54.480 sorry i understand well the question

00:16:55.920 could you repeat yeah

00:16:57.519 i was going to say it too

00:16:59.360 yes so yeah good point

00:17:08.799 if you have a mutation that affects all

00:17:12.000 the interaction the downstream

00:17:13.760 interaction

00:17:15.039 and can you identify this different path

00:17:18.640 if it's not already defined in some in

00:17:22.240 some other with some other method i

00:17:24.720 don't know yeah okay it's a very good

00:17:26.799 question so

00:17:28.240 first mutations if you know the effect

00:17:30.080 you can include them in the network you

00:17:31.440 can say okay i know this

00:17:33.360 kindness activity is disrupted so in my

00:17:35.200 network i can

00:17:36.480 remove its activity because it's not

00:17:38.720 happening

00:17:39.919 if you don't know it you can try to i

00:17:42.400 mean we haven't done this really but i

00:17:44.000 think yes you could try to um identify

00:17:47.919 effects also because

00:17:50.000 one thing that these methods often allow

00:17:51.919 you to see is data that you cannot

00:17:54.000 explain

00:17:55.200 so if you would see you know this this

00:17:57.280 doesn't make sense then it could also

00:17:59.200 point you that there is something

00:18:00.559 strange going on like

00:18:02.160 a mutation is changing something so we

00:18:04.400 didn't do this systematically but i i

00:18:06.240 think such a framework can allow you to

00:18:07.840 contextualize

00:18:09.600 the effect of mutations but also

00:18:12.160 estimate what could be the effect of a

00:18:13.600 mutation given that you see something

00:18:15.679 different to what you would expect

00:18:20.880 okay great um

00:18:24.640 so

00:18:26.320 so the next thing that we did and that's

00:18:27.600 what we work mostly now uh but um

00:18:31.120 depends on the interest we can go deeper

00:18:32.720 or not

00:18:33.679 is whether we can take these methods for

00:18:36.559 bulk into single cell and i told you at

00:18:38.320 the beginning

00:18:39.520 and now there is

00:18:41.120 a lot of technologies that allow us to

00:18:43.360 look at rna or proteins

00:18:46.240 individual cells in a sample

00:18:48.720 and

00:18:50.320 the first thing that that we ask

00:18:51.760 ourselves is if we can use these

00:18:53.440 footprint methods in the context of

00:18:54.880 single cell because what happens when

00:18:56.640 you measure single cells is that you

00:18:58.000 don't get so many

00:18:59.360 measurements so if you do

00:19:00.880 transcriptomics in bulk you get 20 000

00:19:03.280 genes measure if you do

00:19:05.440 single cell for each cell maybe you get

00:19:07.440 a few thousands or even less

00:19:09.679 and technically anyway we did some

00:19:11.520 benchmarks to show that this still

00:19:14.180 [Music]

00:19:15.760 works

00:19:16.720 and then this means that we can

00:19:19.039 then take a complex sample of a tissue

00:19:21.520 and characterize at the pathway or

00:19:23.120 transcription factor level

00:19:25.280 all the cells or all these groups of

00:19:27.120 cells anyone here works with single cell

00:19:30.080 data

00:19:31.280 probably not

00:19:32.799 so i will not get too much deeper into

00:19:34.720 detail but anyway this is just an

00:19:37.039 example of how

00:19:38.559 you can apply these methods to look at

00:19:40.160 subpopulations of cells and find

00:19:42.160 specific pathways and this guide some

00:19:44.080 follow-up experiment but

00:19:45.760 uh uh it's not so important but one

00:19:48.640 thing that

00:19:49.840 is quite interesting for us is that when

00:19:52.799 um first

00:19:54.400 i mean when you have single cell you can

00:19:56.320 look at subgroups of cells or so-called

00:19:58.640 cell types imagine if you measure

00:20:01.039 a sample from

00:20:02.720 from a tissue of a patient you you will

00:20:05.120 see there is

00:20:06.559 um maybe if it's skin some epithelial

00:20:09.280 cells you also will have immune cells

00:20:10.880 you have many different types of cells

00:20:12.799 and what now you can do is to analyze

00:20:15.200 the different cell types in that sample

00:20:17.120 so in each cell type you can look at the

00:20:18.960 things i saw before like pathways

00:20:20.480 transcription factors

00:20:22.000 but also there are ways to understand

00:20:23.760 how cells talk to each other so you can

00:20:25.520 use

00:20:26.320 the single cell data

00:20:28.400 to study

00:20:29.840 communication as an approximation

00:20:32.320 and the idea is very rough so you will

00:20:34.640 say

00:20:35.440 okay i again from prior knowledge so in

00:20:37.600 our case from omnipot i know

00:20:40.000 which ligand can bite with receptor in

00:20:43.039 in in general

00:20:45.280 and then i look in cell type i if the

00:20:48.559 ligand is very expressed highly spreads

00:20:51.360 and if another cell type the cognate

00:20:53.760 receptor is highly expressed

00:20:55.919 and if they are both highly expressed

00:20:58.159 it is very likely that these cells are

00:21:00.400 communicating because one is shooting

00:21:02.000 out the ligands the one has the

00:21:03.280 corresponding receptor

00:21:05.520 and this is very

00:21:07.039 broadly used in the context of single

00:21:08.720 cell

00:21:09.600 and

00:21:11.440 because we were not sure which method to

00:21:13.120 use and

00:21:14.559 um

00:21:15.600 how well they work we put quite a lot of

00:21:18.240 effort in bringing together many

00:21:19.520 different methods for doing this

00:21:21.840 different databases and we did some

00:21:23.760 benchmarks i will not go into the detail

00:21:26.960 but what we found is that

00:21:30.080 in general these methods give you

00:21:32.080 different results and also that

00:21:35.520 they all get

00:21:36.960 some signal

00:21:38.320 but

00:21:39.120 um

00:21:40.240 it's far from perfect and i just want to

00:21:41.919 make this point

00:21:43.360 that

00:21:46.080 the same way i told you before when when

00:21:47.840 you look at signaling pathway

00:21:50.000 people look at the expression of the

00:21:51.840 components and they say okay if the

00:21:53.520 expression is higher the protein is

00:21:55.039 higher and the protein is more active

00:21:57.600 so in the context of cell cell

00:21:58.799 communication there are also many

00:22:00.080 assumptions

00:22:01.360 and in in this case if you think

00:22:03.600 biologically

00:22:04.880 what happens if two cells are

00:22:06.159 communicating is

00:22:07.840 one let's say sender cell has to express

00:22:10.400 the corresponding legal it has to be of

00:22:13.039 course this rna has to be translated

00:22:14.799 into the peptide protein has been

00:22:16.799 secreted it has to diffuse to the media

00:22:19.120 it has to reach the target cell and then

00:22:22.159 it has to elicit a response

00:22:25.200 uh there are many steps that

00:22:27.120 are in between that are assumed to

00:22:30.080 not to be so

00:22:31.919 critical or or is enough to look at the

00:22:34.720 first step here and kind of the the last

00:22:36.559 step here for what's happening in

00:22:38.480 between

00:22:39.520 and in this case i think this is

00:22:42.320 clear but in many cases when people does

00:22:44.880 analysis of omics data or in

00:22:46.240 bioinformatics we all do a lot of these

00:22:47.919 assumptions

00:22:49.039 and maybe as a word of caution

00:22:50.960 if you use any of these methods

00:22:52.799 think hard try to understand all the

00:22:54.880 assumptions biological assumptions

00:22:56.559 behind the method because this will

00:22:58.559 propagate and is the point i also said

00:23:00.720 before

00:23:01.760 that most of these methods only give you

00:23:03.440 hypothesis that

00:23:04.960 you need to really validate and

00:23:06.799 and part of the problem is to know if

00:23:08.559 methods work is if it's hard to have

00:23:10.559 ground truths it's hard to have a

00:23:12.880 condition where you really know what's

00:23:14.240 happening and

00:23:15.679 we spend a lot of time trying to

00:23:17.120 understand you know

00:23:19.440 how can we know in a system if two cells

00:23:21.120 are truly talking are truly

00:23:22.480 communicating and we also have

00:23:24.640 rna-seq from both cells so that we can

00:23:27.440 prove or check if the methods are really

00:23:29.280 capturing

00:23:30.480 whether cells are talking or not and

00:23:32.320 this is very hard to get such data

00:23:36.559 okay

00:23:38.799 uh

00:23:41.120 yeah

00:23:42.799 so the last thing that is very exciting

00:23:44.799 in the field and i will only touch it

00:23:46.400 but i can tell you more if you're

00:23:47.520 interested is that so i told you now we

00:23:49.520 can measure single cells but what you

00:23:51.440 would do is you take a tissue you will

00:23:52.720 separate the cells

00:23:54.240 and look molecularly every one of them

00:23:57.679 but even newer technologies allow you to

00:23:59.760 do this

00:24:01.039 without having to lose the information

00:24:03.200 of which where was each cell in the

00:24:05.360 tissue right so if you separate the

00:24:07.200 cells and you see and you analyze them

00:24:09.600 okay this cell has disappeared down but

00:24:12.080 it matters who was the neighbor no so

00:24:14.159 how are these cells working together now

00:24:15.919 there are technologies

00:24:17.360 that also allow you to look at the

00:24:20.559 what's happening in specific places in

00:24:22.240 the tissue

00:24:24.000 and

00:24:25.440 we have worked a lot on this

00:24:27.840 we have project which is also part of

00:24:30.080 informatics for life where we try to use

00:24:31.840 it to understand

00:24:33.520 uh

00:24:34.320 when there is infection in the heart so

00:24:36.400 the heart adapts it's called this is

00:24:38.000 called

00:24:38.720 remodeling and involves

00:24:40.640 inflammation of febrosis and

00:24:42.480 we're interested in trying to understand

00:24:44.080 how how this happens molecularly because

00:24:45.919 of course it underlies a lot of

00:24:48.320 deaths in in our society

00:24:51.279 and what we did is

00:24:53.600 to use one of these technologies that

00:24:55.440 give us a special resolution it's called

00:24:57.039 spatial transcriptomics

00:24:58.720 combining with other technologies that

00:25:00.640 look at the dissociated cells and i will

00:25:03.039 i will just not get into the details

00:25:05.679 um but anyway this was work work also

00:25:08.000 with colleagues in in achen and

00:25:09.679 bidenhouse and raphael graham and even

00:25:11.279 constant hendrik milting where we look

00:25:12.720 at different samples from people with an

00:25:15.200 infarction healthy ones and and people

00:25:18.400 with chronic heart failure

00:25:20.320 and what we were planning or what we aim

00:25:22.640 to do is to combine different uh these

00:25:24.640 three different molecular data

00:25:27.360 to understand what's regulated in

00:25:29.200 infection both intra and intercellularly

00:25:33.520 and for this we use the tools that i i

00:25:35.840 saw you before the biological knowledge

00:25:38.000 to try to better understand

00:25:41.840 uh

00:25:43.679 happening on on the different places so

00:25:46.080 for example we can look at activity of

00:25:48.400 pathways or of transcription factors now

00:25:50.799 like as maps like with special

00:25:53.039 resolution so not only what's in each

00:25:55.520 individual cell but what's activity in

00:25:57.520 particular places in in a certain tissue

00:26:00.080 sample

00:26:01.760 and also we can use the information

00:26:04.000 about

00:26:05.679 the localization to try to extract

00:26:08.080 effects of interactions so

00:26:10.480 that's one cell or one place in my

00:26:12.640 tissue

00:26:13.840 depends on the neighboring

00:26:15.840 cells and what can this mean

00:26:18.400 and for this we develop a method called

00:26:20.320 misti to just allow us to to extract

00:26:23.600 this information so

00:26:25.919 basically

00:26:26.960 if i have a tissue and i have a place

00:26:29.360 how much of what happens in that place

00:26:30.880 depends on itself what's called an

00:26:32.480 intrinsic view

00:26:33.840 how much depends on the direct neighbors

00:26:35.840 what we call the yukester view or more

00:26:38.240 distant neighbors called the paraview

00:26:39.760 and this is a machine learning model

00:26:41.039 with a multiview system so you can try

00:26:43.440 to look at these different aspects

00:26:46.080 so this part i'm going fast because

00:26:47.679 since none of you works on single cell

00:26:49.760 or transcriptomic data just to give you

00:26:51.679 a flavor of

00:26:53.440 uh what can be done

00:26:56.400 uh but now i i want to go back to in the

00:26:59.440 last part to

00:27:01.679 to the question of

00:27:03.600 compulsive computational models to

00:27:05.440 predict

00:27:06.480 drug response

00:27:07.919 and in my

00:27:09.279 beginning of the talk i told you

00:27:11.840 how

00:27:12.799 we there is these large data sets where

00:27:15.919 where you have um

00:27:20.000 omix data on one side and drag response

00:27:22.000 in the other and you can throw machine

00:27:23.919 learning methods in between and try to

00:27:26.000 play the drug response

00:27:28.080 and i told you that uh our idea is about

00:27:31.039 using biological knowledge and all these

00:27:32.559 features that i saw you before

00:27:34.559 we can help the algorithm

00:27:37.200 so when we did this and other group did

00:27:39.279 this uh what we found is that these

00:27:41.679 mechanistic signatures indeed improve

00:27:43.520 the predictability

00:27:45.039 but still is like far from perfect there

00:27:46.799 is a large room for improvement

00:27:50.399 and and to really understand better this

00:27:52.320 problem and how you can best predict

00:27:54.480 drug response from molecular data

00:27:57.360 we leverage this dream that also rebecca

00:28:00.159 mentioned an introduction which is

00:28:02.399 a way to involve the

00:28:04.480 global scientific community to find

00:28:06.159 solutions to important computational

00:28:08.159 problems

00:28:09.120 so any one of you knows about

00:28:10.559 crowdsourcing

00:28:13.440 some a bit

00:28:14.840 so i mean there are many flavors of

00:28:17.200 crowdsourcing there's different ways to

00:28:18.480 bring together people to solve a problem

00:28:21.279 and the the variant that we use is to do

00:28:24.080 a challenge um so the idea of a

00:28:26.320 challenge is

00:28:27.440 that there is a problem that

00:28:29.440 you want to know how to solve it and of

00:28:31.760 course you can try to solve it yourself

00:28:33.200 like we do in our own research but you

00:28:35.440 want to see if other people can come up

00:28:37.039 with better methods

00:28:39.200 so let's say

00:28:40.799 the question of can i predict the

00:28:42.159 efficacy of a drug from molecular data

00:28:44.480 so what you do is then you you

00:28:46.480 make public

00:28:47.840 uh

00:28:48.640 some data

00:28:50.000 that people can use to build algorithms

00:28:51.679 and you withhold some test data

00:28:56.240 and this allows you to uh assess

00:29:00.000 um the methods in the bias manner

00:29:02.080 because the one who has the answer is

00:29:03.600 not the one who develops the method and

00:29:05.279 if you enforce that people must submit

00:29:06.720 their methods

00:29:07.919 to improve reproducibility

00:29:10.159 uh

00:29:11.039 and and this can be done also if if you

00:29:13.520 want to use some data from for example

00:29:15.840 the sensitive from patients by

00:29:18.159 by using containers

00:29:21.039 and then you can bring together the

00:29:23.440 information or many methods called the

00:29:25.600 wisdom of the cross and this has been

00:29:26.960 applied in many contexts so

00:29:28.880 in the field of

00:29:30.000 protein structures there was this casp

00:29:32.640 initiative for many years

00:29:34.799 uh and then

00:29:36.480 uh deepmind came along and give it a big

00:29:39.760 boost

00:29:40.640 as you probably know if you work on that

00:29:42.880 field so it's the same idea as casp if

00:29:45.279 you are from that field

00:29:48.880 and what is useful also is that by by

00:29:51.440 bringing many methods you can gather

00:29:53.039 what's called the wisdom of the crowds

00:29:55.200 but i don't know if you know what this

00:29:57.120 is

00:29:58.080 uh but so the other wisdom of the crowds

00:30:00.640 goes long way back to balton so

00:30:03.520 was one of the fathers of modern

00:30:05.200 statistics and he also was

00:30:07.279 wondering himself about um

00:30:10.159 uh

00:30:11.360 how crowds work and also you know how

00:30:13.520 even in the context of

00:30:15.440 the value of boating and and so forth

00:30:17.919 and and then he went to a market in

00:30:21.520 plymouth in england where the farmers

00:30:24.000 used to have like a game so there was

00:30:26.000 like

00:30:26.960 an oaks

00:30:28.080 and then the farmers will try to guess

00:30:30.080 the weight of the oaks and the one

00:30:32.240 who the closest would win win

00:30:34.399 win the price so what galton did is to

00:30:36.880 ask the 800 farmers what what they

00:30:39.279 thought was the weight and compute the

00:30:40.720 median

00:30:42.559 and then it turned out the median was

00:30:43.840 almost exactly the weight so even though

00:30:46.799 maybe no single farmer was right on by

00:30:50.159 bringing together

00:30:51.520 you you get a better prediction and this

00:30:53.520 was called the wisdom of the crowds it's

00:30:54.960 also used in machine learning like if

00:30:56.559 you think of ensemble models that you're

00:30:58.720 kind of doing that as well

00:31:00.320 so

00:31:01.120 trying to

00:31:02.159 compensate the bias of different methods

00:31:04.399 so this idea is

00:31:05.840 something that you can leverage in this

00:31:07.519 crowd sourcing efforts because you are

00:31:09.039 bringing many methods for many people

00:31:11.200 and then you can do like average models

00:31:12.880 of all the methods and this is typically

00:31:14.640 better than any single method

00:31:17.120 so in the context of dream we have

00:31:19.360 applied this

00:31:21.200 in many contexts of progress of

00:31:26.559 molecular systems

00:31:28.240 related to the things i i told you

00:31:30.000 before like gene regulation like

00:31:32.000 transcription factors or signaling

00:31:33.760 pathways

00:31:34.880 or other type of biological networks or

00:31:37.200 even the

00:31:38.320 looking at samples in tissues

00:31:41.279 and also we have look at the context of

00:31:43.200 drug response not the problem that

00:31:45.600 i've been discussing earlier on how do

00:31:47.120 you predict drug response from molecular

00:31:49.679 data

00:31:50.880 and

00:31:52.559 so the way you would do that in that

00:31:53.919 context is you will say okay i have from

00:31:56.799 some cell lines

00:31:58.159 i have information on the molecular

00:32:00.159 level

00:32:01.279 i have information of how some drugs

00:32:04.240 affect these cell lines as i said before

00:32:06.880 you keep some data the test data

00:32:09.679 and then you make public the training

00:32:11.120 data so

00:32:13.440 people have let's say three months to

00:32:14.960 build algorithms

00:32:16.480 and then to provide

00:32:17.919 the answer for this data and then you

00:32:19.679 can score them now in in this case this

00:32:21.760 was

00:32:22.480 a few years back now but we had 44

00:32:24.880 different teams

00:32:26.159 and then we could score them just in

00:32:27.600 some metrics

00:32:28.960 and the point i want to make in the

00:32:30.480 context of drag response is that

00:32:32.559 here below you can see each column is

00:32:34.799 each of the methods and and

00:32:36.960 each of the participants and they are

00:32:38.960 grouped by the type of algorithm they

00:32:40.720 use like from random forest or

00:32:42.799 supervisor machines

00:32:44.480 and then you could uh

00:32:47.360 score them and according to this metric

00:32:50.399 a random prediction so just trying to

00:32:52.720 guess

00:32:53.840 would be this line

00:32:56.320 and a perfect which will be one is kind

00:32:58.480 of out of scale up here no so what this

00:33:01.679 figure tells us is that even though most

00:33:04.000 methods got some signal so this is

00:33:05.600 better than random

00:33:07.200 again there is still a lot of room for

00:33:08.720 improvement so the same thing that we

00:33:10.640 found our research and also here we

00:33:12.720 found that teams that would use

00:33:14.080 biological knowledge

00:33:15.519 pathways and so forth would do better

00:33:17.039 than those that didn't

00:33:19.200 so

00:33:21.840 uh

00:33:23.360 so why is it so hard to predict this

00:33:25.200 drug response no so

00:33:27.360 uh i saw you our own research

00:33:29.679 how far we got

00:33:31.039 then we also

00:33:32.880 tried to ask the community

00:33:34.960 how to solve this problem

00:33:36.960 and

00:33:37.510 [Music]

00:33:39.200 i don't know anyone has any idea why is

00:33:40.720 it so hard to predict the drug response

00:33:53.200 so viability across the

00:33:56.320 individuals or

00:33:58.640 drugs

00:34:00.880 yeah exactly

00:34:03.760 every day

00:34:30.960 [Music]

00:34:46.719 and

00:34:47.520 having the right output and all of these

00:34:49.760 are good arguments i normally have a

00:34:52.000 list to show that includes those uh but

00:34:54.960 there is one more that

00:34:57.119 we think it's maybe a bit under

00:34:59.040 appreciated and and this is very nicely

00:35:01.040 summarized in uh

00:35:03.040 in a review of years back from danila

00:35:04.800 thai it's an oncologist in boston

00:35:06.640 bennett farmer

00:35:08.480 he says consider trying to predict what

00:35:10.160 happens if you poke a dog with a stick

00:35:12.480 by analogy with hitting a cancer cell

00:35:14.720 with drug

00:35:15.760 so the analogy would be then you kill

00:35:17.520 the dog you look at all its omics

00:35:19.440 molecularly

00:35:20.800 you measure all these terabytes of data

00:35:23.119 and then you try to use this to make the

00:35:24.880 prediction

00:35:26.079 but then he says okay there is the other

00:35:28.000 way which is

00:35:29.200 hit the dog with the stick and see what

00:35:30.800 happens

00:35:32.000 and what is meant by this is so we are

00:35:34.000 trying to predict how cells responds to

00:35:36.560 a perturbation

00:35:38.400 from basal data right all these

00:35:40.079 molecular measurements are done

00:35:43.119 in kind of static steady state

00:35:45.359 and it's hard to predict from a static

00:35:47.040 measurement how

00:35:48.800 it will respond to a perturbation

00:35:51.920 uh and

00:35:53.359 for this reason we

00:35:55.119 and others also use the kind of other

00:35:57.839 type of approaches to look at this

00:35:59.280 problem so besides throwing a lot of big

00:36:03.200 large data sets into machine learning we

00:36:05.440 try to be more detailed mechanistic

00:36:07.520 models that

00:36:09.200 try to explain the dynamics of these

00:36:10.800 systems so how they evolve over time

00:36:12.480 when they are prepared

00:36:15.040 and for this we used

00:36:17.680 logic models so we tried to simulate

00:36:19.920 what happens in

00:36:21.359 cellular systems

00:36:22.800 by analogy with

00:36:24.640 what happens in kind of in an electrical

00:36:27.280 circuit so you have

00:36:29.119 nodes and edges which are connected by

00:36:31.520 logic gates

00:36:33.200 and this is very simple of course

00:36:34.800 biochemistry is much more complex but

00:36:36.320 this allows us to scale it up to the

00:36:38.320 largest networks and also we can use

00:36:40.720 different different formalisms and you

00:36:42.240 can even convert them into differential

00:36:43.839 equations so this allows us to take data

00:36:46.079 so you have a focus system you perturb

00:36:48.400 with different ways

00:36:50.079 you use prior knowledge to build a model

00:36:52.480 and then you train the model to the data

00:36:54.160 and then you can construct dynamic

00:36:55.520 models dynamic models that you can

00:36:56.960 simulate you can use to predict the

00:36:59.119 effect of new perturbations and think

00:37:00.480 how happens how things happen over time

00:37:03.520 you can use different types of molecular

00:37:05.200 data it doesn't matter so much

00:37:07.119 but then we have used these models to go

00:37:08.720 back to this question of how you can

00:37:10.240 better predict track response

00:37:12.480 and

00:37:14.480 so the way you would go about this is

00:37:16.800 of course these models are much more

00:37:18.079 detailed and more laborious and they

00:37:19.440 also need this perturbation data so you

00:37:21.599 cannot do them for

00:37:23.599 all potential drugs and cell lines but

00:37:26.480 you can go deeper and and in particular

00:37:29.280 cases where you are really bad in

00:37:30.720 predicting in a pure machine learning

00:37:32.480 manner

00:37:33.359 you can try to go deeper to dedicate

00:37:36.320 experiments with perturbations with

00:37:37.920 drugs and stressor ligands across

00:37:40.000 different cell lines and measure if you

00:37:42.640 can the activity of proteins with

00:37:44.160 phosphoproteomics

00:37:46.000 and then by doing this for each of the

00:37:47.599 cell lines

00:37:48.960 you can train a generic network of of

00:37:52.160 the cell to data of each of the cell

00:37:54.240 lines and at the end build cellular

00:37:56.079 specific models

00:37:58.160 and we have done this in the past in in

00:38:00.320 different contexts and we have shown

00:38:02.560 that these models allow us to

00:38:04.960 to

00:38:06.000 to better predict um

00:38:08.160 a drug response and to really identify

00:38:10.000 mechanisms that you will not find by

00:38:12.079 purely looking at

00:38:14.480 at the basal data

00:38:16.640 so if i don't lose my voice in the next

00:38:18.480 five minutes

00:38:19.680 uh i will just show you the latest

00:38:21.680 things we are doing around this

00:38:23.680 which is it goes about the question of

00:38:26.079 how we can translate these ideas of

00:38:28.720 models

00:38:29.760 to patients

00:38:31.440 and the the reason is that these type of

00:38:33.599 experiments like the ones i saw you here

00:38:36.160 are laborious and you need a lot of

00:38:37.520 material because you need to take cells

00:38:39.119 in different conditions prepare for

00:38:40.800 measure so use this you can do in the

00:38:42.480 laboratory but

00:38:43.839 at the end of the day what you want to

00:38:45.119 model and what you want to predict is if

00:38:47.200 a drug is going to work on a patient

00:38:49.520 and in patients it's much harder of

00:38:51.200 course to get material

00:38:53.359 so one context where this is still

00:38:55.119 possible is if you look at blood because

00:38:58.079 blood you can get more amount so for

00:39:00.320 example in a study we did recently we

00:39:02.800 look at multiple sclerosis patients the

00:39:05.040 blood we can study the immune cells and

00:39:07.440 it's always this idea of taking the

00:39:09.440 sample perturbed ligands and drugs

00:39:12.400 in our case 150 donors built for its

00:39:15.599 donor a model and use these models to

00:39:17.920 simulate and to predict potential

00:39:19.920 combination therapies

00:39:21.920 and just to give you an example so this

00:39:23.760 is

00:39:24.480 what such a model looks like so you have

00:39:26.480 all these kinases and proteins covering

00:39:28.720 different aspects of biology and using

00:39:30.880 this model we come up with a combination

00:39:32.640 therapy so two drugs that together were

00:39:34.400 more effective

00:39:36.160 and then going back to a mouse model we

00:39:38.640 could validate this

00:39:40.240 this is just to show that the validation

00:39:41.920 work

00:39:43.280 uh but if you don't look at the blood if

00:39:45.200 you look at solid tissues is you don't

00:39:47.680 get so much material

00:39:49.200 so for this what we did is to work

00:39:50.960 together with christopher merton who was

00:39:52.720 at the mbl he's not epsilon lausanne

00:39:58.640 to use microfluidics to do

00:40:00.800 mini experiments so the idea is that you

00:40:02.800 encapsulate in a very small droplet a

00:40:04.640 few cells with a drug and a barcode so

00:40:07.280 you want to know what you put in each of

00:40:09.440 these little droplets so it's like mini

00:40:11.599 experiments and because they're so small

00:40:14.079 you need first very small amount of

00:40:15.839 material per experiment

00:40:17.760 and second you you you can automatize it

00:40:20.960 no it was if microfluidics and robots

00:40:23.040 this is done super fast and

00:40:25.440 relatively easily

00:40:28.000 and

00:40:28.960 so we did this

00:40:33.119 as our case proof of concept using

00:40:35.680 pancreatic tumor biopsies showing that

00:40:37.440 from patient biopsy you can do drug

00:40:39.839 screenings and then

00:40:41.280 you can use them to to to look at the

00:40:43.599 effect of different drugs and which one

00:40:45.280 induce apoptosis so which one kills the

00:40:47.040 tumors

00:40:48.319 and then also

00:40:49.760 we use the same dynamic models the logic

00:40:51.599 models i told you before to take this

00:40:53.760 data and and to build

00:40:56.319 uh for each of the pages a dynamic model

00:40:58.400 and again use the model to predict drug

00:41:00.640 combinations uh

00:41:02.400 that we could not validate in the

00:41:04.240 patients but we could validate in cell

00:41:06.560 lines or in mouse models

00:41:10.960 yeah so i went a bit fast in this last

00:41:12.640 part but if you want we can talk more

00:41:14.480 about dynamic modeling but

00:41:16.240 i just wanted to show you this other

00:41:17.680 part of the work in the lab and first

00:41:21.839 thank the people in our lab

00:41:24.880 so

00:41:26.480 from what i saw you this omnipot this

00:41:28.560 all knowledge is mostly driven by

00:41:30.160 dentists today

00:41:31.520 they already ended this multi-home

00:41:33.440 integration with cosmos

00:41:35.599 martin did the example the multiple

00:41:37.680 sclerosis federica the one the last one

00:41:39.920 was the microfluidics and jovan and

00:41:42.079 ricardo do a lot of the special work

00:41:46.240 we have a lot of great collaborators i

00:41:48.000 mentioned them along the way um

00:41:50.720 and our founders

00:41:52.240 we are always looking for talented

00:41:54.160 students and postdocs if you like this

00:41:56.079 type of things or you have friends who

00:41:57.839 are looking for positions

00:41:59.680 and i'll just in the last minute i'll

00:42:01.599 try to summarize but

00:42:04.240 i hope i i

00:42:06.480 managed to to

00:42:08.160 explain you today

00:42:09.599 so first

00:42:10.800 so machine learning

00:42:13.119 is

00:42:13.839 a very powerful toolbox

00:42:15.760 that can be applied in the context of

00:42:17.359 personalized medicine to analyze all

00:42:19.040 these large omics data

00:42:21.040 but i think there still is

00:42:22.800 a lot of work to be done and we think

00:42:24.880 that in this context it's very helpful

00:42:26.960 to use this biological knowledge to help

00:42:29.200 the machine learning

00:42:31.599 in particular explain this idea of the

00:42:33.440 footprint so how

00:42:36.240 specific

00:42:38.000 changes downstream can tell you about

00:42:40.079 what happened in your key process of

00:42:41.680 interest more upstream

00:42:43.680 i very briefly mentioned that in the

00:42:45.520 field of single cell and spatial

00:42:46.960 technologies there are

00:42:48.480 new opportunities for this type of

00:42:49.760 approaches but they're also challenges

00:42:52.160 and and at the end

00:42:54.240 how we think there is value in combining

00:42:56.400 this machine learning with more dynamic

00:42:57.920 models with

00:42:59.200 a models that capture how system over

00:43:01.839 time evolves under perturbations and so

00:43:03.920 on

00:43:04.640 and in summary what we do in the in our

00:43:06.960 group is to take biological knowledge

00:43:08.960 from databases this knowledge is

00:43:10.560 typically generic so you don't know how

00:43:13.920 much of that is relevant for your liver

00:43:16.480 your kidney cancer what not

00:43:18.640 but then by training it to data that is

00:43:20.640 more specific to a particular organ a

00:43:23.520 particular disease context we can make

00:43:25.839 these spacing specific models

00:43:27.839 and then these models we use them to

00:43:29.280 understand what's going on get

00:43:30.400 mechanistic insight into disease

00:43:32.640 and also to predict the effect of new

00:43:34.319 therapies

00:43:35.599 and with that yeah i'm

00:43:38.480 happy to take any questions and thank

00:43:39.920 you for listening