Everyone, thank you, from the Cancer Center leadership for giving me this opportunity to share my latest work. I have been working my entire research career, for almost 15 years, on cancer. But the presentation I’m giving today, it’s not much about cancer and not much about the single cell analysis I have been working on for almost 10 years. This is something we haven’t published, it just came out in my lab. I’m happy to hear feedback from you guys. So, I think that largely the anomaly in the omics area recently is, you really don’t have the spatial information anymore. So the spatial omics kind of came out, or emerged, to address this challenge. Over the past couple years, I think largely, you’ll see many different technologies, but largely, they are all based on just FISH. The more specific and more precise FISH, being a single molecule level FISH. So the shortcomings here, using FISH is, it’s difficult, even my lab work and technology, I just cannot do it.
This requires very advanced imaging technology, single-molecule fluorescence. You need to image over some time for a very sort of high volume and genome-scale data you want to collect from one sample, you probably need to image over days, repeatedly, to get this sort of large number of genes analyzed on the same sample. And also, that’s not a sort of unbiased genome-scale, you really need to know the sequence you want to analyze. And also, so far, I think no one else talks about spatial omics and another terminology people use in this field is this spatial transcriptomics. It’s not so obvious, how you can extend to other omics measurements using FISH. So I think the latest breakthrough came out actually this year, the two papers published, I think one just came out last week in Nature Methods, another paper a couple of months ago in Science, to really use the power of Next Generation Sequencing for spatial omics mapping, or spatial transcriptome mapping. So an approach they took actually is quite similar.
So they create sort of a barcoded surface using the packed beads. So whoever working in this space probably know no matter text genomics on the DropSeq technology, you need a DNA barcode beads. So each bead has this thing, the DNA barcode, or whether or not they are from the same cell. They’re basically packing the beads on a monolayer on a glass slide. And they need to decode the beads, they need to know which bead has what sequence. This decoding process was done by either SOLiD sequencing, or again, very much like FISH, you do repeated cell hybridization and imaging to decode the beads. That is a very tedious process as well. But afterwards, you get a pretty good preferential capture in the proximity of the specific bead. It should be captured only by that bead, but I don’t think the lateral sort of diffusion can be really avoided. But at least they saw a pretty good preferential capture.
of the messengers from the adjacent cells.

I think this technology published or released in Science, demonstrate you can do 10 micron resolution spatial mapping of mRNA transcriptome by sequencing.

And this paper came out last week demonstrating you can actually use even smaller beads, like two micron beads, to further sort of reduce the pixel size and increase the resolution. But two microns really (mumbles), the data analysis becomes even more complicated. And it turns out there have to be multiple beads to get a quality image.

So interesting, when we visited their data, we found although they can see sort of anatomic or histological structure of different cells in a tissue, but it is almost impossible to visualize individual genes because the number of genes they can detect per pixel is extremely sparse, about like 100, 200 genes per spot. If you tried to image on individual genes across on pixel’s entire tissue, the data totally is sort of not that meaningful at all.

So what we can do is fundamentally different, I’m not about to say too much in the technical details,
but this is totally different. We don’t use beads and we just need a bunch of reagents with this device. And although we have been working on microfluids for years, but I don’t like complicate microfluids like you guys. So this device, basically, you just place PDMS on top of your tissue and your clamp it, that’s it. That’s everything you need to do to deal with the microfluids. Afterwards, you just pipette your reagent to the host. So in the data, the validation data we have shown is we use sort of pan-messenger RNA FISH to visualize the individual tissue pixels we eventually are able to sequence with the spatial resolution. So we found we can get a very nice 10 micron pixel, as shown here if you zoom in. And then also interestingly, we saw sort of in the tissues after we process with our barcoding strategy, our barcoding approach, show some topological features. Even under optical microscope you can see where your individual pixels are located on the tissue. And worth noting, so this is sort of exactly the same tissue
we're gonna take for sequencing, rather than the previous methods that always have to compare to an adjacent tissue. They are not able to get any good image from the same tissue at all. Also, the tissue sample we analyzed, they are just a formaldehyde-fixed tissue sample on a glass slide. So if you have a freezer of those samples banked in your freezer, we can look at those samples as well. We don’t have to use sort of frozen tissue block and a fresh section to put on our slide. So we did some quantitative analysis of how many cells we can get per pixel, using this DAPI staining. And also, we were also concerned whether or not each pixel is distinct molecular barcode, we can put on or some sort of diffusion between the pixel that might cause cross contamination. We quantified a diffusion distance, we found it using the fluorophores basically. So we found the diffusion distance is actually just one micro meter, which suggests we can potentially further reduce the pixel size and increase the resolution
to about like two micron using our technology.

So the feature size matched the sort of the microfluid design very well.

And the number of cells we can get in the 10 micron pixel size device is about 1.7 cells.

we’re really getting close to single cell level spatial omics.

As I kinda alluded a little bit earlier, so the qualitative data, very important.

So we compared our data to the Slide-seq data published earlier this year.

So for the number of genes they can detect per pixel,

about the size, 10 micron and then the number of genes we detected by using our technology.

So really all that (mumbles) increase, in terms of how many genes,

how many transcripts we can detect.

About two years, three years ago,

similar technology, sort of barcoded surface,

basically capture of messenger RNAs for spatial transcriptome mapping

was published in Science 2016.

But that was very low spatial resolution,

about 150 micron, but in that data,

when you look at how many genes they can detect,

that’s about the same as what we can do.

But the resolution is much, much lower.

Or if you calculated sort of an area per pixel,
it’s 100 times larger than what we have.

So I was very excited about this sort of data quality,

which really enabled on the following slides,

we can really visualize individual genes rather than using extremely sophisticated informatics.

to identify genes just to visualize the different cells types.

We can actually interrogate every single genes across the entire tissue map.

So when we first start with this, I’m extremely excited about tumor micro environment feature.

But we decide to pick something that’s well characterized,

people know what cell types are there.

So we used mouse embryo in the earlier stage of organogenesis, it’s about 10 days.

We were able to map out, actually, I wanna talk about a messenger RNA, actually, we can do also about 22 types of protein simultaneously mapped out using the same barcoding strategy.

microfluid barcoding strategy.

Showing here, is sort of pan-messenger RNA, but done by sequencing.

So you can see actually the intensity of the total signal of the messenger does reflect (mumbles) in the tissue on the embryo slides.
And here, this average signal of over 22 proteins
we're able to look at as a panel. That doesn’t really correlate that very well,
but I think that makes sense,
because you’re not looking at it globally on all proteins,
but the sub panel, it really depends
on what proteins you put in your panel.
Then we did a cluster analysis.
When we look at single cells, we used tSNE,
but here, it does make sense you have to use tSNE.
because you know exactly where the spatial location
of every single pixel is.
But the computational algorithm for clustering
is identical, so, but after clustering,
we just put it back on the tissue histological.
The spatial map, we see sort of about eight clusters over here.
And they pretty much match the anatomic annotation
we got from the eMouseAtlas.
And more interestingly, I think in the eMouseAtlas
you’re now able to kind of resolve a wide stripe the tissue here,
but we saw a very distinct stripe of sort of cell type.
We’re still unclear what those cells are,
but probably associated with the mouse
sort of major aorta around the area.

As I mentioned, we are able to visualize individual genes

or individual proteins at a very high quality across the entire tissue section.

Showing here a couple of genes and couple of proteins.

And overall, I think the protein signal way higher, it’s not a big surprise, this is because you measure

only like 22 rather than genome scale.

But when you compare, you see consistence,

you see concordance and also discordance between the gene and proteins

people have seen over and over.

And very interestingly, when we look at Ep-CAM,

it’s a very nice concordance

between the protein and messenger RNA

in the EpCAM expression right here.

And this one, I think, this is a microvascular tissue,

microvascular tissue already developed

in mouse embryo at this stage all over the whole body,

we can see they are expressed everywhere,

but we don’t see a distinct structure at this resolution,

because this resolution is about 50 micron, not 10 micron.

I will get down to the high resolution data later.

And then we did a sort of validation
to compare our data to immunofluorescence staining for several selected genes. And this vasculature, again, you see extensive everywhere. You see EpCAM exactly the same pattern as we saw using sequencing. So just a couple of those locations showing the expression of the EpCAM. And another validation is we've done the sequencing data and the paper published earlier this year by Jason Du, from the University of Washington, they used single cell sequencing to map out several mouse embryos over different stages. And then you can basically do a tissue, a sort of sample tSNE, or sample UMap, this is not a single cell UMAP, but a sample UMap. So we found a four sample sequence actually mapped very well to this sort of differential or developmental trajectory. So in here, from their data, this is sort of the E9.5 and that this is E10.5 and we are right in the middle. Those are kind of a little bit later stages of the developmental mouse embryos. And then we used a little bit higher resolution to look at the embryonic brain. This is about the entire brain
and a little bit other tissues in the head and the neck.
And also, this one, we didn’t know what that is,
but after data analysis, we found that actually it’s a piece of the heart.
And what we see from the protein
and from the messenger RNA is,
again, the messenger RNA atlas does reflect in the tissue histology very well.
And the protein now, is much higher resolution of 25 micron, you do see some sort of correlation
between tissue histology and protein expression atlas,
but not as so distinct compared to the messenger RNA.
So we were able to visualize individual proteins essentially,
here are four of them, I think are very interesting.
Again, EPCAM, this is a very high resolution,
you can see very tight clusters of EpCAM expression
in specific tissue regions right here and here
and there’s two or three or four.
And the microvasculature, we can see the microvasculature
by sequencing very well.
And when you go to look on the tissue histology,
I just cannot identify where the microvasculature
307 00:16:12.798 --> 00:16:16.440 are located based on the tissue histology.
308 00:16:16.440 --> 00:16:18.890 And the two other proteins, very interesting as well.
309 00:16:18.890 --> 00:16:21.940 This MAdCAM, we found it is a highly enriched
310 00:16:21.940 --> 00:16:25.550 in part of the forebrain, but not entire forebrain.
311 00:16:25.550 --> 00:16:29.350 And we see in CD63 it’s widely implicated
312 00:16:29.350 --> 00:16:31.580 in the early stage mouse development.
313 00:16:31.580 --> 00:16:36.580 It’s kinda anti-correlated with MAdCAM in other areas,
314 00:16:36.840 --> 00:16:38.170 so we kinda put them together,
315 00:16:38.170 --> 00:16:43.170 you can see their relative correlation each other.
316 00:16:43.768 --> 00:16:46.584 So, again, this technology where we want to validate
317 00:16:46.584 --> 00:16:50.530 to make sure what we saw using sequencing
318 00:16:50.530 --> 00:16:53.670 does match immunofluorescence staining.
319 00:16:53.670 --> 00:16:57.001 So this is from sequencing, this is from sequencing,
320 00:16:57.001 --> 00:17:00.710 this is about microvasculature, this is EpCAM,
321 00:17:00.710 --> 00:17:03.367 this immuno staining, you’ll se almost a perfect match.
322 00:17:03.367 --> 00:17:08.367 I was very surprised, this is really a perfect match
323 00:17:08.550 --> 00:17:11.480 of distinct clusters right here, a little bit right here
324 00:17:11.480 --> 00:17:14.050 from immuno staining and we can pick up.
325 00:17:14.050 --> 00:17:18.730 It’s only a few, so one single pixel layer thickness
326 00:17:18.730 --> 00:17:20.530 we can pick up very well.
327 00:17:20.530 --> 00:17:22.040 And so now here, you can see
328 00:17:22.040 --> 00:17:26.950 those microvascular network using immuno staining,
329 00:17:26.950 --> 00:17:31.763 which was also observed in our sequencing map atlas.
330 00:17:33.070 --> 00:17:34.300 So I got an interested in,
331 00:17:34.300 --> 00:17:37.750 this particular protein called MAdCAM and asked my poster
332 00:17:37.750 --> 00:17:40.790 to do some differential gene expression sort of.
333 00:17:40.790 --> 00:17:43.790 But the MAdCAM transcripts, it’s difficult to see
334 00:17:45.325 --> 00:17:47.690 the sort of spatially distinct expression,
335 00:17:47.690 --> 00:17:50.450 but in the protein data, you can it see very well.
336 00:17:50.450 --> 00:17:52.670 Then we decided to use our sort of high quality spatial protein data
337 00:17:52.670 --> 00:17:55.350 to guide the differential gene expression
338 00:17:55.350 --> 00:17:56.902 across the entire transcriptome
339 00:17:56.902 --> 00:17:58.563 for different tissue reagents.
340 00:18:00.510 --> 00:18:03.400 So in this case, we’re looking at MAdCAM-positive
341 00:18:03.400 --> 00:18:06.550 and a MAdCAM-negative and mapped out the top ranked genes
342 00:18:06.550 --> 00:18:08.270 for MAdCAM-positive region.
343 00:18:08.270 --> 00:18:11.800 This is still ongoing, since I’m still in the stages
344 00:18:11.800 --> 00:18:14.040 of learning developmental pathology,
345 00:18:14.040 --> 00:18:16.910 but what we can see some interesting features.
346 00:18:16.910 --> 00:18:19.143 But in the negative region, clearly,
347 00:18:19.143 --> 00:18:22.210 so this is the heart, turns out, this is kind of heart,
348 00:18:22.210 --> 00:18:25.150 kind of microtube associated proteins.
349 00:18:25.150 --> 00:18:26.960 And this is interesting thing,
350 00:18:26.960 --> 00:18:29.800 we don’t really see this protein showed up extensively
in the brain, but some how look like in this local area.
And I have no idea what that is, but later we figure out that’s actually the eye, here.
And then we decided to do even higher resolution,
which is a 10 micron resolution mapping of a particular region of the brain.
And again, we had no idea where to map now, we just randomly placed our device on top and then mapped out this region.
And the red color actually real data, this basically just pan-messenger RNA data.
You can see the signal relatively uniformed and not perfect, but that’s totally okay,
just like when we do single cellular sequencing, we always do normalizations.
Then that gives you, as long as your sequencing quality,
sequencing data quality, number of genes you can read out (mumbles) genes, you can always do normalization
and compare across different pixels.
And as I told you, actually, we can see in the same tissue
sort of after the barcoding and before the sequencing,
we can even just under optical microscope, we can see individual pixels over here.
And then when my poster showed me this image, it’s okay, you got a key wide fiber over there
very likely, because we saw this when we used microfluids before. And I thought that’s unfortunate but anyhow, let’s go ahead and process the sequencing data. But turns out that’s not a key wide fiber that’s really a very thin layer, actually it’s a single cell layer of melanocytes lining a round the eye field. At this stage, the eye field actually, it’s a very, very early stage only, called the eye vesicle an even no optical caps, it’s the optical vesicle. So we can see, very distinctly, a group of genes strongly enriched inside the eye and also lining around the eye, optical vesicle. And then when we put them together, a little bit more structures you can see. For example in Pax6 enriched pretty much in an entire eye field but also in this region is optical nerve fiber. But here this protein, only expressed in the eye, but also other tissue type but not so much optical fiber. You can see this very well at a very high resolution, it’s really about a single cell resolution. So, okay, when you look at it carefully, you see some yellow spots over here. That means the Pax6 and the Pmel are actually co-expressed
in those kinda melanoblast cells but this one is not.

The Six6 is not expressed, only within the eye, optical vesicle. If you further zoom in, you can see the sort of gene expression within the vesicle and also individual pixels, every little square here.

So we can overlay the tissue image and the transcriptome data. So we noticed one gene which is strongly enriched right here, very strongly differential expression spatially. We’re all curious what this gene does.

We did sort of, this time they’re still global, gene differential analysis. We saw only top ranked genes and these two showed up.

But we found their functioning on a top ranked pathways, to some degree, okay, except those ones, are mutually exclusive. And then later we realized that has never been observed before, I don’t have sort of last year’s data to support. But it seems like those cells sort of characterized by this particular gene, later on are gonna determine the development of the lens.

And those cells, even at this stage, you don’t see any morphological difference, they already predetermined to develop
the retina and the photo receptor cells.

And then we were able to basically just put out those pictures obviously and compare it to those to perform a differential gene expression analysis.

And another surprise, now this gene just showed up extremely differentially expressed. But we see many other genes that were very interesting.

We still try to look into the details. So they are kinda enriched on the left side. Eventually, very likely, they will contribute to the photo receptor cell development.

Okay, so even though we’re able to visualize individual genes, we don’t have to use the gene cell enrichment to identify different tissue types, but we had a challenge in particular in this kind of eye field region, due to our lack of knowledge in mouse embryonic development.

But it’ll be great if some computational pipeline can automatically identify different features.

That’s what we demonstrate as well. So using this automatic automated feature identification pipeline, we were able to identify actually 20 different features in this very small region of the brain.
around the eye field. I just will show you some of those, you can see not just the eye, actually you can see very already development of the ear based on the sort of gene expression, but histologically, you cannot see any difference at all. But we also look at entire mouse embryo the E10. We’re able to identify about 20 different features. But we’re asking, so if at later stage many other organs begin to develop, whether or not this pipeline can identify many more tissue features or tissue subtypes. That turns out that that’s right. And using E12, we’re now able to cover entire embryo actually just the lower part of the body, we identify about 40 different features already. So this is a very high resolution as well. Okay, I’m gonna just summarize back to my sort of, the main interest in cancer. So I believe this enabling platform, we demonstrate can do protein and the transcripts. But actually, in my lab, another post I’m working on, so spatial, high spatial resolution epigenomics. I believe we can do high res, high spatial resolution ATAC,
high spatial resolution CHIP-seq.

And the application is extremely broad and the cancer is put right in the middle because that’s really my main focus.

I will like to thank people in my lab who work on this and thank you for your attention.