WEBVTT

1 00:00:00.000 --> 00:00:00.833 <v Rong>Everyone, thank you,</v>

2 00:00:00.833 --> 00:00:03.390 from the Cancer Center leadership for giving me

3 00:00:03.390 --> 00:00:07.120 this opportunity to share my latest work.

4 00:00:07.120 --> 00:00:11.530 I have been working my entire research career,

5 00:00:11.530 --> 00:00:14.900 for almost 15 years, on cancer.

6 00:00:14.900 --> 00:00:19.210 But the presentation I'm giving today,

7 00:00:19.210 --> 00:00:20.550 it's not much about cancer

8 00:00:20.550 --> 00:00:22.930 and not much about the single cell analysis

9 00:00:22.930 --> 00:00:25.740 I have been working on for almost 10 years.

10 00:00:25.740 --> 00:00:27.370 This is something we haven't published,

11 00:00:27.370 --> 00:00:28.900 it just came out in my lab.

12 00:00:28.900 --> 00:00:32.430 I'm happy to hear feedback from you guys.

13 00:00:32.430 --> 00:00:37.200 So, I think that largely the anomaly

14 00:00:37.200 --> 00:00:39.559 in the omics area recently is,

15 00:00:39.559 --> 00:00:43.120 people can do single cell omics

16 00:00:43.120 --> 00:00:47.550 and multi-omics to understand tumor heterogenetics,

17 00:00:49.200 --> 00:00:53.000 but you really don't have the spatial information anymore.

18 00:00:53.000 --> 00:00:56.900 So the spatial omics kind of came out, or emerged,

19 00:00:56.900 --> 00:00:58.890 to address this challenge.

20 00:00:58.890 --> 00:01:01.356 Over the past couple years, I think largely,

21 00:01:01.356 --> 00:01:03.740 you'll see many different technologies,

22 00:01:03.740 --> 00:01:07.140 but largely, they are all based on just FISH.

23 00:01:07.140 --> 00:01:10.380 The more specific and more precise FISH,

24 00:01:10.380 --> 00:01:12.273 being a single molecule level FISH.

25 00:01:13.291 --> 00:01:18.100 So the shortcomings here, using FISH is,

26 00:01:18.100 --> 00:01:21.550 it's difficult, even my lab work and technology,

27 00:01:21.550 --> 00:01:22.860 I just cannot do it.

28 00:01:22.860 --> 00:01:26.890 This requires very advanced imaging technology,

29 00:01:26.890 --> 00:01:28.710 single-molecule fluorescence.

30 00:01:28.710 --> 00:01:31.530 You need to image over some time

31 00:01:31.530 --> 00:01:34.270 for a very sort of high volume

32 00:01:34.270 --> 00:01:38.220 and genome-scale data you want to collect from one sample,

33 00:01:38.220 --> 00:01:41.860 you probably need to image over days, repeatedly,

34 00:01:41.860 --> 00:01:46.160 to get this sort of large number of genes

35 00:01:46.160 --> 00:01:49.290 analyzed on the same sample.

36 00:01:49.290 --> 00:01:53.052 And also, that's not a sort of unbiased genome-scale,

37 00:01:53.052 --> 00:01:57.950 you really need to know the sequence you want to analyze.

38 00:01:57.950 --> 00:02:01.720 And also, so far, I think no one else talks about

39 00:02:01.720 --> 00:02:04.700 spatial omics and another terminology

40 00:02:04.700 --> 00:02:09.600 people use in this field is this spatial transcriptomics.

41 00:02:09.600 --> 00:02:11.830 It's not so obvious,

42 00:02:11.830 --> 00:02:16.703 how you can extend to other omics measurements using FISH.

43 00:02:18.180 --> 00:02:21.300 So I think the latest breakthrough

44 00:02:21.300 --> 00:02:24.370 came out actually this year,

45 00:02:24.370 --> 00:02:26.340 the two papers published, I think one

46 00:02:26.340 --> 00:02:29.430 just came out last week in Nature Methods,

47 00:02:29.430 --> 00:02:33.809 Another paper a couple of months ago in Science,

48 00:02:33.809 --> 00:02:38.809 to really use the power of Next Generation Sequencing

49 00:02:39.060 --> 00:02:40.770 for spatial omics mapping,

50 00:02:40.770 --> 00:02:42.780 or spatial transcriptome mapping.

51 00:02:42.780 --> 00:02:47.570 So an approach they took actually is quite similar.

52 00:02:47.570 --> 00:02:50.930 So they create sort of a barcoded surface

53 00:02:50.930 --> 00:02:54.150 using the packed beads.

54 00:02:54.150 --> 00:02:56.150 So whoever working in this space

55 00:02:56.150 --> 00:02:58.970 probably know no matter text genomics on

56 00:03:01.460 --> 00:03:05.760 the DropSeq technology, you need a DNA barcoder beads.

57 00:03:05.760 --> 00:03:08.740 So each bead has this thing, the DNA barcode,

58 00:03:08.740 --> 00:03:13.050 to really tell you which messenger is from which cell,

59 00:03:13.050 --> 00:03:15.730 or whether or not they are from the same cell.

60 00:03:15.730 --> 00:03:17.040 They're basically packing the beads

61 00:03:17.040 --> 00:03:20.450 on a monolayer on a glass slide.

62 00:03:20.450 --> 00:03:21.457 And they need to decode the beads,

63 00:03:21.457 --> 00:03:25.810 they need to know which bead has what sequence.

64 00:03:25.810 --> 00:03:27.650 So this decoding process was done

65 00:03:27.650 --> 00:03:30.300 by either SOLiD sequencing, or again,

66 00:03:30.300 --> 00:03:34.039 very much like FISH, you do repeated cell hybridization

67 00:03:34.039 --> 00:03:36.060 and imaging to decode the beads.

68 00:03:36.060 --> 00:03:39.400 That is a very tedious process as well.

69 00:03:39.400 --> 00:03:41.960 But afterwards, you get

70 00:03:41.960 --> 00:03:45.700 sort of a freshly micro-sectioned tissue sample

71 00:03:45.700 --> 00:03:49.320 and you place it on top and you lyse the tissue section

72 00:03:49.320 --> 00:03:54.320 and hopefully, the messenger is released from the cells

73 00:03:54.371 --> 00:03:58.010 in the proximity of the specific bead.

74 00:03:58.010 --> 00:04:00.250 It should be captured only by that bead,

75 00:04:00.250 --> 00:04:03.280 but I don't think the lateral sort of diffusion

76 00:04:03.280 --> 00:04:05.600 can be really avoided.

77 00:04:05.600 --> 00:04:09.410 But at least they saw a pretty good preferential capture

78 00:04:09.410 --> 00:04:12.343 of the messengers from the adjacent cells.

79 00:04:13.870 --> 00:04:17.033 I think this technology published or released in Science,

80 00:04:17.033 --> 00:04:19.900 demonstrate you can do 10 micron resolution

81 00:04:21.210 --> 00:04:26.210 spatial mapping of mRNA transcriptome by sequencing.

82 00:04:26.220 --> 00:04:29.000 And this paper came out last week

83 00:04:29.000 --> 00:04:31.492 demonstrating you can actually use even smaller beads,

84 00:04:31.492 --> 00:04:35.805 like two micron beads, to further sort of reduce

85 00:04:35.805 --> 00:04:38.265 the pixel size and increase the resolution.

86 00:04:38.265 --> 00:04:41.380 But two microns really (mumbles),

87 00:04:41.380 --> 00:04:44.387 the data analysis becomes even more complicated.

88 00:04:44.387 --> 00:04:49.387 And it turns out there have to be multiple beads

89 00:04:50.240 --> 00:04:54.120 to get a quality image.

90 00:04:54.120 --> 00:04:56.610 So interesting, when we visited their data,

91 00:04:56.610 --> 00:05:00.654 we found although they can see sort of anatomic or

92 00:05:00.654 --> 00:05:04.500 histological structure of different cells in a tissue,

93 00:05:04.500 --> 00:05:07.781 but it is almost impossible to visualize individual genes

94 00:05:07.781 --> 00:05:10.330 because the number of genes they can detect per pixel

95 00:05:10.330 --> 00:05:15.330 is extremely sparse, about like 100, 200 genes per spot.

96 00:05:15.580 --> 00:05:18.830 If you tried to image on individual genes

97 00:05:18.830 --> 00:05:21.810 across on pixel's entire tissue,

98 00:05:21.810 --> 00:05:26.100 the data totally is sort of not that meaningful at all.

99 00:05:26.100 --> 00:05:30.270 So what we can do is fundamentally different,

100 00:05:30.270 --> 00:05:33.583 I'm not about to say too much in the technical details,

101 00:05:33.583 --> 00:05:35.680 but this is totally different.

102 00:05:35.680 --> 00:05:38.100 We don't use beads and we just need

103 00:05:38.100 --> 00:05:40.830 a bunch of reagents with this device.

104 00:05:40.830 --> 00:05:44.010 And although we have been working

105 00:05:44.010 --> 00:05:45.350 on microfluids for years,

106 00:05:45.350 --> 00:05:50.090 but I don't like complicate microfluids like you guys.

107 00:05:50.090 --> 00:05:53.490 So this device, basically, you just place PDMS

108 00:05:53.490 --> 00:05:55.770 on top of your tissue and your clamp it, that's it.

109 00:05:55.770 --> 00:05:58.360 That's everything you need to do

110 00:05:58.360 --> 00:05:59.880 to deal with the microfluids.

111 00:05:59.880 --> 00:06:03.940 Afterwards, you just pipette your reagent to the host.

112 00:06:03.940 --> 00:06:07.460 So in the data, the validation data we have shown

113 00:06:07.460 --> 00:06:11.370 is we use sort of pan-messenger RNA FISH

114 00:06:11.370 --> 00:06:14.430 to visualize the individual tissue pixels

115 00:06:14.430 --> 00:06:17.750 we eventually are able to sequence

116 00:06:17.750 --> 00:06:21.130 with the spatial resolution.

117 00:06:21.130 --> 00:06:24.250 So we found we can get a very nice 10 micron pixel,

118 00:06:24.250 --> 00:06:26.340 as shown here if you zoom in.

119 00:06:26.340 --> 00:06:28.616 And then also interestingly,

120 00:06:28.616 --> 00:06:30.820 we saw sort of in the tissues

121 00:06:30.820 --> 00:06:33.843 after we process with our barcoding strategy,

122 00:06:33.843 --> 00:06:38.432 our barcoding approach, show some topological features.

123 00:06:38.432 --> 00:06:40.930 Even under optical microscope

124 00:06:40.930 --> 00:06:43.530 you can see where your individual pixels

125 00:06:43.530 --> 00:06:45.280 are located on the tissue.

126 00:06:45.280 --> 00:06:49.770 And worth noting, so this is sort of exactly the same tissue

127 00:06:49.770 --> 00:06:52.960 we're gonna take for sequencing,

128 00:06:52.960 --> 00:06:56.150 rather than the previous methods

129 00:06:56.150 --> 00:06:59.930 that always have to compare to an adjacent tissue.

130 00:06:59.930 --> 00:07:02.030 They are not able to get any good image

131 00:07:02.030 --> 00:07:03.700 from the same tissue at all.

132 00:07:03.700 --> 00:07:07.500 Also, the tissue sample we analyzed,

133 00:07:07.500 --> 00:07:11.920 they are just a formaldehyde-fixed tissue sample

134 00:07:11.920 --> 00:07:12.950 on a glass slide.

135 00:07:12.950 --> 00:07:16.180 So if you have a freezer of those samples banked

136 00:07:16.180 --> 00:07:21.020 in your freezer, we can look at those samples as well.

137 00:07:21.020 --> 00:07:23.895 We don't have to use sort of frozen tissue block

138 00:07:23.895 --> 00:07:28.895 and a fresh section to put on our slide.

139 00:07:31.450 --> 00:07:33.300 So we did some quantitative analysis

140 00:07:33.300 --> 00:07:37.170 of how many cells we can get per pixel,

141 00:07:37.170 --> 00:07:39.330 using this DAPI staining.

142 00:07:39.330 --> 00:07:44.020 And also, we were also concerned whether or not

143 00:07:44.020 --> 00:07:46.800 each pixel is distinct molecular barcode,

144 00:07:46.800 --> 00:07:50.700 we can put on or some sort of diffusion between the pixel

145 00:07:51.810 --> 00:07:53.613 that might cause cross contamination.

146 00:07:53.613 --> 00:07:56.150 We quantified a diffusion distance,

147 00:07:56.150 --> 00:07:58.710 we found it using the fluorophores basically.

148 00:07:58.710 --> 00:08:00.159 So we found the diffusion distance

149 00:08:00.159 --> 00:08:03.880 is actually just one micro meter,

150 00:08:03.880 --> 00:08:07.410 which suggests we can potentially

151 00:08:07.410 --> 00:08:11.990 further reduce the pixel size and increase the resolution

152 00:08:11.990 --> 00:08:15.523 to about like two micron using our technology.

153 00:08:16.910 --> 00:08:21.110 So the feature size matched

154 00:08:21.110 --> 00:08:24.100 the sort of the microfluid design very well.

155 00:08:24.100 --> 00:08:26.510 And the number of cells we can get

156 00:08:26.510 --> 00:08:31.510 in the 10 micron pixel size device is about 1.7 cells,

157 00:08:32.030 --> 00:08:33.950 we're really getting close

158 00:08:33.950 --> 00:08:37.383 to single cell level spatial omics.

159 00:08:38.440 --> 00:08:41.690 As I kinda alluded a little bit earlier,

160 00:08:41.690 --> 00:08:44.370 so the qualitative data, very important.

161 00:08:44.370 --> 00:08:48.480 So we compared our data to the Slide-seq data

162 00:08:48.480 --> 00:08:49.960 published earlier this year.

163 00:08:49.960 --> 00:08:52.407 So for the number of genes they can detect per pixel,

164 00:08:52.407 --> 00:08:54.460 about the size, 10 micron

165 00:08:54.460 --> 00:08:56.592 and then the number of genes we detected

166 00:08:56.592 --> 00:08:57.580 by using our technology.

167 00:08:57.580 --> 00:09:00.910 So really all that (mumbles) increase,

168 00:09:00.910 --> 00:09:02.590 in terms of how many genes,

169 00:09:02.590 --> 00:09:04.810 how many transcripts we can detect.

170 00:09:04.810 --> 00:09:07.190 About two years, three years ago,

171 00:09:07.190 --> 00:09:11.230 similar technology, sort of barcoded surface,

172 00:09:11.230 --> 00:09:12.900 basically capture of messenger RNAs

173 00:09:12.900 --> 00:09:14.870 for spatial transcriptome mapping

174 00:09:14.870 --> 00:09:18.540 was published in Science 2016.

175 00:09:18.540 --> 00:09:21.040 But that was very low spatial resolution,

176 00:09:21.040 --> 00:09:23.558 about 150 micron, but in that data,

177 00:09:23.558 --> 00:09:26.200 when you look at how many genes they can detect,

178 00:09:26.200 --> 00:09:28.203 that's about the same as what we can do.

179 00:09:29.220 --> 00:09:32.150 But the resolution is much, much lower.

180 00:09:32.150 --> 00:09:35.303 Or if you calculated sort of an area per pixel,

181 00:09:35.303 --> 00:09:38.700 it's 100 times larger than what we have.

182 00:09:38.700 --> 00:09:43.114 So I was very excited about this sort of data quality,

183 00:09:43.114 --> 00:09:46.420 which really enabled on the following slides,

184 00:09:46.420 --> 00:09:48.490 we can really visualize individual genes

185 00:09:48.490 --> 00:09:52.354 rather than using extremely sophisticated informatics

186 00:09:52.354 --> 00:09:57.100 to identify genes just to visualize

187 00:09:57.100 --> 00:09:58.620 the different cells types.

188 00:09:58.620 --> 00:10:02.430 We can actually interrogate every single genes

189 00:10:02.430 --> 00:10:04.603 across the entire tissue map.

190 00:10:06.430 --> 00:10:09.843 So when we first start with this,

191 00:10:11.270 --> 00:10:12.670 I'm extremely excited about

192 00:10:12.670 --> 00:10:14.023 tumor micro environment feature.

193 00:10:14.023 --> 00:10:16.017 But we decide to pick something

194 00:10:16.017 --> 00:10:17.560 that's well characterized,

195 00:10:17.560 --> 00:10:19.610 people know what cell types are there.

196 00:10:19.610 --> 00:10:21.620 So we used mouse embryo

197 00:10:21.620 --> 00:10:26.620 in the earlier stage of organogenesis, it's about 10 days.

198 00:10:26.710 --> 00:10:30.380 We were able to map out, actually, I wanna talk about

199 00:10:30.380 --> 00:10:33.280 a messenger RNA, actually, we can do also

200 00:10:33.280 --> 00:10:38.280 about 22 types of protein simultaneously mapped out

201 00:10:38.350 --> 00:10:40.590 using the same barcoding strategy,

202 00:10:40.590 --> 00:10:42.420 microfluid barcoding strategy.

203 00:10:42.420 --> 00:10:45.170 Showing here, is sort of pan-messenger RNA,

204 00:10:45.170 --> 00:10:46.430 but done by sequencing.

205 00:10:46.430 --> 00:10:48.460 So you can see actually the intensity

206 00:10:48.460 --> 00:10:51.970 of the total signal of the messenger

207 00:10:51.970 --> 00:10:56.970 does reflect (mumbles) in the tissue on the embryo slides.

208 00:10:58.250 --> 00:11:02.490 And here, this average signal of over 22 proteins

209 00:11:02.490 --> 00:11:06.380 we're able to look at as a panel.

210 00:11:06.380 --> 00:11:08.570 That doesn't really correlate that very well,

211 00:11:08.570 --> 00:11:10.480 but I think that makes sense,

212 00:11:10.480 --> 00:11:13.960 because you're not looking at it globally on all proteins,

213 00:11:13.960 --> 00:11:16.330 but the sub panel, it really depends

214 00:11:16.330 --> 00:11:18.737 on what proteins you put in your panel.

215 00:11:18.737 --> 00:11:21.040 Then we did a cluster analysis.

216 00:11:21.040 --> 00:11:24.650 When we look at single cells, we used tSNE,

217 00:11:24.650 --> 00:11:26.860 but here, it does make sense you have to use tSNE

218 00:11:26.860 --> 00:11:31.170 because you know exactly where the spatial location

219 00:11:31.170 --> 00:11:33.381 of every single pixel is.

220 00:11:33.381 --> 00:11:36.650 But the computational algorithm for clustering

221 00:11:36.650 --> 00:11:39.360 is identical, so, but after clustering,

222 00:11:39.360 --> 00:11:42.420 we just put it back on the tissue histological.

223 00:11:42.420 --> 00:11:45.820 The spatial map, we see sort of

224 00:11:47.860 --> 00:11:49.850 about eight clusters over here.

225 00:11:49.850 --> 00:11:54.850 And they pretty much match the anatomic annotation

226 00:11:54.870 --> 00:11:56.683 we got from the eMouseAtlas.

227 00:11:58.080 --> 00:12:00.464 And more interestingly, I think in the eMouseAtlas

228 00:12:00.464 --> 00:12:03.215 you're now able to kind of resolve

229 00:12:03.215 --> 00:12:06.129 a wide stripe the tissue here,

230 00:12:06.129 --> 00:12:10.236 but we saw a very distinct stripe of sort of cell type.

231 00:12:10.236 --> 00:12:14.800 We're still unclear what those cells are,

232 00:12:14.800 --> 00:12:18.311 but probably associated with the mouse

233 00:12:18.311 --> 00:12:21.853 sort of major aorta around the area.

234 00:12:23.700 --> 00:12:26.800 As I mentioned, we are able to visualize individual genes

235 00:12:26.800 --> 00:12:30.848 or individual proteins at a very high quality

236 00:12:30.848 --> 00:12:34.519 across the entire tissue section.

237 00:12:34.519 --> 00:12:39.440 Showing here a couple of genes and couple of proteins.

238 00:12:39.440 --> 00:12:42.580 And overall, I think the protein signal way higher,

239 00:12:42.580 --> 00:12:45.830 it's not a big surprise, this is because you measure

240 00:12:45.830 --> 00:12:48.876 only like 22 rather than genome scale.

241 00:12:48.876 --> 00:12:51.640 But when you compare, you see consistence,

242 00:12:51.640 --> 00:12:53.930 you see concordance and also discordance

243 00:12:53.930 --> 00:12:55.900 between the gene and proteins

244 00:12:55.900 --> 00:12:57.860 people have seen over and over.

245 00:12:57.860 --> 00:13:02.043 And very interestingly, when we look at EpCAM,

246 00:13:03.370 --> 00:13:05.950 it's a very nice concordance

247 00:13:05.950 --> 00:13:07.877 between the protein and messenger RNA

248 00:13:07.877 --> 00:13:10.620 in the EpCAM expression right here.

249 00:13:10.620 --> 00:13:15.620 And this one, I think, this is a microvascular tissue,

250 00:13:16.220 --> 00:13:18.016 microvascular tissue already developed

251 00:13:18.016 --> 00:13:22.100 in mouse embryo at this stage all over the whole body,

252 00:13:22.100 --> 00:13:24.618 we can see they are expressed everywhere,

253 00:13:24.618 --> 00:13:28.070 but we don't see a distinct structure at this resolution,

254 00:13:28.070 --> 00:13:31.810 because this resolution is about 50 micron, not 10 micron.

255 00:13:31.810 --> 00:13:35.160 I will get down to the high resolution data later.

256 00:13:35.160 --> 00:13:37.110 And then we did a sort of validation

257 00:13:37.110 --> 00:13:40.820 to compare our data to immunofluorescence staining

258 00:13:40.820 --> 00:13:43.070 for several selected genes.

259 00:13:43.070 --> 00:13:47.360 And this vasculature, again, you see extensive everywhere.

260 00:13:47.360 --> 00:13:50.120 You see EpCAM exactly the same pattern

261 00:13:50.120 --> 00:13:52.120 as we saw using sequencing.

262 00:13:52.120 --> 00:13:55.362 So just a couple of those locations

263 00:13:55.362 --> 00:13:59.530 showing the expression of the EpCAM.

264 00:13:59.530 --> 00:14:01.040 And another validation is

265 00:14:01.040 --> 00:14:02.790 we've done the sequencing data

266 00:14:02.790 --> 00:14:04.700 and the paper published earlier this year

267 00:14:04.700 --> 00:14:07.030 by Jason Du, from the University of Washington,

268 00:14:07.030 --> 00:14:10.370 they used single cell sequencing to map out

269 00:14:10.370 --> 00:14:13.320 several mouse embryos over different stages.

270 00:14:13.320 --> 00:14:16.830 And then you can basically do a tissue,

271 00:14:16.830 --> 00:14:20.080 a sort of sample tSNE, or sample UMap,

272 00:14:20.080 --> 00:14:23.040 this is not a single cell UMAP, but a sample UMap.

273 00:14:23.040 --> 00:14:25.100 So we found a four sample sequence

274 00:14:25.100 --> 00:14:28.320 actually mapped very well to this

275 00:14:28.320 --> 00:14:31.530 sort of differential or developmental trajectory.

276 00:14:31.530 --> 00:14:36.437 So in here, from their data, this is sort of the E9.5

277 00:14:37.680 --> 00:14:42.000 and that this is E10.5 and we are right in the middle.

278 00:14:42.000 --> 00:14:45.500 Those are kind of a little bit later stages

279 00:14:45.500 --> 00:14:50.500 of the developmental mouse embryos.

280 00:14:52.050 --> 00:14:54.670 And then we used a little bit higher resolution

281 00:14:54.670 --> 00:14:58.443 to look at the embryonic brain.

282 00:14:58.443 --> 00:15:01.180 This is about the entire brain

283 00:15:01.180 --> 00:15:04.150 and a little bit other tissues in the head and the neck.

284 00:15:04.150 --> 00:15:08.550 And also, this one, we didn't know what that is,

285 00:15:08.550 --> 00:15:11.070 but after data analysis, we found that actually

286 00:15:11.070 --> 00:15:13.260 it's a piece of the heart.

287 00:15:13.260 --> 00:15:15.776 And what we see from the protein

288 00:15:15.776 --> 00:15:17.560 and from the messenger RNA is,

289 00:15:17.560 --> 00:15:20.090 again, the messenger RNA atlas

290 00:15:20.090 --> 00:15:22.620 does reflect in the tissue histology very well.

291 00:15:22.620 --> 00:15:24.620 And the protein now, is much higher resolution

292 00:15:24.620 --> 00:15:28.410 of 25 micron, you do see some sort of correlation

293 00:15:28.410 --> 00:15:32.960 between tissue histology and protein expression atlas,

294 00:15:32.960 --> 00:15:37.960 but not as so distinct compared to the messenger RNA.

295 00:15:39.088 --> 00:15:40.640 So we were able to visualize

296 00:15:40.640 --> 00:15:42.160 individual proteins essentially,

297 00:15:42.160 --> 00:15:45.970 here are four of them, I think are very interesting.

298 00:15:45.970 --> 00:15:49.260 Again, EPCAM, this is a very high resolution,

299 00:15:49.260 --> 00:15:53.510 you can see very tight clusters of EpCAM expression

300 00:15:53.510 --> 00:15:55.590 in specific tissue regions right here and here

301 00:15:55.590 --> 00:15:58.083 and there's two or three or four.

302 00:15:58.083 --> 00:16:02.201 And the microvasculature, we can see the microvasculature

303 00:16:02.201 --> 00:16:04.010 by sequencing very well.

304 00:16:04.010 --> 00:16:06.930 And when you go to look on the tissue histology,

305 00:16:06.930 --> 00:16:09.370 or maybe I'm not pathology by training,

306 00:16:09.370 --> 00:16:12.798 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

337 00:17:52.670 --> 00:17:55.350 high quality spatial protein data

338 00:17:55.350 --> 00:17:56.902 to guide the differential gene expression

339 00:17:56.902 --> 00:17:58.563 across the entire transcriptome

340 00:17:58.563 --> 00:18:00.510 for different tissue reagents.

341 00:18:00.510 --> 00:18:03.400 So in this case, we're looking at MAdCAM-positive

342 00:18:03.400 --> 00:18:06.550 and a MAdCAM-negative and mapped out the top ranked genes

343 00:18:06.550 --> 00:18:08.270 for MAdCAM-positive region.

344 00:18:08.270 --> 00:18:11.800 This is still ongoing, since I'm still in the stages

345 00:18:11.800 --> 00:18:14.040 of learning developmental pathology,

346 00:18:14.040 --> 00:18:16.910 but what we can see some interesting features.

347 00:18:16.910 --> 00:18:19.143 But in the negative region, clearly,

348 00:18:19.143 --> 00:18:22.210 so this is the heart, turns out, this is kind of heart,

349 00:18:22.210 --> 00:18:25.150 kind of microtube associated proteins.

350 00:18:25.150 --> 00:18:26.960 And this is interesting thing,

351 00:18:26.960 --> 00:18:29.800 we don't really see this protein showed up extensively

352 00:18:29.800 --> 00:18:34.501 in the brain, but some how look like in this local area.

353 00:18:34.501 --> 00:18:37.170 And I have no idea what that is,

354 00:18:37.170 --> 00:18:40.223 but later we figure out that's actually the eye, here.

355 00:18:41.249 --> 00:18:44.050 And then we decided to do even higher resolution,

356 00:18:44.050 --> 00:18:45.980 which is a 10 micron resolution mapping

357 00:18:45.980 --> 00:18:48.587 of a particular region of the brain.

358 00:18:48.587 --> 00:18:53.020 And again, we had no idea where to map now,

359 00:18:53.020 --> 00:18:55.910 we just randomly placed our device on top

360 00:18:55.910 --> 00:18:57.670 and then mapped out this region.

361 00:18:57.670 --> 00:18:59.840 And the red color actually real data,

362 00:18:59.840 --> 00:19:02.820 this basically just pan-messenger RNA data.

363 00:19:02.820 --> 00:19:04.930 You can see the signal relatively uniformed

364 00:19:04.930 --> 00:19:07.210 and not perfect, but that's totally okay,

365 00:19:07.210 --> 00:19:08.990 just like when we do single cellular sequencing,

366 00:19:08.990 --> 00:19:10.377 we always do normalizations.

367 00:19:10.377 --> 00:19:12.529 Then that gives you, as long as your sequencing quality,

368 00:19:12.529 --> 00:19:17.058 sequencing data quality, number of genes you can read out

369 00:19:17.058 --> 00:19:19.680 (mumbles) genes, you can always do normalization

370 00:19:19.680 --> 00:19:22.680 and compare across different pixels.

371 00:19:22.680 --> 00:19:26.988 And as I told you, actually, we can see in the same tissue

372 00:19:26.988 --> 00:19:30.950 sort of after the barcoding and before the sequencing,

373 00:19:30.950 --> 00:19:33.947 we can even just under optical microscope,

374 00:19:33.947 --> 00:19:36.340 we can see individual pixels over here.

375 00:19:36.340 --> 00:19:39.663 And then when my poster showed me this image,

376 00:19:39.663 --> 00:19:42.339 it's okay, you got a key wide fiber over there

377 00:19:42.339 --> 00:19:45.350 very likely, because we saw this

378 00:19:45.350 --> 00:19:48.260 when we used microfluids before.

379 00:19:48.260 --> 00:19:50.300 And I thought that's unfortunate

380 00:19:50.300 --> 00:19:51.440 but anyhow, let's go ahead

381 00:19:51.440 --> 00:19:53.640 and process the sequencing data.

382 00:19:53.640 --> 00:19:55.820 But turns out that's not a key wide fiber

383 00:19:55.820 --> 00:19:58.150 that's really a very thin layer,

384 00:19:58.150 --> 00:20:02.320 actually it's a single cell layer of melanocytes

385 00:20:02.320 --> 00:20:04.610 lining a round the eye field.

386 00:20:04.610 --> 00:20:06.950 At this stage, the eye field actually,

387 00:20:06.950 --> 00:20:09.320 it's a very, very early stage only,

388 00:20:09.320 --> 00:20:13.130 called the eye vesicle an even no optical caps,

389 00:20:13.130 --> 00:20:15.140 it's the optical vesicle.

390 00:20:15.140 --> 00:20:19.327 So we can see, very distinctly, a group of genes

391 00:20:19.327 --> 00:20:22.230 strongly enriched inside the eye

392 00:20:22.230 --> 00:20:27.000 and also lining around the eye, optical vesicle.

393 00:20:27.000 --> 00:20:30.600 And then when we put them together,

394 00:20:30.600 --> 00:20:32.930 a little bit more structures you can see.

395 00:20:32.930 --> 00:20:36.250 For example in Pax6 enriched pretty much

396 00:20:36.250 --> 00:20:38.180 in an entire eye field

397 00:20:38.180 --> 00:20:42.700 but also in this region is optical nerve fiber.

398 00:20:42.700 --> 00:20:47.170 But here this protein, only expressed in the eye,

399 00:20:47.170 --> 00:20:50.630 but also other tissue type but not so much optical fiber.

400 00:20:50.630 --> 00:20:53.723 You can see this very well at a very high resolution,

401 00:20:53.723 --> 00:20:57.050 it's really about a single cell resolution.

402 00:20:57.050 --> 00:20:59.340 So, okay, when you look at it carefully,

403 00:20:59.340 --> 00:21:01.220 you see some yellow spots over here.

404 00:21:01.220 --> 00:21:05.050 That means the Pax6 and the Pmel are actually co-expressed

405 00:21:05.050 --> 00:21:08.877 in those kinda melanoblast cells but this one is not.

406 00:21:08.877 --> 00:21:10.880 The Six6 is not expressed,

407 00:21:10.880 --> 00:21:14.900 only within the eye, optical vesicle.

408 00:21:14.900 --> 00:21:16.870 If you further zoom in, you can see

409 00:21:18.160 --> 00:21:20.770 the sort of gene expression within the vesicle

410 00:21:20.770 --> 00:21:24.160 and also individual pixels, every little square here.

411 00:21:24.160 --> 00:21:26.340 So we can overlay the tissue image

412 00:21:26.340 --> 00:21:29.083 and the transcriptome data.

413 00:21:29.083 --> 00:21:31.390 So we noticed one gene which

414 00:21:33.141 --> 00:21:35.820 is strongly enriched right here,

415 00:21:35.820 --> 00:21:39.278 very strongly differential expression spatially.

416 00:21:39.278 --> 00:21:42.150 We're all curious what this gene does.

417 00:21:42.150 --> 00:21:45.500 We did sort of,

418 00:21:45.500 --> 00:21:49.350 this time they're still global, gene differential analysis.

419 00:21:49.350 --> 00:21:54.350 We saw only top ranked genes and these two showed up.

420 00:21:54.590 --> 00:21:59.590 But we found their functioning on a top ranked pathways,

421 00:22:00.567 --> 00:22:03.970 to some degree, okay, except those ones,

422 00:22:03.970 --> 00:22:06.290 to some degree, are mutually exclusive.

423 00:22:06.290 --> 00:22:09.760 And then later we realized

424 00:22:09.760 --> 00:22:12.200 but that has never been observed before,

425 00:22:12.200 --> 00:22:15.490 I don't have sort of last year's data to support.

426 00:22:15.490 --> 00:22:17.910 But it seems like those cells

427 00:22:20.550 --> 00:22:23.090 sort of characterized by this particular gene,

428 00:22:23.090 --> 00:22:26.670 later on are gonna determine the development of the lens.

429 00:22:26.670 --> 00:22:29.740 And those cells, even at this stage,

430 00:22:29.740 --> 00:22:32.450 you don't see any morphological difference,

431 00:22:32.450 --> 00:22:35.180 they already predetermined to develop

432 00:22:35.180 --> 00:22:38.990 the retina and the photo receptor cells.

433 00:22:38.990 --> 00:22:41.110 And then we were able to basically

434 00:22:41.110 --> 00:22:42.870 just put out those pictures obviously

435 00:22:42.870 --> 00:22:44.990 and compare it to those to perform

436 00:22:44.990 --> 00:22:46.960 a differential gene expression analysis.

437 00:22:46.960 --> 00:22:50.570 And another surprise, now this gene just showed up

438 00:22:50.570 --> 00:22:53.660 extremely differentially expressed.

439 00:22:53.660 --> 00:22:58.210 But we see many other genes that were very interesting.

440 00:22:58.210 --> 00:23:00.653 We still try to look into the details.

441 00:23:00.653 --> 00:23:03.701 So they are kinda enriched on the left side.

442 00:23:03.701 --> 00:23:07.402 Eventually, very likely,

443 00:23:07.402 --> 00:23:12.273 they will contribute to the photo receptor cell development.

444 00:23:13.450 --> 00:23:14.930 Okay, so even though we're able

445 00:23:14.930 --> 00:23:16.860 to visualize individual genes,

446 00:23:16.860 --> 00:23:18.810 we don't have to use the gene cell enrichment

447 00:23:18.810 --> 00:23:21.057 to identify different tissue types,

448 00:23:21.057 --> 00:23:23.860 but we had a challenge in particular

449 00:23:23.860 --> 00:23:25.510 in this kind of eye field region,

450 00:23:26.845 --> 00:23:31.470 due to our lack of knowledge in mouse embryonic development.

451 00:23:31.470 --> 00:23:34.550 But it'll be great if some computational pipeline

452 00:23:34.550 --> 00:23:37.210 can automatically identify different features,

453 00:23:37.210 --> 00:23:38.043 tissue features.

454 00:23:38.043 --> 00:23:41.170 That's what we demonstrate as well.

455 00:23:41.170 --> 00:23:44.124 So using this automatic automated

456 00:23:44.124 --> 00:23:46.540 feature identification pipeline,

457 00:23:46.540 --> 00:23:49.310 we were able to identify actually 20 different features

458 00:23:49.310 --> 00:23:52.240 in this very small region of the brain

459 00:23:52.240 --> 00:23:54.420 around the eye field.

460 00:23:54.420 --> 00:23:57.814 I just will show you some of those,

461 00:23:57.814 --> 00:24:00.480 you can see not just the eye, actually you can see

462 00:24:00.480 --> 00:24:02.520 very already development of the ear

463 00:24:02.520 --> 00:24:06.510 based on the sort of gene expression,

464 00:24:06.510 --> 00:24:10.593 but histologically, you cannot see any difference at all.

465 00:24:12.460 --> 00:24:17.460 But we also look at entire mouse embryo the E10.

466 00:24:17.620 --> 00:24:20.720 We're able to identify about 20 different features.

467 00:24:20.720 --> 00:24:24.560 But we're asking, so if at later stage

468 00:24:24.560 --> 00:24:26.460 many other organs begin to develop,

469 00:24:26.460 --> 00:24:29.374 whether or not this pipeline can identify many more

470 00:24:29.374 --> 00:24:33.230 tissue features or tissue subtypes.

471 00:24:33.230 --> 00:24:35.840 That turns out that that's right.

472 00:24:35.840 --> 00:24:39.470 And using E12, we're now able to cover entire embryo

473 00:24:39.470 --> 00:24:42.570 actually just the lower part of the body,

474 00:24:42.570 --> 00:24:45.493 we identify about 40 different features already.

475 00:24:46.517 --> 00:24:50.940 So this is a very high resolution as well.

476 00:24:50.940 --> 00:24:53.663 Okay, I'm gonna just summarize

477 00:24:53.663 --> 00:24:58.663 back to my sort of, ,the main interest in cancer.

478 00:24:59.260 --> 00:25:02.260 So I believe this enabling platform,

479 00:25:02.260 --> 00:25:04.880 we demonstrate can do protein and the transcripts.

480 00:25:04.880 --> 00:25:08.440 But actually, in my lab, another post I'm working on,

481 00:25:08.440 --> 00:25:11.220 so spatial, high spatial resolution epigenomics.

482 00:25:11.220 --> 00:25:13.060 I believe we can do high res,

483 00:25:13.060 --> 00:25:14.610 high spatial resolution ATAC,

484 00:25:14.610 --> 00:25:16.900 high spatial resolution CHIP-seq.

485 00:25:16.900 --> 00:25:19.007 And the application is extremely broad

486 00:25:19.007 --> 00:25:21.310 and the cancer is put right in the middle

487 00:25:21.310 --> 00:25:25.180 because that's really my main focus.

488 00:25:25.180 --> 00:25:28.360 I will like to thank people in my lab who work on this

489 00:25:28.360 --> 00:25:30.083 and thank you for your attention.