WEBVTT

NOTE duration:"00:43:23.7970000"

NOTE recognizability:0.951

NOTE language:en-us

NOTE Confidence: 0.979803799999999

00:00:00.000 --> 00:00:02.304 Because now the the main focus of today

NOTE Confidence: 0.979803799999999

00:00:02.304 --> 00:00:04.661 will be the analysis and interpretation

NOTE Confidence: 0.979803799999999

00:00:04.661 --> 00:00:06.806 of single cell sequencing data.

NOTE Confidence: 0.979803799999999

00:00:06.810 --> 00:00:08.838 So we won't cover everything today

NOTE Confidence: 0.979803799999999

00:00:08.838 --> 00:00:11.777 and so it will take at least another.

NOTE Confidence: 0.96613234

00:00:14.070 --> 00:00:16.100 Another meeting for covering everything.

NOTE Confidence: 0.96613234

00:00:16.100 --> 00:00:18.120 But today we covered the

NOTE Confidence: 0.96613234

00:00:18.120 --> 00:00:19.736 introduction on the methodologies,

NOTE Confidence: 0.96613234

00:00:19.740 --> 00:00:21.770 some technical and experimental issues,

NOTE Confidence: 0.96613234

00:00:21.770 --> 00:00:24.158 and some issues also with the

NOTE Confidence: 0.96613234

00:00:24.158 --> 00:00:26.629 with the analysis of this data.

NOTE Confidence: 0.96613234

00:00:26.630 --> 00:00:29.072 So single cell analysis as a

NOTE Confidence: 0.96613234

00:00:29.072 --> 00:00:31.489 definition is the study of omics.

NOTE Confidence: 0.96613234

00:00:31.490 --> 00:00:34.318 At least that's what we're speaking about.

NOTE Confidence: 0.96613234

00:00:34.320 --> 00:00:37.372 Today is the study of omics so

NOTE Confidence: 0.96613234

00:00:37.372 --> 00:00:38.680 genomics transcriptomics proteomics

NOTE Confidence: 0.96613234

00:00:38.748 --> 00:00:40.398 at the single cell level.

NOTE Confidence: 0.96613234

00:00:40.400 --> 00:00:43.388 So the advantage is that these.

NOTE Confidence: 0.96613234

00:00:43.390 --> 00:00:46.282 Family of methods allowed to capture

NOTE Confidence: 0.96613234

00:00:46.282 --> 00:00:49.282 a cellular diversity of tissues with

NOTE Confidence: 0.96613234

00:00:49.282 --> 00:00:52.234 the with the single cell resolution.

NOTE Confidence: 0.96613234

00:00:52.240 --> 00:00:55.345 Uh, so they feel there is a bursting is

NOTE Confidence: 0.96613234

00:00:55.345 --> 00:00:58.844 like exploding with a lot with a number

NOTE Confidence: 0.96613234

00:00:58.844 --> 00:01:01.969 of novel experimental techniques every year.

NOTE Confidence: 0.96613234

00:01:01.970 --> 00:01:04.080 But there are also many

NOTE Confidence: 0.96613234

00:01:04.080 --> 00:01:04.924 computational challenges,

NOTE Confidence: 0.96613234

00:01:04.930 --> 00:01:06.121 so these methods,

NOTE Confidence: 0.96613234

00:01:06.121 --> 00:01:08.503 the single cell methods require the

NOTE Confidence: 0.96613234

00:01:08.503 --> 00:01:10.430 development of appropriate analysis.

NOTE Confidence: 0.96613234

00:01:10.430 --> 00:01:14.180 And so we will see that.

NOTE Confidence: 0.96613234

00:01:14.180 --> 00:01:16.490 Common workflows are an employee,

NOTE Confidence: 0.96613234

00:01:16.490 --> 00:01:17.460 some generic.

NOTE Confidence: 0.96613234

00:01:17.460 --> 00:01:19.885 For example clustering analysis that

NOTE Confidence: 0.96613234

00:01:19.885 --> 00:01:22.938 June spoke about in our first meeting.

NOTE Confidence: 0.96613234

00:01:22.940 --> 00:01:26.167 Some of the methods for the normalization,

NOTE Confidence: 0.96613234

00:01:26.170 --> 00:01:27.208 for example,

NOTE Confidence: 0.96613234

00:01:27.208 --> 00:01:30.841 or for the calculation of differential gene

NOTE Confidence: 0.96613234

00:01:30.841 --> 00:01:33.997 expression are taken from the bulk RNA seek,

NOTE Confidence: 0.96613234

00:01:34.000 --> 00:01:37.688 but it is not always the best choice.

NOTE Confidence: 0.96613234

00:01:37.690 --> 00:01:40.917 And since the field is rapidly moving,

NOTE Confidence: 0.96613234

00:01:40.920 --> 00:01:44.728 there is no gold standard I would say.

NOTE Confidence: 0.96613234

00:01:44.730 --> 00:01:47.010 In any step of the analysis,

NOTE Confidence: 0.96613234

00:01:47.010 --> 00:01:50.050 so you will find a lot of methods,

NOTE Confidence: 0.96613234

00:01:50.050 --> 00:01:51.634 a lot of applications.

NOTE Confidence: 0.96613234

00:01:51.634 --> 00:01:54.010 You can find the literature compare

NOTE Confidence: 0.96613234

00:01:54.081 --> 00:01:56.127 for each step of the analysis,

NOTE Confidence: 0.96613234

00:01:56.130 --> 00:01:56.958 alternative approaches,

NOTE Confidence: 0.96613234

00:01:56.958 --> 00:01:59.856 but there is no like gold reference

NOTE Confidence: 0.96613234

00:01:59.856 --> 00:02:02.206 that you that that you can choose.

NOTE Confidence: 0.96613234

00:02:02.210 --> 00:02:02.984 For example,

NOTE Confidence: 0.96613234

00:02:02.984 --> 00:02:06.080 there is a sort of called pipelines in

NOTE Confidence: 0.96613234

00:02:06.162 --> 00:02:09.050 the bulk RNA seek and the single cell.

NOTE Confidence: 0.96613234

00:02:09.050 --> 00:02:10.570 It's not so established.

NOTE Confidence: 0.95873636

00:02:12.680 --> 00:02:14.624 This is a comparison of the

NOTE Confidence: 0.95873636

00:02:14.624 --> 00:02:16.290 method single cell versus bulk,

NOTE Confidence: 0.95873636

00:02:16.290 --> 00:02:18.918 so in the back analysis you take a tissue

NOTE Confidence: 0.95873636

00:02:18.918 --> 00:02:21.436 or population of cells and you extract

NOTE Confidence: 0.95873636

00:02:21.436 --> 00:02:23.896 DNA from the whole population so that

NOTE Confidence: 0.95873636

00:02:23.896 --> 00:02:28.680 you mix up the RNA content in the same.

NOTE Confidence: 0.95873636

00:02:28.680 --> 00:02:31.414 Yeah, in the same container, let's say,

NOTE Confidence: 0.95873636

00:02:31.414 --> 00:02:34.053 and then when you prepare the library

NOTE Confidence: 0.95873636

00:02:34.053 --> 00:02:37.100 and you sequence DNA from the whole from

NOTE Confidence: 0.95873636

00:02:37.100 --> 00:02:39.739 from the whole population of cells.

NOTE Confidence: 0.95873636

00:02:39.740 --> 00:02:42.971 This means that you get for each library from

NOTE Confidence: 0.95873636

00:02:42.971 --> 00:02:46.055 each collection of cells from each tissue.

NOTE Confidence: 0.95873636

00:02:46.060 --> 00:02:47.440 Only one measurement.

NOTE Confidence: 0.95873636

00:02:47.440 --> 00:02:50.200 And this measurement of genes represents

NOTE Confidence: 0.95873636

00:02:50.200 --> 00:02:52.786 the average expression of these genes

NOTE Confidence: 0.95873636

00:02:52.786 --> 00:02:55.749 across all the details of your tissue.

NOTE Confidence: 0.95873636

00:02:55.750 --> 00:02:58.431 So obviously you cannot use the back

NOTE Confidence: 0.95873636

00:02:58.431 --> 00:03:01.080 command code if you want, for example,

NOTE Confidence: 0.95873636

00:03:01.080 --> 00:03:04.120 to see this cellular it originality in your

NOTE Confidence: 0.95873636

00:03:04.191 --> 00:03:07.175 tissue with a single cell analysis you yeah,

NOTE Confidence: 0.95873636

00:03:07.180 --> 00:03:09.861 you first perform a a step that

NOTE Confidence: 0.95873636

00:03:09.861 --> 00:03:12.130 is the isolation of the cells.

NOTE Confidence: 0.95873636

00:03:12.130 --> 00:03:14.416 So this is kind of tricky,

NOTE Confidence: 0.95873636

00:03:14.420 --> 00:03:16.784 especially in solid tissues because you

NOTE Confidence: 0.95873636

00:03:16.784 --> 00:03:18.990 need to mechanically separated each cell.

NOTE Confidence: 0.95873636

00:03:18.990 --> 00:03:21.657 It's easier with the liquid that issues.

NOTE Confidence: 0.95873636

00:03:21.660 --> 00:03:23.352 It's easiest, for example,

NOTE Confidence: 0.95873636

00:03:23.352 --> 00:03:25.890 when you consider the analysis of.

NOTE Confidence: 0.95873636

00:03:25.890 --> 00:03:27.912 Democratic cells and so inside each

NOTE Confidence: 0.95873636

00:03:27.912 --> 00:03:30.267 single cell you you perform the

NOTE Confidence: 0.95873636

00:03:30.267 --> 00:03:32.147 quantification of gene expression.

NOTE Confidence: 0.95873636

00:03:32.150 --> 00:03:35.382 Because you have a way to create a

NOTE Confidence: 0.95873636

00:03:35.382 --> 00:03:37.729 library where you can keep track

NOTE Confidence: 0.95873636

00:03:37.729 --> 00:03:40.749 of the cell of origin of each RNA,

NOTE Confidence: 0.95873636

00:03:40.750 --> 00:03:42.230 and so that's why,

NOTE Confidence: 0.95873636

00:03:42.230 --> 00:03:44.941 then you can quantify for each gene

NOTE Confidence: 0.95873636

00:03:44.941 --> 00:03:47.783 at the expression in each single cell,

NOTE Confidence: 0.95873636

00:03:47.790 --> 00:03:50.527 so that each cell has a distinct

NOTE Confidence: 0.95873636

00:03:50.527 --> 00:03:51.309 expression profiles.

NOTE Confidence: 0.95873636

00:03:51.310 --> 00:03:52.090 For example,

NOTE Confidence: 0.95873636

00:03:52.090 --> 00:03:54.430 this cell expresses only one gene.

NOTE Confidence: 0.95873636

00:03:54.430 --> 00:03:56.514 These other cells express.

NOTE Confidence: 0.95873636

00:03:56.514 --> 00:03:59.119 Different multiple genes and with

NOTE Confidence: 0.95873636

00:03:59.119 --> 00:04:01.442 different amounts and so that you

NOTE Confidence: 0.95873636

00:04:01.442 --> 00:04:02.474 can use that.

NOTE Confidence: 0.95873636

00:04:02.480 --> 00:04:04.955 This difference in the expression

NOTE Confidence: 0.95873636

00:04:04.955 --> 00:04:07.430 between different cells in order

NOTE Confidence: 0.95873636

00:04:07.505 --> 00:04:09.969 to see how much cells are similar

NOTE Confidence: 0.95873636

00:04:09.969 --> 00:04:12.079 to each other or different.

NOTE Confidence: 0.95873636

00:04:12.080 --> 00:04:13.277 So for example,

NOTE Confidence: 0.95873636

00:04:13.277 --> 00:04:15.671 you can perform clustering analysis of

NOTE Confidence: 0.95873636

00:04:15.671 --> 00:04:18.470 cells based on their expression profiles

NOTE Confidence: 0.95873636

00:04:18.470 --> 00:04:20.800 and also other downstream analysis.

NOTE Confidence: 0.95873636

00:04:20.800 --> 00:04:24.650 So obviously you have a richer data.

NOTE Confidence: 0.95873636

00:04:24.650 --> 00:04:28.166 That you can see and UM.

NOTE Confidence: 0.95873636

00:04:28.170 --> 00:04:31.668 And you have multiple more options

NOTE Confidence: 0.95873636

00:04:31.668 --> 00:04:35.870 in India in the final analysis.

NOTE Confidence: 0.95873636

00:04:35.870 --> 00:04:36.478 So yes,

NOTE Confidence: 0.95873636

00:04:36.478 --> 00:04:38.606 when it was launched is like there

NOTE Confidence: 0.95873636

00:04:38.606 --> 00:04:40.832 was this kind of comparison between

NOTE Confidence: 0.95873636

00:04:40.832 --> 00:04:43.220 bike and array seek that Vulcan.

NOTE Confidence: 0.95873636

00:04:43.220 --> 00:04:45.670 Alesis is like the analysis of through.

NOTE Confidence: 0.95873636

00:04:45.670 --> 00:04:48.120 It seems Moody and the single cell.

NOTE Confidence: 0.95873636

00:04:48.120 --> 00:04:50.591 It's like the analysis of a fruit

NOTE Confidence: 0.95873636

00:04:50.591 --> 00:04:53.051 salad where you can distinguish the

NOTE Confidence: 0.95873636

00:04:53.051 --> 00:04:55.691 contribution of each fluid for each

NOTE Confidence: 0.95873636

00:04:55.691 --> 00:04:58.807 fruit is a different cell type or subtype.

NOTE Confidence: 0.95873636

00:04:58.810 --> 00:05:01.282 Now the main application for single

NOTE Confidence: 0.95873636

00:05:01.282 --> 00:05:04.148 cell RNA sequencing when we're speaking

NOTE Confidence: 0.95873636

00:05:04.148 --> 00:05:06.998 about discrimination among different cells.

NOTE Confidence: 0.95873636

00:05:07.000 --> 00:05:09.316 There are multiple. I divided the.

NOTE Confidence: 0.95873636

00:05:09.320 --> 00:05:11.250 These are in two branches,

NOTE Confidence: 0.95873636

00:05:11.250 --> 00:05:13.560 so why is it so cold?

NOTE Confidence: 0.95873636

00:05:13.560 --> 00:05:14.718 The discrete analysis?

NOTE Confidence: 0.95873636

00:05:14.718 --> 00:05:16.648 So you you have a,

NOTE Confidence: 0.95873636

00:05:16.650 --> 00:05:18.386 you collect the expression,

NOTE Confidence: 0.95873636

00:05:18.386 --> 00:05:20.556 abundance of transcripts of genes

NOTE Confidence: 0.95873636

00:05:20.556 --> 00:05:22.896 inside each cell and you want to

NOTE Confidence: 0.95873636

00:05:22.896 --> 00:05:25.033 cluster cells in order to identify

NOTE Confidence: 0.95873636

00:05:25.033 --> 00:05:26.296 different cell types.

NOTE Confidence: 0.95873636

00:05:26.300 --> 00:05:27.012 For example,

NOTE Confidence: 0.95873636

00:05:27.012 --> 00:05:29.148 this cell types that compose the

NOTE Confidence: 0.95873636

00:05:29.148 --> 00:05:30.929 tissue that you're studying.

NOTE Confidence: 0.95873636

00:05:30.930 --> 00:05:33.576 So this is a discrete analysis because

NOTE Confidence: 0.95873636

00:05:33.576 --> 00:05:36.325 you are assuming that your tissue is

NOTE Confidence: 0.95873636

00:05:36.325 --> 00:05:38.635 composed by different types of cells

NOTE Confidence: 0.9781691

00:05:38.710 --> 00:05:39.880 that are clearly.

NOTE Confidence: 0.9781691

00:05:39.880 --> 00:05:41.700 Distinguishable from each other,

NOTE Confidence: 0.9781691

00:05:41.700 --> 00:05:43.975 and so these analysis has,

NOTE Confidence: 0.9781691

00:05:43.980 --> 00:05:46.255 for example, something to do

NOTE Confidence: 0.9781691

00:05:46.255 --> 00:05:48.530 with the class with clustering,

NOTE Confidence: 0.9781691

00:05:48.530 --> 00:05:50.845 because ultimately you want to

NOTE Confidence: 0.9781691

00:05:50.845 --> 00:05:52.697 identify separate clusters of

NOTE Confidence: 0.9781691

00:05:52.697 --> 00:05:54.984 cells based on their expression

NOTE Confidence: 0.9781691

00:05:54.984 --> 00:05:57.169 profile on the right question.

NOTE Confidence: 0.9856172

00:05:57.170 --> 00:05:59.445 Every question this is super

NOTE Confidence: 0.9856172

00:05:59.445 --> 00:06:00.810 relevant to hematopoiesis,

NOTE Confidence: 0.9856172

00:06:00.810 --> 00:06:02.534 so there's even controversy.

NOTE Confidence: 0.9856172

00:06:02.534 --> 00:06:05.120 I don't know that it should

NOTE Confidence: 0.9856172

00:06:05.202 --> 00:06:07.587 be a controversy about whether

NOTE Confidence: 0.9856172

00:06:07.587 --> 00:06:09.972 there are discrete cell states.

NOTE Confidence: 0.9856172

00:06:09.980 --> 00:06:11.976 Versus everything being continuous

NOTE Confidence: 0.9856172

00:06:11.976 --> 00:06:14.970 and logic tells me that there's

NOTE Confidence: 0.9856172

00:06:15.046 --> 00:06:17.584 going to be a continuous change

NOTE Confidence: 0.9856172

00:06:17.584 --> 00:06:19.810 in a bazillion different genes,

NOTE Confidence: 0.9856172

00:06:19.810 --> 00:06:22.145 because every cell is going

NOTE Confidence: 0.9856172

00:06:22.145 --> 00:06:24.013 to be slightly different.

NOTE Confidence: 0.9856172

00:06:24.020 --> 00:06:26.820 So do you have to actually ask

NOTE Confidence: 0.9856172

00:06:26.820 --> 00:06:28.990 the algorithm to analyze the

NOTE Confidence: 0.9856172

00:06:28.990 --> 00:06:31.642 data to find discrete sets versus

NOTE Confidence: 0.9856172

00:06:31.642 --> 00:06:34.320 find a continuous analysis?

NOTE Confidence: 0.9816257

00:06:36.320 --> 00:06:41.270 So I personally don't know if there is a way.

NOTE Confidence: 0.9816257

00:06:41.270 --> 00:06:43.900 If there is a tool so I never use that

NOTE Confidence: 0.9816257

00:06:43.978 --> 00:06:46.866 tool that tells you if the best analysis

NOTE Confidence: 0.9816257

00:06:46.866 --> 00:06:49.710 is discrete or continuous. OK, I think

NOTE Confidence: 0.98393106

00:06:49.710 --> 00:06:52.027 that probably if we looked at different

NOTE Confidence: 0.98393106

00:06:52.027 --> 00:06:54.211 papers where they claim it's discrete

NOTE Confidence: 0.98393106

00:06:54.211 --> 00:06:55.719 versus claiming it's continuous

NOTE Confidence: 0.98393106

00:06:55.719 --> 00:06:58.020 that we would find differences in

NOTE Confidence: 0.98393106

00:06:58.020 --> 00:06:59.820 how they analyzed it, yes,

NOTE Confidence: 0.98393106

00:06:59.820 --> 00:07:02.459 so the the priority knowledge of the

NOTE Confidence: 0.98393106

00:07:02.459 --> 00:07:04.878 sample is something that you can user,

NOTE Confidence: 0.98393106

00:07:04.880 --> 00:07:07.323 and for example if you take for

NOTE Confidence: 0.98393106

00:07:07.323 --> 00:07:08.850 example a peripheral blood,

NOTE Confidence: 0.98393106

00:07:08.850 --> 00:07:11.153 if you take single cell data sets

NOTE Confidence: 0.98393106

00:07:11.153 --> 00:07:12.851 of peripheral blood are where

NOTE Confidence: 0.98393106

00:07:12.851 --> 00:07:14.789 most of these cells are mature

NOTE Confidence: 0.98393106

00:07:14.789 --> 00:07:16.430 and already differentiated,

NOTE Confidence: 0.98393106

00:07:16.430 --> 00:07:19.118 then you see clearly that you have.

NOTE Confidence: 0.98393106

00:07:19.120 --> 00:07:20.904 Very separated discrete clusters,

NOTE Confidence: 0.98393106

00:07:20.904 --> 00:07:24.556 and so it makes more sense to perform

NOTE Confidence: 0.98393106

00:07:24.556 --> 00:07:26.786 a discrete analysis or clustering

NOTE Confidence: 0.98393106

00:07:26.786 --> 00:07:29.816 analysis if you take them on marrow

NOTE Confidence: 0.98393106

00:07:29.816 --> 00:07:32.228 or a population that isn't reached

NOTE Confidence: 0.98393106

00:07:32.230 --> 00:07:34.420 for stem cells or progenitors,

NOTE Confidence: 0.98393106

00:07:34.420 --> 00:07:38.508 then you expect to have a more continuous

NOTE Confidence: 0.98393106

00:07:38.508 --> 00:07:40.660 representation of your sample.

NOTE Confidence: 0.98393106

00:07:40.660 --> 00:07:43.324 And so this is important because

NOTE Confidence: 0.98393106

00:07:43.324 --> 00:07:45.510 whatever tool that you use,

NOTE Confidence: 0.98393106

00:07:45.510 --> 00:07:47.138 any clustering analysis will

NOTE Confidence: 0.98393106

00:07:47.138 --> 00:07:49.580 give you clustering and any like

NOTE Confidence: 0.98393106

00:07:49.647 --> 00:07:52.022 continuous analysis such as like

NOTE Confidence: 0.98393106

00:07:52.022 --> 00:07:53.447 inference of trajectory.

NOTE Confidence: 0.98393106

00:07:53.450 --> 00:07:55.198 We'll find the trajectory.

NOTE Confidence: 0.98393106

00:07:55.198 --> 00:07:59.179 So if you submit your sample to any analysis,

NOTE Confidence: 0.98393106

00:07:59.180 --> 00:08:00.944 you will obtain result,

NOTE Confidence: 0.98393106

00:08:00.944 --> 00:08:03.590 but the result can be meaningless.

NOTE Confidence: 0.98393106

00:08:03.590 --> 00:08:06.030 For example a continuous analysis

NOTE Confidence: 0.98393106

00:08:06.030 --> 00:08:08.470 can be minutes meaningless if

NOTE Confidence: 0.98393106

00:08:08.557 --> 00:08:11.037 your sample is biologically not.

NOTE Confidence: 0.98393106

00:08:11.040 --> 00:08:11.541 Uhm,

NOTE Confidence: 0.98393106

00:08:11.541 --> 00:08:12.543 for example,

NOTE Confidence: 0.98393106

00:08:12.543 --> 00:08:14.547 something that is differentiating

NOTE Confidence: 0.98393106

00:08:14.547 --> 00:08:15.549 or developing.

NOTE Confidence: 0.9756925

00:08:18.950 --> 00:08:21.520 So yeah, yeah, so yeah.

NOTE Confidence: 0.9756925

00:08:21.520 --> 00:08:24.090 And that's that's the parallel.

NOTE Confidence: 0.9756925

00:08:24.090 --> 00:08:28.689 So whenever you see they they like.

NOTE Confidence: 0.9756925

00:08:28.690 --> 00:08:30.730 Courses or tutorials on continuous analysis.

NOTE Confidence: 0.9756925

00:08:30.730 --> 00:08:33.110 That's something you need to be careful.

NOTE Confidence: 0.9756925

00:08:33.110 --> 00:08:35.150 You will always get a graph.

NOTE Confidence: 0.9756925

00:08:35.150 --> 00:08:37.572 You will always get like a sort

NOTE Confidence: 0.9756925

00:08:37.572 --> 00:08:39.230 of differentiation trees the tree,

NOTE Confidence: 0.9756925

00:08:39.230 --> 00:08:41.568 but you have to be careful because

NOTE Confidence: 0.9756925

00:08:41.568 --> 00:08:43.310 sometimes it doesn't make sense.

NOTE Confidence: 0.98708284

00:08:46.020 --> 00:08:48.690 To make the analysis at all.

NOTE Confidence: 0.98708284

00:08:48.690 --> 00:08:50.330 Because it's one of these,

NOTE Confidence: 0.98708284

00:08:50.330 --> 00:08:52.220 one of the assumption of a

NOTE Confidence: 0.98708284

00:08:52.220 --> 00:08:53.878 continuous analysis is that you

NOTE Confidence: 0.98708284

00:08:53.878 --> 00:08:55.876 have a sampling of the continuous

NOTE Confidence: 0.98708284

00:08:55.876 --> 00:08:57.868 process that you're trying to model.

NOTE Confidence: 0.98708284

00:08:57.870 --> 00:08:59.510 For example, development or differentiation.

NOTE Confidence: 0.98708284

00:08:59.510 --> 00:09:01.150 If this is not true,

NOTE Confidence: 0.98708284

00:09:01.150 --> 00:09:02.790 you don't have an assumption

NOTE Confidence: 0.98708284

00:09:02.790 --> 00:09:04.760 to do the analysis at all.

NOTE Confidence: 0.97320217

00:09:06.800 --> 00:09:09.726 So tomorrow you so so I think this

NOTE Confidence: 0.97320217

00:09:09.726 --> 00:09:11.964 is a very important question that

NOTE Confidence: 0.97320217

00:09:11.964 --> 00:09:14.614 I raised because you know in real

NOTE Confidence: 0.97320217

00:09:14.614 --> 00:09:16.856 life situations we will get samples

NOTE Confidence: 0.97320217

00:09:16.856 --> 00:09:19.560 sequenced and and how do we tell if

NOTE Confidence: 0.97320217

00:09:19.560 --> 00:09:21.720 this is reasonable or not reasonable

NOTE Confidence: 0.97320217

00:09:21.720 --> 00:09:24.681 so so just wonder if anyone has done

NOTE Confidence: 0.97320217

00:09:24.681 --> 00:09:27.180 a very careful analysis to sort of,

NOTE Confidence: 0.97320217

00:09:27.180 --> 00:09:29.728 you know something ground truth for example,

NOTE Confidence: 0.97320217

00:09:29.730 --> 00:09:32.187 you have two discrete cell states you

NOTE Confidence: 0.97320217

00:09:32.187 --> 00:09:34.445 already isolated or somehow maintained and

NOTE Confidence: 0.97320217

00:09:34.445 --> 00:09:37.112 put them into a single cell sequencing.

NOTE Confidence: 0.97320217

00:09:37.120 --> 00:09:39.647 And then you force it to assume

NOTE Confidence: 0.97320217

00:09:39.647 --> 00:09:41.408 trajectory based methodology and do

NOTE Confidence: 0.97320217

00:09:41.408 --> 00:09:43.256 cause many major artifacts or not.

NOTE Confidence: 0.97320217

00:09:43.260 --> 00:09:45.647 I think that's one of the ways

NOTE Confidence: 0.97320217

00:09:45.647 --> 00:09:46.670 to think about.

NOTE Confidence: 0.9500493

00:09:48.000 --> 00:09:51.112 Yeah, so, uh, so I'm not aware meaning

NOTE Confidence: 0.9500493

00:09:51.112 --> 00:09:54.828 that I never use the like tools that

NOTE Confidence: 0.9500493

00:09:54.828 --> 00:09:57.749 explicitly tell you which one which

NOTE Confidence: 0.9500493

00:09:57.749 --> 00:10:00.533 branch of the analysis is better.

NOTE Confidence: 0.9500493

00:10:00.540 --> 00:10:02.332 So by exploratory analysis,

NOTE Confidence: 0.9500493

00:10:02.332 --> 00:10:05.520 for example, we will see why when

NOTE Confidence: 0.9500493

00:10:05.520 --> 00:10:08.974 you do the preprocessing and then the

NOTE Confidence: 0.9500493

00:10:08.974 --> 00:10:11.639 dimensionality reduction and you have

NOTE Confidence: 0.9500493

00:10:11.639 --> 00:10:15.860 a lot like a on a hyperplane of cells.

NOTE Confidence: 0.9500493

00:10:15.860 --> 00:10:18.086 That you can like a guess,

NOTE Confidence: 0.9500493

00:10:18.090 --> 00:10:20.680 depending on the structure of your sample,

NOTE Confidence: 0.9500493

00:10:20.680 --> 00:10:22.725 whether it's more reasonable to

NOTE Confidence: 0.9500493

00:10:22.725 --> 00:10:25.154 proceed with the discrete cluster or

NOTE Confidence: 0.9500493

00:10:25.154 --> 00:10:27.359 to perform like a trajectory or boss.

NOTE Confidence: 0.9500493

00:10:27.360 --> 00:10:30.328 So sometimes, for example, if you it would,

NOTE Confidence: 0.9500493

00:10:30.330 --> 00:10:33.482 it could make sense to start with an

NOTE Confidence: 0.9500493

00:10:33.482 --> 00:10:35.520 exploratory analysis on all the cells.

NOTE Confidence: 0.9500493

00:10:35.520 --> 00:10:38.047 I think this as an example because

NOTE Confidence: 0.9500493

00:10:38.047 --> 00:10:39.600 it's on this life,

NOTE Confidence: 0.9500493

00:10:39.600 --> 00:10:43.247 so this seems to be like separate

NOTE Confidence: 0.9500493

00:10:43.247 --> 00:10:44.810 cluster of cells.

NOTE Confidence: 0.9500493

00:10:44.810 --> 00:10:46.206 It could be reasonable,

NOTE Confidence: 0.9500493

00:10:46.206 --> 00:10:47.951 then to select only this

NOTE Confidence: 0.9500493

00:10:47.951 --> 00:10:49.619 cluster within this cluster.

NOTE Confidence: 0.9500493

00:10:49.620 --> 00:10:51.470 We don't see clear subclasses,

NOTE Confidence: 0.9500493

00:10:51.470 --> 00:10:53.690 so within this cluster it may.

NOTE Confidence: 0.9500493

00:10:53.690 --> 00:10:55.922 It could make sense to perform

NOTE Confidence: 0.9500493

00:10:55.922 --> 00:10:58.192 a trajectory analysis to see if

NOTE Confidence: 0.9500493

00:10:58.192 --> 00:10:59.977 there is a continuous process,

NOTE Confidence: 0.9500493

00:10:59.980 --> 00:11:01.830 but not at the beginning

NOTE Confidence: 0.9500493

00:11:01.830 --> 00:11:02.570 taking consideration.

NOTE Confidence: 0.9500493

00:11:02.570 --> 00:11:04.820 Also, these two clusters here

NOTE Confidence: 0.9500493

00:11:04.820 --> 00:11:06.620 because they're clearly separated.

NOTE Confidence: 0.9500493

00:11:06.620 --> 00:11:09.116 So sometimes I think the that

NOTE Confidence: 0.9500493

00:11:09.116 --> 00:11:11.810 the workflow can be also mixed.

NOTE Confidence: 0.9500493

00:11:11.810 --> 00:11:14.316 So you start with all these cells,

NOTE Confidence: 0.9500493

00:11:14.320 --> 00:11:16.456 so you remove clear outlier clusters.

NOTE Confidence: 0.9500493

00:11:16.460 --> 00:11:18.250 Maybe you annotate the cluster

NOTE Confidence: 0.9500493

00:11:18.250 --> 00:11:19.682 so that you know,

NOTE Confidence: 0.9500493

00:11:19.690 --> 00:11:20.402 for example,

NOTE Confidence: 0.9500493

00:11:20.402 --> 00:11:22.894 that inside your population you have a

NOTE Confidence: 0.9500493

00:11:22.894 --> 00:11:25.060 mixture of progenitors or stem cells,

NOTE Confidence: 0.9500493

00:11:25.060 --> 00:11:27.960 and inside that cluster.

NOTE Confidence: 0.9500493

00:11:27.960 --> 00:11:29.708 Perform the trajectory analysis.

NOTE Confidence: 0.98311144

00:11:32.640 --> 00:11:34.674 I see so that would be

NOTE Confidence: 0.98311144

00:11:34.674 --> 00:11:36.030 my yes tentative answer.

NOTE Confidence: 0.98311144

00:11:36.030 --> 00:11:38.064 Now I don't know if anyone

NOTE Confidence: 0.98311144

00:11:38.064 --> 00:11:39.420 else has other suggestions.

NOTE Confidence: 0.9858118

00:11:46.290 --> 00:11:47.690 That maybe we can leave.

NOTE Confidence: 0.9858118

00:11:47.690 --> 00:11:49.412 I can find those some material

NOTE Confidence: 0.9858118

00:11:49.412 --> 00:11:51.639 for next time to see if I can

NOTE Confidence: 0.9858118

00:11:51.639 --> 00:11:52.710 answer more, like extensively.

NOTE Confidence: 0.98225963

00:11:54.940 --> 00:11:56.200 I think it's a tough question

NOTE Confidence: 0.98225963

00:11:56.200 --> 00:11:57.332 because I don't think there's

NOTE Confidence: 0.98225963

00:11:57.332 --> 00:11:58.567 a consensus necessarily in the

NOTE Confidence: 0.98225963

00:11:58.567 --> 00:11:59.889 field and people just to see.

NOTE Confidence: 0.98225963

00:11:59.890 --> 00:12:01.666 OK, if it makes sense or it doesn't

NOTE Confidence: 0.98225963

00:12:01.666 --> 00:12:03.040 make sense to their own eyes.

NOTE Confidence: 0.97921542

00:12:04.670 --> 00:12:06.980 Yep. Yeah, again I I don't know

NOTE Confidence: 0.97921542

00:12:06.980 --> 00:12:09.490 if someone is trying to to build

NOTE Confidence: 0.97921542

00:12:09.490 --> 00:12:11.650 some tools that yeah you know,

NOTE Confidence: 0.97921542

00:12:11.650 --> 00:12:13.806 yeah that like kind of quantify the.

NOTE Confidence: 0.98804873

00:12:16.350 --> 00:12:17.970 Reasonableness of each

NOTE Confidence: 0.98804873

00:12:17.970 --> 00:12:19.950 of the approaches. Yeah.

NOTE Confidence: 0.88182193

00:12:23.460 --> 00:12:25.908 But yes, it's an important distinction.

NOTE Confidence: 0.88182193

00:12:25.910 --> 00:12:30.470 Also, uh, yes. Also also later, 'cause it.

NOTE Confidence: 0.88182193

00:12:30.470 --> 00:12:31.562 Uhm, some history.

NOTE Confidence: 0.88182193

00:12:31.562 --> 00:12:34.110 So this is the first publication on

NOTE Confidence: 0.88182193

00:12:34.186 --> 00:12:37.356 single cell sequencing, so it's a 20.

NOTE Confidence: 0.88182193

00:12:37.356 --> 00:12:39.376 No sorry 12 years ago,

NOTE Confidence: 0.88182193

00:12:39.380 --> 00:12:42.215 so it was a nice seat cover.

NOTE Confidence: 0.88182193

00:12:42.220 --> 00:12:44.512 The whole transcriptome of a single

NOTE Confidence: 0.88182193

00:12:44.512 --> 00:12:47.737 cell so it was really one single cell

NOTE Confidence: 0.88182193

00:12:47.737 --> 00:12:50.656 because it was a mouse blaster that

NOTE Confidence: 0.88182193

00:12:50.656 --> 00:12:53.540 was isolated with a microscope so it

NOTE Confidence: 0.88182193

00:12:53.540 --> 00:12:55.854 was manually picked under my screw.

NOTE Confidence: 0.88182193

00:12:55.854 --> 00:12:58.080 A microscope then lies and then

NOTE Confidence: 0.88182193

00:12:58.156 --> 00:13:00.910 sequenced and together with the blaster.

NOTE Confidence: 0.88182193

00:13:00.910 --> 00:13:03.268 Also, 50 sites were also analyzed

NOTE Confidence: 0.88182193

00:13:03.268 --> 00:13:06.116 and so so basically the trick here

NOTE Confidence: 0.88182193

00:13:06.116 --> 00:13:08.516 to reach the single cell resolution

NOTE Confidence: 0.88182193

00:13:08.516 --> 00:13:11.071 was the isolation of these cells

NOTE Confidence: 0.88182193

00:13:11.071 --> 00:13:13.566 and then that the procedure was

NOTE Confidence: 0.88182193

00:13:13.566 --> 00:13:15.646 standard Lisa and then library

NOTE Confidence: 0.88182193

00:13:15.646 --> 00:13:18.858 preparation as in a as in balcony seek.

NOTE Confidence: 0.88182193

00:13:18.860 --> 00:13:23.466 But from the starting from 1 cell.

NOTE Confidence: 0.88182193

00:13:23.470 --> 00:13:25.816 So from that to the fielder,

NOTE Confidence: 0.88182193

00:13:25.820 --> 00:13:27.296 as I told you,

NOTE Confidence: 0.88182193

00:13:27.296 --> 00:13:30.510 an exploded and so in this plot here.

NOTE Confidence: 0.88182193

00:13:30.510 --> 00:13:34.029 So this is from a review that was 2018,

NOTE Confidence: 0.88182193

00:13:34.030 --> 00:13:36.767 so it was ten years after these

NOTE Confidence: 0.88182193

00:13:36.767 --> 00:13:37.549 first publication.

NOTE Confidence: 0.88182193

00:13:37.550 --> 00:13:40.000 And what you can see is the

NOTE Confidence: 0.88182193

00:13:40.000 --> 00:13:41.568 release of multiple approaches

NOTE Confidence: 0.88182193

00:13:41.568 --> 00:13:44.190 for single cell at any seeker.

NOTE Confidence: 0.88182193

00:13:44.190 --> 00:13:46.075 UM, that increase the number

NOTE Confidence: 0.88182193

00:13:46.075 --> 00:13:48.490 of cells that you can study.

NOTE Confidence: 0.88182193

00:13:48.490 --> 00:13:50.445 So obviously that one alone

NOTE Confidence: 0.88182193

00:13:50.445 --> 00:13:52.400 was a proof of concept,

NOTE Confidence: 0.88182193

00:13:52.400 --> 00:13:53.516 but the real.

NOTE Confidence: 0.88182193

00:13:53.516 --> 00:13:55.004 Single cell explosion happened

NOTE Confidence: 0.88182193

00:13:55.004 --> 00:13:57.618 when you put when you will be

NOTE Confidence: 0.88182193

00:13:57.618 --> 00:13:59.398 able to parallelize the process.

NOTE Confidence: 0.88182193

00:13:59.400 --> 00:14:01.388 So where you were able to capture

NOTE Confidence: 0.88182193

00:14:01.388 --> 00:14:03.424 a single cell expression level of

NOTE Confidence: 0.88182193

00:14:03.424 --> 00:14:05.294 first hundreds and then thousands

NOTE Confidence: 0.88182193

00:14:05.294 --> 00:14:07.198 and then millions of cells.

NOTE Confidence: 0.88182193

00:14:07.200 --> 00:14:09.896 So here you see the publication data of

NOTE Confidence: 0.88182193

00:14:09.896 --> 00:14:11.949 the techniques and the single cells.

NOTE Confidence: 0.88182193

00:14:11.950 --> 00:14:13.640 The number of single cells

NOTE Confidence: 0.88182193

00:14:13.640 --> 00:14:14.654 that were analyzed.

NOTE Confidence: 0.88182193

00:14:14.660 --> 00:14:17.020 So this is our first with only one

NOTE Confidence: 0.88182193

00:14:17.020 --> 00:14:19.768 cells and then you see that the trend

NOTE Confidence: 0.88182193

00:14:19.768 --> 00:14:22.024 is to release techniques that allow

NOTE Confidence: 0.88182193

00:14:22.024 --> 00:14:24.388 you to increase the high throughput.

NOTE Confidence: 0.88182193

00:14:24.390 --> 00:14:26.748 In terms of the number of

NOTE Confidence: 0.88182193

00:14:26.748 --> 00:14:29.220 cells that you can quantify,

NOTE Confidence: 0.88182193

00:14:29.220 --> 00:14:31.848 you can consider in each experiment.

NOTE Confidence: 0.988861

00:14:33.300 --> 00:14:34.680 Question, yeah, I don't know if

NOTE Confidence: 0.988861

00:14:34.680 --> 00:14:36.009 you're going to get to this.

NOTE Confidence: 0.988861

00:14:36.010 --> 00:14:39.674 So if you are just saying never mind.

NOTE Confidence: 0.988861

00:14:39.680 --> 00:14:41.983 As the number of cells that are

NOTE Confidence: 0.988861

00:14:41.983 --> 00:14:43.680 being sequenced has increased,

NOTE Confidence: 0.988861

00:14:43.680 --> 00:14:46.025 the number of reads per cell that

NOTE Confidence: 0.988861

00:14:46.025 --> 00:14:48.779 people get and report on has decreased,

NOTE Confidence: 0.988861

00:14:48.780 --> 00:14:50.760 and I'd like to understand is

NOTE Confidence: 0.988861

00:14:50.760 --> 00:14:52.502 that just because that's what's

NOTE Confidence: 0.988861

00:14:52.502 --> 00:14:54.302 convenient in terms of putting

NOTE Confidence: 0.988861

00:14:54.302 --> 00:14:56.420 it onto an illuminous sequencer,

NOTE Confidence: 0.988861

00:14:56.420 --> 00:14:59.157 or is there something about the various

NOTE Confidence: 0.988861

00:14:59.157 --> 00:15:01.403 techniques where you reach the limit

NOTE Confidence: 0.988861

00:15:01.403 --> 00:15:03.734 of your detection after X number of

NOTE Confidence: 0.988861

00:15:03.800 --> 00:15:08.180 reads and it's not worth getting more?

NOTE Confidence: 0.9299402

00:15:08.180 --> 00:15:09.704 Yes, so there is a tradeoff

NOTE Confidence: 0.9299402

00:15:09.704 --> 00:15:11.440 in in these two parameters.

NOTE Confidence: 0.9299402

00:15:11.440 --> 00:15:14.365 One is the number of cells that you consider

NOTE Confidence: 0.9299402

00:15:14.365 --> 00:15:16.986 and the other is the number of reads

NOTE Confidence: 0.9299402

00:15:16.986 --> 00:15:19.568 there that you obtained for each cell.

NOTE Confidence: 0.9299402

00:15:19.570 --> 00:15:23.112 The trend for the techniques has been

NOTE Confidence: 0.9299402

00:15:23.112 --> 00:15:26.530 mainly to increase the number of cells,

NOTE Confidence: 0.9299402

00:15:26.530 --> 00:15:29.010 and obviously these was against.

NOTE Confidence: 0.9299402

00:15:29.010 --> 00:15:31.495 This is against the number

NOTE Confidence: 0.9299402

00:15:31.495 --> 00:15:33.980 of reads for each cell.

NOTE Confidence: 0.9299402

00:15:33.980 --> 00:15:37.580 So for example and.

NOTE Confidence: 0.9299402

00:15:37.580 --> 00:15:39.680 So let's say that the fielder and

NOTE Confidence: 0.9299402

00:15:39.680 --> 00:15:41.889 the the most popular techniques,

NOTE Confidence: 0.9299402

00:15:41.890 --> 00:15:43.650 for example Tenax, have been

NOTE Confidence: 0.9299402

00:15:43.650 --> 00:15:45.840 increasing more the number of cells,

NOTE Confidence: 0.9299402

00:15:45.840 --> 00:15:49.864 then the number of reads for each cell.

NOTE Confidence: 0.9299402

00:15:49.870 --> 00:15:52.210 Uh, these are depends on the

NOTE Confidence: 0.9299402

00:15:52.210 --> 00:15:54.420 application of the method I guess.

NOTE Confidence: 0.9299402

00:15:54.420 --> 00:15:56.628 So obviously if you're interested in

NOTE Confidence: 0.9299402

00:15:56.628 --> 00:15:59.349 the cell as your unit of interest,

NOTE Confidence: 0.9299402

00:15:59.350 --> 00:16:01.240 so if you're interested in

NOTE Confidence: 0.9299402

00:16:01.240 --> 00:16:02.752 like more cellular biology,

NOTE Confidence: 0.9299402

00:16:02.760 --> 00:16:04.650 you're interested more in capturing

NOTE Confidence: 0.9299402

00:16:04.650 --> 00:16:06.162 cells and separating cells,

NOTE Confidence: 0.9299402

00:16:06.170 --> 00:16:08.065 you're not so interesting looking

NOTE Confidence: 0.9299402

00:16:08.065 --> 00:16:10.405 with in great detail on thousands

NOTE Confidence: 0.9299402

00:16:10.405 --> 00:16:12.679 of genes that are expressed within

NOTE Confidence: 0.9299402

00:16:12.679 --> 00:16:14.879 each cell on the other side,

NOTE Confidence: 0.9299402

00:16:14.880 --> 00:16:16.424 if you're more interested,

NOTE Confidence: 0.9299402

00:16:16.424 --> 00:16:18.354 for example in the molecular

NOTE Confidence: 0.9299402

00:16:18.354 --> 00:16:19.540 biology rather than.

NOTE Confidence: 0.9299402

00:16:19.540 --> 00:16:22.432 Just separating cells so it would

NOTE Confidence: 0.9299402

00:16:22.432 --> 00:16:25.408 be more interesting to increase the

NOTE Confidence: 0.9299402

00:16:25.408 --> 00:16:28.754 depth of the sequencing in each cell.

NOTE Confidence: 0.9299402

00:16:28.760 --> 00:16:32.060 There are techniques where these is

NOTE Confidence: 0.9299402

00:16:32.060 --> 00:16:35.173 maximized and obviously the trade off

NOTE Confidence: 0.9299402

00:16:35.173 --> 00:16:38.453 is that you cannot get so many cells.

NOTE Confidence: 0.9299402

00:16:38.460 --> 00:16:41.760 As for the other method.

NOTE Confidence: 0.9299402

00:16:41.760 --> 00:16:42.470 Uhm?

NOTE Confidence: 0.98397849

00:16:44.710 --> 00:16:46.639 I. Think I'm

NOTE Confidence: 0.98394567

00:16:46.640 --> 00:16:48.500 just asking also and maybe June

NOTE Confidence: 0.98394567

00:16:48.500 --> 00:16:50.490 knows is there a maximum number

NOTE Confidence: 0.98394567

00:16:50.490 --> 00:16:52.524 of reads you want per cell?

NOTE Confidence: 0.98394567

00:16:52.530 --> 00:16:54.165 Because after that you don't

NOTE Confidence: 0.98394567

00:16:54.165 --> 00:16:57.290 get any additional information.

NOTE Confidence: 0.92274123

00:16:57.290 --> 00:17:00.322 So we will see we can measure when

NOTE Confidence: 0.92274123

00:17:00.322 --> 00:17:03.374 you reach the like the the plateau

NOTE Confidence: 0.92274123

00:17:03.374 --> 00:17:06.563 when you reach the plateau of the

NOTE Confidence: 0.92274123

00:17:06.563 --> 00:17:09.713 sequencing using tricks such as the UMI.

NOTE Confidence: 0.92274123

00:17:09.720 --> 00:17:11.832 So if you append to each

NOTE Confidence: 0.92274123

00:17:11.832 --> 00:17:13.980 read like a random barcode,

NOTE Confidence: 0.92274123

00:17:13.980 --> 00:17:16.716 you can see when a it doesn't make sense

NOTE Confidence: 0.92274123

00:17:16.716 --> 00:17:19.692 to sequence more depth because all the

NOTE Confidence: 0.92274123

00:17:19.692 --> 00:17:22.489 additional reads that you are detecting.

NOTE Confidence: 0.92274123

00:17:22.490 --> 00:17:24.990 Other PCR duplicates of what

NOTE Confidence: 0.92274123

00:17:24.990 --> 00:17:26.790 you already sequenced. Right,

NOTE Confidence: 0.97971755

00:17:26.790 --> 00:17:28.010 got it? Yeah I

NOTE Confidence: 0.97971755

00:17:28.010 --> 00:17:29.134 agree with the Tommaso.

NOTE Confidence: 0.97971755

00:17:29.134 --> 00:17:31.522 So I think there are some studies was

NOTE Confidence: 0.97971755

00:17:31.522 --> 00:17:33.454 down those kind of things before,

NOTE Confidence: 0.97971755

00:17:33.460 --> 00:17:35.075 but they were using different

NOTE Confidence: 0.97971755

00:17:35.075 --> 00:17:36.690 technologies compared to what you're

NOTE Confidence: 0.97971755

00:17:36.742 --> 00:17:38.308 going to use probably right now,

NOTE Confidence: 0.97971755

00:17:38.310 --> 00:17:40.375 so I'm not sure whether for every

NOTE Confidence: 0.97971755

00:17:40.375 --> 00:17:41.640 single technology out there,

NOTE Confidence: 0.97971755

00:17:41.640 --> 00:17:43.754 there has already been a paper published.

NOTE Confidence: 0.97971755

00:17:43.760 --> 00:17:45.428 Maybe for 10X there's already paper

NOTE Confidence: 0.97971755

00:17:45.428 --> 00:17:47.090 published on the standard procedures,

NOTE Confidence: 0.97971755

00:17:47.090 --> 00:17:49.330 but in your own data you can actually

NOTE Confidence: 0.97971755

00:17:49.330 --> 00:17:50.925 analyze yourself to see whether

NOTE Confidence: 0.97971755

00:17:50.925 --> 00:17:52.550 it's approaching saturation or not,

NOTE Confidence: 0.97971755

00:17:52.550 --> 00:17:54.727 and you can re sequence more from

NOTE Confidence: 0.97971755

00:17:54.727 --> 00:17:56.489 same library if you want to.

NOTE Confidence: 0.933664227272727

00:17:57.190 --> 00:17:59.992 Yeah. Yes, yes using Umm eyes

NOTE Confidence: 0.933664227272727

00:17:59.992 --> 00:18:02.440 and you can measure that.

NOTE Confidence: 0.933664227272727

00:18:02.440 --> 00:18:05.052 OK, I understand, thank you then.

NOTE Confidence: 0.933664227272727

00:18:05.052 --> 00:18:06.378 The same techniques,

NOTE Confidence: 0.933664227272727

00:18:06.378 --> 00:18:08.588 for example 10X at every

NOTE Confidence: 0.933664227272727

00:18:08.588 --> 00:18:10.690 release like increase or the.

NOTE Confidence: 0.974990600000001

00:18:13.170 --> 00:18:15.030 Increase the detection of multiple

NOTE Confidence: 0.974990600000001

00:18:15.030 --> 00:18:17.258 molecules inside each cell so that

NOTE Confidence: 0.974990600000001

00:18:17.258 --> 00:18:19.372 the saturation limit is higher and so

NOTE Confidence: 0.974990600000001

00:18:19.372 --> 00:18:21.583 it really depends also on the depends

NOTE Confidence: 0.974990600000001

00:18:21.583 --> 00:18:24.441 on the technique and then and on the

NOTE Confidence: 0.974990600000001

00:18:24.441 --> 00:18:27.500 version of the of the technique itself.

NOTE Confidence: 0.974990600000001

00:18:27.500 --> 00:18:29.908 But in general, I would say that

NOTE Confidence: 0.974990600000001

00:18:29.908 --> 00:18:32.574 the number of cells that you can

NOTE Confidence: 0.974990600000001

00:18:32.574 --> 00:18:34.908 measure has increased the more in

NOTE Confidence: 0.974990600000001

00:18:34.985 --> 00:18:37.230 the average of the techniques.

NOTE Confidence: 0.974990600000001

00:18:37.230 --> 00:18:39.558 Then the depth that then the,

NOTE Confidence: 0.974990600000001

00:18:39.560 --> 00:18:42.480 then the within cell depth.

NOTE Confidence: 0.974990600000001

00:18:42.480 --> 00:18:44.485 With an exception that they

NOTE Confidence: 0.974990600000001

00:18:44.485 --> 00:18:47.050 show you in your Indies slide,

NOTE Confidence: 0.974990600000001

00:18:47.050 --> 00:18:49.125 that is the smart speaker

NOTE Confidence: 0.974990600000001

00:18:49.125 --> 00:18:50.370 family of technical.

NOTE Confidence: 0.974990600000001

00:18:50.370 --> 00:18:52.884 So this family of techniques is

NOTE Confidence: 0.974990600000001

00:18:52.884 --> 00:18:55.848 the ideal family when you are not

NOTE Confidence: 0.974990600000001

00:18:55.848 --> 00:18:58.669 interested in capturing a lot of cells,

NOTE Confidence: 0.974990600000001

00:18:58.670 --> 00:19:01.764 but you want to maximize the analysis

NOTE Confidence: 0.974990600000001

00:19:01.764 --> 00:19:05.076 within each cell and the advantage of

NOTE Confidence: 0.974990600000001

00:19:05.076 --> 00:19:07.926 dictating of these techniques is that

NOTE Confidence: 0.974990600000001

00:19:08.009 --> 00:19:11.449 you can have 1,000,000 read for each cell.

NOTE Confidence: 0.974990600000001

00:19:11.450 --> 00:19:14.339 So it has a high coverage and also and

NOTE Confidence: 0.974990600000001

00:19:14.339 --> 00:19:17.501 it's one of the techniques that allow you

NOTE Confidence: 0.974990600000001

00:19:17.501 --> 00:19:20.708 to capture reads from the whole transcript.

NOTE Confidence: 0.974990600000001

00:19:20.710 --> 00:19:23.412 So we will see that the majority

NOTE Confidence: 0.974990600000001

00:19:23.412 --> 00:19:24.570 of commercial techniques,

NOTE Confidence: 0.974990600000001

00:19:24.570 --> 00:19:27.910 such as the 10X A.

NOTE Confidence: 0.974990600000001

00:19:27.910 --> 00:19:29.680 Do not allow you to cover

NOTE Confidence: 0.974990600000001

00:19:29.680 --> 00:19:30.565 the full transcript,

NOTE Confidence: 0.974990600000001

00:19:30.570 --> 00:19:32.346 but they are like three prime

NOTE Confidence: 0.974990600000001

00:19:32.346 --> 00:19:33.530 end or five prime.

NOTE Confidence: 0.974990600000001

00:19:33.530 --> 00:19:34.508 End the libraries.

NOTE Confidence: 0.974990600000001

00:19:34.508 --> 00:19:36.790 So that means that you can capture

NOTE Confidence: 0.974990600000001

00:19:36.860 --> 00:19:39.172 only the fragment that is near to the

NOTE Confidence: 0.974990600000001

00:19:39.172 --> 00:19:41.374 palie for the three prime end or near

NOTE Confidence: 0.974990600000001

00:19:41.374 --> 00:19:43.446 to the cap for the five prime end.

NOTE Confidence: 0.974990600000001

00:19:43.446 --> 00:19:45.616 This is one of the few methods where

NOTE Confidence: 0.974990600000001

00:19:45.616 --> 00:19:48.237 you can add in bulk and in most by

NOTE Confidence: 0.974990600000001

00:19:48.237 --> 00:19:50.271 Karen Acq can capture reads from

NOTE Confidence: 0.974990600000001

00:19:50.271 --> 00:19:52.406 the full transcript and this is an

NOTE Confidence: 0.974990600000001

00:19:52.406 --> 00:19:53.801 advantage because for example if

NOTE Confidence: 0.974990600000001

00:19:53.801 --> 00:19:55.730 you want to do splicing analysis,

NOTE Confidence: 0.974990600000001

00:19:55.730 --> 00:19:57.836 that's the only way you can.

NOTE Confidence: 0.974990600000001

00:19:57.840 --> 00:20:00.717 You know that that that that's the

NOTE Confidence: 0.974990600000001

00:20:00.717 --> 00:20:03.927 only method you can use to have a like.

NOTE Confidence: 0.974990600000001

00:20:03.930 --> 00:20:05.982 To perform splicing and a full

NOTE Confidence: 0.974990600000001

00:20:05.982 --> 00:20:07.013 splicing analysis, otherwise,

NOTE Confidence: 0.974990600000001

00:20:07.013 --> 00:20:08.728 you can prefer splicing analysis

NOTE Confidence: 0.974990600000001

00:20:08.728 --> 00:20:10.430 only on the initial exon,

NOTE Confidence: 0.974990600000001

00:20:10.430 --> 00:20:12.796 five prime end or on the terminal

NOTE Confidence: 0.974990600000001

00:20:12.796 --> 00:20:15.218 axons and also for a Allen Alesys.

NOTE Confidence: 0.974990600000001

00:20:15.220 --> 00:20:17.422 So analysis of variations analysis of

NOTE Confidence: 0.974990600000001

00:20:17.422 --> 00:20:19.887 Snips from Renee see from RNA seek

NOTE Confidence: 0.974990600000001

00:20:19.887 --> 00:20:22.057 A if you're a mutation of interest.

NOTE Confidence: 0.974990600000001

00:20:22.060 --> 00:20:24.610 If your variation of interest is

NOTE Confidence: 0.974990600000001

00:20:24.610 --> 00:20:27.761 inside the body of the gene and not

NOTE Confidence: 0.974990600000001

00:20:27.761 --> 00:20:30.899 at the five prime or the three prime.

NOTE Confidence: 0.974990600000001

00:20:30.900 --> 00:20:34.295 So this has all the advantages of

NOTE Confidence: 0.974990600000001

00:20:34.295 --> 00:20:36.720 allowing analysis within each cell

NOTE Confidence: 0.974990600000001

00:20:36.720 --> 00:20:39.354 that is comparable to the biker,

NOTE Confidence: 0.974990600000001

00:20:39.360 --> 00:20:41.880 any SQL.

NOTE Confidence: 0.974990600000001

00:20:41.880 --> 00:20:44.491 It has a limitation that is shared

NOTE Confidence: 0.974990600000001

00:20:44.491 --> 00:20:45.610 with other techniques.

NOTE Confidence: 0.974990600000001

00:20:45.610 --> 00:20:48.235 Is that most of the single cell

NOTE Confidence: 0.974990600000001

00:20:48.235 --> 00:20:50.796 techniques right now allow you to

NOTE Confidence: 0.974990600000001

00:20:50.796 --> 00:20:53.066 detect the only polyadenylated RNA

NOTE Confidence: 0.974990600000001

00:20:53.066 --> 00:20:56.000 because they're based on quality selection.

NOTE Confidence: 0.974990600000001

00:20:56.000 --> 00:20:58.484 And you have a low number of cells that

NOTE Confidence: 0.974990600000001

00:20:58.484 --> 00:21:00.880 you can sequence in each experiment,

NOTE Confidence: 0.974990600000001

00:21:00.880 --> 00:21:02.164 so less than 1000.

NOTE Confidence: 0.974990600000001

00:21:02.164 --> 00:21:04.779 Then the smart Seeker has already 3 version,

NOTE Confidence: 0.974990600000001

00:21:04.780 --> 00:21:07.048 so it was released first in 2012.

NOTE Confidence: 0.974990600000001

00:21:07.050 --> 00:21:09.466 Then here is a smart seek to release

NOTE Confidence: 0.974990600000001

00:21:09.466 --> 00:21:11.686 the one year after and then the

NOTE Confidence: 0.974990600000001

00:21:11.686 --> 00:21:14.034 latest is March 6th 3 that was

NOTE Confidence: 0.974990600000001

00:21:14.034 --> 00:21:15.498 released the last year.

NOTE Confidence: 0.974990600000001

00:21:15.500 --> 00:21:19.210 So each of these kind of increase.

NOTE Confidence: 0.974990600000001

00:21:19.210 --> 00:21:21.610 Then then the number of usable

NOTE Confidence: 0.974990600000001

00:21:21.610 --> 00:21:23.210 reads with smart smart

NOTE Confidence: 0.9669231

00:21:23.294 --> 00:21:25.786 seek two didn't allow to use the.

NOTE Confidence: 0.9669231

00:21:25.790 --> 00:21:29.660 Umm, I but Smartsilk 3 allows to to use also.

NOTE Confidence: 0.9669231

00:21:29.660 --> 00:21:32.117 Umm, I and this is a comparison

NOTE Confidence: 0.9669231

00:21:32.117 --> 00:21:33.920 between the two versions.

NOTE Confidence: 0.9669231

00:21:33.920 --> 00:21:36.302 Smartsilk 2 smartest see where you

NOTE Confidence: 0.9669231

00:21:36.302 --> 00:21:39.616 see the box block with the number of

NOTE Confidence: 0.9669231

00:21:39.616 --> 00:21:42.088 genes that are detected within each

NOTE Confidence: 0.9669231

00:21:42.169 --> 00:21:44.857 cell and as you see with the mastic

NOTE Confidence: 0.9669231

00:21:44.857 --> 00:21:47.134 tree you can foreach seller cover

NOTE Confidence: 0.9669231

00:21:47.134 --> 00:21:49.540 detector from 10,000 to 12,000 jeans.

NOTE Confidence: 0.9669231

00:21:49.540 --> 00:21:51.962 And also this number it is comparable

NOTE Confidence: 0.9669231

00:21:51.962 --> 00:21:54.886 to buy her any seek if you compare

NOTE Confidence: 0.9669231

00:21:54.886 --> 00:21:57.171 this number with the other method

NOTE Confidence: 0.9669231

00:21:57.171 --> 00:21:59.156 cells such as SYNNEX, internex.

NOTE Confidence: 0.9669231

00:21:59.156 --> 00:22:02.899 I think the average is 3 to 5000 genes for

NOTE Confidence: 0.9669231

00:22:02.899 --> 00:22:05.819 each cell when when this value is high.

NOTE Confidence: 0.95787907

00:22:10.900 --> 00:22:14.032 OK, so this is an example of high coverage,

NOTE Confidence: 0.95787907

00:22:14.040 --> 00:22:15.348 but low throughput are.

NOTE Confidence: 0.95787907

00:22:15.348 --> 00:22:18.112 On the other hand there you have methods

NOTE Confidence: 0.95787907

00:22:18.112 --> 00:22:20.308 where you have low coverage inside

NOTE Confidence: 0.95787907

00:22:20.308 --> 00:22:22.806 each cell and but high throughput and

NOTE Confidence: 0.95787907

00:22:22.806 --> 00:22:25.206 a family of these methods they are

NOTE Confidence: 0.95787907

00:22:25.206 --> 00:22:26.946 the so-called droplet based methods.

NOTE Confidence: 0.95787907

00:22:26.950 --> 00:22:31.081 These was one of the first set that was

NOTE Confidence: 0.95787907

00:22:31.081 --> 00:22:34.836 really that was released and it is the.

NOTE Confidence: 0.95787907

00:22:34.840 --> 00:22:37.336 I had the drop seeker analysis,

NOTE Confidence: 0.95787907

00:22:37.340 --> 00:22:41.806 so the principle is to isolate cells.

NOTE Confidence: 0.95787907

00:22:41.810 --> 00:22:43.720 Single cells in single droplet.

NOTE Confidence: 0.95787907

00:22:43.720 --> 00:22:46.789 Will you have your cell and you have a

NOTE Confidence: 0.95787907

00:22:46.789 --> 00:22:49.430 barcode that beats are barcoded beads.

NOTE Confidence: 0.95787907

00:22:49.430 --> 00:22:52.776 Allow you to attach a cellular barcode

NOTE Confidence: 0.95787907

00:22:52.776 --> 00:22:55.965 that is unique for each beat down

NOTE Confidence: 0.95787907

00:22:55.965 --> 00:22:59.000 and so it's unique for each cell.

NOTE Confidence: 0.95787907

00:22:59.000 --> 00:23:01.975 And so that's the trick that is

NOTE Confidence: 0.95787907

00:23:01.975 --> 00:23:04.025 used in order to, uh,

NOTE Confidence: 0.95787907

00:23:04.025 --> 00:23:06.000 associated the content of each

NOTE Confidence: 0.95787907

00:23:06.000 --> 00:23:08.260 cell with a single barcode.

NOTE Confidence: 0.95787907

00:23:08.260 --> 00:23:11.750 That is the cell barcode.

NOTE Confidence: 0.95787907

00:23:11.750 --> 00:23:15.506 Uhm it, so it allows to map 1000 or 10s of

NOTE Confidence: 0.95787907

00:23:15.506 --> 00:23:18.810 thousands of cells in the same experiments.

NOTE Confidence: 0.95787907

00:23:18.810 --> 00:23:21.498 Uhm, it's so the drops eater is

NOTE Confidence: 0.95787907

00:23:21.498 --> 00:23:23.512 only three prime end sequencing

NOTE Confidence: 0.95787907

00:23:23.512 --> 00:23:26.249 and it allows you the use of,

NOTE Confidence: 0.95787907

00:23:26.250 --> 00:23:27.986 Umm unique molecular identifiers.

NOTE Confidence: 0.95787907

00:23:27.986 --> 00:23:31.336 So I will have some slides later to

NOTE Confidence: 0.95787907

00:23:31.336 --> 00:23:33.695 show what is the meaning of that.

NOTE Confidence: 0.95787907

00:23:33.700 --> 00:23:35.564 This is the pipeline.

NOTE Confidence: 0.95787907

00:23:35.564 --> 00:23:37.428 Are the experimental pipeline

NOTE Confidence: 0.95787907

00:23:37.428 --> 00:23:40.199 of a drop seek experiment.

NOTE Confidence: 0.95787907

00:23:40.200 --> 00:23:42.588 So the the principle is twice.

NOTE Confidence: 0.95787907

00:23:42.590 --> 00:23:44.990 Let some point in a droplet,

NOTE Confidence: 0.95787907

00:23:44.990 --> 00:23:46.980 one cell with one microparticle.

NOTE Confidence: 0.95787907

00:23:46.980 --> 00:23:50.388 Inside this droplet you have the capture of

NOTE Confidence: 0.95787907

00:23:50.388 --> 00:23:52.968 the polyadenylated RNA with a polety probe,

NOTE Confidence: 0.95787907

00:23:52.970 --> 00:23:55.756 and then you have the little transcript,

NOTE Confidence: 0.95787907

00:23:55.760 --> 00:23:57.980 the reverse transcription and the generation

NOTE Confidence: 0.95787907

00:23:57.980 --> 00:24:01.350 of the C DNA and the library preparation.

NOTE Confidence: 0.95787907

00:24:01.350 --> 00:24:04.368 This is kind of similar to

NOTE Confidence: 0.95787907

00:24:04.368 --> 00:24:06.380 also buy currency approaches.

NOTE Confidence: 0.95787907

00:24:06.380 --> 00:24:08.345 Very similar to drop seek

NOTE Confidence: 0.95787907

00:24:08.345 --> 00:24:10.310 is also the 10X approach.

NOTE Confidence: 0.95787907

00:24:10.310 --> 00:24:11.882 That is the commercial

NOTE Confidence: 0.95787907

00:24:11.882 --> 00:24:13.847 development of the drug seeker,

NOTE Confidence: 0.95787907

00:24:13.850 --> 00:24:15.810 so send it to next.

NOTE Confidence: 0.95787907

00:24:15.810 --> 00:24:18.234 You have the same strategy of

NOTE Confidence: 0.95787907

00:24:18.234 --> 00:24:19.850 dividing dividing cells so

NOTE Confidence: 0.95787907

00:24:19.931 --> 00:24:22.097 that you have droplets in oil,

NOTE Confidence: 0.95787907

00:24:22.100 --> 00:24:25.052 in this case with a single cell and a

NOTE Confidence: 0.95787907

00:24:25.052 --> 00:24:27.996 single barcode with the cellular barcode.

NOTE Confidence: 0.95787907

00:24:28.000 --> 00:24:30.744 And as you can see the barcode attached

NOTE Confidence: 0.95787907

00:24:30.744 --> 00:24:33.477 to each bid have standard adapter that

NOTE Confidence: 0.95787907

00:24:33.477 --> 00:24:36.330 you can use in Illumina sequencing.

NOTE Confidence: 0.95787907

00:24:36.330 --> 00:24:37.950 You have the cellular barcode,

NOTE Confidence: 0.95787907

00:24:37.950 --> 00:24:39.560 you have the Umm I,

NOTE Confidence: 0.95787907

00:24:39.560 --> 00:24:41.835 and then you have a positive probe

NOTE Confidence: 0.95787907

00:24:41.835 --> 00:24:44.406 that is used to capture Poly a RNA.

NOTE Confidence: 0.9411849

00:24:47.110 --> 00:24:48.745 Uhm, this is to remind

NOTE Confidence: 0.9411849

00:24:48.745 --> 00:24:49.726 that different platforms,

NOTE Confidence: 0.9411849

00:24:49.730 --> 00:24:51.355 according to different strategies have

NOTE Confidence: 0.9411849

00:24:51.355 --> 00:24:53.320 different gene coverage is so smart.

NOTE Confidence: 0.9411849

00:24:53.320 --> 00:24:55.282 Seek two that we saw before

NOTE Confidence: 0.9411849

00:24:55.282 --> 00:24:56.590 has a full coverage.

NOTE Confidence: 0.9411849

00:24:56.590 --> 00:24:59.533 So if you consider these are like meta gene,

NOTE Confidence: 0.9411849

00:24:59.540 --> 00:25:01.496 we have the five prime UTR,

NOTE Confidence: 0.9411849

00:25:01.500 --> 00:25:03.130 the body of the gene,

NOTE Confidence: 0.9411849

00:25:03.130 --> 00:25:05.530 the coding sequence and the three prime UTR

NOTE Confidence: 0.9411849

00:25:05.530 --> 00:25:08.037 you have coverage of the full transcript,

NOTE Confidence: 0.9411849

00:25:08.040 --> 00:25:10.140 while with 10X or free payment

NOTE Confidence: 0.9411849

00:25:10.140 --> 00:25:11.901 method you haven't richemond only

NOTE Confidence: 0.9411849

00:25:11.901 --> 00:25:13.980 at the three prime end of the

NOTE Confidence: 0.9411849

00:25:13.980 --> 00:25:15.937 transcript with the five prime method.

NOTE Confidence: 0.9411849

00:25:15.940 --> 00:25:17.144 You haven't richemond all

NOTE Confidence: 0.9411849

00:25:17.144 --> 00:25:18.649 yet to five prime end,

NOTE Confidence: 0.9411849

00:25:18.650 --> 00:25:20.687 so you need to be careful on

NOTE Confidence: 0.9411849

00:25:20.687 --> 00:25:22.259 which library you are using,

NOTE Confidence: 0.9411849

00:25:22.260 --> 00:25:25.140 because if it is for just

NOTE Confidence: 0.9411849

00:25:25.140 --> 00:25:26.580 for gene quantification.

NOTE Confidence: 0.9411849

00:25:26.580 --> 00:25:28.060 Methods can be comparable,

NOTE Confidence: 0.9411849

00:25:28.060 --> 00:25:29.910 but if you are interested,

NOTE Confidence: 0.9411849

00:25:29.910 --> 00:25:32.130 for example in a ice form,

NOTE Confidence: 0.9411849

00:25:32.130 --> 00:25:32.870 expressions, pricing,

NOTE Confidence: 0.9411849

00:25:32.870 --> 00:25:34.720 analysis and so on only.

NOTE Confidence: 0.9411849

00:25:34.720 --> 00:25:36.570 These methods allow you to

NOTE Confidence: 0.9411849

00:25:36.570 --> 00:25:38.050 perform a complete analysis,

NOTE Confidence: 0.9411849

00:25:38.050 --> 00:25:39.160 not these ones.

NOTE Confidence: 0.9613364

00:25:41.520 --> 00:25:43.150 And this is another plot

NOTE Confidence: 0.9613364

00:25:43.150 --> 00:25:44.780 comparing the aging coverage are

NOTE Confidence: 0.9613364

00:25:44.838 --> 00:25:46.438 when you have full coverage.

NOTE Confidence: 0.9613364

00:25:46.440 --> 00:25:48.228 So this plot here is similar

NOTE Confidence: 0.9613364

00:25:48.228 --> 00:25:50.212 to plots that you could obtain

NOTE Confidence: 0.9613364

00:25:50.212 --> 00:25:52.017 from back button a seeker.

NOTE Confidence: 0.9613364

00:25:52.020 --> 00:25:53.976 This method is free prime end.

NOTE Confidence: 0.9613364

00:25:53.980 --> 00:25:55.540 There is a free payment

NOTE Confidence: 0.9613364

00:25:55.540 --> 00:25:57.590 method and so you see there.

NOTE Confidence: 0.9613364

00:25:57.590 --> 00:25:58.902 Richmond at the free

NOTE Confidence: 0.9613364

00:25:58.902 --> 00:26:00.214 prime of the transcript.

NOTE Confidence: 0.9614097

00:26:03.960 --> 00:26:07.767 Now this was for many for the technical part.

NOTE Confidence: 0.9614097

00:26:07.770 --> 00:26:10.704 Now the outlook on the computational

NOTE Confidence: 0.9614097

00:26:10.704 --> 00:26:13.206 analysis of single seller is

NOTE Confidence: 0.9614097

00:26:13.206 --> 00:26:15.038 resumed by these workflow.

NOTE Confidence: 0.9614097

00:26:15.040 --> 00:26:17.693 So most of their most popular methods

NOTE Confidence: 0.9614097

00:26:17.693 --> 00:26:20.073 that allow you to generate libraries

NOTE Confidence: 0.9614097

00:26:20.073 --> 00:26:22.782 and the result will be read the

NOTE Confidence: 0.9614097

00:26:22.857 --> 00:26:25.157 sequence with a standard platform

NOTE Confidence: 0.9614097

00:26:25.157 --> 00:26:27.838 such as illuminum some single cell

NOTE Confidence: 0.9614097

00:26:27.838 --> 00:26:29.728 methods have been published also

NOTE Confidence: 0.9614097

00:26:29.728 --> 00:26:32.010 that user full length sequencing.

NOTE Confidence: 0.9614097

00:26:32.010 --> 00:26:35.040 They think they're so probably

NOTE Confidence: 0.9614097

00:26:35.040 --> 00:26:38.520 in the future they would be.

NOTE Confidence: 0.9614097

00:26:38.520 --> 00:26:39.576 Use that more,

NOTE Confidence: 0.9614097

00:26:39.576 --> 00:26:41.688 but right now the standard is

NOTE Confidence: 0.9614097

00:26:41.688 --> 00:26:44.068 to use short read sequencing.

NOTE Confidence: 0.9614097

00:26:44.070 --> 00:26:46.464 Couple to see what self analysis so

NOTE Confidence: 0.9614097

00:26:46.464 --> 00:26:49.198 we will see how RO data are obtained

NOTE Confidence: 0.9614097

00:26:49.198 --> 00:26:52.240 and how the raw data reads can be

NOTE Confidence: 0.9614097

00:26:52.240 --> 00:26:54.012 transformed into count matrices

NOTE Confidence: 0.9614097

00:26:54.012 --> 00:26:56.478 that are similar to the count

NOTE Confidence: 0.9614097

00:26:56.478 --> 00:26:59.110 matrixes of the bike and a secret.

NOTE Confidence: 0.9614097

00:26:59.110 --> 00:27:01.672 But you have instead of having samples

NOTE Confidence: 0.9614097

00:27:01.672 --> 00:27:04.254 and jeans you have single cells and

NOTE Confidence: 0.9614097

00:27:04.254 --> 00:27:06.836 genes in your matrix and the numbers

NOTE Confidence: 0.9614097

00:27:06.836 --> 00:27:09.200 correspond to the number of reads

NOTE Confidence: 0.9614097

00:27:09.200 --> 00:27:11.930 mapping to the gene in the cell.

NOTE Confidence: 0.9614097

00:27:11.930 --> 00:27:14.240 Then there are quality control methodologies.

NOTE Confidence: 0.9614097

00:27:14.240 --> 00:27:17.210 Is well isation methodology's class ring,

NOTE Confidence: 0.9614097

00:27:17.210 --> 00:27:19.458 uh identification of trajectories.

NOTE Confidence: 0.9614097

00:27:19.458 --> 00:27:22.830 So like analysis that assume your

NOTE Confidence: 0.9614097

00:27:22.919 --> 00:27:25.604 sample your population of cells

NOTE Confidence: 0.9614097

00:27:25.604 --> 00:27:28.289 user continues and methods assume

NOTE Confidence: 0.9614097

00:27:28.380 --> 00:27:31.110 that your population is discrete.

NOTE Confidence: 0.9614097

00:27:31.110 --> 00:27:32.946 So let's start from the beginning.

NOTE Confidence: 0.9614097

00:27:32.950 --> 00:27:35.198 So usually in most of the methods row

NOTE Confidence: 0.9614097

00:27:35.198 --> 00:27:37.859 in in the row reads that you receive.

NOTE Confidence: 0.9614097

00:27:37.860 --> 00:27:39.395 There are three important parts

NOTE Confidence: 0.9614097

00:27:39.395 --> 00:27:40.316 that you have,

NOTE Confidence: 0.9614097

00:27:40.320 --> 00:27:43.416 and there are three parts of this sequence.

NOTE Confidence: 0.9614097

00:27:43.420 --> 00:27:45.448 And so the first important part

NOTE Confidence: 0.9614097

00:27:45.448 --> 00:27:46.800 is the cell barcode.

NOTE Confidence: 0.9614097

00:27:46.800 --> 00:27:49.560 So this is a an oligonucleotides there that

NOTE Confidence: 0.9614097

00:27:49.560 --> 00:27:52.876 can be like 8 to 12 or more nucleotide long.

NOTE Confidence: 0.9614097

00:27:52.880 --> 00:27:55.085 This depends on the on the technique

NOTE Confidence: 0.9614097

00:27:55.085 --> 00:27:57.571 and the so it's these sequence the

NOTE Confidence: 0.9614097

00:27:57.571 --> 00:28:00.659 cell barcode is unique for each of the bids.

NOTE Confidence: 0.9614097

00:28:00.660 --> 00:28:03.750 For example that you used.

NOTE Confidence: 0.9614097

00:28:03.750 --> 00:28:05.320 That when your cell was

NOTE Confidence: 0.9614097

00:28:05.320 --> 00:28:06.890 in the in the droplet,

NOTE Confidence: 0.9614097

00:28:06.890 --> 00:28:09.716 so it's what you use to identify the cell,

NOTE Confidence: 0.9614097

00:28:09.720 --> 00:28:11.638 meaning that one of the first step

NOTE Confidence: 0.9614097

00:28:11.638 --> 00:28:14.071 is to look at this region of the

NOTE Confidence: 0.9614097

00:28:14.071 --> 00:28:16.053 reader that correspond to the cell

NOTE Confidence: 0.9614097

00:28:16.053 --> 00:28:17.878 barcode and the group together.

NOTE Confidence: 0.9614097

00:28:17.880 --> 00:28:20.120 All the reads that have the same

NOTE Confidence: 0.9614097

00:28:20.120 --> 00:28:22.277 barcode and that's what you see here.

NOTE Confidence: 0.9614097

00:28:22.280 --> 00:28:24.752 So all the reads with the same barcode

NOTE Confidence: 0.9614097

00:28:24.752 --> 00:28:26.987 here in red belong to sell one,

NOTE Confidence: 0.9614097

00:28:26.990 --> 00:28:28.970 because this is the cell barcode

NOTE Confidence: 0.9614097

00:28:28.970 --> 00:28:30.749 of these cells and so on.

NOTE Confidence: 0.9614097

00:28:30.750 --> 00:28:32.778 So that all reads are grouped

NOTE Confidence: 0.9614097

00:28:32.778 --> 00:28:34.130 according to the value.

NOTE Confidence: 0.9614097

00:28:34.130 --> 00:28:35.606 Off the barcode.

NOTE Confidence: 0.9614097

00:28:35.606 --> 00:28:38.558 So obviously here there are some

NOTE Confidence: 0.9614097

00:28:38.558 --> 00:28:41.356 methodology to account for possible errors

NOTE Confidence: 0.9614097

00:28:41.356 --> 00:28:44.948 in the sequencing of the barcode so that.

NOTE Confidence: 0.9614097

00:28:44.950 --> 00:28:47.000 Barcodes are realized in a

NOTE Confidence: 0.9614097

00:28:47.000 --> 00:28:49.050 way that they have multiple,

NOTE Confidence: 0.9614097

00:28:49.050 --> 00:28:50.280 multiple different nucleotides,

NOTE Confidence: 0.9614097

00:28:50.280 --> 00:28:53.560 so that if you make one error only,

NOTE Confidence: 0.9614097

00:28:53.560 --> 00:28:54.380 you don't.

NOTE Confidence: 0.9614097

00:28:54.380 --> 00:28:57.660 You don't switch from one cell to another,

NOTE Confidence: 0.9614097

00:28:57.660 --> 00:28:59.710 but you need at least,

NOTE Confidence: 0.9614097

00:28:59.710 --> 00:29:00.530 for example,

NOTE Confidence: 0.9614097

00:29:00.530 --> 00:29:02.580 three errors in this sequencing.

NOTE Confidence: 0.9614097

00:29:02.580 --> 00:29:05.331 In the bar code to identify at

NOTE Confidence: 0.9614097

00:29:05.331 --> 00:29:07.910 the wrong self for the reader.

NOTE Confidence: 0.90901923

00:29:10.110 --> 00:29:11.960 The second part that is

NOTE Confidence: 0.90901923

00:29:11.960 --> 00:29:13.810 important is the so called.

NOTE Confidence: 0.90901923

00:29:13.810 --> 00:29:16.770 Umm I so this is not a while.

NOTE Confidence: 0.90901923

00:29:16.770 --> 00:29:19.730 The cell barcode is unique for each cell.

NOTE Confidence: 0.90901923

00:29:19.730 --> 00:29:22.658 The UMI is unique for each of the

NOTE Confidence: 0.90901923

00:29:22.658 --> 00:29:24.540 original molecule in your sample,

NOTE Confidence: 0.90901923

00:29:24.540 --> 00:29:26.760 and that's because there in the

NOTE Confidence: 0.90901923

00:29:26.760 --> 00:29:27.870 library preparation strategies.

NOTE Confidence: 0.90901923

00:29:27.870 --> 00:29:31.222 This is A is a non legal nucleotide

NOTE Confidence: 0.90901923

00:29:31.222 --> 00:29:34.896 that is included is appended to the.

NOTE Confidence: 0.90901923

00:29:34.900 --> 00:29:39.040 Library during the cDNA.

NOTE Confidence: 0.90901923

00:29:39.040 --> 00:29:40.100 Transgeneration before

NOTE Confidence: 0.90901923

00:29:40.100 --> 00:29:41.690 the amplification steps.

NOTE Confidence: 0.90901923

00:29:41.690 --> 00:29:44.122 So before the PCR.

NOTE Confidence: 0.90901923

00:29:44.122 --> 00:29:47.162 So that this means that.

NOTE Confidence: 0.90901923

00:29:47.170 --> 00:29:49.678 These can be used this stretch.

NOTE Confidence: 0.90901923

00:29:49.680 --> 00:29:52.648 This random bar code can be used to

NOTE Confidence: 0.90901923

00:29:52.648 --> 00:29:54.410 discriminate between PCR duplicates

NOTE Confidence: 0.90901923

00:29:54.410 --> 00:29:56.780 and the real biological duplicates.

NOTE Confidence: 0.90901923

00:29:56.780 --> 00:29:59.966 So in early seat you can expect to see

NOTE Confidence: 0.90901923

00:29:59.966 --> 00:30:03.139 two reads that are the same because

NOTE Confidence: 0.90901923

00:30:03.139 --> 00:30:05.980 they were derived from two copies,

NOTE Confidence: 0.90901923

00:30:05.980 --> 00:30:07.231 two different transcripts

NOTE Confidence: 0.90901923

00:30:07.231 --> 00:30:09.316 transcribed from the same gene.

NOTE Confidence: 0.90901923

00:30:09.320 --> 00:30:10.502 So some genes,

NOTE Confidence: 0.90901923

00:30:10.502 --> 00:30:12.472 such as ribosomal the transcript

NOTE Confidence: 0.90901923

00:30:12.472 --> 00:30:13.920 of ribosomal proteins,

NOTE Confidence: 0.90901923

00:30:13.920 --> 00:30:16.500 are expected to be in the

NOTE Confidence: 0.90901923

00:30:16.500 --> 00:30:18.630 range of 1000 to 10.

NOTE Confidence: 0.90901923

00:30:18.630 --> 00:30:21.216 1000 copies in a single cell,

NOTE Confidence: 0.90901923

00:30:21.220 --> 00:30:24.076 so you can expect to have more

NOTE Confidence: 0.90901923

00:30:24.076 --> 00:30:26.410 molecules captured in your library,

NOTE Confidence: 0.90901923

00:30:26.410 --> 00:30:28.948 but these are true biological sequences

NOTE Confidence: 0.90901923

00:30:28.948 --> 00:30:32.010 because at the origin you have two

NOTE Confidence: 0.90901923

00:30:32.010 --> 00:30:34.175 discrete are different RNA molecules.

NOTE Confidence: 0.90901923

00:30:34.180 --> 00:30:37.018 This is different from a PCR

NOTE Confidence: 0.90901923

00:30:37.018 --> 00:30:39.632 duplicates because these are created

NOTE Confidence: 0.90901923

00:30:39.632 --> 00:30:42.156 during the amplification step.

NOTE Confidence: 0.90901923

00:30:42.160 --> 00:30:43.648 So that's why the,

NOTE Confidence: 0.90901923

00:30:43.648 --> 00:30:44.020 UM,

NOTE Confidence: 0.90901923

00:30:44.020 --> 00:30:46.174 I are important because this discrimination

NOTE Confidence: 0.90901923

00:30:46.174 --> 00:30:47.610 between biological duplicates and

NOTE Confidence: 0.90901923

00:30:47.664 --> 00:30:49.579 technical duplicates is really important.

NOTE Confidence: 0.90901923

00:30:49.580 --> 00:30:51.064 When you perform many

NOTE Confidence: 0.90901923

00:30:51.064 --> 00:30:52.177 rounds of amplification,

NOTE Confidence: 0.90901923

00:30:52.180 --> 00:30:54.035 and this happens when you

NOTE Confidence: 0.90901923

00:30:54.035 --> 00:30:55.890 have a low input material,

NOTE Confidence: 0.90901923

00:30:55.890 --> 00:30:58.110 such as in some tricky libraries.

NOTE Confidence: 0.90901923

00:30:58.110 --> 00:31:00.822 When your sample you have a low amount

NOTE Confidence: 0.90901923

00:31:00.822 --> 00:31:03.872 of sample and this is the case in

NOTE Confidence: 0.90901923

00:31:03.872 --> 00:31:05.900 single cell approaches because again,

NOTE Confidence: 0.90901923

00:31:05.900 --> 00:31:08.679 you're starting from the amount of RNA

NOTE Confidence: 0.90901923

00:31:08.679 --> 00:31:11.358 that is extracted from one single set.

NOTE Confidence: 0.90901923

00:31:11.360 --> 00:31:13.940 So you can imagine to have

NOTE Confidence: 0.90901923

00:31:13.940 --> 00:31:15.660 a lot of amplification.

NOTE Confidence: 0.90901923

00:31:15.660 --> 00:31:19.454 Occurring in order to detect the gene.

NOTE Confidence: 0.90901923

00:31:19.460 --> 00:31:21.580 So this is the definition of the UMI

NOTE Confidence: 0.90901923

00:31:21.580 --> 00:31:24.180 is a randomized nucleotide sequence.

NOTE Confidence: 0.90901923

00:31:24.180 --> 00:31:26.628 Again depending on the library preparation

NOTE Confidence: 0.90901923

00:31:26.628 --> 00:31:29.262 and on the technique that you use

NOTE Confidence: 0.90901923

00:31:29.262 --> 00:31:31.439 that it can be 8 nucleotide longer.

NOTE Confidence: 0.90901923

00:31:31.440 --> 00:31:33.974 12 Nook tight long of the longest,

NOTE Confidence: 0.90901923

00:31:33.980 --> 00:31:34.636 the better.

NOTE Confidence: 0.90901923

00:31:34.636 --> 00:31:37.260 It's incorporated into the C DNA and the

NOTE Confidence: 0.90901923

00:31:37.327 --> 00:31:39.787 initial steps of their native protocol.

NOTE Confidence: 0.90901923

00:31:39.790 --> 00:31:41.600 So before the amplification step,

NOTE Confidence: 0.90901923

00:31:41.600 --> 00:31:44.088 so the goal of the UMI is to

NOTE Confidence: 0.90901923

00:31:44.088 --> 00:31:45.634 distinguish between amplified copies

NOTE Confidence: 0.90901923

00:31:45.634 --> 00:31:47.764 of the same earning molecule.

NOTE Confidence: 0.90901923

00:31:47.770 --> 00:31:50.738 Because these have the same C DNA sequence.

NOTE Confidence: 0.90901923

00:31:50.740 --> 00:31:53.076 But it is and they have the same,

NOTE Confidence: 0.90901923

00:31:53.080 --> 00:31:53.457 Umm,

NOTE Confidence: 0.90901923

00:31:53.457 --> 00:31:55.342 so they are technical duplicates

NOTE Confidence: 0.90901923

00:31:55.342 --> 00:31:56.850 and they are removed.

NOTE Confidence: 0.90901923

00:31:56.850 --> 00:31:57.119 Well,

NOTE Confidence: 0.90901923

00:31:57.119 --> 00:31:59.002 what you want to keep his reads

NOTE Confidence: 0.90901923

00:31:59.002 --> 00:32:00.310 from separator marine molecules

NOTE Confidence: 0.90901923

00:32:00.310 --> 00:32:02.085 transcribed from the same gene?

NOTE Confidence: 0.90901923

00:32:02.090 --> 00:32:03.824 Because these will have the same

NOTE Confidence: 0.90901923

00:32:03.824 --> 00:32:06.088 C DNA but will have a different.

NOTE Confidence: 0.90901923

00:32:06.090 --> 00:32:06.413 Umm,

NOTE Confidence: 0.90901923

00:32:06.413 --> 00:32:08.351 I so these are biological duplicates

NOTE Confidence: 0.90901923

00:32:08.351 --> 00:32:10.398 and they are kept so they are.

NOTE Confidence: 0.90901923

00:32:10.400 --> 00:32:13.538 My is a method to reduce

NOTE Confidence: 0.90901923

00:32:13.538 --> 00:32:15.107 the amplification noise.

NOTE Confidence: 0.90901923

00:32:15.110 --> 00:32:16.710 This is a graphical example

NOTE Confidence: 0.90901923

00:32:16.710 --> 00:32:17.670 of the importance,

NOTE Confidence: 0.90901923

00:32:17.670 --> 00:32:19.416 so this is an example where

NOTE Confidence: 0.90901923

00:32:19.416 --> 00:32:21.190 you have a reference sequence,

NOTE Confidence: 0.90901923

00:32:21.190 --> 00:32:23.750 so this is a region of a gene,

NOTE Confidence: 0.90901923

00:32:23.750 --> 00:32:25.670 for example, and in your experiment,

NOTE Confidence: 0.90901923

00:32:25.670 --> 00:32:27.272 for example in the same cell

NOTE Confidence: 0.90901923

00:32:27.272 --> 00:32:28.340 you get 10 reads

NOTE Confidence: 0.97816813

00:32:28.404 --> 00:32:29.829 with identical sequence,

NOTE Confidence: 0.97816813

00:32:29.830 --> 00:32:32.710 and so that they align to the same region.

NOTE Confidence: 0.97816813

00:32:32.710 --> 00:32:35.270 So if we assume that they are all

NOTE Confidence: 0.97816813

00:32:35.270 --> 00:32:37.170 PCR duplicates, we have to remove

NOTE Confidence: 0.97816813

00:32:37.170 --> 00:32:39.430 all of them and keep only one.

NOTE Confidence: 0.97816813

00:32:39.430 --> 00:32:41.698 And that means that when we calculate

NOTE Confidence: 0.97816813

00:32:41.698 --> 00:32:44.389 the abundance of the gene, if we don't.

NOTE Confidence: 0.97816813

00:32:44.389 --> 00:32:47.142 Remove the duplicate. We will say this.

NOTE Confidence: 0.97816813

00:32:47.142 --> 00:32:49.590 Gina is his account of 10.

NOTE Confidence: 0.97816813

00:32:49.590 --> 00:32:52.271 After the duplication we would say that

NOTE Confidence: 0.97816813

00:32:52.271 --> 00:32:55.378 the gene has account of one because by

NOTE Confidence: 0.97816813

00:32:55.378 --> 00:32:58.253 by using this approach we assume that

NOTE Confidence: 0.97816813

00:32:58.253 --> 00:33:01.013 all the duplicates are PCR duplicates.

NOTE Confidence: 0.97816813

00:33:01.020 --> 00:33:03.468 If we include the UM eyes,

NOTE Confidence: 0.97816813

00:33:03.470 --> 00:33:05.098 we can separate technical

NOTE Confidence: 0.97816813

00:33:05.098 --> 00:33:06.319 from biological duplicates.

NOTE Confidence: 0.97816813

00:33:06.320 --> 00:33:09.983 So we can use the Umm I hear different.

NOTE Confidence: 0.97816813

00:33:09.990 --> 00:33:12.252 Umm eyes are different colors in

NOTE Confidence: 0.97816813

00:33:12.252 --> 00:33:14.560 order to group technical duplicates.

NOTE Confidence: 0.97816813

00:33:14.560 --> 00:33:15.368 For example,

NOTE Confidence: 0.97816813

00:33:15.368 --> 00:33:18.600 these four at these three and these two,

NOTE Confidence: 0.97816813

00:33:18.600 --> 00:33:21.018 but we keep the biological duplicates.

NOTE Confidence: 0.97816813

00:33:21.020 --> 00:33:22.640 So in the end,

NOTE Confidence: 0.97816813

00:33:22.640 --> 00:33:24.260 instead of collapsing everything,

NOTE Confidence: 0.97816813

00:33:24.260 --> 00:33:26.666 we can keep four reads because

NOTE Confidence: 0.97816813

00:33:26.666 --> 00:33:28.700 they having four different mice,

NOTE Confidence: 0.97816813

00:33:28.700 --> 00:33:31.082 they probably correspond to four different

NOTE Confidence: 0.97816813

00:33:31.082 --> 00:33:33.140 original molecules in our sample.

NOTE Confidence: 0.97642732

00:33:38.540 --> 00:33:43.246 Is everything clear? Do you? Hear me

NOTE Confidence: 0.97642732

00:33:43.246 --> 00:33:45.150 I don't have interaction. Yes, yes.

NOTE Confidence: 0.99074006

00:33:47.220 --> 00:33:47.900 Yes, OK.

NOTE Confidence: 0.963889

00:33:50.020 --> 00:33:52.540 OK, so that's why you use the cell

NOTE Confidence: 0.963889

00:33:52.540 --> 00:33:55.268 barcode to identify the cell you use it.

NOTE Confidence: 0.963889

00:33:55.270 --> 00:33:57.388 Umm, I to remove technical duplicates

NOTE Confidence: 0.963889

00:33:57.388 --> 00:33:59.800 and that's why instead of counter matrix

NOTE Confidence: 0.963889

00:33:59.800 --> 00:34:01.984 you can find also instead of number

NOTE Confidence: 0.963889

00:34:02.046 --> 00:34:04.494 of reads you can find in single cell

NOTE Confidence: 0.963889

00:34:04.494 --> 00:34:06.354 experiments the number of UM eyes,

NOTE Confidence: 0.963889

00:34:06.354 --> 00:34:07.909 because basically what you are

NOTE Confidence: 0.963889

00:34:07.909 --> 00:34:09.699 doing you are collapsing reads,

NOTE Confidence: 0.963889

00:34:09.700 --> 00:34:11.350 the transcribed so mapping to

NOTE Confidence: 0.963889

00:34:11.350 --> 00:34:13.640 the same gene and with the same.

NOTE Confidence: 0.963889

00:34:13.640 --> 00:34:17.880 Umm I. And after you do all these

NOTE Confidence: 0.963889

00:34:17.880 --> 00:34:19.611 steps that you can, uh,

NOTE Confidence: 0.963889

00:34:19.611 --> 00:34:21.597 you can arrive to your account

NOTE Confidence: 0.963889

00:34:21.597 --> 00:34:23.479 metrics in the single cell.

NOTE Confidence: 0.963889

00:34:23.480 --> 00:34:25.230 It's called digital expression matrix,

NOTE Confidence: 0.963889

00:34:25.230 --> 00:34:28.520 because it represents the number of reads

NOTE Confidence: 0.963889

00:34:28.520 --> 00:34:32.378 mapping to 1 gene in each of your cells.

NOTE Confidence: 0.963889

00:34:32.380 --> 00:34:34.288 And the other all sequence data

NOTE Confidence: 0.963889

00:34:34.288 --> 00:34:36.403 are always end in the balcony

NOTE Confidence: 0.963889

00:34:36.403 --> 00:34:38.348 seeking the fast queue format.

NOTE Confidence: 0.963889

00:34:38.350 --> 00:34:40.800 So that's how you receive your sequence.

NOTE Confidence: 0.963889

00:34:40.800 --> 00:34:42.204 One of these steps,

NOTE Confidence: 0.963889

00:34:42.204 --> 00:34:44.310 in order to quantify the expression,

NOTE Confidence: 0.963889

00:34:44.310 --> 00:34:47.190 is that you align not the UMI and the bar

NOTE Confidence: 0.963889

00:34:47.261 --> 00:34:49.925 code because those are only technical,

NOTE Confidence: 0.963889

00:34:49.930 --> 00:34:50.632 but you.

NOTE Confidence: 0.963889

00:34:50.632 --> 00:34:52.387 You align the read corresponding

NOTE Confidence: 0.963889

00:34:52.387 --> 00:34:53.790 to your C DNA.

NOTE Confidence: 0.963889

00:34:53.790 --> 00:34:56.058 The alignment tool for single cell

NOTE Confidence: 0.963889

00:34:56.058 --> 00:34:58.890 RNA seek are on most all are the

NOTE Confidence: 0.963889

00:34:58.890 --> 00:35:01.509 same as the one used for bulk RNA.

NOTE Confidence: 0.963889

00:35:01.510 --> 00:35:02.330 Seek a.

NOTE Confidence: 0.963889

00:35:02.330 --> 00:35:03.560 So here again,

NOTE Confidence: 0.963889

00:35:03.560 --> 00:35:05.522 you see that there are multiple

NOTE Confidence: 0.963889

00:35:05.522 --> 00:35:07.377 options and multiple align alignment

NOTE Confidence: 0.963889

00:35:07.377 --> 00:35:09.317 tools for different applications.

NOTE Confidence: 0.963889

00:35:09.320 --> 00:35:10.472 So for example,

NOTE Confidence: 0.963889

00:35:10.472 --> 00:35:13.160 here you see the years of publication.

NOTE Confidence: 0.963889

00:35:13.160 --> 00:35:14.580 It's not really updated.

NOTE Confidence: 0.963889

00:35:14.580 --> 00:35:17.174 The methods that you see in red

NOTE Confidence: 0.963889

00:35:17.174 --> 00:35:19.508 are the ones that were developed

NOTE Confidence: 0.963889

00:35:19.508 --> 00:35:21.230 specifically for any seeker,

NOTE Confidence: 0.963889

00:35:21.230 --> 00:35:24.278 and the star that you see here was

NOTE Confidence: 0.963889

00:35:24.278 --> 00:35:26.987 developed like it was released in 2012.

NOTE Confidence: 0.963889

00:35:26.990 --> 00:35:30.131 So almost ten years ago is one of the

NOTE Confidence: 0.963889

00:35:30.131 --> 00:35:32.457 standard alignment tools for back.

NOTE Confidence: 0.963889

00:35:32.460 --> 00:35:35.309 Kearney seek, we saw this with Everett.

NOTE Confidence: 0.963889

00:35:35.310 --> 00:35:37.340 I think two weeks ago.

NOTE Confidence: 0.963889

00:35:37.340 --> 00:35:39.788 It's also the most common tool.

NOTE Confidence: 0.963889

00:35:39.790 --> 00:35:42.226 The default tools in many single

NOTE Confidence: 0.963889

00:35:42.226 --> 00:35:43.038 cell pipelines.

NOTE Confidence: 0.963889

00:35:43.040 --> 00:35:45.735 So almost all of them will use

NOTE Confidence: 0.963889

00:35:45.735 --> 00:35:48.329 star or high SAT or another.

NOTE Confidence: 0.963889

00:35:48.330 --> 00:35:49.506 Ernie Caecum, splines,

NOTE Confidence: 0.963889

00:35:49.506 --> 00:35:51.466 aware aligner tool in order

NOTE Confidence: 0.963889

00:35:51.466 --> 00:35:53.219 to perform the alignment,

NOTE Confidence: 0.963889

00:35:53.220 --> 00:35:55.662 so this is not so different

NOTE Confidence: 0.963889

00:35:55.662 --> 00:35:57.290 from the balcony sick.

NOTE Confidence: 0.963889

00:35:57.290 --> 00:35:57.700 Also,

NOTE Confidence: 0.963889

00:35:57.700 --> 00:35:59.750 the alignment output that you

NOTE Confidence: 0.963889

00:35:59.750 --> 00:36:02.526 receive will be a bum file, so.

NOTE Confidence: 0.963889

00:36:02.526 --> 00:36:05.018 This is a file where each original

NOTE Confidence: 0.963889

00:36:05.018 --> 00:36:06.528 reader containers the information

NOTE Confidence: 0.963889

00:36:06.528 --> 00:36:09.097 on the alignment so it contains the.

NOTE Confidence: 0.963889

00:36:09.100 --> 00:36:10.900 This file contains the coordinate,

NOTE Confidence: 0.963889

00:36:10.900 --> 00:36:13.054 so the chromosome and the genomic

NOTE Confidence: 0.963889

00:36:13.054 --> 00:36:14.490 coordinator of the alignment,

NOTE Confidence: 0.963889

00:36:14.490 --> 00:36:16.583 and you need to use this file

NOTE Confidence: 0.963889

00:36:16.583 --> 00:36:18.672 in order to calculate the number

NOTE Confidence: 0.963889

00:36:18.672 --> 00:36:21.276 of reads that map to each gene

NOTE Confidence: 0.963889

00:36:21.353 --> 00:36:22.739 or each transcript.

NOTE Confidence: 0.963889

00:36:22.740 --> 00:36:24.882 So also this is not different

NOTE Confidence: 0.963889

00:36:24.882 --> 00:36:27.049 from the bunker and a secret.

NOTE Confidence: 0.963889

00:36:27.050 --> 00:36:28.980 The only difference is for

NOTE Confidence: 0.963889

00:36:28.980 --> 00:36:31.450 example you can have a different

NOTE Confidence: 0.963889

00:36:31.450 --> 00:36:33.745 band files for each cell.

NOTE Confidence: 0.963889

00:36:33.750 --> 00:36:35.660 Instead of only one bonfire.

NOTE Confidence: 0.95616823

00:36:38.160 --> 00:36:40.680 OK, so we what we covered so far is the

NOTE Confidence: 0.95616823

00:36:40.751 --> 00:36:43.330 first data preprocessing, so again we

NOTE Confidence: 0.95616823

00:36:43.330 --> 00:36:45.940 have cell barcode you MI and the RNA.

NOTE Confidence: 0.95616823

00:36:45.940 --> 00:36:48.292 You cluster cell according to your cluster

NOTE Confidence: 0.95616823

00:36:48.292 --> 00:36:50.799 reads according to the cell you simplify you,

NOTE Confidence: 0.95616823

00:36:50.800 --> 00:36:52.558 you remove technical duplicates and you

NOTE Confidence: 0.95616823

00:36:52.558 --> 00:36:54.679 arrive to your gene expression matrix.

NOTE Confidence: 0.95616823

00:36:54.680 --> 00:36:56.228 Your digital expression matrix.

NOTE Confidence: 0.95616823

00:36:56.228 --> 00:36:58.550 Now a big difference between account

NOTE Confidence: 0.95616823

00:36:58.616 --> 00:37:00.936 data that you can obtain it back versus

NOTE Confidence: 0.95616823

00:37:00.936 --> 00:37:02.777 single cell is what you see here.

NOTE Confidence: 0.95616823

00:37:02.780 --> 00:37:05.700 So this is a typical account matrix from

NOTE Confidence: 0.95616823

00:37:05.700 --> 00:37:08.835 a balcony sick and you can see the number.

NOTE Confidence: 0.95616823

00:37:08.840 --> 00:37:12.818 Are very high and rarely you see zero values.

NOTE Confidence: 0.95616823

00:37:12.820 --> 00:37:14.806 The single cell RNA seek are.

NOTE Confidence: 0.95616823

00:37:14.810 --> 00:37:17.798 These is what you obtain most of the time.

NOTE Confidence: 0.95616823

00:37:17.800 --> 00:37:20.456 I would say this is a very good.

NOTE Confidence: 0.95616823

00:37:20.460 --> 00:37:21.676 It is very high.

NOTE Confidence: 0.95616823

00:37:21.676 --> 00:37:24.769 Is an example with a low amount of zeros.

NOTE Confidence: 0.95616823

00:37:24.770 --> 00:37:28.460 So what you can see is that the numbers are

NOTE Confidence: 0.95616823

00:37:28.550 --> 00:37:31.894 lower and most of the values are zeros.

NOTE Confidence: 0.95616823

00:37:31.900 --> 00:37:34.820 So the fact that you have lower counts,

NOTE Confidence: 0.95616823

00:37:34.820 --> 00:37:37.298 it means that in all your analysis

NOTE Confidence: 0.95616823

00:37:37.298 --> 00:37:39.627 you will have a higher contribution

NOTE Confidence: 0.95616823

00:37:39.627 --> 00:37:42.413 of noise and this will bring you

NOTE Confidence: 0.95616823

00:37:42.489 --> 00:37:45.069 to a higher uncertainty in results.

NOTE Confidence: 0.95616823

00:37:45.070 --> 00:37:47.625 And if this is a big problem,

NOTE Confidence: 0.95616823

00:37:47.630 --> 00:37:49.814 it means that when you choose

NOTE Confidence: 0.95616823

00:37:49.814 --> 00:37:50.906 among different pipelines,

NOTE Confidence: 0.95616823

00:37:50.910 --> 00:37:54.036 you will have very different results.

NOTE Confidence: 0.95616823

00:37:54.040 --> 00:37:56.176 And but the origin of these is that.

NOTE Confidence: 0.966423

00:37:58.560 --> 00:37:59.910 Your original values,

NOTE Confidence: 0.966423

00:37:59.910 --> 00:38:02.160 your original quantification of expression

NOTE Confidence: 0.966423

00:38:02.160 --> 00:38:04.714 values were generally very low and so

NOTE Confidence: 0.966423

00:38:04.714 --> 00:38:06.790 the contribution of noise is higher.

NOTE Confidence: 0.966423

00:38:06.790 --> 00:38:09.662 So this problem is one of the main

NOTE Confidence: 0.966423

00:38:09.662 --> 00:38:12.301 problem in single cell RNA seek and

NOTE Confidence: 0.966423

00:38:12.301 --> 00:38:15.419 at the moment is kind of unavoidable.

NOTE Confidence: 0.966423

00:38:15.420 --> 00:38:18.138 I would say the second probably

NOTE Confidence: 0.966423

00:38:18.138 --> 00:38:20.580 is you have several zeros.

NOTE Confidence: 0.966423

00:38:20.580 --> 00:38:24.408 And some of these zeros are.

NOTE Confidence: 0.966423

00:38:24.410 --> 00:38:26.420 Real zeros, meaning that in fact

NOTE Confidence: 0.966423

00:38:26.420 --> 00:38:28.790 sell the gene is not expressed,

NOTE Confidence: 0.966423

00:38:28.790 --> 00:38:30.640 so this corresponds to through

NOTE Confidence: 0.966423

00:38:30.640 --> 00:38:32.120 biological zeros they represent

NOTE Confidence: 0.966423

00:38:32.120 --> 00:38:33.900 the true lack of expression,

NOTE Confidence: 0.966423

00:38:33.900 --> 00:38:35.904 but many times the zeros represent

NOTE Confidence: 0.966423

00:38:35.904 --> 00:38:37.920 a technical lack of detection,

NOTE Confidence: 0.966423

00:38:37.920 --> 00:38:39.740 meaning that the gene was

NOTE Confidence: 0.966423

00:38:39.740 --> 00:38:41.196 present in your cell,

NOTE Confidence: 0.966423

00:38:41.200 --> 00:38:44.120 but it was not detected because it was.

NOTE Confidence: 0.966423

00:38:44.120 --> 00:38:46.675 It was not captured by your Beda,

NOTE Confidence: 0.966423

00:38:46.680 --> 00:38:49.326 and so you don't have a way to see

NOTE Confidence: 0.966423

00:38:49.326 --> 00:38:52.040 your gene because you didn't detect

NOTE Confidence: 0.966423

00:38:52.040 --> 00:38:54.380 it in your library preparation.

NOTE Confidence: 0.966423

00:38:54.380 --> 00:38:55.072 And obviously,

NOTE Confidence: 0.966423

00:38:55.072 --> 00:38:57.494 in methods where you have a a

NOTE Confidence: 0.966423

00:38:57.494 --> 00:38:59.797 low coverage inside each cell,

NOTE Confidence: 0.966423

00:38:59.800 --> 00:39:01.730 the probability of these technical

NOTE Confidence: 0.966423

00:39:01.730 --> 00:39:03.274 detection lack of detection.

NOTE Confidence: 0.966423

00:39:03.280 --> 00:39:05.210 This is also called dropout

NOTE Confidence: 0.966423

00:39:05.210 --> 00:39:06.754 effect is very high,

NOTE Confidence: 0.966423

00:39:06.760 --> 00:39:10.243 so I think in the 10X approaches the dropout.

NOTE Confidence: 0.966423

00:39:10.250 --> 00:39:13.362 If you can expect your jeans to be

NOTE Confidence: 0.966423

00:39:13.362 --> 00:39:15.680 not detected with 80% of probability.

NOTE Confidence: 0.966423

00:39:15.680 --> 00:39:17.655 Obviously this depends also on

NOTE Confidence: 0.966423

00:39:17.655 --> 00:39:20.013 whether the gene is highly expressed

NOTE Confidence: 0.966423

00:39:20.013 --> 00:39:21.853 or as low expression level.

NOTE Confidence: 0.966423

00:39:21.860 --> 00:39:25.076 So if a gene has high expression level.

NOTE Confidence: 0.966423

00:39:25.080 --> 00:39:27.060 Capability to be detected at least

NOTE Confidence: 0.966423

00:39:27.060 --> 00:39:28.860 with one molecule is higher,

NOTE Confidence: 0.966423

00:39:28.860 --> 00:39:30.930 but jeans with low expression levels,

NOTE Confidence: 0.966423

00:39:30.930 --> 00:39:32.302 for example transcription factors,

NOTE Confidence: 0.966423

00:39:32.302 --> 00:39:33.674 will rarely be detected,

NOTE Confidence: 0.966423

00:39:33.680 --> 00:39:36.080 but most of the times it will be

NOTE Confidence: 0.966423

00:39:36.080 --> 00:39:38.148 because of a detection problem,

NOTE Confidence: 0.966423

00:39:38.150 --> 00:39:40.214 not because they are not expressed

NOTE Confidence: 0.966423

00:39:40.214 --> 00:39:41.246 in the cell.

NOTE Confidence: 0.966423

00:39:41.250 --> 00:39:43.170 And also this is inherent problem

NOTE Confidence: 0.966423

00:39:43.170 --> 00:39:45.379 with a single cell data analysis.

NOTE Confidence: 0.966423

00:39:45.380 --> 00:39:47.810 So it's very important to.

NOTE Confidence: 0.966423

00:39:47.810 --> 00:39:48.290 Doodle

NOTE Confidence: 0.89605373

00:39:50.780 --> 00:39:51.730 uh no.

NOTE Confidence: 0.97740066

00:39:54.800 --> 00:39:57.747 No, yes, so now we wouldn't be

NOTE Confidence: 0.97740066

00:39:57.747 --> 00:40:01.358 ready at the end of the of the time.

NOTE Confidence: 0.97740066

00:40:01.360 --> 00:40:03.850 So one thing we could do is I I could

NOTE Confidence: 0.97740066

00:40:03.925 --> 00:40:06.217 continue and finish the next time.

NOTE Confidence: 0.9750521

00:40:09.040 --> 00:40:10.660 With the remaining of the analysis

NOTE Confidence: 0.9750521

00:40:10.660 --> 00:40:12.280 steps, sure, yeah, I think Tom,

NOTE Confidence: 0.9750521

00:40:12.280 --> 00:40:13.630 it's your own judgment to

NOTE Confidence: 0.9750521

00:40:13.630 --> 00:40:14.980 how you want to proceed.

NOTE Confidence: 0.9750521

00:40:14.980 --> 00:40:17.410 Like do you think it is a natural stuff?

NOTE Confidence: 0.9750521

00:40:17.410 --> 00:40:19.146 Then you can stop if you think

NOTE Confidence: 0.9750521

00:40:19.146 --> 00:40:21.190 you want to cover 5 more minutes,

NOTE Confidence: 0.9750521

00:40:21.190 --> 00:40:22.810 go ahead and do that so.

NOTE Confidence: 0.8701443

00:40:24.210 --> 00:40:26.937 I can give you like a sort of anticipation

NOTE Confidence: 0.8701443

00:40:26.937 --> 00:40:29.130 on the on the following steps.

NOTE Confidence: 0.8701443

00:40:29.130 --> 00:40:33.106 So uhm, many of these steps that.

NOTE Confidence: 0.8701443

00:40:33.110 --> 00:40:36.334 I mean, many, many of these steps II

NOTE Confidence: 0.8701443

00:40:36.334 --> 00:40:38.856 took inspiration from this review that

NOTE Confidence: 0.8701443

00:40:38.856 --> 00:40:41.930 was recently published in a true method.

NOTE Confidence: 0.8701443

00:40:41.930 --> 00:40:44.450 So it covers the main trials.

NOTE Confidence: 0.8701443

00:40:44.450 --> 00:40:47.084 So the successes and also the

NOTE Confidence: 0.8701443

00:40:47.084 --> 00:40:48.840 limitations of the computational

NOTE Confidence: 0.8701443

00:40:48.913 --> 00:40:51.121 methods for the single cell RNA

NOTE Confidence: 0.8701443

00:40:51.121 --> 00:40:53.994 seek analysis and so next time we'll

NOTE Confidence: 0.8701443

00:40:53.994 --> 00:40:56.204 cover the key preprocessing steps.

NOTE Confidence: 0.8701443

00:40:56.210 --> 00:40:59.150 We have seen this the molecular counting,

NOTE Confidence: 0.8701443

00:40:59.150 --> 00:41:00.718 but we will see.

NOTE Confidence: 0.8701443

00:41:00.718 --> 00:41:03.930 So how we can do quality control?

NOTE Confidence: 0.8701443

00:41:03.930 --> 00:41:06.426 Remove Excel said that are suspicious,

NOTE Confidence: 0.8701443

00:41:06.430 --> 00:41:09.146 for example because they are dying or

NOTE Confidence: 0.8701443

00:41:09.146 --> 00:41:11.714 because they represent the empty droplets

NOTE Confidence: 0.8701443

00:41:11.714 --> 00:41:13.939 or because they represented tablets.

NOTE Confidence: 0.8701443

00:41:13.940 --> 00:41:16.698 So doublets occur when you didn't really

NOTE Confidence: 0.8701443

00:41:16.698 --> 00:41:19.359 manage to separate physically the cells.

NOTE Confidence: 0.8701443

00:41:19.360 --> 00:41:21.808 So there for example in the

NOTE Confidence: 0.8701443

00:41:21.808 --> 00:41:24.359 same droplet you have two cells,

NOTE Confidence: 0.8701443

00:41:24.360 --> 00:41:27.174 or for some reason the cell barcode

NOTE Confidence: 0.8701443

00:41:27.174 --> 00:41:29.778 of two different cells was shared.

NOTE Confidence: 0.8701443

00:41:29.780 --> 00:41:31.548 For some technical problem,

NOTE Confidence: 0.8701443

00:41:31.548 --> 00:41:34.200 so we'll see methods to remove.

NOTE Confidence: 0.8701443

00:41:34.200 --> 00:41:37.072 Dying cell and those who tablets then there

NOTE Confidence: 0.8701443

00:41:37.072 --> 00:41:39.719 are problems related to the normalization.

NOTE Confidence: 0.8701443

00:41:39.720 --> 00:41:42.107 So how to consider how to consider

NOTE Confidence: 0.8701443

00:41:42.107 --> 00:41:44.745 the fact that you have a different

NOTE Confidence: 0.8701443

00:41:44.745 --> 00:41:47.049 read different number of reads in

NOTE Confidence: 0.8701443

00:41:47.125 --> 00:41:50.212 different cells and this is a problem

NOTE Confidence: 0.8701443

00:41:50.212 --> 00:41:52.326 because the biologically speaking you

NOTE Confidence: 0.8701443

00:41:52.326 --> 00:41:55.077 expect the cell of different types to

NOTE Confidence: 0.8701443

00:41:55.077 --> 00:41:57.835 have a different amount over in Asia,

NOTE Confidence: 0.8701443

00:41:57.840 --> 00:42:00.493 so expect some cells you have more

NOTE Confidence: 0.8701443

00:42:00.493 --> 00:42:02.570 any molecules than other cells,

NOTE Confidence: 0.8701443

00:42:02.570 --> 00:42:04.760 but most of the methods.

NOTE Confidence: 0.8701443

00:42:04.760 --> 00:42:07.644 Assume that you needed to have to

NOTE Confidence: 0.8701443

00:42:07.644 --> 00:42:10.956 have from each cell the same number of

NOTE Confidence: 0.8701443

00:42:10.956 --> 00:42:14.160 reads or UMI. And then we will see how.

NOTE Confidence: 0.8701443

00:42:14.160 --> 00:42:16.720 So how to remove a jeans that are

NOTE Confidence: 0.8701443

00:42:16.720 --> 00:42:18.519 not important in the analysis?

NOTE Confidence: 0.8701443

00:42:18.520 --> 00:42:20.440 This is very important because as

NOTE Confidence: 0.8701443

00:42:20.440 --> 00:42:22.856 you can imagine in single cell you

NOTE Confidence: 0.8701443

00:42:22.856 --> 00:42:24.998 have thousands of genes and also

NOTE Confidence: 0.8701443

00:42:24.998 --> 00:42:26.220 thousands of cells are.

NOTE Confidence: 0.8701443

00:42:26.220 --> 00:42:28.677 So your account matrix is very highly

NOTE Confidence: 0.8701443

00:42:28.677 --> 00:42:31.004 dimensional and so all of these methods

NOTE Confidence: 0.8701443

00:42:31.004 --> 00:42:33.217 try to reduce the number of cells

NOTE Confidence: 0.8701443

00:42:33.217 --> 00:42:35.251 are keeping only the high quality

NOTE Confidence: 0.8701443

00:42:35.251 --> 00:42:37.678 sales but also the number of genes.

NOTE Confidence: 0.8701443

00:42:37.678 --> 00:42:39.438 So reducing like the dimensionality

NOTE Confidence: 0.8701443

00:42:39.438 --> 00:42:40.628 of of your data.

NOTE Confidence: 0.87640333

00:42:45.560 --> 00:42:46.360 Leash.

NOTE Confidence: 0.9874245

00:42:47.930 --> 00:42:50.898 That sounds amazing. I look forward to

NOTE Confidence: 0.9874245

00:42:50.900 --> 00:42:53.140 next week. Yeah. And also next week

NOTE Confidence: 0.9874245

00:42:53.140 --> 00:42:55.415 when you see the downstream analysis

NOTE Confidence: 0.9874245

00:42:55.415 --> 00:42:58.313 that for example class ring for single

NOTE Confidence: 0.9874245

00:42:58.381 --> 00:43:01.051 cell approaches and also trajectory

NOTE Confidence: 0.9874245

00:43:01.051 --> 00:43:03.187 possibly also trajectory estimation.

NOTE Confidence: 0.93677676

00:43:04.840 --> 00:43:07.045 Yeah, thanks so much and nothing really.

NOTE Confidence: 0.93677676

00:43:07.050 --> 00:43:09.493 Want to comment. Is your your slides

NOTE Confidence: 0.93677676

00:43:09.493 --> 00:43:12.073 look very beautiful and I wish all my

NOTE Confidence: 0.93677676

00:43:12.073 --> 00:43:14.290 slides and possibly on people in my life.

NOTE Confidence: 0.93677676

00:43:14.290 --> 00:43:16.495 There's as well looks as nice when

NOTE Confidence: 0.93677676

00:43:16.500 --> 00:43:19.036 I try and try to put a course

NOTE Confidence: 0.93677676

00:43:19.036 --> 00:43:21.216 on this so it's kind of a

NOTE Confidence: 0.93677676

00:43:21.220 --> 00:43:23.796 maybe we should use your slides as template.