

WEBVTT

NOTE duration:"00:44:07"

NOTE recognizability:0.962

NOTE language:en-us

NOTE Confidence: 0.987230371666667

00:00:04.780 --> 00:00:07.966 So today we will finish our.

NOTE Confidence: 0.987230371666667

00:00:07.970 --> 00:00:11.696 Session on single cell data analysis.

NOTE Confidence: 0.987230371666667

00:00:11.700 --> 00:00:15.633 So far we arrived at a this step in

NOTE Confidence: 0.987230371666667

00:00:15.633 --> 00:00:19.855 the analysis, so after all the quality

NOTE Confidence: 0.987230371666667

00:00:19.855 --> 00:00:21.987 controller and the normalization,

NOTE Confidence: 0.987230371666667

00:00:21.990 --> 00:00:25.266 a necessary step to reduce the complexity

NOTE Confidence: 0.987230371666667

00:00:25.266 --> 00:00:29.054 of the data set is the reduction

NOTE Confidence: 0.987230371666667

00:00:29.054 --> 00:00:31.350 of dimension of dimensionality.

NOTE Confidence: 0.987230371666667

00:00:31.350 --> 00:00:32.670 So reduction of features.

NOTE Confidence: 0.987230371666667

00:00:32.670 --> 00:00:34.650 So there are two possible ways.

NOTE Confidence: 0.987230371666667

00:00:34.650 --> 00:00:36.862 Feature selection extract only

NOTE Confidence: 0.987230371666667

00:00:36.862 --> 00:00:40.070 relevant genes and also methods

NOTE Confidence: 0.987230371666667

00:00:40.070 --> 00:00:42.170 for dimensionality reduction.

NOTE Confidence: 0.987230371666667

00:00:42.170 --> 00:00:43.800 So last time we saw.
NOTE Confidence: 0.987230371666667

00:00:43.800 --> 00:00:46.032 Uh, the principal component analysis and
NOTE Confidence: 0.987230371666667

00:00:46.032 --> 00:00:48.867 other tools that are used for single cell,
NOTE Confidence: 0.987230371666667

00:00:48.870 --> 00:00:51.528 especially for the visualization that are
NOTE Confidence: 0.987230371666667

00:00:51.528 --> 00:00:54.860 they Disney method and the UMAP methods?
NOTE Confidence: 0.987230371666667

00:00:54.860 --> 00:00:58.920 They're both nonlinear and the graph based.
NOTE Confidence: 0.987230371666667

00:00:58.920 --> 00:01:01.332 So today we will see briefly
NOTE Confidence: 0.987230371666667

00:01:01.332 --> 00:01:02.538 the remaining steps,
NOTE Confidence: 0.987230371666667

00:01:02.540 --> 00:01:05.160 downstream steps of the analysis.
NOTE Confidence: 0.987230371666667

00:01:05.160 --> 00:01:08.688 And we will start from the clustering.
NOTE Confidence: 0.934274135

00:01:12.710 --> 00:01:14.358 Of single cell data.
NOTE Confidence: 0.934274135

00:01:14.358 --> 00:01:17.363 So we have our cells are there
NOTE Confidence: 0.934274135

00:01:17.363 --> 00:01:20.389 mapped in our low dimensional space
NOTE Confidence: 0.934274135

00:01:20.389 --> 00:01:23.827 and we want to identify clusters,
NOTE Confidence: 0.934274135

00:01:23.830 --> 00:01:26.410 meaning cells that have a similar
NOTE Confidence: 0.934274135

00:01:26.410 --> 00:01:28.565 expression signature so that they

NOTE Confidence: 0.934274135
00:01:28.565 --> 00:01:30.707 are very similar to each other.
NOTE Confidence: 0.934274135
00:01:30.710 --> 00:01:33.200 So since this problem of clustering
NOTE Confidence: 0.934274135
00:01:33.200 --> 00:01:35.979 is quite general and it was also
NOTE Confidence: 0.934274135
00:01:35.979 --> 00:01:38.067 covered by June in the first
NOTE Confidence: 0.934274135
00:01:38.067 --> 00:01:40.417 lesson of these didactic section,
NOTE Confidence: 0.934274135
00:01:40.420 --> 00:01:42.530 many of the methods that.
NOTE Confidence: 0.934274135
00:01:42.530 --> 00:01:45.466 Can be used for single cell data are
NOTE Confidence: 0.934274135
00:01:45.466 --> 00:01:48.407 the same that he covered last time.
NOTE Confidence: 0.934274135
00:01:48.410 --> 00:01:49.526 So here is what you see.
NOTE Confidence: 0.934274135
00:01:49.530 --> 00:01:52.330 Here is an example of a data set.
NOTE Confidence: 0.934274135
00:01:52.330 --> 00:01:53.914 Uh of cells,
NOTE Confidence: 0.934274135
00:01:53.914 --> 00:01:57.610 and I think they are divided because
NOTE Confidence: 0.934274135
00:01:57.721 --> 00:02:02.120 of a different developmental stage.
NOTE Confidence: 0.934274135
00:02:02.120 --> 00:02:04.832 And so these clusters that you
NOTE Confidence: 0.934274135
00:02:04.832 --> 00:02:08.342 see here are a derived are known
NOTE Confidence: 0.934274135

00:02:08.342 --> 00:02:10.572 because of the experiments was
NOTE Confidence: 0.934274135

00:02:10.572 --> 00:02:13.703 that these cells were isolated on
NOTE Confidence: 0.934274135

00:02:13.703 --> 00:02:15.416 different different differentiation
NOTE Confidence: 0.934274135

00:02:15.416 --> 00:02:18.154 States and there are already
NOTE Confidence: 0.934274135

00:02:18.154 --> 00:02:20.266 mapped in a low dimensional space.
NOTE Confidence: 0.934274135

00:02:20.266 --> 00:02:23.479 So what you see here is that they are
NOTE Confidence: 0.934274135

00:02:23.479 --> 00:02:25.524 the first two principal components.
NOTE Confidence: 0.934274135

00:02:25.530 --> 00:02:28.035 So usually clustering always per
NOTE Confidence: 0.934274135

00:02:28.035 --> 00:02:30.039 calculated on principal components
NOTE Confidence: 0.934274135

00:02:30.039 --> 00:02:32.838 or on a reduced dimensional space.
NOTE Confidence: 0.934274135

00:02:32.840 --> 00:02:34.345 This doesn't mean that you
NOTE Confidence: 0.934274135

00:02:34.345 --> 00:02:35.850 only take the first two.
NOTE Confidence: 0.934274135

00:02:35.850 --> 00:02:37.740 You have to select the
NOTE Confidence: 0.934274135

00:02:37.740 --> 00:02:39.252 first step 1020 thirty.
NOTE Confidence: 0.934274135

00:02:39.260 --> 00:02:43.610 There are methods to select for
NOTE Confidence: 0.934274135

00:02:43.610 --> 00:02:45.916 a certain number of dimensions

NOTE Confidence: 0.934274135
00:02:45.916 --> 00:02:48.752 so that you keep what could be
NOTE Confidence: 0.934274135
00:02:48.752 --> 00:02:50.712 worthy of information and you
NOTE Confidence: 0.934274135
00:02:50.712 --> 00:02:53.414 remove the lower dimensions that
NOTE Confidence: 0.934274135
00:02:53.414 --> 00:02:56.294 are associated with less violence.
NOTE Confidence: 0.934274135
00:02:56.300 --> 00:02:59.050 And the assumption is that
NOTE Confidence: 0.934274135
00:02:59.050 --> 00:03:01.250 they mostly capture noise.
NOTE Confidence: 0.934274135
00:03:01.250 --> 00:03:02.874 But here the example in the examples.
NOTE Confidence: 0.934274135
00:03:02.880 --> 00:03:05.340 Here is a simplified example,
NOTE Confidence: 0.934274135
00:03:05.340 --> 00:03:08.364 so you see only PC1 and PC2.
NOTE Confidence: 0.934274135
00:03:08.370 --> 00:03:10.800 So in order to class so we you have
NOTE Confidence: 0.934274135
00:03:10.800 --> 00:03:13.369 a the two approaches that June
NOTE Confidence: 0.934274135
00:03:13.369 --> 00:03:15.594 covered in the clustering lesson
NOTE Confidence: 0.934274135
00:03:15.670 --> 00:03:17.858 was a hierarchical clustering.
NOTE Confidence: 0.934274135
00:03:17.860 --> 00:03:20.200 So these methods try to
NOTE Confidence: 0.934274135
00:03:20.200 --> 00:03:21.136 connect progressively.
NOTE Confidence: 0.934274135

00:03:21.140 --> 00:03:25.568 A cells that are similar to each other.

NOTE Confidence: 0.934274135

00:03:25.568 --> 00:03:28.166 So also in this case if you remember

NOTE Confidence: 0.934274135

00:03:28.166 --> 00:03:30.398 there is the concept of distance.

NOTE Confidence: 0.934274135

00:03:30.400 --> 00:03:32.095 So all clustering methods are

NOTE Confidence: 0.934274135

00:03:32.095 --> 00:03:34.402 based on the fact that you have

NOTE Confidence: 0.934274135

00:03:34.402 --> 00:03:36.454 to calculate a sort of distance

NOTE Confidence: 0.934274135

00:03:36.454 --> 00:03:38.020 or similarity measure between.

NOTE Confidence: 0.934274135

00:03:38.020 --> 00:03:39.019 Pairs of cells.

NOTE Confidence: 0.934274135

00:03:39.019 --> 00:03:40.018 So for example,

NOTE Confidence: 0.934274135

00:03:40.020 --> 00:03:42.678 here you can measure the distance

NOTE Confidence: 0.934274135

00:03:42.678 --> 00:03:44.796 as the Euclidean distance in

NOTE Confidence: 0.934274135

00:03:44.796 --> 00:03:46.308 this principle component space.

NOTE Confidence: 0.934274135

00:03:46.308 --> 00:03:48.388 Otherwise you can use other

NOTE Confidence: 0.934274135

00:03:48.388 --> 00:03:49.860 measures of distance,

NOTE Confidence: 0.934274135

00:03:49.860 --> 00:03:52.948 for example correlation and so on and.

NOTE Confidence: 0.934274135

00:03:52.948 --> 00:03:56.288 The choice of the distance,

NOTE Confidence: 0.934274135
00:03:56.290 --> 00:03:58.528 the choice of the distance will
NOTE Confidence: 0.934274135
00:03:58.528 --> 00:04:00.020 change the clustering results.
NOTE Confidence: 0.934274135
00:04:00.020 --> 00:04:01.640 So these old stands true.
NOTE Confidence: 0.934274135
00:04:01.640 --> 00:04:03.992 So hierarchical clustering try
NOTE Confidence: 0.934274135
00:04:03.992 --> 00:04:06.344 to connect progressively similar
NOTE Confidence: 0.934274135
00:04:06.344 --> 00:04:10.588 entities from app until arriving
NOTE Confidence: 0.934274135
00:04:10.588 --> 00:04:12.444 to unifying everything.
NOTE Confidence: 0.934274135
00:04:12.444 --> 00:04:15.276 And then the problem of hierarchical
NOTE Confidence: 0.934274135
00:04:15.276 --> 00:04:18.360 clustering is to decide when to cut the tree.
NOTE Confidence: 0.934274135
00:04:18.360 --> 00:04:20.296 So depending on where you cut the tree,
NOTE Confidence: 0.934274135
00:04:20.300 --> 00:04:22.064 you can separate like if you cut
NOTE Confidence: 0.934274135
00:04:22.064 --> 00:04:23.792 this tree here, you separate.
NOTE Confidence: 0.934274135
00:04:23.792 --> 00:04:25.250 Two clusters here.
NOTE Confidence: 0.934274135
00:04:25.250 --> 00:04:27.080 You separate three and so on.
NOTE Confidence: 0.934274135
00:04:27.080 --> 00:04:28.388 In this example,
NOTE Confidence: 0.934274135

00:04:28.388 --> 00:04:31.098 you know that there are 12345
NOTE Confidence: 0.934274135

00:04:31.098 --> 00:04:32.682 clusters main clusters,
NOTE Confidence: 0.934274135

00:04:32.682 --> 00:04:36.759 but this information is not always a obvious.
NOTE Confidence: 0.934274135

00:04:36.760 --> 00:04:39.064 So always in clustering the number
NOTE Confidence: 0.934274135

00:04:39.064 --> 00:04:41.040 of cluster the optimal number
NOTE Confidence: 0.934274135

00:04:41.040 --> 00:04:42.890 of clusters that represent your
NOTE Confidence: 0.934274135

00:04:42.890 --> 00:04:45.910 data is always a tricky choice
NOTE Confidence: 0.934274135

00:04:45.910 --> 00:04:49.090 and ultimately subjective.
NOTE Confidence: 0.934274135

00:04:49.090 --> 00:04:50.945 The second approach that Jonah
NOTE Confidence: 0.934274135

00:04:50.945 --> 00:04:52.800 explained last in the lesson
NOTE Confidence: 0.887313348833333

00:04:52.869 --> 00:04:55.365 about clustering, is K means.
NOTE Confidence: 0.887313348833333

00:04:55.365 --> 00:04:58.690 Clustering is based on the fact that
NOTE Confidence: 0.887313348833333

00:04:58.798 --> 00:05:01.354 you select a priority before beginning
NOTE Confidence: 0.887313348833333

00:05:01.354 --> 00:05:04.435 a number of cluster that is key in
NOTE Confidence: 0.887313348833333

00:05:04.435 --> 00:05:06.752 which you want to divide your data,
NOTE Confidence: 0.887313348833333

00:05:06.760 --> 00:05:11.058 and then you apply a sort of

NOTE Confidence: 0.887313348833333
00:05:11.058 --> 00:05:13.762 iterative procedures that is based
NOTE Confidence: 0.887313348833333
00:05:13.762 --> 00:05:16.348 on the definition of a centroid.
NOTE Confidence: 0.887313348833333
00:05:16.350 --> 00:05:17.688 Centroid is the.
NOTE Confidence: 0.887313348833333
00:05:17.688 --> 00:05:19.918 Average point of a cluster
NOTE Confidence: 0.887313348833333
00:05:19.918 --> 00:05:22.220 so it's not real point,
NOTE Confidence: 0.887313348833333
00:05:22.220 --> 00:05:23.940 but it's a appointed represents
NOTE Confidence: 0.887313348833333
00:05:23.940 --> 00:05:26.091 the average of all the points
NOTE Confidence: 0.887313348833333
00:05:26.091 --> 00:05:27.896 that belong to the cluster,
NOTE Confidence: 0.887313348833333
00:05:27.900 --> 00:05:31.351 and so the procedure is to iteratively
NOTE Confidence: 0.887313348833333
00:05:31.351 --> 00:05:34.178 assign each cell to the nearest
NOTE Confidence: 0.887313348833333
00:05:34.180 --> 00:05:37.720 centroid until you reach convergence.
NOTE Confidence: 0.887313348833333
00:05:37.720 --> 00:05:38.869 So until you.
NOTE Confidence: 0.990699398
00:05:40.970 --> 00:05:42.440 In consecutive iterations,
NOTE Confidence: 0.990699398
00:05:42.440 --> 00:05:45.024 for example, you don't have any
NOTE Confidence: 0.990699398
00:05:45.024 --> 00:05:47.430 change of lab labels between sets.
NOTE Confidence: 0.990699398

00:05:47.430 --> 00:05:50.124 A family of methods of clustering

NOTE Confidence: 0.990699398

00:05:50.124 --> 00:05:52.379 methods that are widely used

NOTE Confidence: 0.990699398

00:05:52.379 --> 00:05:54.647 in a single cell approaches is

NOTE Confidence: 0.990699398

00:05:54.647 --> 00:05:57.360 they are the graph based family.

NOTE Confidence: 0.990699398

00:05:57.360 --> 00:05:59.292 So this is something that John

NOTE Confidence: 0.990699398

00:05:59.292 --> 00:06:01.730 didn't talk about, so the principle

NOTE Confidence: 0.990699398

00:06:01.730 --> 00:06:05.160 here is to build them a graph.

NOTE Confidence: 0.990699398

00:06:05.160 --> 00:06:08.408 Uh, on this space and they usually

NOTE Confidence: 0.990699398

00:06:08.408 --> 00:06:11.560 mean it is a so called key.

NOTE Confidence: 0.990699398

00:06:11.560 --> 00:06:12.949 Nearest neighbor graph.

NOTE Confidence: 0.990699398

00:06:12.949 --> 00:06:15.727 That means that for every cell

NOTE Confidence: 0.990699398

00:06:15.727 --> 00:06:18.570 you draw a line that connected the

NOTE Confidence: 0.990699398

00:06:18.570 --> 00:06:21.166 seller with a top nearest cells

NOTE Confidence: 0.990699398

00:06:21.166 --> 00:06:23.616 and that's the key parameter.

NOTE Confidence: 0.990699398

00:06:23.620 --> 00:06:26.385 So for example in this example here,

NOTE Confidence: 0.990699398

00:06:26.390 --> 00:06:30.660 this is a 10 nearest neighbor graph,

NOTE Confidence: 0.990699398
00:06:30.660 --> 00:06:32.886 so it means that each cell
NOTE Confidence: 0.990699398
00:06:32.886 --> 00:06:35.079 here is connected to the top.
NOTE Confidence: 0.990699398
00:06:35.080 --> 00:06:36.748 10 nearest excells.
NOTE Confidence: 0.990699398
00:06:36.748 --> 00:06:40.139 So after you do this basically
NOTE Confidence: 0.990699398
00:06:40.139 --> 00:06:43.793 Europe map becomes a graph and
NOTE Confidence: 0.990699398
00:06:43.793 --> 00:06:47.438 so a graph is a set of nodes
NOTE Confidence: 0.990699398
00:06:47.438 --> 00:06:49.720 that and each node here is a
NOTE Confidence: 0.990699398
00:06:49.803 --> 00:06:52.269 cell with a set of connections.
NOTE Confidence: 0.990699398
00:06:52.270 --> 00:06:54.566 So these connections can
NOTE Confidence: 0.990699398
00:06:54.566 --> 00:06:57.436 can also be waited so.
NOTE Confidence: 0.990699398
00:06:57.440 --> 00:07:00.135 A weight can be assigned to eat
NOTE Confidence: 0.990699398
00:07:00.135 --> 00:07:02.043 connection depending on how similar
NOTE Confidence: 0.990699398
00:07:02.043 --> 00:07:04.648 the two cells are and the method.
NOTE Confidence: 0.990699398
00:07:04.648 --> 00:07:07.560 Once you build that the graph is
NOTE Confidence: 0.990699398
00:07:07.656 --> 00:07:10.314 to identify inside this graph that
NOTE Confidence: 0.990699398

00:07:10.314 --> 00:07:13.161 can be also seen as a network.

NOTE Confidence: 0.990699398

00:07:13.161 --> 00:07:15.725 Basically to identify communities

NOTE Confidence: 0.990699398

00:07:15.725 --> 00:07:19.590 so to identify inside this graph.

NOTE Confidence: 0.990699398

00:07:19.590 --> 00:07:22.348 Communities of nodes are so clusters of

NOTE Confidence: 0.990699398

00:07:22.348 --> 00:07:24.992 nodes that are highly interconnected

NOTE Confidence: 0.990699398

00:07:24.992 --> 00:07:29.222 among themselves and with low

NOTE Confidence: 0.990699398

00:07:29.222 --> 00:07:30.966 interconnections with other clusters.

NOTE Confidence: 0.990699398

00:07:30.970 --> 00:07:32.298 So obviously if like,

NOTE Confidence: 0.990699398

00:07:32.298 --> 00:07:33.958 such as in this case,

NOTE Confidence: 0.990699398

00:07:33.960 --> 00:07:36.972 you obtain not a single network

NOTE Confidence: 0.990699398

00:07:36.972 --> 00:07:40.414 but a different networks that are

NOTE Confidence: 0.990699398

00:07:40.414 --> 00:07:42.588 completely separated, it's easier.

NOTE Confidence: 0.990699398

00:07:42.588 --> 00:07:44.658 It's obviously easier to separate

NOTE Confidence: 0.990699398

00:07:44.658 --> 00:07:45.900 these three clusters,

NOTE Confidence: 0.990699398

00:07:45.900 --> 00:07:48.147 because the in the graph they don't

NOTE Confidence: 0.990699398

00:07:48.147 --> 00:07:50.844 share any connections, but sometimes the.

NOTE Confidence: 0.990699398
00:07:50.844 --> 00:07:52.578 Graph based approach.

NOTE Confidence: 0.990699398
00:07:52.580 --> 00:07:56.076 Tried tried also to cut to within uh,

NOTE Confidence: 0.990699398
00:07:56.080 --> 00:07:59.350 these networks in order to yes.

NOTE Confidence: 0.976010583333333
00:08:01.800 --> 00:08:03.588 Sorry it was there a question.

NOTE Confidence: 0.956195296666667
00:08:06.430 --> 00:08:07.474 That was veins veins.

NOTE Confidence: 0.956195296666667
00:08:07.474 --> 00:08:08.779 Did you have a question?

NOTE Confidence: 0.983169928571429
00:08:11.200 --> 00:08:14.784 Sorry, I accidentally had my microphone on.

NOTE Confidence: 0.983169928571429
00:08:14.790 --> 00:08:19.680 Sorry. And so these measures try to try to

NOTE Confidence: 0.983169928571429
00:08:19.680 --> 00:08:23.447 divide a network inside the communities,

NOTE Confidence: 0.983169928571429
00:08:23.450 --> 00:08:25.921 so they try to cut the networks

NOTE Confidence: 0.983169928571429
00:08:25.921 --> 00:08:28.800 in order to increase the density,

NOTE Confidence: 0.983169928571429
00:08:28.800 --> 00:08:32.550 increase the density of the chunks

NOTE Confidence: 0.983169928571429
00:08:32.550 --> 00:08:35.988 in which the network is divided.

NOTE Confidence: 0.983169928571429
00:08:35.990 --> 00:08:38.727 So here I have a slide that

NOTE Confidence: 0.983169928571429
00:08:38.727 --> 00:08:40.620 explains these a better,

NOTE Confidence: 0.983169928571429

00:08:40.620 --> 00:08:42.908 but the principle is that you built at
NOTE Confidence: 0.983169928571429

00:08:42.908 --> 00:08:45.335 the graph where each node is a cell
NOTE Confidence: 0.983169928571429

00:08:45.335 --> 00:08:46.928 connected to the nearest neighbor
NOTE Confidence: 0.983169928571429

00:08:46.928 --> 00:08:49.266 and then you try to identify the
NOTE Confidence: 0.983169928571429

00:08:49.266 --> 00:08:51.611 communities by creating the cuts
NOTE Confidence: 0.983169928571429

00:08:51.611 --> 00:08:55.510 inside the network and the cuts have
NOTE Confidence: 0.983169928571429

00:08:55.615 --> 00:08:59.595 to isolate the parts of the network so
NOTE Confidence: 0.983169928571429

00:08:59.595 --> 00:09:02.866 that you don't remove a lot of links
NOTE Confidence: 0.983169928571429

00:09:02.866 --> 00:09:06.097 and you increase the density of the links.
NOTE Confidence: 0.983169928571429

00:09:06.100 --> 00:09:08.800 Offer what remains so there are
NOTE Confidence: 0.983169928571429

00:09:08.800 --> 00:09:11.581 many approaches to do that in
NOTE Confidence: 0.983169928571429

00:09:11.581 --> 00:09:13.293 the single cell pipelines,
NOTE Confidence: 0.983169928571429

00:09:13.293 --> 00:09:15.648 especially in the most popular
NOTE Confidence: 0.983169928571429

00:09:15.648 --> 00:09:18.876 methods you will see always the logon
NOTE Confidence: 0.983169928571429

00:09:18.876 --> 00:09:21.086 method for the community detection.
NOTE Confidence: 0.983395506875

00:09:24.120 --> 00:09:26.478 The advantage of this is that

NOTE Confidence: 0.983395506875
00:09:26.478 --> 00:09:28.993 many other methods for single cell
NOTE Confidence: 0.983395506875
00:09:28.993 --> 00:09:31.197 analysis uses the same approach,
NOTE Confidence: 0.983395506875
00:09:31.197 --> 00:09:34.119 so they first build that these
NOTE Confidence: 0.983395506875
00:09:34.119 --> 00:09:36.731 graph and then they try to
NOTE Confidence: 0.983395506875
00:09:36.731 --> 00:09:39.960 either a identify the clusters as
NOTE Confidence: 0.983395506875
00:09:39.960 --> 00:09:42.480 communities of the interconnections.
NOTE Confidence: 0.983395506875
00:09:42.480 --> 00:09:43.999 And then we will see that also.
NOTE Confidence: 0.983395506875
00:09:44.000 --> 00:09:46.985 For example, many trajectory tools
NOTE Confidence: 0.983395506875
00:09:46.985 --> 00:09:50.135 will use this kind of graph in
NOTE Confidence: 0.983395506875
00:09:50.135 --> 00:09:52.090 order to build the trajectory.
NOTE Confidence: 0.987324051428571
00:09:55.240 --> 00:09:57.949 So this is for the class ring.
NOTE Confidence: 0.987324051428571
00:09:57.950 --> 00:10:00.278 Uhm? Another important step,
NOTE Confidence: 0.987324051428571
00:10:00.278 --> 00:10:04.060 once you identify the clusters is to
NOTE Confidence: 0.987324051428571
00:10:04.060 --> 00:10:08.154 perform A to identify the genes that
NOTE Confidence: 0.987324051428571
00:10:08.154 --> 00:10:12.364 are characterizing these clusters and.
NOTE Confidence: 0.987324051428571

00:10:12.364 --> 00:10:15.836 That means that you want to identify the
NOTE Confidence: 0.987324051428571

00:10:15.836 --> 00:10:18.716 so-called marker genes for each cluster.
NOTE Confidence: 0.987324051428571

00:10:18.720 --> 00:10:20.880 So these are ideally jeans that
NOTE Confidence: 0.987324051428571

00:10:20.880 --> 00:10:23.004 are expressed only in the cluster
NOTE Confidence: 0.987324051428571

00:10:23.004 --> 00:10:25.146 in in a single cluster and not
NOTE Confidence: 0.987324051428571

00:10:25.146 --> 00:10:27.220 in the in the other cluster.
NOTE Confidence: 0.987324051428571

00:10:27.220 --> 00:10:29.320 And here you see an example.
NOTE Confidence: 0.987324051428571

00:10:29.320 --> 00:10:31.144 So this is a.
NOTE Confidence: 0.987324051428571

00:10:31.144 --> 00:10:33.424 Representation are you map representation
NOTE Confidence: 0.987324051428571

00:10:33.424 --> 00:10:37.187 of the first two dimension of a single cell,
NOTE Confidence: 0.987324051428571

00:10:37.190 --> 00:10:40.880 analysis of peripheral blood cells.
NOTE Confidence: 0.987324051428571

00:10:40.880 --> 00:10:43.088 So you see there are clustering.
NOTE Confidence: 0.987324051428571

00:10:43.090 --> 00:10:45.146 Louvain clustering has been
NOTE Confidence: 0.987324051428571

00:10:45.146 --> 00:10:46.688 applied and identified,
NOTE Confidence: 0.987324051428571

00:10:46.690 --> 00:10:50.092 eight cluster a cluster of cells that
NOTE Confidence: 0.987324051428571

00:10:50.092 --> 00:10:55.119 you see here from zero to 7 and the

NOTE Confidence: 0.987324051428571
00:10:55.119 --> 00:10:58.402 identification of the marker Gina is a
NOTE Confidence: 0.987324051428571
00:10:58.402 --> 00:11:01.629 identification of genes that will be useful.
NOTE Confidence: 0.987324051428571
00:11:01.629 --> 00:11:04.467 In order to annotate the cluster.
NOTE Confidence: 0.987324051428571
00:11:04.470 --> 00:11:05.758 Uh, because for example,
NOTE Confidence: 0.987324051428571
00:11:05.758 --> 00:11:07.046 as you see here,
NOTE Confidence: 0.987324051428571
00:11:07.050 --> 00:11:08.835 if you look at the at the
NOTE Confidence: 0.987324051428571
00:11:08.835 --> 00:11:10.230 expression level of this gene,
NOTE Confidence: 0.987324051428571
00:11:10.230 --> 00:11:13.710 you you only see this kind of representation.
NOTE Confidence: 0.987324051428571
00:11:13.710 --> 00:11:16.342 Also in like paper or analysis of single
NOTE Confidence: 0.987324051428571
00:11:16.342 --> 00:11:18.877 cell you can represent the same map,
NOTE Confidence: 0.987324051428571
00:11:18.880 --> 00:11:21.484 but instead of coloring the cells according
NOTE Confidence: 0.987324051428571
00:11:21.484 --> 00:11:24.300 to the cluster you call on the cell
NOTE Confidence: 0.987324051428571
00:11:24.300 --> 00:11:26.449 according to the expression of 1 gene.
NOTE Confidence: 0.987324051428571
00:11:26.450 --> 00:11:30.178 So here are cells are Gray if the
NOTE Confidence: 0.987324051428571
00:11:30.178 --> 00:11:32.500 gene MS4A1 is not expressed and
NOTE Confidence: 0.987324051428571

00:11:32.500 --> 00:11:34.870 they are violent if the genie.
NOTE Confidence: 0.987324051428571

00:11:34.870 --> 00:11:37.182 Is expressed a lot and you see that
NOTE Confidence: 0.987324051428571

00:11:37.182 --> 00:11:39.584 this is a good marker for cluster
NOTE Confidence: 0.987324051428571

00:11:39.584 --> 00:11:42.145 three because you see that the marker
NOTE Confidence: 0.987324051428571

00:11:42.145 --> 00:11:44.365 is highly expressed only in this
NOTE Confidence: 0.987324051428571

00:11:44.365 --> 00:11:46.865 cluster and most of the other cells
NOTE Confidence: 0.987324051428571

00:11:46.865 --> 00:11:49.256 do not express these genes at all.
NOTE Confidence: 0.987324051428571

00:11:49.256 --> 00:11:52.140 So the task here is to is to
NOTE Confidence: 0.987324051428571

00:11:52.140 --> 00:11:54.528 identify genes that are like this,
NOTE Confidence: 0.987324051428571

00:11:54.530 --> 00:11:56.686 so genes that are highly expressed only
NOTE Confidence: 0.987324051428571

00:11:56.686 --> 00:11:58.866 in one cluster and not in the other.
NOTE Confidence: 0.987324051428571

00:11:58.870 --> 00:12:01.047 So basically this task is very similar
NOTE Confidence: 0.987324051428571

00:12:01.047 --> 00:12:03.163 to a task of differential expression
NOTE Confidence: 0.987324051428571

00:12:03.163 --> 00:12:06.599 because what you want to do is to identify.
NOTE Confidence: 0.987324051428571

00:12:06.600 --> 00:12:08.304 And for each cluster,
NOTE Confidence: 0.987324051428571

00:12:08.304 --> 00:12:10.434 genes that are differentially expressed,

NOTE Confidence: 0.987324051428571
00:12:10.440 --> 00:12:11.194 in particular,
NOTE Confidence: 0.987324051428571
00:12:11.194 --> 00:12:13.079 more expressed in the cluster.
NOTE Confidence: 0.987324051428571
00:12:13.080 --> 00:12:14.940 Then in all the other cells.
NOTE Confidence: 0.987324051428571
00:12:14.940 --> 00:12:19.600 So you divide your cell in two sets.
NOTE Confidence: 0.987324051428571
00:12:19.600 --> 00:12:21.130 It's belonging to the cluster
NOTE Confidence: 0.987324051428571
00:12:21.130 --> 00:12:22.660 and all the other cells,
NOTE Confidence: 0.987324051428571
00:12:22.660 --> 00:12:24.690 and then you try to identify genes
NOTE Confidence: 0.987324051428571
00:12:24.690 --> 00:12:26.120 that are differentially expressed,
NOTE Confidence: 0.987324051428571
00:12:26.120 --> 00:12:28.630 differentially expressed.
NOTE Confidence: 0.987324051428571
00:12:28.630 --> 00:12:31.051 So the aim of this task is to identify
NOTE Confidence: 0.987324051428571
00:12:31.051 --> 00:12:32.980 genes with different expression,
NOTE Confidence: 0.987324051428571
00:12:32.980 --> 00:12:35.360 usually among clusters of cells,
NOTE Confidence: 0.987324051428571
00:12:35.360 --> 00:12:37.270 and because these are important
NOTE Confidence: 0.987324051428571
00:12:37.270 --> 00:12:39.180 because they are cluster markers.
NOTE Confidence: 0.987324051428571
00:12:39.180 --> 00:12:41.484 Now this is something that also
NOTE Confidence: 0.987324051428571

00:12:41.484 --> 00:12:44.072 we covered that in the second
NOTE Confidence: 0.987324051428571

00:12:44.072 --> 00:12:46.016 lesson Everett spoke about.
NOTE Confidence: 0.987324051428571

00:12:46.020 --> 00:12:49.140 These are a task differential expression
NOTE Confidence: 0.987324051428571

00:12:49.140 --> 00:12:52.459 analysis will buy with bulk RNA seek.
NOTE Confidence: 0.987324051428571

00:12:52.460 --> 00:12:54.112 So the approaches them.
NOTE Confidence: 0.987324051428571

00:12:54.112 --> 00:12:57.570 That are used in single cell are different
NOTE Confidence: 0.987324051428571

00:12:57.570 --> 00:13:01.746 and this is a an area where there is no.
NOTE Confidence: 0.987324051428571

00:13:01.750 --> 00:13:03.520 Everything makes my favorite methods,
NOTE Confidence: 0.987324051428571

00:13:03.520 --> 00:13:05.536 so there are two papers that try to
NOTE Confidence: 0.987324051428571

00:13:05.536 --> 00:13:08.255 do a benchmark and compare different
NOTE Confidence: 0.987324051428571

00:13:08.255 --> 00:13:10.880 methodologies to do differential expression,
NOTE Confidence: 0.987324051428571

00:13:10.880 --> 00:13:13.372 but the result is that no method
NOTE Confidence: 0.987324051428571

00:13:13.372 --> 00:13:15.050 is better than others.
NOTE Confidence: 0.987324051428571

00:13:15.050 --> 00:13:19.050 And the problem is that when you consider
NOTE Confidence: 0.982103065294118

00:13:19.050 --> 00:13:20.658 the expression, the distribution
NOTE Confidence: 0.982103065294118

00:13:20.658 --> 00:13:23.070 of expression level of genes of

NOTE Confidence: 0.982103065294118
00:13:23.139 --> 00:13:25.337 single genes in a single cell data,
NOTE Confidence: 0.982103065294118
00:13:25.340 --> 00:13:27.280 they're very thorough genius.
NOTE Confidence: 0.982103065294118
00:13:27.280 --> 00:13:30.190 So here you see three examples
NOTE Confidence: 0.982103065294118
00:13:30.272 --> 00:13:32.284 of the expression densities
NOTE Confidence: 0.982103065294118
00:13:32.284 --> 00:13:34.799 across cells of three genes.
NOTE Confidence: 0.982103065294118
00:13:34.800 --> 00:13:37.930 You see that these ambience
NOTE Confidence: 0.982103065294118
00:13:37.930 --> 00:13:40.434 17 has these density,
NOTE Confidence: 0.982103065294118
00:13:40.440 --> 00:13:42.834 so that could be like a approximated
NOTE Confidence: 0.982103065294118
00:13:42.834 --> 00:13:45.498 with a like a normal distribution.
NOTE Confidence: 0.982103065294118
00:13:45.500 --> 00:13:50.930 Or less these A is a has a 2 pics
NOTE Confidence: 0.982103065294118
00:13:51.095 --> 00:13:54.872 one pick a one pica of cells
NOTE Confidence: 0.982103065294118
00:13:54.872 --> 00:13:58.408 where the gene is not expressed.
NOTE Confidence: 0.982103065294118
00:13:58.410 --> 00:14:00.769 This is quite common in single cell
NOTE Confidence: 0.982103065294118
00:14:00.769 --> 00:14:02.870 because of the dropout events.
NOTE Confidence: 0.982103065294118
00:14:02.870 --> 00:14:04.640 Sorry if you remember we.
NOTE Confidence: 0.982103065294118

00:14:04.640 --> 00:14:06.446 That's one of the main problem of
NOTE Confidence: 0.982103065294118

00:14:06.446 --> 00:14:08.566 single cell in a lot in a lot of cells.
NOTE Confidence: 0.982103065294118

00:14:08.570 --> 00:14:10.730 At the absence of a gene is not
NOTE Confidence: 0.982103065294118

00:14:10.730 --> 00:14:12.209 biological but it's technical,
NOTE Confidence: 0.982103065294118

00:14:12.210 --> 00:14:13.490 so the genie is simplest,
NOTE Confidence: 0.982103065294118

00:14:13.490 --> 00:14:16.024 was not captured in a library preparation.
NOTE Confidence: 0.982103065294118

00:14:16.030 --> 00:14:18.412 So that's why for many genes
NOTE Confidence: 0.982103065294118

00:14:18.412 --> 00:14:20.000 you have this situation.
NOTE Confidence: 0.982103065294118

00:14:20.000 --> 00:14:22.520 You in the same cluster of cells you
NOTE Confidence: 0.982103065294118

00:14:22.520 --> 00:14:24.933 have some cells that express the gene
NOTE Confidence: 0.982103065294118

00:14:24.933 --> 00:14:28.160 and stem cells that do not express the gene.
NOTE Confidence: 0.982103065294118

00:14:28.160 --> 00:14:30.407 And then in other in other cases,
NOTE Confidence: 0.982103065294118

00:14:30.410 --> 00:14:32.342 for example, you have most of the
NOTE Confidence: 0.982103065294118

00:14:32.342 --> 00:14:34.149 cells with the zero expression.
NOTE Confidence: 0.982103065294118

00:14:34.150 --> 00:14:36.796 So this means that there is no.
NOTE Confidence: 0.982103065294118

00:14:36.800 --> 00:14:38.684 Unique distribution that allow

NOTE Confidence: 0.982103065294118
00:14:38.684 --> 00:14:40.568 you to a model.
NOTE Confidence: 0.982103065294118
00:14:40.570 --> 00:14:43.098 The expression of genes.
NOTE Confidence: 0.982103065294118
00:14:43.100 --> 00:14:46.220 So that's why.
NOTE Confidence: 0.982103065294118
00:14:46.220 --> 00:14:48.215 Popular way of approaches for
NOTE Confidence: 0.982103065294118
00:14:48.215 --> 00:14:50.210 a do for doing differential
NOTE Confidence: 0.982103065294118
00:14:50.280 --> 00:14:52.450 expression with single cell data
NOTE Confidence: 0.982103065294118
00:14:52.450 --> 00:14:54.620 is to use nonparametric test.
NOTE Confidence: 0.982103065294118
00:14:54.620 --> 00:14:57.062 So nonparametric tests do not make
NOTE Confidence: 0.982103065294118
00:14:57.062 --> 00:14:59.382 any assumption on the underlying
NOTE Confidence: 0.982103065294118
00:14:59.382 --> 00:15:01.826 distribution of the expression.
NOTE Confidence: 0.982103065294118
00:15:01.830 --> 00:15:02.574 For example,
NOTE Confidence: 0.982103065294118
00:15:02.574 --> 00:15:05.550 probably the most used is the Wilcoxon rank.
NOTE Confidence: 0.982103065294118
00:15:05.550 --> 00:15:08.777 Sum tests. A lot of these tests.
NOTE Confidence: 0.982103065294118
00:15:08.780 --> 00:15:10.084 Do not use that.
NOTE Confidence: 0.982103065294118
00:15:10.084 --> 00:15:12.040 Do not compare the real values,
NOTE Confidence: 0.982103065294118

00:15:12.040 --> 00:15:14.596 but they compare ranks so they
NOTE Confidence: 0.982103065294118

00:15:14.596 --> 00:15:16.300 transform numbers into ranks.
NOTE Confidence: 0.982103065294118

00:15:16.300 --> 00:15:19.300 Once you order numbers from
NOTE Confidence: 0.982103065294118

00:15:19.300 --> 00:15:22.300 the highest to the lower.
NOTE Confidence: 0.982103065294118

00:15:22.300 --> 00:15:22.587 Uh,
NOTE Confidence: 0.982103065294118

00:15:22.587 --> 00:15:25.170 so they can be used with single cell also,
NOTE Confidence: 0.982103065294118

00:15:25.170 --> 00:15:25.774 because uh,
NOTE Confidence: 0.982103065294118

00:15:25.774 --> 00:15:28.190 when you have a lot of cells you
NOTE Confidence: 0.982103065294118

00:15:28.265 --> 00:15:30.203 have a lot of measurements and
NOTE Confidence: 0.982103065294118

00:15:30.203 --> 00:15:32.590 so these kind of tests work well
NOTE Confidence: 0.982103065294118

00:15:32.590 --> 00:15:34.970 when you have a lot of replicates
NOTE Confidence: 0.982103065294118

00:15:34.970 --> 00:15:37.047 because you can consider each cell
NOTE Confidence: 0.982103065294118

00:15:37.047 --> 00:15:40.322 as a replicate when you do the UM,
NOTE Confidence: 0.982103065294118

00:15:40.322 --> 00:15:43.244 try to establish the difference,
NOTE Confidence: 0.982103065294118

00:15:43.244 --> 00:15:45.844 they are problematic since they
NOTE Confidence: 0.982103065294118

00:15:45.844 --> 00:15:46.960 work on ranks,

NOTE Confidence: 0.982103065294118
00:15:46.960 --> 00:15:48.568 their problematic when you have a
NOTE Confidence: 0.982103065294118
00:15:48.568 --> 00:15:50.370 lot of values that are the same.
NOTE Confidence: 0.982103065294118
00:15:50.370 --> 00:15:52.386 So these are tide values and
NOTE Confidence: 0.982103065294118
00:15:52.386 --> 00:15:53.730 that's exactly what happens.
NOTE Confidence: 0.982103065294118
00:15:53.730 --> 00:15:54.615 With the zeros,
NOTE Confidence: 0.982103065294118
00:15:54.615 --> 00:15:56.680 so this could be a problem of
NOTE Confidence: 0.982103065294118
00:15:56.749 --> 00:15:58.217 applying nonparametric test when
NOTE Confidence: 0.982103065294118
00:15:58.217 --> 00:16:00.810 you have a lot of cells where
NOTE Confidence: 0.982103065294118
00:16:00.810 --> 00:16:03.050 the gene is not expressed at all,
NOTE Confidence: 0.982103065294118
00:16:03.050 --> 00:16:05.892 then the other methods are the same
NOTE Confidence: 0.982103065294118
00:16:05.892 --> 00:16:09.138 as the one used in bulk RNA seek,
NOTE Confidence: 0.982103065294118
00:16:09.140 --> 00:16:14.242 so these were the one covered by Everett H.
NOTE Confidence: 0.982103065294118
00:16:14.242 --> 00:16:17.737 R&D C2 and they are based on
NOTE Confidence: 0.982103065294118
00:16:17.737 --> 00:16:19.882 modeling the gene expression with
NOTE Confidence: 0.982103065294118
00:16:19.882 --> 00:16:22.480 a negative binomial distribution.
NOTE Confidence: 0.982103065294118

00:16:22.480 --> 00:16:24.397 And then you have a lot of methods that
NOTE Confidence: 0.982103065294118

00:16:24.397 --> 00:16:26.119 were developed for the single cell,
NOTE Confidence: 0.982103065294118

00:16:26.120 --> 00:16:28.448 and so instead of the negative
NOTE Confidence: 0.982103065294118

00:16:28.448 --> 00:16:30.501 binomial they used other distributions
NOTE Confidence: 0.982103065294118

00:16:30.501 --> 00:16:32.651 that dealer with the accessor
NOTE Confidence: 0.982103065294118

00:16:32.651 --> 00:16:35.432 of zeros that you have in single
NOTE Confidence: 0.982103065294118

00:16:35.432 --> 00:16:36.692 cell data set again.
NOTE Confidence: 0.982103065294118

00:16:36.692 --> 00:16:38.052 So these are the three
NOTE Confidence: 0.982103065294118

00:16:38.052 --> 00:16:39.140 main families that you
NOTE Confidence: 0.9857805625

00:16:39.196 --> 00:16:42.212 will find. There is no clear winner or an
NOTE Confidence: 0.9857805625

00:16:42.212 --> 00:16:44.844 approach that is more used than others.
NOTE Confidence: 0.945627303333333

00:16:47.680 --> 00:16:50.518 Uhm, so finding the marker genes,
NOTE Confidence: 0.945627303333333

00:16:50.520 --> 00:16:51.528 we said it was.
NOTE Confidence: 0.945627303333333

00:16:51.528 --> 00:16:52.788 It's very important because uh,
NOTE Confidence: 0.945627303333333

00:16:52.790 --> 00:16:56.286 it's necessary to understand,
NOTE Confidence: 0.945627303333333

00:16:56.286 --> 00:16:58.666 uh, understand that, uh,

NOTE Confidence: 0.94562730333333
00:16:58.666 --> 00:17:02.066 the identity of each uh, set cluster.
NOTE Confidence: 0.94562730333333
00:17:02.066 --> 00:17:04.724 And it's a important to label
NOTE Confidence: 0.94562730333333
00:17:04.724 --> 00:17:07.280 cell clusters with the cell types.
NOTE Confidence: 0.94562730333333
00:17:07.280 --> 00:17:10.409 So this is a problem that is
NOTE Confidence: 0.94562730333333
00:17:10.410 --> 00:17:12.510 called the cell type annotation.
NOTE Confidence: 0.94562730333333
00:17:12.510 --> 00:17:15.147 So the aim is that you want to annotate
NOTE Confidence: 0.94562730333333
00:17:15.147 --> 00:17:16.998 cluster with a known cell types.
NOTE Confidence: 0.94562730333333
00:17:17.000 --> 00:17:18.824 Depending on the system.
NOTE Confidence: 0.94562730333333
00:17:18.824 --> 00:17:20.648 That you are studying.
NOTE Confidence: 0.94562730333333
00:17:20.650 --> 00:17:21.578 So you want to.
NOTE Confidence: 0.94562730333333
00:17:21.578 --> 00:17:22.738 If you're speaking about that,
NOTE Confidence: 0.94562730333333
00:17:22.740 --> 00:17:23.654 for example,
NOTE Confidence: 0.94562730333333
00:17:23.654 --> 00:17:25.939 peripheral blood you want to
NOTE Confidence: 0.94562730333333
00:17:25.939 --> 00:17:28.687 associate a these clusters with a
NOTE Confidence: 0.94562730333333
00:17:28.687 --> 00:17:31.285 known population of blood cells that
NOTE Confidence: 0.94562730333333

00:17:31.285 --> 00:17:34.047 you find that you expect to find.

NOTE Confidence: 0.945627303333333

00:17:34.050 --> 00:17:36.114 So T cell B cell and so on.

NOTE Confidence: 0.938228128833333

00:17:38.480 --> 00:17:41.174 There are the main approaches are

NOTE Confidence: 0.938228128833333

00:17:41.174 --> 00:17:43.702 obviously the manual approach you look

NOTE Confidence: 0.938228128833333

00:17:43.702 --> 00:17:46.326 at the marker gene and you know which

NOTE Confidence: 0.938228128833333

00:17:46.399 --> 00:17:49.045 are the genes that should be highly

NOTE Confidence: 0.938228128833333

00:17:49.045 --> 00:17:50.666 expressed in each cell population,

NOTE Confidence: 0.938228128833333

00:17:50.666 --> 00:17:53.060 so you know which are the B cell markers,

NOTE Confidence: 0.938228128833333

00:17:53.060 --> 00:17:56.252 the T cell markers and you use your

NOTE Confidence: 0.938228128833333

00:17:56.252 --> 00:17:58.909 personal knowledge to annotate the cluster.

NOTE Confidence: 0.938228128833333

00:17:58.910 --> 00:18:01.862 This is probably has this has been in

NOTE Confidence: 0.938228128833333

00:18:01.862 --> 00:18:04.698 the past analysis of single cells.

NOTE Confidence: 0.938228128833333

00:18:04.700 --> 00:18:06.228 The most used method.

NOTE Confidence: 0.938228128833333

00:18:06.228 --> 00:18:08.983 And that's why since it's manual is

NOTE Confidence: 0.938228128833333

00:18:08.983 --> 00:18:11.342 based on personal knowledge is also

NOTE Confidence: 0.938228128833333

00:18:11.342 --> 00:18:13.814 very time consuming because you need

NOTE Confidence: 0.93822812883333
00:18:13.814 --> 00:18:16.875 to review all the clusters and to
NOTE Confidence: 0.93822812883333
00:18:16.875 --> 00:18:19.742 assign annotate each cluster manually
NOTE Confidence: 0.93822812883333
00:18:19.742 --> 00:18:22.681 based on your subjective knowledge.
NOTE Confidence: 0.93822812883333
00:18:22.681 --> 00:18:25.381 There is a big development,
NOTE Confidence: 0.93822812883333
00:18:25.381 --> 00:18:28.136 a huge development of automatic
NOTE Confidence: 0.93822812883333
00:18:28.136 --> 00:18:30.630 tools to perform this step.
NOTE Confidence: 0.93822812883333
00:18:30.630 --> 00:18:33.540 So to perform cell type annotation.
NOTE Confidence: 0.93822812883333
00:18:33.540 --> 00:18:36.762 And these automatic,
NOTE Confidence: 0.93822812883333
00:18:36.762 --> 00:18:42.060 uh procedures AR of can be divided into two.
NOTE Confidence: 0.93822812883333
00:18:42.060 --> 00:18:44.811 So there are some procedures that are
NOTE Confidence: 0.93822812883333
00:18:44.811 --> 00:18:47.188 based on databases of marker genes.
NOTE Confidence: 0.93822812883333
00:18:47.190 --> 00:18:50.134 So what they do is that what this
NOTE Confidence: 0.93822812883333
00:18:50.134 --> 00:18:52.384 procedure do is that they they
NOTE Confidence: 0.93822812883333
00:18:52.384 --> 00:18:55.228 compare the list of marker genes
NOTE Confidence: 0.93822812883333
00:18:55.228 --> 00:18:58.570 of each cluster with a database
NOTE Confidence: 0.93822812883333

00:18:58.570 --> 00:19:03.238 of marker genes that were found.

NOTE Confidence: 0.938228128833333

00:19:03.240 --> 00:19:03.731 Experimentally,

NOTE Confidence: 0.938228128833333

00:19:03.731 --> 00:19:06.677 in a known population of sensor,

NOTE Confidence: 0.938228128833333

00:19:06.680 --> 00:19:08.640 so the comparison is between

NOTE Confidence: 0.938228128833333

00:19:08.640 --> 00:19:10.600 different lists of marker gene.

NOTE Confidence: 0.938228128833333

00:19:10.600 --> 00:19:13.090 The advantages that you don't know

NOTE Confidence: 0.938228128833333

00:19:13.090 --> 00:19:15.197 you don't necessarily need another

NOTE Confidence: 0.938228128833333

00:19:15.197 --> 00:19:17.815 single cell data set as a reference.

NOTE Confidence: 0.938228128833333

00:19:17.820 --> 00:19:20.662 You just need a list of genes

NOTE Confidence: 0.938228128833333

00:19:20.662 --> 00:19:24.006 and we will see there are there

NOTE Confidence: 0.938228128833333

00:19:24.006 --> 00:19:26.536 are databases that try to.

NOTE Confidence: 0.938228128833333

00:19:26.540 --> 00:19:28.640 Cover all the marker genes

NOTE Confidence: 0.938228128833333

00:19:28.640 --> 00:19:29.900 for cell populations,

NOTE Confidence: 0.938228128833333

00:19:29.900 --> 00:19:32.768 at least in human and mouse.

NOTE Confidence: 0.938228128833333

00:19:32.770 --> 00:19:35.644 Another family of approaches require not

NOTE Confidence: 0.938228128833333

00:19:35.644 --> 00:19:38.798 only the unknown list of marker gene,

NOTE Confidence: 0.93822812883333
00:19:38.800 --> 00:19:44.708 but require a unannotated.

NOTE Confidence: 0.93822812883333
00:19:44.710 --> 00:19:47.608 Did expect single cell RNA seek experiments,

NOTE Confidence: 0.93822812883333
00:19:47.610 --> 00:19:48.930 so they they.

NOTE Confidence: 0.93822812883333
00:19:48.930 --> 00:19:52.010 Their strategy is what is represented here.

NOTE Confidence: 0.93822812883333
00:19:52.010 --> 00:19:54.258 You have a query data set that is

NOTE Confidence: 0.93822812883333
00:19:54.258 --> 00:19:56.525 your it's your data set where you

NOTE Confidence: 0.93822812883333
00:19:56.525 --> 00:19:58.555 have classes but you don't have

NOTE Confidence: 0.93822812883333
00:19:58.555 --> 00:20:00.844 labels and then you have a reference

NOTE Confidence: 0.93822812883333
00:20:00.844 --> 00:20:03.344 data set so someone else already

NOTE Confidence: 0.93822812883333
00:20:03.344 --> 00:20:06.110 did perform their analysis of single

NOTE Confidence: 0.93822812883333
00:20:06.193 --> 00:20:09.035 cell and label the clusters of cells.

NOTE Confidence: 0.93822812883333
00:20:09.040 --> 00:20:12.337 So the strategy is to try to

NOTE Confidence: 0.93822812883333
00:20:12.337 --> 00:20:15.189 identify which of the clusters.

NOTE Confidence: 0.93822812883333
00:20:15.190 --> 00:20:17.782 Of the query data set are more similar

NOTE Confidence: 0.93822812883333
00:20:17.782 --> 00:20:20.875 to the reference and and this is

NOTE Confidence: 0.93822812883333

00:20:20.875 --> 00:20:22.600 a problem of classification basically,
NOTE Confidence: 0.938228128833333

00:20:22.600 --> 00:20:25.343 so they try to classify a data set
NOTE Confidence: 0.938228128833333

00:20:25.343 --> 00:20:27.520 with unknown labels using a data set
NOTE Confidence: 0.938228128833333

00:20:27.591 --> 00:20:30.048 set of single cell with known labels
NOTE Confidence: 0.938228128833333

00:20:30.048 --> 00:20:32.427 and those Indies are in this family.
NOTE Confidence: 0.938228128833333

00:20:32.430 --> 00:20:35.750 Obviously you have many possible
NOTE Confidence: 0.938228128833333

00:20:35.750 --> 00:20:37.950 math methods to do this,
NOTE Confidence: 0.938228128833333

00:20:37.950 --> 00:20:41.212 some some of the methods are based
NOTE Confidence: 0.938228128833333

00:20:41.212 --> 00:20:45.080 on correlation. Try to calculate.
NOTE Confidence: 0.938228128833333

00:20:45.080 --> 00:20:47.430 The similarity through correlation measures.
NOTE Confidence: 0.938228128833333

00:20:47.430 --> 00:20:50.552 Other approaches try to use a supervised
NOTE Confidence: 0.938228128833333

00:20:50.552 --> 00:20:52.729 classification and methods that are
NOTE Confidence: 0.938228128833333

00:20:52.729 --> 00:20:55.027 commonly used in in machine learning.
NOTE Confidence: 0.938228128833333

00:20:55.030 --> 00:20:57.950 So this is one of the field where,
NOTE Confidence: 0.938228128833333

00:20:57.950 --> 00:21:02.774 like speaking about 2021 there are there are.
NOTE Confidence: 0.938228128833333

00:21:02.780 --> 00:21:05.482 Very huge developments and a lot of

NOTE Confidence: 0.93822812883333
00:21:05.482 --> 00:21:08.571 tools that are published or in either
NOTE Confidence: 0.93822812883333
00:21:08.571 --> 00:21:12.808 Inbox Ivorian on journals right now.
NOTE Confidence: 0.93822812883333
00:21:12.810 --> 00:21:15.242 And there is this metaphor that once we
NOTE Confidence: 0.93822812883333
00:21:15.242 --> 00:21:19.530 have a lot of once, we have a lot of.
NOTE Confidence: 0.93822812883333
00:21:19.530 --> 00:21:23.430 Uh, datasets that are annotated uh.
NOTE Confidence: 0.97908853454545
00:21:23.430 --> 00:21:25.265 These Tasker will become a
NOTE Confidence: 0.97908853454545
00:21:25.265 --> 00:21:27.470 such as that will become like
NOTE Confidence: 0.97908853454545
00:21:27.470 --> 00:21:31.450 mapping reads to unknown genome.
NOTE Confidence: 0.97908853454545
00:21:31.450 --> 00:21:33.598 So performing a single cell analysis
NOTE Confidence: 0.97908853454545
00:21:33.598 --> 00:21:36.089 within a new data set will become
NOTE Confidence: 0.97908853454545
00:21:36.089 --> 00:21:38.406 as simple as that because you have
NOTE Confidence: 0.97908853454545
00:21:38.478 --> 00:21:40.386 a lot of reference populations and
NOTE Confidence: 0.97908853454545
00:21:40.386 --> 00:21:42.839 so it will be easier to annotate
NOTE Confidence: 0.97908853454545
00:21:42.839 --> 00:21:45.590 your cell once you have a collection
NOTE Confidence: 0.97908853454545
00:21:45.670 --> 00:21:47.890 of references that is reliable.
NOTE Confidence: 0.97991250333333

00:21:57.530 --> 00:22:00.659 These are two resources to databases that
NOTE Confidence: 0.979912503333333

00:22:00.659 --> 00:22:03.104 collect cell type annotation markers
NOTE Confidence: 0.979912503333333

00:22:03.104 --> 00:22:07.724 and so this can be used to compare the
NOTE Confidence: 0.979912503333333

00:22:07.724 --> 00:22:10.299 markers identifying your cluster with
NOTE Confidence: 0.979912503333333

00:22:10.385 --> 00:22:13.927 a known collection of markers that were
NOTE Confidence: 0.979912503333333

00:22:13.927 --> 00:22:16.842 identified based on a single cell data.
NOTE Confidence: 0.979912503333333

00:22:16.842 --> 00:22:19.806 There is this sort of so if you
NOTE Confidence: 0.979912503333333

00:22:19.806 --> 00:22:22.208 look at flow cytometry, the. The.
NOTE Confidence: 0.979912503333333

00:22:22.208 --> 00:22:25.480 The best markers are considered to be the
NOTE Confidence: 0.979912503333333

00:22:25.561 --> 00:22:29.005 proteins that are expressed on the surface.
NOTE Confidence: 0.979912503333333

00:22:29.010 --> 00:22:32.160 The problem is that the transcripts of
NOTE Confidence: 0.979912503333333

00:22:32.160 --> 00:22:35.064 these proteins of surface markers are
NOTE Confidence: 0.979912503333333

00:22:35.064 --> 00:22:38.494 not always among the top expressed genes,
NOTE Confidence: 0.979912503333333

00:22:38.500 --> 00:22:42.054 and so they may be subjected to dropout
NOTE Confidence: 0.979912503333333

00:22:42.054 --> 00:22:44.980 events and so the best collection of
NOTE Confidence: 0.979912503333333

00:22:45.055 --> 00:22:47.989 cell markers for a transcriptomic study.

NOTE Confidence: 0.979912503333333
00:22:47.990 --> 00:22:50.461 So based on gene expression is different
NOTE Confidence: 0.979912503333333
00:22:50.461 --> 00:22:54.218 from the best collection of markers based on.
NOTE Confidence: 0.979912503333333
00:22:54.218 --> 00:22:56.090 Some surface proteins?
NOTE Confidence: 0.931627771111111
00:22:58.640 --> 00:23:01.045 And, uh, these two databases
NOTE Confidence: 0.931627771111111
00:23:01.045 --> 00:23:02.969 collected single cell signatures,
NOTE Confidence: 0.931627771111111
00:23:02.970 --> 00:23:05.650 single cell markers in different
NOTE Confidence: 0.931627771111111
00:23:05.650 --> 00:23:07.794 samples and tissues mainly,
NOTE Confidence: 0.931627771111111
00:23:07.800 --> 00:23:11.340 or human and mouse.
NOTE Confidence: 0.931627771111111
00:23:11.340 --> 00:23:13.937 So also looking at at comparing the
NOTE Confidence: 0.931627771111111
00:23:13.937 --> 00:23:17.025 two species is important to know which
NOTE Confidence: 0.931627771111111
00:23:17.025 --> 00:23:19.400 markers are conserved across species
NOTE Confidence: 0.931627771111111
00:23:19.400 --> 00:23:22.249 and which one are species specific.
NOTE Confidence: 0.978613616666667
00:23:25.920 --> 00:23:31.112 Then the last step for two days
NOTE Confidence: 0.978613616666667
00:23:31.112 --> 00:23:33.808 is the trajectory analysis.
NOTE Confidence: 0.978613616666667
00:23:33.808 --> 00:23:37.931 So why is clustering tries to divide
NOTE Confidence: 0.978613616666667

00:23:37.931 --> 00:23:41.430 the cells into discrete clusters?
NOTE Confidence: 0.978613616666667

00:23:41.430 --> 00:23:43.495 The idea of trajectory analysis
NOTE Confidence: 0.978613616666667

00:23:43.495 --> 00:23:46.540 is that you are not monitoring.
NOTE Confidence: 0.978613616666667

00:23:46.540 --> 00:23:48.884 You are not capturing
NOTE Confidence: 0.978613616666667

00:23:48.884 --> 00:23:51.228 sales in discrete states,
NOTE Confidence: 0.978613616666667

00:23:51.230 --> 00:23:54.107 but you're capturing a sort of continuous.
NOTE Confidence: 0.978613616666667

00:23:54.110 --> 00:23:57.146 Processor for example,
NOTE Confidence: 0.978613616666667

00:23:57.146 --> 00:24:00.294 differentiation, uh for example,
NOTE Confidence: 0.978613616666667

00:24:00.294 --> 00:24:01.418 yes differentiation.
NOTE Confidence: 0.978613616666667

00:24:01.418 --> 00:24:05.141 So these kind of methods try to place
NOTE Confidence: 0.978613616666667

00:24:05.141 --> 00:24:08.508 sells a longer a continuous path that
NOTE Confidence: 0.978613616666667

00:24:08.508 --> 00:24:11.457 represents the evolution of a process.
NOTE Confidence: 0.978613616666667

00:24:11.460 --> 00:24:12.540 This could be differentiation,
NOTE Confidence: 0.978613616666667

00:24:12.540 --> 00:24:15.830 but for example, if you imagine a another
NOTE Confidence: 0.978613616666667

00:24:15.830 --> 00:24:19.000 simple example is the cell cycle.
NOTE Confidence: 0.978613616666667

00:24:19.000 --> 00:24:20.935 And so instead of dividing

NOTE Confidence: 0.978613616666667
00:24:20.935 --> 00:24:22.483 cells into separate cluster,
NOTE Confidence: 0.978613616666667
00:24:22.490 --> 00:24:26.170 you try to construct a sort of trajectory
NOTE Confidence: 0.978613616666667
00:24:26.170 --> 00:24:29.527 that models this progression through a,
NOTE Confidence: 0.978613616666667
00:24:29.530 --> 00:24:31.780 for example, differentiation.
NOTE Confidence: 0.978613616666667
00:24:31.780 --> 00:24:35.260 And, uh, uh, this sort of, UM,
NOTE Confidence: 0.978613616666667
00:24:35.260 --> 00:24:36.784 tools, trajectory, inference,
NOTE Confidence: 0.978613616666667
00:24:36.784 --> 00:24:40.340 analysis are also named sometimes like a
NOTE Confidence: 0.978613616666667
00:24:40.412 --> 00:24:43.166 term desktop sealed the time analysis,
NOTE Confidence: 0.978613616666667
00:24:43.170 --> 00:24:46.579 because still the timer is a basically,
NOTE Confidence: 0.978613616666667
00:24:46.580 --> 00:24:49.800 a measure is an abstract measure of
NOTE Confidence: 0.978613616666667
00:24:49.800 --> 00:24:52.199 the progression through the process.
NOTE Confidence: 0.978613616666667
00:24:52.200 --> 00:24:54.769 So from when the program when the
NOTE Confidence: 0.978613616666667
00:24:54.769 --> 00:24:58.098 process starts to where it ends.
NOTE Confidence: 0.978613616666667
00:24:58.100 --> 00:25:00.115 So the important assumption is
NOTE Confidence: 0.978613616666667
00:25:00.115 --> 00:25:02.652 that in order to perform trajectory
NOTE Confidence: 0.978613616666667

00:25:02.652 --> 00:25:05.292 analysis is that we are capturing
NOTE Confidence: 0.978613616666667

00:25:05.292 --> 00:25:07.940 with our single cell experiments.
NOTE Confidence: 0.978613616666667

00:25:07.940 --> 00:25:10.478 All the snapshot of the process
NOTE Confidence: 0.978613616666667

00:25:10.478 --> 00:25:13.487 that we want to model and this
NOTE Confidence: 0.978613616666667

00:25:13.487 --> 00:25:16.121 means that we are capturing also
NOTE Confidence: 0.978613616666667

00:25:16.121 --> 00:25:19.227 the intermediates because all the.
NOTE Confidence: 0.978613616666667

00:25:19.230 --> 00:25:20.946 The analysis it spins on the
NOTE Confidence: 0.978613616666667

00:25:20.946 --> 00:25:22.809 assumption that we have a continuum,
NOTE Confidence: 0.978613616666667

00:25:22.810 --> 00:25:23.452 not discreet,
NOTE Confidence: 0.978613616666667

00:25:23.452 --> 00:25:26.020 and so we need to have some to
NOTE Confidence: 0.978613616666667

00:25:26.101 --> 00:25:28.706 capture some cells that represent
NOTE Confidence: 0.978613616666667

00:25:28.706 --> 00:25:30.269 the transition between,
NOTE Confidence: 0.978613616666667

00:25:30.270 --> 00:25:31.998 for example two differentiation.
NOTE Confidence: 0.97169673

00:25:34.510 --> 00:25:38.142 Haynes so the assumption is
NOTE Confidence: 0.97169673

00:25:38.142 --> 00:25:39.957 that we're capturing all the
NOTE Confidence: 0.97169673

00:25:39.957 --> 00:25:41.870 snapshots and we don't have holes,

NOTE Confidence: 0.97169673
00:25:41.870 --> 00:25:43.798 and so we have a lot of intermediates
NOTE Confidence: 0.97169673
00:25:43.798 --> 00:25:45.568 and the warning is that, uh,
NOTE Confidence: 0.97169673
00:25:45.568 --> 00:25:48.144 any data set, so these tools will
NOTE Confidence: 0.97169673
00:25:48.144 --> 00:25:50.565 will will capture a trajectory for
NOTE Confidence: 0.97169673
00:25:50.565 --> 00:25:53.750 each data set that you use as input.
NOTE Confidence: 0.97169673
00:25:53.750 --> 00:25:56.417 But this doesn't mean that the trajectory
NOTE Confidence: 0.97169673
00:25:56.417 --> 00:25:59.498 that you find has any biological meaning.
NOTE Confidence: 0.97169673
00:25:59.500 --> 00:26:01.880 So the common approach to do these
NOTE Confidence: 0.97169673
00:26:01.880 --> 00:26:04.538 are there are a lot of methods also
NOTE Confidence: 0.97169673
00:26:04.538 --> 00:26:08.170 for these common and simply to explain
NOTE Confidence: 0.97169673
00:26:08.170 --> 00:26:11.570 approach is are represented here so.
NOTE Confidence: 0.97169673
00:26:11.570 --> 00:26:14.666 These are is a PCA plot and here
NOTE Confidence: 0.97169673
00:26:14.666 --> 00:26:17.330 instead of only two dimensions,
NOTE Confidence: 0.97169673
00:26:17.330 --> 00:26:19.350 you see the three dimensions,
NOTE Confidence: 0.97169673
00:26:19.350 --> 00:26:21.694 so PC1PC20 PC three.
NOTE Confidence: 0.97169673

00:26:21.694 --> 00:26:24.624 Each dot is a cell.
NOTE Confidence: 0.97169673

00:26:24.630 --> 00:26:27.360 What you do is you first perform
NOTE Confidence: 0.97169673

00:26:27.360 --> 00:26:29.936 a clustering of cells with K
NOTE Confidence: 0.97169673

00:26:29.936 --> 00:26:32.226 means with graph based approach.
NOTE Confidence: 0.97169673

00:26:32.230 --> 00:26:35.263 So this depends on the on the tool and
NOTE Confidence: 0.97169673

00:26:35.263 --> 00:26:38.406 so identify these clusters of cells.
NOTE Confidence: 0.97169673

00:26:38.410 --> 00:26:41.798 Inside your population now if you remember,
NOTE Confidence: 0.97169673

00:26:41.800 --> 00:26:44.272 each cluster can be associated with
NOTE Confidence: 0.97169673

00:26:44.272 --> 00:26:46.825 a centroid where the centroid is
NOTE Confidence: 0.97169673

00:26:46.825 --> 00:26:49.309 the central point of the cluster.
NOTE Confidence: 0.97169673

00:26:49.310 --> 00:26:51.462 So it it is in the position in
NOTE Confidence: 0.97169673

00:26:51.462 --> 00:26:53.645 the mean position with respect to
NOTE Confidence: 0.97169673

00:26:53.645 --> 00:26:56.003 the elements of all the cluster.
NOTE Confidence: 0.97169673

00:26:56.010 --> 00:26:58.734 And these dots here represent the
NOTE Confidence: 0.97169673

00:26:58.734 --> 00:27:02.199 centroids of the cluster that you identified.
NOTE Confidence: 0.97169673

00:27:02.200 --> 00:27:05.300 Now what you do is you try to build a tree.

NOTE Confidence: 0.97169673
00:27:05.300 --> 00:27:07.365 That connects these centroid and
NOTE Confidence: 0.97169673
00:27:07.365 --> 00:27:09.430 those who build these three.
NOTE Confidence: 0.97169673
00:27:09.430 --> 00:27:11.130 There are many strategies.
NOTE Confidence: 0.97169673
00:27:11.130 --> 00:27:14.255 One of the most simple is to
NOTE Confidence: 0.97169673
00:27:14.255 --> 00:27:16.705 build the minimum spanning trees.
NOTE Confidence: 0.97169673
00:27:16.710 --> 00:27:19.132 So you're trying to connect all these
NOTE Confidence: 0.97169673
00:27:19.132 --> 00:27:22.356 points in a way that minimizes the length.
NOTE Confidence: 0.97169673
00:27:22.360 --> 00:27:24.736 The total length of the branch.
NOTE Confidence: 0.97169673
00:27:24.740 --> 00:27:25.366 The branches,
NOTE Confidence: 0.97169673
00:27:25.366 --> 00:27:27.870 so if you have a set of points,
NOTE Confidence: 0.97169673
00:27:27.870 --> 00:27:32.075 you can find a solution with a tree
NOTE Confidence: 0.97169673
00:27:32.075 --> 00:27:34.630 that minimizes the length of the sum
NOTE Confidence: 0.97169673
00:27:34.715 --> 00:27:37.228 of all the branches of your tree,
NOTE Confidence: 0.97169673
00:27:37.230 --> 00:27:39.670 and this is called the minimum spanning tree.
NOTE Confidence: 0.97169673
00:27:39.670 --> 00:27:42.491 So the assumption is that the minimum
NOTE Confidence: 0.97169673

00:27:42.491 --> 00:27:45.554 spanning tree is the correct tree that
NOTE Confidence: 0.97169673

00:27:45.554 --> 00:27:49.058 models the trajectory in this data,
NOTE Confidence: 0.97169673

00:27:49.060 --> 00:27:51.069 and this is not always the case,
NOTE Confidence: 0.97169673

00:27:51.070 --> 00:27:53.318 so a warning is that it's not always
NOTE Confidence: 0.97169673

00:27:53.318 --> 00:27:55.188 the minimum. Spanning tree is not.
NOTE Confidence: 0.97169673

00:27:55.188 --> 00:27:57.751 Always the best solution and, uh, uhm.
NOTE Confidence: 0.97169673

00:27:57.751 --> 00:27:59.419 Once you do this,
NOTE Confidence: 0.97169673

00:27:59.420 --> 00:28:02.759 you have your trip and some tools
NOTE Confidence: 0.97169673

00:28:02.759 --> 00:28:06.212 like try to assign a route to
NOTE Confidence: 0.97169673

00:28:06.212 --> 00:28:09.378 this tree or or they may give you
NOTE Confidence: 0.97169673

00:28:09.378 --> 00:28:11.633 also the possibility to select the
NOTE Confidence: 0.97169673

00:28:11.633 --> 00:28:14.449 root so the user can say that this
NOTE Confidence: 0.97169673

00:28:14.531 --> 00:28:16.409 is the root of the tree,
NOTE Confidence: 0.97169673

00:28:16.410 --> 00:28:18.582 because you know that these cells
NOTE Confidence: 0.97169673

00:28:18.582 --> 00:28:20.823 for example are most similar to
NOTE Confidence: 0.97169673

00:28:20.823 --> 00:28:22.983 what you expect from stem cells,

NOTE Confidence: 0.97169673
00:28:22.990 --> 00:28:25.020 and so once you uh,
NOTE Confidence: 0.97169673
00:28:25.020 --> 00:28:27.556 once you define the root of the tree,
NOTE Confidence: 0.97169673
00:28:27.560 --> 00:28:28.637 then you can.
NOTE Confidence: 0.97169673
00:28:28.637 --> 00:28:31.150 Define and smooth your tree and you
NOTE Confidence: 0.97169673
00:28:31.231 --> 00:28:33.849 can calculate the time for each cell
NOTE Confidence: 0.97169673
00:28:33.849 --> 00:28:36.884 where the time of the cell will be
NOTE Confidence: 0.97169673
00:28:36.884 --> 00:28:39.088 the distance from the position of
NOTE Confidence: 0.97169673
00:28:39.088 --> 00:28:42.032 the cell to the root of your tree
NOTE Confidence: 0.97169673
00:28:42.032 --> 00:28:44.487 following the topology of the tree.
NOTE Confidence: 0.946273925714286
00:28:49.640 --> 00:28:51.845 Is this a more or less clear?
NOTE Confidence: 0.98396079
00:28:54.190 --> 00:28:57.241 Yes to me. OK,
NOTE Confidence: 0.98396079
00:28:57.241 --> 00:29:00.427 so many methods used this approach,
NOTE Confidence: 0.98396079
00:29:00.430 --> 00:29:02.430 so first step dimensional reduction,
NOTE Confidence: 0.98396079
00:29:02.430 --> 00:29:04.955 then clustering, then the construction
NOTE Confidence: 0.98396079
00:29:04.955 --> 00:29:08.763 of a of a tree that at the beginning
NOTE Confidence: 0.98396079

00:29:08.763 --> 00:29:12.317 the tree was built on the single cell,
NOTE Confidence: 0.98396079

00:29:12.320 --> 00:29:13.868 but that's unstable and so that
NOTE Confidence: 0.98396079

00:29:13.868 --> 00:29:15.734 there was a switch from the single
NOTE Confidence: 0.98396079

00:29:15.734 --> 00:29:17.282 cell to the to the centroids.
NOTE Confidence: 0.98396079

00:29:17.290 --> 00:29:20.008 Because the center is more stable,
NOTE Confidence: 0.98396079

00:29:20.010 --> 00:29:22.014 uh, and so if you have
NOTE Confidence: 0.98396079

00:29:22.014 --> 00:29:23.350 deviations of the position,
NOTE Confidence: 0.98396079

00:29:23.350 --> 00:29:25.884 the overall tree will remain the same,
NOTE Confidence: 0.98396079

00:29:25.890 --> 00:29:26.954 so it's more that.
NOTE Confidence: 0.98396079

00:29:26.954 --> 00:29:28.738 That's why it's more stable and
NOTE Confidence: 0.98396079

00:29:28.738 --> 00:29:31.874 then depending on the choice of the
NOTE Confidence: 0.98396079

00:29:31.874 --> 00:29:33.180 dimensionality reduction approach
NOTE Confidence: 0.98396079

00:29:33.180 --> 00:29:35.852 and on the tree that you build it,
NOTE Confidence: 0.98396079

00:29:35.860 --> 00:29:38.310 you have different outcomes and
NOTE Confidence: 0.98396079

00:29:38.310 --> 00:29:40.721 different trajectories and this is
NOTE Confidence: 0.98396079

00:29:40.721 --> 00:29:44.266 from a paper from a paper that

NOTE Confidence: 0.98396079
00:29:44.266 --> 00:29:47.640 two years ago was trying to compare a
NOTE Confidence: 0.956023311666667
00:29:49.750 --> 00:29:53.380 12345678 different methods to perform basa.
NOTE Confidence: 0.956023311666667
00:29:53.380 --> 00:29:57.730 So dimensional reduction and also.
NOTE Confidence: 0.956023311666667
00:29:57.730 --> 00:30:00.654 Trajectory inference so you see, for example.
NOTE Confidence: 0.956023311666667
00:30:00.654 --> 00:30:04.100 So each row here is a different data set,
NOTE Confidence: 0.956023311666667
00:30:04.100 --> 00:30:07.068 so you have five data set from.
NOTE Confidence: 0.956023311666667
00:30:07.070 --> 00:30:08.690 The simplest are because
NOTE Confidence: 0.956023311666667
00:30:08.690 --> 00:30:10.882 it has the smallest size,
NOTE Confidence: 0.956023311666667
00:30:10.882 --> 00:30:14.284 so this is the number of cells.
NOTE Confidence: 0.956023311666667
00:30:14.290 --> 00:30:17.346 And here you have an example where you
NOTE Confidence: 0.956023311666667
00:30:17.346 --> 00:30:20.479 have almost a quarter of million of cells.
NOTE Confidence: 0.956023311666667
00:30:20.480 --> 00:30:21.890 And with the same data set,
NOTE Confidence: 0.956023311666667
00:30:21.890 --> 00:30:24.122 you try to capture the data
NOTE Confidence: 0.956023311666667
00:30:24.122 --> 00:30:25.610 set with different approaches.
NOTE Confidence: 0.956023311666667
00:30:25.610 --> 00:30:27.806 So this is principal component analysis.
NOTE Confidence: 0.956023311666667

00:30:27.810 --> 00:30:29.286 With this disease Disney,
NOTE Confidence: 0.956023311666667

00:30:29.286 --> 00:30:31.908 this is the human approach and these
NOTE Confidence: 0.956023311666667

00:30:31.908 --> 00:30:34.302 others as methods like mono that tries
NOTE Confidence: 0.956023311666667

00:30:34.302 --> 00:30:36.970 also to perform trajectory inference.
NOTE Confidence: 0.956023311666667

00:30:36.970 --> 00:30:40.466 And you see how much the same data
NOTE Confidence: 0.956023311666667

00:30:40.466 --> 00:30:43.440 set is represented in different ways
NOTE Confidence: 0.956023311666667

00:30:43.440 --> 00:30:46.720 depending on the method that you use.
NOTE Confidence: 0.956023311666667

00:30:46.720 --> 00:30:48.533 And what you see here is also
NOTE Confidence: 0.956023311666667

00:30:48.533 --> 00:30:50.518 that when the data set is bigger,
NOTE Confidence: 0.956023311666667

00:30:50.520 --> 00:30:53.962 some of the methods do not even finish.
NOTE Confidence: 0.956023311666667

00:30:53.962 --> 00:30:56.076 So because the time is too much,
NOTE Confidence: 0.956023311666667

00:30:56.080 --> 00:30:59.206 and so you're not sure that some, uh,
NOTE Confidence: 0.956023311666667

00:30:59.206 --> 00:31:02.122 iterative methods that require a lot
NOTE Confidence: 0.956023311666667

00:31:02.122 --> 00:31:05.260 of steps converge because you finish,
NOTE Confidence: 0.956023311666667

00:31:05.260 --> 00:31:05.914 for example,
NOTE Confidence: 0.956023311666667

00:31:05.914 --> 00:31:08.203 the memory or the available time before

NOTE Confidence: 0.956023311666667
00:31:08.203 --> 00:31:10.840 they complete the necessary number of steps.
NOTE Confidence: 0.956023311666667
00:31:10.840 --> 00:31:12.130 And that's what you see here.
NOTE Confidence: 0.956023311666667
00:31:12.130 --> 00:31:14.878 But the point here is that, as usual,
NOTE Confidence: 0.956023311666667
00:31:14.878 --> 00:31:17.248 uh, depending on the choice.
NOTE Confidence: 0.956023311666667
00:31:17.250 --> 00:31:18.330 Of the tool.
NOTE Confidence: 0.956023311666667
00:31:18.330 --> 00:31:20.130 You have a different representation
NOTE Confidence: 0.956023311666667
00:31:20.130 --> 00:31:21.580 of the same data.
NOTE Confidence: 0.914453670833333
00:31:23.660 --> 00:31:26.724 And there is no clear way a automatic
NOTE Confidence: 0.914453670833333
00:31:26.724 --> 00:31:29.630 way to understand which is better.
NOTE Confidence: 0.917559690142857
00:31:32.360 --> 00:31:34.405 These also are some guidelines
NOTE Confidence: 0.917559690142857
00:31:34.405 --> 00:31:37.075 for selecting a like your best
NOTE Confidence: 0.917559690142857
00:31:37.075 --> 00:31:39.484 trajectory analysis tools based on.
NOTE Confidence: 0.917559690142857
00:31:39.484 --> 00:31:42.733 Also on the fact that some tools, uh,
NOTE Confidence: 0.917559690142857
00:31:42.733 --> 00:31:45.598 make assumptions on the trajectory.
NOTE Confidence: 0.917559690142857
00:31:45.600 --> 00:31:49.362 So some tools for example are only trying to
NOTE Confidence: 0.917559690142857

00:31:49.370 --> 00:31:52.660 model linear trajectories without branches.
NOTE Confidence: 0.917559690142857

00:31:52.660 --> 00:31:55.308 Some tools allow branches,
NOTE Confidence: 0.917559690142857

00:31:55.308 --> 00:31:57.294 but only bifurcation,
NOTE Confidence: 0.917559690142857

00:31:57.300 --> 00:31:59.400 so you can only have two choices
NOTE Confidence: 0.917559690142857

00:31:59.400 --> 00:32:01.089 when you have a decision.
NOTE Confidence: 0.917559690142857

00:32:01.090 --> 00:32:02.730 Some tools allow also the.
NOTE Confidence: 0.917559690142857

00:32:02.730 --> 00:32:06.430 Multiplication so from one, uh,
NOTE Confidence: 0.917559690142857

00:32:06.430 --> 00:32:08.936 let's say from one crossroads that you
NOTE Confidence: 0.917559690142857

00:32:08.936 --> 00:32:10.682 have multiple roads, possible roads,
NOTE Confidence: 0.917559690142857

00:32:10.682 --> 00:32:14.257 and some of the tools also allow you to have
NOTE Confidence: 0.917559690142857

00:32:14.257 --> 00:32:16.723 cycles inside the inside your trajectories.
NOTE Confidence: 0.917559690142857

00:32:16.730 --> 00:32:19.980 So depending on your assumptions,
NOTE Confidence: 0.917559690142857

00:32:19.980 --> 00:32:22.950 uh, these are different families
NOTE Confidence: 0.917559690142857

00:32:22.950 --> 00:32:26.450 of methods that you can use.
NOTE Confidence: 0.917559690142857

00:32:26.450 --> 00:32:27.806 And that available online.
NOTE Confidence: 0.917559690142857

00:32:27.806 --> 00:32:31.133 So this is all based on again a review

NOTE Confidence: 0.917559690142857
00:32:31.133 --> 00:32:33.168 of a trajectory analysis choices.
NOTE Confidence: 0.917559690142857
00:32:33.170 --> 00:32:34.112 It's quite recent.
NOTE Confidence: 0.917559690142857
00:32:34.112 --> 00:32:35.368 So two years ago,
NOTE Confidence: 0.917559690142857
00:32:35.370 --> 00:32:38.034 so it it captured for sure it contains
NOTE Confidence: 0.917559690142857
00:32:38.034 --> 00:32:40.550 for sure the most popular methods.
NOTE Confidence: 0.917559690142857
00:32:40.550 --> 00:32:44.350 That are also used now.
NOTE Confidence: 0.917559690142857
00:32:44.350 --> 00:32:47.563 As so, uh, one exception that is not here,
NOTE Confidence: 0.917559690142857
00:32:47.570 --> 00:32:48.898 and it's quite a.
NOTE Confidence: 0.917559690142857
00:32:48.898 --> 00:32:51.820 It's a method that had a lot of
NOTE Confidence: 0.917559690142857
00:32:51.820 --> 00:32:54.370 popularity and it's has some unique.
NOTE Confidence: 0.917559690142857
00:32:54.370 --> 00:32:57.250 It offers some unique insight,
NOTE Confidence: 0.917559690142857
00:32:57.250 --> 00:33:02.104 is called the air nevelocity and this is
NOTE Confidence: 0.917559690142857
00:33:02.104 --> 00:33:04.910 the last thing I will speak about today,
NOTE Confidence: 0.917559690142857
00:33:04.910 --> 00:33:08.830 so the paper was published three years ago,
NOTE Confidence: 0.917559690142857
00:33:08.830 --> 00:33:12.514 and so it's a method that
NOTE Confidence: 0.917559690142857

00:33:12.514 --> 00:33:14.970 analyzes a single seller.
NOTE Confidence: 0.917559690142857

00:33:14.970 --> 00:33:18.970 Using a biological insight,
NOTE Confidence: 0.917559690142857

00:33:18.970 --> 00:33:23.090 uh, that concerns splicing so.
NOTE Confidence: 0.917559690142857

00:33:23.090 --> 00:33:23.498 Uhm?
NOTE Confidence: 0.917559690142857

00:33:23.498 --> 00:33:26.354 You know that in human and in
NOTE Confidence: 0.917559690142857

00:33:26.354 --> 00:33:28.589 Elkhart excels there is the uh,
NOTE Confidence: 0.917559690142857

00:33:28.590 --> 00:33:30.250 when the RNA is transcribed,
NOTE Confidence: 0.917559690142857

00:33:30.250 --> 00:33:31.520 it has to be processed.
NOTE Confidence: 0.917559690142857

00:33:31.520 --> 00:33:33.728 One of the main steps of the processing
NOTE Confidence: 0.917559690142857

00:33:33.728 --> 00:33:36.496 is is a splicing that removes the intern.
NOTE Confidence: 0.917559690142857

00:33:36.500 --> 00:33:38.680 The entrance from your gene,
NOTE Confidence: 0.917559690142857

00:33:38.680 --> 00:33:40.105 and once the genie splice
NOTE Confidence: 0.917559690142857

00:33:40.105 --> 00:33:41.770 them is exported and so on.
NOTE Confidence: 0.917559690142857

00:33:41.770 --> 00:33:43.816 So the basic principle of RNA
NOTE Confidence: 0.917559690142857

00:33:43.816 --> 00:33:46.126 velocity is that when you perform
NOTE Confidence: 0.917559690142857

00:33:46.126 --> 00:33:47.866 single cell araneae seeker,

NOTE Confidence: 0.917559690142857
00:33:47.870 --> 00:33:50.390 you have some reads that are
NOTE Confidence: 0.917559690142857
00:33:50.390 --> 00:33:52.548 captured from unspliced RNA and
NOTE Confidence: 0.917559690142857
00:33:52.548 --> 00:33:54.548 some reads the tag captured.
NOTE Confidence: 0.917559690142857
00:33:54.550 --> 00:33:55.672 From splicer,
NOTE Confidence: 0.917559690142857
00:33:55.672 --> 00:33:59.599 RNA and the you can distinguish between
NOTE Confidence: 0.917559690142857
00:33:59.599 --> 00:34:02.388 unspliced reads and spliced reads.
NOTE Confidence: 0.917559690142857
00:34:02.390 --> 00:34:05.148 When you align the reader to the
NOTE Confidence: 0.917559690142857
00:34:05.148 --> 00:34:07.455 to your reference genome because
NOTE Confidence: 0.917559690142857
00:34:07.455 --> 00:34:10.050 spliced reads will not contain,
NOTE Confidence: 0.917559690142857
00:34:10.050 --> 00:34:11.886 will not contain introns.
NOTE Confidence: 0.917559690142857
00:34:11.886 --> 00:34:13.722 Basically while unspliced reads
NOTE Confidence: 0.917559690142857
00:34:13.722 --> 00:34:15.765 that will contain partially
NOTE Confidence: 0.917559690142857
00:34:15.765 --> 00:34:17.865 or totally intron sequences.
NOTE Confidence: 0.917559690142857
00:34:17.870 --> 00:34:20.299 So you can divide the reads into
NOTE Confidence: 0.917559690142857
00:34:20.299 --> 00:34:22.600 splice and and and splice them.
NOTE Confidence: 0.917559690142857

00:34:22.600 --> 00:34:24.880 So that's the basic assumption.

NOTE Confidence: 0.917559690142857

00:34:24.880 --> 00:34:29.347 And the second assumption is that you

NOTE Confidence: 0.917559690142857

00:34:29.347 --> 00:34:31.426 can calculate the ratio for each gene.

NOTE Confidence: 0.917559690142857

00:34:31.430 --> 00:34:33.670 The ratio between unspliced

NOTE Confidence: 0.917559690142857

00:34:33.670 --> 00:34:35.350 and spliced reads.

NOTE Confidence: 0.917559690142857

00:34:35.350 --> 00:34:38.886 And the assumption is that if you

NOTE Confidence: 0.917559690142857

00:34:38.886 --> 00:34:41.239 have a lot of unspliced reads,

NOTE Confidence: 0.917559690142857

00:34:41.240 --> 00:34:44.432 that means that the gene has a

NOTE Confidence: 0.917559690142857

00:34:44.432 --> 00:34:46.860 high transcription at the moment,

NOTE Confidence: 0.917559690142857

00:34:46.860 --> 00:34:49.996 and so it means that in the future

NOTE Confidence: 0.917559690142857

00:34:49.996 --> 00:34:53.069 probably that gene will be more abundant.

NOTE Confidence: 0.917559690142857

00:34:53.070 --> 00:34:54.670 In the splice the state,

NOTE Confidence: 0.917559690142857

00:34:54.670 --> 00:34:57.310 because if you are capturing a cell at

NOTE Confidence: 0.917559690142857

00:34:57.310 --> 00:34:59.755 time zero in the and you're capturing

NOTE Confidence: 0.917559690142857

00:34:59.755 --> 00:35:02.510 a lot of unspliced RNA in the future,

NOTE Confidence: 0.917559690142857

00:35:02.510 --> 00:35:05.373 that RNA will be spliced and so

NOTE Confidence: 0.917559690142857
00:35:05.373 --> 00:35:07.890 they spliced RNA will increase.
NOTE Confidence: 0.917559690142857
00:35:07.890 --> 00:35:08.988 On the opposite,
NOTE Confidence: 0.917559690142857
00:35:08.988 --> 00:35:11.550 if you have a lot of spliced
NOTE Confidence: 0.973415957272727
00:35:11.634 --> 00:35:14.968 RNA, you can predict that and not an slicer.
NOTE Confidence: 0.973415957272727
00:35:14.970 --> 00:35:17.721 Any you can predict the depth transcription
NOTE Confidence: 0.973415957272727
00:35:17.721 --> 00:35:20.791 at the moment is not shut off and
NOTE Confidence: 0.973415957272727
00:35:20.791 --> 00:35:24.570 displays there in a in the future will
NOTE Confidence: 0.973415957272727
00:35:24.570 --> 00:35:27.576 will be reduced because of degradation.
NOTE Confidence: 0.973415957272727
00:35:27.580 --> 00:35:29.974 So based on the ratio between unspliced
NOTE Confidence: 0.973415957272727
00:35:29.974 --> 00:35:32.156 and spliced that if you have a high
NOTE Confidence: 0.973415957272727
00:35:32.160 --> 00:35:35.030 proportion of a slice of unspliced St,
NOTE Confidence: 0.973415957272727
00:35:35.030 --> 00:35:36.596 you predict that in the future
NOTE Confidence: 0.973415957272727
00:35:36.596 --> 00:35:38.540 the gene will be more expressed.
NOTE Confidence: 0.973415957272727
00:35:38.540 --> 00:35:41.048 If you have no unspliced reads,
NOTE Confidence: 0.973415957272727
00:35:41.050 --> 00:35:43.864 you predict that adage in will
NOTE Confidence: 0.973415957272727

00:35:43.864 --> 00:35:46.849 be less expressed in the future.

NOTE Confidence: 0.973415957272727

00:35:46.850 --> 00:35:49.109 And so if you're measuring, uh,

NOTE Confidence: 0.973415957272727

00:35:49.109 --> 00:35:52.742 the present state of the cell using

NOTE Confidence: 0.973415957272727

00:35:52.742 --> 00:35:55.176 this concept, you can predict where

NOTE Confidence: 0.973415957272727

00:35:55.176 --> 00:35:57.900 the cell will be in the future,

NOTE Confidence: 0.973415957272727

00:35:57.900 --> 00:36:00.413 and so you can infer the position

NOTE Confidence: 0.973415957272727

00:36:00.413 --> 00:36:02.688 of the cell in the future,

NOTE Confidence: 0.973415957272727

00:36:02.690 --> 00:36:05.246 and you can connect the observed

NOTE Confidence: 0.973415957272727

00:36:05.246 --> 00:36:07.410 cell with the predicted cell,

NOTE Confidence: 0.973415957272727

00:36:07.410 --> 00:36:09.384 and that's the principle of RNA velocity.

NOTE Confidence: 0.973415957272727

00:36:09.390 --> 00:36:12.102 So the plots that you see with RNA

NOTE Confidence: 0.973415957272727

00:36:12.102 --> 00:36:15.258 velocity are these, where each cell.

NOTE Confidence: 0.973415957272727

00:36:15.258 --> 00:36:16.935 He's a, uh,

NOTE Confidence: 0.973415957272727

00:36:16.935 --> 00:36:19.855 connect is linked to an arrow and the

NOTE Confidence: 0.973415957272727

00:36:19.855 --> 00:36:22.164 arrow basically connects the cell

NOTE Confidence: 0.973415957272727

00:36:22.164 --> 00:36:25.376 with the prediction of where the cell

NOTE Confidence: 0.973415957272727
00:36:25.376 --> 00:36:28.064 will be in the future based on the
NOTE Confidence: 0.973415957272727
00:36:28.070 --> 00:36:30.650 rates of unspliced versus spliced
NOTE Confidence: 0.973415957272727
00:36:30.650 --> 00:36:34.200 genes that you capture in this cell.
NOTE Confidence: 0.973415957272727
00:36:34.200 --> 00:36:36.573 The advantage of these so these can
NOTE Confidence: 0.973415957272727
00:36:36.573 --> 00:36:39.447 be seen and also as a trajectory
NOTE Confidence: 0.973415957272727
00:36:39.447 --> 00:36:40.337 prediction tool,
NOTE Confidence: 0.973415957272727
00:36:40.340 --> 00:36:42.852 because once you see the map of all
NOTE Confidence: 0.973415957272727
00:36:42.852 --> 00:36:45.254 the errors you can visually see where
NOTE Confidence: 0.973415957272727
00:36:45.254 --> 00:36:47.792 the cells are going and you can and
NOTE Confidence: 0.973415957272727
00:36:47.792 --> 00:36:49.766 you can sort of infer traject trajectory
NOTE Confidence: 0.973415957272727
00:36:49.766 --> 00:36:52.176 and the advantage of this method is that,
NOTE Confidence: 0.973415957272727
00:36:52.180 --> 00:36:54.160 for example you don't have to
NOTE Confidence: 0.973415957272727
00:36:54.160 --> 00:36:56.440 select that that you know from the
NOTE Confidence: 0.973415957272727
00:36:56.440 --> 00:36:58.410 arrows where there is the root.
NOTE Confidence: 0.973415957272727
00:36:58.410 --> 00:37:01.332 So these arrows have a directionality
NOTE Confidence: 0.973415957272727

00:37:01.332 --> 00:37:04.210 so differently from these approaches.
NOTE Confidence: 0.973415957272727

00:37:04.210 --> 00:37:05.518 Yeah, you can link.
NOTE Confidence: 0.973415957272727

00:37:05.518 --> 00:37:08.352 You can not only have a trajectory but
NOTE Confidence: 0.973415957272727

00:37:08.352 --> 00:37:11.808 also have a trajectory with a direction.
NOTE Confidence: 0.92163537

00:37:13.850 --> 00:37:16.634 The problems are that arises obviously
NOTE Confidence: 0.92163537

00:37:16.634 --> 00:37:20.384 that all of this is based on the idea
NOTE Confidence: 0.92163537

00:37:20.384 --> 00:37:23.720 that you capture a lot of unspliced RNA,
NOTE Confidence: 0.92163537

00:37:23.720 --> 00:37:25.430 and this could be problematic,
NOTE Confidence: 0.92163537

00:37:25.430 --> 00:37:27.194 especially with certain libraries.
NOTE Confidence: 0.92163537

00:37:27.194 --> 00:37:30.018 For example, we have seen that three prime
NOTE Confidence: 0.92163537

00:37:30.018 --> 00:37:32.725 and enrich or 5:00 PM and then reached
NOTE Confidence: 0.92163537

00:37:32.725 --> 00:37:35.220 libraries do not cover all the genes,
NOTE Confidence: 0.92163537

00:37:35.220 --> 00:37:37.257 the gene, and so you can have
NOTE Confidence: 0.92163537

00:37:37.257 --> 00:37:39.458 some biases in the original paper.
NOTE Confidence: 0.92163537

00:37:39.460 --> 00:37:41.345 They show that this method
NOTE Confidence: 0.92163537

00:37:41.345 --> 00:37:43.696 works also with the 10X data.

NOTE Confidence: 0.92163537
00:37:43.696 --> 00:37:45.786 That are free prime based,
NOTE Confidence: 0.92163537
00:37:45.790 --> 00:37:49.750 but at least for the data set that they show.
NOTE Confidence: 0.92163537
00:37:49.750 --> 00:37:52.114 And this is part of their
NOTE Confidence: 0.92163537
00:37:52.114 --> 00:37:53.296 daily original paper.
NOTE Confidence: 0.92163537
00:37:53.300 --> 00:37:55.515 And always say this cells
NOTE Confidence: 0.92163537
00:37:55.515 --> 00:37:57.730 are visualized in a hour.
NOTE Confidence: 0.986790723333333
00:37:59.870 --> 00:38:01.582 Multi dimensional reduction space.
NOTE Confidence: 0.986790723333333
00:38:01.582 --> 00:38:05.010 For example, here you see again principal
NOTE Confidence: 0.986790723333333
00:38:05.010 --> 00:38:07.820 component want and principal component 2.
NOTE Confidence: 0.986790723333333
00:38:07.820 --> 00:38:10.670 Come I have two questions, yeah.
NOTE Confidence: 0.986790723333333
00:38:10.670 --> 00:38:14.905 Uhm, how would a for example splicing
NOTE Confidence: 0.986790723333333
00:38:14.905 --> 00:38:17.503 gene mutation would affect this?
NOTE Confidence: 0.986790723333333
00:38:17.503 --> 00:38:20.910 Hey you guys, your lab does do that.
NOTE Confidence: 0.986790723333333
00:38:20.910 --> 00:38:22.830 Yeah, this is a interesting.
NOTE Confidence: 0.986790723333333
00:38:22.830 --> 00:38:27.116 I don't recall that anybody, uh?
NOTE Confidence: 0.986790723333333

00:38:27.116 --> 00:38:29.380 This is interesting, obviously,
NOTE Confidence: 0.986790723333333

00:38:29.380 --> 00:38:33.628 because you expect a mutation in a splice
NOTE Confidence: 0.986790723333333

00:38:33.628 --> 00:38:37.798 factor to deviator like the the trajectory.
NOTE Confidence: 0.986790723333333

00:38:37.798 --> 00:38:41.312 I don't know if anybody did this.
NOTE Confidence: 0.986790723333333

00:38:41.312 --> 00:38:44.119 Uh, I have to say it didn't
NOTE Confidence: 0.986790723333333

00:38:44.119 --> 00:38:47.168 run like an exhaustive search,
NOTE Confidence: 0.986790723333333

00:38:47.170 --> 00:38:49.090 so it may be that in by oxide there is
NOTE Confidence: 0.986790723333333

00:38:49.145 --> 00:38:51.070 something about this but not not them.
NOTE Confidence: 0.986790723333333

00:38:51.070 --> 00:38:53.538 I I'm aware of.
NOTE Confidence: 0.986790723333333

00:38:53.540 --> 00:38:57.284 OK, so the other question is how many,
NOTE Confidence: 0.986790723333333

00:38:57.290 --> 00:38:58.538 roughly in proportion,
NOTE Confidence: 0.986790723333333

00:38:58.538 --> 00:39:01.842 how many genes can be seen or can
NOTE Confidence: 0.986790723333333

00:39:01.842 --> 00:39:04.649 be used to to project these maps?
NOTE Confidence: 0.986790723333333

00:39:04.650 --> 00:39:07.074 'cause I'm imagining only the highly
NOTE Confidence: 0.986790723333333

00:39:07.074 --> 00:39:09.136 highly expressed genes that normally
NOTE Confidence: 0.986790723333333

00:39:09.136 --> 00:39:11.410 contain introns can be used here,

NOTE Confidence: 0.98679072333333
00:39:11.410 --> 00:39:13.966 which shouldn't be that many right?
NOTE Confidence: 0.98679072333333
00:39:13.970 --> 00:39:15.470 Yeah, yes, so that's true.
NOTE Confidence: 0.98679072333333
00:39:15.470 --> 00:39:17.526 So here again in the paper they compare,
NOTE Confidence: 0.98679072333333
00:39:17.530 --> 00:39:19.690 uh, so this is mark two that that
NOTE Confidence: 0.98679072333333
00:39:19.690 --> 00:39:21.837 that this is the technology that is
NOTE Confidence: 0.98679072333333
00:39:21.837 --> 00:39:24.258 idea because you have a lot of reads.
NOTE Confidence: 0.98679072333333
00:39:24.260 --> 00:39:26.476 For each cell and you have the full
NOTE Confidence: 0.98679072333333
00:39:26.476 --> 00:39:29.300 coverage and they capture 22% of
NOTE Confidence: 0.98679072333333
00:39:29.300 --> 00:39:32.820 a set of reads that are unspliced.
NOTE Confidence: 0.98679072333333
00:39:32.820 --> 00:39:36.960 This is with the chromium, so this is 10XL.
NOTE Confidence: 0.98679072333333
00:39:36.960 --> 00:39:37.780 So yeah.
NOTE Confidence: 0.98679072333333
00:39:37.780 --> 00:39:40.300 So it seems that here the ratio is the same,
NOTE Confidence: 0.98679072333333
00:39:40.300 --> 00:39:42.892 but obviously they will be mostly
NOTE Confidence: 0.98679072333333
00:39:42.892 --> 00:39:45.934 the last team Trump and so yeah,
NOTE Confidence: 0.98679072333333
00:39:45.934 --> 00:39:49.000 so the fact that you capture more,
NOTE Confidence: 0.98679072333333

00:39:49.000 --> 00:39:51.502 uh, a selection of genes that
NOTE Confidence: 0.986790723333333
00:39:51.502 --> 00:39:52.753 are highly expressed.
NOTE Confidence: 0.986790723333333
00:39:52.760 --> 00:39:55.610 That's the underlying.
NOTE Confidence: 0.986790723333333
00:39:55.610 --> 00:39:57.140 Bias of all the analysis
NOTE Confidence: 0.986790723333333
00:39:57.140 --> 00:39:58.670 in at the single cell,
NOTE Confidence: 0.986790723333333
00:39:58.670 --> 00:40:02.079 and I assume that this is amplified
NOTE Confidence: 0.986790723333333
00:40:02.079 --> 00:40:04.860 in this sort of analysis.
NOTE Confidence: 0.986790723333333
00:40:04.860 --> 00:40:05.700 So, uh,
NOTE Confidence: 0.986790723333333
00:40:05.700 --> 00:40:09.060 but I I cannot give you a number
NOTE Confidence: 0.986790723333333
00:40:09.060 --> 00:40:12.828 of jeans because I I never used it.
NOTE Confidence: 0.986790723333333
00:40:12.830 --> 00:40:15.610 And so I don't have a in
NOTE Confidence: 0.986790723333333
00:40:15.610 --> 00:40:18.410 hand experience with this.
NOTE Confidence: 0.986790723333333
00:40:18.410 --> 00:40:18.700 K.
NOTE Confidence: 0.984464456875
00:40:21.140 --> 00:40:23.188 But yes, for sure limitation is on the
NOTE Confidence: 0.984464456875
00:40:23.188 --> 00:40:26.790 number of genes and on the number of, uh?
NOTE Confidence: 0.984464456875
00:40:26.790 --> 00:40:30.380 Yeah, and the hot on the also on the length

NOTE Confidence: 0.984464456875
00:40:30.467 --> 00:40:33.851 of the re read and how much coverage you
NOTE Confidence: 0.984464456875
00:40:33.851 --> 00:40:37.070 have from the Poly A tail for example.
NOTE Confidence: 0.984464456875
00:40:37.070 --> 00:40:38.500 I think that shorter jeans,
NOTE Confidence: 0.984464456875
00:40:38.500 --> 00:40:40.540 for example, jeans with few axons
NOTE Confidence: 0.984464456875
00:40:40.540 --> 00:40:42.730 and few interns that are shorter,
NOTE Confidence: 0.984464456875
00:40:42.730 --> 00:40:46.874 will be also like a. More captured,
NOTE Confidence: 0.984464456875
00:40:46.874 --> 00:40:49.968 more more present in this analysis than
NOTE Confidence: 0.984464456875
00:40:49.968 --> 00:40:52.900 than long jeans with long introns. Yeah.
NOTE Confidence: 0.966466048571429
00:40:54.910 --> 00:40:57.955 Uhm, OK, my last slide is about,
NOTE Confidence: 0.966466048571429
00:40:57.960 --> 00:40:59.965 uh, this collection of resources
NOTE Confidence: 0.966466048571429
00:40:59.965 --> 00:41:02.550 about single cell C can alesys,
NOTE Confidence: 0.966466048571429
00:41:02.550 --> 00:41:04.726 so the website is called the single seller.
NOTE Confidence: 0.966466048571429
00:41:04.730 --> 00:41:09.414 Any tools? And, uh, uh, so this is, uh,
NOTE Confidence: 0.966466048571429
00:41:09.414 --> 00:41:11.693 uh, like the trend of the number of
NOTE Confidence: 0.966466048571429
00:41:11.693 --> 00:41:13.989 tools that you can find that in these,
NOTE Confidence: 0.966466048571429

00:41:13.990 --> 00:41:14.800 uh, collection?

NOTE Confidence: 0.966466048571429

00:41:14.800 --> 00:41:18.040 So right now they are over 1000 of

NOTE Confidence: 0.966466048571429

00:41:18.129 --> 00:41:20.764 computational tools for the analysis

NOTE Confidence: 0.966466048571429

00:41:20.764 --> 00:41:23.949 of single cell, an Ernie silica.

NOTE Confidence: 0.966466048571429

00:41:23.949 --> 00:41:28.620 And here you see the stats on the platform.

NOTE Confidence: 0.966466048571429

00:41:28.620 --> 00:41:31.245 So on the languages that are

NOTE Confidence: 0.966466048571429

00:41:31.245 --> 00:41:33.120 mainly used by these tools,

NOTE Confidence: 0.966466048571429

00:41:33.120 --> 00:41:35.115 so most of them right now are

NOTE Confidence: 0.966466048571429

00:41:35.115 --> 00:41:36.760 in our but obviously.

NOTE Confidence: 0.966466048571429

00:41:36.760 --> 00:41:39.100 Almost every every tool is either

NOTE Confidence: 0.966466048571429

00:41:39.100 --> 00:41:40.660 in R or Python.

NOTE Confidence: 0.966466048571429

00:41:40.660 --> 00:41:43.000 Uhm, then you have C++.

NOTE Confidence: 0.966466048571429

00:41:43.000 --> 00:41:45.260 Probably these are covers some

NOTE Confidence: 0.966466048571429

00:41:45.260 --> 00:41:47.520 tools that are at being.

NOTE Confidence: 0.966466048571429

00:41:47.520 --> 00:41:49.830 That that needs to be performed.

NOTE Confidence: 0.966466048571429

00:41:49.830 --> 00:41:52.200 A complicated with efficiency from

NOTE Confidence: 0.966466048571429
00:41:52.200 --> 00:41:54.570 the computational point of view,
NOTE Confidence: 0.966466048571429
00:41:54.570 --> 00:41:57.559 and then you have some tools with
NOTE Confidence: 0.966466048571429
00:41:57.559 --> 00:42:00.906 Mark Lab and others and hear what
NOTE Confidence: 0.966466048571429
00:42:00.906 --> 00:42:04.603 you see is a divide these resources
NOTE Confidence: 0.966466048571429
00:42:04.603 --> 00:42:07.308 in categories depending on so.
NOTE Confidence: 0.966466048571429
00:42:07.310 --> 00:42:10.592 Some tools cover the full pipeline
NOTE Confidence: 0.966466048571429
00:42:10.592 --> 00:42:13.791 from at least from the digital gene
NOTE Confidence: 0.966466048571429
00:42:13.791 --> 00:42:15.813 expression from once you have the
NOTE Confidence: 0.966466048571429
00:42:15.813 --> 00:42:18.047 gene expression to all these steps.
NOTE Confidence: 0.966466048571429
00:42:18.050 --> 00:42:20.048 Jenna Alesys so they mention reduction,
NOTE Confidence: 0.966466048571429
00:42:20.050 --> 00:42:21.538 clustering and so on,
NOTE Confidence: 0.966466048571429
00:42:21.538 --> 00:42:23.770 and some tools are more specific.
NOTE Confidence: 0.966466048571429
00:42:23.770 --> 00:42:26.194 So if you look at the frequency you
NOTE Confidence: 0.966466048571429
00:42:26.194 --> 00:42:29.166 have most of the tools there are about
NOTE Confidence: 0.966466048571429
00:42:29.166 --> 00:42:31.110 visualization of single cell data.
NOTE Confidence: 0.966466048571429

00:42:31.110 --> 00:42:33.162 40% are about visualization.

NOTE Confidence: 0.966466048571429

00:42:33.162 --> 00:42:36.240 Then second position you have clustering

NOTE Confidence: 0.966466048571429

00:42:36.314 --> 00:42:38.468 and dimensionality reduction.

NOTE Confidence: 0.966466048571429

00:42:38.470 --> 00:42:39.830 I didn't speak about this,

NOTE Confidence: 0.966466048571429

00:42:39.830 --> 00:42:42.742 but it's also very important if the

NOTE Confidence: 0.966466048571429

00:42:42.742 --> 00:42:44.760 integration of different data sets.

NOTE Confidence: 0.966466048571429

00:42:44.760 --> 00:42:46.755 So this means the integration

NOTE Confidence: 0.966466048571429

00:42:46.755 --> 00:42:48.750 of different single cell RNA.

NOTE Confidence: 0.966466048571429

00:42:48.750 --> 00:42:49.183 Yes,

NOTE Confidence: 0.966466048571429

00:42:49.183 --> 00:42:51.348 very much and also integration

NOTE Confidence: 0.966466048571429

00:42:51.348 --> 00:42:52.647 of multiple modalities.

NOTE Confidence: 0.966466048571429

00:42:52.650 --> 00:42:55.479 So for example one yeah there

NOTE Confidence: 0.966466048571429

00:42:55.479 --> 00:42:58.342 are a lot of techniques now that

NOTE Confidence: 0.966466048571429

00:42:58.342 --> 00:43:00.584 enables to capture for example

NOTE Confidence: 0.966466048571429

00:43:00.584 --> 00:43:03.254 the RNA levels and also the.

NOTE Confidence: 0.966466048571429

00:43:03.260 --> 00:43:05.615 Some chromatin, uh,

NOTE Confidence: 0.966466048571429
00:43:05.615 --> 00:43:09.156 open versus closed state uh,
NOTE Confidence: 0.966466048571429
00:43:09.156 --> 00:43:10.740 in the same cell,
NOTE Confidence: 0.966466048571429
00:43:10.740 --> 00:43:14.506 and so there are about how to
NOTE Confidence: 0.966466048571429
00:43:14.506 --> 00:43:17.390 integrate these multiple sources of
NOTE Confidence: 0.966466048571429
00:43:17.390 --> 00:43:19.930 information and multiple datasets.
NOTE Confidence: 0.966466048571429
00:43:19.930 --> 00:43:21.650 And then you have over there engine actors,
NOTE Confidence: 0.966466048571429
00:43:21.650 --> 00:43:22.410 differential expressions,
NOTE Confidence: 0.966466048571429
00:43:22.410 --> 00:43:25.450 so a lot of topics that we that
NOTE Confidence: 0.966466048571429
00:43:25.519 --> 00:43:27.517 we covered them in these last.
NOTE Confidence: 0.907150720909091
00:43:30.240 --> 00:43:33.950 Session so if you go and if you look at uh,
NOTE Confidence: 0.907150720909091
00:43:33.950 --> 00:43:35.100 you, you find the tool.
NOTE Confidence: 0.907150720909091
00:43:35.100 --> 00:43:38.780 Set the platform a platform,
NOTE Confidence: 0.907150720909091
00:43:38.780 --> 00:43:40.260 then the number of citations.
NOTE Confidence: 0.907150720909091
00:43:40.260 --> 00:43:41.760 For example, you can see
NOTE Confidence: 0.907150720909091
00:43:41.760 --> 00:43:43.260 which tools are more popular.
NOTE Confidence: 0.907150720909091

00:43:43.260 --> 00:43:45.396 We we we respect to others.

NOTE Confidence: 0.907150720909091

00:43:45.400 --> 00:43:47.670 If I have to make a choice and

NOTE Confidence: 0.907150720909091

00:43:47.670 --> 00:43:51.020 the advantage that is quite

NOTE Confidence: 0.907150720909091

00:43:51.020 --> 00:43:53.610 comprehensive and updated weekly.

NOTE Confidence: 0.9908934

00:44:00.840 --> 00:44:01.330 So.

NOTE Confidence: 0.95894715

00:44:05.290 --> 00:44:07.000 These eyes.