WEBVTT

NOTE duration:"00:16:49"

NOTE recognizability:0.942

NOTE language:en-us

NOTE Confidence: 0.97942126125

00:00:02.820 --> 00:00:05.774 OK, so today's talk will be mainly

NOTE Confidence: 0.97942126125

00:00:05.774 --> 00:00:07.604 about the analysis of single

NOTE Confidence: 0.97942126125

00:00:07.604 --> 00:00:10.328 cell RNA seekers, so we had.

NOTE Confidence: 0.97942126125

00:00:10.328 --> 00:00:13.008 Last time, like the last three

NOTE Confidence: 0.97942126125

00:00:13.008 --> 00:00:15.360 times with dealt with the analysis

NOTE Confidence: 0.97942126125

00:00:15.434 --> 00:00:18.350 of biker NEC classic methods and

NOTE Confidence: 0.97942126125

00:00:18.350 --> 00:00:20.294 pathway richemond analysis method,

NOTE Confidence: 0.97942126125

00:00:20.300 --> 00:00:22.508 I have some slides because last

NOTE Confidence: 0.97942126125

00:00:22.508 --> 00:00:25.310 time I remember there was there were

NOTE Confidence: 0.97942126125

00:00:25.310 --> 00:00:27.395 questions more specific about the

NOTE Confidence: 0.97942126125

00:00:27.400 --> 00:00:30.340 multiple test correction and so I

NOTE Confidence: 0.97942126125

00:00:30.340 --> 00:00:33.688 have these lines at least to like.

NOTE Confidence: 0.97942126125

00:00:33.690 --> 00:00:36.791 Try to explain two of the methods

NOTE Confidence: 0.97942126125

00:00:36.791 --> 00:00:40.107 that are used in order to do that.

NOTE Confidence: 0.97942126125

00:00:40.110 --> 00:00:40.916 So here.

NOTE Confidence: 0.97942126125

00:00:40.916 --> 00:00:43.737 So we start with the assumption that

NOTE Confidence: 0.97942126125

00:00:43.737 --> 00:00:46.796 we are bound are in our statistical

NOTE Confidence: 0.97942126125

00:00:46.796 --> 00:00:50.470 analysis to that P value threshold of 0.05.

NOTE Confidence: 0.97942126125

00:00:50.470 --> 00:00:52.990 So whenever we run a test on

NOTE Confidence: 0.97942126125

00:00:52.990 --> 00:00:56.197 whatever we want to test and they,

NOTE Confidence: 0.97942126125

00:00:56.197 --> 00:00:59.911 it seems that the most important

NOTE Confidence: 0.97942126125

00:00:59.911 --> 00:01:03.666 thing is that you have a P value of

NOTE Confidence: 0.97942126125

00:01:03.670 --> 00:01:07.782 less than 0.00 less point than 0.05.

NOTE Confidence: 0.97942126125

00:01:07.782 --> 00:01:11.260 So RP value of 0.05 means that

NOTE Confidence: 0.97942126125

00:01:11.260 --> 00:01:13.060 the probability of rejecting.

NOTE Confidence: 0.97942126125

00:01:13.060 --> 00:01:17.085 You're not hypothesis when it's true is 5%.

NOTE Confidence: 0.97942126125

00:01:17.085 --> 00:01:18.540 So for example,

NOTE Confidence: 0.97942126125

00:01:18.540 --> 00:01:21.425 when we are comparing gene expression,

NOTE Confidence: 0.97942126125

00:01:21.425 --> 00:01:23.800 we're performing a test of

NOTE Confidence: 0.97942126125

00:01:23.800 --> 00:01:25.371 differential expression of 1

NOTE Confidence: 0.97942126125

00:01:25.371 --> 00:01:27.033 gene between condition A and B.

NOTE Confidence: 0.97942126125

00:01:27.040 --> 00:01:29.987 It means that they are the null

NOTE Confidence: 0.97942126125

00:01:29.987 --> 00:01:33.070 hypothesis is that the Gina is

NOTE Confidence: 0.97942126125

00:01:33.070 --> 00:01:34.798 not differentially expressed.

NOTE Confidence: 0.97942126125

00:01:34.800 --> 00:01:37.150 So if we have a P value of 0.05,

NOTE Confidence: 0.97942126125

00:01:37.150 --> 00:01:40.430 it means that we can reject the null

NOTE Confidence: 0.97942126125

00:01:40.430 --> 00:01:43.465 hypothesis so that the gene is not changing.

NOTE Confidence: 0.97942126125

00:01:43.470 --> 00:01:43.790 Yeah,

NOTE Confidence: 0.97942126125

00:01:43.790 --> 00:01:46.030 and and the probability of that and

NOTE Confidence: 0.97942126125

00:01:46.030 --> 00:01:48.206 Dolly Parton is to be true is only 5%,

NOTE Confidence: 0.97942126125

00:01:48.210 --> 00:01:50.066 so that's why the lower the P value,

NOTE Confidence: 0.97942126125

00:01:50.070 --> 00:01:53.634 the more confident we are in

NOTE Confidence: 0.97942126125

00:01:53.634 --> 00:01:56.010 rejecting the null hypothesis.

NOTE Confidence: 0.97942126125

00:01:56.010 --> 00:01:57.438 One of the limitations of this

NOTE Confidence: 0.97942126125

00:01:57.438 --> 00:01:59.570 is that the P value doesn't say

NOTE Confidence: 0.97942126125

00:01:59.570 --> 00:02:01.058 anything about our hypothesis,

NOTE Confidence: 0.97942126125

00:02:01.060 --> 00:02:04.402 so that the gene is really

NOTE Confidence: 0.97942126125

00:02:04.402 --> 00:02:05.918 differentially expressed UM.

NOTE Confidence: 0.97942126125

00:02:05.918 --> 00:02:09.980 So when we run analysis in high to put way,

NOTE Confidence: 0.97942126125

00:02:09.980 --> 00:02:12.140 we usually have multiple comparison.

NOTE Confidence: 0.97942126125

00:02:12.140 --> 00:02:14.858 So it means that we run a test of

NOTE Confidence: 0.97942126125

00:02:14.858 --> 00:02:16.126 differential expression differential

NOTE Confidence: 0.97942126125

00:02:16.126 --> 00:02:18.808 expression for each gene and we

NOTE Confidence: 0.97942126125

00:02:18.808 --> 00:02:21.110 usually have 10 to 20,000 genes.

NOTE Confidence: 0.97942126125

00:02:21.110 --> 00:02:24.140 For example, in our NGS experiments,

NOTE Confidence: 0.97942126125

00:02:24.140 --> 00:02:26.905 or when we do pathway richemond analysis.

NOTE Confidence: 0.97942126125

00:02:26.910 --> 00:02:28.770 We're testing their retreat of

NOTE Confidence: 0.97942126125

00:02:28.770 --> 00:02:30.630 hundreds or thousands of pathways.

NOTE Confidence: 0.97942126125

00:02:30.630 --> 00:02:33.492 That means that we run a lot of tests

NOTE Confidence: 0.97942126125

00:02:33.500 --> 00:02:36.722 and For these reasons that probability

NOTE Confidence: 0.97942126125

00:02:36.722 --> 00:02:39.780 that is 5% to have false positives.

NOTE Confidence: 0.97942126125

00:02:39.780 --> 00:02:42.718 So to say that the gene is

NOTE Confidence: 0.97942126125

00:02:42.718 --> 00:02:44.118 differentially expressed,

NOTE Confidence: 0.97942126125

00:02:44.120 --> 00:02:47.472 when in reality it is not rises with

NOTE Confidence: 0.97942126125

00:02:47.472 --> 00:02:50.174 the number of tests that we run.

NOTE Confidence: 0.97942126125

00:02:50.180 --> 00:02:51.688 And so, for example,

NOTE Confidence: 0.97942126125

00:02:51.688 --> 00:02:53.950 here you see a numeric example,

NOTE Confidence: 0.97942126125

00:02:53.950 --> 00:02:56.929 so if you come assume that we you run

NOTE Confidence: 0.97942126125

00:02:56.929 --> 00:02:59.640 a differential expression analysis

NOTE Confidence: 0.97942126125

00:02:59.640 --> 00:03:02.696 on next generation sequencing,

NOTE Confidence: 0.97942126125

00:03:02.700 --> 00:03:03.903 you have a,

NOTE Confidence: 0.97942126125

00:03:03.903 --> 00:03:06.710 you have a collection of data on

NOTE Confidence: 0.97942126125

00:03:06.809 --> 00:03:10.365 10,000 genes and you select 1000 genes,

NOTE Confidence: 0.97942126125

00:03:10.370 --> 00:03:15.420 so 10% with a P value that is less than 0.05.

NOTE Confidence: 0.97942126125

00:03:15.420 --> 00:03:16.025 Now,

NOTE Confidence: 0.97942126125

00:03:16.025 --> 00:03:19.898 since you run 10,000 test and you

NOTE Confidence: 0.97942126125

00:03:19.898 --> 00:03:22.990 accepted the 5% probability to make an error,

NOTE Confidence: 0.97942126125

00:03:22.990 --> 00:03:25.755 that means that you expected by chance

NOTE Confidence: 0.97942126125

00:03:25.755 --> 00:03:31.410 to have a 5% of full 5% of tests that

NOTE Confidence: 0.97942126125

00:03:31.410 --> 00:03:36.998 are true that are significant but not true.

NOTE Confidence: 0.97942126125

00:03:37.000 --> 00:03:41.830 So we expect 500 out of the 10,000 genes to

NOTE Confidence: 0.974495814

00:03:41.952 --> 00:03:44.537 be picked up by chance.

NOTE Confidence: 0.974495814

00:03:44.540 --> 00:03:48.482 If we compare so 500 is what we expect

NOTE Confidence: 0.974495814

00:03:48.490 --> 00:03:50.290 is the number of genes we expect it

NOTE Confidence: 0.974495814

00:03:50.290 --> 00:03:52.399 to be in our list of differentially

NOTE Confidence: 0.974495814

00:03:52.399 --> 00:03:53.994 expressed when they are not,

NOTE Confidence: 0.974495814

00:03:54.000 --> 00:03:58.617 and it's 50% of the 1000 that we find.

NOTE Confidence: 0.974495814

00:03:58.620 --> 00:04:04.140 So if we calculate 500 out of 1000,

NOTE Confidence: 0.974495814

00:04:04.140 --> 00:04:07.112 that means that we have a 50% of genes

NOTE Confidence: 0.974495814

00:04:07.112 --> 00:04:09.989 that we expect to be false positive.

NOTE Confidence: 0.974495814

00:04:09.990 --> 00:04:13.071 So this ratio is 50% is what we

NOTE Confidence: 0.974495814

00:04:13.071 --> 00:04:14.956 call the false discovery rate,

NOTE Confidence: 0.974495814

00:04:14.960 --> 00:04:17.907 and in this case using just these

NOTE Confidence: 0.974495814

00:04:17.907 --> 00:04:20.634 values is 50%. So it's very high.

NOTE Confidence: 0.974495814

00:04:20.634 --> 00:04:23.235 It means that potentially one out of

NOTE Confidence: 0.974495814

00:04:23.235 --> 00:04:25.517 two genes can be a false positive.

NOTE Confidence: 0.974495814

00:04:25.520 --> 00:04:28.310 So multiple test correction methods are

NOTE Confidence: 0.974495814

00:04:28.310 --> 00:04:32.060 ways that to modify the original P values,

NOTE Confidence: 0.974495814

00:04:32.060 --> 00:04:34.592 particularly to increase the original P

NOTE Confidence: 0.974495814

00:04:34.592 --> 00:04:37.879 values so that this probability of having

NOTE Confidence: 0.974495814

00:04:37.879 --> 00:04:40.424 false positives is ultimately less.

NOTE Confidence: 0.974495814

00:04:40.430 --> 00:04:41.420 It is reduced.

NOTE Confidence: 0.98081213

00:04:44.820 --> 00:04:46.890 Yeah, I just want to make a quick

NOTE Confidence: 0.981403573333333

00:04:46.890 --> 00:04:50.106 addition to as this is a very important

NOTE Confidence: 0.981403573333333

00:04:50.106 --> 00:04:52.340 concept and the first time I see

NOTE Confidence: 0.981403573333333

00:04:52.340 --> 00:04:54.629 encounters a word not have processes.

NOTE Confidence: 0.981403573333333

00:04:54.630 --> 00:04:56.706 Always got confused initially or so.

NOTE Confidence: 0.981403573333333

00:04:56.710 --> 00:04:59.166 Why is the heck is not hypothesis also

NOTE Confidence: 0.981403573333333

00:04:59.166 --> 00:05:01.479 not purposes we can consider As for

NOTE Confidence: 0.981403573333333

00:05:01.479 --> 00:05:03.496 example if you're comparing two groups

NOTE Confidence: 0.981403573333333

00:05:03.496 --> 00:05:06.104 of samples and not help us is basically

NOTE Confidence: 0.981403573333333

00:05:06.104 --> 00:05:07.582 there's there's a 21 gene the gene

NOTE Confidence: 0.981403573333333

00:05:07.582 --> 00:05:08.916 you're looking at has no difference

NOTE Confidence: 0.981403573333333

00:05:08.916 --> 00:05:10.524 between these two groups, for example,

NOTE Confidence: 0.981403573333333

00:05:10.524 --> 00:05:13.060 so this will be done on hypothesis and the.

NOTE Confidence: 0.981403573333333

00:05:13.060 --> 00:05:15.870 5% P value oh point. 5.050 point

NOTE Confidence: 0.985787001666667

00:05:18.340 --> 00:05:20.775 05 means you have a 5% chance

NOTE Confidence: 0.985787001666667

00:05:20.775 --> 00:05:23.505 to make a wrong call basically,

NOTE Confidence: 0.985787001666667

00:05:23.510 --> 00:05:25.118 so that's that's eventually.

NOTE Confidence: 0.985787001666667

00:05:25.118 --> 00:05:27.128 II interpreted that way myself.

NOTE Confidence: 0.98298246

00:05:27.700 --> 00:05:29.280 Yeah, exactly. So that's the

NOTE Confidence: 0.98298246

00:05:29.280 --> 00:05:31.220 first positive is a wrong code.

NOTE Confidence: 0.98298246

00:05:31.220 --> 00:05:34.524 So basically yes, and when you do

NOTE Confidence: 0.98298246

00:05:34.524 --> 00:05:36.805 pathway richemond Denali part is

NOTE Confidence: 0.98298246

00:05:36.805 --> 00:05:39.528 is that the pathway is not reached?

NOTE Confidence: 0.98298246

00:05:39.530 --> 00:05:43.175 So those are the two examples that I'm doing.

NOTE Confidence: 0.98298246

00:05:43.180 --> 00:05:45.922 Uhm, so all the multiple test

NOTE Confidence: 0.98298246

00:05:45.922 --> 00:05:47.750 correction methods increase the

NOTE Confidence: 0.98298246

00:05:47.832 --> 00:05:50.352 original key values with the aim

NOTE Confidence: 0.98298246

00:05:50.352 --> 00:05:52.860 of reducing these false positives.

NOTE Confidence: 0.98298246

00:05:52.860 --> 00:05:54.940 What they do not do is that they

NOTE Confidence: 0.98298246

00:05:54.940 --> 00:05:57.335 do not swap the order of P values,

NOTE Confidence: 0.98298246

00:05:57.340 --> 00:05:59.052 so you can imagine if you have your

NOTE Confidence: 0.98298246

00:05:59.052 --> 00:06:01.206 list of genes and your rank your genes

NOTE Confidence: 0.98298246

00:06:01.206 --> 00:06:02.820 according to the original P values.

NOTE Confidence: 0.98298246

00:06:02.820 --> 00:06:05.105 When you perform the transformation

NOTE Confidence: 0.98298246

00:06:05.105 --> 00:06:06.476 or the correction,

NOTE Confidence: 0.98298246

00:06:06.480 --> 00:06:09.198 you don't have changes in the

NOTE Confidence: 0.98298246

00:06:09.198 --> 00:06:11.010 rank of the jeans.

NOTE Confidence: 0.98298246

00:06:11.010 --> 00:06:13.965 So this simplest correction is

NOTE Confidence: 0.98298246

00:06:13.965 --> 00:06:16.329 the Bonferroni correction method.

NOTE Confidence: 0.98298246

00:06:16.330 --> 00:06:18.262 It's simple because you just take

NOTE Confidence: 0.98298246

00:06:18.262 --> 00:06:20.238 the original P value and you

NOTE Confidence: 0.98298246

00:06:20.238 --> 00:06:22.098 multiply the original P value for

NOTE Confidence: 0.98298246

00:06:22.098 --> 00:06:24.336 the number of tests that you perform.

NOTE Confidence: 0.98298246

00:06:24.340 --> 00:06:25.630 So in our case, for example,

NOTE Confidence: 0.98298246

00:06:25.630 --> 00:06:27.716 if you run tests for 10,000 genes,

NOTE Confidence: 0.98298246

00:06:27.720 --> 00:06:29.850 that means you take your original

NOTE Confidence: 0.98298246

00:06:29.850 --> 00:06:32.108 P values and you multiply each

NOTE Confidence: 0.98298246

00:06:32.108 --> 00:06:34.400 of these P values for 10,000.

NOTE Confidence: 0.98298246

00:06:34.400 --> 00:06:35.720 So that this is the formula,

NOTE Confidence: 0.98298246

00:06:35.720 --> 00:06:36.250 there just

NOTE Confidence: 0.980613493333333

00:06:36.950 --> 00:06:38.290 have a question I I'm

NOTE Confidence: 0.980613493333333

00:06:38.290 --> 00:06:39.362 understanding what you're saying,

NOTE Confidence: 0.980613493333333

00:06:39.370 --> 00:06:41.986 but there's adjusted P minus value.

NOTE Confidence: 0.896916326666667

00:06:42.810 --> 00:06:44.100 I know it's a P value,

NOTE Confidence: 0.896916326666667

00:06:44.100 --> 00:06:45.452 it's inequation, no, no.

NOTE Confidence: 0.896916326666667

00:06:45.452 --> 00:06:48.349 It's up evalue with a very large space,

NOTE Confidence: 0.896916326666667

00:06:48.350 --> 00:06:52.728 minus yeah. Oh yeah, yeah.

NOTE Confidence: 0.896916326666667

00:06:52.730 --> 00:06:55.818 OK, it's devalue so.

NOTE Confidence: 0.896916326666667

00:06:55.820 --> 00:06:59.490 I will correct this so it's just evalue.

NOTE Confidence: 0.896916326666667

00:06:59.490 --> 00:07:01.326 Yeah, it's the original P value.

NOTE Confidence: 0.896916326666667

00:07:01.330 --> 00:07:03.836 Multiply that by the number of tests.

NOTE Confidence: 0.896916326666667

00:07:03.840 --> 00:07:08.118 OK, thank you and this belongs to a family

NOTE Confidence: 0.896916326666667

00:07:08.118 --> 00:07:10.810 of correction methods that control for

NOTE Confidence: 0.896916326666667

00:07:10.810 --> 00:07:13.576 the so called familywise error rate.

NOTE Confidence: 0.896916326666667

00:07:13.580 --> 00:07:16.460 So after the correction when you

NOTE Confidence: 0.896916326666667

00:07:16.460 --> 00:07:19.465 take everything that is below 0.05,

NOTE Confidence: 0.896916326666667

00:07:19.465 --> 00:07:22.855 interpretation of that is that you're

NOTE Confidence: 0.896916326666667

00:07:22.855 --> 00:07:26.395 relying 5% of probability to have at least.

NOTE Confidence: 0.896916326666667

00:07:26.400 --> 00:07:27.663 One false positive.

NOTE Confidence: 0.896916326666667

00:07:27.663 --> 00:07:29.768 So you're controlling for the

NOTE Confidence: 0.896916326666667

00:07:29.768 --> 00:07:32.102 probability of having in your final

NOTE Confidence: 0.896916326666667

00:07:32.102 --> 00:07:34.232 list at least one false positive.

NOTE Confidence: 0.896916326666667

00:07:34.240 --> 00:07:36.580 That's why it's a very

NOTE Confidence: 0.896916326666667

00:07:36.580 --> 00:07:37.516 conservative correction,

NOTE Confidence: 0.896916326666667

00:07:37.520 --> 00:07:39.360 so it's very stringent because

NOTE Confidence: 0.896916326666667

00:07:39.360 --> 00:07:41.896 basically you are not allowing to have

NOTE Confidence: 0.896916326666667

00:07:41.896 --> 00:07:44.244 any false positive at all almost.

NOTE Confidence: 0.896916326666667

00:07:44.244 --> 00:07:45.180 And so,

NOTE Confidence: 0.896916326666667

00:07:45.180 --> 00:07:46.540 especially when you have a

NOTE Confidence: 0.896916326666667

00:07:46.540 --> 00:07:47.628 large number of tests.

NOTE Confidence: 0.896916326666667

00:07:47.630 --> 00:07:49.730 Since this is the number that you

NOTE Confidence: 0.896916326666667

00:07:49.730 --> 00:07:51.388 multiply your original P values for,

NOTE Confidence: 0.896916326666667

00:07:51.390 --> 00:07:54.720 it can be very not rewarding,

NOTE Confidence: 0.896916326666667

00:07:54.720 --> 00:07:56.450 meaning that after the correction.

NOTE Confidence: 0.896916326666667

00:07:56.450 --> 00:07:58.730 Would ever would be values

NOTE Confidence: 0.896916326666667

00:07:58.730 --> 00:08:00.554 basically reduced to 1.

NOTE Confidence: 0.896916326666667

00:08:00.560 --> 00:08:02.564 And so that no gene after

NOTE Confidence: 0.896916326666667

00:08:02.564 --> 00:08:03.900 the correction is selected.

NOTE Confidence: 0.896916326666667

00:08:03.900 --> 00:08:06.260 So that's why this is simple to explain,

NOTE Confidence: 0.896916326666667

00:08:06.260 --> 00:08:07.740 but it's rarely used.

NOTE Confidence: 0.896916326666667

00:08:07.740 --> 00:08:10.529 The most common method that is used

NOTE Confidence: 0.896916326666667

00:08:10.529 --> 00:08:13.019 is the Benjamini Hochberg correction,

NOTE Confidence: 0.896916326666667

00:08:13.020 --> 00:08:15.210 so this is the most popular

NOTE Confidence: 0.896916326666667

00:08:15.210 --> 00:08:16.670 multiple test correction method

NOTE Confidence: 0.896916326666667

00:08:16.741 --> 00:08:18.847 was introduced there 25 years ago,

NOTE Confidence: 0.896916326666667

00:08:18.850 --> 00:08:20.985 so this is belongs to a family

NOTE Confidence: 0.896916326666667

00:08:20.985 --> 00:08:22.825 of methods that designed to

NOTE Confidence: 0.896916326666667

00:08:22.825 --> 00:08:24.960 control the false discovery rate.

NOTE Confidence: 0.896916326666667

00:08:24.960 --> 00:08:30.008 So the proportion of false positives that we.

NOTE Confidence: 0.896916326666667

00:08:30.010 --> 00:08:32.298 One thing that we expect in our data,

NOTE Confidence: 0.896916326666667

00:08:32.300 --> 00:08:35.492 so we're not if we select

NOTE Confidence: 0.896916326666667

00:08:35.492 --> 00:08:37.088 after the correction.

NOTE Confidence: 0.896916326666667

00:08:37.090 --> 00:08:41.922 0.05 it means that we allow 5% of our genes

NOTE Confidence: 0.896916326666667

00:08:41.922 --> 00:08:45.360 to be wrong calls or false positives.

NOTE Confidence: 0.896916326666667

00:08:45.360 --> 00:08:47.030 So it's a stepwise method,

NOTE Confidence: 0.896916326666667

00:08:47.030 --> 00:08:49.352 so it's it's a more slightly

NOTE Confidence: 0.896916326666667

00:08:49.352 --> 00:08:51.370 more complex than the bond,

NOTE Confidence: 0.896916326666667

00:08:51.370 --> 00:08:52.606 then the Bonferroni,

NOTE Confidence: 0.896916326666667

00:08:52.606 --> 00:08:55.097 but the formula is quite straightforward,

NOTE Confidence: 0.896916326666667

00:08:55.097 --> 00:08:58.940 so it requires a first year that you sort

NOTE Confidence: 0.896916326666667

00:08:59.016 --> 00:09:02.327 all your original values in increasing order.

NOTE Confidence: 0.896916326666667

00:09:02.330 --> 00:09:06.274 So from the smallest to the biggest value,

NOTE Confidence: 0.896916326666667

00:09:06.280 --> 00:09:08.856 so that you can calculate the rank.

NOTE Confidence: 0.896916326666667

00:09:08.860 --> 00:09:11.149 So this more length is the smallest

NOTE Confidence: 0.896916326666667

00:09:11.149 --> 00:09:14.416 has rank one, and then 2-3 and so on,

NOTE Confidence: 0.896916326666667

00:09:14.420 --> 00:09:15.830 so they adjusted.

NOTE Confidence: 0.896916326666667

00:09:15.830 --> 00:09:18.650 Failure is basically the original P

NOTE Confidence: 0.896916326666667

00:09:18.650 --> 00:09:21.734 value multiplied by the number of tests

NOTE Confidence: 0.896916326666667

00:09:21.734 --> 00:09:24.780 divided by the rank of the variable.

NOTE Confidence: 0.896916326666667

00:09:24.780 --> 00:09:27.306 So this means that you don't

NOTE Confidence: 0.896916326666667

00:09:27.306 --> 00:09:29.809 multiply your original value for a

NOTE Confidence: 0.896916326666667

00:09:29.809 --> 00:09:32.023 fixed number as in the Bonferroni,

NOTE Confidence: 0.896916326666667

00:09:32.030 --> 00:09:34.575 but the multiplication the amount

NOTE Confidence: 0.896916326666667

00:09:34.575 --> 00:09:37.120 of the multiplication depends on

NOTE Confidence: 0.896916326666667

00:09:37.204 --> 00:09:39.829 the rank of your original P value.

NOTE Confidence: 0.896916326666667

00:09:39.830 --> 00:09:43.142 And I have these examples to show the

NOTE Confidence: 0.896916326666667

00:09:43.142 --> 00:09:45.449 difference between the two approaches,

NOTE Confidence: 0.896916326666667

00:09:45.450 --> 00:09:47.670 so it's a simplified example.

NOTE Confidence: 0.896916326666667

00:09:47.670 --> 00:09:49.498 We run an analysis.

NOTE Confidence: 0.896916326666667

00:09:49.498 --> 00:09:52.240 Let's assume we run an analysis

NOTE Confidence: 0.896916326666667

00:09:52.333 --> 00:09:55.010 of Honor 5353 jeans.

NOTE Confidence: 0.896916326666667

00:09:55.010 --> 00:09:57.182 So these are our genes we're

NOTE Confidence: 0.896916326666667

00:09:57.182 --> 00:09:58.630 testing for differential expression

NOTE Confidence: 0.896916326666667

00:09:58.693 --> 00:10:00.448 of these genes into conditions,

NOTE Confidence: 0.896916326666667

00:10:00.450 --> 00:10:02.328 so these are the original P

NOTE Confidence: 0.896916326666667

00:10:02.328 --> 00:10:03.580 values that we get.

NOTE Confidence: 0.896916326666667

00:10:03.580 --> 00:10:06.520 For example from let's say at Test

NOTE Confidence: 0.896916326666667

00:10:06.520 --> 00:10:09.076 or something that is more specific

NOTE Confidence: 0.896916326666667

00:10:09.076 --> 00:10:11.272 for next generation sequencing data.

NOTE Confidence: 0.896916326666667

00:10:11.272 --> 00:10:12.168 So it's

NOTE Confidence: 0.948966472

00:10:12.170 --> 00:10:14.930 any test of differential expression.

NOTE Confidence: 0.948966472

00:10:14.930 --> 00:10:17.762 So this is the original P value and

NOTE Confidence: 0.948966472

00:10:17.762 --> 00:10:20.351 I ranked the genes in increasing

NOTE Confidence: 0.948966472

00:10:20.351 --> 00:10:23.051 values so that these genes level

NOTE Confidence: 0.948966472

00:10:23.138 --> 00:10:25.228 4 is the most significant.

NOTE Confidence: 0.948966472

00:10:25.230 --> 00:10:27.636 And so on. This, the laster,

NOTE Confidence: 0.948966472

00:10:27.640 --> 00:10:29.170 has no significance at all,

NOTE Confidence: 0.948966472

00:10:29.170 --> 00:10:31.516 because the P value is 1.

NOTE Confidence: 0.948966472

00:10:31.520 --> 00:10:33.356 So this is the Bonferroni formula.

NOTE Confidence: 0.948966472

00:10:33.360 --> 00:10:35.880 So since we have a 53.

NOTE Confidence: 0.948966472

00:10:35.880 --> 00:10:39.264 Jeans we run 53 tests and so every

NOTE Confidence: 0.948966472

00:10:39.264 --> 00:10:42.442 number has to be multiplied by 53.

NOTE Confidence: 0.948966472

00:10:42.442 --> 00:10:44.210 So after this formula,

NOTE Confidence: 0.948966472

00:10:44.210 --> 00:10:46.989 that's the result that we get them.

NOTE Confidence: 0.948966472

00:10:46.990 --> 00:10:50.410 And so if before the correction all these

NOTE Confidence: 0.948966472

00:10:50.410 --> 00:10:53.498 jeans were below the 0.05 threshold.

NOTE Confidence: 0.948966472

00:10:53.498 --> 00:10:55.100 After the correction,

NOTE Confidence: 0.948966472

00:10:55.100 --> 00:10:58.628 only the first jeans is selected.

NOTE Confidence: 0.948966472

00:10:58.630 --> 00:11:01.170 Because it's the only one that is below 0.05.

NOTE Confidence: 0.941178177857143

00:11:03.450 --> 00:11:06.434 Uhm below here you see the same analysis

NOTE Confidence: 0.941178177857143

00:11:06.434 --> 00:11:09.848 but with the benjamini Hochberg correction.

NOTE Confidence: 0.941178177857143

00:11:09.850 --> 00:11:13.170 So the original P values are the same.

NOTE Confidence: 0.941178177857143

00:11:13.170 --> 00:11:15.645 It's a stepwise procedure because

NOTE Confidence: 0.941178177857143

00:11:15.645 --> 00:11:18.872 you start from the from the bottom.

NOTE Confidence: 0.941178177857143

00:11:18.872 --> 00:11:21.770 And so you multiply that these are,

NOTE Confidence: 0.941178177857143

00:11:21.770 --> 00:11:23.863 that is value one for the number

NOTE Confidence: 0.941178177857143

00:11:23.863 --> 00:11:26.158 of tests are divided by the rank.

NOTE Confidence: 0.941178177857143

00:11:26.160 --> 00:11:27.966 So this is multiplied by one,

NOTE Confidence: 0.941178177857143

00:11:27.970 --> 00:11:29.470 so it stays the same.

NOTE Confidence: 0.941178177857143

00:11:29.470 --> 00:11:30.590 And that's why it's one.

NOTE Confidence: 0.941178177857143

00:11:30.590 --> 00:11:33.890 Also here, then you multiply this

NOTE Confidence: 0.941178177857143

00:11:33.890 --> 00:11:36.318 value for the number of tests,

NOTE Confidence: 0.941178177857143

00:11:36.318 --> 00:11:37.646 53 for the rank.

NOTE Confidence: 0.941178177857143

00:11:37.650 --> 00:11:39.645 So this is slightly more than one,

NOTE Confidence: 0.941178177857143

00:11:39.650 --> 00:11:43.969 so you're a little bit increasing the.

NOTE Confidence: 0.941178177857143

00:11:43.970 --> 00:11:47.458 The value here from one this is the

NOTE Confidence: 0.941178177857143

00:11:47.458 --> 00:11:49.727 result and what I didn't tell you before

NOTE Confidence: 0.941178177857143

00:11:49.727 --> 00:11:51.938 is that it's not simply this formula,

NOTE Confidence: 0.941178177857143

00:11:51.940 --> 00:11:54.138 but once you get your these results,

NOTE Confidence: 0.941178177857143

00:11:54.140 --> 00:11:57.241 you have to check whether this is

NOTE Confidence: 0.941178177857143

00:11:57.241 --> 00:12:00.134 higher than the value of the corrected

NOTE Confidence: 0.941178177857143

00:12:00.134 --> 00:12:02.738 P value of the genes that precedes.

NOTE Confidence: 0.941178177857143

00:12:02.740 --> 00:12:04.060 In this case, it's lower,

NOTE Confidence: 0.941178177857143

00:12:04.060 --> 00:12:05.440 so we keep this.

NOTE Confidence: 0.941178177857143

00:12:05.440 --> 00:12:08.332 But if this was higher than we

NOTE Confidence: 0.941178177857143

00:12:08.332 --> 00:12:11.110 would have kept these this value

NOTE Confidence: 0.941178177857143

00:12:11.203 --> 00:12:13.936 and you see this here so here.

NOTE Confidence: 0.941178177857143

00:12:13.936 --> 00:12:17.240 I proceed and so we multiply this value

NOTE Confidence: 0.941178177857143

00:12:17.331 --> 00:12:20.964 here for 53 / 3 and this is the result.

NOTE Confidence: 0.941178177857143

00:12:20.970 --> 00:12:22.918 Now these results here,

NOTE Confidence: 0.941178177857143

00:12:22.918 --> 00:12:25.898 the multiplication would give you 0.04.

NOTE Confidence: 0.941178177857143

00:12:25.898 --> 00:12:30.202 This is higher than what you obtain here.

NOTE Confidence: 0.941178177857143

00:12:30.210 --> 00:12:33.399 0.035 and so that's why instead of instead,

NOTE Confidence: 0.941178177857143

00:12:33.399 --> 00:12:34.077 instead of,

NOTE Confidence: 0.941178177857143

00:12:34.077 --> 00:12:36.111 the result will not be the

NOTE Confidence: 0.941178177857143

00:12:36.111 --> 00:12:37.989 exact result of this formula,

NOTE Confidence: 0.941178177857143

00:12:37.990 --> 00:12:40.684 but these jeans will take the

NOTE Confidence: 0.941178177857143

00:12:40.684 --> 00:12:43.919 value of the gene that precedes.

NOTE Confidence: 0.941178177857143

00:12:43.920 --> 00:12:46.118 And so that's where the final P

NOTE Confidence: 0.941178177857143

00:12:46.118 --> 00:12:48.420 value adjusted P value will be 0.035,

NOTE Confidence: 0.941178177857143

00:12:48.420 --> 00:12:51.940 and that's why when you use this method.

NOTE Confidence: 0.941178177857143

00:12:51.940 --> 00:12:53.116 You know, if you look now at.

NOTE Confidence: 0.941178177857143

00:12:53.120 --> 00:12:53.754 I mean,

NOTE Confidence: 0.941178177857143

00:12:53.754 --> 00:12:56.290 I noticed you can have a lot of

NOTE Confidence: 0.941178177857143

00:12:56.290 --> 00:12:58.840 adjusted values that are the same.

NOTE Confidence: 0.969053

00:13:00.150 --> 00:13:02.250 Ask a question comma. This is great.

NOTE Confidence: 0.969053

00:13:02.250 --> 00:13:05.000 I'm so appreciating your clarifying

NOTE Confidence: 0.969053

00:13:05.000 --> 00:13:07.058 everything just on the bottom

NOTE Confidence: 0.969053

00:13:07.058 --> 00:13:09.867 where it says BH adjusted P value

NOTE Confidence: 0.969053

00:13:09.867 --> 00:13:12.187 and then in parentheses FDR.

NOTE Confidence: 0.969053

00:13:12.190 --> 00:13:15.696 Q value clarify all those

NOTE Confidence: 0.969053

00:13:15.696 --> 00:13:17.268 different things that FDR.

NOTE Confidence: 0.969053

00:13:17.270 --> 00:13:19.110 I know it's false discovery rate in Q,

NOTE Confidence: 0.969053

00:13:19.110 --> 00:13:22.410 but is this point 027?

NOTE Confidence: 0.969053

00:13:22.410 --> 00:13:24.687 Could it be referred to as the P value,

NOTE Confidence: 0.969053

00:13:24.690 --> 00:13:26.510 the FDR and the Q?

NOTE Confidence: 0.88450238036

00:13:27.160 --> 00:13:29.008 So, uh, yeah, this is a

NOTE Confidence: 0.88450238036

00:13:29.008 --> 00:13:30.240 little bit of terminology,

NOTE Confidence: 0.88450238036

00:13:30.240 --> 00:13:32.550 so this notation here tells you how

NOTE Confidence: 0.88450238036

00:13:32.550 --> 00:13:35.021 the P value has been adjusted so

NOTE Confidence: 0.88450238036

00:13:35.021 --> 00:13:37.754 it has been adjusted with the with

NOTE Confidence: 0.88450238036

00:13:37.754 --> 00:13:39.946 the benjamini Hochberg correction.

NOTE Confidence: 0.88450238036

00:13:39.950 --> 00:13:42.080 So since this method belongs to

NOTE Confidence: 0.88450238036

00:13:42.080 --> 00:13:44.828 a family of methods that are so

NOTE Confidence: 0.88450238036

00:13:44.828 --> 00:13:46.918 called false discovery rate methods,

NOTE Confidence: 0.88450238036

00:13:46.920 --> 00:13:49.671 so the the you can interpret the

NOTE Confidence: 0.88450238036

00:13:49.671 --> 00:13:52.860 result also as a false discovery rate.

NOTE Confidence: 0.88450238036

00:13:52.860 --> 00:13:57.164 So that's why sometimes you will not find.

NOTE Confidence: 0.88450238036

00:13:57.170 --> 00:13:58.442 BH adjusted P value,

NOTE Confidence: 0.88450238036

00:13:58.442 --> 00:14:00.032 but false discovery rate and

NOTE Confidence: 0.88450238036

00:14:00.032 --> 00:14:01.829 also any adjusted P value is.

NOTE Confidence: 0.88450238036

00:14:01.830 --> 00:14:05.320 I think it can be also called the Q value.

NOTE Confidence: 0.88450238036

00:14:05.320 --> 00:14:06.100 Got it, thank you.

NOTE Confidence: 0.88450238036

00:14:06.100 --> 00:14:08.782 So that means that FDR can be used also

NOTE Confidence: 0.88450238036

00:14:08.782 --> 00:14:10.875 for with other corrections methods

NOTE Confidence: 0.88450238036

00:14:10.875 --> 00:14:13.455 that belong to the same family,

NOTE Confidence: 0.88450238036

00:14:13.460 --> 00:14:15.940 but they are not benjamini.

NOTE Confidence: 0.88450238036

00:14:15.940 --> 00:14:19.428 Yes, Benjamin Hochberg corrected.

NOTE Confidence: 0.88450238036

00:14:19.430 --> 00:14:21.836 So usually in publication you use

NOTE Confidence: 0.88450238036

00:14:21.836 --> 00:14:24.609 FDR for example and then you specify

NOTE Confidence: 0.88450238036

00:14:24.609 --> 00:14:27.150 in the methods that you use the

NOTE Confidence: 0.88450238036

00:14:27.232 --> 00:14:29.577 Benjamini occupied in order to.

NOTE Confidence: 0.88450238036

00:14:29.580 --> 00:14:32.280 Calculate the FDR.

NOTE Confidence: 0.88450238036

00:14:32.280 --> 00:14:34.188 But sometimes it's left ambiguous most

NOTE Confidence: 0.88450238036

00:14:34.188 --> 00:14:37.027 of the time it will be the Benjamin IAL,

NOTE Confidence: 0.88450238036

00:14:37.030 --> 00:14:40.909 but in any case. Perfect thank you.

NOTE Confidence: 0.88450238036

00:14:40.910 --> 00:14:41.885 And, uh, yeah.

NOTE Confidence: 0.88450238036

00:14:41.885 --> 00:14:44.500 And finally the first gene as you see,

NOTE Confidence: 0.88450238036

00:14:44.500 --> 00:14:46.243 only the first gene has the same

NOTE Confidence: 0.88450238036

00:14:46.243 --> 00:14:47.590 correction as the Bonferroni,

NOTE Confidence: 0.88450238036

00:14:47.590 --> 00:14:50.222 because this is the only case where

NOTE Confidence: 0.88450238036

00:14:50.222 --> 00:14:51.786 these multiplication since the

NOTE Confidence: 0.88450238036

00:14:51.786 --> 00:14:53.666 rank is one corresponds exactly

NOTE Confidence: 0.88450238036

00:14:53.666 --> 00:14:55.170 to the Bonferroni formula.

NOTE Confidence: 0.88450238036

00:14:55.170 --> 00:14:55.882 Unless so,

NOTE Confidence: 0.88450238036

00:14:55.882 --> 00:14:58.374 unless the value of these is higher

NOTE Confidence: 0.88450238036

00:14:58.374 --> 00:15:00.898 than the value of the second jeans.

NOTE Confidence: 0.88450238036

00:15:00.900 --> 00:15:01.604 Because remember,

NOTE Confidence: 0.88450238036

00:15:01.604 --> 00:15:04.068 in this case you take the you

NOTE Confidence: 0.88450238036

00:15:04.068 --> 00:15:06.388 take the minor of the two values,

NOTE Confidence: 0.88450238036

00:15:06.390 --> 00:15:08.575 the formula or the corrected

NOTE Confidence: 0.88450238036

00:15:08.575 --> 00:15:11.600 values of the gene that precedes.

NOTE Confidence: 0.88450238036

00:15:11.600 --> 00:15:13.504 And as you see, in this case,

NOTE Confidence: 0.88450238036

00:15:13.510 --> 00:15:15.465 after you apply the Benjamini

NOTE Confidence: 0.88450238036

00:15:15.465 --> 00:15:17.354 awkward after the correction,

NOTE Confidence: 0.88450238036

00:15:17.354 --> 00:15:20.840 four of the genes are selected because

NOTE Confidence: 0.88450238036

00:15:20.930 --> 00:15:23.630 the adjusted P value is below 5.

NOTE Confidence: 0.88450238036

00:15:23.630 --> 00:15:25.660 So this is an example showing also

NOTE Confidence: 0.88450238036

00:15:25.660 --> 00:15:28.262 that the Bonferroni is much less much

NOTE Confidence: 0.88450238036

00:15:28.262 --> 00:15:30.790 more stringent than the Benjamini awkward.

NOTE Confidence: 0.88450238036

00:15:30.790 --> 00:15:32.870 Because here you you accept

NOTE Confidence: 0.88450238036

00:15:32.870 --> 00:15:35.002 5% of a false positive.

NOTE Confidence: 0.88450238036

00:15:35.002 --> 00:15:37.540 Here you accept the 5% probably

NOTE Confidence: 0.88450238036

00:15:37.540 --> 00:15:39.490 to have one false positive,

NOTE Confidence: 0.88450238036

00:15:39.490 --> 00:15:42.158 and that's the difference

NOTE Confidence: 0.88450238036

00:15:42.158 --> 00:15:44.159 in the interpretation.

NOTE Confidence: 0.88450238036

00:15:44.160 --> 00:15:44.410 No,

NOTE Confidence: 0.9489760044

00:15:44.420 --> 00:15:46.244 it is not the one that I mentioned

NOTE Confidence: 0.9489760044

00:15:46.244 --> 00:15:48.488 on the nomenclature is truly a big

NOTE Confidence: 0.9489760044

00:15:48.488 --> 00:15:50.213 issue in the scientific literature

NOTE Confidence: 0.9489760044

00:15:50.277 --> 00:15:51.977 because different people use different

NOTE Confidence: 0.9489760044

00:15:51.977 --> 00:15:53.963 ways to refer to these things.

NOTE Confidence: 0.9489760044

00:15:53.963 --> 00:15:55.661 For example, in some papers you

NOTE Confidence: 0.9489760044

00:15:55.661 --> 00:15:57.489 will see the original P value,

NOTE Confidence: 0.9489760044

00:15:57.490 --> 00:15:59.020 which is what Thomas listed

NOTE Confidence: 0.9489760044

00:15:59.020 --> 00:16:00.244 on the third column.

NOTE Confidence: 0.9489760044

00:16:00.250 --> 00:16:01.390 Here as P value.

NOTE Confidence: 0.9489760044

00:16:01.390 --> 00:16:02.815 Some people refer this to

NOTE Confidence: 0.9489760044

00:16:02.815 --> 00:16:04.440 this as a nominal P value,

NOTE Confidence: 0.9489760044

00:16:04.440 --> 00:16:06.988 and some people just refer directly as

NOTE Confidence: 0.9489760044

00:16:06.988 --> 00:16:09.910 P value and on the adjusted P value.

NOTE Confidence: 0.9489760044

00:16:09.910 --> 00:16:11.590 Some people refer to as FDR.

NOTE Confidence: 0.9489760044

00:16:11.590 --> 00:16:14.174 Some people refer to as a Q value.

NOTE Confidence: 0.9489760044

00:16:14.180 --> 00:16:16.394 Some people refer to as like

NOTE Confidence: 0.9489760044

00:16:16.394 --> 00:16:18.539 a tomasetta pH adjust P value.

NOTE Confidence: 0.9489760044

00:16:18.540 --> 00:16:20.376 Some people even will just tell

NOTE Confidence: 0.9489760044

00:16:20.376 --> 00:16:22.729 you that it's FDR adjusted P value.

NOTE Confidence: 0.9489760044

00:16:22.730 --> 00:16:24.995 So there are many different

NOTE Confidence: 0.9489760044

00:16:24.995 --> 00:16:26.807 normal creatures for basically

NOTE Confidence: 0.9489760044

00:16:26.807 --> 00:16:29.238 the same things and different.

NOTE Confidence: 0.9489760044

00:16:29.240 --> 00:16:30.248 Authors use different ways

NOTE Confidence: 0.9489760044

00:16:30.248 --> 00:16:31.508 to refer to those things.

NOTE Confidence: 0.98594059

00:16:32.330 --> 00:16:33.680 Yeah, yeah there is no.

NOTE Confidence: 0.98594059

00:16:33.680 --> 00:16:34.910 Yeah I think yeah there

NOTE Confidence: 0.98594059

00:16:34.910 --> 00:16:36.140 is a lot of redundancy.

NOTE Confidence: 0.98594059

00:16:36.140 --> 00:16:39.390 Let's say now in terminology.

NOTE Confidence: 0.98594059

00:16:39.390 --> 00:16:41.534 And no specific rules.

NOTE Confidence: 0.98594059

00:16:41.534 --> 00:16:44.750 That depends on the reviewers mail.

NOTE Confidence: 0.98594059

00:16:44.750 --> 00:16:48.528 I see OK, so this was like this,

NOTE Confidence: 0.98594059

00:16:48.530 --> 00:16:48.998 uh, an introduction.