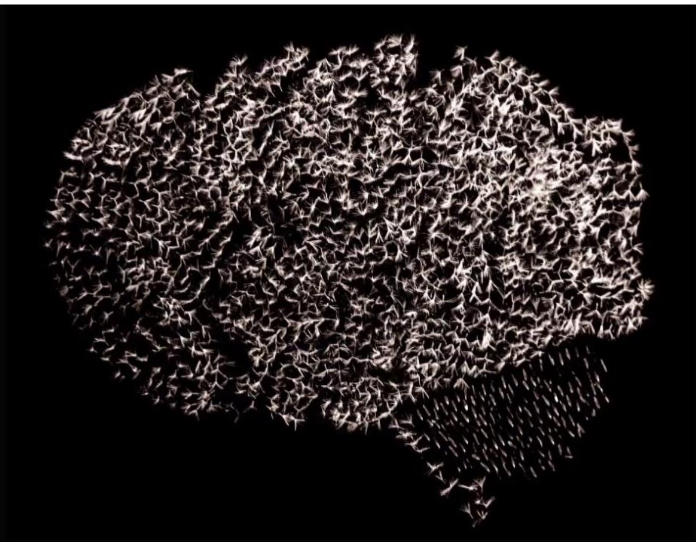
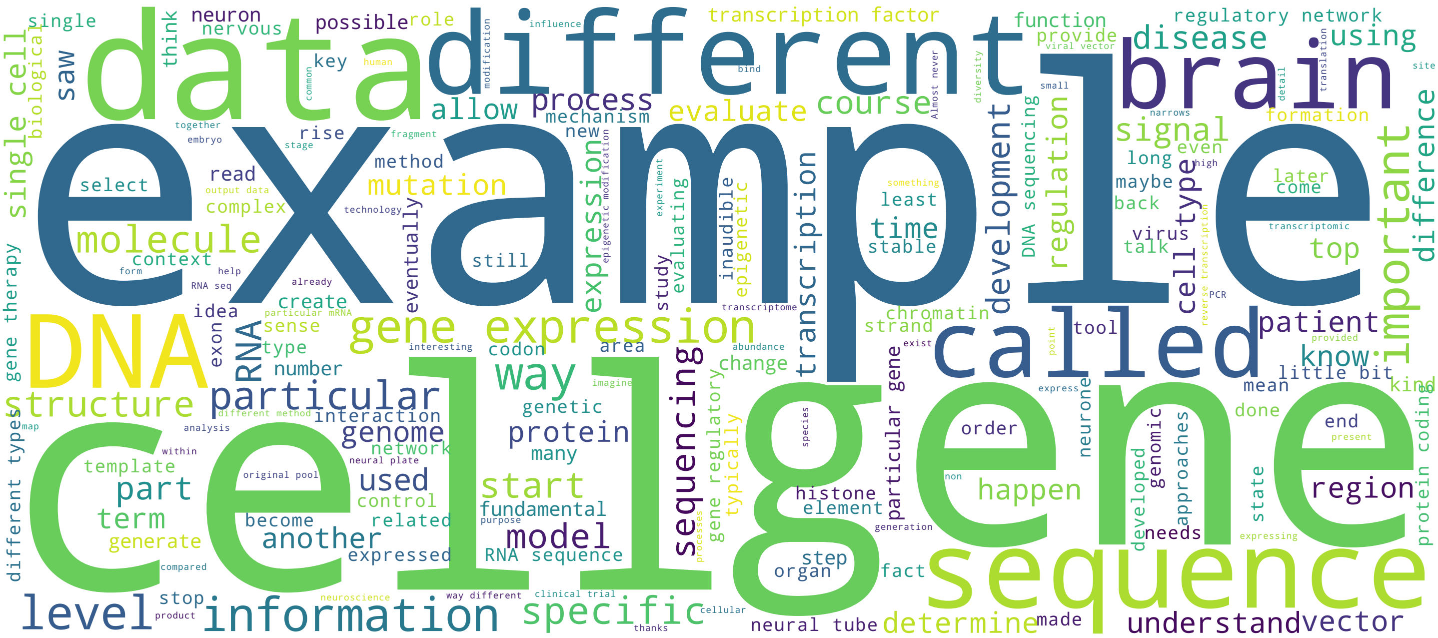


NEUROSCIENCE RECONSTRUCTED

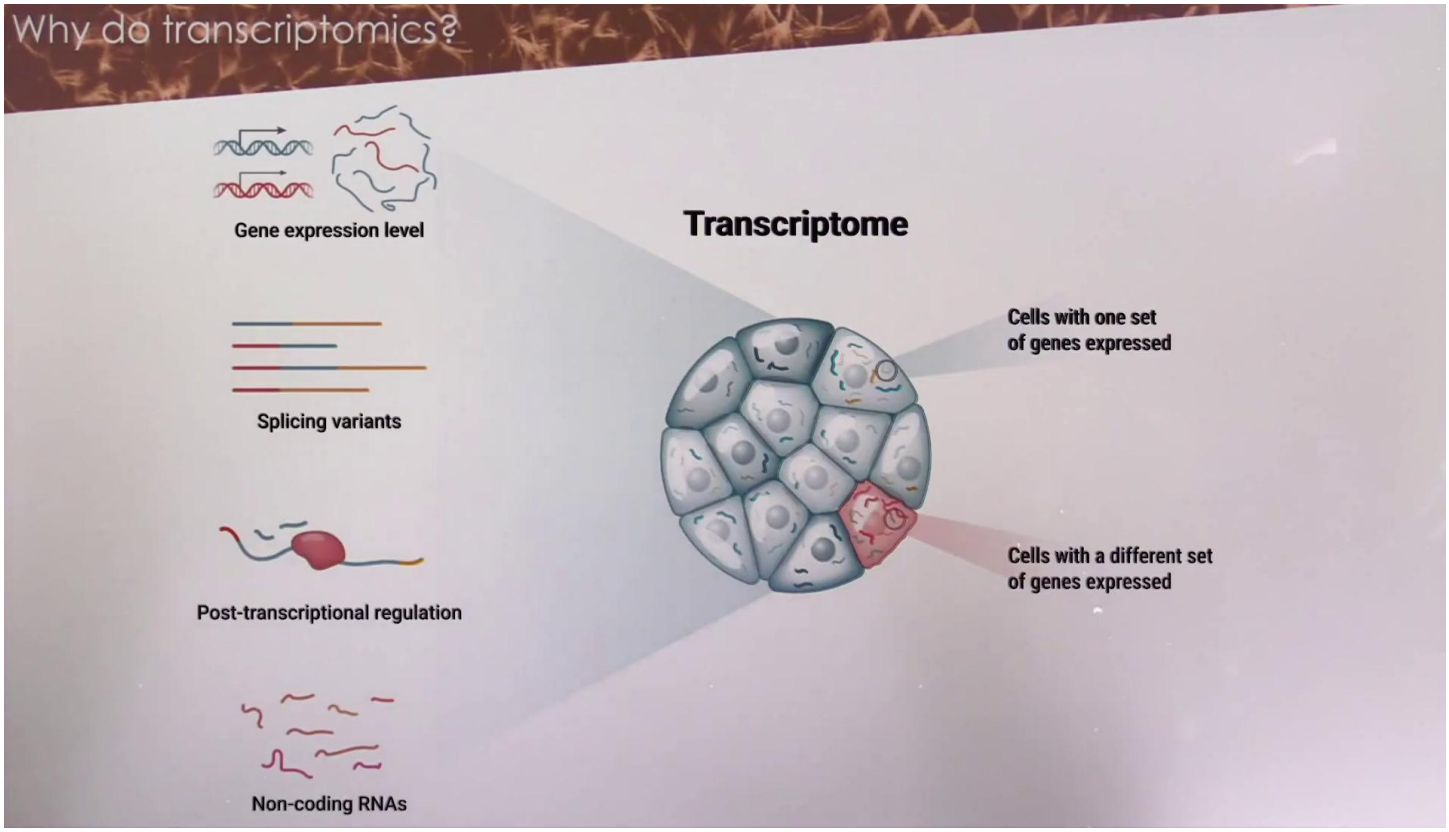
Bulk transcriptome anaylsis

Presented by Gioele La Manno



Search MOOC

Video

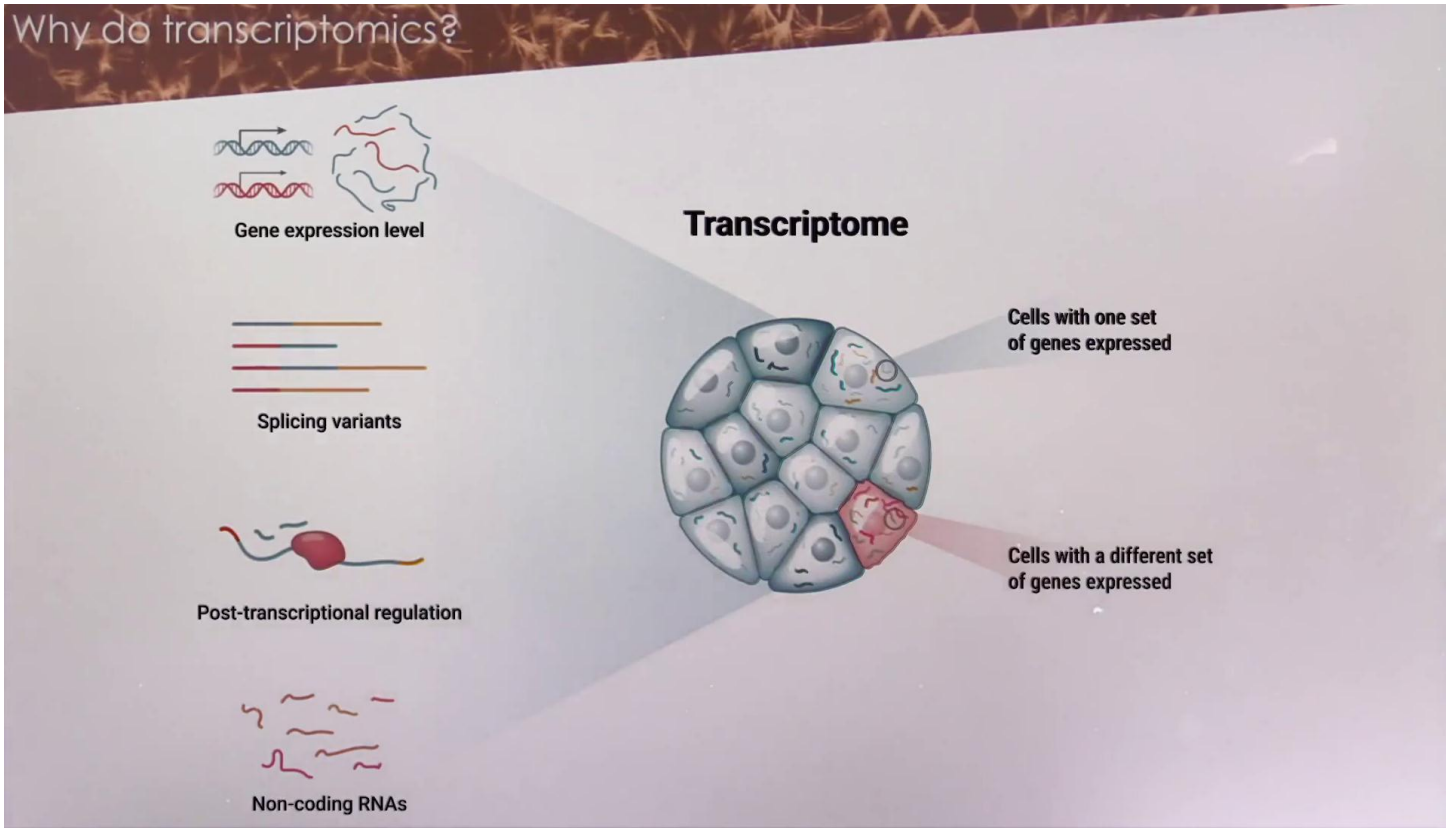


Another very important and fundamental aspect sequencing is used for is to evaluate and measure gene expression. It's what is called doing transcriptomics, evaluating the transcriptome. Transcriptomics, it has a double nature. On one end, we want to discover which kind of transcript exists. It's related to evaluating, for example, the splicing variants that exist of a particular gene, in which way different axon cut and pasted together to generate different types of among the same gene. It has a nature of determining, again, also the sequence, what are the possible isoform, enumerating them. But it's also another less usual aim as compared to what we saw before, at least, where we will be sequencing the RNA as you see converted to cDNA for the purpose of evaluating the concentration of a particular mRNA species into the original pool of mole. Not anymore to determine the sequence per se, but to evaluate the abundance of that particular mRNA molecules that has a particular sequence associated to it in the original pool or sample. This is important because this allows, for example, to tell us the difference between different tissues, a case of a control.

Notes

Summary

0m 05s

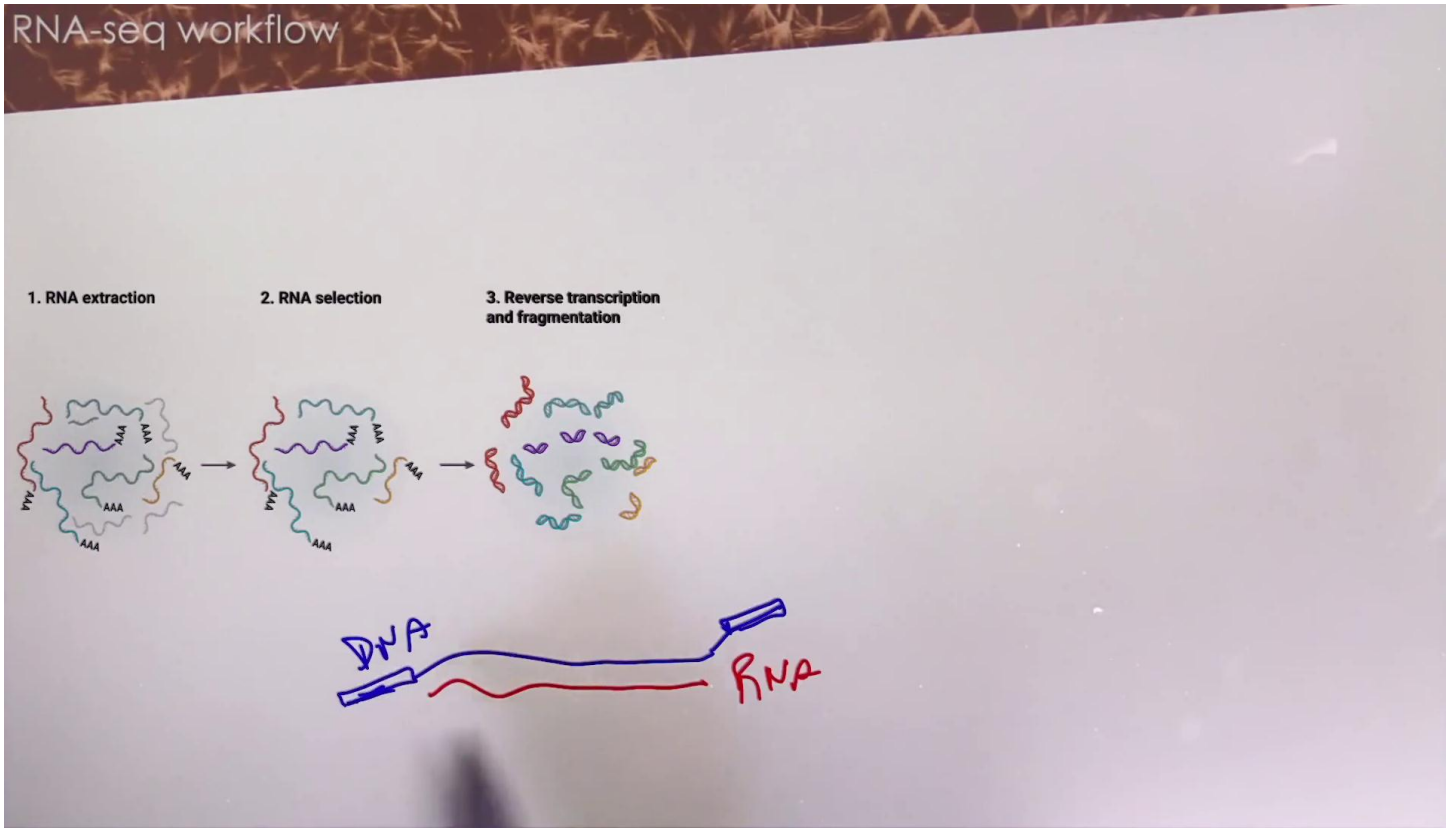


This is a typical setup for what is called bulk transcriptomics. We sequence an entire tissue, entire specimen, and compare it, for example, between healthy condition and the disease case. Or thanks to the single cell revolution, we can nowadays even sequence the transcriptome of a single cell. Therefore, we can evaluate the difference between the gene expression associated with different cell types. For example, in the context of neurobiology, we can ask in which way different types of neurons are regulating their gene expression. What are the genes that make effectively these different kind of neurons different? Of course, thanks to transcriptomics, we can also... As you see, for example, for genomics device, different method that indirectly get information about, for example, transcriptional regulation or get information on other species of RNA that are not necessarily mRNA protein coding, but for example, are noncoding RNAs.

Notes

Summary

1m 28s

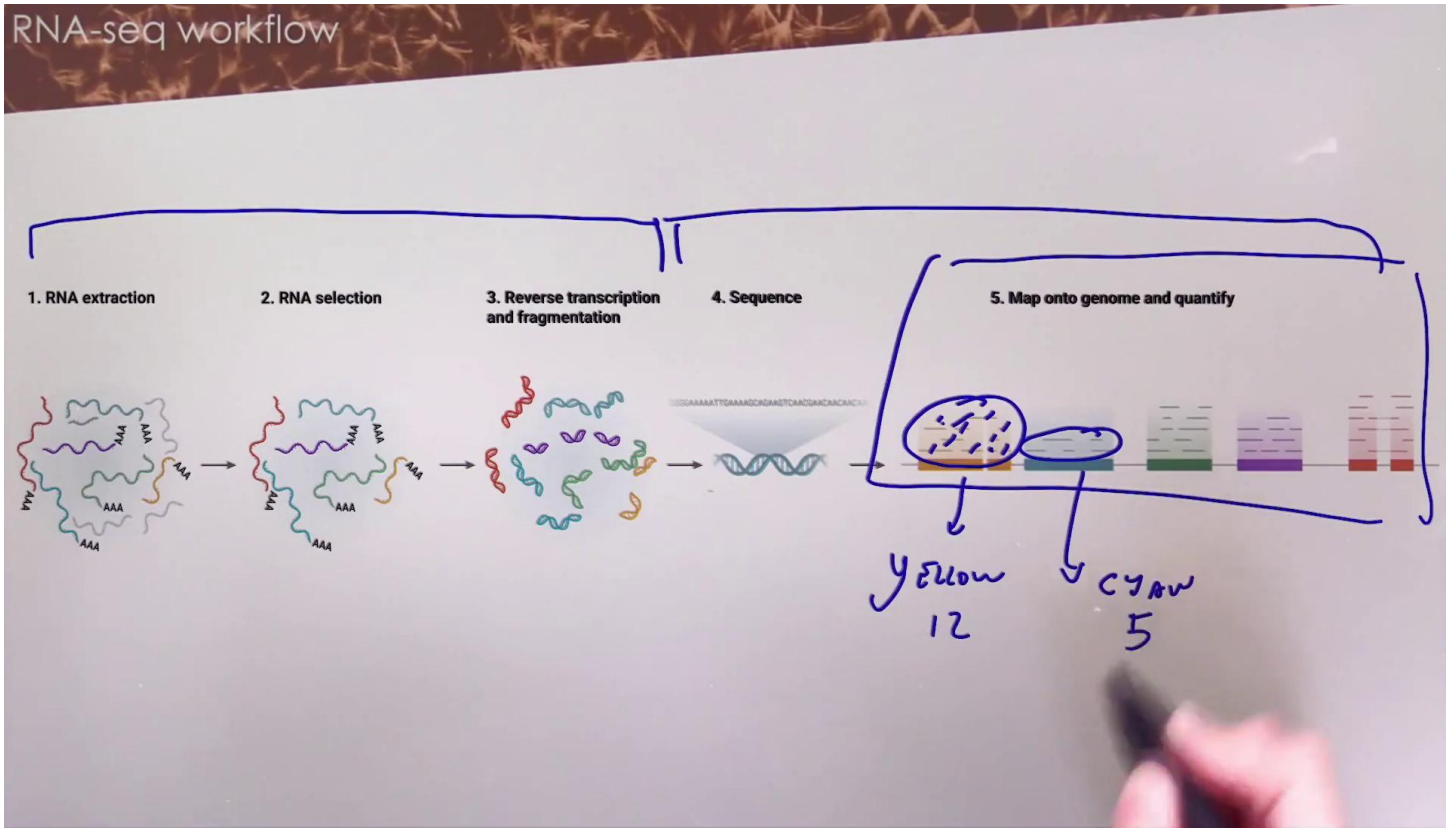


What is the core idea of a transcriptomic RNA sequence workflow? It starts from the structure, from the structure of RNA. This will include not only mRNA, but also other kinds of RNAs usually. This is like we have our RNA library. It comes, for example, from the lysis of a single cell or a set of cells. In different methods, we have a way to select, an active or maybe more passive way to select for typically the most informative one in terms of at least the regulation that is mRNA. MRNA have in common the fact that are all polyethylene layer. We can use this polyethylenelation either by targeting those molecules with a specific oligonucleotide primer that is polyethylene, this is a poly T primer, or we can use some other negative selection approaches that get rid of other ribosomal RNA, other RNAs present in the suspension. Then we have a fundamental step, reverse transcription. It will create on top of template of RNA, it will synthesize a reverse complement strain of DNA and create a DNA-RNA teraduplex. After reverse transcription and usually linking some adapter, we are back in the similar situation of DNA sequencing. Now we have a stretch of a fragment of DNA, and that's its extremities.

Notes

Summary

2m 37s

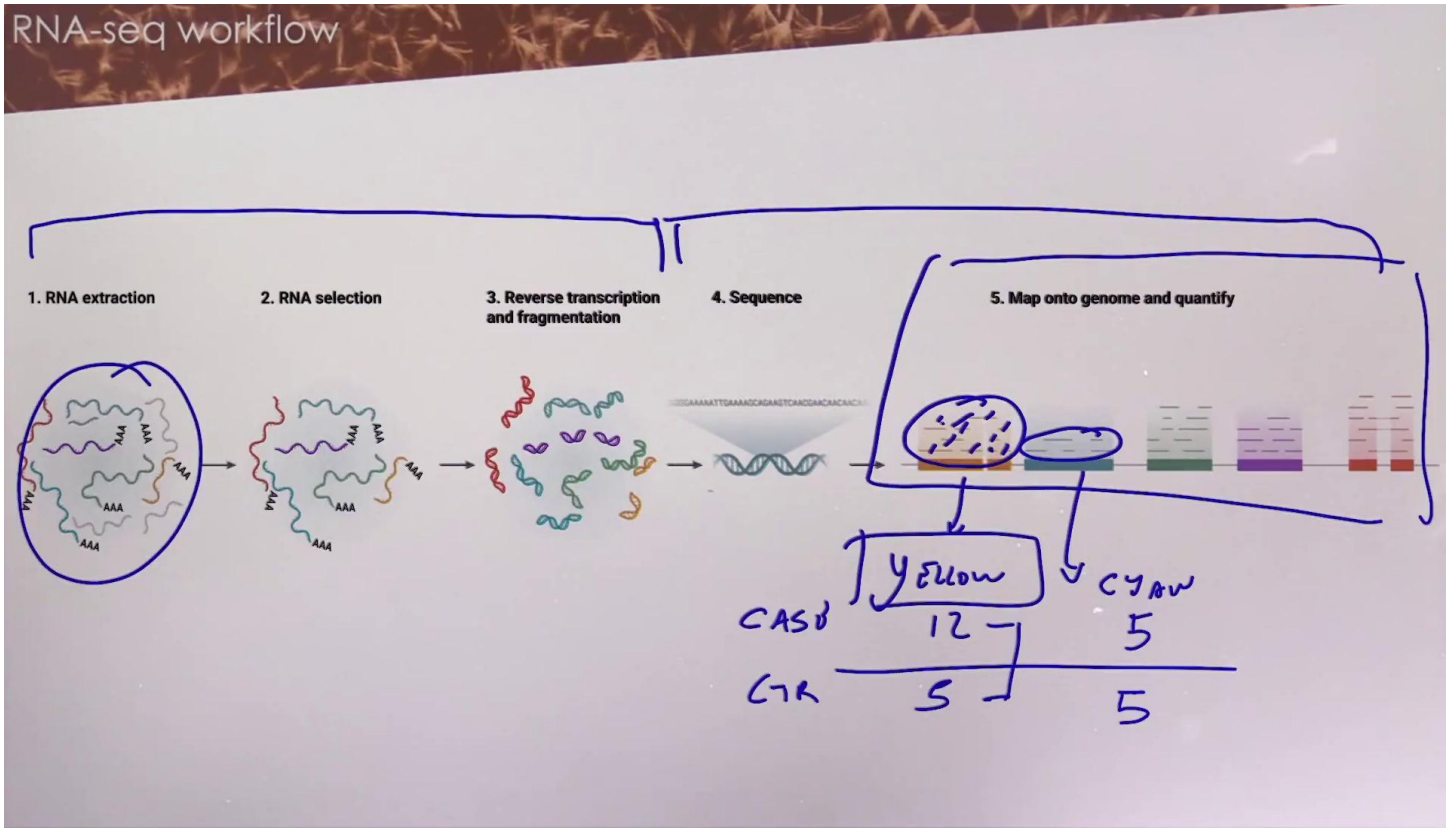


We have sequences that we know. Then we can just proceed with typical approaches for DNA sequencing. From the mentally transcriptomics and RNA seq, exploit a different initial phase, but eventually it narrows down after the resubscription and PCR of the resubscription product, it narrows down to the same sequencing method that we've been seeing before for DNA. Almost never, there are some few exceptions with nanopore sequencing, but almost never, RNA is directly sequenced, but it's always converted to DNA that is more stable, and where all the technology has been developed for DNA sequence can be directly exploited. Typically, the analysis part of RNA sequencing experiment relates to mapping as we saw also for genomics, but here we're going to typically mapping two axons to annotated protein-coding genes. Here I think the key and fundamental conceptual difference from genome sequencing is that in RNA sequence, we're going to care a lot about how many times we saw reads that map to a particular gene. This example, for example, where we have this yellow gene versus this cyan one. It would be important to count how many reads map to this gene, in this case, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 12, while for this other one, we see only five.

Notes

Summary

4m 27s

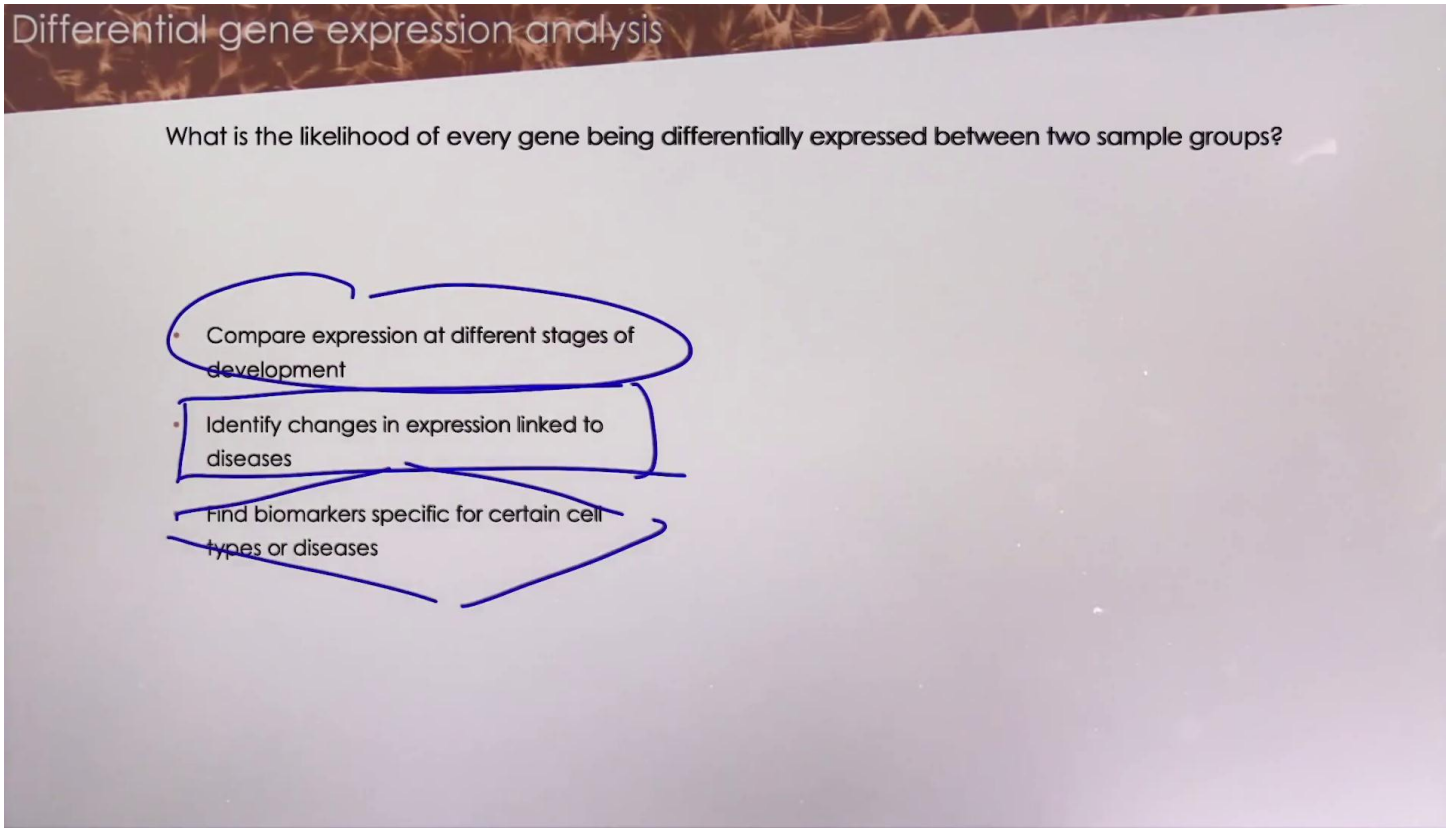


This information is important because it's going to be proportional to the original number of molecules that we provided, that we started with at this RNA structure level. Typically, an RNA seq-based study will evaluate the abundance of this transcript between, say, a case and a control that, for example, will have a different number of molecules. For example, in this case, we could identify a difference between a case and control for the level of expression of a particular gene. This is a little bit why while exploiting mostly the same platform from DNA sequencing, transcriptomics' difference can be different in terms of the way of reasoning about the data and using the data to quantify rather than determine the sequence.

Notes

Summary

6m 21s

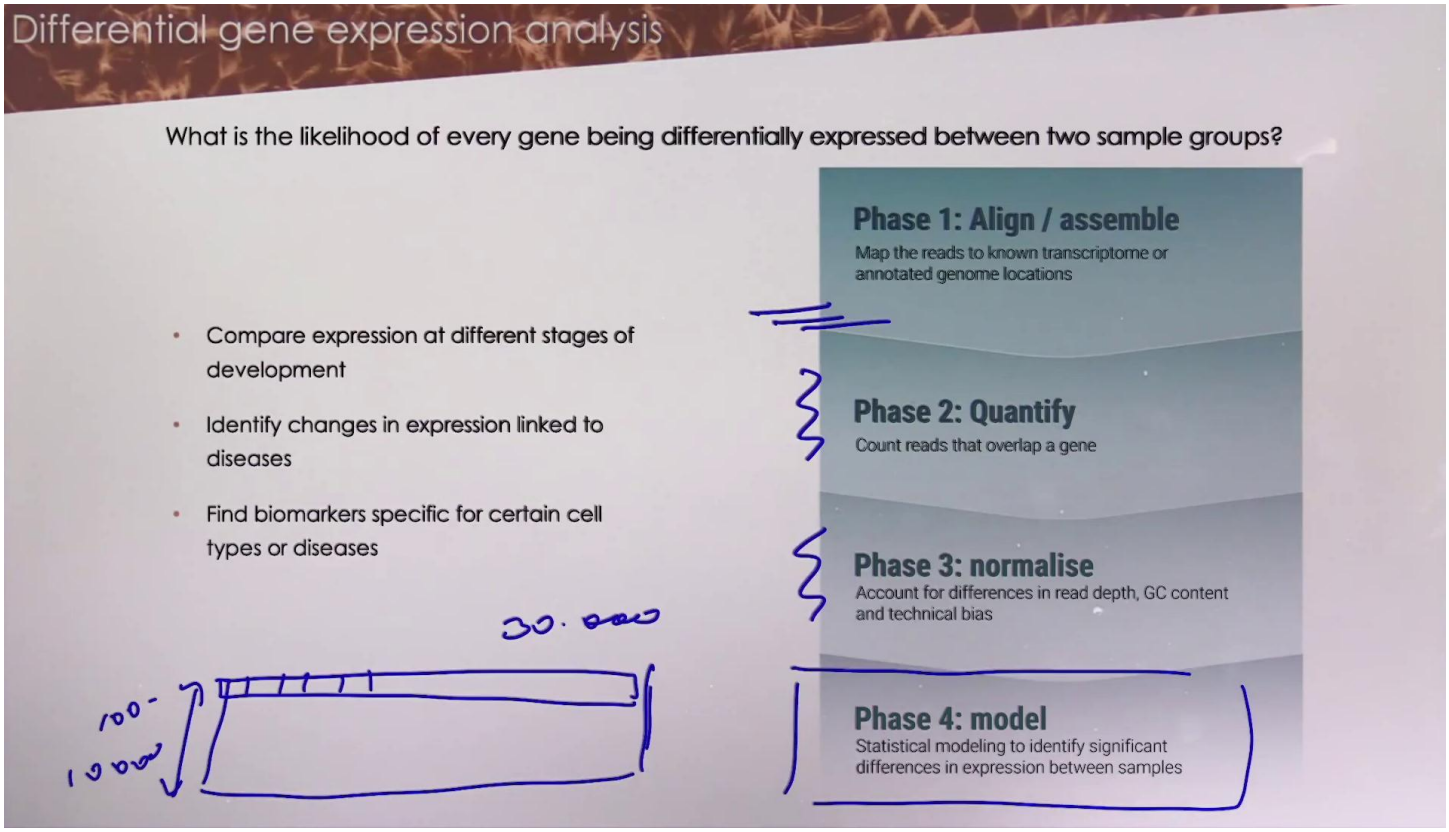


Here we just wanted to list some example building on top of this concept I just made. What one can study with these approaches, one can, for example, compare different expression, gene expression at different stages of development of an organ, for example. As the example I just made before, one can identify changes in gene expression linked to disease, compare basically specimens from healthy controls individuals with patients for a particular disease, or even have approaches that are more targeted with, for example, the identification of biomarkers, locate genes that are expressed only in a particular condition as opposed to many other different settings.

Notes

Summary

7m 20s



The part of the analysis now [inaudible 00:08:16] reached by different other steps. Don't stop in DNA and genomics to this alignment step, but they're more involving, they involve the quantification step. A normalisation step where, for example, the quantification gets adjusted by the depth of sequencing, so to make it relative to the number of reads that have been sequenced by the machine. Then an important part and becoming more and more articulated. Nowadays, the use of tools from statistical inference, from machine learning, and to understand and dive in this output data. It's an output data and it's like a multi-dimensional data because now we're going to have information of the level of expression of each single gene that the gene is expressing. So potentially, we're going to be working with gene expression vectors that have values that have up to 30,000 values potentially. And we will have this for maybe hundreds of samples or even thousands of cells with nowadays single-cell RNA sequence. This generates big matrices of data that are impossible to analyze with maybe more traditional methods and differently data that one cannot really make sense of in a simple Excel sheet that needs to use more computation-involving tools.

Notes

Summary

8m 10s