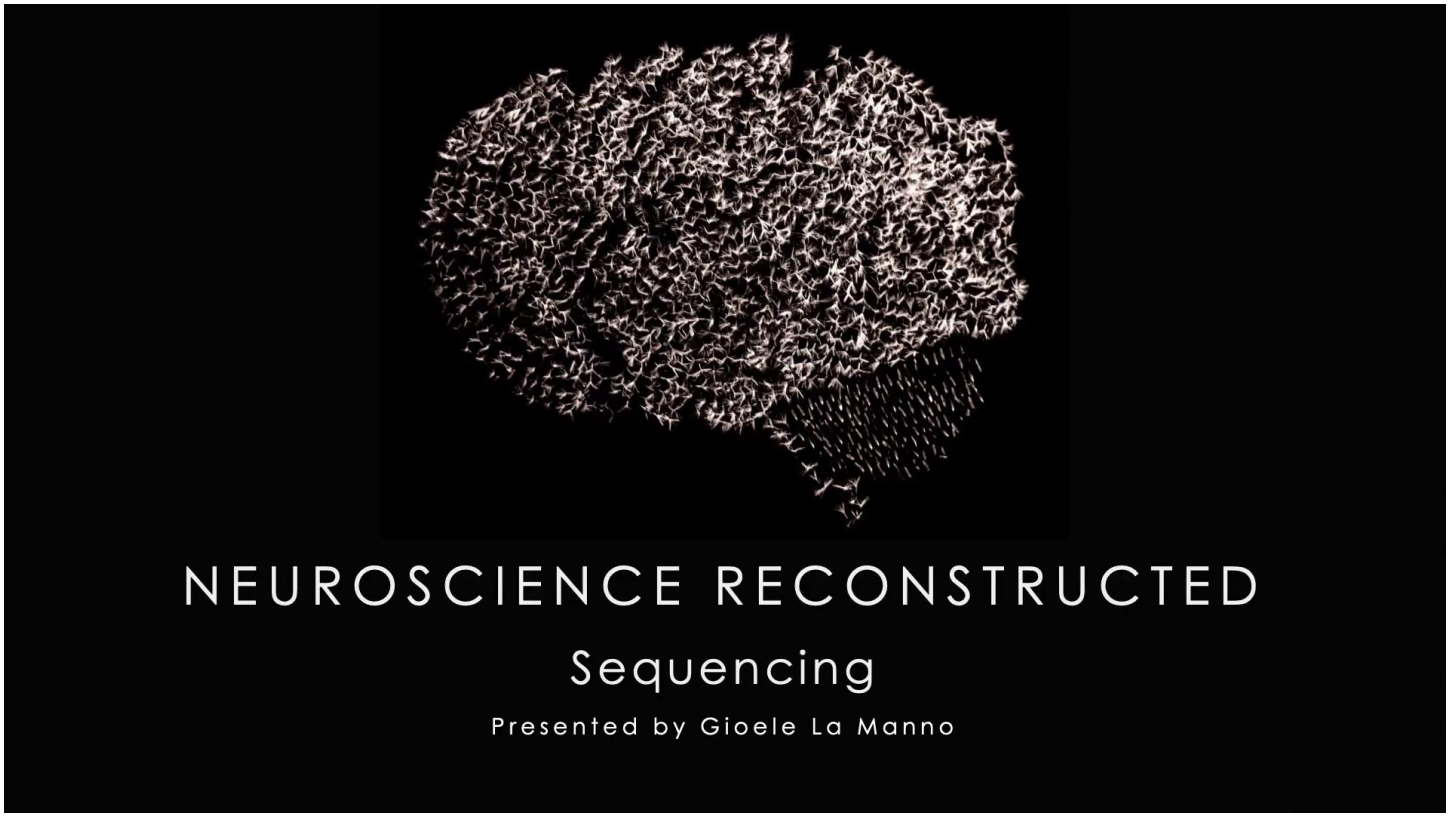
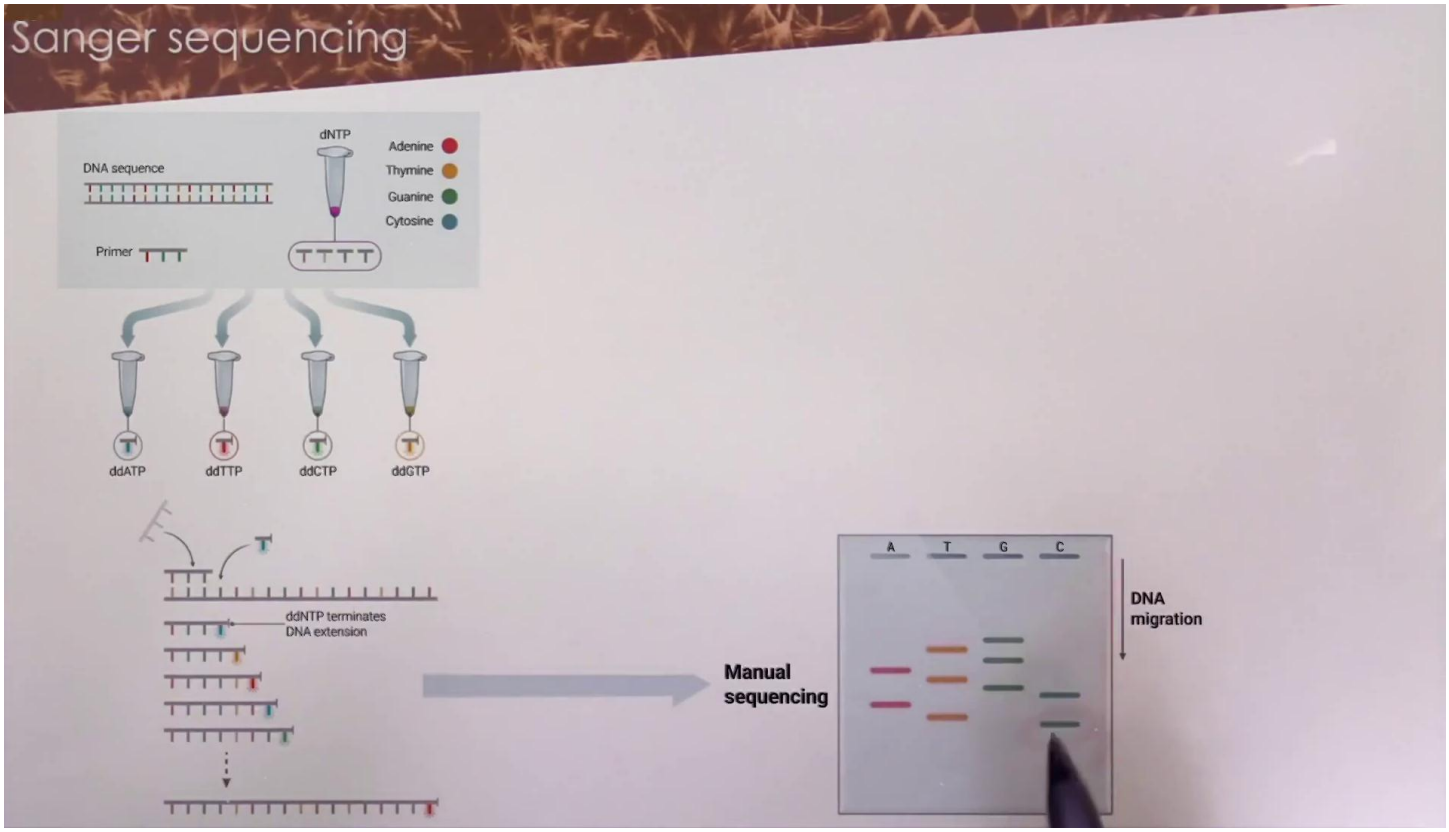


Search MOOC

Video

EPFL

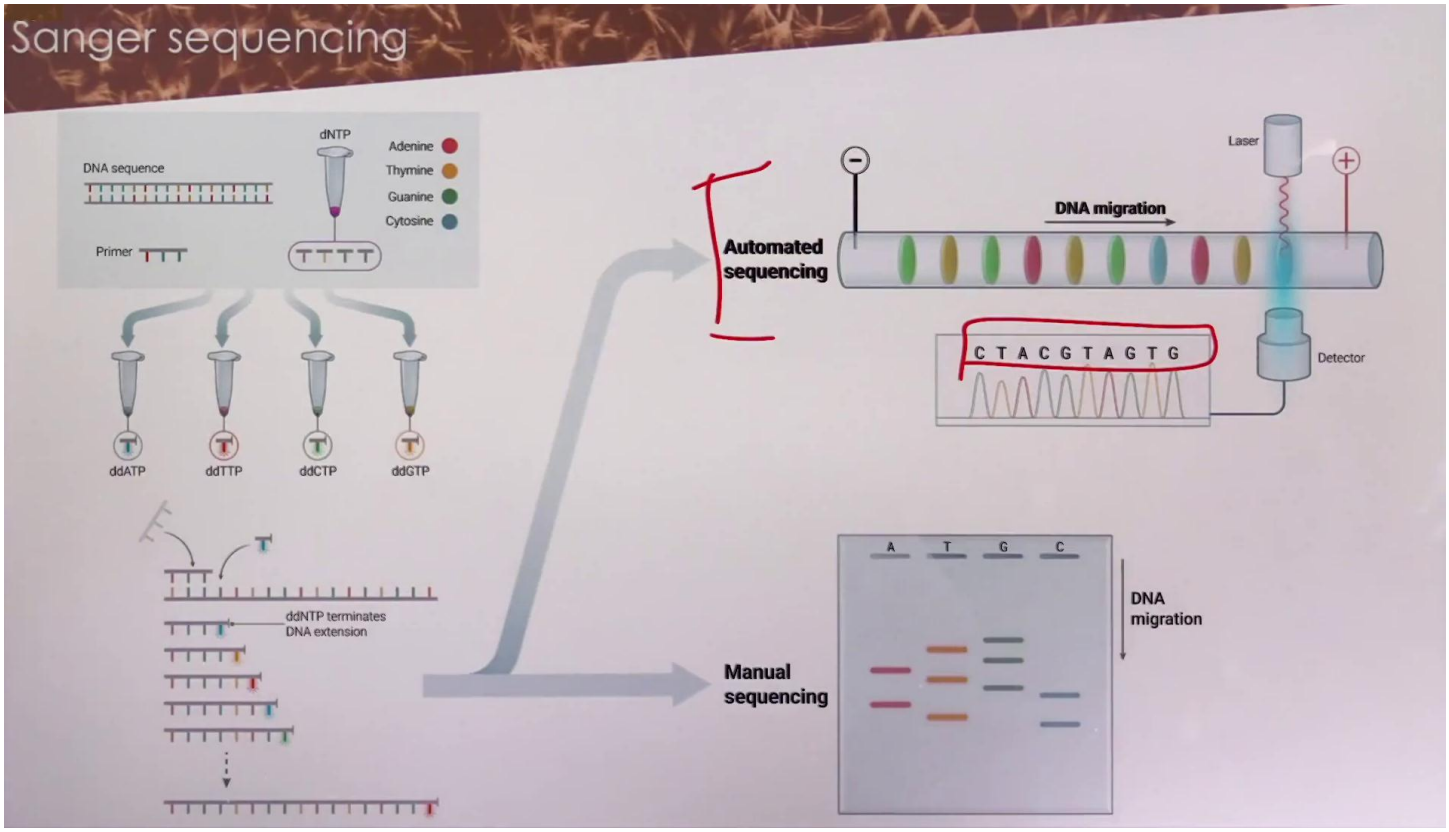


The first sequencing method that became really popular with something that actually fueled the human genome project and still used nowadays for different validation of sequencing validation and experiments is Sanger sequencing or Chain termination-based sequencing. The idea with Sanger sequencing is possible to include into a mix for an elongation step of basically a cycle of PCR, reaction where we have DNA polymerase that is elongating on top of a template. We can add a small amount of specific nucleotides that do not allow elongation anymore. In different cases, a bit randomly, you're going to have a termination of the sequence. Sometimes happens here and it happens randomly in different positions. Then you get a collection out of a step of elongation with random termination of differently sized basically sequences of different lengths. Now you can exploit the property that DNA can be separated on basis of difference in mass and then therefore charge of its molecule and then run electrophoresis. Now, how do we read from this output which sequence was originally the template we have been sequencing? Well, we reason that the smallest fragments will migrate more, so they will be the one on the bottom.

Notes

Summary

0m 05s

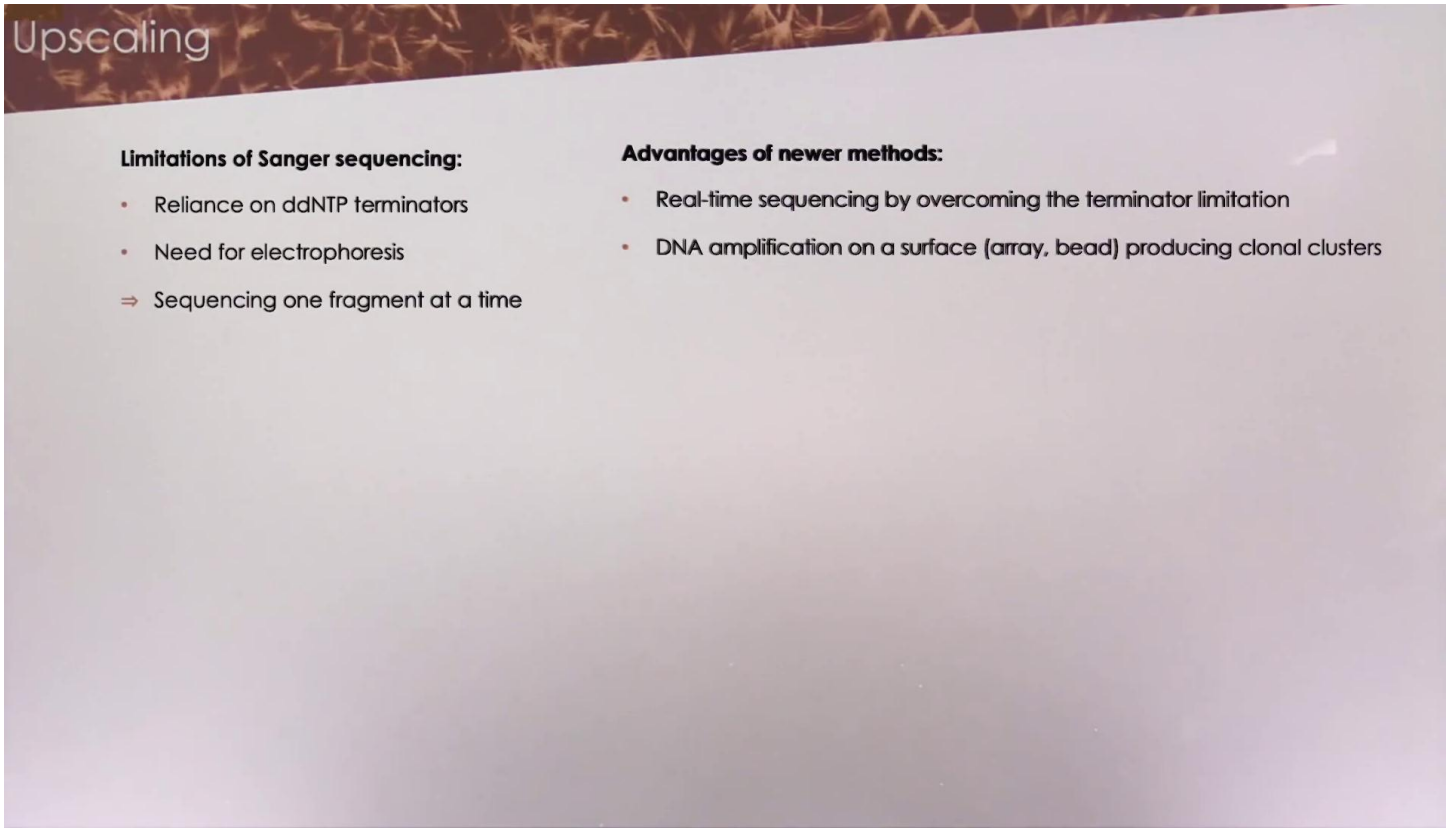


The larger fragments, one on the top, and the ditch color here tell us what is the base that was added at the termination step. We can basically then go and read the sequence. We take this into account. It's CTACGTATGTG. A really big part, not all of the original human genome project, has been seeing basically people doing with a more or less automated fashion similar sequence determination using some more sequencing. This, of course, can be a bit automated using, for example, continuous system where the electrophoresis is coupled with a detector, and therefore we see the signal digitized in a more, a little bit nicer stretch of signals as the one that you see here, where you really get it in a more streamlined way, and this is mostly what is used nowadays. Nonetheless, while this method is definitely based on a cool idea, it's highly accurate, it still has a set of downsides and it's not the most used general purpose sequencing method that is used, nowadays it's still used for different validation steps, but definitely we're sequencing genomes and transcriptome. Nowadays, using other methods, they're much more I tripled.

Notes

Summary

1m 39s

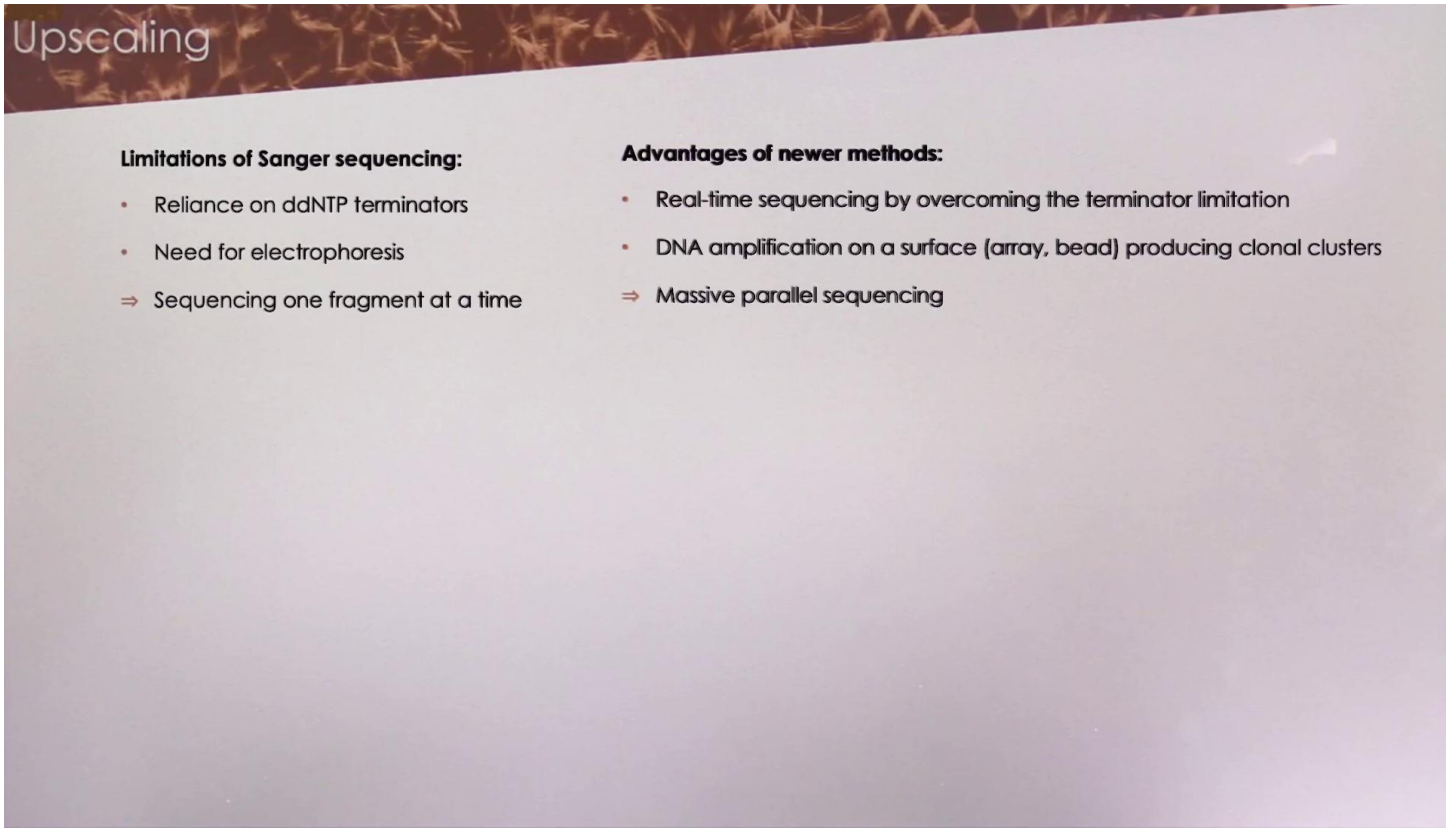


What are the limitations of Songer sequencing? It relies on termination, so random termination of doing the elongation. It needs electrophoresis to read it out. It's definitely limited by how long a stream of gel you can run. Eventually, it also, maybe that's the most important thing that set this apart from newer method is that you can sequence one fragment at a time. The way you go about it is really start from one sequence, you amplify it a lot, then you have enough material to be able to see it on the gel of electrophoresis, and you're actually really sequencing only that original molecules through, thanks to the amplification provided by PCR. However, what you would like to do for more interesting projects is sequence a lot of DNA in the DNA molecule at the same time. You also want, often, a real-time processing. You also want to avoid to use a full entire dedicated instrumentation for each sequence. New methods exploit the use of different DNA amplification of surfaces, for example, on beads or onto a flow cell where some clonal clusters can be produced.

Notes

Summary

3m 29s

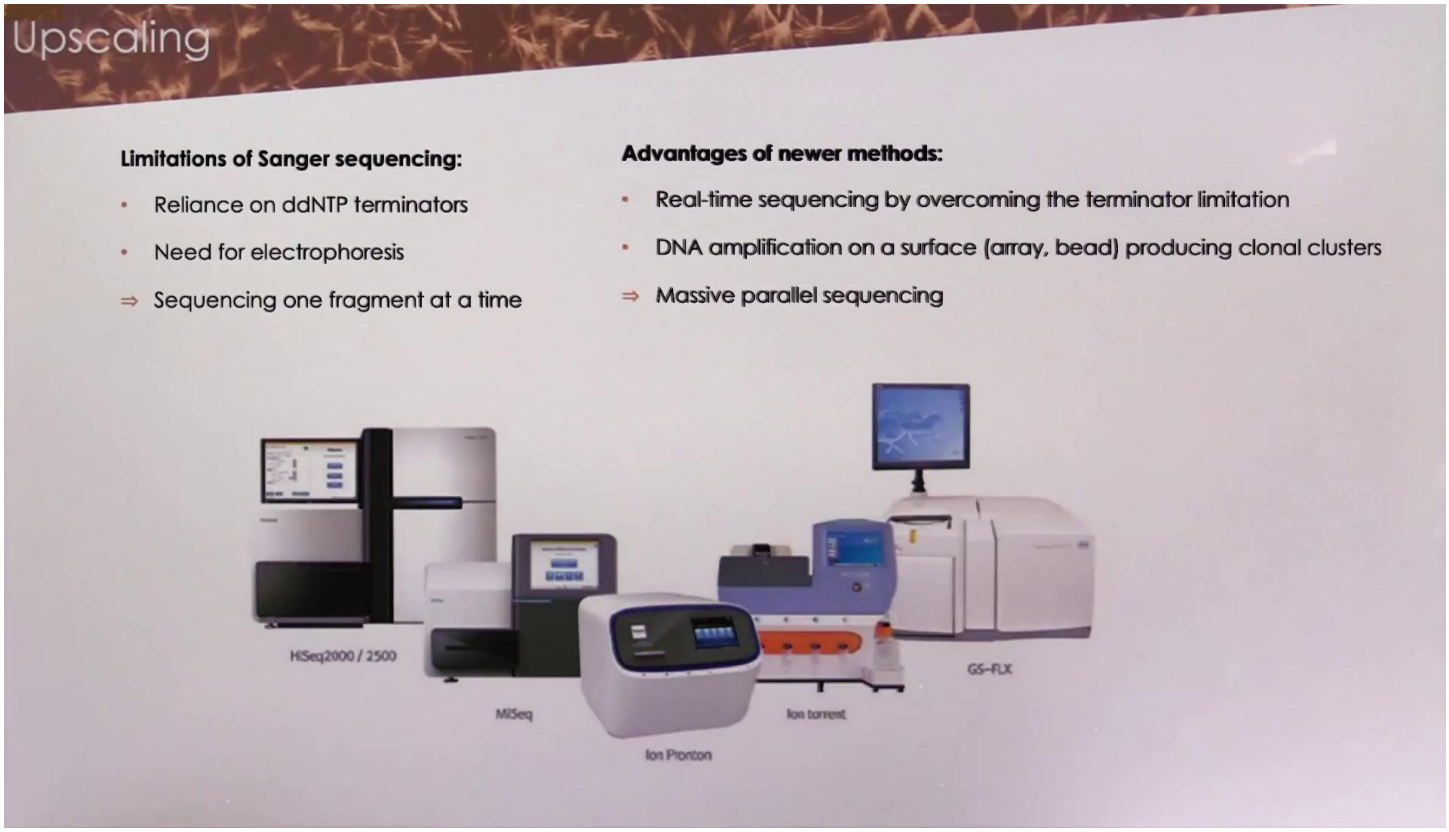


This leads to what is called massively parallel sequencing, the possibility to sequence many stretches of DNA at the same time in a very atypical fashion in the order really of hundreds of millions of reads, each one of them composed of different hundreds of nucleotides in one single experiment. This completely sets us apart from some other sequencing.

Notes

Summary

4m 55s



What is the technology? How is this done? There is not one single answer. The core ideas are definitely producing clonal clusters, addressing each one of them separately through some fluorescent readout. Then there are different technologies have been proposed over the years. These different technologies, I think, can be broadly categorized in two groups. One that is related to short-treat sequencing, where somehow DNA is fragmented in short treats and it's very efficient to get treats of 200 tons base pair sequence in a very [inaudible 00:06:00] fashion. Another family that is long treat sequencing, another family that is long-read sequencing and the other set of technologies that may work in different ways, they're more focused on getting very long stretch, paying off maybe the price of sequencing less number of molecules just for the sake of getting the extra information we're going to gain by sequencing a very long DNA molecule.

Notes

Summary

5m 24s

Upscaling

Limitations of Sanger sequencing:

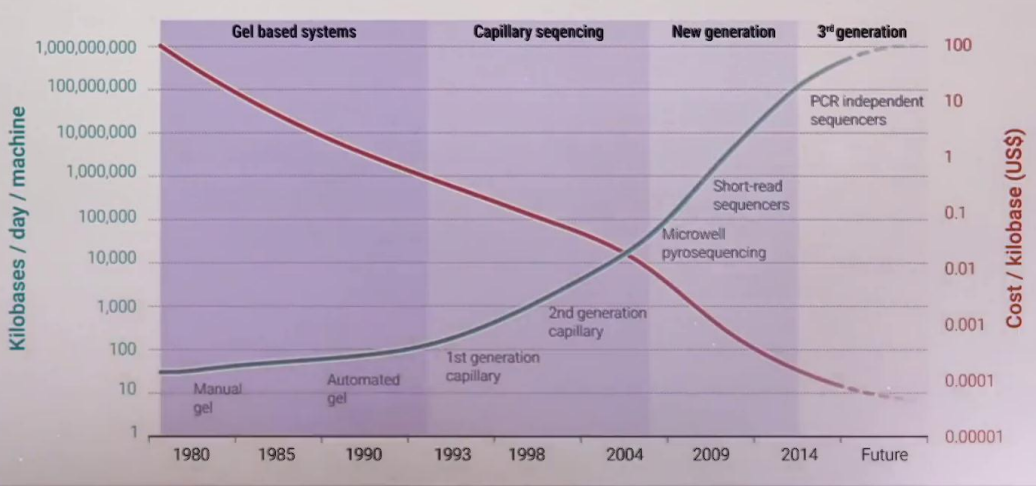
- Reliance on ddNTP terminators
- Need for electrophoresis

⇒ Sequencing one fragment at a time

Advantages of newer methods:

- Real-time sequencing by overcoming the terminator limitation
- DNA amplification on a surface (array, bead) producing clonal clusters

⇒ Massive parallel sequencing

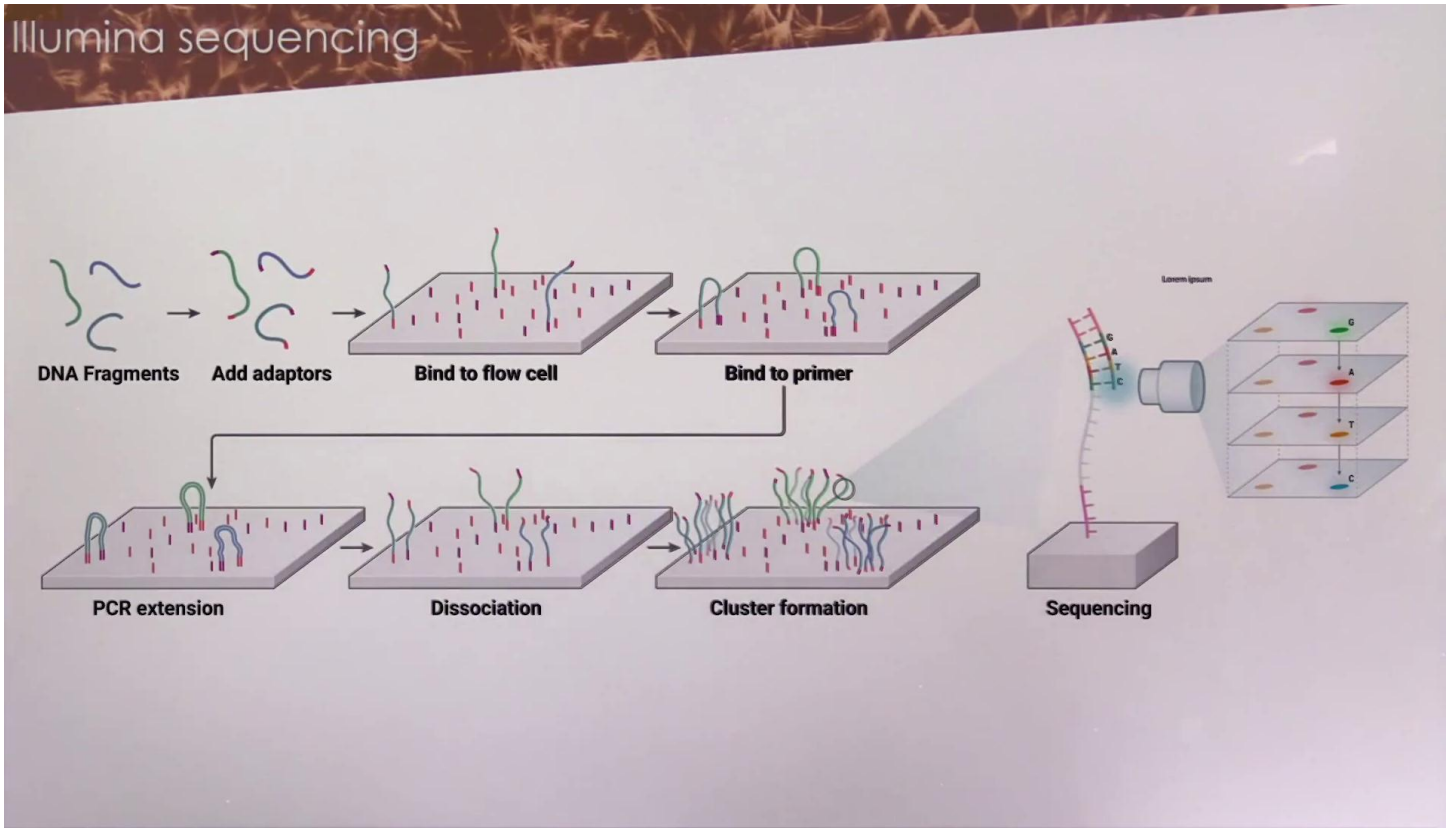


Now in terms of historical trends, record of what has been happening over the years, there's been an important increase number of kilobases of data that can be produced per day. How much we can sequence in a typical experiment? The cost has dropped significantly. It keeps going down, but maybe nowadays it's lower pace than it used to. Nowadays, the sequencing, hundreds of millions of reads is something very approachable for NL.

Notes

Summary

6m 22s

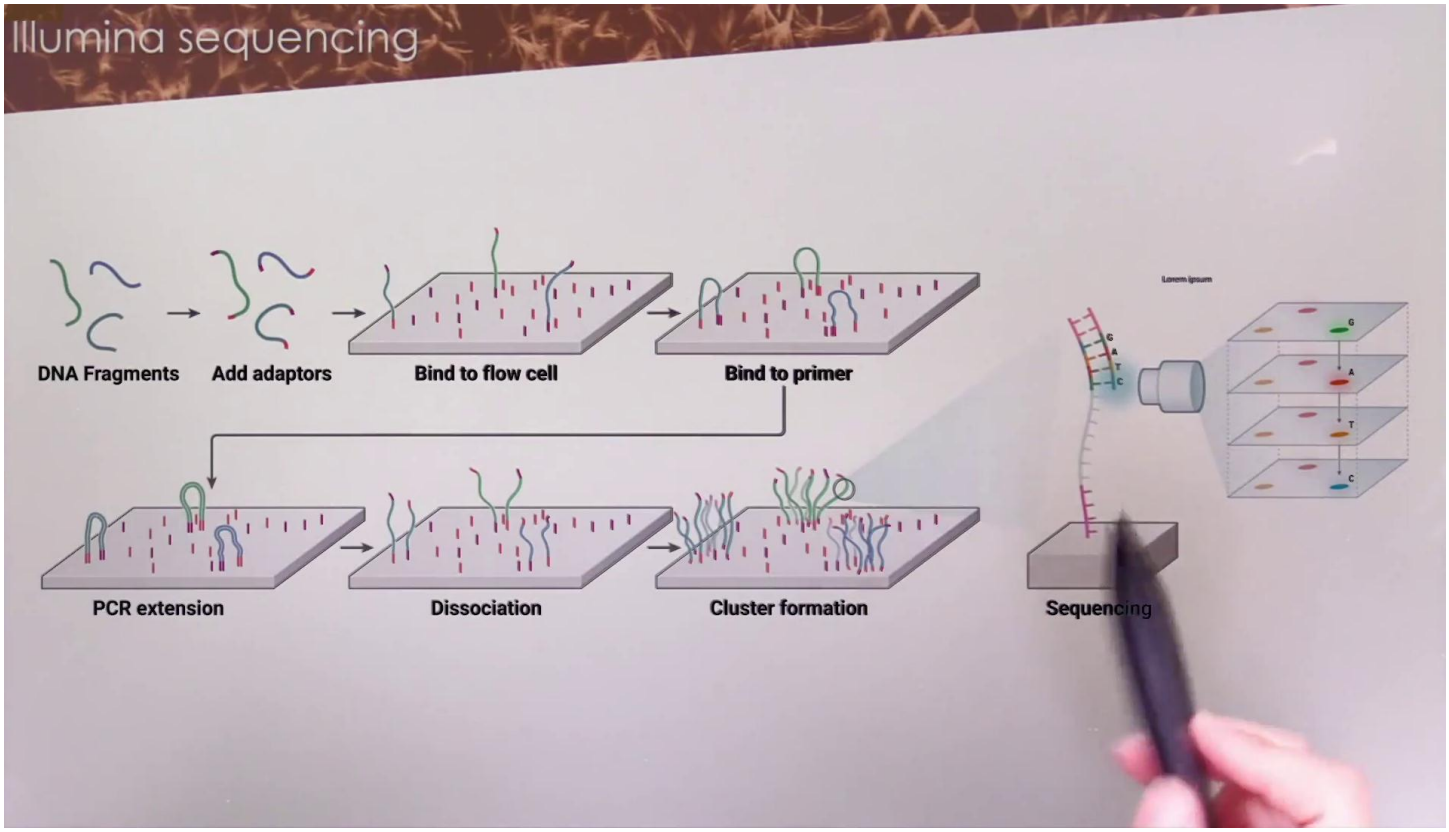


How does this go? Let's go now, let's take it step by step. We start from a pool of DNA fragments. We want to really know the sequence of each one of them in one single measurement process. The first thing to do is to add what we call adaptors. It's to get to a tremendous sequences that we know that we can control, because, of course, we don't know the sequence yet of each of the molecules, but we can get to the tremendous some sequence that we know about. Those help us, for example, to make it possible to hybridise some of those molecules on top of a particularly coated solid support inside the flow cell. That has been coated very densely with a set of the two possible primers that are the same to the one that have been ligated. They're reverse complement to each other. Then what happens is that once you hybridise the molecules here on top, you can decrease the temperature to the point, it allows the kneeling of one of these adaptors to one of the other one, the reverse complement that are arrayed on the solid support. This create this typical bridge and that's why the step of the process is called a Bridge PCR.

Notes

Summary

6m 58s

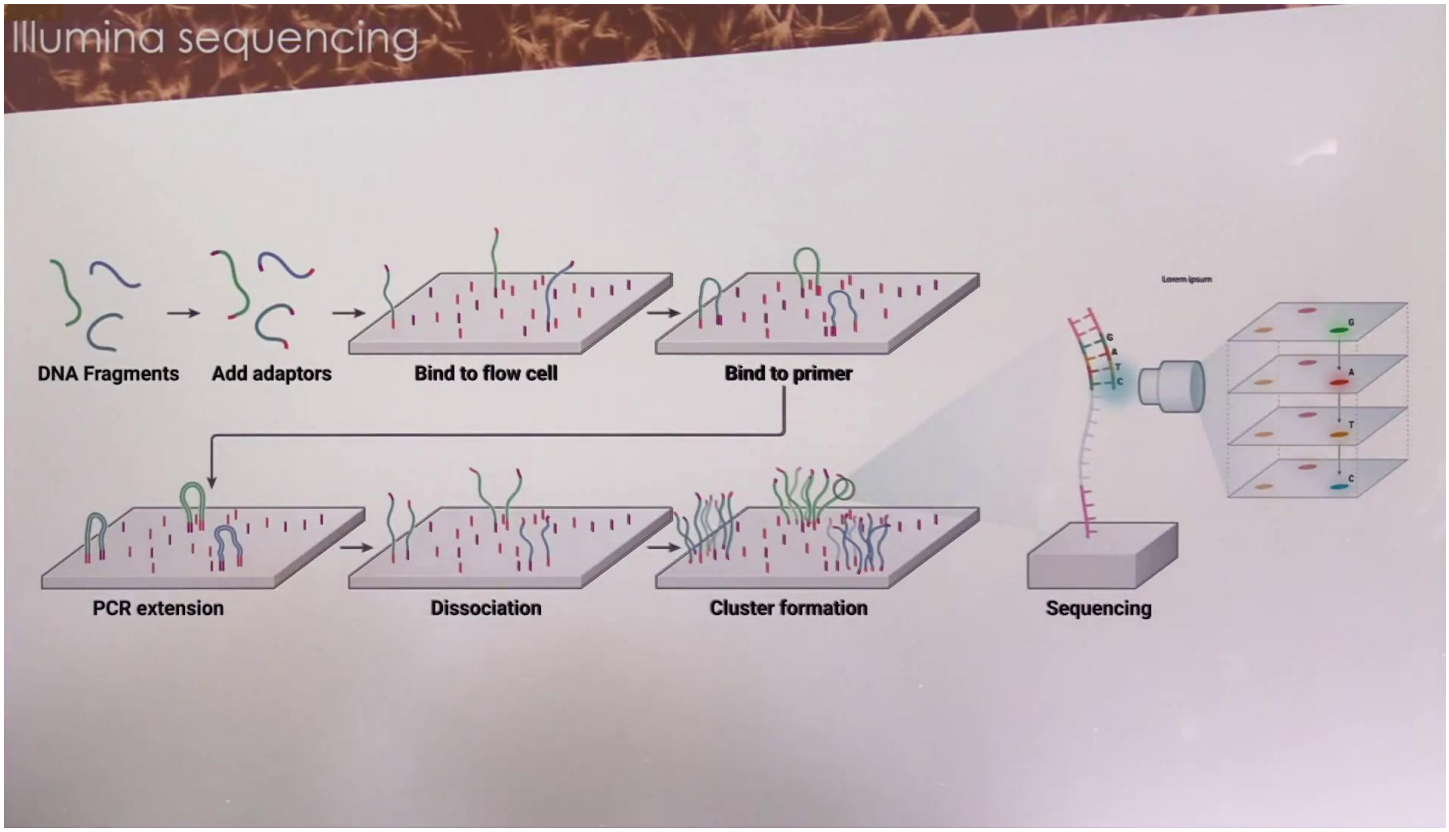


What can happen from the formation of the bridge is that now you can elongate starting from this primary that was in the solid support and you can generate double-stranded molecule, of which both of them actually have an extremity that is covalently bound to the surface of the solid support. Now, what has happened after one cycle of this procedure is that now you get molecules that are the PCR product of an original molecule that you started from that are now clustered together. Now, all the products are one template are close together. You can keep cycling over and over this reaction as you would do for normal PCR. You have now this special effect because of this mechanism is mechanically basically constrained. You're not going to have the molecules product PCR free in solution, but all of them are going to be anchored to the surface and they're going to cluster close to each other, basically similar, all the daughters of original template form a cluster, very densely packed. This is the key because now with a simple elongation termination chemistry, we can generate on top of this spot that has really a micron size. This is, of course, really small dots on top of this surface.

Notes

Summary

8m 29s



We can get enough signal that can be detected by a microscope and can be sequentially read out directly onto the flow cell. This is how Illumina sequence works. It's been perfected to the level where the surface can be very highly packed and really the sequences can get information of hundreds of millions of sequences, each one of them composed of hundreds of base pairs.

Notes

Summary

10m 06s