11/12

1st 2nd GENERATION SEQUENCING / SANGER.

DNA, RNA sequencing means to define the order of nucleotide base of DNA/RNA molecule genomic → composed using an order → of bases

SANGER (1st G.S)
to define the order. 1980 by F. Sanger, Nobel prize using a method in which you stop the reaction at a specific point.
↓
then you have to recognize the point where you stopped!

- Revolutionizing was the introduction of modified deoxynucleotides: one hydrogen substitutes the 3'-OH group of ddNTP

During the sequence, another nucleotide in solution can't make phosphodiester bond
↓
bc there is the lack of 3'-OH group in ddNTP
↓
that was included before in the chain.
=
the reaction is blocked

deoxyribose generating dideoxynucleotides

If you add a free dideoxynucleotide, the chain is "locked" because there is the lack of the bond (3'-OH).
To understand where it's the point of the stop, Sanger at the beginning used a radioactive dNTP as an emission source, then he did 4 ≠ reactions to compare the results of each.

fluorescent method

But with the development of the fluorophore (= absorb light at a specific wavelength)

↑ wavelength = - energy

bc some energy is absorbed by the molecule -

Each fluorophore has a spectral → the result of absorbition wavelenght and emission wavelenght

measurable with a spectrophotometer.

If you can measure the spectrofluorophore, you can even understand where that spectral is produced. If you conjugate ddNTP + fluorophore, you can identify the ddNTP that arbor the fluorophore
≠ fluorophore to conjugate ≠ ddNTPs

4 ≠ fluorophores you can specifically label the 4 ≠ ddNTPs, and that allows you to perform 1 reaction (with 4 ≠ fluo simultaneously)

later, after the development of CE (capill. electroph) to automate the method.

Now, the current Sanger sequencing method is composed by 2 consecutive steps:

1) CHAIN (or dye) TERMINATOR REACTION
2) CE → you read the sequence and define the sequence you produced into the 1st step

RECIPE:

1. You need a template (piece of DNA you want to know the sequence)
   - presented as a PCR product or bacterial clone.

You use a PCR product, which is an amplification of a molecule (= + copies of DNA).
The same base produces a signal at the same time (in danger)
- you amplify it, recognizable from a device that is higher than the background noise of the machine you used.

then you use a DNA primer that is complementary to the DNA you want to sequence.
- So you use an enzyme DNA polymerase to start the polymerisation.
  - then a mix of 4 nucleotides (ACTG) that represent the monomer that allows you to sythetize your DNA
  - and a mix of 4 dideoxynucleotides (a little amount of them) with 4 fluorescent tags are attached to detect.

?Why a little amount of ddNTPs?
The reaction starts with heat denaturation with a termocycler. (During the sequence you use only one primer).
By doing that, you obtain a single strain DNA
then you decrease the temperature that allowes you the primer to anneale the specific region of your target and then you reach the work temperature of your enzyme so that it can start to add nucleotides to the growing strain.

Now, it starts the primer extention.

In the cDNA there is a G.
The DNA polymerase can choice a C, so it choices a T (bc of the G), when ddNPTs is inserted in that position, it can go on further and is blocked at this step.
Starting from another plan, the DNA polymerase can choice the ddGTP and the reaction is stopped at this position.
So at the reaction you have a lot of molecules that are stopped at the first position.
The important is the label of the position

↓

some first, some second, some 3rd

etc

At the end of the DYE TERMINATION SEQUENCING all the positions of the target are represented and labelled with a specific fluorophore.

But why you have to use a limiting amount of ddNTPs compared to the mix of the natural ddNTPs?
If you use only a mix containing ddNTPs, the DNA polymerase can chose only between the ddNTPs so elongation will stop at the 1st base everytime, because the DNA polym. choose for ex the C.
If you use a small amount of ddNTPs, the stop DNA polymerisation has to happen casually. DNA polym. has to choice at them randomly during the extention, only in that way you can represent all the bases of the target with fluorescently labelled with ddNTPs

② You use the capillary of 10 cm, that is filled with gene matrix that allows the separation of your DNA molecules.
That is done just for one base in lenght and an electric field is applied to both ends of the capillary, so the DNA molecules ⊖ are forced to move to the ⊕ pole - The speed of the movement depends on the lenght of the molecules.
Shorter molecules = faster movement.
At the bottom of the capillary, there is a laser source that hit each molecule exciting the fluorophore that labels the fragment.
The fluorophore emits a light that is captured by a detector, which can also recognize the fluorophore that is emitting this signal and so it recognizes the base that is connected to that fluorophore.
A software can give a result and reading the signal, can transform it in the base that gives the signal -
At the end of CE you have a sequence
=
Electropherogram
(represented by a coloured peak. Each peack corresponds to a base.
=
you know the exact order of your target.

SANGER SEQUENCING APPLICATIONS

Genetics
To resequence ⇒ you sequence a DNA or RNA to compare it to a reference into a database

You want to compare the sequence you extracted from a tumor cell with the one present into the database (→ in it you will find the wild type)

You can look for variants → SNPs, point mutations, indel

or in RNA you can look for alternative splicing events or fusions.

Epigenetics ✳ (fra 2 pag)

Check the methylation DNA, used by cells to regolate gene expression.

Some genes have regions that reach the CpG residues.

↓

they are used by cells to regolate gene expression by methylation and to study it, you have to fix the methylation by changing the bases that are involved in methylation.

DNA polymerase reads C and it is able to see if the C is methylated or not, so you have to modify this C and treat the DNA with sodium bisulfite.

↓

Converts the non-methylated cytosine into Uracile, that during the synthesis are read as a A, so the DNA polym. inserts a T in the complementary strain, so that the methylated cytosine is protected from bisulfite conversion.

So after the sodium bisulfite treatment, the cytosine remains cytosine

→ EX

Tumor cell: BRAF → see if the patient has the mutation that cause the substitution in sixthhundred position of BRAF protein. With glutammic acid.

(coding for variants)

The aminoacid substitution is caused by the substitution wild type T base with the mutated A base, so you extract the DNA, then sequence and see if the sequence is the same of the reference or not.
So I perform a DNA resequencing.
You copy your sequence and paste on a tool of the database to compare it.
You have a gap between your reference and your sequence (to be it arbors a base that the software doesn't recognise.
In that position you have 2 bases, you have two peaks in the electrophore-gram, because you are in an eterozygosis condition. So it means that the sequence is the same for both allele, but in this position one allele arbor the wild type base (EX T) and the other allele arbor the mutated base A.

NB You have to inspect the peaks of the bases of electropherogram

This happens because the software converts the raw datas, the instrument reads a fluorescence and then it is converted into bases.
This conversion is called BASE CALLING.

During that process, that step has to pass some filters.
The quality of the BASE CALLING is very low, so the software isn't able to sign a specific base to its position. In this case you can see a grey bar, that represent the quality of the base calling.

higher bar = higher base calling (quality)

So if the base is low in quality, the software assigns the base in that position to the peaks that have the higher quality.

EX if the quality is given by T or C ⇒ you'll have 2 peaks

↓

So that's why you have to inspect your electropherogram, so you'll be able to see that

(*) - (2 pag fa)

ES

If you want to study the effect of a drug on the promoter methylation of a specific gene in a cell line.

↓

You have it. You split it in control cell (my control) that is not treated with drug

↓

in a pool of treated cells (my samples)

After a time point, you check the effect of that drug.
So you extract genomic DNA from both and you treat (for both) a portion of genomic DNA with Sodium bisulfite and the other not

then you SANGER sequence both portions; the sequence gives a result for the control sample (no drug) and you compare it with the one (yes drug) treated with the S. bisulfite. The sequences are the same because it represents the same piece of DNA.

C (green) → in no bisulfite
not methylated T instead of C → in yes bisulfite
it is red → cytosine methylated

- In the other case (yes drug), the 2 sequences have: T → in yes bisulfite
some of the control but without bisulfite → in no bisulfite

Here no one of the C is methylated

At the end, comparing ALL sequences you say that in the CONTROL CASE → the red C are methylated
the green C not methyl.

in the sample (treated case with DRUG) → demethylation is caused by the drug.
↳ demeth. of C (red)

2nd GENERATION SEQUENCING (next generation seq.)
(massively parallel sequencing)
(high-throughput sequencing)
allow you to produce a huge quantity of datas faster and cheaper vs. SANGER

impossible with sanger

PRO:
Sequence a lot of target (even the whole genome) at same time

Sequencing a lot of samples at the same time
≠ samples → = targets or ≠ targets simultaneously

It is the set of all the old technologies that allow you to produce enormous amount of sequences, to analyze them and interpretate.
On the market there are ≠ platforms with ≠ characteristics, but require DNA or cDNA (RNA) libraries as sequencing input.

- What you need:
  An input → DNA or cDNA library
  Instrument able to generate a signal for every nucleotidic base and record the signal.
- Software to decode signal and produce final datas.

1. LIBRARY PREPARATION
   from DNA or RNA source
   → fragments of nucleic acid with ADAPTERS bound to both ends.

ADAPTERS: necessary for the cluster amplification and for sequencing
synthetic oligonucleotides with specific (non si sente!!) sequence

Fragmentation of DNA/RNA
using sonication or enzymatic digestion
- Could be done only on short pieces of DNA/RNA

Ex

- If you want to sequence all exomes that belong to the genome of the organism. (in human it is 2%)

↓

then you fragment.

↓

you attach, at both ends, the adapters

↓

then you separate the fragments containing the target you want. For doing that you use PROBES, then you separate them from the rest.

↓

At the end you have your library composed only by what you want. (in the case are the exomes)

- If you want to sequence RNA (the trasncriptomes)

Separate the target composed by mRNA from the total RNA of the pool

↓

capture your target and attach it to adapters

↓

At the end you have produced RNA sequencing

- If you want to sequence a target composed by few genes of interest, you sequence the region of interest.

In a disease there is a pool of genes that are mutated. In breast cancer, there are some genes that are mutated.

You design a "panel" to study the disease
↓
it represents the target.

Using some commercial kits containing the probes that allow you to sequence only the target you want:
↓
targeted sequencing amplicon-based.

The probes contained into the pause, anneals around the region of interest
↓
with an extention and a ligation → you can produce a complementary region that links the probes.
↓
At the end you have fragment containing only your target
↓
you attach the adapters to the fragments and amplify using PCR.
↓
At the end you completed the library
↓
So you check the lenght of your fragment and start to sequence!

ADAPTERS → linked to both ends of your fragment
↳ synthetic oligos composed by portions that are used during different steps of sequencing (P5) (P7)
There are some red and green colored portions at both codes, that allow the library to adhere to a solid support
↓
CLONAL AMPLIFICATION

(Rd1 SP, Rd2 SP)

The portions around the DNA insert are the portions of the adapters that are used to anneale primers from which start the insert sequencing

↓

The sequence can start from Rd1 SP or from Rd2 SP.

The back parts (index) are unique sequences, short, of 6 or 8 bases that allow you to sequence more samples at the same time -

Ex

Sequence DNA of 2 patients immediatly that are having the same disease - You prepare the library of both and to tell at the end which library belongs to one or the other patient, you use the index! You mark only the fragment of your library with a specific index from the ones that you used to mark the other library

↓

So during the sequencing you can even sequence the index that belongs to that library => so you're able to recognise the ≠ libraries.

(2) CLUSTER GENERATION

You have the region of interest, at both ends of the fragment of the library, you have the adapters.

You immobilize the library on a solid support. (=ours adapters)

Then the fragments will be clonally amplified.

You can do that using:
A) Emulsion PCR (using beads)
B) Bridge amplification

A) A mixture of emulsion oil (= PCR mix)
↓
with beads covered with oligo that are complementary to the adapters of the library

inside there are even primers and DNA POLYMERASE

↓
you make emulsion. In each droplet of it, you should have one base covered with the oligo (like a primer) and one beads covered with one fragment.
In the PCR mix you have to denaturate the ds fragment, then it anneals to adapters on the beads, then you have the extention
↓
polymerase amplifies the forward strand starting from the beads towards the primer site
(sequence of the other strand)
↓
the sequence that you produced is a denaturated strand.
The original strand can anneal to another oligo on the beads and the strand that you have synthesized has on an end the adapters that is able to anneal to the primer in the reaction.
Both strands are duplicated and this is repeated several times. At the end the beads are covered with a lot of copies of the original fragment = A CLUSTER GENERATION (clonal amplification of a strand)

B) The solid support is a FLOW CELL (fragment)
with a surface coated with a lawn of oligo pairs
(orange and green)

both are complementary to both ends of the adapters.

The double stranded fragment of the library are first denaturated and a SS fragment can adhere to one of the two oligo pairs.

↓

The can be used as primer to syntetize a complementary copy of the SS fragment that is annealed to the oligo.

↓

So you produce a ds fragment.
The origiuae template is discarded and the newly synthesized strand can flip over (↑) and anneal to the oligo near to the complementary to the adapter on the opposite hand.
So you generate a ds bridge that is linearised, can flip over, can anneal to the near oligo and this is repeated again a eat.

↓

generating a cluster.
Also in this case it is composed by 1000 copies of the origiuae sample.

③ ≠ methods of library sequencing

- sequencing by ligase enzyme (sequencing by ligation or hybridization)
- sequencing by DNA polymerase enzyme (+ common)
  - prosequencing
  - ion semiconductor sequencing (by a company: ThermoFisher Scientific)
  - cycle reversible termination (by Illumina)

We have our cluster with fragments, and we have a chip covered with wells.
↓
in which you have to flow your clusters and each well has to contain only one beads.
Below the well there is a sensing layer connected to a silicon substrate.

In this technology the sequence is performed by measuring the changing of voltage.
During the synthesis of DNA, a free deoxynucleotide is incorporated into a growing strand of DNA, the incorporation causes the release of a pyrophosphate and also an H^+
↓
that changes the pH of the solution.
↓
the instrument converts it in a changing of voltage.

* To perform this kind of sequence you have to flow

the 4 nucleotides sequentially → first you have to flow a reaction containing only one nucleotide.

Each well contains a fragment in which the template has a complementary base to the C for example and a dCTP is incorporated, an hydrogen ion is released and a signal is produced. → changing voltage measured

The well in which there is no complementary base to the nucleotide that is inserted in that moment, so any signal is produced.

If in the well, two bases are incorporated, two hydrogen ions are released → double signal.

At the end you change the solution that contains nucleotide and substitute it with another solution that contains another nucleotide

→ We have generated the cluster containing bridge amplification, then we linearize the clusters, you wash of the reverse strand and sequence only the forward strand using a primer that can anneal the region of the adapter near the DNA insert. (Rd1 SP)

The mix solution is composed by all 4 nucleotides and they're labelled with a fluorophore and each nucleotide get a blocking molecule, so when during the sequencing a nucleotide gets incorporated, the blocking molecule interfers with the incorporation of another nucleotide and only one nucleotide for each step can be incorporated.

All the other are washed away.

the nucleotide are excited by a light source and the emission is recordered by a camera. The fluorophore and the blocking molecule are cleaved off and the cycle is repeated.

the n. of cycle is determining the lenght of the sequence.

The camera using a filter can acquire the signal and can convert into a base.

If you want to sequence a complementary straud you have to perform a PAIRED-END SEQUENCING

The complem. straud is discard, there is a bridge formation that allow the extention of complementary sequence.

DS is linearized and the original forward straud is cleaved off.

So you can start the sequencing of your insert starting from the opposite end.

through the use of a specific primer, that allows you to do it into the opposite end.

If you want to sequencing more than one samples simultaneously, you have to mark your library with a index, but you have also to read and sequence the index to assign it to the proper library and to do that there is a step that must be done.

How to use indexes?

- Library preparation (index 1, index 2) to each fragment
- Pool all fragments of the two libraries
- Sequence simultaneously the 2 libraries
- A software is able to produce the sequence of the index, then it is able to sort the sequence belonging to one sample, reading the index that you have attached to the library of the sample 1 or 2.