

* The DNA base sequence carries the information a cell needs to assemble protein & RNA molecules

DNA sequencing is important for scientists to investigate the function, evolution, & structure of a gene or protein. Also important in Genetic disorders, diagnostics.

• Sequencing techniques by principle

1. chain Termination ex Sanger sequencing

2. Sequencing by synthesis

Single nucleotide Addition (SNaP)
ex Roche 454, Ion Torrent

cyclic Reversible Termination (CRT)
ex Illumina (Solexa) & Qiagen (GeneReader)

3. Sequencing by ligation ex Applied Biosystem (Solid)

complete Genomics (BGII)

4. Single - Molecule Real - Time (SMRT)

ex Pacific Biosciences (PacBio), Oxford Nanopore

* Sanger method →

- Reaction mixture contain :

1. Target DNA to be sequenced
2. Nucleotides - dNTP -
3. Nucleotides - ddNTP -
4. Buffer
5. DNA Poly.
6. primer

- sequence short DNA fragments - 1000 bp -

• Involves use of special kind of nucleotides called dideoxy that cause chain elongation stop

no oxygen on carbon 2', 3'

Synthesis of DNA to be terminated

Note OH at 3' end is important for DNA poly. because it recognize it & use it to add new nucleotide & to make a phosphodiester bond with a phosphate group

- formates → original
new

Original

New

- involves sequencing the DNA fragment using gel electrophoresis (polyacrylamide gel)

- involves sequencing by synthesis (DNA fragment that we want to sequence will be sequenced while it's made)

- involves preparing four separate sequencing reactions in four separate tubes, in each tube the DNA poly. will be added, 1 primer, template, but each reaction receive an extra dideoxy nucleotide

- choice will be random (ddG-dT)

- when we do sequencing reaction we add thousands of identical template molecules

- doesn't need 4 reactions (only 1 tube, 1 reaction)

- High throughput computer automated sequencing using Sanger method, very helpful for completing human genome project

- dNTPs, DNA poly., primer template, ddNTPs (fluorescent with different color) are mixed in one reaction

- then we allow polymerase to synthesize all possible new strands & all these strands will have one type of dideoxy

- we separate samples on the single lane capillary gel,

We have a very thin tube filled with gel material & we force nucleotides to migrate through capillary by electricity

Original

- The newly synthesized strand that will receive dideoxy their chain elongation will be stopped, adding dideoxy this mean no free-OH-for the polymerase to add the next nucleotide (creating fragments of varying lengths)

Ex → 1000 template
600 dATP 400 ddATP
synthesis will continue

ddNTP → Fluorescently labeled
radioactively labeled
radioisotopes (P^{32} , P^{50})

- The fragments are separated on polyacrylamide gel & Autoradiography is used to identify radioactive fragments.

New

- Samples will be added at the side of cathode as migrating to anod

- The speed of migration will depend on size

- In some point the capillary will be subjected, scanned with laser

- laser emits a different wavelength of light for each different colored ddNTP, it's job allowing the fluorescent to generate a color that will be collected & detected by detector

- detector amplifies & feeds this information to a computer that can run multiple capillary gels at one time = 900 bp sequence

- Now we have peaks are shown (chromatogram)

* Roche 454 →

- based on pyrosequencing

↳ Rely on the Generation
of Pyrophosphate.

- produces highly accurate & long stretches of DNA sequence (greater than 1 gigabase) of DNA per reaction
- use Fluorescence imaging techniques
- Could be used to sequence the whole Genome

- Before sequencing we start with thousands of genome molecules & they are cut to smaller fragment & these billions of small fragment will be mixed with microscope beads (each bead will receive one of the DNA fragment to be sequenced)

~~genome
molecules~~
 10^9 of
small fragment

& these beads will be in SLH
called oil emulsion (water mixed
with oil)

Then → each bead carrying one or few DNA fragment will be separated from each other by adding billions of these beads to billions of well
can fit only 1 bead:

Then → the sequencing reaction will happen

- The first thing will happen each fragment on the bead will be amplified by PCR, so each bead will be carrying hundreds of PCR amplified products.

- Polymerase will remove 2 phosphate (only 1 of the phosphate incorporated to the newly synthesized strand) as Pyrophosphate.

The release of it
could be detected
by 2 enzymes

1. Luciferase
2. Sulfurylase

* The Sulphydrolase can use pyrophosphate to produce ATP

* The luciferase will convert luciferin into oxyluciferin

* ATP, luciferase, luciferin will be used to produce light

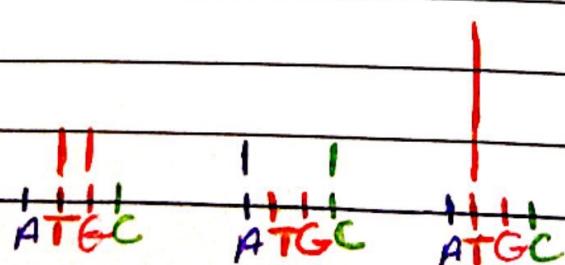
- Add 1 nucleotide at a time & after that it will take a picture of all wells & see which well produce light, then there will be a washing step that will remove nucleotides, polymerases that is added.

signals A T G C
well

* Sometimes you might have more than 1 nucleotide must be added

Polymerase synthesis — AAA local the is
3T well 2 is T

which means the amount of light will be 3 times



* Ion semiconductor sequencing →

- Based on the detection of hydrogen ions (proton, H^+)

• Addition of nucleotides at a specific order & the machine detect the release of H^+ after each addition

• nucleotides washed away & next nucleotide is added

- A microwell containing a template DNA strand to be sequenced is flooded with single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand.

- This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence, multiple nucleotides will be incorporated in a single cycle.

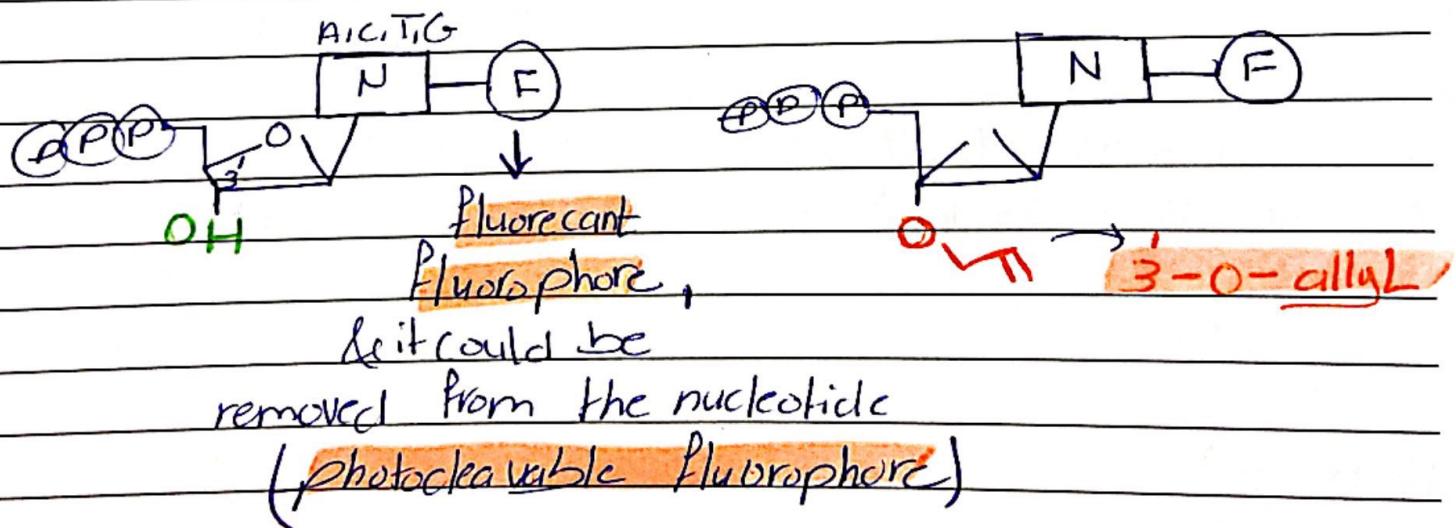
* Cipagen Gene Reader →

- Similar to 454 approach in which DNA fragments are linked to beads & amplified but different sequencing technologies used
- Machine that carry sequencing is called Gene Reader
- use special types of nucleotides

structure of nucleotides

Normal

special



- 4 different nucleotide could be used if they have different nitrogen bases & different Fluorophore

* After the fragmentation of the genome, DNA fragments will receive adaptors & they will be added to beads, & these fragments on bead will be enriched, amplified (bead-based template enrichment)

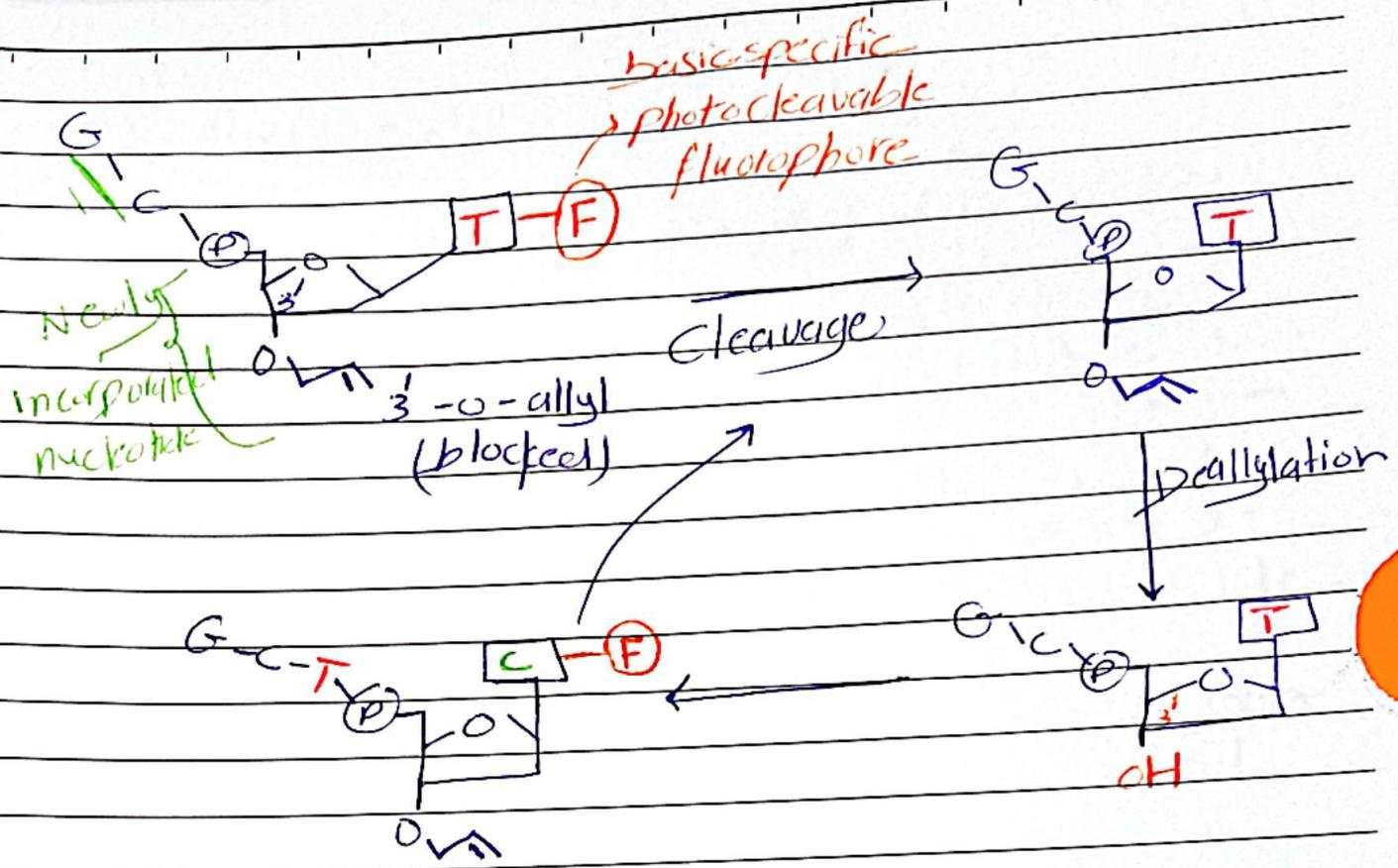
Then primers & polymerases & modified nucleotide will be added to the flow cell

- A mixture of fluorophore-labelled, terminally blocked nucleotides & unlabelled

↳ means that nucleotides have 3' allyl group, no free OH to allow the addition of the next nucleotide

so → the idea is that when a nucleotide is added (fluorophore attach to it), a laser will excite this dye & the color will indicate which nucleotide is added.

Next → the fluorophore is cleaved & washed away & the free OH group is regenerated, so that another nucleotide in the next cycle can be added according to the sequence of template DNA



G C T C
C G A G T - - -

unincorporated bases are washed away

- After the addition of nucleotide, laser will take the color of the added nucleotide (imaged)

- Deallylation → to remove the Group ($\text{O} \swarrow$) & then the next nucleotide could be added

So then the next step of cleavage & deallylation will be repeated over & over

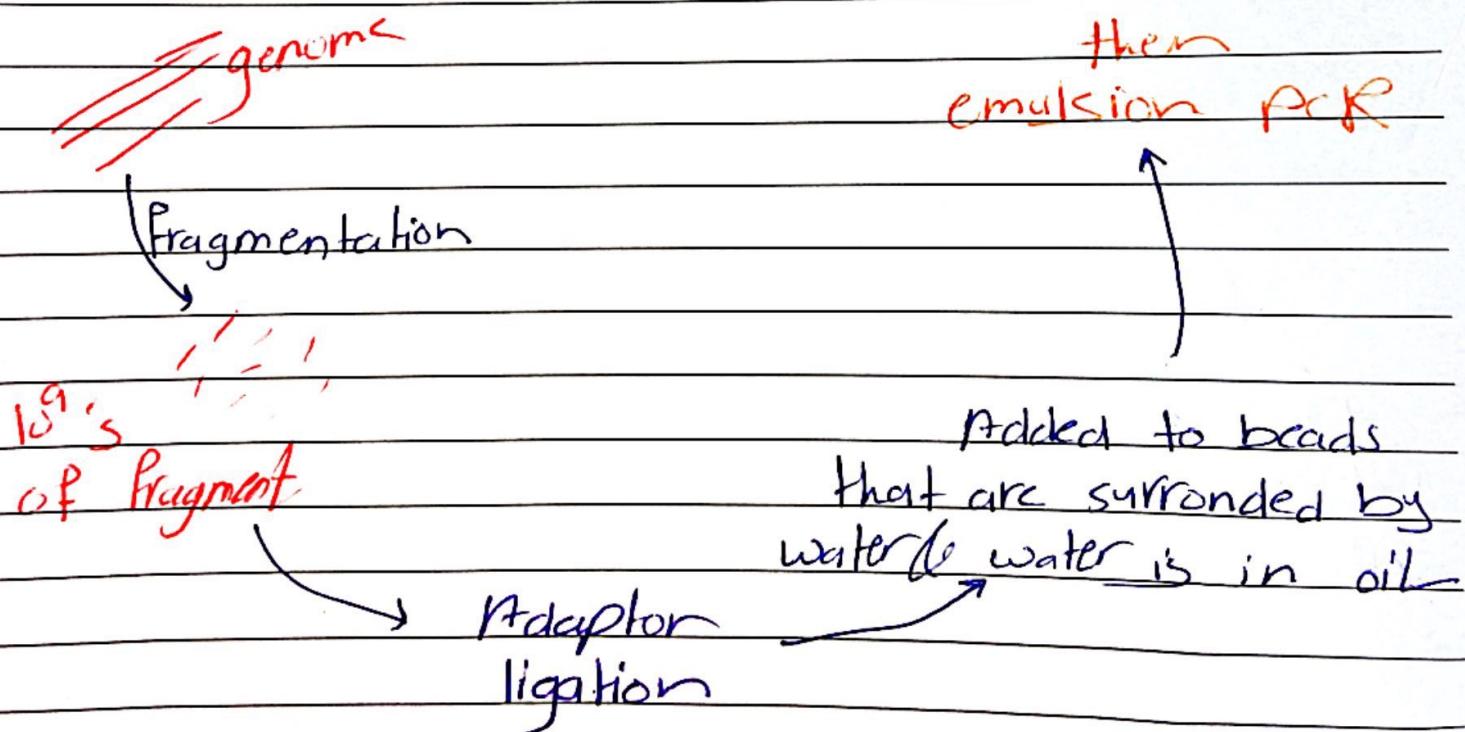
* ABI Solid →

- developed by company called Applied BioSystem

Solid: supported oligonucleotide ligation & detection

Ligases & probes
oligonucleotide

- Similar to 454 approach in which DNA fragments are linked to beads & amplified but different sequencing technologies used



Then → the fragments will be amplified on the surface of the bead. Then the next step is to take these beads (it will be spread on flow cells - solid surface - (454 beads were added to wells))

→ 454 will yield -

before sequencing, the DNA is amplified by emulsion PCR. The resulting bead, each containing only copies of the same DNA molecule, are deposited on a glass slide. The result is sequence of quantities & lengths comparable to Illumina sequencing.

Then sequencing could happen

use thousands of probes, ligases, phosphatases

* Beads will be carrying template DNA sequence (this template will be attached to the bead through adaptor)

Now → to sequence these template we use Di base probes: consist of 8 nucleotides of the first 2 nucleotides that we are interested in & allow us to sequence this template.

- we have 4 different types of probes, each one end with a fluorescent group

* How sequencing happen?

It universal seq. primer will be added
& it's complementary to the adaptor & there
is a perfect match from the 5' end of primer
to the PI adaptor

Now based on the type of
the first 2 nucleotide of template
a specific probe can bind,

Now → if this binding happen, aligase
will seal this gap (make a
phosphodiester bond), now the
Fluorescent group will be excited to make
a signal

Notice

between n & z we have
a cleavage site

* So after the binding of the first probe & after capturing the signal for the first color, acknowledge agent will be used to remove Fluorophore from the probe.

What will remain is 2 ~~de~~ bases, now another probe will be added.

After several cycles → of adding, ligating, receiving the signal, the color of the bead keep changing based on the type of probe.

Now the next step → is to denature this double stranded probe (adaptor, probe, template, sequencing primer)

The top strand will be removed & degraded & then a newly primer will be added.

- the location of the binding primer is shifted by 1 nucleotide

* First primer was called n

because it was perfectly bind to the end of adaptor

but next primer is added is called $n-1$

shorter by 1 nucleotide from its 5' end, so it will^{n't} bind with the adaptor

Now we have cycle of addition of probes & ligases

now the probes here it will hybridize to last nucleotide in adaptor

Note

- Each round we have 7 colors
- unligated probes are washed out

1
2

Now → we have double stranded DNA,
— template completely hybridized to probe
& what will happen that we will repeat
this steps (primer repeat ...)

1 Then a universal primer in round
— 3 will be added ($n-2$)

then we repeat the record of
the colors of the bead

Then → we use a software to determine
— the sequence of the template.

Decoding

→ determining the sequence of the
template based on the colors generated

• Glass cell contain flow cell &
each cell we have bead

• ~~multiple gel analysis~~ using log 10 ~~and~~ *

* Illumina Solexa →

- involves Fragmentation
- sequencing a whole genome
- Based on reversible dye-terminations

~~genome~~

Fragmentation

!!!

→

we have fragments & then we have adaptor ligation (we have 2 different adaptors attach to the different ends)

- we don't add them to beads, we add on the surface & one this surface we have also some single adaptors

Note → before they are added they will be denature

- Single stranded DNA fragments with adaptors on both ends will attach on the surface

Also have primers that have the same sequencing of adaptors

* Amplification happens on the top of the surface by Bridge amplification

The adaptors on the fragments will be complementary to the primers on the surface

so this molecule (single DNA) will bend to make like a bridge

So now we have complementarity between adaptor & primer

Next → what will happen that nucleotides, poly. Will be added & now we have sth called Bridge

↳ repeated many times & now we end up with enrichment of these DNA fragments

So now we have clusters

↳ each cluster

Contains hundreds of the same DNA molecules having the same sequence,

- Certain conditions will allow the annealing of adaptors

poly., dNTP will be added so that the polymerase can extend this primer & make a new strand

so → now we have double stranded & it will be separated

Cluster ~~size, where, time, cost~~

- After the generation of cluster, the next step is carrying out sequencing procedure

↳ Requires polymerases & dNTP's labelled with different color

So sequencing begins after cluster growth, there will be millions of clusters & now the next step is to determine the sequencing of fragment in each cluster

For ex → If we added $-T-$, we will have a laser that will excite the fluorescent tag that is attached to $-T-$ nucleotide & then the $-T-$ will produce green color

next we add another nucleotide $-G-$, so if it will be hit by laser it will produce blue color

• Each dot with a color represent a cluster & the color of the cluster keep changing after the addition of each nucleotide

→ within the sequencing machine there is a scanner

* Why do we have hundreds of fragments within the same cluster have identical sequence?

because the light produced from a fragment mightn't be enough to be captured by the camera.

so for ex → all these fragments will first incorporate the $-T-$, so all fragments will generate a green $-T-$ & the amount of the green color will be enough to capture by camera/scanner

* base calling

↳ based on the color, we can determine the order of the bases added while the newly synthesized is being made based on the template sequencing within the cluster

* Why do we need to remove the other nucleotides?

because the next step is to hit nucleotide with laser, if these remains when the laser hit the flow cell, all these nucleotides will produce different color

Flow Cells and Imaging

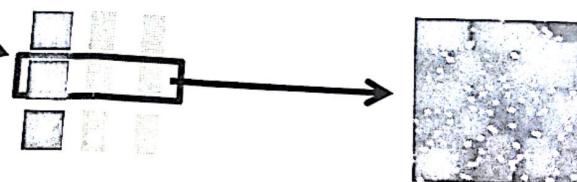
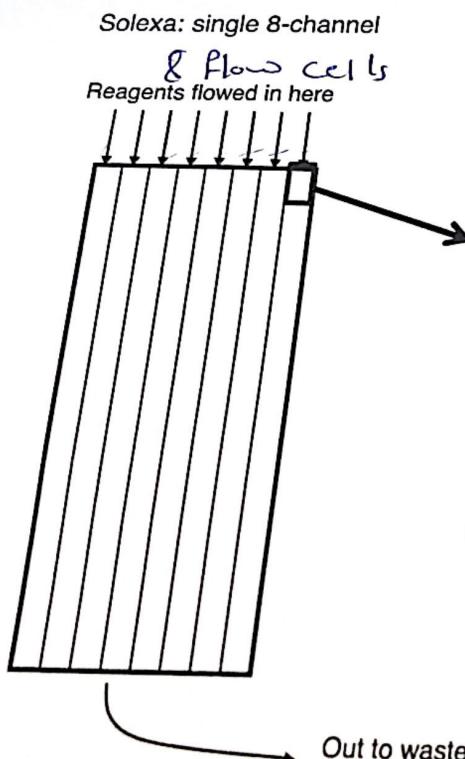


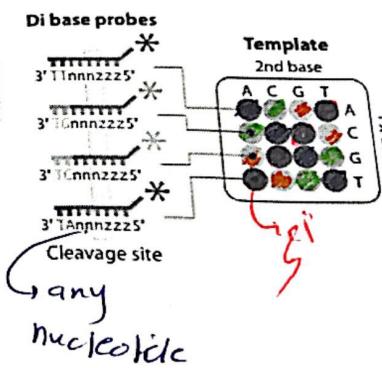
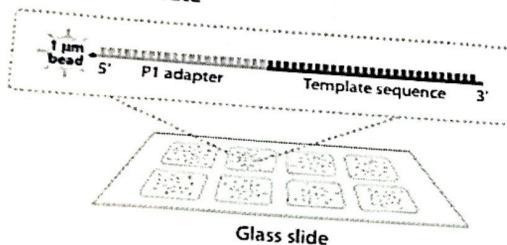
Image for each channel is actually tiled (four images per panel, one for each color)

The reagents will be added then their gonna be removed from the other end & then the scanner (laser) will excite the fluorescent labeled & the camera will capture the colors of clusters & this is repeated over & over again

From Huttenthaler Lab, Dept. of Biostatistics
Harvard University, USA

ABI SOLID

a
SOLiD™ substrate



10^{24} di probes

$\begin{matrix} & \text{u} & \text{u} & \text{u} \\ & | & | & | \\ \text{G} & \text{C} & \text{N} & \text{N} & \text{N} & \text{Z} & \text{Z} & \text{Z} \\ \text{C} & \text{G} & - & - & - & - & - & - \\ \text{A} & \text{T} & - & - & - & - & - & - \end{matrix}$ → 64 different di probes

256 different di probes with red fluorescent Group
ABI SOLID