Karolinska Institutet

Overview of available sequencing technologies

EDCTP ENNEA training on data management

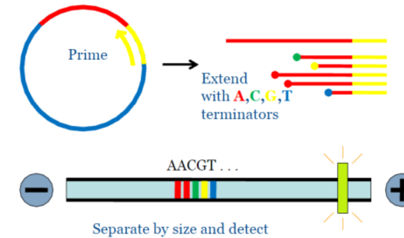
November 14-16 2011, Muhimbili hospital, Dar Es Salaam

Irene Bontell

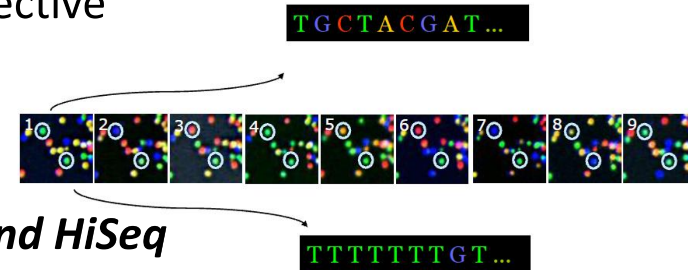
(irene.bontell@ki.se)

Sequencing – 3 generations

1. Capillary sequencing (ABI3700)
   One template -> one capillary -> 0,5-1 kb

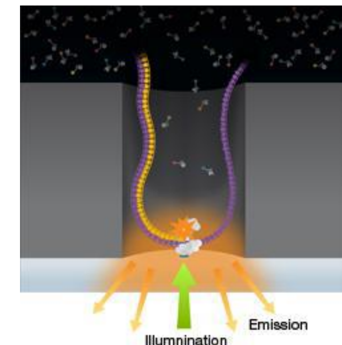


2. ”Next generation sequencing” or ”Ultra-deep sequencing”
   Short reads, extremely high throughput , cost-effective
   – but requires lots of computer power!



   - ***Roche (454): GS FLX Titanium and GS Junior***
   - ***Illumina (Solexa): Genome Analyzer (GAIIx) and HiSeq***
   - ***ABI: SOLiD***

3. ”Next-next generation sequencing” or ”Single-molecule sequencing”
   Longer reads, less throughput, fast cheap runs, single-molecule: easier analysis
   - ***Pacific Biosciences: PacBio RS – SMRT sequencing***
   - ***Ion Torrent***
   - ***Oxford Nanopore***
   - Visigen/Starlight
   - Helicos
   - Complete Genomics



High throughput sequencing

- Extremely high output at low per base cost (but expensive per run!)
- Massively parallell – need to design projects carefully
- For small genomes sample identification essential
- Requires lots of computer power (and knowledge) for analysis
- Ideal for large-scale genome projects and metagenomics
- Also for ultra-deep sequencing to discover minor populations

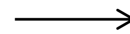
IT infrastructure:

- For every TB of sequence -> need 15TB for analysis
- Need to be ruthless – no storage of primary data (images converted on the instruments)
- One Illumina run – 64 cores needs 2-3 days for processing
- Approx $100,000 needed per sequencing machine to cope with IT

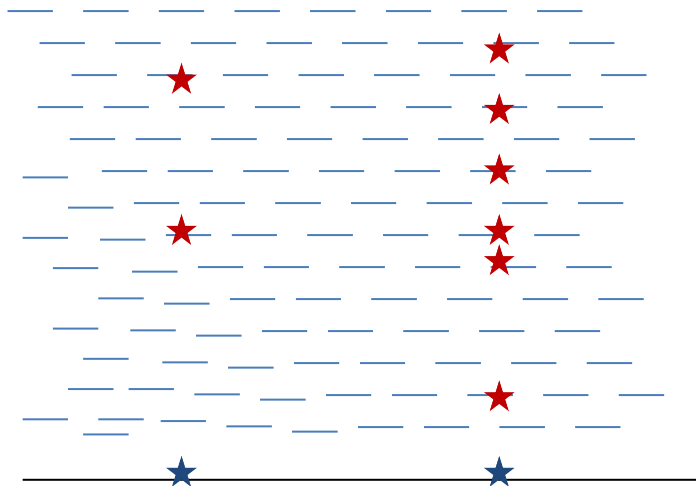
High throughput sequencing for HIV

Roche, 454

Illumina



Millions of parallell reactions (short reads) detected by extremely high resolution cameras



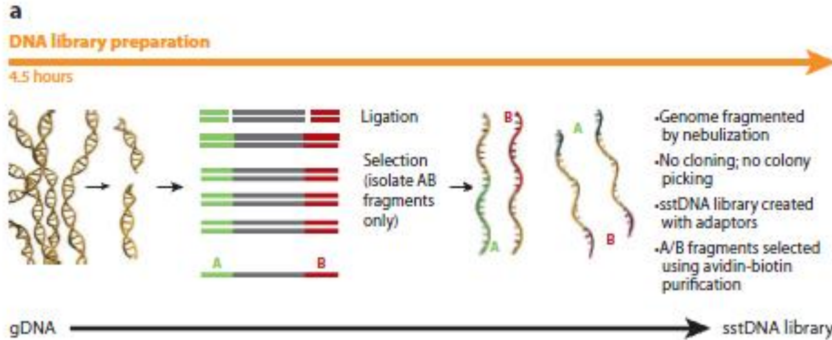
Consensus sequence generated from short fragments aligned to reference. Mutation prevalence can be accurately quantified!

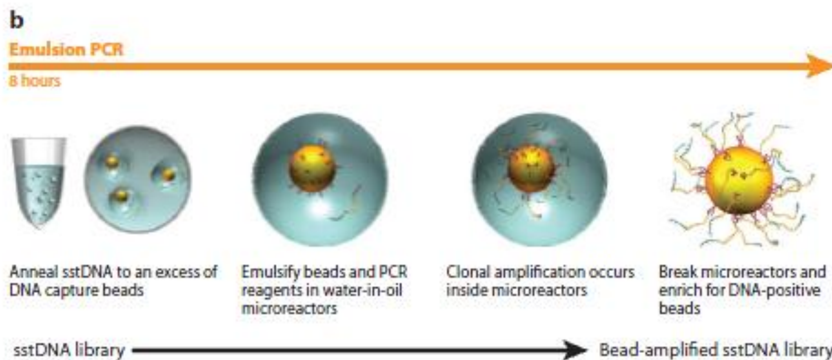
With longer reads – defining minor quasispecies

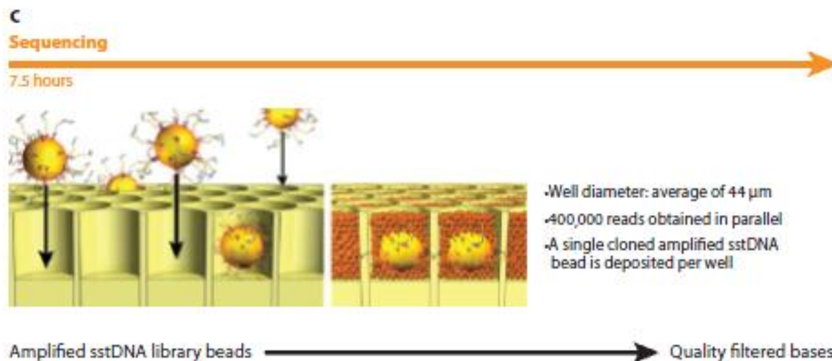
Consensus sequence

Roche 454

www.my454.com

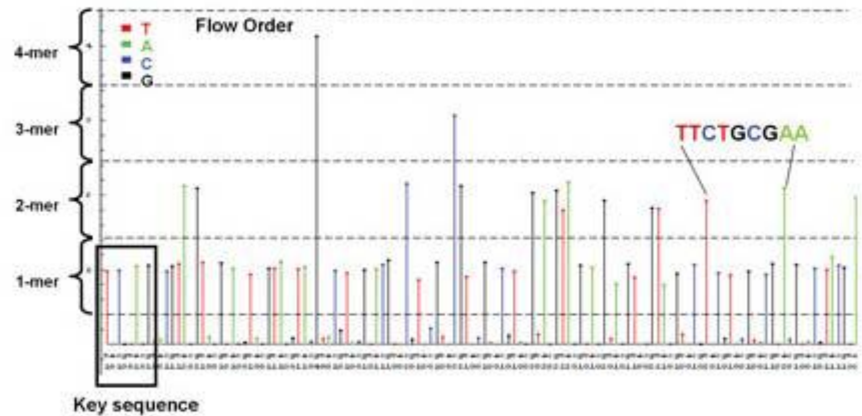






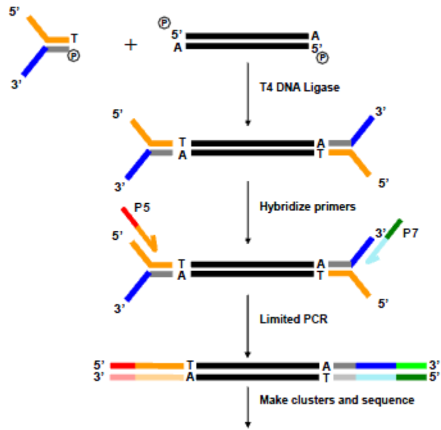
Sequencing by synthesis – pyrosequencing (detection of pyrophosphate release on nucleotide incorporation)

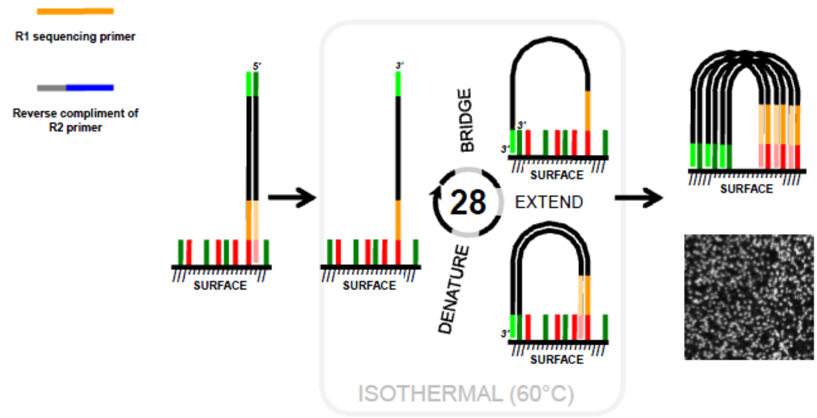
Sequential flow of four bases - picture after each flush



Illumina HiSeq

www.illumina.com/systems.ilmn



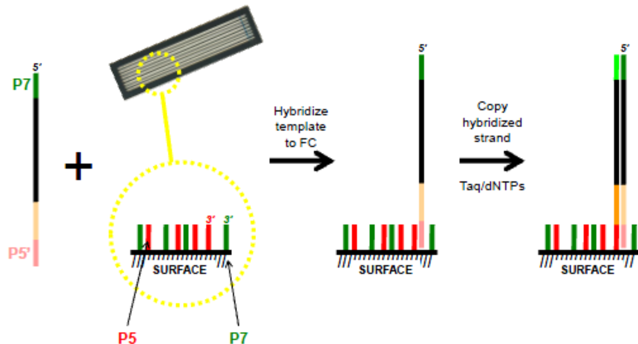


~ 60,000 clusters per image in each colour

4 colours = 240,000 clusters per image (tile)

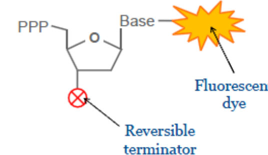
120 tiles per lane (~1 hr imaging per cycle)

2*108bp -> 10 days



Sequencing by synthesis with reversible terminator and fluorescent dye

All four dNTPs at once –> competition –> higher accuracy
No problem with homopolymers



Paired-end sequencing:

- Select fragments of certain length (up to 500 bp)
- Regeneration of cluster after 1st round (without moving chip relative to camera)
- Matching pairs – greatly facilitates genome assembly

ABI SOLiD

www.appliedbiosystems.com

Read position 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

Primer round
1 Universal seq primer (n) 3'
2 Universal seq primer (n−1) 3'
3 Universal seq primer (n−2) 3' Bridge probe
4 Universal seq primer (n−3) 3' Bridge probe
5 Universal seq primer (n−4) 3' Bridge probe

● Indicates positions of interrogation   Ligation cycle 1 2 3 4 5 6 7

SOLiD™ substrate

1 μm bead 5' P1 adapter Template sequence 3'

Glass slide

Di base probes

3' TTnnnzzz 5'
3' TGnnnzzz 5'
3' TCnnnzzz 5'
3' TAnnnzzz 5'

Cleavage site

Template
2nd base
A C G T
1st base
A C G T

1. Prime and ligate

Primer round 1

POH

+ Ligase

Universal seq primer (n) 3' AT

1 μm bead P1 adapter TA Template sequence 3'

2. Image

Excite Fluorescence

TA 3'

3. Cap unextended strands

Phosphatase

PO₄

3'

4. Cleave off fluor

Cleavage agent

HO

AT P

TA 3'

8. Repeat Reset with , n−2, n−3, n−4 primers

5. Repeat steps 1–4 to extend sequence

Ligation cycle 1 2 3 4 5 6 7 ... (n cycles)

3'

6. Primer reset

Universal seq primer (n−1)

3'

2. Primer reset

1. Melt off extended sequence

1 μm bead 3'

7. Repeat steps 1–5 with new primer

Primer round 2

1 base shift

−1

Universal seq primer (n−1) AA CA CG TC AA TA CC

3'

1 μm bead T GT GC AG TT AT GG 3'

Sequencing by ligation

Data collection and image analysis

Collect color image

Identify beads

Identify bead color

Collect color image

Identify bead color

Identify beads

Glass slide

Primer round 1, ligation cycle 1

Primer round 2, ligation cycle 1

Primer round 3, ligation cycle 1

Primer round 4, ligation cycle 1

Color space for this sequence

Possible dinucleotides encoded by each color

2nd base
A C G T
1st base
A C G T

Template sequence

AT	AC	AA	GA
CG	CA	CC	TC
GC	GT	GG	AG
TA	TG	TT	CT

Double interrogation

With 2 base encoding each base is defined twice

A T G G A

Decoding

Color space sequence

TA	AC	AA	GA
GC	CA	CC	TC
CG	GT	GG	AG
AT	TG	TT	CT

Possible dinucleotides

Base zero

AT TG GG GA Decoded sequence

A T G G A Base space sequence

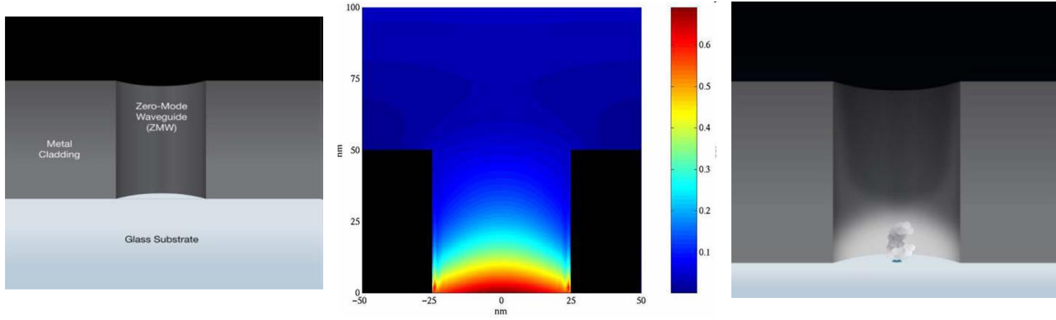
Single molecule sequencing

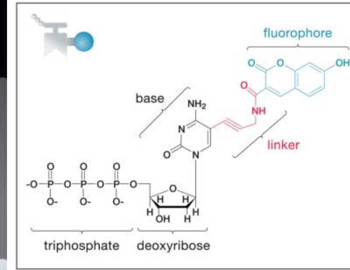
- No amplification needed (less error)
- Longer reads, know the haplotype immediately
- Fast and more cost-effective for smaller projects
- Emerging technologies allow for direct sequencing from RNA

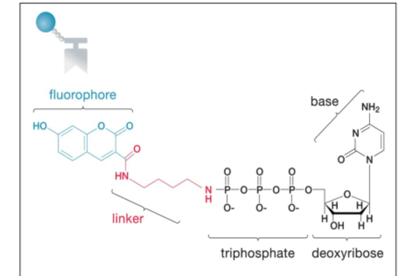
PacBio

www.pacificbiosciences.com

Single-molecule sequencing in real-time

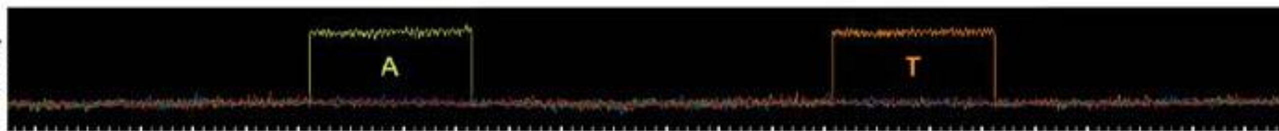




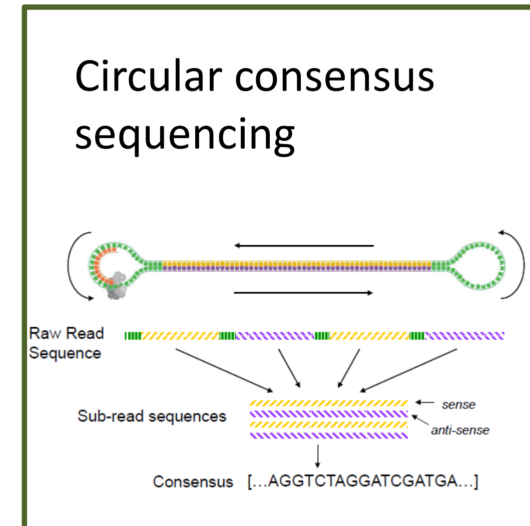


DNA polymerase as sequencing engine – but needs to be slowed down (750 bp/s -> 1-3 bp/s)

Fluorophore naturally cleaved off by polymerase
DNA synthesized is uninterrupted
Eliminates steric hindrance and noise



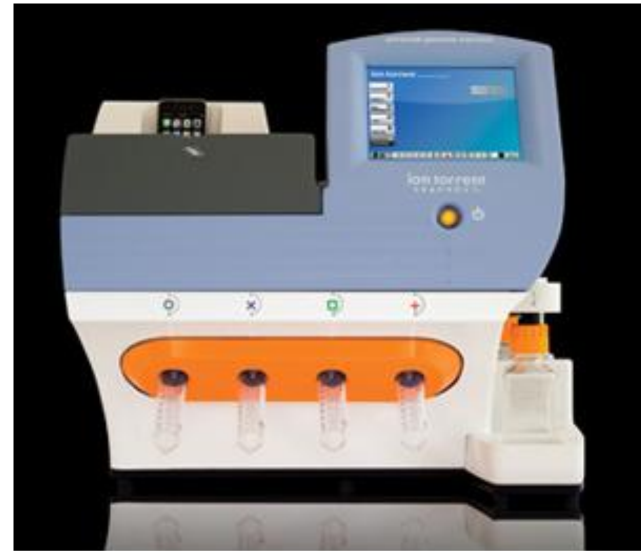
Step 1: Fluorescent phospholinked labeled nucleotides are introduced into the ZMW.
Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light.
Step 3: The phosphate chain is cleaved, releasing the attached dye molecule.
Step 4-5: The process repeats.

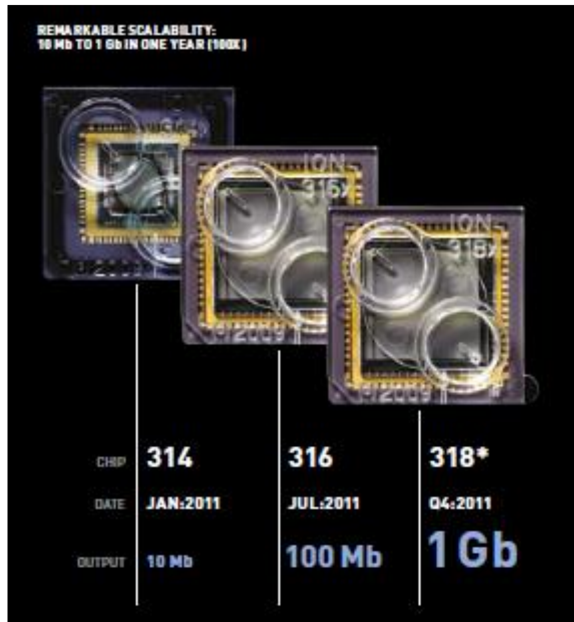


80,000 ZMWs monitored simultaneously (1/3 have one polymerase)

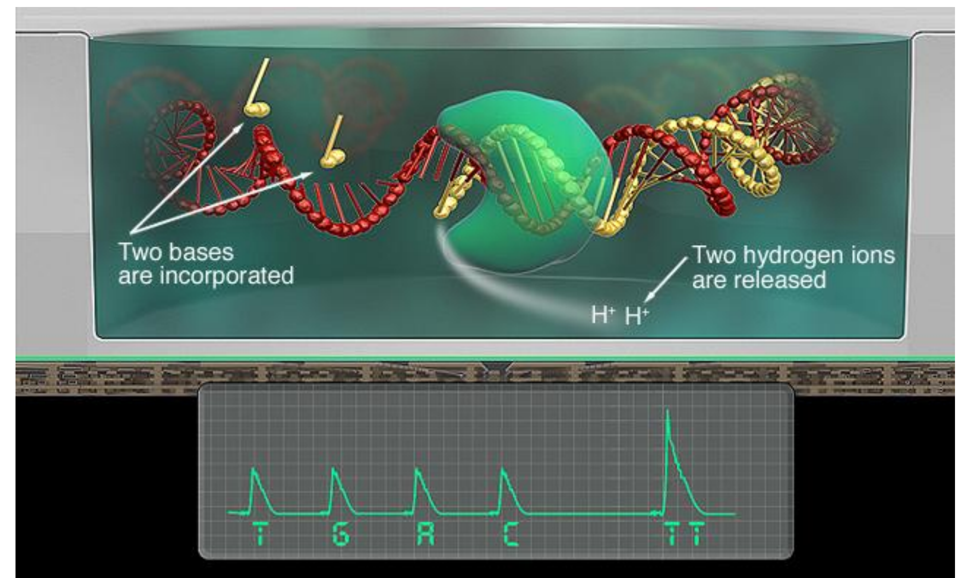
ion torrent

www.iontorrentcom





"Ion Torrent uses the simplest sequencing chemistry including natural nucleotides, no enzymatic cascade, no fluorescence, no chemiluminescence, no optics, no light: The Chip is the Machine"



Oxford **NANOPORE** Technologies®

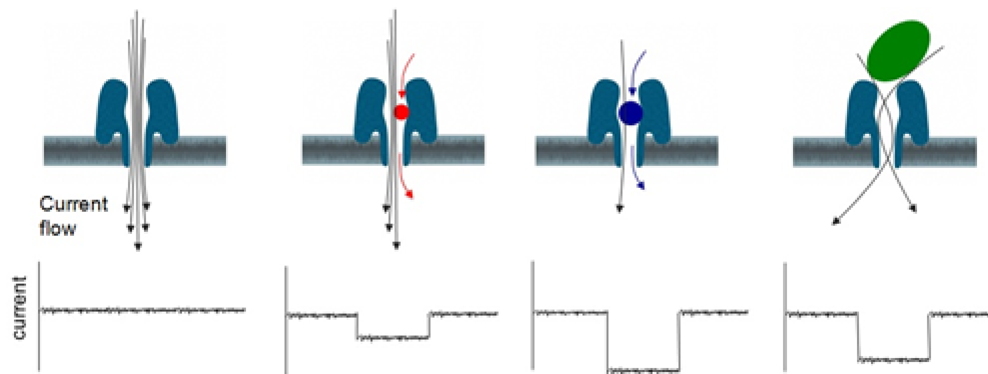
www.nanoporetech.com

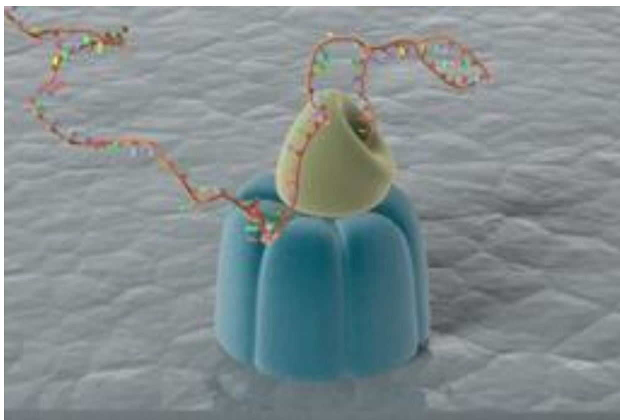
About nanopores

A nanopore is, essentially, a very small hole. This hole may be formed either by a protein pore set into a membrane (biological nanopores), or by artificially creating a hole in solid materials (solid state nanopores).

A nanopore may be used to identify a target analyte as follows:

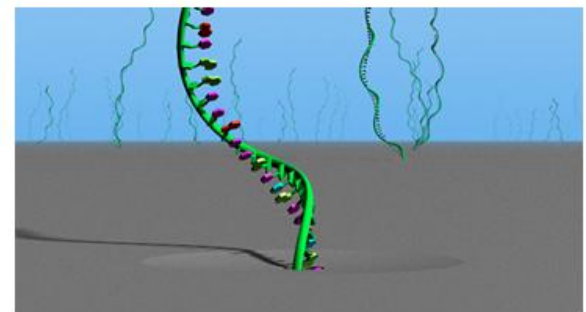
A current is passed through the nanopore. This diagram shows a pore created by the protein α-hemolysin in an electrically resistant lipid bilayer.
If an analyte passes through the pore or past its aperture, this event creates a characteristic disruption in current. By measuring that current it is possible to identify the molecule in question. For example, this system can be used to distinguish the four standard DNA bases and G, A, T and C, and also methylated C. It can be used to tell the difference between enantiomers of ibuprofen or to identify reactive compounds such as explosives.





Can be applied DNA, RNA, proteins. No amplification needed

DNA SEQUENCING (SOLID STATE NANOPORES)



Graphene sheet