AZ Sint-Jan Brugge

MB&C2024

Workshop NGS

8-2-2024, Diepenbeek

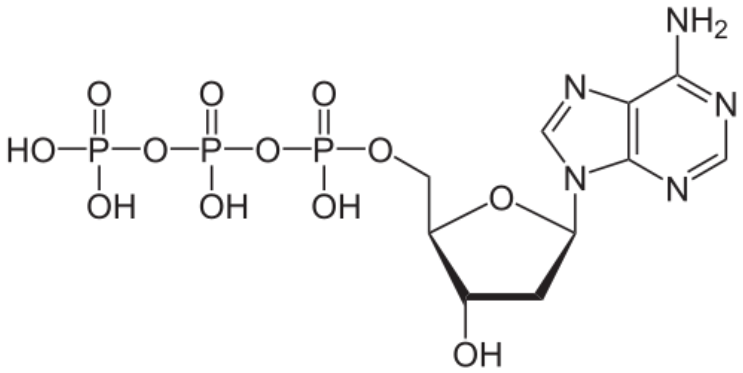
Friedel Nollet, Ph.D.
Molecular Biologist
Laboratory Medicine
AZ Sint-Jan Brugge-Oostende AV

Time line

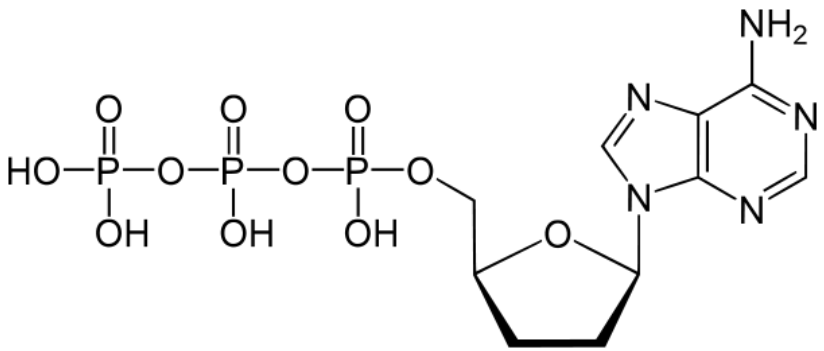
- 1953 ontdekking van de structuur van DNA (Watson en Crick)
- **1975 Publicatie Sanger sequencing methode**
- 1982 Opstart Genbank
- 1987 Eerste automatische sequencer (gel based, ABI373)
- 1996 Eerste capillaire sequencer (ABI310)
- 2003 publicatie humaan genoom project
- 2005 Eerste Next Generation Sequencing platform (Roche/454)
- 2006 Genome Analyzer (Illumina)
- 2007 SoliD Sequencer (Applied Biosystems)
- 2009 1ste single molecule sequencer (Helicos)
- 2011 Ion semi conductor sequencer (Thermo Fisher Scientific)
- 2011 SMRT sequencer (Pacific Bioscience)
- 2015 Nanopore MinION (Oxford Nanopore)

AZ Sint-Jan Brugge

Sanger sequencing



2’-deoxynucleotide trifosfaat



2’,3’-dideoxynucleotide trifosfaat

Sanger F, Coulson AR. “A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase” *J Mol Biol.* 25;94(3):441–448 (**1975)**

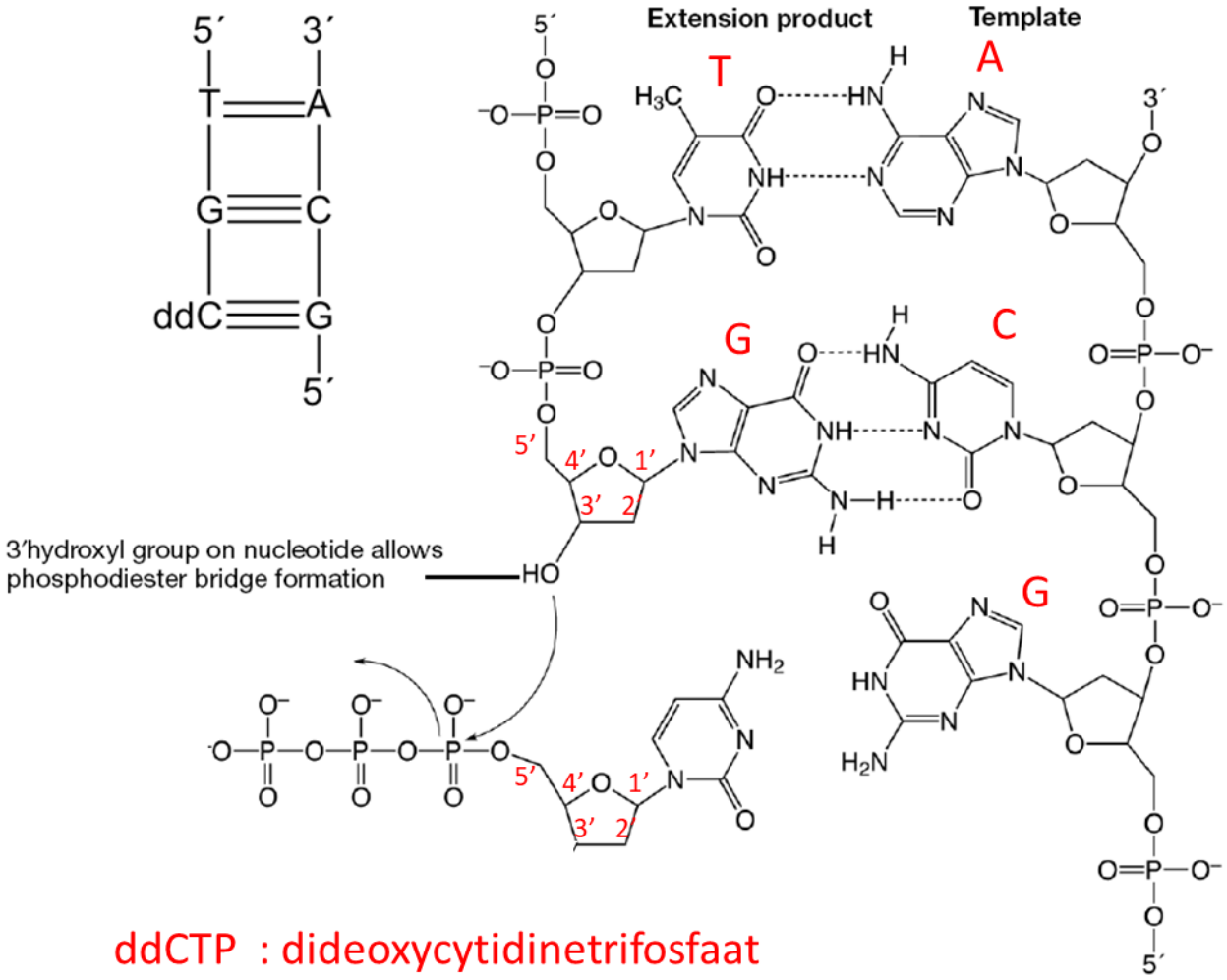
Frederick Sanger :

1958 : Nobelprijs voor structuur eiwitten (oa. insuline)

1980 : 2de Nobelprijs voor DNA sequencing methode

AZ Sint-Jan Brugge

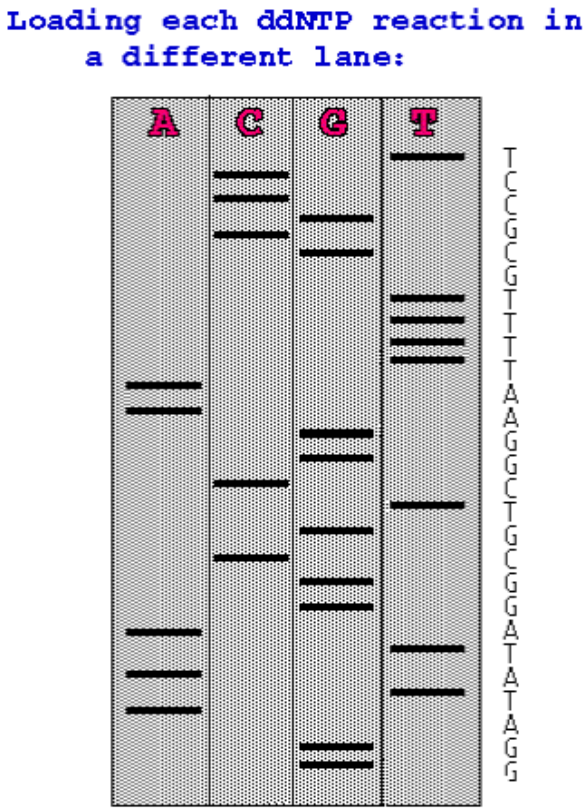
Dye Terminator Sequencing

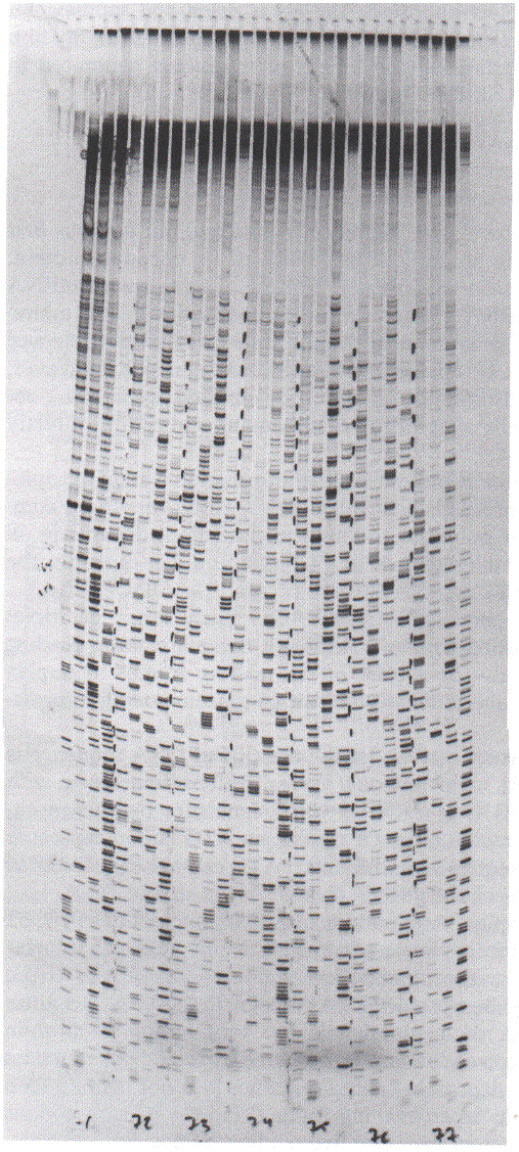


ddCTP : dideoxycytidinetrifosfaat

AZ Sint-Jan Brugge

Sanger sequencing



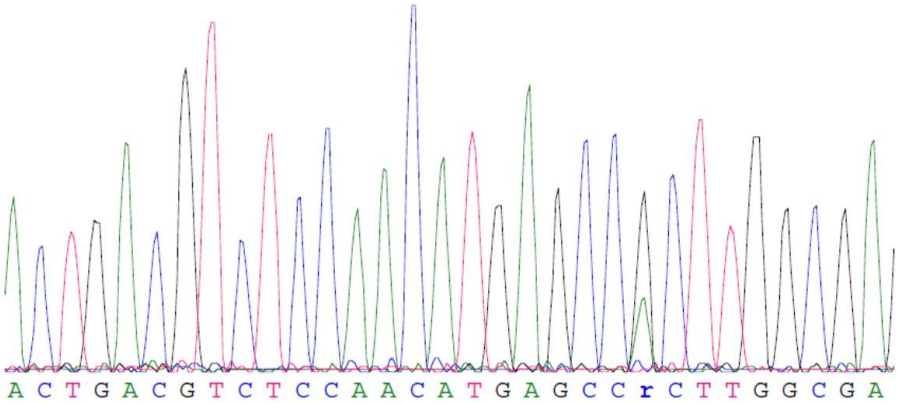


Obv radioactieve isotopen
(beta-emitters, electronen)
^{35}S ipv O
^{32}P isotoop ipv ^{31}P

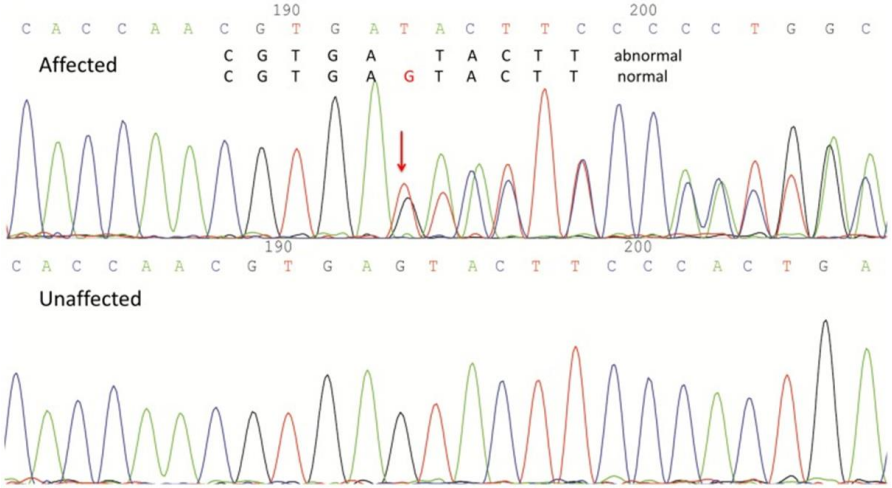
AZ Sint-Jan Brugge

Sanger SEQUENCING met Dye terminators

- Substitutie variant



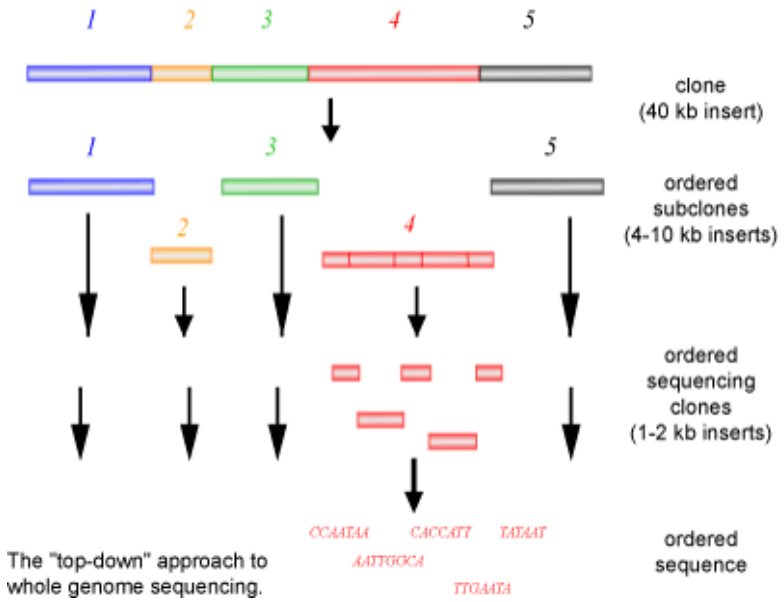
- Indel variant



AZ Sint-Jan Brugge

Human genome project (2001 (90%) / 2003 (99%))

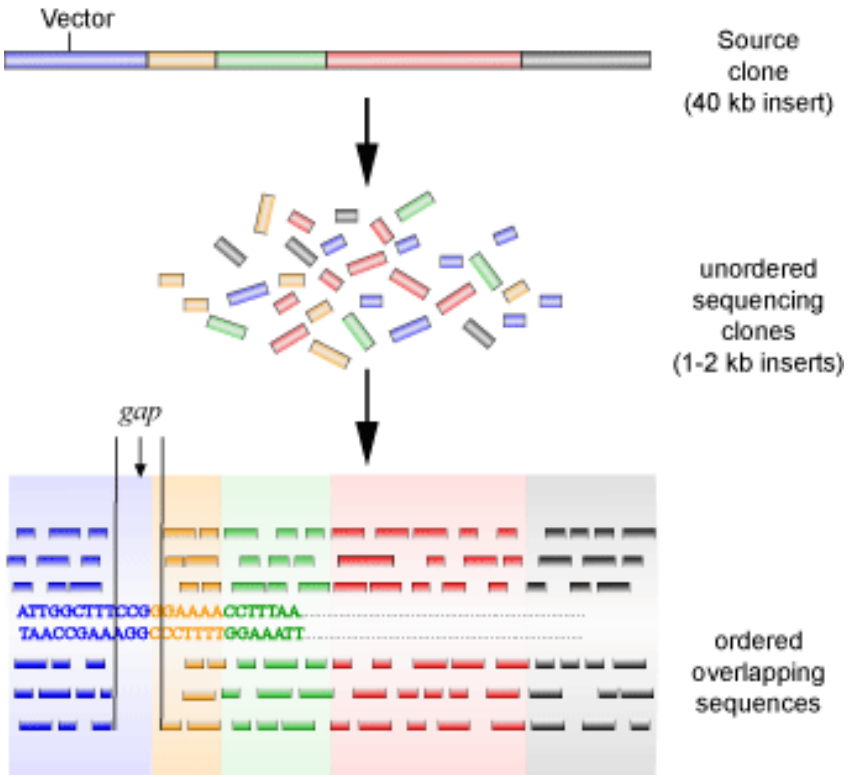
BAC-to-BAC sequencing



The "top-down" approach to whole genome sequencing.

strategie publiek consortium

Shotgun sequencing



strategie Celera Genomics

AZ Sint-Jan Brugge

Time line

- 1953 ontdekking van de structuur van DNA (Watson en Crick)
- **1975 Publicatie Sanger sequencing methode**
- 1982 Opstart Genbank
- 1987 Eerste automatische sequencer (gel based, ABI373)
- 1996 Eerste capillaire sequencer (ABI310)
- 2003 publicatie humaan genoom project
- **2005 Eerste Next Generation Sequencing platform (Roche/454)**
- 2006 Genome Analyzer (Illumina)
- 2007 SoliD Sequencer (Applied Biosystems)
- 2009 1ste single molecule sequencer (Helicos)
- 2011 Ion semi conductor sequencer (Thermo Fisher Scientific)
- 2011 SMRT sequencer (Pacific Bioscience)
- 2015 Nanopore MinION (Oxford Nanopore)

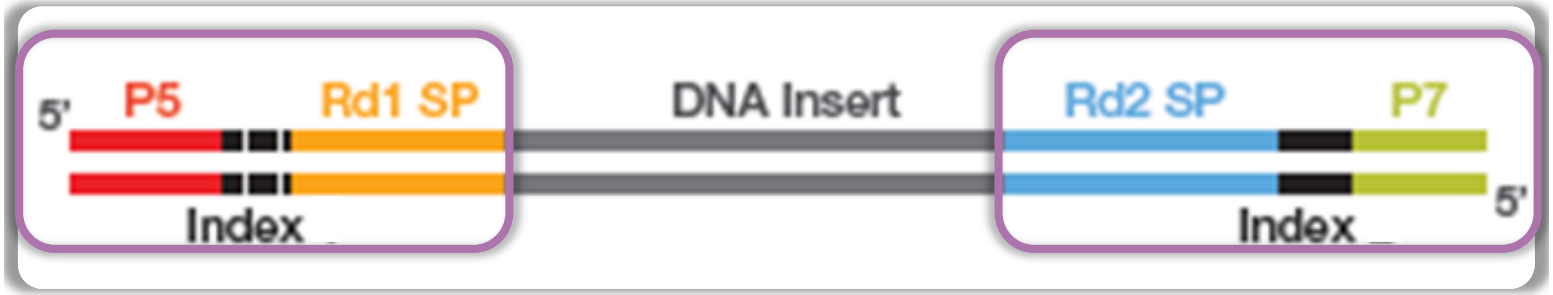
AZ Sint-Jan Brugge

Time line

- 1953 ontdekking van de structuur van DNA (Watson en Crick)
- **1975 Publicatie Sanger sequencing methode**
- 1982 Opstart Genbank
- 1987 Eerste automatische sequencer (gel based, ABI373)
- 1996 Eerste capillaire sequencer (ABI310)
- 2003 publicatie humaan genoom project
- **2005 Eerste Next Generation Sequencing platform (Roche/454)**
- **2006 Genome Analyzer (Illumina)**
- 2007 SoliD Sequencer (Applied Biosystems)
- 2009 1ste single molecule sequencer (Helicos)
- **2011 Ion semi conductor sequencer (Thermo Fisher Scientific)**
- 2011 SMRT sequencer (Pacific Bioscience)
- 2015 Nanopore MinION (Oxford Nanopore)

AZ Sint-Jan Brugge

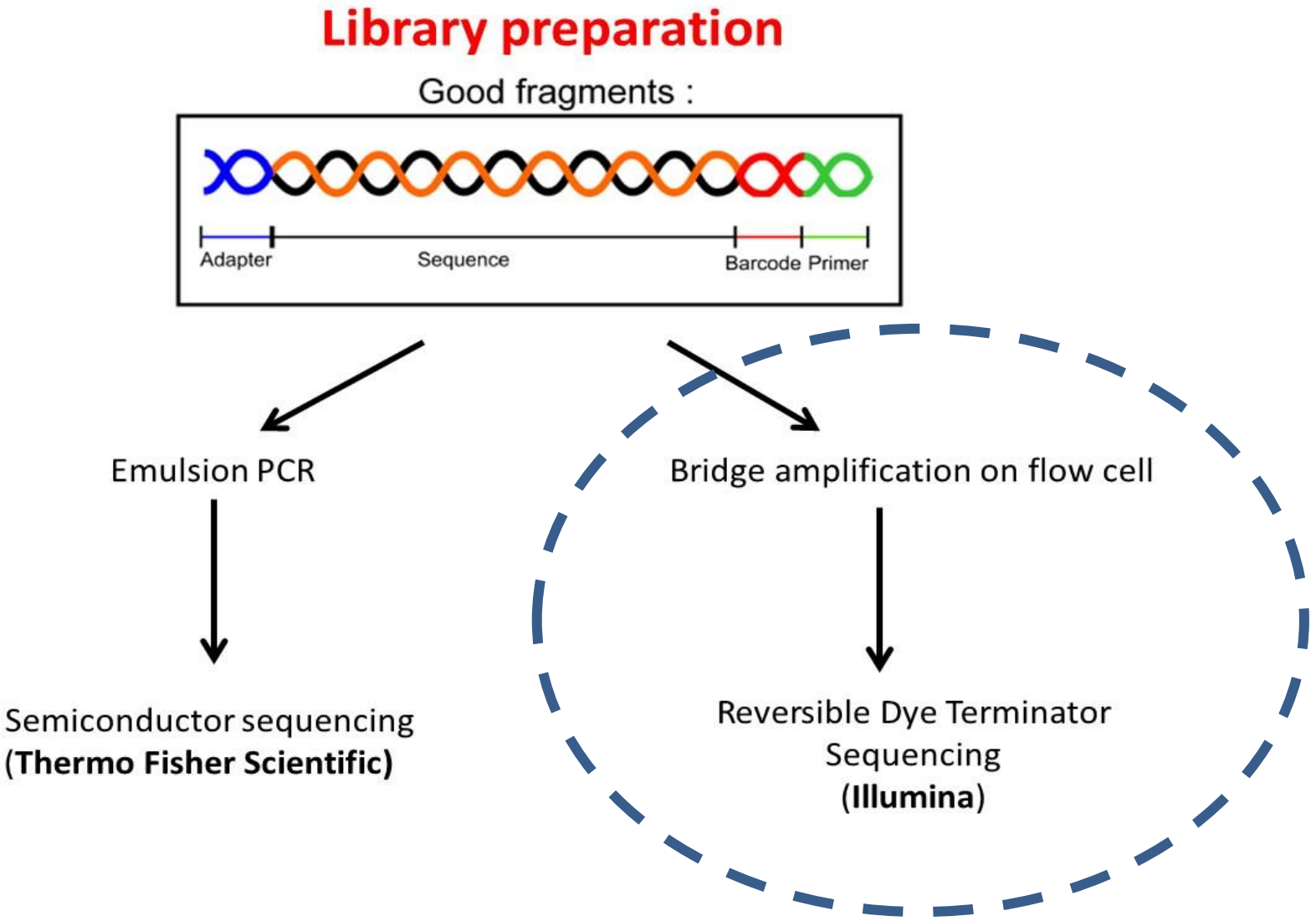
Library molecules voor illumina



- **P5 and P7** sequenties aan uiteinden van de adapters binden aan de flowcell
- **Index** sequentie : identificeren van stalen
- **Rd1 SP en Rd2 SP** : Read-1 en -2 sequencing primer sequentie
- **DNA insert** meestal ~200bp tot 1kb

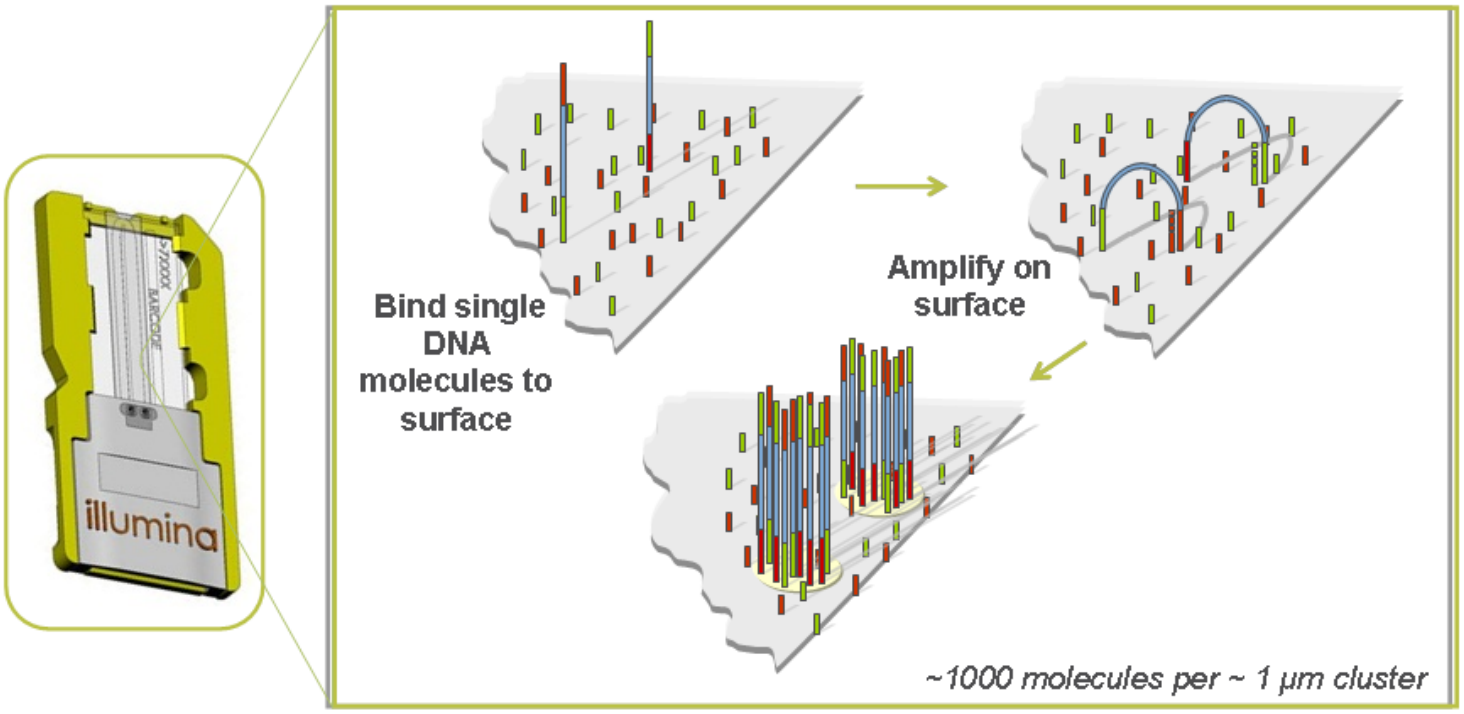
AZ Sint-Jan Brugge

Amplified Single Molecule Sequencing



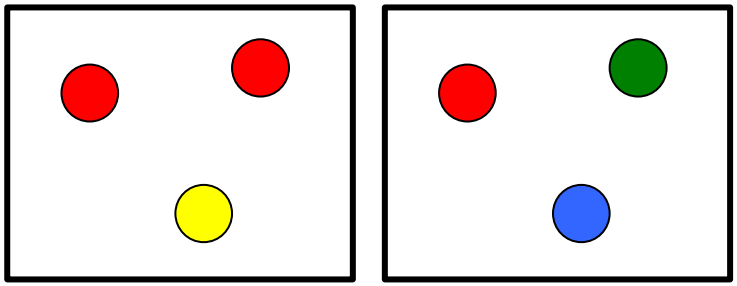
AZ Sint-Jan Brugge

Illumina - Bridge amplification

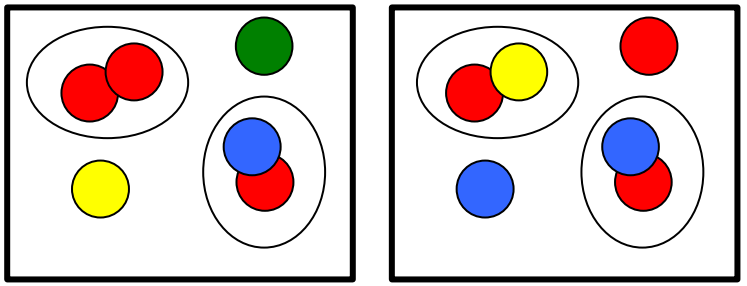


AZ Sint-Jan Brugge

Cluster densiteit

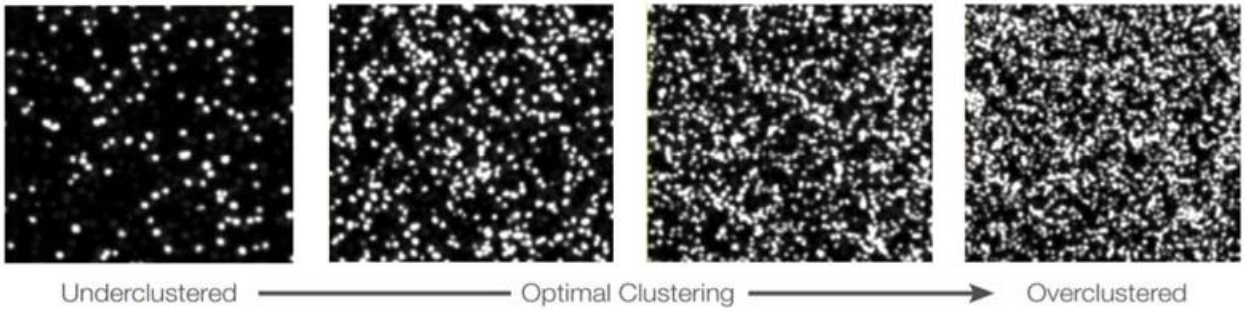


Well-spaced clusters easier to call



Densely-packed clusters difficult to call

- Optimale cluster densiteit
  - Miseq v2 1000-1200 K/mm2; v3 1200-1400 K/mm2
  - NextSeq550 : 170-220 K/mm2
- Accurate kwaliteitscontrole en kwantificatie (pM) van de library is essentieel!

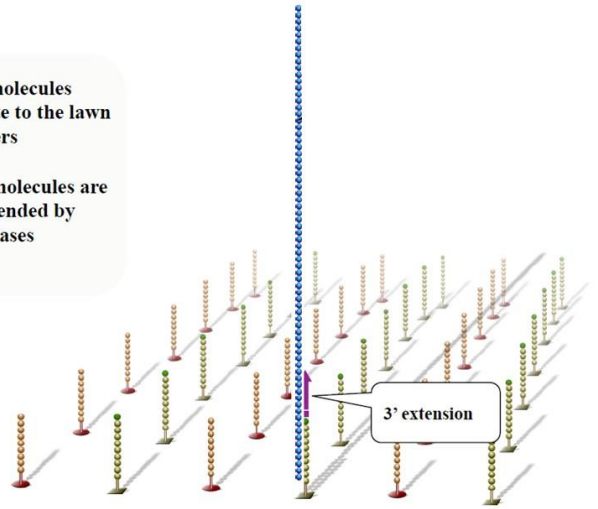


AZ Sint-Jan Brugge

Cluster generation:
Hybridize fragment & extend

single molecules hybridize to the lawn of primers

Bound molecules are then extended by polymerases

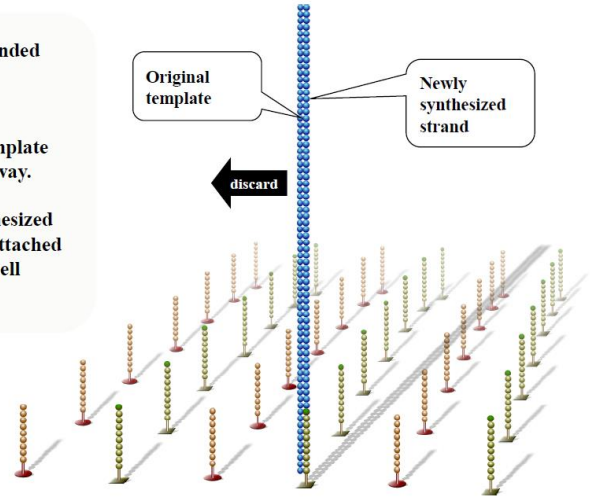


Cluster generation:
Denature double-stranded DNA

Double-stranded molecule is denatured.

Original template is washed away.

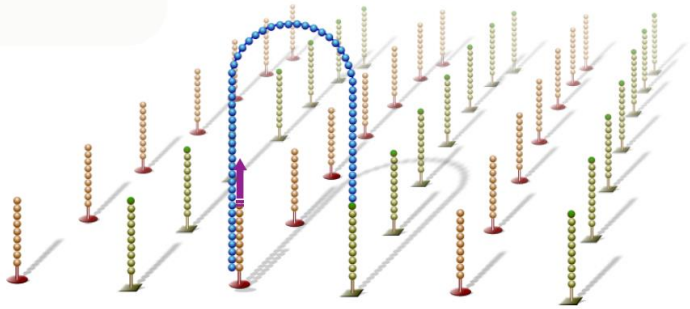
Newly synthesized covalently attached to the flow cell surface.



Cluster generation: Bridge amplification

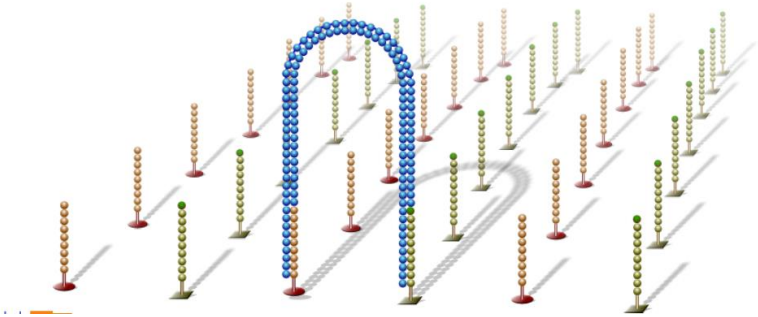
Single-strand flips over to hybridize to adjacent primers to form a bridge.

Hybridized primer is extended by polymerases.



Cluster generation: Bridge amplification

→ double-stranded bridge is formed.

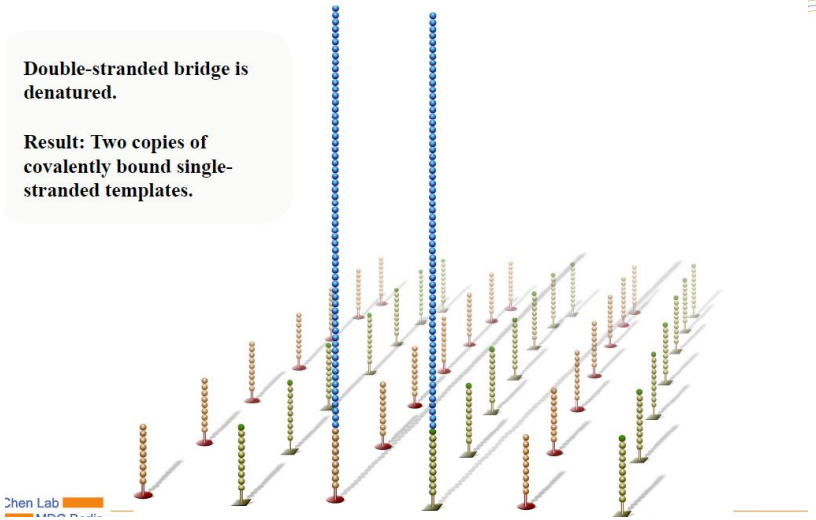


AZ Sint-Jan Brugge

Cluster generation: Bridge amplification

Double-stranded bridge is denatured.

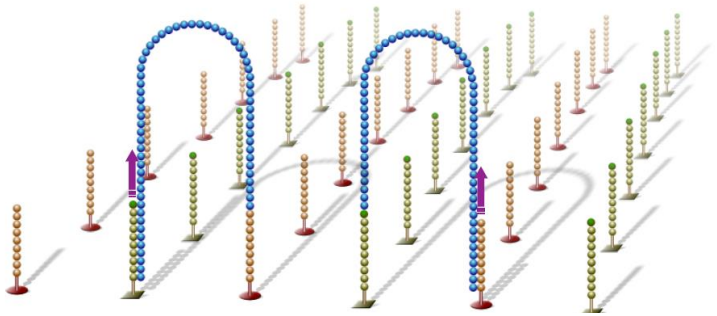
Result: Two copies of covalently bound single-stranded templates.



Cluster generation: Bridge amplification

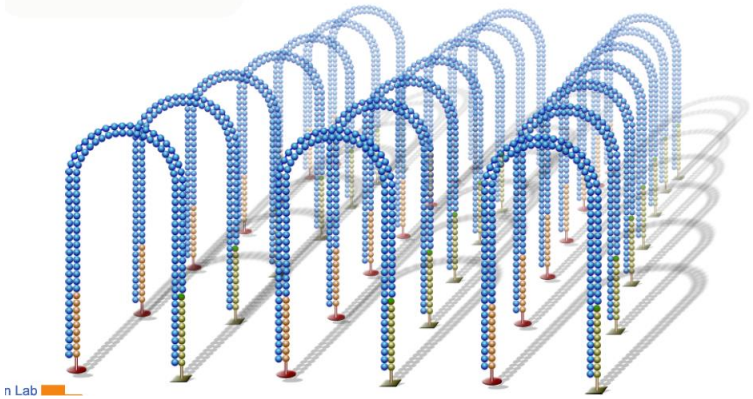
Single-strands flip over to hybridize to adjacent primers to form bridges.

Hybridized primer is extended by polymerase.



Cluster generation: Bridge amplification

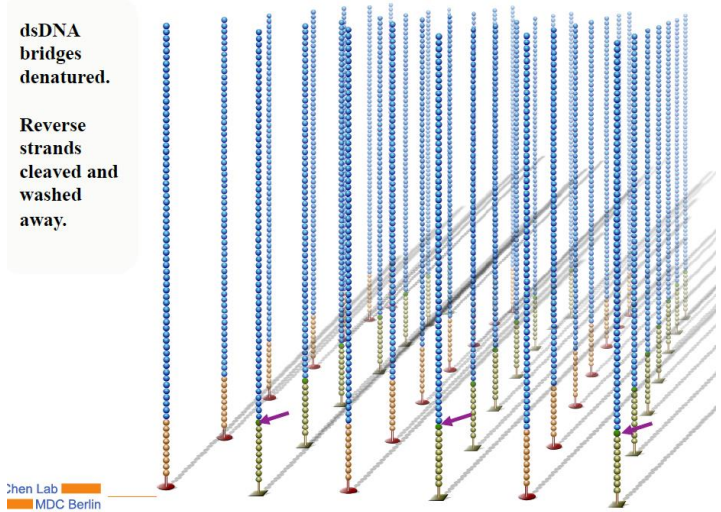
Bridge amplification cycle repeated till multiple bridges are formed



Cluster generation

dsDNA bridges denatured.

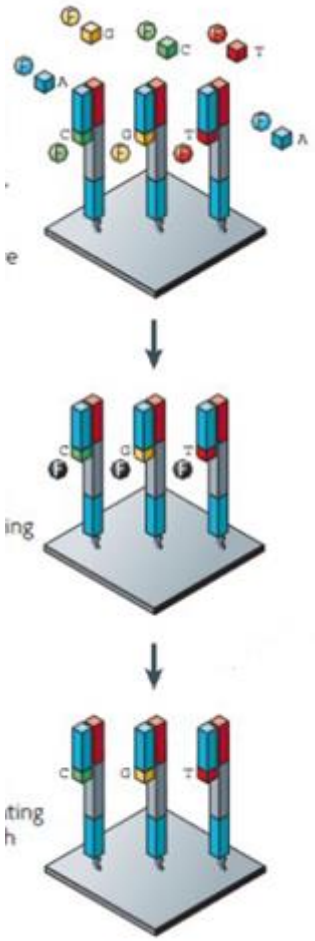
Reverse strands cleaved and washed away.



AZ Sint-Jan Brugge

Illumina reversible dye terminator sequencing

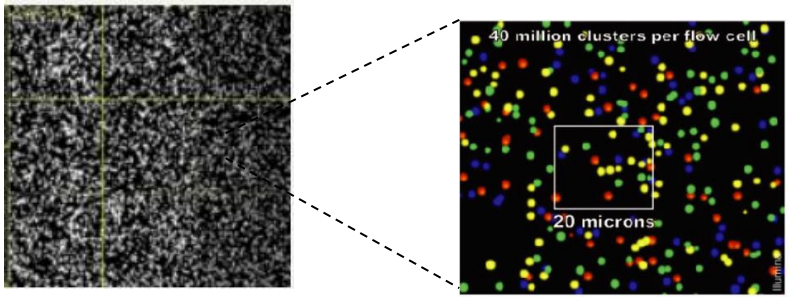
Sequencing cycle



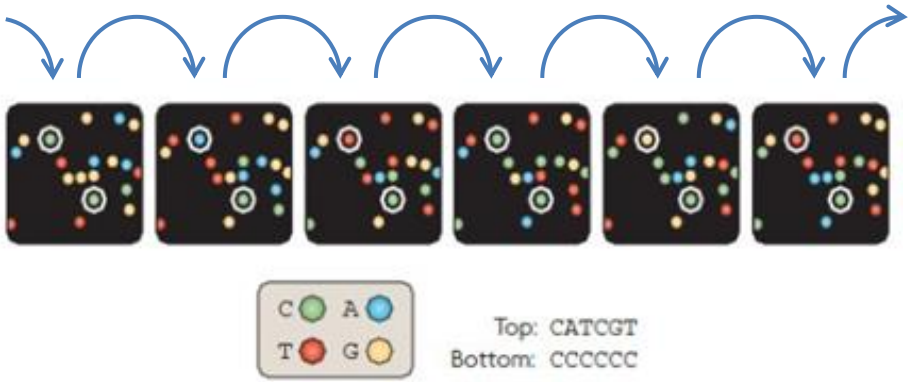
Incorporate all four nucleotides, each labeled with a different dye

Wash and 4-color image

Cleave dye and terminating groups, wash



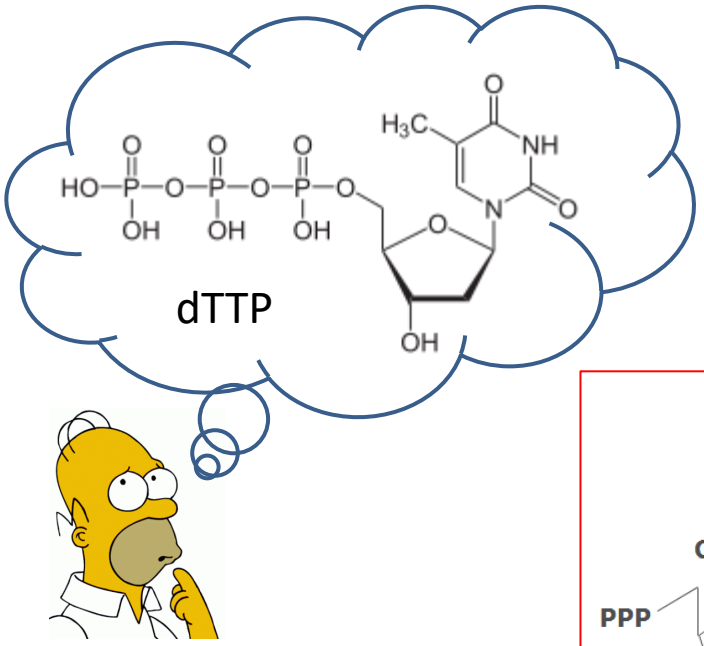
10^6 clusters/mm2

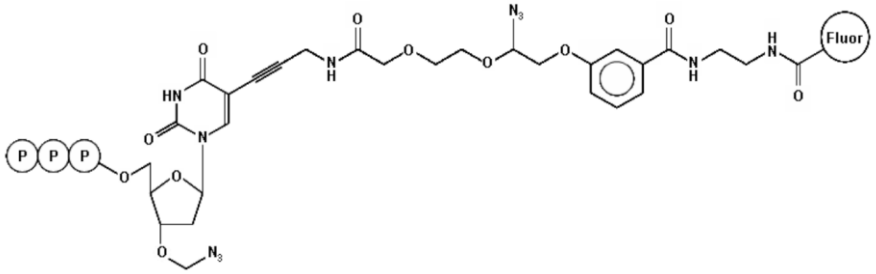


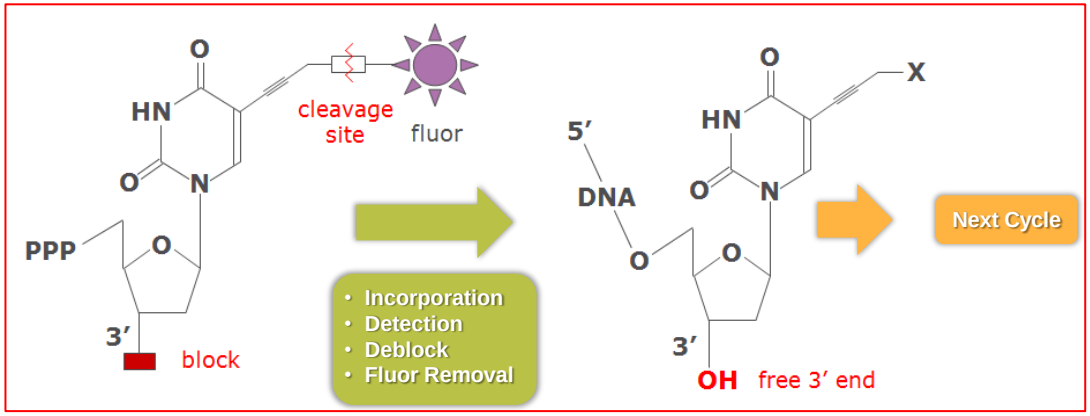
AZ Sint-Jan Brugge

Reversible dye terminator

39-O-azidomethyl 29-deoxynucleoside triphosphates (A, C, G and T), each labelled with a different removable fluorophore







Bentley DR, Balasubramanian S, et al. (2008). *Nature*. **456** (7218): 53–59

AZ Sint-Jan Brugge

Sequencing by synthesis

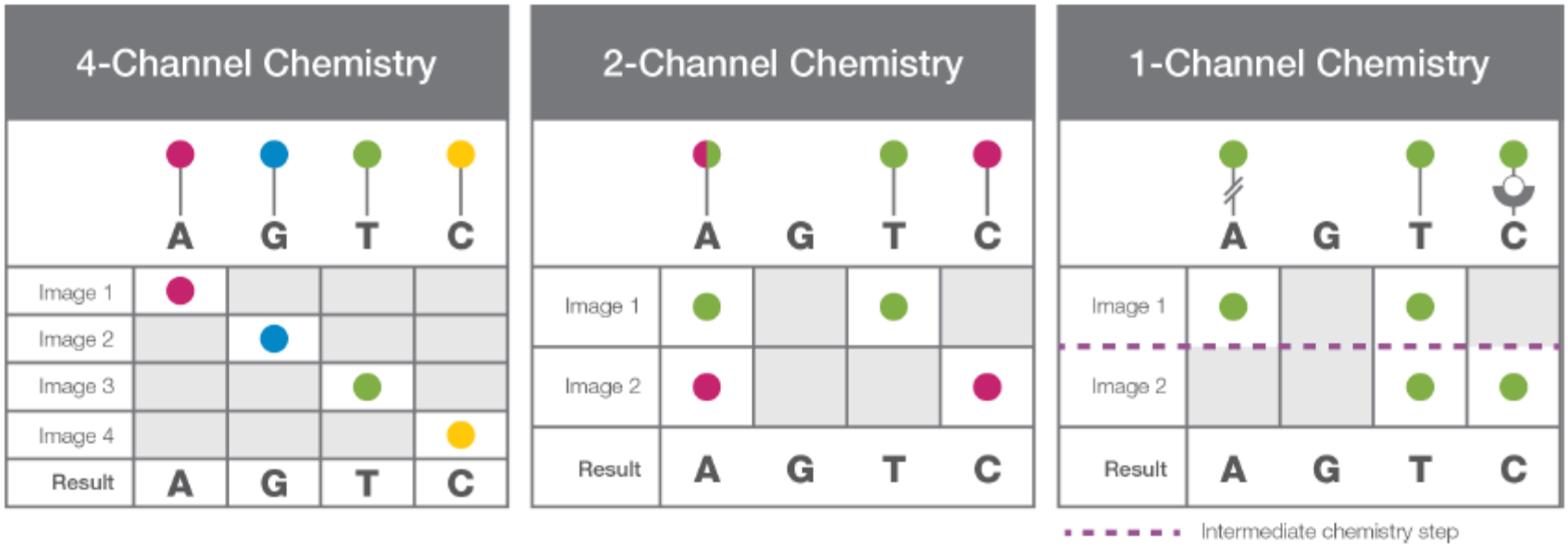


Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software to determine nucleotide identity.

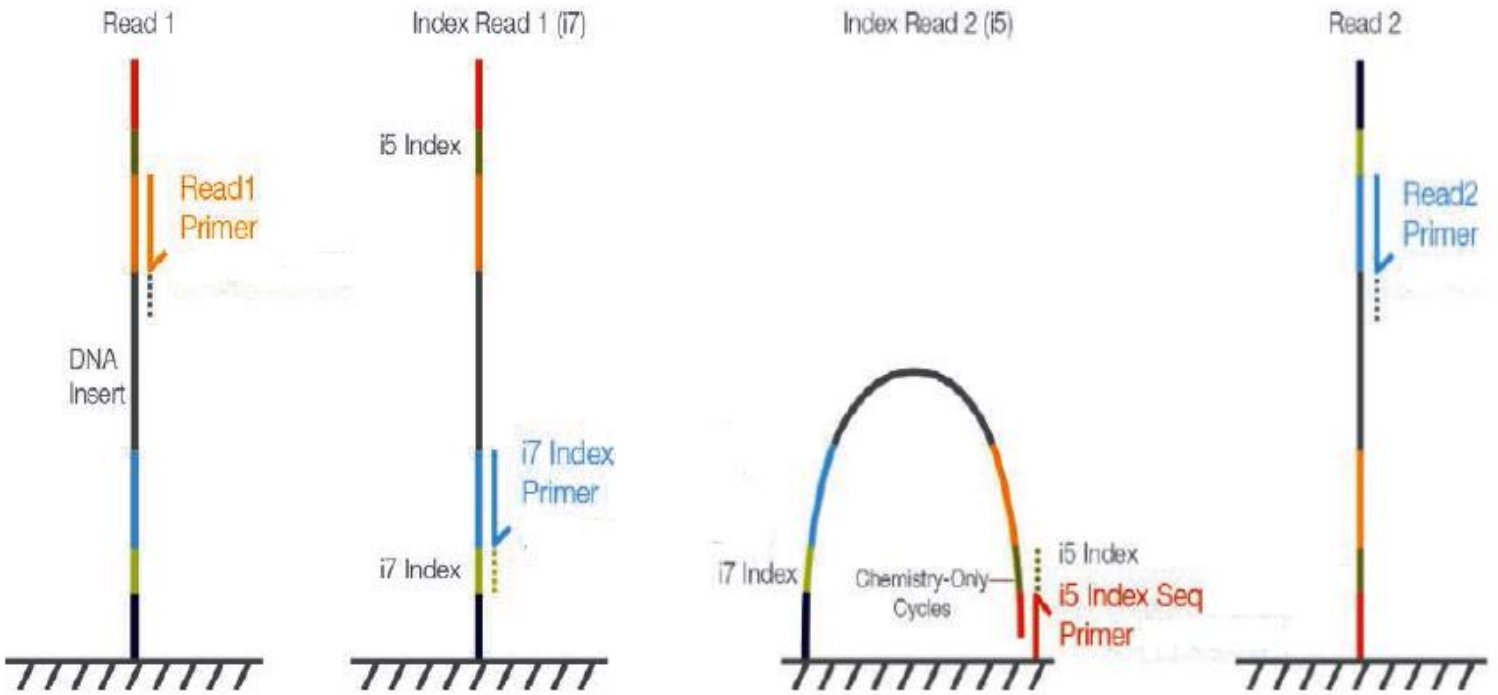
GAIIx, Miseq and HiSeq

NextSeq, Miniseq, Novaseq

iSeq

https://emea.illumina.com/content/dam/illumina-marketing/documents/products/techspotlights/cmos-tech-note-770-2013-054.pdf

Paired-end sequencing



AZ Sint-Jan Brugge

Illumina

iSeq 100

MiniSeq

MiSeq Series

NextSeq 550 Series

NextSeq 1000 & 2000

Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours	11-48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	330 Gb*
Maximum Reads Per Run	4 million	25 million	25 million †	400 million	1.1 billion*
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp

NextSeq 1000 & 2000

NovaSeq 6000 Series

NovaSeq X Series

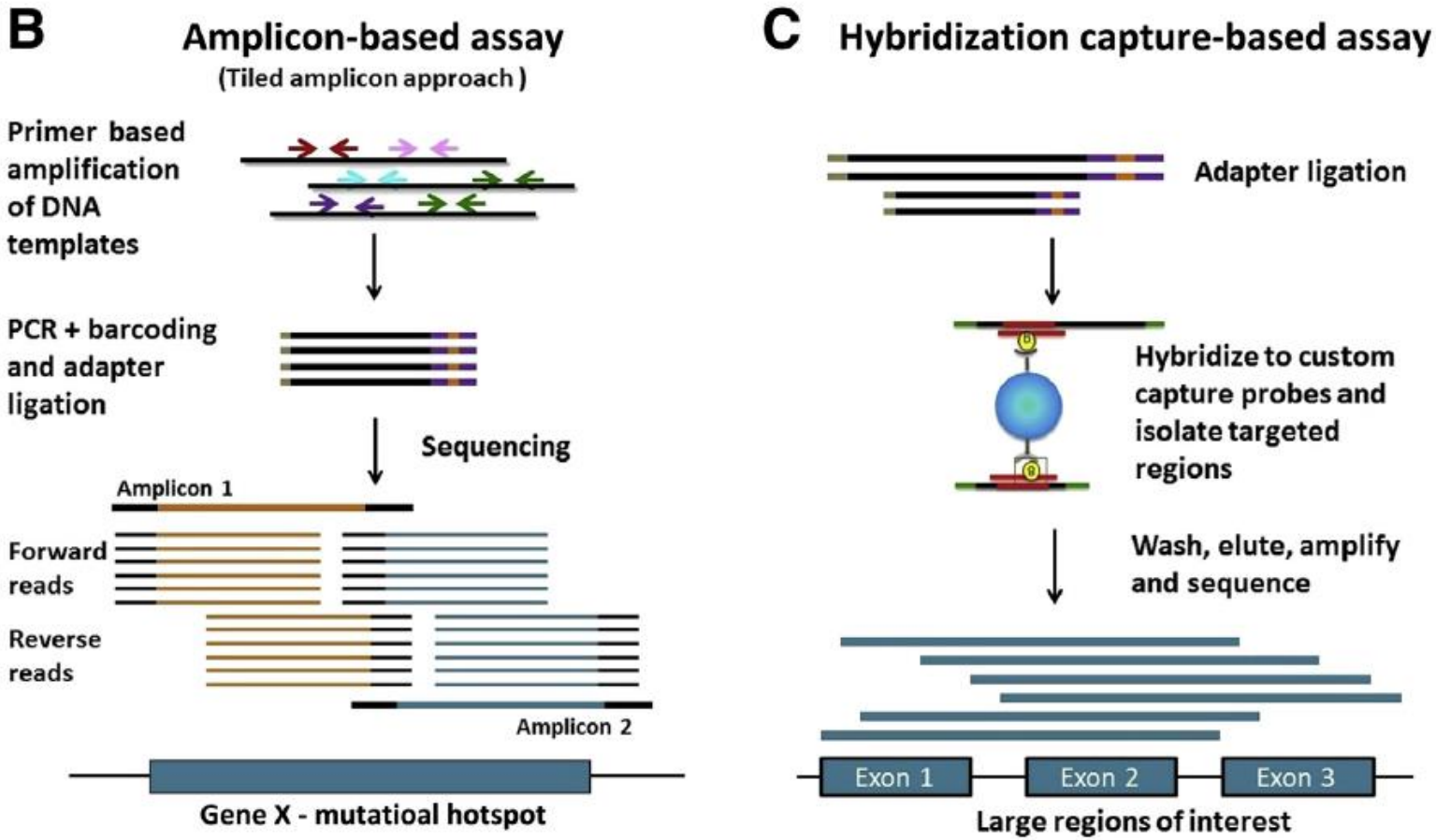
Run Time	11-48 hours	~13–38 hours (dual SP flow cells) ~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)	~13–21 hours (1.5B flow cells‡) ~18–24 hours (10B flow cells‡) ~48 hours (25B flow cells‡)
Maximum Output	360 Gb *	6000 Gb	16 Tb
Maximum Reads Per Run	1.2 billion *	20 billion	26 billion (single flow cells) 52 billion (dual flow cells)
Maximum Read Length	2 × 150 bp	2 x 250 bp**	2 × 150 bp

AZ Sint-Jan Brugge

AZ Sint-Jan Brugge

Library prep

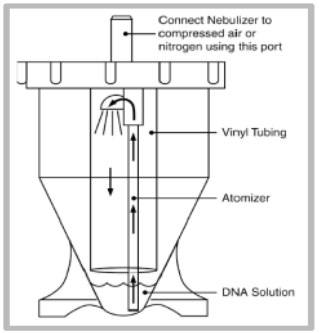
Library prep approaches

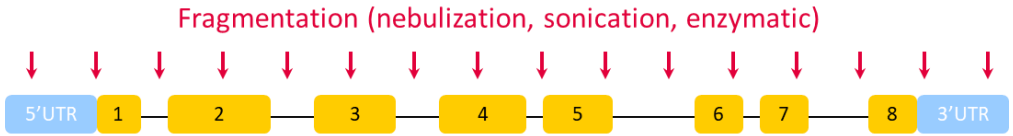


AZ Sint-Jan Brugge

Library prep - VOORBEELD

- Stap-1 : Fragmentatie
  - Nebulizer, goedkoop, OK voor 1-5 µg starting input
  - Andere methodes : Covaris™, Sonication, HydroShear®, Enzymatic!,…



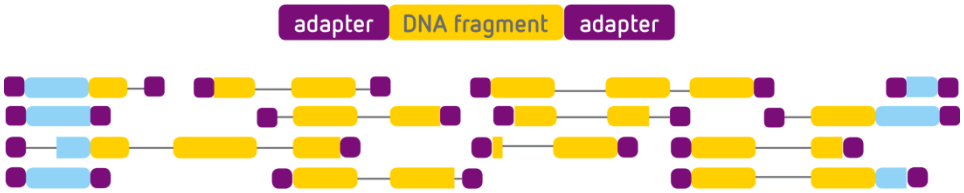


- Stap-2 : End-repair, 5' fosforylatie en dA-tailing
  - Creeërt blunt ends : dNTP, T4 DNA polymerase, Klenow polymerase
  - 5' end fosforylatie : kinase, ATP
  - Converts 3' A-overhangs: dNTP, T4 DNA polymerase, Klenow polymerase

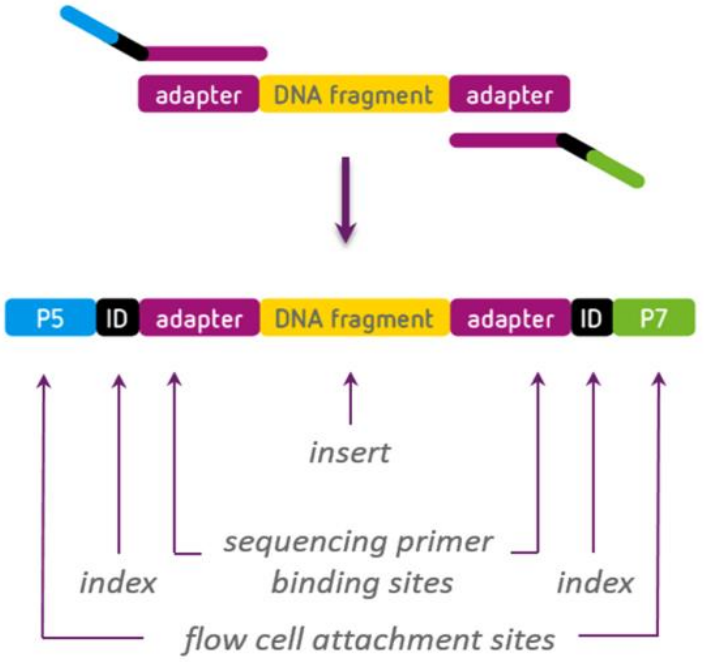
AZ Sint-Jan Brugge

Library prep - VOORBEELD

- Stap-3 : Adapter ligatie
  - DNA ligase
  - (10:1 molar ration of adaptor to insert DNA)



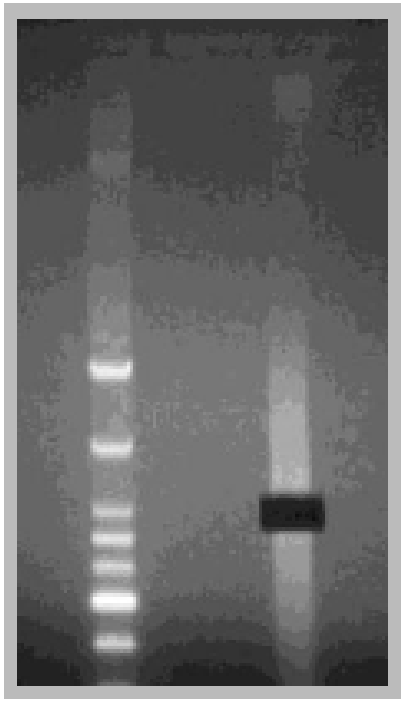
- Stap-4 : Index PCR
  - Single of dual barcoding



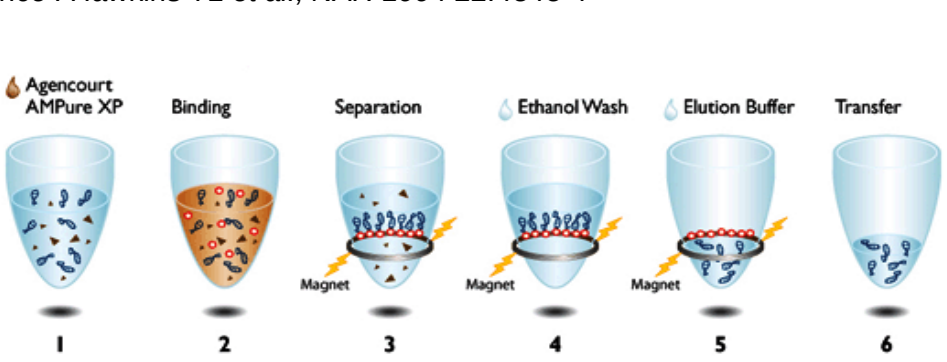
AZ Sint-Jan Brugge

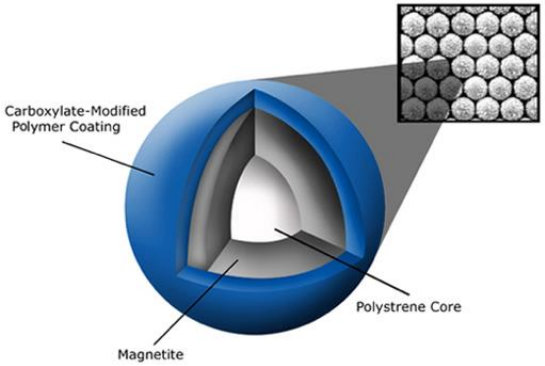
Size selection

- On gel :
- <u>SPRI beads (AMPure beads)</u>
  Solid Phase Reversible Immobilization
  (crowding agent Polyethylene glycol, PEG)
  1. PCR reactie
  2. Binding van de PCR amplicons aan de magnetische beads
  3. PCR amplicons gebonden aan de magnetische beads wordt gescheiden van de contaminanten
  4. Wassen van de PCR amplicons met ethanol
  5. Elutie van PCR amplicons van de magnetische beads
  6. Transfer van de PCR amplicons in een nieuwe reactietube.



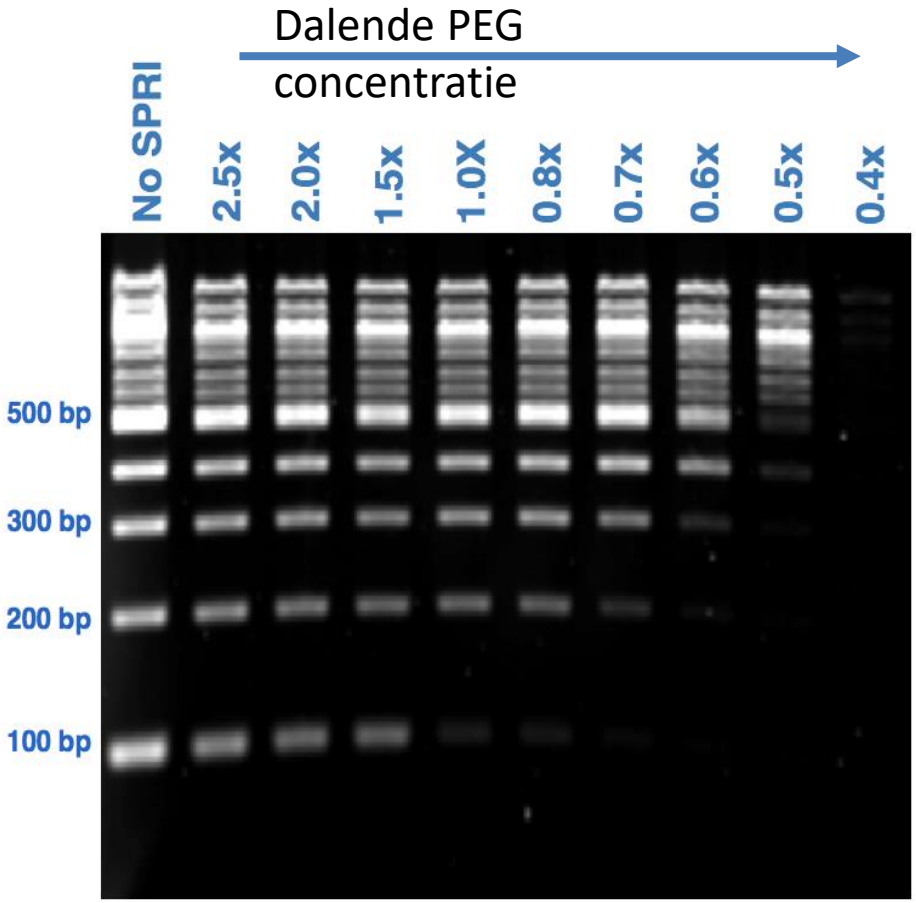
Reference : Hawkins TL et al., NAR 1994 22:4543-4





AZ Sint-Jan Brugge

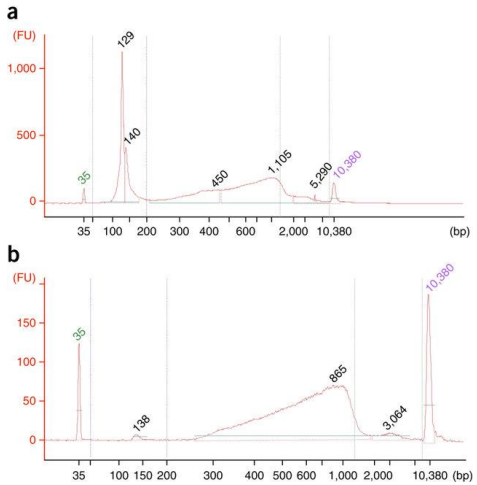
SPRI bead:DNA ratio



AZ Sint-Jan Brugge

Library zuiveren, poolen en sequencen

- Library clean-up en size selectie met 0.6x SPRI beads (bead:DNA ratio)
- kwantificatie van de stalen met Qubit fluorometer
- Pooling
- Library kwantificatie met Qubit fluoromet
- Eventueel lengte bepaling met Bio-Analyzer (electroforese)
- Verdunnen naar een geoptimaliseerde concentratie (bvb. 10 pM)
- Library denaturatie (met NaOH)
- Laden op de MiSeq



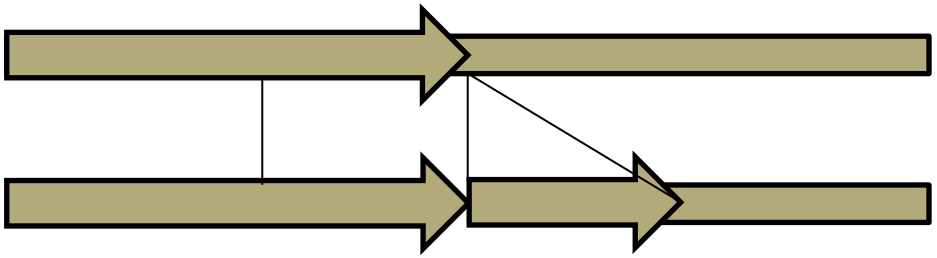
AZ Sint-Jan Brugge

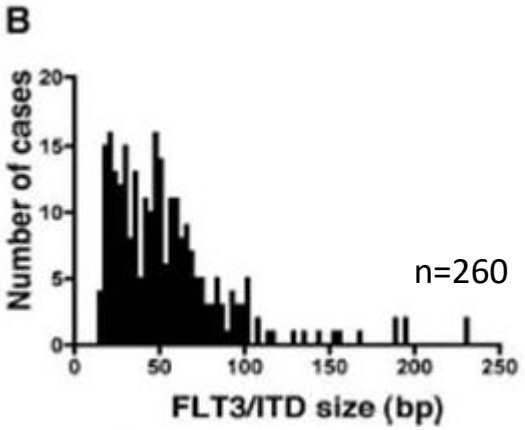
AZ Sint-Jan Brugge

Uitdagingen

Uitdaging-1 : moeilijke genen

- Hoog GC gehalte (hoge smelttemperatuur)
  - → Promotors van genen (TERT promotor,…)
  - → CEBPα : 75% GC-gehalte in coderende regio
- ASXL1 homopolymer 8 G nucleotide stretch : c.1934dupG (p.Gly646fs)
- FLT3-ITD (interne tandem duplicatie)





Gale et al., Blood 2008

Uitdaging-2 : klein of groot genen panel

Klein genen panel
- goedkoper
- Minder complexe bioinformatica
- Minder variant, eenvoudiger interpretatie
- hogere coverage

Groot genen panel
- Meer (alle) mutaties
- 'onverwachte' mutaties
- mutaties voor klinische trial of off-label behandelingsopties

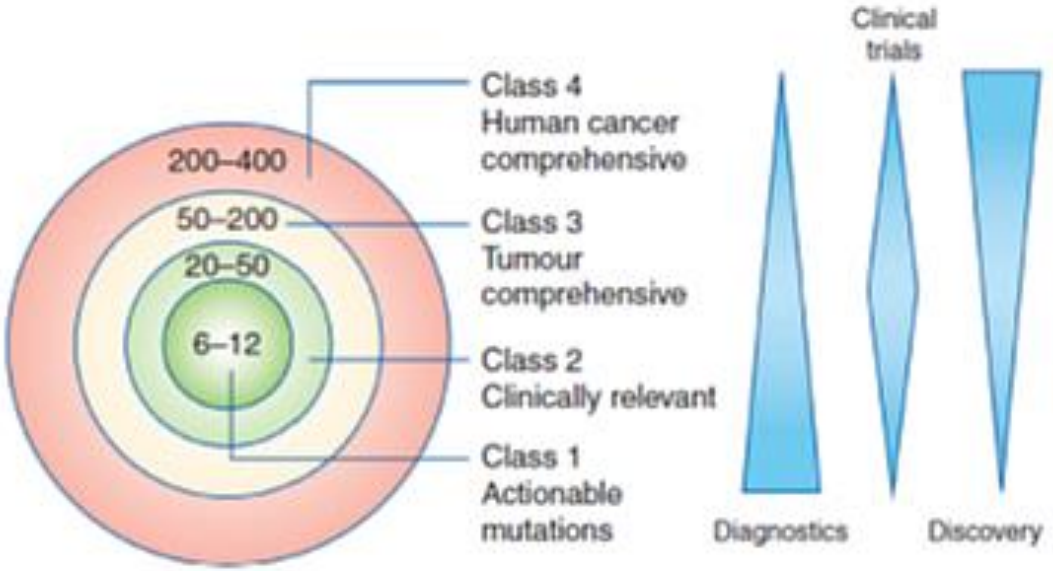
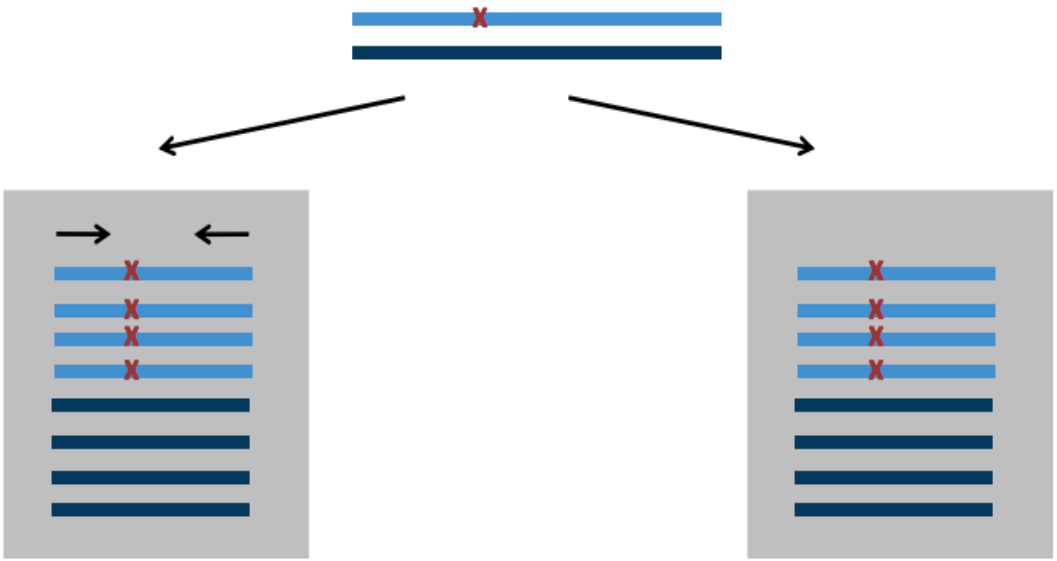


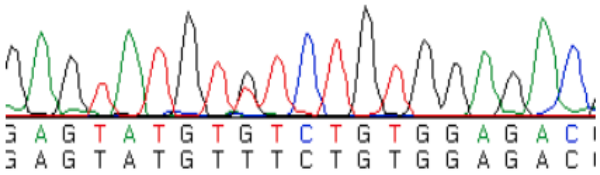
Figure 3. Choice of gene panel based on the paradigm being investigated by NGS.

Hynes et al., Brit J Cancer 2017

AZ Sint-Jan Brugge

Uitdaging-3 : gevoeligheid





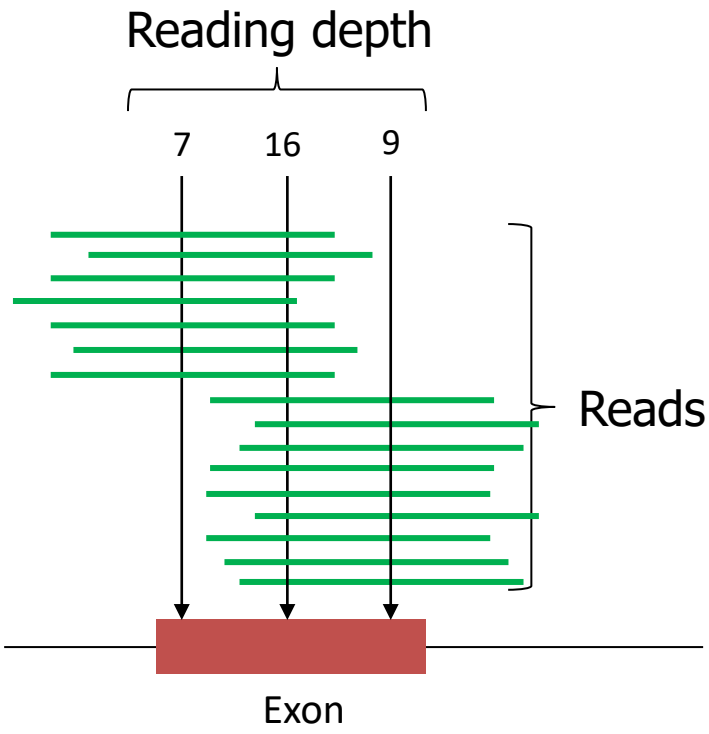
GAGTATGT**T**TCTGTGGAG
GAGTATGT**T**TCTGTGGAG
GAGTATGT**G**TCTGTGGAG
GAGTATGT**G**TCTGTGGAG
GAGTATGT**G**TCTGTGGAG

Sanger sequencing
Sensitiviteit ~20%

Next Generation Sequencing
sensitiviteit 2-5% (UMI's 0,5%)

AZ Sint-Jan Brugge

Uitdaging-3 : gevoeligheid



AZ Sint-Jan Brugge

Uitdaging-3 : gevoeligheid (errors in NGS)

Step	Sample →	DNA extraction →	Fragmentation →	PCR amplification →	Flow cell hybridization →	Cluster generation →	Sequencing by Synthesis
Error source	Mutagenesis	Oxidative damage	Oxidative damage	Polymerase mistakes		Cluster PCR errors	Phasing, Fluorophore crosstalk
Error category	Biological variants	DNA damage	DNA damage	PCR errors		Sequencing errors	Sequencing errors

DNA Polymerase	Substitution rate [a]	Accuracy [b]	Fidelity, rel. to *Taq* [c]	Total bases
Taq	1.5×10^{-4} ($\pm 0.2 \times 10^{-4}$)	6,456	1	98,396,789
Q5	5.3×10^{-7} ($\pm 0.9 \times 10^{-7}$)	1,870,763	280	112,619,228
Phusion	3.9×10^{-6} ($\pm 0.7 \times 10^{-6}$)	255,118	39	118,262,939
Deep Vent	4.0×10^{-6} ($\pm 2.0 \times 10^{-6}$)	251,129	44	106,217,940
Pfu	5.1×10^{-6} ($\pm 1.1 \times 10^{-6}$)	195,275	30	79,614,976
PrimeSTAR GXL	8.4×10^{-6} ($\pm 1.1 \times 10^{-6}$)	118,467	18	118,964,566
KOD	1.2×10^{-5} ($\pm 0.2 \times 10^{-5}$)	82,303	12	121,234,438
Kapa HiFi HotStart ReadyMix	1.6×10^{-5} ($\pm 0.3 \times 10^{-5}$)	63,323	9.4	101,742,963
Deep Vent (exo-)	5.0×10^{-4} ($\pm 0.1 \times 10^{-4}$)	2,020	0.3	60,218,605

[a] Reported error rates are per base per doubling as detailed in Materials and Methods. Standard deviations were determined based on sequencing several samples and are given here in brackets.

[b] Accuracy is calculated as 1 over substitution rate such that accuracy is a number of bases over which 1 substitution error is expected.

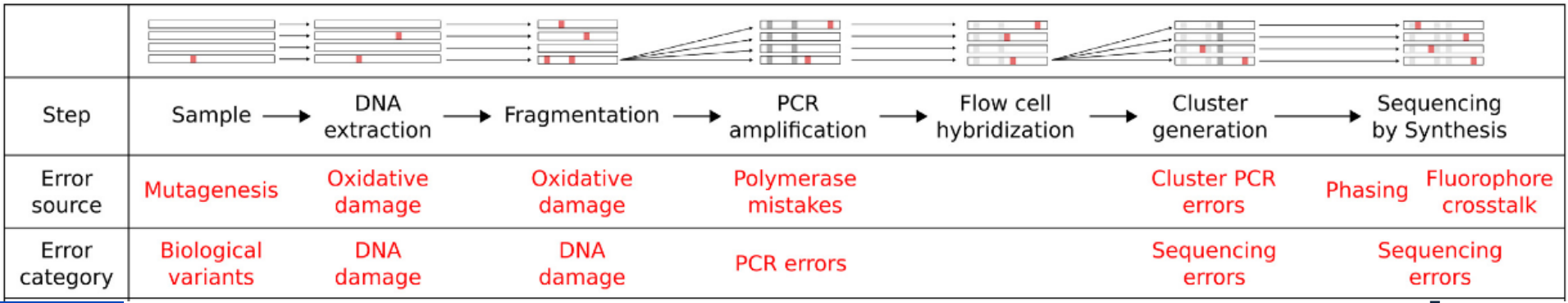
[c] Fidelity relative to *Taq* numbers are computed separately for each amplicon (LacZ-1, LacZ-2, DNA-1, DNA-2) and the average number is reported per DNA polymerase. Individual values are available in S1 Table.

*Potapov, V. & Ong, J. L. Examining Sources of Error in PCR by Single-Molecule Sequencing. Plos One **12**, e0169774 (2017).*

Stoler, N. et al. Nar Genom Bioinform (2021)

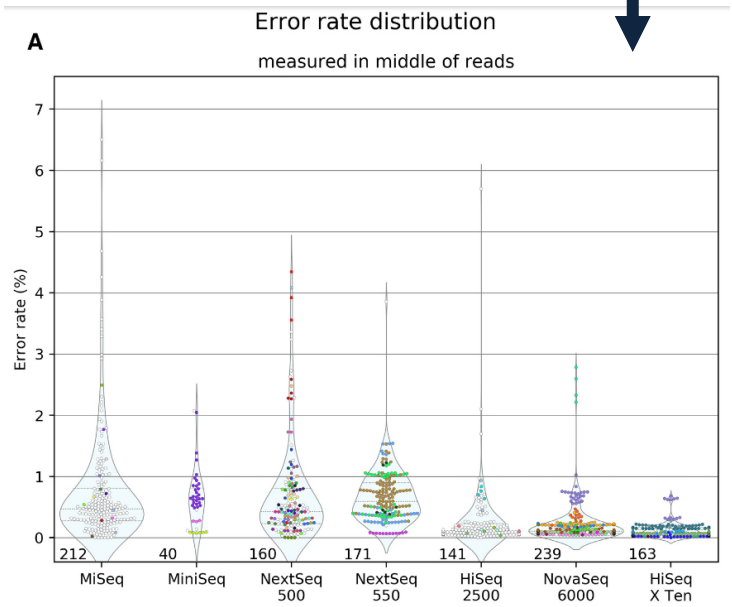
AZ Sint-Jan Brugge

Uitdaging-3 : gevoeligheid (errors in NGS)



Step	Sample →	DNA extraction →	Fragmentation →	PCR amplification →	Flow cell hybridization →	Cluster generation →	Sequencing by Synthesis
Error source	Mutagenesis	Oxidative damage	Oxidative damage	Polymerase mistakes		Cluster PCR errors	Phasing, Fluorophore crosstalk
Error category	Biological variants	DNA damage	DNA damage	PCR errors		Sequencing errors	Sequencing errors

- Based on paired-end sequence read data
- PhiX control in Miseq run error rate mean ~0,5%



Stoler, N. et al. Nar Genom Bioinform (2021)

AZ Sint-Jan Brugge

Massive parallel sequencing

- **Roche/454**
  - GS FLX+ System
  - GS Junior System
- **Illumina**
  - Genome Analyzer IIx
  - HiSeq, Nextse, MiSeq,
  - Novaseq

- Applied Biostems – Life Technologies
  - SOLiD 5500
- **Ion semi conductor sequencing** (ThermoFisher Scientific)
  - Gene Studio S5, Genexus, …

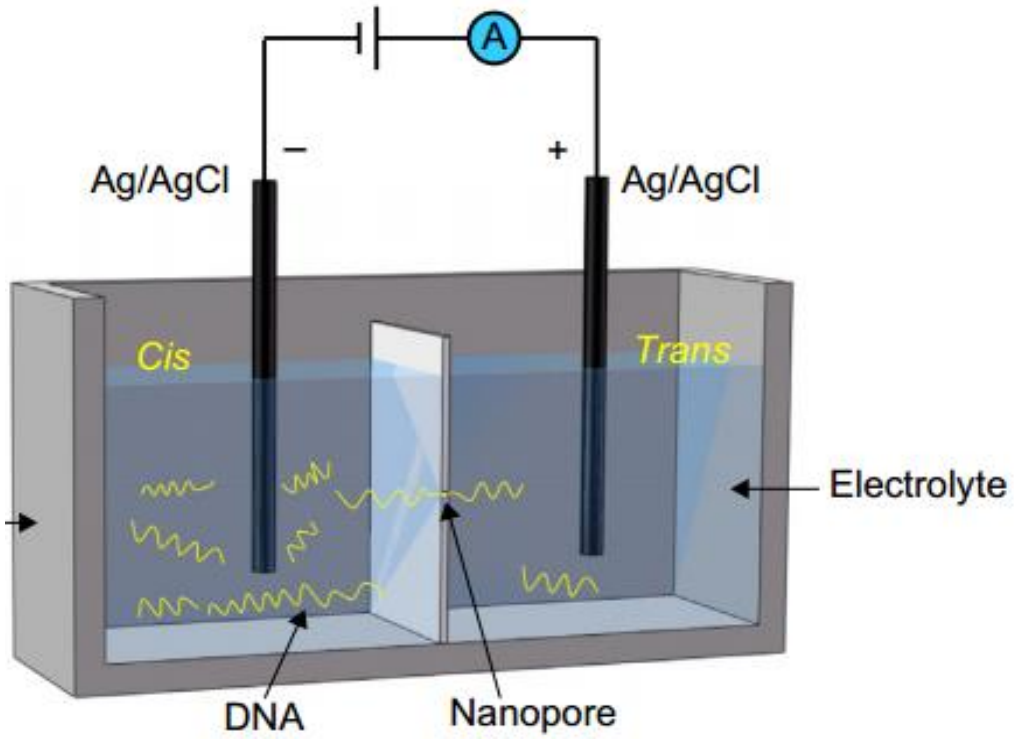
- Helicos
  - Helicos Genetic Analysis System
- **Single Molecule Real Time sequencing** (PacBio) - Sequel
- **Nanopore sequencing** (Oxford Nanopore)

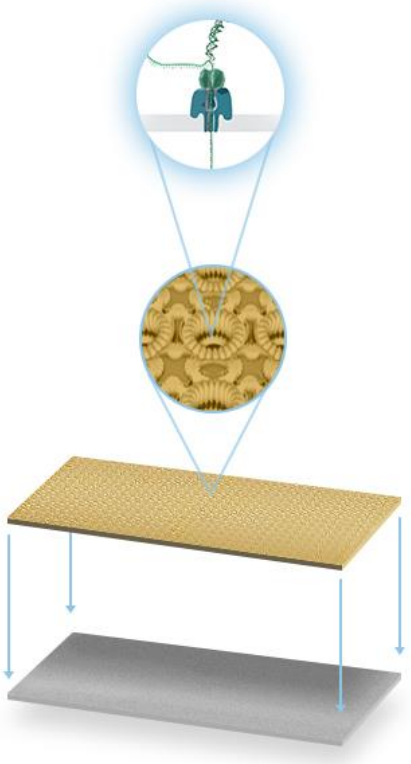
Next Generation Sequencing
Amplified Single Molecule Sequencing

Third Generation Sequencing
Next Next Generation Sequencing
Single Molecule Sequencing

AZ Sint-Jan Brugge

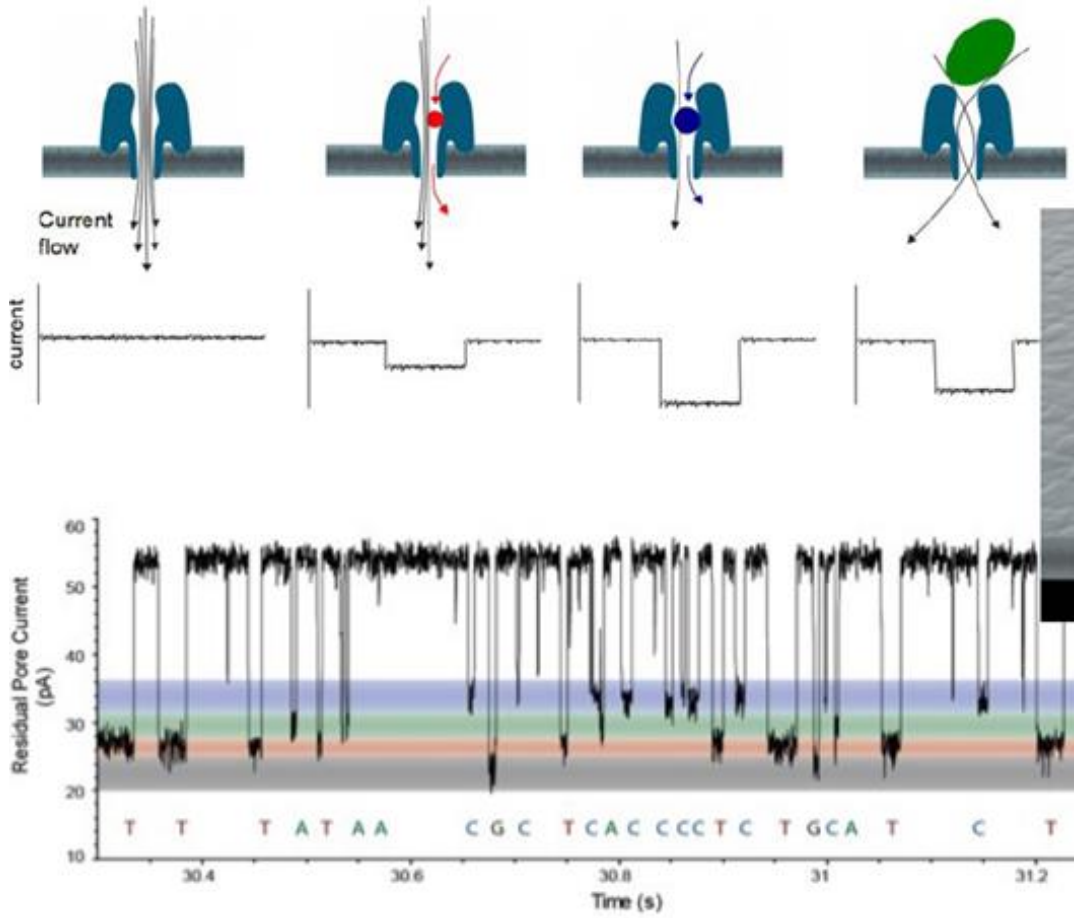
Nanopore sequencing

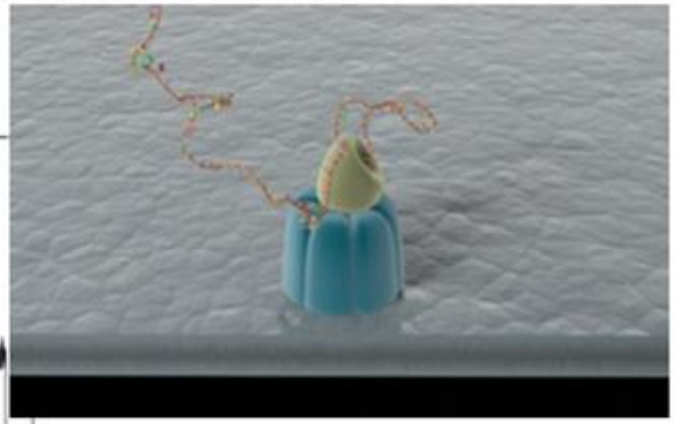




AZ Sint-Jan Brugge

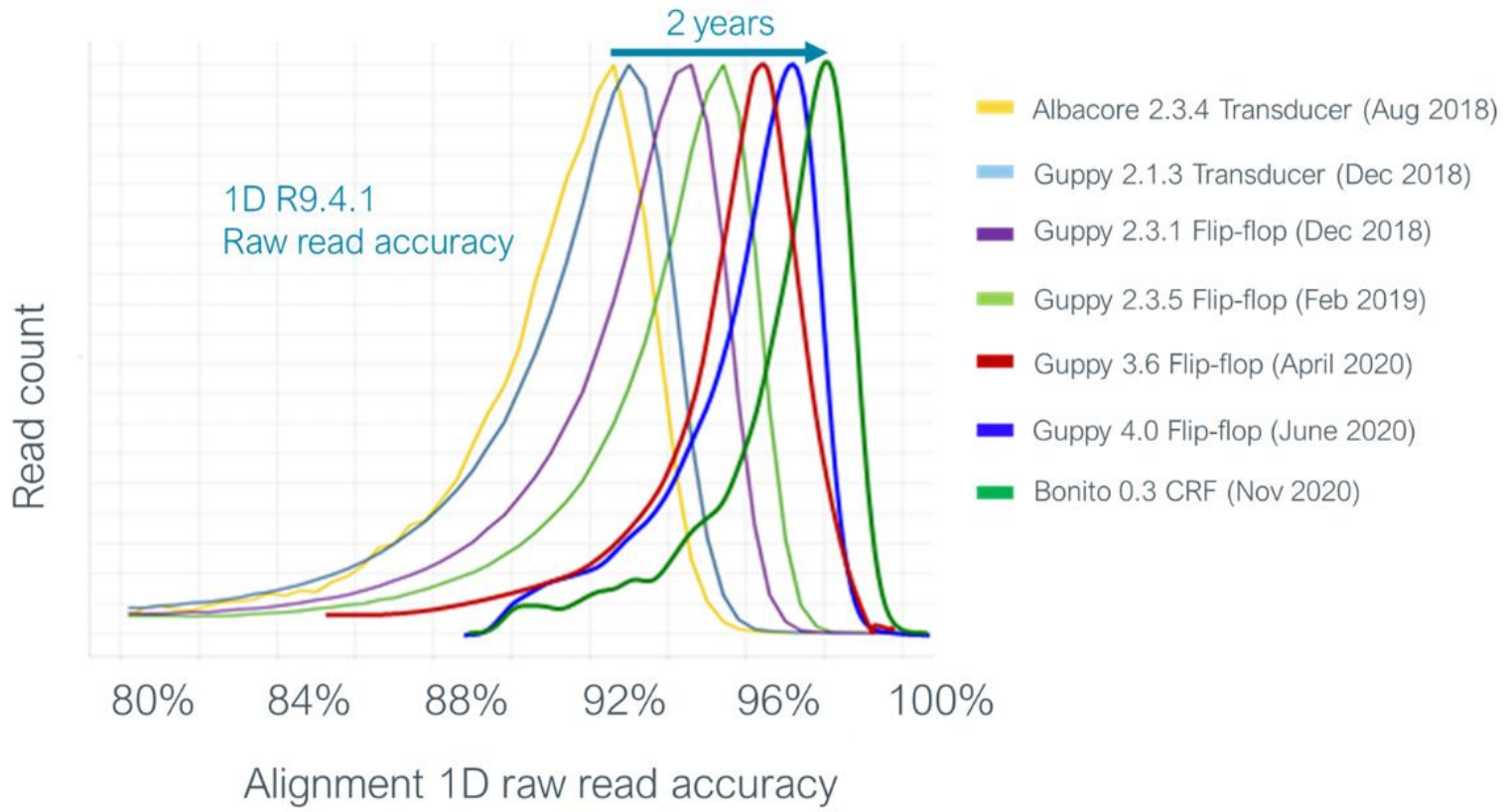
Nanopore sequencing





AZ Sint-Jan Brugge

Accuracy Nanopore sequencing



AZ Sint-Jan Brugge

AZ Sint-Jan Brugge

Toepassingen

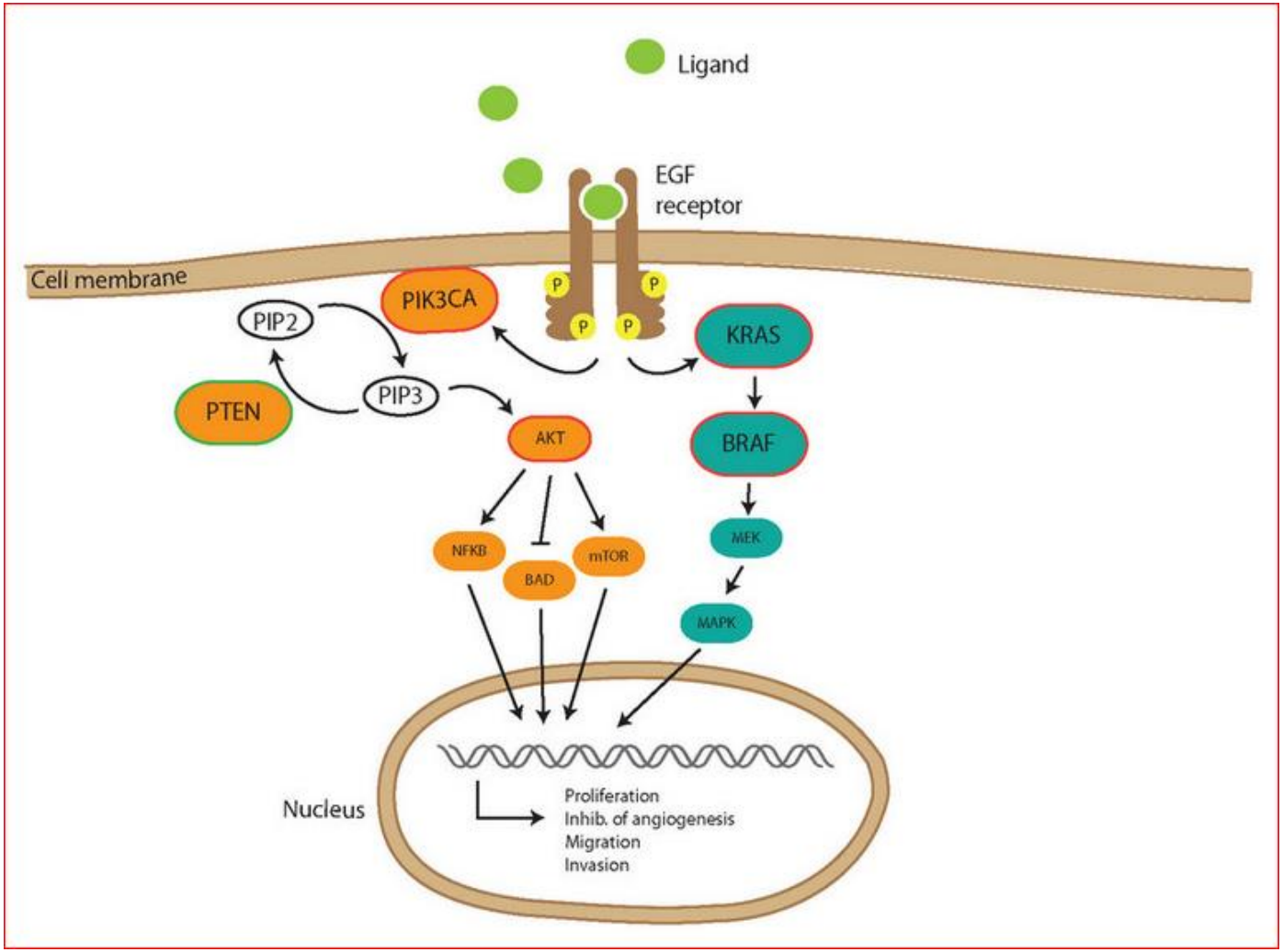
Opsporen van overerfbare en verworven mutatie

- Overerfbare mutaties (erfelijke ziektes)
- Verworven mutaties (sporadische kanker)

<u>Kanker cellen ontstaat na :</u>

- mutaties in proto-oncogenen:
  - groei stimulatie
  - één gemuteerd allel is voldoende
  - meestal door puntmutatie of chromosoomtranslocatie
- mutaties in tumor suppressor genen:
  - controleren of remmen de groei af
  - 2 allelen moeten geïnactiveerd worden
  - meestal door mutatie en/of deletie (Knudson two-hit model)

Oncogene *Epithelial Growth Factor pathway*



AZ Sint-Jan Brugge

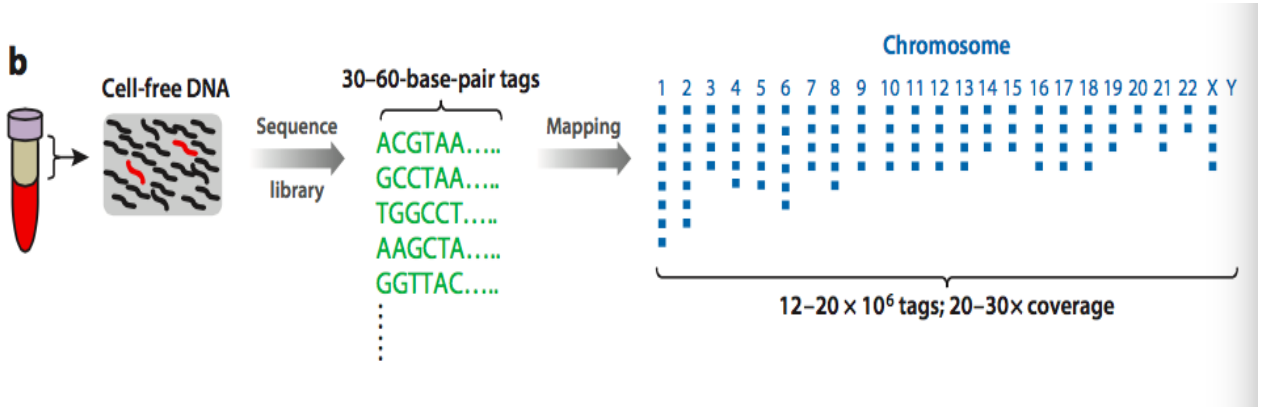
Gerichte therapie : EGF-R inhibitor

AZ Sint-Jan Brugge

NIPT (non invasive prenatal testing)

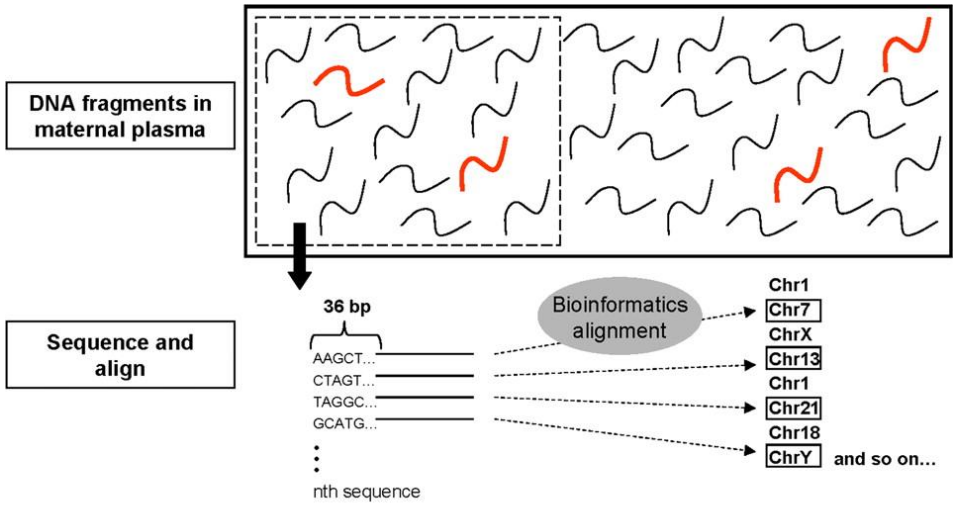
- Foetaal cfDNA
  - Afkomstig van apoptotische cellen van de placenta
  - Korte DNA fragmenten (± 150 bp)
  - Concentratie foetale cfDNA stijgt tijdens zwangerschapsduur
  - Foetale fractie ±10% bij 10 w zwangerschap
  - Verdwijnt 24 uur na de bevalling

Whole genome NIPT



AZ Sint-Jan Brugge

NIPT

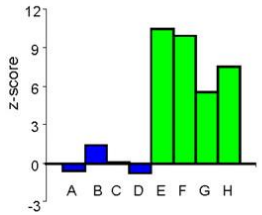


% representation of unique sequences mapped to a chromosome

$$\%\ \text{chrN} = \frac{\text{Unique count for chrN}}{\text{Total unique count}}$$

Disease status determination

$$\text{chrN z-score for test sample} = \frac{\%\ \text{chrN}_{\text{sample}} - \text{mean}\ \%\ \text{chrN}_{\text{reference}}}{\text{S.D.}\ \%\ \text{chrN}_{\text{reference}}}$$

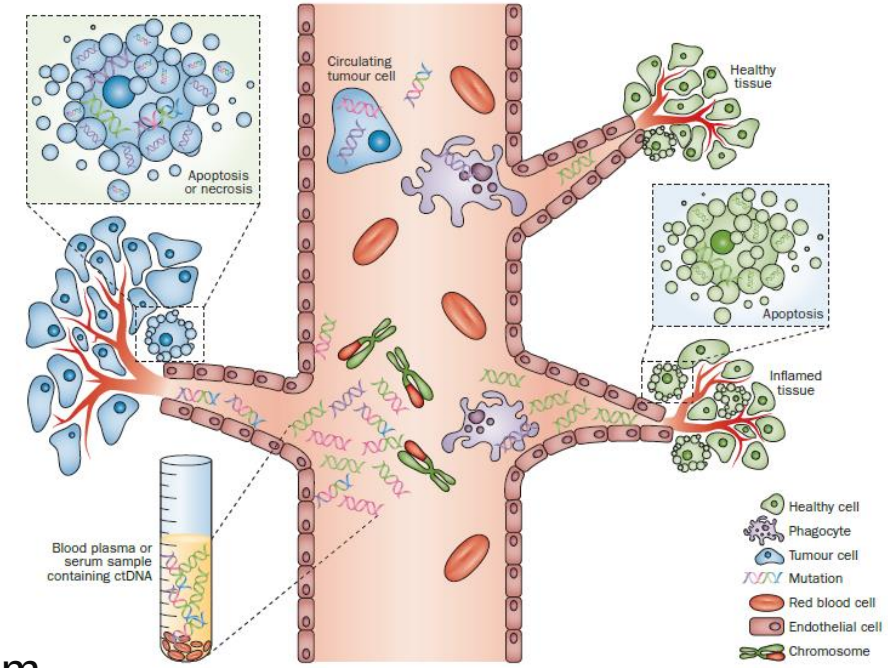


AZ Sint-Jan Brugge

Circulating cell-free DNA (cfDNA)

cfDNA ontstaat na apoptose

- gezonde weefsels,
- infecties
- sport
- orgaanfalen
- chirugie
- trauma



kanker : circulerend

cell-free tumorDNA (ctDNA)

- afhankelijk van tumor stadium
- bevat de kankermutaties

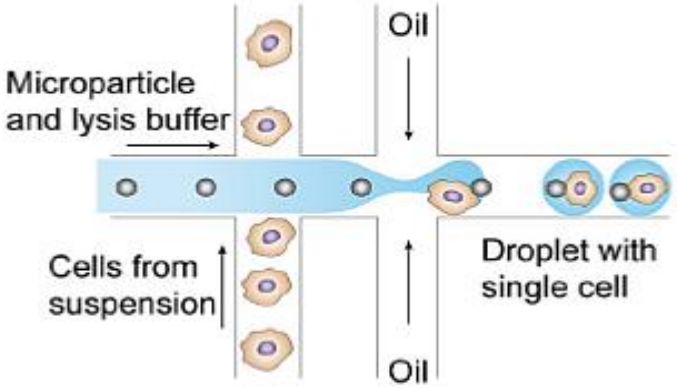
cfDNA = cfDNA (non-tumoraal) + ctDNA

AZ Sint-Jan Brugge

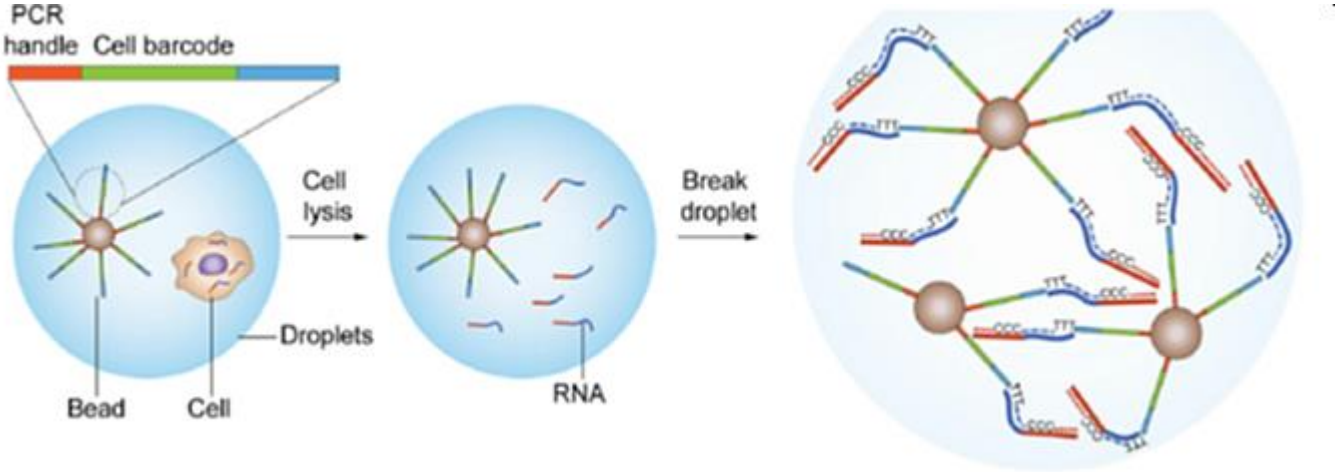
Whole genome sequencing

Single cell DNA en RNA sequencing

Microfluidic technologie voor single cell isolatie



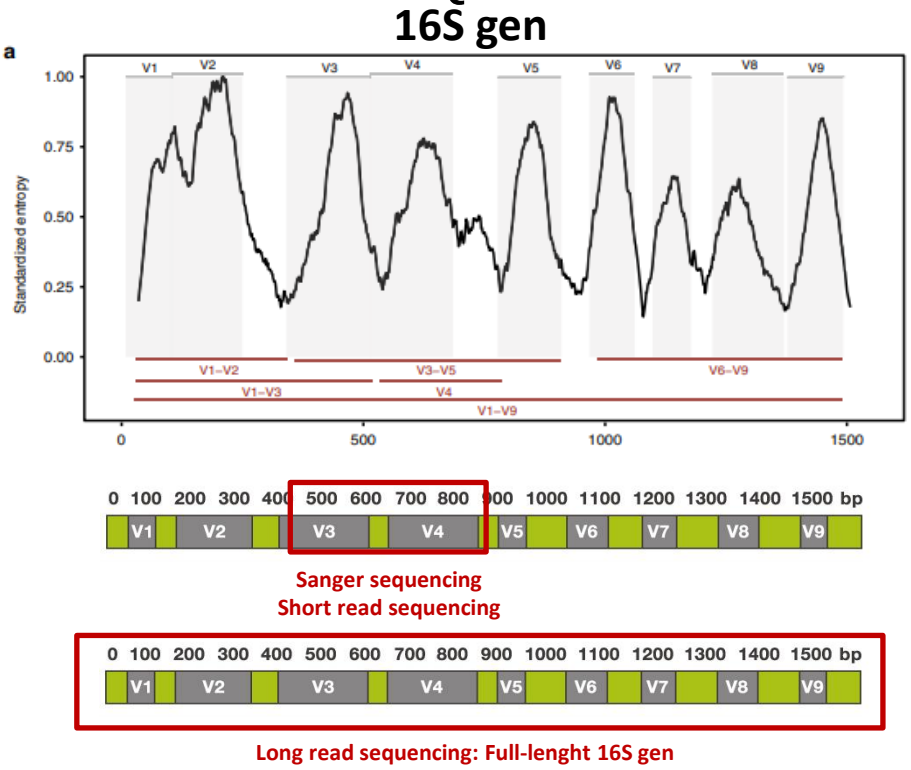
Droplet-gebaseerde library bereiding



AZ Sint-Jan Brugge

Identificatie

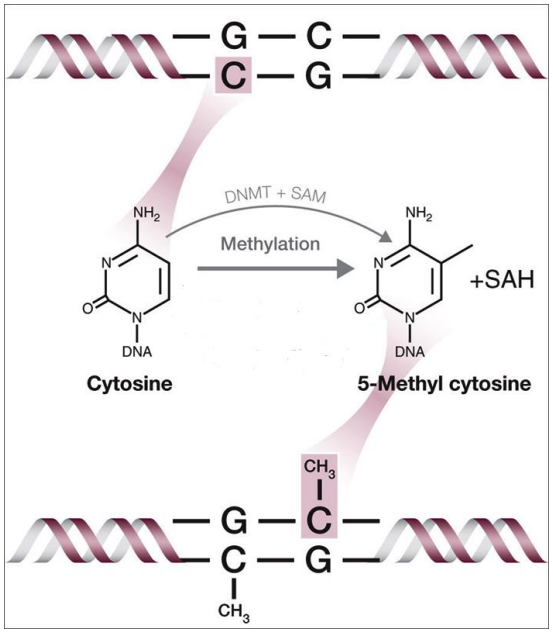
- BACTERIAL 16S RNA SEQUENCING

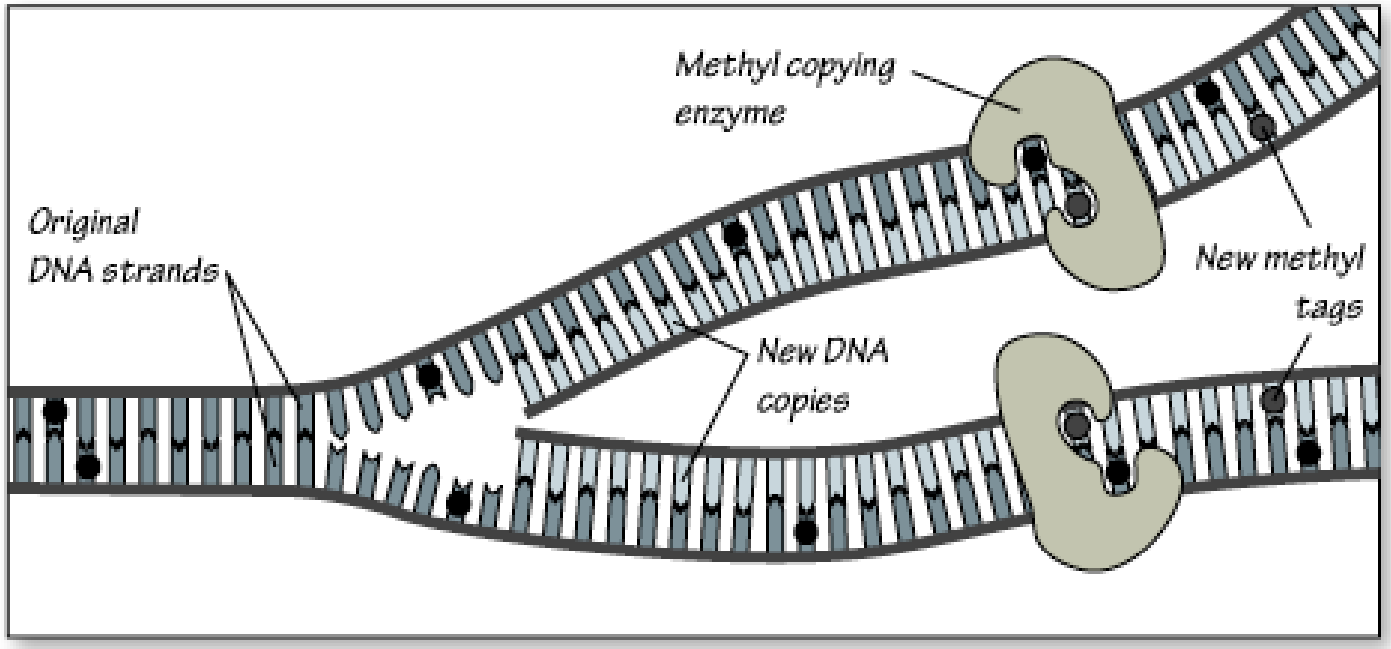


- EUKARYOTEN : MITOCHONDRIAL DNA SEQUENCING
- METAGENOMICS

AZ Sint-Jan Brugge

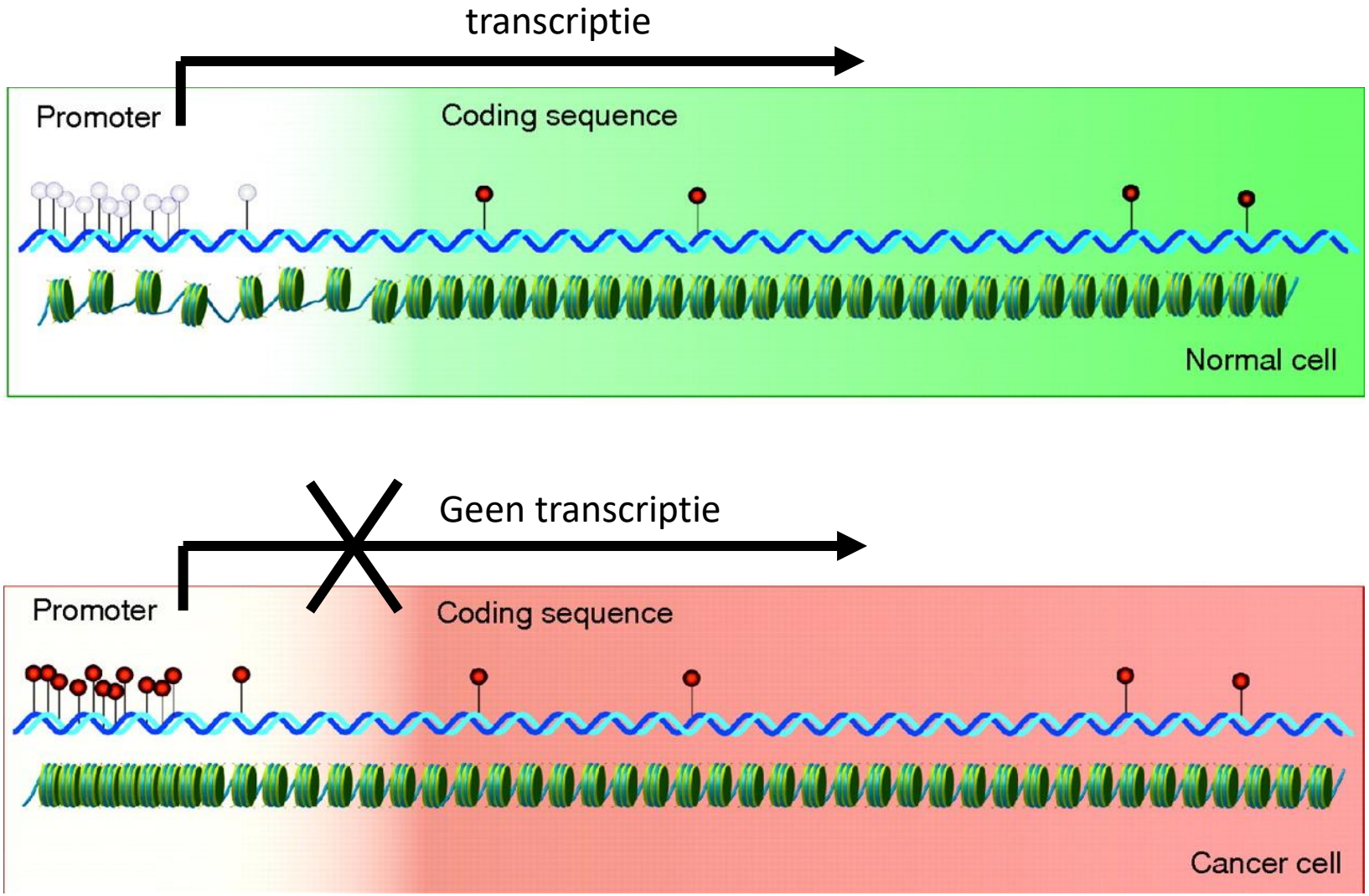
Epigenetica





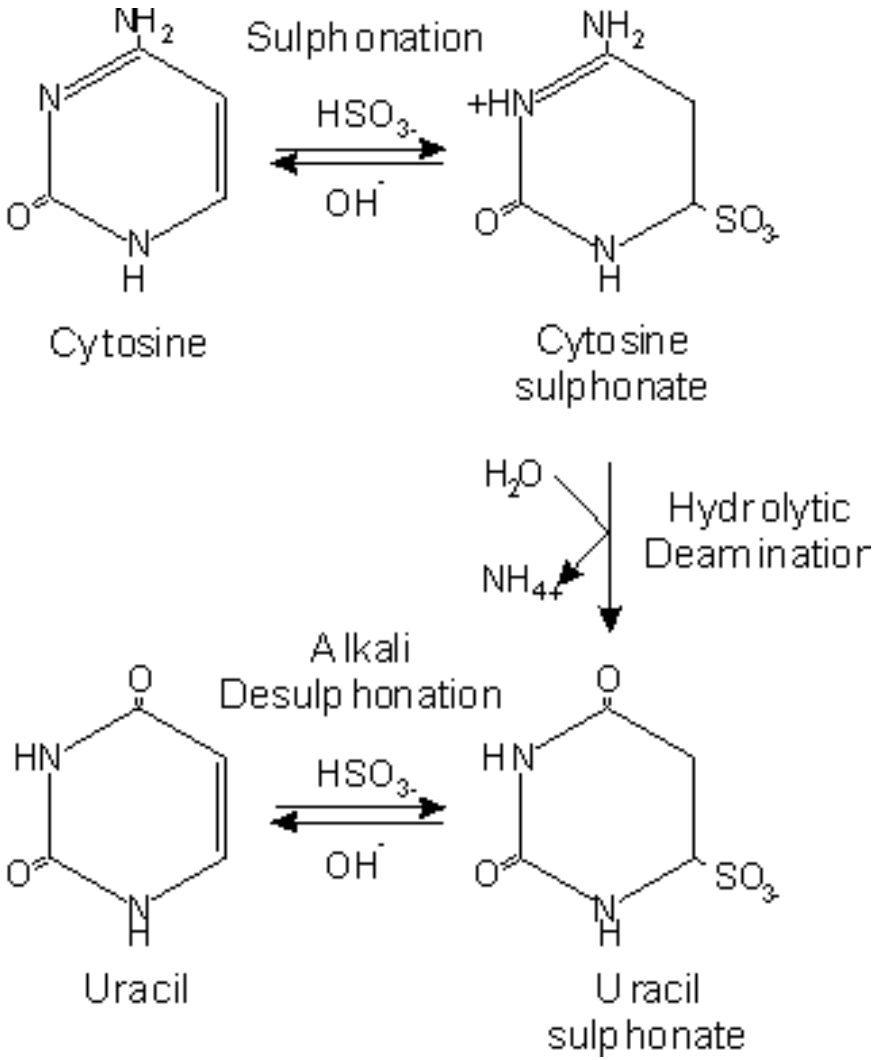
AZ Sint-Jan Brugge

CpG islands



AZ Sint-Jan Brugge

Detectie van DNA methylatie : Bisulfiet



AZ Sint-Jan Brugge