

ATBC - BIO-INFORMATICA - OPSGEM - GENETICA

BIO-INFORMATICA

**OPSPOREN VAN GENETISCHE MUTATIES BIJ ERFELIJKE
ZIEKTES**

GELELEKTROFORESE
SANGER SEQUENCING
GENSTRUCTUUR
GENETISCHE VARIANTEN

OVERZICHT BI2B BIO-INFORMATICA

Week	Onderwerpen theorieles	Onderwerpen werkcollege
1	Introductie, biologische databases	Stamboom analyse, zoeken in OMIM
2	PCR, primer design, primer eigenschappen	Zoeken in NCBI, gen naar mRNA sequentie, primer design
3	Gelelektroforese, Sanger sequencing, genstructuur, genetische varianten	Tools voor primer design en primer analyse, blast resultaten primer analyse, bekende mutaties op het kandidaatgen
4	Introductie + toepassingen next generation sequencing & third generation sequencing, genome assemblies, genome browsers	Primer design, zoeken naar RP genen in genome browsers
5	Genetische markers, mutaties & genetische varianten	Targeted next generation sequencing data analyse, weektaak doornemen
6	Onderzoek doornemen, oefenen met de stof	Exome sequencing data analyse
7	Oefentoets, vragen	Oefentoets thematoets, weektaak afmaken, vragen

DOELEN

Na deze week:

- Weet je hoe gelelektroforese werkt
- Kun je aan de hand van je gelelektroforese resultaat bepalen wat de lengte van je PCR fragment is
- Kun je beschrijven hoe Sanger sequencing werkt
- Kun je de structuur van een gen of mRNA bepalen aan de hand van RefSeq data
- Weet je waar je genetische varianten kan vinden, en weet je wat de gevonden informatie betekend

LEERMIDDELEN

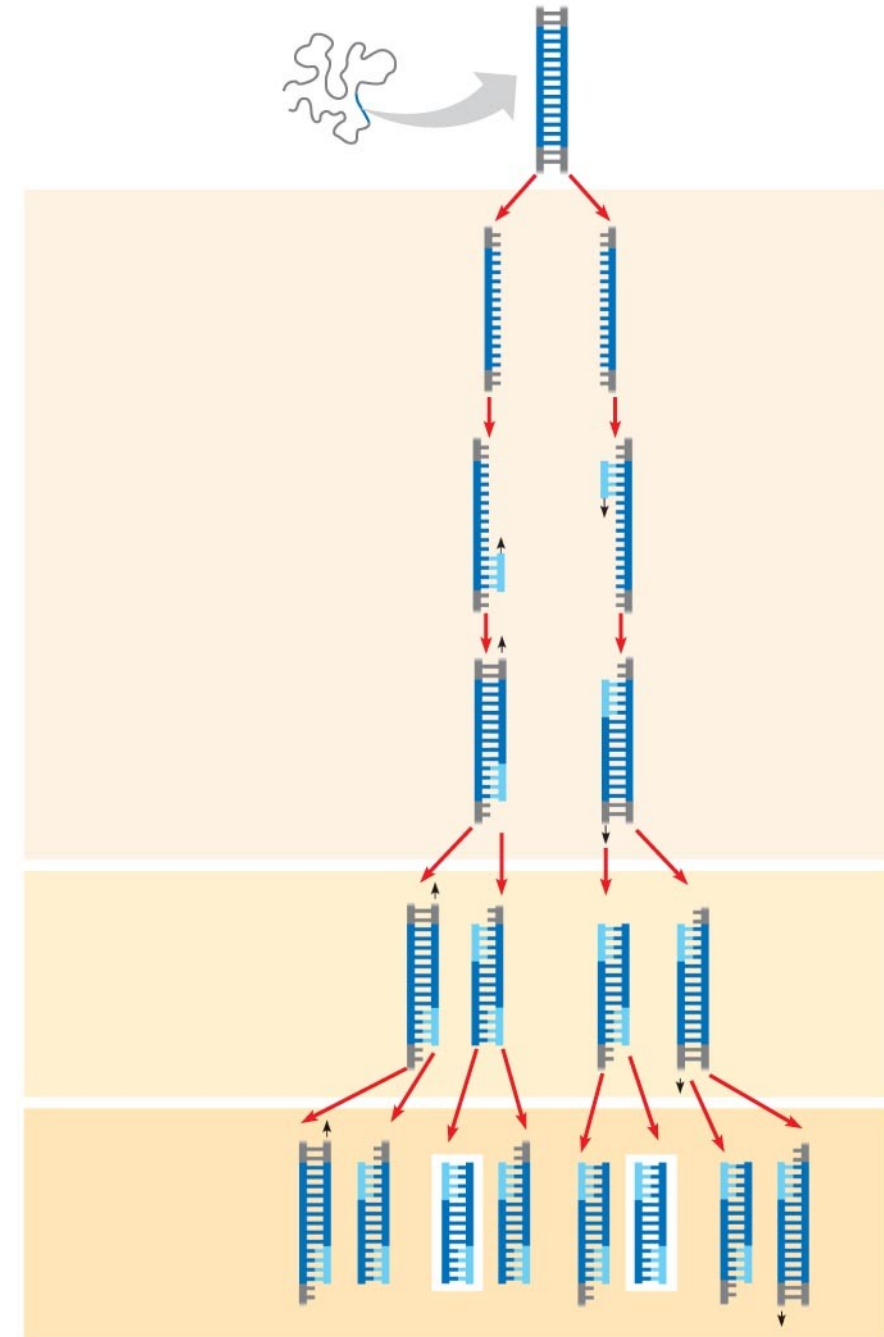
- Pevsner, *Bioinformatics and functional genomics*, third edition, Hoofdstuk 9, pagina 379-382, Hoofdstuk 21, pagina 1037-1042
- Campbell, *Biology*, twelfth edition, Paragraaf 19.1

AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

HERHALING WEEK 2

- Doel PCR: amplificeren van DNA
- Één PCR cyclus bestaat uit drie stappen: denaturatie, hybridisatie en elongatie
- Een PCR experiment bestaat uit meerdere cycli
- Primer design: forward en reverse primer
- Belangrijkst: primers moeten uniek zijn!
- Ontwerpen van primers:
 - Lengte
 - Hybridisatie- en smelt temperatuur
 - GC content
 - Secundaire structuren en repeats



AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

OPDRACHT

**Homo sapiens forkhead box P2 (FOXP2),
transcript variant 1, mRNA**

NCBI Reference Sequence: NM_014491.4

https://www.ncbi.nlm.nih.gov/nuccore/NM_014491.4

gene: 1..6543

CDS: 557..2704

Je wilt een mutatie analyse doen op het FOXP2 gen.
Dit gen heeft te maken met taalontwikkeling.

1. Ontwerp primers om dit gen te amplificeren.
Denk hierbij aan de juiste notatie.
2. Bereken de T_m van je primers.
3. Bereken de ΔT_m van je primers.
4. Bereken het GC percentage van je primers.
5. Hebben je primers een GC klem?

```
301 ttaattacca gcacaaaatg ccatcagctt gggacgtgat cgggcagagg tgtactcaca
361 gtagtgtaaa tactgctgta aatagttgtc tgatgggtggc tttgacagtg agctagcttc
421 tgagttttcc cttcttttta tactgttttc tgtgctggct ttttgaatc ttcctaattt
481 ttcattcttt taacaaactc ctatgaagtt gaaaccggga agtttgctct aacatttcca
541 gagaaggat taagtcatga tgcaggaatc tgcgacagag acaataagca acagttcaat
601 gaatcaaat ggaatgagca ctctaagcag ccaattagat gctggcagca gagatggaag
661 atcaagtggg gacaccagct ctgaagtaag cacagtagaa ctgctgcac tgcaacaaca
721 gcaggctctc caggcagcaa gacaacttct tttacagcag caaacaagt gattgaaatc
781 tcctaagagc agtgataaac agagaccact gcagggtgct gtgtcagtg ccatgatgac
841 tccccagggt atcacccttc agcaaatgca gcagatcctt cagcaacaag tctgtctcc
901 tcagcagcta caagcccttc tccaacaaca gcagggtgtc atgctgcagc agcaacaact
961 acaagagttt tacaagaaac agcaagagca gttacatctt cagcttttgc agcagcagca
1021 gcaacagcag cagcagcaac aacagcagca acaacagcag cagcaacaac aacaacaaca
1081 gcagcaaca cagcagcagc agcagcaaca gcagcagcag cagcaacagc atcctggaaa
1141 gcaagcgaag gagcagcagc agcagcagca gcagcaacag caattggcag ccagcagcct
1201 tgtcttcag cagcagcttc tccagatgca acaactccag cagcagcagc atctgctcag
1261 ctttcagcgt cagggactca tctccattcc acctggccag gcagcacttc ctgtccaatc
1321 gctgcctcaa gctggcttaa gtctgtctga gattcagcag ttatggaaag aagtgcattg
1381 agttcacagt atggaagaca atggcattaa acatggaggg cttagacctc ctactaaca
1441 ttcctcctcg actacctct ccaacacttc caaagcatca ccaccaataa ctcatcattc
1501 catagtgaat ggacagtctt cagttctaag tgcaagacga gacagctcgt cacatgagga
1561 gactggggcc tctcacactc tctatggcca tggagtttgc aaatggccag gctgtgaaag
1621 catttgtgaa gattttggac agtttttaaa gcaccttaac aatgaacacg cattggatga
1681 ccgaagcact gctcagtgtc gagtgcaaat gcagggtgtg caacagttag aaatacagct
1741 ttctaaagaa cgcgaacgtc ttcaagcaat gatgaccac ttgcacatgc gacctcaga
1801 gcccacaaca tctcccaaac ctctaaatct ggtgtctagt gtcaccatgt cgaagaatat
1861 gttggagaca tccccacaga gcttacctca aaccctacc acaccaacgg cccagtcac
1921 cccgattacc cagggaccct cagtaatcac cccagccagt gtgccaatg tgggagccat
1981 acgaaggcga cattcagaca aatacaacat tccatgtca tcagaattg ccccaacta
2041 tgaattttat aaaaatgcag atgtcagacc tccatttact tatgcaactc tcataaggca
2101 ggctatcatg gagtcatctg acaggcagtt aacacttaac gaaatttaca gctggtttac
2161 acggacattt gcttacttca ggcgtaatgc agcaacttgg aagaatgcag tacgtcataa
2221 tcttagcctg cacaagtgtt ttgttcgagt agaaaatgtt aaaggagcag tatggactgt
2281 ggatgaagta gaataccaga agcgaaggct acaaaaagata acaggaagtc caaccttagt
2341 aaaaaatata cctaccagtt taggctatgg agcagctctt aatgccagtt tgaggctgc
2401 cttggcagag agcagtttac ctttgctaag taatcctgga ctgataaata atgcatccag
2461 tggcctactg caggccgtcc acgaagacct caatgggtct ctggatcaca ttgacagcaa
2521 tggaacagat agtccgggct gctcacctca gccgcacata cattcaatcc acgtcaagga
2581 agagccagtg attgcagagg atgaagactg cccaatgtcc ttagtgacaa cagctaatac
2641 cagtcacagaa ttagaagacg acagagagat tgaagaagag cctttatctg aagatctgga
2701 atgagaactg acttgtgaaa cctcagcgtg aaggacata tctactgacct tcataaccac
2761 tccacaacca tgaatatgtt acaaatgtt actgtgacta tttattaagc atggataaag
2821 gagacagccc taaaggaaat tactaagcca gccctttggg attcagtacc aacaggcaaa
2881 ttgcttgttt tcttcttctt cttcttcttt tttttttttt ttttagaaaa aaagacaaaa
```

OPDRACHT

Sequentie gen E (NCBI accession number: NC_004718):

ATGTACTCATTTCGTTTCGGAAGAGA **CAGGTACGTTAATAGTTAATAGCGT** ACTTCTTTTTCTTGCTTTCGTGGT
ATTCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATT **GTGTGCGTACTGCTGCAATAT** TGTTAACGTGA
GTCTTGTAACCTTCTTTTTACGTTTACTCTCGTGTTAAAAATCTGAATTCTTCTAGAGTTCCTGATCTTCTGG
TCTAA

Primers ontworpen op de geel gearceerde stukken.

Opdracht:

1. Wat is de sequentie notatie van de forward en reverse primer voor gen E?
2. Wat is de lengte van het PCR product?

AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

PETRA DIAGNOSTICEREN: WELKE ZIEKTE, WELK GEN, WELKE MUTATIE



- Stamboom analyse
- Diagnose erfelijke ziekte Petra

- Kandidaat gen kiezen
- Primer design (handmatig) voor amplificatie van kandidaat gen

- Primer design (m.b.v. tools) voor amplificatie van kandidaat gen
- Uitzoeken van bekende mutaties in het kandidaatgen

Onderzoek naar één enkel kandidaat gen

AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

GELELEKTROFORESE

- Scheidingstechniek
- Agarose gel wordt gebruikt als zeef om moleculen (DNA, RNA, eiwitten) te scheiden op grootte



<https://www.addgene.org/protocols/gel-electrophoresis/>, 19-11-2021

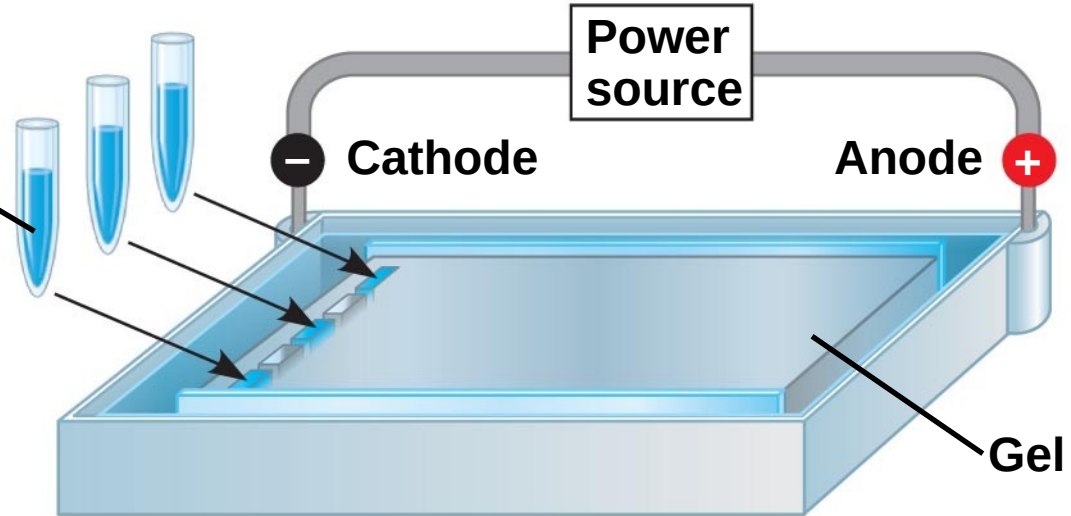
GELELEKTROFORESE

- Elektrisch veld wordt toegepast → geladen moleculen gaan door de gel bewegen
 - pool bevindt zich bovenin de gel
 - + pool onderin de gel
- Negatief geladen moleculen bewegen naar de + pool
 - Hoe kleiner het molecuul, hoe sneller
 - Hoe groter het molecuul, hoe langzamer

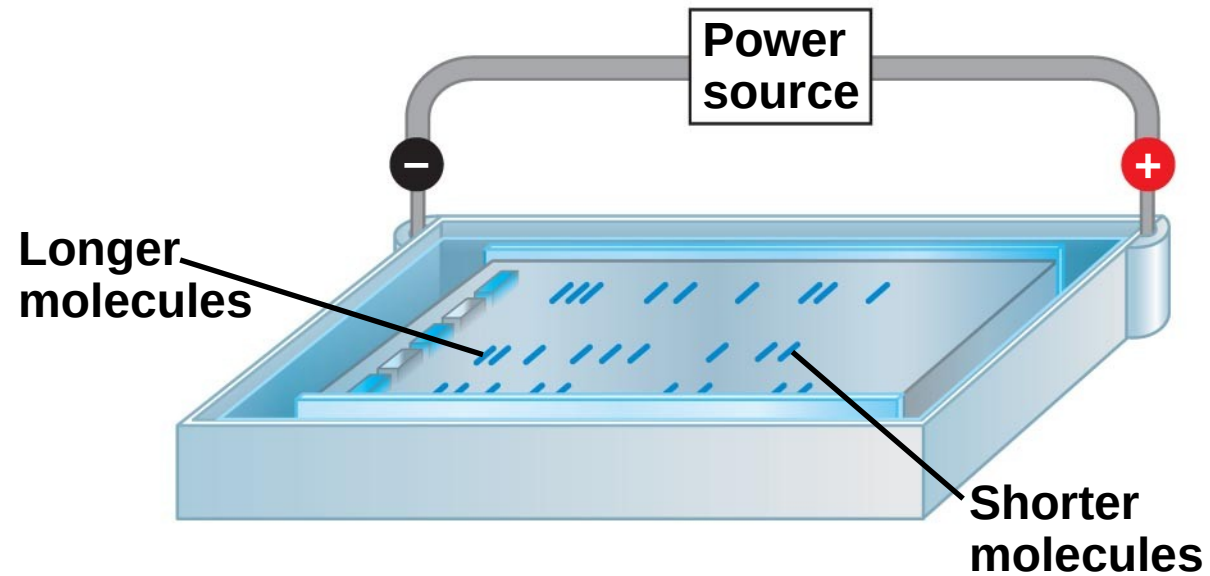
TECHNIQUE

Mixture of
DNA mol-
ecules of
different
sizes

1

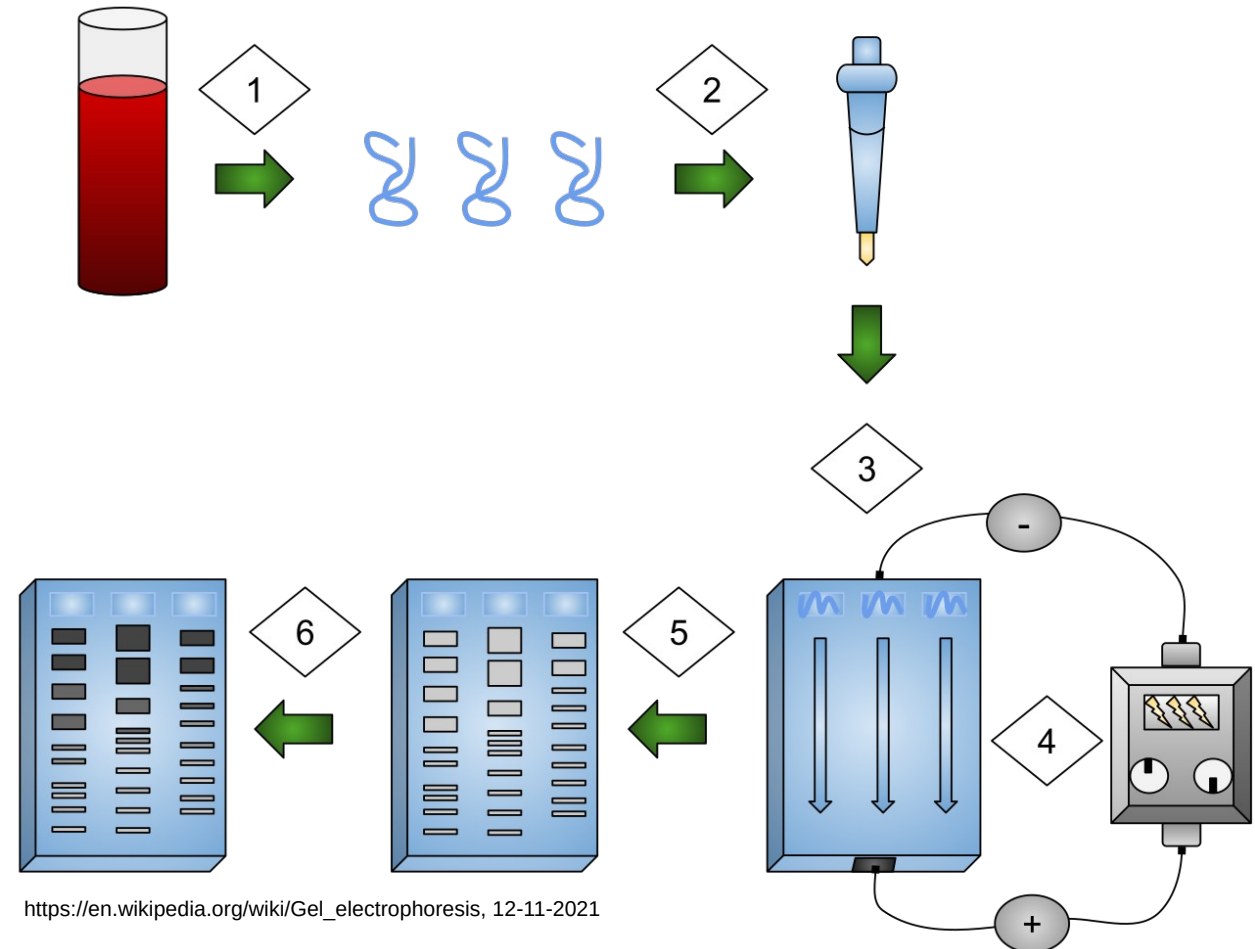


2



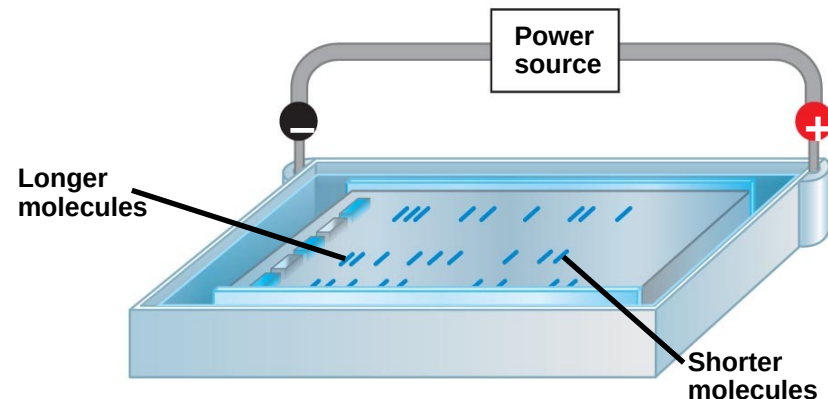
STAPPEN GELELEKTROFORESE

1. DNA wordt geïsoleerd uit sample
2. DNA wordt geamplificeerd m.b.v. PCR
3. DNA wordt in de welletjes van de gel gepipeteerd
4. Op gel wordt elektrisch veld gezet
5. DNA wordt gescheiden op grootte
6. DNA bandjes worden gekleurd



BEPALEN GROOTTE DNA FRAGMENTEN

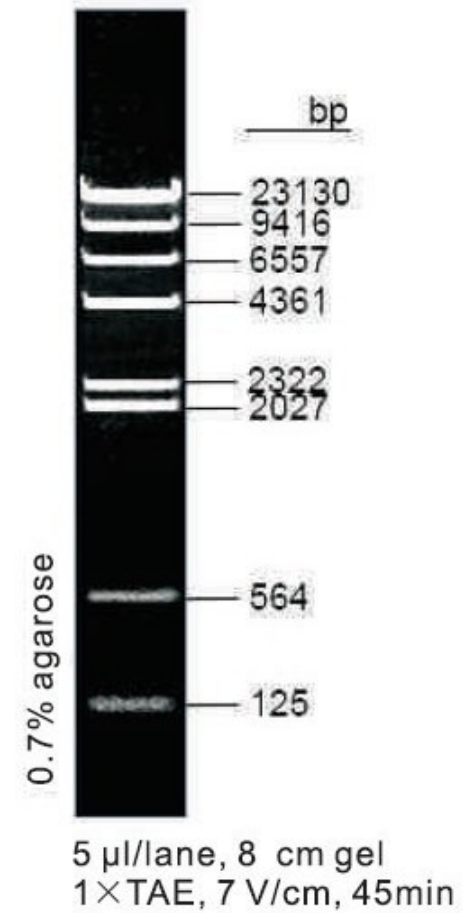
- Een DNA ladder wordt toegevoegd om de grootte van de DNA moleculen te kunnen bepalen
- DNA wordt gekleurd om het te kunnen visualizeren
 - Bijv. met SBR Safe + blauw/UV licht



Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

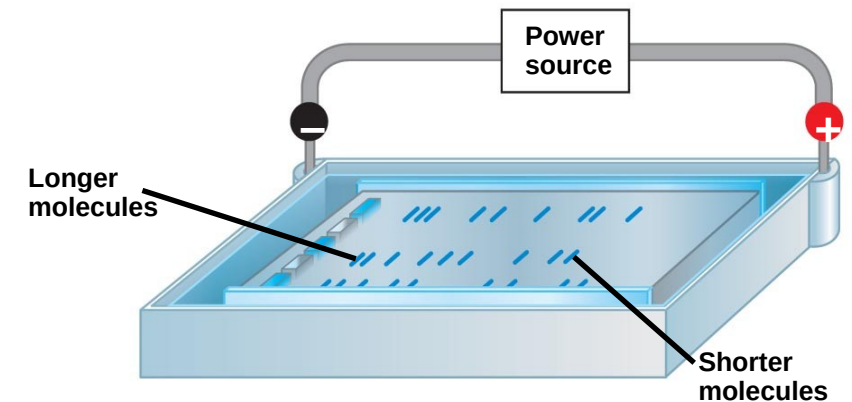
Lambda DNA/*Hind* III

M1201/M1202



WAAROM GELELEKTROFORESE NA PCR?

- Detecteren van de aanwezigheid van PCR product
- Bepalen van de lengte van het PCR fragment



Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

AGENDA

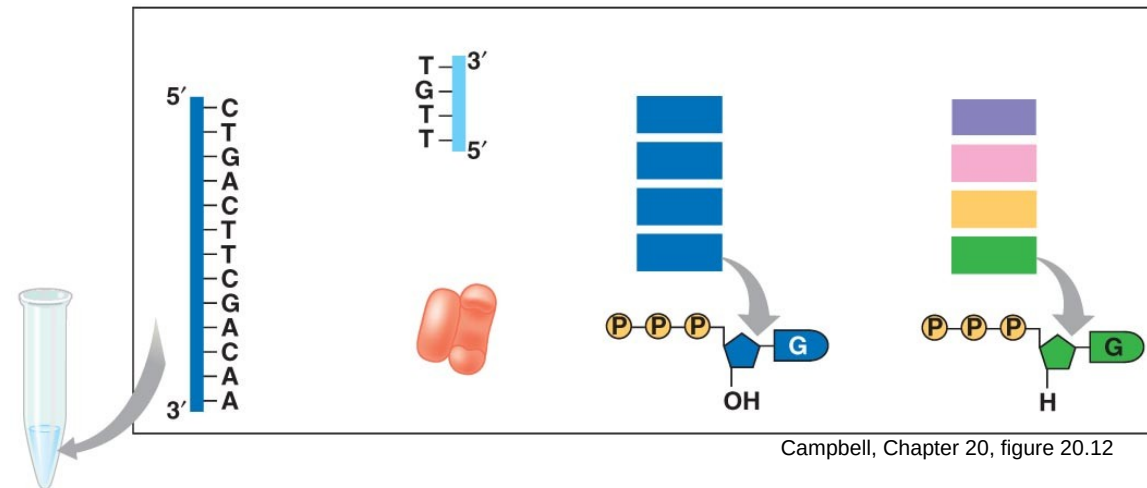
- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

SANGER SEQUENCING

- Methode om DNA te sequencen
- Ontworpen door meneer Sanger en zijn collega's in 1977
- Ook wel **dideoxynucleotide sequencing** genoemd
- Relatief korte DNA fragmenten worden gesequenced met de **dideoxy chain termination methode**

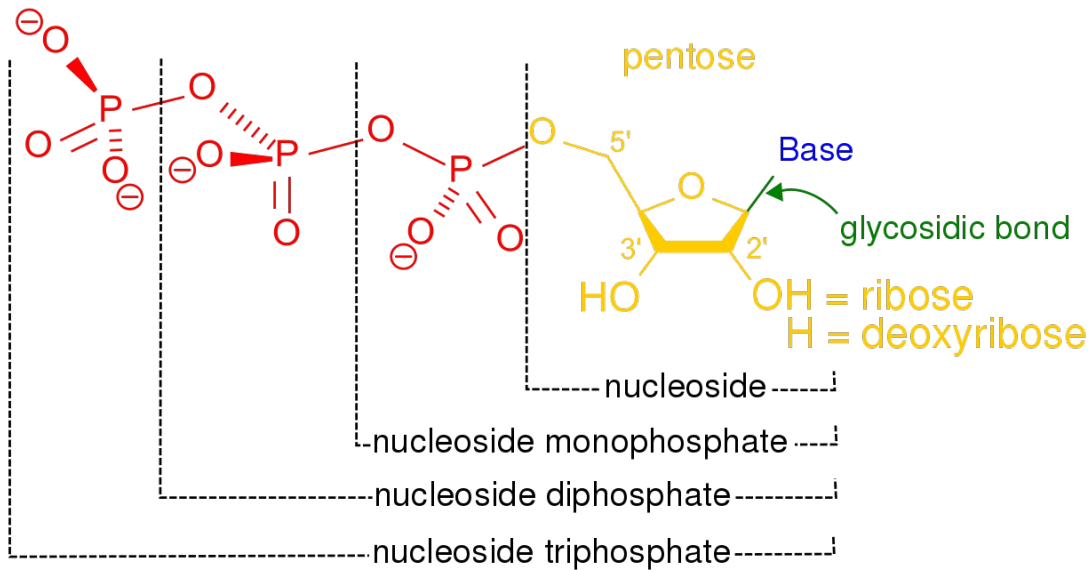
SANGER SEQUENCING – WAT HEB JE NODIG?

- DNA template
- Nucleotiden: dNTPs
dATP, dTTP, dGTP, dCTP
- Gemodificeerde nucleotiden: dideoxynucleotiden (ddNTPs)
ddATP, ddTTP, ddGTP, ddCTP
Gelabeld met verschillende kleuren fluorescerende labels
- Primer
- DNA polymerase

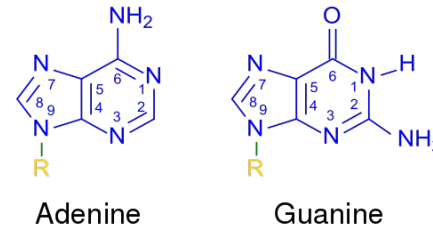


Campbell, Chapter 20, figure 20.12

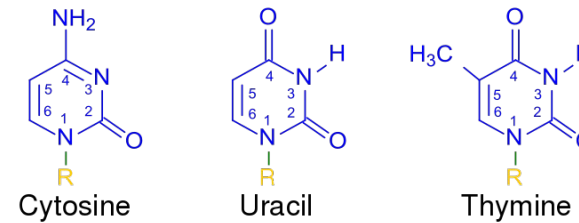
NUCLEOTIDEN



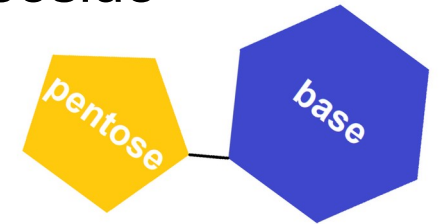
Purines



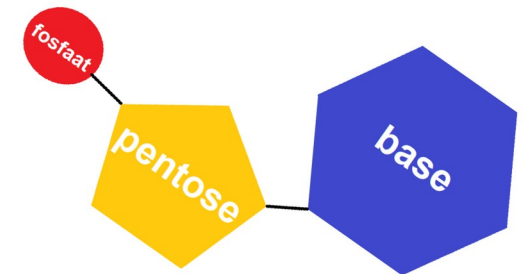
Pyrimidines



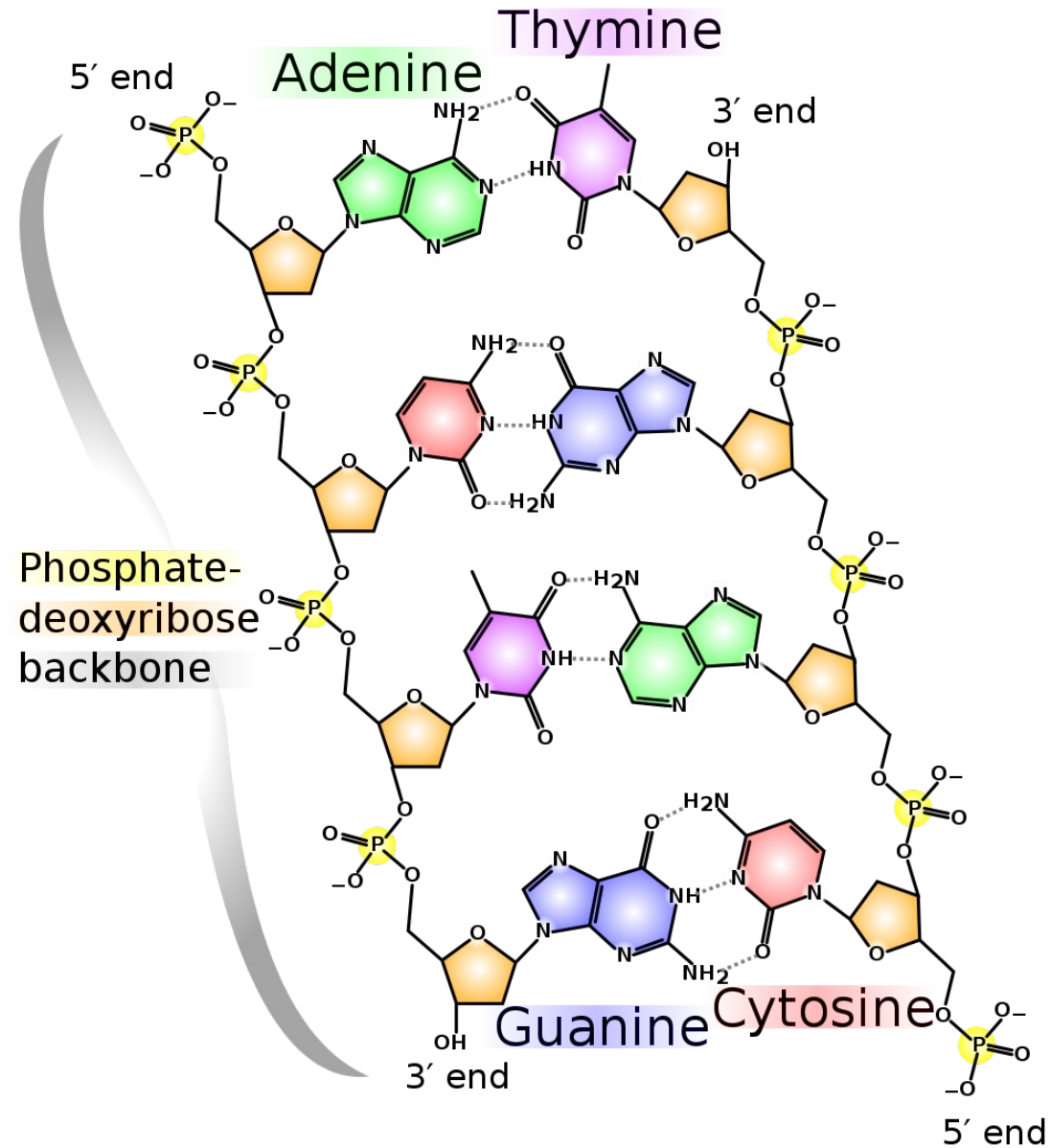
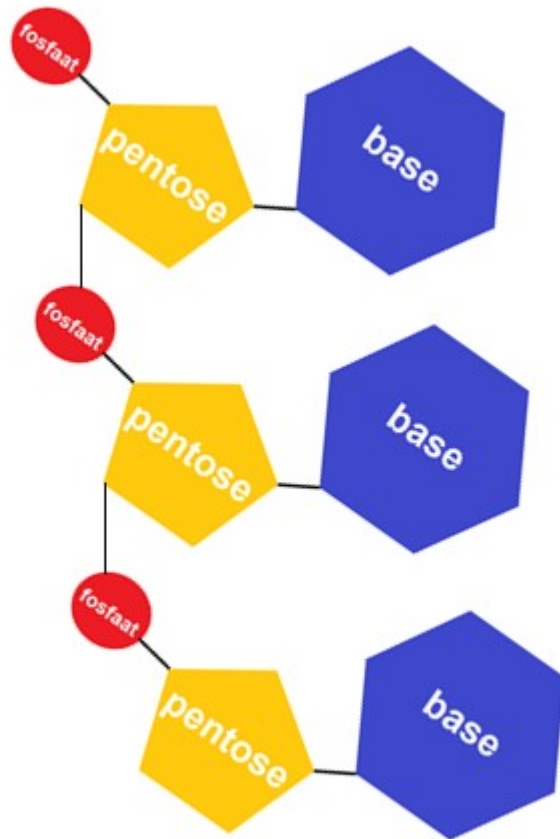
Nucleoside



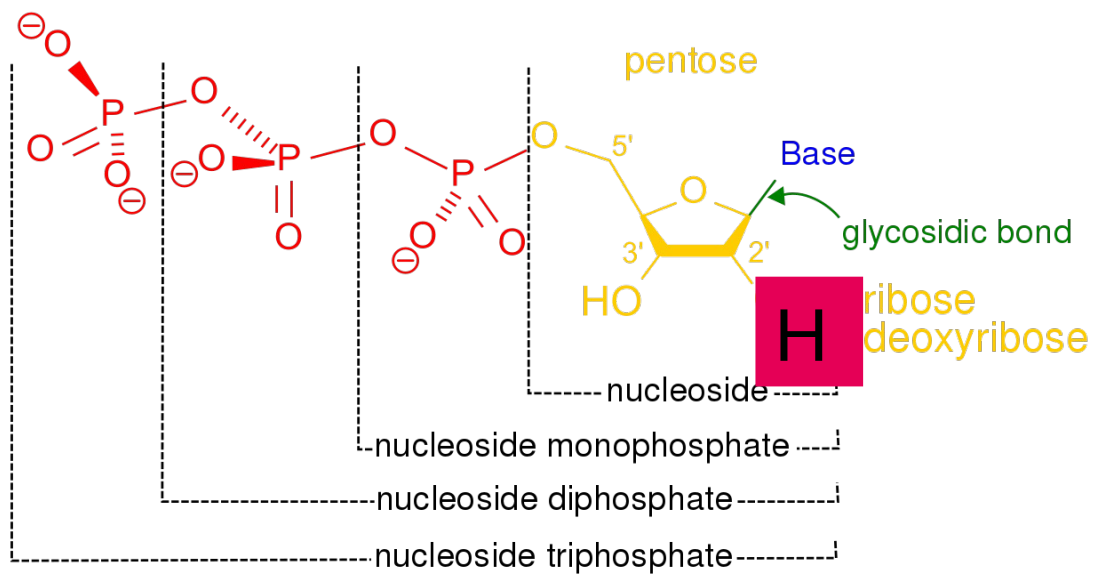
Nucleotide



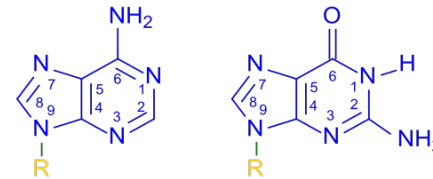
DNA 5' → 3'



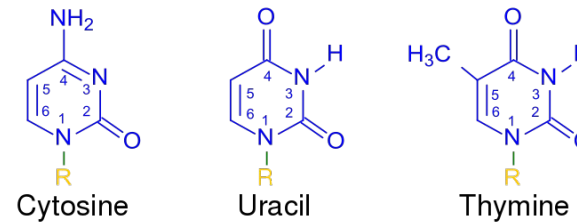
DNTPS – BOUWSTENEN DNA - GEBRUIKT IN PCR



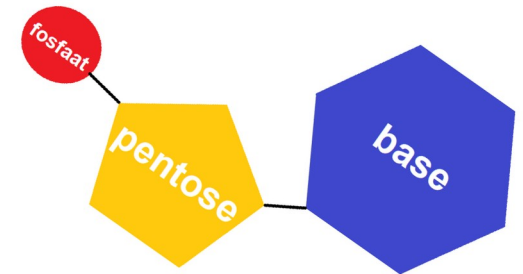
Purines



Pyrimidines

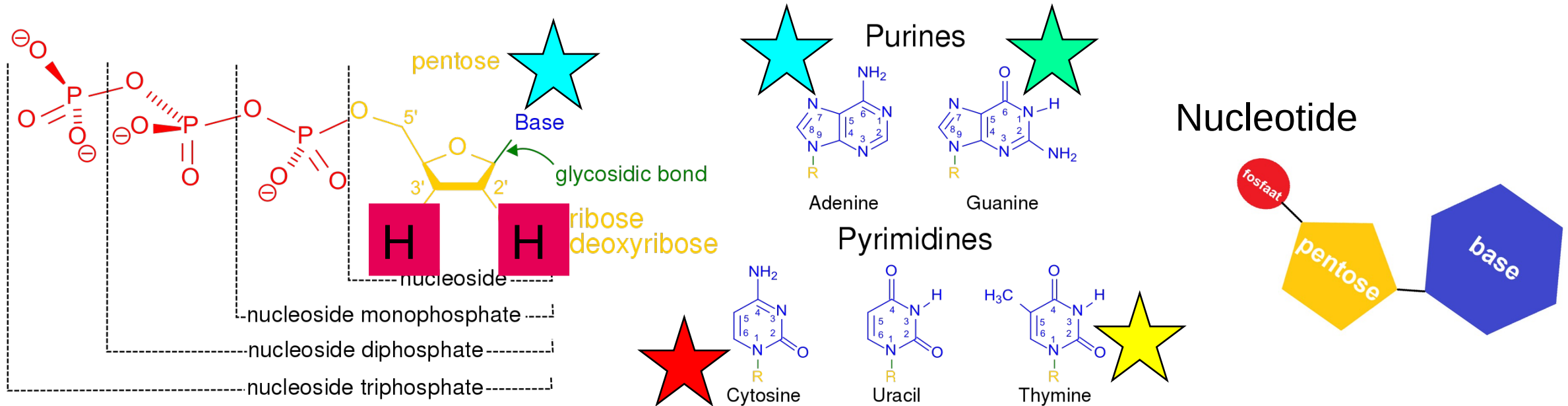


Nucleotide



dNTP: OH-groep op 2' positie in ribose vervangen door H-groep

DDNTPS – GEBRUIKT IN SANGER SEQUENCING



ddNTP: OH-groepen op zowel 2' en 3' positie in ribose vervangen door H-groep
In Sanger sequencing zijn deze ddNTPs gelabeld met een fluorescerend label

SANGER SEQUENCING - STAPPEN

1. DNA template wordt gedenateerd → enkelstrengs DNA
2. Enkele primer wordt toegevoegd
Primer: ~20 nt lang, complementair aan enkelstrengs DNA
3. dNTPs en ddNTPs worden toegevoegd
4. DNA polymerase synthetiseert een complementaire DNA streng met de dNTPs
5. DNA synthese wordt gestopt wanneer DNA polymerase er een ddNTP inzet
→ Waarom stopt de DNA synthese?

SANGER SEQUENCING

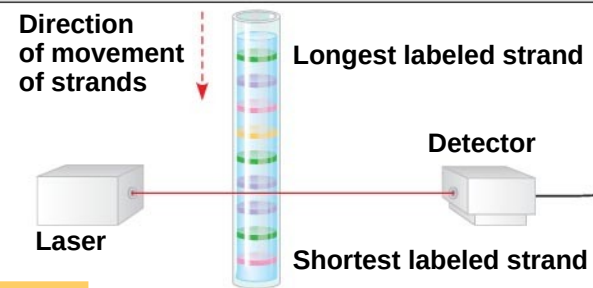
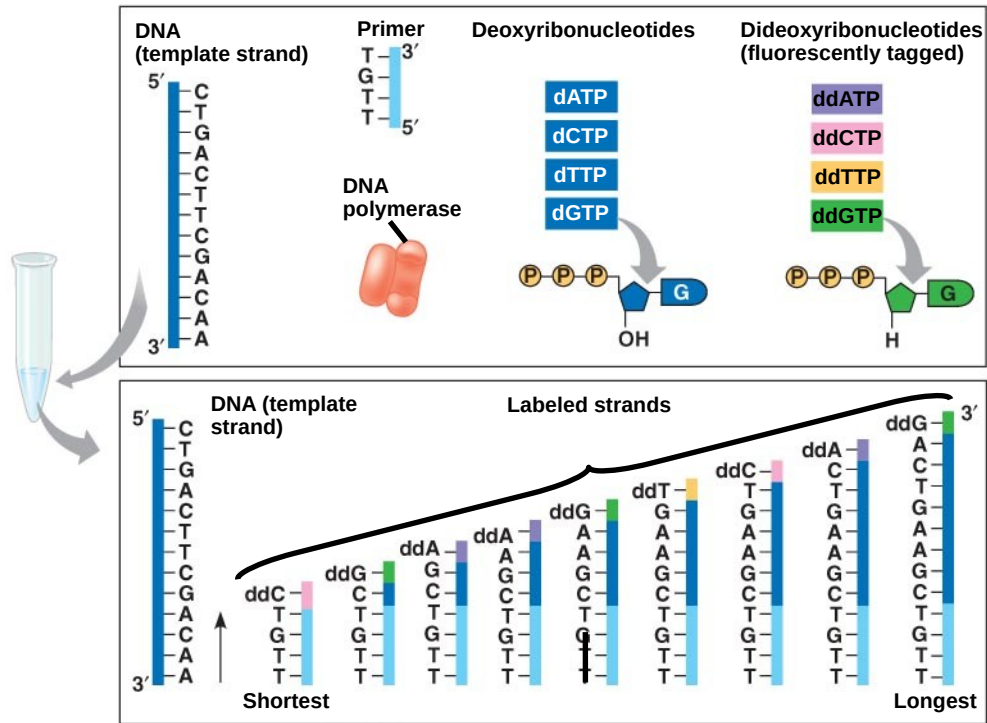
De DNA synthese stopt wanneer er een ddNTP in wordt gezet:

- Omdat ddNTPs de 3' OH-groep missen
- Daarom: **dideoxy chain termination methode**

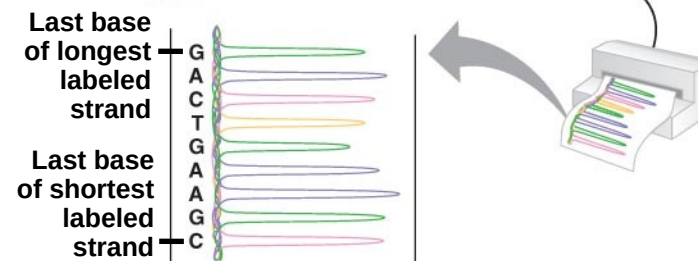
Wanneer de DNA synthese stopt, kan het fluorescerende licht worden gemeten

- Geeft informatie over welke nucleotide er op die plek in het DNA is gezet.

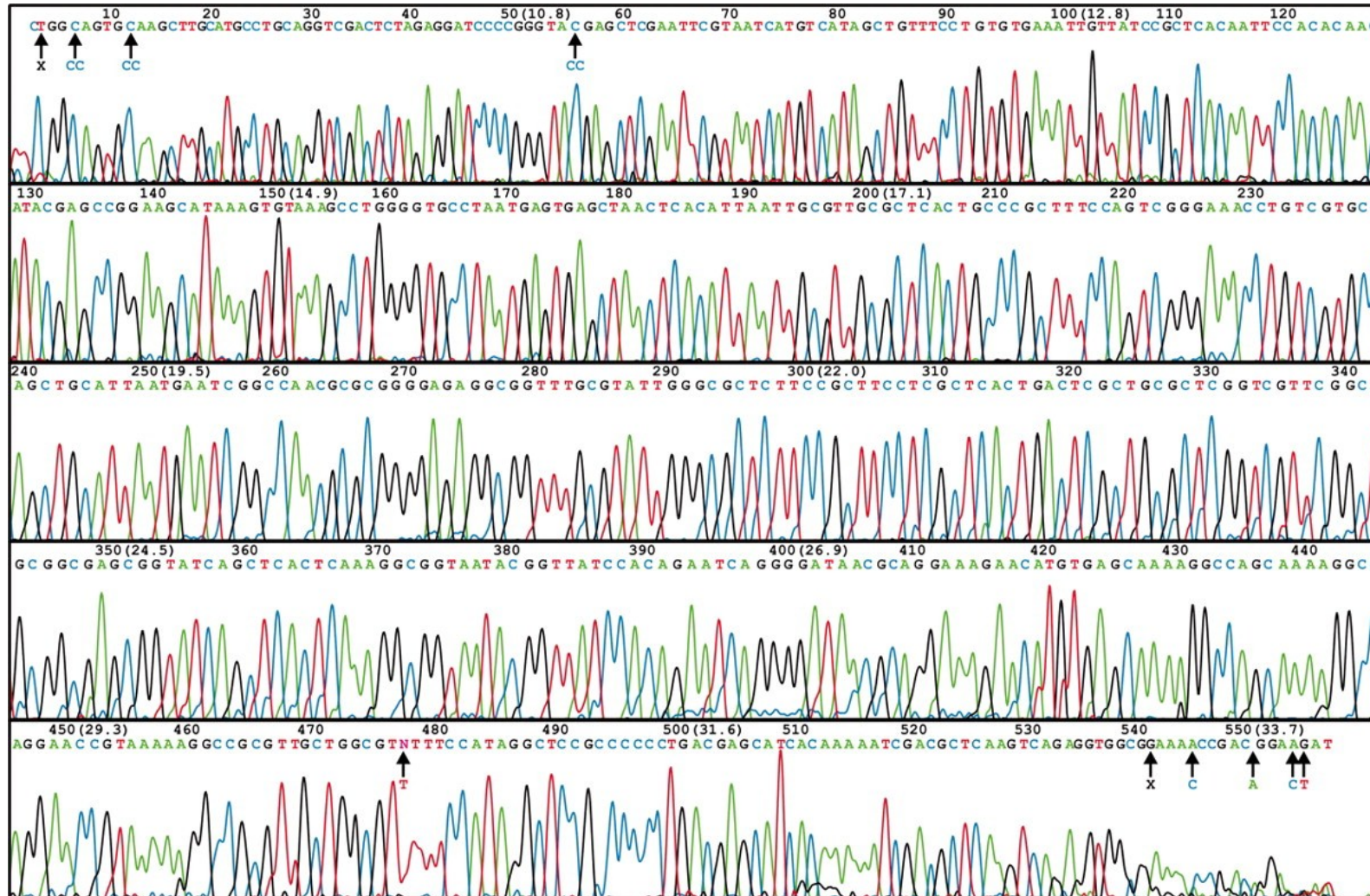
TECHNIQUE



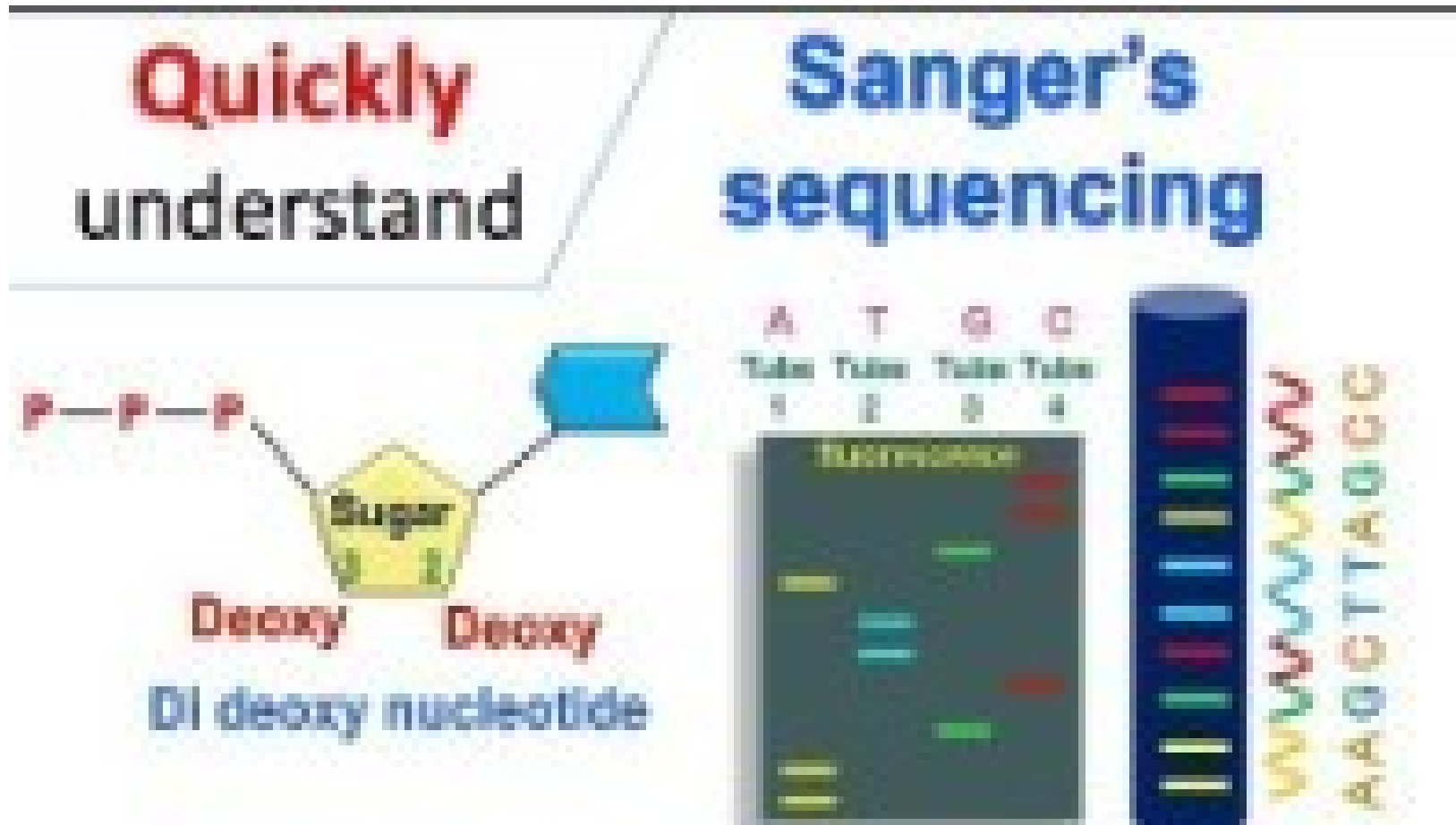
RESULTS



SANGER SEQUENCING - OUTPUT



SANGER SEQUENCY FILMPJE



SANGER SEQUENCING

Sanger sequencing kan je toepassen op je PCR product.

Waarom PCR vóór Sanger sequencing?

AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

BRUGGETJE

- Gen gesequenced → sequentie annoteren en opslaan in database (bijv. NCBI RefSeq database)
- Voorbeelden van annotatie:
 - Positie exonen
 - Positie intronen
 - Positie coding sequence (CDS)
- Gen visualiseren m.b.v. genome browser

Homo sapiens insulin (INS), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000207.3

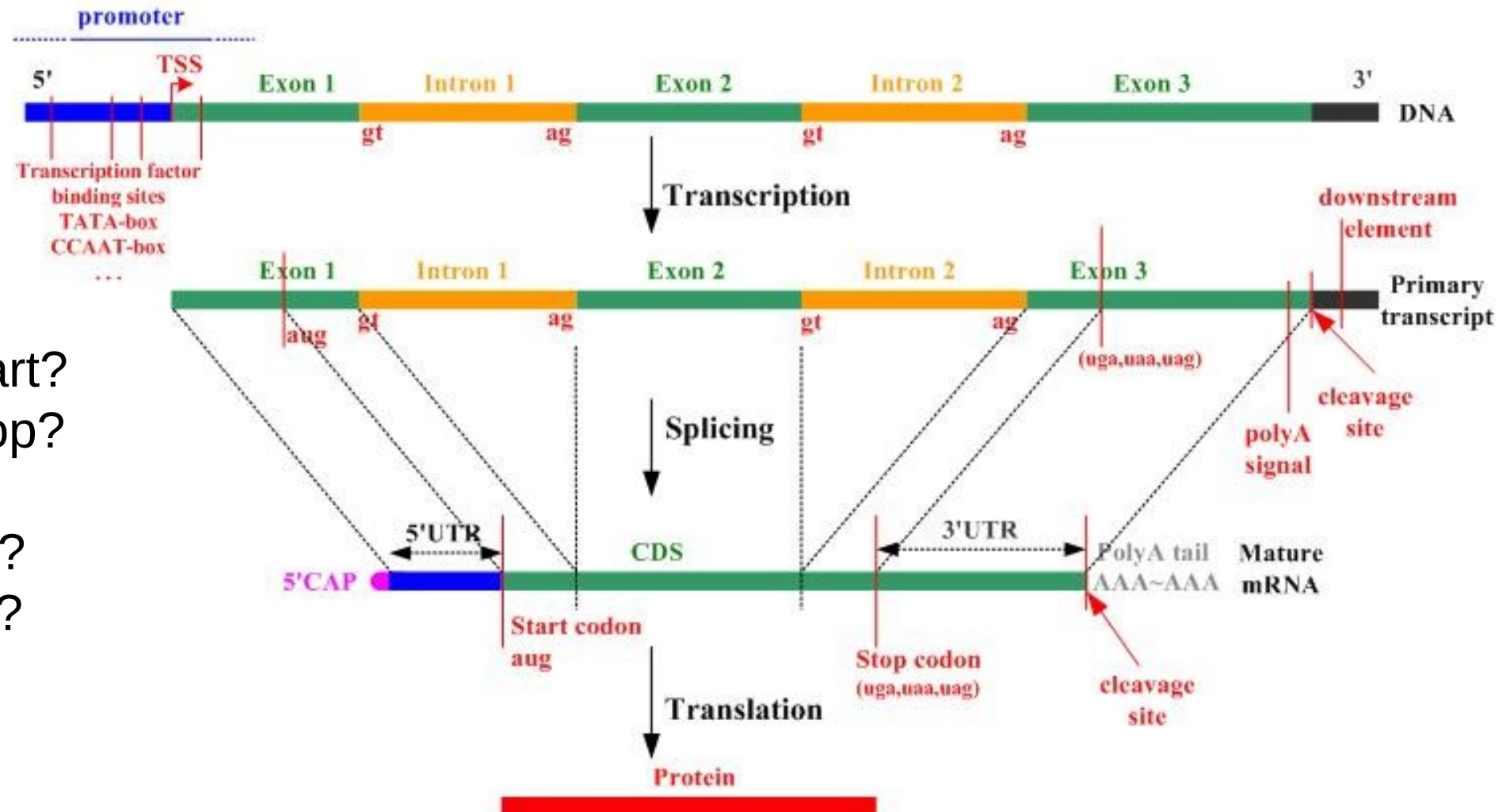
ORIGIN

```
1 agccctccag gacaggctgc atcagaagag gccatcaagc agatcactgt ccttctgcc
61 tggccctgtg gatgcgcctc ctgccctgc tggcgctgct ggccctctgg ggacctgacc
121 cagccgcagc ctttgtgaac caacacctgt gcggctcaca cctggtggaa gctctctacc
181 tagtgtgcgg ggaacgaggc ttcttctaca cacccaagac ccgccgggag gcagaggacc
241 tgcaggtggg gcaggtggag ctgggcgggg gccctggtgc aggcagcctg cagcccttgg
301 ccctggaggg gtccctgcag aagcgtggca ttgtggaaca atgtgttacc agcatctgct
361 ccctctacca gctggagaac tactgcaact agacgcagcc cgcaggcagc cccacaccgc
421 ccgcctcctg caccgagaga gatggaataa agcccttgaa ccagc
```

<u>gene</u>	1..465
<u>exon</u>	1..42
<u>exon</u>	43..246
<u>CDS</u>	60..392
<u>exon</u>	247..465



GEN & MRNA STRUCTUUR



Transcriptie start?
Transcriptie stop?

Translatie start?
Translatie stop?

Mentimeter!

HOE KUNNEN WE DE GENSTRUCTUUR BEPALEN AAN DE HAND VAN NCBI DATA?

Wat is belangrijk om te weten als we de structuur van een gen willen bepalen aan de hand van NCBI annotatie data? Waar moeten we naar op zoek?

- Hoeveel exonen
- Hoeveel intronen
- Positie transcriptie start en stop
- Positie translatie start en stop
- Positie CDS
- Positie 5' en 3' UTR (untranslated region)
= deel van gen dat niet vertaald wordt = deel dat niet CDS is

HOE KUNNEN WE DE GENSTRUCTUUR BEPALEN AAN DE HAND VAN NCBI DATA?

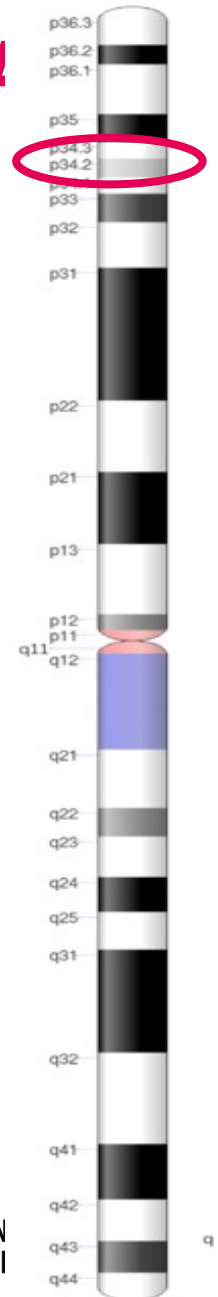
We nemen het MPL gen in de RefSeq database als voorbeeld:

https://www.ncbi.nlm.nih.gov/nucore/NG_007525.1

Interessante informatie:

```
source      1..23661
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="1p34.2"
gene        5001..21661
            /gene="MPL"
            /gene_synonym="C-MPL; CD110; MPLV; THCYT2; THPOR; TPOR"
            /note="MPL proto-oncogene, thrombopoietin receptor"
            /db_xref="GeneID:4352"
            /db_xref="HGNC:HGNC:7217"
            /db_xref="MIM:159530"
```

- 1: Chromosoom nummer 1
- p: gen ligt op de korte arm van het chromosoom
- 34.2: cytogenetische band



HOE KUNNEN WE DE GENSTRUCTUUR BEPALEN AAN DE HAND VAN NCBI DATA?

We nemen het MPL gen in de RefSeq database als voorbeeld:

https://www.ncbi.nlm.nih.gov/nucore/NG_007525.1

Interessante informatie:

```
source      1..23661
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="1"
             /map "1p34.2"
gene        5001..21661
             /gene="MPL"
             /gene_synonym="C-MPL; CD110; MPLV; THCYT2; THPOR; TPOR"
             /note="MPL proto-oncogene, thrombopoietin receptor"
             /db_xref="GeneID:4352"
             /db_xref="HGNC:HGNC:7217"
             /db_xref="MIM:159530"
```

Het gen ligt op positie
5001-21661 van chromosoom 1

HOE KUNNEN WE DE GENSTRUCTUUR BEPALEN AAN DE HAND VAN NCBI DATA?

We nemen het MPL gen in de RefSeq database als voorbeeld:

https://www.ncbi.nlm.nih.gov/nucore/NG_007525.1

Interessante informatie:

mRNA

```
join(5001..5124,5296..5428,5739..5917,6468..6766,  
7161..7323,7584..7710,13642..13826,13989..14131,  
16040..16199,16460..16556,19413..19500,19715..21661)
```

Het *pre*-mRNA bestaat uit 12 exonen.

HOE KUNNEN WE DE GENSTRUCTUUR BEPALEN AAN DE HAND VAN NCBI DATA?

We nemen het MPL gen in de RefSeq database als voorbeeld:

https://www.ncbi.nlm.nih.gov/nucore/NG_007525.1

Interessante informatie:

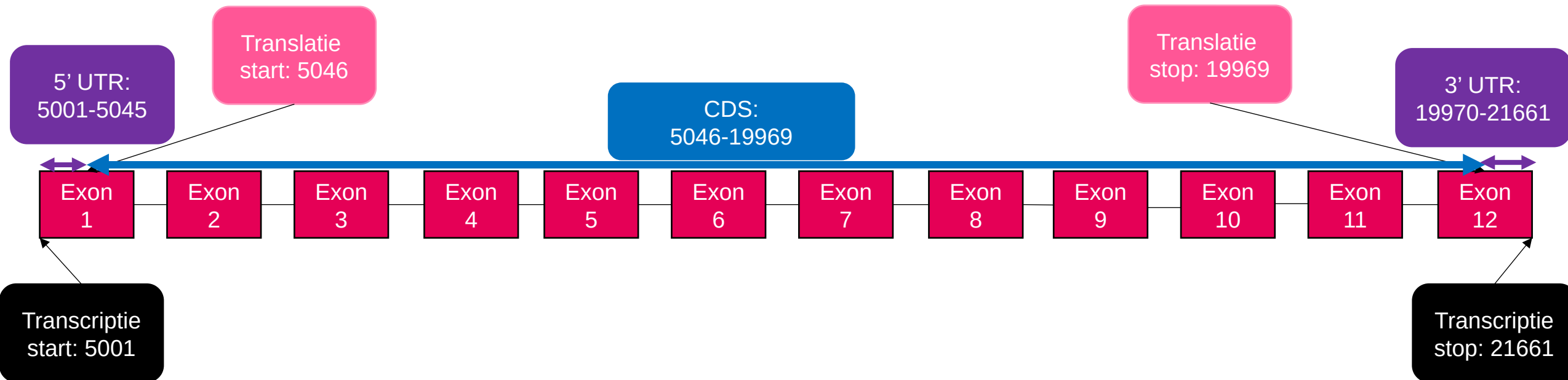
CDS join(5046..5124,5296..5428,5739..5917,6468..6766,
7161..7323,7584..7710,13642..13826,13989..14131,
16040..16199,16460..16556,19413..19500,19715..19969)

De coding sequence (CDS) bestaat ook uit 12 exonen.

Let op: exon 1 begint op een andere positie en exon 12 eindigt op een andere positie in vergelijking met het pre-RNA .

Waarom? → 5' & 3' UTR.

WAT IS NU DE GENSTRUCTUUR VAN HET MPL GEN?



OPDRACHT

Teken de structuur van het coagulation factor IX (F9) gen.
Geef in je tekening de volgende onderdelen aan (geef ook het getal van de positie):

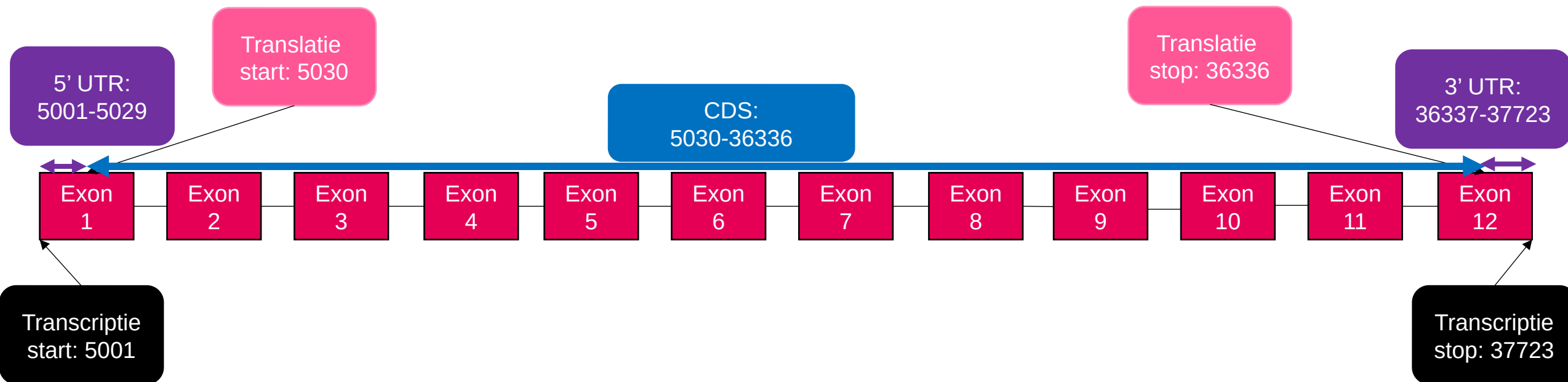
- Exonen
- Intronen
- Positie transcriptie start en stop
- Positie translatie start en stop
- Positie 5' en 3' UTR

Teken de structuur aan de hand van de annotatie die je kan vinden over het F9 gen in de RefSeq database: https://www.ncbi.nlm.nih.gov/nucore/NG_007994.1

AGENDA

- Herhaling week 2
- Huiswerk bespreken
- Weektaak 1 t/m 3
- Gelelektroforese
- Sanger sequencing
- Genstructuur
- Genetische varianten

WAAR IN DIT GEN ZIT VARIATIE?



We kunnen dit opzoeken in de NCBI gene, dbSNP, ClinVar databases.

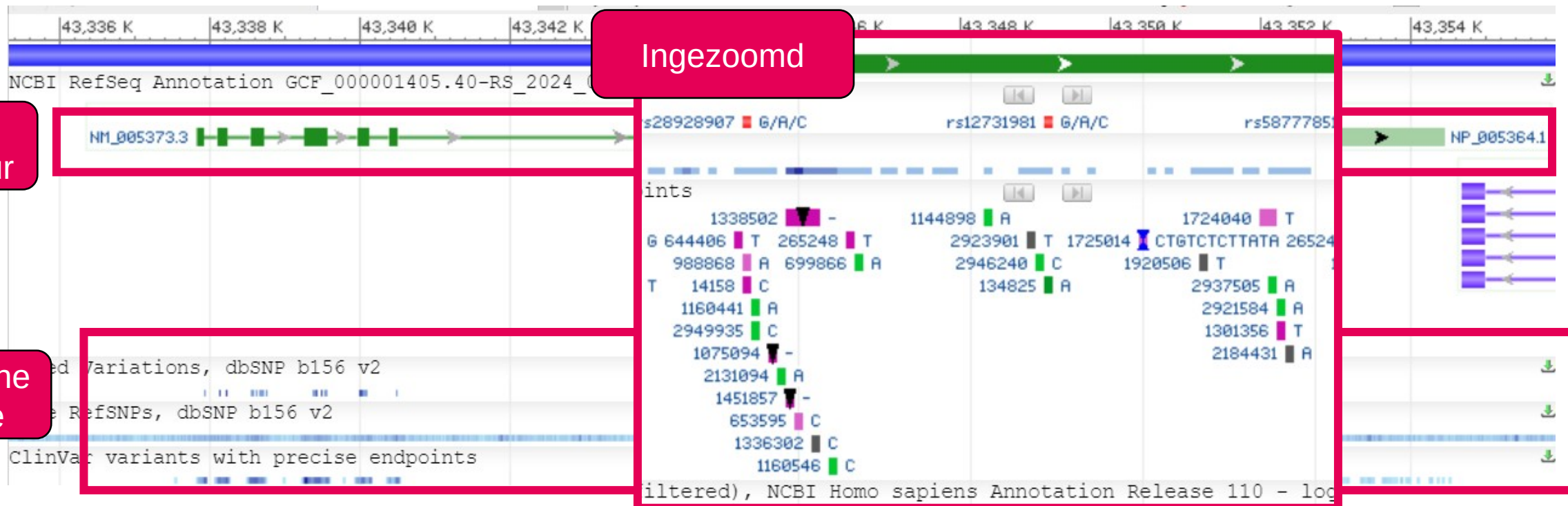
NCBI Gene database:

<https://www.ncbi.nlm.nih.gov/gene/4352>

Niet op schaal

WAAR IN DIT GEN ZIT VARIATIE?

NCBI Gene database: <https://www.ncbi.nlm.nih.gov/gene/4352>



SAMENVATTING

- Gelelektroforese:
 - Methode om moleculen op grootte te scheiden
 - Gel waar elektrisch veld op wordt gezet, negatief geladen moleculen bewegen naar positieve pool.
 - DNA ladder om grootte DNA fragment te bepalen
- Sanger sequencing:
 - Maakt gebruik van ddNTPs met fluorescerend label
 - Wanneer polymerase er een ddNTP in zet stopt de reactie. Er kan geen dNTP meer ingezet worden door polymerase → chain termination
 - Volgorde van fluorescerende kleuren geeft volgorde van DNA
- Genstructuur bepalen aan de hand van NCBI RefSeq data
- Genetische variatie in gen opzoeken in de NCBI Gene, dbSNP en ClinVar databases

BRONNEN

- Pevsner, *Bioinformatics and functional genomics*, Wiley-Blackwell, third edition
- Campbell, *Biology*, Pearson, twelfth edition

VERANTWOORDING

- In deze uitgave is géén auteursrechtelijk beschermd werk opgenomen
- Alle teksten © Ingrid Paffen en Emma Vos / HAN tenzij expliciet externe bronnen zijn aangegeven
- Screenshots op basis van eigen werk auteur en/of vernoemde sites
- Eventuele images zijn opgenomen met vermelding van bron