

**Biologie voor informatici
(in zes eenvoudige lessen)**

Marcel Reinders

TU Delft Delft University of Technology
Challenge the future

The Delft Bioinformatics Lab

Overzicht

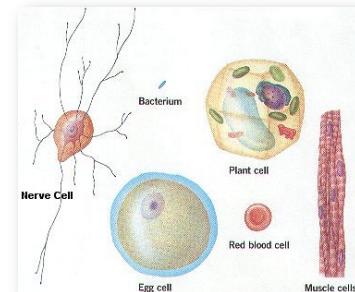
- Celbiologie
 - Eiwitten
 - Metabolisme
 - DNA
- Genen
 - Transcriptie
 - Translatie
- Moleculaire biologie
 - Experimenten
 - Metingen

Overzicht

- **Celbiologie**
 - Eiwitten
 - Metabolisme
 - DNA
- Genen
 - Transcriptie
 - Translatie
- Moleculaire biologie
 - Experimenten
 - Metingen

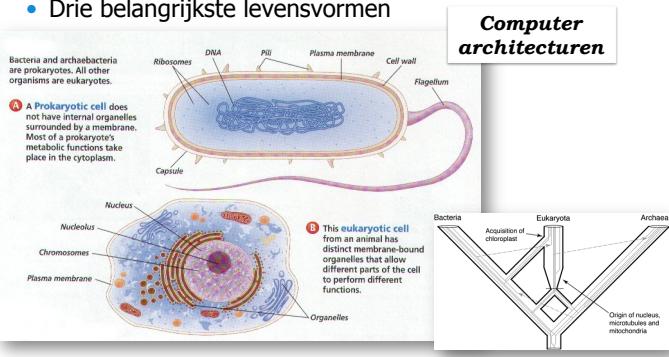
Celbiologie

- De cel is de bouwsteen van (bijna) al het leven



Celbiologie (2)

- Drie belangrijkste levensvormen

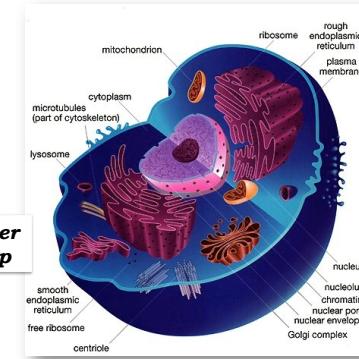


Computer
architecturen

Celbiologie (3)

- Belangrijkste onderdelen van de cel:

- DNA
- RNA
- proteins
- metabolites



Computer
ontwerp

Vraag #1

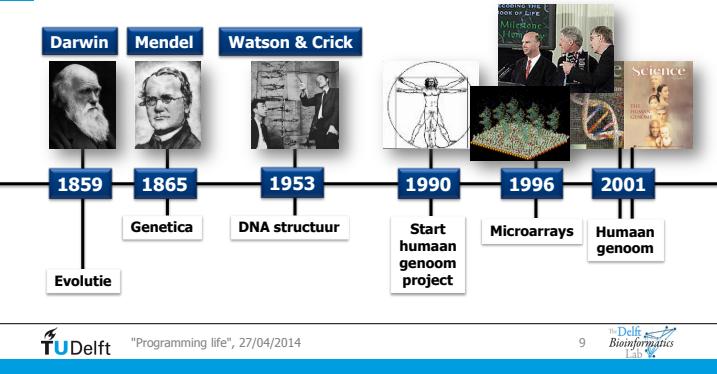
- Goed of fout:
eiwitten doen het meeste werk in de cel



PLAY "The Hidden Life of a Cell (BBC)"

Moleculaire biologie

- Het meeste in deze film is pas de laatste 60 jaar ontdekt!



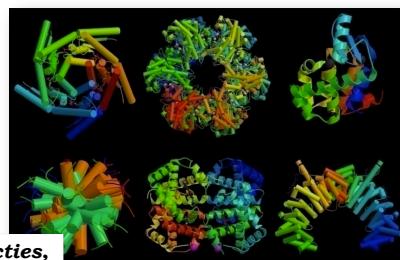
Biologie voor informatici

- De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties**

10 "Programming life", 27/04/2014

Eiwitten

- Eiwitten en eiwitcomplexen zorgen voor:
 - regulatie (DNA, RNA, andere eiwitten)
 - constructie en onderhoud
 - transport binnen en tussen cellen
 - communicatie in en tussen cellen
 - catalyse van chemische reacties (enzymes)



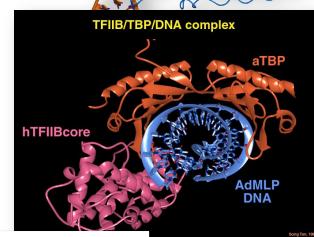
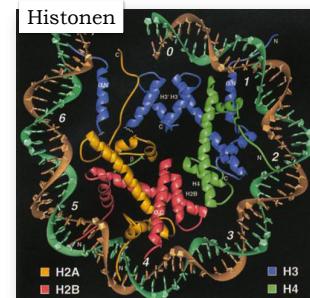
Functies,
operators

TU Delft "Programming life", 27/04/2014

11

Eiwitten en DNA

Nucleases, ligases,
polymerases,
topoisomerase...



Transcriptie
factoren

12

TU Delft Programming life, 15/2/2012

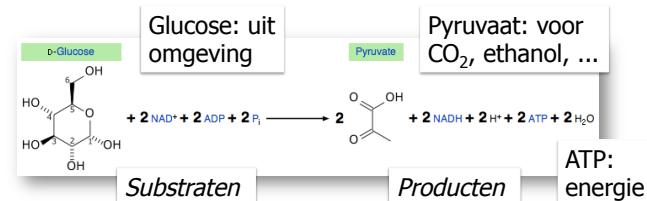
Biologie voor informatici

1. De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties
2. **Schrik niet van het Griekse/Latijnse jargon**

I/O

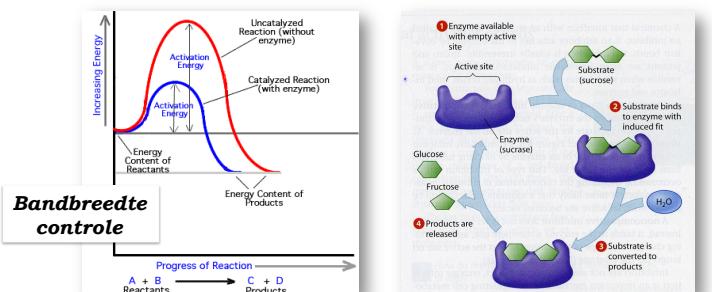
Metabolisme

- Chemische reacties
 - Catabolisme: afbraak van grondstoffen en giftige stoffen
 - Anabolisme: productie van bouwstenen, bv. aminozuren
 - Productie en transport van energie



Enzymen

- Eiwitten beïnvloeden de snelheid van chemische reacties



Vraag 2

- **Goed of fout:**
de functie van een eiwit wordt volledig bepaald door de sequentie van zijn aminozuren



Eiwitstructuur

PGNSAITAGG
VGGXLQVNQL
LYWALTTIGF
QTTMLGGYFA

Primary protein structure is sequence of a chain of amino acids
Amino Acid

Secondary protein structure occurs when the sequence of amino acids are turned by hydrogen bonds
Alpha helix
Pleated sheet

Tertiary protein structure occurs when certain attractions are present between alpha helices and pleated sheets.
Alpha helix
Pleated sheet

Quaternary protein structure is a protein consisting of more than one amino acid chain.

TU Delft "Programming life", 27/04/2014

17 **Delft Bioinformatics Lab**

Eiwitten

- Eiwitten kunnen samenwerken en worden veranderd door andere moleculen

Inactive IRK (Hubbard et al., 1994)

Active IRK (Hubbard, 1997)

U1 snRNP
Cryo-EM
RNA modelling and cryo-EM fitting
U1-A'
U1-A
U1-70k
U1-C
U1-B
Stem I
Stem II
Stem IV
Sm-protein ring complex
Sm-B and U1-C
X-ray and crosslinking

TU Delft "Programming life", 27/04/2014

18 **Delft Bioinformatics Lab**

Chromosomen

Operating system (ROM)

- Mensen zijn *diploide* organismen:
22 chromosoomparen + 2X / XY

Vrouw **Man**

TU Delft "Programming life", 27/04/2014

19 **Delft Bioinformatics Lab**

Vraag 3

- Waar ligt volgens het "centrale dogma van de moleculaire biologie" alle informatie voor het leven opgeslagen?

DNA

TU Delft "Programming life", 27/04/2014

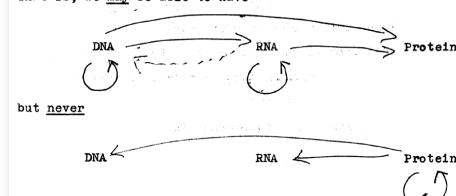
20 **Delft Bioinformatics Lab**

Central dogma of molecular biology

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we may be able to have



Francis Crick,
1958

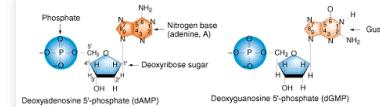
DNA

Assembly code

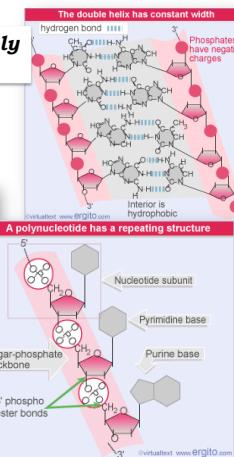
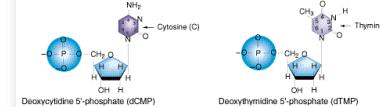
Desoxyribonucleic acid

ACGTATGACCTATATAACGTAC ...

Purine nucleotides

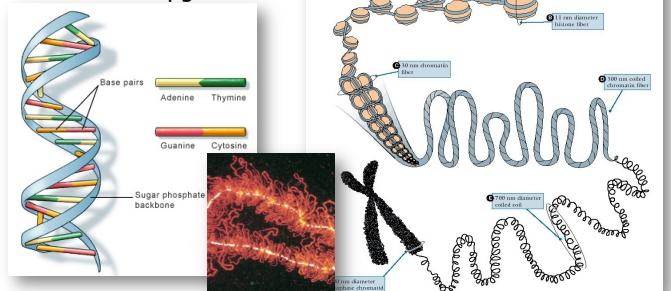


Pyrimidine nucleotides



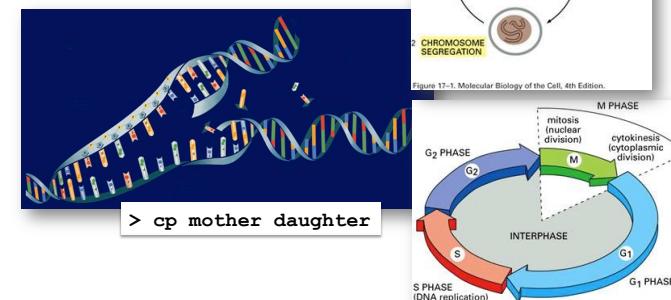
DNA (2)

- In eukaryoten: extrem opgerold



DNA (3)

- DNA kan repliceren:



wehi.edu.au

Molecular visualizations of

DNA

1. *DNA Wrapping*

Vraag 4

- Hoeveel fouten in het menselijk genoom worden er naar verwachting gemaakt bij elke celdeling?

±3
($\sim 10^{-9}$ fout/base)

Overzicht

- Celbiologie
 - Eiwitten
 - Metabolisme
 - DNA
- **Genen**
 - Transcriptie
 - Translatie
- Moleculaire biologie
 - Experimenten
 - Metingen

Genen

- Eenheid van erfelijkheid
- 99.99% hetzelfde tussen individuen
- Single nucleotide polymorphisms (SNPs)

Registry

The diagram illustrates the Human Registry, mapping specific genetic conditions to specific chromosomal regions. The regions are color-coded by chromosome number (1-22, X, Y) and include labels for various diseases and their associated genes:

- Chromosome 1:** OpticGerb98 (MIM# 60010), PIP5K2 deficiency.
- Chromosome 2:** NLGN (Autism).
- Chromosome 3:** TDZ3 (Intellectual disability).
- Chromosome 4:** Coflin-Lowry & MRX19.
- Chromosome 5:** ARX (Autism), TSHZ3 (Autism), LIG4 (Autism), MRX19 (Autism).
- Chromosome 6:** TSHZ3 (Autism).
- Chromosome 7:** EGRAPL (Autism).
- Chromosome 8:** GTC deficiency.
- Chromosome 9:** WDR62 deficiency, March disease (MIM# 248000).
- Chromosome 10:** TSHZ3 (Autism).
- Chromosome 11:** FGD1 (Angiogenesis & MRX).
- Chromosome 12:** SPHIN (Autism, cerebellar hypoplasia, hypotonia, hyporeflexia).
- Chromosome 13:** NLGN (Autism).
- Chromosome 14:** ATXN (Autism).
- Chromosome 15:** Meier disease (PTPA), POF deficiency.
- Chromosome 16:** NME1 (Autism).
- Chromosome 17:** TMEM138 (Autism), PIP5K2 deficiency.
- Chromosome 18:** FGZ2 (Autism, BPD), LRRK2 (DOPA).
- Chromosome 19:** MRX19 (Autism, 47).
- Chromosome 20:** DNM2 (Autism).
- Chromosome 21:** DNM2 (Autism).
- Chromosome 22:** SGB (Autism), RBBP8 (Autism).
- Chromosome X:** Leish-Ahmad (Autism), PRPF8 (Autism).
- Chromosome Y:** PRPF8 (Autism).

Chromosome Viewer

Pythagoras registryviewer (status: IM1 | DSM 1523 | ATCC 51768) chromosome (2222400 bp, g=0.31) (coordinate 244715 - 508646).

hints: Mouse over a gene to see details.

Loaded

Vraag 5

- Goed of fout:
**de mens heeft (samen met mensapen)
het grootste genoom**



Genomen

- HIV: 10kb
- E. coli: 5Mb
- Gist: 12Mb
- Fruitylieg: 50Mb
- Worm: 100Mb
- **Mens, muis: 3Gb**
- Tarwe: 16Gb
- Lelie: 90Gb
- Longvis: 133Gb



Genomes vary greatly in size		
Genome	Gene Number	Base Pairs
Organisms		
Plants	<50,000	$<10^{11}$
Mammals	30,000	$\sim 3 \times 10^9$
Worms	14,000	$\sim 10^8$
Flies	12,000	1.6×10^8
Fungi	6,000	1.3×10^7
Bacteria	2-4,000	$<10^7$
Mycoplasma	500	$<10^6$
dsDNA Viruses		
Vaccinia	<300	187,000
Papova (SV40)	~6	5,226
Phage T4	~200	165,000
ssDNA Viruses		
Parvovirus	5	5,000
Phage fX174	11	5,387
dsRNA Viruses		
Reovirus	22	23,000
ssRNA Viruses		
Coronavirus	7	20,000
Influenza	12	13,500
TMV	4	6,400
Phage MS2	4	3,569
STNV	1	1,300
Viroids		
PSTV RNA	0	359

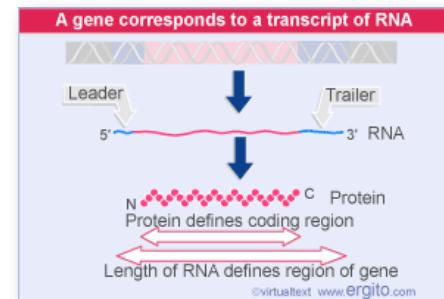
Vraag 6

- Wat betekent het als een gen
"tot expressie komt"?

**Het gen wordt afgeschreven
in messenger RNA (mRNA)**

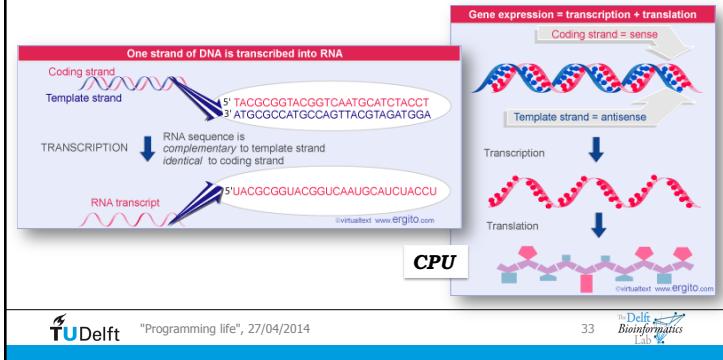
Genen (2)

- Genen: gebieden op het DNA die coderen voor eiwitten



Genen (3)

- Genes worden omgeschreven in messenger RNA ("transcribed") en vertaald ("translated") in eiwit



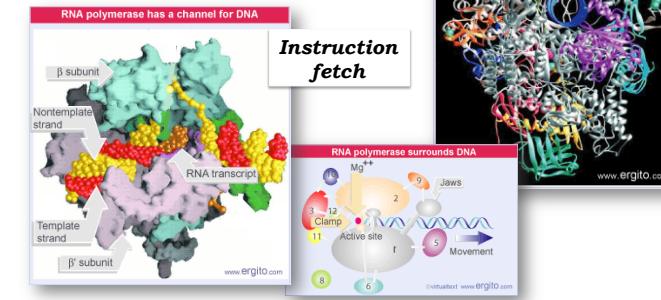
TU Delft "Programming life", 27/04/2014

33

Delft
Bioinformatics
Lab

Transcriptie

- RNA polymerase bindt aan DNA en maakt messenger RNA (mRNA)

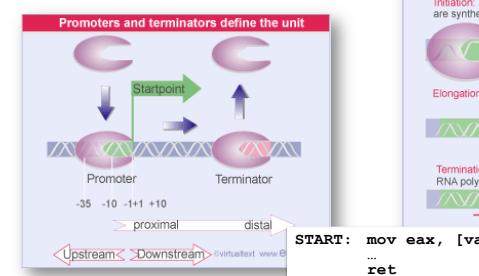


TU Delft "Programming life", 27/04/2014

34 Delft
Bioinformatics
Lab

Transcriptie (2)

- Het af te schrijven deel wordt omsloten door een *promoter* en een *terminator*

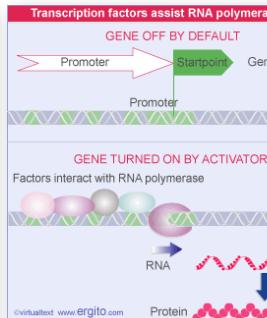
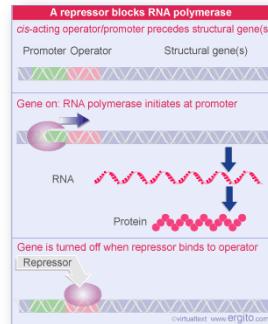


TU Delft "Programming life", 27/04/2014

36 Delft
Bioinformatics
Lab

Transcriptie factoren

- Onderdrukken ("repress") of activeren transcriptie

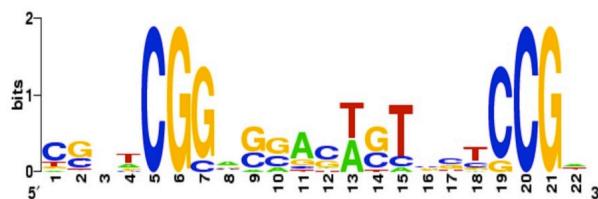


Vraag 7

- Goed of fout:
transcriptie factoren binden
("herkennen") unieke stukken DNA

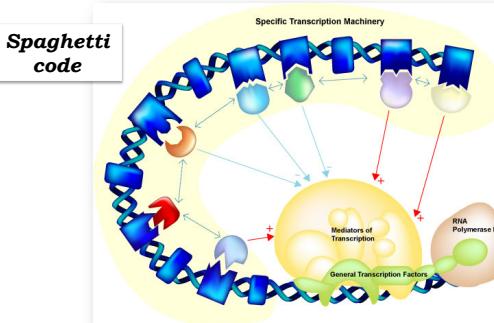
Transcriptie factoren (2)

- TFs hebben, net als polymerase, een "motif": ze binden op een voorkeurs-sequentie, maar kunnen ook binden op afwijkende sequenties (met een bepaalde kans)



Regulatie

- Regulatieprogramma's kunnen erg complex zijn



Translatie

- mRNA bevat de code om eiwitten, reeksen van aminozuren, te maken
- Elk triplet basen (DNA) is een codon voor een van de 20 aminozuren

**CPU
microcode
lookup table**

The genetic code is triplet			
First base		Second base	
U	C	A	G
UUU Phe	UCU Ser	UAU Tyr	UGU Cys
UUC	UCC	UAC	UGC
UUA	UCA	UAA	UGA STOP
UUG Leu	UCG	UAG	UGG Trp
C			
CUU Leu	CCC Pro	CAU His	CGU Arg
CUC	CCG	CAC	CGC
CUA Gln	CCA	CAA	CGA
CUG CGG	CGG	CAG	CGG
A			
AUC Ile	ACU Thr	AAU Asn	AGU Ser
AUA	ACC	AAC	AGC
AUG Met	ACA	AAA Lys	AGA Arg
G			
GUU Val	GCU Ala	GAU Asp	GGU Gly
GUC	GCC	GAC	GGC
GUA Gln	GCA	GAA	GGA
GUG GCG	GCG	GAG Glu	GGG

Vraag 8

- Wat is het maximale effect van een fout van 1 base in het DNA?

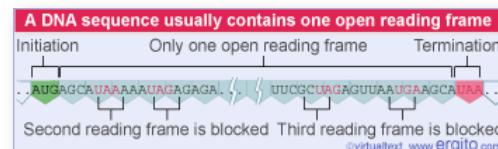
SNP: 1 ander aminozuur

insertie/deletie: compleet ander eiwit

Translatie (2)

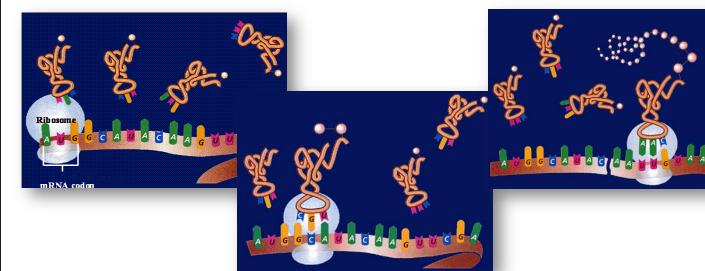
- Reading frame: drie mogelijke vertalingen, bv.

A	C	G	A	C	G	A	C	G
ACG	ACG	ACG	ACG	→ Thr	Thr	Thr	Thr	
CGA	CGA	CGA	CGA	→ Arg	Arg	Arg	Arg	
GAC	GAC	GAC	GAC	→ Asp	Asp	Asp	Asp	



Translatie (3)

- Messenger RNA wordt vertaald door het ribosoom
- Transfer RNAs dragen een aminozuur en een "sleutel" van 3 basen



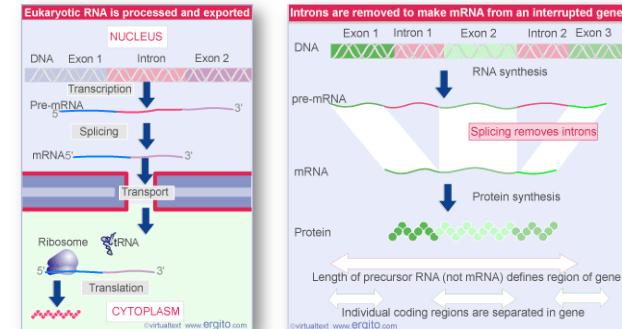
Vraag 9

- Goed of fout:
het menselijk genoom codeert voor
ongeveer 30,000 verschillende eiwitten



Translatie (4)

- In eukaryoten geeft *splicing* meerdere mogelijke eiwitten



**Operator
overloading**

Non-coding RNA

- Genen die voor eiwitten coderen vormen ongeveer 2% van het menselijk genoom
- De rest werd vroeger (2000) "junk DNA" genoemd, maar we denken nu dat het codeert voor:
 - transfer RNAs (tRNA)
 - ribosomaal RNA (rRNA): delen van ribosomen
 - micro RNA (miRNA): regulatie, ...?
 - double-stranded RNA (dsRNA):
 - small interfering RNA (siRNA)
 - small activating RNA (saRNA)
- Misschien heeft RNA ook erfelijke eigenschappen

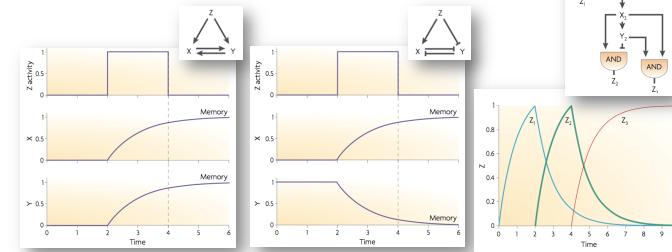
**Self-modifying
code?**

Biologie voor informatici

1. De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties
2. Schrik niet van het Griekse/Latijnse jargon
- 3. Biologen doen alsof ze alles weten, maar er valt nog heel veel te ontdekken**

Complexiteit

- Complex gedrag komt voort uit relatief simpele "schakelingen" van genen/eiwitten die interacties met elkaar aangaan



Biologie voor informatici

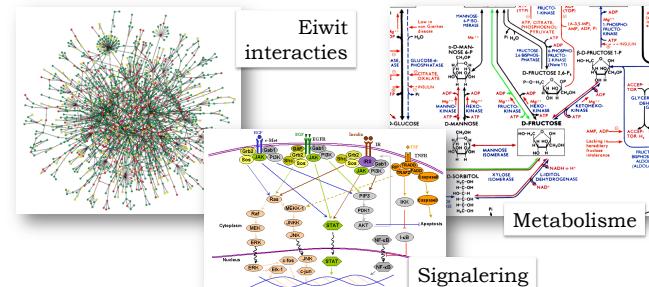
1. De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties
2. Schrik niet van het Griekse/Latijnse jargon
3. Biologen geven de indruk dat ze alles weten, maar er valt nog heel veel te ontdekken
- 4. Het genoom is de broncode van het leven, maar de executie is complex, parallel, analoog en ruizig**

Overzicht

- Celbiologie
 - Eiwitten
 - Metabolisme
 - DNA
 - Genen
 - Transcriptie
 - Translatie
 - **Moleculaire biologie**
 - Experimenten
 - Metingen

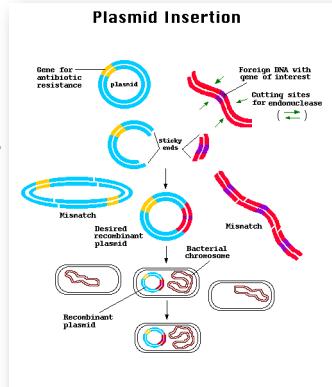
Moleculaire biologie

- Bestudeert interacties tussen moleculen op verschillende niveaus



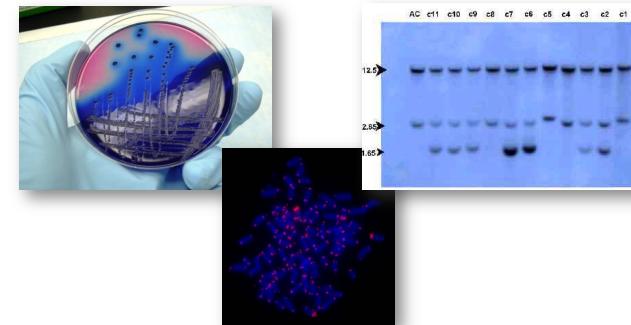
Experimenten

- **Interventies:**
 - Verander de omgeving waarin de cel leeft
 - Kopieer of muteer DNA, bv.
 - breng een nieuw gen in, bv. op een plasmide ("cloning")
 - verwijder een gen, bv. m.b.v. een virus ("knockout")



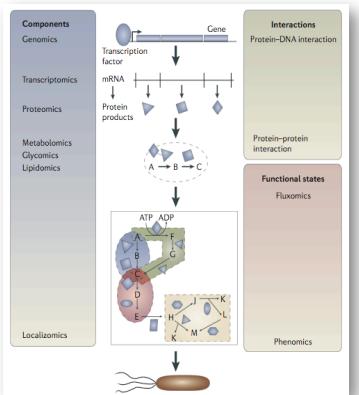
Metingen

- Traditioneel: groei/overleven, microscopen, gels



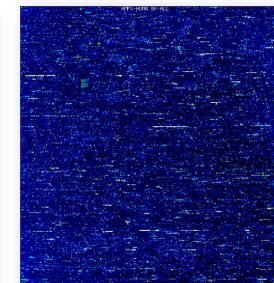
Metingen (2)

- De -omics revolutie: meet in de cel alle...
 - sequenties (DNA, RNA)
 - mRNA transcripten
 - metabole reacties
 - eiwit/metaboliet niveaus
 - eiwit locaties in de cel
 - eiwit-eiwit interacties
 - eiwit-DNA interacties
 - condities ("phenotype")
 - ...
- Moleculair biologen genereren heel veel data



Microarrays

- Meten concentraties van 10,000'en mRNA of DNA moleculen tegelijk

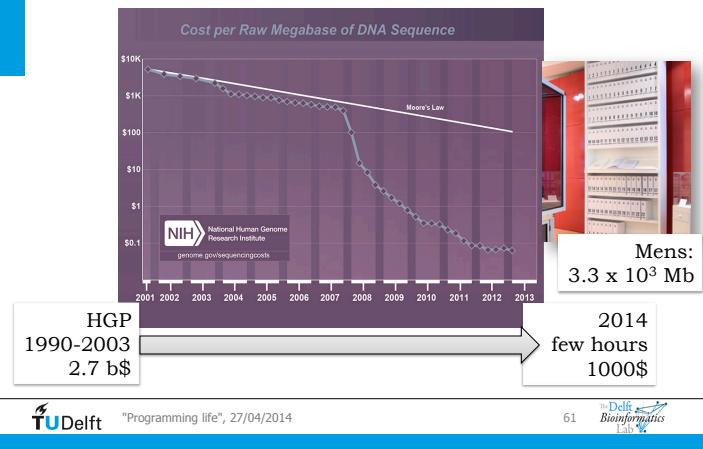


Next generation sequencing

- Nieuwe ontwikkeling: betaalbaar meten van sequenties



Next generation sequencing (3)



Biologie voor informatici

1. De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties
2. Schrik niet van het Griekse/Latijnse jargon
3. Biologen geven de indruk dat ze alles weten, maar er valt nog *heel* veel te ontdekken
4. Het genoom is de broncode van het leven, maar de executie is complex, parallel, analoog en ruizig
5. **Moleculaire biologie wordt voortgedreven door de ontwikkeling van nieuwe apparatuur**

Vraag 10

- Hoeveel basen zijn er opgenomen in de laatste release van GenBank, de sequentie-database?

**150,141,354,858
(150 Gb)**

Biologie voor informatici

1. De cel is een soort soep van moleculen; moleculair biologen bestuderen die moleculen en hun interacties
2. Schrik niet van het Griekse/Latijnse jargon
3. Biologen geven de indruk dat ze alles weten, maar er valt nog *heel* veel te ontdekken
4. Het genoom is de broncode van het leven, maar de executie is complex, parallel, analoog en ruizig
5. Moleculaire biologie wordt voortgedreven door de ontwikkeling van nieuwe apparatuur
6. **Als het *echt* lastig wordt, moeten (bio)informatici te hulp schieten**

Voor volgende keer (02 mei)

- Lees materiaal over "Discovering disease mutations" te vinden op blackboard.
- Maak samenvatting, lever in vóór begin college

Source material

- Wikipedia, <http://www.wikipedia.org/>
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. *Molecular biology of the cell*, 4th ed. Garland Science, 2002.
- Lemon B and Tjian R. Orchestrated response: a symphony of transcription factors for gene control. *Genes and Development* 14:2551-69, 2000.
- Lewin B. *Genes VIII*. Prentice Hall, 2004.