

**Biologie voor informatici  
(in zes eenvoudige lessen)**  
*Dick de Ridder*

**TU Delft** Delft University of Technology  
Challenge the future

## Overzicht

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

**TU Delft** "Programming life", 20/2/2013

2 Bioinformatics

**Overzicht**

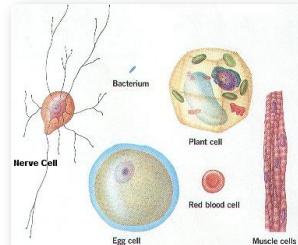
- **Celbiologie**
  - Eiwitten
  - Metabolisme
  - DNA
- Genen
  - Transcriptie
  - Translatie
- Moleculaire biologie
  - Experimenten
  - Metingen

**TU Delft** "Programming life", 20/2/2013

3 Bioinformatics

## Celbiologie

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



**TU Delft** "Programming life", 20/2/2013

4 Bioinformatics

**Celbiologie (2)**

- Drie belangrijkste levensvormen

Bacteria and archaeabacteria are prokaryotes. All other organisms are eukaryotes.

**Computer architectuur**

**Computer ontwerp**

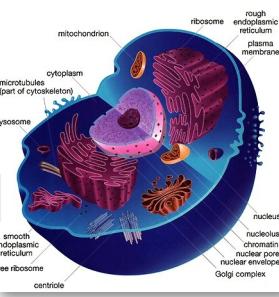
**TU Delft** "Programming life", 20/2/2013

5 Bioinformatics

## Celbiologie (3)

- Belangrijkste onderdelen van de cel:
- DNA
  - RNA
  - proteins
  - metabolites

**Computer ontwerp**



**TU Delft** "Programming life", 20/2/2013

6 Bioinformatics

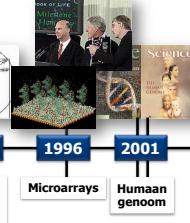
## Vraag #1

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



## Moleculaire biologie

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

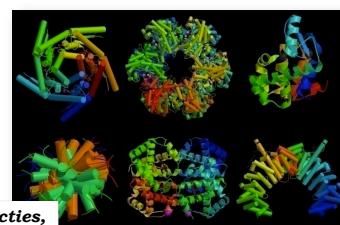


## Biologie voor informatici

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

## Eiwitten

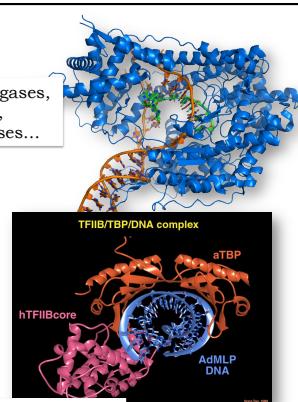
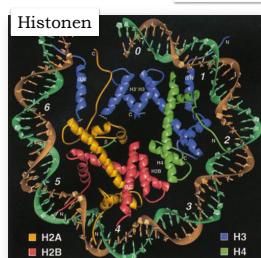
- Eiwitten en eiwitcomplexen zorgen voor:
  - regulatie (DNA, RNA, andere eiwitten)
  - constructie en onderhoud
  - transport binnen en tussen cellen
  - communicatie in en tussen cellen
  - catalyse van chemische reacties (enzymes)



Functies,  
operators

## Eiwitten en DNA

Nucleases, ligases,  
polymerases,  
topoisomerase...



Transcriptie  
factoren

## Biologie voor informatici

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

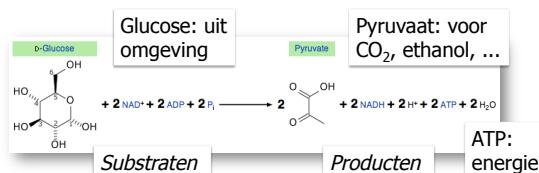
### 2. Schrik niet van het Griekse/Latijnse jargon

## Metabolisme

### I/O

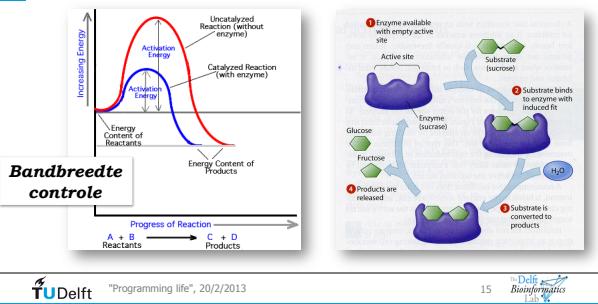
- Chemische reacties

- 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

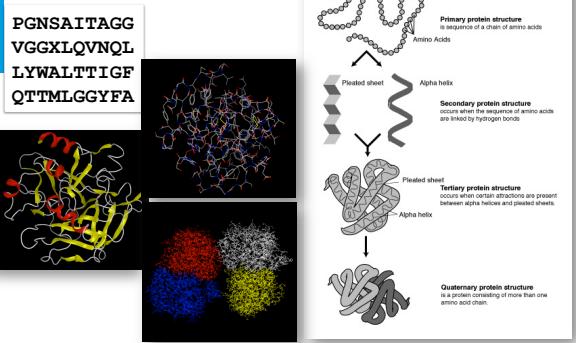


## Vraag 2

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

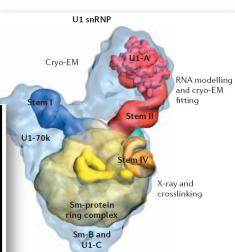
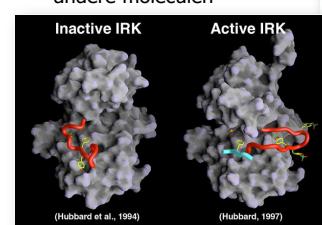


## Eiwitstructuur



## Eiwitten

- Eiwitten kunnen samenwerken en worden veranderd door andere moleculen



## Chromosomen

*Operating system (ROM)*

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

Vrouw                                  Man

TU Delft "Programming life", 20/2/2013      19      TU Delft Bioinformatics 1.1b 14

## Vraag 3

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

# DNA

TU Delft "Programming life", 20/2/2013      20      TU Delft Bioinformatics 1.1b 14

## Central dogma of molecular biology

The Doctrine of the Tried.  
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

but never

where the arrows show the transfer of information.

Francis Crick, 1958

TU Delft "Programming life", 20/2/2013      21      TU Delft Bioinformatics 1.1b 14

## DNA

*Assembly code*

- Desoxyribonucleic acid
- ACGTATGACCTATATAACGTAC ...

Purine nucleotides: Phosphate, Nitrogen base (adenine, A), Deoxyribose sugar, Guanine (G). Deoxyadenosine 5'-phosphate (dAMP), Deoxyguanosine 5'-phosphate (dGMP).

Pyrimidine nucleotides: Cytosine (C), Thymine (T). Deoxycytidine 5'-phosphate (dCMP), Deoxythymidine 5'-phosphate (dTMP).

The double helix has constant width: hydrogen bonds between nitrogenous bases. Interior is hydrophobic. Nucleotide subunit: Pyrimidine base, Purine base.

A polynucleotide has a repeating structure: Sugar-phosphate backbone, 5'-3' phosphodiester bonds.

Source: www.BIG10.com

TU Delft "Programming life", 20/2/2013      22      TU Delft Bioinformatics 1.1b 14

## DNA (2)

- In eukaryoten: extreem oopererd

Base pairs: Adenine (A), Thymine (T), Guanine (G), Cytosine (C). Sugar-phosphate backbone. 3 km length of DNA in a cell. 100 cm radius of chromatin fiber. 1 nm diameter of DNA molecule. > cp mother daughter.

TU Delft "Programming life", 20/2/2013      23      TU Delft Bioinformatics 1.1b 14

## DNA (3)

- DNA kan repliceren:

Figure 17-1. Molecular Biology of the Cell, 4th Edition.

1 CHROMOSOME REPLICATION AND CELL GROWTH  
2 CHROMOSOME SEGREGATION  
3 CELL DIVISION

M PHASE: mitosis (nuclear division) and cytokinesis (cytoplasmic division). INTERPHASE: G<sub>1</sub> PHASE, S PHASE (DNA replication), and G<sub>2</sub> PHASE.

> cp mother daughter

TU Delft "Programming life", 20/2/2013      24      TU Delft Bioinformatics 1.1b 14



## Vraag 4

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

**±3**

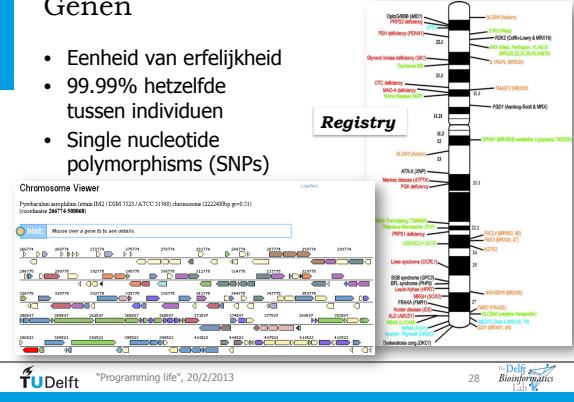
( $3\text{Gb} \times 10^{-9}$  fout/base)

## 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)



## Vraag 5

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

## Genomen

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



Genome	Gene Number	Base Pairs
<b>Organisms</b>		
Reindeer	<50 000	< $10^{11}$
Mammals	30 000	$\sim 3 \times 10^9$
Worms	14 000	$\sim 10^8$
Flies	12 000	$1.6 \times 10^8$
Fungi	6 000	$1.3 \times 10^7$
Bacteria	2-4 000	< $10^7$
Mycoplasma	500	< $10^6$
<b>dsDNA Viruses</b>		
Adeno virus	<300	187 000
Papova (SV40)	~6	5 226
Phage T4	~200	165 000
<b>ssDNA Viruses</b>		
Panavirus	5	5 000
Phage TX174	11	5 387
<b>dsRNA Viruses</b>		
Reovirus	22	23 000
<b>ssRNA Viruses</b>		
Coronavirus	7	20 000
Influenza	12	13 500
TMV	4	5 400
Phage MS2	4	3 569
STNV	1	1 300
<b>Viroids</b>		
PSTV RNA	0	359

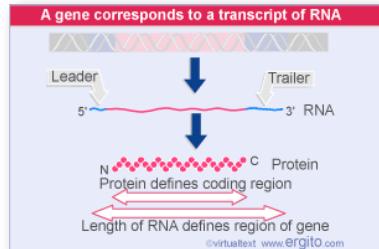
## Vraag 6

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

Het gen wordt afgeschreven in messenger RNA (mRNA)

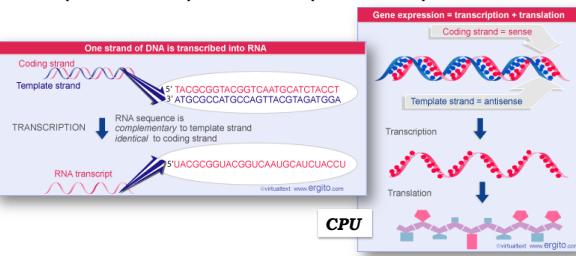
## Genen (2)

- Genen: gebieden op het DNA die coderen voor eiwitten



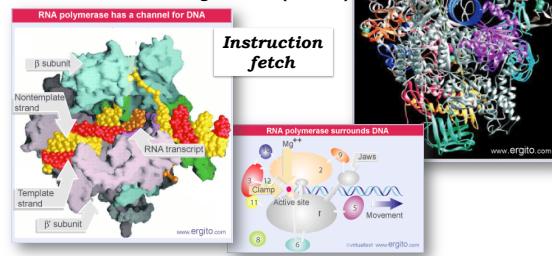
## Genen (3)

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



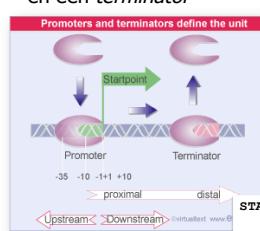
## Transcriptie

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



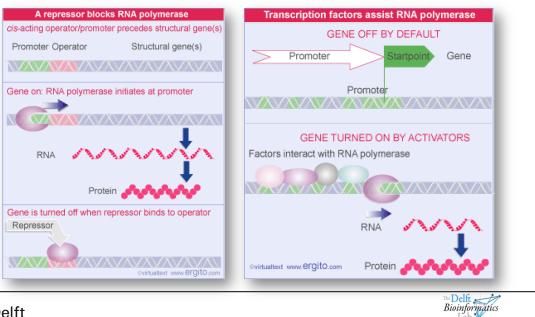
## Transcriptie (2)

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



## Transcriptie factoren

- Onderdrukken ("repress") of activeren transcriptie



TU Delft

```
if (a && !b)
{
    ...
}
```

## Vraag 7

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

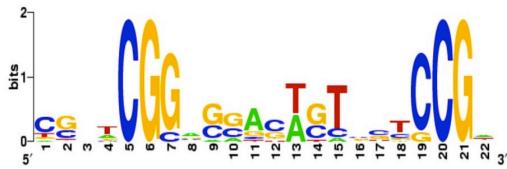


TU Delft "Programming life", 20/2/2013

38 TU Delft Bioinformatics 1.3b

## Transcriptie factoren (2)

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

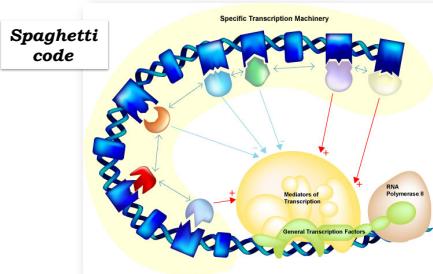


TU Delft

TU Delft Bioinformatics 1.3b

## Regulatie

- Regulatieprogramma's kunnen erg complex zijn



TU Delft "Programming life", 20/2/2013

40 TU Delft Bioinformatics 1.3b

## 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 }	UAU }	UGU }
UUC }	UCG }	UAC }	Cys
UUA }	Ser	UAA }	UGC }
UUG }	Leu	UAG }	STOP
UCA }	UCA }	UAA }	UGA }
CUU }	Pro	CAC }	Trp
CUU }	Leu	CAA }	CGU }
CUA }	Leu	GAA }	CGC }
CUG }	Leu	GAG }	CGA }
CGU }	Arg	CAG }	CGG }
GUU }	Ile	ACU }	His
GUU }	Ile	ACC }	CGU }
GUU }	Ile	ACA }	CGC }
GUU }	Ile	ACG }	CGA }
GUU }	Met	AAC }	Arg
GUU }	Met	AAA }	Arg
GUU }	Met	AAG }	Arg
GUU }	Val	GAC }	Ser
GUU }	Val	GAA }	Asn
GUU }	Val	GAG }	Lys
GUU }	Gly	GAG }	Gly
GUU }	Gly	GAA }	Gly
GUU }	Gly	GAG }	Gly

TU Delft

"Programming life", 20/2/2013

41 TU Delft Bioinformatics 1.3b

## 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**

TU Delft "Programming life", 20/2/2013

42 TU Delft Bioinformatics 1.3b

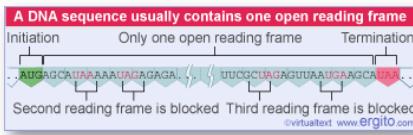
## Translatie (2)

- Reading frame: drie mogelijke vertalingen, bv.

```

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

```



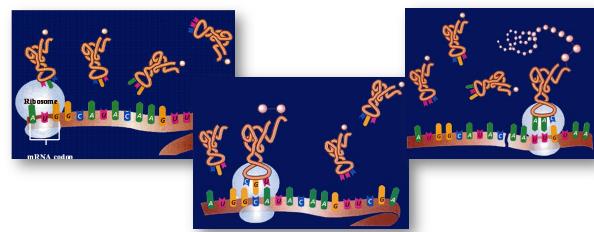
## Vraag 9

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



## Translatie (3)

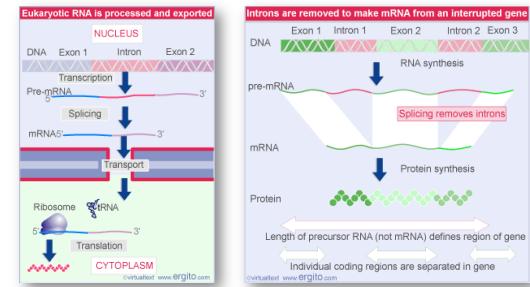
- Messenger RNA wordt vertaald door het ribosoom
- Transfer RNAs dragen een aminozuur en een "sleutel" van 3 basen



## Translatie (4)

**Operator overloading**

- In eukaryoten geeft *splicing* meerdere mogelijke eiwitten



## 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

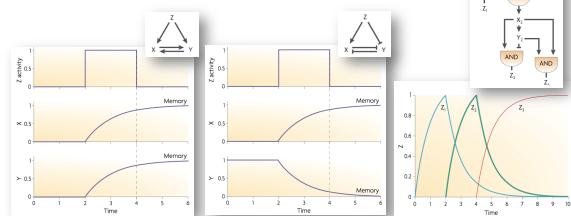
**Self-modifying code?**

## 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

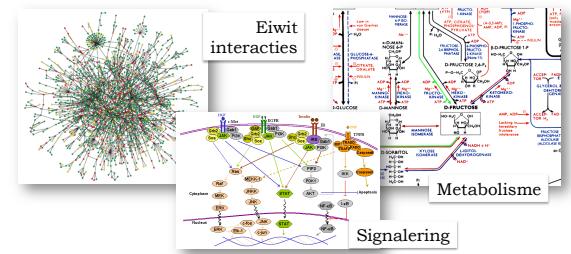


## Overzicht

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

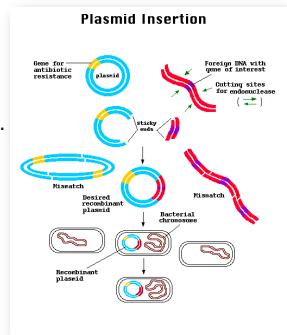
## Moleculaire biologie

- Bestudeert interacties tussen moleculen op verschillende niveaus



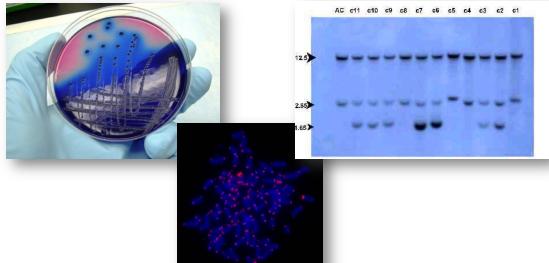
## Experimenten

- 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")



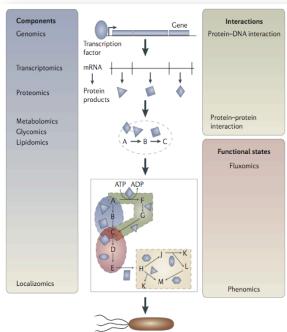
## Metingen

- Traditioneel: groei/overleven, microscopen, gels**



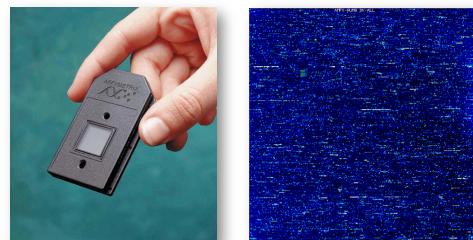
## Metingen (2)

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



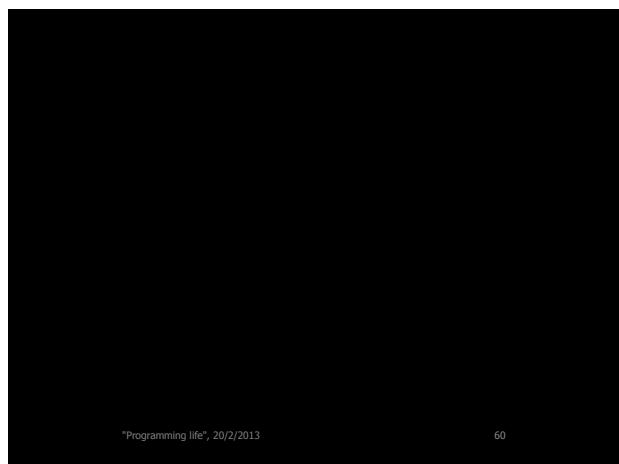
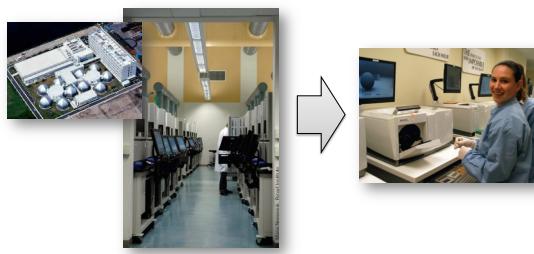
## Microarrays

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

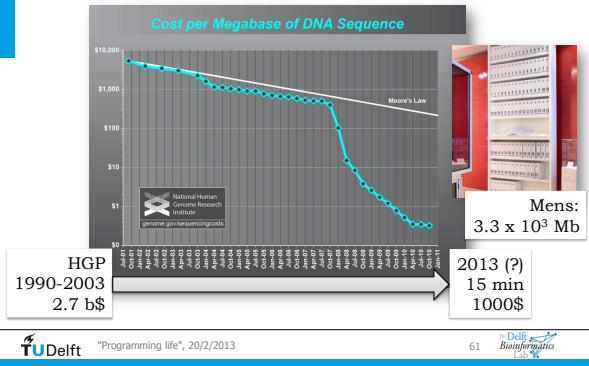


## Next generation sequencing

- Nieuwe ontwikkeling: betaalbaar meten van sequenties



### Next generation sequencing (3)



TU Delft "Programming life", 20/2/2013

61 TU Delft Bioinformatics Lab

### 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 ruzig
5. **Moleculaire biologie wordt voortgedreven door de ontwikkeling van nieuwe apparatuur**

TU Delft "Programming life", 20/2/2013

62 TU Delft Bioinformatics Lab

### Vraag 10

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

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

TU Delft "Programming life", 20/2/2013

63 TU Delft Bioinformatics Lab

### 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 ruzig
5. Moleculaire biologie wordt voortgedreven door de ontwikkeling van nieuwe apparatuur
6. **Als het echt lastig wordt, moeten (bio)informatici te hulp schieten**

TU Delft "Programming life", 20/2/2013

64 TU Delft Bioinformatics Lab

TU Delft "Programming life", 20/2/2013

65 TU Delft Bioinformatics Lab

### Voor volgende keer (27/2)

- Lees materiaal over gen-netwerken en modellen, te vinden op Blackboard: Course Information -> Programming life: Synthetic biology -> Leesmateriaal
- Maak samenvatting, lever in vóór begin college
- Breng een laptop mee, liefst met Matlab!

TU Delft "Programming life", 20/2/2013

66 TU Delft Bioinformatics Lab

## 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.