

Bioinformatica: een nieuw vakgebied dat drijft op toegepaste wiskunde en statistiek

Antwoorden

Opdracht 1

Het menselijk DNA bevat 3.2 miljard ($=3.2 \times 10^9$) nucleotiden, maar omdat DNA dubbelstrengs is zijn er 6.4×10^9 nucleotiden per cel. Gemeen, he? Varkens-DNA heeft ongeveer dezelfde lengte. Slimmerds kunnen over mitochondriaal DNA beginnen, maar dat is verwaarloosbaar op deze getallen. De mens/varken discussie is vrij (dwz: er zijn redenen voor ja en nee. Qua hoeveelheid lijken mensen en varkens dus heel veel op elkaar. Maar qua betekenis, kun je hier vraagtekens bij zetten. Er zijn verschillen, maar die zijn wel weer zo klein dat de varken wordt gezien als portentieel donor voor organen. Hier kun je een hele biologies aan wijden.)

Opdracht 2

If our strands of DNA were stretched out in a line, the 46 chromosomes making up the human genome would extend close to two metres. If the length of the 100 trillion cells could be stretched out, it would be 182 billion kilometres. That is enough material to reach to the sun and back 610 times." *Source:* [Centre for Integrated Genomics](#)

Opm. De schattingen over het aantallen cellen in het menselijke lichaam lopen nogal uiteen. Er kan ook met 10 of 50 triljoen gerekend worden.

Opdracht 3

Ongeveer 24.000.

Opdracht 4

$7 \text{ exp} * 2 \text{ arrays} * 24.000 \text{ genen} * 2 \text{ gen/chip} = 672.000 \text{ data punten}$

Opdracht 5

Aantal te berekenen ratio's = $7*8/2 = 28$ ratio's per gen * 24.000 genen * 2 gen/chip = 1.344×10^6 ratio's = $1.344 \times 10^6 \text{ sec} = 1.344 \times 10^6 / (60*60*8) = 47$ werkdagen van 8 uur. Dat ga je dus niet met een rekenmachine doen. Hiervoor gebruik je een computer.

Opdracht 6

Zeven

Opdracht 7

De Euclidische afstand d tussen gen G_1 met zeven expressieratio's ($a_1 \dots g_1$) en gen G_2 met de zeven expressieratio's ($a_2 \dots g_2$) is:

$$d = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2 + \dots + (g_2 - g_1)^2}$$

Opdracht 8

Voor een set van 24000 genen moeten er $(24.000 * 23999)/2 = 287.988.000$ afstanden berekend worden.

Opdracht 9

$287.988.000 * 0.0001 \text{ sec} = 28798.8 \text{ sec} = 8 \text{ uur}$

Opdracht 10

Stappenplan:

- Bereken eerste de afstand tussen de verschillende genen met de formule uit Opdracht 7
- Begin dan te clusteren volgens de methode beschreven op wikipedia
- Nu zal moeten blijken dat BRCA2 en BRMS1 heel dicht bij elkaar zitten
- Zie verder Opdracht 11

Opdracht 11

De twee genen BRCA2 en BRMS1 spelen allebei een rol bij borstkanker.