Bio-Informatica Juni 2015

1. Beschouw het humane gen JAK2

1. Hoeveel transcripts zijn er volgens UCSC, Ensembl en RefSeq?
2. Zijn er volgens de UCSC CpG Island track CpG-eilanden in de promoter van JAK2? Zo ja, hoe lang is die en wat is het percentage GC
3. Voorspel zelf CpG-eilanden in een regio van 1 Mb rond het JAK2-gen adhv EMBOSS d) Zet je voorspelling als custom track op UCSC. Komt je voorspelling overeen met de CpG-track van UCSC? Hoeveel hebben meer dan 80% overeenkomst?

2. Bekijk in de MySQL-database van de UCSC Genome Browser de tabel over CpG-eilanden.

1. Hoeveel kolommen bevat deze tabel?
2. Hoeveel rijen?
3. Schrijf een mysql query om alle CpG's te vinden langer dan 500 bp
4. Kan je deze info ook krijgen uit de table browser?

3. Bekijk het volgende Python-script dat de smelttemperatuur van een DNA-streng kan voorspellen. Haal de fouten er uit. (lower/upper, integer/float, indentatie…)

4. Schrijf een pythonscript dat het aantal genen in een lijst telt. (het aantal lijnen dus)

5. Beschouw het gen TP53, download de sequentie ervan en voorspel zelf de genen met geneID en vergelijk ze met de RefSeq genen. Hoeveel exonen had je correct voorspeld? Hoeveel exonen kwamen er overeen met de RefSeq genen? Kijk na in welke biologische pathway TP53 werkt, bekom in 1 van deze pathways al de menselijke genen en in dezelfde ook de muis genen. Als je van deze mensgenen de muishomologen neemt, kom je dan op dezelfde lijst muis genen uit? Schrijf een MYSQL command waarbij je het aantal exonen bekomt van de transcripten van tp53 vanuit RefSeq.

6. Beschouw het humaan TBX2 eiwit. Kan je met fuzzpro de T-box vinden in deze eiwitsequentie, (de prosite verwijzing krijg je)? Op welke positie zit deze T-box en welke AZ sequentie heeft die?