Course 1b – Excel opdracht

Maak de opdracht en lever de sheets in. De opdrachten waar BONUS bij staat die zijn optioneel en alleen voor talentvolle studenten.

Zorg dat je 1 mailtje maakt met alle sheets die je gemaakt hebt. Stuur deze mail naar [hanbioinformatics@gmail.com](mailto:hanbioinformatics@gmail.com). **Deadline: donderdag 12 oktober 23:59 uur.**  Deadline herkansing: donderdag 26 oktober 23:59 uur.

Opdrachten basisvaardigheden excel

1. Op OnderwijsOnline vind je het tekstbestand 21213\_hcdiffs bij blok 1 – workshops. In dit bestand staan mutaties (let op, het makkelijkst is dit te doen door eerst Excel te openen en dan het bestand vanuit excel te openen).
2. Zorg dat alle tekst in de kolommen goed leesbaar is
3. Stel een autofilter in zodat je per kolom makkelijk een keuze kunt maken wat je in die kolom wilt zien.
4. Beantwoord de volgende vragen met behulp van draaitabellen
   1. Maak een draaitabel van de kolom “Gene component”. Zet de draaitabel op een apart tabblad. Zorg ervoor dat je de aantallen telt dat dit Gene component voorkomt in je sheet. Je ziet dan o.a. exon region, intron region. Een exon codeert voor een eiwit en een intron niet. Waar vind je de meeste mutaties? Maak deze cel rood.
   2. Maak een draaitabel van de kolom “Mutation”. Je ziet dan een overzicht van de stikstofbasen na mutatie.
   3. BONUS: Maak een draaitabel van “Abberation”, dit is het soort mutatie. Er kan een DNA stiktofbase toegevoegd zijn in de sequentie (insertion), weggehaald zijn (deletion) en er kan een DNA stikstofbase vervangen zijn (substitution). Maak binnen deze draaitabel een overzicht de mutaties per Abberation en dan per Gene component. Hint: zet Gene component bij het vakje “kolommen”.
5. Maak een staafdiagram van de draaitabel van opdracht 4a. Zet titels bij de assen.
6. Bewaar het bestand zodat je het straks kunt mailen.

Opdrachten zoeken in sequenties

Open het bestand:  
<https://www.google.nl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=17&cad=rja&uact=8&ved=0ahUKEwiGjJSsxa7WAhWQLFAKHV_0BfU4ChAWCEcwBg&url=http%3A%2F%2Fwww.heliconius.org%2Fwp-content%2Fuploads%2F2013%2F05%2FHGC_Nature_2012_CSP_OBP_OR_Gene_Predictions.xls&usg=AFQjCNED5EboUfYjr5BKwfbcD7t_0DWZBg>

1. In welke kolom staan de DNA sequenties?

Kolom H

1. Wat betekent CDS sequence?

CDS sequence is een stuk van een gen dat codeert voor een eiwit.

1. Verwijder de kolomen C,D,E,F en de kolommen I t/m T
2. Voeg een kolom toe achter de kolom met de CDS sequences. Deze kolom gaan we gebruiken.
3. Opdracht: schrijf een functie die telt hoe vaak het codon ATG voorkomt in de sequentie.  
   Hint: gebruik de functies VIND.SPEC of VIND.ALLES.   
   Check: heb je nu het aantal ATG’s gevonden of de positie van de eerste ATG? In het laatste geval: bedenk een oplossing om het totaal aantal ATG in de sequentie te vinden.

Met VIND.ALLES("ATG";D2;1) OF VIND.ALLES("ATG";D2) vindt je alleen de positie van het eerste ATG. Met VIND.ALLES("ATG";D2;2) vind je het aantal ATG;s is die sequentie.

1. Bewaar het bestand zodat je het straks kunt mailen.

BONUS opdracht (optioneel, alleen voor talentvolle studenten)

Kopieer de sequenties uit het bestand van de vorige opdracht in een nieuwe excel sheet.

Maak een Alignment tool in excel waarbij je de sequenties met elkaar vergelijkt. De tool geeft een score voor de alignment van twee sequenties. Onderzoek welke sequenties het meest overeenkomen.  
Bewaar het bestand zodat je het kunt mailen.  
HINT: Splits een sequentie op in losse letters waarbij elke stikstofbase (A,C,G,T) van een sequentie in een aparte cel komt te staan.

**STUUR EEN MAIL NAAR** [**HANBIOINFORMATICS@GMAIL.COM**](mailto:HANBIOINFORMATICS@GMAIL.COM) **ZET IN HET ONDERWERP/SUBJECT JE NAAM + EXCELOPDRACHT. VOEG ALS BIJLAGE DE EXCELSHEETS TOE.**