

Bloktoets : **5DT06 Medical Genomics**
Datum : 21 november 2014
Aanvang : 13.00 uur

**Deze tentamenset kunt u na afloop meenemen
Het ANDERE deel ingevuld inleveren bij uw surveillant(e)**

**Het betreft een gesloten boek tentamen, het gebruik van een rekenmachine van het type CASIO
fx-82MS is wel toegestaan.**

ALGEMENE AANWIJZINGEN:

- Dit tentamen bestaat uit **11** open vragen.
- De beschikbare tijd is **2** uur.
- Controleer of uw tentamenset compleet is.
- Vermeld op het antwoordformulier duidelijk uw naam en studentnummer.
- Beantwoord de vragen op de antwoordformulieren in de daarvoor open gelaten ruimten.
- Lees de vragen zorgvuldig alvorens uw antwoord te formuleren.
- Beantwoord de vragen volledig, maar zo beknopt mogelijk; vermijd onnodige uitweidingen.
- Voor beantwoording van de vragen eventueel de achterkant van het formulier gebruiken.
- Schrijf duidelijk leesbaar en gebruik geen afkortingen, het gebruik van een potlood is ongewenst.
- Onleesbaar beantwoorde vragen worden fout gerekend.
- Het gebruik van alle audiovisuele en technische hulpmiddelen is niet toegestaan, tenzij expliciet vermeld elders op dit voorblad. Mocht u dergelijke apparatuur toch gebruiken, dan zal dit als fraude worden aangemerkt. Op uw tafel mogen uw studenten- en registratiekaart en los schrijfmateriaal liggen. Etui's moeten van tafel.
- **Lever na afloop het antwoordformulier in. Indien u commentaar heeft op de vragen, verwijzen we u naar de hyperlink die is opgenomen bij uw toetsindeling in uw webdossier t.b.v. het digitaal studentcommentaarformulier voor deze toets.**

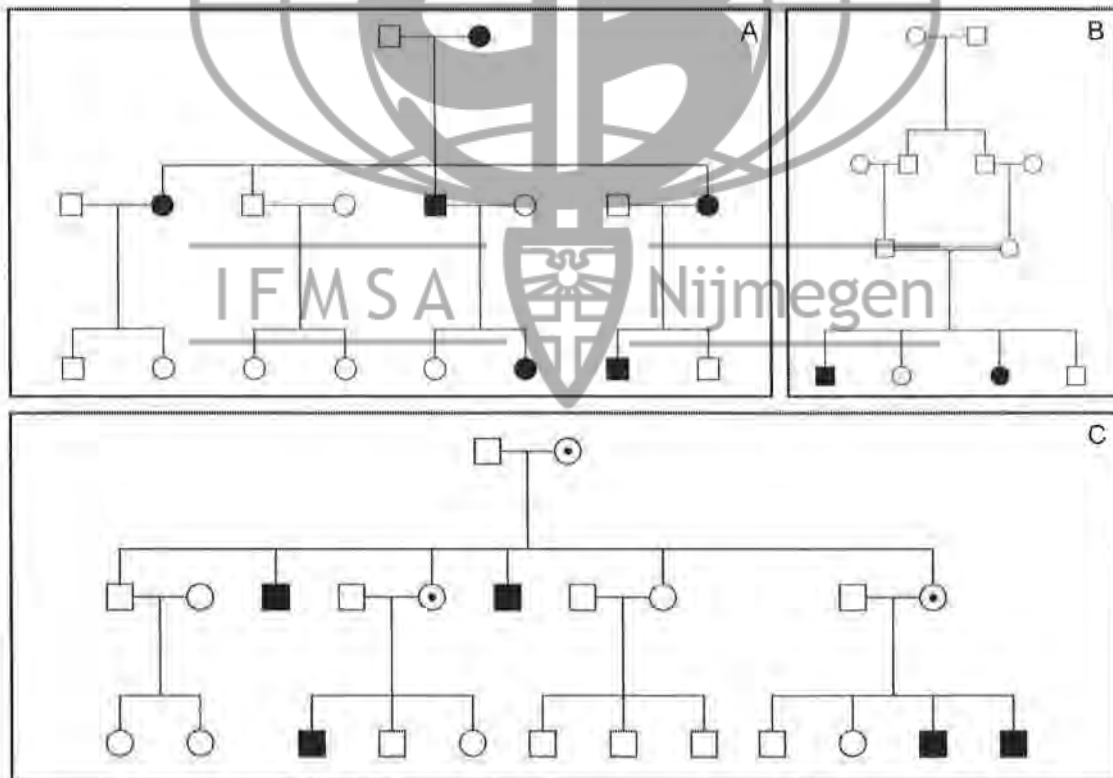
VEEL SUCCES!

LET OP !!

ZET EERST UW NAAM EN STUDENTNUMMER OP ELK ANTWOORDFORMULIER!

Vragen Genetica (Diederik de Bruijn)

1. Next Generation Sequencing (NGS) is een verzamelnaam voor nieuwe technologieën die het sequencen van DNA gemakkelijker en sneller hebben gemaakt. Hieronder vallen onder andere 'whole genome sequencing' en 'exome sequencing'.
 - 1a. Geef in vijf stappen duidelijk aan hoe DNA sequencing met NGS wordt gedaan. Gebruik de juiste terminologie en betrek in uw antwoord ook welk type genetische variatie het makkelijkst met deze technologie kan worden bepaald. (7 punten)
 - 1b. Welk type genetische variatie is lastig te detecteren met NGS? Leg uit waarom. (3 punten)
2. CASUS: Middels exome sequencing heeft u een dominant-negatieve nonsense mutatie in het *P63* gen gevonden in een familie met het EEC syndroom.
 - 2a. Leg uit hoe een nonsense mutatie in het *P63* gen kan leiden tot een dominant-negatief effect op de functies van het p63 eiwit, en betrek in uw antwoord wat een nonsense mutatie is. (6 punten)
 - 2b. Welke van de onderstaande stambomen (A-C) past het best bij de familie waarin u deze *P63* nonsense mutatie hebt gevonden? Licht uw antwoord toe. (2 punten)



- 2c. De, door u bij 2b gekozen, stamboom is onvolledig geannoteerd. Noem twee soorten markeringen die nog aan deze stamboom zouden moeten worden toegevoegd. (2 punten)

De volgende 3 vragen (2d-2f) gaan over het traject dat vooraf is gegaan aan de identificatie van de bovengenoemde *P63* mutatie. Hoewel deze vragen onderling enigszins afhankelijk zijn, gaat het hier om

Meeneemset

een consistente redenering. Een fout antwoord bij 2b betekent dus niet automatisch dat de andere antwoorden ook fout zullen zijn.

- 2d. Ga bij het beantwoorden van deze vraag uit van de keuze die u bij 2b hebt gemaakt. Stel dat u bij aanvang voldoende financiële middelen had om exome sequencing uit te voeren bij vier mensen uit de stamboom, wie had u hiervoor dan het beste kunnen kiezen? Licht uw antwoord zodanig toe dat het duidelijk is *waarom* u voor deze vier personen heeft gekozen. (4 punten)
- 2e. Geef voor elk van de (bij 2d) geselecteerde familieleden aan hoeveel gemuteerde en hoeveel normale P63 allelen u verwacht te vinden in de exome sequencing data. (2 punten)
- 2f. Welk vervolgonderzoek heeft u in de door u gekozen familie moeten doen om de exome sequencing data te verifiëren? Licht uw antwoord toe en schenk hierbij aandacht aan de volgende elementen: 1. De toegepaste moleculaire techniek; 2. Toegepast op welke persoon/personen; 3. Welke uitkomst(en); 4. Verwerking en interpretatie van deze uitkomst(en). (5 punten)

Bepaalde mutaties in de *P53* en *P63* genen worden vaker en op overeenkomstige posities in deze homologe genen gevonden, daarom spreekt men van 'hotspot' mutaties. Eén verklaring voor het ontstaan van zulke mutaties is dat deze genen op die posities blijkbaar gemakkelijk gemuteerd kunnen raken.

- 2g. Welk ander effect zorgt ervoor dat deze mutaties vaker worden gevonden? Licht uw antwoord toe. (4 punten)
- 3a. Onderzoek heeft aangetoond dat de cilia van patiënten met cranioectodermale dysplasie (CED) korter zijn dan de cilia van hun niet-aangedane familieleden. Welk proces verloopt niet goed in de cilia van CED patiënten? Leg tevens uit hoe dit leidt tot het ontstaan van kortere cilia. (4 punten)
- 3b. Bij ciliopathieën, zoals CED, komt vaak intrafamiliaire fenotypische variatie voor. Dit betekent dat twee familieleden beiden dezelfde mutatie dragen, maar dat de ene veel ernstiger is aangedaan dan de andere. Leg uit hoe dat kan en benoem in uw antwoord tenminste twee factoren die (behalve de causale mutatie) een belangrijke rol spelen bij de ernst van de ciliopathie. (4 punten)
4. Warfarine zorgt voor verminderde bloedstolling door het remmen van het *VKORC1* eiwit dat de synthese van bloedstollingsfactoren in de lever reguleert. Het hydrofobe warfarine molecuul wordt actief opgenomen in de darm en getransporteerd naar de lever. Daar wordt warfarine gemetaboliseerd door de *CYP2C9* enzymen en vervolgens via de nieren uitgescheiden. De genotypes van de *VKORC1* en *CYP2C9* zijn daarom belangrijk bij het bepalen van een adequate warfarine dosering.
- 4a. Zal een persoon met een duplicatie van het volledige *VKORC1* gen een verhoogde, een normale of een verlaagde warfarine dosis moeten krijgen? Licht uw antwoord toe. (4 punten)
- 4b. Zal een persoon met een heterozygote aminozuur verandering in de actieve site van *CYP2C9* een verhoogde, een normale of een verlaagde warfarine dosis moeten krijgen? Licht uw antwoord toe. (4 punten)

Vragen Epidemiologie (Sita Vermeulen)

5. Populatiegenetica van EPAS1

Op 2 juli 2010 kopte ScienceDaily.com: "Tibetan adaptation to high altitude occurred in less than 3,000 years". Deze kop is gebaseerd op een studie van Xin Yi *et al.* (Science, 329:75 (2010)), waarvan het abstract hieronder te lezen is:

"Residents of the Tibetan Plateau show heritable adaptations to extreme altitude. We sequenced 50 exomes of ethnic Tibetans, encompassing coding sequences of 92% of human genes, with an average coverage of 18x per individual. Genes showing population-specific allele frequency changes, which represent strong candidates for altitude adaptation, were identified. The strongest signal of natural selection came from endothelial Per-Arnt-Sim (PAS) domain protein 1 (EPAS1), a transcription factor involved in response to hypoxia. One single-nucleotide polymorphism (SNP) at EPAS1 shows a 78% frequency difference between Tibetan and Han samples, representing the fastest allele frequency change observed at any human gene to date. This SNP's association with erythrocyte abundance supports the role of EPAS1 in adaptation to hypoxia. Thus, a population genomic survey has revealed a functionally important locus in genetic adaptation to high altitude."

- 5a. De auteurs wijten het 78% verschil in allelfrequentie voor de genoemde EPAS1 SNP tussen de Tibetaanse populatie (wonende op meer dan 4000 meter hoogte) en de Han populatie (wonende op zeer lage hoogte / zeeniveau) aan 'natuurlijke selectie'. Noem 2 ongerelateerde aspecten uit het abstract die deze hypothese van natuurlijke selectie ondersteunen. (4 punten)

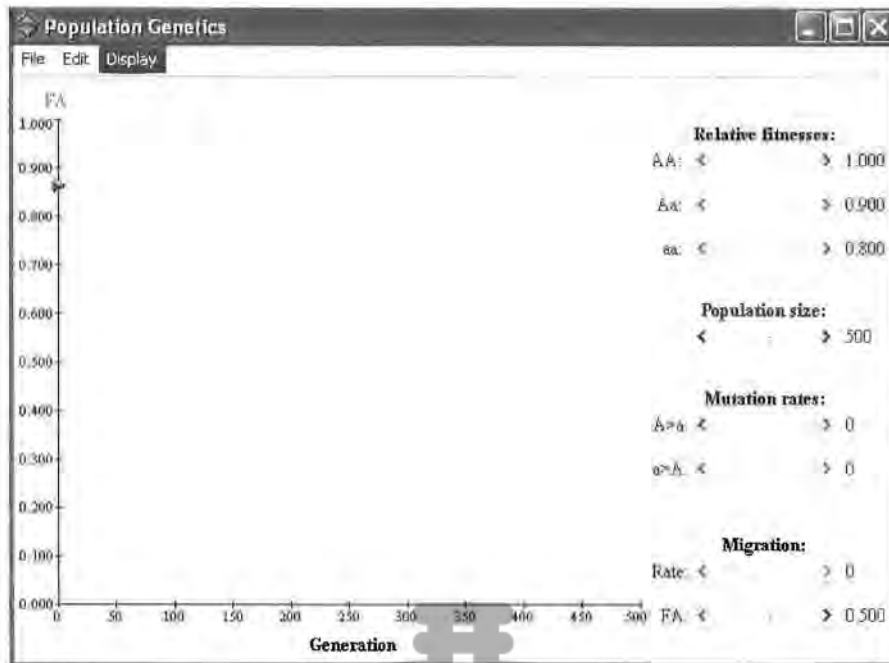
De onderstaande tabel S4 uit deze publicatie beschrijft de frequentie van de EPAS1 SNP en een aantal bijbehorende fenotypes voor drie verschillende populaties.

Table S4: Population frequencies and mean phenotypes at the focal EPAS1 SNP

Allele/genotype	Tibetan frequency	Han frequency	Danish frequency	mean hemoglobin concentration	mean erythrocyte count	mean oxygen saturation
C	0.13	0.9125	1	n/a	n/a	n/a
G	0.87	0.0875	0	n/a	n/a	n/a
CC	10	n/a	n/a	178	5.3	87.5
CG	84	n/a	n/a	178.9	5.6	86.68
GG	272	n/a	n/a	167.5	5.2	86.42

- 5b. De genotype verdeling voor de Han (en Deense) populatie is niet gegeven. Onder welke aanname kunnen we op basis van de allel frequenties de genotype frequenties berekenen? Geef vervolgens deze genotype frequenties voor de Han populatie (afgerond op twee decimalen). (3 punten)
- 5c. In tabel S4 is verder te zien dat het G allel in de Deense populatie geheel afwezig is. Kan 'non-random mating' in de Deense populatie wel of niet hebben bijgedragen aan de afwezigheid van het G allel? Licht het antwoord toe. (4 punten)
- 5d. Hieronder staat een screenshot van het, jullie bekende, programma Selection 3.1. De initiële frequentie van het G allel van 0.87 is aangegeven met de pijlpunt links in beeld. Bestudeer de populatie genetische instellingen voor een simulatie van het verloop van de allelfrequentie van het G allel van de EPAS1 voor de komende 500 Tibetaanse generaties. Teken in deze figuur het meest waarschijnlijke resultaat van deze simulatie. (5 punten)

Meeneemset



6. **Genetische epidemiologie van type 1 diabetes**
Lees onderstaand abstract.

A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene

Hakon Hakonarson^{1,2}, Stuart F. A. Grant^{2,3,4}, Jonathan P. Bradfield^{1,2}, Luc Marchand⁵, Cecilia E. Kim¹, Joseph T. Glessner¹, Rosemarie Grabs¹, Tracy Casalunovo¹, Shayne P. Taback⁶, Edward C. Frackelton¹, Margaret L. Lawson¹, Luke J. Robinson¹, Robert Skraban¹, Yang Lu¹, Rosetta M. Chiavacci¹, Charles A. Stanley¹, Susan E. Kirsch⁷, Eric F. Rappaport⁸, Jordan S. Orange¹⁰, Dimitri S. Monos¹¹, Marcella Devoto¹², Hui-Qi Qu¹ & Constantin Polychronakos¹

Type 1 diabetes (T1D) in children results from autoimmune destruction of pancreatic beta cells, leading to insufficient production of insulin¹. A number of genetic determinants of T1D have already been established through candidate gene studies, primarily within the major histocompatibility complex²⁻⁴ but also within other loci⁵⁻¹². To identify new genetic factors that increase the risk of T1D, we performed a genome-wide association study in a large paediatric cohort of European descent. In addition to confirming previously identified loci²⁻⁹, we found that T1D was significantly associated with variation within a 233-kb linkage disequilibrium block on chromosome 16p13. This region contains *KIAA0350*, the gene product of which is predicted to be a sugar-binding, C-type lectin. Three common non-coding variants of the gene (rs2903692, rs725613 and rs17673553) in strong linkage disequilibrium reached genome-wide significance for association with T1D. A subsequent transmission disequilibrium test replication study in an independent cohort confirmed the association. These results indicate that *KIAA0350* might be involved in the pathogenesis of T1D and demonstrate the utility of the genome-wide association approach in the identification of previously unsuspected genetic determinants of complex traits.

Kijk vervolgens naar de onderstaande tabel uit deze publicatie, hierin staan de topresultaten weergegeven van de GWAS in 1,143 ongerelateerde controlepersonen en 561 ongerelateerde patiënten en in 467 patiënt-ouder trio's.

Meeneemset

Table 1 | TDT and case-control association study results for GWA significant markers

Chr.	SNP	Allele	Case-control cohort				Trio cohort (n = 467)				Locus
			Aff. allele freq. (n = 561)	Ctrl allele freq. (n = 1,143)	OR (95% CI)	P-value	Alleles	Trans:untrans	TDT P-value	P-value combined	
1	rs2476601	A	0.1471	0.08757		1.32×10^{-7}	A:G	137:64	2.62×10^{-7}	1.11×10^{-12}	PTPN22
11	rs1004446	T	0.254	0.3539	0.62 (0.53, 0.73)	4.38×10^{-10}	T:C	160:228	5.56×10^{-8}	6.75×10^{-14}	INS
16	rs2903692	A	0.2834	0.3782	0.65 (0.56, 0.76)	4.77×10^{-10}	A:G	170:251	7.89×10^{-8}	1.03×10^{-10}	KIAA0350
11	rs6356	A	0.4602	0.3593	1.52 (1.31, 1.76)	1.78×10^{-10}	A:G	255:197	0.00637	2.70×10^{-9}	INS
16	rs725613	C	0.3004	0.3898	0.67 (0.58, 0.78)	3.24×10^{-10}	C:A	178:248	6.95×10^{-8}	5.23×10^{-10}	KIAA0350
7	rs10255021	A	0.06667	0.1095	0.58 (0.44, 0.77)	1.16×10^{-11}	A:G	18:57	6.69×10^{-6}	1.71×10^{-8}	COL1A2
11	rs10770141	A	0.2799	0.373	0.65 (0.56, 0.76)	7.20×10^{-10}	A:G	186:234	0.01917	2.95×10^{-8}	INS
1	rs672797	T	0.2257	0.1589	1.54 (1.29, 1.85)	2.67×10^{-11}	T:G	177:119	7.49×10^{-4}	4.20×10^{-6}	IPIN2
16	rs17673553	G	0.2023	0.2791	0.66 (0.55, 0.78)	1.30×10^{-10}	G:A	146:203	0.00228	6.12×10^{-6}	KIAA0350
11	rs7111341	T	0.1843	0.2631	0.63 (0.53, 0.76)	3.77×10^{-10}	T:C	138:185	0.008919	6.90×10^{-6}	INS
11	rs10743152	T	0.271	0.3574	0.67 (0.57, 0.78)	4.73×10^{-10}	T:C	179:233	0.007805	7.53×10^{-6}	INS

Minor allele frequencies, P-values and odds ratios (OR) are shown. The ORs shown are for the minor alleles (as shown in the controls). Combined P-values are also shown. Together with the gene in which the markers reside in which they are closest to. P-values are two-sided (except as noted). Aff. allele freq. allele frequency in affected individuals, Ctrl. allele freq. allele frequency in unaffected individuals, Trans:untrans. ratio of transmitted to untransmitted alleles.

6a. De auteurs rapporteren hier Odds Ratios van 0.65, 0.67, en 0.66 voor respectievelijk SNPs rs2903692, rs725613, en rs17673553 (betrouwbaarheidsintervallen staan tevens in de tabel vermeld). Geef voor de onderstaande uitspraken aan of ze juist of onjuist zijn. Indien u de uitspraak onjuist acht, licht dan toe waarom. (9 punten)

1. **Uitspraak:** "De OR van 0.65 voor rs2903692 geeft aan dat het A allel gepaard gaat met een lager risico op type 1 diabetes dan het G allel; dit is vanuit statistisch oogpunt gezien zeker omdat het 95% betrouwbaarheidsinterval de 1 niet bevat." (3 punten)

- Deze uitspraak is juist.
- Deze uitspraak is onjuist, omdat:

2. De drie SNPs zijn in sterk linkage disequilibrium (in termen van r^2). **Uitspraak:** "Het is dan ook verrassend dat niet slechts één maar alle drie de SNPs een OR van rond de 0.65 laten zien." (3 punten)

- Deze uitspraak is juist.
- Deze uitspraak is onjuist, omdat:

3. De ORs volgen uit een *associatieanalyse* op basis van de genotype metingen in de patiënt-controle studie. **Uitspraak:** "In theorie hadden de auteurs echter ook een linkage-analyse kunnen uitvoeren op basis van de genotype metingen in de patiëntcontrole studie." (3 punten)

- Deze uitspraak is juist.
- Deze uitspraak is onjuist, omdat:

6b. De associatie tussen variatie in KIAA0350 en diabetes wordt niet alleen gezien in de patiëntcontrole studie, maar ook in de patiënt-ouder trio studie. Dit maakt het waarschijnlijker dat de associatie in de patiëntcontrole studie daadwerkelijk een terecht positieve bevinding betreft. Geef hiervoor 2 redenen. (6 punten)

Vragen Bioinformatica (Martijn Huynen)

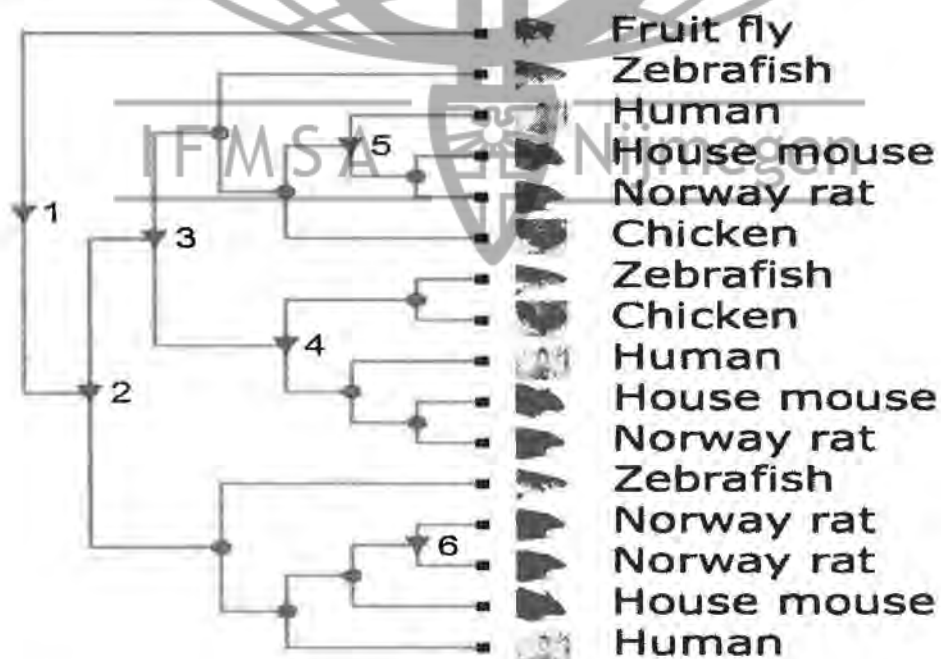
Bij het sequencen van het genoom van een patiënt vinden we in veel genen mutaties die leiden tot de verandering van de aminozuursequentie waarvoor die genen coderen. Sommige van zulke mutaties zullen de functie van het eiwit beïnvloeden, en dus verantwoordelijk voor het ziektebeeld (ofwel pathogeen) zijn, maar de meesten zullen geen effect op de eiwitfunctie hebben en daardoor neutraal zijn. Met behulp van de bioinformatica is het mogelijk om betrouwbare voorspellingen te doen over de vraag welke mutaties het meest waarschijnlijk een effect op de functie van een eiwit hebben.

7. Geef drie onderling verschillende bioinformatische methoden die elk kunnen leiden tot een betrouwbare voorspelling of een mutatie wel of niet verantwoordelijk is voor een ziektebeeld en beargumenteer uw keuze voor elk van deze methoden. (6 punten)

N.B.: ik ben niet op zoek naar de naam van een programma (e.g. "Blast"), maar naar een methode (e.g. "homologie detectie"), en waarom die methode geschikt is voor het doen van (betrouwbare) voorspellingen over het effect van een mutatie op een eiwit.

8. Een van de typen genomics data die ruim voorhanden is, is informatie over de expressie van genen. Hoe kan zulke informatie over de hoeveelheid messenger RNA van genen worden gebruikt om betrouwbare voorspellingen te doen over het biologische systeem waarin een gen dat u wil bestuderen een rol speelt? (3 punten)

9. Hieronder staat een phylogenie van de P63 genfamilie uit de Treefam database. Hierin zijn zes (genummerde) gen duplicaties met een driehoek aangegeven, en negen speciaties met een rondje. De vragen 9a t/m 9c gaan over deze phylogenie.



- 9a. Welke gen duplicatie(s) is/zijn NIET consistent met de gegevens in deze phylogenie (2 punten)?

- 9b. Uit één van de soorten is een P63 homologo verloren gegaan. Uit welke soort? (1 punt)

- 9c. Welke soort heeft P63 homologen die inparalogen van elkaar zijn (1 punt)

Meeneemset

10. Het humane genoom codeert voor een groot aantal functionele elementen. Hoewel we niet van alle functionele elementen uit het menselijk genoom weten wat de functie is, kunnen mutaties in zulke elementen wel ziektes veroorzaken. Hoe is het mogelijk dat we weten dat een stuk humaan DNA een belangrijke functie heeft zonder dat we die functie zelf kennen? (2 punten)

11. In het artikel van PH Dear (2009) Copy-number variation: the end of the human genome, Trends in Biotechnology, 27:448-54 wordt onder andere de grote hoeveelheid copy number variatie in kanker cellen besproken. Hoewel er over de copy number variatie in kanker cellen heel veel data beschikbaar is, blijkt het erg moeilijk om te bepalen welk deel van deze copy number variatie een rol speelt bij het ontstaan van kanker. Waarom is dat zo? (3 punten)

