**Stap 1:** Vinden van de regulerende transcriptiefactoren in je metabolische route

Je moet aan de hand van de motifs van transcriptiefactoren op zoek gaan naar transcriptiefactorbindingsplaatsen (TFBS’en) op en rond promotorsequenties van de genen van het humane genoom in jóuw metabolische route. Op die manier bepaal je door welke transcriptiefactor(en) de genen in jouw metabolische route worden gereguleerd.

* 1. **genen in jouw metabolische route identificeren**
     1. **OPMERKINGEN:**
* **Zoek een manier om een KEGG pathway uit te lezen en alle bijbehorende genen in een lijst te krijgen**
  1. **promotorsequenties genen 1.1 identificeren**
     1. **OPMERKINGEN:**
* **Van een gen uit een (Swissprot?) database moet de lijst met promotorsequenties geladen worden**
  1. **motifs in sequentie 1.2 identificeren**
     1. **OPMERKINGEN:**
* **Van iedere promotorsequentie moet de lijst met motifs geladen worden**
  1. **bijbehorende TFBS’es zoeken**
     1. **OPMERKINGEN:**
* **Het voor mij onduidelijke deel van deze stap.**
* **Eerste ingeving: Genereer de lijst van TFBS’es waar deze motifs bij horen.** 
  + 1. **TODO:**
* **Wat is de relatie tussen een TFBS en de lijst met motifs?**

Stap 2: Vinden van door de transcriptiefactor(en) medegereguleerde genen buiten jouw metabolische route

Je moet met de door jou gevonden transcriptiefactor(en) gaan onderzoeken welke andere nieuwe genen door deze transcriptiefactoren worden gereguleerd buiten jouw metabolische route. De genen die worden gevonden moeten natuurlijk wel de beste genen zijn van alle gevonden genen. De manier waarop je bepaalt welke genen de beste zijn, dien je in je code en documentaties te omschrijven.

* 1. **Uitzoeken welke genen door de TF’s gereguleerd worden**
     1. **OPMERKINGEN:**
* **De lijst met TFBS’es heeft waarschijnlijk een attribuut met genen waarop deze TFBS werkt. Vind ze, 007.**
  1. **Hits filteren op aanwezigheid pathway**
     1. **OPMERKINGEN:**
* **We zijn alleen geïnteresseerd in de genen die niet op onze pathway zitten. Deze worden in het filter opgevangen en verwijderd.**
  1. **Hits filteren op kwaliteit**
     1. **OPMERKINGEN:**
* **The best are only barely good enough. Alleen de beste genen worden geselecteerd voor ons automatische verslag.**

Stap 3: Meer biologische achtergrondinformatie vinden van door jouw gevonden medegereguleerde genen

Na stap 2 heb je nieuwe genen gevonden die ook worden gereguleerd door dezelfde transcriptiefactor(en) van jouw metabolische route. Het is belangrijk dat je over deze nieuwe genen meer informatie probeert te vinden door de volgende biologische vragen erover te kunnen beantwoorden.

1. **Biologische vragen beantwoorden**

**3.1 Enzymen**

**3.1.1 Hoeveel van de 100 best gevonden medegereguleerde genen zijn enzymen?**

**3.1.2 Hoeveel van deze enzymen hebben een bekende functie?**

**3.1.3 Geef deze verhoudingen in een Venndiagram weer**

**OPMERKINGEN:**

* + - * **Enzym/geen-enzym? Verschil is E-code (of hoe die ID ook heet)**
      * **Zou er een attribuut zijn met “Functie” die leeg is als deze niet bekend is?**
      * **Hoe maak je een Venn-diagram in R? Shouldn’t be too hard. (Deel1(Overlap)Deel 2).**

**3.2 MSA**

**3.2.1 Maak een MSA van de 50 beste genen**

**3.2.2 Bereken het conservatie % van het langste gen**

**3.2.3 Hoe heet dit gen?**

**OPMERKINGEN:**

* + - **Dit zal niet de grootste uitdaging zijn.. Denk ik.. Een MSA van 50 genen zal daarentegen niet in een minuutje gedaan zijn. Beware!**

**3.3 Eiwitten**

**3.3.1 Bepaal de eiwitten van de beste 10 genen**

**3.3.2 Bepaal voor elk eiwit het percentage hydrofoob/hydrofiel/neutraal**

**3.3.3 Geef 3.3.2 weer in een staafdiagram met 10 staven (1 per eiwit)**

**3.4 Fylogenie**

**3.4.1 Bepaal de 4 beste paralogen voor de 10 beste genen**

**3.4.2 Maak een fylogenetische boom**

**3.5 Splicing**

**3.5.1 Wat is de cumulatieve exonlengte voor de 20 beste genen?**

**3.5.2 Wat is de cumulatieve intronlen1gte voor de 20 beste genen?**

**3.5.3 Maak scatterplot van lengten genen**

**(x-as: cumulatieve intronlengte y-as: cumulatieve exonlengte)**

1. **Automatisch genereren van het pdf document (LaTeX)**

**OPMERKINGEN:**

* **Is misschien een grotere klus dan van tevoren bedacht. Beware! Dag inplannen voor beide voor dit onderdeel!**