

ITBC Bi5 Proteomics & Programmeren in Java

Weektaak 3-7

OWE5 Proteomics

Taak 3-7 – Onderzoeksopdrachten Project

Centrale Opdracht

Mitogen-activated protein kinases (MAPKs) fosforyleren een groot aantal eiwitten die betrokken zijn bij cellulaire signaleringsprocessen in eukaryoten. In planten is het aantal geïdentificeerde MAPK substraten die de plant defense response reguleren beperkt. MPK3 en MPK6 zijn stress-geactiveerde MAPKs.

Lassowskat et al (2014) heeft transgene *Arabidopsis thaliana* planten gemaakt met een induceerbaar systeem om *in vivo* de activatie van MPK3 en MPK6 te simuleren. Hiervoor zijn wild-type Colombia (Col-0) planten getransformeerd met een constitutief actieve variant van MKK5 (MKK5^{DD}) uit *Petroselinum crispum* onder controle van een dexamethason(DEX)-induceerbare promotor. Als controle is het Col-0 genotype ook getransformeerd met een kinase-inactieve MKK5 mutant (MKK5^{KR}) Metaboloom analyse liet zien dat deze artificiële MPK3/6 activatie, zonder blootstelling aan pathogenen of andere stress-factoren, genoeg is om de productie van belangrijke defense-gerelateerde metabolieten te activeren.

Om (gefosforyleerde) eiwitten te identificeren die mogelijk zich downstream in de signaleringsroute van MPK3 en MPK6 bevinden is door Lassowskat een MS proteomics experiment uitgevoerd. Hiervoor is het construct met constitutief actieve variant van MKK5 (MKK5^{DD}) ook getransformeerd in *mpk3* en *mpk6* mutant planten. Met behulp van proteomics (MS-MS) zijn in elk genotype de aanwezige eiwitten geïdentificeerd en is de relatieve concentratie van deze eiwitten bepaald. De concentratie van de eiwitten is steeds vergeleken met de concentratie van eiwitten voor DEX stimulatie om te bepalen welke eiwitten up-gereguleerd zijn en welke down-gereguleerd zijn als gevolg van MKK5 activatie. Om te bepalen of de gevonden eiwitten inderdaad MAPK substraat kunnen zijn is bekeken met een voorspellings tool of ze de typische MAPK-target fosforyleringssite bevatten. Met MS-MS is vervolgens bepaald of deze peptides inderdaad worden gefosforyleerd.

Een aantal van de gevonden mogelijk MAPK gefosforyleerde eiwitten blijken geassocieerd te zijn met de biosynthese van anti-microbiële defense metabolieten (bv WRKY transcriptie factoren en eiwitten gecodeerd door de genen van de “*PEN pathway*” nodig voor “*penetration resistance to filamentous fungi*”).

De proteomics data is beschikbaar onder de accessiecode PXD001252 via EBI Pride Archive, <http://www.ebi.ac.uk/pride/archive/>). Van deze dataset heb je alleen de files nodig die staan beschreven in Data-tabel.xlsx.

I. Lassowskat, C. Böttcher, L. Eschen-Lippold, D. Scheel and J. Lee. (2014). Sustained mitogen-activated protein kinase activation reprograms defense metabolism and phosphoprotein profile in *Arabidopsis thaliana*. *Frontiers in plant science*. Vol 5: 554.

Onderzoeksplan

Het onderzoek van Lassowskat stamt uit 2014. Ondertussen is er meer bekend over de eiwitten van *Arabidopsis*. Het herhalen van de resultaat analyse door bio-informatici met de huidige kennis kan wellicht beter tot betere resultaten leiden.

Schrijf een onderzoeksplan om een deel van de data van Lassowskat (zie [Data-tabel.xlsx](#)) opnieuw te analyseren met als doel om eiwitten te identificeren die zich mogelijk downstream in de signaleringsroute van MPK3 en MPK6 bevinden. Baseer je onderzoeksplan op de volgende paragrafen.

In elke paragraaf hieronder staat een aantal opdrachten, aangegeven met een bolletje. En een aantal vragen, aangegeven met een vierkantje. De vragen zijn bedoeld om te helpen bij het analyseren van de data.

Geef in het onderzoeksplan de volgende onderdelen duidelijk aan:

- Aanleiding/context
 - Welke biologische data ga je onderzoeken?
 - Hoe is deze data verkregen?
- Onderzoeksvragen: Hoofdvraag en deelvragen.
- Materiaal en methode
 - Op welke manier ga je de data onderzoeken? Werk de methode zo ver mogelijk uit (databases, tools, parameters, wat gaat het script dat je gaat schrijven precies doen, etc.).
 - Geef aan waarom je de gekozen methode gaat gebruiken. Welke keuzes heb je gemaakt? Welke resultaten kun je verwachten? Wat zijn waarschijnlijk de knelpunten in het onderzoek.
 - Vermeld referenties naar wetenschappelijke publicaties aan behorend bij de gebruikte tools.
- Tijdplanning en takenverdeling
 - Maak een gedetailleerd schema van de planning.
 - Voor elke stap/onderdeel wordt een persoon uit de projectgroep verantwoordelijk gesteld. Deze persoon zorgt ervoor dat de activiteiten worden uitgevoerd en dat de tussenresultaten op tijd bij de expert worden ingeleverd. Zorg er voor dat je niet op elkaar gaat zitten wachten. Dat is zonde van de tijd. Misschien is het mogelijk om taken te paralleliseren.
 - Visualiseer de planning in een flowschema inclusief een duidelijk onderschrift.

Begin met het bedenken hoe je met je projectgroep het onderzoek moet gaan uitvoeren (te gebruiken tools/databases, instellingen, parameters, automatisering etc) en maak een werkschema. Stel vragen aan de expert en bediscussieer de manier waarop je het onderzoek gaat uitvoeren en het werkschema met de andere projectgroepen in de werkbespreking tijdens het eerstvolgende tutoruur. Werk na het tutoruur het plan verder uit en lever het plan uiterlijk op vrijdag (30 september 2016) in bij de expert

De tutoruren zullen de rest van het project fungeren als werkbesprekingen. Laat elke week zien wat je hebt gedaan, wat de resultaten zijn, en geef aan waar je moeilijkheden hebt ondervonden en eventueel hoe je deze moeilijkheden hebt opgelost.

De tijdsplanning is erg belangrijk. Je hebt dus bijna twee weken om het onderzoeksplan gedetailleerd uit te werken. Dat betekent natuurlijk niet dat je daarna pas kunt beginnen met het onderzoek. Je hebt tot en met week 6 de tijd om aan het onderzoek te werken en de resultaten te produceren. In week 7 heb je dan de tijd om de resultaten van het onderzoek vast te leggen in een onderzoeksverslag (uiterste inleverdatum vrijdag 4 november 2016 12:00 uur). Een goede planning met vastgestelde tijdstippen voor deelresultaten (mijlpalen) zorgt ervoor dat je aan het eind op tijd klaar bent met het onderzoek en het onderzoeksverslag.

Inwerken op een nieuw project: Data en onderzoeksvragen opstellen

- Download het artikel van Lassowskat et al. 2014 en download de supplementary data (presentation 1, presentation 2 en presentation 3).
- Download een deel van de dataset PDX001252 via EBI Pride Archive. In de [Data-tabel.xlsx](#) kan je vinden welke samples je voor het project nodig hebt.
- Visualiseer de data in Pride Inspector. Te downloaden via Pride Archive.
- Beschrijf de samples. Wat zijn de samples die je gaat analyseren. Waaruit bestaan de samples. (Zie [Data-tabel.xlsx](#) tab 2 en M&M in het artikel van Lassowskat et al 2014 of beschrijving van de data in Pride Archive).
- De eiwitten die zich bevinden in de samples worden upgereguleerd, downgereguleerd of irregular gereguleerd. Wat betekent dit?
- Welke deel-onderzoeksvragen kun je op basis van deze data stellen om eiwitten te identificeren die mogelijk anderszins zich downstream in de signaleringsroute van MPK3 en MPK6 bevinden? Voor het opstellen van de deelvragen kan je hieronder spieken, maar misschien kan je ook nog andere vragen bedenken die je wilt gaan onderzoeken.

Inwerken op een nieuw project: MAPK signalering in *Arabidopsis*. Wat is er al bekend?

Bestudeer de MAPK cascade in *Arabidopsis* in KEGG ([ath04626](#), ([ath04140](#))). Deze informatie kan je gebruiken voor de inleiding van je verslag om te beschrijven wat er al bekend is. In de discussie kun je je resultaten vergelijken met deze informatie.

- Wat zijn de signalen in de plant-pathogen interaction pathway?
- Wat zijn de receptoren?
- Hoe wordt een MAPK geactiveerd?
- Hoe wordt het signaal doorgegeven in de MAPK-cascade?
- Wat voor soort eiwitten worden door de MAPK-cascade geactiveerd, i.e. downstream van de MAPK cascade?
- Is bekend wat de respons is van het signaal waarmee de MAPK-cascade wordt geactiveerd? Hoe komt deze respons tot uiting?
- Wat is er verder (buiten KEGG) bekend over de substraten van de *Arabidopsis* MAPK cascade?

Onderzoek: Consensusproteoom vs overlap tussen twee samples

De eiwitten die differentieel tot expressie komen zijn niet in alle samples gelijk. Het consensusproteoom is de set proteïnen die in alle samples differentiële expressie laat zien. Daarnaast bestaat er natuurlijk een overlap in proteïnen tussen de samples.

In figuur 4B (Lassowskat et al. 2014) is een overzicht gegeven van het aantal eiwitten met een verschil in expressie tussen MKK5^{DD} planten en MKK5^{KR} planten na DEX inductie (0-24 hr).

Eiwit concentratie is bepaald met behulp van Progenesis QI LC-MS software*, door te kijken naar het aantal peptides dat per eiwit wordt gemeten. Hiervoor worden alleen peptides gebruikt die maar aan één eiwit kunnen worden gelinkt (relative quantitation using non-conflicting peptides). De concentratie is genormaliseerd ten opzichte van de maximum concentratie (Supplement tabellen S03, S04, S05 en S06). Per eiwit is de maximum concentratie, gemeten in één van de replicaties van de 0-24 h samples, op 1 gezet en alle andere waarden zijn hieraan gerelateerd. Een waarde van 0 komt overeen met geen peptide van dit eiwit gemeten.

** Om Progenesis QI LC-MS software te kunnen gebruiken is helaas een licentie nodig. Wij kunnen dus niet zelf eiwitconcentraties berekenen maar wel proberen de geanalyseerde data van Lassowskat te begrijpen.*

- Waarom kan je gebruik maken van het aantal peptides om de relatieve concentratie van het eiwit te bepalen?
- Waarom relatieve concentratie en niet de absolute concentratie?
- Waarom is het nodig om te normaliseren per eiwit om uiteindelijk de eiwitconcentraties tussen de eiwitten onderling te kunnen vergelijken?
- Bepaal de proteïne overlap tussen de samples (Col-0 KR, Col-0 DD, DD/*mpk3*, DD/*mpk6*).
- Bepaal het consensus proteoom van Col-0 KR, Col-0 DD, DD/*mpk3* en DD/*mpk6*
- Bepaal welke proteïnen uniek zijn voor een sample.
- Maak een Venn-diagram om de grootte van het consensus proteoom, de overlap en het aantal unieke eiwitten aan te geven.

- Kun je de verschillen tussen het aantal eiwitten per sample biologisch verklaren? Zie ook Lassowskat et al. 2014.
- Zie je verschillen in concentratie tussen de samples van bepaalde eiwitten? Wat zijn dit voor eiwitten?
- Welke lijstjes van eiwitten bevatten eiwitten die zich mogelijk downstream van MKK5 en MPK3 of MKK5 en MPK6 bevinden?

Onderzoek: Protein-protein interactie netwerken

De DEX behandeling van de verschillende MKK5DD planten heeft een groot effect op eiwitconcentraties. Uit de MS-data komen lijsten van eiwitten die aanwezig zijn per sample. Hiervan heb je een overzicht gemaakt in een Venn diagram. De aantallen eiwitten die in de lijsten staan zijn er zoveel dat we nog niet duidelijk weten wat er nu downstream van MKK5 en MPK3/MPK6 gebeurt. Om inzicht te krijgen in wat de functie is van de eiwitten in de samples gaan we de data visualiseren.

Een mooie manier om de data te visualiseren is in de vorm van een eiwit-eiwit interactie netwerk. Eiwitten die interactie met elkaar aangaan zijn vaak betrokken bij hetzelfde proces. In dit interactienetwerk kan je daarom bekijken mbv GO-termen wat de functies zijn van de eiwitten die in het netwerk aanwezig zijn of bij welke processen ze zijn betrokken.

In het Venn diagram dat je gemaakt hebt staan de aantallen eiwitten die differentieel tot expressie komen tussen verschillende samples. Welke eiwitten dat zijn zal je uit de lijsten moeten halen. De *Arabidopsis* locus identifiers van de eiwitten in een lijst kunnen worden geupload in de STRING database. Gevisualiseerd worden dan de eiwit-eiwit interacties tussen de eiwitten uit de lijst.

- Visualiseer de eiwitten geïdentificeerd in de Col0-DD, Col0-KR, DD/mpk3 en DD/mpk6 samples (Supplement tabellen S03, S04, S05 en S06) in STRING eiwit-eiwit interactie netwerken op zo'n manier dat je inzichtelijk maakt wat er downstream van MKK5 en MPK3/MPK6 gebeurt.
- Lassowskat heeft STRING interactome netwerken gemaakt van de up-gereguleerde proteïnen van Col-0 en de *mpk3* mutant (Figuur 5 in van Lassowskat et al. 2014). In de discussie van het artikel wordt gepoogd de verschillen biologisch te verklaren. Vergelijk de biologische functie van jouw gevonden eiwit-eiwit interactie clusters met de resultaten van Lassowskat.

Analyseren van de MS-MS data

Voor het begrip van MS-MS en ter verificatie van de data deze opdracht. We doen dit tijdens de les. Dit onderdeel hoeft niet beschreven te worden in het verslag.

MS-MS data bestaat uit spectra met y-type ionen en b-type ionen. De data die jullie in Pride Inspector hebben gedownload is reeds geanalyseerde data. Elk sample is geanalyseerd mbv Matrix Science Mascot version 2.4.0 om aan de hand van de gemeten peptiden de bijbehorende eiwitten te identificeren. De gebruikte settings voor de analyse staan onder de tab overview – instrument&processing, let ook goed op de titel van het sample dat je bekijkt. In de titel staat met welke variabele modificatie er rekening is gehouden. Er zijn samples die gefilterd zijn (Progenesis ANOVA p-value<0,05; Max. fold change >2.0) en samples die niet gefilterd zijn (Progenesis no parameters).

Onder de tab Protein vind je de geïdentificeerde eiwitten (#PSM is totaal aantal peptiden geïdentificeerd van het eiwit, #Distinct peptides is aantal verschillende peptiden geïdentificeerd van het eiwit. Sommige peptiden zijn dus meerdere keren geïdentificeerd. #PTMs is het aantal post-translationele modificaties.

Onder de tab spectrum een overzicht van alle MS spectra op basis waarvan de peptiden zijn geïdentificeerd. Via Export in de spectrum tab is het mogelijk om een .mgf bestand op te slaan met daarin alle massa's en intensiteiten van de gemeten fragmenten per spectrum. Elk MS-MS spectrum hoort bij één peptide (precursor).

- Hoe is de MS-MS data verkregen? (Zie M&M artikel, Pride Archive beschrijving, Data-tabel.xlsx)
- Kies een eiwit (tabblad protein). Kies van dat eiwit één peptide (nog steeds tabblad protein).
- Zoek bij het uitgekozen peptide de ID van het bijbehorende ms/ms spectrum. Dit kan je doen op basis van het spectrum ID die je kunt zien via de knop "Display Spectrum Properties".
- Exporteer de spectrum data van het sample in een .mgf bestand (tabblad spectrum).
- Met het ms/ms spectrum ID kan je in het .mgf bestand zoeken naar het spectrum van jouw gekozen peptide mbv TITLE= .
- Kopieer het ms/ms spectrum van dat ene peptide (begin ions t/m end ions) naar een nieuwe file en gebruik deze file om het ms/ms spectrum van je peptide te analyseren in Mascot. Kan je het goede peptide terugvinden met een significante score? Welke settings moet je daarvoor gebruiken?
- Wat betekenen de resultaten? Score, Expect (E-value) etc.
- Per sample is een bepaalde variable modification gebruikt om peptiden te identificeren. Waarom wordt per Mascot-search maar één variable modification gebruikt en niet alle vier de gebruikte variable modificatie opties tegelijkertijd?
- Het is in Mascot mogelijk om in één keer meerdere spectra te analyseren? Hoe zien de resultaten er dan uit?

Onderzoeksverslag

Schrijf een onderzoeksverslag met een wetenschappelijke opbouw. Beschrijf je methoden, presenteer je resultaten en bediscussieer de data-analyse en de biologische interpretatie van de data. Verwerk de antwoorden op de vragen in het verslag. **Uiterste inleverdatum: nog nader te bepalen**

Copyright HAN, I.Paffen 2018