**Proteomics Extravaganza**

**Experimentele setting/flowthrough**

* MZmine3
  + looking for raw data
  + install via windows
  + Data inladen 🡪 Raw data overview
    - O.b.v. 1 m/z waarde alle pieken eruit halen
    - Andere kleine pieken = wss isoptopen
* Inleiding
  + EWn = grootste # i/d cel 🡪 MC-MS = methode om EWn te onderzoeken
  + AZn 🡪 ≠ klassen
    - +, -, hydrofiel, hydrofoob, “special cases” (cys, gly, pro)
* LC-MS recap
  + EWn 🡪 digestie 🡪 peptide 🡪 scheiding (LC) 🡪 ionizatie 🡪 MS\*
    - \*ion mobility (IM) (scheiding op charge & size)
  + 🡪 MS1 (precursors vinden) 🡪 MS2 (fragmentatie)
  + Trypsine 🡪 EWn digesteren na R of K (OPM: K of R kan ook geskipt worden)
    - Min # AZn = 5-6
    - OPM: amino terminal = ook + geladen
      * Na trypsine knip 🡪 meestal 2 + ladingen op peptide
  + MS
    - Blg. dat EWn = geladen
    - ESI 🡪 solvent + peptide in cylinder + hoog voltage op cylinder
      * Peptide + solvent 🡪 bubbelvorming 🡪 perifeer + lading 🡪 bubbels worden altijd kleiner 🡪 naar MS sturen
    - IM 🡪 o.b.v. lading & vorm worden peptiden verder gescheiden
      * Via EM golf 🡪 grote peptiden worden meer naar achter geduwd + alles wat 2-3-4 ladingen heeft voelt EM veld sterker -🡪 Hoe groter & hoe meer ladingen peptide heeft, hoe meer weerstand
      * Alles wat meer wind vangt 🡪 wordt verder getrapt dan alles wat minder wind vangt
* LC
  + Stationary phase
    - Vaste beads in kolom 🡪 ftionele groep op beads
      * Reverse phase 🡪 acyl chain (veel C’s 🡪 lipid achtige vorm)
      * Hydrofoob
  + Mobile phase
    - 2 solventen + sample 🡪 onder hoge druk door kolom geperst
      * Solvent A = vaak formic acid of LC grade H2O
      * Solvent B = FA, LC grade H2O, ACN, MeOH (hydrofoob solvent)
        + Zorgt dat peptide makkelijk uit kolom kunnen
* Chromatogram
  + Meten via (UV detector of via) MS
  + ≠ elutie gradiënten 🡪 +/- totale som van alle gradiënten
  + Vaak meerdere pieken 🡪
    - niet meteen goed resultaat
    - Wijst dat er niet veel peptiden in sample zitten
* MS1
  + Isotopen 🡪 ≠ abundanties
  + m/z kan gemeten worden via
    - quadrupole
      * Iets minder accuraat
      * 4 staven
      * Peptiden naar boven & beneden sturen
      * Alles wat niet voldoet aan bepaalde m/z waarde wordt eruit gedrukt 🡪 uit balans & uit quadrupole gestuurd
      * Enkel peptiden met geschikte m/z waarden bereiken einde
    - ion trap
      * Iets minder accuraat
      * Quasi zelfde principe als quadrupole
      * Peptiden trachten stil te zetten in het midden via EM velden
      * Hierna peptiden hard heen & weer bewegen tot deze eruit vliegen
      * Kracht van heen & weer schudden kan veranderd/gemanipuleerd worden
    - time of flight
      * accurater
      * zwaar peptide zal later aankomen dan kleiner/lichter peptiden indien zelfde kracht wordt uitgevoerd
      * probleem:
        + Manier waarmee peptiden vooruit worden gedrukt kan variëren
        + Oplossing 🡪 gebruik van reflectrons 🡪 peptiden laten afbuigen 🡪 zorgen dat peptiden van zelfde grootte op zelfde moment aankomen
      * T = blg. factor, zeker niet te warm
    - Orbitrap
      * Accurater (meest gekozen)
      * Huls met spindle 🡪 peptiden rond spindle laten draaien met constante kracht
      * m/z bepaalt hoe peptide draait rond spindle
      * Sinus signaal voor 1 peptide 🡪 bij vele ≠ peptiden krijgen we een meer chaotischer signaal 🡪 fourier transformation 🡪 piekensignaal
  + Data
    - Piekenpatroon 🡪 Intense & minder intense pieken
    - Isotopen worden ook weergegeven
    - Afstand tss. pieken laat lading zien
      * Vb. afstand van 0,5 tss. pieken 🡪 wijst op lading 2
      * m/z = m/2 🡪 effect van 0,5
    - Soms bep. moleculen die meer isotopen bevatten dan “normale”
    - Bij zeer grote moleculen 🡪 grotere kans dat deze isotopen bevat 🡪 monoisotoop piek zal redelijk klein zijn
* MS2
  + 5-6 pieken uit MS1 scan selecteren om verder te bestuderen 🡪 fragmenteren
  + Fragmentatiespectra 🡪 meest intense pieken uit MS1 worden geselecteerd
  + 1 massa eruit selecteren 🡪 door quadrupole 🡪 fragmentatie
    - OPM: duizenden kopieën worden door quadrupole gestuurd 🡪 fragmentatie met HCD 🡪 creëert ≠ fragmenten
      * a,b,c & x,y,z fragmenten

**Software verwerking voor deze (proteomics) data? (search engines)**

* Proteomics datasets 🡪 combo van MS1 & MS2 spectra
* PRIDE
  + Publieke data
  + Online repository
  + PXD identifiers 🡪 uniek nummer voor elk project
  + Meta-data klopt meestal niet 🡪 terug naar manuscript om te checken
  + Data kan gedownload worden
* XI Tandem & OMSSA 🡪 frequent gebruikte search engines
* Doel MS2 spectra teruglinken aan originele peptiden
  + Kan handmatig, maar zeer lastig
  + Computationeel
    - Sequence DB search
      * Lijst met EW-seqs 🡪 processing 🡪 theoretische spectra 🡪vgln met experimentele spectra 🡪 scoren 🡪 beste matchen uithalen
    - Spectral library search
      * DB creëren met spectra waarvan je zeker bent dat deze een peptide voorstellen
    - Sequence tag-assisted search
      * Niet echt blg. 🡪 de novo methode
    - De novo
      * Hele seq meten
* Seq. DB search
  + EW DB 🡪 digestion 🡪 fragmentation 🡪 theoretical spectra
  + Peptide lijst wordt bekomen
* 1. EW DB 🡪 tekst file + header + EW seq
  + Search space 🡪 bevat alle EW die verwacht zijn om tegen te komen
    - Vb. alle EWn in NFkB pathway
    - OPM: common contaminants EWn best ook opnemen (vb. Keratine of trypsine)
  + DB maken?
    - Compomics.com 🡪 tutorials 🡪 bioinformatics for proteomics 🡪 identification 🡪 1. DB generation
    - UniProt 🡪 proteomes 🡪 zoeken op ref. proteome van mens 🡪 opslaan als FASTA file (opslaan als .fasta file)
      * OPM: keratine moet niet extra toegevoegd worden want zit sws al in proteoom van mens
  + OPM: vragen oplossen op site
* 2. In silico digestion
  + EWn na elke K & R knippen
    - OPM: Indien K door P gevolgd wordt 🡪 NIET knippen
    - OPM: biologie ≠ perfect 🡪 miscleavages toelaten (+/- tss 0 & 2 toelaten)
  + Filteren op lengte
    - Alles korter dan 6-8 en groter dan 30-50 eruit
* 3. Peptide search space
  + Meta-proteomics 🡪 vaak vlas + en/of –
  + PSM = peptide to spectrum match (= +/- peptide)
    - OPM: Bep peptiden die meermaals voorkomen, komen maar 1 maal in lijst voor
* 4. Fragment m/z calculation
  + Peptiden knippen in meerdere stukken + massa hier uit halen
    - Vaak vaste b & y waarden
* 5. Theoretical MS2 spectra matching to experimental MS2
  + Via scoring functions
    - Explained intensity
      * Theoretisch mappen aan experimenteel
      * Overal waar overlap is 🡪 waarde/intensiteit van exp spectrum nemen + optellen
    - Peak counting
      * Indien piek overlap 🡪 +1
    - Combinations
      * Aantal gevonden pieken & gevonden intensiteiten
  + Best scorende ftie gebruiken
  + MS2 lijsten genereren
    - .mgf (mascot generic format) files
      * Titel
      * # scans
      * (retentive tijd)
      * m/z
      * abundantie warden
      * …
    - Mzml file
      * “standard” format
  + PTMs 🡪 effect op zoekruimte?
    - Worden exponentieel groter 🡪 te groot voor search engines

**Peak list generation**

* Raw files conversion 🡪 mgf conversion

**Peptide to Spectrum Matching**

* Searchgui op compomics website downloaden
* Java installatie = nodig
* Zie handleiding 1.3 op proteomics voor manual
* FDR 🡪 aantal valse resultaten trachten terug te vinden
  + EW omdraaien & knippen
  + Decoys erbij 🡪 databank grootte verdubbelen
* Decoy & target hits plotten
  + Aantal vals + toelaten
  + Treshold toelaten
  + FDR v. 1% = vaak voorkomend
    - 1% beste v. decoys toelaten
    - 1% vals + matches 🡪 1% v. targets zullen niet juist gemapt worden
* Q-values = laagst mogelijk FDR die teruggevonden wordt in PSMs
* Probleem 🡪 terugmappen van peptiden naar EWn 🡪 meerdere EWn kunnen dezelfde peptiden bevatten
  + Occam’s razor theory 🡪 meest simpele uitleg die de meeste info weergeeft
* # peptiden < # spectra
* HCD 🡪 vooral b & y ionen