Sintetinės biologijos projektas

Acinetobacter baumannii bakterijų, paveiktų toksinu CheT, proteominių duomenų analizė

Beatričė Radavičiūtė

December 20, 2018

Contents

Įvadas

Mano nagrinėjami duomenys gauti toksinu CheT veiktų bakterijų *Acinetobacter baumannii* bei kontrolinių bakterijų (nepaveiktų toksinu) lizatus tyrus masių spektrometrijos metodu. Visi tyrimai atlikti Gyvybės mokslų centre. Darbe pateikiami jau apdotori duomenys, kuriuose nurodoma nustatytų bakterijose peptidų pavidinimai, jų raiškos pokytis, lyginant su kontrolinėmis bakterijomis bei kitos charakteristikos. Šio darbo tikslas yra atlikti intamųjų aprošomosios statiskos analizę. Taip pat darbo metu bus bandoma nustatyti, ar tarp kintamųjų (nustatytų baltmų charakteristikų) yra koreliacija, priežastinis ryšys ir kt.

Duomenų įsikėlimas, peržiūra, pradinė analizė

Duomenų peržiūra

Įsikeliami duomenys, nustatomas charakteristikų (variables) ir nustatytų baltymų (observations) skaičius, kintamųjų tipas.

```
## Observations: 1,344
## Variables: 21
## $ description         <chr> "Acyl-CoA dehydrogenase OS=Acinetobacter b...
## $ IEP                 <dbl> 5.54, 5.30, 4.69, 4.84, 4.53, 9.87, 5.24, ...
## $ mw                  <dbl> 65872.10, 76792.25, 37183.58, 22315.70, 36...
## $ `max score`         <dbl> 3658.1450, 779.0961, 255.2396, 1830.3930, ...
## $ accession           <chr> "A0A1G5LU08", "A0A241YAV2", "A0A1C9CQK6", ...
## $ `reported peptides` <dbl> 3, 16, 3, 7, 13, 7, 11, 6, 25, 9, 7, 11, 6...
## $ `sequence coverage` <dbl> 7.49, 38.96, 13.86, 55.72, 45.59, 55.35, 5...
## $ `FDR level`         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ entry               <chr> "A0A1G5LU08_ACIBA", "A0A241YAV2_ACIBA", "A...
## $ K1                  <dbl> 7336.7442, 4498.4030, 6098.0000, 4677.4877...
## $ K2                  <dbl> 8147.7455, 5594.2167, 6321.5000, 4949.2173...
## $ K3                  <dbl> 8346.3076, 12322.5575, 4043.0000, 4994.310...
## $ T1                  <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0....
## $ T2                  <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0....
```

```
## $ T3                  <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0....
## $ `AVERAGE K10_4-6`   <dbl> 7943.5991, 7471.7258, 5487.5000, 4873.6717...
## $ `AVERAGE T10_4-6`   <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0....
## $ logFC               <dbl> -12.953526, -12.724876, -12.394068, -12.25...
## $ t                   <dbl> -158.9174436, -34.6682401, -68.1900159, -1...
## $ P.Value             <dbl> 4.390731e-12, 3.946895e-08, 6.940879e-10, ...
## $ adj.P.Val           <dbl> 1.180228e-09, 1.360161e-06, 3.109514e-08, ...
```

Iš viršuje pateiktos lentelės matyti, jog nustatyta 1344 peptidai, įvertintos 21 charakteristikos, kurių dauguma - skaitinės. Svarbu paminėti, kad likusios neskaitinės (šiuo atveju kategorinės) baltymų charakteristikos yra jų pavadinai bei identifikacijos numeriai (description, accession ir entry) duomenų bazėse (šiuo atveju UniProt). Su jais tolesni veiksmai nebus daromi, taigi šie kintamieji bus pašalinti o baltymai bus išrikiuoti pagal jų pavadinimą abėcėlės tvarka ir užkoduoti skaičiais.

```
## Observations: 1,344
## Variables: 18
## $ IEP                 <dbl> 4.90, 5.44, 4.87, 4.49, 7.44, 5.80, 5.59, ...
## $ mw                  <dbl> 10129.64, 41645.43, 26205.95, 19170.71, 20...
## $ `max score`         <dbl> 35042.1500, 6539.5190, 7178.4810, 444.8390...
## $ `reported peptides` <dbl> 5, 15, 13, 3, 9, 8, 15, 14, 16, 18, 23, 18...
## $ `sequence coverage` <dbl> 71.88, 46.92, 70.37, 27.91, 78.07, 41.21, ...
## $ `FDR level`         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ K1                  <dbl> 135483.333, 4461.333, 37535.667, 4089.667,...
## $ K2                  <dbl> 149256.000, 4091.667, 43926.000, 3772.333,...
## $ K3                  <dbl> 167503.667, 10768.000, 45422.667, 4156.000...
## $ T1                  <dbl> 132090.333, 5321.667, 45934.000, 4588.333,...
## $ T2                  <dbl> 175413.333, 8561.333, 49163.000, 4381.000,...
## $ T3                  <dbl> 188138.667, 16920.333, 48734.000, 4185.667...
## $ `AVERAGE K10_4-6`   <dbl> 150747.667, 6440.333, 42294.778, 4006.000,...
## $ `AVERAGE T10_4-6`   <dbl> 165214.111, 10267.778, 47943.667, 4385.000...
## $ logFC               <dbl> 0.121326863, 0.657090910, 0.185100563, 0.1...
## $ t                   <dbl> 0.75310774, 1.21195200, 1.85983591, 1.5971...
## $ P.Value             <dbl> 0.4799392164, 0.2711783419, 0.1123523767, ...
## $ adj.P.Val           <dbl> 0.73216607, 0.58152015, 0.37845011, 0.4587...
```

Žemiau pateiktoje lentelėje išvardintos kitų kintamųjų pavadinimai ir jų reikšmės.

Table 1: Kintamųjų pavadinimai ir jų reikšmės.

Pavadinimai	Reikšmės
IEP	Izoelektrinis taškas
mw	Molekulinis svoris
max score	Didžiausia suminė jonų krūvio vertė atitinkamam peptidui
reported peptides	Nustatytų peptidų skaičius
sequence coverage	Sekos perdengimas
K1	1 kontrolinis mėginys
K2	2 kontrolinis mėginys
K3	3 kontrolinis mėginys
T1	1 tiriamasis mėginys
T2	2 tiriamasis mėginys
T3	3 tiriamasis mėginys
AVERAGE K10_4-6	Kontrolinių mėginių vidurkis
AVERAGE T10_4-6	Tiriamųjų mėginių vidurkis
logFC	Pokyčio logaritmas, kurio pagrindas 2
t	T testo tarp kontrolinių ir tiriamųjų mėginių vertė

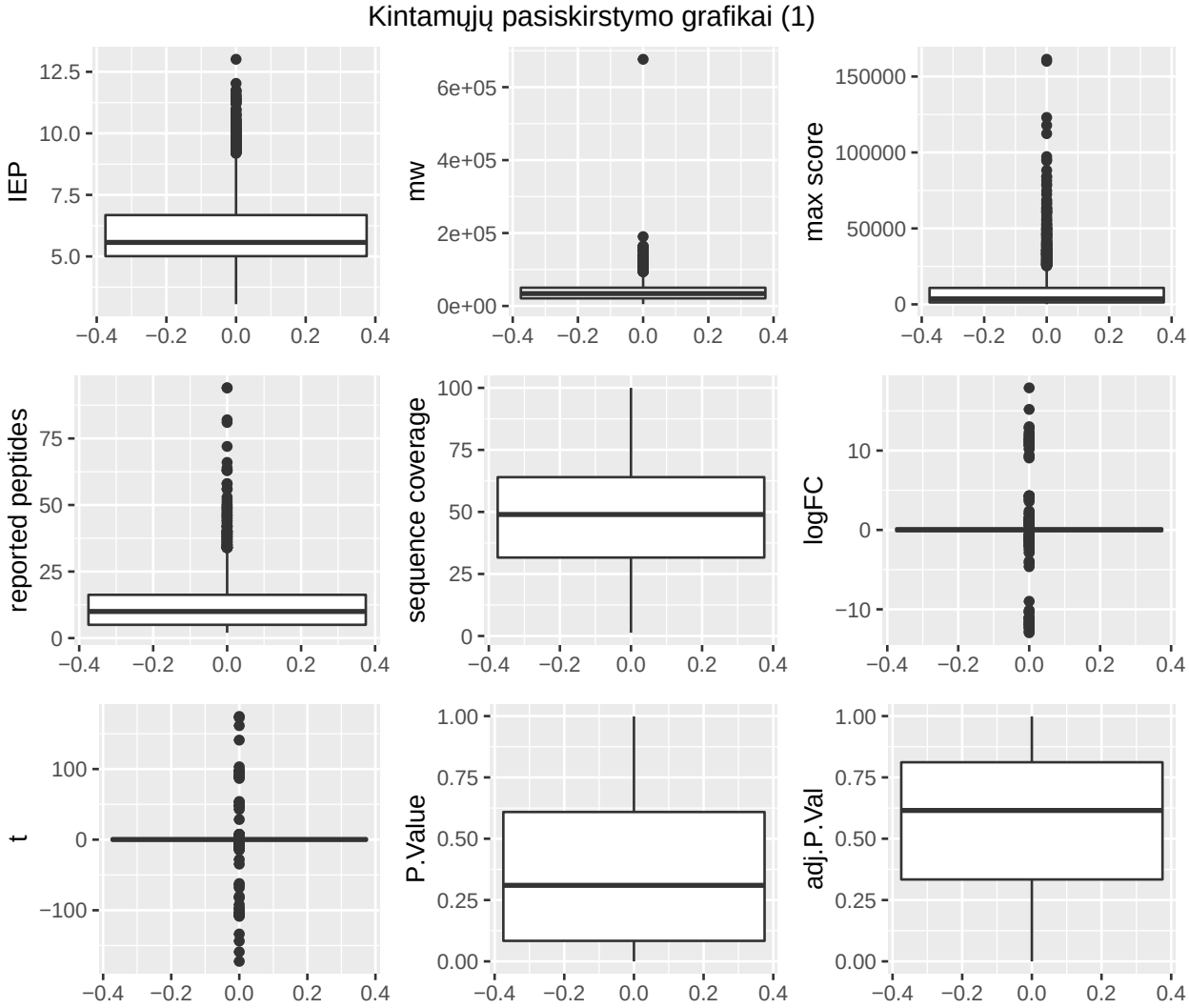
Pavadinimai	Reikšmės
P.Value	Pvertė, nusako statistinų duomenų patikimumą
adj.P.Val	Koreguota P vertė

Pradinė duomenų analizė

Pašalinus kategorinius kintamuosius, galima atlikti dar keletą veiksmų, kad su duomenimis būtų galima dirbti paprasčiau. Visų pirma, **FDR level** charakteristika nusako masių spektrometrijos metu nustatytų peptidų patikimumą. Kai FDR level = 0, peptidai nustatyti teisingai. Šiame darbe nagrinėjami visi peptidai, kurių FDR level = 0, taigi šią charakteristiką galima pašalinti, kadangi tai konstanta.

log FC charakteristika nusako nustatyrų peptidų raiškos pokytų tarp toksinu veiktų bakterijų ir kontrolinių bakterijų mėginių bei gali būti išreikšta formule 2^n, kur n = log FC vertė. Todėl pravartu šią vertę apsiskaičiuoti ir pridėti naują stulpelį į duomenų lentelę, pavadinimu “expression”.

Aprašomoji statistika



```
grid.arrange(k1, k2, k3, t1, t2, t3, avg_k10, avg_t10, top = "Kintamųjų pasiskirstymo grafikai (2)")
```