

Číslo súťažiaceho:

Biologická olympiáda

Ročník: 60.

Školský rok: 2025/2026

Kolo: Celoštátne

Kategória: B

Teoreticko-praktická časť

Praktická úloha č. 1 (90 minút, 45 bodov)

Téma: Rastlinná výpočtová biológia - Analýza odpovede rastlín na stres

Podstatná časť výskumu v rastlinnej biológii sa v súčasnosti zameriava na schopnosť rastlín vysporiadať sa s meniacimi sa klimatickými podmienkami a možnosťami, ako zvýšiť ich odolnosť voči stresu. Jedným z prístupov v tejto výskumnej oblasti je analýza génovej expresie rastlín v rôznych podmienkach. Na základe tejto analýzy možno identifikovať gény, ktorých aktivita sa pri strese mení a prípadne aj upraviť ich aktivitu tak, aby rastlina na stres dokázala lepšie reagovať. Nenechajte sa odradiť tým, že bude v úlohe pre vás mnoho vecí nových. Na jej riešenie nebudete potrebovať predošlú znalosť prezentovaných konceptov, len pozorné čítanie a logické uvažovanie.

Pomôcky: Počítač s priečinkom s číslom súťažiaceho na ploche (obsahuje súbory *normalization_data.xlsx*, *expression_data.xlsx* a súbor *excel_návod.pdf*), kalkulačka

Úvod do analýzy génovej expresie rastlín

Rastliny, tak ako ostatné známe organizmy, vo svojom genóme prostredníctvom DNA kódujú informáciu pre produkciu proteínov (bielkovín), ktoré v ich organizme vykonávajú rozličné funkcie. Napríklad gén *rbcL* kóduje informáciu pre výrobu proteínu RuBisCO, ktorý umožňuje fixáciu oxidu uhličitého v procese Calvinovho cyklu. Rastlinný genóm však obsahuje obrovské množstvo rôznych génov (napríklad kukurica má približne 50,000 génov). Keby boli všetky nimi kódované proteíny produkované neustále v rovnakej miere, rastlina by musela vynaložiť neudržateľne veľké množstvo zdrojov. V záujme rastlinného organizmu je však využívať zdroje čo najefektívnejšie, preto sa oplatí vyrábať jednotlivé proteíny len vtedy, keď sú potrebné. A to je zabezpečené reguláciou génovej expresie. Ide o veľmi zložitý súbor dejov, ktorý upravuje expresiu jednotlivých génov podľa potrieb organizmu.

Doteraz sme hovorili o produkcii proteínov na základe DNA. Proteíny sa však nevyrábajú priamo na základe predlohy DNA, ale v tomto procese existuje medzikrok v podobe RNA. V DNA je zakódovaná informácia pre výrobu proteínu, ak je "vydaný signál" na jeho produkciu, z DNA sa v jadre bunky vyrobí RNA v procese transkripcie, potom je RNA transportovaná do cytoplazmy a až na základe nej sa vyrába proteín (v procese translácie).

Číslo súťažiaceho:



Takéto usporiadanie prináša viacero výhod, v porovnaní s hypotetickým prípadom, v ktorom by boli proteíny vyrábané priamo na základe predlohy DNA. Označte tieto výhody v úlohe nižšie.

Úloha 1:

- A) Proces produkcie proteínu je vďaka prítomnosti RNA medzikroku rýchlejší.
- B) Produkciu proteínu možno vďaka prítomnosti RNA presnejšie regulovať.
- C) Ak na základe jednej molekuly DNA možno vyrobiť viacero molekúl RNA a na základe jednej molekuly RNA viacero molekúl proteínu, tak toto usporiadanie umožňuje zosilnenie signálu v procese exprese.
- D) Ak existuje proteín, ktorý je vyrábaný neustále bez ohľadu na podmienky v rovnakej miere, usporiadanie s RNA medzikrokom minie na jeho produkciu menej zdrojov ako usporiadanie bez neho.

Na základe predchádzajúceho textu by sme mohli očakávať, že ak vystavíme rastlinu stresovým podmienkam a následne budeme analyzovať proteíny, ktoré produkuje (expresný profil), nájdeme väčšie zastúpenie proteínov, ktoré rastline pomáhajú vysporiadať sa s týmto stresom, v porovnaní s expresným profilom kontrolnej rastliny (takej, ktorá stresu vystavená nebola). Skúste sa zamyslieť a označiť správne tvrdenia.

Úloha 2:

- A) Ak je rastlina vystavená soľnému stresu, bude produkovať viac proteínov akumulujúcich ióny v jej koreňoch.
- B) Stres vysokou teplotou a vysokým ožiarením bude pravdepodobne spúšťať odlišnú odpoveď.
- C) Vo všetkých prípadoch vystavenia rastliny stresu možno očakávať plošné zvýšenie produkcie proteínov všetkého druhu.
- D) Vystavenie rastliny stresu pravdepodobne zapríčiní zvýšenie produkcie proteínov, ktoré sú schopné stabilizovať iné proteíny.

Práve analýza expresného profilu rastliny vystavenej stresu a jeho porovnanie s expresným profilom kontrolnej rastliny (tzv. metóda diferenciálnej génovej exprese) sa používa na odhalenie génov asociovaných s odpoveďou na stres. Z praktických dôvodov sa ale vo väčšine prípadov nepoužíva analýza rastlinou produkovaných proteínov (proteómu), ale produkovaných RNA (transkriptómu). Hladina mRNA a zodpovedajúceho proteínu sú skoro vždy priamo úmerné. Ak teda dokážeme analyzovať zastúpenie jednotlivých RNA v bunke a porovnať ho medzi rastlinami vystaveným rozličným podmienkam, vieme identifikovať tie

Číslo súťažiaceho:

RNA (transkripty), ktorých produkcia je odlišná medzi jednotlivými rastlinami. Gény, ktoré tieto RNA kódujú, budú teda pravdepodobne spojené s odpoveďou rastliny na daný stres.

Normalizácia dát

Na ploche vášho počítača vidíte priečinok s vaším číslom. Keď ho otvoríte, vidíte súbor *normalization_data.xlsx*. Po otvorení tohto súboru budete vidieť hárok *data* obsahujúci tabuľku s hypotetickými génmi (*gene A - O*), hladinou ich transkriptu (množstva RNA) pre kontrolné rastliny (*control_raw*) a hladinou transkriptov hypotetických génov rastlín vystavených stresu suchom (*drought_raw*). V praxi vedie k získaniu týchto čísel dlhá cesta, my však máme už pripravené číslo vyjadrujúce množstvo transkriptov prítomných vo vzorkách.

Na prvý pohľad sa zdá, že je vo vzorkách z rastlín vystavených suchu vyššia hladina jednotlivých transkriptov. To však môže byť spôsobené množstvom vstupného materiálu, alebo inými technickými odchýlkami. Preto potrebujeme dáta previesť do formy, v ktorej ich vieme vzájomne porovnať medzi jednotlivými súbormi, tzv. normalizovať. Existuje množstvo rôznych spôsobov normalizácie transkriptomických dát, vy si vyskúšate tie najjednoduchšie.

Úloha 3

Začneme normalizáciou podľa priemeru. Tú vypočítame podľa vzorca *norm = konkrétna hodnota/priemer všetkých hodnôt v skupine*. Napríklad pre gén A bude normalizovaná hodnota množstva transkriptu v kontrolnej rastline vypočítaná takto:

$$norm(\text{gene A}; C) = B2 / average(B2:B16) = 115 / 157,7 = 0,73$$

Vypočítajte v stĺpci *Mean C* **normalizované hodnoty** podľa priemeru (podľa vzorca uvedeného vyššie) pre ostatné gény pre kontrolnú rastlinu a v stĺpci *Mean D* normalizované hodnoty pre rastlinu vystavenú stresu suchom. Môžete používať kalkulačku, ale odporúčame používať samotný Excel. Návod na používanie tohto softvéru máte v súbore vo vašom priečinku *excel_návod*, prípadne môžete požiadať o radu dozor. Odteraz vždy súbor po výpočte preventívne uložte, môžete si to zjednodušiť klávesovou skratkou *Ctrl + S*.

Úloha 4

Okrem normalizácie podľa priemeru môžeme dáta normalizovať podľa mediánu (*norm = konkrétna hodnota/medián všetkých hodnôt v skupine*; Pozn.: Medián je stredná hodnota dát súboru zoradených podľa veľkosti. Ak je súbor dát 1, 2, 9; tak medián bude 2) alebo sumy súboru dát (*norm = hodnota/súčet hodnôt*).

V stĺpcoch *Median C*, *Median D*, *Total C* a *Total D* spočítajte **normalizované hodnoty podľa mediánu (Median) alebo súčtu (Total)** pre súbory dát z kontrolnej rastliny (C) aj rastliny vystavenej stresu suchom (D). Funkcie Excelu, ktoré môžete použiť, nájdete opäť v súbore *excel_návod*.

Po úlohách 3 a 4 by ste mali dostať tabuľku s rôzne normalizovanými hodnotami. Teraz už máte k dispozícii hodnoty, na základe ktorých viete porovnať mieru expresie génov medzi kontrolnými rastlinami a rastlinami vystavenými stresu suchom. Aby ste toto porovnanie

Číslo súťažiaceho:

vedeli urobiť jednoduchšie, je vhodné vypočítať **relatívnu zmenu expresie génov** rastlín vystavených stresu voči kontrolným rastlinám. Tú viete vypočítať ako **podiel** normalizovanej hodnoty expresie génu rastliny vystavenej stresu a normalizovanej hodnoty pre kontrolnú rastlinu.

Ak by bola napríklad normalizovaná hodnota podľa priemeru pre gén A pre kontrolnú rastlinu $norm(gene\ A; C) = 2$ a normalizovaná hodnota podľa priemeru pre gén A pre stresovanú rastlinu $norm(gene\ A; D) = 3$, tak relatívnu zmenu expresie vypočítame ako $norm(gene\ A; D) / norm(gene\ A; C) = 3 / 2 = 1,5$.

Túto hodnotu možno interpretovať tak, že miera expresie génu A v rastlinách vystavených stresu suchom je 1,5-krát vyššia ako v kontrolných rastlinách - hovoríme, že gén je upregulovaný. Tento gén bude teda pravdepodobne asociovaný s odpoveďou rastliny na stres.

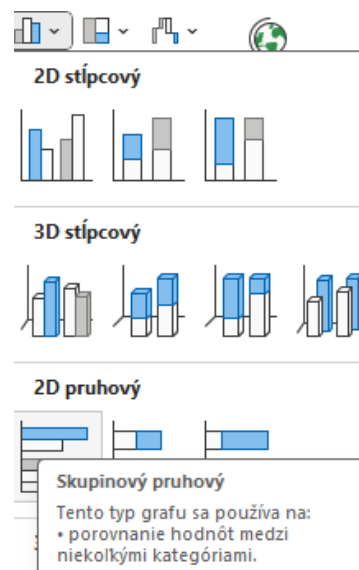
Úloha 5

Do stĺpcov *Fold Change (mean - total)* vypočítajte hodnoty pre **relatívne zmeny v génovej expresii** pre normalizáciu podľa priemeru, mediánu a súčtu.

Úloha 6

Vytvorte graf, v ktorom zobrazíte relatívne zmeny pre normalizáciu podľa súčtu.

1. Držte klávesu *ctrl* a označte stĺpec *Gene* (s hlavičkou).
2. Označte stĺpec *Fold change total* počas toho ako stále držíte klávesu *ctrl*.
3. Klávesu môžete pustiť, v Exceli choďte do *Vložiť* a v grafoch vyberte 2D skupinový pruhový graf (viď obrázok). Obrázok nechajte pod tabuľkou v súbore.



Úloha 7

Na základe grafu a predchádzajúceho textu rozhodnite, ktoré tri gény sú zjavne upregulované, a ktoré tri gény sú zjavne downregulované. Zapište do políček nižšie:

Upregulované gény:

Downregulované gény:

Číslo súťažiaceho:

Mierne odlišný prístup k normalizácii využíva normalizácia pomocou referenčného génu. V tejto metóde sa použije pre kontrolnú aj sledovanú rastlinu miera expresie tzv. housekeeping génu ako normalizačný faktor. Housekeeping gény sú väčšinou stabilne exprimované nezávisle na podmienkach a ich relatívna hladina by sa naprieč rôznymi vzorkami nemala meniť. Môžu to byť napríklad gény pre aktín, gény pre DNA polymerázy, či práve RuBisCO.

Úloha 8

Použite mieru expresie génu B (stĺpce *Control_raw*, *Drought_raw*) na normalizáciu miery expresie ostatných génov. Postup je podobný ako pri predchádzajúcich normalizáciách.

1. Dopočítajte **normalizované hodnoty** pre všetky gény v stĺpci *Reference C* pre kontrolné rastliny.
2. Vypočítajte normalizované hodnoty v stĺpci *Reference D* pre rastliny vystavené suchu.
3. Vypočítajte **relatívnu zmenu expresie** podľa týchto normalizovaných hodnôt v stĺpci *Fold change reference* (Pozn.: Dajte si pozor, aby ste ako normalizačný faktor v prípade stresu vystavených rastlín použili hodnotu expresie housekeeping génu v rastline vystavenej stresu. Ako dobrá kontrola, či ste výpočet urobili správne bude hodnota relatívnej zmeny expresie (*Fold change reference*) housekeeping génu rovná 1).

Jednotlivé prístupy k normalizácii majú svoje výhody v rôznych prípadoch. Zamyslite sa a vyberte správne odpovede v úlohe nižšie.

Úloha 9

- A) V prípade, že referenčný gén reaguje pri odpovedi na stres zmenou svojej expresie, normalizácia podľa referenčného génu je najvhodnejšia metóda úpravy dát.
- B) Ak sú v súbore dát prítomné veľmi odľahlé hodnoty (napr. pri umelo zvýšenej génu), normalizácia pomocou priemeru môže výrazne podhodnocovať mieru expresie ostatných génov.
- C) Výpočet relatívnej zmeny expresie (*Fold change*) voči kontrolnej rastline na základe dát normalizovaných podľa priemeru a súčtu musí v prípade porovnávania miery expresie rovnakého počtu génov vyústiť v rovnaké výsledky.
- D) Normalizácia podľa súčtu je zo všetkých metód najmenej ovplyvnená prítomnosťou niekoľkých veľmi silne downregulovaných génov.

Identifikácia génov spojených s odpoveďou na stres

Už ste sa naučili základné metódy normalizácie transkriptomických dát. Teraz môžete prísť k analýze skutočného súboru dát. Otvorte súbor *expression_data.xlsx* a prezrite si hárok označený *data*. Tentoraz máte pred sebou vybrané gény z *A. thaliana* (označené identifikátorom v databáze, stĺpec *Entrez ID*) a mieru ich expresie v kontrolných rastlinách (stĺpec *control*), v rastlinách vystavených stresu teplom (stĺpec *heat_stress*) a stresu chladom (stĺpec *cold_stress*). Tieto dáta už boli normalizované, preto ich možno analyzovať priamo.

Číslo súťažiaceho:

Úloha 10

Vypočítajte relatívnu zmenu v expresii génov pre rastliny vystavené stresu teplom (*heat_fold_change*), resp. chladom (*cold_fold_change*).

Ak sa pokúsite z týchto dát urobiť graf, zistíte, že je to pomerne náročné, keďže sa v tomto prípade dáta navzájom rádovo líšia. Na ich jednoduchšiu vizualizáciu ich bude potrebné transformovať pomocou funkcie logaritmus. Tá nám umožňuje vizualizovať aj dáta, ktoré sa líšia rádovo a v podstate ich previesť do lineárnej podoby.

Úloha 11

V stĺpcoch *log_heat* a *log_cold* spočítajte transformáciu dát zo stĺpcov *Fold change* pomocou funkcie logaritmus so základom 2. Vzorec pre výpočet tejto funkcie je $\log(\text{číslo};2)$.

Teraz to už s vizualizáciou dát pôjde lepšie.

Úloha 12

Vytvorte graf rovnakým postupom ako v úlohe 6. Opäť ho umiestnite pod tabuľku. Ak vám robí problém označenie osí, skúste zmeniť nastavenia v možnostiach grafu alebo privolajte dozor.

Teraz viete identifikovať, ktoré gény sú upregulované a ktoré downregulované. Takisto by ste mali vedieť povedať, s akým stresom sú asociované. Overiť to viete pomocou databázy [Entrez](#).

1. Zo súboru *expression_data.xlsx* stačí skopírovať Entrez ID (napr. pomocou klávesovej skratky *Ctrl + C*).
2. Prilepte ho do vyhľadávacieho okna na stránke Entrez (skratka *Ctrl + V*). Vedľa vyhľadávacieho okna je potrebné zvoliť, že hľadáte gén namiesto prehľadávania všetkých databáz (viď obrázok).



Úloha 13

Nájdite dva gény, ktoré sú upregulované špecificky pri strese chladom v databáze Entrez. Do políček nižšie napíšte ich celé názvy (*Official full name*).

Číslo súťažiaceho:

Úloha 14

Do políčka nižšie napíšte Entrez ID génu, ktorý pravdepodobne reaguje na rôzne druhy stresu nešpecificky.

Úloha 15

Skúste identifikovať dva gény, ktoré by mohli byť housekeeping génmi. Overte si ich funkciu v databáze Entrez. Riad'te sa tým, že by to mali byť gény, ktoré sú potrebné v bunke neustále, nezávisle na podmienkach. Do políčok nižšie napíšte ich oficiálny symbol (*Official Symbol*).

Koniec praktickej úlohy :)