

Průchod aplikací krok za krokem - záložky 1–6

1) Upload

- Vybrat XLSX soubor
- Zadat Output folder (lokální cesta - např. F:/UOCHB/Proteomics_data/MS25289184)
- Zvolit Analysis mode: t-test (precomputed sloupce v XLSX) a/nebo limma+DEqMS. Limma as DEqMS je typ analýzy, který používáme od začátku
- Zvolit Master sheet (typicky první sheet nebo *_ttest).
- Kliknout na Load data. Tím se nic nezapisuje, pouze se načtou data a ověří se struktura dat v xlsx souboru.

2) Loaded data

- V tabulce Detected sample columns and grouping je dobré zkontrolovat, jestli sedí názvy vzorků a skupin. Hmotníkáři ten format rádi mění a tak je dobré na to dívat pozor, také jsou tam občas chyby v názvech skupin.
- V Detected comparisons (pokud je používán t-test režim), je vidět pro jaké skupiny byl proveden t-test.
- Pokud je vše v pořádku, pokračuje se do Preprocess.

3) Preprocess

V této záložce je připravována matice pro QC a případně pro výpočty:

Doskrolovat až na místo s Global Imputation, v případě, že jsou v souboru chybějící hodnoty. To je možné zkontrolovat tak, že se nejdřív udělá QC:

- Global imputation: volitelně se zvolí MNAR imputation (MinProb / QRILC) nebo None.
- Run QC: vygenerují se QC tabulky a grafy (PCA, missing values, boxplot intenzit, korelační heatmapy).

Advanced, nepoužívat:

- Protein filtering: nastaví se práh pro vyřazení proteinů s příliš velkým podílem chybějících hodnot (Preview proteins to drop).
- Targeted floor imputation: volitelně se nastaví seznam genů nebo regex a vyberou se vzorky; Tohle se může hodit pro virové proteiny
- NA/0 se nahradí minimální baseline.

4) Comparisons + Enrichment

- Vybrat porovnání (A vs B). Pokud je zapnuté Mirror, automaticky se vytvoří i B vs A (logFC krát -1).
- Vybrat režim výpočtu (Use XLSX precomputed t-test / Compute limma+DEqMS).
- Nastaví se BH/FDR (doporučeno, pokud by někdo chtěl použít t-testy).
- Vyberou se enrichment typy (GO_BP/GO_CC/GO_MF/KEGG/Reactome). “Select all” slouží k hromadnému zaškrtnutí/odškrtnutí.
- **Doporučuji** zapnout Parallelize enrichment a nastavit Workers. Může se stát, když se nastaví víc threadů, že dojde ke zpomalení PC, pak nastavit méně “workers”. Počítání enrichmentu je výpočetně náročné, proto je dobré paralelizaci používat, 2 thready /workers znamenají dvojnásobné zrychlení.. atd.

5) Run analysis

- Kliknout na Run (export + enrichment + QC outputs).
- V Run logu je možné sledovat průběh: načtení, QC export, porovnání, export CSV, enrichment, navíc to je vidět i v cmd okně (okno s černým pozadím)-
- V Created files je seznam nově vytvořených souborů v této relaci.

6) Volcano Plot

Volcano plotter je určen pro rychlou vizualizaci a export grafu:

- Data source: buď z nahraného hlavního XLSX (pair sheets, ale pozor v xlsx jsou t-testy), nebo separátní XLSX, nebo CSV – klasické použití. Po vygenerování výsledků v bode 5 je možné načíst příslušná CSV.

- Select pair + Plot selected pair: načte data a vykreslí graf.
- Apply BH/FDR correction: přepíná korekci p-hodnot, opět pro t-testy
- Flip logFC sign: prohodí levou/pravou stranu (změní znaménko logFC).
- Brush selection: myš vybere region v grafu; výběr se objeví v tabulce Selected points.
- Kliknutí na bod se objeví jméno Genu příslušející k bodu, přepíná popisek v Annotations.
- Lze také použít Mark gene pro vyhledání bodu v grafu.
- Export: CSV a SVG.

Výstupy a adresářová struktura

Do výstupní složky (zadáva se v záložce 1) jsou zapisovány:

- results/res_...csv: výsledky porovnání
- results/QC_clustering/: QC tabulky a grafy (PCA, heatmaps, missingness, report multi-gene řádků). To jsou výsledky především pro kontrolu kvality dat.
- results/enrichment_analysis//: CSV tabulky enrichmentu + SVG dotploty pro každé porovnání.
- results/run_log.txt: textový log běhu (užitečné pro debug)