

Porovnání proteotypů nádorových buněčných linií pomocí kvantitativní hmotnostní spektrometrie

Teoretický úvod

Proteomika představuje soubor experimentálně náročných metod využívaných ke studiu mnoha (stovek až tisíců) proteinů současně. Proteomika přispívá mimo jiné k charakterizaci signálních drah zapojených v různých biologických procesech, ke studiu biomarkerů a identifikaci potenciálních terapeutických cílů. Klíčovou metodou proteomiky je **kapalinová chromatografie on-line spojená s tandemovou hmotnostní spektrometrií (LC-MS/MS)**. V rámci tohoto cvičení využijeme proteomický přístup na bázi LC-MS/MS k **porovnání aktuálních proteinových profilů (proteotypů) tří buněčných linií** odvozených od různých typů karcinomů: luminálního nádoru prsu, trojitě negativního nádoru prsu a skvamocelulárního karcinomu.

Nádory prsu patří k nejčastějším nádorovým onemocněním u žen, z hlediska molekulární podstaty jsou přitom velmi heterogenní. Na základě genové exprese je lze rozdělit na několik subtypů, mezi které patří **luminální A**, které jsou hormonálně závislé, a proto léčitelné antihormonální léčbou, a **trojitě negativní**, které jsou velmi agresivní, nevykazují expresi hormonálních receptorů a nejsou proto zatím běžně biologicky léčitelné. **Skvamocelulární karcinom** je pak závažným typem nádorového onemocnění dlaždicobuněčného epitelu (např. horní vrstva kůže, výstelka plic, epikardium) - jedná se o maligní nádor vytvořený z dlaždicovitých buněk, který má tendenci metastázovat.

Při studiu nádorů lze kromě analýzy nádorových tkání využít **buněčné modely**, především **nádorové buněčné linie**. Ty byly derivovány z nádorů konkrétních pacientů a od té doby jsou udržovány v bankách biologického materiálu ve zmraženém stavu. Odtam jsou komerčně dostupné, mohou být kultivovány v laboratorním prostředí a využity například k **funkčním experimentům** zahrnujícím studium úlohy konkrétních proteinů v nádorových procesech. Nevýhodou nádorových linií kultivovaných *in vitro* je **absence nádorového mikroprostředí**, zejména nepřítomnost dalších typů buněk (např. fibroblastů), které mohou růst nádorových buněk v živém systému ovlivňovat.

V úloze se pracuje se třemi nádorovými buněčnými liniemi: s **buněčnou linií MCF-7** odvozenou od **luminálního A nádoru prsu**, s buněčnou linií **MDA-MB-231** odvozenou od **trojitě negativního nádoru prsu** a s buněčnou linií **A431** odvozenou od **skvamocelulárního karcinomu**.

Buněčná linie **MCF-7** je odvozena od luminálního, estrogen receptor (ER) pozitivního nádoru prsu. S hladinou ER silně koreluje hladina anterior-gradient proteinu 2 (**AGR2**), který bychom měli v proteomovém profilu kvantifikovat. Linie MCF-7 se vyznačuje **adhezivními vlastnostmi**, s nimiž koreluje hladina adhezivního markeru Epithelial Cell Adhesion Molecule (**EPCAM**). Všechny tři linie mají epiteliální charakter, což lze potvrdit přítomností **epiteliálních markerů** jako je **E-cadherin (CDH1)** a **cytokeratin 18 (KRT18)**. Linie **MDA-MB-231 a A431**, které mají vyšší schopnost metastázovat, však mají v důsledku **epiteliálně-mezenchymální tranzice** fenotyp částečně posunutý směrem k mezenchymálnímu, což indikuje např. zvýšená hladina **vimentinu (VIM)** a proteinu **CD44**. Linie **A431** se dále vyznačuje zvýšenou expresí **receptoru pro epidermální růstový faktor (EGFR)**, na němž je růst buněk A431 závislý a který aktivuje **STAT** dráhu podporující buněčnou proliferaci. Zatímco linie MCF-7 je z hlediska statusu genu **TP53** wild type, linie MDA-MB-231 a A431 mají gen **TP53** mutovaný, což se obvykle projevuje zvýšenou hladinou kódovaného proteinu **p53**. Tyto a další fenotypové rozdíly bychom měli pozorovat v proteinových profilech všech tří linií.

Princip experimentu

Vstupním biologickým materiálem jsou buněčné kultury MCF-7, MDA-MB-231 a A431. Buněčné linie jsou kultivovány za přítomnosti kompletního růstového média Dulbecco's Modified Eagle's Medium (DMEM) obsahujícího 10 % fetálního hovězího séra (FBS), které obsahuje hormony a růstové faktory. Buněčné linie jsou kultivovány v inkubátoru při 37 °C v 5 % atmosféře CO₂ na Petriho miskách. Buňkám odsajeme médium, promyjeme je pomocí „phosphate buffer saline“ (PBS), sklídíme je do denaturujícího lyzačního pufru a provedeme jejich desintegraci ultrazvukem.

Ultrazvuková desintegrace pracuje na principu tzv. mikrokavitace: Ultrazvuk (jako forma mechanického vlnění) vyvolává vznik dutin v kapalině (=„bublin vakua“), což uvnitř buněk vede k rychlým a opakovaným objemovým změnám buněk a jejich desintegraci (popraskání buněčných membrán). Pro srovnání: kavitace je jev, který nastává např. při mechanickém otáčení lodního šroubu, kdy se rovněž tvoří „bubliny vakua“.

Lyzační pufr obsahuje:

denaturační činidlo – 6 M guanidin hydrochlorid – zajišťuje denaturaci proteinů

detergent – 1 % Triton X-100 – pomáhá rozrušit buněčné membrány

pufr – 100 mM Na-fosfát pH 6,6 – udržuje konstantní pH a definovaný náboj disociabilních skupin proteinů

Centrifugace desintegrované buněčné směsi umožní oddělit zbylé části buněk - membrány (pelet) a solubilizovaný proteinový extrakt (supernatant).

Stanovení celkového proteinu provedeme pomocí kitu RC-DC Protein Assay (Bio-Rad). Postup využívá precipitace proteinů pomocí RC Reagentů I a II k odstranění látek interferujících s vlastním stanovením (např. detergenty). Po rozpuštění precipitátu následuje stanovení modifikovanou Lowryho metodou, při níž Cu²⁺ ionty vzniklé oxidací peptidové vazby redukují v alkalickém prostředí Folin-Ciocalteuovo činidlo za vzniku barvení měřitelného při 750 nm, které je přímo úměrné koncentraci proteinu.

Štěpení proteinů trypsinem na peptidy (digeste) – příprava vzorku a štěpení proteinů na filtru („Filter Aided Sample Preparation“ (FASP)) proběhne na kolonkách Microcon s 30 kDa filtrem. Filtr na kolonkách nám pomocí centrifugace umožní proteiny zakoncentrovat, menší molekuly (včetně potenciálních kontaminantů) navíc kolonkou protečou. Dále na kolonkách provedeme redukci disulfidických můstků cysteinů v prostředí pufru obsahujícího vysokou koncentraci močoviny pomocí tris(2-carboxyethyl)fosfinu (TCEP) a následně alkylaci zredukovaných -SH skupin cysteinů alkylačním činidlem iodacetamidem (IAA). Ke konci FASP provedeme enzymatické štěpení proteinů na peptidy pomocí trypsinu, který štěpí peptidické vazby proteinů na karboxylových koncích aminokyselin lysinu a argininu, pokud za nimi nenásleduje prolin. Takto získané peptidy eluujeme z filtru pomocí centrifugace.

Odsolení peptidů provedeme na kolonkách na principu reverzně fázové chromatografie. Vzorky jsme zpracovávali v pufrch obsahující soli, které mohou potlačit ionizaci peptidů při analýze pomocí hmotnostní spektrometrie, a proto je musíme z peptidových vzorků důkladně odstranit. Sorbent na kolonkách (stacionární fáze) obsahuje imobilizované nepolární oktadecylové (C18) řetězce, s nimiž interagují nepolární skupiny peptidů ze vzorku. Kolonku poté promýváme vodou (polární mobilní fáze), čímž se uvolní ve vodě rozpustné kontaminanty. Přečištěné peptidy poté postupně eluujeme gradientem acetonitrilu (nepolární mobilní fáze) a lyofilizujeme pomocí vakuové centrifugy.

LC-MS/MS – peptidové vzorky budou následně analyzovány pomocí kapalinové chromatografie ve spojení s tandemovou hmotnostní spektrometrií (LC-MS/MS). To bude provedeno na pracovišti „Core

Facility Proteomics“ CEITEC MU. V první řadě je při LC-MS/MS analýze provedena separace peptidů pomocí reverzně fázové kapalinové chromatografie (C18) v kyselém prostředí na analytické koloně s přímým napojením do nanoelektrospreje, kde jsou vlivem vysokého napětí peptidy ionizovány, čímž získávají kladný náboj. Ionizované peptidy poté vstupují do hmotnostního spektrometru.

Ve hmotnostním spektrometru pracujícím v režimu datově závislého sběru dat („data dependent acquisition (DDA)) je u peptidů vystupujících aktuálně z chromatografické kolony nejprve změřeno MS1 spektrum, což je závislost intenzity signálu na poměru hmotnost/náboj (m/z). Prekurzorové ionty poskytující v MS1 spektru nejintenzivnější signály jsou poté postupně fragmentovány v kolizní cele, vzniklé produktové ionty jsou změřeny v MS2 spektrech. Většina moderních hmotnostních spektrometrů dokáže opakovaně provádět DDA cykly s jedním MS1 skenem a alespoň dvaceti MS2 závislými skeny v čase okolo tří sekund, což typicky umožňuje v jedné analýze identifikovat a kvantifikovat tisíce proteinů. Nevýhodou tohoto přístupu je omezená reprodukovatelnost daná do jisté míry náhodnou selekcí prekurzorových iontů pro fragmentaci. Toto do značné míry řeší režim tzv. datově nezávislého sběru dat (**“data independent acquisition” (DIA)**).

V rámci tohoto cvičení provedeme proteomovou analýzu pomocí hmotnostního spektrometru s analyzátozem Orbitrap v **DDA** módu s využitím “high collision dissociation” (HCD) cely, kde se prekurzorové ionty fragmentují, a analyzátoru typu Orbitrap, kde se detekují jak ionty peptidů při měření MS1 spekter, tak jejich produktové ionty při měření MS2 spekter.

Identifikaci a kvantifikaci peptidů a proteinů ze surových LC-MS/MS dat provedeme ve volně dostupném akademickém software **MaxQuant**. Každý datový soubor obsahuje všechna MS1 a MS2 spektra získaná analýzou daného vzorku. Při **databázovém prohledávání (např. proti databázi UniProt)** se experimentální informace o m/z peptidového prekurzoru z MS1 spektra společně s informacemi z MS2 spektra porovnávají s teoretickými informacemi z *in silico* naštěpeného proteomu daného organismu. Na základě shody s databází (vyjádřené jako Posterior Error Probability“ (PEP)) odpovídající pravděpodobnosti falešně pozitivní identifikace) je identifikován peptid odpovídající danému MS2 spektru (hovoříme o “Peptide Spectrum Match“ (PSM)). Díky současnému prohledání proti falešné (**“decoy“=návnada**) **databázi** obsahující reverzní sekvence téhož proteomu se stanoví míra falešně pozitivních identifikací v souboru: Identifikované peptidy jsou seřazeny dle PEP, systém následně nastaví hranici PEP tak, aby soubor obsahoval právě 1% falešně pozitivních identifikací. PSMs s horším PEP jsou ze souboru vyřazeny, “False Discovery Rate“ (FDR) je tedy 1%.

Peptidy mohou být **proteotypické** (unikátní pro určitý protein) nebo **neproteotypické** (sdílené více proteiny – tzv. “**razor**“ peptidy). Pokud jsou všechny identifikované peptidy daného proteinu součástí více proteinů, jsou tyto proteiny sloučeny do tzv. **proteinové skupiny**. Na základě PEP peptidů v rámci proteinové skupiny je vypočítáno FDR na úrovni proteinových skupin. Následně je provedena **kvantifikace**. Jelikož jsme nepoužili žádné chemické nebo metabolické značení proteinů či peptidů, použijeme ke kvantifikaci peptidové signály z MS1 spekter (MS1 “label-free“ kvantifikace (LFQ)). Z nich se zrekonstruují chromatografické eluční profily peptidů (píky), z nichž je vypočtena plocha píky, která se použije ke kvantifikaci peptidu. Pro kvantifikaci proteinových skupin jsou sečteny intenzity peptidů přiřazených k proteinové skupině.

Statistické vyhodnocení rozdílů mezi proteomovými profily všech tří nádorových buněčných linií provedeme v software **Perseus**. V rámci kvantitativních dat na úrovni proteinových skupin nejprve provedeme analýzu rozptylu (**ANOVA**), pomocí které určíme proteiny, jejichž hladiny jsou statisticky významně změněny v rámci souboru všech tří buněčných linií a jejich replikátů. Následně provedeme sérii tří **Studentových t-testů**, jimiž určíme proteiny, jejichž hladiny jsou statisticky významně změněny mezi jednotlivými dvojicemi buněčných linií (MCF-7 vs. MDA-MB-231, A431 vs. MCF-7, MDA-MB-231

vs. A431). Statistickou významnost určíme na základě **p-hodnoty korigované na mnohonásobné testování (q-hodnota <0,05)**, míru změny hladin proteinů mezi jednotlivými vzorky (“Fold Change“ (FC)) určíme pomocí rozdílu LFQ intenzit proteinových skupin v log₂ tvaru (log₂FC). S využitím **shlukovací analýzy sestrojíme heatmapu** pro vizualizaci podobnosti proteinových profilů buněčných linií a rozložení hladin jednotlivých proteinů mezi vzorky. Z heatmapy uvidíme, do jaké míry jsou proteinové profily biologických replikátů téže buněčné linie podobné a jak se shlukují proteiny do klastrů na základě jejich korelace.

Pomocí metody **Gene Set Enrichment Analysis (GSEA)** provedeme analýzu zapojení proteinů do regulačních drah. Vstupem do této analýzy bude seznam proteinů a hodnot log₂FC, který v rámci porovnání profilů dvou buněčných linií importujeme ze software Perseus. Protože je GSEA primárně nástrojem pro zpracování genomických dat, budeme v GSEA pracovat s názvy genů kódujících námi identifikované proteiny. GSEA mezi nimi vyhledá geny, které jsou součástí definovaných molekulárních drah databáze HALLMARK. Míra obohacení je dána tzv. Enrichment score (ES), statistická významnost p-hodnotou. Z výstupu uvidíme, které dráhy jsou obohaceny pozitivně (ES>0), nebo naopak negativně (ES<0) o námi kvantifikované proteiny. Tento nástroj nám umožní interpretovat rozdíly v proteinových profilech ve vztahu k fenotypu nádorových buněčných linií.

POSTUP PRÁCE

Abychom zamezili kontaminaci vzorků keratiny, po celou dobu laboratorní práce budeme mít oblečený čistý laboratorní plášť a rukavice, rukavicemi se nebudeme dotýkat částí těla (vlasů, tváře), klik, madel a dalších zdrojů keratinové kontaminace.

Sklízení buněčných linií (1. cvičení 19.10.2021)

Zaškrtněte vzorek, se kterým pracuje vaše skupina:

Buněčná linie MCF-7

Buněčná linie MDA-MB-231

Buněčná linie A431

1. Připravíme 1,5 ml mikrozkuhavku popsanou B+číslem skupiny (např. B1). Z misky odsajeme růstové médium, napipetujeme 5 ml 1xPBS (0,137 M NaCl, 2,68 mM KCl, 1,47 mM KH₂PO₄ a 6,45 mM Na₂HPO₄) a odsajeme. Přidáme dalších 5 ml 1xPBS a opět odsajeme.
2. Misku umístíme na led a přidáme 200 µl lyzačního pufru o složení 6 M guanidin hydrochlorid, 100 mM Na-fosfát pH 6,6 a 1% Triton X-100.
3. Misku nakloníme na stranu a **seškrábeme buňky** stěrkou z celého povrchu směrem dolů. Misku pootočíme o 90° a opět seškrábeme směrem dolů. Seškrábané buňky v lyzačním pufru přepipetujeme do 1,5 ml připravených mikrozkuhovek a umístíme na led.
4. Provedeme **desintegraci buněk jehlovým ultrazvukem** při nastaveném výkonu 50 W 30 pulzy po 0.1 s – provádějte pod dohledem vyučujícího s použitím chráničů sluchu, vzorek ochlazujte ledem.
5. Vzorek necháme stát 60 min při laboratorní teplotě.
6. Směs **centrifugujeme** při 14000 x g a 4°C po dobu 20 minut. Supernatant (solubilizovaný proteinový lyzát) přepipetujeme do čisté mikrozkuhavky označené L+číslem skupiny (např. L1) a uložíme na ledu. Mikrozkuhavku s peletem (nedesintegrované zbytky buněk) zlikvidujeme.

Stanovení koncentrace proteinů metodou RC-DC Protein Assay (1. cvičení-pokračování)

7. Připravíme si tři mikrozkušavky: do dvou napipetujeme vždy 5 μ l vzorku L1 (dvě paralelní stanovení – obě označíme S+číslem skupiny, např. S1), do třetí nic (blank označíme např. B1).
8. Se všemi třemi mikrozkušavkami pracujeme paralelně jak je popsáno dále:
9. Přidáme 125 μ l RC reagentu I, zvortexujeme a inkubujeme 1 minutu při laboratorní teplotě.
10. Přidáme 125 μ l RC reagentu II, zvortexujeme.
11. Centrifugujeme 14000 x g, 4 min, 20 °C.
12. Odlijeme supernatant, mikrozkušavku pak obrátíme dnem vzhůru a zbytek supernatantu necháme vytéct.
13. K peletu (sraženina proteinů) přidáme 127 μ l reagentu A', vortexujeme a inkubujeme při laboratorní teplotě 5 minut nebo déle, dokud se sraženina nerozpustí. Před dalším krokem zvortexujeme.
14. Přidáme 1 ml reagentu B, okamžitě zvortexujeme a inkubujeme 15 minut.
15. **Změříme absorpenci při 750 nm** proti blanku, vypočítáme průměrnou hodnotu absorbance z paralelních měření.
16. Koncentraci proteinu určíme z kalibrační závislosti (poskytne vyučující).
17. Vypočítáme objem lyzátu (např. L1) obsahující 100 μ g celkového proteinu (abychom následně nanášeli stejné množství proteinu na všechny kolonky). Naměřené absorbance i výpočet předložíme ke kontrole vyučujícímu.
18. Lyzáty označené písmenem L zmrazíme při -80°C do příštího cvičení.

Štěpení proteinů na peptidy (2. cvičení 26.10.2021)

19. Lyzáty označené písmenem L z předchozího cvičení rozmrazíme uložením na ledu.
20. **Štěpící kolonky** (Microcon 30 kDa, Merck Millipore) s 1,5 ml "low-binding" mikrozkušavkou (=s nízkou mírou interakce s peptidy) označíme číslem skupiny.
21. Napipetujeme 200 μ l denaturačního pufru „UA“ o složení 8 M močovina, 0,1 M Tris/HCl pH 8,5 na kolonku. Poté **přidáme proteinový lyzát** o vypočteném objemu obsahujícím 100 μ g celkového proteinu (bod 17). Na kolonce vše důkladně promícháme špičkou a centrifugujeme při 14000 g/30 min/20 °C. Tím proteiny zakoncentrujeme. Eluát z mikrozkušavky pod kolonkou poté vylijeme.
22. Na kolonku přidáme 100 μ l UA a 20 μ l 100 mM TCEP, promícháme špičkou na filtru. Disulfidické můstky proteinů necháme **redukovat** 30 min/600 rpm/37°C na termomixeru. Následovat bude centrifugace při 14000 g/15 min/20 °C. Eluát z mikrozkušavky pod kolonkou poté vylijeme.
23. Na kolonku přidáme 100 μ l UA a 20 μ l 300 mM IAA, promícháme špičkou na filtru. –SH skupiny cysteinů uvolněné redukcí v předchozím kroku **alkylujeme** 1 min/600 rpm/25°C na termomixeru. Poté vzorky ponecháme bez míchání 20 min v temnu a provedeme centrifugaci při 14000 g/15 min/20 °C. Eluát z mikrozkušavky pod kolonkou poté vylijeme.
24. Na kolonku přidáme 100 μ l 100 mM NH_4HCO_3 a budeme centrifugovat 14000 g/20 min/20 °C. Tím proteiny na kolonce **promyjeme**. Eluát z mikrozkušavky pod kolonkou poté vylijeme.

25. Předchozí krok ještě jednou zopakujeme. Během tohoto kroku připravíme roztok trypsinu (Promega V5111 Trypsin sequencing grade): do každé vialky obsahující 20 µg trypsinu přidáme 20 µl “Resuspension buffer” (50 mM kyselina octová). Zvortexujeme, necháme bez míchání 15 min ve 30°C na termomixeru.
26. Vyměníme spodní mikrozkušavku za čistou 1,5 ml “low-binding” mikrozkušavku (=s nízkou mírou interakce s peptidy) označenou písmenem D a číslem skupiny.
27. Na filtr přidáme 100 µl 50 mM NH₄HCO₃ a 3,33 µl roztoku **trypsinu** (1 µg/µl) (poměr trypsin: protein 1:30), víčka mikrozkušavek překryjeme parafilmem, abychom zabránili vysychání kolony v průběhu štěpení. Promícháme v termomixeru 1 min/600 rpm/37°C
28. Vzorky **ponecháme štěpit přes noc** při 37°C v termomixeru bez třepání (vlhká komůrka).
29. Vyučující ráno provede eluci peptidů z filtru pomocí centrifugace při 14000 g/15 min/20 °C. Naštěpené neodsolené peptidy zmrazí při -20 °C.

Odsolení peptidů (3. cvičení 2.11.2021)

30. Rozmrazíme peptidové vzorky z předchozího cvičení, poté **sonikujeme 5 min na ultrazvukové lázni**. **Odsolovací kolony** MicroSpin C-18 (Nest Group, USA) označíme číslem skupiny.
31. Na odsolovací kolonku nanese 200 µl acetonitrilu (ACN) obsahujícího 0,1% kyseliny trifluoroctovou (TFA), odstraníme případnou bublinu mezi kapalinou a sorbentem a centrifugujeme 100g/3 min/RT. Tento krok provedeme celkem 2x. Tímto kolonku **promyjeme** od případných nečistot **nepolární povahy** z výroby.
32. Na kolonku nanese 200 µl vody + 0,1% TFA, centrifugujeme 300g/3 min/RT. Tímto kolonku **promyjeme** od případných nečistot **polární povahy** z výroby.
33. Na kolonku nanese znovu 200 µl vody + 0,1% TFA, necháme **hydratovat** (stát) 15 min, pak centrifugujeme 300g/3 min/RT.
34. Ze sběrací mikrozkušavky odstraníme obsah, na kolonku **nanese peptidový vzorek** (bod 30), centrifugujeme 500 g/3 min/RT. Tím zachytíme peptidy na kolonce.
35. Přidáme 200 µl vody + 0,1% TFA, centrifugujeme 500g/3 min/RT. Tento krok provedeme celkem 3x. Tím **promýváme peptidy** zachycené na kolonce.
36. Vyměníme sběrací mikrozkušavku za novou “low-binding” mikrozkušavku o objemu 2 ml, původní sběrací mikrozkušavku vyhodíme. Na kolonku přidáme 200 µl 50% ACN + 0,1% TFA, centrifugujeme 500g/3 min/RT. Tím **eluujeme peptidy z kolony**.
37. Přidáme 200 µl 80% ACN + 0,1% TFA, centrifugujeme 500g/3 min/RT. Tím **eluujeme** z kolony peptidy, které se neuvolnily v předchozím kroku.
38. Sběrací mikrozkušavku vyměníme za novou 2 ml “low-binding” mikrozkušavku (původní mikrozkušavku necháme stát na stole i s obsahem), přidáme 200 µl 100% ACN + 0,1% TFA, centrifugujeme 500g/3 min/RT. Tím **eluujeme** z kolony peptidy, které se neuvolnily v předchozím kroku, abychom dosáhli maximální výtěžnosti.
39. Eluát z bodu 38 přidáme k eluátu z bodu 37, vše přepipetujeme do čisté 1,5 ml “low-binding” mikrozkušavky, kterou označíme písmenem P a číslem skupiny.

40. Eluát **lyofilizujeme ve vakuové centrifuze (SpeedVac)** do sucha a zamrazíme při -20°C . Vzorky budou následně analyzovány pomocí LC-MS/MS.

LC-MS/MS analýza (provede laboratoř v průběhu listopadu)

LC-MS/MS měření bude provedeno v “Core Facility Proteomics” CEITEC MU.

41. Peptidy budou rozpuštěny ve 20 μl 0.1% kyseliny mravenčí, pro vlastní analýzu budou použity cca 2 μl 10x zředěného vzorku odpovídající cca 2.5 μg peptidového materiálu.

42. Peptidy budou analyzovány na kapalinovém chromatografu RSLC nano (Thermo Fisher Scientific, USA) s využitím trap kolony 100 μm \times 30 mm plněné sorbentem 3.5- μm X-Bridge BEH 130 C18 (Waters, USA) a **separační kolony Acclaim Pepmap100 C18 column (3 μm částice, 75 μm \times 50 cm; Thermo Fisher Scientific, USA) temperované na 40°C a **online propojené do hmotnostního spektrometru Orbitrap ELITE nebo Orbitrap LUMOS** (Thermo Fisher Scientific, USA). LC gradient bude mít efektivní délku **120 min** (0-5 min 5% mobilní fáze B, 500 nl/min; **5-75 min 5-30% B, 300 nl/min**; 75-110 min 30-80% B, 300 nl/min; 110-120 min 80% B, 300 nl/min). Mobilní fáze A bude sestávat z 0.1% kys. mravenčí ve vodě a mobilní fáze B z 0.1% kys. mravenčí v 80% acetonitrilu.**

43. Akvizice dat v DDA módu bude probíhat s využitím “high collision dissociation” (HCD) fragmentace a detekce v analyzátoru typu Orbitrap s MS1 rozlišením 60000 při 400 m/z a s MS2 rozlišením 15000 při 400 m/z.

Analýza LC-MS/MS dat v programu MaxQuant (nastavení parametrů-termín bude upřesněn)

Pro provedení analýzy LC-MS/MS dat si na cvičení přineste vlastní notebook. Analýza bude provedena vzdáleným přístupem z vašeho notebooku k virtuálnímu počítači.

44. Z Microsoft Store si stáhněte programy Remote Desktop a WinSCP a nainstalujte je na svůj počítač. Při připojení k virtuálnímu počítači postupujte dle pokynů vyučujícího. Na virtuálním počítači máte již stažený program **MaxQuant** z <https://www.maxquant.org/> a surová LC-MS/MS data.

45. Na základě **společného zhodnocení “total ion current” (TIC) profilů** všech LC-MS/MS analýz s vyučujícím přesuňte případně vyloučené analýzy do nové podsložky “Excluded”.

46. Z databáze UniProt (<https://www.uniprot.org/>) si **stáhněte kompletní manuálně revidovanou (Swiss-Prot) databázi lidského proteomu** ve formátu .fasta (cca 20 tis. sekvencí). Označte jako “swissprot_human_(datum stažení např. 20221130)” a pomocí WinSCP ji přesuňte na virtuální počítač do složky home\ubuntu\fasta.

47. Na virtuálním počítači otevřete program MaxQuant a v sekci Surové údaje vyberte soubory k analýze kliknutím na tlačítko “Load”. Pro **označení podle buněčné linie a replikátu** použijte tlačítko Set experiment (např. MCF7_1).

48. V sekci „Parametry specifické pro skupinu“ v nabídce “**Digestion**”, která obsahuje enzymy používané pro digesci proteinů, zvolte trypsin/P. Hodnotu maximálního počtu vynechaných štěpných míst (“Max. missed”) ponechte na 2.

49. V nabídce “**Modifications**” ponechte jako variabilní modifikace oxidaci methioninu (“Oxidation (M)”) a acetylaci N-konců proteinů (“Acetyl (Protein N-term)”). V nastavení fixních modifikací ponechte karbamidometylacii cysteinu (“Carbamidomethyl (C)”).

50. V nabídce “**Label Free Quantification**” aktivujte možnost “LFQ”.
51. V sekci „Globální parametry“ v nabídce “Sequences” vyberte .fasta soubor s lidským proteomem (bod 46).
52. Nastavte 7 vláken (“Number of threads”) dostupných pro MaxQuant při analýze dat. Uložte parametry nastavení: “Save parameters”.
53. Spusťte analýzu tlačítkem “Start”. Jakmile analýza doběhne, vyskočí okno s nápisem “Done”, výsledky naleznete ve složce home\ubuntu\combined \txt.

Statistická analýza dat v programu Perseus (v rámci prezenční nebo online schůzky)

54. Stáhněte si na svůj notebook program **Perseus** z odkazu <https://www.maxquant.org/perseus/> a umístěte jej do složky \Perseus.
55. Přejděte do složky s výsledky MaxQuant analýzy na virtuálním počítači: \home\ubuntu\combined\txt. Pomocí programu WinSCP si zkopírujte složku txt na svůj notebook do \C7195_proteomika\combined\txt. V Excelu otevřete soubor proteinGroups.txt, data importujte s využitím tabulátoru jako oddělovače. Soubor obsahuje seznam identifikovaných proteinových skupin. Projděte si význam jednotlivých sloupců.
56. Obdobně otevřete a projděte soubor peptides.txt. **Počet identifikovaných peptidů** zapište do Tab. 2 v protokolu.
57. Otevřete program Perseus, klikněte na “**Generic matrix upload**” a vyberte soubor proteinGroups.txt ve složce \C7195_proteomika\combined\txt. Vybereme “LFQ intensity” pro všechny analyzované vzorky a přesuneme je do výběru “Main”. Tyto sloupce obsahují LFQ intenzity jednotlivých proteinových skupin v rámci všech analýz, které budeme statisticky hodnotit. Dále přesuňte sloupec “Fasta headers” do výběru “Text” Matrix 1. Analýzu uložte pomocí “Soubor” -> “Uložit jako” jako \C7195_proteomika\combined\txt\Perseus\PersAnal.sps a nezapomeňte průběžně ukládat.
58. Odstraníme proteinové skupiny, které jsou **potenciálními kontaminanty**: Vyberte “Filter rows based on categorical column”, postupně zvolte “Potential contaminant”, “Remove matching rows” a “Reduce matrix”. Matrix 2.
59. Odstraníme proteinové skupiny, jejichž peptidy byly **identifikovány v “decoy” databázi** (falešně pozitivní identifikace): Vyberte “Filter rows based on categorical column”, postupně zvolte “Reverse”, “Remove matching rows” a “Reduce matrix”. Matrix 3.
60. Odstraníme proteinové skupiny, které byly identifikovány pouze na základě **posttranslačních modifikací** (PTM). Vyberte “Filter rows based on categorical column”, postupně zvolte “Only identified by site”, “Remove matching rows” a “Reduce matrix”. Matrix 4. **Počet identifikovaných proteinových skupin** zapište do Tab. 2 v protokolu.
61. LFQ intenzity převedeme do log2 tvaru: “Basic” -> “Transform” -> LFQ intenzity. Matrix 5.
62. Za chybějící hodnoty (“**missing values**”) dosadíme hodnoty hladin šumu podle normálního rozdělení: “Imputation” -> “Replace missing values from normal distribution”. Tím zabráníme dělení nulou v dalším průběhu analýzy. Matrix 6.
63. Sloupce obsahující LFQ intenzity **přiradíme do skupin** podle buněčných linií: “Annot. rows” -> “Categorical annotation rows”. Pro přiřazení vzorku do skupiny klikněte v tabulce na řádek

s uvedeným označením vzorku, poté klikněte na zaškrtnuté políčko ve výběru “Values” a smažte poslední dva znaky v označení vzorku. [Matrix 7](#).

64. **ANOVA** testem vyhodnotíme statisticky významně změněné proteinové skupiny v rámci souboru vzorků všech tří buněčných linií: “Tests” → “Multiple-sample tests” → “ANOVA”. Ponecháme základní nastavení, pouze deaktivujte možnost “-Log 10”. Test proběhne s korekcí na mnohonásobné testování - bude aplikována tzv. “false discovery rate” (FDR) korekce, která zajistí, že soubor statisticky významně změněných proteinových skupin bude obsahovat maximálně 5% falešně pozitivních výsledků (FDR<0.05). [Matrix 8](#).

65. Pro statistické srovnání proteotypů dvojic buněčných linií provedeme **sérii tří Studentových t-testů**: “Tests” → “Two-sample tests”. Provedeme pro všechna porovnání: MFC-7 vs. MDA-MB-231, MCF-7 vs. A431 a MDA-MB-231 vs. A431. Ponecháme základní nastavení, pouze deaktivujte možnost “-Log 10 p-value”. Výstupem je q-hodnota, tedy p-hodnota opět korigovaná na mnohonásobné testování (q<0.05). [Matrix 13](#). *Poznámka: Analýzu lze provést též pomocí post-hoc testu v návaznosti na ANOVA, software Perseus však ve stávající verzi neposkytuje p-/q-hodnotu výsledku.*

66. Vypočítáme **průměr (“Mean”) LFQ intenzit** v rámci vzorků jednotlivých buněčných linií: provedeme pomocí “Average groups”. Hodnoty intenzit individuálních měření nebudeme mazat: vybereme “keep original data”. [Matrix 15](#).

67. Odečteme průměry log₂ LFQ intenzit proteinových skupin mezi dvojicemi buněčných linií, čímž **získáme hodnoty “Log₂ Fold Change” (log₂FC)** udávající míru změny hladin proteinových skupin: “Basic” → “Combine main columns” → x-y. Toto provedeme celkem 3x pro všechna porovnání: MCF-7 vs. MDA-MB-231, MCF-7 vs. A431 a MDA-MB-231 vs. A431. log₂FC=1 udává dvojnásobné zvýšení (2¹=2) hladiny proteinové skupiny, log₂FC=0 udává nulovou (2⁰=0) změnu a log₂FC=-1 udává dvojnásobné snížení (2⁻¹=0.5) hladiny proteinové skupiny ve vzorcích jedné buněčné linie oproti vzorkům druhé buněčné linie. [Matrix 18](#).

68. Provedeme **normalizaci**: “Normalization” → “Z-score”. V nabídce “Matrix access” vybereme “Rows”, v nabídce “Grouping” ponechte “No grouping” a zaškrtněte “Use median”. Tím bude provedena normalizace log₂ LFQ intenzit každé proteinové skupiny napříč všemi vzorky, a to odečtením mediánu a vydělením výsledného rozdílu LFQ intenzit hodnotou směrodatné odchylky LFQ intenzit proteinové skupiny ve všech vzorcích, čímž se odstraní rozdíly dané například různými dávkováními vzorků z technických příčin. [Matrix 19](#).

69. Pro shlukovací analýzu **vybereme** statisticky významně změněné proteinové skupiny na základě výsledků ANOVA analýzy: “Filter rows” → “Filter rows based on categorical column” – v nabídce “Column” vyberte “ANOVA Significant”. “Mode” nastavte na “Keep matching rows” a “Filter mode” na “Reduce matrix”. [Matrix 20](#).

70. Provedeme vlastní **shlukovací analýzu vzorků a proteinových skupin**: “Clustering/PCA” → “Hierarchical clustering”. V nabídkách “Rows tree” a “Columns tree” nastavte “Distance” na “Spearman correlation” a klikněte na “OK”. Zobrazení vygenerované heatmapy si přizpůsobte nástrojem “Properties” → v sekci “Row clustering” nastavte “Header width” na 200 a “Names width” na 100, a v sekci “Column clustering” nastavte “Header height” na 300 a “Names height” na 200. Legendu heatmapy popisující barevnou škálu hodnot log₂ LFQ intenzit si zobrazíte kliknutím na “Change color gradient”. Dále vyberte nástroj “Define row clusters”, “Number of clusters” nastavte na 6 a klikněte na “Apply”. Tím se proteiny v heatmapě rozdělí do šesti barevně rozlišených klastrů. Pomocí “Export image” exportujte obrázek heatmapy a uložte jej jako \C7195_proteomika\combined\txt\Perseus\heatmap.jpg. V programu PowerPoint do heatmapy

vložte čísla klastrů proteinových skupin podle **Tabulky 3 v části Protokol**. Shlukovací analýza na základě Spearmanovy korelace ukazuje klastry korelujících (i) vzorků a (ii) proteinových skupin, přičemž vzorky/proteinové skupiny ve stejném klastru mají nejvyšší pozitivní korelaci. V našem případě tedy klastry vypovídají o tom, jestli má daná skupina proteinů zvýšené nebo snížené hladiny v jedné buněčné linii oproti ostatním. (V případě potřeby lze obohacení jednotlivých klastrů o funkční kategorie (GO) proteinů analyzovat pomocí Fisherova exaktního testu; návod pro tento a další postupy lze najít na <http://coxdocs.org/doku.php?id=perseus:start>).

71. Exportujeme výsledky porovnání proteotypů buněčných linií – analýzy ANOVA a t-testů do Excelu: Z Matrix 19 exportujte tabulku pomocí “Generic matrix export” a uložte ji jako `\C7195_proteomika\combined\txt\Perseus\ANOVA_ttests.txt`. V Excelu tento soubor otevřete, data importujte s využitím tabulátoru jako oddělovače a uložte jako `\C7195_proteomika\combined\txt\Perseus\ANOVA_ttests.xlsx`. Odstraňte druhý a třetí řádek a v prvním řádku aplikujte filtr: Data/Filtr. Za diferenciálně regulované považujeme proteinové skupiny s $q < 0,05$ a \log_2FC vyšší než 1 nebo nižší než -1. Pro zjištění počtu deregulovaných proteinů po aplikaci příslušných filtrů klikněte ve sloupci “Protein IDs” na buňku na řádku 2 a stiskněte levé ctrl + levý shift + šipka dolů, v pravém dolním rohu se zobrazí počet označených buněk (= deregulovaných proteinů).

72. S využitím odpovídajících kombinací filtrů v Excelu vyplňte Tabulky 4 a 5 v části Protokol. Pro zjištění počtu diferenciálně regulovaných proteinů lze použít kombinaci kláves SHIFT+CTRL. **Na základě tabulky 5 společně diskutujte, co lze z profilů klíčových proteinových markerů usuzovat o rozdílených fenotypech analyzovaných buněčných linií a jejich schopnostech metastázovat.**

Analýza obohacených regulačních drah pomocí GSEA

73. Stáhneme a nainstalujeme software **GSEA**. Přejděte na následující odkaz (<http://software.broadinstitute.org/gsea/index.jsp>), klikněte na “Download” a na “Click here”. Uvedte svůj email a jako organizaci uveďte “Masaryk University”. Klikněte na “Continue” a v liště “Downloads”, stáhněte instalační soubor, spusťte jej a nainstalujte software GSEA.

74. Připravíme **3 vstupní datové soubory** pro GSEA: V Excelu vytvořte z `\C7195_proteomika\combined\txt\Perseus\ANOVA_ttests.xlsx` samostatné tabulky pro každé porovnání (MFC-7 vs. MDA-MB-231, MCF-7 vs. A431, a MDA-MB-231 vs. A431), které v jednom sloupci obsahují zkratky genů a ve druhém sloupci hodnoty \log_2FC . Sloupec „Fasta headers“ zkopírujte do prvního sloupce prázdného listu a použijte nástroj „Text do sloupců“, jako oddělovač použijte „=“. Poté označte sloupec D a opět použijte „Text do sloupců“, jako oddělovač použijte mezeru. Obsah sloupce D, který představuje seznam názvů genů, pak zkopírujte do samostatných tabulek pro každé porovnání a vložte zkopírovaný sloupec obsahující \log_2FC hodnoty pro dané porovnání. Geny seřadte sestupně podle hodnot \log_2FC . U hodnot \log_2FC je pro další zpracování nutno použít desetinnou tečku místo čárky. Pro odstranění řádků s chybějícími názvy genů označte první sloupec, na kartě „Domů“ ve skupině „Úpravy“ klikněte na položku „Najít a vybrat“ → „Přejít na – jinak“. Zvolte „Prázdné buňky“, klikněte na tlačítko „OK“, na kartě „Domů“ ve skupině „Buňky“ klepněte na tlačítko „Odstranit“ → „Odstranit řádky listu“. Každou z tabulek uložte do složky `\C7195_proteomika\combined\txt\GSEA\` jako textový soubor s daty oddělenými tabulátory: `GSEA_MCF7vsMDAMB231.txt`, `GSEA_MCF7vsA431.txt` a `GSEA_MDAMB231vsA431.txt`. Poté příponu souboru ručně přepište na .rnk.

75. V programu GSEA **nahrajte** pomocí “Load data” všechny tři vstupní .rnk soubory.

76. Provedeme **vlastní GSEA analýzu**: “Tools” → “Run GSEA Preranked” – v nabídce “Gene sets database” vyberte nejnovější verzi databáze HALLMARK, která obsahuje definované biologické dráhy. V nabídce “Ranked list” vyberte jeden ze vstupních .rnk souborů. Protože nepracujeme

s daty z cDNA mikročipů, v nabídce “Collapse/Remap to gene symbols” vyberte “No_Collapse” a nastavení “Chip platform” ponechte prázdné. V rámci “Analysis name” postupně uveďte názvy složek, do kterých GSEA uloží výsledky analýz: “GSEA_MCF7vsMDAMB231”, “GSEA_MCF7vsA431” nebo “GSEA_MDAMB231vsA431”. Pole “Save results in this folder” nastavte na “C:\C7195_proteomika\combined\txt\GSEA\”. Hodnotu “Min size: exclude smaller sets”, která udává minimální počet obohacených genů v dráze, nastavte na 2 a hodnotu “Plot graphs for the top sets of each phenotype”, která udává počet drah v jedné analýze, pro které GSEA vygeneruje detailní výsledky zahrnující seznamy obohacených genů, nastavte na 50. Analýzu spustíme tlačítkem “Run”. Tento postup provedeme pro všechna tři porovnání, tzn. se všemi třemi vstupními .rnk soubory.

77. Uspořádáme výsledky GSEA analýz: Ve složce C:\C7195_proteomika\combined\txt\GSEA\ máme nyní tři nové složky pojmenované GSEA_MCF7vsMDAMB231, GSEA_MCF7vsA431 a GSEA_MDAMB231vsA431, které obsahují výsledky GSEA analýz pro každé porovnání. V každé z těchto tří složek otevřeme výsledky GSEA analýz přes soubor index.html. V každé z těchto složek se také nachází soubor gsea_report_for_na_pos.xlsx, který obsahuje seznam pozitivně obohacených drah pro dané porovnání, a soubor gsea_report_for_na_neg.xlsx, který obsahuje seznam negativně obohacených drah pro dané porovnání. Soubory gsea_report_for_na_pos.xlsx a gsea_report_for_na_neg.xlsx si zkopírujete do složky \C7195_proteomika\combined\txt\GSEA\ a přejmenujte na

“gsea_report_for_na_pos_MCF7vsMDAMB231.xlsx”, “gsea_report_for_na_neg_MCF7vsA431.xlsx”,
“gsea_report_for_na_neg_MCF7vsMDAMB231.xlsx”, “gsea_report_for_na_pos_MDAMB231vsA431.xlsx”,
“gsea_report_for_na_pos_MCF7vsA431.xlsx”, “gsea_report_for_na_neg_MDAMB231vsA431.xlsx”.

U každé deregulované dráhy budeme sledovat dva parametry – “Enrichment Score” (ES), které udává míru obohacení genů dané dráhy, a “Nominal p-value” (NOM p-val), které udává statistickou významnost. Relevantní budou dráhy s NOM p-val nižší než 0,05. Obohacené geny jednotlivých drah jsou uvedeny ve výběru “Details” pro každou dráhu. Na prvním řádku aplikujte filtr: Data/Filtr.

78. S využitím odpovídajících kombinací filtrů v Excelu (Data/Filtr) vytvořte tabulky 6-11 v části Protokol.

79. HALLMARK dráhy typické pro každou z analyzovaných buněčných linií doplňte do tabulky 12, uveďte také obohacené geny v rámci těchto drah (viz “Details” ve výsledcích GSEA). Diskutujte, zda tyto geny mohou představovat potenciální cíle biologické terapie.

Jméno a příjmení	Číslo skupiny

Protokol

Elektronický protokol bude sestávat z 10 souborů ve formátu Excel vytvořených v průběhu analýzy, Perseus datasetu a vyplněné části Protokol ve formátu Word společně zazipovaných v jednom archivu označeném názvem úlohy, číslem skupiny a příjmením autora (ve tvaru Proteomika_sk1_Novák.zip). Tento archiv vložte do Odevzdávrny předmětu v IS MUNI do složky C7195_B_Proteomika.

Dataset 1. Seznam všech identifikovaných peptidů

soubor \C7195_proteomika\combined\txt\peptides.txt uložený ve formátu Excel (peptides.xlsx)

Dataset 2. Seznam všech identifikovaných proteinových skupin

soubor \C7195_proteomika\combined\txt\proteinGroups.txt uložený ve formátu Excel (proteinGroups.xlsx)

Dataset 3. Seznam diferenciálně regulovaných proteinových skupin – porovnání proteotypů buněčných linií

soubor \C7195_proteomika\combined\txt\Perseus\ANOVA_ttests.xlsx

Dataset 4. Kompletní Perseus dataset

soubor \C7195_proteomika\combined\txt\Perseus\PersAnal.sps

Dataset 5-10. Seznam drah obohacených v GSEA analýze -soubory

\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_pos_MCF7vsMDAMB231.xlsx,
\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_neg_MCF7vsMDAMB231.xlsx,
\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_pos_MCF7vsA431.xlsx,
\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_neg_MCF7vsA431.xlsx,
\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_pos_MDAMB231vsA431.xlsx a
\C7195_proteomika\combined\txt\GSEA\gsea_report_for_na_neg_MDAMB231vsA431.xlsx.

Vlastní protokol:

a. Zpracování vzorku

Podtrhněte vzorek, se kterým pracovala vaše skupina:

- Buněčná linie MCF-7
- Buněčná linie MDA-MB-231
- Buněčná linie A431

V tabulce 1 uveďte absorbance, koncentrace proteinových lyzátů a vypočítaný objem proteinového lyzátu, který odpovídá 100 μg proteinu.

A ₇₅₀ – 1. replikát	
A ₇₅₀ – 2. replikát	
Koncentrace proteinu ($\mu\text{g}/\mu\text{l}$)	
Objem lyzátu odpovídající 100 μg proteinu (μl)	

Tab. 1. Výsledky stanovení koncentrace celkového proteinu

b. Souhrnné počty identifikovaných peptidů a proteinových skupin

Z **Datasetu 1** uveďte celkové počty identifikovaných peptidů a z **Datasetu 2** celkové počty identifikovaných proteinových skupin:

Počet peptidů	(FDR=0.01)
Počet proteinových skupin	(FDR=0.01)

Tab. 2. Počty identifikovaných peptidů a proteinových skupin při FDR 1%

c. Výsledky shlukovací analýzy

Vložte heatmapu klastrování vzorků a proteinových skupin vytvořenou na základě analýzy ANOVA z \C7195_proteomika\combined\txt\Perseus\heatmap.jpg:

Do jaké míry se shlukují biologické replikáty jednotlivých buněčných linií?

Uveďte v obrázku heatmapy čísla šesti proteinových klastrů tak, aby bylo jasné, jaký typ porovnání a deregulace každý klastr reprezentuje.

Číslo klastru	Typ deregulace
1	MCF-7 (up) vs. MDA-MB-231
2	MCF-7 (up) vs. A431
3	MDA-MB-231 (up) vs. MCF-7
4	A431 (up) vs. MDA-MB-231
5	A431 (up) vs. MCF-7
6	MDA-MB-231 (up) vs. A431

Tab. 3. Význam klastrů ve výsledcích shlukové analýzy

d. Diferenciálně regulované proteinové skupiny

Z **Datasetu 3** uveďte počty proteinových skupin se statisticky významně zvýšenými nebo sníženými hladinami (podmínky viz záhlaví tabulky) v proteotypech buněčných linií na základě tří t-testů:

Porovnání	Počet proteinových skupin se zvýšenou hladinou $\log_2FC > 1, q < 0.05$	Počet proteinových skupin se sníženou hladinou $\log_2FC < -1, q < 0,05$
MFC-7 vs. MDA-MB-231		
MCF-7 vs. A431		
MDA-MB-231 vs. A431		

Tab. 4. Počty diferenciálně regulovaných proteinových skupin

Z **Datasetu 3** doplňte log2FC a q-hodnoty pro klíčové markerové proteiny AGR2, EPCAM, CDH1, KRT18, VIM, CD44, EGFR, STAT1, P53.

	Gen	MFC-7 vs. MDA-MB-231		MCF-7 vs. A431		MDA-MB-231 vs. A431	
		log2FC	q-hodnota	log2FC	q-hodnota	log2FC	q-hodnota
ER status	<i>AGR2</i>						
adheze	<i>EPCAM</i>						
epiteliální markery	<i>CDH1</i>						
	<i>KRT18</i>						
mezenchymální marker	<i>VIM</i>						
Wnt signalizace Metastazování	<i>CD44</i>						
EGFR dráha	<i>EGFR</i>						
proliferace	<i>STAT1</i>						
p53 hladina	<i>TP53</i>						

Tab. 5. Rozdíly v hladinách klíčových proteinových markerů mezi vybranými buněčnými liniemi

e. Obohacené dráhy – GSEA

Z **Datasetů 5-10** obsahujících obohacené dráhy v GSEA extrahujte ty, které jsou obohacené statisticky významně: “NOM p-val” < 0,05 a současně “Enrichment score (ES)” > 0 (pozitivně obohacené), nebo “NOM p-val” < 0,05 a současně ES < 0 (negativně obohacené). Výsledky po filtraci v Excelu zkopírujte jako tabulky 6-11 přímo z Excelu (nepotřebné sloupce skryjte) pro jednotlivá porovnání: MFC-7 vs. MDA-MB-231, MCF-7 vs. A431 a MDA-MB-231 vs. A431, vždy zvlášť pozitivně a negativně obohacené dráhy, podle vzoru viz níže.

NAME	SIZE	ES	NES	NOM p-val
HALLMARK_ESTROGEN_RESPONSE_LATE	87	0.5770732	1.9327933	0
HALLMARK_OXIDATIVE_PHOSPHORYLATION	168	0.4408747	1.616541	0
HALLMARK_ESTROGEN_RESPONSE_EARLY	72	0.49181032	1.6173242	0.005772006
HALLMARK_KRAS_SIGNALING_DN	19	0.6805452	1.7204058	0.008064516
HALLMARK_XENOBIOTIC_METABOLISM	86	0.41837338	1.3944362	0.03079179
HALLMARK_CHOLESTEROL_HOMEOSTASIS	49	0.45950735	1.4193991	0.031866465

Tab. 6. Pozitivně obohacené dráhy v porovnání MCF-7 vs. MDA-MB-231. Vzor (smažte), Vaši tabulku vložte zkopírováním tabulky z MS Excel, nepotřebné sloupce před kopírováním skryjte.

Tab. 7. Negativně obohacené dráhy v porovnání MCF-7 vs. MDA-MB-231.

Tab. 8. Pozitivně obohacené dráhy v porovnání MCF-7 vs. A431.

Tab. 9. Negativně obohacené dráhy v porovnání MCF-7 vs. A431.

Tab. 10. Pozitivně obohacené dráhy v porovnání MDA-MB-231 vs. A431.

Tab. 11. Negativně obohacené dráhy v porovnání MDA-MB-231 vs. A431.

Které dráhy jsou obohacené v konkrétní buněčné linii ve srovnání proti dvěma zbývajícím? Vypište obohacené geny/proteiny v rámci GSEA analýzy.

MCF7:

Dráha Obohacené geny/proteiny

MDA-MB-231:

Dráha Obohacené geny/proteiny

A431:

Dráha Obohacené geny/proteiny

Tabulka 12. Dráhy typické pro jednotlivé buněčné linie a geny obohacené v rámci GSEA analýzy.

f. Souhrnná Interpretace výsledků

1. Co říkají profily klíčových proteinových markerů a obohacené dráhy o rozdílných fenotypech analyzovaných buněčných linií?
2. Na základě získaných poznatků sestavte teoretické pořadí buněčných linií podle snižujícího se metastatického potenciálu.
3. Na které dráhy (na základě GSEA) a konkrétní genové produkty byste cílili biologickou léčbu?