

Szegedi Tudományegyetem
Szent-Györgyi Albert Orvostudományi Kar
Interdiszciplináris Orvostudományi Doktori Iskola

**KIS VEZIKULÁKBAN REJLŐ NAGY LEHETŐSÉG:
AZ EXTRACELLULÁRIS VEZIKULÁK MOLEKULÁRIS UJJLENYOMATÁNAK
TUMORDIAGNOSZTIKAI CÉLÚ VIZSGÁLATA GÉPI TANULÁSI MÓDSZEREKKEL**

TÉZISFÜZET

Ph.D. értekezés

Bukva Mátyás

Témavezető

Buzás Krisztina, Ph.D.

Immunológiai Tanszék, Szent-Györgyi Albert Orvostudományi Kar, Szegedi Tudományegyetem
Természettudományi és Informatikai Kar, Szegedi Tudományegyetem
Biokémiai Intézet, Biológiai Kutatóközpont, Szeged

2024



**HUN
REN**



A PhD disszertáció az alábbi közleményeken alapszik:

Bukva, M., Dobra, G., Gyukity-Sebestyén, E., Boroczky, T., Korsos, M. M., Meckes, D. G., Jr, Horvath, P., Buzas, K., & Harmati, M. (2023). Machine learning-based analysis of cancer cell-derived vesicular proteins revealed significant tumor-specificity and predictive potential of extracellular vesicles for cell invasion and proliferation - A meta-analysis. *Cell communication and signaling: CCS*, 21(1), 333. <https://doi.org/10.1186/s12964-023-01344-5>. (IF: 8.444, Q1)

Bukva, M., Dobra, G., Gomez-Perez, J., Koos, K., Harmati, M., Gyukity-Sebestyén, E., Biro, T., Jenei, A., Kormondi, S., Horvath, P., Konya, Z., Klekner, A., & Buzas, K. (2021). Raman Spectral Signatures of Serum-Derived Extracellular Vesicle-Enriched Isolates May Support the Diagnosis of CNS Tumors. In *Cancers* (Vol. 13, Issue 6, p. 1407). MDPI AG. <https://doi.org/10.3390/cancers13061407> (IF: 6.575, Q1)

További, a tématerülethez kötődő közlemények:

Dobra, G., Gyukity-Sebestyén, E., **Bukva, M.**, Harmati, M., Nagy, V., Szabó, Z., Pankotai, T., Klekner, Á., & Buzás, K. (2023). MMP-9 as prognostic marker for brain tumours: A comparative study on serum-derived small extracellular vesicles. *Cancers*, 15(3), 712. <https://doi.org/10.3390/cancers15030712> (IF: 5.207, Q1)

Böröczky, T., Dobra, G., **Bukva, M.**, Gyukity-Sebestyén, E., Hunyadi-Gulyás, É., Darula, Z., Horváth, P., Buzás, K., & Harmati, M. (2023). Impact of experimental conditions on extracellular vesicles' proteome: A comparative study. *Life* (Basel, Switzerland), 13(1), 206. <https://doi.org/10.3390/life13010206> (IF: 3.212, Q2)

Harmati, M., **Bukva, M.**, Böröczky, T., Buzás, K., & Gyukity-Sebestyén, E. (2021). The role of the metabolite cargo of extracellular vesicles in tumor progression. *Cancer Metastasis Reviews*, 40(4), 1203–1221. <https://doi.org/10.1007/s10555-021-10014-2> (IF: 9.237, Q1)

Dobra, G., Bukva, M., Szabo, Z., Bruszel, B., Harmati, M., Gyukity-Sebestyén, E., Jenei, A., Szucs, M., Horvath, P., Biro, T., Klekner, A., & Buzas, K. (2020). Small extracellular vesicles isolated from serum may serve as signal-enhancers for the monitoring of CNS tumors. *International Journal of Molecular Sciences*, 21(15), 5359. <https://doi.org/10.3390/ijms21155359> (IF: 5.924, Q1)

Kumulatív IF: 60.69

1. BEVEZETÉS

Az alábbi téziszfüzet a teljes PhD értekezés leglényegesebb elemeit tartalmazza. A *Bevezetésben* a téziszfüzet rövid áttekintést nyújt a multidiszciplináris kutatás elméleti alapjairól. Ennek során több kulcsfontosságú terület kerül bemutatásra: az extracelluláris vezikulák, azok szerepe a tumoros folyamatokban, jelentőségük a biomarker kutatásban, a Raman spektroszkópia potenciális felhasználási lehetősége a tumordiagnosztikában, valamint a gépi tanulás előnyei a rákkutatásban.

1.1. Az extracelluláris vezikulák biológiai tulajdonságai

Az extracelluláris vezikulák (EVk) minden élő sejt által az extracelluláris térbe szekretált lipid kettősréteggel körülhatárolt részecskék, amelyek két alapvető típusba sorolhatók: exoszómák és ektoszómák. Az exoszómák endoszómális eredetűek, és az endoszómális membránok invaginációjával jönnek létre. Ez a folyamat intraluminális vezikulákat magába záró multivezikuláris testek (MVB-k) kialakulásához vezet. A keletkező MVB-k összeolvadnak a sejtek membránjával, ezzel felszabadítva a bennük szállított intraluminális vezikulákat. Az így az extracelluláris térbe kerülő vezikulákra ettől kezdve exoszómákként hivatkozunk. Az ektoszómák ezzel szemben közvetlenül plazmamembránból származó EV-k, melyek a membrán lefűződésével keletkeznek.

A különböző keletkezési útvonalakra vonatkozóan azonban még nem állnak rendelkezésünkre kizárólagos molekuláris markerek; ehelyett operatív kifejezések javasoltak az EV-típusok megkülönböztetésére például biofizikai (pl. sűrűség, méret), vagy biokémiai tulajdonságaik (jelenlévő marker) alapján. Jelenleg a méretszerinti csoportosítás a leggyakrabban alkalmazott, mely szerint megkülönböztethetünk kisméretű EV-eket (sEV) (50-200 nm), közepes méretű EV-eket (200-1000 nm) és nagyméretű EV-eket ($\geq 1 \mu\text{m}$).

Ezen dolgozat, amennyiben más kutatásokat idéz, az idézett munkában használt nevezéktant, csoportosítást alkalmazza. Saját kutatásainkban a méretalapú osztályozást használtunk.

Bár az EV-k keletkezési útjuk szerint különbözőek lehetnek, molekuláris összetételük hasonló: lipideket, fehérjéket, nukleinsavakat és kis molekulatömegű metabolitokat tartalmaznak. A minden EV-ben megtalálható molekulákon túl sejtípus-specifikus komponenseket is hordoznak, amelyek a donorsejtek (EV-eket kibocsátó sejtek) jellemzőit tükrözik.

Átfogó jelentőségük abban rejlik, hogy az általuk szállított molekulákon keresztül képesek információt közvetíteni a donorsejtek a recipiens sejtek (az EV-kel kapcsolatba kerülő

sejtek) között. Az EV-ken keresztül lezajló kommunikáció szabályozza mind a normál élettani (pl. terhesség, immunválasz, neuronális fejlődés, angiogenezis, vérárvadás), mind a pathológiai folyamatokat (pl. daganatos betegségek, neurodegeneratív kórképek).

1.2 Az extracelluláris vezikulák a biomarkerek kutatásában

Eddig számos klinikai tanulmány rávilágított az EV-k széleskörű egészségügyi felhasználási lehetőségeire, beleértve az onkoterápiát, a vakniációt, az immunmoduláló vagy regeneratív terápiákat, valamint a gyógyszerek céljajuttatását. Az EV-k különösen a folyadékbiopszián alapuló biomarkerkutatás területén válhatnak kulcsfontosságú szereplővé a jövőben. Korábbi kutatások kimutatták, hogy a daganatos betegek vérében jelentősen magasabb az EVk átlagos koncentrációja a nem daganatos páciensekhez képest, amelyek a tumorszövetből és a tumorról kapcsolatos immunfolyamatokból származnak. Tekintettel arra, hogy — molekuláris tartalmuknál fogva — számos daganatos folyamatban részt vesznek, a betegek szérum- vagy plazmamintáiból izolált EV-k a tumoros állapot egészét jellemző biomarkerek gazdag forrásai lehetnek.

1.3. Megközelítések a biomarkerek kutatásában

Az EV-k által hordozott molekuláris tartalom vizsgálatára jelenleg két megközelítést alkalmaznak: (I) az egyes molekulatípusok külön-külön történő elemzése omikai módszerekkel (pl. proteomika, lipidomika, genomika), és (II) az EV-k teljes molekuláris komplexitásának jellemzése rezgési spektroszkópiai módszerekkel (pl. Raman-spektroszkópiával).

Az első módszer jól standardizált, és számos tanulmányban bizonyította sikerességét, amikor specifikus molekulák meghatározásáról van szó, de bizonyos esetekben korlátokba ütközik. Például a különböző omikai megközelítésekből származó eredmények sok esetben nincsenek átfedésben, ami a rák biológiájának hiányos megértéséhez vezethet. Példaként szolgálnak erre azok a tumoros betegségek, melyek kifejezetten nagy variabilitást mutatnak a páciensek között (például a glioblasztóma multiforme, GBM). Ennél fogva egy adott protein vagy nukleinsav biomarker, mely a vizsgált betegmintánkon klinikai jelentőséggel bír, elveszítheti relevanciáját új betegminták vizsgálatakor.

A második megközelítés, a rezgési spektroszkópia, például a Raman-spektroszkópia használata új és ígéretes módot kínál az EV-k teljes molekuláris összetételének elemzésére. A Raman-spektroszkópia képes karakterizálni egy minta teljes molekuláris tartalmát a fényszóródás jelenségére támaszkodva. A teljes molekuláris összetétel egyidejű vizsgálata

lehetővé teszi, hogy ne korlátozódjunk csak egyféle molakulatípusra; így egy sokkal általánosíthatóbb diagnosztikai modell hozható létre.

Bár a két megközelítés lényegesen különböző, az általuk generált jelentős mennyiségű adat fejlett elemzési stratégiákat igényel.

1.5 A gépi tanulási módszerek elterjedése a rákkutatásban

A gépi tanulás —mely a mesterséges intelligencia egyik ága — egyre inkább elterjedt a rákkutatásban, mivel képes a komplex és nagy adahalmazokban megbúvó összefüggések elemzésére. A hagyományos statisztikai módszereknek nehézséget okoz a rákkutatásban jelen lévő nagy dimenziójú és változatos adattípusok kezelése. A gépi tanulási algoritmusok ezzel szemben önállóan, előzetes feltételezések nélkül tárják fel a változók közötti bonyolult kapcsolatokat. Előnyös tulajdonságaiknak köszönhetően jól alkalmazhatóak a nem lineáris, bonyolult kapcsolatok leírására, melyek kifejezetten gyakoriak a modern rákkutatásból származó, biológiai „big data-ban”. Az egyre fejlődő gépi tanulási algoritmusok alkalmazása a rákkutatásban, különösen a biomarkerek kutatásában, jelentőségek javíthatják a daganatos betegek kilátásait.

1.6 Kutatási igény, a doktori értekezés témája

Napjainkig számos kutatás rávilágított, hogy az EVk molekuláris tartalmának vizsgálata lehetővé teszi nemcsak a tumoros és egészséges páciensek megkülönböztetését, de az egyes daganattípusok további alcsoportokra való bontását (például a kemoszenzitivitás/rezisztencia szerint).

Az EV-k potenciális klinikai felhasználhatóságát illetően azonban még mindig akadnak kevésbé ismert, feltérképezett területek.

Például nem ismert, hogy az EV-k molekuláris összetételének vizsgálatával megjósolható-e a donorsejtek invázióképesége és osztódási sebessége, mely tulajdonságok központi szerepet játszanak a daganatok terjedésében.

Továbbá, mivel a legtöbb tanulmány korlátozott számú betegcsoportot vizsgál, nem teljesen tisztázott, milyen mértékben lehet specifikus a különböző tumortípusok által kiocátott EV-k molekuláris tartalma.

Végül, bár számos kutatás rávilágított, hogy az EV-k Raman-spektroszkópiai vizsgálata kimagasló potenciállal bír a tumordiagnosztikában, a módszer hatékonysága a rendkívül heterogén vagy jelentős diagnosztikai kihívásokat jelentő tumortípusok, például a központi idegrendszeri (CNS) tumorok elemzésében továbbra sem egyértelmű.

Ezen hiányosságok tükrében a jelen dolgozat két tanulmány eredményeit foglalja össze. Az első rész egy több adatbázist felölelő meta-analízis keretén belül kilenc tumortípusból származó EV-k proteomjának specificitását, valamint prognosztikai potenciálját tárja fel leginkább *in vitro*, részben pedig *in vivo* adatokat felhasználva.

A második részt egy klinikai tanulmány eredményei alkotják, melyben CNS tumorban szenvedő betegek szérummintáiból izolált sEV minták Raman spektroszkópiai elemzését végeztük el tumordiagnosztikai célból.

A két tanulmány közös jellemzője, hogy a hagyományos statisztikai módszerek mellett nagymértékben támaszkodnak gépi tanulási módszerekre az EV-kből származó adatok feldolgozásához, ezzel aknázva ki az általuk hordozott lehetőségeket.

2. CÉLKITŰZÉS

A dolgozat elsődleges célja, hogy rávilágítson az EV-k által hordozott tumorasszociált molekuláris tartalomban rejlő lehetőségekre. Ez magában foglalja az EV-k potenciális szerepét a tumordiagnosztikában, a differenciáldiagnosztikában, a prognózisban és a gyógyszercélpontok meghatározásában a rák jellemzőinek megértése révén.

E célból a dolgozat két tanulmány, egy *in vitro* adatokon végzett meta-analízis és egy klinikai széruminatokon végzett vizsgálat eredményeit foglalja össze.

A kilenc tumortípusból származó 60 különböző sejtvonal (NCI-60 panel) felülűszójából izolált EV-k proteomjának *meta-analízise* során a következő célokat tűztük ki:

1. A teljes EV proteom, valamint mind a 60 EV-mintában közös fehérjék tumorspecifikusságának felmérése, illetve a kilenc tumortípust a leghatékonyabban megkülönböztető diszkriminatív fehérjepanel összeállítása.
2. A diszkriminatív fehérjepanel biológiai funkcióinak feltására tumorspecifikus mintázatok keresve.
3. A donorsejtek inváziós képességét és osztódási idejét előrejelező fehérjepanel összeállítása.

A *klinikai vizsgálat* során glioblastoma multiforme-ban, agyi metasztázisban, meningeomában és lumbális porckorongsérvtben szenvedő betegek széruminataiból származó sEV-dúsított izolátumok Raman-spektroszkópiai elemzését végeztük el az alábbi célt tűzve ki:

4. Egy olyan osztályzási modell létrehozása, mely képes különbséget tenni a különböző betegcsoportok között a sEV-dúsított izolátumok Raman spektrumai alapján.

3. ANYAGOK ÉS MÓDSZEREK

3.1. Az *in vitro* adatok metaanalízise

A meta-analízis alapjául szolgáló proteomikai adatállományt *Hurwitz és munkatársai* szolgáltatották. Kutatásukban a sejtvonalak felülűszójából izolált vezikulákra „EV-ként” hivatkoztak. Következésképpen ez a dolgozat az „EV” kifejezést használja az ezen adathalmazhoz kapcsolódó eredmények tárgyalásakor. Az 1. táblázat összefoglalja a meta-analízis során felhasznált anyagokat és módszereket.

1. Táblázat. A meta-analízis során felhasznált anyagok és módszerek.

Anyag/módszer	Részletes ismertetés/cél
Proteomikai adatkészlet	Az NCI-60 sejtvonalak felülűszójából izolált EV-k proteomját tartalmazza (6 701 fehérje). Az intenzitásokat logaritmizáltuk. Az NCI-60 sejtvonal panel kilenc tumortípust képvisel: emlő, CNS, vastagbél, vese, leukémia, tüdő, melanoma, petefészek, prosztatata.
Inváziós kapacitásra vonatkozó adatok	Magában foglalja az NCI-60 sejtvonalak CIM Plate-16 segítségével mért inváziós képességét.
Osztódási sebességre vonatkozó adatok	Tartalmazza az NCI-60 sejtvonalak osztódási sebességét.
RNS-expressziós adatok	Az NCI-60 sejtvonalak génexpressziós mikroarray adatai.
<i>In vivo</i> szöveti expressziós adatok	Egyes fehérjék mRNS-ének tumorszövetben mért expressziója és az ehhez kapcsolódó túlélési adatok.
Az EV minták klasszifikációja	Többváltozós logisztikus regressziót alkalmaztunk az egyes EV minták kilenc tumortípusba való besorolásához. A diszkriminatív fehérjék kiválasztása Least Absolute Shrinkage and Selection Operator (LASSO) módszerrel történt.
Az inváziós és proliferációs kapacitás predikciója	Többváltozós lineáris regressziót alkalmaztunk az inváziós és proliferációs kapacitás predikciójához. A releváns fehérjék szelektálása LASSO módszerrel történt.
Útvonal elemzés	A túlreprezentált biológiai folyamatok, molekuláris funkciók és sejtkomponensek azonosítása a proteomikai adatokban a Gene Ontology és a Reactome segítségével történt.
Hierarchikus klaszterizálás	A Reactome analízis eredményeit hierarchikus klaszterizációt végeztünk el, melynek eredményét heatmap-ként ábrázoltuk.
t-elosztott sztochasztikus szomszédbeágyazás (t-SNE)	Az EV-minták két-dimenziós térben való megjelenítéséhez t-SNE algoritmust alkalmaztunk.
Az EV-k proteomja és a sejtek RNS-mintázata közötti hasonlóság vizsgálata	Az EV minták proteomja és a sejtek RNS-profiljai közötti hasonlóságot Spearman-féle korrelációs elemzés és RV együtthatók segítségével.

3.2. A klinikai szérummintákból izolált extraelluláris vezikulák Raman-spektroszkópiai vizsgálata

A 2. táblázat összefoglalja a klinikai vizsgálat során használt anyagokat és módszereket.

2. Táblázat. A klinikai vizsgálat anyagai és módszerei.

Anyag/módszer	Részletes ismertetés/cél
Páciensek	138 szérummintát gyűjtöttünk glioblasztómás (GBM), agyi áttétes (BM) és gerincvelői porckorongsérvekben (CTRL) szenvedő betegektől, beleegyezéssel és etikai jóváhagyással.
sEV izoláció	A kisméretű extracelluláris vezikulákat (sEV) a szérummintákból differenciálcentrifugálással izoláltuk. A karakterizálás NanoSight NS300 és Western blot segítségével történt méreteloszlást, valamint klasszikus EV markereket vizsgálva. A jellemzési eredmények alapján sEV-vel dúsított izolátumoknak tekintettük őket, nem pedig tiszta izolátumoknak.
Raman spektroszkópia	A sEV-dúsított izolátumok Raman spektroszkópiai mérését Senterra II Microscope műszerrel végeztük.
Adatok transzformálása	A spektrumokat alapvonal korrekciónak vetettük alá. Ezt követően normalizáltunk Standard Normal Variate (SNV) módszerrel, majd leredukáltuk a spektrális adatsor dimenziót főkomponens elemzéssel (PCA).
Betegek klasszifikációja	A betegek osztályozását elvégző modellt Support Vector Machine (SVM) algoritmus segítségével hoztuk létre. Ennek hatékonyságát szenzitivitással, specificitással, klasszifikációs sikerrel és a Receiver Operating Characteristic (ROC) elemzésből származó görbe alatti terület (AUC) értékével értékeltük.
Releváns spektrális különbségek meghatározása	A csoportok megkülönböztetéséhez hozzájáruló spektrális különbségek feltárása a PCA és FreeViz módszerek kombinálásával.

4. ÚJ MEGÁLLAPÍTÁSOK

A következő fejezet a doktori értekezésben részletesen ismertetett új megállapításokat (keretes írás) és a kapcsolódó eredményeket foglalja össze.

1. Gépi tanulási módszerek eredményei alapján az extracelluláris vezikulák molekuláris mintázata magas fokú tumorspecificitást mutat.

A megállapításhoz kötődő eredmények:

A meta-analízis során vizsgált proteomikai adathalmaz 6 071 fehérjét tartalmazott (Az NCI-60 sejtvonal panelbe tartozó 60 sejtvonal felülúszójából izolált EV minták proteomja). Ezek közül 5 908 fehérjét számszerűsítettünk intenzitás szerint, amelyek együttesen alkotják azt, amelyre a tanulmányunkban teljes proteomként hivatkoztunk. Az 5 908 fehérje közül 213 fehérjét az összes EV mintában azonosítottunk, melyre attól kezdve magproteomként hivatkoztunk. Kezdetben csupán a magproteom alapján osztályozási modelleket építettünk többváltozós logisztikus regresszió segítségével, mely átlagosan 49,14%-os hatékonyságot eredményezett. Ez a hatékonyság 69,10%-ra javult, a teljes proteom bevonásával. Ezt követően a LASSO szelekciós módszer alkalmazásával a teljes proteomból összeállítottunk egy diszkriminatív fehérjepanelt (172 fehérje), melyet felhasználva 91,67%-os osztályozási hatékonyságot értünk el. A kapott osztályozási hatékonyság kilenc tumortípus esetében jelentősen meghaladta a 11,11%-os véletlenszerű osztályozási hatékonyságot.

2. Az extracelluláris vezikulák proteomikai mintázata tumorspecifikus jelutakról árulkodhat.

A megállapításhoz kötődő eredmények:

A 172 protein kiválasztása után feltételeztük, hogy — tekintettel a nagyfokú csoportok közötti különbségekre — az általuk befolyásolt biológiai jelátviteli útvonalak is megkülönböztető mintázatot mutatnak. Annak érdekében, hogy a 172 kiválasztott fehérjét biológiai kontextusba helyezzük, Reactome útvonal elemzést végeztünk el. A Reactome eredményeit hierarchikus klaszterizáció alá vetettük, és heatmap-ként ábráztuk. Csak azokat az útvonalakat vettük számításba, melyek esetében a $p < 0,05$ -nek adódott.

A kiválasztott 172 fehérje az extracelluláris mátrixhoz, a nukleáris folyamatokhoz és a sejtosztódással kapcsolatos jelátviteli útvonalakhoz kapcsolódik.

Bár eredményeink szerint az emlő- és prosztaták nem rendelkeztek jellegzetes jelátviteli útvonalakkal, az EV-minták többsége a tumortípusuk szerint klasztereződött, ami karakterisztikus jelátviteli útvonal-mintázatot mutatott.

Például a kollagénmátrixhpcz, a TGF- β receptorhoz és az ERB4 enzimhez köthető jelátviteli útvonalakat közös jellemzőként azonosították mind a vese-, mind a központi idegrendszeri tumorok esetében. Más daganatokhoz képest a leukémiás mintákban a hiszton- és kromatinmódosítással kapcsolatos nukleáris folyamatok dominálnak. A tüdődaganatokat más tumortípusoktól a vérlemezkékkel kapcsolatos biológiai folyamatok és az integrin jelátviteli útvonalak különböztették meg.

3. Az extracelluláris vezikulák proteomjából összeállított panelek képesek prediktálni a sejtek inváziós képességét és proliferációs idejét.

A megállapításhoz kötődő eredmények:

Az NCI-60 sejtvonal panel változatos, különböző szöveti eredetű tumorokat tartalmaz, amelyek mindegyike eltérő inváziós képességgel és osztódási sebességgel rendelkezik. Ez arra ösztönzött, hogy megvizsgáljuk, vajon lehet-e olyan fehérjepaneleket létrehozni, melyek képesek az inváziós képesség és az osztódási sebesség prediktálására.

Az ilyen panelek összeállítása a LASSO szelekciós módszerrel kombinált többváltozós lineáris regressziót alkalmazásával történt. Az adathalmazt egyenlően két félre osztottuk Tanító és Teszt halmazra, és a Tanító halmazt használtuk a kezdeti elemzéshez. Ebben a halmazban a LASSO algoritmus azonosította azokat a fehérjéket, amelyek potenciálisan előrejelezhetik az inváziós kapacitást és a proliferációs sebességet. Az így kiválogatott fehérjék prediktív erejét ezt követően a Teszt halmazon igazoltuk.

A folyamat eredményeként a Tanító halmazban 20, az inváziós kapacitást prediktáló és 15, a proliferációs kapacitást előrehelző fehérjét azonosítottunk. A továbbiakban ezekre inváziós és proliferációs panelekként hivatkoztunk. Ezen panelek a prediktív értékét többváltozós lineáris regresszió segítségével értékeltük a Teszt halmazon.

A Teszt halmazon való igazolás mindkét panel esetében szignifikáns eredményeket hozott ($p < 0,0001$), magas determinációs együtthatókkal: $R^2 = 0,68$ az invázióra és $R^2 = 0,62$ a proliferációra.

4. A kisméretű vezikulákkal dúsított izolátumok Raman spektrumai alapján felépített gépi tanulási modellek nagy hatékonysággal képesek megkülönböztetni a tumoros betegeket a kontrollcsoporttól.

A megállapításhoz kötődő eredmények:

A klinikai szérummintákból izolált 138 sEV-dúsított minta Raman-spektroszkópiai elemzése során mintánként 5 spektrumot kaptunk. A $799,5\text{ cm}^{-1}$ és $3100,5\text{ cm}^{-1}$ közötti spektrális tartományt vizsgáltuk. Az SNV-normalizálást és a PCA-transzformációt követően a minták osztályozását SVM-algoritmus segítségével végeztük el. Az osztályozás hatékonyságát az klasszifikációs pontosság (CA), szenzitivitás, specificitás és a ROC-analízisből származó görbe alatti terület (AUC) értékével mértük fel. A releváns spektrális különbségeket a PCA-FreeViz módszerekkel tártuk el.

A spektrumok átlagolása után a sorok normalizálását az SNV módszerrel végeztük el. Az SNV-normalizálást követően a négy betegcsoport mintáinak spektrumait páronként összehasonlítottuk (minden betegcsoportot a kontrollhoz hasonlítottunk, és a BM vs. GBM-et hasonlítottuk össze) két célból: egyrészt egy osztályozási algoritmus kifejlesztése és tesztelése céljából, másrészt a releváns spektrális különbségek azonosítása céljából. A páros összehasonlításokra alkalmazott PCA csökkentette a többváltozós adatok dimenzióit az eredeti változók (hullámszámok) kisebb számú új változóvá, azaz főkomponensekké (PC-k) történő átalakításával.

A páros összehasonlításokat a lineáris SVM algoritmus segítségével végeztük el, és minden egyes páros csoportra osztályozási modelleket kaptunk. A tesztmintákra vonatkozó előrejelzések elkészítéséhez meghatároztuk a csoporthoz tartozás pontszámának minimális küszöbértékét. Az e küszöbértéket meghaladó pontszámmal rendelkező vizsgálati mintákat a kívánt célcsoportba soroltuk. Az optimális pontszámküszöböt automatikusan úgy állítottuk be, hogy azok megfeleljenek a legmagasabb osztályozási pontosságnak (CA, a helyesen osztályozott minták aránya az összes mintához viszonyítva).

A CA 85,6% volt a CTRL vs. GBM, 91,4% a CTRL vs. BM, 82,9% a CTRL vs. M és 92,5% a BM vs. GBM esetében. A legjobb osztályozási teljesítményt akkor értük el, ha bizonyos számú PC-t bevontunk a modellekbe: CTRL vs. GBM esetében 30 PC, CTRL vs. BM esetében 38 PC, CTRL vs. M esetében 27 PC, BM vs. GBM esetében 26 PC.

Az osztályozási teljesítmény további mérőszámaiként az érzékenységet és a specificitást értékelték. A páros osztályozási modellek ROC-elemzéseit négy grafikont eredményeztek,

amelyek az automatikusan beállított optimális küszöbértékeket (amelyek a legmagasabb CA-értékkel rendelkeznek) mutatják, a kapcsolódó érzékenységi, specificitási és AUC-értékekkel, valamint a p értékkel.

Az optimális küszöbértékek használatával az osztályozási modellek 90%-os és 80%-os érzékenységgel és specificitással, 93,75%-os és 90%-os, 80%-os és 85%-os érzékenységgel és specificitással tudták megkülönböztetni a GBM, BM és M betegeket a CTRL betegektől. A két rosszindulatú daganatot, a BM-et és a GBM-et 98%-os érzékenységgel és 83,3%-os specificitással lehetett megkülönböztetni egymástól. A páronkénti összehasonlítások azonos sorrendjében (GBM, BM és M betegek vs. CTRL, illetve BM vs. GBM) az AUC értékek 0,87, 0,95, 0,82 és 0,9 voltak ($p < 0,0001$ minden esetben).

5. MEGBESZÉLÉS

A következőkben a vita összefoglalja az alapul szolgáló meta-analízis és klinikai vizsgálatok eredményeit, és következtetéseket von le a gépi tanulás előnyeiről mindkét kutatásban, az EV-k lehetséges szerepéről a tumordiagnosztikában, a differenciáldiagnosztikában, a rák jellemzőinek jobb megértésében. A tézisfüzet a Diskusszióknak csak főbb gondolatmeneteit tartalmazza.

5.1. Előrejelzési és osztályozási problémákra szabott gépi tanulási módszerek

Munkánk során gépi tanulási algoritmusok kombinációját használtuk fel a biológiai "big data" komplex összefüggéseiben való eligazodáshoz. A meta-analízis kulcselemeként lineáris és logisztikus regressziót alkalmaztunk LASSO algoritmussal párosítva, a klinikai tanulmányban pedig PCA-SVM algoritmust. A LASSO azon képessége, hogy a kevésbé fontos változókat eliminálja, felbecsülhetetlen értékűnek bizonyult a proteomikai adathalmazokból származó releváns információk kiszűrésében, megkönnyítve ezzel az értelmezhető és általánosítható modellek létrehozását.

A klinikai tanulmányban alkalmazott SVM módszer stabil elméleti alapjai és robusztussága lehetővé tette, hogy az egyébként kiugró értékekben gazdag spektrális adatokat is megfelelő módon kezeljük a modell felépítésekor; még abban az esetben is, ha a változók száma jelentősen meghaladta a megfigyelési egységek (nálunk a páciensek) számát.

Gépi tanulási módszerekkel való megközelítésünk nemcsak az EV-k által hordozott tumorspecifikus fehérjemintázatokat azonosította, hanem prediktálhatóvá tette az inváziós képességet és a proliferációs időt, feltárta a tumor jellemző útvonalait, valamint lehetővé tette a CNS-tumoros betegek hatékony osztályozását a szérummintákból származó sEV-dúsított izolatumok Raman spektrumai alapján.

5.2. Az EV-k Raman-spektroszkópiája, mint ígéretes diagnosztikai megközelítés

Klinikai kutatásunk arra irányult, hogy a Raman-spektroszkópiát felhasználva feltárjuk az EV-k diagnosztikai potenciálját, különös tekintettel a CNS tumorokra. Szemben a korábbi kutatásokkal, amelyek elsősorban korlátozott számú biomarkerekre (csak fehérjékre, lipidekre vagy nukleinsavakra) összpontosítottak, módszerünk az sEV-vel gazdagított izolatumok teljes molekuláris profilozását tette lehetővé, ezzel egy átfogóbb diagnosztikai stratégiát kínálva. Gépi tanuláson alapuló osztályozási modelljeink — melyeket az ROC elemzési szabványok alapján "kiválónak" és "kimagaslónak" értékelhetünk — kivételes pontossággal különböztették meg a kontroll csoportot és a különböző CNS tumortípusokat. Különösen figyelemre méltó a

GBM klasszifikálása során elért eredmény, mely egy rendkívüli páciensek közötti variabilitást mutató, és emiatt diagnosztikai kihívást jelentő tumortípus.

Továbbá, tanulmányunk jelentős hiányosságot pótol a szakirodalomban, mivel úttörőként végeztük el a CNS tumorok klasszifikációját a szérummintákból származó sEV-dúsított izolátumok Raman-spektrumai alapján.

5.3. Az EV-k által hordozott molekuláris komplexitás magas fokú tumorspecificitása

Mind a klinikai minták Raman-spektroszkópiái, mind az *in vitro* adatok meta-analízise rávilágított az EV-k által hordozott molekuláris tartalom magas fokú tumorspecificitására. Korábbi tanulmányok már bizonyították, hogy az EV-k különböző módszerekkel való vizsgálata lehetővé teszi a tumoros minták megkülönböztetését a kontroll mintáktól, vagy egy adott tumortípust valamilyen tulajdonsága alapján (például kemoszenzitivitás) alcsoportokra való osztását.

Ezzel szemben meta-analízisünkben számos tumortípus egyidejű megkülönböztetését tűztük ki célul. Az analízis eredményei rávilágítottak, hogy az EV-k által hordozott proteom önmagában is kimagaslóan specifikus egyes tumortípusokra nézve, azonban megfelelő módszerekkel nagy klasszifikációs hatékonyság elérését lehetővé tevő diszkriminatív proteinpanelek állíthatók össze.

A specificitás magas fokát nemcsak az *in vitro* eredmények, hanem a klinikai vizsgálatok eredményei is bizonyítják. A Raman-spektroszkópián alapuló modellünk sikeresen megkülönböztette — nem csupán a kontroll csoportot a tumoros csoportoktól —, hanem a különböző tumortípusokat is egymástól. Ez a nagyfokú specificitás különösen fontossá válik, ha egy olyan EV-alapú diagnosztikai platform potenciális fejlesztését vesszük figyelembe, amelynek célja nem csak a tumor potenciális jelenlétének jelzése, hanem a specifikus eredetű szövet azonosítása is.

5.4. Az EV-k, mint ígéretes prognosztikai eszközök

Bár az EV-k diagnosztikai hatékonyságát számos kutatás igazolta, prognosztikai hatékonyságuk kevésbé feltárt.

Meta-analízisünk során olyan proteinpaneleket állítottunk össze, melyek összefüggésben állnak a tumorprogresszió és a metasztázis képződés kulcsfontosságú elemeivel: az inváziós kapacitással és a proliferációs idővel. Ezen fehérjék szöveti expressziója szignifikáns összefüggésben áll a páciensek túlélési idejével, és a hatás iránya megegyezőnek bizonyult a meta-analízis során elvégzett végzett predikciókkal.

Ezen eredmények alapján az EV-k molekuláris tartalmának vizsgálata nem csupán a daganatok diagnosztikájában, hanem a prognosztika területén is hordozhat potenciális lehetőségeket.

5.5. Az EV-k tükrözhetik a daganat jellegzetes jelátviteli útvonalait

A tumorok molekuláris jellemzése alapvető fontosságú a hatékony kezelési stratégiák kidolgozásához; például segít a gyógyszercélpontok azonosításában vagy a személyre szabott kezelési stratégiák kifejlesztésében. Egy tumor molekuláris karakterizálása konvencionálisan szöveti biopsziát igényel, mely korlátokban ütközhet az invazivitása miatt, vagy a tumorszövet heterogenitása miatt. Ezen problémák áthidalásának reményében egyre nagyobb érdeklődés mutatkozik a folyadék-biopszia alapú módszerek iránt.

Újjonnan megjelenő kutatások bizonyították, hogy a tumorok által kibocsátott EV-k specifikus jelátviteli és metabolikus útvonal tagokat hordozhatnak, és az EV-k molekuláris tartalmának elemzése betekintést nyújt a tumor és a mikrokörnyezete közötti kölcsönhatásba, kommunikációba.

Meta-analízisünk eredményei alapján a kilenc tumortípus közötti legnagyobb különbségeket mutató fehérjék jelentősen összefüggnek bizonyos molekuláris mechanizmusokkal és jelátviteli útvonalakkal. Sok esetben az irodalom összhangban volt eredményeinkkel, mivel egyes tumortípusok és a kérdéses jelátviteli utak közötti kapcsolatot már kísérletileg is igazolták.

Eredményeink arra utalhatnak, hogy az EV-k tükrözhetik a tumorok egyedi stratégiáit és jellemző jelútjait, ezzel segítve a gyógyszercélpontok azonosítását, a személyre szabott orvoslás fejlesztését és a tumorbiológia mélyebb megértését.

6. KONKLÚZIÓ

Mint ahogyan előttünk számos tanulmány, a mi eredményeink is azt hangsúlyozzák, hogy az EV-k a biomarkerek kimagaslóan ígéretes forrásai. Munkánk egy olyan Raman-spektroszkópiai megközelítést mutat be, amely gyorsasága, pontossága miatt új, rutinszerűen alkalmazható szűrőeszközként szolgálhatna. Az EV-k Raman-spektroszkópiáján alapuló új diagnosztikai eljárás létrehozása azonban további standardizálást, nemzetközi adatbázis létrehozását igényli. Az EV proteom tumorspecifitása és prognosztikai értéke hangsúlyozza az EV-k potenciális szerepét az eddig feltáratlan klinikai alkalmazásokban. Az eredményeinket összegezve: EV-k a jövő klinikai gyakorlatának nélkülözhetetlen szereplőinek ígérkeznek.

7. KÖSZÖNETNYILVÁNÍTÁS

Őszinte hálámot szeretném kifejezni témavezetőmnek, **Buzás Krisztinának**, a tudományos iránymutatásáért, őszinte bizalmáért, és hogy lehetséges biztosított a csoportjában való munkára.

Köszönet illeti **Horváth Pétert**, a Biokémiai Intézet igazgatóján és a Mikroszkópos Képfeldolgozó és Gépi Tanulás Csoport vezetőjét a tudományos és infrastrukturális támogatásáért. Hasonlóképpen szeretnék köszönetet mondani **Székely Mártának**, az Interdiszciplináris Orvostudományi Doktori Iskola vezetőjének doktori tanulmányaim támogatását.

Szívből jövő köszönetemet fejezem ki kutatócsoportunk minden tagjának — **Greskovics-Dobra Gabriellának**, **Sebestyén Edinának**, **Harmati Máriának**, **Pintér Lillának** és **Böröczkynek Tímeának** — a szakmai és magánéleti támogatásukért, valamint a kivételes munkakörnyezet megteremtéséért.

Hálás vagyok minden velünk együttműködő kutatónak a gyümölcsöző kooperációért. Külön köszönet illeti **Klekner Álmost** a klinikai minták biztosításáért és szakmai éleslátásáért. Hálásak vagyunk **Kónya Zoltánnak** a Raman mikroszkóp biztosításáért és **Juan Gomez-Pereznek** a mérések elvégzéséért.

Köszönetet mondok a Nemzeti Kutatási, Fejlesztési és Innovációs Alapnak, az ÚNKP-20-3-Új Nemzeti Kiválóság Program és az EFOP 3.6.3-VEKOP-16-2017-00009 pályázat által nyújtott pénzügyi támogatásért.

Végül, de nem utolsósorban, szívből jövő köszönetet szeretnék mondani **barátaimnak**, akik jelenléte az egyetemi éveimet igazán örömtelivé és emlékeztetessé tette. Mély hálával tartozom **Nagyszüleimnek**, valamint **Papp Tibornak** - édesanyám férjének -, testvéreimnek, **Katalinnak** és **Miklósnek**, valamint Miklós feleségének, **Eszternek**, szüntelen támogatásukért életem számos vonatkozásában.

Hálával gondolok **Tündére** és **Krisztiánra**, Lehoczki Márk szüleire a támogatásukért és a szívből jövő fogadtatásért. Ezt követően szívből köszönöm magának **Márknak** a rendíthetetlen bizalmat és szeretet, amely mindvégig tartotta bennem a lelket.

A legnagyobb és legőszintébb köszönet **Édesanyámat** illeti meg, akinek rendíthetetlen kitartása és önzetlen szeretete mindezt lehetővé tette.