

Szolid tumorok terápia szempontjából jelentős klinikai alcsoportokba való sorolása génexpressziós adatok alapján

Doktori értekezés

Dr. Sztupinszki Zsófia

Semmelweis Egyetem

Patológia Tudományok Doktori Iskola



Témavezető: Dr. Győrffy Balázs, Ph.D, D.Sc., tudományos főmunkatárs

Hivatalos bírálók: Dr. Rónai Zsolt, Ph.D., egyetemi docens

Dr. Korcsmáros Tamás, Ph.D., csoportvezető

Szigorlati bizottság elnöke:

Dr. Zalatnai Attila, Ph.D., habilitált egyetemi docens

Szigorlati bizottság tagjai:

Dr. Szeltner Zoltán, Ph.D., tudományos főmunkatárs

Dr. Ambrus Attila, Ph.D., egyetemi adjunktus

Budapest

2017

Tartalomjegyzék

1	Rövidítések jegyzéke	4
2	Bevezetés.....	6
2.1	Szolid tumorok kezelése	6
2.1.1	Kemoterápia.....	8
2.1.2	Célzott terápia.....	8
2.2	Vastagbél-daganatok	11
2.2.1	A vastagbél-daganatok kialakulásának molekuláris alapjai	15
2.3	Emlőrák.....	20
2.3.1	Az emlő sebészete - nyirokcsomó áttét	23
2.3.2	Nyirokcsomó áttét kezelése	24
2.4	Daganatok osztályozása	29
2.4.1	Biomarkerek	29
2.4.2	Vastagbél daganatok.....	30
2.4.3	Emlődaganat	31
2.5	Bioinformatika, statisztika	34
2.5.1	Osztályozás, klaszteranalízis	34
2.5.2	Távolságszámítások.....	36
2.5.3	Szignifikáns különbségek keresése génexpressziós adatokon.....	37
2.5.4	Túlélés analízis	39
2.5.5	Többszörös összehasonlítások problémája.....	41
3	Célkitűzések	42
4	Módszerek	43
4.1	Gén-chip módszer	43
4.2	Bioinformatikai, statisztikai elemzés	44
4.2.1	Normalizálás.....	44
4.2.2	Minőségellenőrzés	45
4.2.3	A legjobb próba szettek kiválasztása.....	45
4.2.4	Batch hatás.....	46
4.2.5	Diagnosztikai próbák mutatói, értékelés	46
4.3	Vastagbél-daganat – többgénes osztályozók.....	47
4.3.1	Adatbázis létrehozása	47
4.3.2	Adatok előkészítése	47
4.3.3	Korábbi osztályozó módszerek irodalmi azonosítása.....	48

4.3.4	Osztályozók összehasonlítása.....	48
4.3.5	Gének súlyának megállapítása.....	49
4.3.6	Sejtvonalak	49
4.4	Emlődaganat – rossz prognózisú betegek azonosítása.....	50
4.4.1	Mintagyűjtés	50
4.4.2	Adatbázis létrehozása	51
4.4.3	Adatok előkészítése	51
4.4.4	Betegek osztályozása	51
4.4.5	Optimalizálás	53
4.4.6	Összehasonlítás más osztályozókkal	53
4.4.7	Online felület	54
4.5	Emlődaganat - Nyirokcsomó érintettség előrejelzése.....	55
4.5.1	Mintagyűjtés	55
4.5.2	Adatbázis létrehozása	56
4.5.3	Adatok előkészítése	56
4.5.4	Betegek osztályozása	57
4.5.5	Útvonal elemzés	57
4.6	Biomarkerek reprodukálhatóságának összehasonlítása	58
5	Eredmények	59
5.1	Vastagbélrák - rossz prognózisú betegek azonosítása	59
5.1.1	Adatbázis	59
5.1.2	Algoritmusok azonosítása.....	62
5.1.3	Osztályozók összehasonlítása.....	71
5.1.4	Génlisták összehasonlítása	74
5.1.5	Prekilinikai modellek.....	75
5.2	Emlőrák – rossz prognózisú betegek azonosítása.....	77
5.2.1	Mintagyűjtés – Független validációs adat	77
5.2.2	Adatbázis	77
5.2.3	Optimalizáció.....	79
5.2.4	Összehasonlítás.....	80
5.2.5	Validálás	86
5.3	Emlőrák - nyirokcsomó érintettség előrejelzése	89
5.3.1	Mintagyűjtés és adatbázisunk.....	89
5.3.2	Nyirokcsomó-státusz előrejelzése	91
5.3.3	Útvonal és gén-ontológia elemzés.....	93
5.4	Biomarkerek reprodukálhatóságának összehasonlítása	96

6	Megbeszélés	98
6.1	Vastagbélrák - rossz prognózisú betegek azonosítása	98
6.2	Emlőrák – rossz prognózisú betegek azonosítása	100
6.3	Emlőrák - nyirokcsomó érintettség előrejelzése	101
6.4	Biomarkerek reprodukálhatóságának összehasonlítása	102
7	Következtetések	104
8	Összefoglalás	106
9	Summary	107
10	Irodalomjegyzék	108
11	Saját publikációk jegyzék	138
11.1	Disszertációhoz kapcsolódó publikációk jegyzéke	138
11.2	Disszertációtól független publikációk jegyzéke.....	138
12	Köszönetnyilvánítás.....	140
13	Függelékek.....	141
13.1	Függelék 1.....	141
13.2	Függelék 2.....	141
13.3	Függelék 3. : Géncsendesítés hatékonysága gén-chip vizsgálatokban ..	144

1 Rövidítések jegyzéke

rövidítés	angol kifejezés	magyar kifejezés
5-FU	5-fluorouracil	5-fluorouracil
ATCC	american type culture collection	amerikai sejtbank
AUC	area under the curve	görbe alatti terület
CCLE	cancer cell line encyclopedia	rákos sejtvonal enciklopédia
cDNS	complementer DNA	komplementer DNS
DNS	deoxyribonucleic acid	dezoxiribonukleinsav
FDA	USA Food and Drug Administration	amerikai élelmiszer és gyógyszer ügynökség
FISH	fluorescent in situ hybridisation	fluoreszcens in situ hibridizáció
FDR	false discovery rate	fals találati arány
FN	fals negative	álnegatív
FP	fals positive	álpozitív
GEO	gene expression omnibus	gén expressziós omnibusz
GSE	gene expression omnibus (geo) series	gene expression omnibus (geo) sorozat
HR	hazard ratio	kockázati arány
IHC	immunohistochemistry	immunhisztokémia
MAS5	microarray suite 5.0	microarray csomag 5.1
MMR	mismatch repair	nem illő párok javítása
mRNS	messenger RNA	hírvivő RNS
NCCN	national comprehensive cancer network	nemzeti rák hálózat
PAM	prediction analysis of microarrays	microarray predikciós elemzés
PFS	progression-free survival	progressziómentes túlélés
qRT-PCR	qualitative real-time PCR	kvalitatív valós idejű PCR
RNS	ribonucleic acid	ribonukleinsav
ROC	receiver operating characteristics	receiver operating characteristics

RP	Rank Products	Rank Products
SAM	significance analysis of microarray	microarray szignifikancia elemzés
SH2	src homology 2	src homológia 2
siRNS	small interfering RNA	kis interferáló RNS
TCGA	the cancer genome atlas	a rák genom atlasz

2 Bevezetés

Disszertációmban szolid tumorok, emlő és vastagbél daganat osztályozásával, magas kockázatú, illetve rossz prognózisú betegcsoportok azonosításával foglalkozom. A beteg adott terápiára való válaszána előrejelzésével a terápia hatásossága, költséghatékonysága maximalizálható és a beteget érő gyógyszerterhelés, illetve mellékhatások csökkenthetőek. Ma már tudjuk, hogy az emlő- és a vastagbél daganatok nem tekinthetők homogén betegségnek, hanem a különböző daganat altípusok más-más biológiai és klinikai tulajdonságokkal bírnak. Ezért ezek azonosítása és jellemzése a tumorgenezis mélyebb megértését, új gyógyszer-célpontok azonosítását és megfelelőbb terápiák kiválasztását teszi lehetővé. A prediktív és prognosztikus biomarkerek a jelenlegi terápiás lehetőségek kontextusában a racionális terápia-tervezés fontos elemei és elterjedésük várhatóan sokat finomít a kezelési protokollokon.

Dolgozatomban négy vizsgálat eredményeit szeretném részletesen feldolgozni: A négy téma módszereit és eredményeit külön-külön mutatom be a könnyebb követhetőség kedvéért. Elsőként a vastagbél-daganatok transzkripciós mintázat alapú osztályozásával foglalkozom [1]. A második részben emlődaganatban rossz prognózisú betegek azonosítását végzem génexpresszió alapján [2]. A harmadik téma emlődaganatban a nyirokcsomók érintettségének előrejelzése a primer tumor génexpressziója alapján. Dolgozatom utolsó részében géncsendesítéssel végzett kísérleteket hasonlítok össze [3].

2.1 Szolid tumorok kezelése

A szolid tumorok kezelésében kiemelt fontosságú a prevenció, és a korai felismerés. Napjainkban a tumoros betegek kezelése az esetek többségében empirikus alapokon nyugszik, ám egyre inkább előtérbe kerül a személyre szabott orvoslás szemlélete. A célzott terápiás eljárások széleskörű elterjedéséhez azonban még klinikai vizsgálatok, a gyógyszerek indikációinak bővítése, és további prediktív biomarkerek szükségesek.

A szolid tumoros megbetegedések kezelésében döntő fontosságú a sebészi kezelés, a daganat teljes vagy részleges [4] eltávolítása. A daganat teljes eltávolítása történhet kuratív vagy palliatív céllal.

A sugárterápia az ionizáló sugárzás osztódó sejtekre gyakorolt káros hatásán alapszik, és a daganatos betegeknek a fele részeseül kezelése során valamikor besugárzásban. A daganatok között jelentős különbségek vannak sugárérzékenységüket tekintve. Újabb eljárások az endoradioterápia (a besugárzást a tumorban specifikusan felhalmozódó sugárérzékenyítő kemoterapetikum beadása előzi meg) [5, 6] és a radioizotóp terápia. Utóbbi esetében a radioizotópot egy célmolekula ellenes antitesthez kapcsolják a hatás lokalizációja érdekében, ezt az eljárást a célzott terápiák részeként szokták tárgyalni.

A kemoterapetikumok citotoxikus hatásukat a sejtproliferáció aspecifikus gátlásával érik el, aspecifikus hatásuk miatt alkalmazásuk során súlyos mellékhatásokkal kell számolni. Azonban, fontos kiemelni, hogy a célzott terápiás kezelést több beteg szakítja meg a mellékhatások miatt, mint a kemoterápiás kezelést.

Az endokrin terápiát (hormonterápiát), a célzott terápia legrégebben ismert formáját emlő és prosztatata tumoros betegek esetében alkalmazzák [7, 8]. Ide tartoznak a szelektív ösztrogén receptor modulátorok (SERM, melyek a receptorhoz kötődve csökkentik annak aktivitását), a szelektív ösztrogén receptor degradálók (SERD, melyek csökkentik a receptor szintjét), és az aromatáz gátlók (AI, melyek az ösztrogén szintjét csökkentik). A hormonreceptor-pozitív emlődaganatok esetében használt gyógyszereket az 1. táblázat tartalmazza.

1. táblázat. Emlődaganatok kezelése során alkalmazott hormonterápiás készítmények

Hatóanyag	Típus	Engedélyezés éve (FDA)
tamoxifen	SERM	1973.
toremifen		1997.
fulvestrant	SERD	2002.
anastrozole	AI	1995.
letrozol		1997.
exemestan		2005.

A modern molekulárisan célzott terápiák lényege, hogy a tumor növekedését, egy a progresszióban és növekedésben szerepet játszó specifikus molekula gátlásán keresztül állítják meg.

2.1.1 Kemoterápia

A modern kemoterápia története a 20. század elejére nyúlik vissza, amikor Paul Ehrlich mikrobiológiai kutatások során megalkotta a kemoterápia szót, mely ekkor olyan gyógyító eljárásokat jelölt, melyek kémiai anyagok a szervezetbe juttatásával mikroorganizmusok, vagy sejtek egy csoportjának a pusztulásához vezet [9]. A következő mérföldkő az a megfigyelés volt az első világháború alatt, hogy a vegyi fegyverként bevetett mustárgáz az áldozatokban jelentős fehérvérsejtszám-csökkenést okozott [10]. 1931-ben a mustárgáz antitumorális hatását az első modern értelemben vett klinikai kísérlet során igazolták. Ebben a klinikai vizsgálatban tizenkét beteg esetében alkalmazták a vegyület oldatát intratumorális injekcióban és az egy hónapos megfigyelési idő alatt minden esetben leírták a tumor regresszióját [11]. A kemoterápiás kezeléseknek mára számtalan formáját alkalmazzák a klinikai gyakorlatban, hatékonyságuk és az alacsony kezelési költségek miatt még mindig a gyógyszeres terápia alapjait alkotják. A kemoterápia főbb típusai: anti-metabolitok (pl.: 5-FU, methotrexat, merkaptopurin), alkiláló ágensek (cyclophosphamid, melphalan, carmustin), mikrotubulus gátlók (taxánok és vincristinek), topoizomeráz inhibitorok (doxorubicin, daunorubicin, irinotecan) és a citotoxikus antibiotikumok (doxorubicin, mitomicin). A klasszikus kemoterápia gyógyszerei elsősorban a gyorsan osztódó (daganat)sejtekre fejtik ki hatásukat, mellékhatás-profiljuk is ennek megfelelő: anémia, immunszuppresszió, hajhullás.

2.1.2 Célzott terápia

Bár a célzott terápia (“targeted cancer therapy”) kifejezésre 2000. és 2016. között több mint 23000 hivatkozás szerepel a Pubmed adatbázisban, mégsem létezik egységes definíció a célzott terápia mibenlétére.

Az első hivatalos meghatározást a US Food and Drug Administration (FDA) 2003-as évi ajánlása tartalmazta. Ezen ajánlás szerint a célzott terápiás szerek elrendelését olyan regisztrált diagnosztikus tesztnek kell megelőznie, mely bizonyítja a szer hatékonyságának kulcselemét: a célmolekula létét és hogy adott beteg esetén mindig a leghatékonyabb célzott terápiás szer kerüljön kiválasztásra [12]. Az FDA meghatározás hátrányaira mutatott rá Zhukov és munkatársai [13]. Az ő értelemzésük szerint a célzott terápiás gyógyszerekkel szemben a legfontosabb elvárások a következők.

1. Bizonyított hatás olyan célmolekulán, mely a tumor túlélése szempontjából kritikus szereppel bír, de nincs szerepe a normál sejtek túlélésében.
2. A gyógyszer hatékonyságának előrejelzése a célmolekula függvényében.
3. Az alapvető hatásmechanizmus nem lehet toxikus, és alacsony legyen a nemspecifikus toxicitás.

A célzott terápiával szembeni fenti kritériumok rámutatnak a hagyományos kemoterápia legfőbb hátrányaira: a predikciós tesztek hiányára, és arra, hogy a kezelés megválasztása sokszor empirikus úton történik. A jelenleg használt célzott terápiás készítmények még nem felelnek meg ennek a szigorú kritériumrendszernek, de a rezisztencia mechanizmusok megismerésével [14], új célmolekulák azonosításával egyre közelebb kerülünk a fenti feltételekhez, és a daganatok hatékonyabb gyógyításához.

A célzott terápiás gyógyszereket több szempont szerint is lehet csoportosítani. Szerkezet szerint lehetnek monoklonális antitestek (például: trastuzumab), vagy kis molekulájú gátlók (pl: lapatinib). A hatásmechanizmus alapján a szolid tumorok kezelésében alkalmazott célzott terápiás készítmények főbb csoportjai:

1. **Jelátviteli utakat gátló, az „onkogén addikciót” célzó készítmények:** Igen széles csoport, ide tartoznak például a tirozin-kináz gátlók. Az elsőként alkalmazott kismolekulájú tirozin-kináz inhibitor az imatinib volt, mely kompetitíven köt a CML-re jellemző t(9;22) reciprok kromoszóma transzlokáció eredményeként keletkező az onkogén **BCR-ABL** fúziós tirozin-kináz termékének (p210) az ATP-kötő doménjéhez [15]. Az emlődaganatok kezelésének fontos részét képezi a HER2-ellenes terápia (monoklonális antitest: trastuzumab, pertuzumab, kismolekulájú gátló: lapatinib). Ebbe a gyógyszercsoportba tartoznak továbbá például a PI3K jelátviteli útvonalban szereplő **mTOR** (mammalian target of rapamycin)-t gátló temsirolimus és everolimus, melyeket előrehaladott veserák és emlődaganatok kezelése során használnak [16].
2. **A tumor vérellátását akadályozó, befolyásoló gyógyszerek** közé soroljuk a vaszkuláris endoteliális növekedés faktort (VEGF) (bevacizumab, aflibercept), és a receptoraikat (VEGFR) gátló (ramucirumab, axitinib) hatóanyagokat. Alkalmazásuk széles körű (vastagbél, nem kissejtes tüdődaganatok, glioblasztoma multiforme [17]). A bevacizumab évekig (2007-2011) az amerikai ajánlások része volt paklitaxellel való kombinációban elsővonalbeli kezelésként áttétes emlőtumor

esetében, azonban nem bizonyítható hatékonysága miatt az FDA visszavonta az engedélyét ebben az indikációban [18]. A VEGFR-t gátló tulajdonsággal sok multikináz-inhibitor is bír (sorafenib, sunitinib, pazopanib, cediranib).

3. **Hibajavító útvonalak gátlása** az utóbbi évek kutatásainak köszönhetően kapott lendületet. A Poli (ADP–ribóz) polimeráz (PARP) a DNS hibajavításban résztvevő enzim. Daganatsejtekben a PARP gátlása a kettős DNS-szál törések felhalmozódásához vezethet, ami apoptózist indukálhat. Ezzel párhuzamosan beindulnak a DNS-hibajavítási útvonalak (többek között a CHK1, CHK2, BRCA1, BRCA2 fehérjék révén). Ezért hatékony lehet a PARP gátlása BRCA1 mutációval rendelkező daganatok esetében. Jelenleg három (olaparib [19], rucaparib [20], niraparib [21]) PARP- gátló használata engedélyezett BRCA1/2-mutáns petefészek daganatok kezelésében , valamint emlő és nem kissejtes tüdőrákban is biztató eredményei vannak a fázis III. vizsgálatoknak [22].
4. Az **antitest-gyógyszer konjugátumok** esetében a daganatsejteket célzó monoklonális antitesthez kapcsolnak citotoxikus gyógyszert, mely így lokálisan, nagyobb koncentrációban és kevesebb általános mellékhatással fejtheti ki hatását. Ilyen gyógyszer például az emlődaganatokban alkalmazott ado-trastuzumab emtansine, mely a HER2-receptorhoz kapcsolódó trastuzumab része révén helyben képes nagyobb koncentrációban feldúsulni [23].
5. Az **immunterápiát** egyesek a célzott terápiák közé sorolják. Az utóbbi pár évben a hagyományos immunterápia készítményeit (interferon- α , interleukin-2), felváltották az immunsejteken ható CTLA-4, PD-L1 gátlók. A PD-gátló pembrolizumab az első olyan gyógyszer, ami általános, a daganat lokalizációjától, eredetétől független engedélyt kapott [24], indikációja előrehaladott, mismatch-repair-hibával rendelkező szolid tumor.

A célzott terápias kezelések során a molekuláris célponttól függenek a jellemző mellékhatások is. Például az EGFR-gátlók: (cetuximab, erlotinib) bőr-, nyálkahártya tünetek [25], VEGFR-gátlók (sunitinib): magasvérnyomás, immunterápia [26], PD-L1 gátlók: kimerültség, autoimmun folyamatok [27].

Az emlő és a vastagbél-daganatok kezelése során kemoterápiás készítményekkel kombinációban alkalmazott gyógyszereket a 2. táblázat és a 3. táblázat tartalmazza. (A táblázatban az indikációnál csak az egyszerűsített alcsoport van feltüntetve.)

2. táblázat. Emlődaganatok esetében használt célzott gyógyszerek

Hatóanyag	Célpont	Indikáció (alcsoport)	Engedélyezés éve (FDA)
trastuzumab	HER2	HER2-pozitív	1998.
lapatinib			2007.
pertuzumab			2012.
ado-trastuzumab emtansine			2013.
palbociclib	CDK4/6	hormonreceptor-pozitív, HER2-negatív	2015.
ribociclib			2017.
everolimus	mTOR		2012.

3. táblázat. Vastagbél-daganatok esetében használt célzott gyógyszerek

Hatóanyag	Célpont	Indikáció (alcsoport)	Engedélyezés éve (FDA)
cetuxcimab	EGFR	EGFR-expresszáló	2004.
panitumumab			2006.
bevacizumab	VEGF	-	2004.
ziv-aflibercept		-	2012.
regorafenib	VEGFR2	-	2012.
ramucirumab		-	2015.

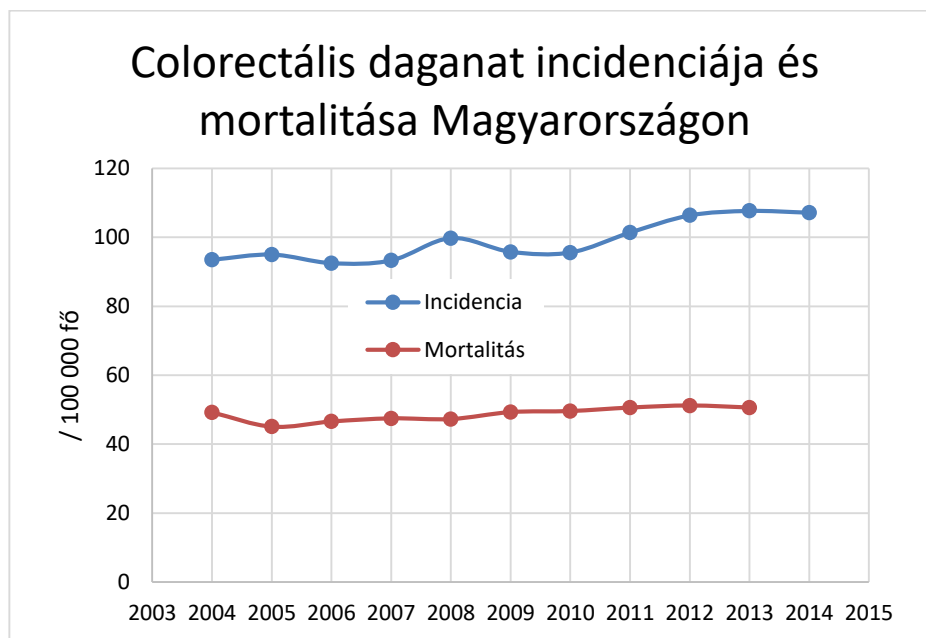
2.2 Vastagbél-daganatok

Vastagbél-daganatok a leggyakoribb gasztrointesztinális tumorok, szövettani megjelenésük az esetek 95 %-ában adenocarcinoma. Az 1950-es években az USA-ban még vezették a daganatos halálokok statisztikáját, mára a megbetegedések számában a 3. helyen állnak, míg a mortalitást tekintve a férfiaknál a második, nőknél a harmadik legtöbb elhalálozásért felelősek [28]. Az incidencia és a mortalitás csökkenése főleg a rizikófaktorok csökkenésének (csökkenő dohányzás, vörös hús fogyasztás), a korai felismerésnek (colonoscopos vizsgálat), és új terápiák elterjedésének köszönhető [29]. A vastagbél-daganatok incidenciáját és prognózisát erősen befolyásolja a szocioökonómiai

státusz; az alacsonyabb keresetű társadalmi csoportokban magasabb az előfordulás és rosszabb a prognózis [30].

A vastagbél-daganatok incidenciája erősen összefügg az ország gazdasági fejlettségével [31]. Az incidenciát és a mortalitást alapján az országok három csoportba sorolhatók. Az első csoportba elsősorban fejlődő országok tartoznak, ahol az incidenciát és a mortalitást is növekszik (például: Kína, Oroszország, Észtország, Spanyolország, Lengyelország), a második csoportban az incidenciát növekedése mellett csökken a mortalitást (például: Kanada, Dánia, Olaszország, Szlovákia), a harmadik csoportban az incidenciát és a mortalitást is csökken (például: USA, Ausztria, Franciaország) [32].

Az elmúlt 10 év hazai statisztikai adatait tekintve vastagbél-daganatok esetében az incidenciát enyhén emelkedett, a mortalitást pedig stagnálást mutat [33]. Az esetszám a Nemzeti Rákregiszter [34] adatbázisból származik, a mortalitást és a 100000 főre vetített arányszámot a Központi Statisztikai Hivatal népesedési adatai alapján határoztam meg. A rosszindulatú colorectális daganat közé a következő diagnózisok sorolhatók (BNO-kód, megnevezés): C18, a vastagbél rosszindulatú daganata; C19, a szigmabél-végbél határ rosszindulatú daganata; C20, a végbél rosszindulatú daganata; C21, a végbélnyílás és anus-csatorna rosszindulatú daganata. (4. táblázat, 1. ábra)

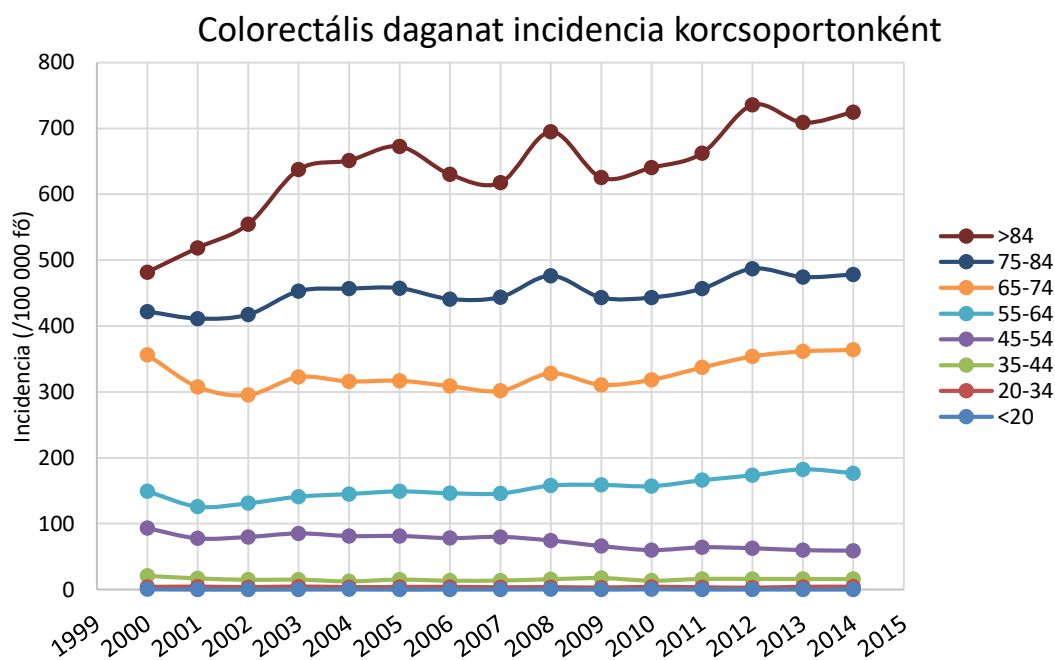


1. ábra. Vastag- és végbél-daganatok incidenciája és mortalitása Magyarországon 2004-2014.

4. táblázat. Vastag- és végbéldaganatok incidenciája és mortalitása Magyarországon 2004-2014.

Év	megbetegedés száma	halálozás száma	Népesség (ezer)	Incidencia /100 000 fő	Mortalitás /100 000 fő
2004	9463	4979	10117		49,2
2005	9598	4557	10098		45,1
2006	9320	4695	10077		46,6
2007	9390	4779	10066	93,4	47,5
2008	10029	4753	10045	99,6	47,3
2009	9610	4949	10031	95,1	49,3
2010	9569	4965	10014	95,3	49,6
2011	10126	5054	9986	106,9	50,6
2012	10569	5084	9932	106,6	51,2
2013	10672	5017	9909	107,6	50,6
2014	10589		9877	108,2	

A magyar vastag- és végbéldaganatok incidenciáját korcsoportonként tekintve a 74 évesnél idősebb korcsoportokban az incidencia növekedése, míg a 45-54 éves korcsoportban kismértékű csökkenés figyelhető meg, a 45 évesnél fiatalabbaknál pedig stagnálás látszik. (2. ábra) Ebben az esetben is fontos kiemelni az amerikai és nyugat-európai trendekhez képesti eltéréseket, hiszen ezekben az országokban az 50 évesnél fiatalabbaknál az incidencia drámai növekedését figyelték meg [35].



2. ábra. Vastag- és végbéldaganatok incidenciája és mortalitása Magyarországon korcsoportonként 2000-2014.

Bár a vastagbél daganatok ellátásában a legfontosabb a korai felismerés és a daganat sebészeti eltávolítása, az elmúlt évtizedben elsősorban a szisztémás terápia fejlődésének következtében jelentősen javult a betegek prognózisa. Míg az 1990-es évek közepéig csak az 5-fluorouracil (5-FU) volt törzskönyvezett, ma már egyéb citotoxikus gyógyszerek (például irinotecan, oxaliplatin) [36], fluoropyrimidinek (például: capecitabine, tegafur), monoklonális antitestek (például: bevacizumab, cetuximab, panitumumab) és érképződést gátló gyógyszerek (például: ziv-aflibercept, regorafenib) is elérhetőek. A kemo- és célzott terápiák ilyen bővülése a betegek pontosabb prognózis meghatározását és részletesebb osztályozását teszi szükségessé.

A vastagbél daganatok stádiumbeosztásánál az egyik első rendszer az 1932-ben publikált Dukes-féle klasszifikáció volt [37]. Ez az osztályozás a következő csoportokat határozta meg: Dukes A (a tumor a bélfalra lokalizálódik), Dukes B (a tumor betör a környező zsírszövetbe, de a nyirokcsomók intaktak), Dukes C (nyirokcsomó-áttétek), Dukes D (távoli áttétek). A Dukes-rendszert, és módosításait (például: Astler-Coller osztályozás[38]) még ma is gyakran alkalmazzák, azonban a használatuk egyre inkább kiszorul. Jelenleg a colorectalis daganatok stádium beosztása során a TNM klasszifikációt (TNM 7. kiadás) ajánlott alkalmazni [39]. A TNM rendszer elterjedéséhez több tényező vezetett: a TNM besorolás abban az esetben is rendelkezik korlátozott prognosztikai értékkel, ha a nyirokcsomók állapota (N) nem megítélhető, míg ebben az esetben a Dukes-féle beosztás nem értékelhető, a Dukes A tumorok közé a submucosába (T1), és a muscularis propriába betörő (T2) daganatok is tartoznak, így a TNM rendszer pontosabb prognózis meghatározást tesz lehetővé. A nyirokcsomók érintettségét is hatékonyabban osztályozza a TNM rendszer. A jelenlegi klinikai gyakorlatban a stádiumfüggő több változatú kombinált terápia alapja a beteg TNM alapú stádiumbesorolása (5. táblázat). Az 1-es stádiumú betegek csak műtétben részesülnek, a 2-es stádiumú betegek magas rizikójú csoportjában kiegészítő kemoterápiás kezelés is lehetséges, a 3-as stádiumú betegek a műtéti eltávolítás és kemoterápia mellett sugárterápiás kezelést is kaphatnak, míg a 4-es stádiumúak esetek célzott (pld. EGFR-gátló, érújdonképződést gátló) terápia alkalmazása is lehetséges.

5. táblázat. Vastagbél-daganatok stádiumbesorolása

Stádium	TNM	Leírás	Dukes
Stage 0	Tis N0 M0	in-situ cacinoma	-
Stage I	T1 N0 M0	a tumor submucosába terjed	A
Stage IIA	T2 N0 M0	a tumor a muscularis propriába terjed (bélfalra lokalizált)	A
Stage IIB	T3 N0 M0	a tumor a bélfal összes rétegét involválja, és a pericorectalis szövetekbe terjed	B
Stage IIC	T4 N0 M0	a tumor terjedése során a közvetlenül szomszédos szöveteken és szerveken túl is kiterjed	B
Stage IIIA	T1-2 N1 M0	1-3 regionális nyirokcsomó, és a tumor mérete T1-2	C
Stage IIIB	T3-4 N1 M0	1-3 regionális nyirokcsomó érintett, és a tumor mérete T3-4	C
Stage IIIC	T1-4 N2 M0	4 vagy több regionális érintett, tumor méretétől független	C
Stage IV	T1-4 N1-2 M1	távoli áttét van jelen, tumor méretétől, nyirokcsomók érintettségétől független	D

Amerikai adatok szerint a korai stádiumban 90%-os, nyirokcsomó áttétek esetén 71%-os, míg előrehaladott, távoli szervi áttéttel rendelkező vastagbélráknál 14% az 5 éves túlélési arány [40]. Az átlagos 5 éves túlélés 2009-ben 66,4% volt ami az 1975-ben tapasztalt 48,6%-hoz képest jelentős javulás, viszont a 2000-ben mért 64,8%-hoz képest gyakorlatilag stagnálást jelent. A 2005-ös hazai adatok szerint a férfiaknál 38%, a nőknél 39% az átlagos 5 éves túlélés [41].

2.2.1 A vastagbél-daganatok kialakulásának molekuláris alapjai

Genetikai instabilitásnak hívjuk azt a jelenséget, amikor nagy számban jelennek meg mutációk a genomban. Ezen mutációk közé tartozhatnak pontmutációk, inzerciók, deléciók, kromoszóma átrendeződések, számbeli kromoszóma-rendellenességek (aneuploiditás). A genetikai instabilitás nagyon fontos szerepet játszhat a tumor

evolúciójában, hiszen a nagyszámú mutáció nagyobb valószínűséggel vezethet rezisztens sejtcsoportok, klónok megjelenéséhez, ugyanakkor a túlzott mértékű instabilitás letális is lehet a daganatsejtek számára. A genetikai instabilitás hátterében többek között a DNS hibajavító folyamataiban szereplő génekben történt mutáció, epigenetikai változás állhat. Attól függően, hogy melyik hibajavító útvonal károsodik egyedi mutációs és expressziós mintázat jellemezheti a tumort. A vastagbél-daganatok esetében a mikroszatellita instabilitás és a kromoszóma instabilitás bír jelentőséggel.

Mikroszatellita instabilitás

A mikroszatelliták 1-6 bázispárnyi repetitív szekvenciák, amelyek 8-50 kópiában ismétlődhetnek. Mikroszatellita instabilitás (MSI) akkor áll fenn, ha a mikroszatellita ismétlődések száma a mismatch repair (MMR) DNS hibajavító rendszer következtében megváltozik. MSI a sporadikus daganatok között leggyakrabban a vastagbél daganataiban fordul elő, ahol az esetek 10-15%-ban mutatható ki [42]. A protokollok szerint a mikroszatellita státusz meghatározására 5 mikroszatellita lókuszt (a BAT25 és a BAT26 mononukleotid, valamint a D5S346, a D2S123 és a D17S250 dinukleotid lókuszt) vizsgálatát kell elvégezni [43]. Ezen vizsgálat alapján háromféle mikroszatellita státuszt különböztetünk meg. Alacsony szintű az instabilitás (MSI-L), ha a vizsgált mikroszatellita lókusztok kevesebb, mint 30%-a mutat instabilitást; magas az instabilitás (MSI-H), ha a vizsgált mikroszatellita lókusztok 30%-nál nagyobb arányban mutatnak instabilitást. A mikroszatellita-régiók stabilitása esetén a tumorokat mikroszatellita-stabilnak (MSS) tekintjük.

A MSI tumorok, főleg a jobb colon félre lokalizáltak, sokszor alacsonyan differenciáltak [44], mucinosus szövettani típusúak [45] és magas fokú intratumoralis lymphoid sejtes infiltrációt mutatnak [46]. Az MSI-H tumorok, jobb prognózisúak az MSI-L és az MSS daganatokhoz képest [47], ami valószínűleg kapcsolatban áll a mutációk nagyobb száma miatti fokozott immunogenitással [48]. Mivel az MSI-H tumorok kevésbé hajlamosak áttétképzésre, ezért arányuk magasabb a 2-es stádiumú (stage II-es), mint a 3-as stádiumú (stage III-as) daganatok között [49].

Az MSI státusz meghatározását abban az esetekben ajánlott elvégezni, ha 1, fennáll a Lynch-szindróma gyanúja: akinél 70 éves kor alatt diagnosztizálnak vastagbél tumort, illetve aki idősebb, de megfelel a Bethesda kritériumoknak [50], 2, stage II-es

beteg, mivel a stage II-es MSI-H betegek jó prognózisúak és nem ajánlott adjuváns 5-FU kezelést kapniuk [51], vagy 3, áttétes betegek.

Kromoszóma instabilitás

A kromoszóma instabilitás (Chromosomal INstability, CIN) esetében kromoszómáriszerveket vagy teljes kromoszómákat érintő számbeli eltérések (duplikációk, deléciók) figyelhetők meg. A CIN következményei az aneuploiditás és a heterozigótaság elvesztésének (loss of heterozygosity, LOH) számbeli emelkedése. A heterozigótaság elvesztése a tumorszupresszorok funkcióvesztésén keresztül a tumor progressziójához vezethet. A szolid tumorok többségét CIN jellemzi [52]. Sokat vizsgált, de egyelőre eldöntetlen kérdés, hogy a CIN a tumor kialakulásának korai eseménye-e, azaz CIN vezet-e a tumorszupresszorok inaktiválódásához, vagy késői esemény, és csak következménye a daganatban keletkező mutációknak [53, 54].

A DNS-metiláció zavarai

A DNS-metiláció a génexpresszió, normális sejtműködés fontos szabályozója, mely során a CpG dinukleotid citozinján a gyűrű 5' pozícióján lévő szénatomhoz metilcsoport kapcsolódik. A CpG dinukleotidok száma az exon és intron régiókban alacsony, a gének promóter régiójában pedig CpG-szigetek formájában halmozódnak fel. Ezen CpG-szigetek hipermetilációja csökkent transzkripció aktivitást eredményez. A daganatokat általános hipometiláció és a promóter régiók hipermetilációja jellemzi. Azokat a vastagbél daganatokat, ahol a CpG-szigetek hipermetilációt mutatnak „CpG island methylator phenotype” (CIMP+) típusú daganatoknak nevezik [55].

A CIMP+ fenotípuson belül, a metiláció mértéke alapján CIMP-magas és CIMP-alacsony daganatokat különböztetünk meg. A CIMP-magas fenotípusú vastagbél daganatokat a BRAF mutáció (70%), a CIMP-alacsonyakat a KRAS mutáció gyakori előfordulása (35%) jellemzi [56].

2.2.1.1 A colorectális adenoma-dysplasia-carcinoma szekvencia modell

A vastagbél adenomacarcinómája egy igen összetett betegség, mely leggyakrabban premalignus adenomából fejlődik ki. Epidemiológiai adatok szerint a lakosság 10-30%-ában található colonoscópos vizsgálattal adenoma a vastagbélben [57], melyből az esetek

5%-ában 10-15 év alatt rosszindulatú daganat fejlődik ki [58]. Ezen adenomatózus polipból invazív daganattá való átalakulás molekuláris lépéseit elsőként Vogelstein jellemezte [59]. Ma már tudjuk, hogy a Vogelstein által leírt „klasszikus” mechanizmus mellett más folyamatok is vezethetnek a colonkarcinóma kialakulásához.

A daganatképződés korai lépése a Wnt-útvonalhoz tartozó APC génben (adenomatous polyposis gene) történő mutáció. A mutáns APC génterméke nem tudja megkötni a β -katenint, amely így - többek között - a ciklin-D1, a c-myc és a c-jun sejtproliferációért felelős gének, valamint az antiapoptotikus survivin gén fokozott expressziójához vezet. Az APC gén mutáció hatására enyhe dysplasiát mutató, tubuláris adenoma jöhet létre. Az APC gént először egy ritka öröklődő betegség, a familiáris adenomatózus polipózis (FAP) kapcsán írták le, melyet az APC csírvonalbeli mutációja okoz, és obligát premalignus állapotnak tekinthető, vagyis minden ilyen beteg 40 éves korára vastagbél-daganat alakul ki [60]. Az APC gén a vastagbél-daganatok 73%-ában mutáns (az esetek döntő többségében trunkáló mutációt szenved) [61].

A daganatképződés következő fontos lépése a genom globális hipometilációja, valamint a p16 tumorszupresszor gén hipermetilációja, és így expressziójának csökkenése [62]. Ezt követő lépés a KRAS onkogén mutációja, mely 1-2 cm-es tubulovillózus közepes dysplasztikus adenoma kialakulását segíti elő.

A következő esemény a 2 cm-nél nagyobb villosus, súlyosan dysplasztikus adenomává alakulás lépése, mely elsősorban a 18q21 kromoszóma régió allélvesztésével (loss of heterozygosity) függ össze, ami a DCC, SMAD2, SMAD4 tumorszupresszor gének inaktiválódásához vezethet.

A súlyosan dysplasztikus adenomából carcinoma kialakulását legtöbbször a TP53 tumorszupresszor gén mutációja és inaktivációja segíti elő. A TP53 mutáció a vastagbél-daganatok több mint 50%-ban figyelhető meg, rossz prognózissal és terápia rezisztenciával társul.

2.2.1.2 Mikroszatellita instabil tumorok kialakulása

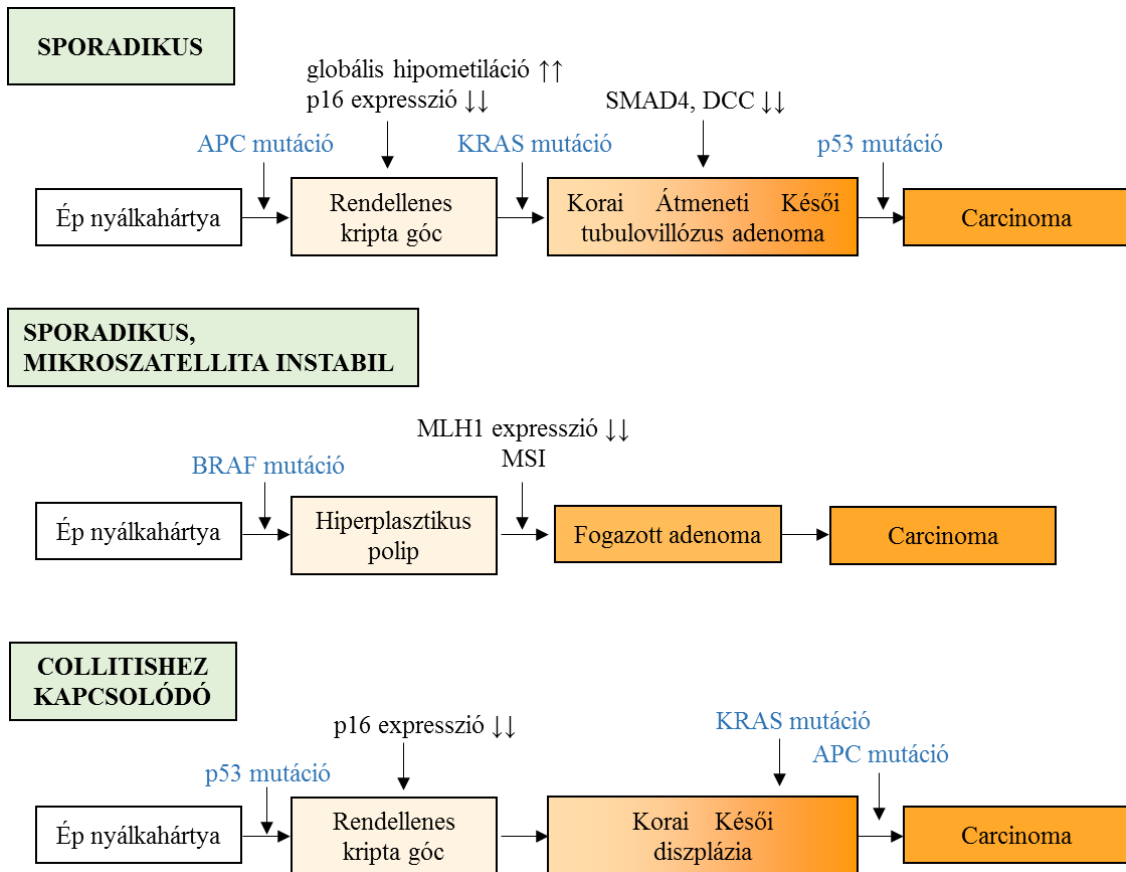
Az elmúlt húsz év kutatómunkája során Vogelstein modellje mellett a vastagbélrák kialakulásának más mechanizmusait, útvonalait is leírták [63]. A vastagbél-daganatok egyik csoportjában a DNS mismatch repair hibát szenved (elsősorban a MSH2, MLH1, PMS2 génekben történt mutációknak, hipermetilációnak

köszönhetően), mely következtében nagy számban fordul elő bennük mutáció és mikroszatellita instabilak. A nem örökletes szindrómához kapcsolódó, azaz sporadikus vastagbél daganatok 10-20%-a mikroszatellita instabil [64]. Ebbe a csoportba tartoznak a vastagbél daganatok 6%-áért felelős, Lynch-szindróma által okozott esetek (örökletes nem polyposisos bélrendszeri daganat, hereditary nonpolyposis colorectal cancer - HNPCC), mely betegeknél a bélrendszeren kívül is fokozott a daganatok előfordulásának kockázata (endometrium, petefészek carcinoma).

2.2.1.3 Gyulladásos bélbetegség talaján kialakuló tumorok

A gyulladásos bélbetegségek növelhetik a vastagbél daganatok kockázatát. A colitis ulcerosa-ban szenvedő betegek 5-10%-ában alakul ki colon adenocarcinoma, míg a Crohn betegség esetében nem fokozott a vastagbél daganatok kockázata. A kutatások szerint a gyulladás a reaktív oxigén gyökök felszaporodása, a megváltozott citokin termelés, a megváltozott micro-RNS mintázat és DNS-metiláció révén vezet a daganatok kialakulásának fokozott kockázatához [65, 66]. Ezekben az esetekben a carcinogenezis korai eseménye a TP53 mutációja [67], melyet gyakran a 6-os kromoszóma egyik alléljának elvesztése [68], és a CDH1 (E-kadherin) promóter hipermetilációja követ [69].

A vastagbél daganatok kialakulásának molekuláris modelljeit a 3. ábrán foglaltam össze.



3. ábra. Vastagbél-daganatok kialakulásának molekuláris háttere

2.3 Emlőrák

Az emlőrák a leggyakoribb rosszindulatú daganat, évente a világon több mint 1,7 millió nőt érint.

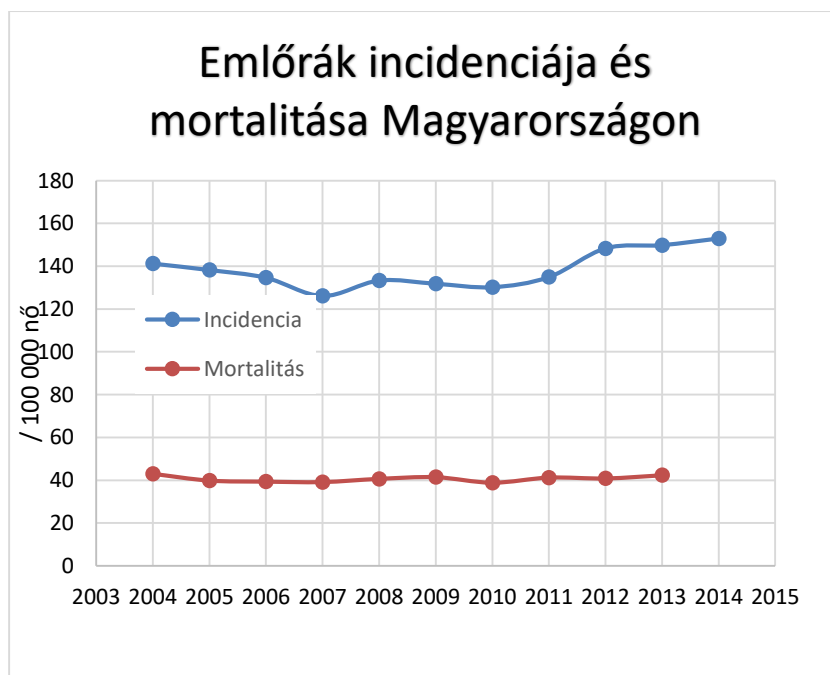
Amerikai adatokat tekintve az incidenciában az 1980-as évek végén a mammográfias szűrés elterjedésének következtében az incidencia gyors emelkedése volt megfigyelhető, mely elsősorban a daganatok 1-3 évvel korábbi diagnosztizálásának köszönhető. A 1990-es évek végén az emelkedés lassú mértékben folytatódott, melynek okai lehetnek a szűrés további terjedése, illetve a rizikófaktorok fokozott megjelenése (túlsúly, menopauzális hormonkészítmények [70]). Az incidencia ezt követően stabilizálódott, miközben a mortalitás lassan emelkedik [71].

Az első magyar szervezett lakosságsszűrés (Bonyhád, 1969) [72] csupán pár évvel követte az első mammográfiai szűrési programok kezdetét [73] a világon, mégis országos szintű szűrőprogram csak 30 évvel később indult. Fontos azonban

megjegyeznünk, hogy az utóbbi időben egyre több tanulmány kérdőjelezi meg ezen szűrőprogramok hasznosságát a felüldiagnózis problémája miatt [74]. (A felüldiagnózis olyan léziók azonosítása, melyekből a beteg élete alatt nem alakult volna ki orvosi beavatkozást igénylő daganat.)

Az emlőrákszűrés következtében a tumorok átlagos mérete, és ezzel összefüggésben a nyirokcsomó-áttétes esetek aránya [75], azon belül pedig a masszíván áttétes esetek aránya jelentősen csökkent [76]. Napjainkban a felfedezett emlőrákok túlnyomó többsége a korai emlőrák kategóriába tartozik, viszonylag kisméretű, és nincs vagy limitált a nyirokcsomó-érintettsége.

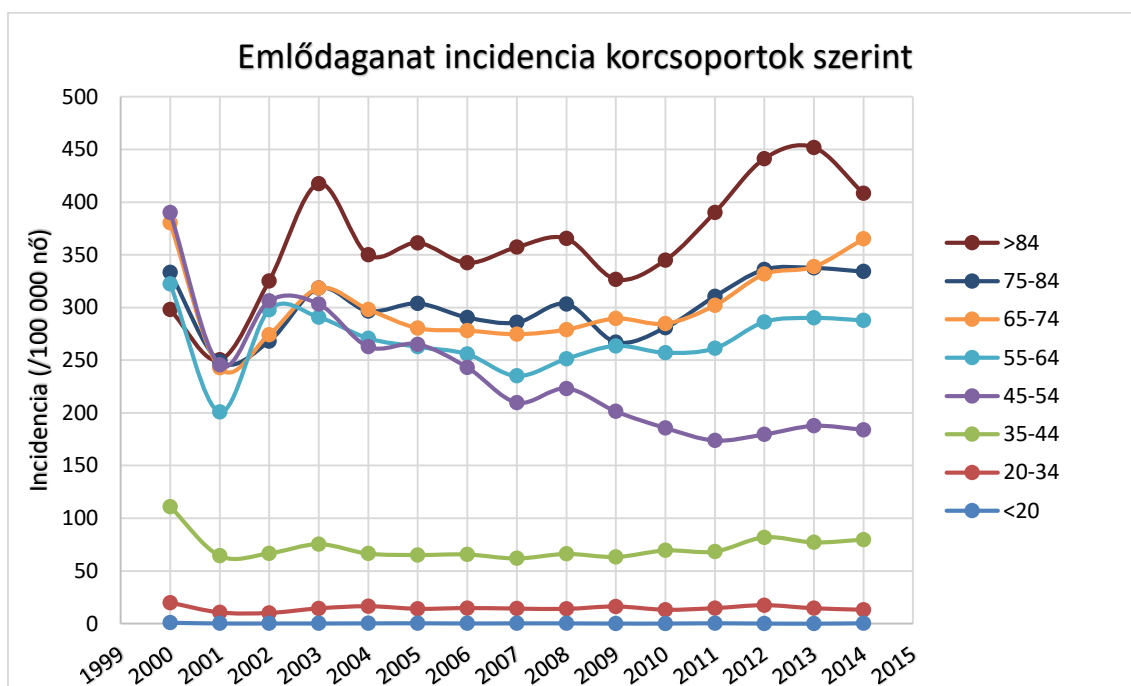
Az elmúlt 10 év hazai statisztikai adatait tekintve az emlődaganatok incidenciája 20%-al emelkedett, míg a mortalitás stagnálást mutat. [33] Ezen adatok pontosabb összehasonlítása érdekében meghatároztam a 100000 főre vetített arányszámot a Központi Statisztikai Hivatal népesedési adatait használva. (6. táblázat, 4. ábra)



4. ábra. Emlődaganatok incidenciája és mortalitása Magyarországon 2004-2014

6. táblázat. Emlődaganatok incidenciája és mortalitása

Év	megbetegedés száma	halálozás száma	Népesség (ezer)	Incidencia /100 000 fő	Mortalitás /100 000 fő
2004	7500	2285	10117		22,6
2005	7330	2109	10098		20,9
2006	7122	2081	10077		20,7
2007	6667	2066	10066	66,5	20,5
2008	7035	2141	10045	70,4	21,3
2009	6941	2183	10031	69,7	21,8
2010	6844	2040	10014	67,0	20,4
2011	7069	2159	9986	79,5	21,6
2012	7720	2123	9932	79,8	21,4
2013	7781	2195	9909	79,9	22,2
2014	7911		9877	81,5	



5. ábra. Emlődaganatok incidenciája és mortalitása Magyarországon korcsoportonként 2000-2014

Hazánkban 1995-ben nyitottak meg az első szűrőközpontok és 2001 őszén indult az a rákszűrő program, melynek keretében a 45 és 65 év közötti nők névre szóló értesítést kapnak a rákszűrés esedékességéről. Ezért ezt követően a szűrt korosztályban az incidencia emelkedését várnánk. Ez azonban nem igen figyelhető meg, melynek oka lehet az alacsony átszűrési arány (5. ábra). A 45-54 éves korosztályban történt csökkenés

hasonló az USA-ban 2002 után tapasztalt incidencia csökkenéshez, mely okának egy klinikai vizsgálat eredményei alapján a menopauzális hormonpótlás visszaszorulását tekinthetjük [77].

Amerikai adatok szerint a korai stádiumban 99%-os, nyirokcsomó áttétek esetén 85%-os, míg előrehaladott, távoli szervi áttéttel rendelkező emlődaganatnál 27% az 5 éves túlélési arány [40]. Az átlagos 5 éves túlélés 2013-ban 89,7% volt ami az 1975-ben tapasztalt 75,2%-hoz képest jelentős fejlődésnek tekinthető, azonban a 2000-es évektől kezdve nem tapasztalható további javulás. A 2005-ös hazai adatok szerint 70% az átlagos 5 éves túlélés [41].

2.3.1 Az emlő sebészete - nyirokcsomó áttét

Az emlő és a daganatok sebészete az orvostudomány és a sebészet történetének fontos mérföldköveit tartalmazza. Az első írásos feljegyzések az emlő sebészetéről már az egyik legrégebbi orvosi írásokban, a Kr.e 1600 körül íródott Edwin-Smith papiruszban megtalálhatóak, melyben az ismeretlen szerző 8 emlőrákos eset bemutatása mellett egy rákos elváltozás cauterisatioját írja le [78]. Már az ókori görög Hippocrates felismerte, hogy a kisebb méretű elváltozások jobb prognózissal társulnak és a kisebb méretű tumoroknak nem javasolta bolygatását. Az első században a római orvosok a tumorok radikális eltávolítását végezték, mely sokszor a pectorális izmokat is érintette. Bár a második században élt Galenus az emlőrákot a „fekete epe” túlsúlyának tulajdonította, de ő volt az első, aki az épen történő vágás jelentőségét leírta [79].

A középkor után Ambroise Pare' (1510–1590) a nagyméretű fekélyes daganatokat kenőcsökkel, a kisebbeket excisioval kezelte, szem előtt tartva az ép határokat. Pare' az elsők közé tartozott, aki felismerte a hónalji nyirokcsomók szerepét.

Az emlősebészet egyik jelentős korszakát (1900-1980) a Halsted elmélet jellemzi. Halsted szerint az emlőrák homogén betegség, terjedése meghatározott sorrendet követ: először lokális majd környéki nyirokcsomók érintettsége révén regionális, majd távoli szervekbe adva áttétet szisztémás betegséggé válik. Ebben a rendszerben a regionális nyirokcsomónak szűrő funkciót tulajdonítottak, a lymphogén áttétképzést tekintették a primer, a haematogént a szekunder folyamatnak, így a lokoregionális terápia határozza meg a prognózist [80]. Ezért a Halsted-féle radikális maszpektómia (1894) során az emlő eltávolítása mellett a környező mellkasi izmokat és a környéki hónalji nyirokcsomókat is

eltávolították. Az ultraradikális masztektómia során ezen felül a supraclavicularis nyirokcsomókat is eltávolították, illetve a 4–5 porcos bordaresectióval a parasternalis nyirokcsomókat is. Azonban ez a fokozott radikalitás nem növelte a túlélést, ugyanakkor növelte a szövődmények gyakoriságát és súlyosságát [81].

Halsted hipotézisét Fisher elmélete követte az 1980-as években. Fisher felismerte, hogy az emlőrák heterogén betegségcsoport, a hematogén metasztatizálás a primer és a lymphogén a szekunder folyamat, és a regionális nyirokcsomók szűrő funkcióval nem bírnak. A prognózist több tényező határozza meg: a szervezet immunstátusza, a disszemináció mértéke, a szövettan, a malignitás foka, a lokoregionális és szisztémás terápia. Fisher elmélete szerint a daganat lokális kezelése eltávolítása csak a helyi tünetek mérséklését szolgálják. Ez a szisztémás elmélet jelentős változást hozott a kezelésben, mivel csökkentett radikalitású és szervmegtartó műtétek kerültek posztoperatív sugárkezeléssel kombinálva az előtérbe [82].

Hellman modellje (1994) szerint az emlőrák heterogén betegségcsoport, mely először lokális, utána a környéki nyirokcsomók érintése révén *regionális*, vagy távoli szervekbe szóródva szisztémás betegség. A távoli áttétképzésben a véráram szerepe a legjelentősebb, ritkán a disszemináció idejében a primer daganat occult is lehet. A regionális nyirokcsomók szűrő funkcióval bírnak, nagyon gyakran ezek az őrszemnyirokcsomók (sentinel node, SN) [83].

Az őrszemnyirokcsomók a tumort drenáló nyirokerek első állomásai, a lymphogén áttétek legvalószínűbb helyei, a régió többi nyirokcsomói csak ezeket követően, másodlagosan lesznek áttétesek. Az őrszemnyirokcsomók azonosítására jelenleg kettős jelölést használnak, mely során a műtét közben az emlő szövetébe fecskendezett kék festéket és radioaktív gamma-sugárzót alkalmaznak.

2.3.2 Nyirokcsomó áttét kezelése

A nyirokcsomó áttétek kezelésének lehetőségei közé tartoznak az őrszemnyirokcsomó(k), illetve az összes hónaljji nyirokcsomó eltávolítása (axillary block dissection, ABD). Az ABD műtét klinikai haszna mellett jelentős morbiditással is járhat, az esetek 15-25%-ában nyiroködéma, 50-75%-ában szenzoros neuropátia 25%-ban a váll mozgásának zavara tapasztalható [84]. Bár az őrszemnyirokcsomók eltávolítása

jelentősen kisebb beavatkozás, ezen műtét kockázatai sem elhanyagolhatóak: 7%-ban figyelhető meg nyiroködéma [85], és 30%-ban neuropátia [86].

A nyirokcsomók eltávolításához kapcsolódó klinikai vizsgálatok az elmúlt két évtizedben elsősorban két kérdéskört vizsgáltak. Egyrészt, hogy mely esetekben lehet elhagyni a hónalji blokkdisszekciót, másrészt, hogy vannak-e olyan esetek, ahol az őrszemnyirokcsomók vizsgálata, és/vagy eltávolítása is elhagyható.

2.3.2.1 Klinikai vizsgálatok

Az emlőrákszűrés terjedésének következtében a felfedezett emlődaganatok átlagos mérete csökkent, és a nyirokcsomók is kevésbé érintettek. Így az ABD szerepének is át kellett értékelődnie, hiszen bár a nyirokcsomók érintettsége az egyik legfontosabb prognosztikai tényező, csupán diagnosztikus céllal végezni nem lenne ajánlott. Ezért volt fontos kérdés az elmúlt 20 évben a hónalji blokkdisszekció terápiás hatásának vizsgálata.

2.3.2.1.1 Axilláris blokkdisszekció terápiás hatása

Az egyik első vizsgálat az axilláris blokkdisszekció elhagyásával kapcsolatban egy 1999-ben, 3000 klinikailag regionális áttétmentes (cN0, akiknél nem tapintható, és képkalkotóval sem igazolható nyirokcsomó áttét), szisztémás kezelésben nem részesült emlőrákos beteg adatain végzett metaanalízis. A vizsgálat eredményei szerint ezeknél, a kemoterápiában nem részesült betegeknél az elektív ABD 5,4%-os javulást eredményezett a túlélésben az ABD elhagyásával szemben [87]. Ezzel szemben tíz évvel később, egy több mint 1500 cN0-s beteget vizsgáló metaanalízis arra a következtetésre jutott, hogy az ABD-nek nincs terápiás hatása. Sem a túlélésben, sem a betegségmentes túlélésben, sem a távoli áttétek gyakoriságában, sem az ipsilaterális emlőrecidívák arányában nem mutatkozott szignifikáns különbség az ABD-ben részesülő és nem részesülő betegek között, csupán a regionális, hónalji recidívák aránya növekedett meg kismértékben [88]. Fontos azonban megjegyezni, hogy ebben a metaanalízisben a vizsgálat végpontja 5 éves túlélés volt, mely emlőráknál rendkívül rövid utánkövetés. Ezeknek az eredményeknek a tükrében azt mondhatjuk, hogy bár az ABD-nek lehet terápiás hatása, de ez a hatást úgy tűnik, hogy adjuváns kezelések révén is el lehet érni [89]. Mivel a korai diagnózisnak köszönhetően a betegek egyre nagyobb hányada cN0 illetve nyirokcsomóáttét-mentes, a jelenlegi ajánlások szerint elfogadott, hogy az

áttétmentes nyirokcsomók eltávolítása csak a morbiditást fokozhatja [90]. Ezért is fontos kérdés az őrszemnyirokcsomó-biopszia, a nyirokcsomó státusz előrejelzése.

Az őrszemnyirokcsomó vizsgálatának elmélete az emlő nyirokelvezetésének anatómiai felépítésén alapul. Az esetek döntő többségében egy vagy néhány nyirokcsomó közvetlen összeköttetésben áll a primer tumor helyével, így először ezekben jelennek meg daganatsejtek és innen szóródhatnak távolabbi nyirokcsomók felé. Az őrszemnyirokcsomók az áttétek legvalószínűbb helyei, ha áttétmentesek, akkor nagy valószínűséggel a nem szentinel nyirokcsomók is áttétmentesek. Úgynevezett skipping („átugró”) áttétek, amikor a szentinel nyirokcsomóban nem, de egyéb nyirokcsomókban találhatóak daganatsejtek, az esetek 3-4%-ában fordulnak elő [91]. Randomizált vizsgálatok támasztják alá, hogy az ABD-t elegendő azon betegekre korlátozni, akiknél áttét van az őrszemnyirokcsomóban [92].

Az őrszemnyirokcsomó pozitív betegek jelentős részében nincs további nyirokcsomó érintettség, így ezekben az esetekben fontos lehet olyan betegcsoportok meghatározása, akiknél a szentinel pozitivitás ellenére nem szükséges ABD-t végezni. Ezért is érdemes a nyirokcsomóáttéteket részletesebben jellemezni. Így került bevezetésre az izolált tumorsejtek és a mikrometasztázis fogalma. Az „izolált tumorsejtek” olyan minimális, 0,2 mm-t meg nem haladó tumoros érintettséget jelöl, amelyet a staging és a kezelés céljából nem kell metasztázisként kezelni (patológiai N0). A mikrometasztázis pedig olyan nyirokcsomóáttétet jelent, ahol a tumorsejtek 0,2 mm és 2 mm közötti csoportokban helyezkednek el. A klinikai vizsgálatok (ACOSOG Z0011 [93] , IBCSG 23-01 [94]) mind izolált tumorsejtek, mind mikrometasztázisok esetében igazolták a várakozást, hogy az ABD biztonsággal elhagyható.

Mivel az őrszemnyirokcsomó pozitivitás mellett a betegek csupán 30-40%-ában figyelhető meg további nyirokcsomó érintettsége, a vizsgálatok következő iránya az volt, hogy pozitív őrszemnyirokcsomó esetén mely betegcsoportban hagyható el az ABD. Az ACOSOG Z-0011 vizsgálat során szentinel-áttétes betegek egy szűk csoportját sorolták be véletlenszerűen ABD- vagy megfigyelés vizsgálati csoportba. A vizsgálatba bevont betegek esetében a T1-T2N0M0 emlőrákot szervmegtartó műtéttel, ép szélekkel távolították el, legfeljebb két őrszemnyirokcsomóban mutatkozott áttét, és a műtétet követően teljes emlőbesugárzás, illetve a betegek többsége szisztémás adjuváns kezelésben is részesült. Az ACOSOG Z-0011 vizsgálat szerint sem 5, sem 10 éves

utánkövetésnél nem volt szignifikáns különbség a lokoregionális recidívák arányában a két vizsgálati kar között, ezért ebben a jól definiált betegcsoportban az ABD biztonsággal elhagyható. Hasonló kérdést vizsgált az EORTC-10981-AMAROS klinikai vizsgálat is, mely szentinel pozitív, T1-T2 stádiumú betegeknél az ABD versus hónalji besugárzást követően az 5 éves lokoregionális recidíva arányát vetette össze. Eredményei szerint a recidívák aránya megegyezik a két vizsgálati karon, azonban a sugárterápia esetében kevesebb a mellékhatás [95]. Érdemes még áttekinteni a Magyarországon, az Országos Onkológia Intézetben folytatott OTOASOR (Optimal Treatment of the Axilla – Surgery or Radiotherapy) klinikai vizsgálatot, mely szintén szentinel-pozitív betegek további kezelését vizsgálta: ABD versus hónalji régió további műtét nélküli célzott besugárzása. Eredményei szerint az ABD a regionális sugárkezeléshez képest nem javítja az őrszemnyirokcsomó-pozitív betegek regionális kontrollját és a betegek teljes túlélési mutatóit [96]. Ezen vizsgálat eredményeit 8 éves utánkövetés is megerősíti [97].

2.3.2.1.2 Az őrszemnyirokcsomó-biopszia elhagyása

Az őrszemnyirokcsomó-biopszia elvégzése a IV. stádiumú emlőrák esetében nem szükséges, más betegcsoportokban az elhagyása azonban ma még merésznek tűnő ötlet. Ezt a kérdést vizsgálja egy jelenleg is zajló klinikai vizsgálat, a SOUND: Sentinel node vs Observation after axillary UltraSOUND, mely emlőmegtartó műtétben részesült korai emlőrák esetében hónalji ultrahangvizsgálattal nyirokcsomó-negatívnak talált betegek között a szentinel-nyirokcsomó biopszia és a megfigyelés közötti különbséget kutatja [98].

Az orvosi gyakorlatban a hónalj ultrahangos vizsgálatát követően az áttétre gyanús betegeknél vékonytű-aspirációs citológiával, hengerbiopsziával nyert minták mikroszkópos vizsgálatával erősítik meg a nyirokcsomók érintettségét. Amennyiben a nyirokcsomó pozitív, ABD ajánlott (kivéve az előbbieken részletezett esetek), míg negativitás esetén őrszemnyirokcsomó-biopszia történik.

Amíg nem sikerül azt a betegcsoportot hatékonyan definiálni, ahol az őrszemnyirokcsomó eltávolítása nem jár sem terápiás, sem diagnosztikus előnnyel, addig a szentinel-biopszia általános műtét marad.

2.3.2.2 *Nemzetközi és hazai ajánlások*

A nyirokcsomó áttétek kezelését tekintve jelenleg ellentmondásosak a nemzetközi ajánlások. Az amerikai NCCN (National Comprehensive Cancer Network) ajánlása szerint a kedvező prognózisú esetekben a hónalji nyirokcsomók vizsgálata elhagyható, illetve a klinikailag nyirokcsomó-negatív esetek jelentős részében (Stage I, II, IIIA betegek) a szentinel nyirokcsomó biopszia a választandó vizsgálati módszer. Mi több, azoknál a neoadjuváns kezelést kapott betegeknél, akik klinikailag nyirokcsomó-pozitívak és tumora T1 vagy T2, és 1 vagy 2 pozitív őrszem nyirokcsomót találtak illetve emlőmegtartó műtétben (pl. lymphectomy) és teljes emlő besugárzásban részesültek nem javallott további hónalji műtét. A neoadjuváns kezelés a pozitív nyirokcsomók számának csökkenését eredményezheti (down-staging), ezért erre a betegcsoportra más ajánlások érvényesek [99].

A 14. Szent Gallen-i konszenzus (2015) ajánlása szerint 1 vagy 2 makrometasztázissal rendelkező pozitív nyirokcsomó esetén a maszpektómiát követően amennyiben nem terveznek radioterápiás kezelést, hónalji disszekció szükséges. Ugyanakkor a bizottság véleménye nem volt egyhangú arra vonatkozóan, hogy a maszpektómiát radioterápiának kell-e követnie. A résztvevők többsége a hónalji műtét elhagyása mellett állt, ha a rezekciós műtétet követően standard tangenciális radioterápia tervezett, és mindenki az ABD elhagyása mellett döntene, ha a tangenciális mezők a hónalji részt is érintik [100].

Az ESMO [101] ajánlásai szerint a szentinel nyirokcsomóban levő makrometasztázis esetén a konvencionális ABD elvégzése ajánlott

Az emlőrák sebészi kezelésével kapcsolatosan hazai ajánlások kiemelkedően naprakészek, progresszívek. A magyar III. Emlőrák Konszenzus Konferencia (2016) [102] ajánlásában a legfontosabb változások a korábbiakhoz képest, hogy ABD-t izolált tumorsejt, illetve mikrometasztázis jelenléte esetén nem, makrometasztázis esetén csak szelektív esetekben kell végezni, illetve hogy a regionális sugárkezelés bizonyos esetekben az ABD alternatívája lehet.

2.4 Daganatok osztályozása

2.4.1 Biomarkerek

Az Amerikai Gyógyszerügyi Hatóság (FDA) szerint a biomarker egy objektíven, pontosan és reprodukálhatóan mérhető indikátora valamilyen egészségügyi állapotnak, biológiai folyamatnak vagy terápiás beavatkozásra adott válasznak [103]. Ezen elég tág definíció szerint bármilyen molekula, a genomban mérhető változások, genotípus biomarkernek tekinthető. A biomarkereknek többféle csoportosítása lehetséges, melyek közül az egyik legfontosabb a diagnosztikus, prognosztikus és prediktív biomarkerek elkülönítése.

2.4.1.1 Diagnosztikus biomarkerek

A diagnosztikus biomarkerek közé sorolhatjuk a korai felismerést, szűrést célzó markereket. Az első tumormarkerek a 170 éve myeloma multiplex-ben szenvedő betegek vizeletében leírt Bence-Jones fehérjét tekintjük. [104] A napjainkban leggyakrabban alkalmazott tumormarkerek (pl.: PSA, CA-125) vérből mért szintje elsősorban utánkövetésre, nem szűrésre alkalmas. A diagnosztikus és a prognosztikus markerek határterületét jelenthetik a tumorból származó szabad DNS, keringő tumorsejtek detektálása szérumból [105], mely kvantifikálható, illetve rezisztens klónok megjelenésének felfedezésére is alkalmas lehet a technikák fejlődését követően [106]. Előrehaladott emlődaganatban a keringő tumorsejtek számának prognosztikus ereje lehet, azonban jelenleg nem ajánlott a vizsgálatuk [107].

2.4.1.2 Prognosztikus biomarkerek

Az onkológiában a prognosztikus biomarkerek daganatprogresszióról, a kezeléstől független túlélési, illetve kiújulási esélyekről nyújtanak tájékoztatást. Ebből kifolyólag az onkológiai prognosztikus markerek nagy része a tumorigenezishez, proliferációhoz, dedifferenciálódáshoz, angiogenezishez, invázióhoz és metasztázisképzéshez kapcsolódik. A daganatterápiában használt klinikai végpontok a teljes túlélés (overall survival, OS), progressziómentes túlélés (progression free survival, PFS), betegség-mentes túlélés (disease free survival, DFS). A prognosztikus biomarkereknek fontos szerepe lehet a terápiatervezésben, hiszen a rossz prognózisú betegek radikálisabb kezelést igényelhetnek, míg a jó prognózisú betegek esetében a felesleges terápia elhagyásával csökkenthető a mellékhatások száma.

2.4.1.3 *Prediktív biomarkerek*

A prediktív biomarkerek alkalmasak olyan betegcsoportok azonosítására, melyek nagy valószínűséggel reagálnak egy adott terápiára. Ezért ezek a biomarkerek lehetnek a személyre szabott terápia alapkövei, a klinikai vizsgálatok tervezését is jelentősen befolyásolják. Az „esernyő” szervezésű klinikai vizsgálatok esetében a daganatokat molekuláris altípusokba sorolják, és minden alcsoport esetében specifikus terápiát alkalmaznak [108].

2.4.2 *Vastagbél daganatok*

A vastagbél daganat szűrésére nem létezik bizonyított, ajánlott szérumszint, azonban a diagnóziskori CEA szérumszint meghatározása igen fontos, mivel olyan erős, független prognosztikus marker, hogy az AJCC (American Joint Committee on Cancer, Amerikai Rákos Betegségeket Vizsgáló Albizottság) a stádiumbeosztásának a részévé tette 2015-től „C-stádium” néven [109].

A vastagbél daganatok alcsoportokba sorolására jelenleg nincs széles körben elterjedt klinikai gyakorlat. Korai stádiumú vastagbélrákban (azaz azokban az esetekben, amikor a daganatot teljes egészében eltávolították, és szomszédos szerv, nyirokcsomó nem érintett, illetve nincs jelen távoli áttét) rossz prognózissal függ össze a T4 stádium, a obstrukció vagy perforáció, perineurális invázió, alacsony differenciáltság (grade3/4) és a magas fokú mikroszatellita instabilitás. További prognosztikus marker jelöltek a kor, a limfocita infiltráció, KRAS mutáció, timidilát szintáz pozitivitás, p53 expresszió, heterozigotáság elvesztése a 18q régióban, SMAD4 kópiaszám csökkenése, és többgénes osztályozók, melyek szerepének igazolására további vizsgálatok szükségesek. Előrehaladott esetekben rossz prognózissal függ össze az 1-nél nagyobb ECOG teljesítmény státusz, 70 évesnél idősebb kor, emelkedett CEA, alkalikus foszfatáz, trombocitaszám, laktát dehidrogenáz szint, csökkent hemoglobin, szérumszint albumin szint, BRAF mutáció [110].

Mivel a 2-es stádiumú betegek 10-15%-ánál jár terápiás haszonnal további kemoterápiás kezelés, ezért a molekuláris osztályozók fő célja ezeknek a betegeknek az azonosítása. A vastagbél daganatok esetében az NCCN ajánlás jelenleg három (Oncotype DX, ColoPrint, ColDX) génextpresszió alapú tesztet tárgyal, melyek alkalmazását elsősorban stage II-es, III-as betegek prognózisának előrejelzésére ajánlja. Sajnos

prediktív erejük csekély, ezért az ESMO nem is javasolja használatukat. A többgénes tesztek, molekuláris osztályozók részletes jellemzését későbbi fejezetben végzem.

Áttétes vastagbél-daganatban az ASCO (American Society of Clinical Oncology) a KRAS és az NRAS 2-es, 3-as, 4-es exonjának vizsgálatát javasolja, és EGFR-gátló terápiában csak az ezekben a régiókban mutációt nem mutató betegek részesülhetnek, a BRAFp.V600 mutáció státusz és mikroszatellita instabilitás meghatározását szintén prognosztikus céllal javasolják [111].

Az utóbbi években egyre nagyobb teret nyer a keringő tumorsejtek, illetve szabad DNS szérumból való izolálása és vizsgálata. Ezen keringő tumorsejteknek a száma összefügg a túléléssel [112], illetve lehetőséget biztosít a betegek utánkövetésére, rezisztens szubklónok megjelenésének korai, szemi-invazív detektálására [113].

2.4.3 Emlődaganat

Az emlődaganatok molekuláris altípusait génexpresszió alapján először 2001-ben írták le [114]. Ezeknek az altípusoknak fontos prognosztikai, prediktív, terápiás jellemzői vannak, meghatározzák egy beteg kezelési stratégiáját. A patológiai gyakorlatban a hormon-, HER2-receptor státuszt immunhisztokémiával, a 13. Szent Gallen Konszenzus alapján határozzák meg [115]. A csoportok jellemzőit, egyszerűsített meghatározását a 7. táblázatban foglaltam össze. Az altípusok gyakorisága erősen függ a földrajzi helytől, rassztól [116], a magyar betegek körében a Luminális A típus a leggyakoribb (52-55%). Gén-chip-en mért génexpressziók alapján osztja be molekuláris szubtípusokba 50 gén expresszióját alkalmazva a PAM50 módszer [117].

7. táblázat. Emlődaganatok altípusai

Altípus	Markerek				Terápia	5 Éves túlélés
	ER	PR	HER2	Ki67		
Luminális A	+	+	-	-	endokrin	95%
Luminális B (Her2-negatív)	+	+/-	-	+	endokrin + kemoterápia	50%
Luminális B (Her2-pozitív)	+	+/-	+	+/-	endokrin + kemot.+ anti-Her2	
Her2-pozitív	-	-	+	+/-	kemoterápia+ anti-Her2	30%
Tripla negatív (Bazális)	-	-	-	+/-	kemoterápia	10%

Újjonnan felfedezett emlődaganatok vizsgálatakor a képalkotó és patológiai értékelés mellett a hormonreceptor- és HER2-státusz vizsgálata ajánlott. Sem a vérkép,

sem a májenzimek szintje nem függ össze tünetmentes betegekben távoli áttét jelenlétével [118]. A korai emlődaganatok követésére jelenleg semmilyen tumormarker nem ajánlott.

Az emlődaganatok esetében a legerősebb prognosztikai tényezők a tumorméret, a differenciáltság foka (grade), az életkor és a nyirokcsomó áttét jelenléte. A magas HER2 expresszió rossz prognózist indikáló prognosztikus biomarkernek tekinthető [119], ugyanakkor a HER2 ellenes monoklonális antitest, trastuzumab, majd később a kismolekulájú lapatinib klinikai bevezetésével a HER2 expresszió prediktív biomarkerré vált, és a HER2-pozitív betegek a hatékony terápiának köszönhetően ma már kedvezőbb prognózisú betegcsoportot alkotnak [120]. A HER2 prediktív az adjuváns antraciklin terápia esetében is [121].

Számos többgénes vizsgálatot fejlesztettek bizonyos betegcsoportok prognózisának előrejelzésére. Az egyik fontos kérdés, hogy a hormonpozitív emlődaganatok esetében az ötéves hormonterápiát követően mely betegeknél magas a kiújulás, késői áttétképzés valószínűsége, azaz kiknek lenne ajánlott tovább folytatni a kezelést. Bár a pre- és postmenopauzális nők között 5 éves endokrin terápiát követően 20-25% az 5-15 éven belüli kiújulás valószínűsége [122], azt már számos klinikai vizsgálat igazolta, hogy minden beteg esetében kiterjeszteni a endokrin terápiát több hátránnyal (például csonttritkulás esetek emelkedett száma) jár, mint előnnyel [123]. A betegeknek csupán 3-5%-ánál hasznos a terápia folytatása [124], többek között ezen betegek azonosítását célozzák meg a következőkben bemutatott tesztek.

Az emlődaganat alcsoportjainál említett PAM50 [117], további fejlesztést követően „Prosigna Breast Cancer Prognostic Gene Signature Assay” [125] néven került forgalomba, de „Risk of Recurrence score”-ként is ismert. Korai (stage I-IIIa), hormonpozitív emlődaganatok esetében használható. 46 gén expresszióját vizsgálva a távoli áttét megjelenésének valószínűségére ad becslést a kezelés megkezdésétől számított 5-10 éven belül, azaz az 5 éves hormonterápiát követően. A valószínűség alapján három rizikócsoportba sorolja a betegeket. A génexpressziót FFPE mintákból izolált RNS „NanoString nCounter Dx Analysis System” [126] platformon határozzák meg.

A „Breast Cancer Index” esetében 7 gén FFPE mintákból izolált RNS expresszióját RT-qPCR-rel határozzák meg [127]. Korai emlődaganatban, nyirokcsomó státusztól függetlenül jelentős prognosztikai erővel rendelkezik. Jó hatékonysággal

azonosítja az N0-N1 tumorok késői kiújulásra hajlamos 35-45%-át, akiknél a meghosszabbított endokrinterápia hasznos lehet.

A főleg Európában, Németországban használt EndoPredict [128] teszt 5 évig endokrin terápiaiban részesült, hormonreceptor pozitív, HER2-negatív betegek esetében ad becslést távoli áttét valószínűségére 5-10 éven belül. Az FFPE mintákból történő RNS izolálást követően a vizsgált 9 gén expresszióját RT-qPCR-rel határozzák meg [129]. Az EndoPredict hatékonyságát randomizált fázis III vizsgálatok során igazolták [130, 131].

Az Oncotype DX tesztje 21 gén expressziója alapján ad a betegeknek „recurrence score”-t egy nullától százig terjedő skálán. A score arányos annak valószínűségével, hogy a daganat tíz éven belül kiújul. A magas vagy alacsony rizikó ismerete segíthet a betegnek és a kezelőorvosnak eldönteni az adjuváns kemoterápia szükségességét. Ezt a tesztet értékelték a legtöbb klinikai vizsgálatban (pl. retrospektív: NSABP B-14, prospektív TAILORx vizsgálat [132]).

A 70 gént használó MammaPrint teszt a távoli áttét valószínűsége alapján jó és rossz prognózisú csoportokba sorolja a betegeket, eredményeit a retrospektív TRANSBIG és a prospektív MINDACT vizsgálat validálta [133].

A Genomic Grade Index (GGI) teszt 97 gén expressziója alapján a grade 2-es tumorokat jó (grade 1-hez hasonló) és rossz prognózisú (grade 3-hoz hasonló) csoportokba sorolja [134]. Szintén a grade 2-esek további csoportosítását szolgálja korábbi publikációnkban leírt 4 génen alapuló osztályozás [135].

A Szent Galleni Kongresszus és az NCCN ajánlása csak az Oncotype DX alkalmazását javasolja. Az ESMO a terápia döntés támogatására 5 módszert tart elfogadhatónak: Oncotype DX, EndoPredict, Prosigna (PAM50), MammaPrint, Genomic Grade Index.

2.5 Bioinformatika, statisztika

2.5.1 Osztályozás, klaszteranalízis

Az orvosi, bioinformatikai kérdések egyik fontos területe objektumok (pl. minták, betegek) csoportokba sorolása különböző tulajdonságaik hasonlósága alapján. A módszerek megértéséhez először néhány a gépi tanulás során alkalmazott alapfogalmat kell áttekinteni [136]. A felügyelt tanulás (supervised learning) esetében ismert, előre definiált csoportokba soroljuk a mintákat, így a modell tanítása során tudjuk, hogy a tanulóhalmazunk mintái melyik osztályba tartoznak. Ezzel szemben klaszteranalízis során nem ismerjük előre a csoportokat, ezért felügyelet nélküli tanulásnak (unsupervised learning) nevezzük. Az osztályozás, klasszifikáció fogalma az angol szaknyelvben több dolgot is kifejezhet. A legtöbb szerző az osztályozás alatt csak a minták nem felügyelt módszerekkel csoportokba való sorolását érti.

A klaszterelemzés, hasonló dolgok csoportjának a létrehozása a minket körülvevő világ megértésének és leírásának nélkülözhetetlen eleme. A tudományos megértésnek is fontos eszköze, régóta alkalmazzák az orvostudomány és a biológia (rendszer) területén is. A klaszteranalízis során a csoportosítás alapját különböző távolság- vagy hasonlóságmértékek képezik. A klaszterező eljárások két általános kategóriája az időigényes hierarchikus és a nemhierarchikus klaszterezés. A hierarchikus klaszterezésnél nem kell előre ismernünk a létrehozandó csoportok számát, míg a nem-hierarchikus eljárásnál már kiinduláskor meg kell azt adnunk.

A felügyelt módszerek közé tartozik a lineáris diszkriminancia-analízis, mely módszer kettő vagy több csoport közti különbséget független változók, jellemzők lineáris kombinációjával képes leírni. Alkalmazható lineáris osztályozóként, de legtöbbször más osztályozó lépések előtti dimenziócsökkentésben van szerepe [137].

A k-legközelebbi szomszéd módszert (kNN, k-nearest neighbours) a lusta klasszifikáló eljárások közé sorolják, mivel nem épít modellt [138]. Azon az elven alapszik, hogy a hasonló csoportba tartozó objektumok hasonló tulajdonságokkal rendelkeznek. A hasonlóság nagyságát távolságfüggvénnyel mérjük. Egy új minta osztályozásához n -dimenziós térben (ahol n a vizsgált jellemzők száma) megkeressük azt a k tanuló mintát, melyek a legközelebb vannak az új, ismeretlen mintához. A kNN alkalmazásának nehézségei, hogy számításigényes, zajos adat igen kedvezőtlenül

befolyásolja a hatékonyságát, ugyanakkor a jól megválasztott távolságmérést alkalmazva hatékony és egyszerű osztályozást lehet elérni.

A „support vector machine” módszer lényege, hogy egy n -dimenziós térben (ahol n a vizsgált tulajdonságok száma) keresi a csoportokat legjobban elválasztó hipersíkot. Azt tekinti a legjobb hipersíknak, amelyik esetében a legközelebbi elemek a hipersíktól vett távolsága a legnagyobb [139].

A döntési fák talán a legismertebb osztályozók, melyek bonyolult összefüggéseket egyszerű döntések sorozatára vezetnek vissza [140]. A döntési fa megjeleníthető egy fa alakú folyamatábraként, mely gyökeréből indulva egy ismeretlen minta osztályozásakor a belső csúcsok egy-egy jellemző (attribútum) értékének vizsgálatát jelentik, a csúcsok közötti élek ennek a vizsgálatnak az eredményei. Mind a bemeneti attribútumok, mind a kimeneti értékek lehetnek diszkrét vagy folytonosak. A döntési fa tanulása során az attribútumok kiválasztására szolgáló eljárás arra irányul, hogy minimalizáljuk az eredményül kapott fa mélységét. Az egyes szabályokat a gyökértől egy-egy levél felé vezető útvonal feltételeinek összeolvasásával egyszerűen nyerhetünk osztályozási szabályokat. A döntési fák előnye, hogy a lényeges változók a gyökér közelébe kerülnek, illetve előfordulhat az is, hogy egyes attribútumok nem befolyásolják a döntést. Így dimenziócsökkentésre is használhatóak. Hatékonyan alkalmazhatóak nagyszámú minta esetén is, azonban amennyiben sok attribútumot használnak, túlérzékenyek lehetnek kicsi változásokra is [141].

A neurális hálózat az egyik legbonyolultabb osztályozási módszer. Neurális hálózatnak nevezzük azt a párhuzamos, elosztott működésre képes információfeldolgozó szoftvert, amely azonos (vagy hasonló típusú), nagyszámú lokális feldolgozást végző műveleti elem, „neuron” többnyire rendezett topológiájú, nagymértékben összekapcsolt rendszeréből áll. A neurális hálózatnak rendelkeznie kell tanulási algoritmussal, és a megtanult információ felhasználását lehetővé tevő előhívási algoritmussal [142]. A neurális hálók továbbá hajlamosak túltanulásra (overtraining), amikor túlzott illeszkedés (overfitting) figyelhető meg. A túlzott illeszkedés esetében bár a tanuló halmaz mintáira jó eredményeket kapunk, de nem tud jól általánosítani már az osztályozó, ezért az új, a tanítás során nem látott, értékelő (teszt) halmaz esetében csökken az osztályozás hatékonysága. A túlilleszkedés veszélye akkor igen nagy ha a tanuló csoport mintáihoz képest magas az attribútumok száma. A döntési fák teljesítményének javítására több

módszer alkalmas. A véletlen erdő (random forest) [143] [144] osztályozó egyszerre több döntési fát épít. Minden fa építése egy rögzített mélységig történik. A véletlen erdő felépítése során az egyes fák egyes csúcsainál véletlenszerűen sorsolt attribútumokól választhatjuk csak ki a döntési attribútumot. Egy új minta osztályozásáról a fák többségi, esetleg súlyozott szavazata dönt. Az erdő hatékonysága elsősorban a generált fák számán (a több fa használata javítja az eredményt), illetve a minőségén múlik (a fák közötti korreláció növekedésével csökken a hatékonyság). A véletlen erdő eljárás gyors, nagyméretű adatokat is képes kezelni, becsléseket ad arra, hogy mely attribútumok fontosak. Hátránya például hogy nem ábrázolható, és nehezebben érthető mint a döntési fák. A döntési fák kombinálásának másik módja a gyorsítás, ami a tanuló halmaz mintáinak (tanulóesetek) eloszlásának adaptív módosítására szolgál. A súlyait minden egyes gyorsítási menet végén módosítjuk. A helytelenül osztályozott esetek súlyait növeljük, mely arra kényszeríti az osztályozót, hogy a következő iterációkban azokra az esetekre összpontosítson, amelyeket nehéz osztályozni. Viszont vigyázni kell, mert a gyorsítási módszer is érzékeny a túlillesztésre, mivel hajlamos a tévesen osztályozott tanulóesetekre fókuszálni.

Abban az esetben, amikor nem definiáltak előre a csoportok, azaz a klaszteranalízis során, a két leggyakrabban alkalmazott módszer a k-közép (k-means) osztályozás és a hierarchikus osztályozás. A k-közép esetében előre meghatározott k-számú csoport azonosítása általában véletlenszerűen felvett középpontokból kiindulva történik. Iteratív módon minden mintát a legközelebbi klaszterbe sorol, míg a csoportok adatai (középpont, átlag) is lépésenként változnak, amíg a besorolás meg nem állapodik. A hierarchikus klaszterezés esetében nem előre definiált a klaszterek száma, a csoportok között fölé- és alárendeltségi viszony is lehetséges. A hierarchikus klaszterezés eredményét gyakran ábrázolják dendrogram formájában.

2.5.2 Távolságszámítások

A csoportok közötti hasonlóság szubjektív fogalom, azonban a klaszterező eljárások megkívánják egy bizonyos távolságmérési módszer alkalmazását. Szükség van két megfigyelés hasonlóságát mérő mérték definiálására, melyre különféle távolságszámítási eljárások alkalmasak. Az *euklideszi-távolság* az egyik legismertebb és legerősebb távolságmérő eljárás, mely a Pitagorasz-tételen alapul: $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$. A Manhattan-távolság (abszolút eltérés, „háztömb” távolság)

$d(x, y) = \sum_{i=1}^n |x_i - y_i|$ magas dimenziójú kérdések esetében általában jobban alkalmazható, mivel az euklidészi számítás ebben az esetben hajlamos túlbecsülni a távolságot. A Manhattan-távolság súlyozott változata a Canberra-távolság: $d(x, y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|}$ [145]. A statisztikai számítások során még gyakran alkalmazzák a Mahalanobis-távolságot is $d(x, y) = \sqrt{\sum_{i=1}^n \frac{(x_i - y_i)^2}{s_i^2}}$, ahol s_i a tapasztalati szórás [146].

2.5.3 Szignifikáns különbségek keresése génexpressziós adatokon

Kettő vagy több csoportot egymástól legjobban megkülönböztető, legjobban leíró tulajdonságok keresésére (feature selection) több algoritmus is használható. A módszer kiválasztásánál figyelembe kell venni többek között annak sajátosságait (pl. mennyire érzékeny kiugró értékekre), az adatok jellemzőit (pl. mennyire zajosak), az elemszámot, és hogy mi a célunk, mennyire fogadhatunk el fals pozitív találatokat [147].

2.5.3.1 SAM – Significance Analysis of Microarrays

A „Significance Analysis of Microarrays” R csomag [148] módszere módosított t-teszten alapul, és a nemparaméteres eljárások közé tartozik.

A SAM analízis számításainak főbb lépései:

1. minden génre a d-érték meghatározása, mely a gén kifejeződésbeli relatív különbségét fejezi ki a két csoport között:

$$d(i) = \frac{\bar{x}_S(i) - \bar{x}_R(i)}{s_i + s_0}$$

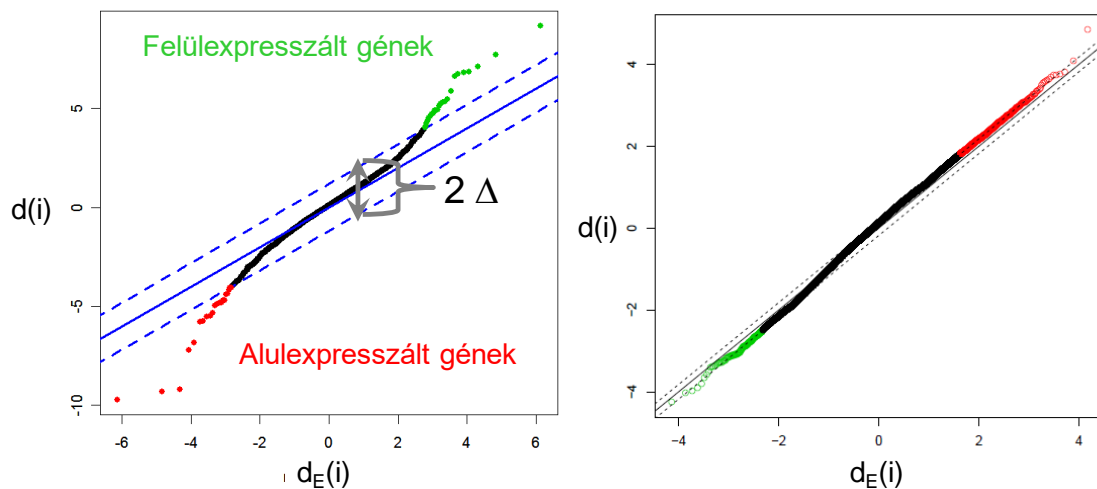
A fenti képletben az $\bar{x}_S(i)$ és az $\bar{x}_R(i)$ az i gén egyik és másik csoportokban mért átlagos expressziója, az s_i a standard hiba, az s_0 egy konstans érték. (A konstans szerepe a $d(i)$ értékek szórásának csökkentése kicsi s_i értékek esetében.) Ezt követi a $d(i)$ -k sorbarendezése: $d_1(1) < d_1(2) < \dots < d_1(g)$

2. Az adatok eloszlásának becslésére a minták p számú permutációjára meghatározza minden génnek a d-értékeit. Adott gén d-értékeinek az átlaga a $d_E(i)$, azaz a várható relatív különbség.

3. Azokat a géneket tekintjük a két csoportban (pl. szenzitív és rezisztens) eltérő mértékben kifejeződőnek, melyekre teljesül, hogy $|d(i) - d_E(i)| > \Delta$, ahol Δ egy tetszőlegesen megválasztott küszöbérték. A várható relatív különbség, $d_E(i)$, és a számított relatív különbség, $d(i)$ együttes ábrázolásával, jól látható, hogy a delta – érték megválasztása miként befolyásolja a szignifikánsnak tekintett gének számát (6. ábra). A d -érték alapján tekintünk egyes géneket jólteljesítő, top géneknek.

A SAM-mel végzett elemzés során a beviteli adatokat logaritmizáltam, 1000-szeres permutációt használtam, a delta-értéket a keresztvalidáció alapján a megfelelő értékre – amelynél legkisebb az álpozitívak száma – állítottam. A permutációk számának ilyen nagyra választása bár a számítást lényegesen meghosszabbítja, a szignifikancia szintek pontos becslésének alapfeltétele.

A SAM a szignifikancia szint kifejezésére a fals találati arányt (false discovery rate, FDR) alkalmazza.



6. ábra. Az első grafikon Tusher eredeti közleményében szereplő ábra, a második egy általunk végzett elemzés eredményeit jeleníti meg. A $d_E(i)$: várható relatív különbség, és a $d(i)$: számított relatív különbség, a Δ küszöbérték feletti és alatti gének az áttekinthetőség kedvéért eltérő színnel jelöltek.

2.5.3.2 Rank Products

A rang produktum (RankProducts) és a vele rokon rang összeg (RankSum) számítás [149] lényege, hogy a géneket mérésenként expresszió szerint sorba rendezi, és a sorban elfoglalt helye szerint rangot oszt a génnek. Egy gén rang-produktum értéke a gén rangjainak a mértani átlaga, a rang-összeg értéke pedig a rangok összege. Ezen érték

alapján tekintünk egyes géneket jóteljesítő, top géneknek. A szignifikancia szint meghatározására a SAM-hez hasonlóan permutáció-alapú becslést alkalmaz. A hibásan pozitívak aránya (proportion of false positives: pfp) az adott gén véletlen körülmények között várható RP értékének, és a RP értékek alapján növekvő sorban elfoglalt helyének a hányadosa.

2.5.4 *Túlélés analízis*

Betegcsoportok túlélési idejének a jellemzése és összehasonlítása speciális vizsgálati módszereket igényel. Ezzel a problémával először az 1950-es években biztosítási statisztikusok kezdtek foglalkozni. A *követési idő* az adott beteg vizsgálatba kerülésétől a végpont teljesítéséig számított idő. Teljes túlélés (Overall survival, OS): a diagnózistól a (daganat-specifikus) halálozásig eltelt idő, progressziómentes túlélési idő: (Progression Free Survival, PFS) a diagnózistól a progresszióig vagy a halálozásig eltelt idő, relapszusmentes-túlélés (Relapse Free Survival, RFS) az első relapszusig eltelt idő, betegségmentes túlélés (Disease Free Survival, DFS) fogalma azokra az esetekre vonatkozik, ahol az eseményt (halál, relapszus) megelőzően kuratív beavatkozás történt, progresszióig eltelt idő (Time to progression, TTP) abban különbözik a PFS-től, hogy nem számít progressziókat a nem daganatos halálok (azaz cenzorálnak tekintik ezeket az eseteket). Az onkológiai, klinikai vizsgálatokban fontos a követési idők egységes és pontos definíciója [150]. A progresszió megítélésére a radiológiai vizsgálatok alapján a WHO- [151], illetve legtöbbször a RECIST kritériumok szerint ítélik meg [152]. Bár a PFS megítélése ellentmondásos, hiszen problémát jelent, hogy a „progresszió idejét” az is befolyásolja, hogy mikor jelenik meg egy beteg az állapotát felmérő vizsgálaton, jelenleg nem létezik alkalmasabb módszer [153].

Egy beteg túlélési idejét cenzorálnak mondjuk, ha az esemény idejét nem ismerjük. Ez több okból megtörténhet, leggyakoribb, hogy a beteg a megfigyelés végén még életben volt, vagy hogy a megfigyelés alatt meghalt, de nem a vizsgált betegség miatt. Azaz ha tovább követtük volna a beteget, illetve nem halálozott volna el más betegség következtében, akkor további idő telt volna el például a progresszióig. Ezt a cenzorálást nevezzük jobb oldali cenzorálásnak. A bal oldali cenzorálást jóval ritkábban, azokban az esetekben alkalmazzuk, amikor az ismert megfigyelési idő előtt korábban, de ismeretlen időpillanatban következett be egy esemény.

2.5.4.1 Kaplan–Meier görbék

A túlélési görbe a túlélési valószínűségeket a követési idő függvényében ábrázolja, meghatározásukra leggyakrabban alkalmazott módszer a Kaplan-Meier elemzés [154]. Kaplan-Meier elemzés során halálozási arányt minden olyan időpontban meghatározzuk, amikor legalább egy halálozás történt. A cenzorált adatokat a függvényt metsző x vagy $|$ jellel szokás jelölni, a függvény függőleges esései azokban a pontokban vannak, ahol ténylegesen bekövetkezett halálozás.

Több vizsgálati csoport esetén a túlélési idők, illetve a túlélési görbék összehasonlítására lograng-próbát (log rank-test [155]) alkalmazunk. Ennek során meghatározzuk azt, hogy a két csoportra eső aktuális halálozási számból mennyi jutna egyre-egyre, ha a csoportok elméleti halandósága minden időpontban megegyezne, azaz nem lenne különbség a két csoport között. Az így kapott várt, becsült értékek és a tényleges, megfigyelt halálozás összehasonlítása khi-négyzet-próbával történik, mely a szignifikanciát p -értékkel fejezi ki.

2.5.4.2 Cox arányos hazard modell

A lograng próbával csak azt vizsgáljuk, hogy van-e szignifikáns eltérés két csoport túlélése között. Az hogy ez „milyen mértékű”, további vizsgálatokat igényel, melynek egyik módja a Cox arányos hazard modell. A hazardfüggvény azt mutatja, hogy ha valaki a t időpontban él, akkor mekkora valószínűséggel fog ott meghalni.

A vizsgált esemény kockázata a hazard rate (a kumulatív túlélési görbe meredeksége egy időintervallumban). A kockázatot úgyis definiálhatjuk, mint adott t időpontban a halál bekövetkezésének valószínűségét, amikor tudjuk, hogy az egyén a t idő előtt még él. A megfogalmazást kérdés formájában is feltehetjük: a beteg milyen valószínűséggel éli meg az öt évet a beavatkozás után, ha már három évet túlélte? Ha az egyik csoportban a halál kockázata háromszorosa a másik csoporténak, akkor a kockázat állandó marad az egész vizsgálat folyamán. Ezt úgy mondjuk, hogy a két csoport hazard függvénye egymással arányos, proporcionális. Az előbbi példa kapcsán ez azt jelenti, hogy az első csoportban a halálesemény valószínűsége háromszor nagyobb, mint a másik csoportban. A Cox-regressziós modell a vizsgált magyarázó változók relatív kockázatát becsli [156]. A hazard rate-ek egy időpillanatbeli hányadosa a hazard ratio.

2.5.5 Többszörös összehasonlítások problémája

Abban az esetben, amikor egyszerre több statisztikai tesztet végzünk, a hagyományos $p=0,05$ szignifikancia szint eredményeképpen statisztikailag minden 20. teszt eredménye még akkor is szignifikánsan eltérhet, ha valójában nincs különbség a vizsgált csoportok között. Így a gén chip alapú elemzések esetében, ahol akár 54000 próba szett vizsgálata is lehetséges, veszélyesen megnő az álpozitívak száma. Például ha egy DNS-chipen 10 ezer gént vizsgálunk, akkor $p=0,05$ esetén két minta között 500 „szignifikáns” eltérést találhatunk még abban az esetben is, ha valójában e mögött semmilyen biológiai különbség nincs. [157] Az egyik leggyakrabban alkalmazott és legegyszerűbb módszer az álpozitív találatok korrigálására Bonferroni-korrekciónak [158], mely esetében a szignifikancia szint határát jelölő α értéket osztjuk az elvégzett tesztek számával, így például ha 10000 tesztet végzünk el, akkor nem az 0,05-nél kisebb p -értékeket tekintjük szignifikánsnak, hanem csak a $0,05/10000=5 \cdot 10^{-6}$ értéknél kisebbeket. A Bonferroni-korrekciónak konzervatívsága mellett számos korlátja van [159], ezért nagy áteresztőképességű vizsgálatok esetében az álpozitívak kontrollálására más módszerek javasoltak. Az egyik ilyen módszer a Benjamini és Hochberg [160] által kidolgozott fals találati arány (False Discovery Rate, FDR) meghatározása. Az FDR a hibásan visszautasított hipotézisek (V) becsült (expected: E) aránya az összes visszautasított hipotézis számához (rejected: R) viszonyítva. A Benjamini és Hochberg által

$$FDR = E [V / R \mid R > 0] \times \text{Prob} (R > 0)$$

–ként definiált FDR tehát annak a valószínűsége, hogy hibás felfedezést teszünk.

Meg kell említenünk a Fernando és mtsai által a többszörös dependens tesztekre kidolgozott módszert [161], amely nem függ a tesztek közötti korrelációtól és a tesztek számától. Ez a $pFDR$ -hez hasonló módon számolandó „hibásan pozitívak aránya” (proportion of false positives: PFP), ahol a V és az R külön is becsült értékek:

$$PFP = E(V)/E(R)$$

A PFP az elvégzett kísérletekben felhalmozódott hibásan pozitívak arányát kontrollálja, míg a $pFDR$ a hibásan pozitívak várt arányát az egyes kísérletben.

3 Célkitűzések

PhD munkám során elsősorban biomarkerek keresésével és validálásával foglalkoztam emlő és vastagbél-daganat magas kockázatú alcsoportjaiban.

Célkitűzéseim a következők az egyes témákkal kapcsolatban:

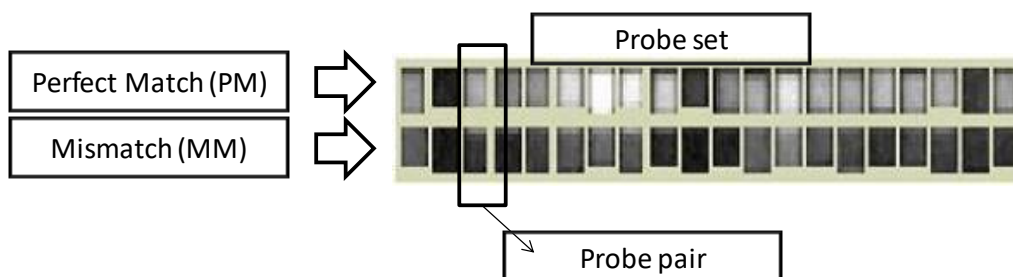
1. Vastagbél-daganatok osztályozóinak összehasonlítása és a legnagyobb prognosztikus értékkel bíró azonosítása.
2. Sejtmodellek összekapcsolása a vastagbél-daganatok altípusaival.
3. Emlődaganatok génexpresszió alapú osztályozása és összehasonlítása korábbi prognosztikus tesztekkel.
4. Emlődaganatok nyirokcsomó-érintettségének előrejelzésére új osztályozó készítése.
5. Experimentális módszerekkel azonosított biomarkerek reprodukálhatóságának összehasonlítása (siRNS).

4 Módszerek

4.1 Gén-chip módszer

A microarray vagy más néven gén chip technológia az 1990-es évek végén jelent meg először. Lehetővé tette több tízezer gén együttes vizsgálatát, utat nyitva ezzel a poligénesen meghatározott tulajdonságok széleskörű vizsgálatának.

Az Affymetrix expressziós microarray-ek esetében egy szilárd hordozóhoz (üveglapka) 25 bázisból álló oligonukleotid próbák vannak rögzítve. Ezekhez a próbákhoz hibridizáltatják a mintából származó RNS-ről készített, fluoreszcensen jelölt egyszálú cDNS vagy cRNS molekulákat. Minden transzkriptumot 11 próbával fednek le. A tökéletesen kapcsolódó próbák (perfect match, PM) mellett a nem specifikus kötődés, és a kereszthibridizáció becslése céljából partnerpróbák (mismatch, MM) helyezkednek el. Ezek a partnerpróbák az eredeti próbákhoz képest a középső, azaz a 13. pozícióban egy nukleotidtranszverziót tartalmaznak. Az egymáshoz tartozó perfect match és mismatch próbákat együtt próba pároknak nevezzük (probe pair), és egy transzkriptumothoz tartozó 11 próba pár alkot egy probe set-et. A 7. ábra a microarrayek felépítését szemlélteti.



7. ábra. A cDNS–chipek felépítése

Az általam vizsgált gén-chipek az Affymetrix Human Genome U133A Array (GPL96), a HGU133 Plus 2.0 (GPL570) és a HGU133A 2.0 (GPL571). Ezek a gén-chipek a leggyakrabban használt génexpressziós microarray-ek, illetve, mivel ugyanazokat a próbákat használják, ezért eredményeik könnyen összevethetőek, harmonizálhatók. A HGU133A chipen 22283 próba szett segítségével 39000 transzkriptumot és variánst, köztük 12298 emberi gént lehet vizsgálni. A HGU133A 2.0

chipen 22277 próba szett segítségével 18400 transzkriptumot és variánst, köztük 14500 emberi gént lehet vizsgálni. A HGU133 Plus 2.0 expressziós chip pedig 47400 transzkriptum, 20517 gén expressziójának egyidejű vizsgálatát teszi lehetővé több mint 54675 próba szett mérésével. A chip leolvasásának az eredménye egy .DAT fájl, ennek a részben processzált, kvantifikált próba intenzitás értékeket tartalmazó származéka a .CEL fájl. A vizsgált gének expressziós szintjére a hibridizáció intenzitások értékeiből következtetnek [162].

4.2 Bioinformatikai, statisztikai elemzés

Az adatok értékelését R statisztikai környezetben [163] végeztem. Az R programozási nyelv és szoftverkörnyezet statisztikai számításokhoz és ábrázoláshoz kifejezetten alkalmas. A programozást segítik az interneten elérhető “csomagok” (R packages), melyek korábban kódolt függvényeket tartalmaznak. Ezeknek a csomagoknak két nagy adatbázisa a CRAN és a Bioconductor.

4.2.1 Normalizálás

Az adatok összehasonlíthatóságának, és további feldolgozásának feltétele a normalizálás. A normalizálás történhet microarray chip-enként (MAS5 [164]) vagy az összes a kísérletben szereplő chip együttes figyelembevételével (RMA [165]).

A MAS5 számításának első lépése a háttér intenzitás, és a zaj becslése. Ehhez a chipet 16 zónára osztja, és minden zónában a legalacsonyabb 2% intenzitás átlagát háttérnek, szórását a zaj értékének tekinti. Egy adott ponthoz tartozó háttér, illetve zaj értékek a zónánkénti értékek súlyozott átlaga. A második lépésben, a jelintenzitás = PM–MM értékeként határozzák meg. Arra az esetre, ha az MM értéke nagyobb, mint a PM, bevezeti az ideális mismatch fogalmát. Ennek meghatározása során a Tukey–féle kétszer súlyozott átlagszámítást is felhasználja, mely értéke kevésbé befolyásolt a kiugró adatok által. Egy próba szett-re vonatkozó jelintenzitás meghatározásához a próbák jelintenzitásának összegzése során szintén Tukey–féle statisztika kerül alkalmazásra. Az utolsó lépésben skálázás történik, azaz a chippek átlagos intenzitását egy előre meghatározott értékre állítja be.

Az RMA normalizálás során csak a PM próbákat veszik figyelembe, a PM értékeket a próba intenzitások tapasztalati eloszlása segítségével korrigálja. A háttérzajra való korrekció és log₂ transzformáció után kvantilis normalizáció történik. Végül az egyes próbák expresszióját lineáris modell segítségével határozza meg.

4.2.2 Minőségellenőrzés

Első lépésben a chip-ek minőségellenőrzését végeztem el. A minőségellenőrzés során értékeljük a háttérjel intenzitását, a zaj mértékét (pixelek közötti variancia alapján számolt, raw Q-nak nevezett), a present call-ok arányát, az RNS-degradáció mértékét a GAPDH és a β-aktin 3' és az 5' végéhez illeszkedő próbák expressziójának arányát, illetve az ismert koncentrációban jelenlevő úgynevezett spike-in próbák, hibridizációs kontrollok jelenlétét (bioB-/C-/D- tüskék). Vizsgálataim során a leggyakrabban alkalmazott határértékeket használtam, és a továbbiakban csak azokkal a mintákkal dolgoztam tovább, melyeknél a háttérjel értéke 19 és 218 között van, $0,5 < \text{raw Q} < 14$, a present call-ok aránya több mint 30%, GAPDH 3':5' arány < 4.3 , β-aktin 3':5' arány < 18 , és a bioB-/C-/D- tüskék kimutathatóak [166]. A minőségellenőrzést a simpleaffy [167] és affyQCReport R csomaggal végeztem.

A GEO adatbázisnak veszélyes hibája, hogy sok esetben ugyanazokat a gén-chip méréseket több azonosító alatt is tartalmazza. Ezen duplikátumok, triplikátumok azonosítása és kiszűrése fontos lépése az elemzésnek. Ezért 100 véletlenszerűen kiválasztott próba expresszióját hasonlítom össze chip-enként, és amennyiben ezek teljesen megegyeznek biztosan mondhatjuk, hogy többszörös közlésről van szó [168].

4.2.3 A legjobb próba szettek kiválasztása

Az Affymetrix gén-chipek redundánsak, azaz egy génhez több próba szett is tartozik. Többféle módszer létezik a génenkénti „legjobb” próba szettek kiválasztására. Lehet azt a próba szettet választani, amelyiket a gyártó ajánlja, vagy amelyiknek a legnagyobb a szórása, vagy a legmagasabb az expressziója, illetve a legszenzitívebbet és specifikusabbat választani. Ez utóbbit végeztem ebben a vizsgálatban a JetSet-módszer segítségével. A JetSet algoritmus a próbák és a gének nukleotidjait összevetve meghatározza egy-egy próba szett génenkénti szenzitivitást, specificitását, és azonosítja azokat, melyek hasítási izoforma lefedettségben a legtökéletesebbek, és degradáció hatását a leghatékonyabban zárják ki [169].

4.2.4 *Batch hatás*

Amikor különböző forrásból, vizsgálatból származó (microarray) adatokat dolgozunk fel, még akkor is, ha a nyers fájlokkal dolgozunk, számolnunk kell a máshol készített vizsgálatok közötti, technikai eltérésekből származó eltérésekkel. Ezt a jelenséget nevezzük batch-hatásnak [170]. Ezeket a változásokat azonosítani és szűrni batch-hatás korrekciónak nevezett módszerekkel lehet [171], melyek célja a nem tervezett eltérések csökkentése, ami a további elemzés során a fals találatok csökkenését eredményezi. Azonban azokban az esetekben, amikor kiegyensúlyozatlanok a vizsgálatok (például nem hasonló a nembeli, életkor béli, stádium összetétele) a batch-hatás korrekció korlátozott értékű, nem ajánlott [172].

4.2.5 *Diagnosztikai próbák mutatói, értékelés*

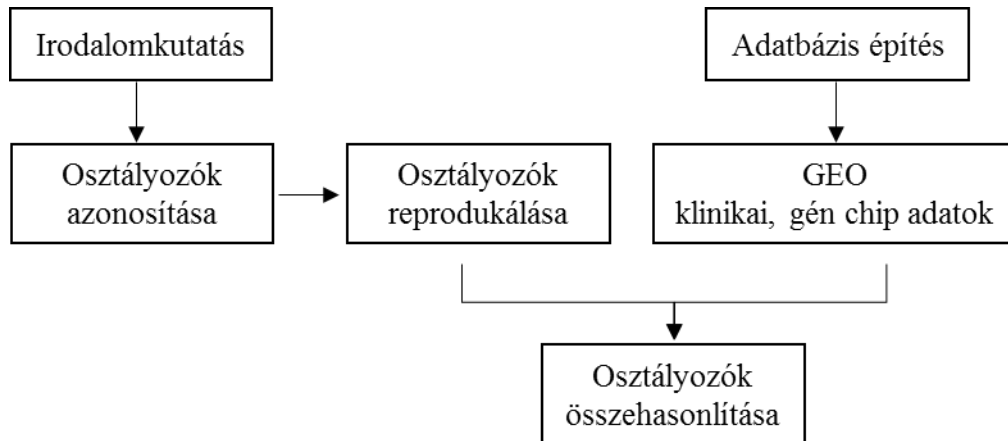
Egy klinikai teszt, osztályozó hatékonyságát klasszikusan a specificitással és a szenzitivitással jellemezzük. Amikor egy módszert, tesztet fejlesztünk, akkor fontos meghatározni, hogy mi a célunk, mely mutatókra optimalizáljuk a módszert. Vannak esetek, amikor a jó szenzitivitásra törekszünk (azaz minél több beteget/ rossz prognózisút megtaláljunk, még az álpozitív emelkedése árán is), vagy a specificitásra helyezük a hangsúlyt. A pontosság: jól osztályozott (valós pozitív + valós negatív) betegek számának aránya az összes vizsgált beteghez képest. A mutatókat meghatározó képleteket a 8. táblázat tartalmazza.

8. táblázat. A diagnosztikai próbák mutatói

		teszt, vizsgálat eredménye		
		negatív	pozitív	
ismert állapot	negatív	valós negatív (VN)	álpozitív (ÁP)	specificitás $\frac{VN}{VN+ÁP}$
	pozitív	álnegatív (ÁN)	valós pozitív (VP)	szenzitivitás $\frac{VP}{VP+ÁN}$
		szegregancia (negatív prediktív érték, NPV) $\frac{VN}{\text{összes negatív}}$	relevancia (pozitív prediktív érték, PPV) $\frac{VP}{\text{összes pozitív}}$	

4.3 Vastagbél-daganat – többgénes osztályozók

A vizsgálatom célja a korábban publikált, vastagbél-daganatokban a prognózist előrejelző tesztek, transzkripció mintázatok független validálása, összehasonlítása volt. A kísérlet egyszerűsített felépítését a 8. ábrán foglalom össze.



8. ábra. Többgénes vastagbél-daganat osztályozók értékelése

4.3.1 Adatbázis létrehozása

A saját adatbázis létrehozásához a GEO-ban olyan adathalmazokat, vizsgálatokat kerestem, melyekben vastag- vagy végbél-daganatokban végeztek Affymetrix génextpressziós microarray-n méréseket. Csak azokat az adathalmazokat gyűjtöttem ki, ahol elérhető volt a nyers adatfájl, mivel ebben az esetben van lehetőség az adatok megfelelő összehasonlítására. A keresést a következőképpen végeztük el: ((GPL96[GEO Accession] OR GPL570[GEO Accession] OR GPL571[GEO Accession]) AND (“colon cancer”[DESC] OR "colorectal cancer"[DESC] OR "CRC"[DESC] OR "colorectal neoplasms"[DESC]) AND cel[Supplementary Files]). A klinikai adatok ellenőrzését és harmonizációját kézi ellenőrzéssel végeztem.

4.3.2 Adatok előkészítése

Mivel a vizsgálat egyik célja az osztályozók reprodukálása volt, ezért az adatok előkészítésénél is az eredeti közlemények részletes leírását követtem. Amennyiben nem volt ismert a normalizáció módja, MAS5 normalizálást használtam, mivel ez az egyik leggyakrabban alkalmazott módszer. Azokban az esetekben, amikor sem a használt próba szettek neve nem szerepelt a leírásban, sem a kiválasztásuk módja, akkor a JetSet módszert alkalmaztam.

4.3.3 *Korábbi osztályozó módszerek irodalmi azonosítása*

A meta-analízisek minőségének biztosítása és reprodukálhatósága érdekében a legtöbb újság kötelezővé tette a szigorú PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) ajánlások követését [173]. Vizsgálatom során én is ezeket az irányelveket követtem.

Korábban publikált osztályozó algoritmusok keresése során a PubMed (<http://www.pubmed.com>) adatbázisban folytattam keresést a következő kulcsszavakkal: ("colon cancer"[TIAB] OR "colorectal cancer" OR "colorectal neoplasms"[TIAB]) AND "gene expression"[TIAB] AND ("prognostic"[TIAB] OR "prognosis"[TIAB] OR "survival analysis"). A találatok feldolgozása során először az absztraktokat, majd a releváns publikációk esetében a teljes szöveget tekintetem át.

4.3.4 *Osztályozók összehasonlítása*

Az osztályozók összehasonlításánál a legjobb és a legrosszabb prognózisú csoport között a „hazard ratio” értékeket Cox proportional hazards regresszióval határoztam meg, a p-értékeket logrank teszttel számoltam ki és az eredményeket Kaplan-Meier görbékkel ábrázoltam. 1405 beteg esetében volt elérhető progressziómentes túlélési adat. A multivariáns elemzés során a következő prognosztikus paramétereket vettem figyelembe: MSI-státusz, nem, MKI67 és CDX2 expresszió [174].

Az osztályozók eredményei közötti korrelációt a Cramer-féle asszociációs együttható - Cramer féle V, illetve Cramer-index néven is ismert - segítségével határoztam meg. Két változó kapcsolata a jól ismert χ^2 -négyzet statisztika alapján kifejezve a következő: $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$, ahol O_i a megfigyelt (observed), E_i a várt (expected) gyakoriságot jelenti, i : az egyes megfigyeléspárok, a kapcsolódó p-értéket a χ^2 -négyzet eloszlás alapján határozzák meg. A Cramer $V = \sqrt{\frac{\chi^2}{N(k-1)}}$, ahol N a minták száma, k pedig a kategóriák száma.

Míg a χ^2 -négyzet teszt csak arról ad információt, hogy van-e szignifikáns kapcsolat a változók között, addig a Cramer-féle V a kapcsolatnak a fokát fejezi ki, értéke 0 és 1 közötti lehet, ahol az 1 erős, a 0 gyenge kapcsolatot jelent.

4.3.5 *Gének súlyának megállapítása*

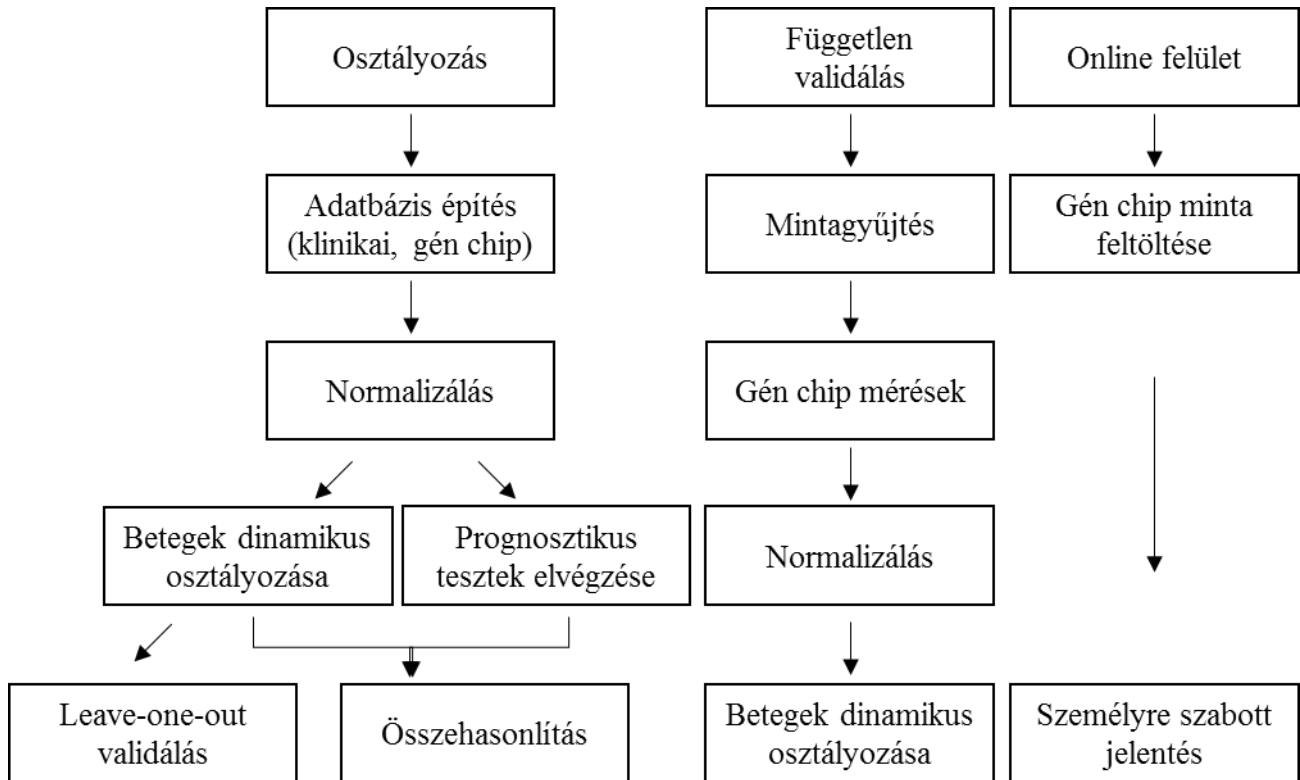
Annak érdekében, hogy összehasonlítsam az osztályozóban szereplő gének jelentőségét, minden gén esetében meghatároztam egy pontszámot, amely azt is figyelembe veszi, hogy hány gént alkalmaz egy osztályozó. Először megállapítottam a gének esetében az arányukat az egyes osztályozókban az alapján, hogy hány gént alkalmaz az adott modell, majd a következő módon meghatároztam a „súlyukat”: $\text{gén-pontszám} = [\text{osztályozók száma, amiben szerepel}] * \sum[\text{aránya az osztályozókban}]$.

4.3.6 *Sejtvonalak*

További célunk volt a vizsgált osztályozókhoz tartozó sejtkultúra modellek azonosítása, mivel így az egyes altípusok tovább vizsgálhatók kísérletes körülmények között is. A GEO és az Array Express adatbázisban olyan vastagbél eredetű sejtvonalat kerestem, melyek génexpressziója Affymetrix HGU 133 plus 2.0 chipen került lemérésre, és elérhető a nyers .CEL fájl. Számos sejtvonal gén expresszióját több kísérlet során is meghatározták, ezekben az esetekben minden méréshez meghatároztuk a sejtvonal osztályát, és ha a sejtvonal 60%-ában ugyanazt az osztályt kaptuk, akkor tekintettük ez eredmény robosztusnak, egyéb esetekben nem meghatározhatóként definiáltuk a vonal altípusát. A sejtvonalak mutációs státuszát a COSMIC (<http://cancer.sanger.ac.uk/cosmic>) és a Cancer Cell Line Encyclopedia (<http://www.broadinstitute.org/ccle>) adatbázisokban publikált eredmények alapján foglaltam össze.

4.4 Emlődaganat – rossz prognózisú betegek azonosítása

A vizsgálat célja rossz prognózisú emlődaganatok génexpresszió alapú azonosítása volt. A kísérletek felépítését és a használt módszereket a 9. ábrán foglalom össze.



9. ábra. Rossz prognózisú emlődaganatok dinamikusan azonosítása

4.4.1 Mintagyűjtés

Eredményeim független validálása érdekében a Frankfurti és a Hamburgi Egyetemi Kórház nőgyógyászati osztályán 325 korai stádiumú emlődaganatban szenvedő beteg friss fagyasztott tumormintáit és klinikai adatait gyűjtöttük ki. A vizsgálatot a helyi etikai bizottság engedélyezte, minden beteg beleegyező nyilatkozatot írt alá. A betegek 2004 és 2007 júliusa között részesültek műtétben, az RNS izolálása friss fagyasztott mintákból RNeasy Midi (Qiagen) oszloppal történt. A transzkripció profil meghatározására Affymetrix Human Genome U133A gén-chipet alkalmaztunk. A klinikai és gén-chip adatok a GSE4611 (Frankfurt dataset) és a GSE46184 (Hamburg dataset) azonosítók alatt kerültek publikálásra.

4.4.2 Adatbázis létrehozása

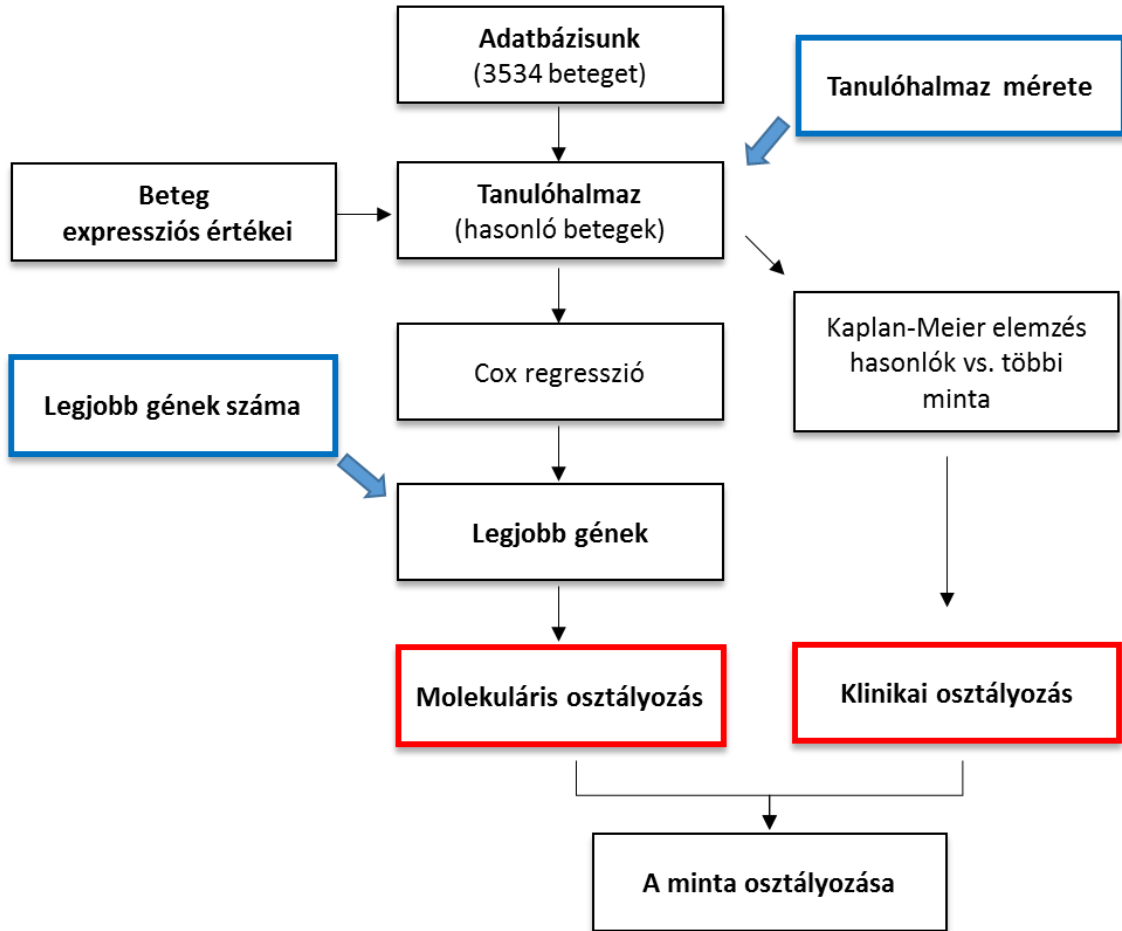
A GEO (Gene Expression Omnibus) (<http://www.pubmed.com/geo>) és a TCGA (The Cancer Genome Atlas) (<http://cancergenome.nih.gov>) adatbázisokat vizsgáltuk át megfelelő adatokat keresve. Olyan publikációkat kerestem, ahol emlődaganattal kezelt betegek microarray adatait közölték. Feltétel volt, hogy legalább húsz beteg szerepeljen a vizsgálatban, legyenek elérhető a kezelési és túlélési adatok és a nyers microarray adatok. Három microarray platformot, a GPL96 (Affymetrix HG-U133A), a GPL570 (Affymetrix HG-U133 Plus 2.0), és a GPL571/GPL3921 (Affymetrix HG-U133A 2.0) használtunk fel. A kinyert microarray és klinikai adatokból felépítettük a saját adatbázisunkat, ahol a beteg azonosítójához a génenkénti expresszió és a klinikai adatok (túlélési, klinikai válasz és kezelési adatok) csatoltam.

4.4.3 Adatok előkészítése

Az adatok normalizálását a korábban tárgyalt MAS5 algoritmussal végeztem. Az elemzés során csak azokat a próba szetteket vettem figyelembe, amelyek mind a 3 fajta chipen megtalálhatóak (n=22277). A batch hatás csökkentése érdekében az egyes gén-chipek átlagos expressziós értékeit a 22277 próba szettre 1000-re állítottam. Mivel a vizsgálat későbbi részében a medián génexpresszió alapján fog történni a betegek csoportosítása, ezért ebből a vizsgálatból kizártam azokat a próba szetteket, ahol egyetlen egy minta esetében sem volt az 1000 felett a próba szett expressziója. A redundáns próba szettek esetében a legjobb kiválasztására a JetSet módszert alkalmaztam. Így a továbbiakban használt gének száma: 9886.

4.4.4 Betegek osztályozása

Vizsgálatomban az osztályozás két fő lépésből áll, egyrészt egy „molekuláris”, másrészt egy „klinikai” osztályozásból. A molekuláris osztályozás lényege hogy a vizsgált beteghez hasonló minták csoportjában minden génnek megvizsgáljuk a prognosztikus hatását. Az így azonosított prognosztikus gének közül kiválasztjuk a „legjobb” géneket, melyek expressziója alapján jó, illetve rossz prognózisú csoportba soroljuk a beteget. A klinikai osztályozás során a kiválasztott hasonló betegek prognózisát vetem össze az adatbázisban szereplő többi beteg prognózisával. Az osztályozás menetét a 10. ábrán szemléltetem, az optimalizálás lehetőségeit kékkel jelölöm.



10. ábra. Az osztályozás lépései, késsel jelölve az optimalizáció lehetőségeit

A molekuláris osztályozás során, az első lépésben a vizsgált beteghez leginkább hasonló betegek csoportját választjuk ki. Ebben a lépésben módszerünk optimalizálására többféle lehetőségünk van. Egyrészt változtathatjuk a távolságszámítás módját (pl. euklidészi-, Manhattan-távolság), illetve a hasonlóság megállapítása során figyelembe vett gének számát. Előzetes vizsgálataim alapján euklidészi-távolságszámítást és az összes ($n=9886$) gén használatát választottam. A kiválasztott hasonló betegek csoportján (tanuló halmaz) minden egyes gén esetében a medián expresszió alapján két csoportra bontva Cox-regressziót végez az osztályozónk, majd a legmagasabb, illetve legalacsonyabb házard ratio-jú gének összesített, átlagos expressziója alapján két csoportra bontja a betegeket. Végül megnézzük, hogy az éppen vizsgált beteg melyik csoportba (jó/rossz prognózisú) tartozik.

A klinikai osztályozás azt hivatott vizsgálni, hogy a kiválasztott hasonló betegek prognózisa hogyan viszonyult az adatbázisunk többi betegéhez. A két csoport összehasonlítása ebben az esetben is Cox-regresszióval történik.

A végső, kombinált osztályozást a molekuláris és a klinikai osztályozás összevetésével érem el (9. táblázat). Itt látszik jól a klinikai osztályozás szerepe, például ha a hasonló betegek csoportja rossz prognózisú, akkor betegünk végső osztályozása nem lehet „jó”. Mivel osztályozónk – szemben a gyakorlattal - nem fix, előre definiált génlistát alkalmaz, ezért neveztük el dinamikus osztályozónak.

9. táblázat. Kombinált osztályozás (N.Sz.: Nem szignifikáns)

		Molekuláris osztályozás		
		Jó	N.Sz.	Rossz
Klinikai osztályozás	Jó	Jó	Közepes	Közepes
	N.Sz.	Jó	Közepes	Rossz
	Rossz	Közepes	Közepes	Rossz

4.4.5 *Optimalizálás*

Az előbbieken leírt módszer optimalizálására több lépésben is lehetőségünk van. Változtatható a kiválasztott hasonló betegek száma, és az osztályozás során alkalmazott legjobb gének száma is.

Leave-one-out keresztvalidáció esetében validációra egyetlen mintát használunk, míg a többi minta részt vehet az osztályozás lépéseiben, és ezt megismételjük az adatbázisunk minden elemére.

A módszer optimalizálását leave-one-out keresztvalidációt alkalmazva a következő paraméterek kombinációjának szisztematikus vizsgálatával végeztük: gének száma: 3-tól 200-ig, tanulóhalmaz mérete: 100-tól 500 betegig. Az eredmények összehasonlítása során a χ^2 -teszt logrank p-értékét hasonlítottuk össze.

4.4.6 *Összehasonlítás más osztályozókkal*

Az általam fejlesztett módszert a klinikai gyakorlatban jelenleg is alkalmazott tesztekkel hasonlítottam össze. A három legelterjedtebb ilyen teszt a Bevezetésben is bemutatott 97-génes Genomic Grade Index (GGI) [175, 176], 70 gént használó

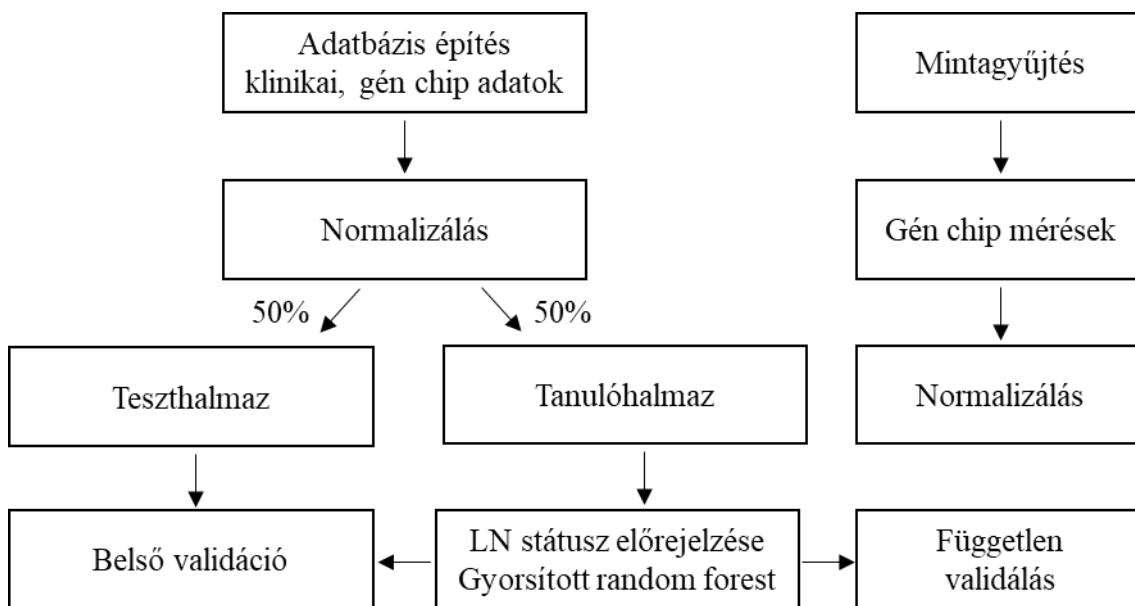
Mammaprint [177], a 21-génes Oncotype DX. A minták GGI és Mammaprint osztályait a genefu R csomag segítségével határoztam meg, az Oncotype DX esetében az RT-PCR alapú teszt gén chipre való átalakítását munkacsoportunk már korábban közölte [166]. Az összehasonlítás során a relapszusmentes túlélést vettem össze prognosztikus tesztenként az összes beteg, és al csoportok esetében. Amennyiben a teszt három csoportba sorolja az eseteket (Oncotype DX, Dinamikus osztályozónk), a legjobb és a legrosszabb prognózisú csoportok közötti eltérést vizsgáltam. A receptorstátuszok (ER, HER2) meghatározásához munkacsoportunk által korábban leírt határértékeket használtam [178]. Az 5 éves túlélési adatokat tekintve szenzitivitás, specificitás, pozitív- és negatív prediktív értékeket határoztam meg.

4.4.7 Online felület

Az osztályozónkat Drupal alapú online felületen tettük elérhetővé: http://www.recurrenceonline.com/?q=Re_training. A .CEL fájl feltöltése után személyre szabott jelentés készül a beteg osztályozásáról.

4.5 Emlődaganat - Nyirokcsomó érintettség előrejelzése

Az emlődaganatok esetében a hónalji nyirokcsomók érintettsége az egyik legfontosabb prognosztikai stádiummeghatározó tényező. A nyirokcsomó-negatív esetekben azonban az őrszem nyirokcsomók eltávolítása nem jár terápiás előnnyel. Jelen vizsgálatban ezért az volt a célom, hogy az elsődleges tumor génexpressziós mintázata alapján előrejelezzem a hónalji nyirokcsomók érintettségét. A kísérletek felépítését és a használt módszereket a 11. ábrán foglalom össze.



11. ábra: Nyirokcsomó érintettség előrejelzésének lépései

4.5.1 Mintagyűjtés

Minden biomarker kutatásnak fontos lépése az eredmények független validálása. Ezért egy nemzetközi együttműködés keretein belül a seattle-i (USA) Atossa cég biobankjából száz 2004 és 2010 között operált beteg paraffinba ágyazott, formalin fixált mintáit és a hozzájuk tartozó klinikai adatokat gyűjtöttük ki. A minták minden esetben a primer daganatból származnak. Az RNS-t Macherey-Nagel reagenssel izoláltuk, a génexpresszió meghatározására Affymetrix Human Genome U133Plus 2.0 gén-chipet (Affymetrix, Santa Clara, CA) alkalmaztunk. A génexpresszió meghatározását a gyártó előírásai szerint végeztük, azaz Affymetrix SensationPlus FFPE Amplification és 3' IVT labeling kit alkalmazásával. Az etikai engedélyt az intézeti etikai bizottság hagyta jóvá. Elemzésem követi a MIAME-útmutató (Minimális információ a microarray kísérletekről)

ajánlásait [179]. A nyers microarray adatok (.CEL fájlok) a GEO (Gene Expression Omnibus database, www.ncbi.nlm.nih.gov/geo) adatbázisba [180] kerültek feltöltésre.

4.5.2 Adatbázis létrehozása

Az előbbi kísérletek során tárgyalt módon a GEO génexpressziós adatbázis alapján saját adatbázist építettem. Olyan vizsgálatokat kerestem, ahol HG-U133 Plus 2.0-n vagy HG-U133A gén-chipen mérték emlődaganat génexpresszóját, és elérhető klinikai információ a nyirokcsomók érintettségéről. Fontos továbbá a jelen vizsgálatból kizárni a primer szisztémás (neoadjuváns) terápiában részesült betegeket, mivel az a hónalji régió downstagingjéhez vezethet.

4.5.3 Adatok előkészítése

Az affy csomag segítségével végzett minőség-ellenőrzést [168], és a duplikátumok eltávolítását követően [166] a nyers .CEL fájlokat fRMA algoritmussal, alternatív annotációt alkalmazva normalizáltam. Az RMA normalizáció hátránya, hogy mivel a próbák torzító hatásának (próba specifikus hatások) becslését több chip alapján végzi, ezért egyedi minták normalizálására, így diagnosztikus gyakorlatban sem alkalmazható, illetve a már RMA normalizált adatokhalmazok sem összehasonlíthatók. Ezért fejlesztették ki a fagyasztott (frozen)-RMA módszerét, ahol a próba specifikus hatásokat előzőleg, sok gén chip adatait elemezve előre meghatározták a különböző chip típusok esetében [181].

Az alternatív annotáció során a próba szettből csak azokat a próbákat vettem figyelembe, amelyek a transzkriptum 5' vége felé helyezkednek el, mivel a validálás során vizsgált FFPE mintákban jelentős RNS degradáció várható [182].

A további elemzés során csak azokat a próba szetteket használtam, amelyek mindkét chipen megtalálhatóak (n=22277). Mivel nőbetegek emlődaganatait vizsgáltam, ezért meghatároztam az Y-kromoszómán levő génekhez tartozó próba szettek átlagos expresszióját, és kizártam azokat a próba szetteket, amelyek átlagos expressziója ezt nem haladta meg. Génenként a legjobb próbák kiválasztására a JetSet módszert alkalmaztam, a további vizsgálatokat 9462 génnel végeztem.

4.5.4 *Betegek osztályozása*

Abból a megfontolásból, hogy kellő méretű tanulólthalmazokkal tudjak a továbbiakban dolgozni, nem a klasszikus négy, hanem három csoportba soroltam a betegeket: ER-negatív, ER-pozitív és MKI67-pozitív, valamint ER-pozitív és MKI67-negatív betegeket különböztettem meg. A gén chip esetében az ER-, MKI67-pozitivitás meghatározására korábban validált határértékeket használtam [166]. A nyirokcsomó-státusz és a prognózis közötti kapcsolatot Cox-regresszióval vizsgáltam, Kaplan-Meier görbékkel jelenítettem meg. Osztályozónk megfelelő tanítása és validálása érdekében a betegek véletlenszerűen kiválasztott fele a tanulólthalmazt, másik fele a belső validációs halmazt képezte.

A nyirokcsomó pozitív és negatív betegek között a különbözően kifejeződő gének azonosítására SAM és RankProduct algoritmusokat használtam. Előzetes vizsgálatok alapján végül a RankProducttal azonosított géneket alkalmaztam a további lépésekben.

Az osztályozást csoportonként, gyorsított döntési fákkal, a caret R csomaggal végeztem [183].

4.5.5 *Útvonal elemzés*

A nyirokcsomó pozitív és negatív esetek közötti különbség jobb megértésének céljából a különbözően expresszálódó géneket gén-ontológiai és útvonalkeresésnek vettem alá. Az elemzést clusterProfiler [184] és pathview [185] R csomagok segítségével végeztem.

4.6 Biomarkerek reprodukálhatóságának összehasonlítása

Az onkológiai kísérleteknek egyik fontos módja a biomarkerek, transzkripció hatások sejtkultúrán történő validálása. Egy gén hatásának vizsgálatának egyik módja a gén csendesítése, és az így létrejövő változások (más gének expressziója, fenotípus) jellemzése. A kis interferáló RNS-sel (siRNS-sel) történő csendesítés az RNS interferencia elvén működik, hatására a DNS-ről mRNS-re átíródó információ a gén által kódolt fehérje szintézisének megkezdődése előtt gátlódik (poszttranszkripció géncsendesítés) [186]. A kísérletek során a csendesítés hatékonyságát qPCR-el, vagy Western blot-tal szokás ellenőrizni.

A géncsendesítés hatékonysága sejtvonalanként eltérhet, ezért a csendesítés hatékonyságának vizsgálatát és összehasonlítását tűztem ki célul olyan kísérletekben, ahol a siRNS-sel történő géncsendesítést és génexpressziós microarray-t együttesen alkalmaztak.

Vizsgálatom első lépésében a GEO adatbázisban azonosítottam olyan kísérleteket, ahol daganat eredetű sejtvonalon végeztek siRNS-sel géncsendesítést, és Affymetrix HGU133A vagy HGU133Plus2.0 chip-en végeztek génexpresszió mérést. Ezt a keresést a GEOquery csomaggal, R programkörnyezetben végeztem. A találatok további szűrése kézi ellenőrzéssel történt. A további vizsgálatomat azokra a sejtvonalakkal korlátoztam, amelyekkel legalább 20 mérés történt, és azokra a génekre, melyekre legalább 5 csendesítéses kísérlet volt. A mintákhoz tartozó génexpressziós adatokat MAS5 algoritmussal normalizáltam. Mivel az Affymetrix gén-chipek redundánsak, azaz egy génhez több próba szett is tartozik, ezért génenként a legmegfelelőbb próba szettek kiválasztásához JetSet módszert alkalmaztam.

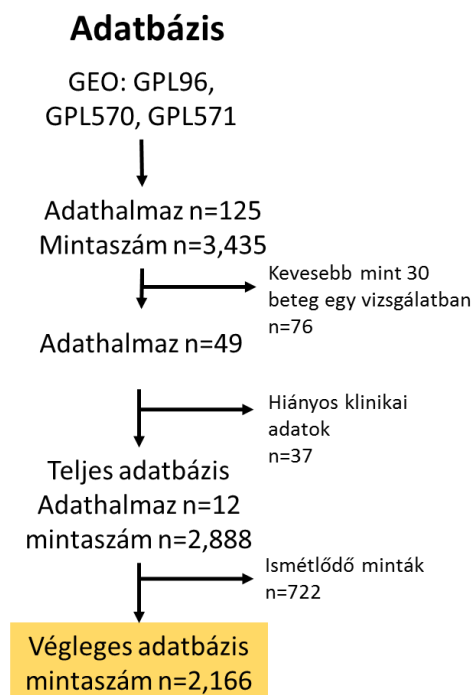
A géncsendesítés hatásosságának a megítélése során t-tesztet használtam, és a génexpressziók hányadosát vizsgáltam ($\text{Fold change} = \frac{\text{célgén expressziója[csendesített]}}{\text{célgén expressziója[kontroll]}}$). Jelen vizsgálatban nagyon megengedő határértéket alkalmaztam, már az expresszió 70% alá való csökkenést hatékonynak csendesítésként ítélt meg.

5 Eredmények

5.1 Vastagbélrák - rossz prognózisú betegek azonosítása

5.1.1 Adatbázis

A GEO adatbázisban GPL96, GPL570, GPL571 chip-eken mért vastagbél daganatos mintákat keresve 125 adathalmazban (dataset-ben) azonosítottam mintákat. A dataset kifejezés a GEO-ban az egy kísérlethez tartozó adatok (például génexpressziós adatok, metilációs adatok és klinikai adatok) összességét jelenti. Az adatbázisépítés lépéseit a 12. ábra foglalja össze. A következő lépésben kizártam a kis esetszámú, kevesebb mint 30 beteget tartalmazó vizsgálatokat (n=76 adathalmaz), majd azokat, amelyek nem tartalmaztak klinikai adatokat (n=37 adathalmaz). Azokban az esetekben, amikor egy mintát különböző azonosítók alatt is közöltek, az időben korábban közölt azonosítóval kerültek az adatbázisunkba, a duplikátumokat pedig kizártam (n=777 minta).



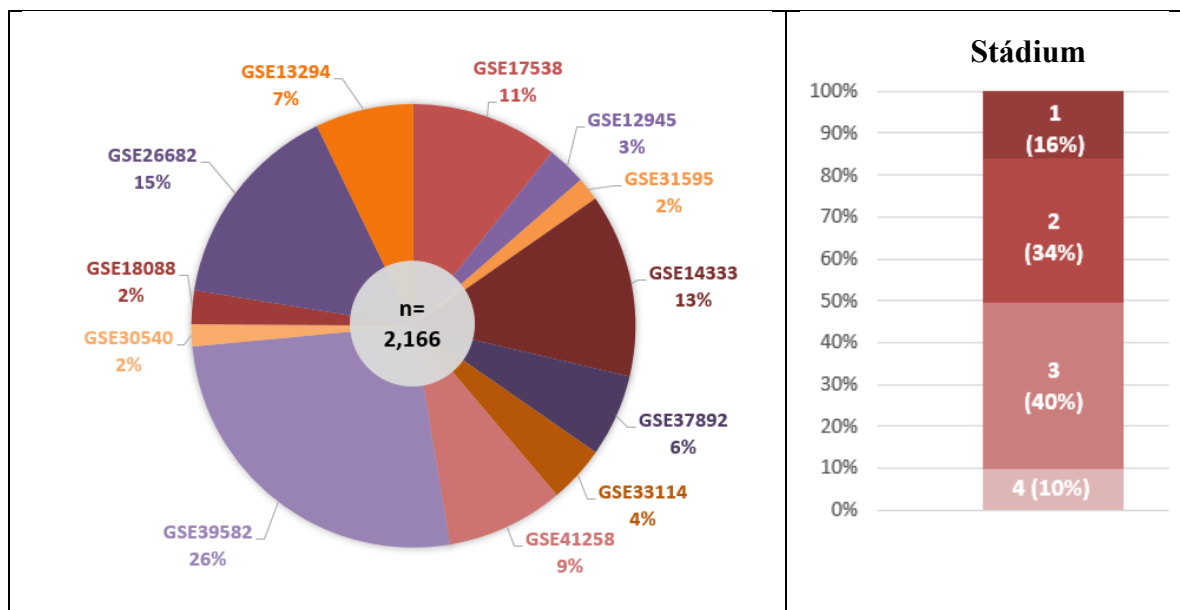
12. ábra. Az adatbázisépítés lépései

A végleges adatbázisunkat a következő adathalmazok alkotják: GSE17538, GSE12945, GSE31595, GSE14333, GSE37892, GSE33114, GSE41258, GSE39582,

GSE30540, GSE18088, GSE26682 és GSE13294, melyek összesen 2166 beteg klinikai és génexpressziós adatait tartalmazzák, közülük 1405 beteghez volt elérhető relapszusmentes túlélési adat.

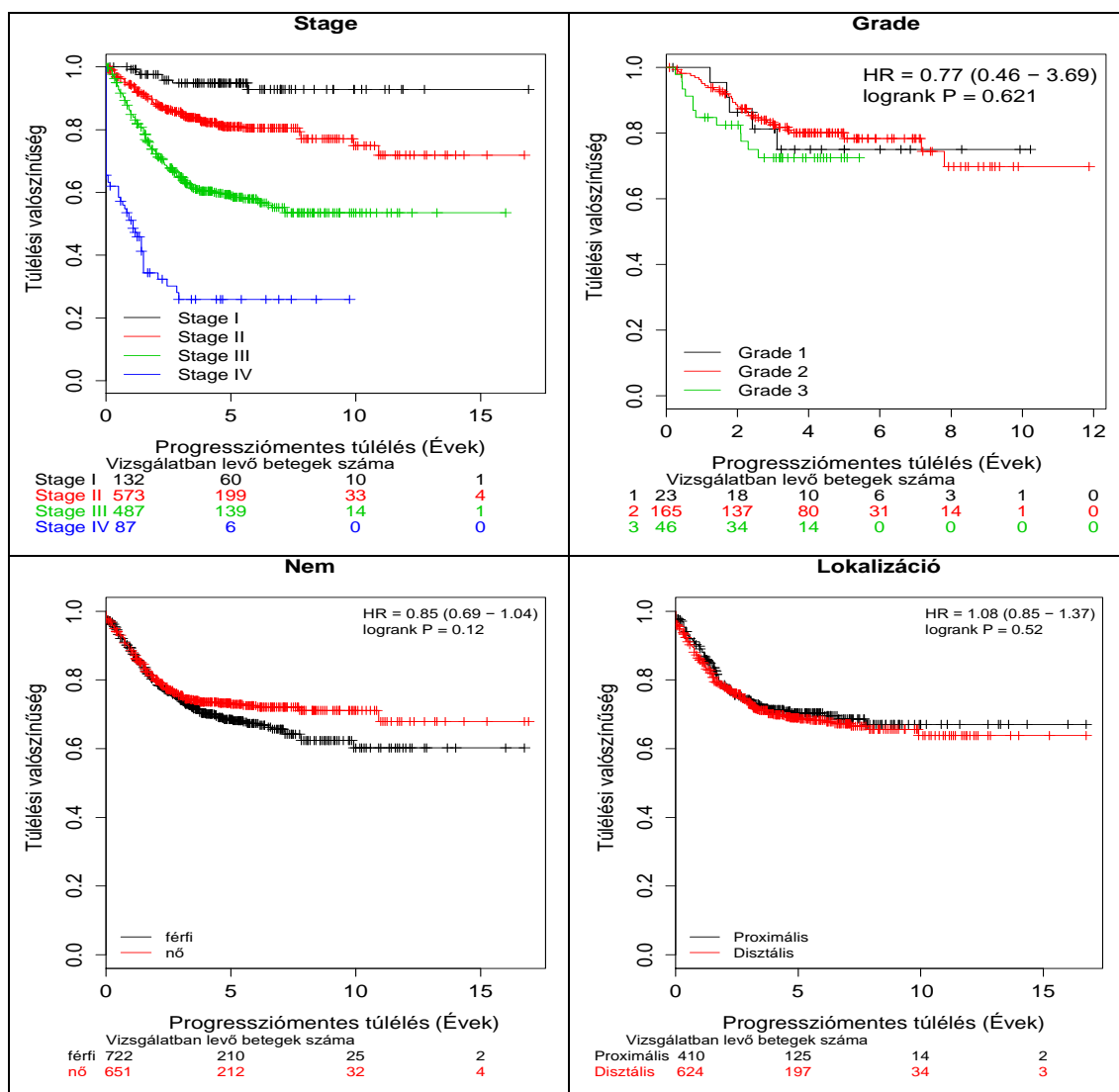
10. táblázat. Az adatbázis betegeinek klinikai jellemzői

Adathalmaz	# betegek	Medián RFS	Nem (nők %)	Grade (1/2/3)	Stage (1/2/3/4)	Medián kor	Lokalizáció (proximal/distal)
GSE17538	232	45,71	47,4	17/166/30	28/72/76/56	65,5	-
GSE12945	62	49	54,8	0/31/31	13/23/21/5	65	0/33
GSE31595	37	38,33	59,5	8/21/6	0/20/17/0	73	14/23
GSE14333	290	38,46	43,4	-	44/94/91/61	67	122/164
GSE37892	130	42,75	46,9	-	0/73/57/0	68	72/57
GSE33114	90	38,67	53,3	-	-	74	-
GSE41258	185	63,5	47	-	28/50/49/58	66	25/65
GSE39582	566	43	45,2	-	33/264/205/60	68,1	224/342
GSE30540	35	-	-	-	-	-	-
GSE18088	53	-	50,9	2/35/16	-	66	28/25
GSE26682	331	-	45,9	-	-	74	-
GSE13294	155	-	-	-	-	-	-
Összes	2166	43,12	42,6	27/253/83	146/596/516/240	69	485/709



13. ábra. Az adatbázis összetétele

A betegek legnagyobb hányada (566 beteg, 26%) a GSE39582 dataset-ből származik. A vizsgált betegek 57,4%-a férfi, és a túlnyomó többségük, 74,2%-uk stage II vagy stage III. Ez azért kiemelkedően fontos mivel a prognosztikus teszteknek ebben a betegcsoportban van a legnagyobb klinikai jelentőségük. A beteg mintavételkori medián életkora 69 év volt, a medián relapszusmentes túlélési idő 43,1 hónap. Ezen klinikai adatok áttekintését szolgálja a 10. táblázat, és a 14. ábra.



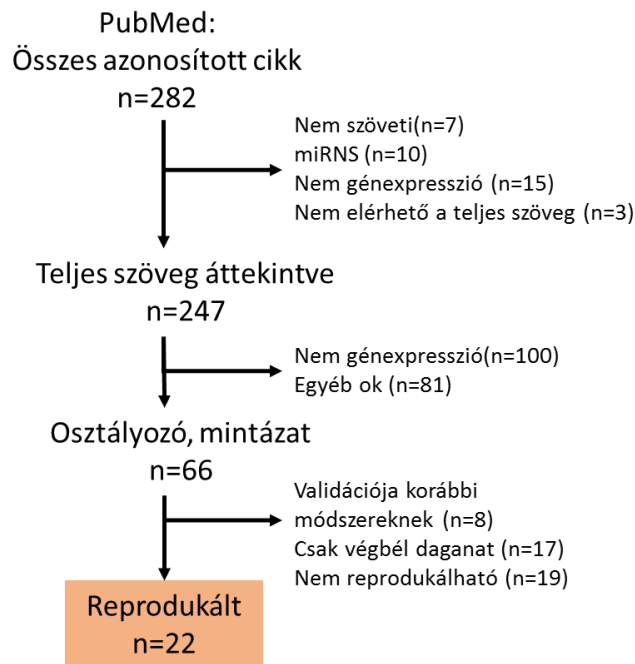
14. ábra. Betegek jellemzői és a progressziómentes túlélés

A stádiumot összevetve a relapszusmentes túléléssel jól látszik a csoportok közötti jelentős prognózisbeli különbség. A differenciáltság fokát (grade) vizsgálva – bár kevés beteg esetében volt elérhető adat – nem látszik szignifikáns különbség a magas és alacsony differenciáltságú csoportok között.

5.1.2 Algoritmusok azonosítása

Az irodalmi adatok keresése során 282 absztraktot tekintettem át, majd kizártam azokat, amelyek nem szöveti, hanem például szérumban mérhető markerekkel foglalkoztak (n=7), melyek mikro-RNS-sel (n=10), vagy nem génexpresszió alapultak (például fehérje alapú) (n=15), valamint azokat a cikkeket is, ahol nem volt elérhető a teljes szöveg (n=3). (15. ábra)

Molekuláris osztályozók



15. ábra. Algoritmusok azonosítása

A kiegészítő anyagok részletes áttekintése, és az eredeti közlemények szerzőivel történő levélváltásokat követően 22 osztályozót tudtam reprodukálni. Az osztályozók egy részének betanításához meghatározott betegcsoportokat használtak a szerzők, a módszerek reprodukálásánál én is ezeket alkalmaztam. További fontos lépés azoknak az adathalmazoknak (dataseteknek) az algoritmusonkénti kihagyása, amelyen azok azonosításra kerültek, hiszen ezen betegcsoportok esetében természetesen az adott módszer jobban teljesítene a többihez képest. A reprodukálhatóságot legtöbbször a hiányos dokumentáció, illetve a tanulólthalmazok és a klinikai adataik elérhetősége akadályozta. A következőekben röviden összefoglalom az azonosított algoritmusokat.

5.1.2.1 Budinska – nem felügyelt osztályozás: 5 altípus

Budinska és mtsai. [187] öt vastagbél-daganat altípust – felszíni kripta-szerű („surface crypt-like”), mélyebb kripta-szerű („lower crypt-like”), metilátor-szerű („CpG island methylator phenotype, CIMP-H-like”), mesenchimális és kevert – azonosított génexpressziós adatok nem-felügyelt (unsupervised) osztályozásával. A lineáris diszkriminancia-analízisen alapuló osztályozó 54 meta-gént (658 gént) alkalmazott.

5.1.2.2 CCHS – prognosztikus értékkel rendelkező hypoxia score

Krónikus hipoxia következtében megváltozott expressziójú gének prognosztikus értékét vizsgálva vastagbél-daganatban a következő 6 gén expresszióváltozását igazolták, és olyan modellt készítettek, melyben ezen gének expressziója súlyozottan szerepel. Az így létrehozott „Colon Cancer Hypoxia Score: CCHS” értékét a következőképpen határozták meg: $CCHS\text{-score} = 1.301 + 0.543*[BCCIP] - 0.416*[BNIP3L] + 0.596*[GADD45B] + 0.538*[INSIG2] - 0.177*[TP53]$. Ezen modellt FFPE mintákban validálva azt találták, hogy a 4,526-nál nagyobb CCHS score-ral rendelkező betegek szignifikánsan rosszabb progresszió-mentes túléléssel rendelkeztek [188].

5.1.2.3 Chang95

Chang és mtsai [189] korábban publikált génlisták hálózati elemzésével egy 95 génből álló osztályozót készített, mely ezen 95 gén átlagos expressziója alapján három csoportba sorolja a betegeket.

5.1.2.4 CIN25 – kromoszómális instabilitás

Hat tumor típus metaanalízise során Carter és mtsai [190] a kromoszóma instabilitás (chromosomal instability, CIN) 25 génből álló expressziós mintázatát azonosította. Ezen 25 gén expressziójának az összegét meghatározva a medián értékek feletti daganatokat tekintik kromoszómális instabilitással rendelkezőnek. Számos daganat esetében igazolódott ezen csoport rossz prognózisa.

5.1.2.5 CMS – konszenzus molekuláris altípusok

Hat korábban publikált vastagbélrák osztályozó szerzői egy konzorciumot alkotva 5000 minta (microarray és RNA-seq adatok) feldolgozásával négy daganat osztályt azonosítottak, melyeket „konszenzus molekuláris altípus”, consensus

molecular subtypes (CMSs)-nak neveztek el [191]. Ezen osztályozó alapja hat korábbi osztályozó: Budinska, Schlicker, CRCassigner-786, DeSousa és Marisa integrálása hálózati analízissel. A meghatározott altípusok a következők: CMS1: mikroszatellita instabil, CMS2: kanonikus (a klasszikus adenoma-dysplasia-carcinoma szekvencia modelhez legközelebb álló), CMS2: metabolikus altípus, CMS4: mesenchymális altípus. A betegek 13%-át nem tudták ezekbe az osztályokba sorolni. A CMS1 a legjobb prognózisú csoport, a CMS3 és CMS4 közepes, a CMS2 pedig a legrosszabb prognózisú csoport. Mivel ebben a konszenzus osztályozóban hat korábbi osztályozót integráltak, ezért az objektív értékelés során ki kellene zárni az összes olyan beteget, akik a 6 osztályozó illetve a CMS osztályozó tanuló halmazában szerepelnek. Ez azonban az összes általunk vizsgált beteg kizárását jelentené. A csoportok jellemzése során a következő megállapításokra jutottak. A CMS1 mikroszatellita instabil, így ebben a csoportban a legmagasabb a mutációk száma, gyakori a BRAF, PTEN mutációi, a CMS2 altípusban gyakoriak a KRAS mutációk, míg a CMS4-ben a TGFβ aktivációja figyelhető meg. Ez utóbbi altípus rendelkezik a colitis alapján kialakult vastagbél-daganatok jellemzőivel [192].

5.1.2.6 ColoGuideEx – magas és alacsony kockázatú betegek

Agesen és mtsai [193] 13 génes, ColoGuideEx nevű osztályozót fejlesztettek stage II-es colorectalis betegek prognózisának meghatározására. Ez az osztályozó génenként rossz prognózisú csoportba sorolja a betegeket az alapján, hogy az adott gén expressziója a 80. percentilis felett (amennyiben a gén magas expressziója rossz prognózissal függ össze), illetve 20. percentilis alatti (amennyiben a gén alacsony expressziója rossz prognózissal függ össze). Ha legalább öt gén esetében a magas rizikójú csoportba tartozik a beteg, akkor magas rizikójúnak tekintendő. A gének 80. és 20. percentilisének megállapítása során ugyanazt a tanulóhalmazt alkalmaztuk, mint az eredeti közleményben.

5.1.2.7 ColoGuidePro – rossz és jó prognózis

Sveen és mtsi [194] stage II és III betegekre 7-génes RT-PCR alapú ColoGuidePro-nak nevezett tesztet írtak le, mely a ColoGuideEx egy változatának is tekinthető, hiszen az elv és a szerzők is megegyeznek. Ebben a tesztben négy gén (DMBT1, NT5E, SEMA3A, WNT11) magas és 3 gén (CXCL9, OLFM4, UGT2B17)

alacsony expressziója függ össze rossz prognózissal. Egy beteg akkor tekintendő magas kockázatúnak, hogy ha legalább három gén esetében a rossz prognózisú csoporthoz tartozik.

5.1.2.8 CRCassigner – öt, felügyelet nélküli tanulással azonosított altípus

A nem-felügyelt (irányítatlan, unsupervised) osztályozás esetében nincsenek előre definiált csoportok, így alkalmas új, klinikai és molekuláris faktoroktól független szubtípusok azonosítására. Sadanandam és mtsai nemnegatív mátrix faktorizációt microarray-en mért génexpressziós adatokon alkalmazva a betegek öt csoportját különítette el. Majd Significance analysis of Microarrays-t [195] alkalmazva a csoportok között 786 eltérő expressziójú gént azonosítottak, végül az osztályozást Prediction analysis of Microarray-jel (PAM) [196] végezték. Az öt csoportot utólagos jellemzés alapján így nevezték el: goblet-sejt szerű, enterocita, összejszerű, gyulladásoos, progenitor-szerű (transit amplifying - átmenetileg osztódó). Az összejszerű altípus rendelkezik a legrosszabb prognózissal [197].

5.1.2.9 DeSousa – CIN, MSI és CIMP-pozitív altípus

De Sousa és mtsai szintén nem-felügyelt klaszteringet alkalmazva három betegcsoportot azonosítottak HGU-133 Plus 2.0 géncsipek elemzésével, majd az osztályozást ők is Prediction Analysis of Microarrays-szel végezték. Ezt a három szubtípust úgy jellemezték, mint 1: kromoszomális instabil, 2: Mikroszatellita instabil, 3: CIMP-pozitív. Az osztályozójukat elérhetővé tették a DeSousa2013 R csomagban, azonban a HGU133A minták esetében a PAM újratanítására volt szükség [198].

5.1.2.10 Marisa – 6 molekuláris altípus

Marisa és mtsai nem-felügyelt megközelítést alkalmaztak, azonban nem nemnegatív mátrix faktorizációt használtak, hanem konszenzus hierarchikus klaszteringet a csoportok azonosítására. A hierarchikus klaszterezésnél nem kell előre ismernünk a létrehozandó klaszterek számát, míg a nem-hierarchikus eljárásnál már kiinduláskor meg kell adnunk azt (particionáló eljárás). A másik fontos különbség, az eljárás időigénye. A hierarchikus klaszterezés lényegesen nagyobb időigény. Az osztályozásra a PAM-hez hasonló egyszerű középpont-alapú (centroid) távolság meghatározást alkalmaztak (azaz klaszterközéppontoktól való távolság alapján). Az osztályozót a citccmst R csomagban (citccmst: CIT Colon Cancer Molecular Subtypes)

tették elérhetővé. A hat alcsoportot későbbi jellemzésük alapján a következőképpen nevezték el: kromoszomális instabilitást mutató és immunreakciókban szereplő gének expressziója csökkent (CIN-ImmuneDown), mismatch repair deficiens (dMMR), gyakran KRAS mutációt hordozó (*KRAS*-mutant), őssejt-szerű (cancer stem cell), kromoszomális instabilitást mutató és a Wnt-útvonalban szereplő gének emelkedett expressziójúak (CIN-WntUp), valamint kromoszomális instabilitást mutató (CIN-normal) csoport [199].

5.1.2.11 *MDA114 – jó és rossz prognózisú csoportok*

A tanulóhalmaz betegeinek a gén chippel meghatározott transzkripció mintázatán hierarchikus klaszteringet alkalmazva két csoportot különítettek el. A két betegcsoportot t-tesztel összehasonlítva 114 szignifikánsan ($p < 0.001$) eltérően expresszáló gént azonosítottak. A HGU133Plus2.0 chip-ek esetében 114 gént használtam, míg a HGU133A chip esetében a két platformon megtalálható, az eredeti közleményben is alkalmazott 80 gént alkalmaztam. Az osztályozó tanítását az eredeti közleménnyel megegyező módon, a GSE17536 adathalmazon végeztem [200].

5.1.2.12 *Merlos-Suarez – őssejt génextpressziós mintázat*

A daganatban a szomatikus intesztinális őssejtek jelenléte, illetve az ezekre jellemző génextpressziós mintázat összefügg a túléléssel. Merlos-Suarez és munkacsoportja is ezt a jelenséget vizsgálva határozott meg a szomatikus intesztinális őssejtekre jellemző génextpressziós mintázatokat. Az EphB2 receptor felülexpresszáló sejtpopuláció esetében 29 génes, az Lgr5-t felülexpresszáló sejteknél 64 génes mintázatot írtak le, majd ezeket vizsgálták humán minták esetében. A mintázatban szereplő gének átlagát meghatározva a betegeket három egyenlő méretű csoportba sorolták (alacsony, közepes, magas expressziójú). Mindkét őssejtcsoportra jellemző mintázat prognosztikusnak bizonyul 2-es stádiumú vastagbél-daganatok esetében [201].

5.1.2.13 *Meta163 –Dukes B és C daganatok osztályozása*

A Dukes B és C stádiumú tumorokat vizsgálva meghatározhatunk a korai és a késői, áttétes vastagbél-daganathoz hasonló csoportokat. Ehhez Dukes A és D minták MAS5 normalizált génextpressziós profilját vetették össze SAM-mel, így 128 különbözően kifejeződő gént (163 próba szettet) azonosítottak. Az osztályozást PAM-mel végezték [202].

5.1.2.14 ODXcolon – 3 prognosztikai csoport

Az Oncotype DX teszt RT-PCR alapú, a stage II, III-as betegnél a kiújulás kockázatát hivatott előrejelezni. Az FFPE mintákból 7 rákkal összefüggő, és 5 háztartási gén expresszióját méri, melyek súlyozott összegéből határozza meg a „Kiújulási pontértéket” (Recurrence score, RS). Munkacsoportunk már a korábbiakban bizonyította, hogy a hasonló emlődaganatra specifikus Oncotype DX teszt jól reprodukálható gén chip vizsgálatokban [166]. Az átalakítás módja a következő: $\Delta Ct(\text{génX}) = -15 - (\log_2(\text{átlag}(\text{háztartási gének})) - \log_2(\text{génX}))$. Az RS pont értékek alapján a 30-nál kisebbel rendelkező tumorok alacsony, a 30-40 közöttiek közepes, a 40 felettiak magas kockázati csoportba tartoznak [203, 204]. Ezt a tesztet – mint a bevezetőben részleteztem – több retrospektív és egy prospektív vizsgálatban validálták 2-es és 3-as stádiumú betegeknél [205] [206].

5.1.2.15 Oncofender –kiújulás előrejelzése

Az Oncofender FFPE mintákra optimalizált 5 gént és 5 háztartási gént (B2M, GUSB, POLR2L, PSMB6, UBC) használó RT-PCR alapú teszt, mely a nyirokcsomó-negatív vastagbél-daganatok prognózisának az előrejelzését célozza. A teszt prognosztikus értékét 1-es és 2-es stádiumú vastagbélrák esetében validálták. Az RMA normalizációt, \log_2 transzformációt követően az öt gén expresszióját kivontuk a háztartási gének átlagos expressziójának átlagából. Az Oncofender pont = $\text{abs}(BMI1 \times VEGFA \div H3F3B) - \text{abs}(ETV6 \times H3F3B \div RPS10)$. Két csoport elkülönítésére nem az eredeti közleményben szereplő küszöbértéket használtam, hanem a pontértékek mediánját, mivel ez a vágópont jobban összefüggött a túléléssel [207].

5.1.2.16 Popovici – BRAF-mutáció mintázata

A prognosztikus biomarker jelölt BRAF-mutációval összefüggő mintázatot írtak le Popovici és mtsai [208]. A BRAF-mutáció prediktív markerként való meghatározását jelenleg csak klinikai vizsgálatokban használják, a korábbi eredmények arra mutatnak, hogy a BRAF-mutáns tumorok az EGFR-gátló terápiával (pl: cetuximab) szemben rezisztensek [209]. Ez az osztályozó modell két géncsoport (G1, G2) expressziójának átlagát hasonlítja össze mintánként. Azokat a mintákat, ahol

a G1 csoportokba tartozó gének expressziója alacsonyabb a G2-nél, BRAF-mutáns-szerű mintáknak nevezik.

5.1.2.17 Schetter – gyulladásoos válasz és prognózis

Schetter és mtsai gyulladásoos reakciókkal kapcsolatos mintázat azonosítását követően RT-PCR-alapú, prognosztikus tesztet fejlesztett: gyulladásoos rizikó érték (inflammatory risk score, IRS). A génextpressziók 18S RNS-hez való normalizálását követően két modell: „nem-daganatos rizikó modell” és „tumor rizikó modell” értékelése után azon betegek, akik mindkét teszten a mediánnál nagyobb értéket értek el magas IRS-ű betegként osztályozottak [210].

5.1.2.18 TCA19 – 3-as stádiumú betegek osztályozása

Egészséges vastagbél, primer colorectalis daganatok és máj áttétek RNAseq adatainak elemzése alapján Kim és mtsai 19 gén expresszióján alapuló rizikó score-t írt le [211]. Ezen kockázati pontszám betegenkénti meghatározása a gén expressziók Cox-regressziós együtthatóval súlyozott összege. A meghatározás során használt regressziós együtthatók a CIT kohorton (GSE39582) lettek meghatározva. A betegek a pontszám medián értéke alapján jó és rossz prognózisú csoportokra oszthatóak. Magas kockázati pontszámmal rendelkező 3-as stádiumú betegeknek szignifikánsan rosszabb a betegség-mentés túlélése, míg ez a 2-es stádiumú betegekre nem teljesült.

5.1.2.19 Yuen3 – 3 gén alapján 4 betegcsoport

Yuen és mtsai három gén (TAZ, AXL és CTGF) expresszióját kombinálta prognosztikus markerében. A három gén expresszióját vizsgálva a medián expressziós értékek alapján négy csoportba sorolták a betegeket, az alapján, hogy hány (0,1,2 vagy 3) gén expressziója magasabb a mediánnál. Akiknek mindhárom génben emelkedett expressziója, azok rendelkeznek a legrosszabb prognózissal [212].

5.1.2.20 V7RHS – magas, alacsony kockázatu csoportok

Jiang és mtsai gén chipen mért adatok elemzésével RT-PCR alapú 7-génes tesztet fejlesztett, mely prognosztikus értékét FFPE mintákon validálták stage II betegek esetében. A 7 gén ΔC_t értéke alapján betegenként meghatározták a „relapszus-rizikó pont”-nak nevezett értéket „relapse hazard score (RHS)”. A ΔC_t értékek gén chipre való konvertálása során az RMA normalizált, log₂ transzformált értékeket kivontuk az

eredeti vizsgálatban is használt három háztartási gén (ACTB, HMBS, RPL13A) átlagos expressziójából. A 0 RHS-nél nagyobb értékkel rendelkező betegeket tekintettem magas rizikójúnak [213].

5.1.2.21 Watanabe-MSI és Watanabe-CIN – MSI és CIN státusz előrejelzése

Ennek a munkacsoportnak nem az új alcsoportok azonosítása volt a célja, hanem a már jól ismert, prognosztikus mikroszatellita- és kromoszóma instabilitás génextpresszióra gyakorolt hatásának vizsgálata. (Az ismert, hogy az MSI-sal rendelkező tumoroknak jobb a prognózisa a MSS-ekhez képest.) A MSI és az MSS tumorokat gén chipen összehasonlítva 177 eltérően kifejeződő próba szettet azonosítottak, mely alkalmas volt az MSI státuszt nagy hatékonysággal megállapítani. Vizsgálatomban az osztályozó reprodukálása során a Watanabe és mtsai által alkalmazott lépéseket követtem, a kNN osztályozót is ugyanazon az adathalmazon (GSE5445) tanítottam be. Hasonló elemzést követve a magas és az alacsony kromoszomális instabilitással rendelkező mintákat összevetve 112 gén alapú osztályozót hoztak létre. Az így meghatározott csoportok szignifikánsan összefüggtek a 2-es, 3-as stádiumú betegek esetében a progressziómentes túléléssel [214].

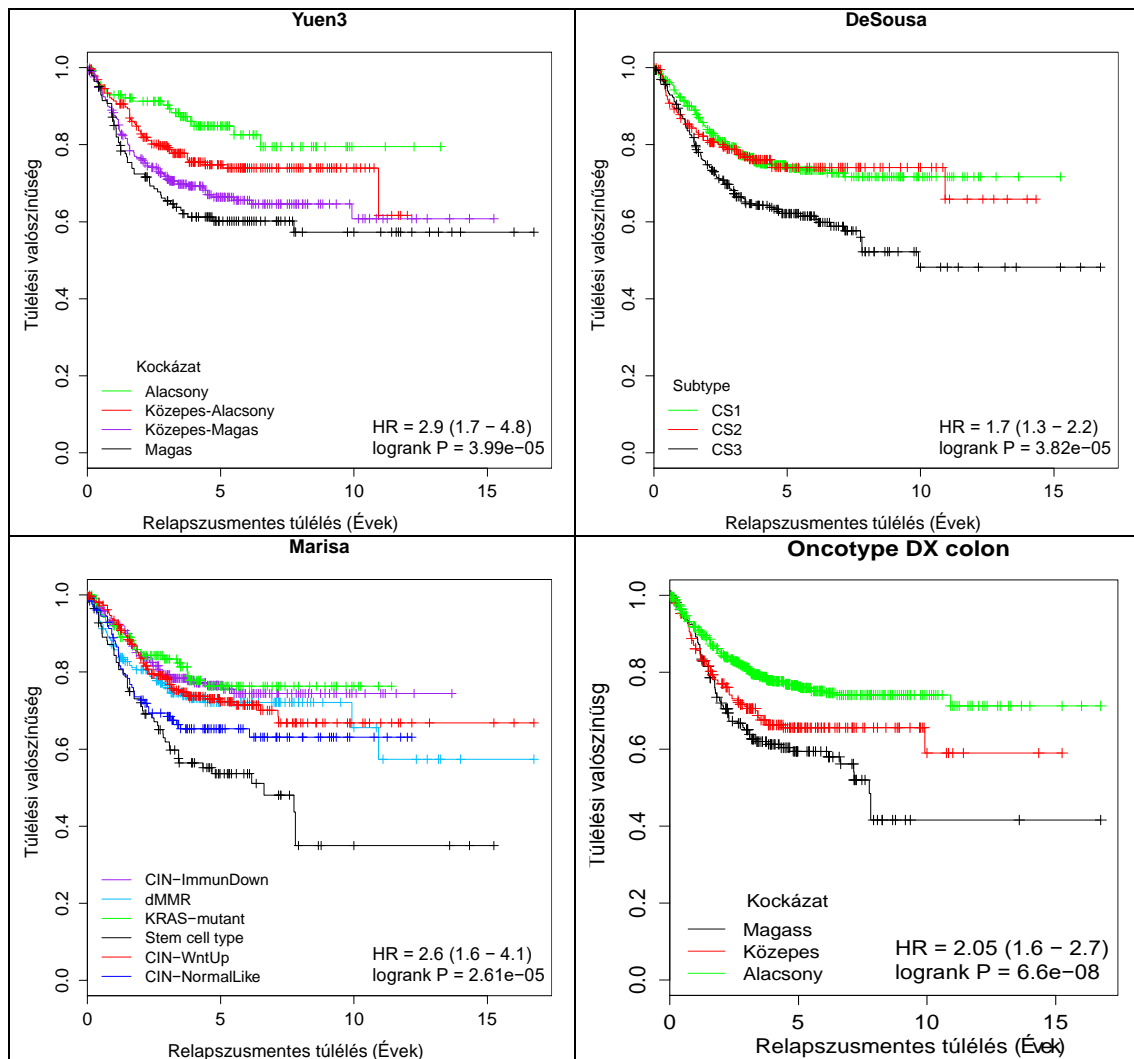
A 11. táblázat a vizsgált osztályozók tulajdonságait foglalja össze.

11. táblázat. Az osztályozók legfontosabb tulajdonságainak összehasonlítása

Teszt				Osztályozás módja				1.vali dáció
Név	Hivatkozás	FF / FP	stádium	csoportok száma	osztályozás módja	gének száma	Eredeti platform	minták száma
Budinska	[187]	FF, FP	II-III	5	lineáris diszkriminancia-analízis	658	ALMAC, U133+2, U133A	720
CCHS	[188]	FF, FP	II-III	2	súlyozott összeg, meghatározott határérték	6	U133+2	126
Chang95	[189]	FF	I-III	3	átlagos expresszió mediánja	95	ALMAC, U133+2, U133A, HuEx-1_0-st	682
CIN25	[190]	FF	-	2	átlagos expresszió mediánja	25	U133+2, U133A	NA
CMS	[191]	FF, FP	-	4	középpont-alapú	693	Agilent, RNA-seq, Affymetrix	1721
ColoGuideEx	[193]	FF	II	2	80. percentilis feletti/20.percentil is alatta gének számae	13	HuEx-1_0-st	203
ColoGuidePro	[194]	FF	II-III	2		7	HuEx-1_0-st	290
CRCassinger-786	[197]	FF	-	5	PAM	786	U133+2	744
DeSousa	[215]	FF	-	3	PAM	146	U133+2	1074
Marisa	[199]	FF	-	6	középpont-alapú	57	U133+2	1181
MDA114	[200]	FF	II-III	2	súlyozott összeg	114	U133+2	213
Merlos-EphB2-ISC	[201]	FF	-	3	átlagos expresszió	29	Affymetrix mouse4302	340
Merlos-Lgr5-ISC	[201]	FF	-	3	átlagos expresszió	64	Affymetrix mouse4302	340
Meta163	[202]	FF	II-III	2	PAM	128	U133+2	99
ODXcolon	[204]	FP	II- III	3	súlyozott összeg	12	RT-PCR	711
Oncodefender	[207]	FP	I-II	2	multiplication of signature genes	5	RT-PCR	264
Popovici	[208]	FP, FF	II-III	2	closest mean expression	64	ALMAC	114
Schetter	[210]	FF	-	2	súlyozott összeg	9	RT-PCR	139
TCA19	[211]	FF	III	2	súlyozott összeg	19	U133+2	449
Yuen3	[212]	FF	II-III	4	génenkénti medián	3	U133+2	-
V7RHS	[213]	FP	II	2	súlyozott összeg	7	RT-PCR	-
Watanabe-CIN	[214]	FF	II-III	2	SVM	112	U133+2	290

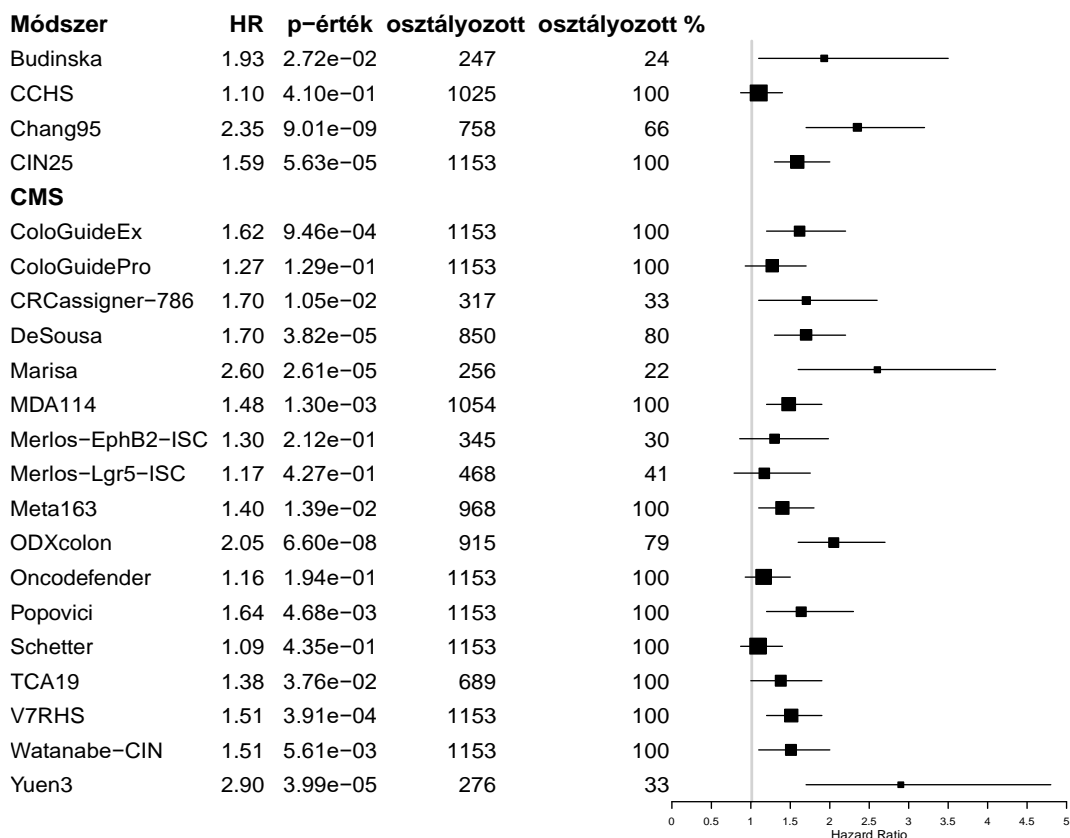
5.1.3 Osztályozók összehasonlítása

Az eredmények összehasonlítása során több módszert alkalmaztam. Cox-regresszióval meghatároztam a legjobb és a legrosszabb prognózisú betegcsoportok közötti prognózisbeli eltérést, házard ratio-t. Az osztályozók eredményeit Kaplan-Meier görbéken ábrázoltam. A 16. ábra négy teszt eredményeit mutatja.

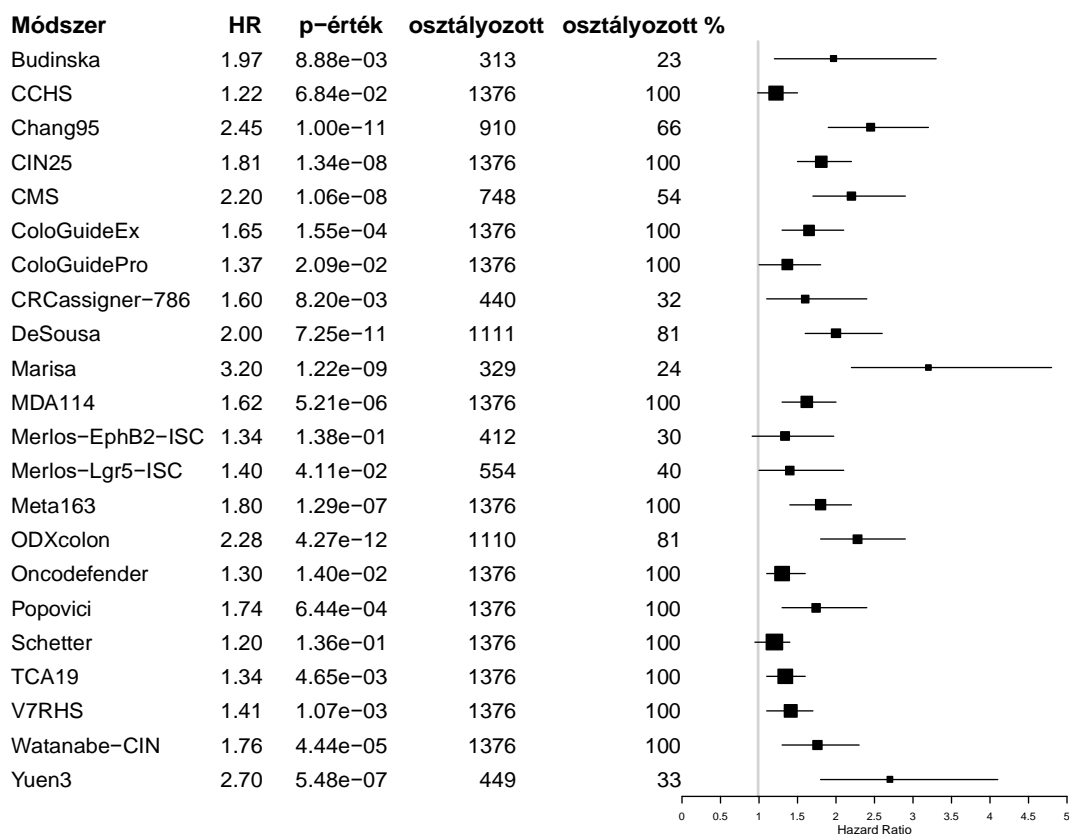


16. ábra. Betegek túlélése alcsoportonként (A 2-es és 3-as stádiumú betegekre korlátozva, az eredeti tanulóhalmazok kizárásával.)

A Módszerek fejezetben leírtak alapján, az osztályozók befolyásolatlan összehasonlítása érdekében kizártuk mindegyik módszernél azokat a betegeket, amin a teszt eredetileg kifejlesztésre került (tanulóhalmazuk). A minden betegre vonatkozó eredményt a 17. ábra, a csak a 2-es és 3-as stádiumúakra vonatkozókat pedig a 18. ábra szemlélteti. Az osztályozott betegek száma tehát azért lehet kisebb, mint az egész adatbázisunk, mivel kizártam a tanulóhalmazt. Az osztályozott betegek százaléka



17. ábra. Minden beteg osztályozása

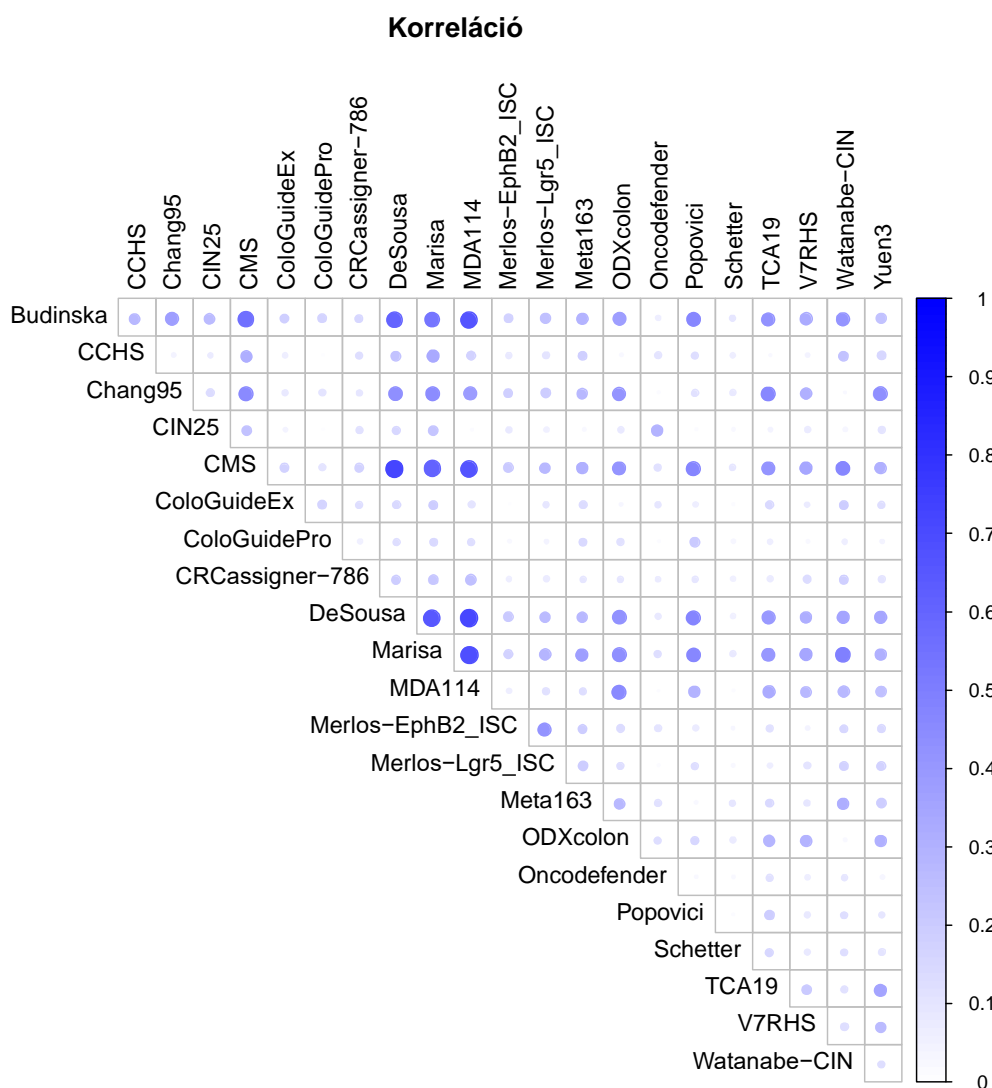


18. ábra. 2-es és 3-as stádiumú betegek osztályozása

azt fejezi ki, hogy a nem kizárt betegek közül hány százalék került a legrosszabb és a legjobb prognózisú csoportba összesen.

A 2-es és 3-as stádiumú betegek esetében a Yuen és munkatársai által leírt, 3 gén expresszióján alapuló osztályozó bír a legnagyobb prognosztikai erővel (HR=2,9). Ezt követi Marisa osztályozója (HR=2,60), mely ha minden stádiumú beteget vizsgálunk, a legjobban teljesített (HR=3,20). A harmadik leghatékonyabb módszer a Chang95 (HR=2,35).

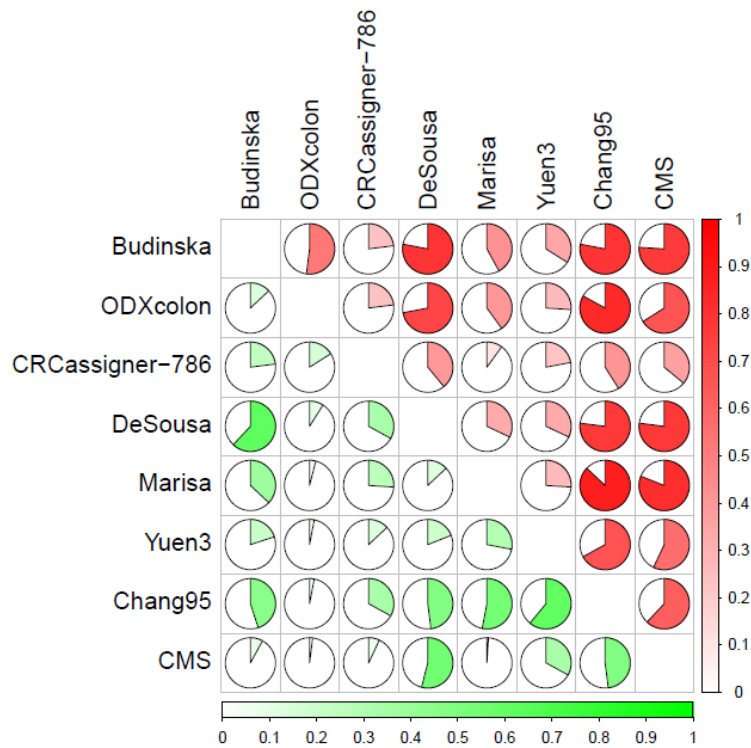
Az osztályozók csoportjai közötti átfedést a 19. ábra mutatja. Jól látszik, hogy a CMS eredményei jól korrelálnak az általa beépített osztályozókkal, a DeSousa féle osztályozás eredményével az asszociáció mértéke 0.72. Fontos azonban azt is



19. ábra. Korreláció az osztályozók között

megjegyezni, hogy a kereskedelmi forgalomban kapható tesztek egy része elég gyenge összefüggést mutat. Például az Oncotype DX és a ColoGuideEx között az asszociáció mértéke csupán 0,03. A Cramer-V teszt eredményei a legjobb tíz osztályozó esetében a Függelék 1-ben található.

Az osztályozók közötti összefüggést szemlélteti a 20. ábra is, mely azt mutatja meg, hogy mennyire fednek át a jó (zöld) és rossz (piros) prognózisúak előrejelzett betegeket a legjobb 8 osztályozóban, a 2-es és 3-as stádiumú betegeknél. Megállapítható, hogy a rossz prognózisúként osztályozott betegek között magasabb az átlagos átfedés az osztályozók között.



20. ábra. Korreláció a jó és rossz prognózisúak között. A körök összterülete az egy pár osztályozó által jó/rossz prognózisúak osztályozott betegeket összesített számát fejezi ki. Minél nagyobb része színes a körnek, annál nagyobb az átfedés a csoportok között.

5.1.4 Génlisták összehasonlítása

Az osztályozók által alkalmazott géneket összehasonlítva azt találjuk, hogy a 22 osztályozó összesen 2001 gént használt. Csupán 5 olyan gén volt (REG4, ASCL2, VAV3, C10orf99 és CYPB1), ami 6 osztályozóban is előfordult. A géneket egyenként,

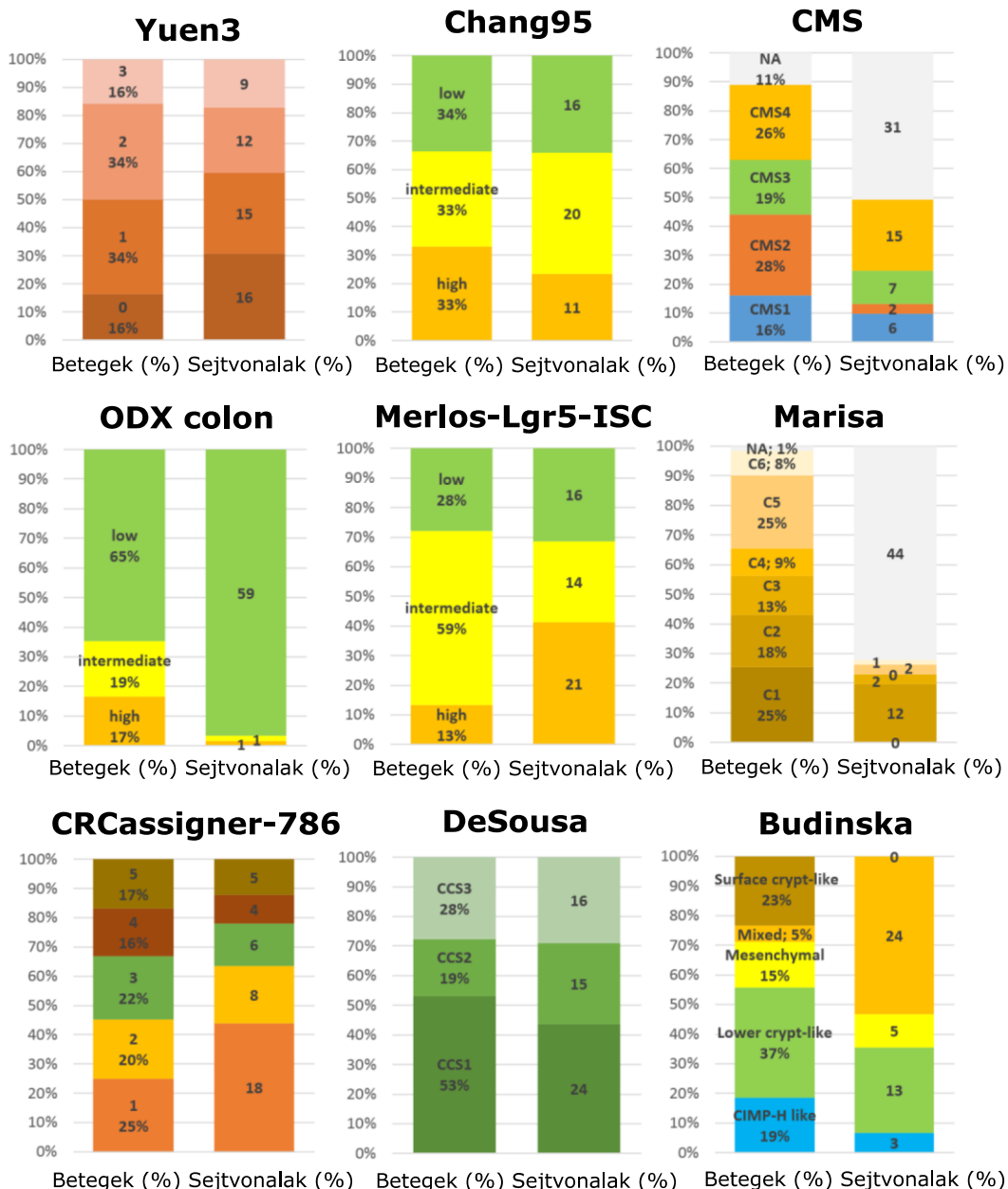
univariáns elemzéssel is vizsgáltam, illetve meghatároztam a „gén-pontszámukat” az osztályozókban alkalmazott gének számát figyelembe véve. A legfontosabb géneket a 12. táblázat tartalmazza.

12. táblázat. A legfontosabb gének

Gén	p-érték (univariáns)	osztályozók száma	arány - pontszám	gén-pontszám
CTGF	0	4	0,39	1,56
GADD45B	0,00007	4	0,35	1,38
FAP	0	5	0,11	0,56
INHBA	0,00431	5	0,1	0,51
ANXA1	0,00017	4	0,13	0,5
SEMA3A	0,08568	2	0,22	0,44
CXCL9	0,00002	3	0,15	0,44
VEGFA	0,70597	2	0,21	0,42
BGN	0,00021	4	0,1	0,42
BNIP3L	0,59288	2	0,2	0,4
ENPP3	0,00645	4	0,09	0,36
HLA-DRA	0,30654	3	0,11	0,34
TAZ	0,18042	1	0,33	0,33
AXL	0	1	0,33	0,33
CXCL10	0,00131	4	0,08	0,32
PTPRO	0,12132	5	0,06	0,32

5.1.5 Preklinikai modellek

Összesen 61 sejtvonal 151 génexpressziós vizsgálatát azonosítottam; 51 származik a CCLE-ből (Cancer Cell Line Encyclopedia, GSE36133), 60 a Cancer Cell Line Project-ből (E-MTAB-37), 15 a GSE8332 és 21 a GSE32474 adathalmazból. Ezen minták között nem volt többszörösen publikált minta. A sejtvonalakat génexpressziójuk alapján a reprodukált osztályozókkal alcsoportokba soroltam. A legfontosabb eredményeket a Függelék 1-ben csatoltam. Ez lehetőséget ad további kutatások során az éppen vizsgálni kívánt alcsoportnak megfelelő sejtvonal kiválasztására. A daganatminták osztályozásával összevetve jól látszik, hogy vannak olyan osztályozók, melyek hasonló számban sorolják csoportokba a sejtvonalakat, míg mások nem képesek rá (21. ábra).



21. ábra. Alcsoportok előfordulása sejtvonalakban

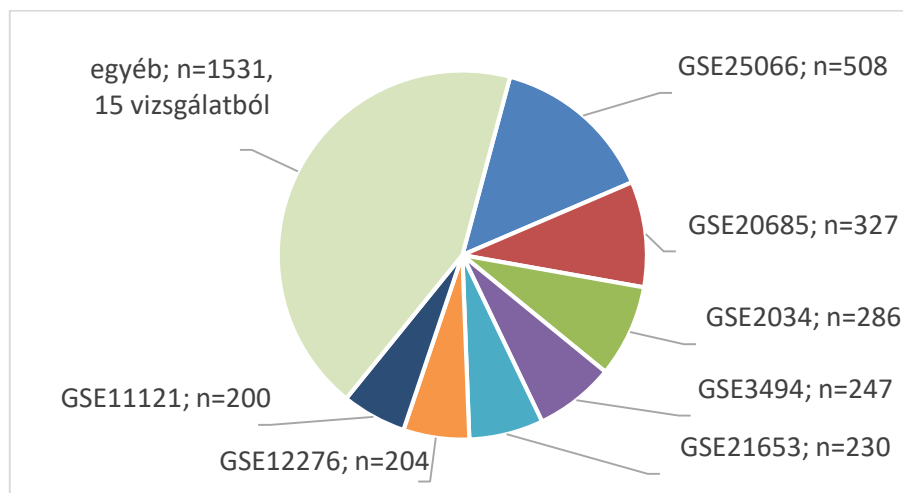
5.2 Emlőrák – rossz prognózisú betegek azonosítása

5.2.1 Mintagyűjtés – Független validációs adat

A független validálás céljából gyűjtött 325 emlődaganat minta esetében a betegek átlagos utánkövetési ideje 66 hónap volt. Ezen időszak alatt 97 esemény (relapszus vagy haláleset) történt. A betegek 81,1%-a (n=206) ER-pozitív volt, 39,4%-ának volt nyirokcsomó érintettsége. A tumorok többsége (n=186) grade 2-es differenciáltságú volt.

5.2.2 Adatbázis

A GEO adatbázisban végzett keresésem eredménye alapján 22 adathalmazban (GSE1456, GSE4922, GSE5327, GSE6532, GSE7390, GSE9195, GSE11121, GSE12093, GSE12276, GSE2034, GSE16391, GSE16446, GSE17705, GSE17907, GSE19615, GSE2603, GSE20685, GSE20711, GSE21653, GSE25066, GSE2990, GSE31519 és GSE3494) n=3534 olyan mintát azonosítottam, melyhez elérhető volt a nyers gén-expressziós adat és a relapszusmentes túlélési idő (relapse free survival, RFS). Adatbázisunk egységesítése érdekében relapszusmentes túlélésnek tekintettük azt a túlélési időt, ami a diagnózis idejétől a legkorábbi eseményig (relapszus vagy progresszió) eltelt. A legtöbb beteg (n=508) a GSE25066 adathalmazból származik, ugyanakkor a betegek csaknem fele, 43%-a olyan vizsgálatból származik, amiben 200-nál kevesebb minta szerepel. Az adatbázisunk összetételét a 22. ábra szemlélteti, a betegek klinikai adatait 13. táblázat foglalja össze.

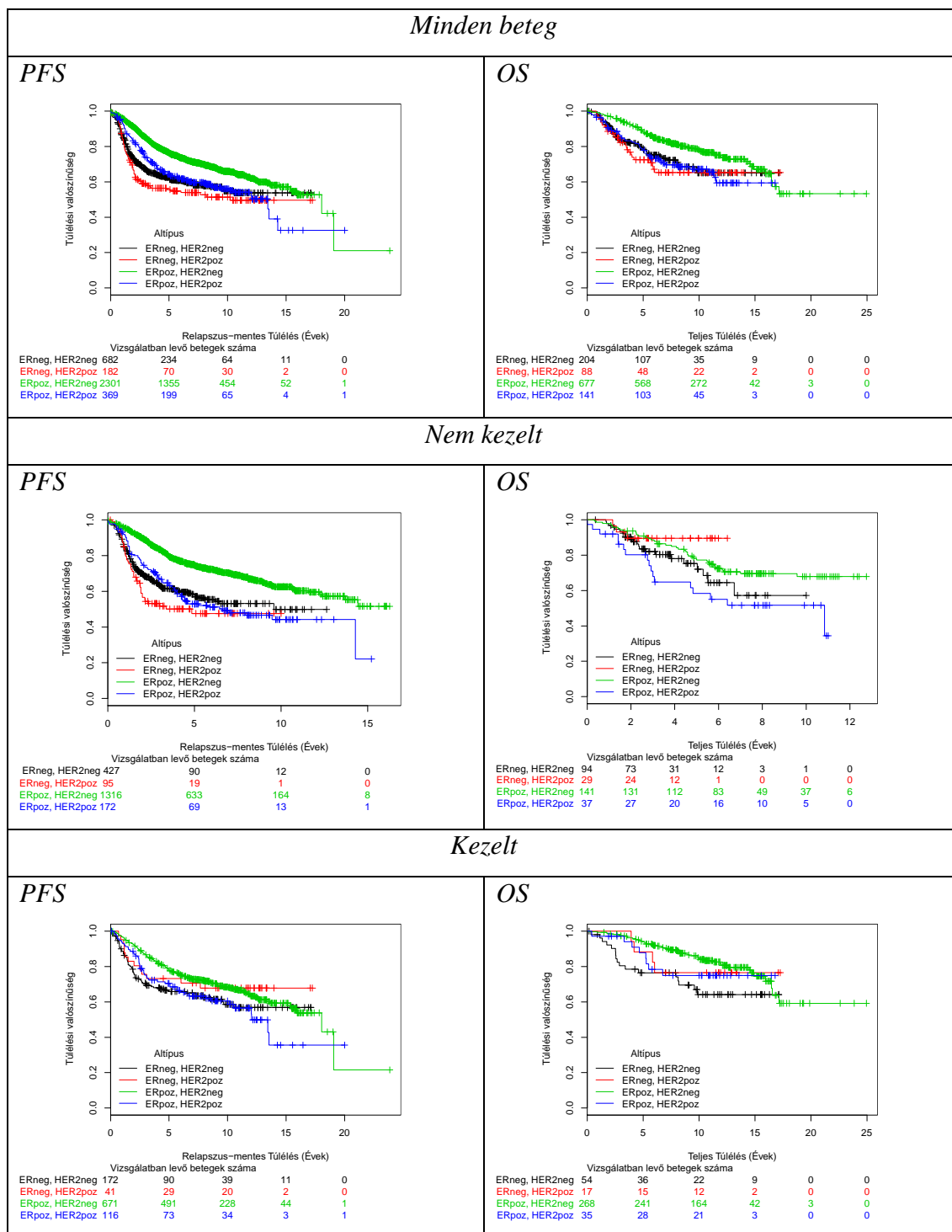


22. ábra. Az adatbázis összetétele

13. táblázat. Az adatbázis betegeink klinikai jellemzői

	Minden beteg	HER2–			HER2+
		ER+		ER–	(ER+ és ER–)
		Nem történt szisztémás kezelés	adjuváns kezelés	adjuváns kezelés	adjuváns kezelés
betegszám	3534	672	1316	427	551
Medián kor (évek)	53,2	55,2	55,5	49,9	51,5
Medián tumorméret (cm)	2,3	2	2,49	2	2,35
Medián RFS (évek)	5,85	7,85	5,42	3,44	5,51
Szövetteni grade					
1	329	143	132	10	17
2	842	306	355	45	97
3	964	78	297	271	177
Ismeretlen	1349	144	501	101	260
Nyirokcsomó státusz					
Pozitív	992	3	564	195	147
Negatív	2906	672	850	221	379
Ismeretlen	314	0	233	103	86

Adatbázisunkban a relapszusmentes túlélést vizsgálva, Kaplan-Meier görbékkel ábrázolva jól látszik az a jól ismert jelenség, hogy a HER2-es alcsoport (ER-negatív, HER2-pozitív) az adjuváns terápiában nem részesült betegek esetében a legrosszabb prognózisú, azonban a HER2-ellenes terápia hatására jobb prognózisú betegcsoporttá válik. A kezelt betegek esetében az 5 éves túlélést tekintve a tripla-negatív esetek a legrosszabb prognózisúak. (23. ábra)



23. ábra. Relapszusmentes túlélés klinikai alcsoportokban

5.2.3 Optimalizáció

Az optimalizáció során a kiválasztott hasonló betegek számát, és a molekuláris osztályozás során használt gének számát vettem figyelembe. A log-rank teszt χ^2 értékeit

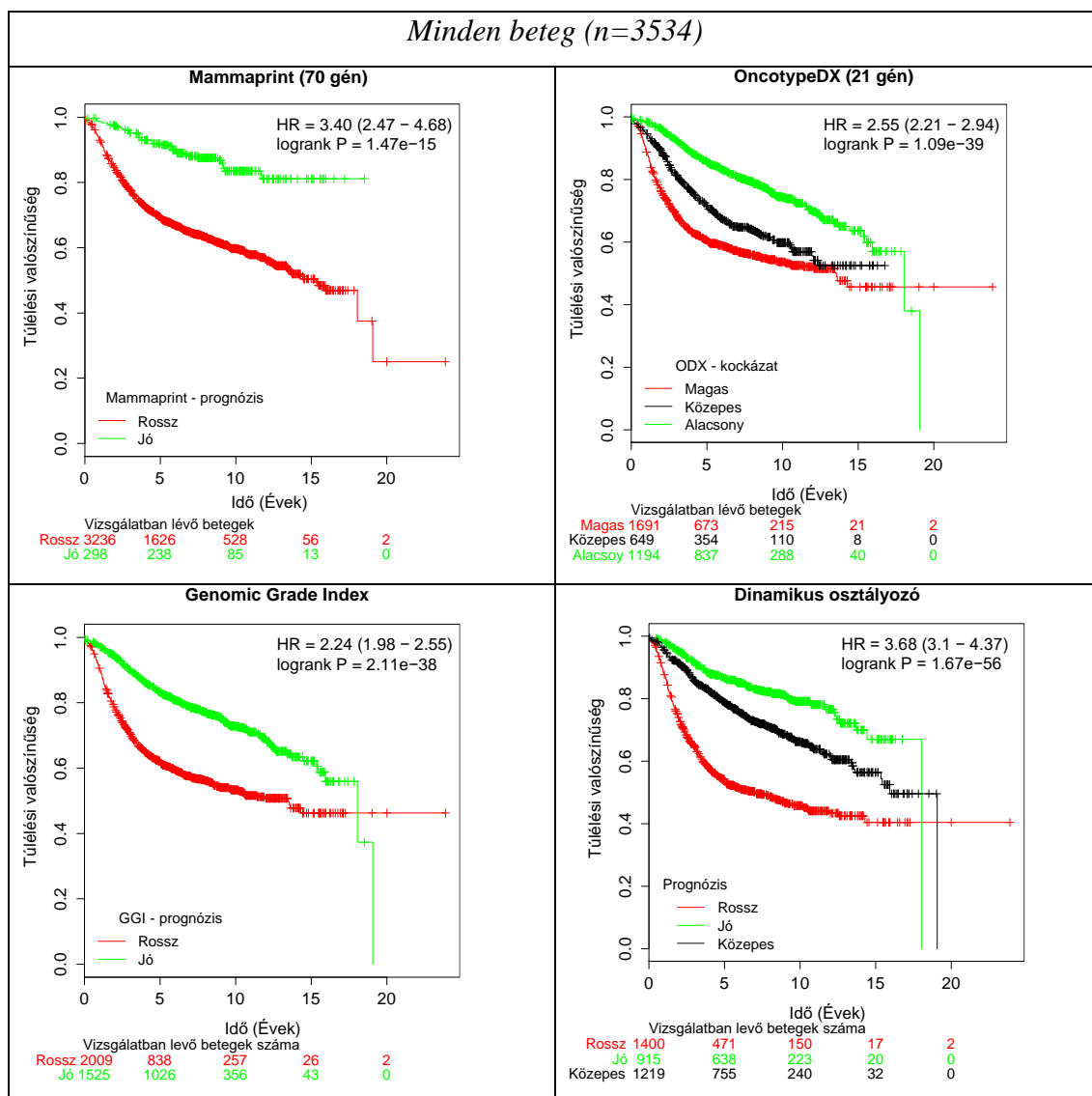
alapul véve a tanulóhalmaz méretét 400 főben, a Cox-regresszióban a kiválasztott legjobb gének számát 25-ben határoztam meg (14. táblázat).

14. táblázat. Az optimalizálás eredménye, a vastag fekete keret jelzi a leghatásosabb paramétereket

Tanulóhalmaz (betegek száma)		Osztályozás során használt, kiválasztott gének száma						
		3	5	10	25	50	100	200
100	chi2	236	232	226	231	287	303	283
	p-érték	4.66E-52	3.33E-51	9.34E-50	5.53E-51	4.26E-63	1.75E-66	3.67E-62
	HR	2.55	2.45	2.48	2.54	3.11	3.07	3.05
200	chi2	284	287	306	298	297	294	315
	p-érték	2.30E-62	4.18E-63	4.48E-67	1.58E-65	2.80E-65	1.29E-64	4.57E-69
	HR	3.03	2.98	3.18	3.18	3.36	3.45	3.68
300	chi2	281	284	300	288	292	269	293
	p-érték	1.01E-61	2.57E-62	8.89E-66	2.98E-63	3.57E-64	3.94E-59	2.23E-64
	HR	2.9	2.9	2.91	3.12	3.35	3.13	3.48
400	chi2	317	303	322	325	310	294	306
	p-érték	1.77E-69	1.84E-66	1.54E-70	2.20E-71	5.78E-68	1.63E-64	4.53E-67
	HR	3.17	3.19	3.52	3.68	3.51	3.54	3.6
500	chi2	302	297	302	313	302	288	290
	p-érték	2.47E-66	3.69E-65	2.46E-66	1.14E-68	2.13E-66	2.93E-63	9.97E-64
	HR	3.18	3.15	3.24	3.74	3.73	3.56	3.7

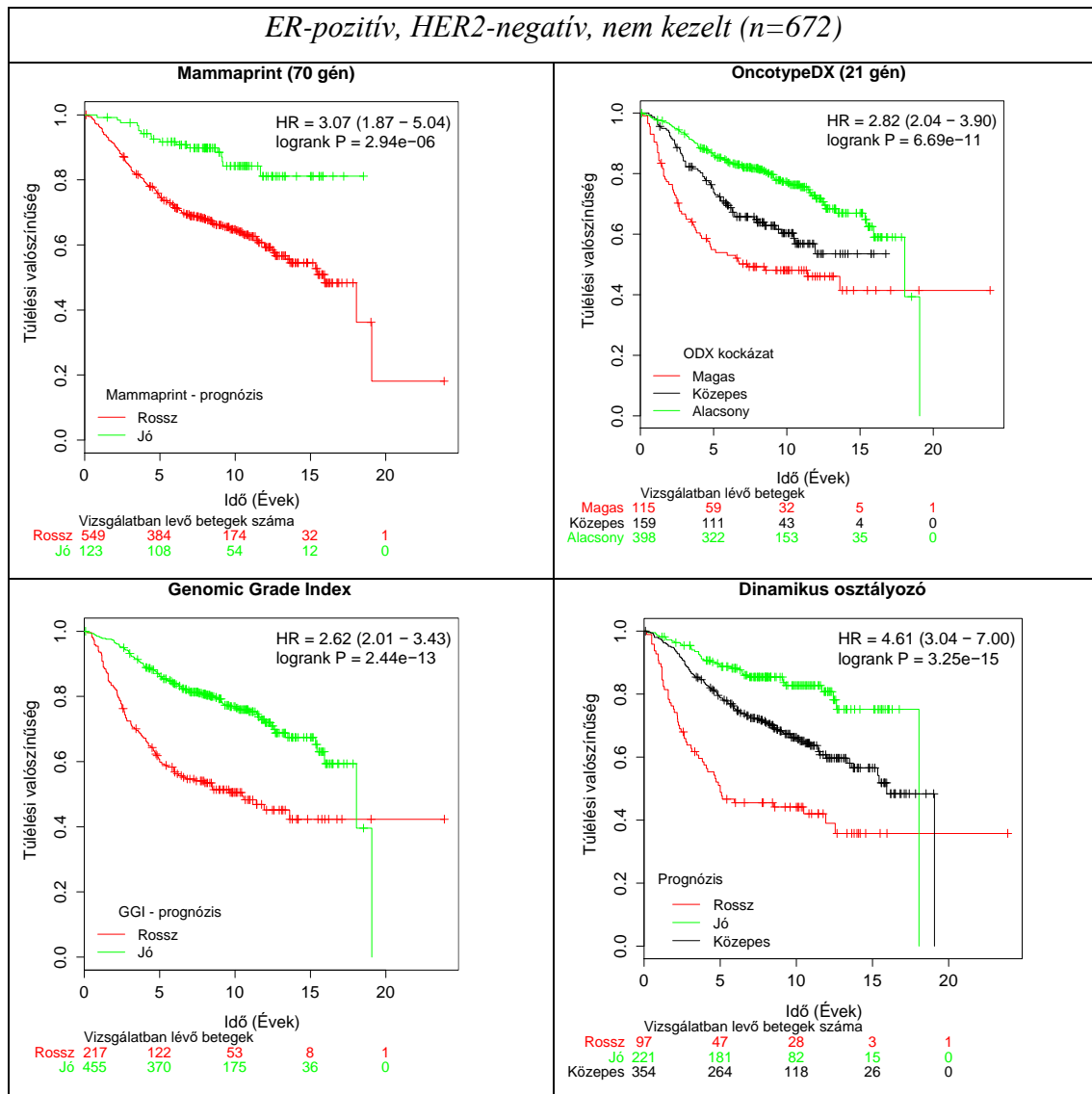
5.2.4 Összehasonlítás

A 70-génes Mammaprint, 21-génes Oncotype DX, és a 97-génes Genomic Grade Index összehasonlítását elvégezve minden beteg, illetve klinikailag releváns alcsoportok esetében láthatjuk, hogy minden esetben az új Dinamikus osztályozónk volt a legjobb. A részletes eredményeket a következő ábrákon mutatom be.



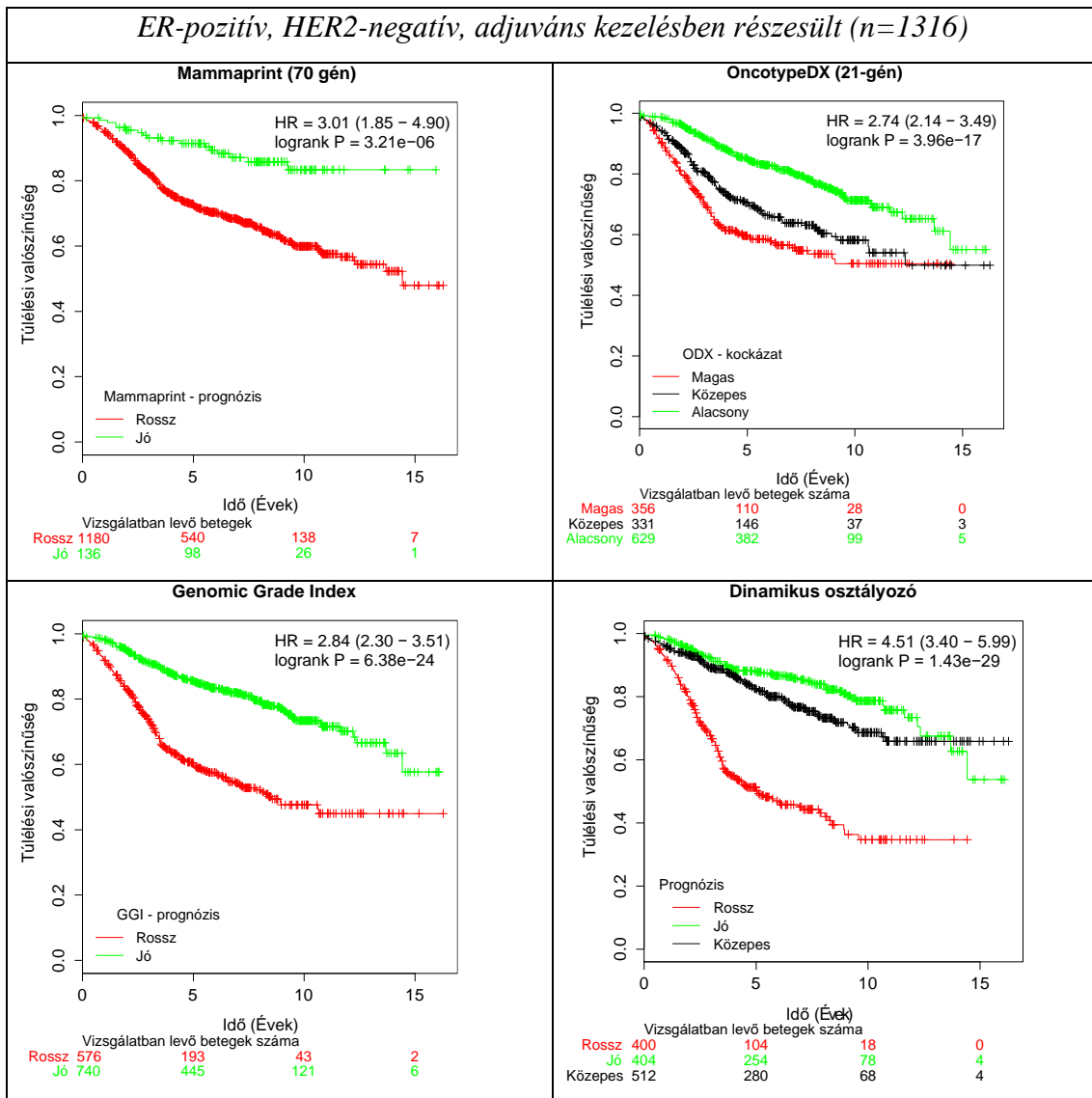
24. ábra. Minden beteg osztályozásának eredménye, relapszusmentes túlélése

Minden beteget vizsgálva a Dinamikus osztályozónk volt a legjobb (HR=3,68) a legrosszabb és a legjobb prognózisú csoportokat összevetve, továbbá fontos azt is kiemelni, hogy a betegek csupán 40%-át soroltuk a rossz prognózisú betegek közé. (24. ábra)



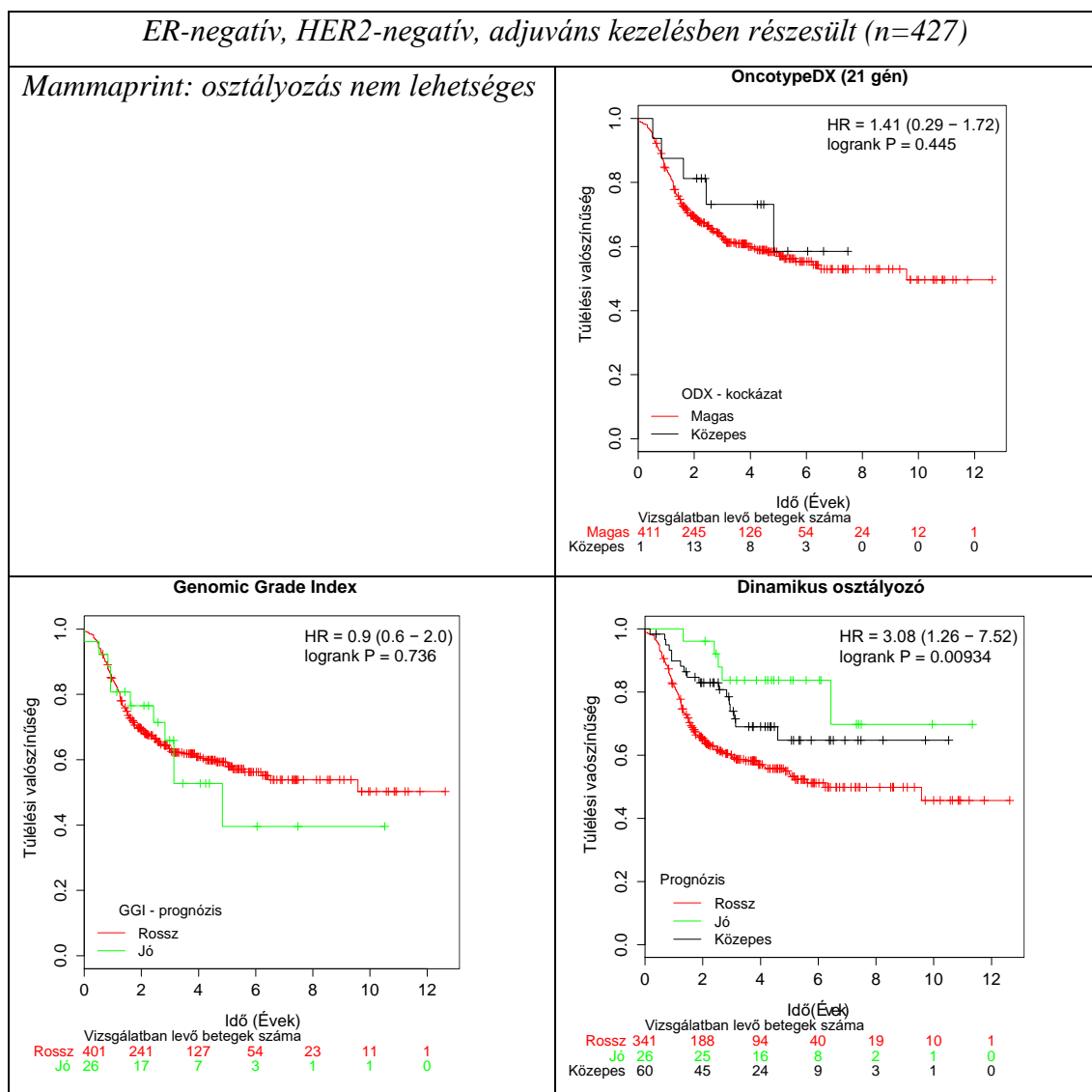
25. ábra. ER-pozitív, HER2-negatív, nem kezelt betegek relapszumentes túlélése

Az ER-pozitív, HER2-negatív, nem kezelt betegek esetében is a Dinamikus osztályozónk bizonyolult a leghatékonyabbnak (HR=4,61). Ebben az esetben is a mi osztályozónk sorolta a legkevesebb beteget a rossz prognózisú csoportba. (25. ábra)



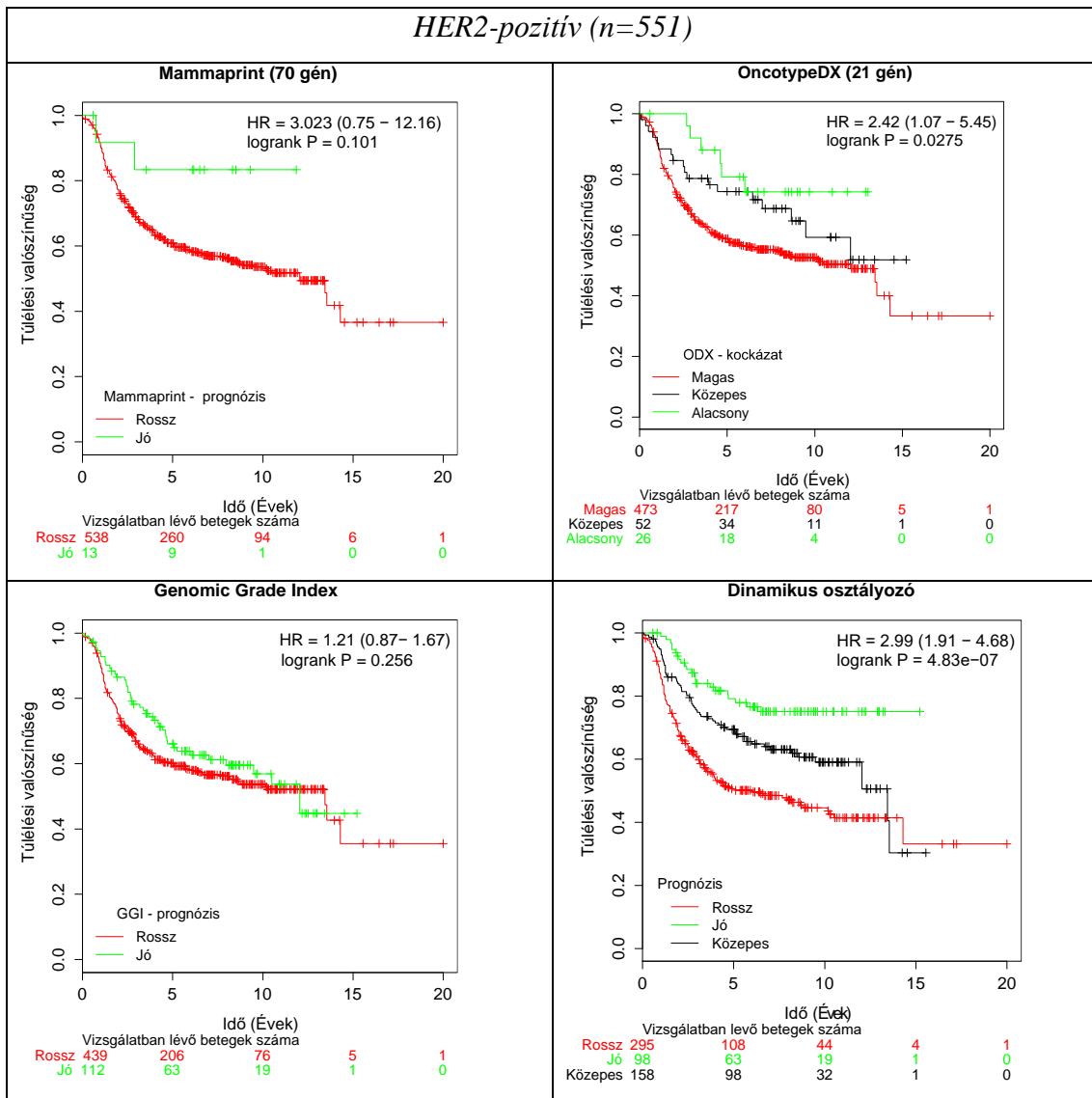
26. ábra. ER-pozitív, HER2-negatív, adjuváns kezelésben részesült betegek relapszusmentes túlélése

Az ER-pozitív, HER2-negatív, adjuváns kezelésben részesült betegek között is a Dinamikus osztályozó volt a leghatékonyabb (HR=4,51) (26. ábra)



27. ábra. ER-negatív, HER2-negatív, adjuváns kezelésben részesült betegek túlélése

Az ER-negatív, HER2-negatív, adjuváns kezelésben részesült betegek osztályozását a többi teszt nem tudta megoldani, míg a Dinamikus osztályozó ebben a betegcsoportban is el tudott különíteni jó és rossz prognózisú populációt (HR=3,0) (27. ábra).



28. ábra. HER2-pozitív betegek osztályozása

A HER2-pozitív betegeket vizsgálva is csak a Dinamikus osztályozónk volt képes hatékonyan (HR=2,99) további prognosztikus csoportokra bontani (28. ábra).

Az 5 éves relapszusmentes túlélést végpontnak tekintve a szenzitivitást, specificitást, pontosságot, pozitív és negatív prediktív értékeket vetettük össze a tesztek további értékelése érdekében (15. táblázat). Bár a 70 génes teszt a legérzékenyebb, specificitása nagyon alacsony. Specificitása, és pozitív prediktív értéke a Dinamikus osztályozónak a legmagasabb, miközben a negatív prediktív érték, és a szenzitivitás elfogadható marad.

15. táblázat. Az osztályozók jellemzői minden beteget (n=3534) vizsgálva (95%-os konfidencia intervallumok)

	Dinamikus osztályozónk	21 génes teszt	70 génes teszt	97 génes teszt
Szenzitivitás	0,84 (0,81-0,86)	0,80 (0,76-0,82)	0,98 (0,96-0,98)	0,81 (0,78-0,84)
Specificitás	0,58 (0,55-0,61)	0,55 (0,53-0,58)	0,14 (0,12-0,16)	0,49 (0,46-0,52)
Pontosság	0,68 (0,66-0,70)	0,64 (0,62-0,65)	0,47 (0,46-0,47)	0,61 (0,59-0,63)
PPV	0,56 (0,54-0,58)	0,53 (0,52-0,55)	0,42 (0,42-0,43)	0,51 (0,49-0,52)
NPV	0,85 (0,82-0,87)	0,81 (0,79-0,83)	0,92 (0,86-0,95)	0,80 (0,77-0,83)

A korábbi prognosztikus tesztek eredményeit összevetve a Dinamikus osztályozónkkal láthatjuk, hogy az egyezés legfeljebb 55%-os (16. táblázat).

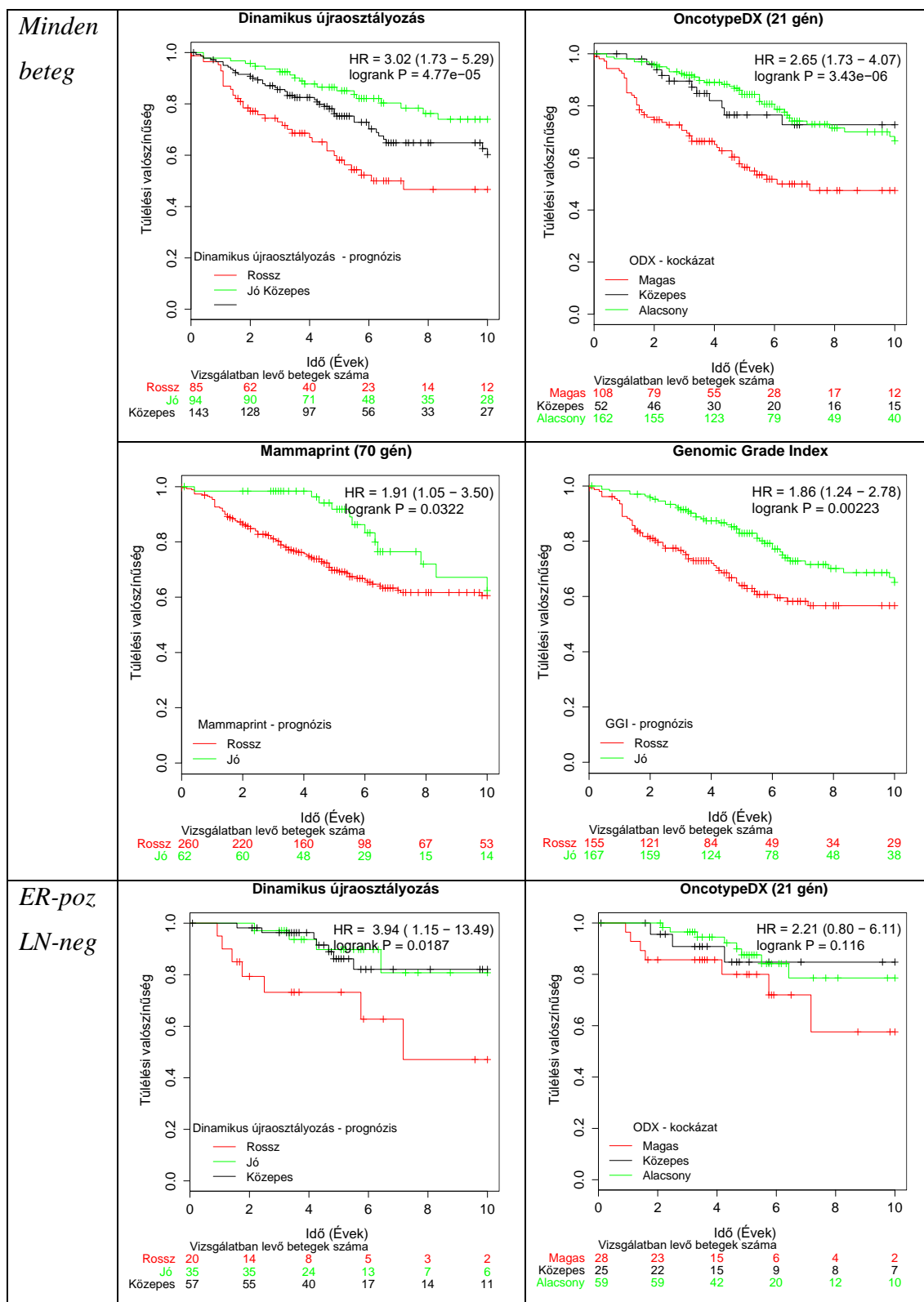
16. táblázat. A Dinamikus osztályozónk és más tesztek közötti korreláció, minden beteg esetében

Dinamikus Osztályozónk	21 génes teszt			97 génes teszt		70 génes teszt	
	Jó	Közepes	Rossz	Jó	Rossz	Jó	Rossz
Jó	527	174	214	651	264	181	734
Közepes	567	295	357	685	534	110	1109
Rossz	100	180	1120	189	1211	7	1393
Egyezés	55.0%			52.7%		44.5%	

5.2.5 Validálás

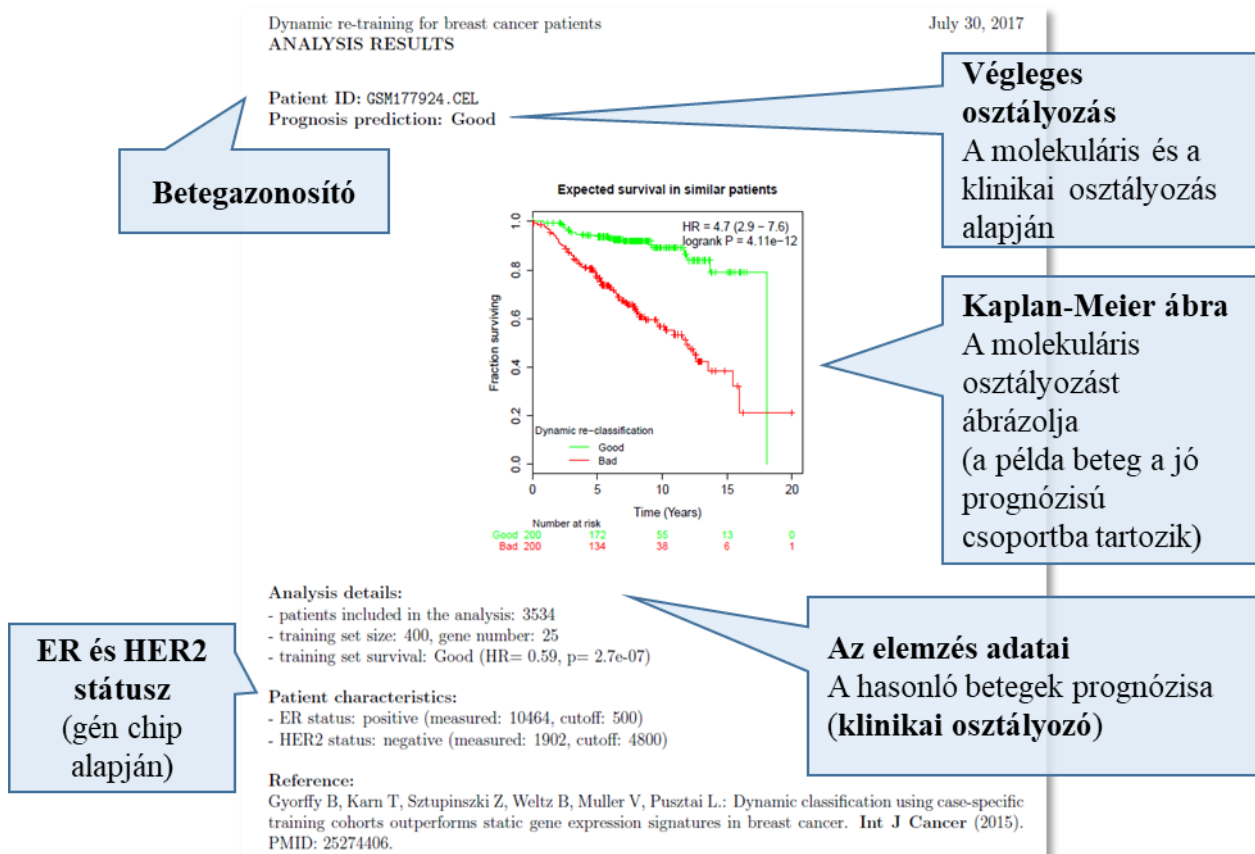
A 325 független mintán elvégzett validálás esetében a korlátozott betegszám miatt csak két csoportosításban (minden beteg, illetve ER-pozitív, nyirokcsomó negatív betegcsoport) vizsgáltuk a tesztek hatékonyságát (29. ábra). Minden beteget vizsgálva a Dinamikus osztályozónk volt a legjobb (HR=3,02), ER-pozitív, nyirokcsomó negatív betegek esetében csak a Dinamikus (HR=3,94) és a 21 génes teszt (HR=2,21) volt képes

osztályozásra, azonban csak a Dinamikus osztályozás eredménye volt szignifikáns (29. ábra).



29. ábra. A független validálás eredménye

A dinamikus osztályozót online felületen tettük elérhetővé (http://www.recurrenceonline.com/?q=Re_training), ahol egy gén chip vizsgálat eredményét feltöltve (.CEL fájl) a rendszer elvégzi a minta osztályozását, és egy jelentést készít (30. ábra).

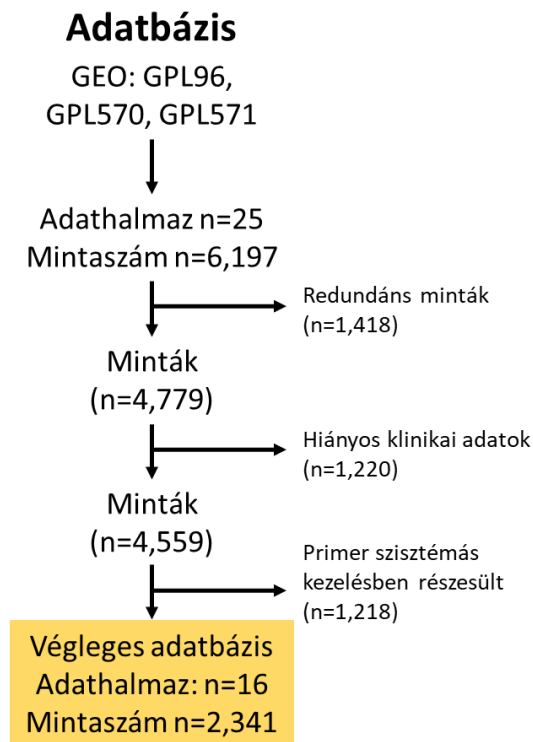


30. ábra. Egy jelentés bemutatása

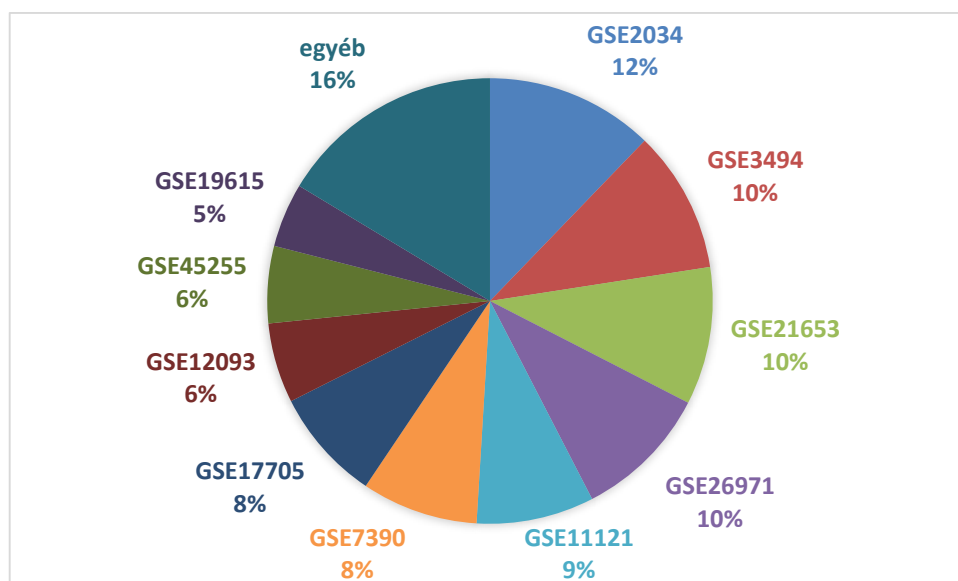
5.3 Emlőrák - nyirokcsomó érintettség előrejelzése

5.3.1 Mintagyűjtés és adatbázisunk

A GEO adatbázisban 16 adathalmazban 2341 beteget azonosítottam (32. ábra), a betegek 21%-volt nyirokcsomó-pozitív, 16 %- ER-negatív. Az adatbázisépítés lépéseit a 31. ábra foglalja össze.



31. ábra. Az adatbázisépítés részletei



32. ábra. Az adatbázis összetétele

A mintagyűjtés során validációs csoportnak 100 beteg adatait és mintáit gyűjtöttük össze. A betegek klinikai jellemzőit a 17. táblázat foglalja össze.

17. táblázat. A vizsgált betegek klinikai adatai

	Adatbázisunk			p-érték	Validációs halmaz			p-érték
	összes beteg	LN			összes beteg	LN		
		Negatív	Pozitív			Negatív	Pozitív	
betegszám	2341	1724	617		97	50	47	
Kor (évek)				0,0065				0,048
≤50	458	324	134		18	5	13	
>50	789	497	292		79	45	34	
Ismeretlen	1094	903	191		0	0	0	
Tumor átmérő (mm)				<0,0001				0,02
≤20	683	573	110		60	37	23	
>20	718	472	246		37	13	24	
Ismeretlen	940	679	261		0	0	0	
Szövetteni				0,0001				0,81
1	260	201	59		24	14	10	
2	628	458	170		34	17	17	
3	537	344	193		33	17	16	
ER				0,026				0,2065
Pozitív	1978	1439	539		74	35	39	
Negatív	363	285	78		23	15	8	
HER2				0,038				1
Pozitív	341	235	106		87	45	42	
Negatív	2000	1489	511		10	5	5	
KI67				0,07				0,8727
Pozitív	739	517	222		76	40	36	
Negatív	1602	1207	395		21	10	11	

5.3.2 Nyirokcsomó-státusz előrejelzése

Az előrejelzés hatékonyságát vizsgáltam a belső validációs halmazon, és a független validáció 100 betegén a három, előzőekben definiált betegcsoport (ER-negatív, ER-pozitív és MKI67-pozitív, ER-pozitív és MKI67-negatív) esetében. Az ER-pozitív és MKI67-negatív betegek esetében nem voltunk képesek a validációs beteghalmaz esetében a hatékony osztályozásra, ezért a továbbiakban ezzel a csoporttal kapcsolatos eredményeimet nem ismertetem.

A gyorsított döntési fa a rang-összeg algoritmus eredményein alapszik, csak a nyirokcsomó pozitívak és negatívak között csoportoként szignifikánsan eltérő géneket használja. A rang-összeg elemzésből az 10-10 legnagyobb expresszió-különbséggel rendelkező géneket a 18. táblázatban tüntettem fel. Néhány gént kiemelve a listából: A VAPA gén és az általa kódolt membránfehérje a sejtvezérlésével, mikro-RNS-ek és a PTEN szabályozásával áll szoros kapcsolatban. A KRT14 (Keratin 14) magas expressziója a bazális emlődaganatokra jellemző, az epitheliális-mesenchimális tranzakció egyik eleme [216]. A CECAM6 [217] és az S100P [218] emelkedett expressziója a migráció is az invázió fokozott mértékéhez vezet.

18. táblázat. Példa az eltérő expressziójú génekre a nyirokcsomó pozitív és negatív betegek között

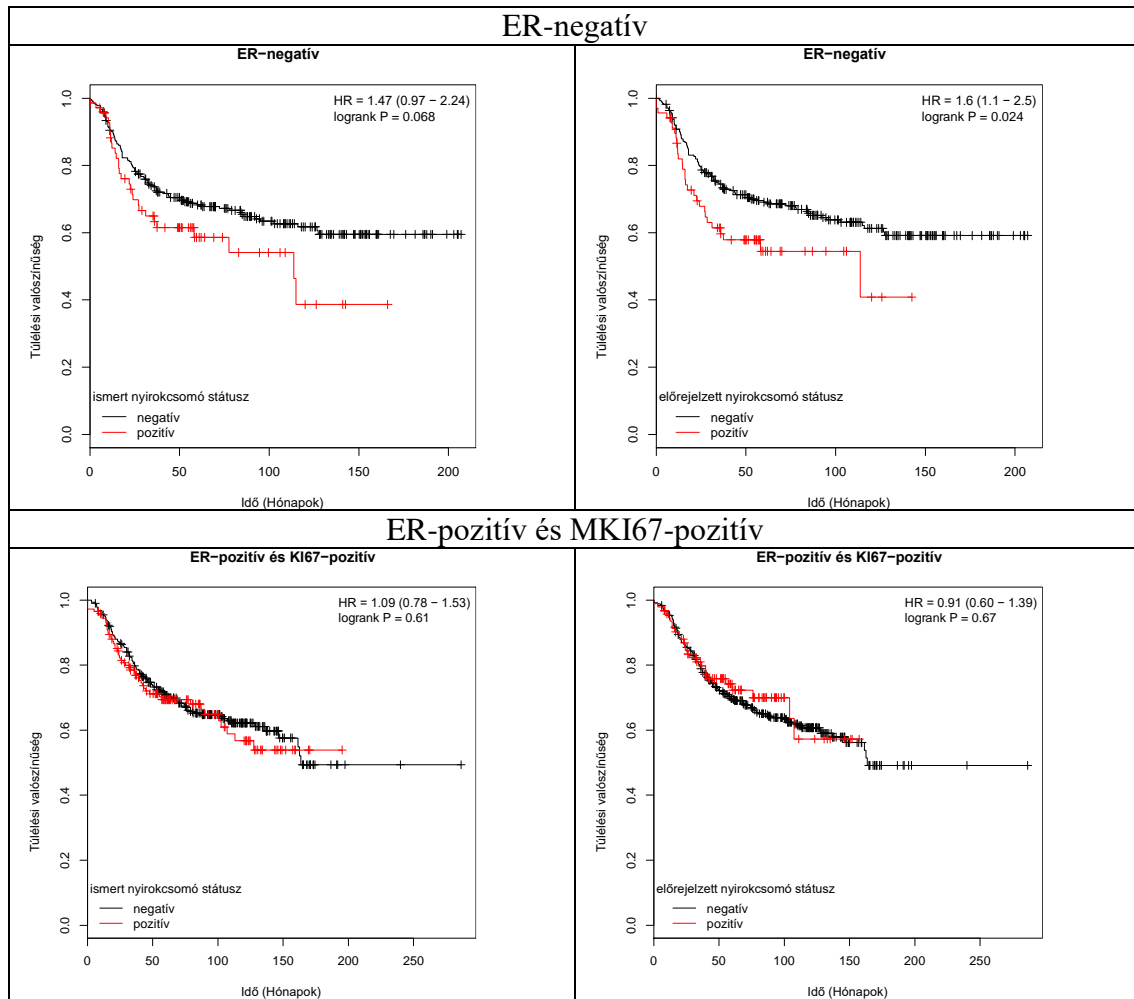
ER-negatív		ER-poz, MKI67-poz.	
gén	Fold change (LN+/LN-)	gén	Fold change (LN+/LN-)
nyirokcsomó pozitívban csökkent			
VAPA	0,48	VAPA	0,52
KRT14	0,70	PIP	0,68
MMP7	0,63	FABP4	0,62
RPL10	0,57	VTCN1	0,64
CYCS	0,58	LTF	0,81
nyirokcsomó pozitívban emelkedett			
SCGB2A2	1,92	SCGB1D2	1,05
CEACAM6	2,16	PTDSS1	1,70
PIP	1,43	S100P	1,41
PTDSS1	1,73	EEF1A2	1,53
TFAP2B	1,72	CEACAM6	1,27

A nyirokcsomó-érintettség előrejelzésének hatékonyságát vizsgálva a belső validációs csoportban mind a hormonreceptor-negatív esetekben (NPV=0,85), mind az ER-pozitív és MKI67-pozitív esetekben (NPV=0,78), magas negatív prediktív értéket értünk el, míg a pontosság is 75% felett maradt. A formalin fixált, paraffinba ágyazott független validációs minták esetében a negatív prediktív érték 0,92 és 1,00 volt, azaz azoknak a betegeknek, akiknek nem jelzünk előre nyirokcsomó-érintettséget nagy valószínűséggel tényleg nincs nyirokcsomó-áttéte. Az osztályozás ilyen módon történő optimalizálása természetesen a pozitív prediktív érték és a pontosság csökkenésével jár, azonban nagyobb a klinikai használhatósága. A belső és a független validáció eredményeit a 19. táblázatban tüntettem fel.

19. táblázat. A nyirokcsomó-státusz előrejelzésének hatékonysága

	Belső validáció		Külső validáció	
	ER-negatív	ER-poz, MKI67-poz	ER-negatív	ER-poz, MKI67-poz
Szentizivítás	0,25	0,41	0,88	1,00
Specifititás	0,88	0,90	0,73	0,50
Pontosság	0,77	0,75	0,78	0,86
PPV	0,31	0,64	0,64	0,83
NPV	0,85	0,78	0,92	1,00

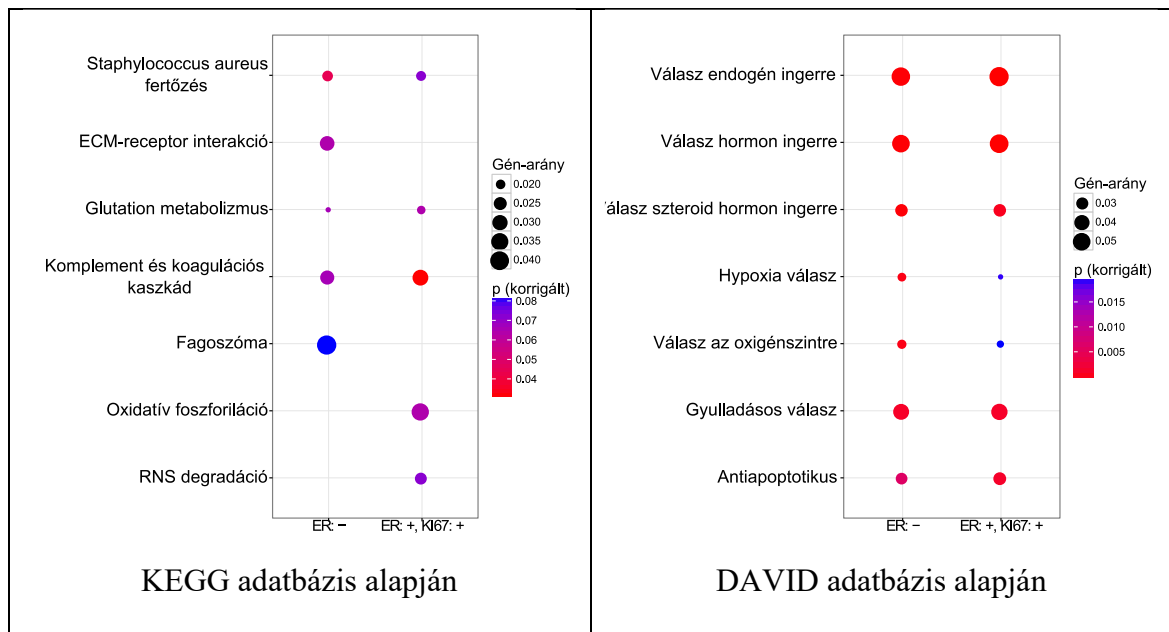
A relapszusmentes túlélés és a nyirokcsomó-érintettség összefüggését vizsgálva látható, hogy az előrejelzett nyirokcsomó-státusz hasonló prognosztikai erővel bír, mint az ismert nyirokcsomó-státusz (33. ábra). Ez azért is fontos mivel ebben a vizsgálatban nem egy prognosztikus teszt fejlesztése volt a célom.



33. ábra. Relapszusmentes túlélés és az ismert illetve előrejelzett nyirokcsomó-érintettség kapcsolata

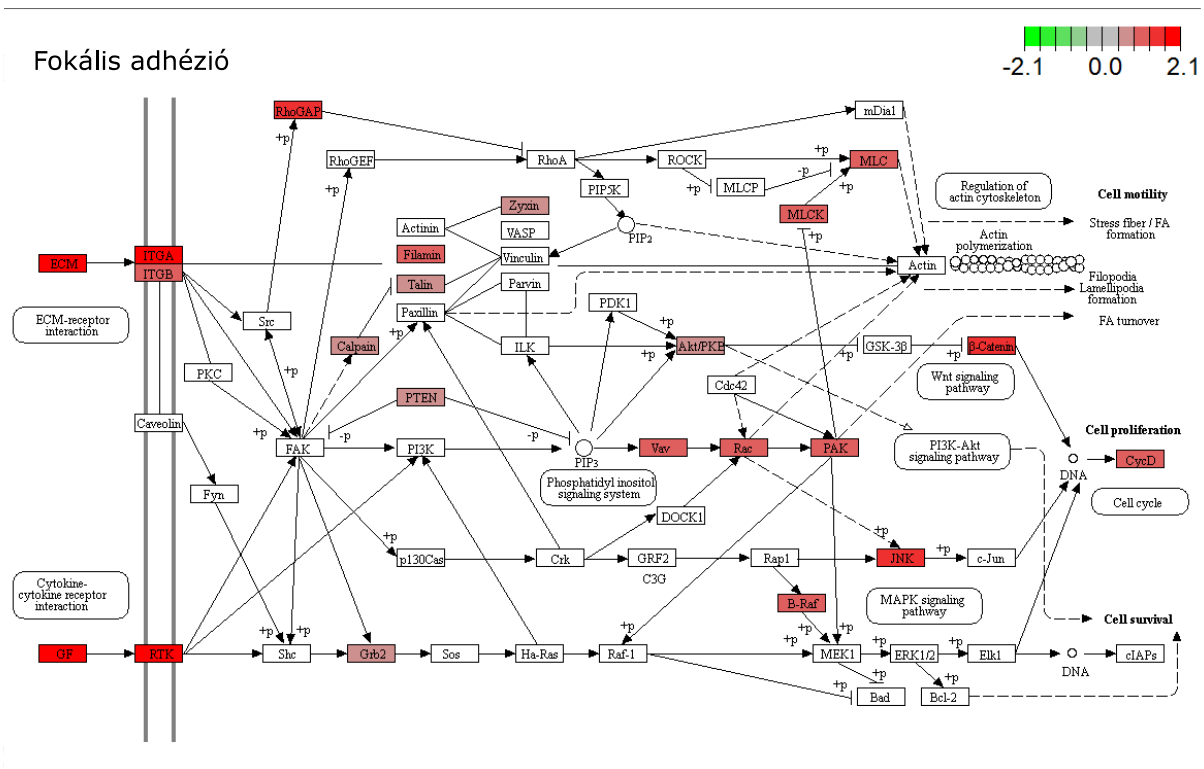
5.3.3 Útvonal és gén-ontológia elemzés

A nyirokcsomó pozitív és negatív betegek csoportjai között szignifikánsan eltérően expresszálódó géneket gén-ontológiai analízisnek vettem alá a DAVID [219] és a KEGG adatbázisokat alkalmazva. Az eredményeket tartalmazó 34. ábra alapján jól látszik, hogy jelentős különbség van a hormonreceptor-negatív és az ER-pozitív, MKI67-pozitív csoportokban a nyirokcsomó áttétképzésben szereppel bíró gének funkciója között. Érdekes, hogy *Staphylococcus Aureus* fertőzéshez kapcsolódó gének felülreprezentáltak a vizsgált génjeink között – más munkacsoport is hasonló eredményre jutott [220] –, valószínűleg a tumorról szembeni immunválaszhoz kapcsolódó génekről van szó.



34. ábra A gén-ontológiai vizsgálat eredménye

A szintén a KEGG adatbázison alapuló útvonalelemzés eredménye-képpen három útvonal volt szignifikánsan felülexpresszált az eltérően kifejeződő gének között: fokális adhézió (hsa04510), extracelluláris mátrix – receptor interakció (hsa04512), koaguláció és komplement kaszkád (hsa04610). A 35. ábra a fokális adhézió útvonalát szemlélteti; az ER-negatív betegek csoportjában a nyirokcsomó pozitív esetekben megnövekedett expressziójú géneket piros színnel jelöltem, a szín intenzitása arányos az expresszióváltozás mértékével.



35. ábra A fokális adhézió útvonalán megváltozott expressziójú gének ER-negatív esetekben

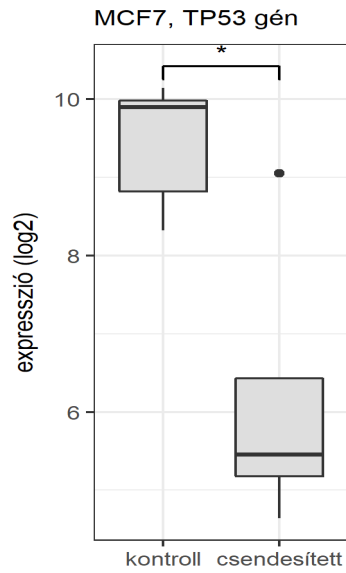
5.4 Biomarkerek reprodukálhatóságának összehasonlítása

A GEO adatbázisban végzett keresést követően 8 sejtvonal esetében összesen 15 gén csendesítéséhez tartozó vizsgálatot azonosítottam, összesen 441 gén-chip elemzését végeztem el. A 441 chip-ből 289 siRNS-sel kezelt, 152 kontroll sejtvonalból származó minta volt. A MAS5 normalizálást követően összehasonlítottam a siRNS-sel kezelt, csendesített gén expresszióját a célgén kontroll, vehikulummal kezelt mintákban mérhető expresszójával (20. táblázat). Látható, hogy 3 esetben nem volt hatékony a csendesítés: HeLa sejtvonal CTNNB1 gén, MCF7 CTNNB1 gén, IMR32 vonal CHAF1A gén. Ez az eredmény is jól mutatja, hogy q-PCR, Western-blot mellett fontos a gén-chipeken is ellenőrizni a célgén expresszióját.

20. táblázat. Csengesítés hatékonysága gén chip alapján

sejtvonal	eredet	csendesített gén	csendesített (átlag)	csendesített (szórás)	kontroll (átlag)	kontroll (szórás)	<i>kezelt kontroll</i>
MCF7	emlő cc.	ESR1	1973	1771	4558	1123	0,43
MCF7	emlő cc.	CTNNB1	839	NA	1187	139	0,71
MCF7	emlő cc.	TP53	161	246	774	363	0,21
MCF7	emlő cc.	TFAP2C	1664	727	3223	803	0,52
MCF7	emlő cc.	CDK4	2883	NA	5478	850	0,53
MCF7	emlő cc.	JMJD6	8	4	24	14	0,33
MCF7	emlő cc.	MAPK14	319	95	732	100	0,44
MCF7	emlő cc.	TOP1	304	31	612	13	0,5
HeLa	cervix cc.	CTNNB1	2495	NA	2585	NA	0,97
HeLa	cervix cc.	DROSHA	884	741	2660	477	0,33
HeLa	cervix cc.	CDK8	206	140	651	102	0,32
HeLa	cervix cc.	CDK19	133	90	452	85	0,29
HeLa	cervix cc.	HIRA	119	41	771	101	0,15
MDAMB231	emlő cc.	TP53	69	19	726	427	0,1
MDAMB231	emlő cc.	JMJD6	36	7	86	22	0,42
IMR32	neuroblastoma	EZH2	1106	727	2145	211	0,52
IMR32	neuroblastoma	CDK4	1131	478	8366	1262	0,14
IMR32	neuroblastoma	CHAF1A	614	200	724	215	0,85
A375	melanoma	CTNNB1	819	NA	1474	369	0,56
A375	melanoma	TP53	280	NA	500	22	0,56
A375	melanoma	CDK4	1872	NA	6870	541	0,27
HCT116	vastagbél cc.	TOP1	93	25	318	127	0,29
SW480	vastagbél cc.	CTNNB1	892	134	2750	608	0,32
LNCAP	prosztata cc.	AR	292	36	1152	659	0,25
LNCAP	prosztata cc.	EZH2	494	32	1394	115	0,35

Érdekes módon a csendesítés átlagos hatékonysága az MCF7 sejtvonalak esetében a legkisebb. A 36. ábra a TP53 gén MCF7 sejtvonalon történt csendesítésének hatékonyságát mutatja be. További eredmények a Függelék 3-ban találhatóak.



36. ábra. MCF7 sejtvonala, TP53 csendesítése

6 Megbeszélés

6.1 Vastagbélrák - rossz prognózisú betegek azonosítása

A vastagbél-daganatok kezelésének fontos, és még eldöntetlen kérdése, azoknak a 2-es stádiumú betegeknek az azonosítása, akiknek további kezelés lenne szükséges. A jelenlegi kezelési sémák szerint a 2-es stádiumú betegek általában nem részesülnek adjuváns kemoterápiában, azonban 15-20%-uknál késői relapszus fordul elő. A 3-as stádiumú betegek bizonyos kivételektől (pl. 75 évesnél idősebb kor) eltekintve adjuváns kezelésben részesülnek. Ezért is szerepel három prognosztikus teszt (Oncotype DX, ColoPrint, ColDX) az NCCN-ajánlásban, melyek közül csak az Oncotype DX-et tudtam reprodukálni. A ColoPrint [221] -et nem tudtam elkészíteni a kézirat alapján, a ColDX pedig egyedi tervezésű Affymetrix chipen készült, betanításához a PETACC3 vizsgálatban résztvevő betegeket alkalmazták, akiknek a klinikai adatai nem hozzáférhetőek. Ezek mellett más tesztek is kaphatóak kereskedelmi forgalomban. Dolgozatomban 22 génexpresszió alapú osztályozót hasonlítottam össze. Az alkalmazott géneket vizsgálva elmondható, hogy kicsi az átfedés a génlisták között, csak öt olyan gén volt, mely hat osztályozóban is szerepelt (REG4, ASCL2, VAV3, C10orf99 és CYPB1). Ezen gének mindegyikét már korábbi vizsgálatokban leírták mint önálló prognosztikai markereket. Az IHC-val kimutatott REG4-pozitivitás hosszabb teljes túléléssel járt nem-mucinózus vastagbél adenokarcinómában [222]. A VAV3 onkogén emelkedett expressziója csökkent betegségmentes túléléssel járt [223]. Az ASCL2 transzkripciós faktor pedig az intesztinális őssejtek fejlődésével és a sejtproliferáció serkentésével hozható kapcsolatba [61]. Az ASCL2 magas szintje rossz prognózissal jár. A C10orf99 citokin szintjének csökkenését írták le vastagbél-daganatokban [224].

A gén pontszám alapján a három legfontosabb, legnagyobb hatással bíró gén a CTGF (Connective Tissue Growth Factor), GADD45B (Growth Arrest and DNA-Damage-inducible, beta), és a FAP (Fibroblast Activation Protein) voltak. Eredményeimhez hasonlóan az emelkedett CTGF-szint rossz teljes túléléssel való kapcsolatát több kohort esetében is megfigyelték [225]. A FAP rossz prognózissal való kapcsolatát ugyancsak leírták már korábban [226]. Az antiapoptotikus GADD45B

expresszióváltozásának hatását a túlélésre eddig nem vizsgálták, annyi ismert, hogy az onkogén KRAS növelheti a GADD45B expresszióját [227].

Az Oncotype DX a vizsgálatomban szignifikánsnak bizonyult, továbbá ennek a tesztnek az előnyei közé tartozik, hogy nem csak kaukázusi, hanem távolkeleti, koreai populáció esetében is validálták [228].

A prognosztikus osztályozás nem mindig jelent prediktív csoportosítást is, például a CMS4 alcsoport igen rossz prognózisú, mégsem javította a 2-es és 3-as stádiumú CMS4 altípusba tartozó betegek progressziómentes túlélését az intenzív oxaliplatin adjuváns kezelés [229].

Fontos eredmény, hogy a prognosztikus tesztek között kicsi volt a korreláció, mely azt jelenti, hogy bár hatékonyan azonosítanak rossz prognózisú betegeket, ezek a betegek nem ugyanazok a különböző osztályozók esetében. Ezért is az egyik logikus lépés bizonyos tesztek kombinálása, amit a „Consensus Molecular Subtypes” esetében is történt. Azonban azt hogy a CMS-be miért azt a bizonyos hat osztályozót integrálták a szerzők semmivel nem indokolják meg. Ezért tekinthető a CMS csak nevében konszenzusnak.

A patológiai, klinikai jellemzőket vizsgálva az adatbázisunkban nem tapasztalunk szignifikáns különbséget a magas és alacsony differenciáltságú csoportok között, ami korábbi ismereteinkhez hasonló [230]. A nem vizsgálatomban nem függött össze a prognózissal. Ez a megfigyelés is egyezik korábbi eredményekkel, miszerint csak a kor- és stádium-korrigálás után látszik a nők prognózisbeli előnye [231]. Az oldaliságot tekintve a proximális (jobb oldali) versus disztális (bal oldali) daganatok között nem láttunk szignifikáns progressziómentes túlélésbeli különbséget, egy nemrég publikált metaanalízis szerint is a teljes túlélést vizsgálva látszik csak a bal oldali tumorok kedvezőbb prognózisa [232].

A vastagbél-daganatok esetében igen jó prognosztikus erővel bír az immunhisztokémiai vizsgálat alapú Immunoscore, mely a különböző immunsejtek számának, arányának tumorbeli, illetve a tumorszéleken történő meghatározásán alapszik. Bár a tumor immunsejtösszetétele jól jellemezhető a génexpresszió alapján, a topológiai szempont miatt nehézkes lenne ezt a tesztet génexpresszió alapú vizsgálattá

alakítani. A CMS altípusok és az immunsejtek jelenlétét, összetételét vizsgálva látható, hogy a különböző altípusok jelentősen eltérő immunológiai profillal rendelkeznek [233].

Elérhetőek továbbá klinikai és patológiai változókon alapuló pontrendszerek is. Ilyen például az Adjuvant! (www.adjuvantonline.com) [234], melyek bár segítik a rossz prognózisú betegek azonosítását, mégsem terjedtek el a klinikai gyakorlatban, nem használtak széles körben.

További kérdésekhez vezet az intratumor heterogenitás, mely a daganat diagnosztika igen fontos és gyakorlati kérdése, ami nem csak mutáció szinten hanem génexpresszió szintjén is megfigyelhető [235]. Ez azt jelenti, hogy a daganat más részből származó minta génexpresszió alapján más-más alcsoportba tartozhat. Különösen igaz lehet ez azokra az alcsoportokra, ahol immunológiai folyamatokban is szereplő gének is részt vesznek a beteg osztályozásában.

Vizsgálatom továbbá az átlátható, reprodukálható kutatás témáját is érintette. Mivel 19 osztályozót vagy a hiányos dokumentáció, vagy a nem elérhető adatok miatt nem lehetett reprodukálni, ezért erről a területről elmondható, szigorúbb dokumentációs előírásokra lenne szükség.

6.2 Emlórák – rossz prognózisú betegek azonosítása

A nagy áteresztőképességű gén chip technológia megjelenésével lehetővé vált sok gén szerepének, expresszióválozásának együttes vizsgálata. A jelenleg legelterjedtebb Affymetrix platformok 2002-ben és 2003-ban jelentek meg. Alapvetően megváltoztatták a biomarker kutatás felépítését. Míg korábban valamilyen biológiai megalapozott hipotézis alapján kerültek célgének kiválasztásra, addig a microarray vizsgálatok korában a nagy mintaszám, a megbízható klinikai adatok, a bioinformatikai elemzésre és a számos biomarker-jelölt kombinálására került a hangsúly. Az emlődaganatok esetében jelenleg alkalmazott többgén teszt - függetlenül attól hogy valamilyen chip vagy RT-PCR alapúak - mind egy-egy előre meghatározott, fix génlistát, és általában a gén expressziók súlyozott lineáris kombinációját alkalmazzák a prognosztikus csoportokba sorolás során. Ezen génlisták meghatározása a 2000-es évek közepén született publikációkban

meglepően kevés beteg (általában kevesebb mint 100) transzkripciós mintázatának összehasonlításával történt.

Mivel a dinamikus osztályozó nem alkalmaz fix génlistát, ezért az alkalmazott gének összehasonlítása más osztályozók génlistáival nem lehetséges. Három gén van, amelyik az osztályozások 20%-ában része a molekuláris osztályozás során alkalmazott 25-25 génnek: CENPE, RACGAP1 és PGK1. A leggyakrabban, csupán az osztályozások 23,7%-ában kiválasztott gén a CENPE, mely szerepét már a sejtmozgás szabályozásában, az invázióban már tripla negatív emlőrákban korábban vizsgálták [236]. A 21-génes (Oncotype DX), a 70-génes (Mammaprint) és a 97-génes Mammaprint génlistáit összehasonlítva négy közös gént találtunk: MYBL2, CCNE2, CENPA és PRC1.

A dinamikus osztályozó egyik lehetséges alkalmazási területe az ER-pozitív antiösztrogén terápiában részesült betegek közül azon rossz prognózisú betegek azonosítása, akiknél a terápia 5 évnél további folytatása ajánlott.

A dinamikus osztályozó előnye, hogy az adatbázis növelésével a prognosztikus erő növekedése várható. A gépi tanulás egyik elve, hogy a modell komplexitásának növelésével a mintán belüli hiba folyamatos csökkenése mellett a mintán kívüli hiba előbb csökken majd nő. A modell tanítása során a mintán kívüli hiba minimalizálása a cél. A mintaszám növelésével is a mintán kívüli hiba csökkenése tapasztalható egy mértékig. Az adatbázis optimális méretét szimulációs vizsgálatokkal lehetne meghatározni.

6.3 Emlőrák - nyirokcsomó érintettség előrejelzése

Az emlődaganatok kezelése során a nyirokcsomó-érintettség meghatározása, így a hónalji blokkdisszekció elhagyásának lehetősége fontos, terápiás következményekkel járó kérdés. Számos (több mint tíz) klinikai (pl. kor, daganat helyzete) és patológia jellemzőn (pl. ép szél, differenciáltság foka) alapuló nomogram létezik [237]. Ezek két kérdésre kereshetik a választ: pozitív őrszem nyirokcsomó esetén a vizsgált beteg esetében mennyi a valószínűsége, hogy további nyirokcsomó nem érintett, illetve hogy a primer daganat tulajdonságai alapján mennyi a szentinel-pozitivitás valószínűsége. A legismertebb ilyen nomogramokat nagy amerikai onkológiai központokban fejlesztették

ki: Memorial Sloan-Kettering Cancer Center (MSKCC) [238], Tenon [239] és Stanford [240]. Ezen nomogramok összehasonlításával ROC-elemzés alapján a görbe alatti terület 0,70 - 0,75 közötti, szenzitivitásuk, negatív prediktív értékük a küszöbértékek megválasztásától függ [241]. Bár több mint 10 éve kerültek kifejlesztésre és nagyobb betegcsoportokon (300 fő) is validálásra kerültek [242], mégsem ajánlott a használatuk egyetlen nemzetközi szervezet szerint sem, mivel eredményeik még nem megbízhatóak. A gén chip technológia megjelenésével ezen a területen is változást hozott. A kezdeti, kis betegszámon végzett vizsgálatok ellentmondásosak voltak, míg az első ilyen vizsgálat leírt a nyirokcsomó-áttétellel összefüggő transzkripciós mintázatot [243], addig az ezt követő vizsgálat nem talált eltérően kifejeződő géneket.[244]. Azonban ezek a csoportok az emlőrákot homogén betegségként kezelték, nem vették figyelembe a molekuláris altípusokat. Jelentős javulás egy mikro-RNS mintázaton [245] és a leggyakoribb, Luminális A betegcsoportra korlátozott génexpressziós [246] vizsgálat ért el, de ezek szenzitivitásai, negatív prediktív értékei is elmaradnak attól, ami a klinikai gyakorlatban elfogadható lenne. A vizsgálatom figyelembe vette a klinikai alcsoportokat, ezért is képes magas negatív prediktív értéket produkálni.

6.4 Biomarkerek reprodukálhatóságának összehasonlítása

A biomarkerek vizsgálatának fontos lépése a preklinikai, sejtkultúrákon végzett vizsgálatok. Eredményeim felhívják rá a figyelmet, hogy – a Western-blot és a RT-PCR elvégzése mellett – minden esetben fontos a csendesített gén expressziójának változását igazolni a gén chipen mért adatokon is.

Az első RNS-interferenciát alkalmazó fázis I klinikai vizsgálat során macula degeneráció esetében a VEGFR1 expresszióját csökkentették [247]. Bár ez a 2006-ban zárult vizsgálat biztató eredményeket mutatott, a hatás célzott elérése és a siRNS lebomlásának megakadályozása jelentős problémát okozott. Ezért jelentős technológiai fejlődésre volt szükség, így a szolid tumorok kezelése területén most jelennek csak meg a siRNS-en alapuló fázis II vizsgálatok eredményei. Jelenleg számos onkológiai készítmény áll fázis I/II vizsgálatok alatt. Az egyik vizsgálatba a PLK1 kinázt célozzák liposzómba csomagolt hatóanyaggal (TKM-080301) [248], szintén liposzómát alkalmazva a PKN3 kináz expresszióját csökkentik [249], a harmadik készítmény esetében a siRNS-eket nanoliposzómába csomagolva alkalmazták az EphA2 kináz

csendesítésére (EPHARNA) [250]. Hepatocelluláris carcinóma esetében pedig lebomló mini-implantátum segítségével bejutattott G12D-mutáns KRAS-t célzó siRNS hatékonyságát vizsgálták [251]. További lehetséges terápias megközelítés az immunterápia és a géncsendesítés kombinálása [252]. Bár jelenleg a génexpresszió csökkentésének még nincs terápias alkalmazása, de továbbra is a biomarker kutatás fontos eszköze lesz.

7 Következtetések

A vastagbél-daganatok osztályozóinak reprodukálása és összehasonlítása 2166 beteg adatai alapján megállapíthatjuk, hogy a 2-es és 3-as stádiumú betegek esetében a Yuen és munkatársai által leírt, 3 gén expresszióján alapuló osztályozó bír a legnagyobb prognosztikai erővel (HR=2,9). Ezt követi Marisa osztályozója (HR=2,60), mely ha minden stádiumú beteget vizsgálunk, a legjobban teljesített (HR=3,20). Fontos megfigyelés, hogy a szignifikáns osztályozók között is korlátozott volt az azonosított jó és rossz prognózisú betegek között az átfedés, azaz nem ugyanazokat a betegeket azonosították. Az alkalmazott géneket összehasonlítva csak öt gén volt, mely hat osztályozóban is előfordult.

Az altípusonként a megfelelő preklinikai modell kiválasztásához 61 vastagbélrák eredetű sejtvonalat (151 génexpressziós vizsgálat alapján) a megfelelő altípusokba soroltam. Eredményeim alapján látszik, hogy vannak olyan osztályozók, melyek nem képesek a sejtvonalakat hatékonyan különböző csoportokra osztani. Eredményeim segíthetik a megfelelő sejtvonal kiválasztását a vastagbél-daganatok vizsgálata során.

Az emlődaganatok esetében 3524 beteg adatait felhasználva egy új, „Dinamikus osztályozót” hoztam létre, mely figyelembe veszi a vizsgált beteg génexpresszió alapú prognózisát, illetve a hozzá hasonló betegek prognózisát is. A Dinamikus osztályozónk három, korábban publikált, klinikai gyakorlatban alkalmazott prognosztikus teszttel összehasonlítva minden beteget vizsgálva is a legjobb eredményt mutatta: Oncotype DX HR=1,4, $p=4,3 \cdot 10^{-39}$, MammaPrint: HR=3,4, $p=1,5 \cdot 10^{-15}$, Genomic Grade Index HR=2,2, $p=2,2 \cdot 10^{-38}$, dinamikus osztályozónk: HR=3,2, $p=7,0 \cdot 10^{-54}$). A klinikai al csoportokat tekintve a dinamikus osztályozó messze meghaladta a korábbi teszteket. Egyik korábbi teszt sem volt képes az ER- és HER2-negatív, gyógyszeres kezelésben részesült betegek osztályozására, míg a dinamikus osztályozó hatékonysága: HR=3,9, $p=4,8 \cdot 10^{-4}$. A dinamikus osztályozónk a 325 fős független validációs kohortunk esetében is felülmúlta a korábbi többgén teszteket hatékonyságát.

Az emlődaganatok nyirokcsomó-érintettségének előrejelzésére a primer daganat génexpressziós mintázata alapján 2341 beteg klinikai és microarray adatait felhasználva döntési fa alapú osztályozót hoztam létre, melyet 100 fős független betegcsoporton

validáltam. A belső validációs csoportunk esetében a negatív nyirokcsomók előrejelzése ER-negatív betegek esetében 85%-os pontosság (ACC) mellett 88%-os negatív prediktív értéket (NPV) ért el, az ER-pozitív / MKI67-pozitív betegek esetében: az ACC=90% és NPV=77% volt. A validációs csoportunkban esetében is hatékonynak bizonyult az előrejelzés: az ER-negatív kohortban: az ACC=73%, a NPV=92% és az ER-pozitív / MKI67-pozitív esetekben pedig: a NPV=100%, az ACC=86% volt.

Az RNS-interferenciával végzett géncsendesítés értékelése során, a géncsendesítés előtti és utáni gén chipek összehasonlítása alapján megállapíthatjuk, hogy a microarray alapú vizsgálatok alkalmasak a géncsendesítés hatásának vizsgálatára.

8 Összefoglalás

A vastagbél daganatok rendkívüli genetikai heterogenitása miatt molekuláris osztályozásuk még nem megoldott. Számottevő az igény a rossz prognózisú, 2-es és 3-as stádiumú betegek azonosítására, melyre számos többgénos osztályozó került kifejlesztésre illetve áll klinikai vizsgálatok alatt.

Kutatásaim alapján a vastagbél-daganatok esetében – 2166 beteg gén chip és klinikai adatainak elemzése alapján – a legjobb és legrosszabb prognózisú csoportokat Cox elemzéssel összevetve a legjobban teljesítő osztályozók a Yuen ($p=3,9*10^{-5}$, HR=2,9), Marisa ($p=2,6*10^{-5}$, HR=2,6) és Chang ($p=9*10^{-9}$, HR=2,35) voltak. Továbbá alcsoportonként meghatároztam a legjobb preklinikai modelleket.

Az emlődaganatok célzott kezelését lehetővé tevő, fix génlistákat használó többgénos tesztek egyre nagyobb szerepet kaptak az elmúlt években. Saját vizsgálatomban azonban egy olyan osztályozó létrehozását tűztem ki célul, mely a vizsgált beteghez hasonló betegek prognózisát is figyelembe veszi.

Az emlődaganatok vizsgálata során 3534 emlőtumorban szenvedő beteg klinikai, túlélési és gén chip adatai alapján új, dinamikus osztályozó algoritmust hoztam létre, mely eredményeit korábbi prognosztikus tesztekkel vettem össze (Oncotype DX, Mammaprint, Genomic Grade Index). Az összes beteg esetében a dinamikus osztályozó jobban teljesített (dinamikus osztályozó: HR=3,2, $p=7,0*10^{-54}$). A klinikai alcsoportokat tekintve a dinamikus osztályozó messze meghaladta a korábbi tesztekét. Eredményeimet 325 fős független betegcsoporton validáltam.

Az emlő rosszindulatú daganatai esetében a nyirokcsomó-státusz alapvetően jelentősen befolyásolja az alkalmazott terápiát. Azonban az őrszem nyirokcsomó biopszia és a hónalji nyirokcsomók blokkdisszekciója is jelentős diagnosztikus és terápiás előnyeik mellett morbiditások forrásai is lehetnek.

Az emlődaganatok nyirokcsomó-érintettségének előrejelzéséhez a primer tumor génexpressziós mintázata alapján 2341 beteg klinikai és génexpressziós adatait felhasználva az ER-negatív és ER-pozitív / MKI67-pozitív betegcsoportok esetében is magas pontosságot és negatív prediktív értéket (NPV) értünk el. A 100 beteget tartalmazó független validációs csoportunkban az ER-negatív kohortban a pontosság 73%, NPV: 92% és az ER-pozitív / MKI67-pozitív esetekben pedig pontosság: 86%, NPV:100% volt.

9 Summary

Colorectal cancer is a genetically heterogeneous disease, thus there is no consensus molecular classification system yet. Because of the increasing need for identifying bad prognostic stage II and III patients several multigene classifiers were published, and some of them are already being examined in clinical trials.

Based on my research regarding colorectal cancer by analyzing microarray and clinical data of 2,166 patients, the highest efficacy to predict progression free survival in stage II-III patients was achieved by the Yuen ($p = 3.9 \times 10^{-5}$, HR = 2.9), Marisa ($p = 2.6 \times 10^{-5}$, HR = 2.6) and Chang ($p = 9 \times 10^{-9}$, HR = 2.35) classifiers. For each of the subtypes the best preclinical models were determined.

To personalize the treatment of breast cancer patients, numerous multigene prognostic tests are available. To overcome limitations of these previous multigene prognostic classifiers, I propose a new dynamic predictor, which utilizes the prognosis of patients with similar gene expression profiles to the investigated sample.

In case of breast cancer based on gene expression and clinical data from 3,534 patients a new classifier was developed. Its performance was compared to three previously published multigene prognostic tests (Oncotype DX, MammaPrint, Genomic Grade Index). Our prognostic discrimination was the highest for all cases (dynamic predictor: HR=3.2, $p=7.0 \times 10^{-54}$). In the clinical subtypes the dynamic classifier also outperformed the others. The model was also validated in 325 independent cases.

Lymph node status is one of the most important clinical parameters of breast cancer. Axillary lymph node dissection and sentinel lymph node biopsy have considerable morbidity associated with them and a method to accurately predict lymph node positivity would be clinically useful.

Prediction of lymph node metastases based on the primary tumor's gene expression was developed using data from 2341 patients. In our internal validation set of ER-negative patients and ER-positive / MKI67-positive group the predictor achieved good accuracy (ACC) and negative predictive value (NPV). In case of our independent validation set of 100 patients our prediction model performed also well: for the ER-negative cohort: ACC=0.73, NPV=0.92, and for the ER-positive and MKI67-positive group: ACC=0.86, NPV=1.0.

10 Irodalomjegyzék

1. Sztupinszki, Z.B. Gyorffy. (2016) Colon cancer subtypes: concordance, effect on survival and selection of the most representative preclinical models. *Sci Rep*, 6: 37169.
2. Gyorffy, B., T. Karn, Z. Sztupinszki, B. Weltz, V. MullerL. Pusztai. (2015) Dynamic classification using case-specific training cohorts outperforms static gene expression signatures in breast cancer. *Int J Cancer*, 136: 2091-8.
3. Munkacsy, G., Z. Sztupinszki, P. Herman, B. Ban, Z. Penzvalto, N. SzarvasB. Gyorffy. (2016) Validation of RNAi Silencing Efficiency Using Gene Array Data shows 18.5% Failure Rate across 429 Independent Experiments. *Mol Ther Nucleic Acids*, 5: e366.
4. Silberman, A.W. (1982) Surgical debulking of tumors. *Surgery, gynecology & obstetrics*, 155: 577-85.
5. Zoller, F., M. Eisenhut, U. HaberkornW. Mier. (2009) Endoradiotherapy in cancer treatment--basic concepts and future trends. *European journal of pharmacology*, 625: 55-62.
6. Morgenroth, A., A.T. Vogg, F.M. MottaghyJ. Schmaljohann. (2011) Targeted endoradiotherapy using nucleotides. *Methods*, 55: 203-14.
7. Garnick, M.B. (1987) Current status of endocrine therapy for prostate cancer. *Oncology*, 1: 19-25, 30-1, 34.
8. Allegra, J.C. (1983) Rational approaches to the hormonal treatment of breast cancer. *Seminars in oncology*, 10: 25-8.
9. Williams, K.J. (2009) The introduction of 'chemotherapy' using arsphenamine - the first magic bullet. *J R Soc Med*, 102: 343-348.
10. Krumbhaar, E.B.H.D. Krumbhaar. (1919) The Blood and Bone Marrow in Yellow Cross Gas (Mustard Gas) Poisoning: Changes produced in the Bone Marrow of Fatal Cases. *J Med Res*, 40: 497-508 3.
11. Adair, F.E.H.J. Bagg. (1931) Experimental and Clinical Studies on the Treatment of Cancer by Dichlorethylsulphide (Mustard Gas). *Ann Surg*, 93: 190-9.

12. US Food and Drug Administration, O.o.C.P. (2002) Annual Report to Congress Federal Food, Drug, and Cosmetic Act as Amended by the Medical Device User Fee Act of 2002. National Press Office.
13. Zhukov, N.V.S.A. Tjulandin. (2008) Targeted therapy in the treatment of solid tumors: practice contradicts theory. *Biochemistry. Biokhimiia*, 73: 605-18.
14. Penzvalto, Z., B. Tegze, A.M. Szasz, Z. Sztupinszki, I. Liko, A. Szendroi, R. SchaferB. Gyorffy. (2013) Identifying resistance mechanisms against five tyrosine kinase inhibitors targeting the ERBB/RAS pathway in 45 cancer cell lines. *PLoS One*, 8: e59503.
15. Deininger, M., E. BuchdungerB.J. Druker. (2005) The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood*, 105: 2640-53.
16. Mihaly, Z., Z. Sztupinszki, P. SurowiakB. Gyorffy. (2012) A comprehensive overview of targeted therapy in metastatic renal cell carcinoma. *Curr Cancer Drug Targets*, 12: 857-72.
17. Grothey, A.L.M. Ellis. (2008) Targeting angiogenesis driven by vascular endothelial growth factors using antibody-based therapies. *Cancer journal*, 14: 170-7.
18. Sasich, L.D.S.R. Sukkari. (2012) The US FDA's withdrawal of the breast cancer indication for Avastin (bevacizumab). *Saudi Pharm J*, 20: 381-5.
19. Kim, G., G. Ison, A.E. McKee, H. Zhang, S. Tang, T. Gwise, R. Sridhara, E. Lee, A. Tzou, R. Philip, H.J. Chiu, T.K. Ricks, T. Palmby, A.M. Russell, G. Ladouceur, E. Pfuma, H. Li, L. Zhao, Q. Liu, R. Venugopal, A. IbrahimR. Pazdur. (2015) FDA Approval Summary: Olaparib Monotherapy in Patients with Deleterious Germline BRCA-Mutated Advanced Ovarian Cancer Treated with Three or More Lines of Chemotherapy. *Clin Cancer Res*, 21: 4257-61.
20. Dockery, L.E., C.C. GundersonK.N. Moore. (2017) Rucaparib: the past, present, and future of a newly approved PARP inhibitor for ovarian cancer. *Onco Targets Ther*, 10: 3029-3037.
21. Scott, L.J. (2017) Niraparib: First Global Approval. *Drugs*, 77: 1029-1034.
22. Reck, M., N. Blais, E. Juhasz, V. Gorbunova, C.M. Jones, L. Urban, S. Orlov, F. Barlesi, E. Kio, U. Keilholz, Q. Qin, J. Qian, C. Nickner, J. Dziubinski, H. Xiong, R.K. Mittapalli, M. Dunbar, P. Ansell, L. He, M. McKee, V. GirandaS.S.

- Ramalingam. (2017) Smoking History Predicts Sensitivity to PARP Inhibitor Veliparib in Patients with Advanced Non-Small Cell Lung Cancer. *J Thorac Oncol*, 12: 1098-1108.
23. Lambert, J.M.R.V. Chari. (2014) Ado-trastuzumab Emtansine (T-DM1): an antibody-drug conjugate (ADC) for HER2-positive breast cancer. *J Med Chem*, 57: 6949-64.
 24. (2017) First Tissue-Agnostic Drug Approval Issued. *Cancer Discov*, 7: 656.
 25. Ehmann, L.M., T. RuzickaA. Wollenberg. (2011) Cutaneous side-effects of EGFR inhibitors and their management. *Skin Therapy Lett*, 16: 1-3.
 26. Donskov, F., M.D. Michaelson, I. Puzanov, M.P. Davis, G.A. Bjarnason, R.J. Motzer, D. Goldstein, X. Lin, D.P. Cohen, R. WiltshireB.I. Rini. (2015) Sunitinib-associated hypertension and neutropenia as efficacy biomarkers in metastatic renal cell carcinoma patients. *Br J Cancer*, 113: 1571-80.
 27. Brahmer, J.R., S.S. Tykodi, L.Q. Chow, W.J. Hwu, S.L. Topalian, P. Hwu, C.G. Drake, L.H. Camacho, J. Kauh, K. Odunsi, H.C. Pitot, O. Hamid, S. Bhatia, R. Martins, K. Eaton, S. Chen, T.M. Salay, S. Alaparthi, J.F. Grosso, A.J. Korman, S.M. Parker, S. Agrawal, S.M. Goldberg, D.M. Pardoll, A. GuptaJ.M. Wigginton. (2012) Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*, 366: 2455-65.
 28. Siegel, R.L., K.D. MillerA. Jemal. (2017) Cancer Statistics, 2017. *CA Cancer J Clin*, 67: 7-30.
 29. Edwards, B.K., E. Ward, B.A. Kohler, C. Ehemann, A.G. Zauber, R.N. Anderson, A. Jemal, M.J. Schymura, I. Lansdorp-Vogelaar, L.C. Seeff, M. van Ballegooijen, S.L. GoedeL.A. Ries. (2010) Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*, 116: 544-73.
 30. Siegel, R.L., K.D. Miller, S.A. Fedewa, D.J. Ahnen, R.G. Meester, A. BarziA. Jemal. (2017) Colorectal cancer statistics, 2017. *CA Cancer J Clin*.
 31. Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J.W. Coebergh, H. Comber, D. FormanF. Bray. (2013) Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*, 49: 1374-403.

32. Arnold, M., M.S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray. (2017) Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, 66: 683-691.
33. Kasler, M., S. Otto, I. Kenessey. (2017) A rákmorbiditás és -mortalitás jelenlegi helyzete a Nemzeti Rákregiszter tükrében. *Orv Hetil*, 158: 84-89.
34. Kasler, M., S. Otto, O. Solyom. (2014) [National Cancer Registry. Significance of a reliable database in the implementation of the required structural changes of cancer care in Hungary]. *Orv Hetil*, 155: 1415-20.
35. Siegel, R.L., S.A. Fedewa, W.F. Anderson, K.D. Miller, J. Ma, P.S. Rosenberg, A. Jemal. (2017) Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst*, 109.
36. Sanoff, H.K., D.J. Sargent, M.E. Campbell, R.F. Morton, C.S. Fuchs, R.K. Ramanathan, S.K. Williamson, B.P. Findlay, H.C. Pitot, R.M. Goldberg. (2008) Five-year data and prognostic factor analysis of oxaliplatin and irinotecan combinations for advanced colorectal cancer: N9741. *J Clin Oncol*, 26: 5721-7.
37. Dukes, C.E. (1932) The classification of cancer of the rectum. *The Journal of Pathology and Bacteriology*, 35: 323-332.
38. Astler, V.B., F.A. Coller. (1954) The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg*, 139: 846-52.
39. Edge, S.B., C.C. Compton. (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*, 17: 1471-4.
40. Howlader, N., A.M. Noone, M. Krapcho, D. Miller, K. Bishop, C.L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D.R. Lewis, H.S. Chen, E.J. Feuer, K.A. Cronin. (2017) SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission, posted to the SEER web site.
41. Tusnady, G., I. Gaudi, L. Rejto, M. Kasler, Z. Szentirmay. (2008) A magyar daganatos betegek túlélési esélye a nemzeti rákregiszter adatai alapján. *Magy Onkol*, 52: 339-49.
42. Boland, C.R., A. Goel. (2010) Microsatellite instability in colorectal cancer. *Gastroenterology*, 138: 2073-2087 e3.

43. Boland, C.R., S.N. Thibodeau, S.R. Hamilton, D. Sidransky, J.R. Eshleman, R.W. Burt, S.J. Meltzer, M.A. Rodriguez-Bigas, R. Fodde, G.N. Ranzani, S. Srivastava. (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res*, 58: 5248-57.
44. Kim, H., J. Jen, B. Vogelstein, S.R. Hamilton. (1994) Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol*, 145: 148-56.
45. Bubb, V.J., L.J. Curtis, C. Cunningham, M.G. Dunlop, A.D. Carothers, R.G. Morris, S. White, C.C. Bird, A.H. Wyllie. (1996) Microsatellite instability and the role of hMSH2 in sporadic colorectal cancer. *Oncogene*, 12: 2641-9.
46. Gafa, R., I. Maestri, M. Matteuzzi, A. Santini, S. Ferretti, L. Cavazzini, G. Lanza. (2000) Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability. *Cancer*, 89: 2025-37.
47. Popat, S., R. Hubner, R.S. Houlston. (2005) Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*, 23: 609-18.
48. Carethers, J.M., B. Murali, B. Yang, R.T. Doctolero, A. Tajima, R. Basa, E.J. Smith, M. Lee, R. Janke, T. Ngo, R. Tejada, M. Ji, M. Kinseth, B.L. Cabrera, K. Miyai, T.O. Keku, C.F. Martin, J.A. Galanko, R.S. Sandler, K.L. McGuire. (2014) Influence of race on microsatellite instability and CD8+ T cell infiltration in colon cancer. *PLoS One*, 9: e100461.
49. Buecher, B., W. Cacheux, E. Rouleau, B. Dieumegard, E. Mitry, A. Lievre. (2013) Role of microsatellite instability in the management of colorectal cancers. *Dig Liver Dis*, 45: 441-9.
50. Moreira, L., F. Balaguer, N. Lindor, A. de la Chapelle, H. Hampel, L.A. Aaltonen, J.L. Hopper, L. Le Marchand, S. Gallinger, P.A. Newcomb, R. Haile, S.N. Thibodeau, S. Gunawardena, M.A. Jenkins, D.D. Buchanan, J.D. Potter, J.A. Baron, D.J. Ahnen, V. Moreno, M. Andreu, M. Ponz de Leon, A.K. Rustgi, A. Castells, E. Consortium. (2012) Identification of Lynch syndrome among patients with colorectal cancer. *JAMA*, 308: 1555-65.

51. Sargent, D.J., S. Marsoni, G. Monges, S.N. Thibodeau, R. Labianca, S.R. Hamilton, A.J. French, B. Kabat, N.R. Foster, V. Torri, C. Ribic, A. Grothey, M. Moore, A. Zaniboni, J.F. Seitz, F. Sinicrope S. Gallinger. (2010) Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol*, 28: 3219-26.
52. Rajagopalan, H., M.A. Nowak, B. Vogelstein C. Lengauer. (2003) The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer*, 3: 695-701.
53. Michor, F., Y. Iwasa M.A. Nowak. (2004) Dynamics of cancer progression. *Nat Rev Cancer*, 4: 197-205.
54. Michor, F., Y. Iwasa, B. Vogelstein, C. Lengauer M.A. Nowak. (2005) Can chromosomal instability initiate tumorigenesis? *Semin Cancer Biol*, 15: 43-9.
55. Nazemalhosseini Mojarad, E., P.J. Kuppen, H.A. Aghdaei M.R. Zali. (2013) The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*, 6: 120-8.
56. Weisenberger, D.J., K.D. Siegmund, M. Campan, J. Young, T.I. Long, M.A. Faasse, G.H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A.J. French, S.N. Thibodeau, J. Jass, R. Haile P.W. Laird. (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*, 38: 787-93.
57. Ferlitsch, M., K. Reinhart, S. Pramhas, C. Wiener, O. Gal, C. Bannert, M. Hassler, K. Kozbial, D. Dunkler, M. Trauner W. Weiss. (2011) Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. *JAMA*, 306: 1352-8.
58. Leslie, A., F.A. Carey, N.R. Pratt R.J. Steele. (2002) The colorectal adenoma-carcinoma sequence. *Br J Surg*, 89: 845-60.
59. Vogelstein, B., E.R. Fearon, S.R. Hamilton, S.E. Kern, A.C. Preisinger, M. Leppert, Y. Nakamura, R. White, A.M. Smits J.L. Bos. (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319: 525-32.

60. Lynch, H.T., J.F. Lynch, P.M. Lynch, T. Attard. (2008) Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. *Fam Cancer*, 7: 27-39.
61. Cancer Genome Atlas, N. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487: 330-7.
62. Yi, J., Z.W. Wang, H. Cang, Y.Y. Chen, R. Zhao, B.M. Yu, X.M. Tang. (2001) p16 gene methylation in colorectal cancers associated with Duke's staging. *World J Gastroenterol*, 7: 722-5.
63. Fleet, J.C. (2014) Animal models of gastrointestinal and liver diseases. New mouse models for studying dietary prevention of colorectal cancer. *Am J Physiol Gastrointest Liver Physiol*, 307: G249-59.
64. Poynter, J.N., K.D. Siegmund, D.J. Weisenberger, T.I. Long, S.N. Thibodeau, N. Lindor, J. Young, M.A. Jenkins, J.L. Hopper, J.A. Baron, D. Buchanan, G. Casey, A.J. Levine, L. Le Marchand, S. Gallinger, B. Bapat, J.D. Potter, P.A. Newcomb, R.W. Haile, P.W. Laird. I. Colon Cancer Family Registry. (2008) Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. *Cancer Epidemiol Biomarkers Prev*, 17: 3208-15.
65. Chiba, T., H. Marusawa, T. Ushijima. (2012) Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology*, 143: 550-63.
66. Westbrook, A.M., B. Wei, J. Braun, R.H. Schiestl. (2011) Intestinal inflammation induces genotoxicity to extraintestinal tissues and cell types in mice. *Int J Cancer*, 129: 1815-25.
67. Burner, G.C., D.A. Crispin, V.R. Kolli, R.C. Haggitt, B.G. Kulander, C.E. Rubin, P.S. Rabinovitch. (1991) Frequent loss of a p53 allele in carcinomas and their precursors in ulcerative colitis. *Cancer Commun*, 3: 167-72.
68. Ezaki, T., M. Watanabe, N. Inoue, T. Kanai, H. Ogata, Y. Iwao, H. Ishii, T. Hibi. (2003) A specific genetic alteration on chromosome 6 in ulcerative colitis-associated colorectal cancers. *Cancer Res*, 63: 3747-9.

69. Azarschab, P., R. Porschen, M. Gregor, N. BlinK. Holzmann. (2002) Epigenetic control of the E-cadherin gene (CDH1) by CpG methylation in colectomy samples of patients with ulcerative colitis. *Genes Chromosomes Cancer*, 35: 121-6.
70. Coombs, N.J., K.A. Cronin, R.J. Taylor, A.N. FreedmanJ. Boyages. (2010) The impact of changes in hormone therapy on breast cancer incidence in the US population. *Cancer Causes Control*, 21: 83-90.
71. Breen, N., J.F. GentlemanJ.S. Schiller. (2011) Update on mammography trends: comparisons of rates in 2000, 2005, and 2008. *Cancer*, 117: 2209-18.
72. Horváth, J., R. Futár, Z. Péntek, B. JuhászJ. Balogh. (1972) Emlőrák lakosságszűrés szervezési tapasztalatai. *Népegészségügy*, 53: 170-172.
73. Shapiro, S., P. StraxL. Venet. (1966) Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. *JAMA*, 195: 731-8.
74. Jorgensen, K.J., P.C. Gotzsche, M. KalagerP.H. Zahl. (2017) Breast Cancer Screening in Denmark: A Cohort Study of Tumor Size and Overdiagnosis. *Ann Intern Med*, 166: 313-323.
75. Tabar, L., S.W. Duffy, B. Vitak, H.H. ChenT.C. Prevost. (1999) The natural history of breast carcinoma: what have we learned from screening? *Cancer*, 86: 449-62.
76. Cady, B., M.D. Stone, J.G. Schuler, R. Thakur, M.A. WannerP.T. Lavin. (1996) The new era in breast cancer. Invasion, size, and nodal involvement dramatically decreasing as a result of mammographic screening. *Arch Surg*, 131: 301-8.
77. Ravdin, P.M., K.A. Cronin, N. Howlader, C.D. Berg, R.T. Chlebowski, E.J. Feuer, B.K. EdwardsD.A. Berry. (2007) The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med*, 356: 1670-4.
78. Mansfield, C.M. (1976) Early breast cancer its history and results of treatment. *Exp Biol Med*, 5: 1-129.
79. Gray, J., K. Brain, P. Norman, C. Anglim, L. France, G. Barton, L. Branston, E. Parsons, A. Clarke, J. Sampson, E. Roberts, R. Newcombe, D. Cohen, C. Rogers, R. ManselP. Harper. (2000) A model protocol evaluating the introduction of genetic assessment for women with a family history of breast cancer. *J Med Genet*, 37: 192-6.

80. Moore, C.H. (1867) On the Influence of Inadequate Operations on the Theory of Cancer. *Med Chir Trans*, 50: 245-80.
81. Halsted, W.S. (1894) I. The Results of Operations for the Cure of Cancer of the Breast Performed at the Johns Hopkins Hospital from June, 1889, to January, 1894. *Ann Surg*, 20: 497-555.
82. Fisher, B., C. Redmond, E.R. Fisher, M. Bauer, N. Wolmark, D.L. Wickerham, M. Deutsch, E. Montague, R. Margolese R. Foster. (1985) Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *N Engl J Med*, 312: 674-81.
83. Hellman, S. (1994) Karnofsky Memorial Lecture. Natural history of small breast cancers. *J Clin Oncol*, 12: 2229-34.
84. Keshtgar, M.R.M. Baum. (2001) Axillary dissection over the years: where to from here? *World J Surg*, 25: 761-6.
85. Huang, T.W., K.N. Kuo, K.H. Chen, C. Chen, W.H. Hou, W.H. Lee, T.Y. Chao, J.T. Tsai, C.M. Su, M.T. Huang K.W. Tam. (2016) Recommendation for axillary lymph node dissection in women with early breast cancer and sentinel node metastasis: A systematic review and meta-analysis of randomized controlled trials using the GRADE system. *Int J Surg*, 34: 73-80.
86. Mejdahl, M.K., K.G. Andersen, R. Gartner, N. Kroman H. Kehlet. (2013) Persistent pain and sensory disturbances after treatment for breast cancer: six year nationwide follow-up study. *BMJ*, 346: f1865.
87. Orr, R.K. (1999) The impact of prophylactic axillary node dissection on breast cancer survival--a Bayesian meta-analysis. *Ann Surg Oncol*, 6: 109-16.
88. Sanghani, M., E.M. Balk B. Cady. (2009) Impact of axillary lymph node dissection on breast cancer outcome in clinically node negative patients: a systematic review and meta-analysis. *Cancer*, 115: 1613-20.
89. Clarke, M., R. Collins, S. Darby, C. Davies, P. Elphinstone, V. Evans, J. Godwin, R. Gray, C. Hicks, S. James, E. MacKinnon, P. McGale, T. McHugh, R. Peto, C. Taylor, Y. Wang G. Early Breast Cancer Trialists' Collaborative. (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, 366: 2087-106.

90. Cserni, G. (2014) Őrszemnyirokcsomó-státus és hónalji blokkdissectio az emlőrák sebészi ellátásában. *Orv Hetil*, 155: 203-15.
91. Wang, H., X.Y. Mao, T.T. Zhao, X.Y. Zheng, F. JinJ.G. Li. (2012) Study on the skip metastasis of axillary lymph nodes in breast cancer and their relation with Gli1 expression. *Tumour Biol*, 33: 1943-50.
92. Lyman, G.H., A.E. Giuliano, M.R. Somerfield, A.B. Benson, 3rd, D.C. Bodurka, H.J. Burstein, A.J. Cochran, H.S. Cody, 3rd, S.B. Edge, S. Galper, J.A. Hayman, T.Y. Kim, C.L. Perkins, D.A. Podoloff, V.H. Sivasubramaniam, R.R. Turner, R. Wahl, D.L. Weaver, A.C. Wolff, E.P. WinerO. American Society of Clinical. (2005) American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol*, 23: 7703-20.
93. Caudle, A.S., K.K. Hunt, H.M. Kuerer, F. Meric-Bernstam, A. Lucci, I. Bedrosian, G.V. Babiera, R.F. Hwang, M.I. Ross, B.W. Feig, K. Hoffman, J.K. Litton, A.A. Sahin, W. Yang, G.N. Hortobagyi, T.A. BuchholzE.A. Mittendorf. (2011) Multidisciplinary considerations in the implementation of the findings from the American College of Surgeons Oncology Group (ACOSOG) Z0011 study: a practice-changing trial. *Ann Surg Oncol*, 18: 2407-12.
94. Galimberti, V., B.F. Cole, S. Zurrada, G. Viale, A. Luini, P. Veronesi, P. Baratella, C. Chifu, M. Sargenti, M. Intra, O. Gentilini, M.G. Mastropasqua, G. Mazzarol, S. Massarut, J.R. Garbay, J. Zgajnar, H. Galatius, A. Recalcati, D. Littlejohn, M. Bamert, M. Colleoni, K.N. Price, M.M. Regan, A. Goldhirsch, A.S. Coates, R.D. Gelber, U. Veronesii. International Breast Cancer Study Group Trial. (2013) Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23-01): a phase 3 randomised controlled trial. *Lancet Oncol*, 14: 297-305.
95. Donker, M., G. van Tienhoven, M.E. Straver, P. Meijnen, C.J. van de Velde, R.E. Mansel, L. Cataliotti, A.H. Westenberg, J.H. Klinkenbijl, L. Orzalesi, W.H. Bouma, H.C. van der Mijle, G.A. Nieuwenhuijzen, S.C. Veltkamp, L. Slaets, N.J. Duez, P.W. de Graaf, T. van Dalen, A. Marinelli, H. Rijna, M. Snoj, N.J. Bundred, J.W. Merkus, Y. Belkacemi, P. Petignat, D.A. Schinagl, C. Coens, C.G. Messina, J. BogaertsE.J. Rutgers. (2014) Radiotherapy or surgery of the axilla after a

- positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-inferiority trial. *Lancet Oncol*, 15: 1303-10.
96. Savolt, A., C. Polgar, P. Musonda, Z. Matrai, F. Renyi-Vamos, L. Toth, M. KaslerG. Peley. (2013) Does the result of completion axillary lymph node dissection influence the recommendation for adjuvant treatment in sentinel lymph node-positive patients? *Clin Breast Cancer*, 13: 364-70.
 97. Savolt, A., G. Peley, C. Polgar, N. Udvarhelyi, G. Rubovszky, E. Kovacs, B. Gyorffy, M. KaslerZ. Matrai. (2017) Eight-year follow up result of the OTOASOR trial: The Optimal Treatment Of the Axilla - Surgery Or Radiotherapy after positive sentinel lymph node biopsy in early-stage breast cancer: A randomized, single centre, phase III, non-inferiority trial. *Eur J Surg Oncol*, 43: 672-679.
 98. Gentilini, O.U. Veronesi. (2012) Abandoning sentinel lymph node biopsy in early breast cancer? A new trial in progress at the European Institute of Oncology of Milan (SOUND: Sentinel node vs Observation after axillary UltraSOUND). *Breast*, 21: 678-81.
 99. Network, N.C.C. NCCN Guidelines Version 2.2017 Breast Cancer. 2017. [cited 2017. 2017.06.02.].
 100. Coates, A.S., E.P. Winer, A. Goldhirsch, R.D. Gelber, M. Gnant, M. Piccart-Gebhart, B. Thurlimann, H.J. SennM. Panel. (2015) Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol*, 26: 1533-46.
 101. Senkus, E., S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, F. CardosoE.G. Committee. (2015) Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 26 Suppl 5: v8-30.
 102. Lázár, G., A. Bursics, Z. Farsang, L. Harsányi, C. Kósa, R. Maráz, Z. Mátrai, A. Paszt, G. PavlovicsR. Tamás. (2016) III. Emlőrák Konszenzus Konferencia – Az emlőrák korszerű sebészi kezelése. *Magyar Onkológia*, Vol 60: 194-207.

103. Firestein, G.S. (2006) A biomarker by any other name... *Nature Clinical Practice Rheumatology*, 2: 635-635.
104. Hajdu, S.I. (2006) A note from history: the first biochemical test for detection of cancer. *Ann Clin Lab Sci*, 36: 222-3.
105. Bettegowda, C., M. Sausen, R.J. Leary, I. Kinde, Y. Wang, N. Agrawal, B.R. Bartlett, H. Wang, B. Luber, R.M. Alani, E.S. Antonarakis, N.S. Azad, A. Bardelli, H. Brem, J.L. Cameron, C.C. Lee, L.A. Fecher, G.L. Gallia, P. Gibbs, D. Le, R.L. Giuntoli, M. Goggins, M.D. Hogarty, M. Holdhoff, S.M. Hong, Y. Jiao, H.H. Juhl, J.J. Kim, G. Siravegna, D.A. Laheru, C. Lauricella, M. Lim, E.J. Lipson, S.K. Marie, G.J. Netto, K.S. Oliner, A. Olivi, L. Olsson, G.J. Riggins, A. Sartore-Bianchi, K. Schmidt, M. Shih 1, S.M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T.T. Harkins, S. Veronese, T.L. Wang, J.D. Weingart, C.L. Wolfgang, L.D. Wood, D. Xing, R.H. Hruban, J. Wu, P.J. Allen, C.M. Schmidt, M.A. Choti, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, N. Papadopoulos L.A. Diaz, Jr. (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*, 6: 224ra24.
106. Spindler, K.L., N. Pallisgaard, R.F. Andersen A. Jakobsen. (2014) Changes in mutational status during third-line treatment for metastatic colorectal cancer--results of consecutive measurement of cell free DNA, KRAS and BRAF in the plasma. *Int J Cancer*, 135: 2215-22.
107. Smerage, J.B., W.E. Barlow, G.N. Hortobagyi, E.P. Winer, B. Leyland-Jones, G. Srkalovic, S. Tejawani, A.F. Schott, M.A. O'Rourke, D.L. Lew, G.V. Doyle, J.R. Gralow, R.B. Livingston D.F. Hayes. (2014) Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J Clin Oncol*, 32: 3483-9.
108. Menis, J., B. Hasan B. Besse. (2014) New clinical research strategies in thoracic oncology: clinical trial design, adaptive, basket and umbrella trials, new end-points and new evaluations of response. *Eur Respir Rev*, 23: 367-78.
109. Thirunavukarasu, P., C. Talati, S. Munjal, K. Attwood, S.B. Edge V. Francescutti. (2015) Effect of Incorporation of Pretreatment Serum Carcinoembryonic Antigen Levels Into AJCC Staging for Colon Cancer on 5-Year Survival. *JAMA Surg*, 150: 747-55.

110. Schmoll, H.J., E. Van Cutsem, A. Stein, V. Valentini, B. Glimelius, K. Haustermans, B. Nordlinger, C.J. van de Velde, J. Balmana, J. Regula, I.D. Nagtegaal, R.G. Beets-Tan, D. Arnold, F. Ciardiello, P. Hoff, D. Kerr, C.H. Kohne, R. Labianca, T. Price, W. Scheithauer, A. Sobrero, J. Tabernero, D. Aderka, S. Barroso, G. Bodoky, J.Y. Douillard, H. El Ghazaly, J. Gallardo, A. Garin, R. Glynne-Jones, K. Jordan, A. Meshcheryakov, D. Papamichail, P. Pfeiffer, I. Souglakos, S. TurhalA. Cervantes. (2012) ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann Oncol*, 23: 2479-516.
111. Sepulveda, A.R., S.R. Hamilton, C.J. Allegra, W. Grody, A.M. Cushman-Vokoun, W.K. Funkhouser, S.E. Kopetz, C. Lieu, N.M. Lindor, B.D. Minsky, F.A. Monzon, D.J. Sargent, V.M. Singh, J. Willis, J. Clark, C. Colasacco, R.B. Rumble, R. Temple-Smolkin, C.B. VenturaJ.A. Nowak. (2017) Molecular Biomarkers for the Evaluation of Colorectal Cancer. *Am J Clin Pathol*.
112. Romiti, A., S. Raffa, R. Di Rocco, M. Roberto, A. Milano, A. Zullo, L. Leone, D. Ranieri, F. Mazzetta, E. Medda, I. Sarcina, V. Barucca, C. D'Antonio, V. Durante, M. Ferri, M.R. TorrisiP. Marchetti. (2014) Circulating tumor cells count predicts survival in colorectal cancer patients. *J Gastrointestin Liver Dis*, 23: 279-84.
113. Kondo, Y., K. Hayashi, K. Kawakami, Y. Miwa, H. HayashiM. Yamamoto. (2017) KRAS mutation analysis of single circulating tumor cells from patients with metastatic colorectal cancer. *BMC Cancer*, 17: 311.
114. Sorlie, T., C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, T. Thorsen, H. Quist, J.C. Matese, P.O. Brown, D. Botstein, P.E. LonningA.L. Borresen-Dale. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98: 10869-74.
115. Goldhirsch, A., E.P. Winer, A.S. Coates, R.D. Gelber, M. Piccart-Gebhart, B. Thurlimann, H.J. Sennm. Panel. (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*, 24: 2206-23.

116. Sweeney, C., P.S. Bernard, R.E. Factor, M.L. Kwan, L.A. Habel, C.P. Quesenberry, Jr., K. Shakespear, E.K. Weltzien, I.J. Stijleman, C.A. Davis, M.T. Ebbert, A. Castillo, L.H. Kushi B.J. Caan. (2014) Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: differences by age, race, and tumor characteristics. *Cancer Epidemiol Biomarkers Prev*, 23: 714-24.
117. Nielsen, T.O., J.S. Parker, S. Leung, D. Voduc, M. Ebbert, T. Vickery, S.R. Davies, J. Snider, I.J. Stijleman, J. Reed, M.C. Cheang, E.R. Mardis, C.M. Perou, P.S. Bernard M.J. Ellis. (2010) A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*, 16: 5222-32.
118. Louie, R.J., J.E. Tonneson, M. Gowarty, P.P. Goodney, R.J. Barth, Jr. K.M. Rosenkranz. (2015) Complete blood counts, liver function tests, and chest x-rays as routine screening in early-stage breast cancer: value added or just cost? *Breast Cancer Res Treat*, 154: 99-103.
119. Press, M.F., L. Bernstein, P.A. Thomas, L.F. Meisner, J.Y. Zhou, Y. Ma, G. Hung, R.A. Robinson, C. Harris, A. El-Naggar, D.J. Slamon, R.N. Phillips, J.S. Ross, S.R. Wolman K.J. Flom. (1997) HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J Clin Oncol*, 15: 2894-904.
120. Dawood, S., K. Broglio, A.U. Buzdar, G.N. Hortobagyi S.H. Giordano. (2010) Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. *J Clin Oncol*, 28: 92-8.
121. Paik, S., J. Bryant, C. Park, B. Fisher, E. Tan-Chiu, D. Hyams, E.R. Fisher, M.E. Lippman, D.L. Wickerham N. Wolmark. (1998) erbB-2 and response to doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breast cancer. *J Natl Cancer Inst*, 90: 1361-70.
122. Goss, P.E., J.N. Ingle, S. Martino, N.J. Robert, H.B. Muss, M.J. Piccart, M. Castiglione, D. Tu, L.E. Shepherd, K.I. Pritchard, R.B. Livingston, N.E. Davidson, L. Norton, E.A. Perez, J.S. Abrams, P. Therasse, M.J. Palmer J.L. Pater. (2003) A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med*, 349: 1793-802.

123. Jinih, M., N. Relihan, M.A. Corrigan, S. O'Reilly, H.P. Redmond. (2017) Extended Adjuvant Endocrine Therapy in Breast Cancer: Evidence and Update - A Review. *Breast J.*
124. Sgroi, D.C., E. Carney, E. Zarrella, L. Steffel, S.N. Binns, D.M. Finkelstein, J. Szymonifka, A.K. Bhan, L.E. Shepherd, Y. Zhang, C.A. Schnabel, M.G. Erlander, J.N. Ingle, P. Porter, H.B. Muss, K.I. Pritchard, D. Tu, D.L. Rimm, P.E. Goss. (2013) Prediction of late disease recurrence and extended adjuvant letrozole benefit by the HOXB13/IL17BR biomarker. *J Natl Cancer Inst*, 105: 1036-42.
125. Wallden, B., J. Storhoff, T. Nielsen, N. Dowidar, C. Schaper, S. Ferree, S. Liu, S. Leung, G. Geiss, J. Snider, T. Vickery, S.R. Davies, E.R. Mardis, M. Gnant, I. Sestak, M.J. Ellis, C.M. Perou, P.S. Bernard, J.S. Parker. (2015) Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*, 8: 54.
126. Geiss, G.K., R.E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D.L. Dunaway, H.P. Fell, S. Ferree, R.D. George, T. Grogan, J.J. James, M. Maysuria, J.D. Mitton, P. Oliveri, J.L. Osborn, T. Peng, A.L. Ratcliffe, P.J. Webster, E.H. Davidson, L. Hood, K. Dimitrov. (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*, 26: 317-25.
127. Sgroi, D.C., J.A. Chapman, T. Badovinac-Crnjevic, E. Zarrella, S. Binns, Y. Zhang, C.A. Schnabel, M.G. Erlander, K.I. Pritchard, L. Han, L.E. Shepherd, P.E. Goss, M. Pollak. (2016) Assessment of the prognostic and predictive utility of the Breast Cancer Index (BCI): an NCIC CTG MA.14 study. *Breast Cancer Res*, 18: 1.
128. Muller, B.M., E. Keil, A. Lehmann, K.J. Winzer, C. Richter-Ehrenstein, J. Prinzler, N. Bangemann, A. Reles, S. Stadie, W. Schoenegg, J. Eucker, M. Schmidt, F. Lippek, K. Johrens, S. Pahl, B.V. Sinn, J. Budczies, M. Dietel, C. Denkert. (2013) The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. *PLoS One*, 8: e68252.
129. Denkert, C., R. Kronenwett, W. Schlake, K. Bohmann, R. Penzel, K.E. Weber, H. Hofler, U. Lehmann, P. Schirmacher, K. Specht, M. Rudas, H.H. Kreipe, P. Schraml, G. Schlake, Z. Bago-Horvath, F. Tiecke, Z. Varga, H. Moch, M. Schmidt, J. Prinzler, D. Kerjaschki, B.V. Sinn, B.M. Muller, M. Filipits, C.

- Petry M, Dietel. (2012) Decentral gene expression analysis for ER+/Her2- breast cancer: results of a proficiency testing program for the EndoPredict assay. *Virchows Arch*, 460: 251-9.
130. Dubsy, P., M. Filipits, R. Jakesz, M. Rudas, C.F. Singer, R. Greil, O. Dietze, I. Luisser, E. Klug, R. Sedivy, M. Bachner, D. Mayr, M. Schmidt, M.C. Gehrman, C. Petry, K.E. Weber, R. Kronenwett, J.C. Brase, M. Gnant, B. Austrian G. Colorectal Cancer Study. (2013) EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol*, 24: 640-7.
131. Fitzal, F., M. Filipits, M. Rudas, R. Greil, O. Dietze, H. Samonigg, S. Lax, W. Herz, P. Dubsy, R. Bartsch, R. Kronenwett M. Gnant. (2015) The genomic expression test EndoPredict is a prognostic tool for identifying risk of local recurrence in postmenopausal endocrine receptor-positive, her2neu-negative breast cancer patients randomised within the prospective ABCSG 8 trial. *Br J Cancer*, 112: 1405-10.
132. Sparano, J.A., R.J. Gray, D.F. Makower, K.I. Pritchard, K.S. Albain, D.F. Hayes, C.E. Geyer, Jr., E.C. Dees, E.A. Perez, J.A. Olson, Jr., J. Zujewski, T. Lively, S.S. Badve, T.J. Saphner, L.I. Wagner, T.J. Whelan, M.J. Ellis, S. Paik, W.C. Wood, P. Ravdin, M.M. Keane, H.L. Gomez Moreno, P.S. Reddy, T.F. Goggins, I.A. Mayer, A.M. Brufsky, D.L. Toppmeyer, V.G. Kaklamani, J.N. Atkins, J.L. Berenberg G.W. Sledge. (2015) Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*, 373: 2005-14.
133. Cardoso, F., L.J. van't Veer, J. Bogaerts, L. Slaets, G. Viale, S. Delaloge, J.Y. Pierga, E. Brain, S. Causeret, M. DeLorenzi, A.M. Glas, V. Golfopoulos, T. Goulioti, S. Knox, E. Matos, B. Meulemans, P.A. Neijenhuis, U. Nitz, R. Passalacqua, P. Ravdin, I.T. Rubio, M. Saghatchian, T.J. Smilde, C. Sotiriou, L. Stork, C. Straehle, G. Thomas, A.M. Thompson, J.M. van der Hoeven, P. Vuylsteke, R. Bernards, K. Tryfonidis, E. Rutgers, M. Piccart M. Investigators. (2016) 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*, 375: 717-29.
134. Liedtke, C., C. Hatzis, W.F. Symmans, C. Desmedt, B. Haibe-Kains, V. Valero, H. Kuerer, G.N. Hortobagyi, M. Piccart-Gebhart, C. Sotiriou L. Pusztai. (2009)

- Genomic grade index is associated with response to chemotherapy in patients with breast cancer. *J Clin Oncol*, 27: 3185-91.
135. Szasz, A.M., Q. Li, A.C. Eklund, Z. Sztupinszki, A. Rowan, A.M. Tokes, B. Szekely, A. Kiss, M. Szendroi, B. Gyorffy, Z. Szallasi, C. Swanton, J. Kulka. (2013) The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. *PLoS One*, 8: e56707.
 136. Cruz, J.A.D.S. Wishart. (2007) Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*, 2: 59-77.
 137. Fisher, R.A. (1936) The use of multiple measurements in taxonomic problems. *Ann Eugen*, 7: 179-188.
 138. Singh, D., P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, P. Tamayo, A.A. Renshaw, A.V. D'Amico, J.P. Richie, E.S. Lander, M. Loda, P.W. Kantoff, T.R. Golub, W.R. Sellers. (2002) Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1: 203-209.
 139. Furey, T.S., N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, D. Haussler. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16: 906-914.
 140. Quinlan, J.R. (1987) Simplifying Decision Trees. *Int J Man Mach Stud*, 27: 221-234.
 141. Esposito, F., D. Malerba, G. Semeraro. (1995) Simplifying decision trees by pruning and grafting: New results. *Machine Learning: Ecml-95*, 912: 287-290.
 142. Schmidhuber, J. (2015) Deep learning in neural networks: An overview. *Neural Netw*, 61: 85-117.
 143. Breiman, L. (2001) Random forests. *Machine Learning*, 45: 5-32.
 144. Shi, T.S. Horvath. (2006) Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics*, 15: 118-138.
 145. Lance, G.N.W.T. Williams. (1971) New Divisive Classificatory Program for Mixed Data. *Computer Journal*, 14: 154-&.
 146. Bedrick, E.J., J. Lapidus, J.F. Powell. (2000) Estimating the Mahalanobis distance from mixed continuous and discrete data. *Biometrics*, 56: 394-401.

147. Jeffery, I.B., D.G. Higgins A.C. Culhane. (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC bioinformatics*, 7: 359.
148. Tusher, V.G., R. Tibshirani G. Chu. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, 98: 5116-21.
149. Breitling, R., P. Armengaud, A. Amtmann P. Herzyk. (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters*, 573: 83-92.
150. Saad, E.D.A. Katz. (2009) Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Ann Oncol*, 20: 460-4.
151. Miller, A.B., B. Hoogstraten, M. Staquet A. Winkler. (1981) Reporting results of cancer treatment. *Cancer*, 47: 207-14.
152. Eisenhauer, E.A., P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe J. Verweij. (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*, 45: 228-47.
153. Booth, C.M.E.A. Eisenhauer. (2012) Progression-free survival: meaningful or simply measurable? *J Clin Oncol*, 30: 1030-3.
154. Dinse, G.E.S.W. Lagakos. (1982) Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. *Biometrics*, 38: 921-32.
155. Mantel, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50: 163-70.
156. Cox, D. (1986) Citation-Classic - Regression-Models and Life-Tables. *Current Contents/Social & Behavioral Sciences*: 16-16.
157. Györfy, B., A. Györfy Z. Tulassay. (2005) [The problem of multiple testing and solutions for genome-wide studies]. *Orv Hetil*, 146: 559-63.
158. Bland, J.M.D.G. Altman. (1995) Multiple significance tests: the Bonferroni method. *BMJ*, 310: 170.
159. Perneger, T.V. (1998) What's wrong with Bonferroni adjustments. *BMJ*, 316: 1236-8.

160. Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi, I. Golani. (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, 125: 279-84.
161. Fernando, R.L., D. Nettleton, B.R. Southey, J.C. Dekkers, M.F. Rothschild, M. Soller. (2004) Controlling the proportion of false positives in multiple dependent tests. *Genetics*, 166: 611-9.
162. Schena, M., D. Shalon, R.W. Davis, P.O. Brown. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270: 467-70.
163. R Development Core Team, (2011), R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
164. Hubbell, E., W.M. Liu, R. Mei. (2002) Robust estimators for expression analysis. *Bioinformatics*, 18: 1585-92.
165. Irizarry, R.A., B.M. Bolstad, F. Collin, L.M. Cope, B. Hobbs, T.P. Speed. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research*, 31: e15.
166. Györfy, B., Z. Benke, A. Lanczky, B. Balazs, Z. Szallasi, J. Timar, R. Schafer. (2012) RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res Treat*, 132: 1025-34.
167. Wilson, C.L., C.J. Miller. (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, 21: 3683-5.
168. Györfy, B., R. Schafer. (2009) Meta-analysis of gene expression profiles related to relapse-free survival in 1,079 breast cancer patients. *Breast Cancer Res Treat*, 118: 433-41.
169. Li, Q., N.J. Birkbak, B. Györfy, Z. Szallasi, A.C. Eklund. (2011) Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*, 12: 474.
170. Leek, J.T., R.B. Scharpf, H.C. Bravo, D. Simcha, B. Langmead, W.E. Johnson, D. Geman, K. Baggerly, R.A. Irizarry. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*, 11: 733-9.

171. Chen, C., K. Grennan, J. Badner, D. Zhang, E. Gershon, L. JinC. Liu. (2011) Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One*, 6: e17238.
172. Nygaard, V., E.A. RodlandE. Hovig. (2016) Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, 17: 29-39.
173. Liberati, A., D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gotzsche, J.P. Ioannidis, M. Clarke, P.J. Devereaux, J. KleijnenD. Moher. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 339: b2700.
174. Dalerba, P., D. Sahoo, S. Paik, X. Guo, G. Yothers, N. Song, N. Wilcox-Fogel, E. Forgo, P.S. Rajendran, S.P. Miranda, S. Hisamori, J. Hutchison, T. Kalisky, D. Qian, N. Wolmark, G.A. Fisher, M. van de RijnM.F. Clarke. (2016) CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med*, 374: 211-22.
175. Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. BryantN. Wolmark. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*, 351: 2817-26.
176. Sotiriou, C., P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D. Nuyten, M. Buyse, M.J. Van de Vijver, J. Bergh, M. PiccartM. Delorenzi. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*, 98: 262-72.
177. van 't Veer, L.J., H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. BernardsS.H. Friend. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415: 530-6.

178. Mihaly, Z.B. Gyorffy. (2013) Improving Pathological Assessment of Breast Cancer by Employing Array-Based Transcriptome Analysis. *Microarrays* (Basel), 2: 228-42.
179. Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. ViloM. Vingron. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*, 29: 365-71.
180. Edgar, R., M. DomrachevA.E. Lash. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, 30: 207-10.
181. McCall, M.N., B.M. BolstadR.A. Irizarry. (2010) Frozen robust multiarray analysis (fRMA). *Biostatistics*, 11: 242-53.
182. Callari, M., A. Lembo, G. Bianchini, V. Musella, V. Cappelletti, L. Gianni, M.G. DaidoneP. Provero. (2014) Accurate data processing improves the reliability of Affymetrix gene expression profiles from FFPE samples. *PLoS One*, 9: e86511.
183. Kuhn, M. (2008) Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28: 1-26.
184. Yu, G., L.G. Wang, Y. HanQ.Y. He. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 16: 284-7.
185. Luo, W.C. Brouwer. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29: 1830-1.
186. Agrawal, N., P.V. Dasaradhi, A. Mohammed, P. Malhotra, R.K. BhatnagarS.K. Mukherjee. (2003) RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev*, 67: 657-85.
187. Budinska, E., V. Popovici, S. Tejpar, G. D'Ario, N. Lapique, K.O. Sikora, A.F. Di Narzo, P. Yan, J.G. Hodgson, S. Weinrich, F. Bosman, A. RothM. Delorenzi. (2013) Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol*, 231: 63-76.
188. Dekervel, J., D. Hompes, H. van Malenstein, D. Popovic, X. Sagaert, B. De Moor, E. Van Cutsem, A. D'Hoore, C. VerslypeJ. van Pelt. (2014) Hypoxia-driven gene

- expression is an independent prognostic factor in stage II and III colon cancer patients. *Clin Cancer Res*, 20: 2159-68.
189. Chang, W., X. Gao, Y. Han, Y. Du, Q. Liu, L. Wang, X. Tan, Q. Zhang, Y. Liu, Y. Zhu, Y. Yu, X. Fan, H. Zhang, W. Zhou, J. Wang, C. FuG. Cao. (2014) Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. *Gut*, 63: 1457-67.
 190. Carter, S.L., A.C. Eklund, I.S. Kohane, L.N. HarrisZ. Szallasi. (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet*, 38: 1043-8.
 191. Guinney, J., R. Dienstmann, X. Wang, A. de Reynies, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster, E. Fessler, E.M.F. De Sousa, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. VermeulenS. Tejpar. (2015) The consensus molecular subtypes of colorectal cancer. *Nat Med*, 21: 1350-6.
 192. Dienstmann, R., L. Vermeulen, J. Guinney, S. Kopetz, S. TejparJ. Tabernero. (2017) Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer*, 17: 79-92.
 193. Agesen, T.H., A. Sveen, M.A. Merok, G.E. Lind, A. Nesbakken, R.I. SkotheimR.A. Lothe. (2012) ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut*, 61: 1560-7.
 194. Sveen, A., T.H. Agesen, A. Nesbakken, G.I. Meling, T.O. Rognum, K. Liestol, R.I. SkotheimR.A. Lothe. (2012) ColoGuidePro: a prognostic 7-gene expression signature for stage III colorectal cancer patients. *Clin Cancer Res*, 18: 6001-10.
 195. Tusher, V.G., R. TibshiraniG. Chu. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, 98: 5116-21.
 196. Tibshirani, R., T. Hastie, B. NarasimhanG. Chu. (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A*, 99: 6567-72.

197. Sadanandam, A., C.A. Lyssiotis, K. Homicsko, E.A. Collisson, W.J. Gibb, S. Wullschleger, L.C. Ostos, W.A. Lannon, C. Grotzinger, M. Del Rio, B. Lhermitte, A.B. Olshen, B. Wiedenmann, L.C. Cantley, J.W. Gray, D. Hanahan. (2013) A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med*, 19: 619-25.
198. De Sousa E Melo, F., X. Wang, M. Jansen, E. Fessler, A. Trinh, L.P.M.H. de Rooij, J.H. de Jong, O.J. de Boer, R. van Leersum, M.F. Bijlsma, H. Rodermond, M. van der Heijden, C.J.M. van Noesel, J.B. Tuynman, E. Dekker, F. Markowitz, J.P. Medema, L. Vermeulen. (2013) Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med*, 19: 614-618.
199. Marisa, L., A. de Reynies, A. Duval, J. Selves, M.P. Gaub, L. Vescovo, M.C. Etienne-Grimaldi, R. Schiappa, D. Guenot, M. Ayadi, S. Kirzin, M. Chazal, J.F. Flejou, D. Benchimol, A. Berger, A. Lagarde, E. Pencreach, F. Piard, D. Elias, Y. Parc, S. Olschwang, G. Milano, P. Laurent-Puig, V. Boige. (2013) Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*, 10: e1001453.
200. Oh, S.C., Y.Y. Park, E.S. Park, J.Y. Lim, S.M. Kim, S.B. Kim, J. Kim, S.C. Kim, I.S. Chu, J.J. Smith, R.D. Beauchamp, T.J. Yeatman, S. Kopetz, J.S. Lee. (2012) Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. *Gut*, 61: 1291-8.
201. Merlos-Suarez, A., F.M. Barriga, P. Jung, M. Iglesias, M.V. Cespedes, D. Rossell, M. Sevillano, X. Hernando-Momblona, V. da Silva-Diz, P. Munoz, H. Clevers, E. Sancho, R. Mangués, E. Batlle. (2011) The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*, 8: 511-24.
202. Jorissen, R.N., P. Gibbs, M. Christie, S. Prakash, L. Lipton, J. Desai, D. Kerr, L.A. Aaltonen, D. Arango, M. Kruhoffer, T.F. Orntoft, C.L. Andersen, M. Gruidl, V.P. Kamath, S. Eschrich, T.J. Yeatman, O.M. Sieber. (2009) Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. *Clin Cancer Res*, 15: 7642-7651.

203. Webber, E.M., J.S. LinP.W. Evelyn. (2010) Oncotype DX tumor gene expression profiling in stage II colon cancer. Application: prognostic, risk prediction. *PLoS Curr*, 2.
204. Clark-Langone, K.M., C. Sangli, J. KrishnakumarD. Watson. (2010) Translating tumor biology into personalized treatment planning: analytical performance characteristics of the Oncotype DX Colon Cancer Assay. *BMC Cancer*, 10: 691.
205. Srivastava, G., L.A. Renfro, R.J. Behrens, M. Lopatin, C. Chao, G.S. Soori, S.R. Dakhil, R.B. Mowat, J.P. Kuebler, G. Kim, M. Mazurczak, M. LeeS.R. Alberts. (2014) Prospective multicenter study of the impact of oncotype DX colon cancer assay results on treatment recommendations in stage II colon cancer patients. *Oncologist*, 19: 492-7.
206. You, Y.N., R.B. RustinJ.D. Sullivan. (2015) Oncotype DX((R)) colon cancer assay for prediction of recurrence risk in patients with stage II and III colon cancer: A review of the evidence. *Surg Oncol*, 24: 61-6.
207. Lenehan, P.F., L.A. Boardman, D. Riegert-Johnson, G. De Petris, D.W. Fry, J. Ohrnberger, E.R. Heyman, B. Gerard, A.A. AlmalW.P. Worzel. (2012) Generation and external validation of a tumor-derived 5-gene prognostic signature for recurrence of lymph node-negative, invasive colorectal carcinoma. *Cancer*, 118: 5234-44.
208. Popovici, V., E. Budinska, S. Tejpar, S. Weinrich, H. Estrella, G. Hodgson, E. Van Cutsem, T. Xie, F.T. Bosman, A.D. RothM. Delorenzi. (2012) Identification of a poor-prognosis BRAF-mutant-like population of patients with colon cancer. *J Clin Oncol*, 30: 1288-95.
209. Bokemeyer, C., E. Van Cutsem, P. Rougier, F. Ciardiello, S. Heeger, M. Schlichting, I. CelikC.H. Kohne. (2012) Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials. *Eur J Cancer*, 48: 1466-75.
210. Schetter, A.J., G.H. Nguyen, E.D. Bowman, E.A. Mathe, S.T. Yuen, J.E. Hawkes, C.M. Croce, S.Y. LeungC.C. Harris. (2009) Association of inflammation-related and microRNA gene expression with cancer-specific mortality of colon adenocarcinoma. *Clin Cancer Res*, 15: 5878-87.

211. Kim, S.K., S.Y. Kim, J.H. Kim, S.A. Roh, D.H. Cho, Y.S. Kim J.C. Kim. (2014) A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol*, 8: 1653-66.
212. Yuen, H.F., C.M. McCrudden, Y.H. Huang, J.M. Tham, X. Zhang, Q. Zeng, S.D. Zhang W. Hong. (2013) TAZ expression as a prognostic indicator in colorectal cancer. *PLoS One*, 8: e54211.
213. Jiang, Y., G. Casey, I.C. Lavery, Y. Zhang, D. Talantov, M. Martin-McGreevy, M. Skacel, E. Manilich, A. Mazumder, D. Atkins, C.P. Delaney Y. Wang. (2008) Development of a clinically feasible molecular assay to predict recurrence of stage II colon cancer. *J Mol Diagn*, 10: 346-54.
214. Watanabe, T., T. Kobunai, Y. Yamamoto, K. Matsuda, S. Ishihara, K. Nozawa, H. Yamada, T. Hayama, E. Inoue, J. Tamura, H. Iinuma, T. Akiyoshi T. Muto. (2012) Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *J Clin Oncol*, 30: 2256-64.
215. De Sousa, E.M.F., X. Wang, M. Jansen, E. Fessler, A. Trinh, L.P. de Rooij, J.H. de Jong, O.J. de Boer, R. van Leersum, M.F. Bijlsma, H. Rodermond, M. van der Heijden, C.J. van Noesel, J.B. Tuynman, E. Dekker, F. Markowitz, J.P. Medema L. Vermeulen. (2013) Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med*, 19: 614-8.
216. Tomaskovic-Crook, E., E.W. Thompson J.P. Thiery. (2009) Epithelial to mesenchymal transition and breast cancer. *Breast Cancer Res*, 11: 213.
217. Lewis-Wambi, J.S., H.E. Cunliffe, H.R. Kim, A.L. Willis V.C. Jordan. (2008) Overexpression of CEACAM6 promotes migration and invasion of oestrogen-deprived breast cancer cells. *Eur J Cancer*, 44: 1770-9.
218. Prica, F., T. Radon, Y. Cheng T. Crnogorac-Jurcevic. (2016) The life and works of S100P - from conception to cancer. *Am J Cancer Res*, 6: 562-76.
219. Huang da, W., B.T. Sherman R.A. Lempicki. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4: 44-57.
220. Chen, D.H. Yang. (2015) Integrated analysis of differentially expressed genes in breast cancer pathogenesis. *Oncol Lett*, 9: 2560-2566.

221. Salazar, R., P. Roepman, G. Capella, V. Moreno, I. Simon, C. Dreezen, A. Lopez-Doriga, C. Santos, C. Marijnen, J. Westerga, S. Bruin, D. Kerr, P. Kuppen, C. van de Velde, H. Morreau, L. Van Velthuysen, A.M. Glas, L.J. Van't Veer, R. Tollenaar. (2011) Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J Clin Oncol*, 29: 17-24.
222. Kaprio, T., J. Hagstrom, H. Mustonen, S. Koskensalo, L.C. Andersson, C. Haglund. (2014) REG4 independently predicts better prognosis in non-mucinous colorectal cancer. *PLoS One*, 9: e109600.
223. Uen, Y.H., C.L. Fang, Y.C. Hseu, P.C. Shen, H.L. Yang, K.S. Wen, S.T. Hung, L.H. Wang, K.Y. Lin. (2015) VAV3 oncogene expression in colorectal cancer: clinical aspects and functional characterization. *Sci Rep*, 5: 9360.
224. Pan, W., Y. Cheng, H. Zhang, B. Liu, X. Mo, T. Li, L. Li, X. Cheng, L. Zhang, J. Ji, P. Wang, W. Han. (2014) CSBF/C10orf99, a novel potential cytokine, inhibits colon cancer cell growth through inducing G1 arrest. *Sci Rep*, 4: 6812.
225. Zhang, S.D., C.M. McCrudden, H.F. Yuen, K.L. Leung, W.J. Hong, H.F. Kwok. (2016) Association between the expression levels of TAZ, AXL and CTGF and clinicopathological parameters in patients with colon cancer. *Oncol Lett*, 11: 1223-1229.
226. Wikberg, M.L., S. Edin, I.V. Lundberg, B. Van Guelpen, A.M. Dahlin, J. Rutegard, R. Stenling, A. Oberg, R. Palmqvist. (2013) High intratumoral expression of fibroblast activation protein (FAP) in colon cancer is associated with poorer patient prognosis. *Tumour Biol*, 34: 1013-20.
227. Serova, M., L. Astorgues-Xerri, I. Bieche, S. Albert, M. Vidaud, K.A. Benhadji, S. Emami, D. Vidaud, P. Hammel, N. Theou-Anton, C. Gespach, S. Faivre, E. Raymond. (2010) Epithelial-to-mesenchymal transition and oncogenic Ras expression in resistance to the protein kinase C beta inhibitor enzastaurin in colon cancer cells. *Mol Cancer Ther*, 9: 1308-17.
228. Jeong, D.H., W.R. Kim, B.S. Min, Y.W. Kim, M.K. Song, N.K. Kim. (2015) Validation of a quantitative 12-multigene expression assay (Oncotype DX((R)) Colon Cancer Assay) in Korean patients with stage II colon cancer: implication of ethnic differences contributing to differences in gene expression. *Onco Targets Ther*, 8: 3817-25.

229. Song, N., K.L. Pogue-Geile, P.G. Gavin, G. Yothers, S.R. Kim, N.L. Johnson, C. Lipchik, C.J. Allegra, N.J. Petrelli, M.J. O'Connell, N. Wolmark S. Paik. (2016) Clinical Outcome From Oxaliplatin Treatment in Stage II/III Colon Cancer According to Intrinsic Subtypes: Secondary Analysis of NSABP C-07/NRG Oncology Randomized Clinical Trial. *JAMA Oncol*, 2: 1162-9.
230. Rosty, C., E.J. Williamson, M. Clendenning, R.J. Walters, A.K. Win, M.A. Jenkins, J.L. Hopper, I.M. Winship, M.C. Southey, G.G. Giles, D.R. English D.D. Buchanan. (2014) Should the grading of colorectal adenocarcinoma include microsatellite instability status? *Hum Pathol*, 45: 2077-84.
231. Majek, O., A. Gondos, L. Jansen, K. Emrich, B. Holleczeck, A. Katalinic, A. Nennecke, A. Eberle, H. Brenner G.C.S.W. Group. (2013) Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. *PLoS One*, 8: e68077.
232. Petrelli, F., G. Tomasello, K. Borgonovo, M. Ghidini, L. Turati, P. Dallera, R. Passalacqua, G. Sgroi S. Barni. (2016) Prognostic Survival Associated With Left-Sided vs Right-Sided Colon Cancer: A Systematic Review and Meta-analysis. *JAMA Oncol*.
233. Sadanandam, A., X. Wang, E.M.F. de Sousa, J.W. Gray, L. Vermeulen, D. Hanahan J.P. Medema. (2014) Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications. *Cell Cycle*, 13: 353-7.
234. Gill, S., C. Loprinzi, H. Kennecke, A. Grothey, G. Nelson, R. Woods, C. Speers, S.R. Alberts, A. Bardia, M.J. O'Connell D.J. Sargent. (2011) Prognostic web-based models for stage II and III colon cancer: A population and clinical trials-based validation of numeracy and adjuvant! online. *Cancer*, 117: 4155-65.
235. Dunne, P.D., D.G. McArt, C.A. Bradley, P.G. O'Reilly, H.L. Barrett, R. Cummins, T. O'Grady, K. Arthur, M.B. Loughrey, W.L. Allen, S.S. McDade, D.J. Waugh, P.W. Hamilton, D.B. Longley, E.W. Kay, P.G. Johnston, M. Lawler, M. Salto-Tellez S. Van Schaeybroeck. (2016) Challenging the Cancer Molecular Stratification Dogma: Intratumoral Heterogeneity Undermines Consensus Molecular Subtypes and Potential Diagnostic Value in Colorectal Cancer. *Clin Cancer Res*, 22: 4095-104.

236. Kung, P.P., R. Martinez, Z. Zhu, M. Zager, A. Blasina, I. Rymer, J. Hallin, M. Xu, C. Carroll, J. Chionis, P. Wells, K. Kozminski, J. Fan, O. Guicherit, B. Huang, M. Cui, C. Liu, Z. Huang, A. Sistla, J. Yang B.W. Murray. (2014) Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther*, 13: 2104-15.
237. Koca, B., B. Kuru, N. Ozen, S. Yoruker Y. Bek. (2014) A breast cancer nomogram for prediction of non-sentinel node metastasis - validation of fourteen existing models. *Asian Pac J Cancer Prev*, 15: 1481-8.
238. Van Zee, K.J., D.M. Manasseh, J.L. Bevilacqua, S.K. Boolbol, J.V. Fey, L.K. Tan, P.I. Borgen, H.S. Cody, 3rd M.W. Kattan. (2003) A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. *Ann Surg Oncol*, 10: 1140-51.
239. Barranger, E., C. Coutant, A. Flahault, Y. Delpech, E. Darai S. Uzan. (2005) An axilla scoring system to predict non-sentinel lymph node status in breast cancer patients with sentinel lymph node involvement. *Breast Cancer Res Treat*, 91: 113-9.
240. Kohrt, H.E., R.A. Olshen, H.R. Bermas, W.H. Goodson, D.J. Wood, S. Henry, R.V. Rouse, L. Bailey, V.J. Philben, F.M. Dirbas, J.J. Dunn, D.L. Johnson, I.L. Wapnir, R.W. Carlson, F.E. Stockdale, N.M. Hansen, S.S. Jeffrey S.L.N.S. Bay Area. (2008) New models and online calculator for predicting non-sentinel lymph node status in sentinel lymph node positive breast cancer patients. *BMC Cancer*, 8: 66.
241. Hessman, C.J., A.M. Naik, N.M. Kearney, A.J. Jensen, B.S. Diggs, M.L. Troxell J.T. Vetto. (2011) Comparative validation of online nomograms for predicting nonsentinel lymph node status in sentinel lymph node-positive breast cancer. *Arch Surg*, 146: 1035-40.
242. Kuo, Y.L., W.C. Chen, W.J. Yao, L. Cheng, H.P. Hsu, H.W. Lai, S.J. Kuo, D.R. Chen T.W. Chang. (2013) Validation of Memorial Sloan-Kettering Cancer Center nomogram for prediction of non-sentinel lymph node metastasis in sentinel lymph node positive breast cancer patients an international comparison. *Int J Surg*, 11: 538-43.

243. Huang, E., S.H. Cheng, H. Dressman, J. Pittman, M.H. Tsou, C.F. Horng, A. Bild, E.S. Iversen, M. Liao, C.M. Chen, M. West, J.R. Nevins, A.T. Huang. (2003) Gene expression predictors of breast cancer outcomes. *Lancet*, 361: 1590-6.
244. Weigelt, B., L.F. Wessels, A.J. Bosma, A.M. Glas, D.S. Nuyten, Y.D. He, H. Dai, J.L. Peterse, L.J. van't Veer. (2005) No common denominator for breast cancer lymph node metastasis. *Br J Cancer*, 93: 924-32.
245. Smeets, A., A. Daemen, I. Vanden Bempt, O. Gevaert, B. Claes, H. Wildiers, R. Drijckoningen, P. Van Hummelen, D. Lambrechts, B. De Moor, P. Neven, C. Sotiriou, T. Vandorpe, R. Paridaens, M.R. Christiaens. (2011) Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs. *Breast Cancer Res Treat*, 129: 767-76.
246. Nakauchi, C., Y. Naoi, K. Shimazu, R. Tsunashima, M. Nishio, N. Maruyama, A. Shimomura, N. Kagara, M. Shimoda, S.J. Kim, S. Noguchi. (2014) Development of a prediction model for lymph node metastasis in luminal A subtype breast cancer: the possibility to omit sentinel lymph node biopsy. *Cancer Lett*, 353: 52-8.
247. Shen, J., R. Samul, R.L. Silva, H. Akiyama, H. Liu, Y. Saishin, S.F. Hackett, S. Zinnen, K. Kossen, K. Fosnaugh, C. Vargeese, A. Gomez, K. Bouhana, R. Aitchison, P. Pavco, P.A. Campochiaro. (2006) Suppression of ocular neovascularization with siRNA targeting VEGF receptor 1. *Gene Ther*, 13: 225-34.
248. Abou-Alfa, G.K., J.H. Yoon, M. Modiano, B.Y. Ryoo, T. Yau, B. Freilich, J. Knox, M. Ly, H. Ahmed, S. Gahir, D. Niforos, M. Kowalski, R.K. Kelley. (2016) An open-label, multi-center, phase I/II, dose escalation study of IV TKM-080301 in subjects with advanced hepatocellular carcinoma. *Eur J Cancer*, 69: S22-S22.
249. Schultheis, B., D. Strumberg, J. Kuhlmann, M. Wolf, K. Link, T. Seufferlein, J. Kaufmann, F. Gebhardt, N. Bruyniks, U. Pelzer. (2016) A phase Ib/IIa study of combination therapy with gemcitabine and Atu027 in patients with locally advanced or metastatic pancreatic adenocarcinoma. *J Clin Oncol*, 34.
250. Wagner, M.J., R. Mitra, M.J. McArthur, W. Baze, K. Barnhart, S.Y. Wu, C. Rodriguez-Aguayo, X.N. Zhang, R.L. Coleman, G. Lopez-Berestein, A.K. Sood.

- (2017) Preclinical Mammalian Safety Studies of EPHARNA (DOPC Nanoliposomal EphA2-Targeted siRNA). *Mol Cancer Ther*, 16: 1114-1123.
251. Golan, T., E.Z. Khvalevsky, A. Hubert, R.M. Gabai, N. Hen, A. Segal, A. Domb, G. Harari, E. Ben David, S. Raskin, Y. Goldes, E. Goldin, R. Eliakim, M. Lahav, Y. Kopleman, A. Dancour, A. ShemiE. Galun. (2015) RNAi therapy targeting KRAS in combination with chemotherapy for locally advanced pancreatic cancer patients. *Oncotarget*, 6: 24560-24570.
252. Salzberg, M.O., P. Triozzi, M. Kooshki, A.T. Alistar, R.L. Bitting, J. Fenstermaker, G. LametschwandtnerH. Loibner. (2016) Adoptive cellular immunotherapy with APN401, autologous Cbl-b-silenced peripheral blood mononuclear cells, in patients with solid tumors. *J Clin Oncol*, 34.

11 Saját publikációk jegyzék

11.1 Disszertációhoz kapcsolódó publikációk jegyzéke

Sztupinszki, Z., B. Gyorffy. (2016) Colon cancer subtypes: concordance, effect on survival and selection of the most representative preclinical models. *Sci Rep*, 6: 37169.

IF=5,228

Munkacsy, G.*, **Z. Sztupinszki***, P. Herman, B. Ban, Z. Penzvalto, N. SzarvasB. Gyorffy. (2016) Validation of RNAi Silencing Efficiency Using Gene Array Data shows 18.5% Failure Rate across 429 Independent Experiments. *Mol Ther Nucleic Acids*, 5: e366 IF=5,048

Gyorffy, B., T. Karn, **Z. Sztupinszki**, B. Wetz, V. MullerL. Pusztai. (2015) Dynamic classification using case-specific training cohorts outperforms static gene expression signatures in breast cancer. *Int J Cancer*, 136: 2091-8. IF=5,53

11.2 Disszertációtól független publikációk jegyzéke

Beres, N.J., Z. Kiss, **Z. Sztupinszki**, G. Lendvai, A. Arato, E. Sziksz, A. Vannay, A.J. Szabo, K.E. Muller, A. Cseh, K. BorosG. Veres. (2016) Altered mucosal expression of microRNAs in pediatric patients with inflammatory bowel disease. *Dig Liver Dis*.

IF=2,719

Lee, S., K. Vargova, I. Hizoh, Z. Horvath, P. Gulacsi-Bardos, **Z. Sztupinszki**, A. Apro, A. Kovacs, I. Preda, E. Toth-ZsambokiR.G. Kiss. (2014) High on clopidogrel treatment platelet reactivity is frequent in acute and rare in elective stenting and can be functionally overcome by switch of therapy. *Thromb Res*, 133: 257-64. IF=2,32

Mihály, Z., **Z. Sztupinszki**, A. Szendrői, M. Szűcs, P. Nyirády, B. Gyórfy. (2013) A metasztatizáló világossejtes veserák prognózisának előrejelzése microarray vizsgálatok alapján. *Uroonkológia*, 10: 78-84.

Szasz, A.M., B. Acs, E. Agoston, **Z. Sztupinszki**, A.M. Tokes, L. Szitty, B. Szekely, M. Szendroi, Q. Li, L. Harsanyi, J. Timar, Z. Szallasi, C. Swanton, B. GyorffyJ. Kulka.

(2013) [Simplified, low-cost gene expression profiling for the prediction of outcome in breast cancer based on routine histologic specimens]. *Orv Hetil*, 154: 627-32. IF= 0,291

Szasz, A.M., Q. Li, A.C. Eklund, **Z. Sztupinszki**, A. Rowan, A.M. Tokes, B. Szekely, A. Kiss, M. Szendroi, B. Gyorffy, Z. Szallasi, C. SwantonJ. Kulka. (2013) The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. *PLoS One*, 8: e56707. IF=3,057

Penzvalto, Z., B. Tegze, A.M. Szasz, **Z. Sztupinszki**, I. Liko, A. Szendroi, R. SchaferB. Gyorffy. (2013) Identifying resistance mechanisms against five tyrosine kinase inhibitors targeting the ERBB/RAS pathway in 45 cancer cell lines. *PLoS One*, 8: e59503. IF=3,057

Mihaly, Z., **Z. Sztupinszki**, P. SurowiakB. Gyorffy. (2012) A comprehensive overview of targeted therapy in metastatic renal cell carcinoma. *Curr Cancer Drug Targets*, 12: 857-72. IF=3,707

12 Köszönetnyilvánítás

Elsőként témavezetőmnek, Dr. Gyórfy Baláznak szeretnék köszönetet mondani, aki lehetőséget biztosított munkám elvégzéséhez, a munkámat javaslataival és tanácsaival segítette. Szeretnék köszönetet mondani az I. Sz. Gyermekgyógyászati Klinika igazgatójának, Dr. Prof. Tulassay Tivadarnak és Dr. Prof. Szabó Andrásnak a II. Sz. Gyermekgyógyászati Klinika igazgatójának, hogy kutatásaimat a Gyermekgyógyászati Klinikán végezhettem. Köszönettel tartozom a munkacsoport tagjainak, Pongor Lőrinc, Pénzváltó Zsófia PhD hallgató társaimnak, illetve a laboratóriumban dolgozó munkatársaimnak, akik munkámat segítették.

Köszönöm továbbá minden szerzőtársamnak, hogy munkájával hozzájárult publikációink kialakításához. A PhD-m alatt több külföldi tanulmányúton részt vehettem, amiért köszönettel tartozom témavezetőmnek, aki ebben mindig támogatott, a fogadó intézményekben dolgozó témavezetőimnek és közvetlen kollégáimnak. Végül köszönöm, szüleimnek és barátaimnak a támogatást.

13 Függelékek

13.1 Függelék 1.

	Budinska	Chang95	CMS	ColoGuideEx	CRCAssigner-786	DeSousa	Marisa	ODXcolon	Popovici	Yuen3
Budinska	1	0.38	0.56	0.18	0.16	0.6	0.54	0.38	0.46	0.24
Chang95		1	0.46	0.09	0.1	0.44	0.45	0.42	0.12	0.44
CMS			1	0.18	0.18	0.72	0.61	0.41	0.48	0.31
ColoGuideEx				1	0.13	0.14	0.19	0.03	0.2	0.12
CRCAssigner-786					1	0.19	0.22	0.1	0.09	0.11
DeSousa						1	0.66	0.43	0.47	0.35
Marisa							1	0.43	0.46	0.3
ODXcolon								1	0.15	0.3
Popovici									1	0.09
Yuen3										1

13.2 Függelék 2.

Bélrendszeri daganatból származó sejtvonalak legfontosabb altípusokba való besorolása

Sejtvonal	Budinska	Chang95	CMS	CRCAssigner-786	DeSousa	Marisa	Merlos-Lg5-ISC	ODXcolon	Popovici	Yuen3	minták száma
HT-29	D	low	-	-	1	-	high	low	WT	low	4
HCT-116	D	low	CMS4	1	2	-	high	low	WT	int-l	4
SW480	-	-	-	-	3	-	int	low	WT	high	5
LoVo	-	low	-	-	2	-	-	low	WT	int-h	2
T84	-	-	-	-	-	-	-	low	WT	high	5
DLD-1	D	low	-	1	2	-	high	low	WT	int-l	1
SW620	D	-	-	-	-	-	-	low	WT	-	8
LS-174T	B	high	-	-	1	-	low	low	WT	int-h	3
COLO-205	D	low	CMS4	1	1	-	high	low	WT	low	4
HCT-8	C	high	-	3	3	-	low	low	WT	high	3
WiDr	D	high	-	4	3	-	int	low	WT	high	3

Sejtvonal	Budinska	Chang95	CMS	CR Cassigner-786	DeSousa	Marisa	Merlos-Lg5-ISC	ODXcolon	Popovici	Yuen3	minták száma
HCT-15	-	-	-	-	-	-	-	low	WT	-	7
RKO	-	-	-	-	2	-	-	-	WT	int-l	2
COLO-320-HSR	D	high	CMS4	1	3	C2	high	low	WT	int-l	4
LS180	B	low	-	2	1	-	low	low	BRAFm	int-l	1
SW1116	-	-	-	-	-	-	-	low	-	high	5
KM12	D	low	-	-	2	-	high	low	-	-	5
SW48	-	high	-	-	3	-	low	low	-	int-h	4
COLO201	D	int	CMS4	1	3	C2	high	low	BRAFm	int-h	3
SW948	-	-	-	-	-	-	int	low	WT	-	5
CW-2	D	int	CMS4	1	2	C6	high	low	WT	low	1
NCI-H716	B	int	-	3	3	-	low	low	WT	int-l	4
SNU-C4	B	int	CMS1	2	1	C2	int	low	WT	int-l	1
SW837	-	-	-	-	3	-	-	low	WT	high	5
HT-115	D	int	CMS3	1	1	-	high	low	BRAFm	low	1
C2BBe1	D	int	CMS4	1	1	C2	high	low	BRAFm	int-l	1
CL-14	D	int	CMS4	1	1	-	high	low	WT	int-h	1
RCM-1	B	int	CMS1	5	1	C2	low	low	WT	int-h	1
CL-11	D	int	CMS4	1	3	C2	high	low	BRAFm	int-h	1
SW403	-	-	-	-	1	-	low	low	-	-	5
LS-513	B	low	-	4	1	-	int	low	WT	low	1
SNU-C5	A	int	CMS3	3	2	-	low	low	BRAFm	int-h	1
LS-1034	-	-	-	-	1	-	low	low	WT	-	5
HCC2998	D	low	CMS4	1	1	-	high	low	WT	low	3
SK-CO-1	D	low	CMS4	5	1	-	int	high	WT	int-l	1
RKO-E6	-	high	-	-	2	-	low	low	BRAFm	high	3
SW1417	-	-	-	-	-	-	low	low	-	high	5
LS-123	A	high	CMS3	3	3	C3	low	low	WT	int-h	1
SNU-C1	C	low	CMS1	4	1	-	int	low	BRAFm	int-h	1
SNU-407	C	int	CMS1	2	2	C2	low	low	BRAFm	int-l	1
SNU-C2A	A	int	CMS3	2	2	C3	int	low	WT	int-l	1
LS-411N	B	int	-	5	2	-	int	low	WT	low	1
SNU-61	B	int	CMS2	3	1	-	low	low	WT	low	1
CCK-81	D	int	CMS4	1	1	C2	high	low	BRAFm	low	1
NCI-H508	B	-	-	5	1	-	int	low	WT	int-l	4
CL-34	D	int	CMS4	1	1	-	high	low	WT	low	1
CL-40	D	int	CMS4	1	1	C2	high	low	WT	int-l	1
GP2d	D	low	CMS4	1	2	C5	high	low	WT	low	1
HT55	-	-	-	-	1	-	-	low	WT	-	2

Sejtvonal	Budinska	Chang95	CMS	CRCassigner-786	DeSousa	Marisa	Merlos-Lg5-ISC	ODXcolon	Popovici	Yuen3	minták száma
SNU-1040	B	low	-	2	2	C2	int	low	WT	low	1
SNU-175	C	int	CMS3	2	2	-	low	low	BRAFm	low	1
SW1463	-	-	-	2	1	-	-	low	WT	-	2
COLO-678	D	int	CMS4	4	3	-	high	low	BRAFm	low	1
OUMS-23	B	low	CMS1	5	3	-	int	low	BRAFm	int-l	1
SNU-503	C	int	CMS1	2	2	C2	int	low	WT	low	1
HCC-56	D	low	-	1	1	C2	high	low	WT	low	1
Hs 675.T	D	high	CMS3	1	3	-	high	low	BRAFm	int-h	1
Hs 698.T	D	high	CMS3	1	3	-	high	low	BRAFm	int-h	1
MDST8	-	high	-	-	3	-	-	low	WT	-	2
NCI-H747	B	high	-	-	3	-	low	low	BRAFm	high	4
SNU-283	B	low	CMS2	3	1	C5	int	low	WT	int-l	1

13.3 Függelék 3. : Géncsendesítés hatékonysága gén-chip vizsgálatokban

