

AZ EMLŐ-ÉS ENDOMETRIUM EREDETŰ ER α POZITÍV SEJTVONALAK META-ANALÍZISE

Készítette: *Bojcsuk Dóra*

Debreceni Egyetem, Általános Orvostudományi Kar
klinikai laboratóriumi kutató mesterképzési szak II. évfolyam

Témavezető:

Dr. Bálint Bálint László, egyetemi adjunktus

Debreceni Egyetem
Biokémiai és Molekuláris Biológiai Intézet
Klinikai Genomikai és Személyreszabott Orvoslási Központ



Bírálatra való beküldéshez hozzájárulok:

Prof. Dr. Tőzsér József
intézetigazgató

Dr. Bálint Bálint László
témavezető

Dr. Tőkés Szilvia
intézeti TDK felelős

Tartalomjegyzék

Rövidítések jegyzéke.....	3
Összefoglalás.....	5
I. Bevezetés.....	6
II. Anyagok és módszerek	14
III. Eredmények és megbeszélés	25
1. Műtermékek szűrése	25
2. DiffBind analízis.....	26
3. Expresszió alapuló korrelációs analízis	29
4. <i>Peak</i> -ek lokalizációja a genomban.....	31
5. SNV-k száma és aránya a sejtvonalakban	31
6. Azonosított gének	32
7. Motívumkeresés	37
IV. Köszönetnyilvánítás	42
V. Irodalomjegyzék.....	43
VI. Függelék.....	47

Rövidítések jegyzéke

AP-1 – activating protein 1 (transzkripció faktor)

BED – Browser Extensible Data

ChIP – chromatin immunoprecipitation (kromatin immunprecipitáció)

ChIP-seq – chromatin immunoprecipitation followed by deep sequencing (kromatin immunprecipitáció és az azt követő mélyszekvenálás)

CSV – comma-separated values (vesszővel tagolt szöveges *file*)

dbSNP – The Single Nucleotide Polymorphism database

DES – diethylstilbestrol (diethylstilbösztrol)

DiffBind – Differential Binding Analysis

E1 – estrone (ösztrol)

E2 – estradiol, 17 β -estradiol (ösztadiol, 17 β -ösztadiol)

E3 – estriol (ösztrol)

ENCODE – Encyclopedia of DNA Elements

ERE – estrogen response element (ösztrogén válaszadó elem)

ESR1 – estrogen receptor alpha gene (ösztrogén receptor α gén)

ESR2 – estrogen receptor beta gene (ösztrogén receptor β gén)

ER α – estrogen receptor alpha (ösztrogén receptor alfa)

ER β – estrogen receptor beta (ösztrogén receptor béta)

ER⁺ – estrogen receptor positive (ösztrogén receptor pozitív)

EtOH – ethanol

FASTA – FAST Alignment

FoxA1 – Forkhead box protein A1 (transzkripció faktor)

GATA3 – GATA-binding protein 3 (transzkripció faktor)

GEO – Gene Expression Omnibus

HER2 – human epidermal growth factor receptor 2 (humán epidermális növekedési faktor receptor 2)

HOMER – Hypergeometric Optimization of Motif EnRichment

IGV – Integrative Genomics Viewer

MACS – Model-based Analysis of ChIP-Sequencing

MAF – minor allele frequency (minor allélfrekvencia)

MCF-7 – Michigan Cancer Foundation-7 (sejtvonal)
MEME – Multiple EM for Motif Elicitation
NFkB – nuclear factor kappa b (transzkripció faktor)
TF – transcription factor (transzkripció faktor)
PCR – polymerase chain reaction (polimeráz láncreakció)
PR – progesterone (progeszteron)
RPKM – Reads Per Kilobase of transcript per Million mapped reads
SNV – single nucleotide variant (egynukleotidos variáns)
SNP – single nucleotide polymorphism (egynukleotidos polimorfizmus)
Sp1 – specificity protein 1 (transzkripció faktor)
SPMR – signal per million reads
SRA – The Sequence Read Archive
TTS – transcription termination site
TSS – transcription start site
3' UTR – 3' untranslated region (3' át nem íródó régió)
5' UTR – 5' untranslated region (5' át nem íródó régió)

Összefoglalás

Az ösztrogén receptor alfa ($ER\alpha$) a magreceptor családhoz tartozó, ligand-aktivált transzkripciós faktor (TF), amely képes a 17β -ösztradiolt, azaz ösztrogént kötni. Az így kialakuló hormon-receptor komplex a DNS transzkripciót szabályozó régióhoz kötődve befolyásolja a specifikus gének expresszióját. Az $ER\alpha$ kiemelt jelentőségű TF az emlőrák háromnegyedében, mivel az általa kialakított kötőesemények és az így létrejövő változások olyan folyamatokat indukálnak, melyek meghatározzák az ösztrogén receptor pozitív (ER^+) emlőrák kimenetelét, illetve a kezelésre adott választ. A kezelés során az emlőrákos nők közel 40%-a válik rezisztenssé.

Munkánk során két emlő-eredetű (MCF-7 és T-47D) és két endometrium-eredetű (ECC-1 és Ishikawa) ER^+ sejtvonal meta-analízisét végeztük el: az $ER\alpha$ transzkripciós faktor kötőhelyek vizsgálatához minden publikált ChIP-seq (kromatin immunprecipitáció, majd azt követő mélyszekvenálás) adatot bioinformatikai módszerekkel újraelemeztünk, és meghatároztuk az egyes sejtvonalak egyedi és közös $ER\alpha$ kötőhelyeit. Feltérképeztük azokat a géneket is, amelyekben a kötőhelyek találhatóak, továbbá azonosítottuk a kötőhelyeken belül található SNV-ket és azt, hogy ezek a variánsok milyen transzkripciós faktorok kötőhelyeit bontják meg jelenlétükkel. Végül arra kerestük a választ, hogy összefüggésben állnak-e az általunk azonosított SNV-k és gének az emlőrákkal.

Külön hangsúlyt fektettünk a különböző eredetű sejtvonalak összehasonlítására: az azonos eredetű sejtvonalak lényegesen több átfedő, közös $ER\alpha$ kötőhelyet mutattak, mint az egyes, különböző eredetű sejtvonalak. 143 olyan közös kötőhelyet azonosítottunk, amely mind a négy sejtvonalban átfedett, továbbá ezek közül azokat, amelyek az intronban helyezkedtek el, 29 génhez társítottuk: az $ER\alpha$ által szabályozott gének expressziója jelentősen csökkent, azonban néhány gén esetében ellentétes hatás volt tapasztalható: kifejeződésüket az $ER\alpha$ jelenléte alapszinten gátolta, expressziójuk az $ER\alpha$ csendesítése után megnőtt. A 143 közös kötőhelyen belül feltérképezett 1250 SNV közül 280 az $ER\alpha$ motívumán belül lokalizálódott, és ezen TF motívumain belül azonosítottuk az ER^+ emlőrákkal már összefüggésbe hozott, az $ER\alpha$ -val kooperáló FoxA1 és GATA3 TF-ok motívumait is.

I. Bevezetés

Az orvostudomány területeinek kétségtelenül a leggyorsabban növekvő *alterülete* az onkológia. Ezen belül is az emlőrák, amely a nők körében a leggyakrabban diagnosztizált rosszindulatú daganat, évente 1,4 millió új esettel világszerte, egy multidiszciplináris problémaként áll a középpontban, mindennemű földrajzi korlátozottságok nélkül [1, 2]. Az emlőrák korai felismerésén túlmenően, amely nagymértékben javítja a betegség prognózisát, ma már rendelkezésünkre áll számos kezelési formája, amely történhet műtéttel, a petefészek eltávolításával (amely hatására annak ösztrogén termelése megszűnik), továbbá kemoterápiával, sugárkezeléssel, monoklonális antitestek alkalmazásával, kóros érnövekedést gátló kezeléssel, valamint hormongátló készítményekkel. Az alkalmazott kezeléseket és/vagy azok sorrendjét, esetleges kombinálásukat a daganat molekuláris jellemzői határozzák meg.

A 19. század végén *Beatson* volt az, aki rájött, hogy azon nők körében, akik petefészkét eltávolították, életük további részében csekélyebb számban alakult ki emlőrák [3]. Ekkor, 1896-ban hozták először összefüggésbe a mellrákot a hormonokkal, még évtizedekkel az ösztrogén és az ösztrogén receptorok felfedezése előtt.

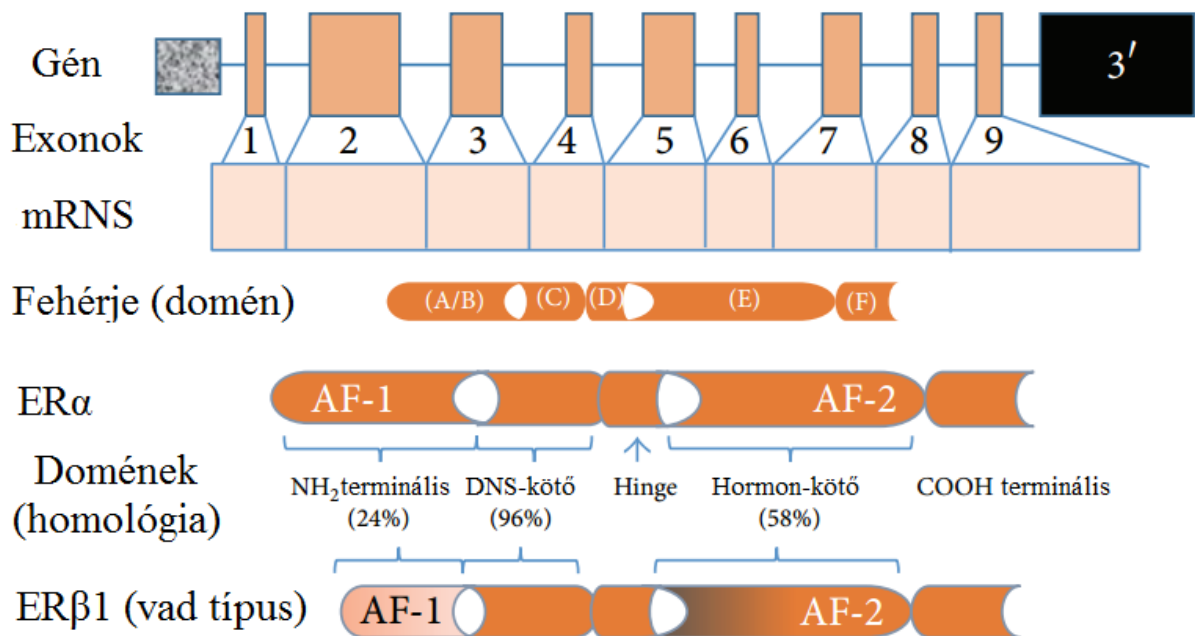
Ösztrogén receptorok

Jensen és *Gorski* kutatócsoportjai voltak azok, akik az 1960-as évek végén elsőként definiálták az ösztrogénkötő receptorok jelenlétét az intracelluláris térben, majd *Chambon* és társai klónozták elsőként magát az ösztrogén receptor gént [4-6].

Az ösztrogén receptoroknak ezidáig két izoformája ismert, név szerint az ER α és az ER β , melyek ligand-függő transzkripciós faktorok. Mindkét izoforma ugyanazon magreceptor család moduláris struktúrájának jellemzőit mutatja, és mindkét receptor hasonló kötőaffinitással rendelkezik természetes ligandjukkal, az ösztrogénnel szemben, szerepük a génexpresszió szabályozásában azonban eltér [7].

Az ER α -t és az ER β -t két különböző gén kódolja: az α izoformát az ESR1 gén, míg a β izoformát az ESR2 gén. Az ER α a humán kromoszóma 6q24–27 lókuszan, az ER β pedig a kromoszóma 14q21-22 lókuszan helyezkedik el. A gének transzkriptjei egyformán 9 exont tartalmaznak, de a két fehérje eltérő számú aminosavból épül fel: az ER α 590, az ER β ettől kevesebb, 530 aminosavból áll. Mindkettő fehérjének hat eltérő doménje van, melyből három

relatíve megegyezik. A domének közötti legnagyobb különbségek az A/B doménekben rejlenek, amelyek felelősek a két ösztrogén receptor között megfigyelt számos antagonistá hatásért [8]. A két magreceptor szerkezetét az 1. ábra taglalja.

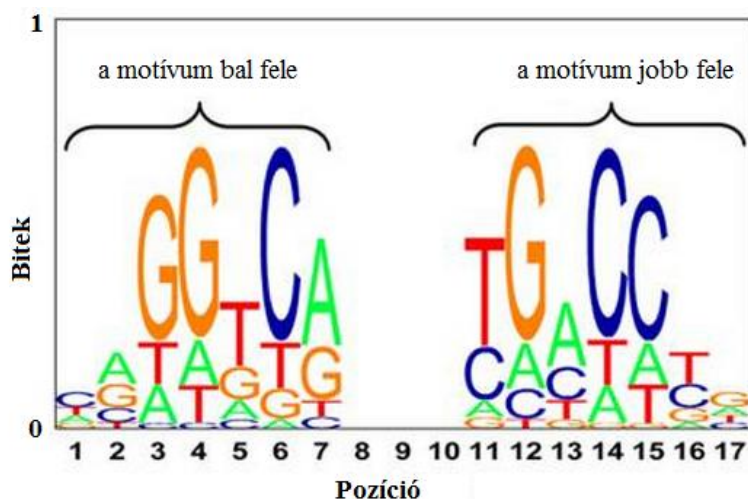


1. ábra. Az ösztrogén receptorok szerkezete. A magreceptor család többi tagjához hasonlóan az ösztrogén receptor hat, evolúciósan konzervált doménből áll (A-F). Bár az NH₂-terminális domén tartalmazza az AF-1 aktiváló faktort, amely lehetővé teszi az egyéb transzkripciós faktorokkal történő kölcsönhatást és a target gén aktiválódását, ezen AF-1 domén aktivitása lényegesen eltér a két ösztrogén receptor között. A DNS-kötő domén két erősen konzervált cink-ujj domént tartalmaz, melyek nélkülözhetetlenek az ösztrogén válaszadó elemhez történő kötődés során, végül a C-terminális domén pedig mediálja a ligand-kötődést, receptor dimerizációt és a magi transzlokációt, továbbá itt lokalizálódik az AF-2 ligandfüggő aktivációs domén is. (Gerald M. Higa and Ryan G. Fell. (2013). Sex Hormone Receptor Repertoire in Breast Cancer. *Int J Breast Cancer* 2013:284036.)

Az ösztrogének

Az ösztrogének a szteroid hormonok csoportjába tartozó kis lipofil molekulák, melyek hidroxilcsoportjuk számát tekintve több típusra oszthatóak: az öszttron (E1) egy, az ösztadiol (E2) kettő, míg az ösztrol (E3) három hidroxilcsoportot tartalmaz. Az ösztadiolt (17 β -ösztdiol) ösztrogénszerű hatásának erőteljes mértéke különíti el a csoport másik két tagjától, és túlnyomórészt ez az ösztrogén felelős a női reprodukív rendszer szabályozásáért, fejlődéséért, ugyanakkor szerepet játszik a kardiovaszkuláris rendszer, a központi idegrendszer és a csontmetabolizmus szabályozásában is [9].

Az ösztrogének természetes ligandjai az ösztrogén receptoroknak és az általuk létrehozott hormon-receptor komplex a kromatinhoz kapcsolódása révén befolyásolja a transzkripció folyamatokat. Az ER α általi transzkripció szabályozás folyamata, mely fontos szereppel bír az emlőrák során, kétféle módon valósulhat meg: az ER α közvetlenül képes kapcsolódni a DNS-hez (klasszikus útvonal), vagy fehérje-fehérje kölcsönhatás révén (nem klasszikus útvonal). A klasszikus útvonal létrejötte során az ER α homodimerek a DNS egy specifikus szekvenciamotívumaihoz kötődnek, az ún. ösztrogén válaszadó elemekhez (ERE). Az ERE-t 1986-ban a *Xenopus vitellogenin* gén promóter régiójában azonosították először [10]. Olyan 13 bázispárnyi palindrom szekvencia, amely egy GGTCA és egy TGACC szekvenciát tartalmaz, közöttük 3 bázispár kihagyásával (2. ábra) [11, 12].



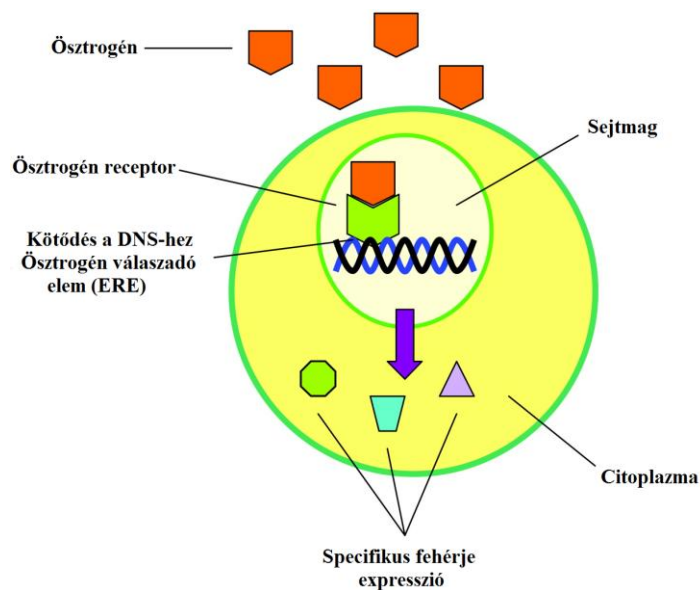
2. ábra. ERE motívum. Az ösztrogén válaszadó elem palindrom szekvenciája, középen három bázispár kihagyással. Az ER α képes kötődni a nem tökéletes ERE motívumhoz is. (Liu ET.(2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor- α . Mol Syst Biol. 21;6:456)

A másik lehetséges módja a kötődésnek a nem klasszikus útvonal. Ennek során az ER α indirekt módon kötődik a DNS-hez, más transzkripciós faktorokhoz kapcsolódva, mint pl. az Sp-1, AP-1 és NF- κ B. Ezen útvonal létrejötte a transzkripciót az ERE-től független módon szabályozza [13, 14].

Genom-szintű ER α ChIP tanulmányokból kiderült, hogy az ER α kötőhelyeknek csak kis töredéke található a promóter régiókban. Többségük túlnyomórészt több, mint 20 kb távolságra helyezkedik a génektől, és ezen távolságok miatt a target gének meghatározása meglehetősen önkényes feladat. Mivel a kötőhelyek mennyisége meghaladja az ER α által regulált gének számát, ez jelzi azt, hogy több kötőhely együttesen szabályozza a transzkripciót. Klasszikus, az ER α által szabályozott gének például a pS2/TFF1, a GREB1, valamint a bcl-2 [11].

Fontos megemlíteni még, hogy az ER α transzkripció faktor kötőhelyekben lévő SNP-k jelenléte egyedenként eltérő lehet. Jelenlétük befolyással bír az adott génexpresszióra, azáltal, hogy megváltoztatják az ER α kötődését. Az ilyen SNP-k azonosítása teszi lehetővé az egyedekre jellemző kezelés lehetőségét, hiszen az ER α meghatározza az emlőrák kimenetelét és a kezelésre adott választ.

Tanulmányunk elsődleges célja a hormonérzékeny, ösztrogén receptor pozitív (ER $^{+}$) emlőrákos sejtekben megjelenő ER α transzkripciós faktorok kötőhelyeinek genomszintű feltérképezése volt. Az endokrin kezelés során az ER α aktiválása több úton is történhet, amelyek mindegyike képes gátolni az ER1 általi sejtproliferációt. Ezen beavatkozások nagymértékben növelik a túlélés esélyét és teljes gyógyuláshoz is vezethetnek, azonban a nők 40%-a a hormonterápia során mégis rezisztenssé válik a kezelésre. Esetünkben azért fontos az ER α kötőhely és a bennük létrejövő változások meghatározása, mert a molekuláris kölcsönhatások a hormon-receptor komplex és a DNS között szabályozzák olyan gének expresszióját, amelyek felelősek a rákos sejtek fenotípusaiért, és ezen *target* gének egyben meghatározzák a sejtnövekedést és az endokrin választ is (3. ábra).



3. ábra. Ösztrogén receptor által kiváltott génaktiváció. Az ösztrogén receptorok normál körülmények között a sejtmagban helyezkednek el. Amikor egy ösztrogén molekula bejut a sejtbe és kötődik a receptorhoz, ez az ösztrogén-receptor komplex a DNS egy specifikus régiójához kötődik, az ún. ösztrogén válaszadó elemekhez, azon gének közelében, melyek ösztrogén által szabályozódnak. Ez a mechanizmus aktiválja a gén transzkripcióját, majd azon specifikus fehérjék szintézisét, melyek befolyásolhatják a sejt viselkedését a sejtípustól függően. (Giacinti L et al. (2006): The Oncologist 11:1-8)

ERα pozitív sejtvonalak

Az ERα kifejeződik többek között a méh és a prosztata kötőszövetes vázában, a here Leydig- és a petefészek tüszőinek külső sejtjeiben, de az emlő szöveteiben is egyaránt. Az ösztrogén indukált ERα kötőesemények modellezésére számos sejtvonal áll rendelkezésünkre, mi összesen négy, két emlő-és két endometrium-eredetű sejtvonalakon végzett kísérletek eredményeit vizsgáltuk.

1) MCF-7

Az emlő-eredetű MCF-7 sejtvonal egy 69 éves, metasztatikus emlőrákban szenvedő beteg mellürében felszaporodó folyadékból (pleurális folyadékgyülem vagy pleurális effúzió) származik, melyet Herbert Soule 1973-ban fedezett fel.

A beteg jobb mellében jóindulatú daganatot, bal mellében pedig rosszindulatú, duktális adenokarcinómát találtak és utóbbin a teljes mellre kiterjedő masztektómiát végeztek el. Soule-ék leírták, hogy a tumor helyi kiújulását 3 évig radioterápiával és hormonterápiával sikerült kontrollált állapotban tartani. Napjainkban a mellrák legáltalánosabb hormonkezelése *tamoxifennel* történik, azidőben valószínűleg nagy dózisú szintetikus, nem-szteroid ösztrogén, a *diethylstilbestrol* (DES) lehetett. A tumorszövet igen érzékeny volt a hormonra, hiszen a várható, átlagosan egy éves túléléshez képest háromszor hosszabb ideig sikerült kontrollált állapotban tartani.

Ezt követően *Marc Lippman* volt, aki bebizonyította, hogy a tamoxifen, mint antiösztrogén gátolja az MCF-7 sejtek növekedését, de ez a gátlási folyamat ösztrogénnel visszafordítható. A tamoxifent bár úgy sorolták be, mint az ösztrogén kompetitív inhibitora, nagy koncentrációban azonban elpusztította a sejteket.

Kathryn Horwitz a glükokortikoid hormonreceptorok, progeszteronok és androgének, valamint az ösztrogén receptor azonosítása után pedig arra a következtetésre jutott, hogy az MCF-7 kiváló in vitro modelljeként szolgálhat a tumor endokrin terápiára adott válaszána, valamint a hormon kötődése és az így létrejövő biológiai hatás kapcsolatának komplex vizsgálatának. A mai napig a legáltalánosabban alkalmazott sejtvonal az emlőrák folyamata során megfigyelhető változások, hormon-receptor kölcsönhatások vizsgálatára [15, 16].

2) T-47D

Az emlő-eredetű T-47D sejtvonal egy 54 éves nő ductalis adenocarcinómájának pleurális effúziójából létrehozott sejtvonal. A sejtekben ösztrogén-és progeszteron receptorok egyaránt kifejeződnek. Mivel p53 mutációval rendelkezik, így alkalmas ezen mutációval rendelkező emlőrákos betegek gyógyszerrezisztens állapotának modellezésére [17].

3) ECC-1

Az ECC-1 az endometrium adenocarcinómájából származó epitheliális sejtvonal. Az ECC-1 sejteken kifejeződnek mind az ösztrogén, progeszteron és androgén receptorok, ezáltal alkalmas az endokrin és parakrin szabályozás in vitro tanulmányozására is [18].

4) Ishikawa

Az Ishikawa egy 39 éves nőből származó, humán endometriális adenocarcinoma sejtvonal. Az ösztrogén-és progeszteron receptorok, illetve androgén receptorok az in vitro tenyésztett sejteken és az indukált tumorban egyaránt kifejeződnek [19].

ChIP-seq

A *ChIP-seq* mozaikszó a kromatin immunprecipitáció és az azt követő mélyszekvenálás folyamatának (chromatin immunoprecipitation followed by deep sequencing) angol kifejezéséből ered.

A kromatin immunprecipitációs módszer alkalmazásával in vivo fehérje – DNS interakciók sejtekben történő vizsgálata válik lehetővé. Az eljárás során a genomiális DNS-hez kötődő fehérjék (transzkripciós faktorok vagy hisztonok) jelenlétének vagy hiányának detektálásához elsőként egy kémiai reakció során kovalensen keresztkötjük azokat pl. formaldehiddel, majd a tisztított kromatint szonikálással vagy emésztéssel történő fragmentáció során 200 – 300 bp nagyságú darabokra törjük. A vizsgált fehérjére specifikus ellenanyag hozzáadásával csak a vizsgálni kívánt fehérje által kötött DNS darabok könyvtárát hozhatjuk létre. A fehérje – DNS komplex keresztkötésének megbontásával a DNS megtisztítható és inputként alkalmazható a további elemzések során a feldúsult DNS azonosításához. Következő lépésben DNS fragmenteket megszekvenáljuk.

A tisztított DNS megszekvenálható bármelyik új generációs platform egyikével. Ezeknek a platformoknak az alapelve hasonló: közös adaptorokat ligálnak a ChIP DNS-hez, és klonálisan csoportosított amplikonokat hoznak létre. A szekvenálási lépés magába foglalja minden templát enzim-alapú meghosszabbítását parallel módon, váltakoztatva a fluoreszcens jelek detektálását.

Az Illumina/Solexa platform esetén a klonális szekvencia csoportjait híd-amplifikációs módszerrel (bridge-PCR) hozhatjuk létre, a szekvenálást pedig szekvenálás-szintézissel.

A 454/SOLiD platform esetében a klonális amplikonokat emulziós PCR-ral állítják elő, az amplikont μ m-es *bead* felszínével fogják ki. Miután a *bead*-eket az amplikonokkal kinyerik és immobilizálják egy síkbeli szubsztráthoz, szekvenálják piroszekvenálással (454) vagy DNS-ligáz alapú szintézissel (SOLiD) [20].

A szekvenálást követően a DNS azon helyei, amelyek kölcsönhatnak transzkripciósfaktorokkal vagy más, kromatin asszociált fehérjékkel, genomszintű vizsgálata válik lehetővé.

II. Anyagok és módszerek

Beválogatási kritériumok

Az ER α transzkripciós faktor kötőhelyeinek bioinformatikai elemzésébe minden olyan, a GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>) és SRA (The Sequence Read Archive, <http://www.ncbi.nlm.nih.gov/sra/>) adatbázisokban elérhető, korábban publikált ChIP-seq kísérletet bevontunk, amelyben a genomszintű ER α kötőeseményt vizsgálták ösztrogén kezelést követően, és elérhetőek voltak a nyers szekvenciaadatai. Továbbá csak azok a kísérletek kerültek beválogatásra, amelyeket az Illumina szekvenáló platformon végeztek. A mintákról információkat az 1. táblázat tartalmaz.

Sejtvonal	GEO azonosító	Replikátum	Antitest	E2 kezelés	Input
MCF-7	GSM1187116	rep1	ER α	10 nM, 45 min	GSM589244
MCF-7	GSM1187117	rep2	ER α	10 nM, 45 min	GSM589244
MCF-7	GSM1187118	rep3	ER α	10 nM, 45 min	GSM589244
MCF-7	GSM1115990	rep1	ER α	100 nM 1 hr	GSM589244
MCF-7	GSM1115991	rep2	ER α	100 nM 1 hr	GSM589244
MCF-7	GSM589237	-	ER α	100 nM 45 min	GSM589244
MCF-7	GSM365926	-	ER α	10 nM 1 hr	GSM589244
MCF-7	GSM470419	-	ER α	10 nM 45 min	GSM589244
MCF-7	GSM614610	-	ER α	100 nM 45 min	GSM614606
T-47D	GSM803539	-	ER α	10 nM 1 hr	GSM589247
ECC-1	GSM589243	-	ER α	100 nM 45 min	GSM589245
ECC-1	GSM803422	-	ER α	10 nM 1 hr	GSM589246
Ishikawa	GSM589241	-	ER α	100 nM 45 min	GSM589246

1. táblázat A beválogatásra kerülő minták. A táblázat tartalmazza, hogy az adott kísérletet milyen sejtvonalon végezték, továbbá a GEO azonosítóikat, azt, hogy egymásnak replikátumai voltak-e a minták (egy kísérletből származtak-e), az alkalmazott antitest nevét, az ösztrogén kezelés körülményeit (koncentráció, időtartam) és a DiffBind analízis során az egyes kísérletekhez használt input minták GEO azonosítóit (Input: kontrollként használat).

Kizárási kritériumok

Kizárára kerültek azok a kísérletek, amelyekben inaktiválták (kiütötték) bármelyik gént, illetve egy előzetes elemzés során azon MCF-7 sejtvonalon végzett kísérletek mintái, amelyekben a *peak*-ek (kötőhelyek) száma tízezernél kevesebb volt. A mintákról információkat a 2. táblázat tartalmaz.

Sejtvonal	GEO azonosító	Replikátum	Kezelés	Kizárási ok
MCF-7	GSM862179	rep1	10nM 1hr	<i>peak</i> -ek száma: 2 624
MCF-7	GSM862180	rep2	10nM 4hr	<i>peak</i> -ek száma: 5 268
MCF-7	GSM862181	rep3	10 nM 24 hr	<i>peak</i> -ek száma: 2 103
T-47D	GSM589239	-	100 nM 45 min	bizonytalan kezelés

2. táblázat A kizárára kerülő minták. *A táblázat tartalmazza, hogy az adott kísérletet milyen sejtvonalon végezték, továbbá a GEO azonosítóikat, azt, hogy egymásnak replikátumai voltak-e a minták (egy kísérletből származtak-e), az ösztrogén kezelés körülményeit (koncentráció, időtartam) és a kizárásuk okát. A GSM862179-80-81-es mintákat egy előzetes elemzés során zártuk ki, mivel a többi MCF-7 sejtvonalon végzett kísérletekhez képest alacsony peak számmal rendelkeztek. A GSM589239 minta E2 kezeltnek volt feltüntetve, leírásában helyett EtOH-os kezelés szerepelt.*

ChIP-seq analízis

Bioinformatikai szempontból a ChIP-szekvenálás többet jelent a leolvasott szekvenciák referenciaszekvenciához történő illesztésénél; kiegészíti ezt számos egyéb lépés még, mint például a kötőhelyek meghatározása, a motívumok keresése. Ehhez ma már számos program, alkalmazás érhető el (HOMER, ChIP-seqer, MACS, MEME).

A GEO adatbázisból származó legtöbb ChIP-seq kísérlet eredményei eredeti elemzésük során a hg18-as humán génomszekvenciához lettek illesztve. Mivel az elemzések folyamán a minták génomszekvenciája közötti átfedést vizsgáltuk, ezért minden kísérletet egy egységes *pipeline* (magyar megfelelője a csőoperátor kifejezés, jelölése a || karakterekkel történik; az „első || második” példa szemlélteti, hogy az „első” kimenete átkerül a „második” bemeneti file-jába) segítségével újraelmeztünk, amely elvégezte minden minta hg19-es humán referenciagenomhoz illesztését és feldolgozását, továbbá meghatározta az egyes mintákban található kötőhelyek lokalizációját a genomban [21]. A *script* (olyan egyszerű program, amely

forráskódja fordítás nélkül futtatható parancssorban a nyelv megfelelő értelmezőjével; magyar megfelelője a szkript kifejezés) a következő szoftverek számunkra szükséges műveleteit végezte el:

1) HOMER

A HOMER (<http://homer.salk.edu/homer>, Hypergeometric Optimization of Motif EnRichment) egy olyan programcsomag, amely eredetileg a motívumok feltérképezését tette lehetővé, majd később a kötőhelyek feldúsulásának azonosítását is. Képes a gének annotálására, klaszterezésére, és lehetővé teszi az eredmények különféle megjelenítését. A ChIP-seq adatok feldolgozásának megkönnyítésére a HOMER egyik első lépése a „tag könyvtár” („Tag Directory”) létrehozása, amely során a program gyakorlatilag egy adatbázist készít, és a kísérlettel kapcsolatos összes lényeges információ ebbe a könyvtárba kerül.

Mi a HOMER által létrehozott *bedGraph file*-t (magyar megfelelője a fájl kifejezés) az IGV-ben (Integrative Genomics Viewer) történő megjelenítésére használtuk [22]. A referencia genom kiválasztása után a *bedGraph*-okat betöltve manuálisan vizsgálható pl. egy kezeletlen és egy E2-vel kezelt mintában a kezelés hatására történő kötőhelyváltozás az új csúcsok megjelenésével. A *bedGraph file* létrehozása:

```
makeUCSCfile <tag-directory> -name <track-name> -o <output-  
file.bedgraph> -norm -fragLength <fragLength>
```

Ha kontroll minta is elérhető:

```
makeUCSCfile </path-to-tagdir> </control-tag-directory> -name-of-  
the-experiment -o <path-to-output-file.bedgraph> -norm
```

A HOMER-on belül a *peak*-ek annotálására az *annotatePeaks.pl* parancs szolgál. Ez egy „minden egy programban” parancs, amely képes illeszteni a csúcsokat a legmegfelelőbb génhez, de végre tudja hajtani a gén ontológiai elemzését, genomi funkciókkal kapcsolatos analíziseket (Genome Ontology), társítja őket a génexpressziós adatokhoz, kiszámítja a ChIP-seq *tag*-ek (leolvasások) sűrűségét a különböző kísérletekben, ill. meghatározza a lehetséges motívumokat. Az *annotatePeak.pl* használható még hisztogramok és *heatmap*-ek (hő térképek) készítésére is:

```
annotatePeak.pl <file.bed> <genome> -noann -hist1 -size <size> -d  
<tagdir> > <output- file.txt>
```

A motívumok keresése a `findMotifsGenome.pl` paranccsal történik, melynek egyszerűen egy, a *peak*-ek lokalizációját tartalmazó *file*-t kell bemenetként megadni, a referenciagenomot (hg19), illetve egy könyvtárat, ahová majd az eredményeink kerülnek:

```
findMotifsGenome.pl <peak-or-bedfile> <genome> <otuput-directory>
```

Végül a motívumokat képfarmátumban tárolhatjuk:

```
motif2Logo.pl <myfile.motif> -pdf vagy -png
```

2) MACS

A MACS (<http://liulab.dfci.harvard.edu/MACS>, Model-based Analysis of ChIP-Sequencing) program a HOMER-hoz hasonlóan a *peak*-ek prediktálására alkalmas. A MACS empirikusan modellezi a szekvenált ChIP-seq *read*-ek eltolódásának mértékét, így javítva a várható transzkripciós faktor kötőhelyek térbeli felbontását. Egy dinamikus Poisson-eloszlást használ a genomban lévő torzítások rögzítésére, lehetővé téve a ChIP-seq csúcsok pontosabb azonosítását. Végezhetünk vele kísérletet kontroll mintával vagy anélkül is. Ez a leggyakrabban használt program, Python programnyelven íródott [23, 24]. Az összehasonlító elemzések során a MACS *peakfile*-ekkel dolgoztunk, melyeket a következő parancs hoz létre:

```
macs2 <arguments> -n <experiment-name> -t <target-file.bam> -g hs -B  
--SPMR -c <control-bamfile>
```

A `-B --SPMR` kapcsolóval a MACS az észlelt fragmenteket `bedGraph` formátumban fogja tárolni minden kromoszómán (*spm*: *signal per million reads*).

Argumentumként az elemzés célja szerint a `callpeak`-et alkalmaztuk, amely a `macs2` fő funkciója; az illesztett eredmények *peak*-jeinek prediktálását végzi el.

Műtermékek szűrése

A minták ER α kötőhelyeinek egységes elemzését követően elsőként a műtermékek kiszűrését végeztük el: a *peakfile*-okat az ENCODE egy artefaktumokat tartalmazó *file*-jához illesztettük [25].

A két lépés, mely során kigyűjtöttük a műtermékeket, majd azokat az eredeti *file*-ból eltávolítottuk, alább látható (4.1, 4.2 ábra):

```
intersectBed -u -a <original-macs-peaks.bed> -b <ENCODE-blacklist.bed> > <ENCODEvsPeakfile.bed>
```



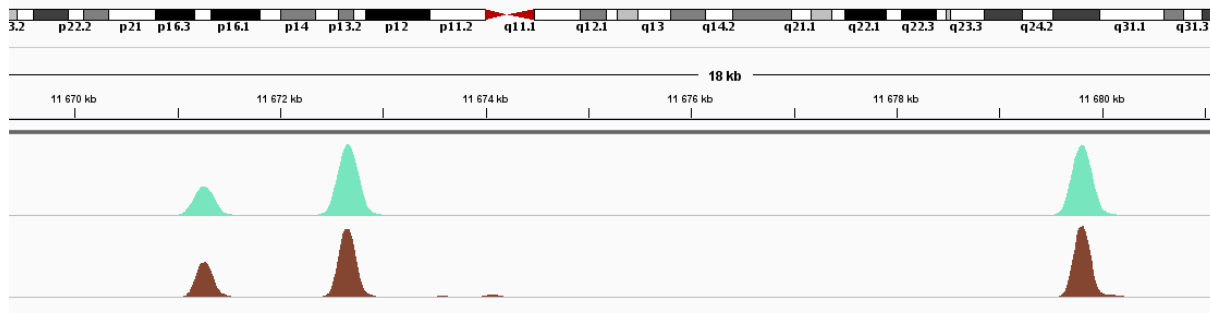
4.1. ábra. Műtermékek kigyűjtése egy *file*-ba. Az *intersectBed* parancs segítségével azokat a csúcsokat gyűjtöttük ki, amelyek átfedést mutattak az ENCODE artefaktumokat tartalmazó *file*-jával. (-a: eredeti peak *file*; -b: ENCODE műtermék *file*-ja; szürke sáv: visszanyert peak-ek).

```
subtractBed -A -a <original-macs-peaks.bed> -b <ENCODEvsPeakfile.bed> > <final-output-file.bed>
```

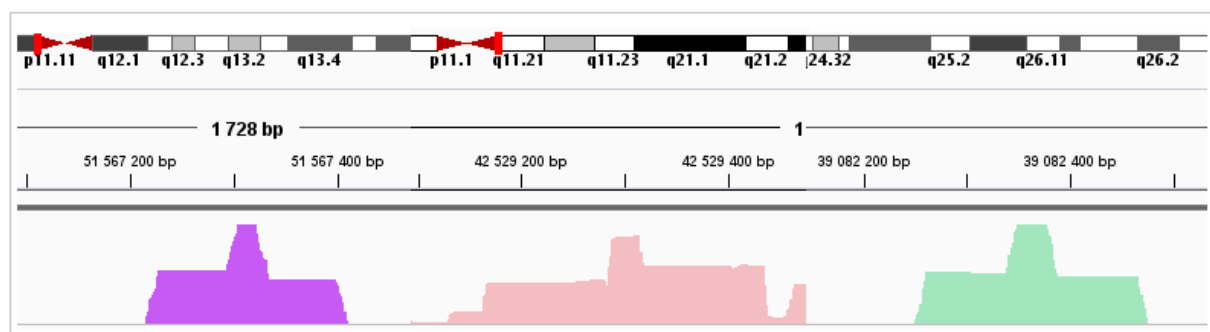


4.2. ábra. Műtermékek eltávolítása. A *subtractBed* parancs segítségével eltávolítottuk az eredeti *file*-ból a műtermékekkel átfedő peak-eket. (-a: eredeti peak *file*; -b: műtermékeinket tartalmazó *file*; szürke sáv: visszanyert peak).

A műtermékek valóságát az IGV-ben ellenőriztük. A 4.3 ábra jól szemlélteti a valódi kötőhelyeket jelölő csúcsokat, míg a 4.4. ábra ehhez képest a műtermékek jellegzetes megjelenési formáját.



4.3. ábra. Valódi peak-ek. Az ábrán két különböző ChIP-seq kísérletből származó ER α transzkripciós faktor kötőhelyek csúcsai láthatóak az IGV (Integrative Genomics Viewer) programban vizualizálva.



4.4. ábra. A műtermékek valóságának ellenőrzése. Az ENCODE műterméklistájához illesztett, saját peak file-okból kiszűrt artefaktumok valóságának ellenőrzése IGV (Integrative Genomics Viewer) használatával. Az ábrán három különböző kísérletből származó műtermék csúcsa látható, és a 4.3. ábrán bemutatott valódi transzkripciós faktorok kötőhelyeit jelölő csúcsokhoz viszonyítva jól látható a műtermékek jellegzetes alakja.

DiffBind analízis

A már műtermékek nélküli mintákat az R Bioconductor DiffBind szoftverével elemeztük. A DiffBind képes a ChIP-seq kísérletek különböző kötőhelyeit meghatározni az affinitási

(kvantitatív/mennyiségi) adatok felhasználásával, emellett lehetővé teszi az átfedő régiók elemzését és vizualizálását is. Képes a genomi lókuszokon feldúsult ChIP-seq adatok feldolgozására ott ahol a specifikus fehérje – DNS kötődés megtörténik, esetünkben azon *peakset*-ek feldolgozására, amelyeket a MACS program prediktált.

Az analízis első lépéseként létrehoztunk egy vesszővel tagolt (csv, comma-separated values) *file*-t a következő információkkal: SampleID, Tissue, Factor, Condition, Replicate, bamReads, bamControl, Peaks. A továbbiakban ezt az összeállított *file*-t olvastuk be.

Első lépésben a replikátumokat összevontuk, minimális átfedésnek egyet megadva, így minden kötőhelyet a replikátumok mintáiból megtartottunk. A továbbiakban az MCF-7 sejtvonal mintáinak kötőhelyeit vontuk össze, és azokat tartottuk meg, amelyek legalább 2 mintában jelen voltak, majd az ECC-1 sejtvonal 2 mintáját. Ennél azonban szintén megtartottunk minden azonosított kötőhelyet. Utolsó lépésben a 4 sejtvonal kötőhelyeinek egyedi halmazát együttesen vizsgáltuk, és arra voltunk kíváncsiak, mennyi olyan régió van, amely mind a 4 sejtvonalban átfed.

A hierarchikus DiffBind elemzés során használt parancssorok a következők voltak:

library(DiffBind)

```
sample_cell.lines <- dba(sampleSheet = "sample.csv", peakFormat = "bed")

consensus.sample.replicates = dba.peakset(sample_cell.lines, consensus =
DBA_REPLICATE, minOverlap = 1)

consensus.sample.tissues = dba.peakset(consensus.sample.replicates,
c(6,7,8,9,14,15), minOverlap = 2)

consensus.sample.tissues2 = dba.peakset(consensus.sample.tissues, 10:11,
minOverlap = 1)

consensus.sample.final = dba.peakset(consensus.sample.tissues2,
c(12,13,16,17), minOverlap = 4, sampID = "final")

sample_cell.lines.count <- dba.count(sample_cell.lines, peaks =
consensus.sample.final$peaks[[18]], bCorPlot = F, bParallel = F)

consensus.sample.final.overlap = dba.overlap(consensus.sample.final, mask =
c(12, 13, 16, 17))
```

Az egyes mintákban azonosított intervallumok számát és a hierarchikus elemzés során visszanyert átfedő régiókat a 3. táblázat tartalmazza.

	Szövettypus	Faktor	Kezelés	Replikátumok	Átfedő intervallumok
1	MCF-7	ER α	E2	1	46 498
2	MCF-7	ER α	E2	1	41 250
3	MCF-7	ER α	E2	1	33 355
4	MCF-7	ER α	E2	2	32 021
5	MCF-7	ER α	E2	2	11 198
6	MCF-7	ER α	E2	3	13 076
7	MCF-7	ER α	E2	4	82 922
8	MCF-7	ER α	E2	5	29 102
9	MCF-7	ER α	E2	6	74 787
10	ECC-1	ER α	E2	7	2 855
11	ECC-1	ER α	E2	8	35 528
12	T-47D	ER α	E2	9	32 621
13	Ishikawa	ER α	E2	10	3 063
14	MCF-7	ER α	E2	1	70 549
15	MCF-7	ER α	E2	2	33 966
16	MCF-7	ER α	E2	3-4-5-6-1-2	51 137
17	ECC-1	ER α	E2	7-8	37 749
18	T-47D-Ishikawa-MCF-7-ECC-1	ERα	E2	9-10-3-4-5-6-1-2-7-8	145

3. táblázat. DiffBind analízis eredményei lépcsről-lépésre. *A különböző sejtvonalak egyedi és közös intervallumai a hierarchikus DiffBind elemzése során. A táblázat 14-15. sora a 2 MCF-7 kísérlet replikátumainak összevonását, a 16. sor ezen összevont replikátumok és az egyedi MCF-7 ChIP-seq kísérletek egy halmazzá redukálását, a 17. sor a két ECC-1 minta összevonását mutatja. Az utolsó, 18. sor a 4 vizsgált sejtvonal kötőhelyeinek egyedi halmazát egyesíti.*

A DiffBind analízist követően az átfedő szakaszok tartalmaztak Y kromoszómán detektált kötőhelyeket – habár a többi kromoszómához képest jóval csekélyebb számban. Mivel nőkből származó sejtvonalakról van szó, ezek az adatok nem relevánsak. Ezen jelenség fennállására magyarázattal szolgálhat egy 2010-es tanulmány, melyben a SOLiD és az Affymetrix szekvenáló platformokon vizsgálták, hogy MCF-7 sejtvonalakon végzett kísérletek során a szekvenálás után mennyi Y kromoszómára vonatkozó adatot nyertek vissza. Az Y kromoszómán detektált kötőhelyek a további elemzésből kizárásra kerültek [26].

Az egyes sejtvonalakban átfedő régiókat a `dba.plotVenn` függvénnyel Venn diagram formájában ábrázoltuk.

Expresszió alapuló korrelációs analízis

A korrelációs analízis során azokat a kötőhelyeket határoztuk meg, amelyek a 13 minta legalább 60%-ában ($n \geq 7$) jelen voltak és azokban minimum 50%-ban ($f = 0.5$) átfedtek. Az eredményeket *heatmap*-en és a korrelációs *plot*-on vizualizáltuk. A korrelációs *plot* a korrelációs mátrix, konfidencia intervallum mátrixának (alsó-és felső tartományának) grafikai megjelenítését szolgálja [27]. A *heatmap* a különböző kötőhelyekben található *read*-ek sűrűségét rajzolja fel és a *pheatmap*, mint „pretty heatmap” szoftvercsomaggal készült [28].

Az egyes minták értékeit a *read*-ek számával normalizáltuk, a klaszterezés pedig a *complete* klaszterezési eljárással történt.

Az adatok feldolgozása során a leolvasásokat 1:1 millióra vonatkoztattuk, azaz a leolvasott egységek száma/1 millió leolvasott egység száma szerint, amit 1 kilobázispárnyi génhosszra vonatkoztattunk, így elkerülve a gének hosszából adódó különbségeket (RPKM, *Reads Per Kilobase of transcript per Million mapped reads*)).

Peak-ek lokalizációja a genomban

Annak meghatározására, hogy az egyes sejtvonalakban található kötőhelyek a genomban hol lokalizálódnak, a *macs peak file*-okat a hg19-es humán referenciagenomhoz illesztettük.

SNV-k azonosítása

A sejtvonalak egyedi, a DiffBind analízis során kapott kötőhelyeinek listáját a hg19 humán referenciagenomjára elérhető dbSNP (GRCh37/hg19) adatbázishoz illesztve meghatároztuk a sejtvonalakban az SNV-k számát. Ezt az értéket a *peak*-ek teljes hosszának összegével elosztva egy-egy sejtvonalra jellemző, normalizált SNV számot kaptunk.

Gének azonosítása

Az MCF-7, ECC-1, T-47D és Ishikawa sejtvonalakban átfedő, konszenzus kötőhelyet lokalizációjuk alapján az Ensembl (<http://www.ensembl.org/index.html>) humán génlistájához illesztve megvizsgáltuk, hogy mely kötőhelyek találhatóak intronban, ill. ezen introni kötőhelyeket az Ingenuity web-alapú szoftver (Ingenuity Pathways Analysis, <http://www.ingenuity.com/products/ipa>) segítségével meghatároztuk, hogy mely géneken helyezkednek el. Továbbá szintén az Ingenuity szoftver segítségével azonosítottuk a feltérképezett gének lehetséges *upstream* regulátorait is.

A géneket és regulátoraikat *circos* ábrán mutattuk be, melyet a Circos programmal készítettünk [29].

Az irodalomban fellelhető publikációk alapján összefüggést kerestünk az azonosított gének, szabályozó fehérjék és az emlőrák között, és két nagyobb tanulmány GEO adatbázisban található microarray kísérletinek eredményeiből hő térképen ábrázoltuk az általunk azonosított gének expressziójának mértékét az ösztrogén kezelést követő 3-6-12. órákban, illetve azt, hogy miként változik ezen gének kifejeződése az ER α lecsendesítését követően [11, 30].

Motívumkeresés

Annak meghatározására, hogy a négy sejtvonal közös kötőhelyeiben feltérképezett SNV-k milyen transzkripciós faktorok motívumait bontják meg jelenlétükkel, a FunSeq (Function based Prioritization of Sequence Variants, funseq.gersteinlab.org) web-alapú alkalmazást használtuk. Elsőként a konszenzus kötőhelyeket lokalizációjuk alapján a dbSNP adatbázisával

összevetve egy olyan kimeneti *file*-t kaptunk eredményül, amely tartalmazta az SNV-k rsID-jait, ezen információ ugyanis szükséges a FunSeq elemzés során.

Minor allélfrekvenciaként (MAF) a FunSeq alkalmazáshoz 2013-ban megjelent cikk alapján egy származtatott allélfrekvencia értéket adtunk meg, melynek értéke 0-1 közötti skálán 0.005 (0,5%) volt [31].

III. Eredmények és megbeszélés

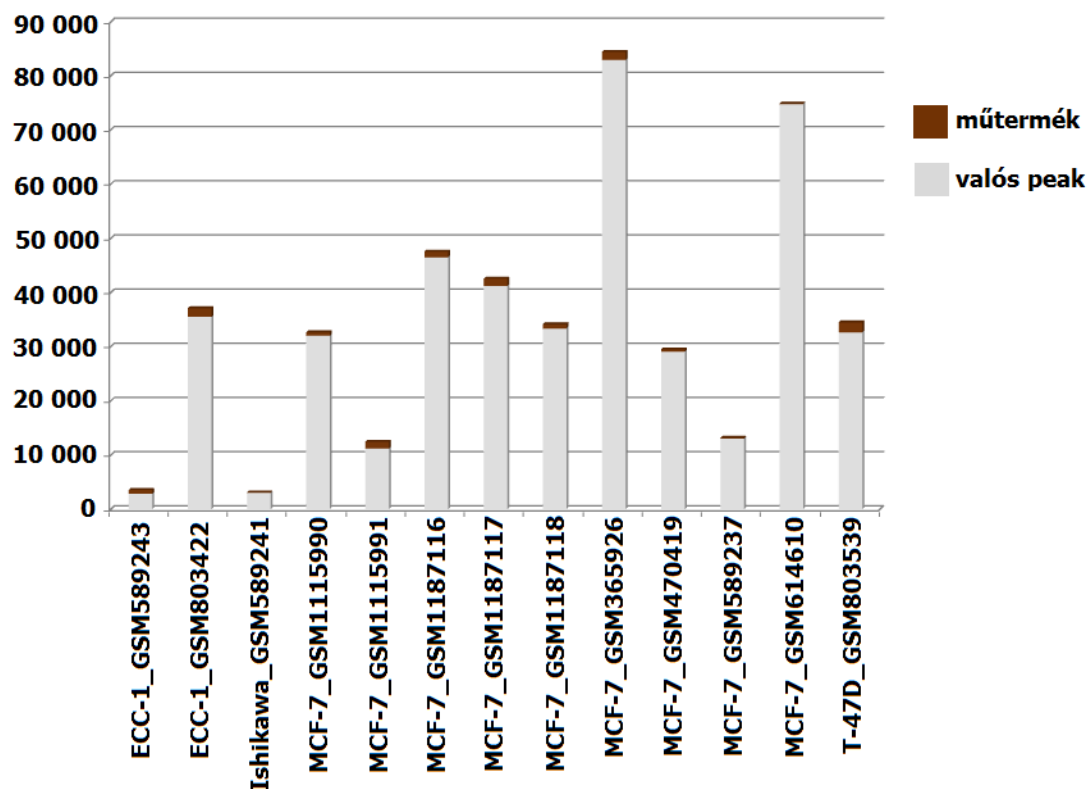
1. Műtermékek szűrése

A ChIP-seq kísérletek egységes újraelemzése, majd a műtermékek és az Y kromoszómára kapott eredmények kiszűrése után az egyes minták valós *peak*-jeinek száma a következőképp alakultak:

Sejtvonal	GEO azonosító	Peak-ek száma	Műtermékek	-Y chr	Valós <i>peak</i> szám
MCF-7	GSM1187116	47 552	1 054	75	46 423
MCF-7	GSM1187117	42 557	1 307	86	41 164
MCF-7	GSM1187118	34 139	784	49	33 306
MCF-7	GSM1115990	32 646	625	57	31 964
MCF-7	GSM1115991	12 431	1 233	24	11 174
MCF-7	GSM589237	13 173	97	24	13 052
MCF-7	GSM365926	84 411	1 489	145	82 777
MCF-7	GSM470419	29 504	402	166	28 936
MCF-7	GSM614610	74 856	69	90	74 697
T-47D	GSM803539	34 498	1 877	-	32 621
ECC-1	GSM589243	3 563	708	-	2 855
ECC-1	GSM803422	37 085	1 557	-	35 528
Ishikawa	GSM589241	3 102	39	-	3 063

4. táblázat. Műtermékek szűrése. *A valós peak-ek száma a mintákban a műtermékek és az Y kromoszómára visszazvert eredmények levonása után.*

A valós csúcsok és műtermékek arányát oszlopdiagramon szemléltetjük (5. ábra)



5. ábra. A valós csúcsok és artefaktumok aránya az egyes mintákban. Az *ENCODE* műterméklistájával összevetett peak file-ok valós peak-jeinek és műtermékeinek aránya.

2. DiffBind analízis

A vizsgált, négy sejtvonal egyedi és konszenzus *peak*-jeinek DiffBind analízise során a már műtermékektől megszűrt *peakset*-eket elemeztük, az elemzés végén azonban Y kromoszómában lokalizálódó kötőhelyek eredményei is szerepeltek. Mivel a sejtvonalak nőkből származnak, ezen adatok nem relevánsak. Egy 2010-es tanulmányban MCF-7 sejtvonalon és több szekvenáló platformon tesztelték az Y kromoszómára visszakapott *peak*-ek számát, a jelenséget egyszerűen technikai artefaktumként magyarázták [26].

Az Y kromoszómára visszanyert adatokat az átfedő szakaszok meghatározása után vontuk ki, a kivont értékeket az 5. táblázat tartalmazza.

Sejtvonal	chrY	Sejtvonal	chrY	Sejtvonal	chrY
MCF-7	51	Ishikawa, MCF-7	1	Ishikawa, MCF-7, ECC-1	0
T-47D	13	Ishikawa, ECC-1	8	T-47D, MCF-7, ECC-1	9
ECC-1	35	MCF-7, ECC-1	11	T-47D, Ishikawa, ECC-1	0
Ishikawa	29	Ishikawa, T-47D	0	T-47D, Ishikawa, MCF-7	0
MCF-7, T-47D	4	T-47D, ECC-1	6	MCF-7, T-47D, ECC-1, Ishikawa	2

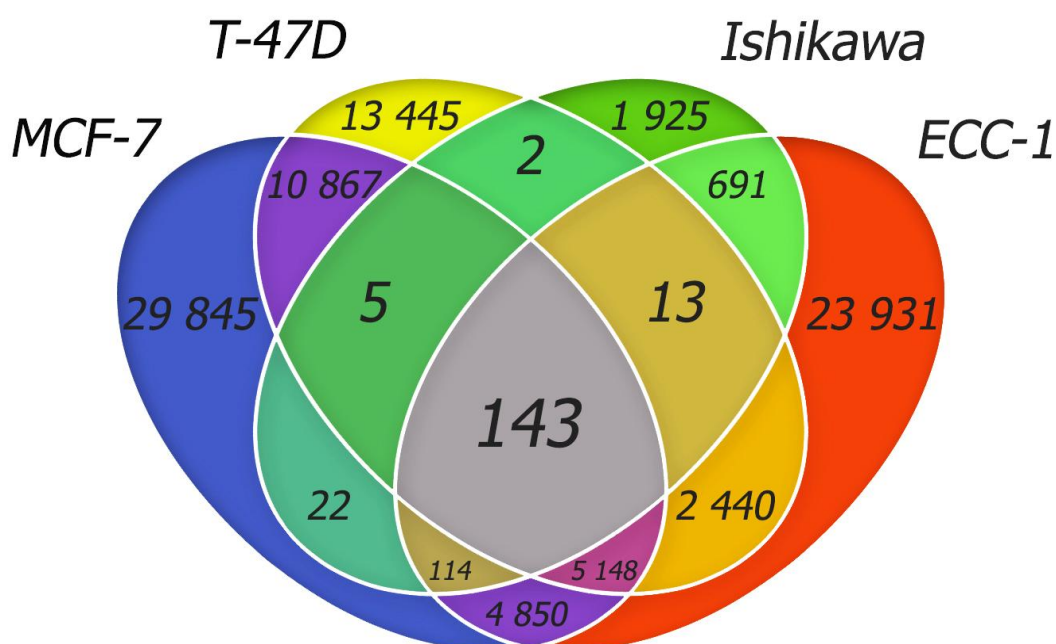
5. táblázat. A DiffBind elemzés után az egyes *peak*-set-ekből kivont Y kromoszómára vonatkozó értékek.

Az emlő-eredetű MCF-7 sejtvonal 29 854, a T-47D sejtvonal pedig 13 445 egyedi *peak*-et tartalmazott, az endometrium-eredetű ECC-1 sejtvonal 23 931, az Ishikawa sejtvonal pedig 1 925 egyedi csúcsot. A mind a 4 sejtvonalban megtalálható, konszenzus kötőhelyek száma 143 volt.

Jól látható, hogy az Ishikawa sejtvonal *peak*-jeinek száma jóval alacsonyabb volt a többihez képest, és a legtöbb közös kötőhelye a szintén endometrium-eredetű ECC-1 sejtvonallal volt (691 darab). A többi, emlő-eredetű sejtvonallal csupán 2, ill. 22 *peak*-ben fedett át. Bár az ECC-1 több (a T-47D-vel 2 440 és az MCF-7-tel 4 850 darab) átfedő *peak*-et mutatott az emlő-eredetű sejtvonalakkal, arányaiban ezek mégsem jelentenek többet, ha az egyes sejtvonalakban találató összes kötőhely számát vesszük alapul.

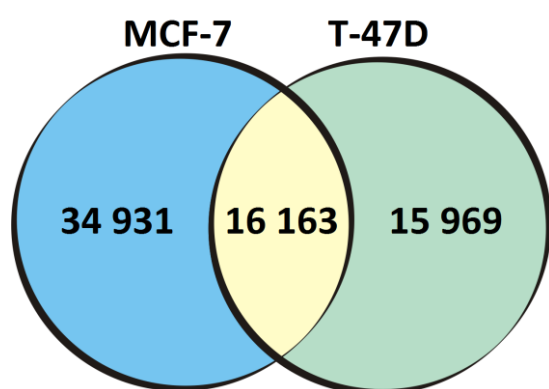
Az MCF-7 és a T-47D sejtvonalak, melyek emlő-eredetűek, szintén több, egymással közös kötőhelyet tartalmaznak (10 867 darab), mint az egyéb eredetű sejtvonalakkal összevetve.

Ezen eredmények alapján elmondható, hogy a különböző eredetű sejtvonalak egymással nem fednek át olyan mértékben, mint az azonos eredetű társaikkal. Az eredményeket Venn diagram formájában a 6. ábra szemlélteti.

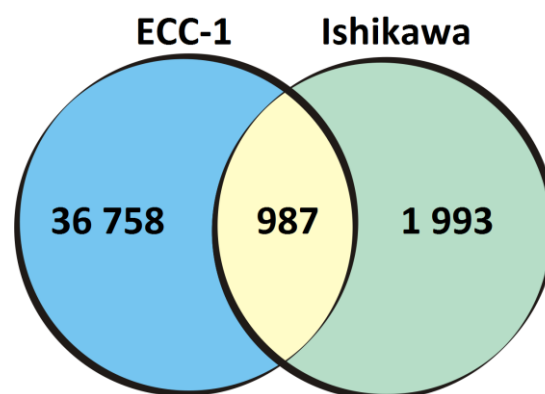


6. ábra. Az emlő-és endometrium-eredetű sejtvonalak egyedi és konszenzus kötőhelyeinek száma. Az ábra jól szemlélteti, hogy az egyes sejtvonalak teljes peak-jeinek számához viszonyítva az azonos eredetű sejtvonalak közötti közös kötőhelyek aránya nagyobb, mint bármely két, eltérő eredetű sejtvonale. A mind a négy sejtvonalba átfedő peak-ek száma 143.

Külön hangsúlyt fektettünk az emlő-és endometrium-eredetű sejtvonalak együttes elemzésére: csak az emlő-eredetű MCF-7 és T-47D sejtvonalak egyedi kötőhelyeit összevetve 16 163 közös intervallumuk volt, míg az endometrium-eredetű sejtvonalaknak 987. Az eredményeket Venn diagramok formájában a 7.1. és 7.2. ábra szemlélteti.



7.1. ábra. Az emlő-eredetű sejtvonalak egyedi és konszenzus intervallumainak száma.

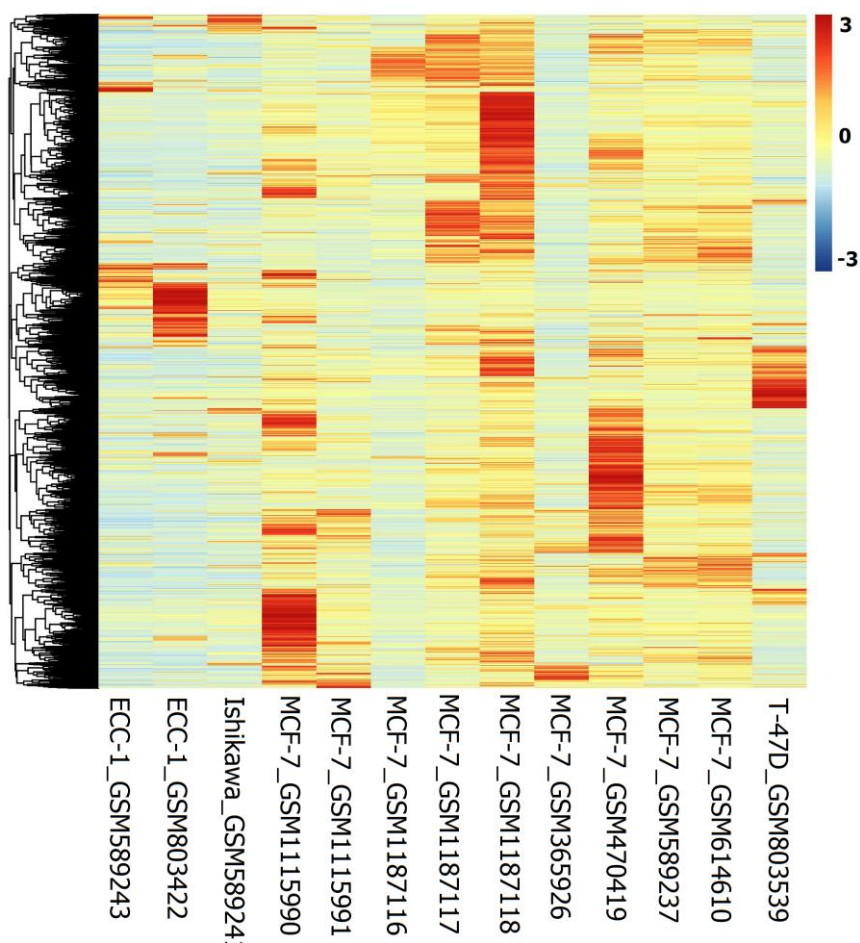


7.2. ábra. Az endometrium-eredetű sejtvonalak egyedi és konszenzus intervallumainak száma.

3. Expresszió alapuló korrelációs analízis

Az elemzés során már azokat a *file*-okat elemeztük, amelyekből az Y kromoszómára vonatkozó eredmények el lettek távolítva és műtermékekre is szűrve lettek.

A normalizált, klaszterezett hőtésképen (8. ábra) látható, hogy a legtöbb kötőhely az MCF-7 sejtvonal mintáiban található, azonban az így vizualizált eredmény azok heterogenitására utal az azonos sejtvonal-eredetű minták, de még azonos kísérletből származó minták replikátumai között is.



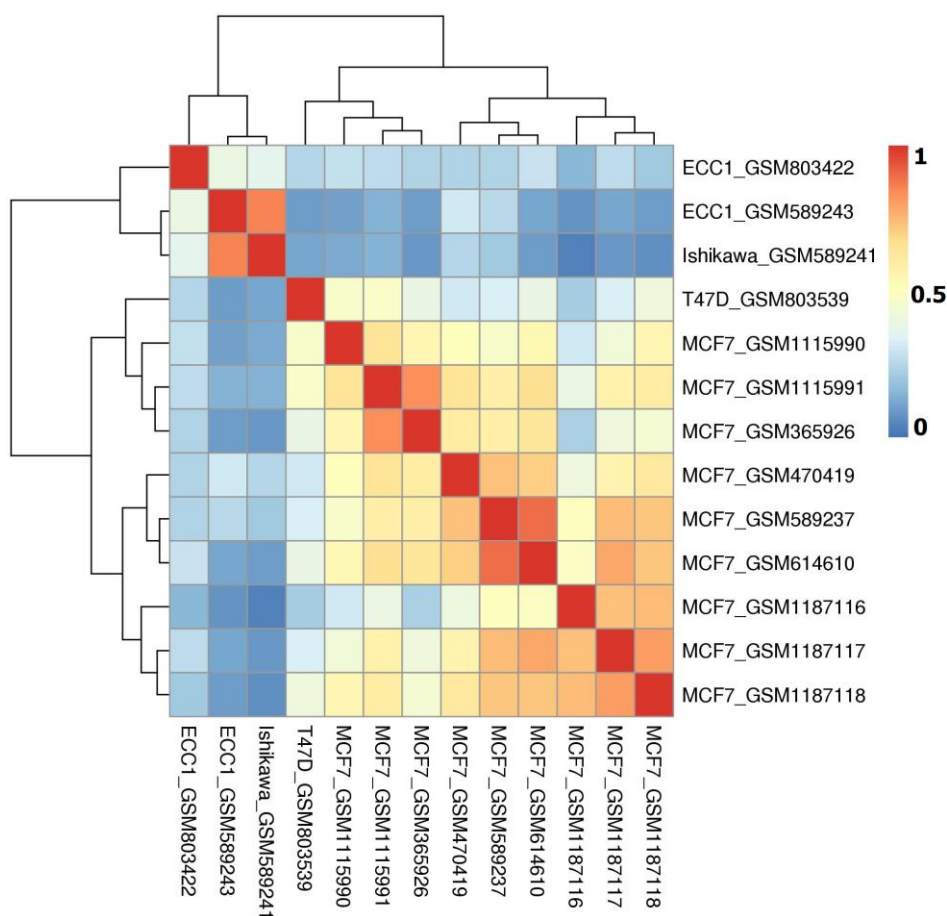
8.ábra. A minták normalizált, klaszterezett hőtésképe. A függőleges sávok egy-egy mintának felelnek meg, a vízszintes sávok az egyes mintákban adott pozícióban található kötőhelyeknek, színezésük pedig a kötőhelyek read-ek számával normalizált méretétől függ.

Az ER α kötőhelyei közötti különbségeket és azok okait számos publikációban tárgyalták. Az kísérleti vagy technikai különbségek, mint a platformok, a ligand koncentrációja, a kezelés

ideje, vagy az alkalmazott statisztikai küszöbértékek egyes tanulmányok között alapja lehet ezen eltérő eredményeknek [32].

Oka lehet továbbá a biológiai különbség, azaz a sejtek eltérő kezelése. Az MCF-7 sejtvonalat évtizedek óta fenntartják, az egyes laboratóriumokban az eltérő módszerek miatt többé nem is lehet ugyanolyan. Különbségek adódhatnak a karotípusukban, a hormonreceptor tartalomukban, de még a növekedésük ütemében is - bár morfológiailag azonosak maradnak [33].

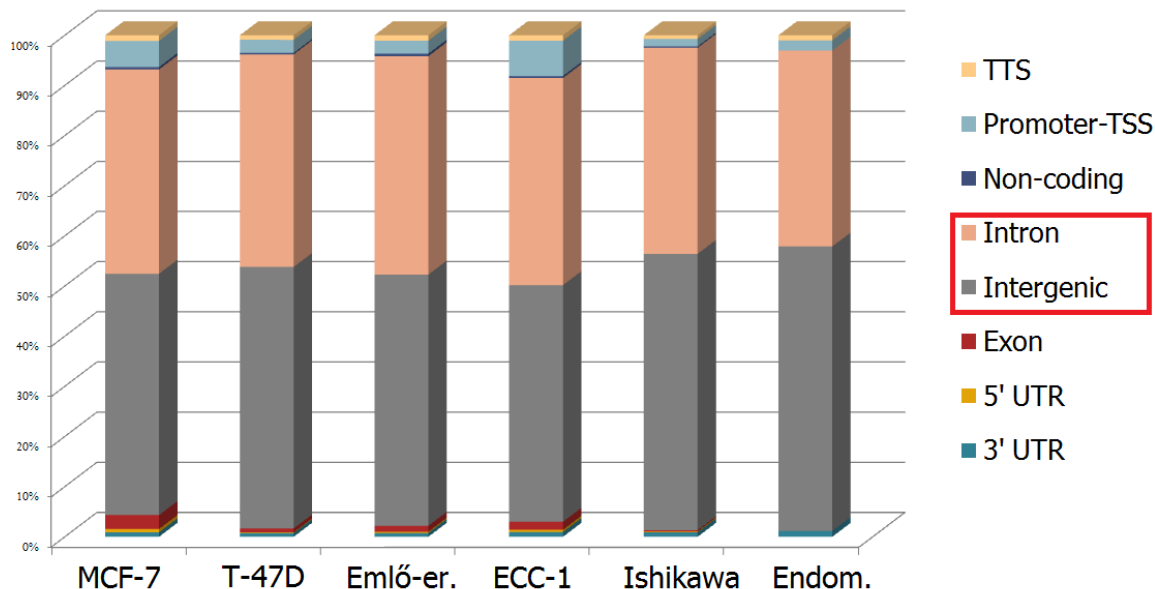
A korrelációs *plot*-on (9. ábra) jól látszik, hogy az emlő-és endometrium-eredetű sejtvonalakon végzett kísérletek mintái két csoportra oszlanak, közöttük a kötőhelyek tekintetében korreláció nincs, vagy nagyon alacsony. Az endometrium-eredetű ECC-1 és Ishikawa minták között nagymértékű korreláció a két különböző sejtvonal egy-egy mintája között rajzolódott ki, a GSM803422-es minta feltehetően a kötőhelyek alacsony száma miatt mutatkozott kevésbé hasonlónak a másik ECC-1 mintával.



9. ábra. Korrelációs plot. Korreláció a különböző eredetű sejtvonalak között nincs, vagy nagyon alacsony és a minták sejtvonal-eredetük szerint két csoportot alkotnak.

4. Peak-ek lokalizációja a genomban

A kötőhelyek kivétel nélkül mindegyik sejtvonal esetén 90-95%-ban a genom introni és intergénikus régióiban helyezkedtek el.



10. ábra. Az egyes sejtvonalak *peak*-jeinek lokalizációja a genomban. Egy 100%-nak megfelelő skálán szemlélítve minden sejtvonal esetében a kötőhelyek 90-95%-a a genom introni és intergénikus régiójában lokalizálódott. Az ábrán a sejtvonal-eredetek szerinti összesített eredmények is láthatóak. (TSS: transcription termination site; TSS: transcription start site; UTR: untranslated region)

5. SNV-k száma és aránya a sejtvonalakban

A különböző sejtvonalakban található SNV-k számát ha normalizáltuk a *peak*-ek hosszának összegével, azt találtuk, hogy bár az Ishikawa sejtvonalat képviselő egy kísérlet kötőhelyeinek teljes hossza volt a legkevesebb, ugyanakkor ezen kevés bázispárnai régió belül találtuk a legtöbb SNV-t. Az eredeti SNV számokat, a kötőhelyek hosszát és a normalizált értékeket a 6. táblázat tartalmazza.

Sejtvonal	MCF-7	T-47D	ECC-1	Ishikawa
SNV-k száma	162 989	40 747	63 245	4 754
Peak-ek hossza	~ 9Mbp	~ 2Mbp	~ 5 Mbp	53 Kbp
Normalizált SNV szám	$18,3 \times 10^{-3}$	17×10^{-3}	$13,1 \times 10^{-3}$	$89,1 \times 10^{-3}$

6. táblázat. A sejtvonalakban található peak-ek teljes hossza, az ezek alatt talált SNV-k száma, és a normalizált SNV értékek. A legtöbb SNV az Ishikawa sejtvonal kötőhelyein belül volt megtalálható.

6. Azonosított gének

Az MCF-7, ECC-1, T-47D és Ishikawa sejtvonalakban átfedő, intronban található konszenzus kötőhelyeket 29 génhez tudtuk társítani, amelyekben a USCS Genome Browser Repeat Masker alkalmazásával megvizsgálva a kötőhelyek pontos lokalizációját, azok nem, vagy a teljes méretük 50%-ánál kevesebb bázispárnai régióban fedtek át repetitív régióval. A gének nevét a 7. táblázat tartalmazza.

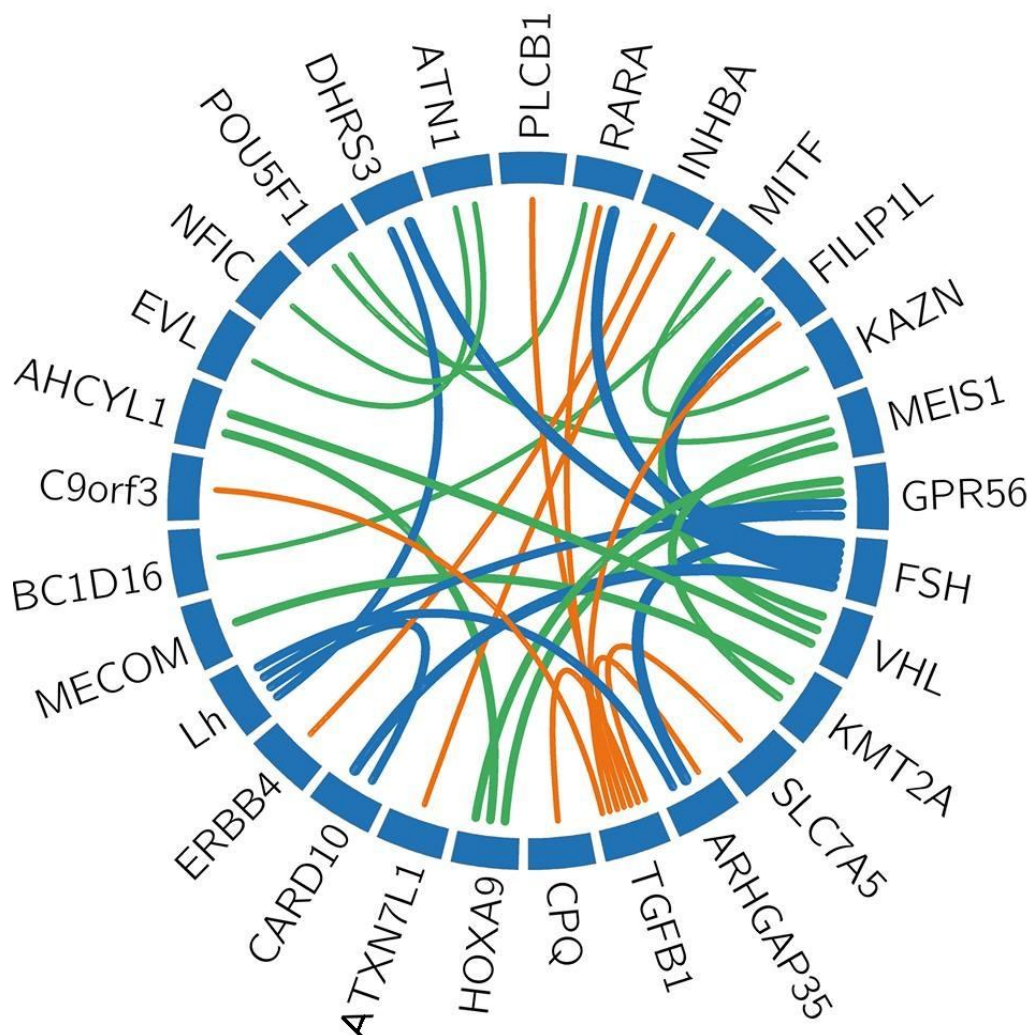
AFF3	CUEDC1	GPR56	POC1B	TEMEM120B
ANO1	DHRS3	KAZN	RBM20	UPP2
ARHGAP35	DIS3L2	MECOM	RNLS	VPS13D
ATXN7L1	EVL	MEIS1	ROBO2	WWC1
C9orf3	FAM86JP	NFIC	SLC22A5	XXYLT1
CELSR1	FSIP1	PLCB1	SLC7A5	-

7. táblázat. Az általunk azonosított gének. A táblázat tartalmazza, hogy a 4 sejtvonalban átfedő, intronban található kötőhelyek mely géneken találhatóak.

A nem intronban lokalizálódó kötőhelyeket sem génekhez, ezáltal regulátorokhoz sem rendeltük. Annak oka, hogy az intronban elhelyezkedő kötőhelyeket vizsgáltuk tovább, az volt, hogy ezeket a kötőhelyeket tudtuk csak biztosan egy-egy adott, biztos génhez társítani.

A gének lehetséges *upstream* regulátorait *circos* ábrán vizualizáltuk (12. ábra). A szabályozók és szabályozottak között nem sikerült olyan tulajdonságot találnunk, ami alapján csoportokba

sorolhattuk volna őket, mivel vannak közöttük olyanok, amelyek mind a szabályozó, mind target szerepet is betöltik. A regulátorok egy kivételével mind kapcsolatba hozhatóak az emlőrákkal az irodalmi adatok alapján, melyekről részletesebb információt a 8. táblázat tartalmaz.



11. ábra. Circos ábra a gének lehetséges upstream regulátorairól. A gének és regulátoraik közötti huzalok jelzik a közöttük fennálló kapcsolatot. A körvonalazódó gének és regulátoraik között nem lehetett különbséget tenni színezésükben, mert egyes elemek mind a szabályozó, mind a szabályozott szerepet is betöltik. A huzalok színezése: narancssárga: növekedési faktorokból kiinduló; kék: hormonokból kiinduló; zöld: transzkripciót szabályozó elemekből kiinduló. Regulátorok: FSH, Lh, POU5F1, VHL, ATN1, KMT2A, MITF, TGFB1, INHBA, HOXA91.

<i>Upstream regulátorok</i>	<i>Kapcsolatuk az emlőrakkal</i>	<i>Referencia</i>
TGFB1 , transforming growth factor, beta 1	az emlő szöveteinek epitheliális – mesenchymális differenciálódásában vesz részt	<i>David JH BMC Cancer 2007</i>
INHBA , inhibin, beta A	expressziója emelkedett szintet mutat az emlődaganatban	<i>Dennis CS Breast Cancer Res 2009</i>
MITF , microphthalmia-associated transcription factor	A BCSC-1 (breast cancer suppressor candidate-1) egy melanoma tumorsupresszor, amely alulszabályozza a MITF-et.	<i>Piguet V Pigment Cell Melanoma Res 2012</i>
KMT2A , lysine (K)-specific methyltransferase 2A	<i>nincs elérhető irodalmi adat a kapcsolatukról</i>	-
VHL , von Hippel-Lindau tumor suppressor	az utólagos VHL kötődés az emlőkarcinóma sejtek sejtproliferációjának csökkenését okozza	<i>Wendy WG Genes&Cancer 2013</i>
Lh , luteinizing hormone	a szérumban a normálhoz képest nincs szignifikáns eltérés a koncentrációjában, de az emlőrakos szövetekben emelkedett értéket mutat	<i>Hortobagyi GN Breast Cancer Res Treat 2002</i>
FSH , follicle stimulating hormone	a luteinizáló hormonnal (Lh) együtt az emlőrak progressziójában játszanak szerepet, HER2 ⁺ posztmenopauzális betegekben szintje emelkedett	<i>Suzhan Zhang Oncology Letters 2013</i>
ATN1 , athropin1	medulláris emlőkarcinómában szintje emelkedett	<i>Max C Cancer Res 2007</i>
POU5F1 , POU class 5 homeobox 1 (Oct-4)	expressziója magas CD44+/CD24– sejtekben, biomarkere az emlőrak progressziójának, differenciálódásának	<i>Lu P Ann Surg 2011</i>
HOXA9 , homeobox A9	szabályozza a BRCA1 (breast cancer 1) gén expresszióját	<i>Weaver VM J Clin Invest 2010</i>

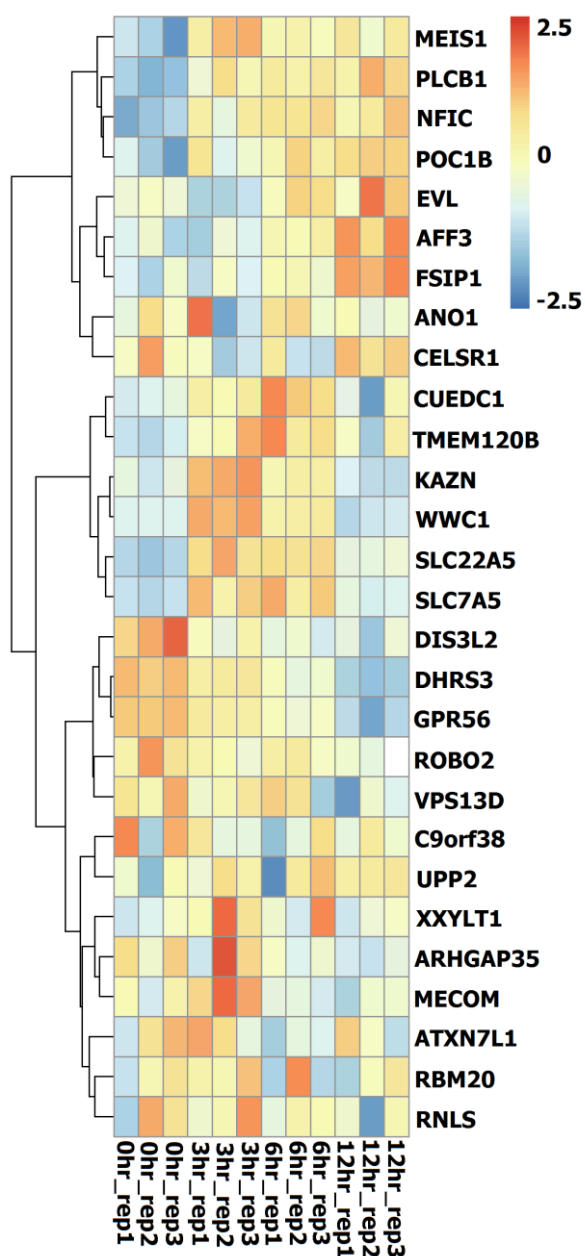
8. táblázat. A 29 gén lehetséges upstream regulátorai. A táblázat tartalmazza a regulátorokat és azok már ismert, az irodalmi adatok alapján alátámasztott szerepüket, kapcsolatukat az emlőrakkal.

Jason S. Carroll és munkatársai 2006-ban publikálták azt a tanulmányukat, amelyben ősszgenom-szinten vizsgálták az ösztrogén receptor kötőeseményeket MCF-7 sejtvonalon [11]. Azonosították a *cis*z kötőhelyeket és a *target* géneket, és ezen egyedi adatok kombinálása a génexpressziós adatokkal alátámasztották a különböző mechanizmusait az ösztrogén mediált génszabályozásnak, különösképpen az ösztrogén által elnyomott gének esetében. Az általunk azonosított gének közül egy kivételével (FAM86JP) mindegyik expressziójának a mértékét vizsgálták ebben a tanulmányban az ösztrogén (E2) kezelést követő 3-6-12. órákban. A géneknek a GEO adatbázisban elérhető microarray adataik alapján az adott időintervallumokban többszöri detektált expressziójuk mértékét hőtérképen szemléltettük (13. ábra).

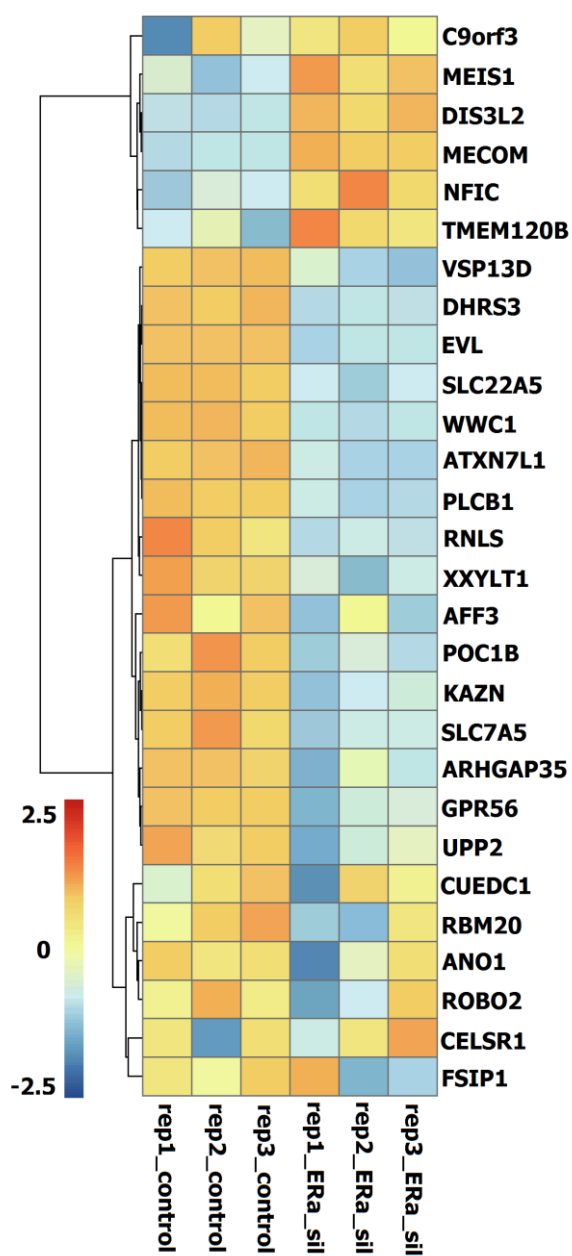
Az eredmények alapján nem lehet általánosítani abból a szempontból, hogy a stimulációt követően két kategóriára oszlanak-e a gének, mert egy csoportjának kifejeződése a 3. órát követően végig emelkedett marad, egy másik csoport kifejeződése hasonlóan megemelkedik és emelkedett marad a 12. órában is, azonban csak a 6. órától kezdve, míg van egy csoport, amely expressziója a 3-6. órában megnő, azonban a 12. órára visszacsökken. Ezen kezdeti alacsony expresszióval bíró gének csoportjával ellentétben a *heatmap*-en alul elhelyezkedő gének nagy része az ösztrogén stimuláció előtt, ún. kontroll állapotban már nagyobb mértékben expresszálódik, mint a stimulációt követő órákban.

Egy másik, 2011-es tanulmány során szintén emlő-eredetű, MCF-7 sejteken azt vizsgálták, hogy az ER α a funkciója elvesztése után, tehát kifejeződésének gátlását követően okoz-e hormonrezisztenciát, azaz eredményez-e transzdifferentiációt az epithelialis fenotípusból mesenchymalissá, amely folyamat felelős az emlőrák fokozott agresszivitásáért és a metasztatikus hajlamáért [30]. Az ER α siRNS által mediált inaktiválása ösztrogénre rezisztens, megváltozott morfológiájú sejteket eredményezett, melyek átrendeződtek és keratin/aktinból egy vimentin-alapú citoskeletonná váltak. Ezen eredmények betekintést nyújtanak az ösztrogén által szabályozott génexpresszió mechanizmusának megértésébe az emlőrák során, és arra utalnak, hogy az ösztrogén receptor funkciójának gátlását követően a korábban ösztrogénre érzékeny sejtek elvesztik hormonfüggő képességüket.

Az 14. ábrán látható, hogy a gének két csoportra oszlanak: az ER α által szabályozott gének expressziója jelentősen csökkent, azonban a gének egy kisebb csoportja esetében épp ellentétes hatás volt tapasztalható: kifejeződésüket az ER α jelenléte alapszinten gátolta, expressziójuk annak inaktiválása után megnőtt.



13. ábra. A gének expressziója az E2 kezelést követő 3-6-12. órákban. *Expressziójuk mértéke az idő függvényében eltérően változik.*



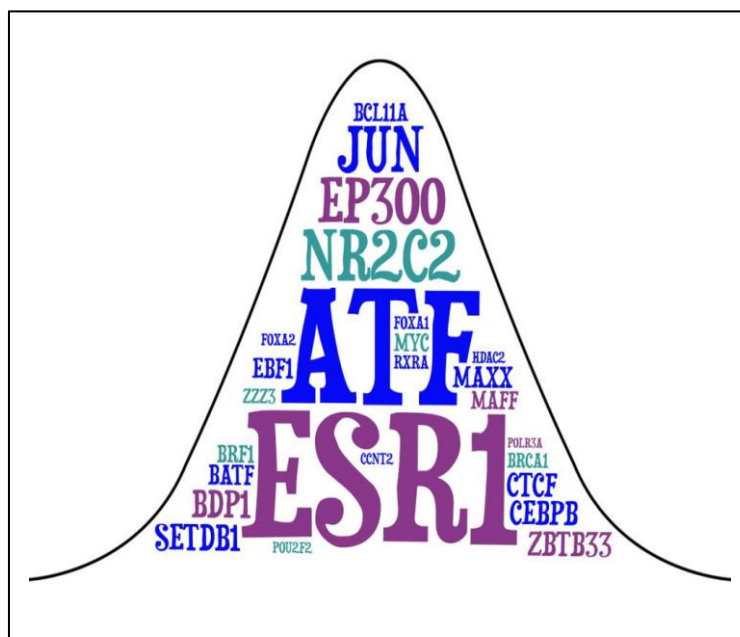
14. ábra. A gének expressziója az ER α gén lecsendesítését követően. Az ER α által szabályozott gének expressziója lecsökkent, a gének egy csoportján pedig pont ellentétes hatás volt tapasztalható.

7. Motívumkeresés

Az MCF-7, T-47D, ECC-1 és Ishikawa sejtvonalakban átfedő, 143 közös kötőhelyben található SNV-eket a FunSeq alkalmazással elemezve összesen 1250 olyan SNV-t térképeztünk fel, amelyek 26 transzkripciós faktor motívumát „bontották meg” jelenlétükkel. Az 1250 SNV közül legnagyobb számban, pontosan 280 darab az ER α motívumán belül volt jelen. A pontos adatokat, hogy az 1250 SNV mely transzkripciós faktorok motívumaiban voltak megtalálható és milyen számban, a 9. táblázat tartalmazza, ezen adatoknak a proporcionális ábrázolását pedig aszerint, hogy milyen gyakorisággal fordultak elő a különböző motívumok, a 15. ábrán látható. Az alkalmazás eredményként az adott transzkripciós faktort kódoló gén nevét adta meg, esetünkben tehát az ER α -t kódoló ESR1 gént.

TF	Db	TF	Db
ATF	100	FOXA2	1
BATF	10	HDAC2	1
BCL11A	7	JUN	27
BDP1	14	MAFF	11
BRCA1	6	MAXX	13
BRF1	4	MYC	4
CCNT2	1	NR2C2	53
CEBPB	15	POLR3A	2
CTCF	12	POU2F2	3
EBF1	7	RXRA	2
EP300	28	SETDB1	20
ESR1	280	SETD	25
FOXA1	2	ZZZ3	2

9. táblázat. Motívumok I. Azon transzkripciós faktorok, amelyek motívumain belül a 143 közös kötőhelyen belül azonosított 1250 SNV elhelyezkedik.

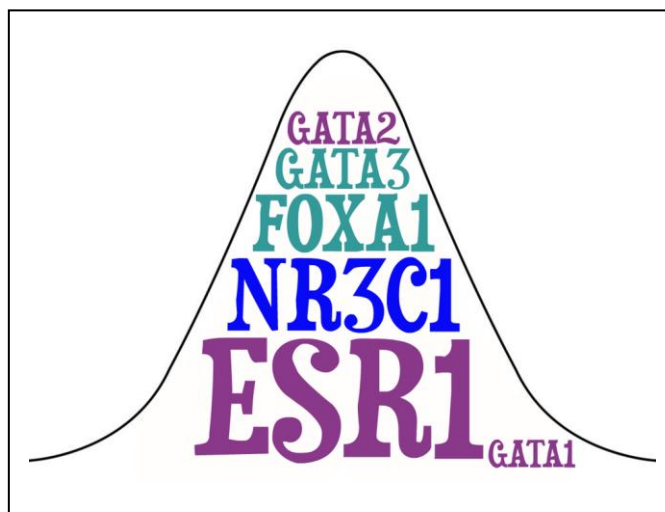


15. ábra. Motívumok I. Azon transzkripciós faktorok méretarányos ábrázolása, amelyek motívumában a 143 közös kötőhelyen belül azonosított 1250 SNV elhelyezkedik.

A következő két lépésben már csak a 280 ER α kötőhely motívumait vizsgáltuk tovább, hiszen az elemzés során mátrixszerűen több transzkripciós faktornak a motívumát is meghatározza a program, amelyeken belül az SNV-k egyaránt jelen voltak. Pontos adatokat a további, második és harmadik lehetséges TF-ok motívumairól és azok számáról a 9. és 10. táblázat tartalmaz, az eredmények proporcionális ábrázolása pedig a 15. és 16. ábrán látható.

TF	SNV
ESR1	219
FOXA1	10
GATA1	1
GATA2	1
GATA3	6
NR3C1	15

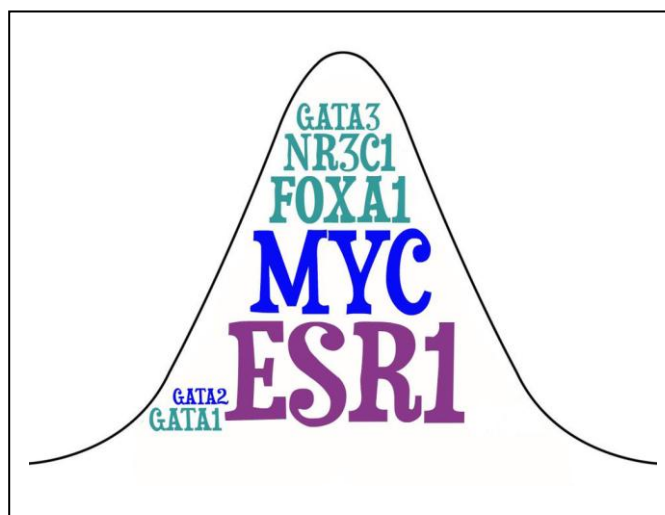
10. táblázat. Motívumok II. A 280 ER α (ESR1) motívumán belül azonosított, második lehetséges transzkripciós faktorok motívumai és a bennük talált SNV-k száma.



16. ábra. Motívumok II. Méretarányos ábrázolása annak a 280 ER α (ESR1) TF motívumai mellett talált TF-oknak, amelyek motívumában a 280 SNV-ből valamennyi jelen volt.

TF	SNV
ESR1	86
FOXA1	24
GATA1	3
GATA2	2
GATA3	5
MYC	38
NR3C1	17

11. táblázat. Motívumok III. A 280 ER α (ESR1) motívumán belül azonosított, harmadik lehetséges transzkripciós faktorok motívumai és a bennük talált SNV-k száma.



17. ábra. Motívumok III. Méretarányos ábrázolása annak a 280 ER α (ESR1) TF motívumai mellett talált TF-oknak, amelyek motívumában a 280 SNV-ből valamennyi SNV szintén jelen volt.

A motívumkeresés első lépésekor az SNV-k az ER α mellett az ATF (activating transcription factor) jelent meg, motívumában összesen 100 SNV-vel. Az ATF az AP-1 transzkripciós faktor dimerek egy családja, amelyről a bevezetésben már kiderült, hogy az ER α képes a indirekt módon más transzkripciós faktorokhoz kötődve, mint például az AP-1-hez kötődve a DNS-hez kapcsolódni [34]. Viszonylag nagy számmal volt jelen SNV az EP300 motívumában, amely az ösztrogén receptor egy koaktivátoraként ismert [35].

A további lehetséges transzkripciós faktorok száma, melyek motívumaiban a korábbi, ER α motívumában jelen lévő SNV-k elhelyezkedhettek, ugyan leeredukálódtak, de – talán nem meglepő módon egy glükokortikoid hormonreceptor (GR), mint az NR3C1 és a már ER α -val számos tanulmányban vizsgált és azonosított kooperáló faktorok, mint a FoxA1 és GATA3 jelentek meg nagyobb számban. Génexpressziós analízisek kimutatták, hogy azok az ER $^{-}$ tumorok, melyek nagymértékben expresszálják glükokortikoid hormonreceptorokat, epitheliális-mesenchymális átalakulást, sejtadhéziót és gyulladásos folyamatokat indítanak el. A közvetlen transzkripciós szerepét a GR-nak a rossz prognózisú ER $^{-}$ emlőrákban azonosították [36].

A Myc szerepét a STAT3-mal való kölcsönhatásán keresztül írták le: az abnormális mértékben expresszálódó STAT3-nak (signal transducer and activator of transcription 3) elsősorban az ER $^{+}$ emlőrák rossz prognózisának előrejelzésében van szerepe. Az aktivált STAT3 kölcsönhatásban van több szabályozási útvonallak, beleértve a STAT3-MYC-et, STAT3-ER α -t és a STAT3-MYC-ER α kölcsönhatást. Ezen útvonalak aktiválódása segít megjósolni a rossz kimenetelű eseményeket és a kezelésre adott választ, továbbá egy újabb oka a tamoxifenre való rezisztens állapot kialakulásának [37].

Az ER α -nak a DNS transzkripciót szabályozó régiójához való kötődése 95%-ban a FoxA1 által mediált. Motívuma az ER α motívumán belül dúsulást mutat, 50-60%-ban azzal át is fed és szükséges az ER α DNS-sel való kölcsönhatásának fenntartásához. Döntő szerepet játszik az emlőrák korai fázisától kezdve a gyógyszerrezisztencia, ill. a metasztatikus állapot kialakulásában is.

A FoxA1 mellett a GATA3 motívum is feldúsul az ER α motívumán belül, és mint azzal kooperáló transzkripciós faktor tartják számon. Az emlőrák expressziós profiljának kombinálása vezetett a FoxA1 és GATA3 felfedezéséhez és karakterizálásához. A FoxA1

kötődésének gátlása leredukálja az ER α kötődésének kapacitását és a transzkripció aktivitást, míg a sejtspecifikus FoxA1 eloszlása megmagyarázza a különféle ER α kötődési profilokat az emlőrákos sejtekben. A FoxA1 és GATA3 túlzott kifejeződése az ER $^{-}$ emlőrákban bizonyos mértékig képes azt ösztrogénre válaszul bírni – mutatva ezzel, hogy ezen képességük hatással van az ER α kötődésre [38-42].

Elmondható, hogy az ER α DNS motívumán belül talált SNV-k legnagyobb számban olyan transzkripció faktorok motívumaiban voltak jelen, melyek kivétel nélkül összefüggésbe hozhatóak az emlőrákkal. Jelenlétük kiemelkedő szereppel bír az ER α általi transzkripció szabályozásban, és a szekvenciájukban található variánsok ezen szabályozást befolyásolni képesek.

Jelen munkánk során négy ER α pozitív sejtvonal ER kötőhelyeinek meta-analízisét végeztük el. A hierarchikusan felépített bioinformatikai elemzések során külön hangsúlyt fektettünk az emlő- és endometrium eredetű sejtvonalakban feltérképezett kötőhelyek esetleges különbözőségeinek vizsgálatára, de fő célunk az összes általunk vizsgált sejtvonalakban átfedő kötőhelyek együttes elemzése volt.

Az MCF-7 és T-47D, mint emlő-eredetű sejtvonalak konszenzus kötőhelyeinek száma nagyobbak mutatkozott, mint bármelyik endometrium-eredetű sejtvonallal átfedő kötőhelyeik száma, és ugyanez a megállítás érvényes volt az ECC-1 és Ishikawa sejtvonalak konszenzus kötőhelyeire is. A négy sejtvonalban átfedő, összesen 143 ER α kötőhely közül azokat, amelyek az intronban lokalizálódtak, 29 génhez sikerült társítani. Az így azonosított gének és az általunk meghatározott regulátorok, illetve maga az emlőrák közötti kapcsolat néhány gén kivételével már szerepel az irodalomban. A FAM86JP gén kivételével ezen ER α által szabályozott gének expresszióját vizsgálva elmondható, hogy kifejeződésük az ER α siRNS általi inaktiválása után csökkent, azonban a gének egy csoportjánál ellentétes hatás volt megfigyelhető. Kifejeződését az ER α alapállapotban gátolta, annak lecsendesítése után expressziójuk megnőtt. A FAM86JP génnel kapcsolatban az irodalomban nem találtam releváns adatot. A GEO adatbázisban egyetlen expressziós adat található a génről, a tripla-negatív emlőrák vonatkozásában (ER $^{-}$, PR $^{-}$, HER2 $^{-}$). A gén expressziós adatai nem mutatnak szignifikáns eltérést a gén kifejeződésének mértékében a tripla-negatív és a nem tripla-negatív emlőrákos betegek között.

A teljes genomi régióban elhelyezkedő 143 közös kötőhelyen belül azonosítva az SNV-eket, majd megvizsgálva, hogy mely transzkripciós faktorok DNS motívumában találhatók meg, talán nem meglepő, hogy az ER α motívumában voltak jelen a legnagyobb számban.

Az ER α kötődése a DNS-hez megvalósulhat mind közvetlen, mint közvetett módon, ugyanakkor a közvetlen folyamat során is más, kooperáló transzkripciós faktorok előzetes kötődése segítheti és/vagy indukálhatja a megfelelő DNS szekvenciához történő kapcsolódását. Ezen TF-ok motívumait azonosítottuk mi is olyan ER α kötőeseményeken belül is, ahol a variánsok jelen voltak.

Az emlőrákban ha ezek a DNS szekvenciáiban létrejövő eltérések - esetleges mutációk - jelen vannak, csökkenthetik a FoxA1 kötődésének kapacitását. Mivel az ER α DNS-hez kötődése függ a FoxA1 kötődésétől, ez arra utal, hogy a tumor kialakulása és növekedése függetlenné válhat az ösztrogéntől. Ezen tumorok valószínűleg agresszívvabbak lesznek, hiszen nem reagálnak az általánosan alkalmazott hormonterápiára [43].

Az ilyen módon történő, különböző ER α kötőesemények és a bennük történő funkcionális változások hozzájárulnak az emlőrák kialakulásának folyamatához, megismerésük pedig a betegség molekuláris szinten történő megértéséhez.

IV. Köszönetnyilvánítás

Köszönöm Prof. Dr. Nagy László hozzájárulását, hogy pályamunkámat a Klinikai Genomikai és Személyreszabott Orvoslási Központban írhattam.

Szeretném megköszönni témavezetőmnek, Dr. Bálint Bálint Lászlónak, hogy legjobb tudásával irányította munkámat és értékes tanácsaival, bizalmával mindvégig támogatott.

Külön köszönettel tartozom Horváth Attilának a több hónapos folyamatos munkájáért, amely idő alatt legjobb szakmai tudása szerint tanított és lépésről-lépésre sajátíthattam el tőle az egyes bioinformatikai módszerek használatát.

Továbbá köszönettel tartozom Dr. Barta Endre, Jónás Dávid, Nagy Gergely, Czipa Erik bioinformatikusoknak, illetve Dr. Póliska Szilárdnak és Erdős Edinának a munkám során nyújtott segítségükért.

V. Irodalomjegyzék

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. (2010) *Int J Cancer*. 127(12):2893-917.
2. Zwart W, Koornstra R, Wesseling J, Rutgers E, Linn S, Carroll JS. (2013) A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions from breast cancer core needle biopsy samples. *BMC Genomics*. 2013 Apr 8;14(1):232.
3. Beatson GT. (1896) On the treatment of inoperable case of carcinoma of the mamma: suggestions for new method of treatment with illustrative cases. *Lancet*. 2:104-7.
4. Gorski J et al. (1966) A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. *Proc Natl Acad Sci USA*. 55(6) 1574–81.
5. EV Jensen, T Suzuki, T Kawashima, WE Stumpf, PW Jungblut, ER DeSombre. (1968) A two-step mechanism for the interaction of estradiol with rat uterus. *Proc Natl Acad Sci USA*. 59(2) 632–8.
6. Kumar V, Green S, Staub A, Chambon P. (1896) Localisation of the oestradiol-binding and putative DNA-binding domains of the human oestrogen receptor. *Embo J*. 5(9) 2231–6.
7. Shao W, Brown M. Advances in estrogen receptor biology: prospects for improvements in targeted breast cancer therapy. *Breast Cancer Res*. 2004;6:39–52
8. Gerald M. Higa, Ryan G. Fell. (2013) Sex Hormone Receptor Repertoire in Breast Cancer. *Int J Breast Cancer*. 2013:284036.
9. Katzenellenbogen BS. (1996) Estrogen receptors: bioactivities and interactions with cell signaling pathways. *Biol Reprod*. Feb;54(2):287-93.
10. Klein-Hitpass L, Schorpp M, Wagner U, Ryffel GU. (1986) An estrogen-responsive element derived from the 5' flanking region of the *Xenopus vitellogenin A2* gene functions in transfected human cells. *Cell*. 26;46(7):1053-61.
11. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF et al. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nature Genetics*. 38 1289–1297.
12. Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL et al. (2007) Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet*. 2007 Jun;3(6):e87.

13. Kaczynski J, Cook T, Urrutia R. (2003) Genome Biol. Sp1- and Krüppel-like transcription factors. 4(2):206.
14. Safe S, Abdelrahim M. (2005) Sp transcription factor family and its role in cancer. Eur J Cancer. 2005 Nov;41(16):2438-48
15. Soule HD, Vazquez J, Long A, Albert S, Brennan M. (1973) A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst. 51(5):1409-16.
16. Anait S. Levenson and V. Craig Jordan. (1997) MCF-7: The First Hormone-responsive Breast Cancer Cell Line. Cancer Res. 57:3071-3078.
17. Keydar I, Chen L, Karby S, Weiss FR, Delarea J, Radu M, Chaitcik S, Brenner HJ. (1979) Establishment and characterization of a cell line of human breast carcinoma origin. European Journal of Cancer **15** (5): 659–670.
18. Bilan Mo et al. (2006) ECC-1 cells: a well-differentiated steroid-responsive endometrial cell line with characteristics of luminal epithelium. Biol Reprod. 75(3):387-94.
19. Nishida M. (2002) The Ishikawa cells from birth to present. Hum Cell.15(3):104-17.
20. Peter J Park. (2009) ChIP-Seq: advantages and challenges of a maturing technology. Nat Rev Genet. 10(10): 669–680.
21. Endre Barta. (2011) Command line analysis of ChIP-seq results. Embnet Journal 17(1) pp. 13-17.
22. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative genomics viewer. Nat Biotechnol. 29(1):24-6.
23. Zhang et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9: R137.
24. Feng J, Liu T, Zhang Y. (2011) Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics. Chapter 2:Unit 2.14.
25. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature. 6;489(7414):57-74.
26. Bradford JR, Hey Y, Yates T, Li Y, Pepper SD, Miller CJ. (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics. 11:282.
27. Murdoch, D.J. and Chow, E.D. (1996) A graphical display of large correlation matrices. The American Statistician 50, 178-180.

28. Kolde R (2012) pheatmap: Pretty Heatmaps. R package version 061.
29. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* 19(9):1639-45.
30. Al Saleh S1, Al Mulla F, Luqmani YA. (2011) Estrogen receptor silencing induces epithelial to mesenchymal transition in human breast cancer cells. *PLoS One.* 2011;6(6):e20610.
31. Gernstein M et al. (2013) Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science.* 4;342(6154):1235587.
32. Osborne CK, Hobbs K, Trent JM. (1987) Biological differences among MCF-7 human breast cancer cell lines from different laboratories. *Breast Cancer Res Treat.* 9(2):111-21.
33. Bahia H, Ashman JN, Cawkwell L, Lind M, Monson JR, Drew PJ, Greenman J. (2002) Karyotypic variation between independently cultured strains of the cell line MCF-7 identified by multicolour fluorescence in situ hybridization. *Int J Oncol.* 20(3):489-94.
34. van Dam H1, Castellazzi M. (2001) Distinct roles of Jun : Fos and Jun : ATF dimers in oncogenesis. *Oncogene.* 20(19):2453-64.
35. Wirtenberger M, Tchatchou S, Hemminki K, Schmutzhard J, Sutter C et al. (2006) Associations of genetic variants in the estrogen receptor coactivators PPARGC1A, PPARGC1B and EP300 with familial breast cancer. *Carcinogenesis.* 27(11):2201-8.
36. Pan D1, Kocherginsky M, Conzen SD. (2011) Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. *Cancer Res.* 71(20):6360-70.
37. Liu LY, Chang LY, Kuo WH, Hwa HL, Lin YS, Jeng MH, Roth DA, Chang KJ, Hsieh FJ. (2014) Prognostic features of signal transducer and activator of transcription 3 in an ER(+) breast cancer model system. *Cancer Inform.* 13:21-45.
38. Ross-Innes CS1, Stark R, Teschendorff AE, Holmes KA, Carroll JS et al. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 481(7381):389-93.
39. Robinson JL, Carroll JS. (2012) FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Front Endocrinol (Lausanne).* 18;3:68.

40. Welboren WJ1, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG. (2009) ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* 2009 May 20;28(10):1418-28.
41. Jason S Carroll. (2013) Understanding estrogen receptor transcription in breast cancer. *Clin Epigenetics.* 5(Suppl 1): S7.
42. Theodorou V, Stark R, Menon S, Carroll JS. (2013) GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* 2013 Jan;23(1):12-22.
43. Hurtado A., Holmes K. A., Ross-Innes C. S., Schmidt D., Carroll J. S. (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* 43, 27–33.

VI. Függelék

A szerző önállóan végezte a következőket:

- ChIP-seq adatok gyűjtése a GEO és SRA adatbázisból
- ChIP-seq adatok teljeskörű bioinformatikai elemzése
- ábrák, táblázatok szerkesztése
- irodalmi háttér tanulmányozása