

Doktori (PhD) értekezés tézisei

**Interferon reguláló faktorok DNS kötőhelyeinek
genomszintű vizsgálata dendritikus sejtekben**

Dr. Nagyné Csumita Mária

Témavezető: Dr. Széles Lajos István



Debreceni Egyetem

Molekuláris Sejt- és Immunbiológia Doktori Iskola

Debrecen, 2022

**Interferon reguláló faktorok DNS kötőhelyeinek genomszintű vizsgálata
dendritikus sejtekben**

Értekezés a doktori (PhD) fokozat megszerzése érdekében
az elméleti orvostudományok tudományágban

Írta: Dr. Nagyné Csumita Mária
okleveles biomérnök

Készült a Debreceni Egyetem Molekuláris Sejt- és Immunbiológia Doktori Iskolája keretében

Témavezető: Dr. Széles Lajos István

A doktori szigorlati bizottság:

elnök: Prof. Dr. Szabó Gábor, az MTA doktora
tagok: Prof. Dr. Boros Imre Miklós, az MTA doktora
Dr. Lányi Árpád, PhD

A doktori szigorlat időpontja: Debreceni Egyetem ÁOK, Élettudományi Központ,
Biokémiai és Molekuláris Biológiai Intézet, 2019. július 01. 12 óra

Az értekezés bírálói:

Prof. Dr. Emri Tamás, az MTA doktora
Dr. Pankotai Tibor, PhD

A bírálóbizottság:

elnök: Prof. Dr. Fésüs László, az MTA rendes tagja
tagok: Prof. Dr. Emri Tamás, az MTA doktora
Dr. Pankotai Tibor, PhD
Gálné Dr. Miklós Ida, PhD
Prof. Dr. Boros Imre Miklós, az MTA doktora

Az értekezés védésének helyszíne és időpontja:

Debreceni Egyetem ÁOK, Sürgősségi Orvostani Tanszék előadóterme
2023. február 3. 11 óra

Bevezetés

1.1. A dendritikus sejtek

Zanvil Cohn és Ralph Steinman 1973-ban publikálták megfigyelésüket, miszerint egér lép sejtjei között mikroszkóppal egy addig ismeretlen sejtípust fedeztek fel, amelyet fa formájú nyúlványai után a görög eredetű „dendron” (fa) szóból dendritikus sejtnak (dendritic cell, DC) neveztek el. Az azóta eltelt időben számos altípust írtak le, és bebizonyosodott, hogy a DC-k kulcsszerepet játszanak többféle immunválasz kiváltásában és szabályozásában. Eredetük alapján megkülönböztetünk mieloid és limfoid DC-eket. A mieloid DC-k az immunrendszer leghatékonyabb hivatásos antigén-prezentáló sejtjei, résztvesznek immunogén és tolerogén immunválaszok kialakításában, míg a plazmacitoid DC-k szerepe elsősorban a vírusellenes immunválasz kialakítása I-es típusú interferon (I-IFN) termelése által. A DC-k aktiválása patogén jelenlétében történik meg, ugyanis olyan mintázatafelismerő receptorokat (PRR) fejeznek ki, amelyekkel detektálják a patogéneken található konzervált struktúrákat (patogén asszociált molekuláris mintázatot(kat), PAMP). A mintázatafelismerő receptorok elsősorban transzmembrán fehérjék, de ismerünk citoplazmikus PRR-okat is.

1.2. A dendritikus sejtek altípusai egérben

Számos dendritikus sejt alpopulációt azonosítottak emberben és egérben. Munkánk során egér DC-ekkel dolgoztunk. A különböző DC populációk differenciálódását citokinek és transzkripciós faktorok (TF) szabályozzák. A csontvelőben kezdődik a prekursor DC-k differenciálódása mieloid és limfoid progenitor sejtekből.

A pDC-k szerepe elsősorban a vírusellenes immunválasz kialakítása és a vírusfertőzés hatására nagy mennyiségű I-IFN termelése, de továbbá rendelkeznek antigén prezentáló képességgel is. A Langerhans sejtek a bőrszövet epidermiszében helyezkednek el. A migratórikus dendritikus sejtekhez hasonlóan az aktivációt és az antigén felvételt követően a nyirokcsomókba vándorolnak, ahol antigént prezentálnak és stimulus-függő módon szabályozzák a génexpressziót.

A monocita eredetű dendritikus sejtek gyulladásos állapotban differenciálódnak a vérben keringő monocitákból. Számos dendritikus sejtekre jellemző tulajdonsággal rendelkeznek. A monociták makrofág kolóniastimuláló faktor receptort (MCSFR) expresszálnak, amely nélkülözhetetlen a differenciálódásuk szempontjából.

A migratórikus DC-k az antigén feldolgozására és bemutatására specializálódnak, a test különböző szöveteiben megtalálhatóak, antigén felvevő „őrszemként” működnek. A különböző alpopulációkat a sejtfelszíni molekulák kifejeződése alapján különböztetjük meg. A migratórikus CD11+ és CD103+ dendritikus sejtek a kórokozók felvétele és patogén asszociált molekuláris

mintázatának (PAMP) felismerése után éresen mennek át, mely elindítja migrációjukat a közeli nyirokcsomókba. Az érés során megváltozik a morfológiájuk, nő az antigén feldolgozó képességük és a sejt-aktiváló kapacitásuk. Eközben a patogének antigénjeit feldolgozzák és a fehérjéket peptidekké darabolják. Az alapvető különbség a migratórikus CD11⁺ DC és a migratórikus CD103⁺ DC között az, hogy az előbbiek az antigént MHC-II-n (direkt prezentáció), míg utóbbiak elsősorban MHC-I-n keresztül (keresztprezentáció) mutatják be.

A nyirokszövet-rezidens DC-k különböző nyirokszövetekben találhatóak meg, például a lépben, a csecsemőmirigyben és a nyirokcsomókban. A nyirokszövet-rezidens DC-keket a CD4 és CD8 α sejtfelszíni markerek expressziója alapján csoportosíthatjuk. A marker CD8⁺ sejtfelszíni molekula az MHC-I és a T-sejt receptor közötti interakciót segítő koreceptor. A nyirokszövet-rezidens DC-keket a vérben található, illetve a migratórikus DC-k által nyirokszövetekbe szállított PAMP-ok aktiválják, illetve így módon veszik fel a bemutásra kerülő antigéneket is. A CD103⁺ migratórikus DC-khez a CD8⁺ sejtek hasonlóak abban a tekintetben, hogy képesek az antigének keresztprezentációjára az MHC-I-membránfehérjéken keresztül. A CD8⁺ DC-k a nyirokszövetekben található prekursor DC-kből differenciálódnak, fejlődésükhöz kulcsfontosságú citokin az Flt3L. Több transzkripciós faktor is nélkülözhetetlen a fejlődésükhöz, ilyen például az PU.1, IRF8 és Batf3. A CD8⁺ DC-k alapállapotban segítenek fenntartani a saját szövetekkel szembeni toleranciát. Az éretlen sejtek főként a patogén felismerését és felvételét elősegítő receptorokat fejeznek ki nagy számban. Legfőbb mintázatfelismerő receptorai a TLR-ok (TLR3 és TLR9), és a C-típusú lektin receptorok (CLR). Stimuláció és érés után megnövekedik az MHC és a kostimulációs molekulák expressziója, továbbá nagy mennyiségben termelnek citokineket, pl. IL-12-t és más gyulladásos citokineket. Eközben a felvett antigéneket MHC-I-en keresztül prezentálják (keresztprezentáció), ezért kiemelt szerepük van a citotoxikus CD8⁺ T sejtek aktiválásában, és az antivirális immunválasz indukálásában.

1.3. A MutuDC CD8⁺ sejtvonala

A DC-k viszonylag kis mennyiségben fordulnak elő egerben és emberben, izolálás után pedig érzékeny sejtek. Az eger csontvelőből (vagy emberi vérből) lehet prekursorokat izolálni, de az „*in vitro*” differenciáltatás több napot vesz igénybe és citokinek használata szükséges hozzá, tehát időigényes és költséges. Ennek kiküszöbölésére Hans Acha-Orbea és munkatársai (Unil, Lausanne, Svájc) eger lépéből származó CD8⁺ dendritikus sejtekből létrehoztak egy immortalizált sejtveget (MutuDC1940), amely egy nagy számban kinyerhető, homogén sejtpopuláció. Közvetlen összehasonlítás során bizonyították, hogy a MutuDC sejtvonala megtartotta a primer CD8⁺ DC fenotípusos és funkcionális tulajdonságait. Ezt a sejtveget használtuk kísérleteinkhez.

1.4. A stimulus-regulált transzkripció faktorok (SRTF-ek)

A DC-k a környezetükből érkező szignálokat képesek érzékelni és integrálni, majd ezek alapján megfelelő immunválaszt indukálni. Az immunválasz kiváltása együtt jár azzal, hogy a DC-ben több száz gén expressziója megváltozik. A transzkripció szabályozást olyan transzkripció faktorok (TF-ok) végzik, melyek aktivitását képesek a szignálok módosítani. Ezeket a TF-okat Glass és Natoli nyomán stimulus-regulált transzkripció faktoroknak (stimulus-regulated transcription factor, SRTF) nevezik. A SRTF-ek kofaktorai a hisztonok és kromatinszerkezet módosításán és egyéb mechanizmusokon keresztül az RNS polimeráz II (Pol-II) aktivitását szabályozzák. Azt, hogy egy sejtípusban az aktivált SRTF egy adott gént milyen mértékben képes szabályozni, alapvetően három tényező határozza meg: (i) a specifikus DNS motívumok jelenléte, (ii) a szabályozó elem „használatossága” és kapcsolata a promóterrel, illetve (iii) egyéb TF-ok kötése.

Az első SRTF-kötést meghatározó tényező a specifikus DNS motívum jelenléte. Egy molekula SRTF DNS-kötő doménje általában egy 4–6 bp hosszúságú DNS szakaszt köt. Az SRTF-ek dimerként vagy trimerként kötnek, ezek teljes felismerőhelye a félhelyek különböző kombinációjából tevődik össze. Az optimális szekvenciát általában kanonikus DNS szekvenciának nevezik. A kanonikus DNS szekvencia jelenléte se nem szükséges, sem nem elégséges feltétele egy adott SRTF kötésének. A kanonikus DNS szekvencia jelenléte azért nem szükséges, mert az SRTF-ek többféle, a kanonikus szekvenciától némileg eltérő szekvenciákat is tudnak kötni, melyeket összegző mátrix-szal (Position weight matrix, PWM) adhatunk meg. A kanonikus DNS szekvencia jelenléte pedig azért nem elégséges, mert a kondenzált kromatinszerkezet és a metilált DNS képes gátolni az SRTF-ek kötődését még a kanonikus DNS szekvenciákhoz is.

A második, SRTF-kötést meghatározó tényező tehát a szabályozó elem „használatossága” (ezt a kromatin nyitottság, hiszton módosítások, DNS metiláltság stb. együttesen határozzák meg) valamint a kromatin 3D szerkezete (enhanszer-promóter kapcsolata). Ezek közül a kromatin nyitottság és a hiszton acetiláció a legáltalánosabban vizsgált tényezők, mert a nyitott, nukleoszómamentes kromatin egyértelműen kedvez az SRTF kötésnek, míg az acetilált hisztonok (elsősorban H3 és H4 lizin oldalláncain) az aktív szabályozó régiókkal mutatnak asszociációt. A különböző sejtípusokban az aktív vagy aktiválható szabályozó elemek globálisan azonosíthatóak ATAC-seq-el (Assay for Transposase Accessible Chromatin sequencing). Hiszton módosítások genom szintű feltérképezését kromatin immunprecipitációt követő szekvenálással (ChIP-seq) és az adott hiszton módosításra specifikus antitest alkalmazásával végzik. A hiszton módosításokon kívül a ChIP-seq technológia alkalmas a különböző sejtípusok promóter és enhanszer régióinak a feltérképezésére is. A sejtípus specifikus enhanszer repertoár kialakításáért a sejtípus mester-regulátor TF-jai (angolban „lineage determining TFs”, LDTFs) a felelősek. Promoterek esetében

egyéb TF-ok is képesek az aktiválhatóság létrehozására. Az LDTF-ek befolyásolják az adott sejttípus kialakulását, a szövetre/sejttípusra jellemző gének expressziójának fenntartását, illetve az enhanszerek nyitottságát. Az SRTF-ek nagyrészt (>70 %) olyan régiókba kötnek, amelyek nyitottak vagy egyéni enhanszer jellegzetességet mutatnak, és csak kisebb részben korábban nem nyitott („látens” vagy „de novo”) régiókba. A nyitottság kialakításában általában több LDTF és kollaborátor TF együttesen vesz részt. Mivel az aktív és aktiválható enhanszer elemek nagy részét LDTF-k hozzák létre, tehát sejttípus függőek, az adott stimulusra kialakuló SRTF-függő válasz nagyrészt szintén sejt-specifikus.

A harmadik, SRTF-kötést meghatározó tényező az egyéb TF-ok kötése. Az egyes szabályozó elemek gyakran többféle SRTF kötőhelyét is tartalmazzák, ezzel lehetővé téve az SRTF-ek kötését. A gyulladásos szignálok hatására például, TLR aktiválást követően, több SRTF egyszerre aktiválódik, és sokszor egymás közelében kötnek.

1.5. Az IRF transzkripciós faktorok

Az Interferon reguláló faktor család tagjai az I-IFN-ok indukálása mellett kiemelt szerepet játszanak a sejten belüli antivirális válasz kialakításában, a gyulladásos immunválaszokban, és bizonyos sejttípusok differenciálódásában. Az IRF-ek kiesése vagy funkciójuk megváltozása specifikus immunválaszok hiányához vagy onkogenezishez vezethet, illetve olyan autoimmun betegségekkel hozható kapcsolatba, mint például a gyulladásos bélbetegség, a szisztémás lupus erythematosus és a rheumatoid arthritis. Az IRF család emberben és egérben 9 fehérjét tartalmaz. Az IRF család tagjai eltérést mutatnak szöveti kifejeződés, aktiválás, dimerizáció és komplexképzés, illetve a szabályozott transzkripciós programok szempontjából.

1.5.1. Az IRF-fehérjék szerkezete

Az IRF-fehérjék alapvető szerkezeti elemei az N-terminális DNS-kötő domén (DNA binding domain, DBD) és a C-terminális asszociációs domén (IRF association domain, IAD). A DBD által felismert DNS szekvencia az 5'-GAAA-3' tetranukleotid, amelyből a kanonikus IFN-stimulált válaszadó elem (ISRE, 5'-GAAANNGAAA-3') kettő található. Az IAD felelős más TF-ral történő kölcsönhatás kialakításáért. Egyes IRF-fehérjék, mint például az IRF1, IRF3 és az IRF7 homo- vagy heterodimerket képez IRF-ekkel.

Vannak IRF-k (IRF4, IRF8 és IRF9), amelyek más családba tartozó TF-ral lépnek kapcsolatba és azokkal alkotott dimerként vagy trimerként kötődnek a DNS-hez. Azok a TF-ok, amelyeket az IRF4 és IRF8 partnereként írtak le, például a PU.1. A STAT1 (Signal transducer and activator of transcription 1) és STAT2 (Signal transducer and activator of transcription 2) TF-okkal alkot heterotrimert az IRF9, így léterhozva az ISGF3 (Interferon-stimulated gene factor 3)

komplexet. A poszttranszlációs módosítások szabályozzák az IRF-fehérjék aktivitását és a közöttük lévő kölcsönhatásokat.

1.5.2. AZ IRF-ek által felismert DNS motívumok

Ebben az alfejezetben azokat a DNS szekvenciákat és motívumokat tekintjük át, amelyekhez IRF-ek kötődnek. Mindegyik szekvencia tartalmaz legalább egy IRF-kötő félhelyet (5'-GAAA-3'), de ezen túl nagy változatosság figyelhető meg. „Kanonikus szekvenciának” nevezzük azokat a szekvenciákat, ahol minden pozícióban csak egy bázis van megadva.

ISRE és TISRE

Az eredeti konszenzus ISRE-t (12-15 bp, néhány I-IFN-stimulált génből) a következőképpen határozták meg: YAGTTTC(A/T)YTTYCC és AGTTTCNNTTTCNC/T. Ezek a jelenlegi fogalmaink szerint tehát az ISGF3 kötőhelyei. A 10 bp hosszú ISRE-re (GAAANNGAAA) a kanonikus ISRE szekvencia. Az egyes IRF-eknek lehet preferenciája az ISRE-n túlnyúló 5' és 3' pozícióban lévő bázisok tekintetében. A lehetséges kötőhelyek összessége sokszor nehézkesen adható meg betűkóddal, ezeket olyan motívum logóval szokták megjeleníteni, ami a PWM alapján készül. Az 5'-GAAA-3' szekvencia 3-4 tandem ismétlődésének jelenlétét számos effektor ISG (interferon-stimulated gene, IFN-stimulált gén) szabályozó elemében dokumentálták. A három egymást követő ISRE-félhelyet (továbbiakban: tripartite ISRE, TISRE) az IRF1, IRF2 és az IRF8 fehérjék kötőhelyeként azonosították.

EICE, EIRE, IECS és AICE

Az IRF4 és IRF8 fehérjék kötőhelyeként azonosítottak egy 10 bázis hosszú ETS-IRF kompozit elemet (ETS-IRF composite element, EICE, 5'-GGAANNGAAA-3') azon gének promóter régióiban, amelyek alapvető fontosságúak a makrofágok és a B-sejtek működéséhez. Az EICE egyik feléhez (GAAA) IRF4 vagy IRF8 kapcsolódik, míg a másik feléhez (GGAA) az ETS (E26 transformation-specific vagy Erythroblast Transformation Specific) TF kötődik. Az EICE motívumhoz hasonlóan az IRF4/8-PU.1 komplex kötődhet egy 11 bázis hosszú ETS-IRF válaszadó elemhez is (ETS-IRF response element, EIRE, 5'-GGAANNNGAAA-3'), amely 3 bázis távközzel rendelkezik. Egy alternatív és kevésbé gyakori elem az IRF-ETS kompozit szekvencia (IECS, 5'-GAAANN(N)GGAA-3'), melyben az ETS és IRF kötőhelyek fordított sorrendben helyezkednek el 2 vagy 3 bázis távközzel. A T-sejtekben (ahol alacsony a PU.1 szint) az IRF4 nem EICE-hez kötődik, hanem ún. AP-1-IRF kompozit elemhez (AP-1-IRF composite element, AICE), mely IRF4 és AP-1 TF-ok közös kötőhelyeként szolgál. Az AICE-t eredetileg TGAnTCA/GAAA szekvenciaként írták le.

1.5.3. IRF3, IRF5 és IRF9 és az általuk szabályozott transzkripció programok

1.5.3.1. IRF3 és az antivirális citokinek

A TLR3 és a TLR4 jelátviteli útvonalak az IRF3 aktiválásához vezetnek. A TLR-en kívül az RLR szignalizáció is aktiválja az IRF3-t. Az aktiválását két rokon protein-kináz, az I κ B kináz (IKK ϵ) és a TANK-kötő kináz 1 (TBK1) szabályozza, amelyek a Ser396 oldalláncon foszforilálják az IRF3 fehérjét, és így elősegítik a fehérje dimerizációját és a nukleáris transzlokációját. Az IRF3 homodimert, vagy más IRF-ekkel heterodimert (pl. IRF3/IRF7) alkot.

A foszforilált IRF3 a sejtmagba transzlokálódik, ahol többek között antivirális citokineket kódoló géneket aktivál.

1.5.3.2. Az IRF5 és az gyulladásos program

Az IRF5 aktiválását a MYD88-on keresztül történő TLR szignalizáció szabályozza a Ser158 és Ser309 oldalláncok foszforilációján keresztül TBK1 által. Az IRF5 fehérje az inflammatorikus válaszok kialakulásában játszik fontos szerepet.

A gyulladásos citokineket kódoló gének képezik a TLR-indukált gyulladásos programok alapját a makrofágokban és a dendritikus sejtekben. Ezek többféle funkcióval bírnak, a szisztémás gyulladás mediációjától kezdve a T-sejt válaszok kialakításáig lokálisan. A gyulladásos citokineket kódoló gének számos más, különböző funkciójú génekkel együtt aktiválódnak, szabályozódnak. Ezeknek a géneknek az aktiválását, amelyet egy adott gyulladásos stimulus aktivál „gyulladásos programnak” nevezik.

1.5.3.3. Az IRF9 és az antivirális interferon-stimulált gének (ISG-k)

Az I-IFN-ok egy heterodimer receptor komplexhez kötődnek, amely az IFN- α receptor 1 (IFNAR1) és IFNAR2 alegységekből áll. Az I-IFN jelátvitel a STAT1 és a STAT2 fehérjék foszforilációjához, heterodimerizációhoz és az IRF9 fehérjével történő kölcsönhatáshoz vezet, ami az ISGF3 komplexet képezi. Az ISGF3 a sejtmagban ISRE elemeket köti, és ezáltal az ISG-k transzkripcióját eredményezi. Az ismertetett útvonalon kívül az I-IFN-ek indukálják a nem kanonikus JAK/STAT jelátvitelt is. Az ISGF3 mellett az IRF9-fehérjét és csak egyféle STAT1 vagy STAT2 fehérjét tartalmazó komplexek szintén szabályozzák a génexpressziót. Az ISGF3 kötődése az ISRE-hez kofaktorok toborzásához és végső soron a szabályozott ISG-k promoterén az RNS polimeráz II fokozott aktiválásához vezet.

Az ISG-k által kódolt fehérjék sokrétű szerepet játszanak. Számos effektor ISG fontos szerepet játszik a vírusok, baktériumok vagy paraziták által okozott fertőzések leküzdésében, ugyanis képesek közvetlenül gátolni a patogének életciklusát.

2. Célkitűzés

A disszertációban ismertetett kísérletek két összefüggő témakörhöz tartoznak, melyek két tudományos közleményben kerültek publikálásra. Az első témakörben azt vizsgáltuk, hogy hogyan valósul meg, hogy a különböző IRF-ek eltérő felismerő-helyekre kötődnek és eltérő transzkripciós programokat szabályoznak. A második témakör alapkérdése pedig az volt, hogy milyen (epi)genetikai jellemzői vannak az effektor ISG-knek, amik lehetővé teszik, hogy az I-IFN-ok számos sejtben, nagymértékben indukálódnak. A kérdések megválaszolásához olyan globális módszereket alkalmaztunk, mint a ChIP-seq, ATAC-seq, és az RNA-seq, valamint bioinformatikai módszereket használtunk az adatok kielemezéséhez.

2.1. IRF3-, IRF5- és IRF9-specifikus kötőhelyek azonosítása és jellemzése

- Az IRF3, IRF5 és IRF9 fehérjék kötőhelyeinek feltérképezése a teljes genomon ChIP-seq felhasználásával.
- Azon régiók azonosítása, ahol jelentős különbség van az IRF3, IRF5 és IRF9 kötésében (IRF3-, IRF5- és IRF9-specifikus kötőhelyek).
- A IRF3, IRF5 és IRF9-specifikus kötőhelyek jellemzése a következő tulajdonságok alapján: ISRE motívum, ISRE 5' és 3' túlnyúló bázisok, a társfaktorok kötése és a kromatin nyitottsága.

A IRF3-, IRF5- és IRF9-specifikus kötőhelyek és transzkripciós programok kapcsolatának vizsgálata.

2.2. Effektor ISG-ek szabályozó régióinak azonosítása és jellemzése

- I-IFN-indukált effektor ISG-k és lehetséges szabályozó régiók azonosítása RNA-seq és IRF9 ChIP-seq segítségével.
- Az effektor ISG gének szabályozó régióinak DNS kötő motívumainak azonosítása.
- A kromatin nyitottsága az effektor ISG gének szabályozó régióiban DC-ben és egyéb immunsejtekben.

3. Anyagok és módszerek

3.1. Sejttenyésztés, ligandkezelés

A CD8⁺ dendritikus sejtvonal (Mutu1940) Prof. Hans Acha-Orbea laboratóriumából származik (Unil, Lausanne, Svájc.). A sejteket 10% hőinaktivált marha szérum, 100 U/ml penicillin és 100 µg/ml sztreptomycin antibiotikum és 50 µM β-merkaptóetanol tartalmú tápoldatban 37°C-on, 5% CO₂ jelenlétében tenyésztettük.

Megfelelő sejtszámból (10-20 millió sejt/minta) kiindulva a sejteket különböző TLR ligandokkal és IFNβ liganddal kezeltük. TLR9 aktiválás esetében 1mM CpG liganddal kezeltük a sejteket, mely egy szintetikus CPG-oligonukleotid. TLR3 aktiválás esetében 5µg/ml Poly (I:C) liganddal kezeltük a sejteket, mely egy szintetikus RNS analóg. Az IFNα/β (IFNAR) receptor aktiválása esetében 100 U/ml IFNβ liganddal kezeltük a sejteket. A ChIP-seq és ChIP-qPCR kísérletekhez a CD8⁺ sejteket 90 percig kezeltük az adott ligandokkal. A ChIP-qPCR kísérletek esetében kombinált kezeléseket is alkalmaztunk. Génexpressziós qPCR mérések esetében a kezelések időtartama 1.5, 3, 6 és 12 óra volt.

3.2. ChIP-seq és ChIP-qPCR (kromatin immunprecipitáció szekvenálással kapcsolva és kromatin immunprecipitáció valós idejű kvantitatív PCR-rel kapcsolva)

A kettős keresztkötés előtt a sejtekről leszívtuk a tápoldatot, majd 2mM-os koncentrációban hígítva a DSG-t tartalmazó PBS-ben 40 percig inkubáltuk a sejteket. A második keresztkötés előtt leszívtuk a DSG oldatot a sejtekről és 1%-os koncentrációban hígítva formaldehidet adtunk a sejtekhez, majd 10 percig inkubáltuk. Ezután a keresztkötést 125 mM végső koncentrációban glicint tartalmazó PBS oldattal leállítottuk. A sejteket PBS-el mostuk majd felkapartuk. Centrifugálás után a felülúszót leszívtuk és a keresztkötött sejtpelletet folyékony nitrogénben fagyasztottuk majd -70°C-on tároltuk az immunprecipitációig.

A keresztkötött és fagyasztott sejteken sejtlízist és sejtmaglízist végeztünk proteáz inhibítort tartalmazó ChIP lízis pufferrel. Ezután szonikálással 100-2000 bp nagyságú szakaszokra fragmentáltuk a DNS-t. A következő antitesteket használtuk: IRF3 (D83B9, Cell Signaling), IRF5 (ab21689, Abcam), IRF9 (AF5629, R&D Systems), STAT2 (07-140, Merck Millipore), STAT1 (sc-346, Santa Cruz Biotechnology) és RNAPII-pS5 (ab5131, Abcam). Ennél a lépésnél 50 µl kromatint elteszünk -20°C-ra háromszoros térfogatú abszolút etanolban, később ezeket a mintákat használjuk a ChIP-qPCR kísérletekhez ún. „input” azaz kiindulási mintaként.

Az immunprecipitációt követően négy mosási lépésben mágneses tisztítással azok a DNS fragmentek maradtak, amelyekhez az általunk vizsgált fehérje kötődik, illetve az arra specifikus

antitest kötődött. Ezután a DNS fragmenteket és az input mintákat RNázsal és Proteináz K-val emésztettük, ezzel eltávolítva róluk az RNS-eket és fehérjéket. Majd a DNS-t a NucleoSpin Gel and PCR Clean-up kit segítségével tisztítottuk a gyártó utasításai alapján. A ChIP-DNS minták koncentrációját Qubit fluorométer segítségével határoztuk meg.

A ChIP-seq esetében 1-10 ng ChIP-DNS mintából kiindulva készítettünk indexelt cDNS könyvtárakat a TruSeq ChIP Library Preparation Kit segítségével a gyártó utasításai alapján. A könyvtárakat Illumina NextSeq 500 vagy HiSeq 2500 rendszeren szekvenáltuk a DE-ÁOK-BMBI-Genomi Medicina és Bioinformatikai Szolgáltató Laboratóriumban. A ChIP-qPCR mérésekhez a kilenc reprezentatív gén promóter régiójára terveztünk primereket, és a feldúsulást valós idejű kvantitatív PCR (polymerase chain reaction-polimeráz láncreakció) mérésekkel detektáltuk nyolc mintában (UT, PIC, CPG, IFNB, PIC+CPG, PIC+IFNB, CPG+IFNB, PIC+CPG+IFNB) triplikában az input mintákra normalizálva. A méréshez SYBR Green festéket használtunk.

A futtatáshoz a következő hőmérsékleti protokollt alkalmaztuk: 1 perc 94°C, 40 ciklus 12 másodperc 94°C, 30 másodperc 60°.

3.3.ChIP-seq analízis

A nyers ChIP-seq adatok elsődleges elemzését az Debreceni Egyetem, ÁOK, Biokémiai és Molekuláris Biológiai Intézetben kifejlesztett ChIP-seq „pipeline” parancssorai segítségével végeztük. A szekvenciaolvasások (read) az egér referencia genomra (mm10) lettek illesztve a Burrows- Wheeler Alignment Tool (BWA) szoftverrel. A feldúsult genomi régiók azonosításához a Model-based Analysis of ChIP-Seq 2 (MACS2) szoftvert használtuk a következő paraméterekkel: q-függvény küszöbértéke $q = 0.001$ és a `-call-summits` kapcsoló használata, amely felbontja a nagyobb peak régiókat kisebb darabokra. A műtermékeket az ENCODE (Encyclopedia of DNA Elements) genomi régiók feketelistája segítségével zártuk ki. A konszenzus kötőhelyek azok a régiók, amelyek mindkét párhuzamos mintában detektálva lettek. Az IRF3, IRF5 és IRF9 cisztrómk meghatározásához az IRF3, IRF5 és IRF9 konszenzus kötőhelyeit összevontuk. A meghatározott kötőhelyek lefedettségét, az RPKM (Reads Per Kilobase per Million mapped reads) értéket a `bamtools` és a `bedtools` (`coverageBed`) szoftverekkel, és `awk` szkriptekkel határoztuk meg. A klaszterezéshez a lefedettségi értékek (`occupancy values`) a medián értékekre lettek normalizálva minden IRF esetében. Az adatok megjelenítéséhez és a reprezentatív ábrák készítéséhez az Integrative Genomics Viewer (IGV) genomböngészőt használtuk. A szekvenciaolvasások lefedettségét ábrázoló fájlokat az ún. `bedgraph`-okat átkonvertáltuk `tdf` fájlokká `igvtools` segítségével, a `'toTDF'` paranccsal. A normalizált szekvenciaolvasások hőtérképen és hisztogramon (read distribution (RD) heat map/histogram) való megjelenítését HOMER (Hypergeometric Optimization of Motif EnRichment) program segítségével készítettük

el, majd R vagy Java TreeView programmal vizualizáltuk. A ChIP-seq adatok nyers fastq fájljai az alábbi GEO azonosítók alatt találhatóak meg: GSE172376, GSE125340.

3.4. A motívum keresés, súlyozott motívum mátrix (PWM) és DNS szekvenciák azonosítása

A *de novo* motívum keresését a HOMER 'findMotifsGenome.pl' programjával végeztük. A p-értéket úgy határoztuk meg, hogy a vizsgált kötőhelyeken feldúsult motívumokat összehasonlítottuk a HOMER által véletlenszerűen generált háttér szekvenciákon feldúsult motívumokkal.

A súlyozott motívum mátrixokat (Position weight matrices, PWMs) a HOMER adatbázisból importáltuk vagy magunk határoztuk meg. Az ISRE, EICE, AICE1, NFKB, TRE, CRE és GAS motívumok PWM-i a HOMER adatbázisból származnak. Az IECS, AICE2 és a TISRE motívumok PWM-jei nem voltak elérhetőek az adatbázisban, így azokat publikus ChIP-seq adatok *de novo* újrakezelésével határoztuk meg. A PWM-ek visszatérképezését a genomra a HOMER 'scanMotifGenomeWide.pl' programmal végeztük.

Konkrét DNS szekvenciák azonosításához a klaszterekben 6 bázis hosszú szekvenciákat (6-mer) és ISRE variánsokat generáltunk. Az egyes kötő régiók DNS szekvenciáját az egér genomból (GRCm38/mm10) azonosítottuk 'fastaFromBed bedtools' program alkalmazásával. Az összes kötő genomi régióban 'grep' paranccsal kerestük az összes általunk meghatározott 6-mer és ISRE variáns szekvenciát és azok reverz komplementerét is. A 5'-GAANNAAA-3' szekvenciát R programban 'Biostrings' csomag segítségével azonosítottuk genom szinten.

3.5. ATAC-seq kísérletek és elemzésük

Az ATAC-seq kísérleteket biológiai duplika alkalmazásával hajtottuk végre. A sejteket hideg PBS oldatban felkapartuk, majd hígítottuk, hogy elérjük a 10 000-15 000/ml sejtszámot. Sejtlízis és sejtmaglízist végeztünk ATAC-lízis pufferrel. A lízált sejtmagon történt a tagmentálás Nextera DNA Library Preparation Kit (Illumina) segítségével. A tagmentálás után a DNS-t MinElute PCR Purification Kittel (QIAGEN) izoláltuk. A tagmentált DNS szakaszokat a Kapa HiFi Hot Start Kit (Kapa Biosystems) segítségével felsokszoroztuk 9 PCR ciklus során. A felsokszorozott könyvtárakat ismét a MinElute PCR Purification Kit (QIAGEN) segítségével izoláltuk. A könyvtárak fragmenteloszlását Agilent Bioanalyzer készülékkel határoztuk meg és Illumina HiSeq 2500 rendszeren szekvenáltuk.

Az ATAC-seq adatok elsődleges analízisét a ChIP-seq elemzés során is használt pipeline parancsai segítségével végeztük. A saját adataink mellett GEO adatbázisból további ATAC-seq nyers adatokat töltöttünk le és elemeztünk. A szekvencialeolvasások az egér referencia genomra (mm10) lettek illesztve a BWA szoftverrel. A feldúsult ATAC-seq régiók azonosításához a

MACS2 szoftvert használtuk ($q\text{-érték} \leq 0.001$). A műtermékeket az ENCODE genomi régiók feketelistája segítségével zártuk ki. Az eredmények megjelenítéséhez a bedgraph-okat 'UCSCfile.pl' paranccsal hoztuk létre, majd átkonvertáltuk tdf fájlokra az igvtools 'toTDF' parancsával. Az adatok megjelenítéséhez az IGV programba importáltuk ezeket. Az ATAC-pozitív régiók kiszámításához a 'bedtools intersect' parancsot használtuk. Az ATAC-seq adatok GEO azonosítója: GSE125340.

3.6. Háromszög diagramok és Random Forest elemzés

A háromszög diagramok („ternary plot”) generálásához az R 'ggtern' nevű könyvtárcsomagot használtuk. A háromszög alakú diagram egy adott pontja három bemeneti változó (ChIP-seq szignál vagy mRNS szint) alapján lett számoltatva. A Random Forest gépi tanuló algoritmust Pythonban alkalmaztuk Random Forest Classifierrel a Scikit-learn könyvtárcsomagból. ChIP-seq, DNS motívum vagy ATAC-seq értékeknek a különböző halmazát használtuk bemeneti változóként és a klaszterek neveit használtuk osztálycímkékként.

3.7. Génlisták

A különböző GO (The Gene Ontology) kategóriákba sorolható géneket az MGI adatbázisból (Mouse Genome Informatics database, <http://www.informatics.jax.org>) töltöttük le. Az antivirális citokinekhez tartoznak a vírusellenes immunválasz (GO:0051607) és a citokin aktivitás (GO:0005125) génjeit tartalmazó GO kategóriák közös elemei. A gyulladásos program génjeihez tartoznak az inflammatorikus válasz (GO:0006954) génjei, amelyeket a citokin aktivitás (GO:0005125) vagy a DNS-kötő transzkripciós faktor aktivitás (GO:0003700) kategóriákban is megtalálhatunk. Az antivirális ISG-k az irodalmi adatok alapján lettek kiválasztva. Két listánál (effektor ISG-k és gyulladásal kapcsolatos citokinek) figyelembe vettük az RNA-seq eredményeket is. Az effektor ISG-k az antivirális ISG-k közül azokat tartalmazza, amelyeket IFN β szignifikánsan indukált DC-ekben, és irodalmi adatok alapján effektor funkciót rendeltek. A gyulladásal kapcsolatos citokinek (inflammation-related cytokines, IR-cytokines) azokat a géneket tartalmazzák, melyeket a CpG szingifikánsan indukált DC-ekben és megtalálhatóak voltak mind az inflammatorikus válasz (GO:0006954) mind a citokin aktivitás (GO:0005125) listákban.

3.8. RNS izolálás és RT-qPCR

Az RNS-t Trizolate reagens (UD-GenoMed) alkalmazásával izoláltuk. A lízis után a szerves-vizes fázis optimális elválasztása kloroform hozzáadásával történt és az RNS-t a vizes fázisból alkoholos kicsapással nyertük vissza. A kicsapás után az RNS csapadékot 10 percig centrifugáltuk 4°C-on 16000 g-vel. A felülúszót eltávolítottuk, és közben figyeltünk arra, hogy a

pellethez ne érjünk hozzá. A pelletet 70%-os etanollal mostuk. Az etanol eltávolítása után az RNS-t szobahőmérsékleten beszárítottuk vákuumkoncentrátorban. Az RNS-t nukleáz mentes vízben (NFW, Affymetrix) oldottuk be. Az RNS mintákat a mérések kezdetéig -20°C-on tároltuk.

A reverz transzkripció során a High Capacity cDNA Reverse Transcription Kit-et (Thermo Fischer Scientific) használtuk a gyártó utasításai alapján.

A PCR reakcióhoz a következő hőmérsékleti protokollt alkalmaztuk: 10 perc 25°C, 120 perc 42°C, 5 perc 72°C. A futtatást Applied Biosystems 2720 Thermal Cycler PCR készüléken végeztük. A keletkezett cDNS-t jégén, illetve -20°C-on tároltuk a QPCR mérések kezdetéig.

A qPCR mérések során a kilenc reprezentatív gén expresszióját határoztuk meg valós idejű kvantitatív PCR mérésekkel UT, PIC, CPG, IFNB mintákban biológiai és technikai triplikában. A méréshez SYBR Green festéket és LightCycler 480 készüléket (Roche) alkalmaztunk.

A futtatáshoz a következő protokollt alkalmaztuk: 1 perc 94°C, 12 másodperc 94°C, 30 másodperc 60°C. Ciklusok száma: 40 (második és harmadik lépés). Az adatok normalizálására az Rplp0 háztartási gént használtuk.

3.9. RNS szekvenálás

Az RNS-szekvenáláshoz (RNS-seq) az RNS-t Trizolate reagenssel (UD-GenoMed) izoláltuk. Az RNS minőségét qPCR technikával és Agilent BioAnalyzer készüléken ellenőriztük Eukaryotic Total RNA Nano Kit segítségével. A könyvtár készítést, szekvenálást és adatelemzést a DE-ÁOK-BMBI-Genomi Medicina és Bioinformatikai Szolgáltató Laboratórium munkatársai végezték. Az RNA-Seq könyvtárakat a teljes RNS-ből Ultra II RNA Sample Prep kit (New England BioLabs) segítségével készítették. A szekvenálást Illumina NextSeq500 készüléken végezték 75 ciklusos szekvenálással. A nyers szekvenciákat az mm10 genomra illesztették HISAT2 segítségével. A későbbi elemzésekhez a BAM fájlokat a Strand NGS programba importálták (Strand Life Sciences Pvt. Ltd., Bangalore, India). A szabályozott gének azonosítására moderált t-teszteket alkalmaztak Benjamini-Hochberg korrekció ($p < 0,05$ és $(FC > 2,0)$ alkalmazásával.

4. Eredmények

4.1. Az IRF3, IRF5 és IRF9 fehérjék kötő régióinak feltérképezése

Az általunk használt CD8⁺ immortalizált sejtvonalban legmagasabb szinten az IRF3, IRF5, IRF8 és IRF9 fejeződik ki az IRF transzkripciós faktorok közül.

Az IRF3, IRF5 és IRF9 fehérjék kötő régióinak feltérképezéséhez ChIP-seq kísérleteket végeztünk. A CD8⁺ DC sejteket 90 percig kezeltük CpG-vel, pIC-vel illetve IFN β -al. Ezek az agonisták hatékonyan aktiválják azokat a receptorokat és jelátviteli útvonalakat, amelyek a vizsgálandó IRF-el vannak kapcsolatban. Ezek az útvonalak: pIC $\rightarrow$ TLR3 $\rightarrow$ IRF3; CpG $\rightarrow$ TLR9 $\rightarrow$ IRF5 és IFN β $\rightarrow$ IFNAR1/2 $\rightarrow$ IRF9.

Azokkal a kötőhelyekkel dolgoztunk tovább, melyek egy adott minta mindkét replikájában azonosíthatóak voltak. Ezzel a módszerrel 25 203 (IRF3), 17 435 (IRF5) és 25 040 (IRF9) kötőrégiót azonosítottunk. Abban az esetben, ha az IRF3, IRF5 és IRF9 lefedettségi értéke egy adott genomi régióra hasonló értéket ($0,5 < \text{lefedettségi hányados} < 2$) ért el, IRF3, IRF5 és IRF9 fehérjék közös kötő genomi régióról beszélünk és közös klaszternek nevezzük. További három klasztert definiáltunk, amelyekben az egyik IRF fehérje kötődése az adott genomi régióhoz gyengébb volt, mint a másik két IRF-fehérje kötődése.

4.2. DNS-motívumok előfordulása az IRF klaszterekben

Három stratégia alkalmazásával elemeztük a DNS-motívumokat a klaszterekben: (1) „*de novo*” motívum kereséssel, (2) súlyozott motívum mátrixok (PWM) feldúsulásainak elemzésével és (3) specifikus DNS-szekvenciák keresésével. Mindhárom stratégiának megvannak a maga előnyei és korlátai, ezért ezeknek a megközelítéseknek a kombinálása átfogóbb képet eredményez.

Az első megközelítést (*de novo* motívum keresés) felhasználva azt találtuk, hogy az IRF3- és IRF9-domináns klaszterekben feldúsult motívumok olyan DNS-motívumok voltak, amelyek hasonlóak az ISRE-hez, vagy ISRE kompozit elemekhez, például az ETS-IRF kompozit elemhez (EICE). Összhangban a korábbi IRF5 ChIP-seq vizsgálatokkal, *de novo* motívum kereséssel nem sikerült azonosítani ISRE motívumot az IRF5-domináns klaszterben, azonban az ISRE félhely és gyenge IRF-Ets kompozit elem (IECS) feldúsulást mutatott. Ezek mellett az IRF-kötő motívumok mellett számos más motívumot is azonosítottunk, köztük az NF-kB-kötő motívumot (kB vagy NFkB), a TPA válasz elemet (TRE), a cAMP válasz elemet (CRE) és az IFN-gamma aktivációs szekvenciát (GAS). Ezek a motívumok, olyan transzkripciós faktorok kötő régiói, amelyek útvonal specifikusan aktiválódnak az IRF-fehérjékkel. TLR aktiválás során az IRF3 vagy IRF5 együtt aktiválódik az AP-1 család tagjaival, melyek a TRE vagy CRE motívumokhoz kötődnek, és az

NFkB család tagjaival, melyek az NFkB motívumhoz kötődnek. Az I. típusú IFN-ok indukálják az ISGF3 komplex kialakulását STAT1 homodimerekkel, melyek a GAS motívumhoz kötődnek.

A második megközelítésünkben motívum mátrixok (position weight matrix, PWM) feldúsulását vizsgáltuk a klaszterekben. A „de novo” motívum keresésekben a hasonló motívumok összevonódnak, továbbá adott motívumok közvetlen összehasonlítása a klaszterek között nem lehetséges. Ebben segít a PWM-ek feldúsulásának vizsgálata. A HOMER adatbázisból és újraelemzett ChIP-seq adatokból gyűjtöttünk IRF faktorok kötőhelyeiként azonosított elemek PWM-jeit. Az ISRE, EICE, IECS és az együttesen aktivált transzkripciós faktor motívumok mellett három hosszabb motívumot is vizsgáltunk az AICE1, AICE2 és TISRE motívumokat. Az egyes motívumok küszöbértékét (cut-off value) szisztematikusan határoztuk meg véletlenszerűen kiválasztott régiók halmaza alapján, és a PWM-ek feldúsulását ezekhez viszonyítva határoztuk meg.

Eredményeink alapján három alapvető megfigyelést tettük. Első megfigyelésünk az volt, hogy az IRF3 és IRF9 domináns klaszterekben dúsult fel a legjobban az ISRE motívum. Ezzel összhangban az IRF3-IRF9 közös klaszterben is magas volt az ISRE frekvenciája. Második megfigyelésünk, hogy a TISRE motívum különösen gyakori volt az IRF3-domináns, az IRF9-domináns és IRF3-IRF9 közös klaszterekben. Harmadik megfigyelésünk, hogy az EICE motívum az ISRE motívumokhoz hasonló dúsulást mutatott a vizsgált klaszterekben. Ezek az eredmények nem jelentik azt, hogy az IRF3-dimerek és az ISGF3 komplex magas frekvenciával kötik az EICE motívumot. Sokkal valószínűbb, hogy az IRF3-dimerek és az ISGF3 komplex előnyben részesíti azokat a hozzáférhető genomi régiókat, amelyek a PU.1 és az IRF8 által kötöttek és ezáltal nyitottak. Mindkét transzkripciós faktor LDTF, azaz felelősek az adott dendritikus sejt alpopuláció sejtípus specifikus kialakulásáért, és együttesen kötik az EICE és IECS motívumot.

A második megközelítésünk (PWM feldúsulások) további elemzéseiben azt vizsgáltuk, hogy milyen előfordulása van az NFkB, TRE, CRE és GAS motívumoknak. Ezek a motívumok olyan transzkripciós faktorok kötőhelyei, melyek együtt aktiválód(hat)nak bizonyos IRF-ekkel. A *de novo* motívum keresés eredményei és PWM feldúsulásának elemzései általában összhangban voltak. Az egyetlen eltérést az IRF5 domináns klaszter esetében figyeltük meg, ahol a TRE és CRE motívumok a *de novo* elemzés során feldúsulást mutattak, azonban a súlyozott motívum mátrix analízis során nem. Az NFkB motívum az IRF3-domináns, az IRF5-domináns, valamint a közös IRF3-IRF5 klaszterben, míg a GAS motívum az IRF9 domináns klaszterben dúsult fel jelentősen. A TRE motívum az IRF3-domináns, a IRF3-IRF5 és IRF3-IRF9 közös klaszterekben, míg a CRE az IRF3-domináns, és IRF3-IRF5 közös klaszterekben fordult elő nagyobb gyakorisággal, mint a többi klaszterben.

A harmadik megközelítésünk során specifikus DNS-szekvenciák előfordulását vizsgáltuk IRF klaszterekben. Először az ISRE-félfhelyeket elemeztük. A 4 bázis hosszú ISRE-félfhely (5'-GAAA-3') és változatai helyett azért választottunk 6 bázis hosszú szekvenciákat, mert az IRF-fehérjék kötődését az ISRE félfhelyek 5' és 3' túlnyúló végükön található bázisok is befolyásolják. Készítettünk egy 6-mer listát (n=535), amely tartalmazta a kanonikus 5'-GAAA-3' ISRE félfhely szekvenciát és két extra bázist három lehetséges pozícióban (5'-NNGAAA-3', 5'-NGAAAN-3' vagy 5'-GAAANN-3'), egy bázis eltérést engedélyezve. Meghatároztuk ezeknek a 6-mer szekvenciáknak a gyakoriságát a klaszterekben. Sok olyan szekvenciát azonosítottunk (n=84), amelyek legalább 1,5-szer gyakrabban fordultak elő az IRF3-, IRF5- és IRF9-domináns klaszterekben, mint a másik két klaszterben. Legtöbb (n=68) ezek közül a legnagyobb feldúsulást az IRF5-domináns klaszterben mutatta. Több 6-mer szekvencia feldúsulása is egybeesett a korábbi „*in vitro*” vizsgálatok eredményeivel. A fehérje-kötő microarray (Protein binding microarray, PBM) vizsgálatok azt mutatták, hogy a 5'-GGAAAC-3'szekvencia nagyobb affinitással kötődik az IRF3 fehérjéhez, mint az IRF5, IRF6 vagy IRF9 fehérjékhez. Ezzel a megfigyeléssel összhangban azt találtuk, hogy a GGAAAC szekvencia feldúsulást mutatott az IRF3-domináns klaszterben. PBM kísérletek bizonyították, hogy az IRF5 fehérje sokkal jobban tolerálja az ISRE-félfhely 3. és 4. pozícióban lévő bázisok eltérését a kanonikus szekvenciáktól, mint az IRF7 vagy más vizsgált IRF fehérjék. Több olyan 6-mer szekvenciát detektáltunk az IRF5 domináns klaszterben melyek bázisai ezekben a pozíciókban eltért a kanonikus szekvenciától (például CGACAC és CGAGAC).

További PBM kísérletek alapján a TGAAAC szekvenciát nagyobb affinitással kötötte az IRF9 fehérje, mint az IRF3, IRF5 vagy IRF6 fehérjék. Elemzésünk során azt találtuk, hogy ez a szekvencia és még további néhány 6-mer szekvencia mutatott feldúsulást az IRF9 domináns klaszterben.

A harmadik megközelítésünk további elemzéseiben olyan speciális ISRE variánsok (n=37) jelenlétét vizsgáltuk a klaszterekben, amelyeket korábbi tanulmányok alapján vagy a saját ISRE félfhely 6-mer szekvenciák alapján választottunk ki. 20 olyan ISRE variánst azonosítottunk, amelyek legalább 1,5-szer gyakrabban fordultak elő az IRF3-domináns klaszterekben, mint a másik két klaszterben. Az IRF9-domináns klaszterekben 4 ilyen ISRE variánst azonosítottunk, viszont nem találtunk olyan ISRE-variánsokat, amelyek gyakrabban fordultak volna elő az IRF5-domináns klaszterben, mint a másik két klaszterben.

Összhangban a 6-bp ISRE félfhely eredményeinkkel, azt találtuk, hogy a GGAAANNNGAAA szekvencia gyakrabban fordul elő az IRF3-domináns klaszterben a többi klaszterhez képest. Azok az ISRE variánsok, melyekben két extra AA bázist tartalmazott az 5' végén szintén nagyobb feldúsulást mutattak az IRF3-domináns klaszterben. Az EMSA és a fehérje

krisztallográfiás kísérletek során azt találták, hogy az ISGF3 komplex által kötött szekvenciák hosszabbak, mint a 10 bp hosszú konszenzus ISRE motívum, az IRF9 és a STAT fehérjék kötési preferenciája miatt. Elemzéseink során azt találtuk, hogy az 5'-GAAANNGAAACT-3' feldúsulást mutatott az IRF9 domináns klaszterben.

4.3. A kromatin hozzáférhetősége az IRF klaszterekben

Ahogy a Bevezetésben áttekintettük, a SRTF-ek kötődésének nagy része a korábban (LDTF-ek által) „kijelölt”, nyitott, hozzáférhető enhanszer régiókban történik. A kromatin nyitottság és SRTF kötés kölcsönösen hat egymásra: a nyitottság jelentős mértékben meghatározza a bekötődést, viszont a kötődés tovább „nyitja” a régiót. Az LPS által stimulált makrofágokban azt találták, hogy az IRF3 fehérje elősegítheti a nukleoszómák átrendeződését és/vagy a korábban zárt/kevésbé hozzáférhető kromatin nyitását. Az ISGF3 komplex által történő kromatin nyitást szintén tanulmányozták. Tudomásunk szerint az IRF5 kromatin átrendező képességét nem vizsgálták átfogó tanulmányokban. A kromatin hozzáférhetősége fontos meghatározó lehet az IRF-specifikus enhanszerek kiválasztásában. Ezért megvizsgáltuk a hozzáférhető kromatin régiókat ATAC-seq kísérletekkel stimulálatlan (UT) DC-kben.

Kiszámoltuk az IRF-ek összes kötőhelye és a klaszterekben a hozzáférhető kromatin régiók százalékos arányát. Az IRF-ek összes kötőhelye esetében azt találtuk, hogy a különbség kisebb volt (~13%) a legalacsonyabb és a legmagasabb frekvencia között (IRF3: 56,8% és IRF5: 69,8%). Az IRF klaszterek esetében sokkal nagyobb különbséget (~50%) detektáltunk a legalacsonyabb és a legmagasabb frekvencia között (IRF3-domináns: 23,8% vs. Közös: 73,8%). A specifikus klaszterek közül az IRF5-domináns klaszterekben találtuk a legtöbb hozzáférhető kromatin régiót (62,4%). Az IRF3-kötő régiók, a hozzáférhető kromatin régiók és a kanonikus ISRE (5'-GAAANNGAAA-3') szekvencia átfedését szintén meghatároztuk.

A kanonikus ISRE szekvenciáknak csak kis részét (<1%) foglalta el az IRF3 *in vivo* aktiválás után. Ezzel szemben a stimuláció előtt hozzáférhető régiók ~35% -át foglalta el az IRF3 fehérje. Egy adott gén többféle IRF kötőhellyel lehet kapcsolatban. A már megjelölt (nyitott) és a kötő régiók sokféleségére reprezentatív példa a *Ccl5* gén szabályozó elemei. Összességében megállapítottuk, hogy az ATAC-negatív régiók feldúsultak az IRF3- és IRF9-domináns régiókban, ami arra utal, hogy a „zárt” nehezen hozzáférhető kromatin hozzájárul az IRF3- és IRF9-domináns (de nem a IRF5-domináns) kötőhelyek specificitásához, más IRF-ek kötődésének kizárásával. Megfigyeléseink összhangban állnak más transzkripció faktorokkal kapcsolatos eredményekkel, amelyek arra utalnak, hogy a kromatin állapota kulcsszerepet játszik a szelektív transzkripció válaszok kialakításában.

4.4. Az IRF dominancia prediktálása gépi tanulási módszerrel

A gépi tanulási módszerek egyik alkalmazása „osztálycímkek” predikciója bemeneti adatok alapján. A Random Forest segítségével megvizsgáltuk, hogy az IRF dominanciát milyen pontosan lehet prediktálni a kötőhelyek jellemzőinek felhasználásával. Páronkénti összehasonlítást végeztünk úgy, hogy a kötőhelyek klaszter-címkéje (IRF3-, IRF5- vagy IRF9-domináns) szolgált „osztálycímkeként”. Bemeneti adatként tíz motívum PWM pontszámait, az ATAC-seq szignálértékeket, ISRE variánsok és 6-mer szekvenciák előfordulási értékeit alkalmaztuk, vagy ezek kombinációját használtuk. A predikció pontosságát a helyes és az összes predikció aránya alapján adtuk meg, a predikció helyes volt, amikor az eredeti és a prediktált osztálycímkek megegyeztek. Jelenetős különbség volt, annak függvényében, hogy melyik jellemzőket használtuk a predikcióhoz.

Összességében az eredmények azt mutatták, hogy minden jellemző figyelembevételével az osztálycímkek nagy pontossággal (~80-85%) prediktálhatóak voltak, amikor az IRF5-domináns klasztert összehasonlítottuk a másik két domináns klaszterrel. Az osztálycímke prediktálása kevésbé volt pontos (~65%) az IRF3-domináns és az IRF9-domináns klaszterek összehasonlításakor. Ez azt jelenti, hogy az IRF5-domináns klaszter jelentősen különbözik a másik két klasztertől, míg a kötőhely jellemzők alapján az IRF3 és IRF9 domináns klaszterek kevésbé váltak el egymástól.

4.5. Az antivirális citokineket, gyulladásos citokineket és antivirális ISG-ket kódoló gének gyakran kapcsolódnak IRF3-, IRF5- és IRF9-domináns régiókkal

Az IRF klaszterek feldúsulását megvizsgáltuk három transzkripciós program génjeinek kibővített promóter régióiban. Az antivirális citokinek, a gyulladásos program génjei (citokinek és transzkripciós faktorok), és az antivirális ISG-k közül sokat az IRF3, az IRF5 és az IRF9 fehérjék célgénjeként azonosítottak. Három génlistát állítottunk össze a GO (The Gene Ontology) listák és a szakirodalom alapján.

Az antivirális citokinek (n=25), a gyulladásos program génjei (n=119) és az antivirális ISG-k (n=63) transzkripciós starthely körül 20 kb méretű régióit (TSS $\pm$ 10 kb) meghatároztuk, majd azonosítottuk ott az IRF kötőhelyeket. Ezután megvizsgáltuk az adott kötőhely típusát, és kiszámítottuk a relatív (klasztermérettel normalizált) feldúsulást. Az elemzés során azt találtuk, hogy az antivirális citokinek TSS $\pm$ 10 kb régióiban az IRF3 domináns klaszter (és kisebb mértékben IRF3-IRF5 közös klaszter) mutatott feldúsulást. A gyulladásos programban részt vevő gének TSS $\pm$ 10 kb régióiban az IRF5-domináns és az IRF3-IRF5 közös klaszterek dúsultak fel, míg az antivirális ISG-k TSS $\pm$ 10 kb régióiban, az IRF9 domináns és az IRF3-IRF9 közös

klaszterek fordultak elő leggyakrabban. Megvizsgáltuk az azonosított, fenti három génlistához kapcsolható 28 IRF3-, 62 IRF5- és 45 IRF9-domináns kötőhelyek jellemző DNS motívumait és kromatin nyitottságát. Az elemzés azt mutatta, hogy a teljes klaszter és azon részhalmazai a legtöbb esetben hasonló mintázatot mutattak. Különösen feltűnő volt az IRF5-domináns kötőhelyek magas NFκB motívum pontértéke és ATAC-seq szignálja, illetve alacsony ISRE motívum pontértéke.

Az IRF3, IRF5 és IRF9 kötődésének és génexpresszió kapcsolatának vizsgálatához reprezentatív géneket választottunk ki. Olyan antivirális citokineket, gyulladásos citokineket és antivirális ISG-eket azonosítottunk, amelyek az IRF3, IRF5 vagy IRF9 domináns csúcsokkal asszociáltak a promóter régiókban. RT-qPCR technika alkalmazásával megvizsgáltuk az mRNS szintjüket különböző ligandok jelenlétében, az idő függvényében (1,5-12 óra).

Megállapítottuk, hogy a kiválasztott antivirális citokineket, az inflammatórikus citokineket és az antivirális ISG-eket a leghatékonyabban rendre a pIC, a CpG és az IFN-β szabályozzák. Fontos kiemelni, hogy mivel a géneket számos enhanszer szabályozhatja, és a TLR jelátviteli útvonalak egyszerre több SRTF-t aktiválhatnak, az adott ligand hatását nem feltétlenül kizárólag csak ezek a régiók és vizsgált IRF fehérjék mediálják. Eredményeink azt mutatták, hogy a régiók, amelyek túlnyomórészt IRF3, IRF5 vagy IRF9 fehérjék által szabályozottak, fontos kötőhelyek az adott transzkripciós válaszokhoz. CHIP-qPCR módszer segítségével megvizsgáltuk, hogy a TLR és IFN útvonalak között kimutatható-e együttműködés vagy gátlás kombinált ligandkezelések esetén. Ezekben a kísérletekben a DC-eket egy agonistával, vagy a ligandok kombinációival stimuláltuk. Megállapítottuk, hogy az IRF3 kötődését az antivirális citokinek IRF3 domináns régióihoz az összes kombinált kezelés gátolja. Az IRF5 kötődése a gyulladásos gének IRF5-domináns régióihoz nagyobb feldúsulást mutatott, amikor a pIC + CpG kombinált kezelést alkalmaztuk. Mivel az IRF5 fehérjét a TLR3 és a TLR9 agonista is aktiválta, a megnövekedett IRF5-kötődésének a két útvonal közötti együttműködés lehet az oka. Végül, az IRF9 kötődése az antivirális ISG-k IRF9-domináns régióihoz hasonló volt, amikor az IFNβ-t önmagában vagy CpG-vel kombinált kezelést alkalmaztuk, ami arra utal, hogy a TLR9 útvonal nem képes gátolni az IRF9 aktivációját vagy kötődését.

Ezzel szemben a TLR3 útvonal gátolta az IRF9 aktivációját vagy kötődését az IRF9-domináns régiókba. Ezek az eredmények együttesen azt sugallják, hogy az IRF-ek és más SRTF-ek közötti verseny vagy együttműködés jelentősen befolyásolja az IRF3, IRF5 és IRF9 fehérjék kötődését.

4.6. Az effektor ISG-k lehetséges szabályozóelemeinek azonosítása

RNA-seq felhasználásával elkészítettük azoknak az ISG-knek a listáját ($n=1936$), amelyek 6 órás IFN β -kezelés után legalább kétszeres indukciót mutattak. Az ISG-k listájából kiválasztottuk azokat a géneket, amelyeket korábbi vizsgálatokban „antivirális effektorként” írtak le. Ily módon 37 effektor ISG-t azonosítottunk.

Az effektor ISG-k szabályozó elemeit az IRF9 kötőhelyeinek feltérképezésével azonosítottuk. Egy adott sejtben aktív (vagy aktiválható) génszabályozó elemek többféle módszerrel azonosíthatóak, de minden módszernek megvannak a korlátai. Az egyik lehetséges módszer az aktivált SRTF-ek feltérképezése. Ennek az lehet a hátránya (ha nem kombináljuk más módszerrel), hogy nincs információnk arról, hogy a bekötődés az enhanszer aktivitásával jár-e. Bizonyos útvonalak esetében probléma lehet továbbá, hogy egy receptor több SRTF-et aktivál (pl. TLR-ok esetében az NF- κ B, AP-1, IRF faktorok), amelyek kötőhelyei csak részben fednek át. Esetünkben ez utóbbi nem volt jelentős probléma, mivel az IRF9 alapvető komponense az IFN-indukált transzkripció programoknak.

A szabályozó elemek és a szabályozott gének összekapcsolása továbbra is jelentős technikai kihívás. Mivel az enhanszer-promoter kölcsönhatások többsége ~50 kb távolságon belül történik, és többnyire a kötőrégiók a legközelebbi gén expresszióját szabályozzák, lehetséges a szabályozott gének közelség alapján történő előrejelzése. Fontos megjegyezni, hogy az enhanszer-promoter párok összerendelése ebben az esetben csak becslés vagy feltételezés, és bizonyos esetekben téves hozzárendelésekhez vezethet. Mindazonáltal, ez a megközelítés kísérleteket nem igényel, egyszerű bioinformatikai módszerrel elvégezhető és az esetek jelentős részében az enhanszer-promoter párok összerendelése helyes, ezért általánosan alkalmazzák és ebben a tanulmányban mi is ezt használtuk.

Az elemzésünkhöz az IRF9 ChIP-seq adatainkat használtuk. További ChIP-seq kísérleteket is végeztünk STAT1, STAT2 és RNS-polimeráz II (Pol-II) ellenes antitestek felhasználásával. Összesen 27546 IRF9 kötőrégiót azonosítottunk IFN β -vel stimulált DC-kben (konszenzus set: replikák közös régiói). 9289 régiót azonosítottunk, amelyekben az összes ISGF3 alegység (IRF9, STAT1 és STAT2) kötődését kimutattuk az IFN β -vel stimulált DC-kben. Összesen 78 IRF9 kötőrégiót azonosítottunk, amelyekhez egy effektor ISG-t rendelt a HOMER algoritmus legközelebbi génként. A 78 kötőrégióból 34 található a transzkripció kezdőhelyétől 1 kb-on belül (TSS-proximális IRF9 kötőrégiók). A többi IRF9 kötőrégió ettől nagyobb távolságra volt, maximálisan 33 kb-ra. Meghatároztuk az ISGF3 alegységek kötését, és azt találtuk, hogy az IRF9 szignálintenzitásokkal korreláltak a STAT1 és STAT2 szignálintenzitások ezekben a régiókban.

Fontos megfigyelésünk volt, hogy az IRF9 szignálintenzitások magasabbak voltak a TSS-proximális régiókban, mint a TSS-disztális régiókban. Az effektor ISG-k többsége ($n = 30$) olyan

IRF9 kötőhellyel volt kapcsolatban, amely kevesebb, mint 1 kb-ra volt a TSS-től. Az effektor ISG-eket a kapcsolódó IRF9 csúcsok száma alapján osztályoztuk. Mint korábban említettük, 3 gén volt, amelyekhez nem társult IRF9 csúcs. A legtöbb effektor ISG (n = 25) egy vagy két IRF9 csúccsal, míg 9 effektor ISG három vagy több IRF9 csúcshoz kapcsolódott.

4.7. Az ISRE és TISRE az effektor ISG-k szabályozó régióiban

Megvizsgáltuk az ISG-khez tartozó IRF9 kötőrégiókban az ISRE (5'-GAAANNGAAA-3') szekvencia jelenlétét. Az ISRE szekvenciát az effektor ISG-khez tartozó 78 IRF9 kötőhely közül 43-ban azonosítottuk. Ez az előfordulási gyakoriság (55,1%) magasabb volt, mint a más ISG-khez tartozó régiókban számolt frekvenciák. A HOMER program segítségével minden IRF9 kötőhelyben azonosíthattuk azokat a szekvenciákat, amelyek leginkább hasonlítottak az ISRE motívumra, és meghatároztuk a motívum pontszámát, amely tükrözi az azonosított szekvencia súlyozott motívum mátrixhoz való hasonlóságát. Megállapítottuk, hogy a motívum pontszámok magasabbak voltak az effektor ISG-k promotor régióiban lokalizált IRF9 kötőhelyekben, mint az effektor ISG-k IRF9 kötőhelyeiben. Kiszámoltuk az TISRE előfordulási gyakoriságát is, és azt tapasztaltuk, hogy több mint háromszor gyakoribb volt az effektor ISG-khez tartozó IRF9-kötő régiókban más ISG-khez képest (29,5%, 78 kötőhelyből 23-ban). A következő 22 effektor ISG tartalmazta a TISRE motívumot az IRF9 kötő régióiban: *Bst2*, *Ddx60*, *Eif2ak2*, *Gbp2*, *Gbp3*, *Ifit1*, *Ifit2*, *Ifit3*, *Isg15*, *Isg20*, *Mx1*, *Mx2*, *Oas1a*, *Oas1c*, *Oas1g*, *Oas3*, *Oasl2*, *Rnasel*, *Samhd1*, *Trim21*, *Trim5*, és *Zc3hav1*. A legtöbb TISRE csak 1-2 bázissal különbözött a kanonikus 5'-GAAANNGAAANNGAAA-3' szekvenciától.

4.8. A kromatin nyitottság vizsgálata az effektor ISG-k szabályozó elemeiben

Három kérdésre kerestük a választ az effektor ISG-k szabályozó régiók nyitottságával kapcsolatban: (1) Milyen mértékben nyitottak az effektor ISG-k TSS-proximális és TSS-disztális IRF9 kötőhelyei a stimuláció előtt a DC-kben? (2) Mennyire nyitottak az effektor ISG-k TSS-proximális és TSS-disztális IRF9 kötőhelyei más sejtekben? (3) Mennyire nyitottak a DC-kben és a különböző sejtípusokban az egyes gén-szettek promotor régiói? Ezeknek a kérdéseknek a megválaszolásához saját ATAC-seq adatainkat és nyilvánosan elérhető ImmGen adatokat használtunk.

Első kérdésünk megválaszolásához meghatároztuk az elérhető régiók gyakoriságát különböző kötőhely listákban. Megállapítottuk, hogy az összes IRF9 kötőhelyek nagy része nyitott (ATAC-seq pozitív) volt a stimuláció előtt (69,0%), ami arra utalt, hogy más SRTF-ekhez hasonlóan az ISGF3 is jelentős mértékben kötődik a már létező nyitott (aktív vagy aktiválható) régiókhoz. Az effektor ISG-khez kapcsolódó TSS-proximális IRF9 csúcsok gyakrabban voltak

elérhető (79,4%), mint a TSS-disztális IRF9 kötőhelyek esetében (65,9%). Az ATAC-seq szignálintenzitásokban nem észleltünk különbséget a TSS-proximális és a TSS-distális IRF9 csúcsok között.

Második kérdésünk megválaszolásához megvizsgáltuk az effektor ISG-khez rendelt IRF9 kötőhelyek nyitottságát letölthető ImmGen ATAC-seq adatokat felhasználva. Kiszámítottuk az ATAC szignál-értékeket 34 TSS-proximális és 44 TSS-disztális IRF9-kötő régióban 92 immunsejttípusban. A variációs együttható (CV) értékeit használtuk a sejtek közötti variancia számszerűsítésére. Azt találtuk, hogy egy adott effektor ISG TSS-proximális IRF9 kötőhelyeinél az ATAC szignál-értékek jellemzően kevésbé változtak (alacsonyabb CV-t eredményezve), mint a TSS-distális IRF9 szignál-értékek. Ezek az adatok azt sugallják, hogy sok TSS-disztális szabályozóelem sejt-specifikus módon működik, az ISGF3 kötődése ezekhez a helyekhez sejt-specifikus kötődési mintákat és célgénaktivációt mutatnak.

Harmadik kérdésünk megválaszolásához elemeztük a DC-k és különböző sejttípusok promotor régióinak nyitottságát. Ezekben az analízisekben két további génlistát, a háztartási gének (housekeeping genes, HK) (n=31) és a gyulladásal kapcsolatos citokinek (IR, inflammation-related, n=36) listáját is elemeztük összehasonlításként. A DC-kben az ATAC szignál-értékek mediánját és a 92 sejttípusban az ATAC-seq szignál-értékek CV-jét kiszámítottuk. Azt találtuk, hogy DC-kben az ATAC szignál-értékek medián értékei hasonlóak az „összes egyéb ISG-k” értékéhez, és magasabbak, mint a legjobban indukálódó ISG-k és az IR-citokinek értékei. A különböző immunsejteket összevetve az ATAC-jelek CV-jében az effektor ISG-k és más ISG-listák között. A HK gének homogénebbek voltak (alacsonyabb CV értékek, kisebb szórással), mint bármely más génlistánál számolt adatok. Az ATAC-jelek CV-je magasabb volt az IR-citokinekben, mint a többi készletben.

5. Megbeszélés

Az első tanulmányunkban vizsgált IRF-ek, az IRF3, IRF5 és IRF9, eltérő, de részben átfedő szerepet játszanak az antivirális citokinek szabályozásában, a gyulladásos reakciókban és az antivirális állapot kialakításában. Eredményeink azt mutatták, hogy az IRF fehérjék által felismert DNS-motívumok és a további jellemzők, például a társfaktorok kötődése és a kromatin hozzáférhetősége szerepet játszanak a specifikus-kötés kialakításában. Fontos megjegyezni, hogy az IRF-domináns klaszterek régiói eltérőek voltak, és nem minden régióra volt igaz az adott klaszter jellemzőinek a mintázata. A régiók változatossága és az a tény, hogy egy adott jellemző nem csak kizárólag egy klaszterben fordul elő, azt jelzi, hogy különböző jellemzők sorozata kombinatorikus módon befolyásolja az IRF-specifikus kötődést, és nem kizárólag az erősen polarizált DNS-szekvenciák. A vizsgálati rendszerünkben azonosítottuk és összehasonlítottuk az IRF3-, IRF5- és IRF9-domináns régiókat. A három domináns klaszter közül az IRF5-domináns klaszter alapvetően különbözött az IRF3- és IRF9-domináns klaszterektől. Az IRF5 fehérje három potenciális mechanizmuson keresztül foglalhat el genomi régiókat: homo- vagy heterodimerként, monomerként és közvetett módon kötődhet az ISRE-hez. Eredményeink azt mutatták, hogy az ISRE motívum feldúsult az IRF3 és IRF5 megosztott és a közös klaszterben, ami arra utal, hogy az IRF5 fehérjék egy része dimerként kötődik a kanonikus ISRE motívumhoz. Azon megfigyelések alapján, hogy a *de novo* motívum kereséssel ISRE félhelyeket azonosítottunk az IRF5 cisztrómban, és egy másik tanulmányban PBM vizsgálatok alapján azonosított IRF5 logó hasonlít az ISRE félhely motívumra (GAAA), nem feltétlenül jelenti azt, hogy az IRF5 fehérje dominánsan vagy kizárólag monomerekként kötődik a DNS-hez. Andrienas és munkatársai szerint a rövidebb logó egy dimer helyet képvisel, aminek az a következménye, hogy egy IRF5 fehérje az ISRE félhelyhez erőteljesebben kapcsolódik. Tehát, a PBM kísérletek alapján az IRF5 dimerként kötődik, és az aszimmetrikus ISRE-félhelyeket preferálja. A harmadik mechanizmus szerint az IRF5 nem köti közvetlenül a DNS-t, hanem más transzkripciós faktorok határozzák meg a kötőhelyét. Ezt a mechanizmust bizonyították az IRF3 fehérje esetében, amely szignál-specifikus kofaktorként funkcionál az NF- κ B-függő gének transzkripciójának aktiválása során anélkül, hogy az ISRE motívumhoz kötődne. Az IRF5 fehérje által közvetített génszabályozás igazolása, amely a fehérje közvetett kötődése révén valósul meg további kísérleti megerősítést igényel.

Vizsgálatunk során azt találtuk, hogy az IRF5-domináns klaszternek három jellemző tulajdonsága van. Az első megállapításunk, hogy az IRF5 domináns régiók többsége nem tartalmaz ISRE motívumokat. Egy nemrégiben elvégzett tanulmányban bizonyították, hogy az ISRE-félhely 5' és 3' végén C-től eltérő bázisok (5-CGAAAC-3) megakadályozhatják az IRF5 fehérje kötődését az IFN promóterének vírus-válaszeleméhez. Elemzésünk során megfigyeltük, hogy nemcsak azok

az ISRE variánsok, amelyek ezeken a helyeken a C-től eltérő bázisokat tartalmaztak (például GAAANGGAAA), hanem az összes ISRE variáns alulreprezentált az IRF5-domináns klaszterben. Következő megállapításunk, hogy a speciális ISRE-félhelyek mint például az 5'-GAGA-3' és az 5'-GACA-3' feldúsultak az IRF5-domináns klaszterben a többi klaszterhez képest. További vizsgálatokra van szükség annak a feltevésnek és molekuláris mechanizmusnak a bizonyítására, amely szerint az IRF5 kötődik a speciális 6-merekhez, és hogy az IRF5 fehérjék monomerekként képesek-e megkötni ezeket az ISRE-félhelyeket. Végül azt találtuk, hogy az NFκB motívum feldúsul az IRF5-domináns klaszterben és ez összhangban volt azzal a megfigyeléssel, hogy az IRF5-kötés nagyrészt az NF-κB-vel való interakcióin alapszik.

Eredményeink azt mutatták, hogy az IRF3 és az IRF9 domináns régiók jobban hasonlítottak egymáshoz, és az IRF3 és az IRF9 dominancia kevésbé volt prediktálható. Számos oka lehet annak, hogy a gépi tanulással az előrejelzés alacsonyabb pontosságú. A klaszterezés pontatlan lehet a ChIP hatékonyságának különbségei vagy más okok miatt. A pontos predikcióhoz olyan jellemzőkre is szükség lehet melyeket nem vizsgáltunk. Az ilyen hiányzó nem vizsgált jellemzők lehetnek a motívumok száma egy adott régióban, a különböző motívumok közötti távolság, a transzkripció faktorok lefedettségi értékei, további motívumok jelenléte vagy az ISRE és a 6-merek további variánsai. A jövőben egy összetettebb elemzés felfedhet további jellemzőket is, amelyekre szükség lehet a pontosabb predikcióhoz. Ugyanakkor, elemzésünk során azonosítottunk olyan mintázatokat, amelyek előnyben részesítik az IRF3-domináns vagy az IRF9-domináns kötődést.

Az IRF3-domináns klaszterben azt találtuk, hogy a kötő régiók közül a stimuláció előtt kevesebb volt a nyitott kromatin régiók száma. Az NFκB, AICE, TRE és CRE motívumok nagy frekvenciája a klaszterben arra enged következtetni, hogy az IRF3 más társfaktorokkal való együttműködést igényel a kromatin „megnyitásához”. Alternatív megoldás lehet három vagy négy IRF3 vagy más IRF tandem-kötődése a TISRE motívumhoz, amely szintén feldúsul a klaszterben és ez elegendő lehet a nukleoszóma eltávolításához. Az IRF9 nem köti nagy affinitással az IRF3-domináns kötő régiókat, valószínűleg azért, mert az NF-κB és AP-1 nem indukálódik az I-IFN útvonalon. Továbbá azt találtuk, hogy sok IRF9-domináns klaszterben feldúsulnak az ISRE és GAS motívumok. Az 5-WBVGGAANNGAAACT-3 konszenzusszekvencia és variánsai is feldúsultak az IRF9-domináns klaszterben, viszont előfordulási gyakoriságuk alacsony volt. A GAS-motívum feldúsulása arra enged következtetni, hogy az I-IFN jelátvitel által aktivált STAT1 homodimerek fontos meghatározói az IRF9-domináns kötés kialakulásának.

A második tanulmányunkban az effektor ISG-k szabályozó elemeit vizsgáltuk. Azt tapasztaltuk, hogy a legtöbb effektor ISG (81,0%, 37 génből 30) promoter régiója olyan ISGF3-kötő régióval rendelkezik, amely erős ISRE-motívumot tartalmaz, és számos sejttypusban

hozzáférhető. A promóterrégiók nemcsak az általános TF-ek számára tartalmazhatnak kötőhelyeket, hanem promóter-specifikus TF-ek (például SP-1, YY-1 és NFY), LDTF-ek és más TF-ek számára is. A DNS-szekvenciák mellett a genomi régiók hozzáférhetősége kulcsfontosságú az SRTF-kötések kialakításánál. Azok a promóterek, amelyeket promóter-specifikus TF-ek tesznek hozzáférhetővé (nyitottá), azok nem függenek az LDTF-ektől, tehát sejttípus-függetlenül aktiválhatóak. Az IRF9 maga is fontos szerepet játszik az ISG-k promóter-régióinak nyitásában, ugyanis számos promóterben (pl. az Mx1, Mx2, Oas1a és Oas2 promóterei), ahol az IRF9 kötődött nyugvó vad típusú (wild type, WT) makrofágokban, a kromatin hozzáférhetősége csökkent az *Irf9*^{-/-} sejtekben.

Az effektor ISG-k csak egy kisebb része (23,0%, 39 génből 9) kapcsolódott két vagy több TSS-disztális elemhez. A TSS-proximális és TSS-disztális IRF9 kötőhelyek bizonyos szempontból különböztek. Az ISRE motívumok erőssége és gyakorisága a TSS-disztális IRF9-kötő régiókban alacsonyabb volt, mint a TSS-proximális régiókban. A „gyengébb” ISRE-motívumok alacsonyabb IRF9-kötéssel párosultak. Ezenkívül nyitottságuk inkább sejttípus-specifikus volt, mint a TSS-proximális csúcsok nyitottsága. Mivel az ISGF3 főként a már meglévő, hozzáférhető szabályozó régiókhoz kötődik, valószínűleg az ISGF3 különböző sejttípusokban különböző TSS-disztális szabályozó régiókat foglal el.

6. Összefoglalás

Első tanulmányunkban azokat a jellemzőket határoztuk meg, amelyek lehetővé teszik, hogy az IRF-ek specifikus genomi helyeken kötődjenek. Az IRF3, IRF5 vagy IRF9 által kötött régiók azonosítására ChIP-seq kísérleteket végeztünk aktivált egér dendritikus sejtekben. Az azonosított régiókat a DNS-motívumok feldúsulása, az interferon-stimulált válaszadó elem (ISRE) és az ISRE félhelyek, valamint a kromatin hozzáférhetősége szempontjából elemeztük. Gépi tanulási módszerrel vizsgáltuk az IRF-dominancia kiszámíthatóságát. Megállapítottuk, hogy az IRF5-domináns régiók alapvetően különböztek az IRF3- és IRF9-domináns régióktól: az ISRE-k ritkák, míg az NF κ B-motívum és a speciális ISRE-félhelyek, mint például az 5'-GAGA-3' és az 5'-GACA-3' gyakrabban fordultak elő. Az IRF3- és IRF9-domináns régiókat a feldúsult ISRE motívum és a nyitott kromatin alacsonyabb gyakorisága jellemezte. Az IRF3- vagy IRF9-domináns régiók kevésbé voltak szétválaszthatóak, de például az ISRE néhány speciális változata, a TISRE és az NF- κ B motívumai az IRF3 esetében, valamint a GAS motívum és bizonyos ISRE változatok az IRF9 esetében nagyobb gyakorisággal fordultak elő.

Második vizsgálatunkban többféle omikai megközelítést alkalmaztunk azon (epi)genetikai jellemzők azonosítására, amelyek lehetővé teszik az IFN-indukálható antivirális effektorok robosztus és általános, több sejtben megvalósuló transzkripció szabályozását. Meghatároztuk a szabályozó elemek elhelyezkedését, a DNS-motívumokat, az ISGF3 alegységeket (IRF9-STAT1-STAT2) és egyéb transzkripció faktorok kötését, valamint a kromatin nyitottságát egér dendritikus sejtekben. Az ISRE és a TISRE elemek az effektor ISG-k szabályozó elemeiben gyakrabban fordultak elő, mint más vizsgált ISG csoportokban. A kromatin nyitottsága a promóter régiókban hasonló volt a legtöbb más ISG-hez, de magasabb, mint a gyulladással kapcsolatos citokinek promótereinél, amelyeket referencia génkészletként használtak. A legtöbb effektor ISG-nek (81,1%) volt legalább egy ISGF3-kötőhelye a transzkripció starthelyéhez (TSS) közel, és az effektor ISG-k csak egy részhalmaza (24,3%) kapcsolódott három vagy több IRF9-kötő régióhoz. Az IRF9 jelek jellemzően magasabbak voltak, az ISRE motívumok pedig „erősebbek” a TSS-proximális, mint a TSS-disztális szabályozó régiókban. Ezen kívül a legtöbb promóter régió több sejttypusban is nyitott volt a stimuláció előtt. Eredményeink azt mutatják, hogy az „erős” ISRE motívumok és az univerzálisan nyitott promóter régiók az effektor ISG-k fontos jellemzői.

7. Új tudományos eredmények összefoglalása

IRF3-, IRF5- és IRF9-specifikus kötőhelyek azonosítása és jellemzése

- Az IRF5-domináns régiók alapvetően különböztek az IRF3- és IRF9-domináns régióktól: az ISRE-k ritkák, míg az NF κ B-motívum és a speciális ISRE-félhelyek, gyakrabban fordultak elő.
- Az IRF3- és IRF9-domináns régiókat a feldúsult ISRE motívum és a nyitott kromatin alacsonyabb gyakorisága jellemezte.
- Az IRF3- vagy IRF9-domináns régiók kevésbé voltak szétválaszthatóak, de például az ISRE néhány speciális változata, a TISRE és az NF- κ B motívumai az IRF3 esetében, valamint a GAS motívum és bizonyos ISRE változatok az IRF9 esetében nagyobb gyakorisággal fordultak elő.

Effektor ISG-ek szabályozó régióinak azonosítása és jellemzése

- A legtöbb effektor ISG-nek volt legalább egy ISGF3-kötőhelye a transzkripció starthelyéhez (TSS) közel.
- Az ISRE és a TISRE elemek az effektor ISG-k szabályozó elemeiben gyakrabban fordultak elő, mint más vizsgált ISG csoportokban.
- A kromatin nyitottsága a promóter régióiban hasonló volt a legtöbb más ISG-hez, de magasabb, mint a gyulladásal kapcsolatos citokinek promótereinél, amelyeket referencia génkészletként használtunk.
- Az IRF9 kötési értékek jellemzően magasabbak voltak, az ISRE motívumok pedig „erősebbek” a TSS-proximális, mint a TSS-disztális szabályozó régiókban. A több sejttípusban detektálható nyitottság jellemző a TSS-proximális kötőhelyekre.



Nyilvántartási szám: DEENK/458/2022.PL
Tárgy: PhD Publikációs Lista

Jelölt: Nagyné Csumita Mária
Doktori Iskola: Molekuláris Sejt- és Immunbiológia Doktori Iskola
MTMT azonosító: 10057535

A PhD értekezés alapjául szolgáló közlemények

1. Göczi, L., **Csumita, M.**, Horváth, A., Nagy, G., Póliska, S., Pigni, M., Thelemann, C., Dániel, B., Mianesaz, H., Varga, T., Sen, K., Raghav, S. K., Schoggins, J. W., Nagy, L., Acha-Orbea, H., Meissner, F., Reith, W., Széles, L.: A Multi-Omics Approach Reveals Features That Permit Robust and Widespread Regulation of IFN-Inducible Antiviral Effectors.
J. Immun. 209 (9), 1-12, 2022.
DOI: <http://dx.doi.org/10.4049/jimmunol.2200363>
IF: 5.426 (2021)
2. **Csumita, M.**, Csermely, A., Horváth, A., Nagy, G., Monori, F., Göczi, L., Orbea, H. A., Reith, W., Széles, L.: Specific enhancer selection by IRF3, IRF5 and IRF9 is determined by ISRE half-sites, 5' and 3' flanking bases, collaborating transcription factors and the chromatin environment in a combinatorial fashion.
Nucleic Acids Res. 48 (2), 589-604, 2020.
DOI: <http://dx.doi.org/10.1093/nar/gkz1112>
IF: 16.971





További közlemények

3. Póliska, S., Besenyei, T., Végh, E., Hamar, A. B., Karancsiné Pusztai, A., Váncsa, A., Bodnár, N., Szamosi, S., **Csumita, M.**, Kerekes, G., Szabó, Z., Nagy, Z., Szűcs, G., Szántó, S., Zahuczky, G., Nagy, L., Szekanecz, Z.: Gene expression analysis of vascular pathophysiology related to anti-TNF treatment in rheumatoid arthritis. *Arthritis Res. Ther.* 21 (1), 94, 2019.
IF: 4.103

A közlő folyóiratok összesített impakt faktora: 26,5

**A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre):
22,397**

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudományometriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2022.10.24.



Köszönetnyilvánítás

Köszönetemet és hálámat szeretném kifejezni témavezetőmnek Dr. Széles Lajosnak, a PhD képzésem során nyújtott folyamatos iránymutatásáért. Köszönöm az értékes elméleti és gyakorlati tanácsait és hogy minden körülmények között támogatta és türelemmel kísérte végig munkámat. Szeretném megköszönni a Molekuláris Sejt- és Immunbiológia Doktori Iskola korábbi vezetőinek Prof. Dr. Fésüs Lászlónak és Prof. Dr. Tózsér Józsefnek, és jelenlegi vezetőjének Prof. Dr. Balogh Istvánnak, hogy PhD tanulmányaimat és kutatómunkámat a doktori iskola keretein belül végezhettem.

Továbbá köszönöm Dr. Zahuczky Gábornak, hogy lehetővé tette a PhD képzés elkezdését és az első két évben témavezetőként nyújtott segítséget. Külön szeretném megköszönni Dr. Pólska Szilárdnak az MSc és PhD képzésem alatt nyújtott értékes szakmai tanácsait. Szeretném köszönetemet kifejezni közeli munkatársaimnak Göczi Lorándnak, Monori Fannynak, Csermely Attilának, Dr. Erdős Edinának és Dr. Nagy-Bojcsuk Dórának, akik folyamatos szakmai segítséget nyújtottak munkáim során. Köszönetemet szeretném kifejezni a Biokémiai és Molekuláris Biológiai Intézet jelenlegi és korábbi munkatársainak: Ixchelt Cuaranta Monroy-nak, Dr. Nagy Gergőnek, Dr. Horváth Attilának, Dr. Ozgyin Lillának, Mátyás Erzsébetnek, Balogh Máriának, Fürtös Ibolyának és Silye-Cseh Tímeának, akik sok hasznos tanáccsal láttak el a kutatómunkám során. Köszönettel tartozom a Debreceni Egyetem Humángenetikai Tanszék korábbi vezetőjének Prof. Nagy Bálintnak és további munkatársainak, Dr. Szilágyi-Bónizs Melindának, Dr. Márton-Deme Évának és Magyarné Trefán Katalinnak a laboratóriumi munkámban nyújtott támogatásukért. Köszönöm Dr. Kádas Jánosnak az UD-GENOMED Medical Genomic Technologies Kutatás-fejlesztési és Szolgáltató Kft. ügyvezető igazgatójának, hogy támogatta a PhD tanulmányaim befejezését.

Hálával tartozom családomnak és barátaimnak. Köszönöm szüleimnek a türelmet, hogy végig kísérték ezen az úton is. Köszönöm gyermekemnek és férjemnek, hogy velem voltak, szeretetükkel támogattak munkáim során és mindig jobbra ösztönöztek.