

EGYETEMI DOKTORI (PhD) ÉRTEKEZÉS TÉZISEI

**A csontvelő eredetű makrofágokban szerepet játszó
transzkripciós faktorok vizsgálata**

Nagy Gergely

Témavezető: Dr. Barta Endre



DEBRECENI EGYETEM

MOLEKULÁRIS SEJT- ÉS IMMUNBIOLÓGIA DOKTORI ISKOLA

Debrecen, 2016

A csontvelő eredetű makrofágokban szerepet játszó transzkripciós faktorok vizsgálata

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

Írta:
Nagy Gergely
okleveles molekuláris biológus

Készült a
Debreceni Egyetem
Molekuláris Sejt- és Immunbiológia
doktori iskolája keretében

Témavezető: Dr. Barta Endre, PhD

A doktori szigorlati bizottság:

elnök: Prof. Dr. Szabó Gábor, az MTA doktora
tagok: Prof. Dr. Nagy Péter, az MTA doktora
Dr. Homolya László, az MTA doktora

A doktori szigorlat időpontja: Debreceni Egyetem ÁOK
In Vitro Diagnosztikai Tömb
2016. június 21. 11:00

Az értekezés bírálói:

Prof. Dr. Boros Imre, az MTA doktora
Dr. Nemoda Zsófia, PhD

A bírálóbizottság:

elnök: Prof. Dr. Szabó Gábor, az MTA doktora
tagok: Prof. Dr. Boros Imre, az MTA doktora
Prof. Dr. Nagy Péter, az MTA doktora
Dr. Homolya László, az MTA doktora
Dr. Nemoda Zsófia, PhD

Az értekezés védésének időpontja: Debreceni Egyetem ÁOK
Belgyógyászati Intézet „A” épület tanterme
2016. június 21. 13:00

1. Bevezetés

1.1. Transzkripció szabályozás

A transzkripció iniciáció, elongáció és termináció a naszcens RNS szintézisének fő lépései. Mivel az RNS polimeráz, ha elindult, végigfut a génen, az iniciáció biztosíthatja a legszélesebb körű szabályozási lehetőségeket. Eukariotákban a fehérjekódoló gének promótereinek transzkripció faktor kötőhelyein (TFBS) felépült pre-iniciációs komplex (PIC) a transzkripció faktorokból (TF), ezek ko-regulátoraiból, valamint az RNS polimeráz II-ből (Pol II) áll. A Pol II elindításához nemcsak a nagyszámú szabályozó fehérje jelenléte, hanem az enzimaktivitásuk is szükséges, pl. az RNS polimeráz II B1 (RPB1) alegység C-terminális doménjének többszörös szerin foszforilációja is.

A hiszton fehérjék számos poszttranszlációs módosítást hordozhatnak, melyek közül az acetilációt és metilációt vizsgálták leginkább. Nagyszámú módosításokat létrehozó és eltávolító enzim van, amelyek végsősoron felépítik a különböző sejttípusok epigenetikai mintázatát, így meghatározva a DNS hozzáférhetőségét. A TF-ok sajátos DNS elemeket kötnek a promóterben és távolabbi régiókban, melyeket enhanszereknek (vagy silencer-eknek) neveznek. E szabályozó régiók működése a DNS-szál hurkolódásával magyarázható, melynek során a promóter és enhanszer(ek) kötőhelyei a 3-dimenziós kromatinszerkezetben közelkerülnek egymáshoz.

A ko-regulátorok, melyek kölcsönhatnak a TF-okkal, de képtelenek közvetlenül DNS-t kötni, felelősek a nukleoszómák módosításáért és eltávolításáért, és végeredményben a génexpresszióért. A ko-aktivátorok, pl. a P300/CBP fehérjék saját hiszton acetiltranszferáz aktivitással rendelkezhetnek, míg a ko-represszorok jellemzően hiszton deacetiláz szereppel bírnak. A H3 és H4 hisztonok lizin oldalláncainak acetilációja általában az aktív enhanszerekre jellemző. A hiszton metiltranszferázok és demetilázok a hisztonmódosító ko-regulátorok másik jobban ismert csoportjai.

1.2. A transzkripció faktorokat irányító promóter és enhanszer szekvenciák

A TF-ok, melyek jeleket továbbítanak a sejtmagba, bent egyedi DNS motívumok (TFBS-ek) által is irányítottak. A promóterek és enhanszerek eltérő motívumokkal rendelkeznek. Az eukariota promóterben a TATA-box volt az első leírt szabályozó elem. Ezt a „TATA-kötő fehérje” (TBP) köti, amely a TAF1-13 fehérjékből is álló TFIID komplex része. A TFIID-t a TFIIA trimer és a TFIIB fehérje közvetlenül köti, mely utóbbi „B felismerő elemeket” (BRE) foglalhat el. Az „általános transzkripció faktorok” közül a TBP és TFIIB tekinthető „valódi”, saját kötőhellyel rendelkező TF-nak, de a TAF-ok közül néhányánál szintén kimutatták bizonyos szekvenciák előnyben részesítését.

Az újgenerációs szekvenálás (NGS) korszaka előtt azt írták le, hogy a legtöbb promóter TATA-mentes, a fele tartalmaz Iniciátort és a fele „downstream” promóter elemmel vagy BRE-vel rendelkezik, lefedve így a transzkripció start helyek (TSS) közel 100%-át. Manapság úgy tűnik, hogy ezek az elemek sokkal ritkábbak. Az NGS megmutatta, hogy kétféle transzkripció iniciációs hely létezik: keskenyek, kevesebb, mint 10 bp szélességgel (TSS) és szélesebbek, 25-250 bp szélességgel, melyeket transzkripció start régióknak (TSR) neveznek. Az összesen 22%-nyi TSS inkább a klasszikus elemeket hordozza, míg a TSR-ek inkább CpG szigetek motívumaival rendelkeznek.

Az elsőként azonosított CpG-kötő TF az SP1 volt, amely az „SP1-szerű” család alapító tagja. Ez a család közeli rokona a „Krüppel-szerű” faktoroknak (KLF1-17), amelyek ugyanolyan típusú DNS-kötő doménnel (DBD) rendelkeznek, így szintén képesek a GC-box kötésére. A további promóterspecifikus motívumfeldúsulások keresése közben különböző megközelítésekkel hasonló motívumok jelentek meg. Az NGS-szel párosított kromatin immunoprecipitáció (ChIP-seq) azoknak a DNS fragmenseknek a kihalászására és megszekvenálására alkalmas módszer, amelyeket köt, amelyekkel így keresztköthető egy adott, specifikus antitesttel kifogható fehérje. Ezzel és más NGS módszerekkel, további

motívumokat határoztak meg: egy ETS kötőhelyet (EBS), a CCAAT-boxot, a CRE-t, TRE-t és az NRF1, GFY, GFX, MYC és YY1 motívumait.

Az első „eritroblaszt transzformáció-specifikus” (ETS) fehérje az E26 retrovírus fúziós onkogénjeként lett azonosítva, amely csirkében leukémiához vezet. A szupercsaládban 27/28, 12 filogenetikai csoportba osztott tagja van humánban, ill. egérben. Promóterekben inkább a ccGGAAgt szekvencia mutat feldúsulást, amelyet általában a minden sejtben előforduló GABP dimer, az ETS2, ETV6 és a TCF család fehérjéi kötnek, és sejttípus-specifikusan a PEA3 és ELF családok tagjai is köthetnek. Két egyedi motívum is létezik: Az SPDEF többnyire a caGGATga szekvenciát, míg az SPI család inkább a gaGGAAgt szekvenciát köti.

Az aktivátor protein 1-et (AP-1) a forbol 12-O-tetradekanoát 13-acetát (TPA) által indukálható TF-ként írták le, amely egy promóterspecifikus TPA válaszadó elemet (TRE) köt. Nem sokkal később, az AP-1 fehérjét JUN-ként és annak interakciós partnereiként, FOS-ként és FOS-szerű antigénekként azonosították. Ma már valószínűleg az összes „alap leucin cipzár” (bZIP) fehérje ismert, így világossá vált, hogy a FOS és JUN fehérjék külön családokat alkotnak, és az ATF/CREB, BATF és MAF család fehérjéivel képeznek heterodimereket. Az ATF-et vagy CREB-et magukban foglaló dimereket CREB fehérjének tekintik, mivel ezek általában a cAMP válaszadó elemet (CRE) kötik. Az AP-1 és CREB fehérjék egy olyan palindrom szekvenciában osztoznak, amely két TGA félhelyből épül fel egy, ill. két elválasztó nukleotiddal.

Az NFY egy három fehérjéből álló komplex, amely a CCAAT-boxot köti, és bár ez az elem gyakoribb a TATA-boxnál, az NFY-nak csak „erősítő” szerepe van. Az NRF1 egy homodimert képző egyedi TF, amely nagyszámú gén promóterében köt palindrom szekvenciát. Mivel eleme csak a középső citozinjainak metilált állapotában köthető, a GFX

ZBTB33-ként való azonosítása nem volt könnyű. A GFY motívum kötésére két jelölt is van, a ZFP143 és THAP11.

1.3. Magreceptorok

A magreceptor egy beszédes név, mivel olyan TF-okat takar, amelyek szignál molekulákat ismernek fel. A ligandkötést követően a konformációjuk megváltozik, ami módosítja a válaszadó elemeikhez és a PIC tagjaihoz való affinitásukat. Ezek a ligandok általában hormonok, beleértve a szteroidokat, retinoidokat, tiroid hormont és a D-vitamint, de a szupercsaládnak számos, ismert ligand nélküli tagja is van, amelyeket ezért árva receptoroknak neveznek. A DBD a variábilis N-terminális A/B és a középső „csukló” domének között, két konzervált cink-ujjból áll, amelyek jellemzően az AGGTCA szekvenciát kötik. A C-terminális ligandkötő domén (LBD) biztosítja a ligandspecifitást és szelektivitást, valamint a dimerizációs és egyéb interakciós felszíneket. Árva receptorok esetében ez felel a folyamatos vagy poszttranszlációs módosítástól függő szabályozásért.

A magreceptorok négy osztályba sorolhatóak. Az első klónozott magreceptorok, a glükokortikoid és ösztrogén receptor a szteroid hormon receptorok osztályának (I. osztály) alapító tagjai. A II. osztály receptorai a retinoid X receptorral (RXR) alkotnak heterodimert, és általában egymást követő RGKTCA félhelyeket (DR) kötnek. Az RXR, bár 9-cisz retinsav és 9-cisz-13,14-dihidroretinsav ligandjait leírták, a III. osztályba tartozik, amely a dimerképző árva receptorokat foglalja magában. A IV. osztály monomer árva receptorai a konszenzus hexamert kötik. A II. osztály magreceptorai sokféle lipidet kötnek. A peroxiszóma proliferátor aktivált receptor (PPAR) sokféle zsírsavat, a retinsav receptor (RAR) transz- és 9-cisz-retinsavat és a tiroid hormonreceptor (THR) trijód-tironint köt. A D-vitamin receptor (VDR), máj X receptor (LXR), farnezoid X receptor (FXR), pregnán X receptor (PXR) és a konstitutív androsztán receptor (CAR) mind szteroidszármazék ligandokra válaszolnak.

1.4. A makrofágok fejlődésében résztvevő transzkripció faktorok

A makrofágok a mieloid vonalból származó „hivatásos” fagociták. Bár az eredetük és szöveti környezetük befolyásolja a működésüket és anyagcseréjüket, úgy tűnik, a makrofágok mindig ugyanazokban a TF-okban osztoznak. A kialakulásuk során, a FLI1 (ETS) által serkentett RUNX1 a TAL1-gyel és C/EBPb-val együtt indukálja a PU.1 (ETS) kifejeződését, amely a C/EBPa-val végül meghatározza a makrofág jelleget. Van egy további, a makrofág funkciót meghatározó TF csoport, az interferon (IFN) szabályozó faktor (IRF) család, amely 9 taggal rendelkezik. Az IFN-stimulált válaszadó elemet (ISRE) számos IRF dimer köti, de bizonyos IRF-ek PU.1-gyel is heterodimerizálnak. Egér csontvelő eredetű makrofág (BMDM) sejtekben az ETS-IRF kompozit elemet (EICE) a PU.1/IRF8 heterodimer köti.

1.5. NGS módszerek a funkcionális genomikában

A ChIP-seq és RNA-seq valószínűleg a legszélesebb körben alkalmazott NGS módszerek többé-kevésbé kifinomult kiértékelő stratégiákkal, bár mindig akadnak olyan kérdések, amelyek egyedi megoldást igényelnek. Ezek egyike a nukleoszóma-mentes régiók (NFR) meghatározása. Csak néhány olyan módszer van, amely lehetővé teszi az NFR-ek hiszton módosításra végzett ChIP-seq adatokból való prediktálását. Ilyen pl. a HOMER-é, amely a ChIP-seq jelekben lévő legnagyobb különbséget mutató régiókat keresi. Létezik egy egyedi módszer a naszcens transzkriptóm feltérképezésére, amelyet teljes genomra kiterjedő „run-on” szekvenálásnak (GRO-seq) hívnak. Ennek elemzésére – a miénk mellett – csak két publikált stratégia van, egy rejtett Markov modell alapú transzkriptum-meghatározó megközelítés, és a HOMER-é, bár ezek nagyrészt a dominánsan átírt szál meghosszabbított transzkriptumaira összpontosítanak, így a másik szálon kicsi a meghatározás pontossága.

2. A tanulmány célja

Az egér BMDM egy jól ismert sejttípus, ezért jó modell a TF-ok által meghatározott epigenetikai „domborzat” kutatására. A hiszton módosítási ChIP-seq adatokban az NFR-ek a funkcionálisan legfontosabb régiók, ezért ezek meghatározására H3K4me2, H3K4me3, H3K27ac és H4ac módosításokra végeztünk kísérleteket, és kifejlesztettünk egy új NFR meghatározási módszert. A következő cél a módszerünk motívumfeldúsulások elemzése általi validálása és azoknak a TF-oknak a meghatározása volt, amelyek képesek kötni a talált motívumoknak megfelelő elemeket. Mivel a visszatérképezett TFBS-ek a különböző TF csoportok számos tagja által elfoglalhatóak, GRO-seq-et és RNA-seq-et hajtottunk végre a lehetséges regulátorok naszcens és érett mRNS szintjének kiderítésére. Emellett célunk volt a makrofág állapotot meghatározó PU.1 DNS-kötésének vizsgálata is.

Az RXR heterodimerizáló partnerei számos útvonalat aktiválnak, ezért az RXR inkább egy passzív segédmolekulának látszik. A BMDM sejtek RXR liganddal való kezelése után az LXR-ek, RAR-ok és PPAR-ok általi génaktiváció várható, ezért kérdés volt, hogy vajon vannak-e csak az RXR-re specifikus útvonalak. Hogy megválaszoljuk ezt a kérdést, ChIP-seq-et végeztünk RXR-re, PU.1-re és P300-ra, a génexpressziós változásokat pedig GRO-seq-kel és RNA-seq-kel követtük. A GRO-seq nemcsak génexpressziós szinten teszi lehetővé a közvetlen szabályozó hatások észlelését, hanem az enhanszerek átírásának szintjén is, ezért felhasználható az aktív enhanszerek szabályozott génekhez való hozzárendeléséhez. Mivel nem volt alkalmas eszköz az átírt szabályozó régiók a kifejeződött génektől való elkülönítésére, kifejlesztettünk egy eszközt, amely képes körülhatárolni és annotálni minden egyes transzkripció eseményt. CTCF és RAD21 fehérjékre is hajtottunk végre ChIP-seq-et, és kifejlesztettünk egy predikciós módszert a szabályozó egységeket határoló inzulátor régiók megtalálására. Végül néhány kiválasztott régióban 3C-seq végrehajtásával erősítettük meg a talált promóter-enhanszer kölcsönhatásokat.

3. Anyagok és Módszerek

3.1. A csontvelő eredetű makrofág (BMDM) sejtek differenciációja

A csontvelő vad típusú hím C57Bl/6J egerek combcsontjából lett kimosva. A sejteket Ficoll-Paque gradiensén (Amersham Biosciences, Arlington Heights, IL) keresztül tisztítottuk és 20% endotoxin-mentes magzati borjúsérumot és 30% (MCSF tartalmú) L929 kondicionált médiumot tartalmazó DMEM-ben tartottuk 5 napig. A sejtek aztán oldószerrel vagy 100 nM LG268 (LG100268) liganddal, M. Leibowitz ajándékával (Ligand Pharmaceuticals) lettek 1 órán keresztül kezelve.

3.2. Az újgenerációs szekvenálással kapcsolt kromatin immunoprecipitáció (ChIP-seq)

A BMDM sejteket 30 percen keresztül DSG-vel (Sigma), majd 10 percig formaldehiddel (Sigma) keresztkötöttük. Fixálás után a kromatint Diagenode Bioruptor-al szonikáltuk 200-1000 bp fragmensek létrehozásához. A kromatint preimmun IgG-vel (Millipore, 12-370) és H4ac (Millipore, 06-866), H3K27ac (ab4729), H3K4me2 (Upstate, 07-030), H3K4me3 (ab8580), RXR (sc-774), P300 (sc-585), PU.1 (sc-352), CTCF (Millipore, 07-729) és RAD21 elleni antitestekkel (ab992) immunoprecipitáltuk. A kromatin-antitest komplexeket protein A-val burkolt paramágneses gyöngyökkel (Life Technologies) csaptuk ki. A komplexeket 6 mosási lépés után eluáltuk, majd megszüntettük a keresztkötéseket. A DNS fragmenseket oszloppal (Qiagen, MinElute) tisztítottuk, és QPCR analízishez vagy könyvtárkészítéshez használtuk őket. A ChIP-seq könyvtárak Ovation Ultralow Library Systems (NuGen) alkalmazásával készültek a gyártó utasításai szerint. A könyvtárakat 16 PCR ciklusban sokszorozítottuk, majd a primerek eltávolítása érdekében E-Gel systems-szel (Life Technologies) géltisztítottuk. A szekvenálás Illumina HiScanSQ szekvenátorral történt.

3.3. ChIP-seq analízis

A ChIP-seq-ből származó nyers szekvencia-leolvasások elsődleges elemzése a parancssoros ChIP-seq elemző eszközünkkel történt. Az mm9 egér genom-összeállításra való illesztést a BWA tools-szal végeztük, a BAM fájlok pedig SAMTools-szal készültek. A „genomlefedettség” fájlok HOMER-rel készültek, és IGV2-vel kerültek megjelenítésre. A genomlefedettség információ felhasználásával a PeakSplitter-t (EMBL-EBI) alkalmaztuk a nukleoszómák által elfoglalt régiók (NOR) meghatározására. Az NFR-eket a NOR-párok közötti régiók alapján határoztuk meg. A pontosabb NFR pozicionálás érdekében végül kombináltunk egy csúcspont és régióhatár alapú predikciós megközelítést.

A ChIP-seq csúcsok MACS2-vel lettek meghatározva. A kontroll és LG268 kezelt RXR minták két párhuzamosát a DiffBind v1.0.9 programmal elemeztük: a közös csúcsok azokból a csúcsokból lettek meghatározva, amelyek négy mintából legalább kettőben megjelentek; a szignifikánsan változó „kötőaffinitást” mutató csúcsok a „teljes könyvtárméret” paraméterrel lettek meghatározva. A különböző minták „csúcsérték-küszöbét” kézzel határoztuk meg. Az NFR-ek közepéhez vagy a csúcspontokhoz viszonyított átlagos „leolvasás-eloszlás” hisztogramok és „hőterképek” HOMER-rel készültek. Az átfedéseket BEDTools-szal határoztuk meg, és VennMaster-0.37.5 programmal ábráztuk. A motívum-feldúsulásokat HOMER-rel határoztuk meg, és az egyedi RXR elemek megkeresésére a fuzznuc-ot (EMBOSS) használtuk.

3.4. A teljes genomra kiterjedő „run-on” szekvenálás (GRO-seq)

A teljes genom szintű „run-on” vizsgálat és könyvtárárkészítés a korábbiakban leírtaknak megfelelően történt (Core et al., Science, 2008; Hah et al., Cell, 2011). A könyvtárak 100 nM LG268 liganddal 0, 30, 60 és 120 percig kezelt BMDM sejtek magjainak két biológiai replikájából készültek. A szekvenálás Illumina HiScanSQ szekvenátorral történt.

3.5. GRO-seq elemzés

A nyers szekvencia adatok elsődleges elemzése a ChIP-seq-nél leírtakhoz hasonlóan történt. A 2x4 makrofág mintából nyert szekvencia-leolvasásokat együttesen használtuk fel a transzkriptum-predikcióhoz és annotáláshoz. A szálspecifikus genomlefedettség fájlok HOMER-rel készültek. Az átírt „egységeket” szálspecifikusan, PeakSplitter-rel (EMBL-EBI) határoztuk meg. Az egységek hossza a csúcspontjukhoz képest +/-250 bp-ban lett korlátozva.

A divergens helyeket a széttartó irányú egységek alapján határoztuk meg. A transzkriptumokat az azonos szálon, egymástól 600 bp-nál közelebb fekvő egységekből építettük fel a BEDTools és egyéb parancssoros programok segítségével. A promóter régiók – közeli és intronikus – enhanszerektől való elkülönítésére a H3K4me3 ChIP-seq feldúsulásait használtuk.

A hosszabb transzkripciós eseményeket az alábbi feltételek alapján csoportosítottuk: egy transzkriptumot akkor annotáltuk ismert génként, ha az irányuk megegyezett és a prediktált, H3K4me3-mal jelölt TSS közelebb volt az ismert gén valamely TSS-éhez, mint 1,5 kb (1. csoport). A további H3K4me3-jelölt intergenikus és antiszensz transzkriptumok a 2. csoportba kerültek. A H3K4me3 jel nélküli divergens vagy egyszerű TSS-sel rendelkező transzkriptumokat a 3. csoportba gyűjtöttük. A további, ismert TSS-sel átfedő H3K4me3-jelölt vagy divergens helyek „full pausing”-ot mutató helyekként lettek összegyűjtve (4. csoport). Az ismeretlen transzkriptumokat újraannotáltuk és a lehetséges, hosszabb génekre eső azonos irányú géneket is azonosítottuk.

Minden ismert transzkriptumot meghatároztunk, és a kb. 45 nukleotid/s polimeráz sebességre való tekintettel ezeknek a legfeljebb 50 kb hosszú 5' végei lettek felhasználva a génexpressziós elemzésekhez, kizárva az „illesztési réseket”, a 3' túlnyúló végeket és bármely divergens helyet. A mintánkénti egyedi leolvasások számát azokon a fragmenseken határoztuk meg, amelyek egyesített hossza nagyobb volt fél kb-nál. A maSigPro-val végzett

elemzéshez az RPKM értékek 1/20-át használtuk. A 30 perc LG268 kezelés hatására szignifikánsan változó divergens helyek meghatározására a DiffBind v1.0.9 az RXR csúcsoknál használtakkal megegyező paraméterekkel került alkalmazásra. A szignifikánsan szabályozott, divergens transzkripciót mutató RXR csúcsokat PeakAnnotator-ral annotáltuk a fél Mb-on belüli legközelebbi, azonos irányban szabályozott gén transzkriptumhoz.

3.6. RNA-seq

A BMDM sejtek két biológiai replikáját 100 nM LG268-cal 0, 30, 60 és 120 percig kezeltük. Az RNA-seq könyvtárak a TruSeq RNA Sample Preparation Kit (Illumina) használatával készültek a gyártó utasításainak megfelelően: a könyvtárkészítéshez 2,5 µg totál RNS-t használtunk. A poli-A farokkal rendelkező RNS molekulák tisztítása poli-T-kötött mágneses gyöngyökkel történt. Az mRNS-t kétértékű kationokkal fragmentáltuk 85 °C-on, a cDNS első szála pedig random primerekkel és SuperScript II reverz transzkriptázzal (Invitrogen, Life Technologies) lett létrehozva. Ezt a második cDNS szál szintézise követte, majd a kétszálú cDNS fragmensek egy „A” nukleotid hozzáadásán, majd „vonalkód” indexelt adapter ligáláson (end repair-en) mentek keresztül. cDNS könyvtár létrehozásához az adapterkapcsolt termékek adapterspecifkus PCR-ral lettek feldúsítva. Az agaróz gélelektroforézist E-Gel EX 2%-os agaróz gélen (Invitrogen, Life Technologies) hajtottuk végre és a könyvtárakat a QIAquick Gel Extraction Kit (Qiagen) használatával tisztítottuk ki a gélből. A könyvtárak Illumina HiScanSQ szekvenátorral lettek megszekvenálva.

3.7. RNA-seq elemzés

A megszakított leolvasások térképezésére, a transzkriptumok összeállítására és a génexpressziós értékek FPKM formátumban való kinyerésére a TopHat és Cufflinks eszközöket használtuk.

3.8. Domén predikció az együtt detektált CTCF és RAD21 csúcsok alapján

A CTCF és RAD21 fehérjék által közösen elfoglalt régiókat (co-peak), amelyek 15 pont fölötti „csúcsértékekkel” és ezek 3 alatti hányadosával rendelkeztek, inzulátoroknak tekintettük. Akkor határoztuk meg az 1 Mb-on belüli legközelebbi inzulátorokat lehetséges funkcionális-domén határokként, ha az egyedi co-peak-ek összesített pontszáma 5/3-nál kisebb hányadost mutatott a párok között. A 100 kb-nál közelebb elhelyezkedő aktív doméneket nagyobb, aktív topológiai doménekké egyesítettük, a köztük lévő régiók pedig inaktív topológiai doménekként lettek meghatározva. A prediktált doméneket a BEDTools használatával annotáltuk a szabályozott génekhez. A RAD21 inzulátorokon és az aktív/passzív (GRO^{+/−}) RXR-kötött régiókon való feldúsulásainak elemzése a korábban leírtakhoz hasonlóan történt.

3.9. A kromatinkonformáció vizsgálata (3C)

A sejteket 10 percig 2%-os formaldehydben fixáltuk. A sejtmagok izolálása 10 mM Tris-HCl pH 7,5, 10 mM NaCl, 0,2% NP40 (Sigma) és proteáz inhibitor tablettát (Roche) tartalmazó pufferben történt. A kromatint 400U HindIII (Fermentas) restrikciós enzimmel emésztettük 37 °C-on 16 órán át, majd további 1 órán át 100U-tal. A kromatin fragmenseit 100U T4 DNS ligázzal (Fermentas) egyesítettük 16 °C-on 4 órán keresztül. A ligálást követően a kromatin keresztkötése egy éjszakán keresztül 65 °C-on lett megszüntetve. A ligált termékeket oszloppal (Roche, High Pure PCR Template Preparation Kit) tisztítottuk. Az egymást követő primerek a restrikciós enzimhasítóhelyek közvetlen közelébe lettek tervezve.

3.10. 3C szekvenálás

A 3C DNS-t fenol/kloroform/izoamil alkohol (25:24:1) (Sigma) eleggyel tisztítottuk. A második restrikciós emésztést DpnII (NEB) használatával végeztük 16 órán keresztül a

gyártó utasításai szerint. A második ligálás 16 °C-on 6 órán keresztül 200U T4 DNS ligázzal történt. A DNS-t aztán ismét fenol/kloroform/izoamil alkohol (25:24:1) eleggyel, majd QIAquick géltisztító oszloppal (Qiagen) tisztítottuk. A „csali”-specifikus inverz PCR-eket az univerzális Illumina adapterekkel és „vonalkód” szekvenciákhoz kapcsolt primerek használatával végeztük. A reakcióelegyeket QIAquick géltisztító oszlopokkal tisztítottuk. Két biológiai replika két technikai replikájának az ampliton könyvtárai Illumina MiSeq és HiSeq2000 szekvenátorral lettek megszekvenálva.

3.11. 3C-seq elemzés

A mintákat FASTX tools-szal válogattuk szét az index szekvenciájuk, majd a HindIII felismerő helyben végződő csali szekvenciájuk alapján. A megmaradó 68-82 nukleotid hosszú (HindIII hellyel kezdődő) fragmensek az mm9 genomösszeállításra való illesztésére a BWA tools-t használtuk. A teljes genomot beborító 1 Mb-os régiók lefedettsége az összes lehetséges kölcsönhatás ezredrészeként lett meghatározva a távoli kölcsönhatások gyakoriságának elemzésére. A korábban leírtak szerint meghatározott inaktív és aktív topológiai doméneket szétválasztottuk, majd meghatároztuk az aktív domének RXR-függő és független régióit.

3.12. Az AP-1/CREB-hez köthető bZIP fehérjék filogenetikai összehasonlítása

A BioMart-on keresztül 44 AP-1/CREB fehérje szekvenciáját töltöttük le az Ensembl adatbázisból. A filogenetikai elemzéshez a MEGA 5.05 program beépített eszközeit használtuk. A központi domének 56 aminosav hosszú szakaszainak többszörös illesztésére a MUSCLE-t alkalmaztuk. A filogenetikus fa „neighbor-joining” statisztikai módszerrel készült. A páronkénti távolságokat a Poisson modell alkalmazásával becsültük és a kapott hasonlósági mátrixot a filogenetikai fa alapján parancssorban rendeztük.

4. Eredmények

4.1. A makrofágok szabályozó régióinak meghatározása hiszton lefedettség alapján

A HOMER által meghatározott több mint 22000 enhanszerspecifikus NFR sokkal kevesebb volt a vártnál, bár a módszer jól működött a H3K4me3-ből származó adatokra. Hogy a lehető legtöbb NFR-t megtaláljuk, a NOR-ok meghatározására PeakSplitter-t használtuk. E NOR-ok lehetővé tették az NFR predikció finomhangolását és a szélesebb NFR-ek megtalálását is. Módszerünkkel a HOMER-rel kapott eredményekhez viszonyítva nagyjából négyszer több NFR volt meghatározható, és 56704 mind acetilációs, mind metilációs jelet mutató régiót találtunk. Ezek lehettek a makrofágjaink legaktívabb régiói.

Az NFR-ek folytonos szélességeloszlása összhangban volt a korábban javasolt „gördülő nukleoszóma” modellekkel. Az enhanszerekre jellemző NFR-ek nagyobb fele 150 +/- 75 bp széles régiókon mutatta a hisztonok eltávolodását, amely hozzávetőleg megegyezik az egy nukleoszómamagra tekeredett DNS hélix hosszával. H3K4me3 esetében az enhanszer jelekhez képest jellemzően szélesebb NFR-ek voltak megfigyelhetőek, amely a promóteren összeszerelődött PIC-nek tulajdonítható. A H3K4me3 intenzitás nagymértékben követte az enhanszer jeleket, de a trimetilált hisztonok általi legnagyobb lefedettséget a legszélesebb régiók mutatták.

4.2. A makrofágokra jellemző transzkripciós faktorok meghatározása NFR predikciókból

A hiszton acetilációs adatokból származó legjobb NFR-ek alkalmasak voltak az összes makrofág-specifikus motívum, a PU.1, AP-1, C/EBP, IRF, CREB és RUNX motívumok meghatározására. Érdekes módon a legjobb H3K4me2-jelölt régiók ugyanazokat a promóter-specifikus motívumokat adták, mint a H3K4me3 által jelöltek (SP1, NFY, ETS, NRF1, GFY, YY1, GFX, CREB és IRF). Végül közel 30000 lehetséges enhanszert azonosítottunk

makrofágokra jellemző elemekkel (H3K4me2 és H4ac és/vagy H3K27ac jellel, H3K4me3 nélkül), és több mint 11000 promótert a jól ismert TSS/TSR-specifikus motívumokkal.

4.3. A hisztonmintázatok vizsgálata a PU.1 kötőhelyek közelében

A hisztonmódosítások intenzitása és a PU.1 feldúsulás között szoros összefüggés volt. A PU.1 csúcsoknak kb. fele fedett át NFR-ekkel, amely meglepő módon azt jelentette, hogy a csúcsok jelentős része NOR-okkal vagy lehetséges heterokromatikus régiókkal fedett át, így további bizonyítékot nyújtottunk arról, hogy a PU.1 olyan régiókat is képes kötni, amelyek elméletileg nukleoszómákba vannak csomagolva. A PU.1 a saját elemén kívül gyakran kötötte az AP-1, C/EBP, RUNX és IRF elemek szoros közelében is a DNS-t, és valószínűleg kötötte a promóterspecifikus EBS-t is, bár a saját kötőhelyéhez képest kisebb gyakorisággal.

4.4. Az NFR adatok együttes használata további lehetséges regulátorok keresésére

Egy szigorúbb, mindkét hisztonacetilációs adatot ötvöző NFR készlet használata javított a korábban megtalált motívumok feldúsulásán. A H3K4me2-val is jelölt, de H3K4me3-val nem jelölt régiók még jobb motívumfeldúsulásokat adtak, és újabb motívumok is felbukkantak: egy HLH-motívum-szerű találat és a MADS fehérje-családra jellemző motívum. Szekvenciahasonlóság alapján az ismeretlen motívum MITF/TFE (MiT) fehérjecsoportra jellemző M-box lehetett. A MiT fehérjék a kizárólag egymással dimerizáló bHLH-ZIP család tagjai, amelyek a fagocita aktivitásért felelősek. A MADS család ősi DBD-je a CArG-boxot köti. A MEF2A-D fehérjék emellett a DBD mellett rendelkeznek egy további MEF2 doménnel, amely a DNS-kötésért, dimerizációért és fehérje-fehérje kölcsönhatásokért is felelős. A MEF2 fehérjéknek érdekes módon valóban szerepe van a monocita/makrofág sejtek kialakulásában.

4.5. Az RXR cisztróm vizsgálata makrofágokban

A számos más szövetspecifikus környezeti jel mellett a makrofágokat lipid molekulák is finomszabályozzák, amelyek magreceptorokon keresztül szabályozzák a génexpressziót. Makrofágokban a legjobban ismert magreceptorok az RXR-ek és heterodimerizáló partnereik, a PPAR-ek, LXR-ek és RAR-ek. Ezek vizsgálatára ChIP-seq-et hajtottunk végre egér BMDM sejtekben RXR-re, LG268 agonistájának hiányában és jelenlétében, és meghatároztuk a ChIP-seq csúcsokat. A használt statisztikai módszer 730 régióban bizonyította a DNS RXR általi elfoglalásának gyakrabba válását, míg 83 mutatott csökkenést. Emellett kaptunk egy reális méretű, 5206 csúcsból álló közös szettet, amelyet felhasználhattunk a további elemzésekhez.

A PU.1 összességében valamivel alacsonyabb lefedettséget mutatott az LG268-kezelt sejtekben, mint a kontroll makrofágokban. A legtöbb PU.1-kötött helyen enyhe növekedés volt megfigyelhető, míg a csúcsok negyede erősebb csökkentést mutatott. Az RXR aktiváció e hatásának megerősítésére bizonyos régiókban ChIP-QPCR kísérleteket végeztünk. LG268 kezelés hatására a P300 általános feldúsulása volt jellemző. A P300-kötött helyek a legtöbb esetben átfedtek az RXR és/vagy PU.1 csúcsokkal. Ezek azt sugallták, hogy a P300, mint ko-aktivátor, szorosan köthető az RXR-hez.

4.6. A makrofágok naszcens transzkriptómjának meghatározása

Mivel az RXR-aktiváció génexpresszióra gyakorolt közvetlen hatásait akartuk kimutatni, a 0, 30, 60 és 120 perces LG268 kezelést követően GRO-seq-et hajtottunk végre. Az elongációtól függetlenül minden lehetséges divergens transzkriptumot prediktáltunk, majd a H3K4me3 lefedettségi adatait használtuk a promóterek és enhanszerek elkülönítésére. A PeakSplitter által generált szálspecifikus „csúcsok” megadták a lehetőséget az összes rövid divergens és meghosszabbított transzkriptum mindkét szálon történő nagy felbontású meghatározására.

10586 gént határoztunk meg „pausing”-gal, H3K4me3 jellel és pontos annotációval, 4601 lehetséges gént H3K4me3 jellel és eddig ismeretlen TSS-sel, 8869 egyéb transzkriptumot H3K4me3 jel nélkül és 1504 alacsonyan kifejeződött gént, amely csak gyenge iniciációs jelet mutatott. Összesen 11235 ismert gént és 12821 eddig ismeretlen transzkriptumot határoztunk meg, beleértve a meghosszabbított enhanszer transzkriptumokat és egyéb hosszú nem-kódoló RNS-eket. 318 fel- és 423 leszályozott gént találtunk az LG268 kezelést követően, bár a serkentett gének sokkal nagyobb mértékben változtak.

4.7. A TF-ok RXR aktivációt követő génexpressziós változásainak vizsgálata

A változások érett mRNS szinten való követésére ugyanazokban az LG268 kezelés utáni időpontokban végeztünk RNA-seq-et, mint GRO-seq esetében. Az NFR predikcióban meghatározott motívumokat kötő TF-okat is azonosítani akartuk, ezért összegyűjtöttük a TF-ok különböző csoportjainak az expressziós szintjeit.

Az Sp1/Klf gének harmada fejeződött ki BMDM-ben. A legmagasabb expressziót a Klf6 mutatta, amelynek fehérjetermékét, mint kulcs faktort írták le proinflammatorikus makrofágokban. A Klf10 az LG268 kezelést követően érdekes módon azonnali indukciót mutatott, amelyet mind a naszcens, mind az érett RNS szintjén gátlás követett. A KLF10 a csontvelő eredetű proangiogenikus sejtekre jellemző, amelyek serkentik az érképződést.

Naszcens RNS szinten az Etv3 még összehasonlítható volt a Pu.1-gyel, de az utóbbi mRNS-e sokkal nagyobb stabilitást mutatott. Az inkább promóterspecifikus Etv6, Elf1, 2 és 4, Elk3 és Fli1 szintén magas szinten fejeződtek ki, de az mRNS szintek alapján nem volt meglepő, hogy a nagy mennyiségű PU.1 képes volt kiszorítani ezek fehérjetermékeit még a promóterekről is. Érdekes módon két indukált családtag is volt, a Fli1 és Ets2, amelyeket korábban érképzéshez kötöttek különböző sejt típusokban.

Mivel a Thap11 magasan fejeződött ki, míg a Zfp143 alacsony expressziót mutatott, eldöntöttnek látszott, hogy a rendszerünkben mi GFY. Bár a promóterspecifikus NFR-ek egyértelmű motívumfeldúsulást adtak, a Zbtb33 (Gfx) mRNS-e alacsony szintet mutatott. Az Nrfl egy kicsit magasabb, a Yy1 sokkal magasabb expressziót mutatott a Gfx-nél, amely összhangban volt a motívumfeldúsulásokkal.

A 44 AP-1/CREB-hez köthető gén közül számos mutatott magas expressziót szinte minden családból. A Jun partnerei közül az Atf3 mutatta a legmagasabb szintet, míg a Mafk, Mafk, Batf, Fos és Jdp2 gének egyre alacsonyabb expresszióval rendelkeztek. Ez utóbbi két gén RXR aktivációra indukálódott, amely a Fos gyors és átmeneti hatását jelentette, míg a Jdp2 mRNS szintjének megemelkedéséhez több idő kellett. Ennek van értelme biológiai szempontból, mivel a FOS inkább egy aktivátor a JUN felé mutatott nagy affinitással, így kizoríthatja az ATF3 és BATF represszorokat a TRE-kről, míg a JDP2 a JUN fehérjék represszora, amely aztán gátlást okozhat ugyanazokon az elemeken. Az ATF4, amelynek a kódoló génje mutatta a legmagasabb szintet az AP-1-hez köthető gének közül, CRE-ken hatva a JUN az előző modellben szereplő összes partnerével képes dimerizálni. Ez azt sugallja, hogy az ATF4 a JUN-éhoz hasonló szerepet játszhat, párhuzamosan más szabályozó helyeken. Érdekes módon az ATF4 aktiválja a Vegfa kifejeződését, így szintén indukálja az érképzést.

A MiT gének közül a Tfe3 mRNS-e volt a legmagasabb szinten feldúsulva, amely, ha nincsenek preferált homo- vagy heterodimerek, azt sugallja, hogy a TFE3 homodimerek dominálnak az M-boxokon. A Mef2 gének közül egy kevésbé ismert, a 2310045N01Rik gén mutatott mRNS szinten magas expressziót, míg a Mef2a, d és c alacsonyabban fejeződtek ki.

A génexpressziós eredményeink alapján a BMDM sejtek a GR, THRA ill. ESR1 által valószínűleg válaszképesek a glükokortikoid, tiroid és akár az ösztrogén hormonokra is. A BMDM sejtek alapvetően LXR/RXR és RAR/RXR heterodimerekkel rendelkezhetnek, de a

PPARD/RXR, THRA/RXR és GR/GR dimerek is jellemzőek lehettek a sejtekre. Az árva receptorok közül mindkét Rev-erb és mindkét Tr gén kifejeződött; és az Ear2, Esrra és Nur77 mRNS-e mutatott indukciót RXR aktivációra.

4.8. A lehetséges szabályozó régiók annotálása

Az aktív és inaktív RXR kötőhelyek aránya a divergens helyekkel való átfedés alapján 53.4 vs. 46.6 volt. A divergens helyek „eltérő kötés” elemzését végrehajtva azt találtuk, hogy az RXR kötés sokkal jellemzőbb volt az indukált enhanszer transzkripcióra (673 vs. 45 csúcs). E helyek annotálásának feltételei a következők voltak: az RXR kötést mutató enhanszer transzkriptumok expressziós változásának iránya azonos kellett legyen egy 1 MB-on belüli TSS-sel rendelkező szabályozott génével. Ezúton 387 enhanszerhez és 27 „silencer”-hez 226 fel, ill. 26 leregulált gént tudtunk rendelni.

25 LG268 által közvetlenül serkentett, érzékeléshez köthető gént találtunk, melyek közül 21-nek volt RXR csúcs a közelében. A FOS/ATF4, ETS2, KLF10 és FLI1 egymást követő indukciója szintén hozzájárulhatott a megfigyelt átmeneti indukciókhoz. A talált 21 gén között szerepelt a Vegfa és Angptl4 is.

4.9. Az annotált RXR csúcsok lehetséges szabályozó elemei

Az annotált RXR csúcsok motívumfeldúsulásai azt mutatták, hogy a csúcsok 57%-a DR1-et, DR4-et vagy más magreceptor elemet tartalmazott. A génexpressziós adatok alapján a PPARD, RARA és RARG voltak a legjobb jelöltek a DR1-kötésre, és a DR4-eket valószínűleg az LXRB/RXR vagy THRA/RXR heterodimerek kötötték. A 387 enhanszer kisebb felénél hiányoztak a magreceptor kötőhelyek: 20%-uk csak PU-boxot hordozott és a maradék részből ezek az elemek is hiányoztak.

4.10. A funkcionális domének vizsgálata

Az inzulátorspecifikus CTCF-re és a kohezin komplex részét képező RAD21-re is hajtottunk végre ChIP-seq-et. E fehérjéket nemrégiben kötötték össze funkcionálisan, mivel együttesen kötik a topológiai domének határait. Eredményeink alapján 30290 CTCF és 24648 RAD21 csúcs volt meghatározható. A 12662 CTCF/RAD21 közös csúcsból 1113 mutatott indukciót LG268 kezelésre, míg 128 mutatott csökkenést a p-értékek alapján. A 2786 divergens transzkriptumot kifejező RXR kötőhelyből 223 szignifikánsan serkentett és csak 21 gátolt volt meghatározható. 10204 lehetséges funkcionális domént határoztunk meg, és ezek egyesítése közel 700 aktív topológiai domént eredményezett. A közvetlenül szabályozott gének 80%-a helyezkedett el a szabályozó régióival közös funkcionális doménen, és ezek 33%-ánál volt indukált RAD21 lefedettség a hozzátartozó RXR-kötött enhanszer(ek)en és/vagy CTCF kötőhely(ek)en.

Néhány kiválasztott szabályozási egységre 3C-seq-et végeztünk. Ez a módszer teljes genom szinten minden interakciós (cél) helyet megad egy kiválasztott (csali) régióra, ezért egy domén két viszonylag távoli csalija jó kontrollt biztosít egymás számára. Olyan csalipárokat használtunk, amelyek a Vegfa és az Abcg1 első, és a Tgm2 gén harmadik intronjában, és a hozzájuk tartozó távoli enhanszerben helyezkedtek el az adott funkcionális doménen belül. A Vegfa lókuszra egy közel 300 kb-os hurkot jósoltunk a promóterközeli upstream és egy nagyon távoli downstream inzulátor között. A célszekvenciák alapján a 2. csali valóban mutatott kölcsönhatást a gén közelében, és a számos közös célszekvencia mellett az 1. csali megtalálta a 2. csali szomszédos restrikciós helyét is. Az Abcg1 lókuszánál egy „klasszikus génhurkot” láttunk, melyben a szabályozó elemek upstream vagy intronikusan helyezkedtek el a génhez képest. Ebben a ~100 kb-os doménben a két csali legtöbb kölcsönhatást mutató régiója közös volt és lefedte a szabályozó régiókat. A Tgm2 lókuszon az előző hurkok közül mindkét típus megvolt: egy génhurok és az „ellentéte”, amely

hasonló a Vegfa-éhoz. Az inzulátorkötés, a szabályozó régiók és a 3C-seq feldúsulások alapján úgy tűnik, hogy a Tgm2 génhurok szabályozó szerepe fontosabb a másikonál, mivel a downstream jelek sokkal gyengébbek.

5. Diskusszió

Az NGS adataink feldolgozása során az NFR predikció mellett kifejlesztettünk egy stratégiát a GRO-seq adatok feldolgozására is, amely magában foglalta a különböző naszcens RNS típusok meghatározását és annotációját. Mivel a rövid divergens és meghosszabbított transzkriptumok átfedtek, elkülönítettük ezeket a polimeráz aktivitásokat. Így több mint 50000 aktív szabályozó helyet, 11235 ismert gént és 12831 még ismeretlen transzkriptumot tudtunk kimutatni, ami azt jelenti, hogy az egér genom ötöde mutatott transzkripciót legalább az egyik szálon. Az RNA-seq adatokkal együtt, az RXR aktivációt követően megfigyelhettük a különböző gének mRNS érésének dinamikáját, és pl. THAP11-ként azonosíthattuk a GFY-t a kísérleti rendszerünkben. Kifejlesztettünk egy további stratégiát az inzulátorok által határolt kromatindomének meghatározására, és egy másikat a 3C-seq adatok elemzésére, hogy validálhassunk néhány domént. Ezek a módszerek segítettek feltérképezni a naszcens és érett transzkriptómot és a makrofágok szabályozási hálózatát, figyelembe véve az RXR-specifikus funkciókat is.

A TFII komplex általános TF-ai és azok, amelyek a CpG-gazdag promótereket kötik, szükségesek minden egyes sejt kialakulásához és fenntartásához. Az egyedi sejtípusok differenciációját a TF-ok egy másik, speciális csoportja, a „mester regulátorok” irányítják. Ezek közül néhányan, az úgynevezett „úttörő faktorok” képesek fellazítani a zárt kromatint. A „forkhead” (FOX) fehérjék a linker hisztonok kiszorításával képesek felnyitni a DNS-t, a PU.1 pedig hasonlóképpen kiszabadíthatja a nukleosómák közötti összekötő DNS-t. Makrofágokban a PU.1 fő, egyben dimerizáló partnere az IRF8, de a C/EBPA is hasonlóan

meghatározó, mivel megemelkedett expressziója pre-B sejtekből makrofágokat tudott létrehozni. Az AP-1 fehérjék szintén képesek a PU.1-hez csatlakozni makrofágokban, de mivel ennek a fehérjecsaládnak számos tagja van, nehéz megmondani, hogy pl. mi a JUN partnere ezekben a komplexekben. A RUNX1, amely a makrofág differenciáció során játszik szerepet, szintén jelen van az érett sejtekben is. A makrofágok további TF-ai valószínűleg specializálódtak a különböző funkciókhoz, az általános vagy a környezeti jeleket felsimerő útvonalakhoz.

A TF-ok ko-regulátorokkal együttműködve hozzák létre a szükséges kromatin-környezetet, hurkolják a DNS-t és odagyűjtik a PIC-et a promóterekre. E kromatinkörnyezet a promóterek és enhanszerek különböző epigenetikai módosításait jelenti. Ennek vizsgálatára kifejlesztettünk egy módszert az összes, aktív hiszton jelekkel körülvett szabályozó régió meghatározására, és szoros összefüggést találtunk a promóter- (H3K4me3) és enhanszer-specifikus (H3K4me2, H4ac, H3K27ac) módosítások között. Az enhanszerspecifikus módosításokra meghatározott leggyakoribb NFR-hossz 125 és 150 bp közé esett, amely nagyjából megegyezik az egy nukleoszómamagra tekeredett DNS hosszával. Ez azt mutatja, hogy használható paramétereket sikerült beállítani a predikcióhoz. A legaktívabb NFR-ek promótereknek látszottak, mivel jellemzően szélesek és minden vizsgált módosítással jelöltek voltak, beleértve a H3K4me3-t is. A H3K4me2 és me3 együttes jelenléte nem volt meglepő, mivel ugyanazt a fehérjét képviselik különböző metilációs fokkal. Az enhanszerjelek jelenléte ellenben meglepő volt a promótereken, mivel ezek erőssége hasonlóan magas volt, mint azokban a régiókon, ahol a H3K4me3-lefedettség sokkal kisebb volt. Az „aktív enhanszer” jelek így minden aktív szabályozó régióra jellemzőek voltak, ezáltal a H3K4me3 vált a legjobb hisztonjellé, amely képes volt elkülöníteni a promótereket az enhanszerektől.

A motívumfeldúsulási elemzések szintén megerősítették az előbbi eredményeket, és a régiók legaktívabb enhanszerekre való szűkítésével néhány nem várt motívumot is

megkaptunk. Az ismert sejtvonal-meghatározó faktorok motívumai mellett kijöttek a ténylegesen makrofágokhoz köthető MiT és MEF2 fehérjecsaládok motívumai is. A génexpressziós adatok alapján valószínűleg a TFE3 dominált az M-boxokon, és a MEF2D mellett egy nem jellemzett MEF2 fehérje kötötte a MADS-boxokat. A TRE/CRE-kötő fehérjék esetében nem volt könnyű meghatározni a különböző heterodimereket, mivel mind a 16 AP-1/CREB családot képviselte legalább egy tagja. A JUN számos partnere kifejeződött, és úgy tűnik, hogy az ATF4 ugyanazokkal a heterodimerizáló partnerekkel foglalta el a CRE-ket, mint a JUN a TRE-ket.

A Pu.1 kiemelkedően magas mRNS szintet mutatott, ami megmagyarázza a nagyszámú elfoglalt EBS-t, beleértve a PU-boxokat és a promóterre jellemző ETS elemeket. A PU.1-kötött helyek fele inaktív szabályozó régióknak látszott, amelyekről hiányoznak a jellegzetes hiszton mintázatok. Érdekes módon a másik fél átlagosan ugyanazt a mintázatot mutatta H4ac-ra, H3K27ac-ra és H3K4me2-ra is, ami azt sugallta, hogy bármely hisztonjel jelenléte elkülönítheti az aktív szabályozó régiókat az inaktívaktól. Ezek a PU.1 csúcsok jelentősen nagyobbak voltak az inaktívaknál, amely azt mutatja, hogy az aktív szabályozó régiókban nagyobb a kötési gyakorisága. Az „együttes kötés” elemzés azt mutatta, hogy a PU.1 együttműködik mind a promóterekre, mind a makrofágokra jellemző TF-okkal. A promóterspecifikus EBS-eken tapasztalt, a PU-boxokon tapasztaltakéhoz viszonyítva kisebb PU.1 feldúsulás az elemek felé mutatott különböző affinitásnak vagy az egyéb ETS fehérjékkel való versengésnek tulajdonítható. De mivel az összes fontosabb promóter-specifikus elem mutatta a PU.1 közelségét, jelenléte egyértelmű volt.

A magreceptor motívumok NFR-ekben való feldúsulásának hiánya valószínűleg technikai hiányosságokból fakadt, mivel a magreceptorok szerepe ismert makrofágokban, azonban jelezte, hogy ezek a TF-ok nem határozzák meg a sejtípust, csak finomhangoló szerepük van a BMDM sejtekben. Az Rxra és Rxrb mellett az Lxrb-t találtuk magasan

expresszálnak, és a II. osztályból a Ppard, Rara/g és Thra mutatott még magasabb kifejeződést. A többi osztályból a Gr, Esrra és Ear2 mutatott említésreméltó expressziós szintet. Ez azt jelenti, hogy az RXR heterodimerek vannak túlsúlyban a többi magreceptorral szemben; és valóban, az RXR több mint 5000 régiót kötött, és kötési gyakorisága jellemzően emelkedett az LG268 agonistájával történő kezelést követően. Ez a feldúsulás úgy tűnt, aktiváló szereppel párosul, mivel a P300 ko-aktivátor is követte a feldúsulást. Mi több, a P300-kötött régióknak több mint 2/3-a fedett át RXR kötéssel, ami egy szorosabb, akár közvetlen kölcsönhatást is mutathat a fehérjék között. Ezzel ellentétben a PU.1 inkább alacsonyabb kötési gyakoriságot mutatott az RXR aktiváció után.

Mivel mind enhanszer-, mind géntranszkripció szinten megmutatja a közvetlen szabályozó hatásokat, a GRO-seq volt a legalkalmasabb módszer az LG268 kezelés utáni expressziós változások követésére. A naszcens RNS expresszió alapján közel kétszer több divergens hely volt szignifikánsan indukált, mint gátolt, és az RXR-kötött régiókra ez az arány közel volt a 15-höz, amely 673 közvetlenül serkentett szabályozó helyet jelentett. Meglepő módon a szignifikánsan gátolt gének száma (423) meghaladta az indukált gének számát (318), bár az indukció átlagos mértéke ~2.8-szor volt nagyobb a gátlásénál. Végül közelség alapján 387 enhanszert és 27 silencer-t rendeltünk 226 fel ill. 26 leregulált génhez. Ezután az inzulátorspecifikus CTCF/RAD21 átfedő csúcsokat használva funkcionális doméneket prediktáltunk, amelyek lefedték a szabályozott gének 80%-át, tovább erősítve az annotáció használhatóságát. A RAD21 érdekes módon ko-aktivátorként viselkedett, megjelölve az RXR-kötött régiókat és indukciót mutatva LG268 kezelés hatására. Azt figyeltük meg összességében, hogy az RXR aktiváció közvetlenül (inkább serkentőleg) hatott több mint 200 génre, és valószínűleg közvetve szabályozott (főként gátolt) további 500 gént.

A 387 RXR-kötött enhanszer több mint fele bírt magreceptor kötőhellyel: a DR1 volt a leggyakoribb, amelyet a PPARD/RXR és RAR/RXR heterodimerek köthettek, ha kizártuk a

COUP-TFG, TR2/4 és RXRA/B dimereket, míg a DR4 elemekről az LXRβ/RXR valószínűleg kiszorította a THRA/RXR heterodimereket. Nehéz megmondani, hogy az RXR mellett melyik magreceptorok köthették a további fél és kompozit elemeket, de biztosnak látszott, hogy a magreceptor elemek nélküli helyeket az RXR közvetve, pl. PU.1-en, C/EBP-n, AP-1-en, RUNX1-en vagy más fehérjéken keresztül kötötte. Mindemellett az RXR jelenléte indukált gének közelében a LG268 által indukálható transzkripciót mutató enhanszerek azt jelezte, hogy az RXR-nek a partnereitől független szerepe is lehet, bár volt átfedés a célgénjeik között, pl. a Tgm2 egy RAR/LXR célgén, az Angptl4 egy PPAR célgén, míg az Abca1 és Abcg1 LXR célgénnek.

Az RXR érdekes módon számos különböző típusú, érképzéshez köthető TF-t is indukált. Egyéb feladataik mellett a KLF10, FLI1, ETS2 és ATF4 angiogenikus szereppel is bírnak. Az ATF4 nem, de heterodimerizáló partnerei mutattak indukciót. Az expressziós eredmények alapján úgy tűnt, hogy az ATF4-nek a JUN-(oké)hoz hasonló központi szerepe van, mivel ezek a fehérjék nem kölcsönhatnak egymással, mindkettő magas expressziót mutatott és számos közös partnerük is kifejeződött. Az LG268 kezelés által a Fos – amelynek a foszforilált fehérjeterméke inkább aktivátor – és a Jdp2 – amelynek a fehérjeterméke inkább represszor – egyszerre indukálódtak. Mivel a Jdp2 sokkal hosszabb gén, a hatása időben eltolódott, ami egy átmeneti FOS/JUN és FOS/ATF4 hatást okozhatott.

Valóban számos angiogenikus gén indukálódott az RXR aktivációt követően. Nemcsak az ATF4 célgén Vegfa, az RAR/LXR célgén Tgm2 és a PPAR célgén Angptl4, hanem a Hbegf, a hem oxigenáz 1 (Hmox1), a trombospondin receptor (Cd36) és számos egyéb gén, amely szerepet játszik az erek képződésében. Ez egy az RXR és downstream regulátorai által vezérelt összehangolt együttműködésnek látszott. A legtöbbet vizsgált és talán legfontosabb angiogenikus gén, a Vegfa látszólag egy szokatlanul távoli, RXR által kötött és valószínűleg RXR által szabályozott enhanszercsoporttal rendelkezett. Mind az

enhanszertranszkripció serkentése, mind a közel 300 kb domén azt mutatta, hogy megtaláltuk a Vegfa legfőbb szabályozó helyeit egér BMDM sejtekben, ezért 3C-seq kísérletekkel is megerősítettük ezt az eredményt, melyek valóban kölcsönhatást mutattak a gén és a felfedezett downstream régió között. A Tgm2 nem mutatott ilyen szélsőséges távolságokat, csak egy egyszerű génhurkot, melyben a szabályozó helyek 3 dimenziós szerkezetben hasonló méretű DNS „ujjakat” rögzíthettek.

6. Kulcsszavak

Makrofág, transzkripciós faktorok, nukleoszóma-mentes régió, PU.1, RXR, érképződés

7. Összefoglalás

A BMDM sejtek transzkripció szabályozásának vizsgálatára számos NGS módszert használtunk, beleértve a ChIP-seq-et, GRO-seq-et, RNA-seq-et és 3C-seq-et. Mivel nagy mennyiségű eszköz van, de az elemzések jelentős részéhez még mindig nincs széleskörben használt standard, az adatfeldolgozáshoz új megközelítésekre és új stratégiák kidolgozására volt szükség a többé-kevésbé egyedi kérdések megválaszolására. A GRO-seq szerteágazó információt biztosított, amely egy összetett rendszert igényelt a transzkriptumok meghatározására, annotálására és az expressziós szintek kiszámítására. A promóterek enhanszerektől való, tehát a gének enhanszer transzkriptumoktól való elválasztásához beépítettük a műveletek közé a H3K4me3 ChIP-seq-ből származó adatokat is. Új stratégiát alkalmaztunk az NFR- és doménpredikciókra ChIP-seq adatokból, és a 3C-seq elemzéshez is.

A makrofágokban működő főbb TF családokat a prediktált NFR-ek motívumfeldúsulásai alapján határoztuk meg, de az egyes TF-ok azonosítása génexpressziós adatokat igényelt. Érdekeltek az RXR ligandkezelés közvetlen hatásai, és azt találtuk, hogy az RXR – a heterodimerizáló partnereivel vagy azok nélkül – inkább aktivátorként működött. P300 és RAD21 feldúsulást mutatott, melyek előkészíthetik a kromatinkörnyezetet a génindukcióra. Számos angiogenikus gént (pl. Vegfa, Angptl4, Tgm2 és Hbegf) és szintén érképzésben érintett TF-t (ATF4, ETS2, KLF10 and FLI1) indukált. A FOS és JDP2, a JUN-ok és ATF4 heterodimerizáló partnerei szintén szabályozódtak, amely egy átmeneti indukciót okozhatott. A legfontosabb angiogenikus génhez, a Vegfa-hoz egy szokatlanul távoli enhanszercsoportot találtunk, amely egy 300 kb-os hurkon keresztül közelíti meg a gént, mely huroknak a végei valóban kölcsönhatást mutattak a 3C-seq adatok alapján.



Nyilvántartási szám: DEENK/67/2016.PL
Tárgy: PhD Publikációs Lista

Jelölt: Nagy Gergely
Neptun kód: AOKH5U
Doktori Iskola: Molekuláris Sejt- és Immunbiológiai Doktori Iskola

A PhD értekezés alapjául szolgáló közlemények

1. Dániel, B., **Nagy, G.**, Hah, N., Horváth, A., Czimmerer, Z., Pólska, S., Gyuris, T., Keirsse, J., Gysemans, C., Van Ginderachter, J.A., Bálint, B.L., Evans, R.M., Barta, E., Nagy, L.: The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages.
Genes Dev. 28 (14), 1562-1577, 2014.
DOI: <http://dx.doi.org/10.1101/gad.242685.114>
IF:10.798
2. **Nagy, G.**, Dániel, B., Jonás, D., Nagy, L., Barta, E.: A novel method to predict regulatory regions based on histone mark landscapes in macrophages.
Immunobiology. 218 (11), 1416-1427, 2013.
DOI: <http://dx.doi.org/10.1016/j.imbio.2013.07.006>
IF:3.18





További Közlemények

3. Cuaranta-Monroy, I., Simándi, Z., Kolostyák, Z., Doan-Xuan, Q., Póliska, S., Horváth, A., **Nagy, G.**, Bacsó, Z., Nagy, L.: Highly efficient differentiation of embryonic stem cells into adipocytes by ascorbic acid.
Stem Cell Res. 13 (1), 88-97, 2014.
DOI: <http://dx.doi.org/10.1016/j.scr.2014.04.015>
IF:3.693
4. Dániel, B., **Nagy, G.**, Nagy, L.: The intriguing complexities of mammalian gene regulation: How to link enhancers to regulated genes. Are we there yet?
FEBS Lett. 588 (15), 2379-2391, 2014.
DOI: <http://dx.doi.org/10.1016/j.febslet.2014.05.041>
IF:3.169
5. Széles, L., Póliska, S., **Nagy, G.**, Szatmári, I., Szántó, A., Pap, A., Lindstedt, M., Santegoets, S.J.A.M., Rühl, R., Dezső, B., Nagy, L.: Research resource: Transcriptome profiling of genes regulated by RXR and its permissive and nonpermissive partners in differentiating monocyte-derived dendritic cells.
Mol. Endocrinol. 24 (11), 2218-2231, 2010.
DOI: <http://dx.doi.org/10.1210/me.2010-0215>
IF:4.889

A közlő folyóiratok összesített impakt faktora: 25,729

A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre): 13,978

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

Debrecen, 2016.03.18.



Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Dr. Barta Endrének, hogy bevezetett a bioinformatika világába.

Ez a tézis nem jöhetett volna létre a korábbi témavezetőm és jelenlegi főnököm, Prof. Nagy László nélkül.

Szeretném megköszönni Dr. Dániel Bencének, hogy óriási mennyiségű és jó minőségű adatot biztosított a munkámhoz.

Hálás vagyok Prof. Fésüs Lászlónak és Prof. Tózsér Józsefnek, a Biokémiai és Molekuláris Biológiai Intézet korábbi és jelenlegi vezetőjének a lehetőségért, hogy egy jól felszerelt környezetben dolgozhattam.

Hálás vagyok a Magreceptor-kutató Csoport minden korábbi és jelenlegi munkatársának a segítségükért és a tudományos megbeszélésekért.

Külön köszönet Dr. Póliska Szilárdnak, Horváth Attilának és Jónás Dávidnak.

Szeretném kifejezni a hálámat Dórinak és a családomnak a támogatásukért és türelmükért.