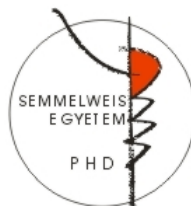


Cas9 nukleázok működési hatékonyságának kísérletes meghatározása és predikciója

Doktori tézisek

Tálas András

Semmelweis Egyetem
Molekuláris Orvostudományok Doktori Iskola



Témavezető: Dr. Welker Ervin, DSc., tudományos tanácsadó

Hivatalos bírálók: Dr. Geiszt Miklós, DSc., egyetemi tanár
Dr. Varga Máté, PhD., egyetemi docens

Szigorlati bizottság elnöke: Dr. Csala Miklós, DSc, egyetemi tanár
Szigorlati bizottság tagjai: Dr. Váradi András, DSc., prof. emeritus
Dr. Rónai Zsolt, PhD, egyetemi docens

Budapest
2021

Bevezetés

A CRISPR/Cas9 rendszer célzott génmódosításra való használata forradalmasította a biológia szinte minden ágát az utóbbi 8 évben, és lehetőséget teremtett öröklődő betegségek gyógyítására is. A Cas9 fehérjét baktériumokban felfedezték fel, amelyről bizonyították, hogy a baktériumok vírusok elleni védekezésében játszik szerepet. A fehérjét később sikeresen használták fel sejtek és élőlények géneinek szerkesztésére is.

A génmódosításra legelterjedtebben használt *Streptococcus pyogenes* Cas9 fehérje (SpCas9) egy irányító RNS-sel (gRNS-sel) komplexben szekvencia-specifikusan képes DNS szekvenciákat (targeteket) felismerni és a felismerés helyén kettős szálú DNS törést létrehozni. A specificitásért a gRNS 5' végén található 20 bázispár felel, melyet spacernek nevezünk. A célszekvencia felismerésének további előfeltétele, hogy a target mellett egy úgynevezett PAM konszenzus szekvenciának kell jelen lennie a DNS szálon. Amennyiben a célzott DNS szakasszal homológ a spacer szekvencia, valamint a PAM is megtalálható a DNS-en a célzott szekvencia mellett, a fehérje a DNS mindkét szálát elvághatja, kettősszálú DNS törést hoz létre. A DNS töréseket követő hibajavítást felhasználva lehet géneket kiütni, vagy bevinni a genomba.

A technológia egyik korlátja volt, hogy néhány esetben a fehérje nem csak a célzott szekvenciával megegyező, hanem néhány bázisban különböző szekvenciákat is hasíthat. Ennek

csökkentésére létrehoztak megnövelt pontosságú SpCas9 variánsokat - mint például az SpCas9-HF1 -, amelyek meglehetősen csökkentik ezt a hatást. Egy másik korlátja a rendszernek, hogy az SpCas9 fehérje az eltérő szekvenciájú targeteket eltérő hatékonysággal hasítja, viszont ennek mibenléte nem igazán ismert. Valamint azt is megfigyelték, hogy a különböző SpCas9 variánsok, mint például az előbb említett SpCas9-HF1, szekvencia-specifitása eltérő a vad típusú SpCas9-hez képest. Munkám során olyan módszereket fejlesztettem ki, amelyekkel kísérletesen meghatározható vagy *in-silico* megjósolható, hogy egy adott SpCas9 variáns mely gRNS-ek segítségével képes jól vagy kevésbé jól működni.

Célkitűzések

Disszertációmban SpCas9 variánsok működési hatékonyságát két féle megközelítéssel szándékoztam meghatározni. A kísérletes megközelítésre egy fluoreszcens módszert kívántam kidolgozni, amellyel elő-tesztelhető a jelölt gRNS-ek működési hatékonysága. A módszert úgy kívántam kifejleszteni, hogy gRNS-ek hatékonysága genomi környezetben, egyszerűen áramlási citometriával vizsgálható legyen. Az SpCas9 által létrehozott DNS kettős törések helyére egy zöld fluoreszcens fehérjét kifejező plazmidot szándékoztam hatékonyan bejuttatni. A törések hatékonysága a beépült plazmidok mennyiségével kell, hogy arányos legyen, így a hatékonyabb és a kevésbé hatékony gRNS-ek kísérletesen kiválaszthatók egy egyszerű fluoreszcens módszer segítségével.

A másik céloom az volt, hogy az elterjedten használt megnövelt pontosságú SpCas9-HF1 variánsra kifejlesszek egy predikációs eljárást, amellyel ezt a fehérjét is rutinszerűen lehetne használni. Ezt a célt egy minden korábbinál nagyobb, kb. 1 millió különböző gRNS-t tartalmazó könyvtár használatával, valamint egy bakteriális szelekciós rendszer kidolgozásával kívántam elérni, amelyben nagy számú szekvenciáról eldönthető, hogy melyek működnek jól, valamint kevésbé hatékonyan.

Módszerek

Expressziós plazmidok létrehozása molekuláris klónozással. Az önvágó plazmidokat NEB HiFi Assembly segítségével hoztam létre. A genomi szekvenciákat célzó Cas9-gRNS plazmidokat a px330 plazmidból állítottam elő egy egy lépéses restriktív enzimmel történő hasítással és a spaceret tartalmazó oligonukleotidok ligálásával (one-pot klónozás). A GFxFP kísérletekhez a célszekvenciákat a pWN100042 plazmidba one-pot klónozással vittem be. A GFxFP plazmidot célzó spaceret a pmCherry-gRNA plazmidba klónoztam one-pot klónozással.

stgRNS könyvtárak létrehozása. Az stgRNS könyvtárak létrehozásához először egy klónozó plazmidot terveztem, ezt megszintetizáltattuk, majd PCR-rel amplifikáltam. A véletlenszerű szekvenciákat tartalmazó oligonukleotidból egy egy lépéses PCR-rel készítettem kettőszálú inszertet, amelyet NEB HiFi Assembly segítségével klónoztam a vektorba. Egy több mint egymillió szekvenciát tartalmazó könyvtárat hoztam

létre (1M) könyvtár, amelyet a predikciós algoritmus tanító adathalmazaként használtunk, valamint egy 100 000 spacert tartalmazó könyvtárat (100K), amelyet teszt adathalmazként használtunk.

A könyvtárak szelekciója és újgenerációs szekvenálás. Az elkészített könyvtárakat az általam készített WT-SpCas9-et és SpCas9-HF-et expresszáló kompetens *E. coli* baktériumokba transzformáltam 50x-es transzformációs lefedettséggel. A transzformált baktériumkolóniákat egy három dimenziós agaróz-LB médiumban növesztettem antibiotikum szelekció mellett. A szelekció után a túlélő baktériumokból plazmidokat izoláltam és újgenerációs szekvenálás segítségével megszekvenáltam. Az újgenerációs szekvenálás eredményeinek kiértékelését Dr. Tusnády Gábor, Varga Júlia, Welker Zsombor, Erdős Gergely, Dr. Pach Péter és Dr. Pongor Lőrinc végezte.

Predikciók kifejlesztése. Dr. Györgypál Zoltán segítségével az 1M könyvtár hasításának adataiból létrehoztunk egy rövid szekvencia-motívumokon alapuló predikciós eljárást, a Hi-CRISPR A-t. A neurális hálózaton alapuló Hi-CRISPR B és C algoritmusokat Welker Ágnes készítette. A predikciók webes felületét Welker Zsombor készítette, amely ezen a linken érhető el: <https://hicrispr.welkergroup.hu>

Sejtenyésztés és transzfekció. Munkáim során az ATTC-től vásárolt N2a, HeLa, NIH/3T3 és HEK293T sejtvonalakat használtam és az ATTC előírásainak megfelelően tenyésztettem.

A transzfekeciót sejtvonalként különböző protokollok alapján Turbofect reagenssel végeztem.

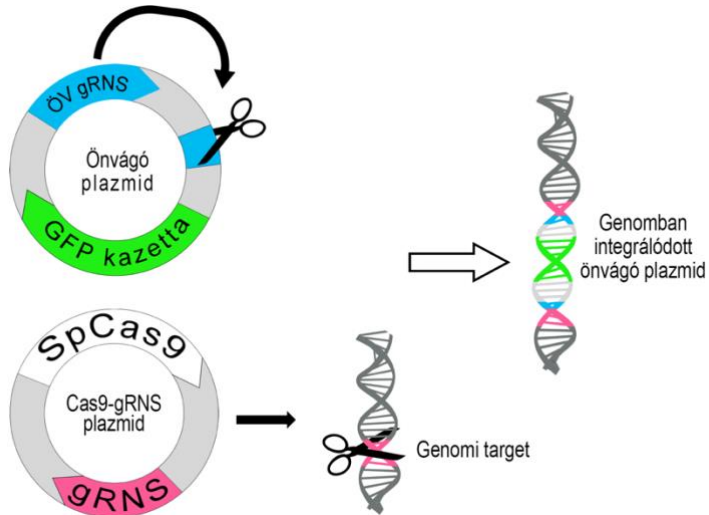
Áramlási citometria. Az áramlási citometriás méréseket egy Attune NxT citométerrel végeztem. A GFP-t a 488 nm-es lézerrel gerjesztettem és az 530/30nm-es szűrővel detektáltam. Az mCherry-t az 561 nm-es lézerrel gerjesztettem és az 615/20 nm-es szűrővel detektáltam.

Eredmények

gRNS-ek hatékonyságának meghatározása önvágó plazmidokkal

A gRNS-ek hatékonyságát kísérletesen úgy szándékoztam meghatározni, hogy a genomot az SpCas9 fehérjével hasítom, és az NHEJ hibajavítás segítségével a törés helyére egy zöld fluoreszcens fehérjét kódoló plazmidot beépítek. A donor plazmid integrációja nem volt hatékony az SpCas9 által létrehozott kettős törések helyére, amennyiben az cirkuláris volt, vagy egy restrikiós enzimmel a transzfekeciót megelőzően linearizáltam. Az integráció hatékonyságát úgy sikerült megnövelni, hogy a cirkulárisan transzfekektált plazmidot SpCas9 segítségével intracellulárisan linearizáltam. A plazmidok sejten belüli hasítása jóval nagyobb célzott integrációt eredményezett a véletlenszerű integrációt jelző kontroll kondícióhoz képest. Létrehoztam egy univerzálisan használható úgynevezett önvágó plazmidot, amely az önmaga linearizálásáért felelős gRNS

expressziós kazettát, valamint a célszekvenciát is hordozza (1. ábra).



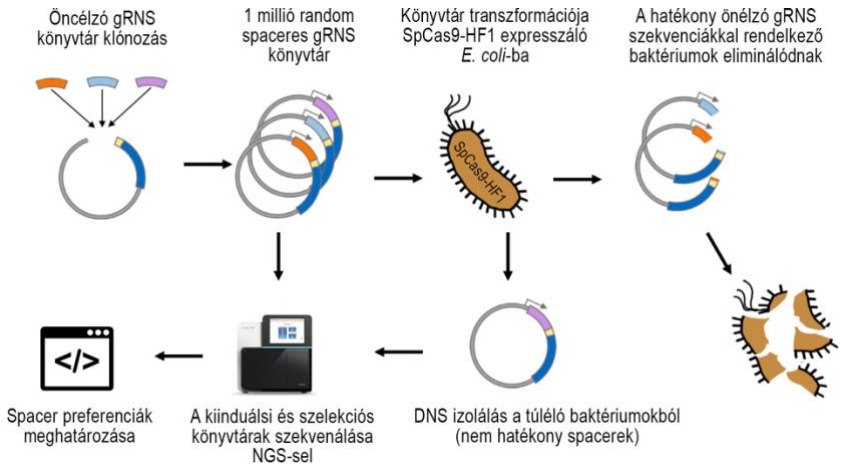
1. ábra. Az önvágó plazmid integrációja a genomban. Az SpCas9 és a genomot célzó gRNS (rózsaszín) a genomi szekvenciát hasítja, valamint az önvágó plazmid az önvágó gRNS (kék) és az SpCas9 segítségével a sejtben belül linearizálódik, és a genomban integrálódik.

Az önvágó plazmidok segítségével sikeresen lehetett detektálni a hatékonyabb és kevésbé hatékony gRNS-eket N2a, NIH/3T3, HeLa és HEK293T sejtekben. A módszerünket két PCR alapú módszerhez is hasonlítottuk, valamint a kezdeti 12-15 napos detekciós időt sikerült lecsökkentenünk 6 napra egy GFP-hez fuzionált degradációs domén használatával. A módszert NHEJ mediált génbevitelre és sejtvonala alapításra is felhasználtuk.

Az SpCas9-HF1 aktivitásának meghatározása egymillió célszekvencián

Disszertációm második felében egy gyakran használt megnövelt pontosságú SpCas9 variáns az SpCas9-HF1 működésének hatékonyságát szerettem volna meghatározni egy predikciós eljárás kifejlesztésével, hiszen a célkitűzések megfogalmazásakor még nem létezett erre a fehérjére ilyen módszer. Az SpCas9 predikciós eljárások kifejlesztésekor a megfelelő mennyiségű SpCas9 hasítási információ létrehozása szokta jelenteni a szűk keresztmetszet. A jelenlegi technológiával pár tízezer féle gRNS-t tartalmazó könyvtár hozható létre, mert a gRNS spacer régióját és annak célszekvenciáját is egy DNS-re kell juttatni, majd a hasítást detektálni. Mi egy sokkal nagyobb körülbelül egymillió féle gRNS-t tartalmazó könyvtárat akartunk létrehozni, amelynek kivitelezésére sikerült egy megoldást találnunk.

Egy úgynevezett öncélzó gRNS-t hoztam létre, amely spacer szekvenciája és célszekvenciája ugyanaz, mivel a spacer mellé egy PAM szekvenciát is bevittünk. Egy öncélzó gRNS-t tartalmazó plazmidot egy sejtbe bejuttatva a gRNS a saját kódoló szekvenciáját célozza így „öngyilkos” módon a plazmidot elvágja. Ilyen öncélzó gRNS-ekből álló egymillió különböző véletlenszerű szekvenciát tartalmazó könyvtárat hoztam létre (**2. ábra**). A könyvtárat olyan *E. coli* baktériumokba transzformáltam, amelyek az SpCas9-HF1 vagy a vad típusú WT-SpCas9 fehérjét expresszálták.



2. ábra. Az SpCas9-HF1 spacer preferenciáinak meghatározása egy szelektív módszerrel

Azok a szekvenciák, amelyek hatékonyan működnek elvágják a saját plazmidjaikat, így a plazmidon lévő antibiotikum rezisztenciát is elvesztik, végül eliminálódnak. Ezzel ellentétben a nem hatékony spacerekkel rendelkező szekvenciák feldúsulnak. A túlélő sejtekből DNS-t izoláltam és a plazmidokat újgenerációs szekvenálással megszekvenáltuk. A kiindulási könyvtárat is megszekvenálva összehasonlítottuk a szelekcióval, így meghatároztuk a két fehérjével jól és kevésbé jól működő szekvenciákat. Az adatok elemzése kimutatott olyan szekvencia motívumokat, melyek mindkét fehérje működését elrontják, valamint olyanokat is, amelyek csak az SpCas9-HF1 működését rontják el. Az 1 millió szekvencia hasítási adataiból 3 különböző módszeren alapuló SpCas9-HF1 predikciós módszert hoztunk létre amelyeket Hi-CRISPR A, B és C-nek nevezünk el. A

predikciók a bakteriális rendszerben létrehozott teszt adatokon kiválóan működtek.

A kísérletek kiértékelése közben megjelent egy SpCas9-HF1 predikciós eljárás DeepHF néven, amely 50000 spaceren alapszik és amelyet emlős sejtekben hoztak létre. Addig nagyon kevés adat segíthette a predikciónk értékelését, így ezeket az adatokat is elemeztük a bakteriális rendszerben kapott információk alapján. Megvizsgáltuk, hogy a korábban bakteriális sejtekben azonosított, az SpCas9-HF1 aktivitását legjobban befolyásoló szekvencia motívumok ugyanúgy megtalálhatóak-e ebben az emlős sejtekben készített adatszettben. Azt találtuk, hogy ebben a függetlenül létrehozott adatszettben is megtalálhatók a csak az SpCas9-HF1-et elrontó, valamint a mindkét fehérje működését elrontó motívumok. Végül, az általunk kifejlesztett 1 millió bakteriális szekvencián tanított Hi-CRISPR predikcióinkat szerettük volna összehasonlítani a DeepHF predikcióval. 5 kisebb adatkészletet gyűjtöttünk össze a szakirodalomból, valamint a DeepHF algoritmus kidolgozásához használt 50 000 spacer tartalmazó adathalmazt használtuk a predikciók összehasonlítására. A négy előrejelzési algoritmus hasonlóan jó teljesítményt mutatott ezeken az adathalmazokon.

Következtetések

A CRISPR/Cas9 rendszerek genommodosításra való használata a modern molekuláris biológia egyik forradalmi újításának tekinthető, hiszen a korábbi módszerekhez képest sokkal

könnyebben és olcsóbban használható gének módosítására. A módszer korlátai közé tartozik, hogy a Cas9 fehérje a különböző célszekvenciákat különböző hatékonyságokkal hasítja, így ennek megismerése kulcsfontosságú az összes alkalmazás esetén.

Az egyik megközelítés, hogy a jelölt gRNS-eket kísérletesen teszteljük, majd a legjobbnak bizonyuló célszekvenciákkal végezzük el az adott génmódosítási feladatot. Itt egy fluoreszcens detekción alapuló módszert dolgoztam ki, amelyben a gRNS-ek hatékonysága genomi környezetben egyszerűen vizsgálható, nem igényel további klónozási lépéseket, valamint PCR amplifikációt és szekvenálást sem. Az SpCas9 és a különböző gRNS-ek által létrehozott DNS törések helyére egy GFP expressziós kazettát építettem be és a fluoreszcens sejtek követésével az adott gRNS hatékonyságát tudtam vizsgálni. A hatékony integrációt a plazmid sejten belüli felnyitásával értem el, majd létrehoztam egy önvágó plazmidot, amely tartalmazza az önmaga linearizálásért felelős gRNS-t, így univerzálisan használható bármely jelölt gRNS-sel. Az önvágó plazmiddal számos sejtvonalban meg tudtam határozni gRNS-ek aktivitását és a megközelítést már létező módszerekkel validáltam. A rendszert a kísérletes tesztelésen túl, NHEJ-mediált génbevitelre is felhasználtam.

Csak néhány Cas9 variáns esetében elérhetők olyan predikciós algoritmusok, amelyek változó hatékonysággal, de képesek előre jelezni a gRNS-ek aktivitásának várható mértékét. A különböző Cas9 ortológok és variánsok különböző szekvencia preferenciákkal rendelkeznek, így ezek rutinszerű használatához

szükség lenne egy predikciós eljárásra. A problémára egy olyan módszert fejlesztettem ki, amellyel nagyszámú Cas9 hasítási információ egyszerűen hozható létre és az így generált adatok alapján létrehozható egy predikciós algoritmus. A módszerrel a gyakran használt megnövelt pontosságú Cas9 variáns az SpCas9-HF1 szekvencia preferenciáit határoztuk meg, majd egy megbízható predikciós algoritmust fejlesztettünk ki, amelyet számos adatszenzen validáltunk.

Saját publikációk

Disszertációhoz kapcsolódó publikációk:

1. **Tálas, A.**, Huszár, K., Kulcsár, P.I., Varga, J.K., Varga, É., Tóth, E., Welker, Z., Erdős, G., Pach, P.F., Welker, Á., Györgypál, Z., Tusnády, E.G., Welker, E. (2021) A method for characterizing Cas9 variants via a one-million target sequence library of self-targeting sgRNAs. *Nucleic Acids Res.*, gkaa1220 (IF: 11,501)

2. **Tálas, A.**, Kulcsár, P.I., Weinhardt, N., Borsy, A., Tóth, E., Szabéniyi, K., Krausz, S.L., Huszár, K., Vida, I., Sturm, Á., Gordos, B., Hoffmann, O.I, Bencsura, P., Nyeste, A., Ligeti, Z., Fodor, E., Welker, E. (2017) A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a ‘self-cleaving’ GFP-expression plasmid. *DNA Res*, 24, 609-621. (IF: 5,415)

Disszertációtól független publikációk

3. Kulcsár, P.I., **Tálas, A.**, Tóth, E., Nyeste, A., Ligeti, Z., Welker, Z., Welker, E. (2020) Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5' G-extended sgRNAs. *Nat Commun*, 11, 1-14. (IF: 12,121)
4. Tóth, E., Varga, É., Kulcsár, P.I., Kocsis-Jutka, V., Krausz, S.L., Nyeste, A., Welker, Z., Huszár, K., Ligeti, Z., **Tálas, A.**, Welker, E. (2020) Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. *Nucleic Acids Res*, 48, 3722-3733. (IF: 11,501)
5. Tóth, E., Czene, B.C., Kulcsár, P.I., Krausz, S.L., **Tálas, A.**, Nyeste, A., Varga, É., Huszár, K., Weinhardt, N., Ligeti, Z., Borsy, A.É., Fodor, E., Welker, E. (2018) Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic Acids Res*, 46, 10272-10285. (IF: 11,147)
6. Kulcsár, P.I., **Tálas, A.**, Huszár, K., Ligeti, Z., Tóth, E., Weinhardt, N., Fodor, E., Welker, E. (2017) Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol*, 18, 1-17. (IF: 13,214)
7. Tóth, E., Weinhardt, N., Bencsura, P., Huszár, K., Kulcsár, P.I., **Tálas, A.**, Fodor, E., Welker, E. (2016) Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol Direct*, 11, 1-14. (IF: 2,856)