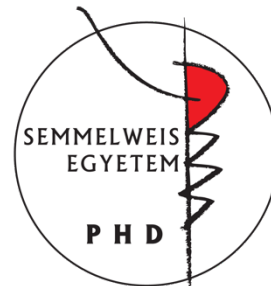


Cas9 nukleázok működési hatékonyságának kísérletes meghatározása és predikciója

Doktori értekezés

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

Tartalomjegyzék

1. Rövidítések jegyzéke.....	4
2. Bevezetés.....	5
2.1. Kezdeti génmódosító technikák.....	5
2.2. Célzott genommódosítás designer nukleázokkal.....	6
2.3. CRISPR - a bakteriális adaptív immunrendszer.....	8
2.4. A CRISPR/Cas9 rendszer működése.....	12
2.5. Célzott genommódosítás a CRISPR rendszerrel.....	16
2.6. A CRISPR/Cas9 rendszer hatékonyságának kísérletes meghatározása.....	19
2.7. A CRISPR/Cas9 rendszer hatékonyságának predikciója.....	23
3. Célkitűzések.....	25
4. Módszerek.....	26
4.1. Felhasznált anyagok.....	26
4.2. Expressziós plazmidok létrehozása.....	26
4.2.1. Önvágó plazmidok létrehozása.....	26
4.2.2. Plazmidok létrehozása az SLS módszerhez.....	27
4.3. <i>stgRNS</i> könyvtárak létrehozása.....	27
4.4. A könyvtárak szelekciója bakteriális sejtekben (SLS szelekció) és NGS.....	28
4.5. Bioinformatikai analízisek.....	29
4.5.1. A hasítás hatékonyság kiszámítása.....	29
4.5.2. A nukleotid összetétel (NucCom) kiszámítása, és a Hi-CRISPR A kifejlesztése.....	30
4.5.3. A Hi-CRISPR B és C kifejlesztése.....	31
4.5.4. Predikciók minőségének értékelése.....	32
4.6. Sejtenyésztés és transzfekció.....	33
4.6.1. Sejtenyésztés.....	33
4.6.2. Önvágó plazmid kísérletek transzfekciója és puromicin szelekció.....	33
4.7. <i>GFxFP</i> esszé.....	34
4.8. Áramlási citometria.....	35
4.9. Konfokális mikroszkópia.....	35
4.10. Genomi DNS izolálás.....	35
4.11. TIDE és T7 endonukleáz I esszé.....	35
4.12. Statisztika.....	36

5.	Eredmények.....	37
5.1.	<i>gRNS-ek hatékonyságának meghatározása önvágó plazmidokkal.....</i>	<i>37</i>
5.1.1.	Egy plazmid sejten belüli hasítása megkönnyíti annak integrációját a genomba	37
5.1.2.	Az önvágó plazmid konstrukciója univerzális spacer és célszekvencia párokkal.....	40
5.1.3.	A módszer összehasonlítása a homológ rekombinációval, valamint működése további sejtvonalakban	47
5.1.4.	Észlelési idő rövidítése egy degradációs domén fúziójával	50
5.1.5.	A módszer hasonlítása két PCR alapú módszerhez.....	51
5.1.6.	NHEJ alapú génbevitel önvágó plazmid segítségével.....	52
5.2.	<i>SpCas9 variánsok hatékonyságának predikciója egy 1 millió szekvenciát tartalmazó gRNS könyvtár segítségével</i>	<i>57</i>
5.2.1.	Egymillió különböző szekvenciát tartalmazó gRNS könyvtár és a szelektív rendszer kifejlesztése	57
5.2.2.	Az SpCas9-HF1 aktivitásának meghatározása egymillió célszekvencián	60
5.2.3.	Az SpCas9-HF1 aktivitásának előrejelzése bakteriális sejtekben	64
5.2.4.	Az SpCas9-HF1 aktivitásának előrejelzése emlős sejtekben	67
5.2.5.	Az SpCas9 alternatív PAM felismerése és az NmCas9 öncélzó gRNS-ekkel	71
6.	Megbeszélés	75
7.	Következtetések.....	82
8.	Összefoglalás	84
9.	Summary	85
10.	Irodalomjegyzék	86
11.	Saját publikációk jegyzéke.....	98
11.1.	<i>Disszertációhoz kapcsolódó publikációk jegyzéke</i>	<i>98</i>
11.2.	<i>Disszertációtól független publikációk jegyzéke.....</i>	<i>98</i>
12.	Köszönetnyilvánítás.....	99
13.	Függelék.....	100
13.1.	<i>Önvágó plazmidokhoz használt primerek listája.....</i>	<i>100</i>
13.2.	<i>Könyvtárkészítéshez, valamint az stgRNS-ek klónozásához használt primerek</i>	<i>101</i>

1. Rövidítések jegyzéke

Cas9	CRISPR associated protein 9, CRISPR asszociált fehérje 9
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats, halmozottan előforduló, szabályos közökkel elválasztott rövid palindrom ismétlődések
crRNS	CRISPR RNS
DD	Degradációs Domén
DMEM	Dulbecco's Modified Eagle Médium
DSB	Double Strand Break, kettős szálú DNS törés
dSpCas9	Dead SpCas9, nukleáz inaktív SpCas9
eSpCas9	Enhanced SpCas9
FBS	Fetal Bovine Serum, magzati marhaszérum
FSC	Forward Scatter, előre irányuló fényszórás
GeCKO	Genome-Scale CRISPR Knock-Out, teljes genomra kiterjedő CRISPR génkiütés
GFP	Green Fluorescent Protein, zöld fluoreszcens fehérje
gRNS	Guide RNS, irányító RNS
HDR	Homology-Directed Repair, homológia-irányított rekombináció
HR	Homológ rekombináció
Indel	Inszerciók és/vagy deléciók
MCC	Matthews Correlation Coefficient, Matthews korrelációs együttható
NHEJ	Non-Homologous End Joining, nem homológ végek összekapcsolása
PAM	Protospacer Adjacent Motif, protospacer mellett található motívum
PCR	Polimerase Chain Reaction, polimeráz láncreakció
RNP	Ribonukleoprotein
SLS	Self-targeting Library Screen, öncélzó gRNS könyvtár teszt
Sp	<i>Streptococcus pyogenes</i>
SpCas9-HF1	SpCas9 high-fidelity variant number 1
SSC	Side Scatter, oldalra irányuló fényszórás
stgRNS	Self-targeting gRNS, öncélzó gRNS
TALEN	Transcription Activator-Like Effector Nucelase, transzkripció aktivátor-szerű effektor nukleáz
TMP	Trimethoprim
tracrRNS	Transzaktiváló CRISPR RNS
WT-SpCas9	Wild-type SpCas9, vad típusú SpCas9
ZFN	Zinc Finger Nuclease, cink-ujj nukleáz

2. Bevezetés

A CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats, halmozottan előforduló, szabályos közökkel elválasztott rövid palindrom ismétlődések) rendszer és ezen belül a Cas9 fehérje (CRISPR associated protein 9, CRISPR asszociált fehérje 9) célzott génmódosításra való használata forradalmasította a biológia szinte minden ágát az utóbbi 8 évben, és lehetőséget teremtett öröklődő betegségek gyógyítására is. Jól mutatja a rendszer potenciálját, hogy a CRISPR genomszerkesztő eljárásért 2020-ban kémiai Nobel-díjat kapott Jennifer Doudna és Emmanuelle Charpentier, az eredeti publikáció után nem egészen 8 évvel később.

2.1. Kezdeti génmódosító technikák

Az első mérföldkövet a DNS szekvenciák célzott szerkesztésében a restriktációs endonukleázok felfedezése jelentette az 1970-es években. Ezek az enzimek rövid konszenzus szekvenciákat ismernek fel a DNS-ben és azt a felismerés helyén, vagy ahhoz közel hasítják. Mivel ezek csak rövid (4-8 bázispár hosszú) szekvenciákat ismernek fel, a genomot ezekkel nem lehet egy kiválasztott helyen precíziósan megcélózni, viszont lehetőséget teremtett, kisebb DNS szekvenciák célzott hasítására és megalapozta a molekuláris klónozás alapjait¹.

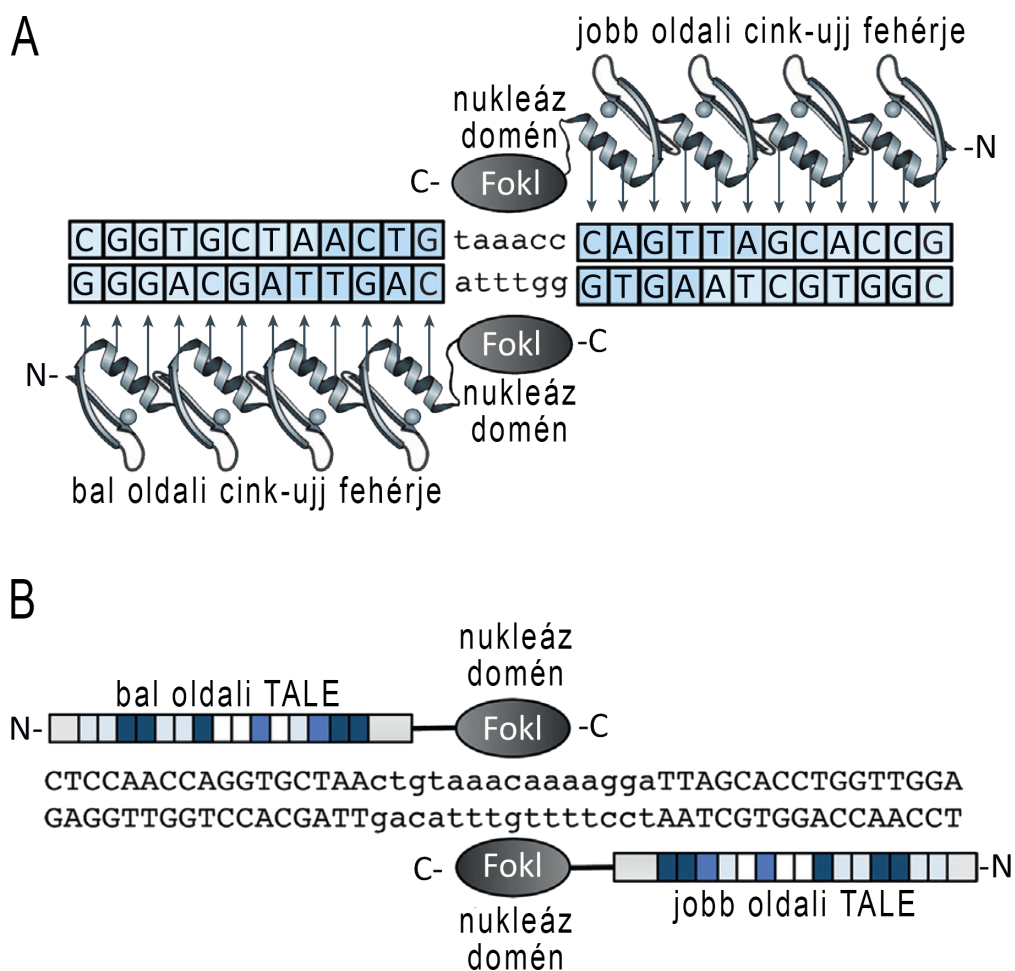
A második, forradalmi változást az a felismerés hozta, hogy a homológ rekombináció (HR) mechanizmusát felhasználva célzottan be lehet vinni DNS szekvenciákat a genomba, melyet először élesztő², majd emlős sejteken³ is demonstráltak. A homológ rekombinációt egy kettős szálú DNS törés (double strand break – DSB) idézi elő (valamint egy homológ donor DNS-nek is jelen kell lennie), amely itt teljesen véletlenszerűen történt meg. Így nem meglepő, hogy ezek a módszerek rendkívül alacsony hatékonysággal működtek: egy DNS bevitelének a gyakorisága 1:1000 – 1:1000000 között változott. Továbbá, a bevinni kívánt donor DNS sok esetben véletlenszerűen integrálódott a genom nem célzott részeire, sokszor hatékonyabban, mint a megcélzott szekvencia helyén⁴. Ezen korlátai ellenére lényeges előrelépés volt a génmódosítás terén, hiszen a HR alkalmazása volt az első ténylegesen célzott génbeviteli módszer, amely alapjaiban változtatta meg a genetikai célú kutatásokat, hiszen gének szerepét, illetve hiányát könnyebben lehetett vizsgálni.

2.2. Célzott genommodosítás designer nukleázokkal

A következő lépcsőfokot az a felfedezés hozta, hogy a genomban a génmódosítás kívánt helyén történő kettős szálú DNS törés a célzott integrációt hatékonyan megnöveli. Ennek első demonstrációja során a korábbiaknál nagyságrendekkel hatékonyabb génbevitelt értek el az I-SceI meganukleázzal létrehozott célzott DSB segítségével⁵. A meganukleázok olyan endonukleázok, amelyek a hagyományos restriktációs endonukleázokkal szemben nagyobb (14-40 bp) felismerő hellyel rendelkeznek, így egy genomban sokkal kevesebbszer tudnak hasítani, specifikusabbak. Ezek használatával lehetővé vált a korábbiaknál jóval hatékonyabb célzott génbevitel, habár a módszert súlyosan limitálta pár tényező. Annak ellenére, hogy többszáz ilyen meganukleáz azonosítottak, csak kevésnek van hasznosítható, a genomban is megfelelő helyen lévő célszekvenciája. Néhány próbálkozás született arra is, hogy megváltoztassák ezeknek a meganukleázokak a felismerő helyét jobban használható szekvenciákra^{6,7}. Sajnos ez rendkívül bonyolult, költséges és hosszadalmas folyamatnak bizonyult és továbbra is rendkívül korlátozottak maradtak az elérhető célszekvenciák és organizmusok. Ezek a megváltoztatott meganukleázok voltak az első úgynevezett designer nukleázok, amelyek a természetben megtalálható endonukleázok módosított, valamilyen célra kialakított változatai voltak.

A cink-ujj motívumok hasznosítása a célzott genommodosítás számára is egy új eszközzel szolgált. A cink-ujj motívumok olyan DNS kötő domének, melyek szerkezetét egy cink ion stabilizálja és képesek szekvencia-specifikusan egy három bázispáros tripletet felismerni a DNS-ben⁸. Ezen motívumok egymás mögé illesztésével létrehozható egy olyan fehérje, amely szekvencia-specifikusan képes felismerni egy adott DNS célszekvenciát (**1/A ábra**). A cink-ujj motívumok, valamint a FokI restriktációs endonukleáz DNS hasításért felelős doménjének fúziójával alkották meg az első cink-ujj nukleázokat (zinc finger nucleases – ZFN)⁹. A FokI nukleáz domének csak dimerben képesek működni, így két ZFN gondos tervezésével a DNS két szálára megfelelő távolságban tervezett célszekvenciákkal, a nukleáz domének dimerizálódni képesek és ott DSB-t képesek létrehozni. Ezen ZFN-eket eredetileg *in-vitro* használatra fejlesztették ki⁹ és csak jóval később használták először a *Drosophila*¹⁰ majd emberi sejtek¹¹ genomjának célzott hasítására. A ZFN-ek elterjedése és egyre kifinomultabb használata a

korábbiakhoz képest jóval több célszekvencia megcélzását tette lehetővé. Azonban a technika azon korlátját sajnos nem sikerült kiküszöbölni, hogy a ZFN-ek újratervezése egy új célszekvenciára igencsak bonyolult és költséges folyamat maradt. Az egyes tripleteket felismerni képes cink-ujj doméneket összekötő linkereket minden új szekvenciára optimalizálni kellett, hogy hatékonyan működjenek. Valamint, az egyes cink-ujj doméneket egymás mellé rakva az eredeti domén szekvencia-specifitása attól függően megváltozhatott, hogy mik a környező domének.



1. ábra. A cink-ujj és TALE nukleázok felépítése

(A) A cink-ujj nukleázok (ZFN) egyrészt a cink-ujj doménekből -melyek egyesével három bázispárt képesek felismerni-, valamint a FokI restrikciós endonukleáz nukleáz doménjének -mely csak dimer formában aktív – fúziójával lettek létrehozva. A DNS két ellentétes szálára megfelelő helyre tervezett ZFN-ekkel a FokI domének dimerizációja megtörténik és szekvencia-specifikus DNS hasítást tudunk velük létrehozni. (B) A TALEN-ek esetében a DNS felismerését TALE domének végzik, melyek egyesével egy bázispárt képesek felismerni. A fehérjéhez a ZFN-hez hasonlóan a FokI nukleáz domén van fuzionálva, mely a DNS-ben kettős szálú DNS törést képes létrehozni. Az ábra Gaj és mtsai. nyomán készült¹².

A ZFN-ek használatát egy még jobban használható technika a TALEN (transcription activator-like effector nuclease, transzkripciós aktivátor szerű effektor nukleáz) követte. A növényi patogén *Xanthomonas* baktériumokban fedezték fel a TALE fehérjét, amelyek DNS kötő doménjeit 33-35 bázispár hosszú szekvencia kódolja^{13,14}. Az ezekben található hipervariábilis régió a cink-uj motívumokkal ellentétben már nem 3, hanem 1 bázispárt képes felismerni. A ZFN-ek tervezésével analóg módon, ezen domének egymás mellé illesztésével, valamint a FokI endonukleáz DNS hasító doménjével fuzionálva egy rendkívül pontos, DNS-t felismerő endonukleáz hozható létre (**1/B ábra**). Így a ZFN-hez hasonló, de már nem 3 hanem 1 bázispárt felismerő alkotóelemekből álló, genommodosításra használható designer nukleázt hoztak létre^{15,16}. A ZFN-el összehasonlítva a TALEN-ek nagy előnye, hogy a TALE doméneket viszonylag szabadon lehet variálni, a ZFN-eknél tapasztalt linker-optimalizálás nélkül. Az is előny, hogy a ZFN domének (az összes tripletet felismerő) 64 variációja helyett itt csak 4, a nukleotidokat felismerő domént kellett használni. További előnye volt a ZFN-hez képest, hogy itt az a hatás sem érvényesült, hogy a különböző szekvenciakörnyezetre megváltozott volna az egyes domének specificitása. Sajnos azonban ennek a módszernek is vannak hátulütői. Az egyik, hogy az egyes TALE domének egymás mellé klónozása egy hosszadalmas és nehéz feladat, hiszen egyrészt, minden új célszekvenciára új fehérjét kell létrehozni, ami itt már bázisonként variált domének sorozatát jelentette. A másik klónozási nehézség, hogy (a variábilis régiók kivételével) a TALE domének repetitív szekvenciákból állnak. Fejlettebb klónozási módszerekkel¹⁷ sikerült hatékonyabbá tenni ezeknek a konstrukcióknak a létrehozását, de az alapvető probléma továbbra is fennállt; hogy minden megcélzott szekvenciához pénz- és időigényes fehérjemérnökségre volt szükség. Megjegyzendő, hogy mivel mindkét technika a FokI dimerizációján alapul, minden nehézség minden célszekvencia esetében kétszeresen jelentkezik. A fenti rendszerekkel ellentétben a CRISPR/Cas9 már egy sokkal könnyebben, gyorsabban és olcsóbban használható technológia.

2.3. *CRISPR - a bakteriális adaptív immunrendszer*

A CRISPR rövidítés az angol clustered regularly interspaced short palindromic repeats (halmozottan előforduló, szabályos közökkel elválasztott rövid palindrom ismétlődések) kifejezésből ered¹⁸. Ezeket a furcsa, ismétlődő szekvenciákat 1987-ben Ishino és mtsai.

fedezték fel az *Escherichia coli* (*E. coli*) baktérium genomjában, de funkciójukat nem sikerült feltárniuk¹⁹. Ezekre sokáig fölösleges, semmilyen funkcióval nem rendelkező (junk) DNS szekvenciaként tekintettek. Időközben több baktérium és archea genomjában hasonló ismétlődő szekvenciákat találtak, de szerepüket sokáig nem sikerült felderíteni.

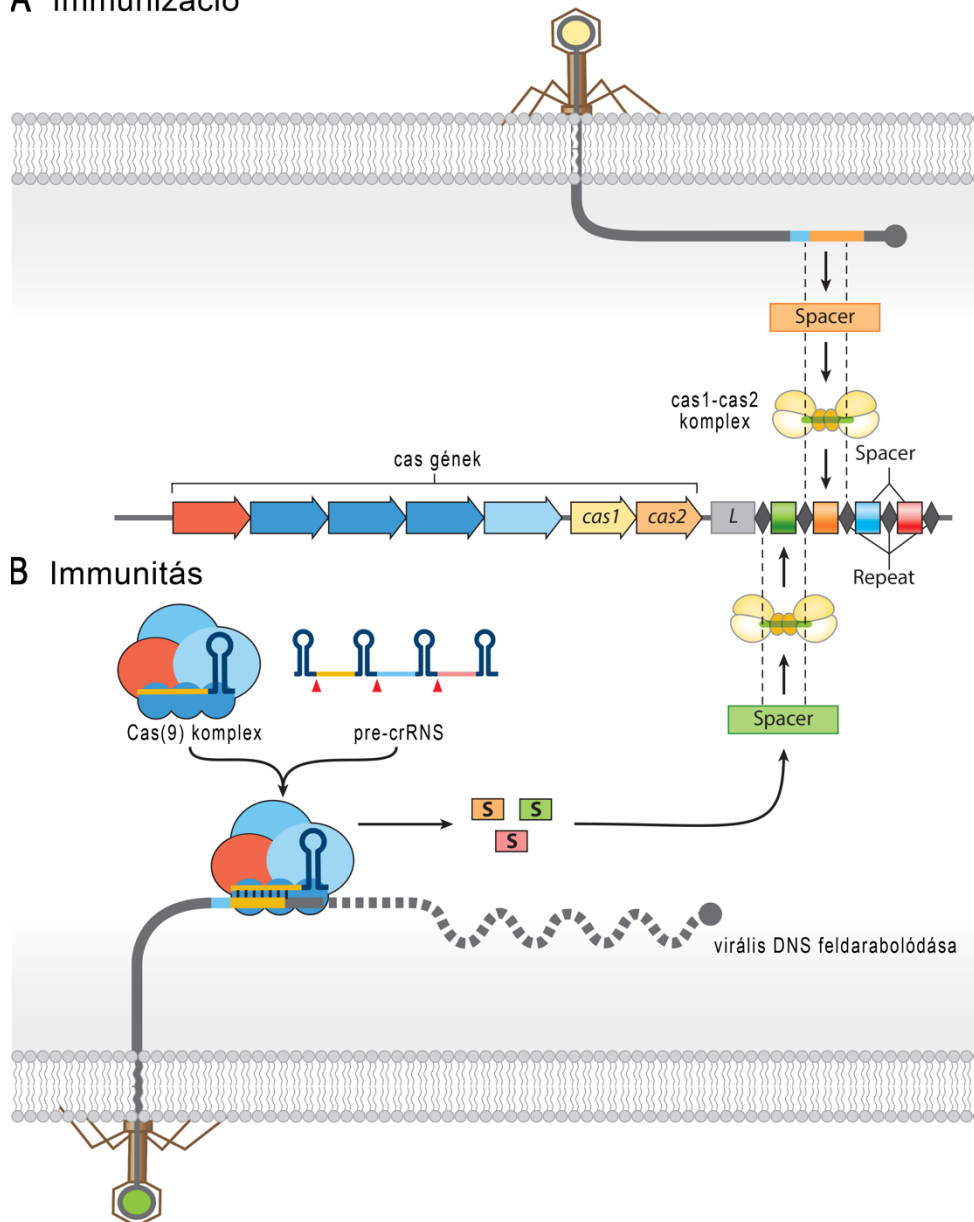
2005-ben három kutatócsoport Mojica, Bolotin és Vergnaud vezetésével, egymástól függetlenül bizonyították, hogy a furcsa ismétlődő szakaszok között lévő szekvenciák (amelyeket *spacer*nek nevezünk, az ismétlődések közti helyre utalva) virális genomokkal mutatnak nagy egyezést²⁰⁻²². Mojica felvetette, valamint Eugene Koonin kutatócsoportja egy modellt is felállított arra, hogy ez valamilyen vírusok elleni védekező mechanizmus lehet²³, amiről akkoriban úgy gondolták, hogy az RNS interferenciához hasonlóan működhet. 2007-ben Rodolphe Barrangou és Philippe Horvath kísérletesen igazolták, hogy *Streptococcus thermophilus* baktériumok egy fágfertőzést túlélő sejtjeiben új spacerek integrálódnak a CRISPR lókuszbán, és ezek szekvencia-komplementaritást mutatnak a fágok genomjával²⁴. Az itt integrálódott szekvenciák eltávolításával vagy hozzáadásával megváltoztatható volt a baktériumok fág-rezisztenciája. Tehát, ez szolgáltatta az első kísérletes bizonyítékot a bakteriális „immunrendszer” működésére. Későbbi munkákban bizonyították, hogy a CRISPR rendszer a fág DNS-ét célozza²⁵, a CRISPR lókusztól RNS-ek íródnak át²⁶, amelyek részt vesznek a védekezésben, a Cas9 enzim felelős a virális DNS hasításáért²⁷, valamint Emmanuelle Charpentier kutatócsoportja bizonyította, hogy egy transz-aktiváló CRISPR RNS (tracrRNS) jelenléte is elengedhetetlen az immunitás kialakulásához²⁸.

A fenti kutatásokat összegezve mai tudásunk szerint a CRISPR-Cas adaptív immunrendszer a következők szerint működik (**2. ábra**). Az immunizáció szakaszában (**2/A ábra**), a virális DNS bejut a bakteriális sejtekbe és alacsony valószínűséggel különböző Cas enzimek segítségével (pl. Cas1-Cas2) a virális DNS egy rövid (kb. 20 bp) szakasza beépül a baktérium genomjának CRISPR lókuszába két ismétlődő szekvencia (repeat) közé. Az egyes fág DNS-ek időrendben épülnek be a lókuszbba, így az is felderíthető, hogy a baktérium a védettségre milyen fásra és milyen időrendben tett szert.

Az immunitás fázisában (**2/B ábra**), egy következő fág fertőzésakor, a CRIPR lókusztól fokozódik a Cas fehérjék, a tracrRNS, valamint a beépült szekvenciát már tartalmazó pre-crRNS-ek átíródása. A pre-crRNS ezután feldarabolódik, és ezekből érett crRNS-ek

lesznek. A Cas9 fehérje (vagy egyes baktériumokban és archeákban más Cas fehérje, vagy fehérje komplex), a tracrNS-sel valamint crNS-sel komplexben (amelyek között számos, az adott fágra specifikus szekvencia is megtalálható), a fág DNS-hez szekvencia-specifikusan kötődik, majd elhasítja, így azt eliminálja. Tehát, ez egy memóriával rendelkező adaptív immunrendszerhez hasonlít, ilyen módon az emlősök immunrendszeréhez hasonlítható, de természetesen evolúciósan teljesen független folyamat. Az egyik nagy különbség a két rendszer között, hogy a bakteriális immunrendszer tovább örökíthető, hiszen az integrált spacerek a baktérium genomjában osztódás után is jelen lesznek, és az utódsejteknek is védettséget jelentenek. Későbbi kutatások kiderítették, hogy a baktériumok legalább 42%-a és az archeák legalább 85%-a rendelkezik ilyen „immunrendszerrel”²⁹.

A Immunizáció



2. ábra. A bakteriális adaptív immunrendszer működési elve

(A) Az immunizáció szakaszában a virális DNS bejut a bakteriális sejtbe, és a rajta található PAM (kék) szekvencia mellett található protospacer (sárga) szekvencia a Cas enzimek segítségével (ekkor már spacernek hívjuk) beépül a bakteriális genom CRISPR lokuszára. (B) Az immunitás fázisában egy már korábbi fertőzést túlélte sejt, egy ismételt fágfertőzés esetén, a támadó fág DNS ellen már hatékonyabban tud védekezni az eltárolt szekvencia segítségével. A Cas9 fehérje (esetenként másik, bizonyos baktériumokban sokszor több fehérjéből álló Cas fehérje komplex), a hozzá asszociált tracrRNA és a specificitásért felelős crRNA segítségével szekvencia komplementaritás alapján felismeri a virális DNS-t és elhasítja, így eliminálja a sejtből. Az ábra Marrafini és mtsai. nyomán készült³⁰.

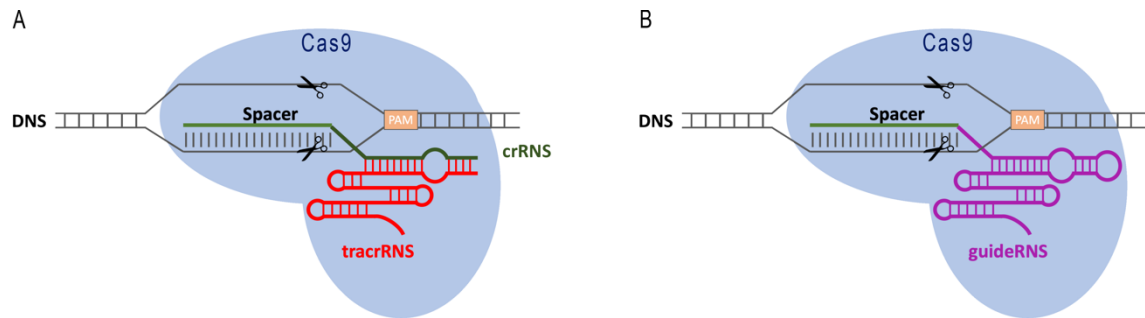
Felvetődhet a kérdés, hogy a baktérium hogyan tudja megkülönböztetni a virális genomot a saját genomjában spacerként „eltárolt” virális DNS szakaszoktól, hiszen azok szekvenciájukban teljesen megegyeznek, így egy „autoimmun” folyamatban a Cas9 a baktérium saját genomját is elhasítaná. A fág genomból csak olyan szakaszok épülhetnek be a bakteriális genomba, ahol a megcélzott DNS (protospacer) egy úgynevezett protospacer melletti szekvencia (protospacer adjacent motif, PAM) mellett helyezkedik el. Ezt a szekvenciát a Cas9 fehérje ismeri fel és baktériumfajonként változó, például a genommodosításra is használt *Streptococcus pyogenes* Cas9 (SpCas9) esetében ez egy NGG szekvencia. A spacer beépülése helyén (amely a crRNS általános szekvenciáját kódolja) a bakteriális genomban nem található PAM, így azt nem képes hasítani a fehérje.

A bakteriális immunrendszer effektor fehérjéről 2012-ben Jennifer Doudna³¹ és Virginijus Šikšnys³² kutatócsoportja is bizonyította, hogy *in vitro* körülmények között is képes DNS szekvenciákat elhasítani. Arra a fontos következtetésre is rájöttek, hogy a crRNS spacer régióját megváltoztatva, gyakorlatilag bármilyen DNS szekvencia célba vehető. Ez a kulcsfontosságú lépés alapozta meg a CRISPR/Cas9 rendszerrel történő génszerkesztést.

2.4. A CRISPR/Cas9 rendszer működése

A Cas9 fehérje tehát egy RNS irányított DNS endonukleáz, amely programozható módon képes egy kívánt DNS szekvencia felismerésére, és kettős szálú hasítására (**3. ábra**). Az *in vitro* koncepciókísérlet után nem sok idő telt el, hogy először azt is megmutatták, hogy az SpCas9 képes emlős sejtek és elölények genomjának módosítására is³³⁻³⁶. A Cas9 fehérje a crRNS-sel és a tracrRNS-sel ribonukleoprotein (RNP) komplexet képez és a felismert DNS szekvencia mindkét szálát hasítja (**3/A ábra**). A crRNS nem variábilis, a tracrRNS-sel komplementer régióján keresztül kapcsolódik tracrRNS-hez, valamint a tracrRNS további részének szerkezetének felismerésével a fehérje kapcsolódik az RNS komplexhez. Az SpCas9 esetében a crRNS első 20 bázisa (spacer) felelős a célszekvencia (protospacer, vagy target) felismeréséért. A spacer szabadon változtatható, így az RNP programozható módon képes felismerni és elhasítani különböző DNS szekvenciákat. A célszekvencia felismerésének előfeltétele, hogy a protospacer mellett PAM motívumnak is jelen kell lennie a DNS szálon. Amennyiben a PAM motívum melletti felismert DNS-sel homológ a spacer szekvenciája, a fehérje két nukleáz (RuvC és HNH) doménjével

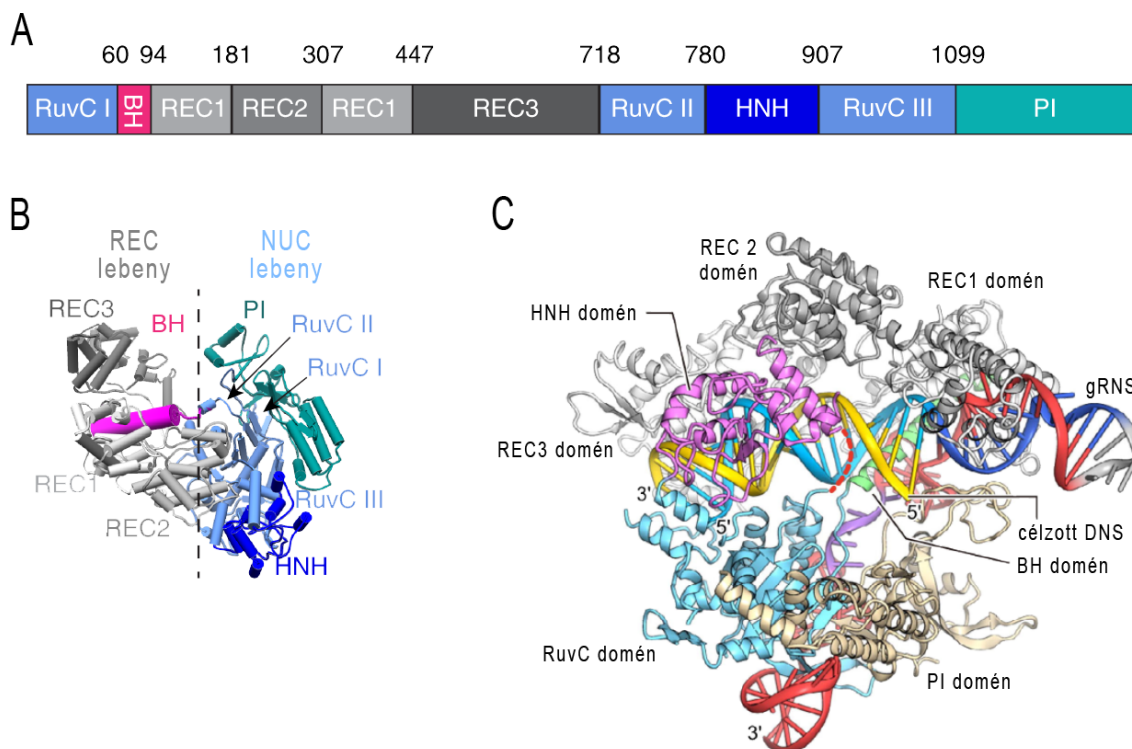
kettős-szálú DNS törést hoz létre. A Cas9 nukleáz genomszerkesztő eszközként való használatának megkönnyítése céljából, a cr- és tracrRNS-eket egy hurokkal kovalensen összekötötték³¹, létrehozva így az irányító RNS-t (guideRNS, gRNS, **3/B ábra**). Ez a rendszer leegyszerűsítésén túl, azért is hasznos, mert a két RNS-t nem kell külön promoterről átírni, valamint ezek (lehetségesen eltérő) expresszióját vizsgálni. Így ezt a megoldást ma már széleskörűen használják.



3. ábra. A CRISPR/Cas9 rendszer működése

(A) A Cas9 fehérje (kék) szekvenca-specifikusan képes egy megcélzott DNS szekvenca (szürke) felismerésére és elhasítására. A Cas9 enzimhez két RNS kapcsolódik a crRNS (zöld) és a tracrRNS (piros). A tracrRNS az RNS komplex fehérjéhez való kapcsolódását segíti, a crRNS 5' végén található spacernek nevezett 20 bázis hosszú régió (világos zöld) megváltoztatható és így a fehérje programozható módon irányítható egy adott DNS szekvenciához. A felismerés további feltétele egy rövid konszenzus szekvenca (PAM – magenta) megléte a DNS-en a célzott szekvenca mellett. (B) A rendszer egyszerűbb működésének céljából az előbbi két RNS-t fuzionálni lehet egy RNS molekulává (gRNS, lila), amely képes ellátni mindkét RNS funkcióit.

Az SpCas9 fehérje kristályszerkezetének felderítése^{37,38} részletesebben feltárta a fehérje doménjeit és azok funkcióit (**4/A ábra**). A fehérje funkcionálisan a nukleáz (nuclease - NUC) és a felismerő (recognition – REC) lebenyre különíthető el (**4/B ábra**). A nukleáz lebenyben található a HNH és a három részből álló RuvC (RuvC-I, -II, -III) endonukleáz domének, melyek a DNS szálak hasításáért felelősek, valamint a PAM felismerésért felelős PAM-mal kölcsönható (PAM interacting – PI) domén. A többnyire alfa hélixekből álló REC lebenyben található a híd-hélix (bridge helix, BH) domén, mely a két lebenyt köti össze, valamint a három REC domén (REC-1, -2, -3), melyek a DNS és/vagy az RNS felismeréséért felelősek (**4/C ábra**).



4. ábra. A Cas9 doménjeinek elhelyezkedése és szerkezete

(A) Az ábrán a Cas9 doménjeinek N-től C terminális irányban való elhelyezkedése látszódik sematikusán. A domének vége az ott elhelyezkedő aminosav számával vannak jelölve. Az ábrán látszódik a 3 aldoméből álló RuvC (világos kék) valamint a HNH (sötétkék) endonukleáz domén; a PAM kölcsönható domén (PI, zöld); a híd hélix domén (BH, rózsaszín); valamint a 3 felismerő REC domén (szürke). (B) Az ábrán a REC és NUC lebenyek és a domének térbeli elhelyezkedése látható a fehérje RNS-t nem kötött állapotában. (C) A fehérje doménjeinek elhelyezkedése a fehérjén belül a gRNS, valamint a célzott DNS szál (egy 20 bp hosszú oligonukleotid) kötött állapotában. Az ábrán a cél DNS sárga, a gRNS spacer régiója világos kék és a fennmaradó része kék és piros. Az ábrán a fehérje REC doménjei szürkék, a BH domén lila, a PI domén vajszínű, a HNH rózsaszín, a RuvC pedig türkiz. Az ábra Nishimasu és mtsai.³⁷ valamint Wilkinson és mtsai.³⁹ nyomán készült.

A különböző körülmények között (gRNS nélkül³⁸, csak gRNS-sel⁴⁰, valamint gRNS-sel és cél DNS-sel³⁷) készült kristályszerkezetek, valamint a Cas9 mechanizmusának felderítésére tervezett kísérletek^{41,42} segítségével jobban megértették a fehérje működésének folyamatát. A jelenleg elfogadott modell szerint a gRNS bekötődése következtében a Cas9 jelentős konformációváltozáson esik át, melynek következtében kialakul a DNS-felismerés-kompetens állapot. Ebben a konformációban a PI domén exponált állapotba kerül, mellyel képes a potenciális PAM-okat felismerni. A Cas9 egy gyors folyamatban a cél DNS-en PAM-okat pásztáz, a megtalált PAM szekvenciához

kötődve elindítja a DNS szálak szétválasztását, majd a PAM-ot követő RNS-DNS szekvencia komplementaritását vizsgálja. A szálak szétválasztása közben, illetve ezt követően a fehérje köti a nem-célzott DNS szálakat is, így azt stabilizálja, valamint ezzel gátolja annak visszakötődését a célzott DNS szálhoz. A DNS szétválasztása és a bázispárosodás folytatódik a gRNS 5' irányában, tehát a gRNS-DNS buborék a PAM-tól a gRNS 5' végéig egyre előrébb halad, amelyet a fehérje további konformációs változásai segítenek elő, a fehérje aktív állapotba kerül. A teljes spacer-target szekvencia kötődésénél újabb szterikus változással kialakul a hasítás kompetens konformáció, a HNH valamint a RuvC endonukleáz domén a DNS elvágásához szükséges állapotba kerül és elhasítják a cél DNS-t a PAM-tól 5' irányban 3 nukleotid távolságra, leggyakrabban tompa végű DNS törést hozva létre. A HNH a gRNS által célzott, míg a RuvC a nem célzott szálakat hasítja.

A Cas9 nukleáz doménjeit lehetséges inaktiválni aminosav mutációk révén^{31,32} (a RuvC-t a D10A, a HNH-t a H840A aminosavak mutációjával), ilyenkor (a kiválasztott domén inaktivációjától függően) csak az egyik szálakat hasítja, DNS bevágást (nick-et) hoz létre, az így létrejött fehérjét egy szálakat hasító, vagy nickáz Cas9-nek (nCas9) nevezzük. Amennyiben mindkét nukleáz domént inaktiváljuk egy katalitikusan inaktív Cas9 (dead, dCas9) jön létre, amely a célszekvenciát felismerni és azt kötni igen, hasítani viszont nem képes.

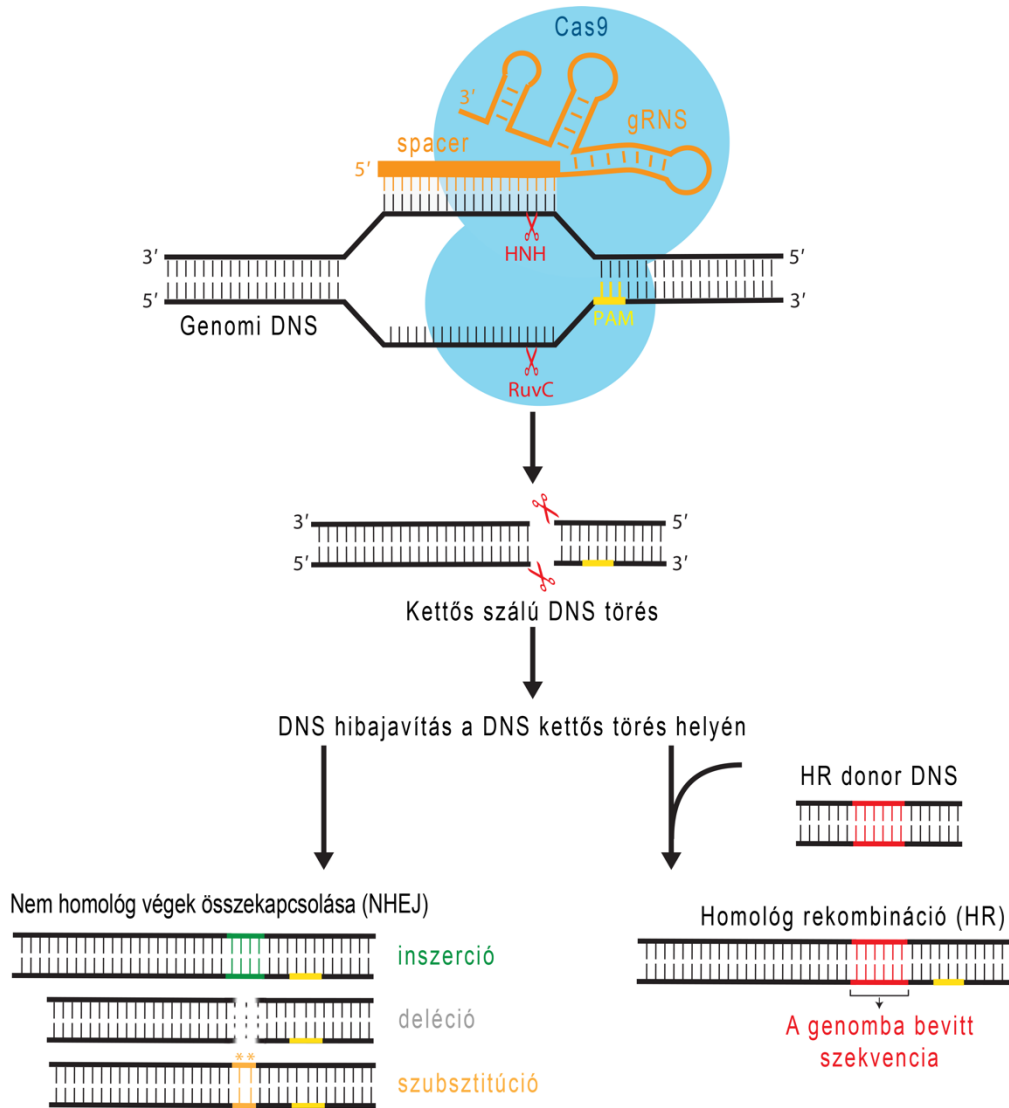
A CRISPR/Cas9 rendszer nagy előnye a korábbi fehérje-DNS kölcsönhatáson alapuló rendszerekkel (meganukleáz, ZFN, TALEN) ellentétben, hogy RNS-DNS interakción alapszik, amely segíti a könnyebb programozhatóságot. Ebben a rendszerben nincs szükség fehérjemérnökségre, hiszen egy szekvencia megcélzásához, csak a cr- vagy gRNS-en található 20 bázis hosszú spacer szekvenciát kell módosítani. Ebből következőleg rendkívül gyorsan létrehozható egy expressziós konstrukció, valamint jóval olcsóbban használható, mint a korábbi, bonyolult fehérjetervezést és optimalizálást igénylő módszerek. Ezek a tulajdonságok segítették elő, hogy rendkívül rövid idő alatt, széles körben elterjedjen a módszer és ma már számos tudományterületen alkalmazzák.

2.5. *Célzott genommódosítás a CRISPR rendszerrel*

A Cas9 fehérjével a célzott génmódosítás helyét jelöljük ki az általa létrehozott kettős szálú DNS töréssel³³⁻³⁶. Ezt követően ez javítható, (leegyszerűsítve) két főbb DNS hibajavító mechanizmuson keresztül: az egyik a nem homológ végek összekapcsolása (Non-Homologous End Joining – NHEJ), a másik a homológia irányított rekombináció (Homology-Directed Repair – HDR, innentől egyszerűsítve csak homológ rekombináció, vagy HR, **5. ábra**).

Az NHEJ hibajavításban résztvevő enzimek a duplaszálú DNS törést újra összeilleszthetik. Ez történhet tökéletesen, vagy hibával terhelt, amely a törés helyén inszerciókat és/vagy deléciókat (indel) eredményezhet (**5. ábra**). Amennyiben a hibajavítás tökéletes a DNS újra szubsztrátja lesz a Cas9-gRNS komplexnek, mely szintén a hiba felé tolja el a folyamatot. Gének kiütésénél ezt a folyamatot használjuk fel, hiszen egy fehérjét kódoló génszakaszban a hibás javítás következtében a nyílt leolvasási keret eltolódhat vagy korai STOP kodon alakulhat ki, mely a fehérje funkcióvesztéséhez, kiütéséhez vezethet³³⁻³⁶.

A homológ rekombináció mechanizmusa során a törött szál a DSB-t követően a testvérkromatidán tárolt homológ információt felhasználva javítható. Ezt a folyamatot egy homológ donor DNS szekvencia felhasználásával tudjuk eltéríteni és az általunk kívánt szekvenciákat bevinni a célzott genomi lókuszra^{34,43} (**5. ábra**). A homológ donor DNS a kívánt szekvencia két oldalán a genommal homológ szakaszokat tartalmaz, mely a testvérkromatidát mimikálva indukálja a homológ szekvenciák közti kívánt szekvencia beépítését. A donort DNS-t főlegben adva a sejtekhez, megnöveli annak az esélyét, hogy ne a testvérkromatidáról, hanem a donorról történjen a javítás.



5. ábra. Az NHEJ és a homológ rekombináció hibajavító rendszerek CRISPR/Cas9 hasítást követően

A Cas9 által létrehozott duplaszálú DNS törések két fő hibajavító rendszerrel javíthatóak. Az NHEJ javítás hibával terhelt, bázisok beépülhetnek (inszerció) vagy kieshetnek (deléció) esetleg lecserélődhetnek másik bázisokra (szubsztitúció). A pontos hibajavításért a homológ rekombináció felel, amelyet felhasználhatunk gének bevitelére egy homológ rekombinációs donor DNS felhasználásával. A donor DNS a megcélzott DNS-re homológ szekvenciákat tartalmaz (fekete), amely középen a bevinni kívánt szakaszt hordozza (piros). Az ábra Jiang és mtsai. nyomán készült⁴⁴.

A rendszer egy másik felhasználási módja, hogy az inaktív dCas9 fehérjét fel lehet használni arra, hogy valamilyen más fehérjét a célzott DNS szakaszhoz irányítson. Például egy dCas9 és egy transzkripció aktivátor fúziójával egy gén átírását lehet növelni⁴⁵, valamint egy represszor fúziójával ideiglenesen csökkenteni⁴⁶. A Cas9

valamint különböző fehérjék fúziójával, mint a hiszton acetil-transzferáz⁴⁷, hiszton deacetiláz⁴⁸, hiszton demetiláz⁴⁹ valamint a DNS metil-transzferáz^{50,51}, irányítottan lehet epigenetikai módosításokat végezni, valamint különböző fluoreszcens markerek használatával kromoszómák helyzetét vizsgálni^{52,53}.

Genetikai vizsgálatokhoz óriási segítség volt, a genom egészére kiterjedő kiütési módszerek kidolgozása, mint például a teljes genomra kiterjedő CRISPR génkiütés (Genome-Scale CRISPR Knock-Out, GeCKO)⁵⁴. Ezekben a kísérletekben egy olyan lentivirális gRNS könyvtárat alkalmaznak, amely a genom összes génjét (génenként átlagosan 3 gRNS-sel) célozzák a Cas9 segítségével. Minden sejt genomjába átlagosan 1 gRNS-t kódoló lentivírus épül be, ahol kiütheti az adott gént, majd a kísérlettől függően ez a sejt feldúsulhat a populációban. Az adott populációban az integrálódott gRNS-t kódoló DNS szekvenálása során kideríthető, hogy mely gén kiütése okozta az adott fenotípust.

Az utóbbi években szélesben elterjedt az úgynevezett bázis-szerkesztő (base editing, BE) enzimek használata, melyek egy nickáz Cas9 valamint egy citozin-⁵⁵ vagy adenin-deamináz⁵⁶ fúziójából állnak. Ezek használatával lehetséges egy megcélzott bázis rendkívül pontos lecserélése (bázis-szerkesztés) C-ről T-re⁵⁵ valamint A-ról G-re⁵⁶. Amennyiben csak egy bázis lecserélése a génszerkesztés célja, ezzel a módszerrel hatékonyan elérhető, valamint alacsony mértékű DSB történik, így a nem kívánt módosításból származó háttér is kisebb, pl. a homológ rekombinációval történő javításhoz hasonlítva. Később létrehozták a prime editing technikát⁵⁷, amely már nem limitált bizonyos bázisokra: egy rövidebb DNS-szakaszon bármilyen bázis bármilyen bázisra lecserélhető (DSB nélkül) egy nCas9 valamint egy reverz-transzkriptáz fúzióját felhasználva.

A Cas9 fehérjével történő genommodosítás különböző korlátait az évek folyamán sikerült jobbnál jobb megoldásokkal lecsökkenteni. Az egyik leküzdendő probléma, hogy az SpCas9 által megcélozható szekvenciák száma limitált, hiszen az NGG PAM szekvenciának a target mellett jelen kell lennie. Ez bizonyos esetekben korlátozó lehet, így más baktériumfajokból izolált Cas9 variánsok, mint a *Neissera meningitidis* Cas9 (NmCas9)⁵⁸, vagy a *Staphylococcus aureus* Cas9 (SaCas9)⁵⁹ eltérő PAM preferenciái miatt használhatók olyan szekvenciákon, ahol adott esetben az SpCas9 nem. Egy másik

CRISPR fehérje a Cas12a (régebbi nevén Cpf-1) génmódosításra történő alkalmazása is színesítette a génszerkesztő eszközök palettáját⁶⁰. A Cas12a az SpCas9-cel szemben nem G-, hanem T-gazdag PAM-mal rendelkezik, így a Cas9-cel nem megcélozható szekvenciák egy része elérhetővé vált ezzel a fehérjével, valamint nem tompa, hanem ragados végű töréseket hoz létre, amely a Cas9-tól eltérő mintázatú hibajavítást indukálhat. A felismerhető PAM szekvenciákat tovább bővítették az SpCas9 racionálisan tervezett mutációival⁶¹⁻⁶⁴, valamint később létrehoztak irányított evolúcióval olyan variánsokat, amelyek PAM-ja már csak 1 bázisban kötött⁶⁵⁻⁶⁷.

Egy másik leküzdendő akadály, hogy az SpCas9 bizonyos esetekben nem csak a gRNS-sel tökéletesen megegyező, hanem néhány bázisban eltérő szekvenciákat is hasíthat. Ezt a jelenséget nevezzük off-target hatásnak. Különböző módszerekkel sikerült ezt lecsökkenteni, mint például két nCas9 használatával (a két ellentétes szálát vágva hozható létre DSB)^{68,69}, vagy a TALEN és ZFN rendszerekhez hasonlóan a FokI endonukleáz dCas9-hez történő fúziójával^{70,71}. Ezeknek a módszereknek hátránya volt, hogy a célzott génmódosítás hatékonysága (a megemelkedett pontosság mellett) sok esetben alacsony volt. A problémára az eSpCas9 (enhanced SpCas9)⁷², az SpCas9-HF1 (SpCas9 high-fidelity variant number 1)⁷³ illetve más, megnövelt pontosságú SpCas9 variánsok⁷⁴⁻⁷⁷ bizonyultak a legjobb megoldásnak, hiszen az off-target hatást csökkentették, viszont a célzott génmódosítás hatékonysága sem csökkent le drámai mértékben.

2.6. A CRISPR/Cas9 rendszer hatékonyságának kísérletes meghatározása

Egy további korlátja maradt a rendszernek, hogy a Cas9-ek a különböző megcélozott szekvenciákat különböző hatékonysággal hasítják. Egy génkiütés, vagy bármilyen Cas9-cel történő génszerkesztési feladathoz fontos lenne tudni, hogy a jelölt gRNS-eken található spacerekkel a fehérje milyen hatékonysággal képes hasítani. Egy célravezető módszer, ha a Cas9 hasításának hatékonyságát a lehetséges célszekvenciákon kísérletesen megvizsgáljuk és kiválasztjuk azt, amely a leghatékonyabb volt. A Cas9 hasítási hatékonyságát számos technikával lehet vizsgálni. A legtöbb módszer a Cas9 hasítása után bekövetkező kettős szál törések által létrehozott indelek gyakoriságát méri. Ezek közös pontja, hogy a megfelelő indeleket tartalmazó régió PCR amplifikálásával kezdődik, majd a PCR terméket különböző módszerekkel elemzik.

A széles körben elterjedt, endonukleázokon alapuló módszerek (melyek a Surveyor⁷⁸ vagy T7 endonukleáz I-t -T7EI⁷⁹- használják) a PCR termékekben lévő inszerciókat vagy deléciókat képesek kimutatni. A PCR termékeket egy gyors fűtési és hűtési reakció során kétszálúvá teszik, így az indeleket tartalmazó egyszálú PCR termékek az ezeket nem tartalmazókkal is párba állhatnak. Ilyenkor, inszerció esetén a módosult, deléció esetén az eredeti szálon egy hurok képződik. Ezeket a hurkokat képes a Surveyor vagy T7EI enzim felismerni és a hurok helyén a DNS-t elhasítani. Az enzimhasítás után a DNS darabok mennyiségét gélelektroforézissel meg lehet határozni és a hasított fragmensek aránya, valamint az ebből következő génszerkesztés hatékonysága kiszámítható. A módszer hátránya, hogy sok optimalizálást igényel, valamint pontatlan, viszont az egyik leggyorsabb a PCR alapú módszerek közül.

Egy másik módszer a PCR termékek közvetlen Sanger szekvenálása. A TIDE algoritmus képes a kromatogramot analizálni és kiszámítani a génszerkesztés hatékonyságát és minőségét⁸⁰. A program dekonvolúción alapul, ahol egy kontroll (nem vágott) és egy Cas9-cel kezelt sejtől származó PCR termék kromatogramját hasonlítja össze. Ez az enzimatikus módszerekkel szemben már az esetleges deléciók méretét, vagy a beépült bázisok minőségét is képes megállapítani. Ez a módszer sem tökéletesen pontos, de jóval megbízhatóbb, mint az endonukleázokat használó módszerek. Előnye, hogy viszonylag gyors és jobban megállapítható a génedítálás minősége, viszont hátránya, hogy drágább, mint az előző módszerek.

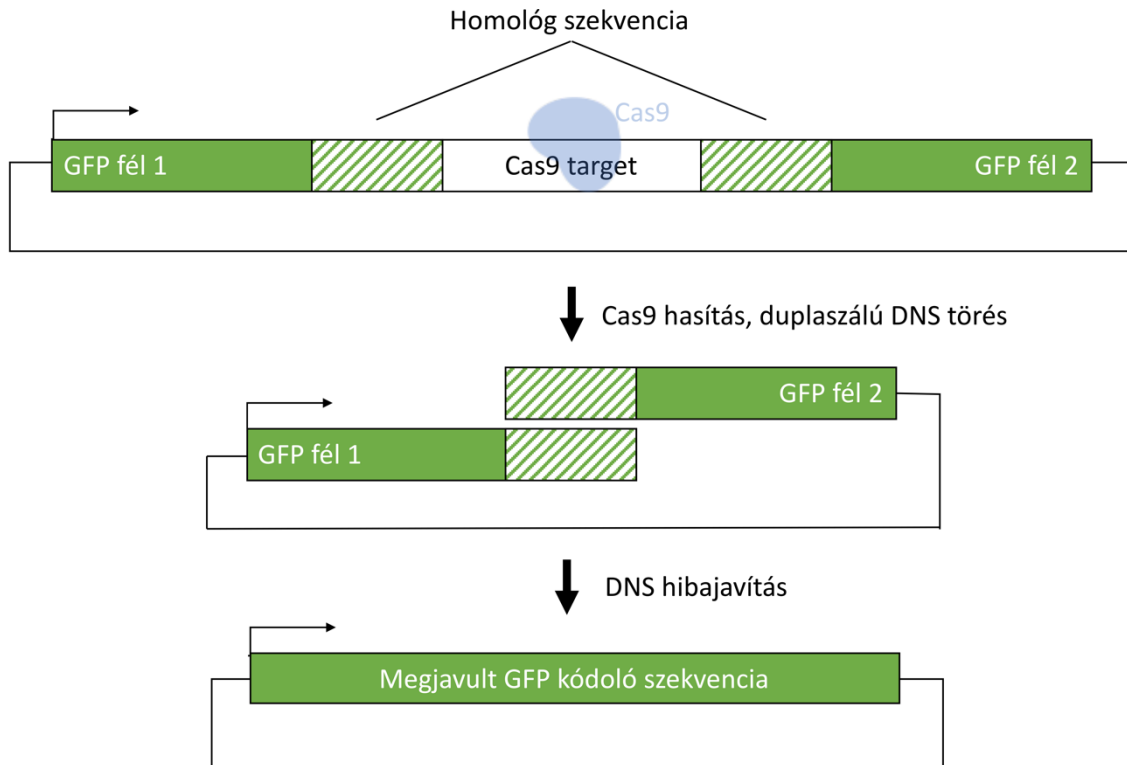
A gRNS-ek hatékonyságának meghatározására a legpontosabb módszer az újgenerációs szekvenálás (next generation sequencing – NGS), amely a legnagyobb felbontásban képes megállapítani az indelek és szubsztitúciók méretét és gyakoriságát^{68,81}. Hátrányai közé tartozik, hogy hosszú előkészítést igényel és ez a legdrágább az itt felsorolt módszerek közül. További kevésbé elterjedt módszerek egy PCR termékben történt indelek meghatározására a fluoreszcens PCR kapilláris gélelektroforézis⁸², a nagy felbontású DNS olvadás elemzés⁸³, vagy nagyszámú génszerkesztett klón létrehozása és egyenkénti genotipizálása⁸⁴.

A PCR alapú módszerekkel szemben a gRNS-ek aktivitásának meghatározására a riporter fehérjét használó módszerek kényelmes alternatívák. Az ilyen rendszerek a célszekvenciánál történő DSB valamilyen fluoreszcens jel megjelenését vagy eltűnését

okozza, így lehet velük követni a génszerkesztés hatékonyságát. Az előnyök közé tartozik, hogy mellőzik a PCR alapú módszerek minden negatívumát és jóval olcsóbbak, mint az újgenerációs szekvenálás.

A GFP-elrontási esszében¹⁷ egy genomban kódolt GFP fehérjét célunk a Cas9 fehérjével és az indelek következtében a GFP kódoló szekvenciája elromolhat, így a GFP-t expresszáló sejtek száma is lecsökken. A módszerrel könnyen és viszonylag gyorsan össze lehet hasonlítani különböző Cas9 variánsok hasítási hatékonyságát. Hátránya, hogy csak a GFP-ben található szekvenciákkal vizsgálható, így más genomi szekvenciák vizsgálatára alkalmatlan. Továbbá, a genom vágásának pillanatában jelenlévő fluoreszcens fehérjék lebomlását meg kell várni a kiértékeléshez, így ez nem a leggyorsabban kiértékelhető módszer.

A GFxFP esszében^{85,86} már nincs megkötve a vizsgált spacer szekvenciája, az szabadon változtatható (**6. ábra**). Ebben a plazmid alapú módszerben, a GFP kódoló szekvenciája két részre bontott, melyek egy rövid homológ szekvenciát tartalmaznak. A két fél között található egy klónozó hely, amelyre bármely általunk kívánt célszekvencia bevihető. A plazmidon található target hasítása a homológ részek javításához, így a GFP szekvenciájának helyreállításához vezet. Ez a módszer nagyon jól használható, amennyiben egymáshoz akarunk hasonlítani szekvenciákat, valamint gyorsan kiértékelhető. Hátránya, hogy mivel nem a genomban, hanem plazmidon történik a hasítás az nem mindig feleltethető meg a genomi környezettel.



6. ábra. A GFxFP esszé működésének sematikus ábrázolása

A GFxFP plazmid egy kettéosztott GFP szekvenciából áll (zöld), melynél mindkét fél 480 bázispár hosszú homológ szekvenciát (fehér, zöld csíkozással) tartalmaz. A Cas9 célszekvenciáját (fehér) a két GFP fél közé lehet molekuláris klónozással beépíteni. A plazmidot emlős sejtekbe juttatva, a Cas9 vágása nélkül nem (vagy alacsony) expresszálja a GFP fehérjét. A célszekvencia Cas9-cel történő hasítása, a homológ szekvenciák átlapolását, majd az ezt követő DNS hibajavítást indukálja, így a GFP kódoló szekvenciája helyreáll. Minél hatékonyabban hasítja a Cas9 a célszekvenciát, annál nagyobb lesz a GFP-pozitív sejtek száma, melyet áramlási citometriával mérhetünk.

Olyan módszerek is léteznek, amelyek a fenti módszerek előnyeit egyesítve genomi környezetben képesek fluoreszcens módszerrel meghatározni gRNS-ek aktivitását. Erre egy megoldás egy fluoreszcens kazetta homológ rekombinációval történő bevitele a genomba⁸⁷. Ennek a módszernek a hátránya, hogy minden vizsgált gRNS esetében a homológ szekvenciákat újra kell tervezni, illetve klónozni a donor plazmidra, így használata rendkívül időigényes.

Mi egy olyan fluoreszcens áramlási citométert használó módszert akartunk kifejleszteni, amely genomi kontextusban képes vizsgálni a gRNS-ek hatékonyságát, valamint nem

igényel további klónozási lépéseket, így könnyebben használható. A módszerrel egy GFP expressziós kazetta építhető be a célzott genomi helyre NHEJ hibajavítás segítségével.

2.7. *A CRISPR/Cas9 rendszer hatékonyságának predikciója*

A gRNS-ek hatékonyságának kísérletes tesztelése mellett, egy jól működő predikciós eljárás is előre jelezhetné a várható aktivitás mértékét. Egy tökéletesen működő előrejelző algoritmus fölöslegessé tenné a kísérletes tesztelést, viszont ezt az elméleti célt még nem sikerült elérni, így a kísérletes megközelítések mellett, a predikciós eljárások fejlesztésére is szükség van. A legtöbb kísérletben használt vad típusú SpCas9 (WT-SpCas9) esetében már elérhetők ilyen predikciós algoritmusok⁸⁸⁻⁹³ és a legújabbak már hatékonyabban használhatók^{94,95}. Ugyanakkor számos ortológ Cas9-et, valamint Cas9 variánst, mint például a megnövelt pontosságú vagy megváltoztatott PAM-ot felismerő SpCas9-et használnak genommérnökségi feladatokra, melyekre - a doktori munka elején még - nem volt elérhető predikciós eljárás. Azért fontos ezen variánsokra is predikciós eljárást fejleszteni, mert ezen variánsok és ortológok szekvencia-specificitása és hasítási hatékonysága eltér a WT-SpCas9-hez képest és ilyen előrejelzési eszközök létrehozása, nagyban megkönnyítené a rutinszerű használatukat.

A pontos előrejelzési eszközök kifejlesztése azonban komoly kihívást jelentő, munkaigényes és rendkívül költséges feladat, elsősorban azért, mert reprodukálható módon, nagyszámú célszekvencián kell hasítási információt gyűjteni, majd ezekből az adatokból megalkotni az algoritmust. Az SpCas9 ortológjaira és variánsaira kevés ilyen információval rendelkezünk. Még a WT-SpCas9 esetében is a legújabb predikciós eljárások is csak néhány tízezer ilyen szekvencián alapulnak, amelynek száma eltörpül a billió ($4^{20} = \text{kb. } 10^{12}$) lehetséges megcélozható szekvenciához képest.

Nagyobb számú szekvenciára kiterjedő hasítási információ előállítását az korlátozza, hogy a gRNS-nek és a célszekvenciának egy sejtben együttesen kell jelen lennie, valamint, az ezt követő hasítás kimenetelét is detektálni kell tudni. Emiatt a teljesen véletlenszerű gRNS könyvtárak használata nehezen megoldható, hiszen a célszekvenciának is azonosnak kell lennie a spacer szekvenciájával. Így legtöbbször a gRNS-célszekvencia párokat egyesével szintetizálják, majd ezekből nagyobb kevert könyvtárakat készítenek. A módszer korlátjaiból következik, hogy csak egy igen limitált

nagyságrendig alkalmazható. Mi ezt a problémát szeretnénk volna áthidalni és létrehozni egy több mint egymillió különböző szekvenciából álló véletlenszerű spacerekből álló gRNS könyvtárat. A könyvtár használatával, valamint egy szelektációs módszer létrehozásával az egyik megnövelt pontosságú SpCas9 variáns szekvencia-specificitását kívántuk meghatározni és egy predikciós algoritmust létrehozni.

3. Célkitűzések

gRNS-ek aktivitásának meghatározására, egy új fluoreszcens riporter rendszer kidolgozását tűztem ki első célomként. A létrehozandó módszernek genomi környezetben kell képesnek lennie a gRNS-ek hatékonyságának vizsgálatára, melyet áramlási citometria segítségével lehet követni. Ezt a célt úgy kívántam elérni, hogy a genomban létrehozott DNS törések helyére hatékonyan beépíték egy GFP fehérjét kódoló expressziós kazettát. Minél hatékonyabb egy gRNS annál hatékonyabban kell beépülnie a kazettának, így a fluoreszcens sejtek megszámlálásával kideríthető, hogy a vizsgált gRNS jelöltek közül mely volt a leghatékonyabb. A hatékony beépülést elősegítendő, a GFP expressziós kazettát egy olyan plazmidon szerettem volna a sejtekbe juttatni, amely a sejteken belül linearizálódik. Azt elkerülendő, hogy minden új szekvenciára a plazmidot újra kelljen tervezni, egy univerzálisan használható önvágó plazmidot terveztem, amely a vizsgált szekvenciától függetlenül képes önmagát linearizálni.

Második célom volt egy gyakran használt megnövelt pontosságú Cas9 variáns az SpCas9-HF1 szekvencia-specifitásának megértése, és egy gRNS hatékonyság előrejelző algoritmus kidolgozása. Ezt a célt úgy kívántam elérni, hogy egy olyan módszert hozok létre, amelyben a korábbiaknál jóval nagyobb számú gRNS-ről nyerhető hasítási információ, így remélhetőleg egy pontos predikciós algoritmust lehet ezen információk alapján létrehozni. Egy egymillió spacer szekvenciából álló gRNS könyvtárat terveztem létrehozni, majd ezt az SpCas9-HF1-el hasítani. A könyvtár hasítás előtt és után történő új generációs szekvenálásával kideríthető, hogy mely szekvenciák működtek a leghatékonyabban és melyek kevésbé. A generált adatokból reményeink szerint kinyerhetők olyan motívumok, amelyek az SpCas9-HF1 számára nem előnyösek vagy ellenkezőleg, hatékonyan működik velük a fehérje. A nagy számú hasítási információ alapján egy predikciós eljárás kidolgozását is terveztem, mellyel ez a variáns is általánosabban használható lenne a jövőben.

4. Módszerek

4.1. Felhasznált anyagok

A következő anyagokat és reagenseket a Thermo Scientific-től vásároltuk: restrikiós endonukleázok, T4 ligáz, hőinaktivált FBS (fetal bovine serum, magzati marhaszérum), Dulbecco's modified Eagle Médium (DMEM), Turbofect, Qubit™ dsDNA HS Assay Kit, penicillin/sztreptomicin. A következőket a Sigma-Aldrichtól vásároltuk: DNS oligonukleotidok, a plazmidok általános izolálására használt GenElute HP Plasmid Miniprep kit, trimethoprim (TMP). A ZymoPure Plasmid Midiprep kit-et Zymo Research-től vásároltuk. A NEBuilder HiFi DNA Assembly Master Mix-et, T7EI enzim mixet, Lipofectamine 2000-t, a NEB5-alpha kompetens sejteket, valamint a Q5 High-Fidelity DNS polimerázt a New England Biolabs-től (NEB) vásároltuk. A DNS fragmenseket a molekuláris klónozások során NucleoSpin Gel and PCR Clean-up kittel tisztítottuk ki, amelyet a Macherey-Nagel-től vásároltunk.

4.2. Expressziós plazmidok létrehozása

Az itt ismertetett klónozásokhoz használt oligonukleotidok valamint az elkészült expressziós plazmidokat a Függelékben részleteztem. Minden elkészített plazmidot Sanger szekvenálással ellenőriztünk.

4.2.1. Önvágó plazmidok létrehozása

A gRNS-ek spacer régiójának klónozásához a px330 plazmidba (amelyen egy gRNS-t és dead-, WT- vagy eSpCas9 expressziós kazetta található) vittem be a spacereket kódoló linker oligonukleotidokat, egy ún. one-pot reakciót segítségével, amelynél a vektor hasítása és az oligonukleotidok ligálása egy lépésben történik. Egy 20 µl végtérfogatú reakcióban 2 egység BpiI enzimet 1,5 egység T4 ligázt, 500 µM ATP-t, 50 ng vektort és 5 µM spacereket kódoló oligonukleotidot 1X Green pufferben összemérem és 37 °C-n 30 percig inkubáltam, majd a reakcióból 5 µl-t 50 µl NEB5-alpha kompetens sejtekbe transzformáltam.

Az önvágó plazmidokat a pEGFP-C1 plazmidból hoztam létre NEB HiFi assembly és restriktációs klónozás segítségével. A puromicint is expresszáló önvágó plazmid létrehozásához a pLKO.1puro plazmidot (Addgene #8453) használtam. A degradációs domént a pBMN DHFR(DD)-YFP (Addgene #29325) plazmidról amplifikáltam. Az itt használt WT-SpCas9, dSpCas9 és eSpCas9 expressziós plazmidokat a 77. referenciában írtuk le. A jellemzett önvágó plazmidok az Addgene nonprofit plazmid megosztó szolgáltatásról a következő hozzáférési számokon érhetőek el: pSc1 (#80436), pSc2 (#80437), pSc1-puro (#80438) and pSc1-DD (#80439).

4.2.2. Plazmidok létrehozása az SLS módszerhez

Az öncélzó RNS (self-targeting gRNS, stgRNS) és a kanonikus gRNS összehasonlításánál a spacereket egy stgRNS klónozó plazmidba (mely a pmCherry-gRNA plazmiddal tökéletesen megegyezik, csak tartalmazza az stgRNS mutációkat, Addgene- pAT-9222 (#124222)), a kanonikus gRNS-eket pedig a pmCherry-gRNA plazmidba klónoztam one-pot klónozással, BpiI helyekre.

A GFxFP kísérletekhez a kanonikus gRNS-ek klónozásához, a pmCherry_gRNA plazmidba klónoztam a spacereket kódoló linker oligonukleotidokat, a fent ismertetett one-pot klónozás segítségével. A GFxFP kísérletekhez a célszekvenciákat kódoló oligonukleotidokat a pWN10042 plazmidba vittem be EcoRI és BamHI enzimekkel, one-pot reakció segítségével. Az emlős SpCas9-HF1 expressziós plazmidot pKH-1699 (#124227) a pX330-Flag-wtSpCas9 felhasználásával hoztuk létre NEB HiFi Assembly segítségével. A bakteriális WT-SpCas9 (pAT9218, Addgene) és SpCas9-HF1 (pAT9251, Addgene) expressziós plazmidokat a pdCas9-bacteria⁹⁶ felhasználásával készítettem el.

4.3. *stgRNS* könyvtárak létrehozása

Az stgRNS könyvtár létrehozásához először egy klónozó plazmidot terveztem meg, melyben a bakteriális promoter mögé lehet a spacer szekvenciákat klónozni és tartalmazza a stgRNS módosításait is. Ezt a plazmidot megszintetizáltattuk (Genscript), PCR amplifikáltam, emésztettem DpnI enzimmal (hogy hasítsam a plazmid templátot) majd gélen izoláltam. A randomizált oligonukleotidot (4 μ M) egy egy lépéses PCR reakció során Q5 polimeráz és egy primer segítségével feltöltöttem, amely egy kettős

szálú inszertet eredményezett. Ezt gélből izoláltam, majd 1:5 vektor:inszert arányban HiFi Assembly segítségével (50 °C, 1h) a vektorral (25 ng) egyesítettem. A reakció végeredményét az (Inoue módszerrel⁹⁷) általam készített ultrakompetens *E. coli* sejtekbe transzformáltam, míg az 1,2 millió kolóniát el nem értem (1M könyvtár). A kolóniákat nagy számuk miatt Petri csészék helyett Bioassay lemezekben inkubáltam. A kolóniákat lemostam majd a plazmidokat ZymoPure Plasmid Midiprep kit-tel izoláltam. A 100K könyvtárat a fentiekkel megegyező módon készítettem el.

4.4. *A könyvtárak szelekciója bakteriális sejtekben (SLS szelekció) és NGS*

A WT-SpCas9 és SpCas9-HF1 plazmidokat NEB5alpha kompetens sejtekbe transzformáltam, majd ezekből az Inoue módszerrel⁹⁷ ultrakompetens sejteket készítettem. Az 1M és 100K könyvtárakat ezekbe a kompetens sejtekbe transzformáltam legalább 50x-es transzformációs lefedettséggel, három párhuzamos kísérletben. A transzformált baktériumokat egy 3D agaróz-LB médiumban (0,3 m/V % agaróz) növesztettem. A médiumot az alkotóelemek összekeverése után autoklávoztam, majd felhasználásig 37 °C-on tartottam. Az antibiotikumokat, valamint 100 µl transzformált baktériumot mágneses keverővel 50 ml médiumhoz adtam majd jégben lehűtöttem. A transzformációt addig ismételtam, ameddig az 50x-es transzformációs lefedettséget el nem értem. A zselé állagú médiumot 37 °C-on növesztettem 24 óráig, majd homogenizáltam és centrifugáltam a baktériumokat (3000 g, 10 perc, 25 °C). A plazmidokat ZymoPure Plasmid Midiprep kit-tel izoláltam.

Az újgenerációs szekvenálás elvégzése előtt a minta-előkészítés a következők szerint zajlott. A célzott régiót (promoter, spacer, gRNS) PCR segítségével amplifikáltam, a kiindulási és a szelektált könyvtárakból, majd az amplikonokat gélen izoláltam. Egy amplikon NGS szekvenálásakor a klasszikus módszerrel két lépéses PCR-t szoktunk végezni, az elsőt az amplifikált régióra specifikus primerekkel, melyek a második lépés primereinek fogadó szekvenciáját (capture site) is hordozzák. Az első lépésre itt nem volt szükség mert ezeket a fogadó szekvenciákat a vektorba előre beépítettem. Ennek előnye, hogy az első PCR esetleges hibáit teljesen ki lehet vele küszöbölni. Ez a stratégia természetesen genomikai szekvenciák esetében nem alkalmazható. A második lépés (itt az első) PCR primerei tartalmazták a plazmidra előre beépített univerzális NGS fogadó szekvenciákat, valamint az indexeket és az adaptereket (a használt NGS primereket lásd

a Függelékben). Az amplikonok pontos koncentrációjának meghatározására a Qubit™ dsDNA HS Assay Kit-et használtam, majd a mintákat összemértem (pooling) és Illumina HiSeq4000 segítségével megszekvenáltuk (BGI Genomics). Minden minta esetén minimum 50x-es szekvenálási lefedettséget értünk el. A szekvenálás nyers adatai az NCBI Sequence Read Archive oldalán PRJNA643977 számon érhető el.

4.5. Bioinformatikai analízisek

4.5.1. A hasítás hatékonyság kiszámítása

Az NGS páros (paired-end) leolvasásait a BBMerge-el (v37.22) egyesítettük alapértelmezett beállításokkal, majd ezeket a referencia stgRNS szekvenciára illesztettük BLASTn-nel (v2.6.0). Azokat a leolvasásokat eltávolítottuk, amelyek a referenciához képest bármilyen hibát (indel, mutáció) tartalmaztak és ezek közül a 10 leolvasásnál kevesebbel rendelkezőket is eltávolítottuk. A nem vágott spacerek nagy számban vannak jelen minden hasított könyvtárban, így ezeket használtuk a hasított és a vágott szekvenciák közötti leolvasások normalizálására. Ezt minden szelektált könyvtárnál megismételtük, a kiindulási nem hasított könyvtárával. A folyamat végén minden párhuzamos szelektált könyvtár spacereire egy 0 és 1 közötti értéket kaptunk, amelynél 1 a teljes 0 pedig az egyáltalán nem hasított szekvenciákat jelölte. Minden szekvenciánál a végső hasítási hatékonysági értéket a három párhuzamos kísérlet átlaga adta meg. A nyers 1 667 787 szekvenciából álló adathalmaz a szűrési lépések után 1 222 805 szekvenciát tartalmazott, ezt neveztük el 1M (vagy tanító) könyvtárnak. A másik (100K, vagy teszt) könyvtár esetében 281 834 szekvenciából 136 556 maradt a szűrések után. Az NGS adatok analízisét, Varga Júlia, Dr. Tusnády Gábor, Welker Zsombor, Welker Ágnes, Dr. Györgypál Zoltán, Erdős Gergely, Dr. Pach Péter és Dr. Pongor Lőrinc végezte.

A predikciók teljesítményét vizsgálándó, a 100K könyvtárból véletlenszerűen kiválasztottunk szekvenciákat úgy, hogy 3 kiegyensúlyozott és 3 em kiegyensúlyozott adathalmazt kapjunk mindkét vizsgált fehérjére. Az SpCas9-HF1 esetében a nem kiegyensúlyozott tesztkészlet 3x 330 darab nem vágó és 3x 4670 darab vágó; míg a WT-SpCas9 esetében 3 x 95 nem vágó és 3x 4905 vágó spacert tartalmazott. Az SpCas9-HF1 esetében a kiegyensúlyozott tesztkészlet 3x 2360 nem vágó, valamint 3x 2360 vágó szekvenciát tartalmazott mivel a 100K adathalmazban csak 3x 2360 nem vágó szekvencia

volt. A WT-SpCas9 esetében a kiegyensúlyozott adatkészlet ugyanebből az okból 3 x 695/695 hasított és nem hasított szekvenciából állt. Minden bakteriális adatkészlet hatékony és nem hatékony szekvenciákra való felosztásához 0,3-as hatékonysági határértéket használtunk.

4.5.2. A nukleotid összetétel (NucCom) kiszámítása, és a Hi-CRISPR A kifejlesztése

A nukleotid összetétel (NucCom) egy olyan index, amelyet az SpCas9 variánsok hatékonyságának predikciójára, a Hi-CRISPR A algoritmusok kifejlesztéséhez használtunk. A NucCom értékek generálásához először egy NucCom szótárt hoztunk létre, amely minden lehetséges permutációban 1-7 nukleotid hosszúságú motívumot tartalmaz a spacer teljes hosszában (pl. NucCom4 esetében AAGC motívum a spacer 1-4 pozíciókban vagy NucCom3 esetében GCC motívum a spacer 18-20 pozícióiban). A NucCom szótár egyes motívumainak kiszámításakor átlagoltuk minden olyan spacer hasítási hatékonyságát az 1M könyvtárban, amely tartalmazta az adott motívumot az adott pozícióban. A könnyebb érthetőségért egy példával illusztrálom a számítást. A példában a 20 bázispár hosszú spacer első 3 bázisainak pozícióiban található „GAA” motívumról szeretnénk eldönteni, hogy az hogyan befolyásolja az SpCas9-HF1 hasítását. Első lépésben kikeressük az SpCas9-HF1-el hasított 1M könyvtárból az összes olyan szekvenciát, amely a „GAA” motívummal kezdődik. Az NGS analízist követően korábban minden spacerre megállapítottunk egy hasítási hatékonyságot jelző értéket. A „GAA” motívummal kezdődő összes spacerhez tartozó hasítási hatékonysági értéket átlagoljuk, és a szótárban rögzítjük ezt az értéket a spacer 1-3-as pozíciójában lévő „GAA” motívumhoz. Ezt a folyamatot megismételjük minden lehetséges pozícióban, minden lehetséges 1-7 nukleotid hosszú motívumra. Így végül kapunk egy NucCom szótárt amelyben szerepel az összes lehetséges 1-7 nukleotid hosszú motívum, melyekhez 2 szám tartozik: az egyik az SpCas9-HF1 hasítási hatékonyságát, a másik a WT-SpCas9 hatékonyságát jelzi.

Ezt követően a szótár felhasználásával ki tudjuk számolni egy konkrét spacer NucCom# értékét. Először a spacerben az adott pozícióban megtalálható motívum értéket kikeressük a szótárból, majd ezeket 1M könyvtár átlagából kivonjuk (átlagtól való eltérés), végül ezeket az értékeket összeadtuk. A jobb érthetőségért itt is egy példával illusztrálom egy spacer NucCom# értékének kiszámítását. A példában a

GAAGCTCATCGTGCGTAGCG szekvenciájú spacer NucCom3 értékét szeretnénk kiszámolni. A NucCom3, három bázis hosszú motívumokon alapszik és a korábban létrehozott szótár segítségével számolhatjuk ki. Először megnézzük, hogy a spacer első három (1-3) pozícióiban milyen szekvencia található, majd kikeressük a szótárból a hozzá tartozó értéket. Itt a „GAA” motívum az első 3 bázis, így a szótárból az ehhez tartozó értéket keressük ki. A következő lépésben a kapott értéket kivonjuk a könyvtár átlagos hasítási értékéből, vagyis abból az értékből, amit akkor kapunk, ha az összes spacer hasítási értékét átlagoljuk. A kapott értéket rögzítjük. A következő lépésben a spacer 2-4 pozícióiban található „AAG” motívumhoz tartozó értékeket keressük ki és kivonjuk a könyvtár átlagából. Így haladunk tovább az utolsó 18-20 pozíciókban lévő „GCG” motívumig. A kapott értékeket összeadjuk így megkapjuk a spacer NucCom3 értékét.

A NucCom értékek kiszámítását R szkriptekkel végeztük, a standard R 3.4.4 verziót alkalmazva külső kódcsomagok nélkül.

Egy adott spacer NucCom4 (vagy Hi-CRISPR A) predikciójára egy Python szkriptet is kifejlesztettünk, amely elérhető a Github-on (<https://github.com/welkergroup/HiCRISPR>). Egy webes felületet is készítettünk az itt kifejlesztett predikciók egyszerűbb használatához a <http://hicrispr.welkergroup.hu/> címen. A Hi-CRISPR A algoritmust Dr. Györgypál Zoltán készítette, a webes felületet pedig Welker Zsombor.

4.5.3. A Hi-CRISPR B és C kifejlesztése

A neurális hálózatok tanításához az 1M könyvtárból az SpCas9-HF1 esetében 1:2 arányban 154147 gyengén és 308294 hatékonyan vágó szekvenciát választottunk ki. A WT-SpCas9 esetén 43204 gyengén és 86408 hatékonyan vágó szekvenciát választottunk.

A Hi-CRISPR B Chuai és mtsai.⁹⁴ által kifejlesztett konvolúciós neurális hálózatra épül, amelyet a bakteriális szelekciós könyvtárak adataival tanítottunk. A kiindulási script itt érhető el: https://github.com/bm2-lab/DeepActiveCRISPR/blob/master/cnn/transfer/ontar_raw_cnn_pretrain.py. A tanítást során a tanító adathalmazt százszor áramoltattuk át (epoch) a hálózaton keresztül.

A Hi-CRISPR C, H. K. Kim és mtsai.⁹⁵ által kifejlesztett neurális hálózatra épül, amelyet a szintén a bakteriális szelekciós könyvtárak adataival tanítottunk. A kiindulási script itt érhető el: <https://github.com/MyungjaeSong/Paired-Library/blob/DeepCRISPR.info/>

DeepCas9/DeepCas9_TestCode.py. Az egyszerre áramoltatott adatpontok száma (batch) 700 volt, amit 2000 iterációval végeztünk. A Hi-CRISPR B és C algoritmus kódja a <https://github.com/welkergroup/HiCRISPR> oldalon érhető el. Ezek az algoritmusok a webes felületen szintén elérhetők a <http://hicrispr.welkergroup.hu/> címen. A Hi-CRISPR B és C algoritmusokat Welker Ágnes készítette.

4.5.4. Predikciók minőségének értékelése

A predikciók minőségét bináris klasszifikációval, a Matthews korrelációs együtthatóval (MCC)⁹⁸ és a G-átlaggal⁹⁹ értékeltük ki. A konfúziós mátrixban (**1. táblázat**), valamint az alábbi képletekben a TP a valódi pozitív (true positive), az FP a hibásan pozitív (false positive), a TN a valódi negatív (true negative) és a FN a hibásan negatív (false negative) predikciókat jelölik. Ezen értékek jobb szemléltetését a konfúziós mátrix mutatja, hogy a prediktált és a valós (mért) értékek hogyan viszonyulnak egymáshoz és melyik jelölés melyik esetre vonatkozik. A pozitív és negatív eseteket egy bizonyos küszöbérték feletti, és alatti értékek alkotják.

1. táblázat. A konfúziós mátrix

A konfúziós mátrixban láthatóak a prediktált és a mért értékek összefüggései. A mért értékeknél meghatározunk egy küszöbértéket, amely felettieket pozitív, alattiakat negatív kimenetelnek határozzuk meg, ebben az esetben azt, hogy az adott szekvencia jól (pozitív) vagy kevésbé jól (negatív) volt hasítható. Az predikciós algoritmusnál is meghatározunk egy küszöbértéket, amely elválasztja a hatékonyak (pozitív) és kevésbé hatékonyak (negatív) prediktált hasítási értéket.

		Mért értékek	
		Pozitív	Negatív
Prediktált értékek	Pozitív	TP	FP
	Negatív	FN	TN

Az érzékenység (szenzitivitás), a specificitás és az ezekből származtatott MCC és G-átlag számítási módját az alábbi egyenletek segítségével számoltuk ki:

$$\text{Érzékenység} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specifititás} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{FP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

$$\text{G} - \text{átlag} = \sqrt{\text{érzékenység} * \text{specifititás}}$$

A HI-CRISPR értékek predikciós küszöbeként azt az értéket választottuk, ahol az érzékenység és specifitás értékei közt a legkisebb különbség található.

Az sgRNA Designer értékek kiszámolásához az Azimuth-ot⁹¹ a DeepCRISPR értékek kiszámolásához a DeepCRISPR v2.0-t használtuk⁹⁴. A CRISPRScan⁸⁹, CRISPRater⁹³, SSC⁹⁰ és sgRNA scorer⁹² értékeket az irodalmi adatokra Haeussler és mtsai. által összegyűjtött adatbázisból¹⁰⁰ gyűjtöttük ki. Dr. Györgypál Zoltán és Welker Ágnes által előzetesen kalkulált és az irodalmi adatok összegyűjtése alapján a predikciókat a szerző határozta meg.

4.6. Sejttenyésztés és transzfekeció

4.6.1. Sejttenyésztés

A munka során felhasznált következő sejtvonalatokat az ATCC-től vásároltuk: N2a (N2a, CCL-131), HeLa (CCL-2), NIH/3T3 (CRL-1658) and HEK293 (CRL-1573). A CST sejtvonalat Dr. Szakács Gergely munkacsoportjától kaptuk. A CST kivételével minden sejtet 10% FBS, 100 U/ml penicillin, valamint 100 µg/ml sztreptomocinnal kiegészített DMEM médiumban tartottam fenn 5%-os széndioxid atmoszférában. A CST sejteket 10% FBS, 100 U/ml penicillin, valamint 100 µg/ml sztreptomocinnal komplettált DMEM-F12 médiumban tartottam fenn 5%-os széndioxid atmoszférában.

4.6.2. Önvágó plazmid kísérletek transzfekeciója és puromicin szelekció

A HeLa és N2a sejteket a transzfekeció előtti nap $1,5 * 10^5$ sejt/well, a CST és NIH/3T3 esetében pedig $3 * 10^5$ sejt/well sejtszámmal ültettem ki 6 well plate-re. Összesen 2 µg DNS-sel transzfekektáltam a sejteket wellenként, 1 µg Cas9-gRNS vektort és 1 µg önvágó vagy HR donor vektort mértem össze a transzfekeciós reagenssel 400 µl szérumentes DMEM-ben. A transzfekeciós reagens a HeLa, N2a és HEK293 sejtek esetében 6 µl

Turbofect/well volt, a NIH/3T3 és CST sejtek esetében 9 μ l Lipofectamine 2000/well. Minden transzfekciót három párhuzamos kísérletben végeztem. A transzfekció hatékonyságát a transzfekciót követő második napon detektáltam a tranziens expresszió mérésével.

A degradációs doménnel (DD) ellátott önvágó plazmid kísérleteknél 1 μ g Cas9-gRNS, 700 ng pSc1-DD és 300 ng mCherry plazmidot használtam a fenti transzfekciós protokollal. Itt az mCherry jelet használtam a transzfekció hatékonyságának monitorozására (a DD miatt a GFP-t itt nem lehetett használni).

A puromicinnel végzett szelekciónál először megállapítottam, hogy mely az a legkisebb koncentrációjú antibiotikum, amely két hét alatt eliminálja az összes nem transzfektált sejtet egy 0-tól 3,1 μ g/ml koncentrációjú hígítási sorral. Ezt 0,7 μ g/ml-ben állapítottam meg. A fenti transzfekciós protokollal transzfektáltam a sejteket, a célzott módosításhoz a pten-gRNS Cas9 plazmidot használtam, a nem célzotthoz a pten-gRNS dCas9 plazmidot. A puromicin szelekció a transzfekciót követő második naptól két hétig folyt, majd a túlélő GFP pozitív sejteket áramlási citométer segítségével detektáltam.

4.7. *GFxFP esszé*

N2a sejteket a transzfekció előtt 1 nappal $3 \cdot 10^4$ sejt/well sűrűségben ültettem ki. A következő mennyiségű plazmidokat 50 μ l szérumentes DMEM-ben kihígítottam: 10 ng GFxFP plazmid, 150 ng Cas9 kódoló plazmid (SpCas9-HF1 vagy dSpCas9) és 90 ng gRNS kódoló plazmid (kanonikus vagy stgRNS). Az elegyhez 1 μ l Turbofect transzfekciós reagenst mértem, és 30 percig szobahőmérsékleten inkubáltam, majd a sejtekhez mértem. Minden transzfekciót 3 párhuzamos kísérletben végeztem el. A transzfekció után 3 nappal a sejteket áramlási citométerrel vizsgáltam. A GFxFP plazmid rendelkezik egy alacsony háttér GFP fluoreszcenciával, ezt mérendő a GFxFP plazmidokat a nukleáz SpCas9-HF1 mellett kontrollként dSpCas9-cel is transzfektáltam. Az ebből mért háttér fluoreszcenciát kivontam az SpCas9-HF1-el kezelt mintákon mért értékekből. Az ábrákon negatív érték is szerepelhet, ez olyankor lehetséges, amikor a háttér fluoreszcencia egy kissé magasabb értéket mutatott, mint az SpCas9-HF1-el mért. A **24/A ábrán** bemutatott 57 spacer a humán vagy egér genomban meglévő szekvenciák, amelyeket a BiSearch¹⁰¹ segítségével választottunk ki az 1M könyvtárból. Ezeket a

szekvenciákat később az 1M könyvtárból eltávolítottuk, hogy ne befolyásolják az algoritmusok tanítását, valamint adjanak ezért fals predikciót. A GFxFP esszé vázlatos működése a **6. ábrán** látható.

4.8. Áramlási citometria

Az áramlási citometriás méréseket egy Attune NxT citométerrel (Life Technologies) végeztük. Az adatok kiértékeléséhez az Attune Cytometric szoftvert használtuk. Minden kísérletben 5000-10000 élő, önálló sejtet mértünk le, amelyeket az előre és oldalra irányuló fényszórási (FSC, SSC) értékek alapján kapuztam. A GFP-t a 488 nm-es lézerrel gerjesztettem és az 530/30nm-es szűrővel detektáltam. A mCherry-t az 561 nm-es lézerrel gerjesztettem és az 615/20 nm-es szűrővel detektáltam.

4.9. Konfokális mikroszkópia

A konfokális mikroszkópos vizsgálatok előtt a CST sejteket 4%-os paraformaldehiddel fixáltam, majd a sejtmagokat DAPI-val festettem. A képet egy Zeiss LSM 710 konfokális mikroszkóppal készítettük Plan-Apochromat 40x/1,4 olajimmerziós objektívvel. A DAPI-t a 405 nm-es lézerrel gerjesztettük és 410-503 nm-es szűrővel detektáltuk, a GFP-t a 488 nm-es lézerrel gerjesztettük és 515-565 nm-es szűrővel detektáltuk. A konfokális mikroszkóppal készült képet (**19. ábra**) Bartos Zsuzsa készítette.

4.10. Genomi DNS izolálás

A genomi DNS-t a PureLink Genomic DNA Mini Kit (Invitrogen) segítségével tisztítottam ki a gyártó protokollja alapján. Az izolált DNS-ről a genomi PCR-eket Q5 polimerázzal végeztem.

4.11. TIDE és T7 endonukleáz I esszé

A genomi DNS-ről készült PCR termékeket agaróz gélen futtattam, majd NucleoSpin Gel and PCR Clean-up kit segítségével tisztítottam a gyártó előírásai szerint. A PCR termékek Sanger szekvenálása után (Microsynth AG.) az indel hatékonyságokat a TIDE webes eszköz segítségével analizáltam (<https://tide.nki.nl/>)⁸⁰. A kezelt minták mellett mindig

készítettem egy Cas9-cel nem kezelt kontroll PCR terméket is, mert a TIDE működése során szükség van egy kontroll kromatogramra.

A T7 endonukleáz I esszéhez 200 ng PCR terméket 5 egység T7EI enzimmal 2 óráig 37 °C-on emésztettem a gyártó protokollja alapján, majd agaróz gélen futtattam meg a mintákat. A minták hasításának mértékét az ImageJ szoftverrel határoztam meg a következő képlet alapján, ahol b és c a hasított, a pedig a kiindulási fragmens mennyiségét jelzi:

$$\text{InDel \%} = 100 * \left(1 - \sqrt{\left(1 - \left(\frac{b+c}{a+b+c} \right) \right)} \right)$$

4.12. Statisztika

Egyenlőtlen varianciájú csoportok átlagai közt lévő eltéréseket egyszempontú Welch Anova analízissel vizsgáltuk Games-Howell post-hoc teszttel, valamint a homoszkedasztikus minták esetében egyszempontú Anova-t használtunk Tukey post-hoc teszttel. A szóráshomogenitás feltételét Levene tesztjével ellenőriztük. A statisztikai próbákat az SPSS v.20 (IBM) programmal végeztük.

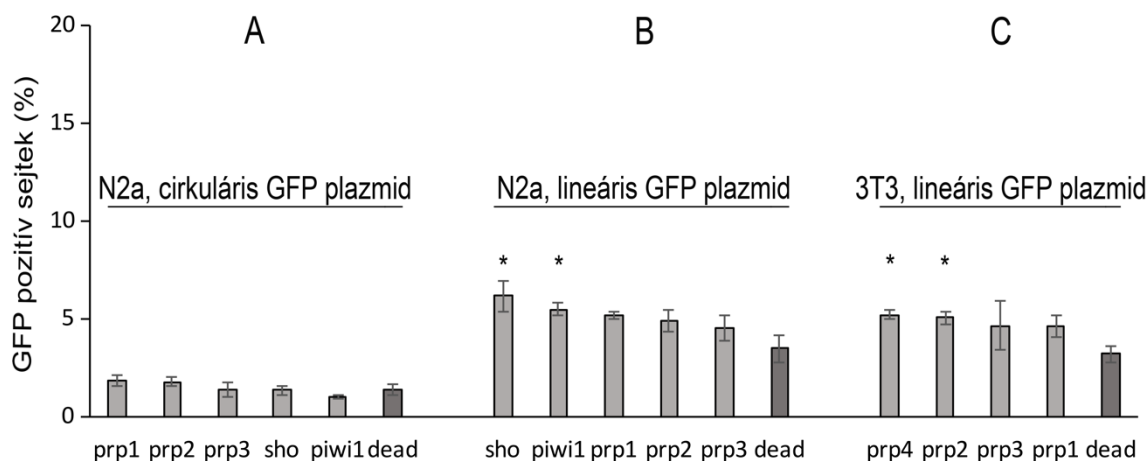
5. Eredmények

5.1. *gRNS-ek hatékonyságának meghatározása önvágó plazmidokkal*

5.1.1. Egy plazmid sejten belüli hasítása megkönnyíti annak integrációját a genomba

A gRNS-ek működésének hatékonyságát úgy szerettem volna követni, hogy egy GFP fluoreszcens fehérje kódoló szekvenciáját tartalmazó plazmidot SpCas9 segítségével beépítsek a genomba az NHEJ DNS hibajavítás segítségével. Minél aktívabb egy gRNS annál több integrációs esemény következhet be és annál több GFP pozitív sejtet kapunk a folyamat végén. A genomban integrált kópia mellett a plazmidon kódolt fehérje a plazmid lebomlásáig kifejeződik, ezt nevezzük tranziens expressziónak. A beépült plazmidokról történő expresszió követéséhez, a tranziens expresszió lecsengését meg kell várnunk. Abban nem lehettünk biztosak, hogy a plazmid integrációja, olyan hatékony lesz az SpCas9 által létrehozott DSB helyén, hogy fluoreszcens módszerekkel elkülöníthető lesz az expressziós szint a háttértől, amely a genomba véletlenszerűen integrálódó plazmidoktól származik. Ennek kiderítésére az SpCas9-et egy genomi célszekvenciát kódoló gRNS-el, valamint egy cirkuláris GFP expressziós plazmiddal ko-transzfectáltam N2a (egér neuroblasztóma) sejtekbe. A transzfectiót követően (a tranziens expresszió lecsengése után) az integrált plazmidokat a GFP pozitív sejtek detekciójával határoztam meg áramlási citometria segítségével. Kontrollként egy nukleáz inaktív dSpCas9-et használtam, melyet a tranziens expresszió lecsengésének követésére használtam. A tranziens expresszió kb. 12-15 nap alatt csengett le, ezután detektáltam az integrált plazmidokat az itt bemutatott kísérletekben. A transzfectió hatékonyságát pedig a tranziens expresszió mérésével határoztam meg a transzfectiót követő 2. napon.

Öt genomi szekvenciát célzó gRNS-t (melyek az egér genomban a *Prnp*, *Sprn* és *Piwil2* géneket célozzák), valamint az SpCas9-et kódoló plazmidokat a GFP plazmiddal ko-transzfectáltam N2a sejtekbe. Az SpCas9-cel célzott kondíciókat nem lehetett elkülöníteni a dSpCas9-cel kezelt kontrolltól, mely a véletlenszerű integrációt mutatta. (7/A ábra).



7. ábra. SpCas9-cel célzott N2a és NIH/3T3 sejtekben a cirkuláris vagy lineáris plazmidok integrációja nem hatékony

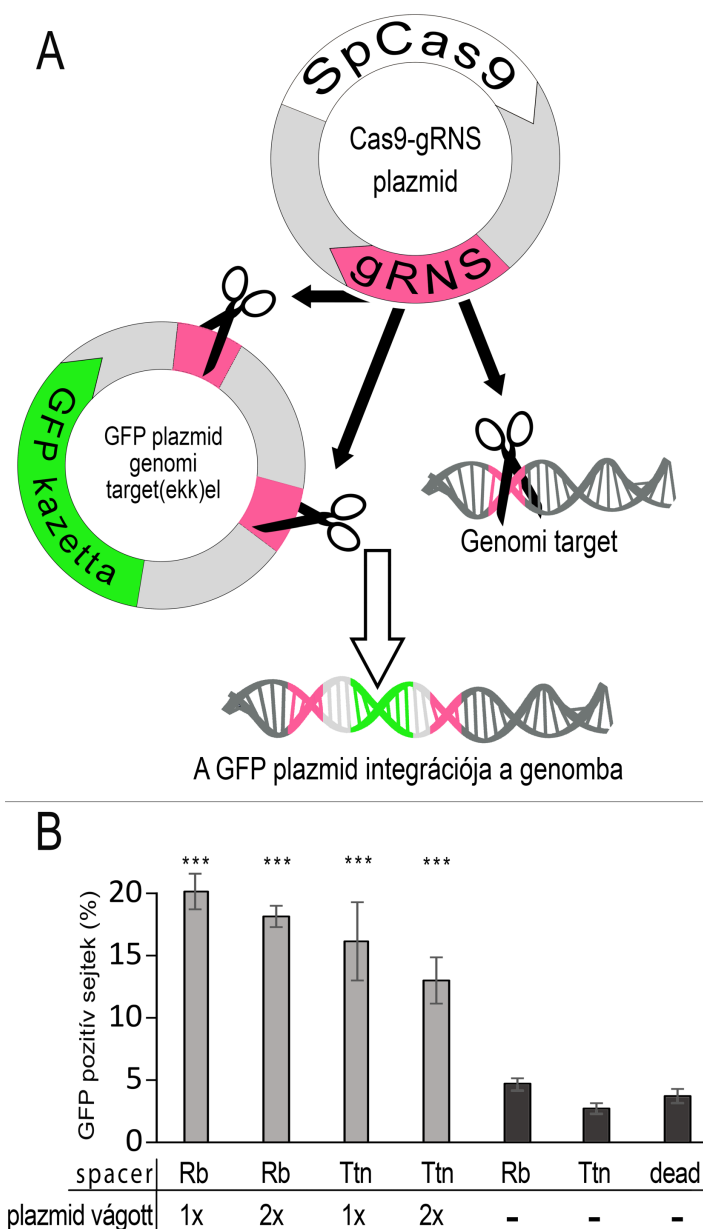
A GFP pozitív sejtek százaléka (a tranziens GFP expresszió lecsengése után) N2a (A, B) és NIH/3T3 sejtekben (C), amennyiben egy cirkuláris (A) vagy lineáris (B, C) GFP plazmidot integráltunk célzott genomi lókuszokra, SpCas9 és különböző gRNS-ek segítségével. Negatív kontrollként egy nukleáz inaktív (dead) SpCas9-et használtam, mely a plazmidok véletlenszerű genomi integrációjából származó GFP pozitív sejtek százalékát adja meg. Az oszlopok $n=3$ független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszívek a szórást jelölik. A csillagok a véletlenszerű integrációs értéktől való eltérést jelzik. *: $p < 0,05$

Azt reméltük, hogyha egy restriktív enzim segítségével linearizált plazmidot juttatunk a sejtbe, akkor az integráció hatékonyabban fog megtörténni, mint a cirkuláris plazmidok esetén. Mivel a szabad DNS végek a linearizált plazmid esetében jobban kitettek a sejtben jelenlévő exonukleázoknak, a restriktív hasítási helyet úgy választottam ki, hogy a GFP expressziós kazetta mindkét végén még kb. 1000-1000 bázis maradjon. Az előzőekhez hasonlóan transzfektáltam a linearizált plazmidokat N2a, valamint NIH/3T3 egér fibroblaszt sejtekbe és követtem a plazmid integrációját. A linearizált plazmid hatékonyabban épült be mindkét sejt genomjába, viszont a cirkuláris plazmidokhoz hasonlóan, a célzott integráció mennyisége csak nehezen vagy egyáltalán nem volt elkülöníthető a háttértől: a véletlenszerű integráció mértéke is megnőtt (7/B, C ábra).

Korábbi tanulmányokban hatékonyabb volt egy donor DNS integrációja, amikor a donor plazmidot a sejtben belül linearizálták designer nukleázokkal¹⁰²⁻¹⁰⁴. Hogy megvizsgáljuk ezt a lehetőséget is, két hatékony gRNS-t expresszáló plazmidot választottunk ki (melyek két gént *Rbl2* és *Ttn* céloznak az egér genomban), és ezek célszekvenciáját a GFP-plazmidra klónoztam a GFP-kazetta elé 5' irányban, vagy mögé 3' irányban, illetve

mindkét helyre (**8/A ábra**). Ezeket, illetve a kontroll (célszekvencia nélküli) GFP-plazmidot SpCas9-cel és a megfelelő gRNS-t (Rb vagy Ttn) kódoló plazmiddal együtt ko-transzfektáltam N2a sejtekbe. A plazmidok sejten belüli hasítása négy-hatszoros célzott integrációt eredményezett a random integráció háttérszintjéhez képest (**8/B ábra**).

Azt vártuk, hogy amikor a GFP-kazettát mindkét oldalán hasítjuk, az több sejten belül linearizált plazmidhoz vezet és így nagyobb integrációt eredményez. Az elképzeléseinkkel szemben nem észleltünk nagyobb fluoreszcenciát a két targetet tartalmazó plazmidokkal, így a további kísérletekben a plazmidot csak az egyik oldalán nyitottuk fel.



8. ábra. Egy plazmid célzott integrációja megnövelhető a plazmid sejten belüli hasításával

(A) A sejten belül hasított és genomban integrált plazmid működésének vázlata látható. Az SpCas9-et és egy genomi célszekvenciát kódoló gRNS, (magenta) mind a GFP-kazettát (zöld) tartalmazó plazmidot, mind a genomban található, ezzel megegyező szekvenciát hasítja. A létrejött DNS törés helyére a felnyitott plazmid hatékonyan beépül.

(B) GFP pozitív sejtek százalékos aránya látszódik, amikor egy sejten belül felnyitott GFP plazmidot az SpCas9 segítségével integrálunk a genomban. Ezt az *Rbl2* vagy a *Ttn* gént célozva is megvizsgáltuk. Az oszlopok alatt látható számok jelölik, hogy az adott kondícióban az integrált plazmid hány helyen lett felnyitva. Negatív kontrollként egy nukleáz inaktív (dead) SpCas9-et használtunk melynél a plazmidok véletlenszerű genomi

integrációjából származó fluoreszcens jel látható. További kontrollként aktív SpCas9-cel céloztuk a két gént (sötétszürke oszlopok, Rb, Ttn), viszont a plazmidokat nem nyitottuk fel a sejtben (nincs rajtuk célszekvencia), ez mutatta a cirkuláris plazmidok beépülését ugyanazon kondíciók mellett. Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszázlók a szórást jelölik. ***: $p < 0,001$.

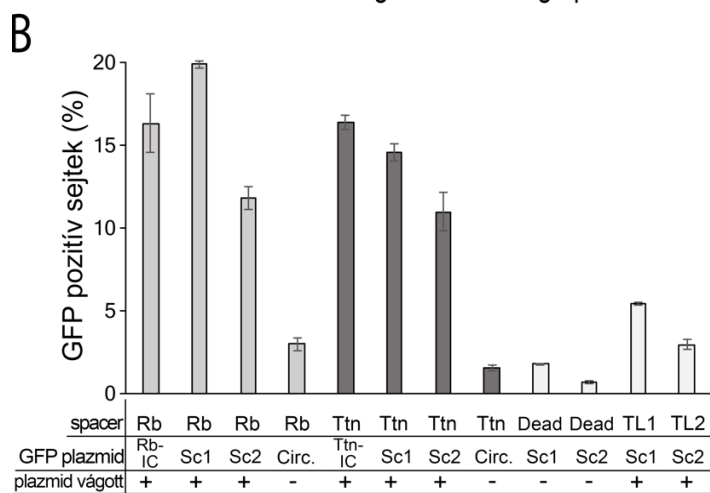
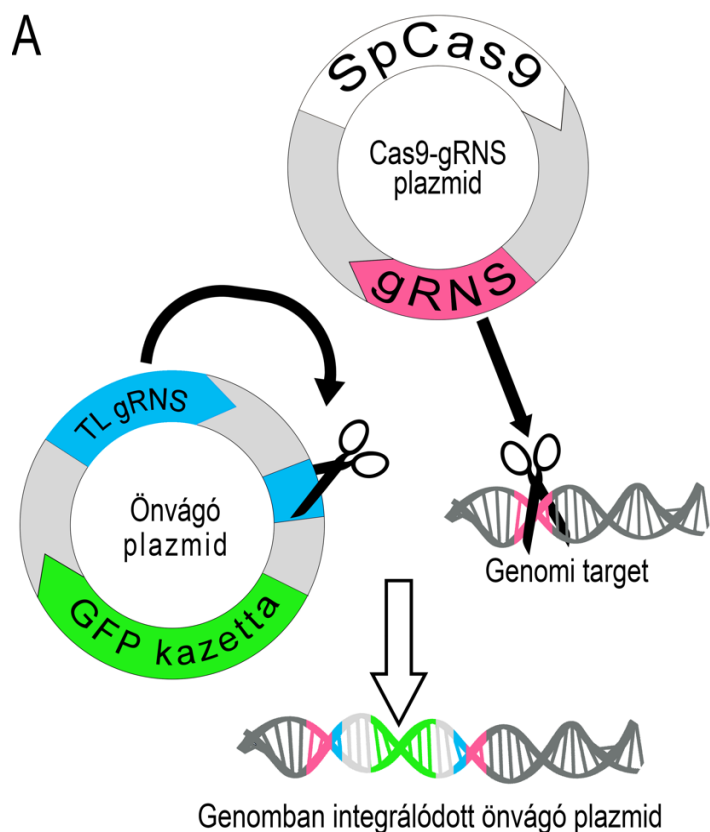
5.1.2. Az önvágó plazmid konstrukciója univerzális spacer és célszekvencia párokkal

Habár a fenti eredmények biztatók és azt mutatták, hogy a sejtben belül linearizált plazmidok homológ karok nélkül is hatékonyan integrálhatók a genomba, a jelenleg alkalmazott megközelítés nem teszi lehetővé, hogy egy gRNS tényleges hatékonyságát mérje. Ennek oka, hogy ugyanaz a gRNS célozza a genomot és a plazmidot, így egy összefüggő eseményt mérünk és ezért azt is nehéz meghatározni, hogy a plazmid sejtben belüli hasításával mekkora a véletlenszerű integráció (háttér). A rendszer következő korlátja, hogy a célzott integráció hatékonysága nem csak a genomot célzó gRNS-től függ, hanem a plazmid vágásának hatékonyságától, amely különböző gRNS használatokor eltérő lehet. További negatívum, hogy a vizsgálandó szekvencia targetét mindig újra kell klónozni a plazmidon és ez limitálja a módszer használhatóságát.

Annak érdekében, hogy a különböző gRNS-ek genomi működése összehasonlíthatóbb legyen, egy olyan általános GFP-plazmidot terveztem, amely a genomi célszekvenciáktól függetlenül felnyitható, minden genomot célzó gRNS esetében ugyanaz, így ezek összehasonlítása könnyebbé válik. Továbbá, egy ilyen általános szekvencia használatával nem kellene minden vizsgált célszekvenciánál egy új GFP plazmidot létrehozni molekuláris klónozással. Mindemellett, hogy még általánosabban használható legyen, a plazmidot célzó gRNS kódoló kazettát beépítettem a GFP-t kódoló plazmidba. Így létrehoztam egy „önvágónak” elnevezett plazmidot, mely a saját linearizálásához szükséges gRNS expressziós kazettát és a célszekvenciáját is tartalmazza (9/A ábra).

Az önvágó plazmidon található univerzális célszekvencia kiválasztásához, véletlenszerűen generáltam 30000 spacer szekvenciát és ezek közül kiválasztottam azokat, amelyek a legkevesebb off-targettel rendelkeztek az eCRISP¹⁰⁵ off-target kereső program szerint, valamint a leghatékonyabban hasítanak az sgRNA Designer program szerint⁹¹. Ezek közül kiválasztottam azt a 2 spacer, amelynek a CasFinder¹⁰⁶ szerint nincs

célszekvenciája az egér, a humán és további 10 élőlény genomjában. Ezeket célszekvencia nélküli 1 és 2-nek (target less – TL1 és TL2) neveztem el, és ezeket felhasználva hoztam létre a két önvágó plazmidot (pSc1, pSc2, Addgene #80436, #80437).



cel hasítottuk és a vágás helyére mindkét önvágó (Sc1, Sc2), valamint a **8. ábrán** bemutatott, a sejtben felnyitott, de nem önvágó (Rb-IC, Ttn-IC) plazmidok integrációjának hatékonyságával hasonlítottuk össze. Az önvágó plazmidok véletlenszerű integrációs háttérének mérésére egy aktív SpCas9-et expresszáló vektort a megfelelő TL gRNS-ekkel (utolsó két oszlop: TL1, TL2)

9. ábra. Az önvágó plazmid célzott integrációja a genomban

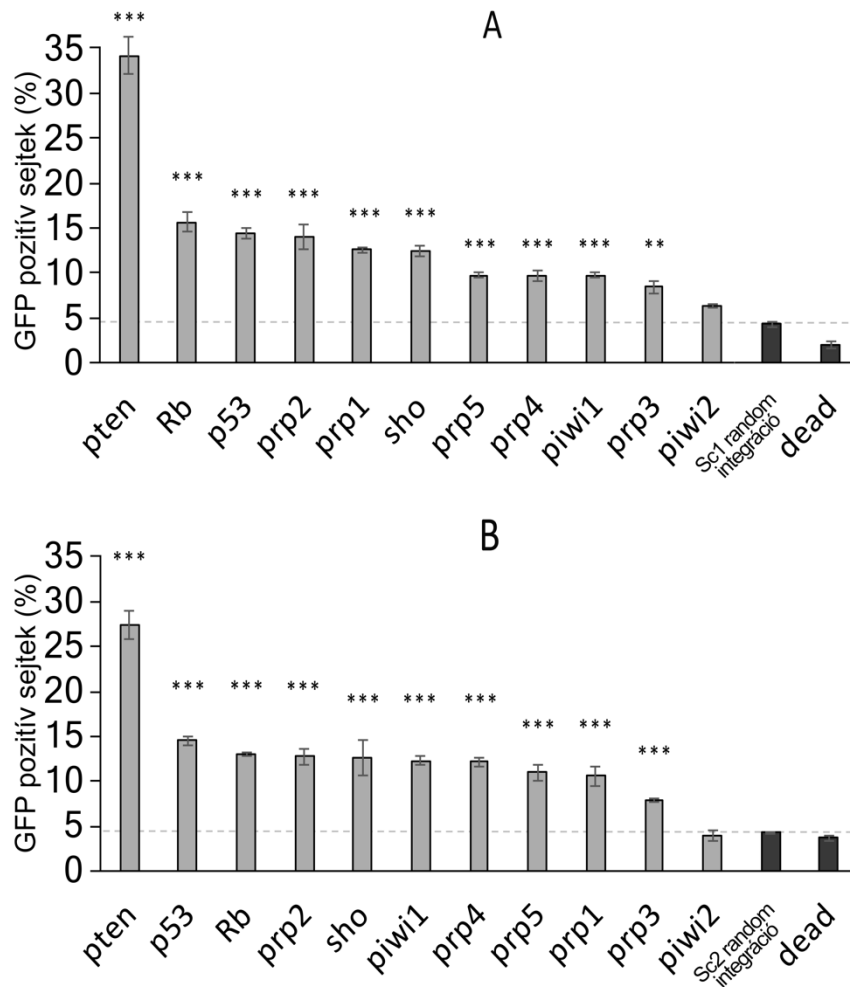
(A) Az ábrán az önvágó plazmid genomi integrációjának vázlatja látható. Az SpCas9-et és a gRNS-t kódoló plazmid mely a genomi szekvenciát (rózsaszín) tartalmazza, a genom szekvencia-specifikus hasításáért felel. Az önvágó, GFP expressziós kazettát (zöld) hordozó plazmidon egy univerzális spacer kódoló gRNS (TL-gRNS, kék), valamint ennek célszekvenciája is megtalálható. A két plazmidot a sejtekbe juttatva, az SpCas9 hasítja a genomi targetet, valamint a TL gRNS segítségével az önvágó plazmidot intracellulárisan linearizálja, így az hatékonyan be tud épülni a genom megcélzott helyére.

(B) Az ábrán GFP pozitív N2a sejtek százalékos aránya látható, amikor az önvágó plazmid, valamint különböző GFP expressziós plazmidok integrálódnak a sejtek genomjába. Az *Rbl2* gént az Rb gRNS segítségével (szürke), valamint a *Ttn* gént a Ttn gRNS segítségével (sötétszürke) valamint az SpCas9-

együtt transzfektáltunk; így az önvágó plazmidok a sejtben fel tudnak nyílni, de a genom nincs célozva. További kontrollként az SpCas9-cel a két gént célozva megvizsgáltuk a cirkuláris GFP plazmid beépülésének hatékonyságát annak felnyitása nélkül (Rb, Ttn, Circ.) Negatív kontrollként egy inaktív SpCas9-et (Dead) alkalmaztunk a két önvágó plazmiddal, amely a cirkuláris plazmid random beépülésének mértékét mutatja, annak célzott felnyitása nélkül (Dead, Sc1, Sc2). Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszázalék a szórást jelölik.

Az önvágó plazmidokat az Rb és Ttn gRNS-sel teszteltem, valamint összehasonlítottam a korábban használt sejtben linearizált plazmidokkal (**9/B ábra**). A kísérletben kontrollként, a felnyílt önvágó plazmidok véletlenszerű genomi integrációját úgy vizsgáltuk, hogy a sejtekbe csak a plazmid linearizálásért felelős TL gRNS-t juttattuk be, de a genom nem volt célozva. Az önvágó plazmidok mindkét genomi target esetében hasonló hatékonysággal integrálódtak a genomban, mint a sejtben felnyitott, de nem önvágó plazmidok. Ebből arra következtettünk, mindkét TL spacer tartalmazó gRNS hasonló hatékonysággal hasítja a hozzá tartozó önvágó plazmidot, mint az Rb vagy Ttn spacerok. Ezért úgy gondoltuk, hogy ezek az önvágó plazmidok hatékonyan integrálhatók, miközben a különböző gRNS-ek hatékonyságát már univerzálisan lehet velük mérni.

A két univerzális önvágó plazmidot (pSc1 és pSc2) további 11 spacer tesztelésére használtam, melyek 7 különböző gént céloztak meg az egér genomban (*Pten*, *Rbl2*, *Tp53*, *Prnp*, *Sprn*, *Piwil2* és *Piwil4* – **10. ábra**). A sejtben belül hasított önvágó plazmid véletlenszerű integrációjából származó háttért meghaladó GFP pozitív sejtek számát vettük célzott integrációnak. Az itt használt kontrollnál csak a plazmidot hasítjuk, de a genomot nem. Ez a 4-5 %-os háttér magasabb volt, mint a cirkuláris (1-2 %) vagy restriktív enzimmel linearizált (2-3,5 %) plazmid véletlenszerű integrációjakor mért érték. A közel 2 %-os növekedés valószínűleg a megnövekedett mértékű véletlenszerű integrációból ered, melyet az intracelluláris hasítás indukál, de az is elképzelhető, hogy a TL gRNS a tervezés ellenére a genomban off-target helyeken is hasít.

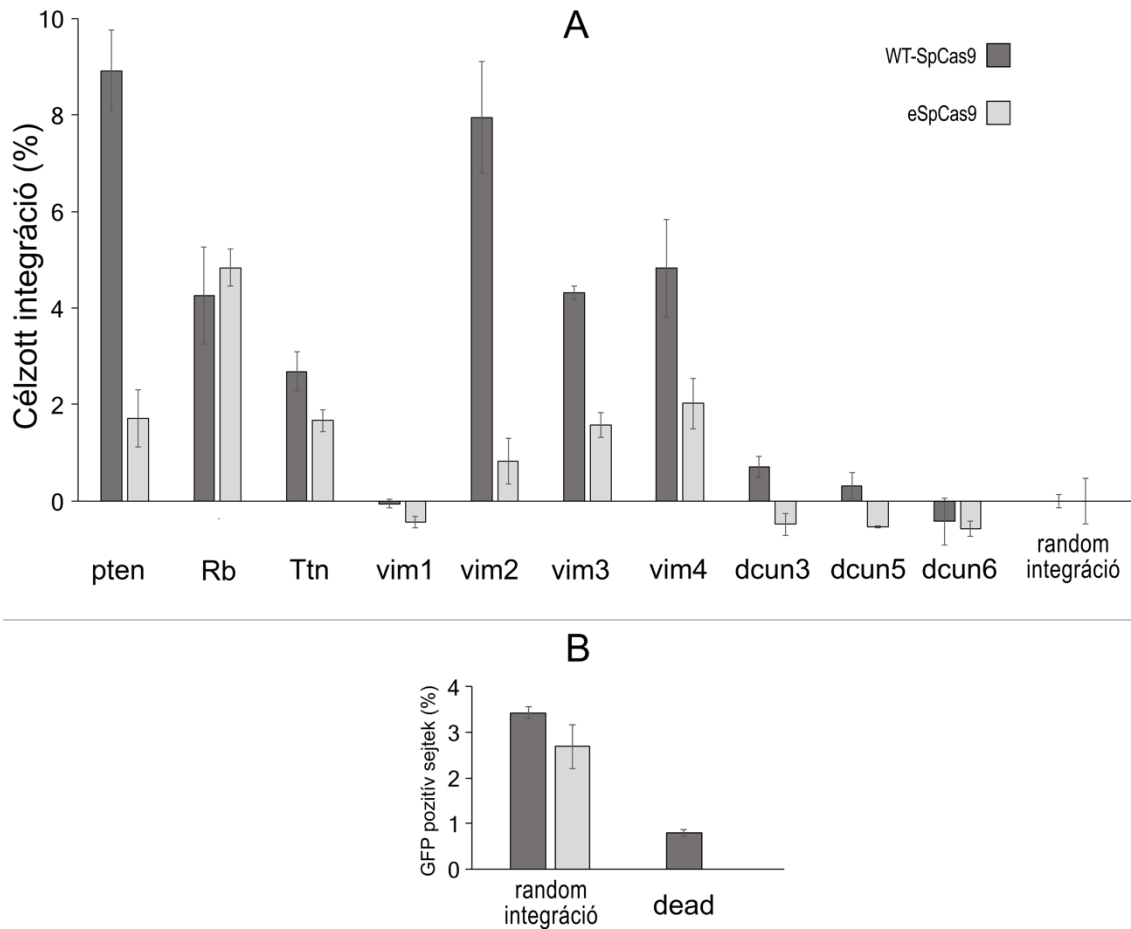


10. ábra. Különböző spacerek hatékonysága mindkét önvágó plazmiddal vizsgálva, N2a sejtekben

Különböző egér genomban található célszekvenciák hasítási hatékonyságát vizsgáltuk az önvágó 1 (Sc1 - **A**) illetve az önvágó 2 (Sc2 - **B**) plazmid segítségével. Az ábrán GFP pozitív N2a sejtek százaléka látható, mely az SpCas9-cel célzott genomi helyekre integrálódott önvágó plazmidokból adódik. Mindkét önvágó plazmiddal hét génben található 11 célszekvenciát vizsgáltunk. A célzott *Pten*, *Tp53*, *Rbl2*, *Prnp*, *Sprn*, *Piwil2* és *Piwil4* génekhez a *pten*, *Rb*, *p53*, *prp1-5*, *sho*, *piwi1* és *piwi2* gRNS-ek tartoznak sorrendben. Az önvágó plazmidok véletlenszerű integrációs háttérnek mérésére egy aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot intracellulárisan linearizáltuk, de a genom nem volt megcélözva (random integráció). Ennek a háttérnek a mértékét szaggatott vonallal jelöltük. Az ezektől szignifikánsan eltérő értékeket vettük célzottan integrált eseményeknek. További negatív kontrollként egy inaktív SpCas9-et alkalmaztunk a két önvágó plazmiddal, amely a cirkuláris plazmid random beépülésének mértékét mutatja. Az oszlopok $n=3$ független transzfekcióban mért GFP pozitív sejtek átlagát, a hibazászlók a szórást jelölik. A csillagok a véletlenszerű integrációs értéktől való eltérést jelzik. **: $p<0,01$ ***: $p<0,001$.

Egy spacer kivételével (piwi2) mindegyik target a véletlenszerű integráció értéke feletti, különböző mértékű célzott integrációt mutatott mindkét önvágó plazmival. A leghatékonyabb spacer mindkét önvágó plazmid esetén a pten spacer volt, amely körülbelül 25 %-os célzott integrációt eredményezett. A többi spacer mindkét önvágó plazmival összehasonlítható módon működött.

A következő kísérletben az előzőek során használt három (pten, Rb, Ttn), valamint további hét gRNS (vim1-4, dcun3,5,6 melyek a *Vim* és *Dcun1d2* géneket célozzák) hatékonyságát teszteltem a megnövelt pontosságú eSpCas9, valamint az eddig használt vad típusú SpCas9-cel (**11/A ábra**). Az irodalmi adatoknak megfelelően^{72,77} az eSpCas9 hatékonysága átlagosan gyengébb volt, mint a WT-SpCas9. Az Rb gRNS-el azonos volt a két nukleáz aktivitása, míg 4 gRNS-sel (vim1 és dcun3, 5 és 6) nem volt mérhető aktivitás. Az intracellulárisan hasított, de a genomban célzottan nem integrált kontrollnál, az eSpCas9-nél kisebb háttérrel lehetett tapasztalni, amely vagy annak köszönhető, hogy ez a nukleáz a plazmidot is kevésbé hatékonyan hasította, vagy pedig az off-target helyekre kevésbé épült be a plazmid, a nagyobb hűségű nukleáznak köszönhetően (**11/B ábra**).



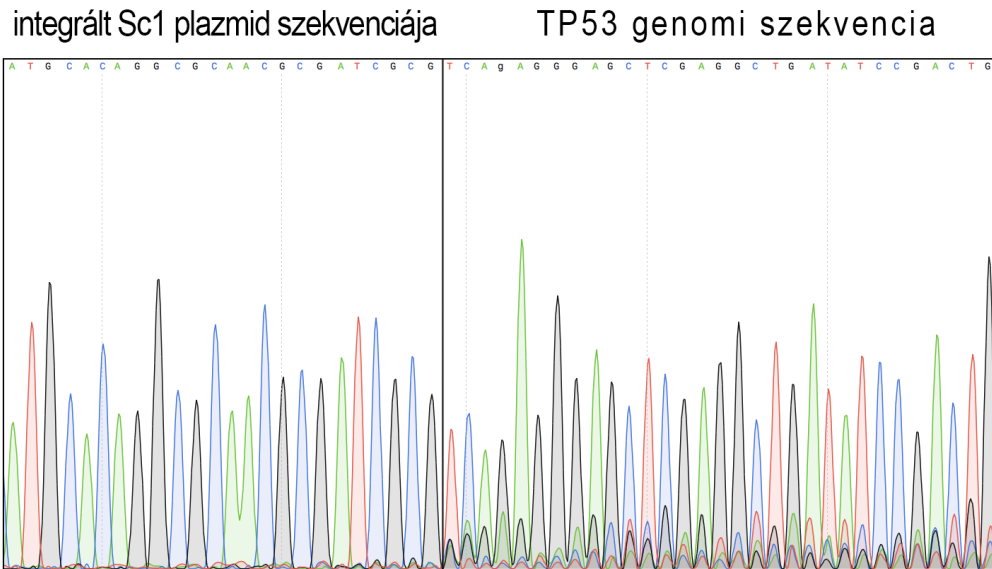
11. ábra. Az önvágó plazmid beépítésének hatékonysága WT- és eSpCas9-cel

(A) Különböző egér genomban található célszekvenciák hasítási hatékonyságát vizsgáltuk WT-SpCas9-cel (sötétszürke) valamint eSpCas9-cel (világosszürke) az önvágó plazmid segítségével. Az ábrán GFP pozitív N2a sejtek százalékos aránya látható, amikor a két SpCas9 variánssal célzott genomi targetekre az önvágó plazmidot integráltuk. Mindkét SpCas9-cel öt génben található tíz célszekvenciát vizsgáltunk. Az önvágó plazmid véletlenszerű integrációs háttérének mérésére mindkét aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélözva (random integráció). Ennek értékét az A ábrán már kivontuk a mért értékekből, így ezek már a célzott integráció mértékét mutatják. (B) Az önvágó plazmid véletlenszerű integrációját ábrázoltuk, amikor a plazmid sejten belül fel van nyitva (random integráció) illetve amikor a plazmid cirkuláris marad, mert inaktív SpCas9-cel volt ko-transzfektálva (dead). Az oszlopok $n=3$ független transzfektációban mért GFP pozitív sejtek átlagát, a hibaszívek a szórást jelölik.

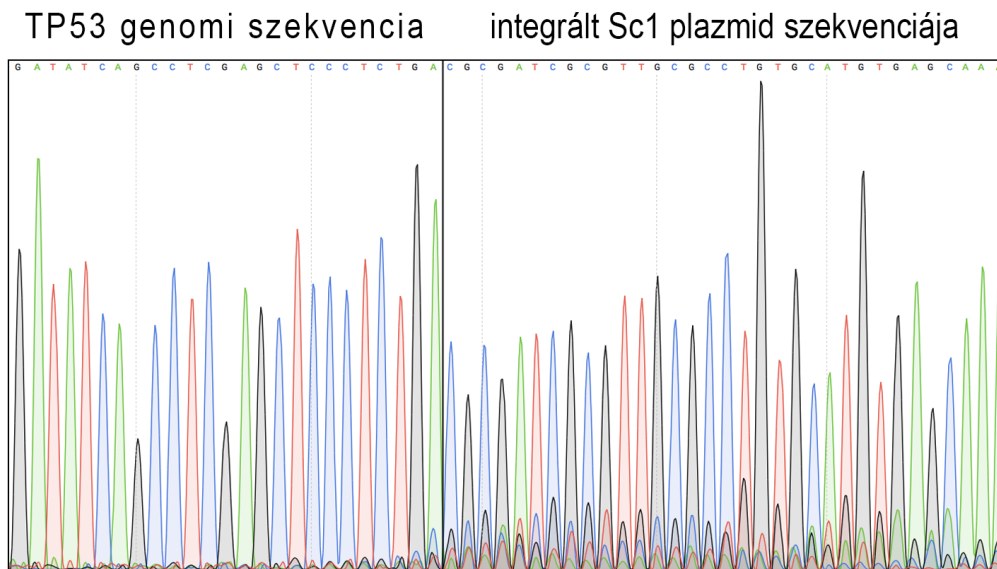
A következőkben megvizsgáltam, hogy a célzott hasítás helyén a plazmid valóban beépült-e a genomba, a hasított plazmid végeire és a genomban célzott helyre tervezett primerek és PCR segítségével (12. ábra). Genomi DNS-t izoláltam olyan sejtpopulációból, ahol az önvágó plazmidot a *Tp53* lókuszra építettem be, PCR-rel

amplifikáltam, majd Sanger szekvenáltam a plazmid és genom csatlakozási helyét. A PCR termék szekvenálása megmutatta, hogy a beépült plazmidok többsége pontosan (hiba nélkül) épült be a genom célzott helyére.

A



B



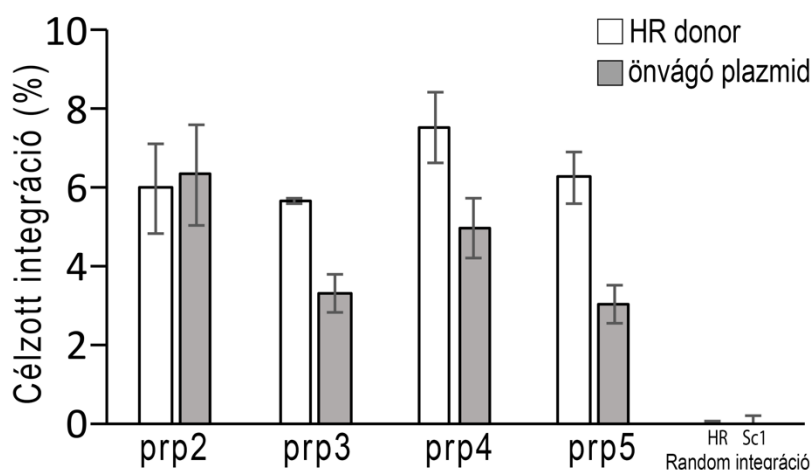
12. ábra. A *Tp53* génbe illesztett önvágó plazmid csatlakozási pontjának szekvenálása

A kromatogramok a *Tp53* gén részleges szekvenciáját mutatják, azon a csatlakozási ponton, ahova az önvágó plazmid (Sc1) integrálódott. Az integráció helyét a génre és az önvágó plazmidra

tervezett PCR primerekkel amplifikáltuk, majd Sanger szekvenáltuk. Az ábrán a plazmid felől (A), illetve a genomi szekvencia felől (B) indított szekvenálásból származó kromatogramok látszódnak. A csatlakozási pontot egy fekete vonal jelöli.

5.1.3. A módszer összehasonlítása a homológ rekombinációval, valamint működése további sejtvonalakban

Négy a prionfehérjét kódoló gént (*Prnp*) célzó spacer (prp2-prp5) relatív hatékonyságát hasonlítottuk össze, egy GFP kazetta integrációjával homológ rekombináció, illetve NHEJ-közvetített beépüléssel (13. ábra). A prionfehérje génjében olyan célszekvenciákat választottam ki, amelyek hasítási pontja néhány bázispáron belül helyezkedik el, így a homológ rekombinációt nem befolyásolhatja az, hogy a homológ kar szekvenciáktól a hasítási pont esetleg messzebb található. Mind a négy szekvencia hatékonyan működött, így mind az önvágó, mind a HR-karokkal rendelkező konstrukció beépült a megcélzott helyekre.



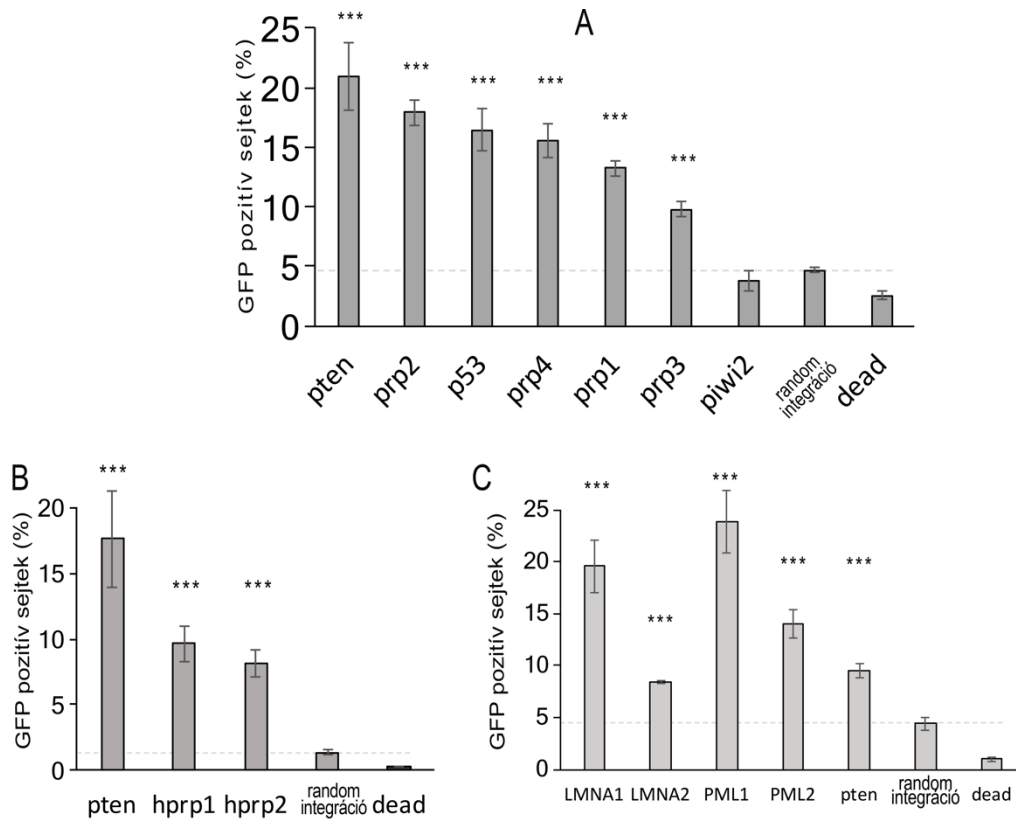
13. ábra. Az önvágó plazmid genomi integrációjának hatékonysága összemérhető a HR-mediált integráció hatékonyságával

Az ábrán GFP pozitív N2a sejtek százalékos aránya látható, amikor egy homológ karokkal rendelkező GFP expressziós kazettát tartalmazó plazmidot homológ rekombináció segítségével (fehér oszlopok); valamint az önvágó plazmid segítségével NHEJ-mediált integrációval juttattunk a *Prnp* gén különböző helyeire. A *Prnp* génben négy különböző gRNS-t használtunk az expressziós kazetták célzott integrációjára. A HR-mediált integrációk kontrolljaként a HR donor plazmidot egy inaktív SpCas9-cel ko-transzfectáltuk (HR, random integráció). Az erre a kontrollra kapott értékeket (0,1 %) a célzott integrációk értékéből kivontuk és így ábrázoltuk. Az önvágó plazmid véletlenszerű integrációs háttérének mérésére az SpCas9-cel, valamint az önvágó

plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélózva (Sc1, random integráció). Ennek értékét (5 %) a célzott integrációk értékéből kivontuk és így ábrázoltuk. Az oszlopok n=3 független transzfeccióban mért GFP pozitív sejtek átlagát, a hibázászlók a szórást jelölik.

Jelenleg az egyetlen nem NHEJ integráción alapuló fluoreszcens módszer, amely képes genomi környezetben vizsgálni egy gRNS hatékonyságát az egy fluoreszcens kazetta HR-rel való integrációja. A HR-mediált módszert azonban sokkal nehezebb rutinszerűen használni, hiszen minden új genomi lókuszra új homológ karokat kell a bevinni kívánt fluoreszcens kazetta két oldalára klónozni. Ezt csak olyan ritka esetben lehet általánosan használni, ha az összes tesztelendő gRNS-hez ugyanazt a HR donort tudjuk használni, mivel a targetek egymáshoz közel találhatók. Ezzel ellentétben, a bemutatott önvágó plazmid módszer esetén nem szükséges a donor plazmid átalakítása, minden megcélózni kívánt genomi szekvencia esetében általánosan használható.

A következő kísérletben egy egér (NIH/3T3) és két humán (HeLa és HEK293) sejtvonalon teszteltük a módszer működését (**14. ábra**). Hét, korábban már N2a sejtekben bemutatott, gRNS aktivitását teszteltük NIH/3T3 sejtekben (**14/A ábra**). A gRNS-ek hasonló aktivitást mutattak, mint korábban N2a sejtekben, itt is a pten spacer bizonyult a legaktívabbnak 15 % célzott integrációval míg a piwi2 spacerrel nem detektáltunk beépülést.



14. ábra. Különböző spacerek hatékonysága önvágó plazmiddal vizsgálva NIH/3T3, HeLa és HEK293 sejtekben

Az ábrákon GFP pozitív NIH/3T3 egér fibroblaszt (A), HeLa (B) és HEK293 (C) sejtek százalékos aránya látható, amikor az SpCas9-cel célzott genomi targetekre az önvágó plazmidot integráltuk. A NIH/3T3 esetén négy gén hét különböző célszekvenciáját, a HeLa esetén egy gén két különböző célszekvenciáját, valamint a HEK293 esetén a genomban három gén öt különböző célszekvenciáját vizsgáltuk. Az összes kísérletben az önvágó plazmidok véletlenszerű integrációs háttérnek mérésére egy aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélzva (random integráció). Ennek a háttérnek a mértékét szaggatott vonallal jelöltük. Az ezektől szignifikánsan eltérő értékeket vettük célzottan integrált eseményeknek. További negatív kontrollként egy inaktív SpCas9-et alkalmaztunk az önvágó plazmiddal, amely a cirkuláris plazmid random beépülésének mértékét mutatja. Az oszlopok $n=3$ független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszázlók a szórást jelölik. A csillagok a véletlenszerű integrációs értéktől való eltérést jelzik. ***: $p<0,001$.

Két humán sejtvonalon is megvizsgáltuk az önvágó plazmid működését (14/B, C ábra). Az egér genomot célzó gRNS-ek többségét itt nem tudtuk használni, viszont a pten gRNS nem csak az egér, hanem a humán genomban is a *PTEN* génre specifikus. Ezzel és másik két humán *PRNP* gént célzó gRNS-sel teszteltük a rendszer működését HeLa sejtekben

(**14/B ábra**). Az intracellulárisan linearizált plazmid háttér 1,4 % maradt, míg a különböző gRNS-ek 5-15%-ig terjedő célzott beépülést okoztak. HEK293 sejtekben a *PTEN* valamint négy – már publikált¹⁰⁷ - gRNS-sel teszteltük a rendszert, melyek az *LMNA* és *PML* géneket célozzák. Itt is minden gRNS hatékony beépülést okozott (**14/C ábra**).

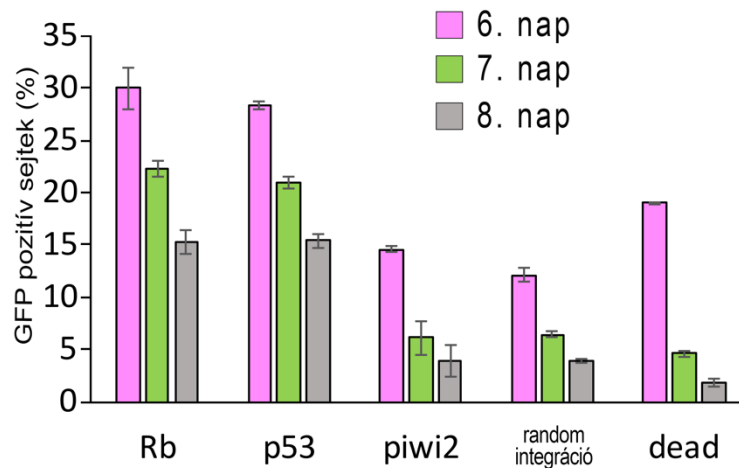
Összességében ezek az eredmények a különböző sejtvonalakon (N2a, NIH/3T3, HeLa és HEK293) azt sugallják, hogy a módszer valószínűleg sejtvonalaktól függetlenül, általánosan használható.

5.1.4. Észlelési idő rövidítése egy degradációs domén fúziójával

Habár a fenti kísérletekből kiderül, hogy ez egy jóval egyszerűbben használható fluoreszcens rendszer, mint a korábbi megoldások, a detekcióhoz szükséges idő korlátozza a módszer használatát, hiszen meg kell várni a tranziens expresszió lecsengését, amely 12-15 nap is lehet. A tranziens jel hosszú lecsengése valószínűleg a GFP molekulák nagy stabilitásából ered, akár napokkal azután is jelen lehetnek, hogy az őket expresszáló plazmid DNS már lebomlott a sejtben. E hatás csökkentése érdekében, ideális esetben a GFP-t a plazmid DNS lebomlása után lenne jó expresszálni, így biztosítva, hogy a GFP jel csak az integrált DNS-ről származhasson.

A GFP fehérje gyorsabb lebomlásának érdekében egy reverzibilis degradációs domént (DD) fuzionáltunk a fehérjéhez, amely az *E. coli* dihidrofolát-reduktáz fehérje destabilizált mutánsa. A degradációs domén stabilizálódik trimetoprim (TMP) jelenlétében, és megakadályozza a fúziós fehérje lebomlását¹⁰⁸. Így, amennyiben a TMP-t a plazmid DNS lebomlása után adjuk a sejtekhez, lehetővé teszi, hogy csak a genomban integrált DNS-ről expresszált fehérjékből származzon a GFP jel. Így a kísérlet időtartama jelentősen lerövidíthető.

A DD-t a GFP fehérjéhez fuzionáltam és egy önvágó plazmidba klónoztam (önvágó-DD plazmid). Három korábban használt gRNS-sel (kettő hatékonyan és egy rosszul működő) 2-10 napig, naponta mértem a GFP jelet, hogy meghatározzam, hogy melyik napon lehet ideálisan mérni azok aktivitását a önvágó-DD plazmiddal. A kísérletek azt mutatták, hogy 6-8 napra le lehet csökkenteni az eredetileg 12-15 napos mérést (**15. ábra**). A gRNS-ek aktivitása hasonló tendenciát mutatott, mint a DD nélküli önvágó plazmidok esetében.

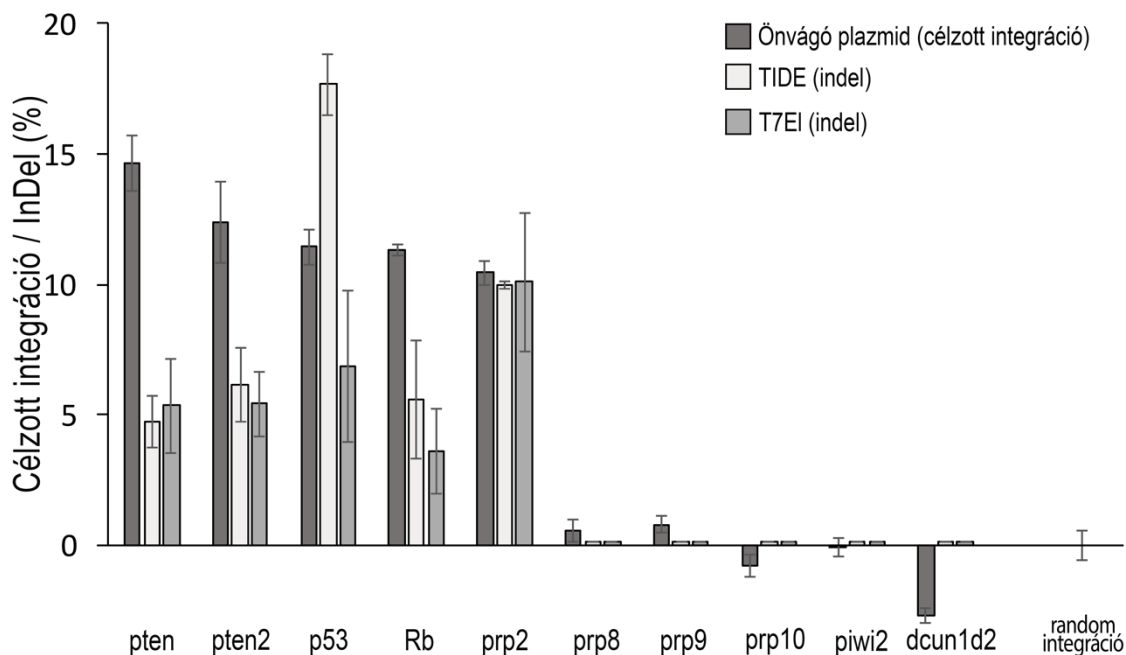


15. ábra. Az észlelési idő lerövidítése egy degradációs domén GFP-hez történő fúziójával

Az ábrán GFP pozitív N2a sejtek százalékos aránya látható a transzfekciót követő 6-8. napokon, amikor az SpCas9-cel célzott három genomi targetre az önvágó-DD plazmidot integráltuk. Az önvágó-DD plazmid egy degradációs doménnel ellátott GFP expressziós kazettát tartalmaz. A degradációs domén stabilizálása érdekében a TMP kismolekulát 24 órával az áramlási citometriás mérés előtt adtuk a sejtekhez. Kontrollként az önvágó plazmid véletlenszerű integrációs hátterének mérésére egy aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélzva (random integráció). További negatív kontrollként egy inaktív SpCas9-et alkalmaztunk az önvágó plazmiddal, amely a cirkuláris plazmid random beépülésének mértékét mutatja (dead). Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszalók a szórást jelölik.

5.1.5. A módszer hasonlítása két PCR alapú módszerhez

A módszerünket két korábban már publikált PCR alapú módszerrel szeretnénk volna összehasonlítani: a T7 endonukleáz I esszével, és a TIDE-al⁸⁰. Tíz gRNS-t transzfektáltam az önvágó-DD plazmiddal. Az áramlási citometriával lemért sejtekből genomi DNS-t izoláltam és PCR-rel amplifikáltam a célzott genomi lókuszt. A keletkező indelek mértékét T7 endonukleáz I esszével, valamint TIDE-al mértem (**16. ábra**). Mindhárom módszer ugyanazt az öt gRNS-t találta jobban működőnek a másik öthöz képest, amelyeknél csak alacsony vagy semmilyen aktivitást nem lehetett detektálni. Természetesen kisebb különbségek adódtak a gRNS-ek relatív hatékonyságában, hiszen ezek a módszerek nem tökéletesen ugyanazoknak a hibajavító mechanizmusoknak a végkifejletjét mérik.



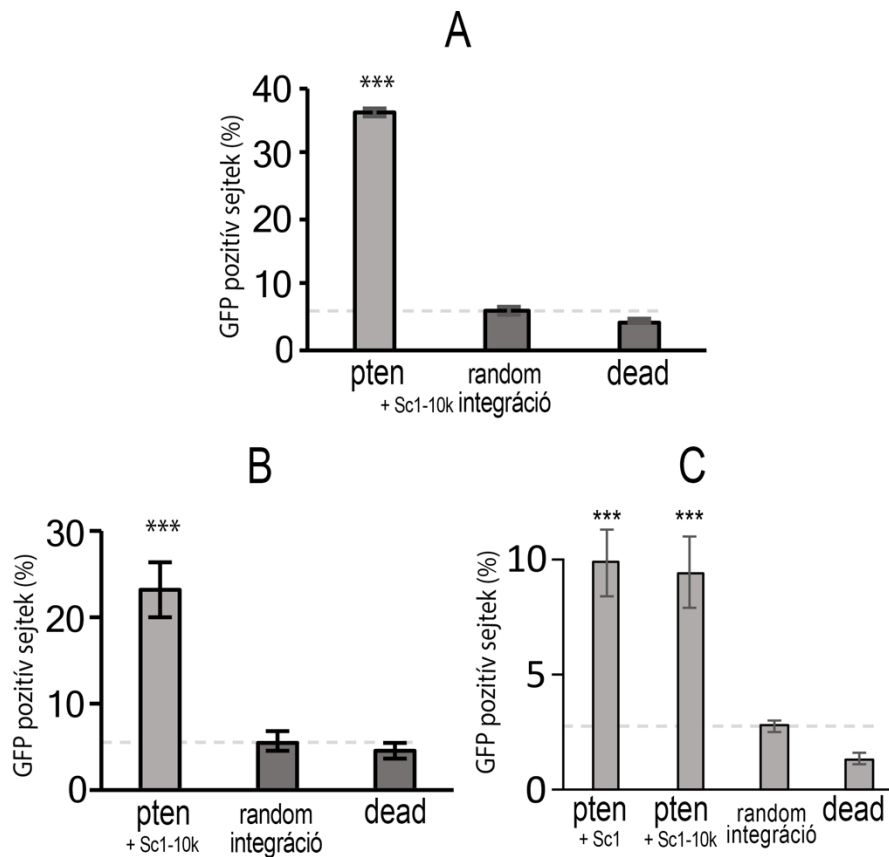
16. ábra. A módszer hasonlítása TIDE és T7 endonukleáz I esszéhez

Tíz gRNS aktivitását mértük a degradációs doménnel ellátott önvágó plazmid integrációjával (sötétszürke), valamint a keletkezett indelek mértékét TIDE (világos szürke) illetve T7EI esszével (szürke) vizsgáltuk. A gRNS-ek a *Pten*, *Tp53*, *Rbl2*, *Prnp*, *piwil4*, és *Dcun1d2* géneket célozzák. Az ábrán GFP pozitív N2a sejtek százalékos aránya látható, amikor az SpCas9-cel célzott tíz genomi targetre az önvágó-DD plazmidot integráltuk. Az áramlási citométerrel lement sejtekből genomi DNS-t izoláltunk, majd a megcélzott genomi szekvenciákat PCR-rel amplifikáltuk. A keletkezett indeleket egyrészt Sanger szekvenálást követően TIDE-al értékeltük ki, másrészt T7EI esszével kvantifikáltuk. Kontrollként az önvágó plazmid véletlenszerű integrációs háttérének mérésére egy aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélózva (random integráció). Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek, illetve a mért indelek átlagát, a hibaszalók a szórást jelölik.

5.1.6. NHEJ alapú génbevitel önvágó plazmid segítségével

Korábban az NHEJ hibajavítást felhasználó génbevitellel (NHEJ klónozás) korlátozott hatékonysággal ugyan, de képesek voltak egy 5 kilobázisos célszekvenciát a genomban integrálni¹⁰⁹. Kíváncsiak voltunk, hogy egy 10 kb-os plazmid bevitele lehetséges-e az NHEJ segítségével és ha igen, milyen hatékonysággal. Egy 5 kb-os kazettát az önvágó plazmidba molekuláris klónozással bejuttatva létrehoztam egy kb. 10 kilobázisos (10805 bp) önvágó plazmidot. A leghatékonyabb spacerrel (pten) 30 és 18 %-os célzott integrációt értünk el N2a (**17/A ábra**) illetve HeLa (**17/B ábra**) sejtekben. Ezek az

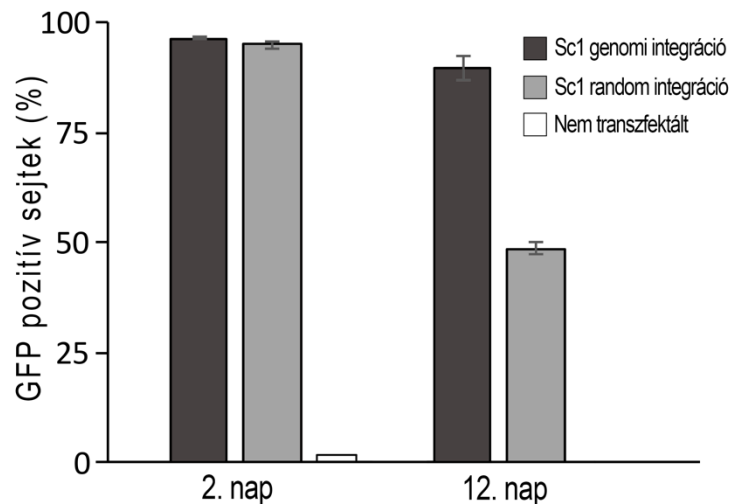
értékek egy nagyságrenddel magasabbak, mint amit korábbi tanulmányokban leírtak HEK293 sejtekben¹⁰⁹. Ezért az eredményeket összevettük a **14. ábrán** bemutatott eredményekkel, valamint egy 10 kb-os önvágó plazmiddal megismételtük a kísérletet ezen a sejtvonalon egy nagyságrenddel jobb eredményt értünk el (**17/C ábra**), mint mások korábban (7 % vagy annál nagyobb integráció vs. 0,17 % a 109. referenciában).



17. ábra. Egy 10 kilobázis hosszú önvágó-plazmid hatékonyan integrálható különböző sejtek genomjába

Az ábrákon GFP pozitív N2a (**A**), HeLa (**B**) és HEK293 (**C**) sejtek százalékos aránya látható, amikor az SpCas9-cel valamint a pten gRNS-sel egy 10 kilobázis hosszú önvágó plazmidot integráltuk. Az összes kísérletben az önvágó plazmidok véletlenszerű integrációs háttérnek mérésére egy aktív SpCas9-cel, valamint az önvágó plazmidhoz tartozó TL gRNS-sel az önvágó plazmidot linearizáltuk intracellulárisan, de a genom nem volt megcélózva (random integráció). Ennek a háttérnek a mértékét szaggatott vonallal jelöltük. Az ezektől szignifikánsan eltérő értékeket vettük célzottan integrált eseményeknek. További negatív kontrollként egy inaktív SpCas9-et alkalmaztunk az önvágó plazmiddal, amely a cirkuláris plazmid random beépülésének mértékét mutatja. Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszázalék a szórást jelölik. A csillagok a véletlenszerű integrációs értéktől való eltérést jelzik. ***: p<0,001.

Kíváncsiak voltunk arra is, hogy két expressziós kazettát tartalmazó plazmid esetén a két kazetta kapcsolatosan integrálódik-e a genomban. DNS szekvenciák véletlenszerű integrálódásánál az egyik legfőbb probléma szokott lenni, hogy pl. egy antibiotikum rezisztenciára való szelekció nem mindig kapcsolódik össze a kívánt gén integrációjával. Létrehoztam egy olyan önvágó plazmidot, amelyre a GFP kazetta mellé egy puromicin expressziós kazettát is klónoztam. Ezt transzfektáltam N2a sejtekbe SpCas9-cel, valamint a *Pten* gént célzó gRNS-sel (**18. ábra**). Kontrollként, a plazmidot dSpCas9-cel ko-transzfektáltam (a véletlenszerű integrációt mérendő). Az önvágó plazmid esetében a szelektált sejtek 90%-a mutatott GFP fluoreszcenciát, tehát ilyen hatékony volt a kapcsolatos integráció. Ez a véletlenszerű integráció során (a várható) 50% volt. Az utóbbi kísérletben két-három nagyságrenddel kevesebb túlélő klón volt, amely a NHEJ klónozás hatékonyságát is mutatja a plazmid véletlenszerű integrációjával szemben.



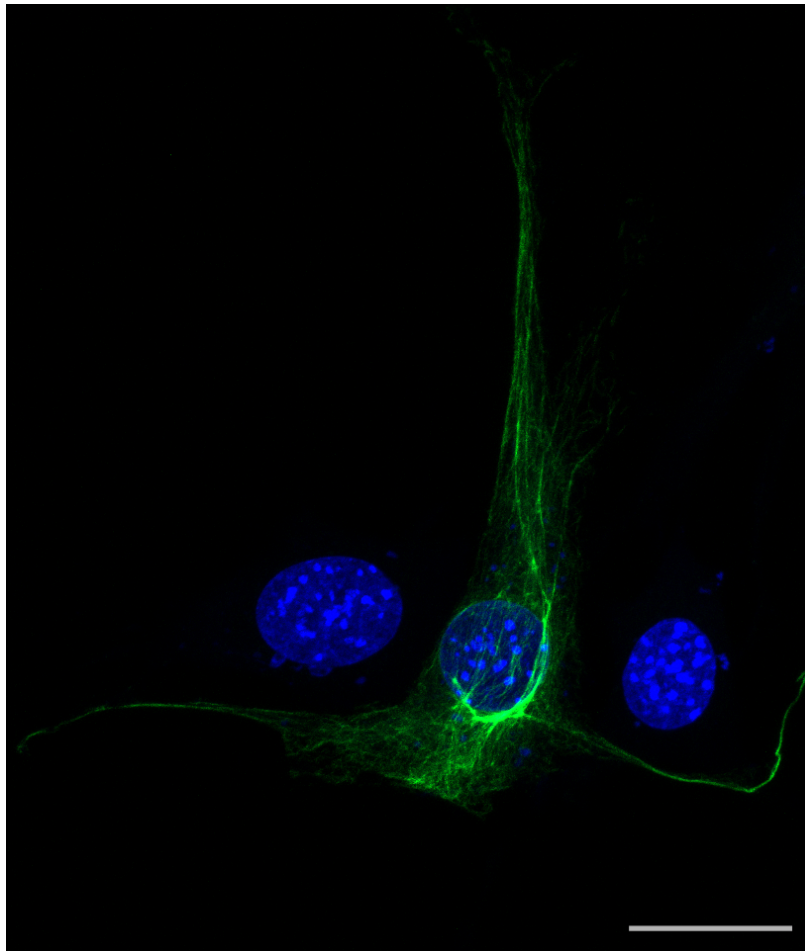
18. ábra. Két expressziós kazetta kapcsolatos integrálása önvágó plazmiddal, N2a sejtekben

Az ábrákon GFP pozitív N2a sejtek százalékos aránya látható a transzfekciót követő 2. és 12. napon, amikor az SpCas9-cel célzott *pten* targetre egy puromicin expressziós kazettával rendelkező önvágó plazmidot integráltuk (fekete oszlopok), puromicin szelekció mellett. Kontrollként a két expressziós kazetta véletlenszerű integrációját követendő, a plazmidot egy nukleáz inaktív SpCas9-cel (dCas9) ko-transzfektáltuk (szürke oszlopok), ilyenkor a plazmidot célzottan nem nyitjuk fel, a DNS törés pontja véletlenszerű. Nem transzfektált sejteket (fehér oszlop) használtunk a puromicin szelekció követésére, melyek a 12. napra már teljesen eliminálódtak. Az oszlopok n=3 független transzfekcióban mért GFP pozitív sejtek átlagát, a hibaszázalék a szórást jelölik.

Későbbi munkáinkban a módszert többféleképpen is felhasználtuk. Egy szakmai bírálóat alatt álló publikációnkban¹¹⁰ egy nem fluoreszcens kazetta bevitele volt a célunk minél hatékonyabban HEK293 sejtekbe. A nem fluoreszcens kazettát az önvágó plazmidba (a GFP szekvencia helyére) klónoztuk, és ezt Cas9 segítségével bevittük HEK293 sejtekbe. A sikeres génbevittelt megnövelendő, a puomicin expressziós kazettával is rendelkező önvágó plazmidot használtuk antibiotikum szelekció mellett, így dúsítani tudtuk a génszerkesztett sejteket. A génbevétel rendkívül hatékony volt, a klónok többsége hordozta a bevinni kívánt szekvenciát, és rendkívül gyors sejtvonalalapítást tett lehetővé.

Egy nemrég megjelent publikációnkban⁷⁴ módosított SpCas9 variánsokkal történő génbevittelt teszteltünk, egy olyan önvágó plazmid segítségével, amelynél a GFP szekvencia előtt nem található promoter, valamint a plazmidon található SpCas9 célszekvencia közvetlenül a GFP előtt található. Az így módosított önvágó plazmidot egy endogén fehérje START kodonja mögé építettük be (74. referencia 7/a, b ábra). A beépülésnek rendkívül pontosnak kell lennie, hiszen más leolvasási keretben az endogén promoter nem hajtja meg a fehérjét. Így teszteltük a génbevétel hatékonyságát különböző SpCas9 variánsok esetében.

Egy másik rendkívül hasznos felhasználása az NHEJ klónozásnak, ha az adott sejtben alacsonyban működik, vagy teljesen ki van ütve homológ rekombinációhoz szükséges gének egyike, így csak az NHEJ klónozás lehetséges. A Hámori és mtsai. által alapított és jellemzett CST sejtvonal¹¹¹ egy olyan egér emlőtumor, amelyben a *Brca1* gén nem funkcionális, így a homológ rekombináció is deficiens a sejtben. Ebbe a sejtvonalba rendkívül nehéz (vagy lehetetlen) a klasszikus, homológ rekombinációval történő, bázispár pontosságú génbevétel. Célunk a vimentin fehérje GFP-vel való jelölése volt. Ehhez egy másik önvágó plazmidot hoztunk létre, melyben a vimentin gén utolsó exonjával fuzionált GFP szekvenciája, valamint az ez előtt lévő intron található. A plazmidot és a genomot is az intronban lévő célszekvenciával vágtuk meg, ahova a plazmid be is épült. A módszer előnye, hogy az NHEJ hibajavítást követő esetleges inszerciók és deléciók az intronban jól toleráltak és az exon megfelelő splicingja így is megtörténik. A módszerrel sikeresen hoztunk létre endogén vimentin-GFP fúziós fehérjét kódoló CST sejtvonalat (**19. ábra**).



19. ábra GFP-vel jelölt vimentin fehérjét kifejező egér emlőtumor sejtvonala

NHEJ-mediált integráció segítségével jelöltük a vimentin fehérjét CST egér emlőtumor sejtekben. A vimentin jellegzetes hálózatos struktúrája megfigyelhető. A vimentin-GFP zölddel, a sejtmagok kék színnel látszódnak, a lépték 20 μm -t jelöl.

5.2. SpCas9 variánsok hatékonyságának predikciója egy 1 millió szekvenciát tartalmazó gRNS könyvtár segítségével

Az SpCas9 nukleázzal való hatékony genomszerkesztési módszer kifejlesztésével egyidős az a megfigyelés is, hogy a nukleáz a különböző célszekvenciákat különböző hatékonysággal képes hasítani. Majdnem minden génszerkesztési feladat során rendkívül hasznos lenne, ha tudnánk, hogy a potenciális célszekvenciák közül (pl. egy gén kiütésénél), melyek a legígéretesebbek. Ez különösen fontos olyan a genom egészére kiterjedő kiütési vizsgálatoknál, mint például a Genome-Scale CRISPR Knock-Out (GeCKO)⁵⁴ könyvtár. Ilyen könyvtárak használatakor rendkívül fontos, hogy előre tudjuk, hogy az összes gRNS hatékony, hiszen nem elkülöníthetők a rosszul működő gRNS-ek a negatív eredményektől és ez rontja a kísérlet hatékonyságát. A WT-SpCas9 esetében már számos olyan predikációs eljárást fejlesztettek ki, amelyek - különböző hatékonysággal, de - képesek előre jelezni, hogy mely gRNS-ekkel fog jobban vagy rosszabbul hasítani a nukleáz^{88-95,112}.

Az utóbbi években nagyszámú Cas9 variánst fejlesztettek ki, mint például a megnövelt pontosságú^{61,72,75,77} vagy alternatív PAM-ot felismerő^{61-63,65-67} SpCas9 nukleázok, viszont ezek szekvencia-specificitása látszólag eltérő és csak nagyon korlátozottan ismert. Ezek rutinszerű használatához ugyanúgy szükség lenne egy hatékonyság-előrejelző algoritmusra, mint a WT-SpCas9 esetében. Egy ilyen, pontosan működő eszköz kifejlesztése viszont általában egy nagy kihívást jelentő, drága és nem utolsó sorban hosszadalmas folyamat, mert megismételhető módon kell nagyon sok szekvencia esetén megmérni, hogy milyen hatékonysággal működnek, majd ezek alapján kifejleszteni a kívánt predikciót.

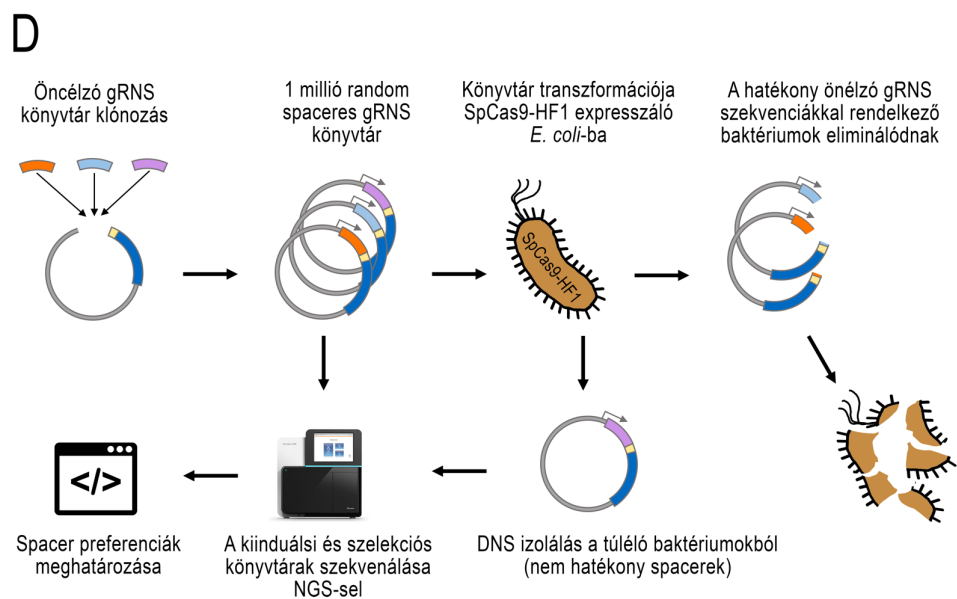
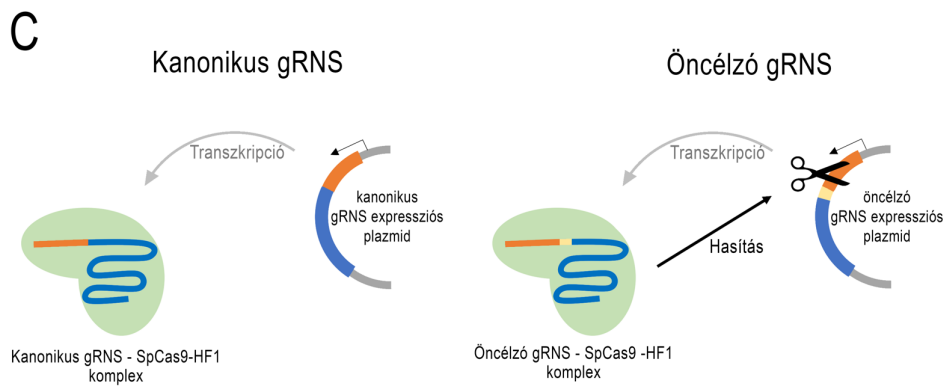
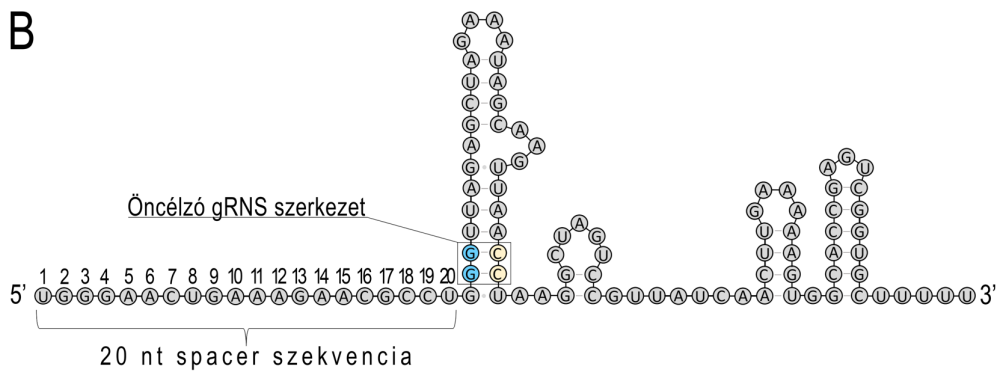
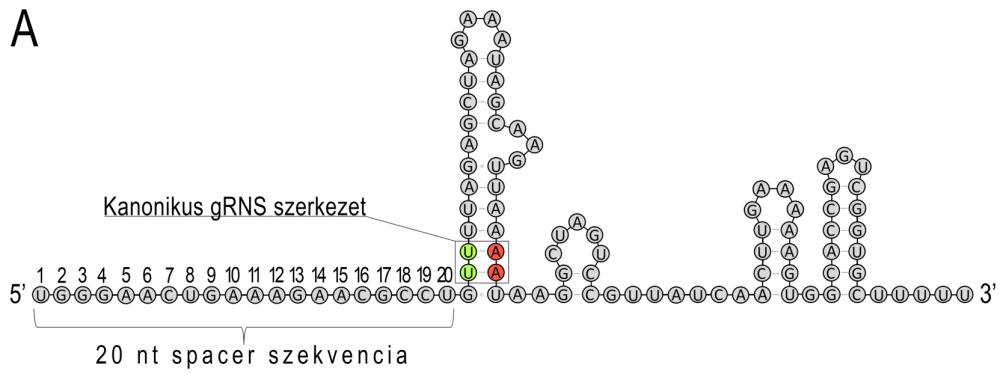
5.2.1. Egymillió különböző szekvenciát tartalmazó gRNS könyvtár és a szelekciós rendszer kifejlesztése

Az általunk kifejlesztett módszerben megoldást találtunk a fenti problémára. Itt azt a jelenséget használjuk ki, hogy az SpCas9 érzéketlen a gRNS-ben a spacer szekvencia mellett található RNS másodlagos szerkezeti hurok konkrét szekvenciájára (**20/A ábra**), annak csak egy intakt, megfelelő másodlagos struktúrát kell felvennie. Korábbi kutatásokban már bizonyították, hogy ezek a bázisok viszonylag szabadon lecserélhetők,

és az SpCas9 hatékonysága nem változik számottevően, ha a struktúra intakt marad^{37,113-116}. A hurokban a spacer közvetlenül követő kanonikus három bázist (**20/A ábra**) egy NGG motívumra cseréltük le (**20/B ábra**), amely megegyezik az SpCas9 PAM szekvenciájával, ezáltal öncélzó gRNS-eket (self-targeting gRNS, stgRNS) hoztunk létre, melyek saját kódoló szekvenciájukat célozzák¹¹⁴⁻¹¹⁶ (**20/C ábra**).

Ha egy plazmidról egy ilyen módosított gRNS-t expresszálunk egy bakteriális vagy emlős rendszerben, az stgRNS az SpCas9 fehérjével komplexet képezve a saját expressziós plazmidját célozza meg és vágja el (**20/C ábra**). Ez a megközelítés hatékonyan oldja meg azt a problémát, hogy a megfelelő gRNS és célszekvencia egy sejtben legyen megtalálható. Valamint a módszer lehetővé teszi teljesen véletlenszerű gRNS könyvtárak használatát, hiszen a véletlenszerűen előállított spacer szekvenciája minden esetben megegyezik a target szekvenciájával. A célunk egy egymillió spacer tartalmazó stgRNS könyvtár létrehozása volt, valamint egy eddig kevésbé jellemzett, megnövelt pontosságú SpCas9, az SpCas9-HF1 szekvencia-specifititásának jellemzése, és egy jól használható predikciós eljárás kifejlesztése.

Három, randomizált spacerrel rendelkező öncélzó gRNS könyvtárat hoztam létre, melyek kb. 500, 700 illetve 100 ezer szekvenciát tartalmaztak. Kombinálva az első kettőt, később a predikciós eljárás tanító adathalmazaként alkalmaztam (~1 millió szekvenciát tartalmazó, 1M könyvtár) a harmadik pedig a predikciót tesztelő adathalmazként funkcionált (100K könyvtár). A könyvtárak újgenerációs szekvenálása az elméleti 25% helyett kissé csökkent timin (T) tartalmat mutatott ki, (lásd 117. referencia, **1. kiegészítő ábra**), amely hasznos lehet, amennyiben ezeket a szekvenciákat emlős rendszerben vizsgáljuk. Ugyanis, emlős expressziós rendszerben a gRNS-eket az RNS polimeráz III-mal expresszáljuk, melynél transzkripció terminációs szignálként működhetnek a hosszabb T-gazdag szekvenciák. Az NGS kiértékelésével az 1M könyvtár 1 223 805 a 100K könyvtár 136,557 szekvenciát tartalmazott (lásd 117. referencia, **1. kiegészítő táblázat**).



20. ábra. Az öncélzó gRNS könyvtár teszt (SLS) működése

A kanonikus (A) és az öncélzó (B) gRNS-ek szekvenciája és módosításai. A kanonikus gRNS 22-23. pozícióiban található TT dinukleotidokat (zöld) GG dinukleotidokra (kék) cseréltük, így egy PAM szekvenciát kialakítva létrehoztuk az öncélzó gRNS-eket (stgRNS). A másodlagos szerkezet megtartása érdekében az ezekkel párban álló nukleotidokat AA-ról CC-re módosítottuk. (C) Az stgRNS-ek, szemben a kanonikus gRNS-ekkel, a saját kódoló DNS-szekvenciájukat célozzák meg. Az spacer szekvencia narancssárgával, a PAM sárgával, a gRNS további részei kézzel, a plazmid gerinc pedig szürkével jelölt. (D) Folyamatára az SLS módszer működéséről, amelyet az SpCas9-HF1 spacer preferenciáinak azonosítására fejlesztettünk ki. Az egymillió spacer szekvenciát tartalmazó stgRNS könyvtárból, a hatékonyan működő célszekvenciákat hordozó plazmidok antibiotikum-szekció hatására elbomlanak. Mind a kiindulási, mind a nukleázzal hasított könyvtárakat megszekvenáljuk új generációs szekvenálással, majd összehasonlítjuk és így a hatékony (hasított) és a kevésbé hatékony szekvenciákat azonosítani tudjuk.

Létrehoztam egy bakteriális negatív szelekciós rendszert, amelyben SpCas9-HF1, valamint kontrollként WT-SpCas9 enzimet expresszáló kompetens *E. coli* sejtekbe stgRNS-eket transzformáltam (20/D ábra). Amennyiben az - antibiotikum rezisztenciát is hordozó - plazmidok hasítása hatékonyabb, mint azok hibajavítása és replikációja, az adott baktérium antibiotikum szelekció mellett nem képes osztódni. Így, ha a létrehozott könyvtárat transzformáljuk ezekbe a sejtekbe, a hatékony stgRNS-ek az SpCas9 segítségével elvágják önmaguk kódoló plazmidját, így ezek a szekvenciák alulreprezentálva lesznek, ha megszekvenáljuk új generációs szekvenálással a kísérlet után a baktériumokból izolált plazmidokat. A nem hatékony szekvenciák ebből következően feldúsulnak a populációban. A hasítás előtti könyvtár szekvenálását összehasonlítva a hasítás utánival, megállapíthatók azok a célszekvenciák, amelyekkel hatékonyan volt képes hasítani az SpCas9 és eltűntek a könyvtárból (vagy számuk jelentősen lecsökkent). Ezt a módszert öncélzó gRNS könyvtár tesztnek (Self-targeting Library Sreen – SLS) neveztük el.

5.2.2. Az SpCas9-HF1 aktivitásának meghatározása egymillió célszekvencián

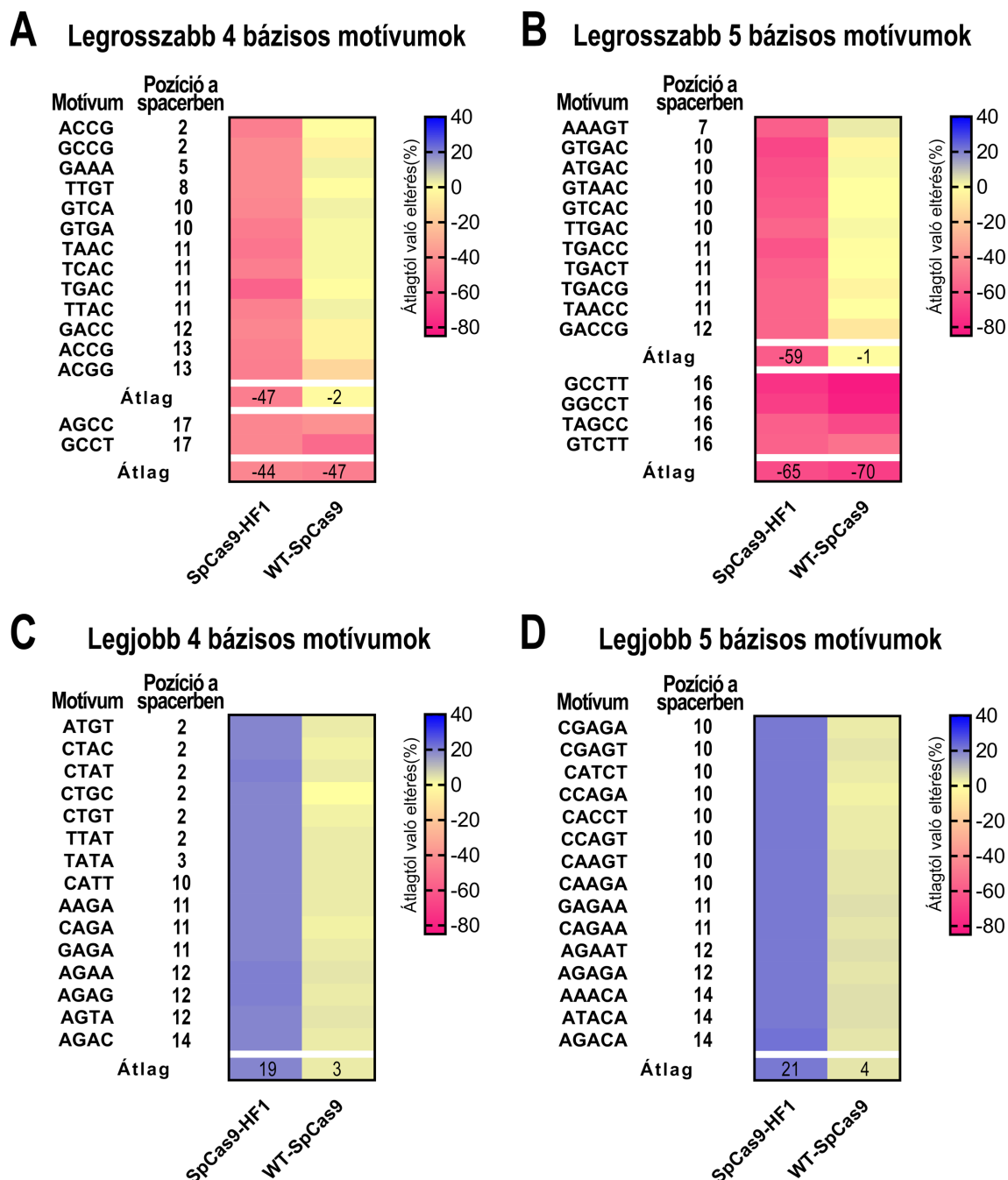
Három párhuzamos kísérletben az 1M könyvtárat, és - egy független kísérletben - a 100K könyvtárat is transzformáltam SpCas9-HF1-et, valamint WT-SpCas9-et expresszáló kompetens sejtekbe is. Egyes klónok esetleges túlszaporodásának elkerülése érdekében folyékony tápfolyadék helyett 3 dimenziós félszilárd agaróz médiumban növesztettem a

transzformált sejteket. A plazmidokat izoláltuk a túlélő klónokból és NGS segítségével megszekvenáltuk.

Az NGS eredményeket kiértékelve megállapítottuk, hogy a párhuzamos kísérletek kiválóan reprodukálhatók voltak (a kísérletek közti Pearson korrelációs együttható $r = 0,92-0,97$ (lásd 117. referencia, **2. kiegészítő táblázat**). Az NGS leolvasások normalizálása után egy hasítási értéket állapítottunk meg minden szekvenciára (a három párhuzamos kísérletet átlagolva): 0 az egyáltalán nem vágott, 1 pedig a teljesen vágott szekvenciákat jelölte. Az 1M és 100K könyvtárakat vágó és nem vágó szekvenciákra osztottuk fel, melyeknél a 0,3 érték alattiakat tekintettük rosszul működőnek. Mind az SpCas9-HF1, mind a WT-SpCas9 a szekvenciák többségét hasította, így egy kiegyensúlyozatlan adathalmazokat kaptunk: előbbi esetén 6,6%, utóbbi esetén 1,8% nem vágó szekvenciával.

Az SpCas9-HF1 szekvencia-preferenciáinak vizsgálata során azt tapasztaltuk, hogy a hasítási hatékonyságot drámaian rontja, amennyiben a 10-14 pozícióban a GTNAC motívummal átfedő az adott szekvencia (**21/A, B ábra**). Ennek a motívumnak a jelenléte érdekes módon, a WT-SpCas9 aktivitását nem befolyásolta. Ezt a jelenséget egyik korábbi tanulmány sem mutatta be a vad típusú SpCas9 esetében. Korábbi tanulmányok egy PAM proximálisan elhelyezkedő GCC tripletről (18-20. pozíció) bizonyították, hogy erősen rontja a gRNS-ek vágási hatékonyságát a WT-SpCas9 esetében¹¹⁸. Mi azt találtuk, hogy nem csak a WT-SpCas9 hanem az SpCas9-HF1 hatékonyságát is negatívan befolyásolja (**21/A, B ábra**). Érdekes módon ez a GCC triplet nem csak a 18-20 hanem 5' irányban elhelyezkedve a 17-19, valamint a 16-18-as pozíciókban is blokkoló hatású mindkét fehéjére nézve. Ezt a tulajdonságot korábbi tanulmányok nem írták le a WT-SpCas9 esetében. A PAM mellett 5' irányban elhelyezkedő TT dinukleotidról korábban leírták a WT-SpCas9 esetében, hogy szintén rontja annak működését¹¹⁸, itt ezt a tulajdonságot nem detektáltuk. Ez azzal magyarázható, hogy a tanulmány emlős sejtekben készült, ahol (a gRNS-ben a spacer 3' végét követő T-gazdag szekvencia miatt) a PAM közeli T bázisok ronthatják a transzkripciót¹¹⁹. Ezt a bakteriális RNS polimerázok vagy jobban tűrik, vagy az stgRNS eltérő szekvenciája (PAM közeli GG) mentheti ezt a tulajdonságot. Egy korábbi publikációval ellentétben mi nem láttuk az eredményeinkben azt, hogy a 18. pozícióban preferált lenne a citozin a WT-SpCas9 esetén⁹⁰. A legjobbnak bizonyuló 15 darab 4 és 5 bázisos motívum (**21/C,D ábra**) vizsgálatakor csak az SpCas9-

HF1-re jellemző motívumokat találtunk, amelyek neutrálisan hatottak a WT-SpCas9-re. Ezek a motívumok az SpCas9-HF1-re sem hatottak annyira hangsúlyosan, mint a legrosszabb motívumok.

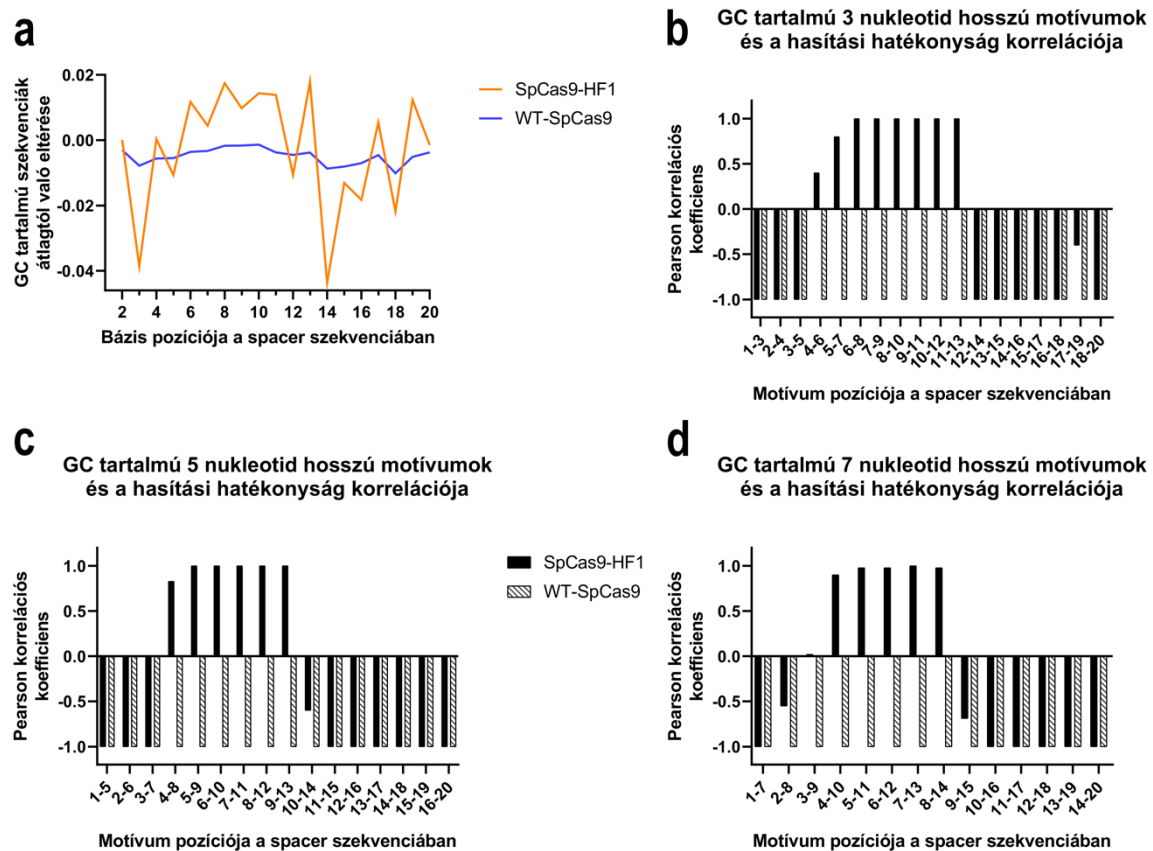


21. ábra. Szekvencia motívumok, amelyek a legkedvezőbb vagy a legkevésbé kedvező hatással vannak az SpCas9-HF1 hasítási hatékonyságára a bakteriális 1M célkönyvtárban

A hőképek az SpCas9-HF1 (az összes ábrán a bal oldali panel) és a WT-SpCas9 (jobb oldali panelek) átlagos aktivitásától való százalékos eltérését mutatják a jelzett motívumokat tartalmazó összes szekvenciában, a bakteriális 1M könyvtár adatai alapján. Az ábrákon 15-15 olyan négy (A, C) és öt (B, D) nukleotid hosszú motívum látható, amelyek az SpCas9-HF1 hasítását a legkevésbé

kedvezően (**A, B**) vagy a legkedvezőbben (**C, D**) befolyásolják, melyeket az SLS szelekcióban azonosítottunk. Ugyanezen motívumok hatását a WT-SpCas9 esetben is ábrázoltuk. A legrosszabb motívumok között egy PAM-proximális GCC motívumot azonosítottunk a 16-20. pozíciók között, ami hasonlóan befolyásolja a WT-SpCas9 és az SpCas9-HF1 működését, valamint a 10-14. pozíciók között található GTNAC motívum csak az SpCas9-HF1 aktivitását befolyásolta negatívan. A legjobb motívumok csak az SpCas9-HF1 aktivitását befolyásolta, viszont nem olyan mértékben, mint a legrosszabbak.

Megvizsgáltuk azt is, hogy milyen korreláció fedezhető fel a spacerok különböző szegmenseinek GC-tartalma és azonos GC-tartalmú szegmenseket tartalmazó spacerok átlagos hasítási hatékonysága között (**22. ábra**). A 6. és 13. pozíciók közötti szegmensben a magasabb GC-tartalommal rendelkező szekvenciák átlagos hasítása nagyobbak bizonyult az SpCas9-HF1 esetében, megjegyzendő, hogy a tulajdonságot korábbi tanulmányok nem azonosították. Ezt a WT-SpCas9 esetében eddigi tanulmányok nem találták igazoltnak, valamint az itt bemutatott munka sem: a magas GC-tartalommal rendelkező szegmensek pozíciótól és hosszától függetlenül a WT-SpCas9-re negatív hatásúak (**22. ábra**).



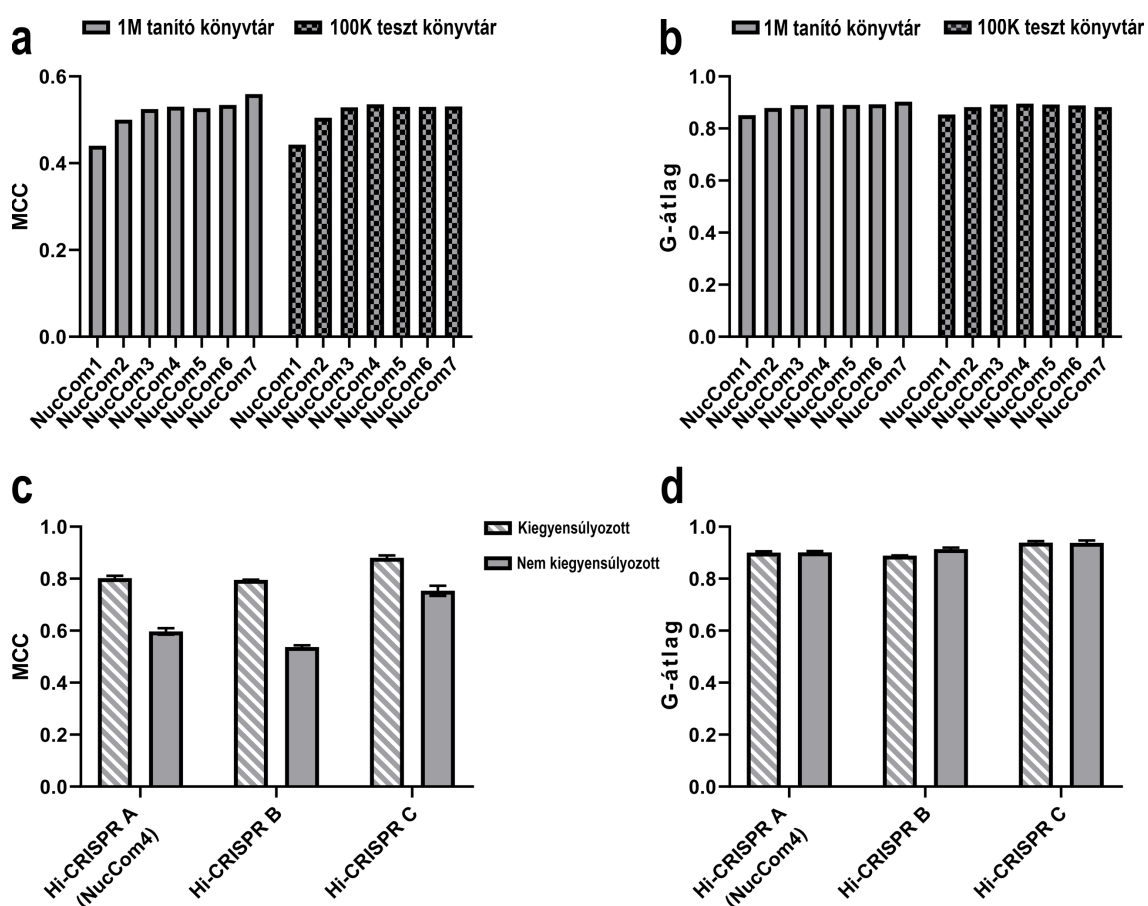
22. ábra. A spacer 6-13. pozícióiban lévő magasabb GC-tartalom korrelál az SpCas9-HF1 magasabb hasítási hatékonyságával

(A) Az adott pozícióban lévő G vagy C bázissal rendelkező spacerek átlagos vágási hatékonyságának eltérése az 1M könyvtár összes szekvenciájának átlagos vágási hatékonyságától, amikor azokat az SpCas9-HF1 (narancsságra) vagy WT-SpCas9 (kék) hasította. A WT-SpCas9 esetében a GC gazdag szekvenciák minden esetben kisebb hasítás hatékonyságot eredményeznek minden pozícióban. Ezzel szemben a G vagy C nukleotidok kedvezően hatnak az SpCas9-HF1 vágási hatékonyságára a 6-11-es pozíciókban. (B-D) Az ábrákon a spacerek lokális GC-tartalmának az átlagos vágási hatékonyságra gyakorolt hatását ábrázolja. Pearson korrelációval vizsgáltuk hasítási hatékonyság és a pozíció függő helyi GC-tartalom összefüggését a spacer 3 (B), 5 (C) illetve 7 (D) nukleotid hosszú szegmenseiben az olyan szekvenciákban melyek azonos GC-tartalommal rendelkeznek az adott szegmensben, mind az SpCas9-HF1 (fekete oszlopok), mind a WT-SpCas9 (sávozott fehér oszlopok) hasítása után.

5.2.3. Az SpCas9-HF1 aktivitásának előrejelzése bakteriális sejtekben

Első lépésben hét különböző predikciót fejlesztettünk mindkét fehérjére, melyek mindegyike 1-7 nukleotid hosszúságú pozíciófüggő motívumokon alapul, ezeket Nukleotid Összetétel 1-7-nek (Nucleotide Composition - NucCom1 - NucCom7) neveztük el. A tanító, 1M könyvtár adatai alapján, először minden pozíciófüggő motívumra kiszámítottunk egy értéket az 1M könyvtár azon szekvenciáinak vágási hatékonyságából, amelyek az adott motívumot tartalmazták a megfelelő pozícióban. Tehát, szekvenciálisan minden lehetséges 1-7 hosszúságú DNS motívumot összegyűjtöttünk a spacer minden pozíciójában és egyesével megvizsgáltuk, hogy az adott motívum, mely szekvenciákban található meg az 1M könyvtárban és ezek vágási hatékonyságát átlagoltuk. Ezeket az értékeket minden motívumhoz egy NucCom „szótárban” összegyűjtöttük. Ezek után ezt a „szótárt” használva az egyes spacerek vágási hatékonyságát az adott spacerben található motívumok alapján prediktáltuk (részletesebben lásd a 4.5.2 fejezetben, itt a számítás módszerét egy példán is bemutatom). Az 1M könyvtár kísérletes és prediktált értékeit összehasonlítva meghatároztunk a hét NucCom paraméter mindegyikére vonatkozó küszöbértéket. Ennek a küszöbnek az értékét olyan értékben határoztuk meg, ahol a predikció specificitása és szelektivitása egyenlő. Ez a küszöb hatékonyan és gyengén hasító osztályokra osztja a prediktált targeteket.

Az 1M könyvtáron tanított predikciók hatékonyságát a korábban függetlenül létrehozott 100K könyvtár adatain kívántam tesztelni. A bináris osztályozáshoz a Matthews korrelációs együtthatót (MCC)⁹⁸ és a G-átlagot (azaz az érzékenység és specificitás geometriai átlagát)⁹⁹ alkalmaztam. A legmagasabb értékeket a négy nukleotid hosszúságú motívumokon alapuló predikcióval (NucCom4) tudtuk elérni 100K könyvtáron mind az MCC (**23/A ábra**) mind a G-átlag esetében (**23/B ábra**), amely egy igen jelentős 0,86-os G-átlag értéket adott. A WT-SpCas9 esetében is hasonló tendenciát figyeltünk meg 0,85-ös G-átlag értékkel (lásd 117. referencia, 5. kiegészítő ábra).



23. ábra. Az SpCas9-HF1 aktivitásának predikciója a Hi-CRISPR algoritmusok segítségével bakteriális tesztadatokon

(A, B) Az SpCas9-HF1 aktivitásának előrejelzése rövid szekvencia-motívumokon alapuló algoritmusok segítségével az 1M tanító (szürke oszlopok) és a 100K teszt (sakktábla mintás oszlopok) könyvtárakon. A NucCom 1-7 pozíciófüggő rövid (egy-hét bázis hosszú) szekvencia motívumokon alapú előrejelzők, melyeket az 1M tanító könyvtár SpCas9-HF1 hasítási adatainak alapján fejlesztettünk ki. Az osztályozás minőségét (A) a Matthews korrelációs együtthatóval

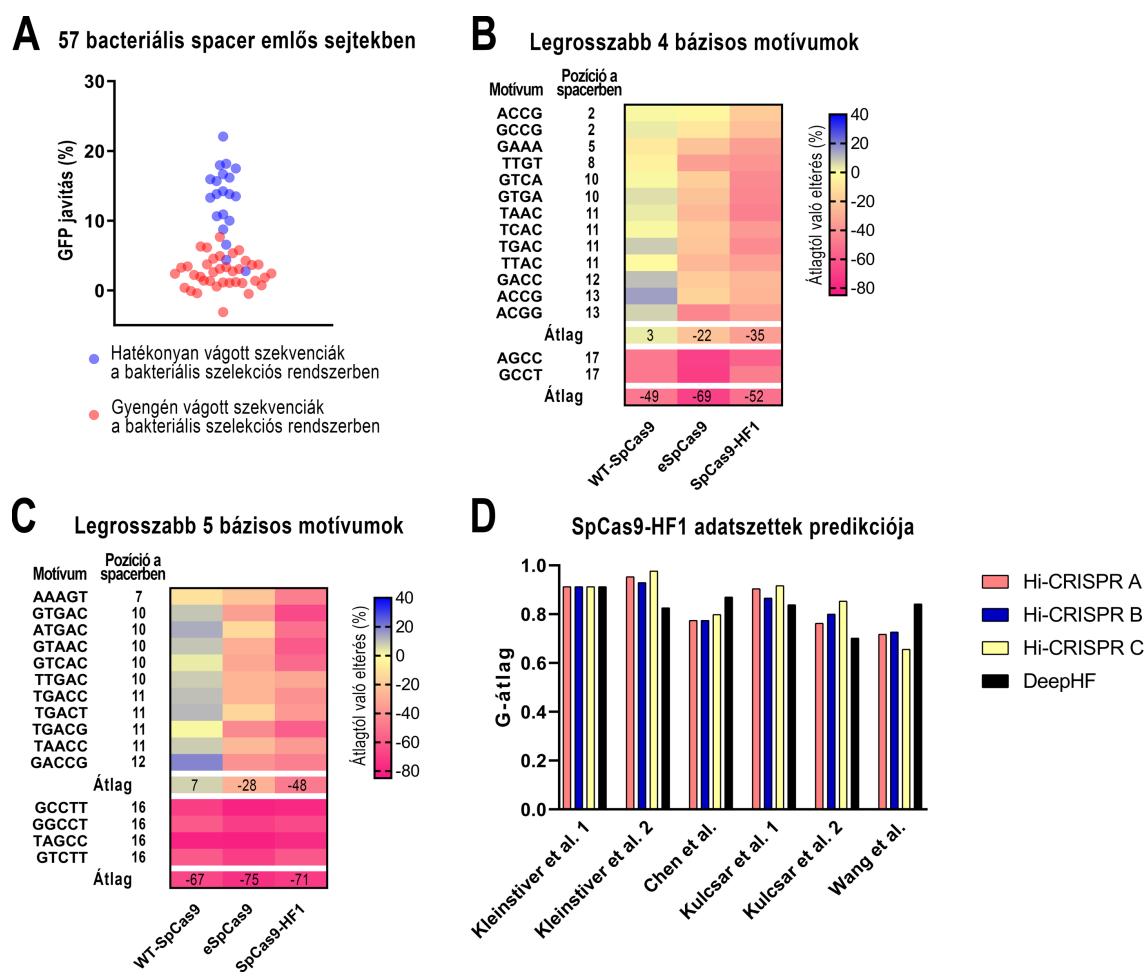
(MCC) és **(B)** a G-átlag segítségével értékeltük. **(C, D)** Az itt kifejlesztett Hi-CRISPR A (NucCom4), valamint két neurális hálózaton alapuló SpCas9-HF1 predikciós algoritmus a Hi-CRISPR B és C teljesítményeit három kiegyensúlyozott (50 % hatékonyan és 50 % gyengén hasított szekvencia– csíkozott oszlopok) és három kiegyensúlyozatlan (93,4 % hatékonyan és 6,6 % gyengén hasított szekvencia– szürke oszlopok) teszt adathalmazon hasonlítottuk össze. A kiegyensúlyozott és nem kiegyensúlyozott adathalmaz szekvenciáit véletlenszerűen válogattuk le a 100K teszt adatkészletből. Az MCC értékek igen **(C)**, de a G-átlag értékek nem **(D)** érzékenyek arra, hogy kiegyensúlyozott vagy kiegyensúlyozatlan adatkészleteket használunk-e. Az oszlopok a három adatkészletre vonatkozó átlagos predikció értékét a hibaszólók a szórást jelölik.

Mivel a G-átlag értékek csökkennek az 100K könyvtáron hosszabb motívumokon alapuló predikciók esetében (>4), miközben azok emelkednek a tanító adathalmazon, arra a következtetésre jutottunk, hogy ez a túltanítás jelensége miatt lehetséges **(23/B ábra)**. Ezért, a későbbi adatok predikciójához a NucCom4-et használtuk melyet Hi-CRISPR A-nak neveztünk el (a **High Fidelity CRISPR prediction** angol elnevezést rövidítve). A Hi-CRISPR A mellett két korábbi publikációban használt neurális hálón alapuló algoritmust (DeepCRISPR⁹⁴, DeepSpCas9¹²⁰) is felhasználtuk az adatok predikciójára, melyeket szintén az 1M könyvtáron tanítottunk. Ezeket Hi-CRISPR B és Hi-CRISPR C-nek neveztük el.

Azt vizsgálándó, hogy a különböző összetételű adathalmazok mennyire befolyásolják a predikció minőségét, a 100K adathalmazból véletlenszerűen kiválasztottunk néhány ezer szekvenciát, úgy, hogy egy kiegyensúlyozott (a fele hatékony, a másik fele pedig rosszul hasított szekvencia) és egy nem kiegyensúlyozott (1,8% illetve, 6,6 %-a rosszul hasított a WT-SpCas9, illetve az SpCas9-HF1 esetében) adathalmazt kapjunk. Ezek az adatok a Hi-CRISPR predikciók maximum 0,94-es G-átlagot értek el az SpCas9-HF1-el függetlenül az adathalmaz összetételétől **(23/C, D ábra)**. A WT-SpCas9 esetében ezek az értékek kissé alacsonyabbak, feltehetően a nagyon kiegyensúlyozatlan WT-SpCas9 tanító adathalmaznak köszönhetően (lásd 117. referencia, 5. kiegészítő ábra). Továbbá a saját algoritmusainkat összehasonlítottuk korábban publikált WT-SpCas9 predikciókkal ezeken az adathalmazokon. Ahogy az korábbi tanulmányokból sejthető volt, az MCC értékek a prediktált adathalmaz összetételétől igencsak függenek, míg a G-átlag kevésbé volt érzékeny (lásd 117. referencia, 5. kiegészítő ábra). Azért, hogy az adatok összetételétől a kiértékelés ne függjön ennyire, a későbbi kísérletekben a G-átlagot használtuk.

5.2.4. Az SpCas9-HF1 aktivitásának előrejelzése emlős sejtekben

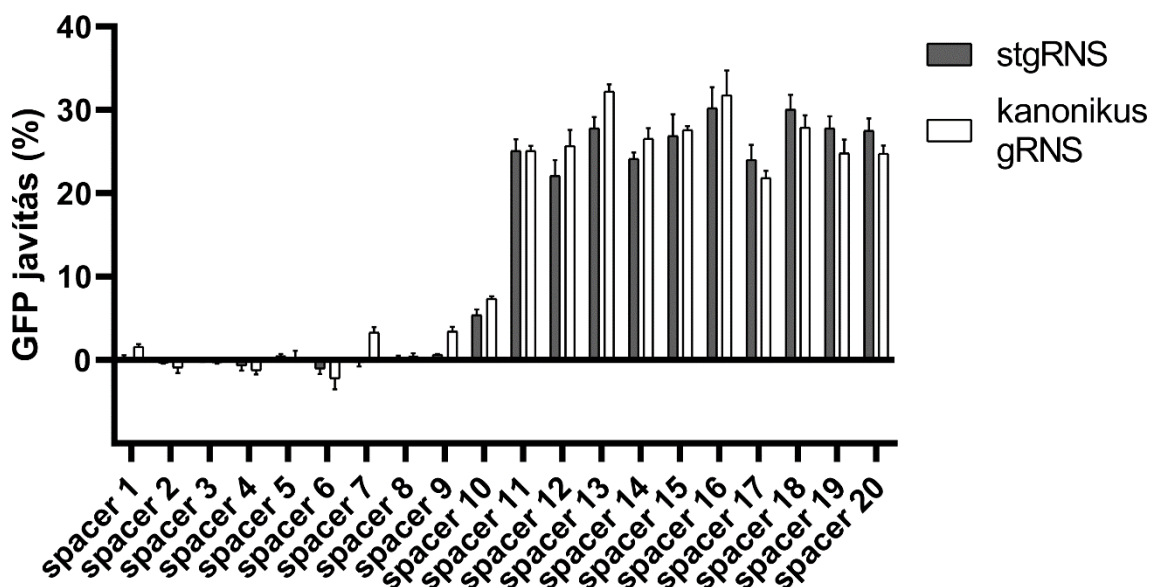
Kíváncsiak voltunk, hogy a bakteriális sejtekben kapott eredmények mennyire relevánsak emlős sejtekben. 57 olyan szekvenciát választottam ki, amelyeket korábban a bakteriális rendszerben már vizsgáltam, és N2a emlős sejtekben tanulmányoztam, hogy az SpCas9-HF1 milyen mértékben hasítja ezeket. Ehhez egy korábban publikált plazmid alapú fluoreszcens módszert a GFxFP esszét használtam^{85,86}, melyben a GFP fluoreszcens fehérje kódoló szekvenciája két félre van bontva homológ szekvenciákkal. A két fél közötti SpCas9 hasítás a homológ részek javítását indukálja a sejtben, ami a GFP szekvenciájának helyreállítását okozhatja (**6. ábra**). Az emlős sejtekben a szekvenciák többsége hasonlóan viselkedett, mint a bakteriális rendszerben. Ez arra engedett következtetni, hogy az SpCas9-HF1 aktivitását a sejt környezete számottevően nem befolyásolja (**24/A ábra**).



24. ábra. Az SLS szelekció eredményeinek validálása emlős sejtekből származó adatokon

(A) 20 hatékonyan és 37 gyengén hasított szekvenciát választottunk ki a bakteriális szelekciós könyvtárból, és ezeket N2a emlős sejtekben, egy GFP helyreállítási esszé (GFxFP esszé) segítségével vizsgáltuk. A bakteriális rendszerben hatékonyabban hasított szekvenciák (kék pontok) szintén hatékonyabbak voltak emlős sejtekben, mint a gyengén hasítottak (piros pontok). A hasított és nem hasított szekvenciák közötti különbség statisztikailag szignifikánsnak bizonyult ($p < 0,0001$; független mintás t-próba Kolmogorov-Smirnov normalitás teszttel). Az ábrán szereplő pontok három párhuzamos transzfekeció átlagát jelölik, amelyek mindegyike esetében a negatív kontrollok háttér GFP jelét kivontuk az SpCas9-HF1 kezelt mintákból. (B, C) A bakteriális könyvtárak segítségével azonosított kedvező és kedvezőtlen motívumok validálása egy 50000 szekvenciát tartalmazó emlős adatkészlet segítségével. A hőértékek az SpCas9-HF1, az eSpCas9 és a WT-SpCas9 átlagos aktivitásától való százalékos eltérését mutatják a jelzett motívumokat tartalmazó összes szekvenciában, az 50000 szekvenciát tartalmazó emlős könyvtárban. Az ábrákon az SpCas9-HF1 aktivitását legjobban elrontó 15-15 olyan négy (B) és öt (C) nukleotid hosszú motívum látható, melyeket a bakteriális SLS szelekcióban azonosítottunk. Az adatok alapján ezek a motívumok emlős sejtekben is kedvezőtlenek az SpCas9-HF1 aktivitására. A motívumok némelyike csak a két megnövelt pontosságú SpCas9 variánst érinti (felső panel), míg a WT-SpCas9-et nem, hasonlóan a bakteriális könyvtárban kapott adatokhoz. A hőértékeken az adott motívum szekvenciája, valamint a spacerben elfoglalt helyzete (a motívum első bázisának száma) is fel van tüntetve. (D) A Hi-CRISPR A (magenta), B (kék) és C (sárga) előrejelzések összehasonlítása a DeepHF (fekete) algoritmussal (G-átlag értékek ábrázolásával) öt független adathalmazon, valamint a DeepHF tanításához használt könyvtáron (Wang et al.).

Továbbá, arra is kíváncsiak voltunk, hogy van-e különbség a kanonikus, valamint az öncélzó gRNS-ek aktivitásában. SpCas9-HF1 fehérjével GFxFP rendszerben tíz (korábban a bakteriális rendszerben) aktív és tíz kevésbé aktív spacer segítségével hasonlítottam össze a két gRNS típust (25. ábra). Az SpCas9-HF1 aktivitása erős korrelációt mutatott a két gRNS típussal ($r=0,98$), ami arra enged következtetni, hogy a fehérje aktivitását emlős sejtekben elsősorban annak jellemzői határozzák meg, nem pedig az öncélzó és a kanonikus gRNS közötti különbségek.



25. ábra. Az SpCas9-HF1 aktivitása 20 különböző célszekvencián kanonikus vagy öncélzó gRNS-ekkel

Az SpCas9-HF1 hasítási aktivitását kanonikus (szürke oszlopok) és öncélzó gRNS-ek használatával (fehér) hasonlítottuk össze GFxFP esszé segítségével, N2a emlős sejtekben. Húsz különböző spacert választottunk ki az 100K bakteriális könyvtárból, amelyek közül tíz kevésbé hatékony (spacer 1-10) és tíz pedig hatékony (spacer 11-20) volt a bakteriális SLS szelekció alapján. A két gRNS típus aktivitása erősen korrelált (Pearson $r=0,98$) a 20 célszekvencián. Az oszlopok három párhuzamos transzferenció átlagát, a hibaszívek a szórást jelölik.

Kísérleteink kiértékelése közben megjelent egy publikáció, amely az SpCas9-HF1 aktivitásának predikciójára fejlesztett ki egy algoritmust (DeepHF)¹¹², amely 50000 genomi célszekvencia hasításán alapul. Ez az adathalmaz segíthette a megközelítésünk további értékelését, hiszen eddig nem állt rendelkezésre emlős sejtekben készült nagyobb adathalmaz, amely SpCas9-HF1 hasítási információt tartalmazott volna. Ebben a publikációban az 50000 gRNS hasítását nem csak a WT-SpCas9-en és az SpCas9-HF1-en vizsgálták, hanem a megnövelt pontosságú eSpCas9-en is.

Először azt vizsgáltuk meg, hogy az SpCas9-HF1 aktivitását legjobban befolyásoló szekvencia motívumok, amelyeket már korábban azonosítottunk bakteriális sejtekben (21. ábra), ugyanúgy megtalálhatóak-e ebben az emlős sejtekben készített adathalmazban. Itt is 15-15 olyan négy (24/B ábra) és öt (24/C ábra) bázis hosszú motívumot választottunk ki, amelyek a leginkább elrontották az SpCas9-HF1 aktivitását bakteriális sejtekben. Amennyiben ezek a motívumok emlős sejtekben is elrontják az SpCas9-HF1 aktivitását ezen motívumokkal rendelkező spacerok hasítási hatékonysága

el kell, hogy térjen az átlagtól. Ezekről azt találtuk, hogy valóban eltérnek. Mint a bakteriális adatok kiértékelésénél itt is két csoportra osztottuk ezeket a motívumokat.

Az első csoportban olyan PAM közeli motívumok találhatóak, amelyek mind az SpCas9-HF1, mind a WT-SpCas9 aktivitását negatívan befolyásolták korábban. A bakteriális adatokkal összhangban ezen GCC tripletet tartalmazó motívumok közül volt, amely 78 %-kal csökkentette az SpCas9-HF1 és 80 %-kal a WT-SpCas9 átlagos hasítási hatékonyságát a spacer 18-20-as pozícióiban. A bakteriális adatoknak megfelelően ez a motívum a 16-18, valamint a 17-19 pozíciókban is rontotta mindkét fehérje aktivitását, melyeket korábbi tanulmányok nem azonosítottak, beleértve a vizsgált adathalmazt közlő publikációt is.

A másik csoportban olyan motívumokat találtunk, amelyek (szintén a bakteriális könyvtárakon kapott adatokkal összhangban) az SpCas9-HF1 aktivitását negatívan befolyásolták, de a WT-SpCas9-re nincsenek hatással. Ezek többsége a GTNAC motívummal fedett át a spacer 10-14-es pozícióiban. Azt is szeretnénk volna kideríteni, hogy ezek a motívumok csak az SpCas9-HF1-re jellemzőek-e vagy pedig általánosabban más megnövelt pontosságú SpCas9-re is hatással lehetnek. Így a DeepHF adathalmaz eSpCas9 hasítási adatait is elemeztük. Az első csoportba tartozó motívumok az eSpCas9 hatékonyságát is (a WT- és SpCas9-HF1-hez hasonlóan) meglehetősen csökkentették. Érdekes módon a második csoportba tartozó motívumok - amelyek a WT-SpCas9 hatékonyságát nem befolyásolták - az eSpCas9 hasítási hatékonyságát is csökkentették, körülbelül SpCas9-HF1-el látottak felére.

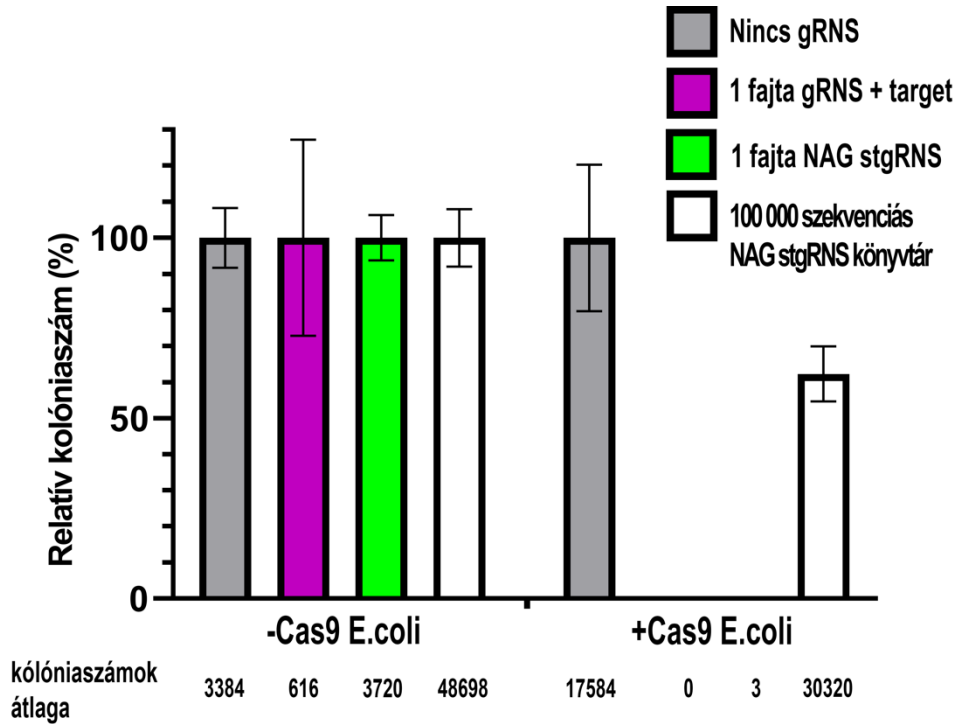
Ezután megvizsgáltuk, hogy az egymillió SpCas9-HF1 bakteriális hasítási adat alapján a legjobbnak bizonyult szekvencia motívumok is észrevehetőek-e a DeepHF emlős adatkészletben. A legjobb 15 négy bázisos motívumok látszólag nem eldönthetően, az öt bázis hosszú motívumok pedig látszólag egyértelműbben a megnövelt hűségű variánsokra voltak pozitív hatással, míg a WT-SpCas9-re nem voltak hatással. A probléma pontosabb körüljárása érdekében egyfajta kontrollként 10 különböző, egyenként 15 véletlenszerű motívumot és az ezekhez tartozó átlagos hasítási értéket tartalmazó adathalmazt hasonlítottuk össze mind a legjobb és a legrosszabb 4 és 5 bázis hosszú motívumkészletet az összes fehérje esetében (lásd 117. referencia, **8. kiegészítő ábra**). A WT-SpCas9 esetében sem a legrosszabb sem a legjobb motívumok nem voltak

statisztikailag megkülönböztethetők a véletlenszerűen generált motívumoktól. Az eSpCas9 és SpCas9-HF1 esetében mind a négy mind az öt bázis hosszúságú motívum-szett szignifikánsan különbözött a véletlenszerű adathalmazoktól. A legjobb motívumok esetében a négy bázis hosszú motívumok egyik fehéjénél sem, az öt bázis hosszúak viszont az eSpCas9 és SpCas9-HF1 esetében a legtöbb véletlenszerű adathalmaztól szignifikánsan különbözött.

Az általunk kifejlesztett 1 millió bakteriális szekvencián tanított Hi-CRISPR predikcióinkat szeretnénk volna összehasonlítani a DeepHF emlős adathalmazon készült predikcióval. Öt kisebb adatkészletet gyűjtöttünk össze a szakirodalomból, valamint a DeepHF algoritmus kidolgozásához használt 50000 spacert tartalmazó adathalmazt használtuk a predikciók teljesítményének összehasonlítására. A négy előrejelzési algoritmus hasonlóan jó teljesítményt mutatott ezeken az adathalmazokon, a Hi-CRISPR-ek valamivel magasabb átlagos G-átlag értékeket értek el az öt kisebb adatkészleten (0,85 összevetve 0,83-mal, **24/D ábra**). Ahogy az sejthető volt, a DeepHF jobban teljesített a nagy adatkészleten, amelyen azt tanították.

5.2.5. Az SpCas9 alternatív PAM felismerése és az NmCas9 öncélzó gRNS-ekkel

A koncepciót igazolandó, megvizsgáltuk, hogy az SpCas9 nem-kanonikus PAM felismerése (NGG helyett NAG) összeegyeztethető-e a mi módszerünkkel^{121,122}. Az stgRNS PAM részét NAG-ra cseréltük, és hatékonyan működött a megváltoztatott RNS *E. coli* sejtekben is egy korábban emlős sejtekben NAG PAM-mal működő spacerrel (**26. ábra**). Létrehoztam egy NAG PAM-mal rendelkező öncélzó gRNS könyvtárat és ezt a WT-SpCas9-cel hasítottam. A kísérlet azt mutatta, hogy a fehérje a szekvenciáknak csak a 37 %-át hasította el. Ezt összevetve a korábbi WT-SpCas9 NGG PAM-mal való hasításával (ahol a szekvenciák kb. 98%-át hasította el), látható, hogy a NAG PAM-mal a fehérje sokkal szelektívebb. Ezekből arra lehet következtetni, hogy ezt a tulajdonságot, valamint más módosított PAM specificitású Cas9 variánsok szekvencia-szelektivitását is lehetséges meghatározni az SLS módszerrel.



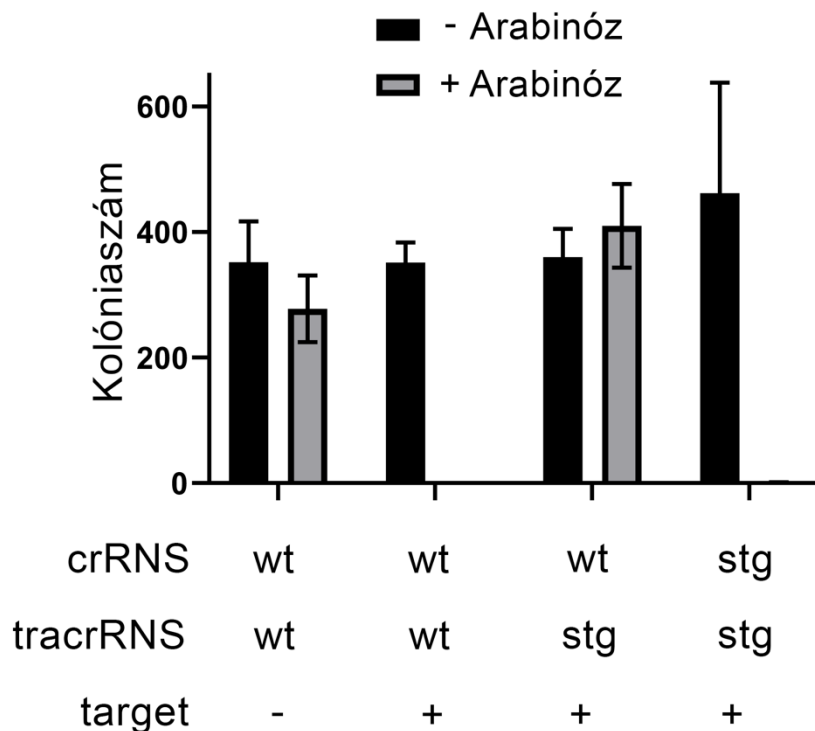
26. ábra. Az SpCas9 nem-kanonikus PAM felismerése stgRNS-ekkel

Az öncélt gRNS-t úgy módosítottuk, hogy a PAM szekvencia részén az SpCas9 kanonikus „NGG” PAM felismerő helyét a nem kanonikus „NAG” PAM-ra cseréltük. A NAG stgRNS aktivitását egy olyan spacerrel vizsgáltuk, mely korábban N2a sejtekben vad típusú gRNS struktúrával és NAG PAM-mal magas aktivitást mutatott. Az „öngyilkos” plazmid Cas9-cel nem rendelkező, és Cas9-et expresszáló *E. coli* sejtekbe transzformálása után a túlélő sejtek száma több mint három nagyságrenddel csökkent (zöld oszlop, a jobb oldali panelen nem látható méretű), a NAG stgRNS aktívan működött. Kontrollként egy olyan plazmidot használtunk, amelyen egy kanonikus gRNS ugyanezzel a spacer szekvenciával rendelkezett, valamint a plazmidon a célszekvencia is jelen volt (magenta, a jobb oldali panelen nem látható méretű). Negatív kontrollként egy hasonló gerinccel rendelkező, de gRNS-t nem kódoló plazmidot használtunk (szürke), amely a két sejt kompetenciájának összehasonlításához is fel tudunk használni. Végül, egy 100000 szekvenciából álló NAG stgRNS könyvtárat hoztunk létre és ezeket is a két sejtbe transzformáltuk (fehér). A kontrollhoz képest az SpCas9 a szekvenciák 37%-át elvágta. Az oszlopok öt párhuzamos transzformáció kólónia-számainak átlagát a hibázászlók a szórást jelölik. A kolóniák valós számát az oszlopok alatt jelöltük.

Az SpCas9 több ortológját is rutinszerűen használják emlős sejtekben, de ezek közül azonban csak a *Staphylococcus aureus* Cas9-re fejlesztettek ki predikációs eljárást¹²³. Kíváncsiak voltunk arra, hogy más ortológok számára is előállíthatók-e öncélt gRNS-ek. Szintén koncepciótervként, a *Neisseria meningitidis* Cas9-cel (NmCas9) asszociált RNS-eket (itt gRNS helyett, crRNS-t és tracrRNS-t használunk, lásd 3/A ábra) a már

korábban bemutatott technikával módosítottuk. A crRNS spacer melletti és a tracrRNS megfelelő nukleotidjait úgy változtattuk meg, hogy tartalmazza az NmCas9 PAM szekvencia motívumát (NNNNGATT).

Azt találtuk, hogy az így módosított öncélzó cr-, és tracrRNS-el a nukleáz a kanonikus RNS-ekhez hasonló aktivitást mutatott *E. coli* baktériumokban (27. ábra). Ha a vad típusú crRNS-t egy öncélzó tracrRNS-sel párosítottuk, nem mutatott aktivitást, még akkor sem, ha a megfelelő PAM-mal rendelkező célszekvenciát is elhelyeztük a plazmidon. Ezekből arra lehet következtetni, hogy az öncélzó gRNS-ekkel számos Cas9 ortológ kompatibilis lehet, és hogy ezek szekvencia-specifitását lehetséges meghatározni a módszerünkkel.



27. ábra. Az SpCas9 egyik ortológja, az NmCas9 esetében is létrehozhatók működő öncélzó RNS-ek

Az NmCas9 crRNS-ét úgy módosítottuk, hogy egy PAM-szekvenciát kódoljon (ugyanúgy, mint korábban az SpCas9 esetében) és a tracrRNS komplementer pozícióit is ezeknek megfelelően mutáltuk, hogy az RNS másodlagos szerkezete ne változzon, így kialakítva az öncélzó cr- és tracrRNS-eket. A konstrukciót egy negatív szelekciós rendszerben vizsgáltuk, hasonlóan az SpCas9-HF1-hez, azzal a különbséggel, hogy ebben a rendszerben az NmCas9 expressziója indukálható arabinóz hozzáadásával. Az ábrán a kolóniák száma látható indukció nélkül (fekete oszlopok) valamint indukcióval (szürke oszlopok) a különböző crRNS/tracrRNA/target

kombinációk esetében. Az NmCas9 indukálása a kolóniaszám jelentős csökkenését eredményezi, ha a vad típusú cr- és tracrRNS-eket a megfelelő célszekvenciával rendelkező plazmiddal használjuk (második oszloppár), valamint akkor is, amikor öncélzó crRNS (stg) használunk a hozzá tartozó módosított tracrRNS-sel (negyedik oszloppár). Ez azt jelzi, hogy az NmCas9 aktív a módosított RNS-ekkel. Érdekes módon, amikor egy vad típusú tracrRNS-t kombináltunk az öncélzó crRNS-el (harmadik oszloppár), az NmCas9 nem hasít, tehát (az SpCas9-HF1 kísérletekkel összehangban) ez az RNS-ek másodlagos szerkezetét megzavarja, ami a fehérjét működésképtelenné teszi. Negatív kontrollként egy vad típusú cr- és tracrRNS-ekkel rendelkező plazmidot használtunk, célszekvencia nélkül (első oszloppár).

6. Megbeszélés

A Cas9 nukleázok használata forradalmasította a genomszerkesztést és ma már a különböző alkalmazások teljes arzenálja segíti az orvosi és alapkutatási munkákat. Mint minden új módszernek, ennek is számos korlátja létezik, amelyek kiküszöbölésével még hatékonyabb és pontosabb lehet ez a technológia.

Az egyik ilyen korlátja a rendszernek, hogy a Cas9 nukleázok szelektívek a célozni kívánt szekvenciára, és a különböző összetételű célszekvenciákat eltérő hatékonysággal hasítják. Mind a klinikai felhasználásban, mind a felfedező kutatásokban is rendkívül fontos, hogy a célzott DNS szekvencia hasítása hatékony legyen. A DNS Cas9 nukleázokkal való kettős szálú törése jelöli ki és indukálja azokat a hibajavító mechanizmusokat, amelyeket gének kiütésénél, vagy egy esetleges kívánt szekvencia genomba való bevitelénél kihasználunk. Minél hatékonyabb a DNS hasítása annál hatékonyabban tudjuk ezeket a feladatokat elvégezni. Ezért rendkívül fontos előre tudni, hogy az esetleges gRNS jelöltekből a leghatékonyabbat válasszuk ki egy későbbi genommodosítási feladatra. Ezt az igényt kielégítendő, számos módszer született a hasítási hatékonyság kísérletes meghatározására.

A PCR alapú módszerek (pl. NGS, TIDE vagy T7EI) hátránya, hogy nem használhatók, ha maga a PCR nem megvalósítható például, mert ismétlődő szekvenciák vannak a célszekvencia körül vagy a célszekvencia egy GC gazdag régióban helyezkedik el. További probléma lehet, hogy ha a DNS hasítás következtében egy nagyobb szakasz esik ki, mint amire az azt detektáló PCR primerek tervezve vannak, akkor ezeket az eseményeket nem detektáljuk. Egy további korlátjuk, hogy a sikeres és specifikus genomi PCR is nagyon sok esetben optimalizálást kíván.

Jó alternatívák a fluoreszcens kiértékelésen alapuló módszerek, melyek nem igénylik a PCR használatát, valamint gyorsan kiértékelhetők. Mi egy olyan áramlási citometrián alapuló fluoreszcens módszert fejlesztettünk ki, amely képes könnyen és megbízhatóan a gRNS-ek hatékonyságát genomi kontextusban vizsgálni. Számos más fluoreszcens riporter rendszer minden vizsgált gRNS-re új klónozási lépést vagy lépéseket igényel^{85,124-126}. Ezzel szemben a mi rendszerünket előzetesen már elkészített bármilyen gRNS-sel további klónozási lépések nélkül egyszerűen lehet használni. További előny, hogy az aktivitás hatékonyságát nem egy plazmidon, hanem a gRNS genomi

kontextusában képes vizsgálni, az oda beépített GFP expressziós kazettával. Minél hatékonyabb egy gRNS annál hatékonyabb lehet a genomi beépülés. A még hatékonyabb beépítést egy önvágó plazmiddal értük el, amely a sejten belül linearizálódva éri el a hatékony integrációt. A módszer publikálásának idejében további leküzdendő problémát jelentett, hogy az elérhető SpCas9 predikciós módszerek rendkívül pontatlanok voltak és praktikusán nem lehetett őket használni^{88-90,92}. Azóta ezek az algoritmusok fejlődésen mentek keresztül^{94,95,112}, de az a probléma továbbra is fennáll, hogy a kifejlesztett Cas9 variánsok^{95,112} és egyéb Cas fehérjék^{59,127} csak töredékére érhető el, megbízható predikciós eljárás.

A módszerünket limitálta, hogy hiába használható jóval egyszerűbben, mint a korábbi módszerek, a GFP kazetta beépülési hatékonysága csak a tranziens expresszió lecsengésével volt kiértékelhető, amely 12-15 napig is tarthat. Egy indukálható degradációs domén fúziójával ez az időt 6-8 napra lehetett csökkenteni, ami körülbelül annyi, mint a szekvenálás alapú módszerek időtartama, viszont valamivel hosszabb, mint a T7EI vagy Surveyor nukleáz használó módszereké. Az eredményeinkből az is nyilvánvalóvá vált, hogy a véletlenszerű integráció a sejtek típusától és állapotától függően kissé eltérhet. Ezért a célzott integráció pontos meghatározásához szükséges egy negatív kontroll használata minden kísérletben, mint a sejten belül hasított, nem célzottan integrált plazmid kontroll. Továbbá, a módszerünk a plazmid integrációját méri, és nem tesz különbséget, a célzott (on-target) és off-target beépülések között. Ez azonban alacsony tartható a spacer szekvenciák gondos kiválasztásával, off-target kereső programok^{105,128}, valamint a megnövelt pontosságú SpCas9 nukleázok használatával⁷²⁻⁷⁷.

A Cas9 által okozott kettős szálú DNS törések és egy exogén szekvencia NHEJ-mediált beépítése közötti korreláció jellegét és mértékét nehéz felmérni, mivel nem nagyon létezik módszer a sejtekben kialakuló kettős száltörések közvetlen mérésére. A legtöbb módszer, mint például az NGS⁷⁴, a TIDE⁸⁰ vagy a GUIDE-seq¹²⁹ a DNS törések nem tökéletes javítása után fellépő véletlenszerűen beépült vagy törlődött (indel) szekvenciákat képes mérni. A sejten történő száltörések, valamint a javított szekvenciák közötti összefüggést sem lehet egyértelműen kideríteni. A DNS törés után fellépő javítás sok tényezőtől függ, mint a hasítás körül lévő szekvenciáktól, valamint az itt megtalálható mikrohomológiák jelenlététől és pozíciójától, amelyekről úgy gondolják, hogy nagyban befolyásolják a hibamentes és a hibával terhelt NHEJ és mikrohomológia-mediált javítás

(MMEJ) arányát¹³⁰⁻¹³². A HR és az NHEJ által közvetített javítás arányát befolyásoló feltételeket és tényezőket szintén nehéz kideríteni. Emellett mindkét fő javítórendszer képes a töréspontokra exogén DNS-t beépíteni, de a tényezők, amelyek ezt befolyásolhatják, szintén kevésbé ismertek, mint például az NHEJ-mediált integráció és az NHEJ-javítás aránya. Ezenkívül az alkalmazott módszerek nem feltétlenül mérik pontosan a javítási rendszer által generált mutációk gyakoriságát. Például nem teljesen ismert, hogy az indelek következtében fellépő mutációk eloszlása és jellege hogyan befolyásolja a Surveyor/T7EI nukleáz-vizsgálatok eredményét^{133,134}. Továbbá a PCR-alapú módszerek általában figyelmen kívül hagyják a nagyobb inszerciókat illetve deléciókat, amelyek előfordulásának aránya szintén befolyásolhatják a kisebb indelek számát. Így nem nyilvánvaló, hogy az gRNS-ek relatív hatékonysága azonos lenne, ha két különböző módszerrel mérnénk őket. Mindazonáltal minden itt megemlített megközelítés alkalmas lehet arra, hogy megkülönböztessük velük a nagyon gyenge és a hatékony spacer szekvenciákat.

A módszerünket először a HR közvetített célzott integrációval hasonlítottuk össze, ugyanazokat a gRNS-eket használva. Ügyeltünk arra is, hogy a gRNS-ek homológ karokhoz viszonyított pozíciója ne befolyásolja a homológ rekombináció hatékonyságát. Habár lehet, hogy más nehezebben kideríthető tényezővel is lehetne számolni, ez lehetővé tette a két javítórendszer által közvetített gRNS-ek hatékonyságának közvetlenebb összehasonlítását. A módszert továbbá - azonos gRNS-ek használatával - a TIDE, és az T7 endonukleáz I esszével is összehasonlítottuk, melyek a Cas9 által kiváltott indel gyakoriságokat mérik. Ezek a kísérletek megerősítették, hogy megközelítésünk képes megkülönböztetni a gyenge és hatékony spacer szekvenciákat, és hasznos, könnyen használható alternatívája lehet ezeknek a módszereknek a hatékony gRNS-ek kimutatására.

A módszerünk elengedhetetlen része, hogy egy GFP-plazmidot a sejten belüli linearizáljunk. Érdekes módon, ha a plazmid nem a sejten belül, hanem restriktációs enzimekkel linearizáltan kerül a sejtbe, az integráció hatékonysága meglehetősen romlik. A sejten belül hasított plazmidok integrálásának pontos mechanizmusa nem jól ismert. Feltételezzük, hogy a különbség abból adódhat, hogy a kör alakú plazmidok a sejten belüli DNázokkal szemben ellenállóbbak lehetnek, mint a lineáris fragmensek. Az is elképzelhető, hogy a lineáris DNS végek olyan módosításokat szenvednek a sejten belül,

amelyek azok integrációját megnehezíti, míg a cirkuláris plazmidok ezeknek a faktoroknak kevesebb ideig vannak kitéve.

Az NHEJ-közvetített célzott génbevitel rendkívül hasznos eszköz lehet, ha nincs szükség a DNS-végek tökéletesen pontos összekapcsolására, valamint, ha el lehet kerülni vele azt a fáradságos munkát, hogy minden egyes genomi szekvenciára új homológ karokat kelljen tervezni és klónozni. További hasznos felhasználása a módszerünknek, hogy könnyű egyszerű funkcióvesztéses allélokat létrehozni, a DNS kazetta célzott beépítésével. A célzott beépítés után a génkiütött fluoreszcens klónok szelekciója rendkívül egyszerű FACS használatával. Továbbá, módszerünket sikeresen alkalmaztuk arra, hogy egy több mint 10 kilobázis nagyságú kazettát szelekció nélkül több mint 30 %-os hatékonysággal a genomba hatékonyan beépítsünk. Azt is demonstráltuk, hogy egy szelekciós marker segítségével, két expressziós kazettát 90 %-os hatékonysággal kapcsoltan lehet beépíteni a genomba. Nem fluoreszcens kazetták beépítésekor is nagyon jó hatékonysággal tudtuk használni a szelekciós markerrel ellátott önvágó plazmid rendszert, mivel itt is egy expressziós kazetta bevitele volt a cél, tökéletesen pontos integráció nélkül.

Az NHEJ klónozást szekvenciák pontos beépítésére, például endogén gének jelölésére is lehetséges használni, habár csökkent hatékonysággal. Mivel a donor DNS két orientációban és három leolvasási keretben is integrálódhat, a célzott beépülési események közül csak néhány származik a donor DNS jó leolvasási keretbe építéséből. Ennek hatékonysága attól is függhet, hogy az adott DNS szekvencia beépítése közben a pontos vagy a hibával terhelt NHEJ vagy MMEJ preferált vagy, hogy az inszerció helyén lévő szekvenciák bizonyos indeleket előnyben részesítenek-e a többivel szemben. A hátrányok ellenére, igaz csökkent hatékonysággal, de képesek voltunk a módszerrel az *SPRN* gén START kodonja mögé promoterral nem rendelkező GFP szekvenciákat hibamentesen bevinni⁷⁴. Megjegyzendő, hogy ilyen génbevitel alkalmazásakor, amennyiben nem akarjuk, hogy az önvágó plazmidon található gRNS expressziós kazetta is beépüljön úgy az **8. ábrán** található módszert kell követni, és egy külön plazmidon elhelyezett univerzális gRNS-sel kell felnyitni a plazmidot, mint ahogy az az idézett publikációban is történt.

A gRNS-ek hatékonyságának felderítésére egy tökéletes előrejelző algoritmus lenne a legjobb megoldás, amely bármilyen spacer szekvenciáról el tudja dönteni, hogy az adott kísérletben mennyire fog hatékonyan működni. Ezt az elméleti célt nem sikerült még elérnie a kutatóknak, de a WT-SpCas9 esetében már léteznek viszonylag hatékonyabbnyire gépi tanulást alkalmazó predikciós eljárások^{94,95,112}. A különböző alkalmazásokra fejlesztett Cas9 variánsok száma több nagyságrenddel meghaladja, az ezeket prediktáló algoritmusok számát. A predikciós eljárások alacsony száma nem meglepő, hiszen ilyen algoritmusok kifejlesztése, nagy költséget és időt jelentő feladat. Mindazonáltal, a különböző megváltoztatott PAM felismerésű, vagy megnövelt pontosságú SpCas9 nukleázok rutinszerű használatához szükség lenne olyan megbízható predikciókra, amely a WT-SpCas9-re már elérhetőek. Célunk volt, az SpCas9-HF1 variánsra is megbízható predikciós eljárást fejleszteni, hiszen a pontosabb nukleázok használata minden alkalmazás esetében preferált, és ez az egyik legtöbbet használt ilyen variáns.

Egy olyan módszert fejlesztettünk ki, amellyel a korábbi módszerekkel ellentétben, teljesen véletlenszerű, és nagyságrendekkel nagyobb mennyiségű szekvencián alapuló predikciós eljárást lehet létrehozni Cas9 variánsokra. Az ilyen előrejelzések kifejlesztésénél, szűk keresztmetszet a nagy mennyiségű és megbízható Cas9 hasítási információk létrehozása, amelyet a célkönyvtár létrehozási módszerével, valamint a szelekciós SLS rendszerrel meg lehetett valósítani. Az SpCas9-HF1-el történt korábbi vizsgálatok, összesen néhány száz hasítási adatot szolgáltatottak⁷³⁻⁷⁵, melyeket jóval megnövelt Wang és mtsai. által nem rég publikált 50000 gRNS¹¹². A mi módszerünkkel generált több mint 1 millió hasítási információ jóval több mint az összes eddigi publikált együttvéve, így azt reméltük ezek alapján megbízható előrejelzést tudunk kifejleszteni erre az SpCas9 variánsra is.

A módszerünk öncélzó gRNS-eket (stgRNS) használ, melyek a saját expressziós plazmidjaikat hasítva „öngyilkos” módon működnek. Egymillió ilyen stgRNS-ből álló könyvtárat készítettünk, majd ezeket WT-SpCas9 és SpCas9-HF1-et expresszáló baktériumokba transzformáltuk (SLS szelekció). A hatékony szekvenciákat a Cas9-ek elvágják, a plazmid lebomlik, így az eredeti, illetve a szelektált könyvtárakat összehasonlítva, megállapíthatók a jól és kevésbé jól vágható szekvenciák. Az SLS megközelítés előnye, hogy a célzott szintézisen alapuló módszerekkel szemben, egy ilyen

hatalmas mennyiségű adat létrehozásához kevesebb időt és költséget igényel. Az előrejelző algoritmusok fejlesztésénél komoly problémát szokott okozni, a nagy számú hasítási információ létrehozása, megismételhetősége, valamint egy olyan rendszer kidolgozása, amelyben egyértelmű az adatok kiolvasása. Ezekre itt mind megoldást találtunk, illetve arra számítottunk, hogy ez a módszer a későbbiekben megkönnyítheti az SpCas9 számos ortológjának és variánsának a szekvencia-specificitásának a feltárását.

Némileg meglepő volt, hogy egy bakteriális adatokon kifejlesztett predikciós eszköz ilyen jól teljesít az emlős adatokból származó hasítási információkon, amely egy emlős adatokon tanított predikcióval¹¹² összemérhető hatékonyságú. Úgy véljük, hogy ez annak tulajdonítható, hogy a megközelítésünkben használt sokkal nagyobb számú szekvencia kedvező hatásai, a bakteriális és emlős rendszerek közötti különbségek negatív hatásait ellensúlyozzák. Az adataink arra utalnak, hogy bakteriális és emlős sejtekben az SpCas9-HF1 aktivitásában való különbség kisebb, mint egyes konkrét kísérletek körülményei közt. Emlős sejtekben soha nem konkrétan az SpCas9-HF1 aktivitását mérjük, hanem az ezt követő hibával terhelt DNS javítás eredményét, hiszen a pontos javítás az eredeti szekvenciától megkülönböztethetetlen, hiába volt elhasítva. Az ilyen javítások eredményét a törések körüli szekvenciák befolyásolhatják. Ez olyannyira igaz, hogy nem rég megjelent több publikáció is bizonyította, hogy ezen javítások eredményét megbízható módon prediktálni is lehet¹³⁵⁻¹³⁷. Ezzel szemben, homológ rekombinációval történő javítást már kevésbé befolyásolhatják a törés körül elhelyezkedő szekvenciák. Egy tökéletes predikciós algoritmustól viszont azt lehetne elvárni, hogy csak a Cas9 hatékonyságát mérje függetlenül a használt sejttypustól, DNS javítástól vagy a transzfekeziós hatékonyságtól. Ezek a faktorok valószínűleg az emlős sejtekben készült predikciókat legalább annyira befolyásolják, mint az általunk fejlesztett algoritmusét a bakteriális környezet.

A szelekciós kísérletek adatainak elemzése azt mutatja, hogy az SpCas9-HF1 szekvencia preferenciája eltér a vad típuséhoz képest, valamint előbbi módosított szekvencia preferenciája a spacer középső régiójában található. Ez a következtetés nem csak bakteriális, de emlős adatok alapján is helytállónak bizonyult, valamint kiderült, hogy érdekes módon ezt a tulajdonságot az eSpCas9 is hordozza. Megállapítottuk, hogy az SpCas9-HF1 a spacer középső régiójában (6-13. pozíciókban) magasabb GC tartalmú szekvenciákat részesít előnyben, más régiókhoz képest. Ezzel szemben a 3-7 bázis hosszú

GC tartalmú motívumok a WT-SpCas9 esetében minden régióban negatívan befolyásolták az aktivitást. A megnövelt pontosságú fehérjék tervezésének szempontjai megmagyarázhatják ezeket a megfigyeléseket. Ezen variánsok tervezésekor a fehérjék azon aminosavait mutálták, amelyek gyengítik RNS-DNS hibrid hélixek kötését és így érik el a megnövelt pontosságot, így a spacer középső régiójában található GC-tartalom előnyben részesítése ellensúlyozhatja ezeket a hatásokat. Általánosságban elmondható, hogy az SpCas9-HF1 DNS kötése a vad típusúhoz képest nem változik^{75,77}, hanem inkább az aminosav mutációk szabályozzák a hasítási aktivitását, és megakadályozzák, hogy a nukleáz a konformációs ellenőrzési ponton áthaladjon a felismerési és hasítási folyamat során¹³⁸. Tehát, az adataink azt sugallják, hogy a spacer szekvenciák középső része kritikus szerepet játszhat abban, hogy a megnövelt pontosságú SpCas9 nukleázoknál ki tudjon alakulni a hasítás-kompetens konformáció. Az adatok elemzése arra is rámutatott, hogy mind bakteriális, mind emlős sejtekben léteznek olyan motívumok a spacerben melyek elrontják a WT-SpCas9 és a megnövelt pontosságú SpCas9-ek aktivitását is; valamint azonosítottuk a GTNAC motívumot, amely csak az eSpCas9 és az SpCas9-HF1 hasítására volt igen negatív hatással, a WT-SpCas9 aktivitását nem befolyásolta.

A módszerünk arra alkalmas, hogy azonosítsuk azokat a szekvencia-jellemzőket, amelyek a legerősebben gátolják az SpCas9-HF1 hasítási hatékonyságát. Ezért inkább azokat a szekvenciákat tudja jobban meghatározni, amelyek kerülendők ahelyett, hogy kifinomultabb különbséget tenne az átlagos, vagy kifejezetten jól működő szekvenciák között.

7. Következtetések

Az itt bemutatott önvágó plazmid módszert sikeresen alkalmaztuk arra, hogy a hatékony és kevésbé hatékony gRNS-eket kísérletesen azonosítani tudjuk. A gRNS-ek hatékonyságának mérése mellett, a módszert fel lehet használni nagyobb DNS szekvenciák homológia-független (NHEJ-mediált) beépítésére is a genomba. A vonatkozó publikáció¹³⁹ elbírálása alatt¹⁴⁰⁻¹⁴², valamint a publikáció megjelenése után¹⁴³⁻¹⁴⁵ is számos hasonló elven működő módszert mutattak be. Ezek közül volt amely kevésbé hatékonyan^{141,144}, de volt amely viszonylag hatékonyan alkalmazta szekvenciák NHEJ-mediált beépítését, mint például a „homology-independent targeted integration” (HITI) módszer¹⁴⁰. Közös jellemzőjük, hogy az integráció hatékonyságát azzal növelik meg, hogy a bevinni kívánt szekvenciát tartalmazó PCR terméket¹⁴⁴ vagy plazmidot intracellulárisan hasítják. Ezeknek a módszereknek előnye, hogy az integráció hatékonysága a homológ rekombinációhoz képest magasabb, valamint az, hogy olyan sejtekben is lehetővé teszik gének bevitelét, ahol a HR nem, vagy csak limitáltan működik, például posztmitotikus sejtekben. A PCR termékek intracelluláris hasítása is a kevésbé hatékony génbeviteli módszerek közé tartozik. A hatékonyság csökkenése abban állhat, hogy nem egy cirkuláris plazmidot, hanem egy lineáris DNS hasítanak. A lineáris fragmensek jobban kitettek lehetnek a sejtben előforduló exonukleázoknak, valamint a két helyen hasított lineáris PCR termék három részre bomlik, így a bevinni kívánt szekvencia 1:2 arányban lesz jelen a másik két fragmenshez képest. A hasítás helyére történő integrációkor ezek a fragmensek kompetálhatnak és csökkenthetik a génbevitel hatékonyságát. A lineáris fragmensek sejtben történő hasítását az eredeti publikációkban mi is körüljártuk¹³⁹, és hasonló megállapításra jutottunk: a cirkuláris plazmidok intracelluláris hasítása jóval hatékonyabb, mint a lineáris fragmenseké.

A módszerünk génbevitelre történő hasznosítását HR deficiens CST sejtekben is kipróbáltuk és sikeresen jelöltünk egy gént a GFP fluoreszcens fehérjével. Itt a célzott gén utolsó intronját hasítottuk, majd a donor plazmidon a kiesett intron részt, valamint az utolsó exont és az ezzel leolvasási keretben lévő GFP fehérjét vittük vissza a sejtekbe. Ez a módszer azért hasznos, mert az intron szekvenciában jobban megtűrtek az esetlegesen keletkező inszerciók és deléciók, így ez nem befolyásolja nagy mértékben a kísérlet

sikerességét. Egy nem rég megjelent publikációban hasonló módszerrel, sejtekben linearizált plazmiddal hajtottak végre intron-jelölést egér neuronokban¹⁴³.

A gRNS-ek hatékonyságának meghatározása a kísérletes megközelítések mellett, megbízható *in silico* módszerekkel is lehetséges. A vad típusú SpCas9 aktivitásának meghatározására már több algoritmus is született a közelmúltban. Az évek során az SpCas9 fehérje mellett, több Cas9 ortológot, valamint más Cas fehérjéket sikerült genommérnökségre alkalmazni. A különböző Cas fehérjék, valamint SpCas9 variánsok specificitása eltérő a vad típusú SpCas9-hez képest. Ez a jelenség nyilvánvalóbb lehet az eltérő PAM-mal rendelkező és/vagy az evolúciósan távolabb álló fehérjék, mint a Cas12a fehérjék, az SaCas9 vagy az NmCas9 esetében, de kísérletesen az SpCas9 variánsokról is kiderült, hogy nem ugyanazokat a szekvenciákat preferálják. Az itt bemutatott munkában az egyik ilyen variánsra fejlesztettünk ki predikciós eljárást. A vonatkozó publikációnk elkészítésével egy időben jelent meg az első olyan predikciós algoritmus, amely kifejezetten egy SpCas9 variáns működését jelzi előre, amely megbízhatónak bizonyult és az általunk kifejlesztett algoritmushoz is hasonlítottuk (**24/B-C ábra**). A publikációnk revíziója során egy másik csoport is bemutatta, hogy különböző, széles körben használt megnövelt pontosságú fehérjére működési hatékonyságát lehetséges előre jelezni¹¹². Mindkét módszer az itt bemutatotthoz képest eltérő módszert alkalmazott: előre tervezett gRNS/célszekvencia könyvtárakat szintetizáltak, majd ezeken a könyvtárakon végezték el a hasítást, a különböző SpCas9 variánsokkal. Ennek a módszernek az alkalmazása a mi módszerünkkel szemben meglehetősen limitálja a vizsgált szekvenciákat, hiszen technológiailag, valamint pénzügyileg is korlátozott szekvenciákat így előállítani. Ezzel szemben a mi módszerünkkel teljesen véletlenszerű gRNS könyvtárak hozhatók létre, sokkal olcsóbban és egyszerűbben. Így úgy gondoljuk, hogy a számos még nem vizsgált Cas9 ortológ, valamint Cas9 variáns működési hatékonyságának meghatározása ezzel a módszerrel elérhetővé válik.

8. Összefoglalás

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 adathalmazon validáltunk.

9. Summary

Repurposing CRISPR/Cas9 systems for genome engineering tasks has revolutionized molecular biology, owing to the ability of the technique to modify genes significantly easier and cheaper than pre-existing methods. A key limitation for the widespread use of the method is that the cutting efficiency of the Cas9 nuclease varies across different target sequences, therefore exploring the behavior of the Cas9 molecule is key for all subsequent applications.

A widely used approach for the experimental testing of suitable target sequences for genome engineering applications, is capturing the best candidate gRNA from a pool. Here, I developed a simple, fluorescence-based method to examine the efficacy of different gRNAs within the genomic context. This method requires no further cloning steps, PCR, or sequencing. In this method, first, a GFP expression cassette is integrated into the DNA break sites introduced by the action of Cas9 and different gRNAs, and then the GFP positive cells can be quantified via flow cytometry to indicate the efficiency of the corresponding gRNA. To enhance the integration of this cassette, the plasmid was linearized within the cell. To further optimize the system as a whole, I have constructed a self-cleaving plasmid which bears a gRNA targeting its own plasmid, thus, this plasmid can be applied universally with any candidate gRNA. gRNA activity can be quantified with this method in numerous cell lines and the system was validated using existing methods. Besides experimental testing, I have also used this method for NHEJ-mediated gene integration.

Prediction algorithms for gRNA activity are available only for few Cas9 variants, and these prediction tools show varying accuracy. Cas9 orthologs and variants possess different sequence preferences, and reliable prediction methods are necessary for their routine use. To overcome this limitation, I have developed a simple method which can be used to generate large amounts of Cas9 cleavage information and therefore can be exploited to create a prediction algorithm based on the acquired data. Using this approach, we have determined the sequence specificity of a widely used high fidelity Cas9 variant (SpCas9-HF1) and developed a reliable prediction method which was then validated on several datasets from the literature.

10. Irodalomjegyzék

1. Cohen, S.N., Chang, A.C., Boyer, H.W., Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci*, 70, 3240-3244.
2. Rothstein, R.J. (1983) One-step gene disruption in yeast. *Methods Enzymol*, 101, 202-211.
3. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., Kucherlapati, R.S. (1985) Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature*, 317, 230-234.
4. Lin, F., Sperle, K., Sternberg, N. (1985) Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc Natl Acad Sci*, 82, 1391-1395.
5. Rouet, P., Smih, F., Jasin, M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 14, 8096-8106.
6. Rosen, L.E., Morrison, H.A., Masri, S., Brown, M.J., Springstubb, B., Sussman, D., Stoddard, B.L., Seligman, L.M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res*, 34, 4791-4800.
7. Seligman, L.M., Chisholm, K.M., Chevalier, B.S., Chadsey, M.S., Edwards, S.T., Savage, J.H., Veillet, A.L. (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res*, 30, 3870-3879.
8. Klug, A., Rhodes, D. (1987), *Cold Spring Harbor Symp Quant Biol*. Cold Spring Harbor Laboratory Press, Vol. 52, pp. 473-482.
9. Kim, Y.-G., Cha, J., Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci*, 93, 1156-1160.
10. Bibikova, M., Golic, M., Golic, K.G., Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 161, 1169-1175.
11. Porteus, M.H., Baltimore, D. (2003) Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science*, 300, 763-763.
12. Gaj, T., Gersbach, C.A., Barbas III, C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 31, 397-405.
13. Moscou, M.J., Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, 326, 1501-1501.

14. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326, 1509-1512.
15. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 29, 143-148.
16. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186, 757-761.
17. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, 30, 460.
18. Jansen, R., Embden, J.D.v., Gaastra, W., Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43, 1565-1575.
19. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 169, 5429-5433.
20. Mojica, F.J., García-Martínez, J., Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 60, 174-182.
21. Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151, 2551-2561.
22. Pourcel, C., Salvignol, G., Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151, 653-663.
23. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 1, 7.
24. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-1712.
25. Marraffini, L.A., Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 322, 1843-1845.

26. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., Van Der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321, 960-964.
27. Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.
28. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471, 602-607.
29. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P. (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*, 18, 67-83.
30. Nussenzweig, P.M., Marraffini, L.A. (2020) Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annu Rev Genet*, 54, 93-120.
31. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.
32. Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V. (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci*, 109, E2579-E2586.
33. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-823.
34. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, 339, 823-826.
35. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., Doudna, J. (2013) RNA-programmed genome editing in human cells. *eLife*, 2, e00471.
36. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.J., Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*, 31, 227-229.
37. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156, 935-949.

38. Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343.
39. Wilkinson, R.A., Martin, C., Nemudryi, A.A., Wiedenheft, B. (2019) CRISPR RNA-guided autonomous delivery of Cas9. *Nat Struct Mol Biol*, 26, 14-24.
40. Jiang, F., Zhou, K., Ma, L., Gressel, S., Doudna, J.A. (2015) A Cas9–guide RNA complex preorganized for target DNA recognition. *Science*, 348, 1477-1481.
41. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507, 62-67.
42. Sternberg, S.H., LaFrance, B., Kaplan, M., Doudna, J.A. (2015) Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, 527, 110-113.
43. Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., Sheen, J. (2013) Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol*, 31, 688-691.
44. Jiang, F., Doudna, J.A. (2017) CRISPR–Cas9 structures and mechanisms. *Annu Rev Biophys*, 46, 505-529.
45. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W. (2013) RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods*, 10, 973-976.
46. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154, 442-451.
47. Hilton, I.B., D'ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., Gersbach, C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 33, 510-517.
48. Kwon, D.Y., Zhao, Y.-T., Lamonica, J.M., Zhou, Z. (2017) Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat Commun*, 8, 1-8.
49. Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., Maehr, R. (2015) Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat Methods*, 12, 401-403.

50. Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., Zoldoš, V. (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res*, 44, 5615-5628.
51. Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., Jaenisch, R. (2016) Editing DNA methylation in the mammalian genome. *Cell*, 167, 233-247. e217.
52. Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S., Pederson, T. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci*, 112, 3002-3007.
53. Knight, S.C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L.B., Bosanac, L., Zhang, E.T., El Beheiry, M., Masson, J.-B., Dahan, M. (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, 350, 823-826.
54. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 343, 84-87.
55. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533, 420-424.
56. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R. (2017) Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature*, 551, 464-471.
57. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576, 149-157.
58. Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.-F., Sontheimer, E.J., Thomson, J.A. (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci*, 110, 15644-15649.
59. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S. (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520, 186-191.
60. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E.V., Zhang, F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163, 759-771.
61. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.-R.J. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523, 481-485.

62. Chatterjee, P., Lee, J., Nip, L., Koseki, S.R., Tysinger, E., Sontheimer, E.J., Jacobson, J.M., Jakimo, N. (2020) A Cas9 with PAM recognition for adenine dinucleotides. *Nat Commun*, 11, 1-6.
63. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., Joung, J.K. (2015) Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol*, 33, 1293-1298.
64. Gao, L., Cox, D.B., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., Zhang, F. (2017) Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol*, 35, 789-792.
65. Huang, T.P., Zhao, K.T., Miller, S.M., Gaudelli, N.M., Oakes, B.L., Fellmann, C., Savage, D.F., Liu, D.R. (2019) Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat Biotechnol*, 37, 626-631.
66. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H. (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 361, 1259-1262.
67. Walton, R.T., Christie, K.A., Whittaker, M.N., Kleinstiver, B.P. (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, 368, 290-296.
68. Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154, 1380-1389.
69. Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X. (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*, 11, 399-402.
70. Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., Joung, J.K. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 32, 569-576.
71. Guilinger, J.P., Thompson, D.B., Liu, D.R. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*, 32, 577.
72. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351, 84-88.
73. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., Joung, J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529, 490-495.

74. 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.
75. Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., Doudna, J.A. (2017) Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*, 550, 407-410.
76. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F. (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*, 36, 265.
77. 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.
78. Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K., Yeung, A.T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, 26, 4597-4602.
79. Mashal, R.D., Koontz, J., Sklar, J. (1995) Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat Genet*, 9, 177-183.
80. Brinkman, E.K., Chen, T., Amendola, M., van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*, 42, e168-e168.
81. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*, 31, 833-838.
82. Ramlee, M.K., Yan, T., Cheung, A.M., Chuah, C.T., Li, S. (2015) High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci Rep*, 5, 1-13.
83. Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J., Yee, J.-K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol*, 33, 175-178.
84. Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X. (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res*, 41, 9049-9061.
85. Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., Ikawa, M. (2013) Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci Rep*, 3, 1-6.

86. Tóth, E., Weinhardt, N., Bencsura, P., Huszár, K., Kulcsár, P.I., Tálás, 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.
87. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKolver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*, 27, 851-857.
88. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., Root, D.E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, 32, 1262-1267.
89. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., Giraldez, A.J. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods*, 12, 982-988.
90. Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C.A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J.S. (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res*, 25, 1147-1157.
91. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 34, 184-191.
92. Chari, R., Yeo, N.C., Chavez, A., Church, G.M. (2017) sgRNA Scorer 2.0: a species-independent model to predict CRISPR/Cas9 activity. *ACS Synth Biol*, 6, 902-904.
93. Labuhn, M., Adams, F.F., Ng, M., Knoess, S., Schambach, A., Charpentier, E.M., Schwarzer, A., Mateo, J.L., Klusmann, J.-H., Heckl, D. (2018) Refined sgRNA efficacy prediction improves large-and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res*, 46, 1375-1385.
94. Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B. (2018) DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol*, 19, 1-18.
95. Kim, N., Kim, H.K., Lee, S., Seo, J.H., Choi, J.W., Park, J., Min, S., Yoon, S., Cho, S.-R., Kim, H.H. (2020) Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat Biotechnol*, 38, 1328-1336.
96. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152, 1173-1183.

97. Inoue, H., Nojima, H., Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96, 23-28.
98. Matthews, B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*, 405, 442-451.
99. Kubat, M., Holte, R.C., Matwin, S. (1998) Machine learning for the detection of oil spills in satellite radar images. *Mach Learn*, 30, 195-215.
100. Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J. (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*, 17, 1-12.
101. Tusnady, G.E., Simon, I., Varadi, A., Aranyi, T. (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res*, 33, e9-e9.
102. Maresca, M., Lin, V.G., Guo, N., Yang, Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res*, 23, 539-546.
103. Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M.C., Urnov, F.D., Gregory, P.D., Cost, G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol Bioeng*, 110, 871-880.
104. Auer, T.O., Duroure, K., De Cian, A., Concordet, J.-P., Del Bene, F. (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res*, 24, 142-153.
105. Heigwer, F., Kerr, G., Boutros, M. (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods*, 11, 122-123.
106. Aach, J., Mali, P., Church, G.M. (2014) CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. *BioRxiv*, 005074.
107. Pinder, J., Salsman, J., Dellaire, G. (2015) Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res*, 43, 9379-9392.
108. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D., Wandless, T.J. (2010) A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem Biol*, 17, 981-988.
109. Bachu, R., Bergareche, I., Chasin, L.A. (2015) CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnol Bioeng*, 112, 2154-2162.
110. Tálas, A., Simon, D., Kulcsár, P., Varga, É., Welker, E. (2021) BEAR reveals that increased fidelity variants can successfully reduce the mismatch-tolerance of

- adenine but not cytosine base editors. Research Square, <https://doi.org/10.21203/rs.3.rs-105846/v1>.
111. Hámori, L., Kudlik, G., Szabényi, K., Kucsma, N., Szeder, B., Póti, Á., Uher, F., Várady, G., Szüts, D., Tóvári, J. (2020) Establishment and Characterization of a *Brcal*^{-/-}, *p53*^{-/-} Mouse Mammary Tumor Cell Line. *Int J Mol Sci*, 21, 1185.
 112. Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F. (2019) Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat Commun*, 10, 1-14.
 113. Briner, A.E., Donohoue, P.D., Goma, A.A., Selle, K., Slorach, E.M., Nye, C.H., Haurwitz, R.E., Beisel, C.L., May, A.P., Barrangou, R. (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell*, 56, 333-339.
 114. Perli, S.D., Cui, C.H., Lu, T.K. (2016) Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science*, 353, aag0511.
 115. Kalhor, R., Mali, P., Church, G.M. (2017) Rapidly evolving homing CRISPR barcodes. *Nat Methods*, 14, 195.
 116. Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., Church, G.M. (2018) Developmental barcoding of whole mouse via homing CRISPR. *Science*, 361, eaat9804.
 117. Tálás, A., Huszár, K., Kulcsár, P.I., Varga, J.K., Varga, É., Tóth, E., Welker, Z., Erdős, G., Pach, P.F., Welker, Á., Györgypál, Z., Tusnady, G.E., Welker, E. (2021) A method for characterizing Cas9 variants via a one-million target sequence library of self-targeting sgRNAs. *Nucleic Acids Res*.
 118. Graf, R., Li, X., Chu, V.T., Rajewsky, K. (2019) sgRNA sequence motifs blocking efficient CRISPR/Cas9-mediated gene editing. *Cell reports*, 26, 1098-1103. e1093.
 119. Arimbasseri, A.G., Rijal, K., Marais, R.J. (2013) Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829, 318-330.
 120. Kim, H.K., Kim, Y., Lee, S., Min, S., Bae, J.Y., Choi, J.W., Park, J., Jung, D., Yoon, S., Kim, H.H. (2019) SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Sci Adv*, 5, eaax9249.
 121. Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, 31, 233.
 122. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31, 827.

123. Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N. (2018) Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat Biotechnol*, 36, 179.
124. Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435, 646-651.
125. Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., Scharenberg, A.M. (2011) Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods*, 8, 671-676.
126. Kim, H., Um, E., Cho, S.-R., Jung, C., Kim, H., Kim, J.-S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods*, 8, 941.
127. Kim, H.K., Min, S., Song, M., Jung, S., Choi, J.W., Kim, Y., Lee, S., Yoon, S., Kim, H.H. (2018) Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. *Nat Biotechnol*, 36, 239.
128. Bae, S., Park, J., Kim, J.-S. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 30, 1473-1475.
129. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., Aryee, M.J., Joung, J.K. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 33, 187-197.
130. McVey, M., Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*, 24, 529-538.
131. Mladenov, E., Iliakis, G. (2011) Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat Res-Fundam Mol Mech Mutag*, 711, 61-72.
132. Bae, S., Kweon, J., Kim, H.S., Kim, J.-S. (2014) Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods*, 11, 705-706.
133. Huang, Y., Zhao, S., Shi, M., Liu, J., Liang, H. (2012) Microchip electrophoresis coupled with on-line magnetic separation and chemiluminescence detection for multiplexed immunoassay. *Electrophoresis*, 33, 1198-1204.
134. Vouillot, L., Thélie, A., Pollet, N. (2015) Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)*, 5, 407-415.
135. Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M. (2019) Predicting the

- mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*, 37, 64-72.
136. Chakrabarti, A.M., Henser-Brownhill, T., Monserrat, J., Poetsch, A.R., Luscombe, N.M., Scaffidi, P. (2019) Target-specific precision of CRISPR-mediated genome editing. *Mol Cell*, 73, 699-713. e696.
 137. Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., Sherwood, R.I. (2018) Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, 563, 646-651.
 138. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., Yildiz, A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv*, 3, eaao0027.
 139. Tálas, A., Kulcsár, P.I., Weinhardt, N., Borsy, A., Tóth, E., Szebényi, K., Krausz, S.L., Huszár, K., Vida, I., Sturm, Á., Gordos, B., Orsolya, H.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.
 140. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z. (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, 540, 144-149.
 141. Schmid-Burgk, J.L., Höning, K., Ebert, T.S., Hornung, V. (2016) CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat Commun*, 7, 1-12.
 142. He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., Feng, B. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res*, 44, e85-e85.
 143. Suzuki, K., Yamamoto, M., Hernandez-Benitez, R., Li, Z., Wei, C., Soligalla, R.D., Aizawa, E., Hatanaka, F., Kurita, M., Reddy, P. (2019) Precise in vivo genome editing via single homology arm donor mediated intron-targeting gene integration for genetic disease correction. *Cell Res*, 29, 804-819.
 144. Manna, P.T., Davis, L.J., Robinson, M.S. (2019) Fast and cloning-free CRISPR/Cas9-mediated genomic editing in mammalian cells. *Traffic*, 20, 974-982.
 145. Yoshimi, K., Oka, Y., Miyasaka, Y., Kotani, Y., Yasumura, M., Uno, Y., Hattori, K., Tanigawa, A., Sato, M., Oya, M. (2020) Combi-CRISPR: combination of NHEJ and HDR provides efficient and precise plasmid-based knock-ins in mice and rats. *Hum Genet*, 1-11.

11. Saját publikációk jegyzéke

11.1. Disszertációhoz kapcsolódó publikációk jegyzéke

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., Szebényi, 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**

11.2. Disszertációtól független publikációk jegyzéke

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**

12. Köszönetnyilvánítás

Mindenekelőtt Dr. Welker Ervinnek szeretnék köszönetet mondani, hogy a laborjában készíthettem el a disszertációm. Nagyon hálás vagyok a sok éves támogatásáért, tanácsáért és bizalomért.

Szeretném megköszönni Kulcsár Péternek a sok éves baráti és szakmai támogatást, amivel aktívan segítette az összes publikációm létrejöttét.

Köszönöm a kollégáknak, akik segítettek a disszertációm alapjául szolgáló publikációk létrejöttét: Dr. Tusnád Gábor, Dr. Szabó Kornélia, Dr. Györgypál Zoltán, Varga Júlia, Welker Ágnes, Welker Zsombor, Dr. Pongor Lőrinc, Dr. Kulcsár Péter, Varga Éva, Huszár Krisztina, Dr. Tóth Eszter.

Köszönöm Dr. Fodor Elfriedának a statisztikai kiértékelésekkel kapcsolatos segítségét, valamint a hasznos észrevételeiért a közös publikációinkban.

Köszönöm a laborunk jelenlegi és volt asszisztenseinek a kiváló munkát, amivel nagyban segítettek a munkáimat: Szűcsné Pulinka Ildikó, Szűcs Judit, Czene Bernadett, Turgyán Luca, Fetter Dávid, Burkus Lilla, Karl Vivien, Bencsura Petra.

Köszönettel tartozom a Welker labor minden volt és jelenlegi munkatársának a sok segítségért és a támogató és vidám légkörért, amiben itt dolgozhattam.

Köszönöm Dr. Várady Györgynek és Dr. Szabó Editnek, hogy a FACS laborban mindig készséggel segítettek.

Köszönöm Dr. Zeke Andrásnak, hogy elvállalta a disszertációm előbírálását, valamint Dr. Kulcsár Péternek, Dr. Welker Ervinnek és Végi Vanesszának, hogy észrevételeikkel javították a doktori értekezésem színvonalát.

Szeretném megköszönni a Semmelweis Egyetemnek, valamint a TTK Enzimológiai Intézet munkatársainak a sok éves támogató légkört.

Nagyon sok köszönet illeti a „szokásos péntek esti mulatozás”, valamint az „Enzim Ulti Kör” csoportok tagjait, akik a sok kihívást jelentő időket segítettek sokkal vidámabbá tenni.

Nagyon sok köszönettel tartozom Bartos Zsuzsának, akitől rengeteg lelki és szakmai segítséget kaptam, és nagyban segítette a PhD munkám létrejöttét.

Végül, de nem utolsó sorban szeretném megköszönni családomnak a rengeteg lelki és anyagi támogatást, nélkülük ma ez a munka nem létezne.

13. Függelék

13.1. Önvágó plazmidokhoz használt primerek listája

Spacerek klónozásához használt primerek önvágó plazmidokkal

spacer neve	oligo 1	oligo 2
OTL1spacer	CACCGGCGCAACGCGATCGCGTAA	AAACTTACGCGATCGCGTTGCGCC
OTL2spacer	CACCGATCAATTTGCGCCGCAACGC	AAACGCGTTGCGGGCGAAATTGATC
prp1spacer	CACCGTTGGCCCCATCCACCGCCAT	AAACATGGCGGTGGATGGGGCCAAC
prp2spacer	CACCGGCCTCCCTCATCCACGATC	AAACGATCGTGGGATGAGGGAGGCC
prp3spacer	CACCGATTTTGAGATCAGTCATCA	AAACTGATGACTGATCTGCAAAATC
prp4spacer	CACCGTCTGATCGTGGGATGAGGG	AAACCCCTCATCCACGATCAGGAC
prp5spacer	CACCGTCAGTCATCATGGCGAACCT	AAACAGGTTGCGCCATGATGACTGAC
piwi1spacer	CACCCGGTACACGACTATCCGTGC	AAACGCACGGATAGTCGTGTACCG
piwi2spacer	CACCCGGATACTAGCCGCATAGC	AAACGCTATGCGGGCTAGTATCCGG
shospace	CACCGCTACGGCTCCTCTGCGCG	AAACCGCGCAGAGAGGAGCCGTAGC
Rbspacer	CACCGCAAAGCCCTACACCCAAA	AAACTTTGGGGTGTAGGGCTTTGC
Ttnspacer	CACCCGAAGAAGTCAAAACTCAG	AAACCGTGAGTTTTGACTTCTTCG
ptenspac	CACCGAGATCGTTAGCAGAAAACAA	AAACTTTGTTTCTGCTAACGATCTC
p53spacer	CACCGCCTCGAGCTCCCTCTGAGCC	AAACGGCTCAGAGGGAGCTCGAGGC
hprp1spacer	CACCGCAGTCATTATGGCGAACCT	AAACAGGTTGCGCCATAATGACTGC
hprp2spacer	CACCGAACCTTGGCTGCTGGATGC	AAACGCATCCAGCAGCCAAGGTTT
pten2spacer	caccGCTAACCTAAGTTCTGCCACG	aaacCGTGGCAGAACTTAGGTTAGC
prp8spacer	CACCGCCTCTGCAAAAAGCGGCCA	AAACTGGCCGCTTTTTGCGAGGGC
prp9spacer	CACCGCCACGAGAATGCGAAGGAAC	AAACGTTCTTTCGATTCTCGTGGC
prp10spacer	CACCGTCTTTCGATTCTCGTGGTC	AAACGACCACGAGAATGCGAAGGAC
dcun1d2spacer	caccgTCATGGACTAGCCTTCAGGG	aaacCCCTGAAGGCTAGTCCATGAC
vim1spacer	caccGTAACCTTTTCTACGTCTTT	aaacAAAGACGTAGAAAAGGTTTAC
vim2spacer	caccGTCAGGAGCGCAAGATAGATT	aaacAATCTATCTTGGCTCCTGAC
vim3spacer	caccGAATAAAAATTGCACACT	aaacAGTGTGTGCAATTTTTATT
vim4spacer	caccGACAGTGCAGTACCAGCAAGA	aaacTCTTGCTGGTACTGCACTGTC
dcun3spacer	caccGAGGGTCATGGACTAGCCTTC	aaacGAAGGCTAGTCCATGACCCTC
dcun5spacer	caccGTGTCACGGACTAGCCTTCA	aaacTGAAGGCTAGTCCGTGACAC
dcun6spacer	caccGAGTGTACGGACTAGCCTTC	aaacGAAGGCTAGTCCGTGACACTC

TIDE-hoz és T7EI-hez használt genomi PCR primerek

primer neve	primer szekvenciája
pten-fwd	TGCGAGGATTATCCGTCTTC
pten-rev	ATAGACGCTTTGGGTTGTGC
p53-fwd	ACGTGGTTGGTTACCTCTGC
p53-rev	CTCTGGGCCTTACCAGGATA
pten2-fwd	GGCCAACTACTCCTGTGCAT
pten2-rev	GCATCTTCCAGACACAACAA
rb-fwd	GCGTGCTACACATGGCTTTA
rb-rev	AGGGAAGGAAAGGGAAGTGA
prpstart-for	TACCAGTTGGGGTAACTCTTGTCA
prpstart-rev	ATGTACCGCTACCCTAACCAAGTG
prpstop-for	GATCCATTTTGGCAACGACT
prpstop-rev	TAGTTGGGGCTTTGTTTTGG
piwi2-fwd	GCCTAGGCCAGGATTGAACT
piwi2-rev	TGGTACTGCAGCAACACAGG

dcun1d2-for	GTGTCCAGTTAAATGGTCATG
dcun1d2-rev	CTGTAGTCATCAACCCTTG

Az integrált önvágó plazmid szekvenálásához használt primerek

primer neve	primer szekvenciája
plasmid 5'-for	CTCGCTCTGCTAATCCTGTT
plasmid 3'-rev	CATGAAGCAGCAGACTTCT
p53-fwd	ACGTGGTTGGTTACCTCTGC
p53-rev	CCTCTGACTTGAGCGTCG

13.2. Könyvtárkészítéshez, valamint az stgRNS-ek klónozásához használt primerek

Primerek az NGG PAM-os stgRNS könyvtárhoz

oligo neve	szekvencia
NGG-könyvtár-random-oligo	GCCTTAGGTTAACTTGCTATTTCTAGCTCTAACCCNNNNNNNNNN NNNNNNNNNNNCGCTAGCATTATACCTAGGACTGAGCTAGCT GTCAA
NGG-könyvtár-feltöltés	TTGACAGCTAGCTCAGTCCTAGGT
NGG-könyvtár-vektor-for	TAGAAATAGCAAGTTAACCTAAG
NGG-könyvtár-vektor-rev	TACCTAGGACTGAGCTAG

Primerek az NGG PAM-os stgRNS könyvtárhoz

oligo neve	szekvencia
NAG-könyvtár-random-oligo	GCCTTAAGTTAACTTGCTATTTCTAGCTCTAACTCNNNNNNNNNN NNNNNNNNNNNCGCTAGCATTATACCTAGGACTGAGCTAGCT GTCAA
NAG-könyvtár-feltöltés	TTGACAGCTAGCTCAGTCCTAGGT
NAG-könyvtár-vektor-for	TAGAAATAGCAAGTTAACTAAG
NAG-könyvtár-vektor-rev	TACCTAGGACTGAGCTAG

NGS primerek

oligo neve	szekvencia
NGS-lib-i5-for	AATGATACGGCGACCACCGAGATCTACAC[i5- index]TCGTCGGCAGCGTC
NGS-lib-i7-rev	CAAGCAGAAGACGGCATACGAGAT[i7- index]GTCTCGTGGGCTCGG

A bakteriális könyvtárak NGS indexelése

minta neve	leírás	i5 index neve	i5 index szekvenciája	i7 index neve	i7 index szekvenciája
lib100	100 000-es könyvtár	NGS_i5_S507	AAGGAGTA	NGS_i7_N703	TTCTGCCT
lib500	500 000-es könyvtár	NGS_i5_S508	CTAAGCCT	NGS_i7_N703	TTCTGCCT

lib600	600 000-es könyvtár	NGS_i5_S510	CGTCTAAT	NGS_i7_N703	TTCTGCCT
WT100/1	WT-SpCas9-cel vágott 100 000-es könyvtár /1	NGS_i5_S502	CTCTCTAT	NGS_i7_N701	TCGCCTTA
WT100/2	WT-SpCas9-cel vágott 100 000-es könyvtár /2	NGS_i5_S503	TATCCTCT	NGS_i7_N701	TCGCCTTA
WT100/3	WT-SpCas9-cel vágott 100 000-es könyvtár /3	NGS_i5_S505	GTAAGGAG	NGS_i7_N701	TCGCCTTA
WT500/1	WT-SpCas9-cel vágott 500 000-es könyvtár /1	NGS_i5_S506	ACTGCATA	NGS_i7_N701	TCGCCTTA
WT500/2	WT-SpCas9-cel vágott 500 000-es könyvtár /2	NGS_i5_S507	AAGGAGTA	NGS_i7_N701	TCGCCTTA
WT500/3	WT-SpCas9-cel vágott 500 000-es könyvtár /3	NGS_i5_S508	CTAAGCCT	NGS_i7_N701	TCGCCTTA
WT600/1	WT-SpCas9-cel vágott 600 000-es könyvtár /1	NGS_i5_S510	CGTCTAAT	NGS_i7_N701	TCGCCTTA
WT600/2	WT-SpCas9-cel vágott 600 000-es könyvtár /2	NGS_i5_S502	CTCTCTAT	NGS_i7_N702	CTAGTACG
WT600/3	WT-SpCas9-cel vágott 600 000-es könyvtár /3	NGS_i5_S503	TATCCTCT	NGS_i7_N702	CTAGTACG
HF100/1	SpCas9-HF1-el vágott 100 000-es könyvtár /1	NGS_i5_S505	GTAAGGAG	NGS_i7_N702	CTAGTACG
HF100/2	SpCas9-HF1-el vágott 100 000-es könyvtár /2	NGS_i5_S506	ACTGCATA	NGS_i7_N702	CTAGTACG
HF100/3	SpCas9-HF1-el vágott 100 000-es könyvtár /3	NGS_i5_S507	AAGGAGTA	NGS_i7_N702	CTAGTACG
HF500/1	SpCas9-HF1-el vágott 500 000-es könyvtár /1	NGS_i5_S508	CTAAGCCT	NGS_i7_N702	CTAGTACG
HF500/2	SpCas9-HF1-el vágott 500 000-es könyvtár /2	NGS_i5_S510	CGTCTAAT	NGS_i7_N702	CTAGTACG
HF500/3	SpCas9-HF1-el vágott 500 000-es könyvtár /3	NGS_i5_S502	CTCTCTAT	NGS_i7_N703	TTCTGCCT
HF600/1	SpCas9-HF1-el vágott 600 000-es könyvtár /1	NGS_i5_S503	TATCCTCT	NGS_i7_N703	TTCTGCCT
HF600/2	SpCas9-HF1-el vágott 600 000-es könyvtár /2	NGS_i5_S505	GTAAGGAG	NGS_i7_N703	TTCTGCCT
HF600/3	SpCas9-HF1-el vágott 600 000-es könyvtár /3	NGS_i5_S506	ACTGCATA	NGS_i7_N703	TTCTGCCT

A pAT-9218 és pAT-9251 plazmidokhoz felhasznált primerek

oligo name	sequence
bactCas9-for	CAGTGATAGAGAAAAGAATTCAAAAAAAGAGGAGAAAGGATCTATGGA CTATAAGGACCACGACGG
bactCas9-rev	TTATTTGATGCCTGGAGATCCTTACCTTACTTTTTCTTTTTTGCTGGCCG

Primerek az 57 spacer klónozásához GFxFP esszéhez

plazmid neve	spacer szekvenciája	oligo1	oligo2
9328	GAAAAAAGAAGGAA AAGCC	CACCGAAAAAAGAAGGAA AAGCC	ACCCGGCTTTTCCT TCTTTTTTC
9336	GAAAAAAGGAAAGGG GAAAG	CACCGAAAAAAGGAAAGGG GAAAG	ACCCCTTTCCCCTT TCCTTTTTTC
9305	GAAACTTGGAGACCCG ACCT	CACCGAAACTTGGAGACCC GACCT	ACCCAGGTCGGGT CTCCAAGTTTC
9350	GAAAGAGAGGACTGG GTTCC	CACCGAAAGAGAGGACTGG GTTCC	ACCCGGAACCCAG TCCTCTTTTC
9300	GAAATGACGAAGAAA ACTGT	CACCGAAATGACGAAGAAA ACTGT	ACCCACAGTTTTTC TTCGTCATTTTC
9319	GAACACAGAGCTCAGG GGAC	CACCGAACACAGAGCTCAG GGGAC	ACCCGTCCCCTGA GCTCTGTGTTTC
9330	GACAAAACGGGGGTG GGCAA	CACCGACAAAACGGGGGTG GGCAA	ACCCTTGCCACC CCCGTTTTGTC
9329	GACCAGAATAGACCCA AACC	CACCGACCAGAATAGACCC AAACC	ACCCGGTTTGGGT CTATTCTGGTC
9339	GACCCAATGTCCACTA AGGA	CACCGACCCAATGTCCACTA AGGA	ACCCTCCTTAGTG GACATTGGGTC
9332	GACCTTGGTACTGGCA CCAC	CACCGACCTTGGTACTGGCA CCAC	ACCCGTGGTGCCA GTACCAAGGTC
9337	GACTATTAATGAACA AAGA	CACCGACTATTAATGAACA AAGA	ACCCTCTTTGTTC TTAATAGTC
9325	GAGAAAGGCAGTACA GGAAC	CACCGAGAAAGGCAGTACA GGAAC	ACCCGTTCCCTGTA CTGCCTTTCTC
9338	GAGAAGAAGAAATCA CAAGT	CACCGAGAAGAAGAAATCA CAAGT	ACCCACTTGTGAT TTCTTCTCTC
9341	GAGAAGCCACAGGGTG GCTG	CACCGAGAAGCCACAGGGT GGCTG	ACCCAGCCACCC TGTGGCTTCTC
9335	GAGACAATAGAGACA AGAAA	CACCGAGACAATAGAGACA AGAAA	ACCCTTTCTTGTCT CTATTGTCTC
9331	GAGATCCTAGAGCCTG GCAC	CACCGAGATCCTAGAGCCTG GCAC	ACCCGTGCCAGGC TCTAGGATCTC
9334	GAGGATCCAGTGCCAC CTTC	CACCGAGGATCCAGTGCCA CCTTC	ACCCGAAGGTGGC ACTGGATCCTC
9355	GAGGCCAGAACACGA GGCCC	CACCGAGGCCAGAACACGA GGCCC	ACCCGGCCTCGT GTTCTGGCCTC
9344	GAGGCTGGGCTGGCCG GGCG	CACCGAGGCTGGGCTGGCC GGGCG	ACCCCGCCGGCC AGCCAGCCTC
9347	GAGGGCGGAAGGCCTG GGCC	CACCGAGGGCGGAAGGCCT GGCC	ACCCGGCCAGGC CTCCGCCCTC
9313	GATGTTAGCACACAGC ACAT	CACCGATGTTAGCACACAGC ACAT	ACCCATGTGCTGT GTGCTAACATC
9352	GATTACATTCAAGGCC AGCC	CACCGATTACATTCAAGGCC AGCC	ACCCGGCTGGCCT TGAATGTAATC
9312	GCAACAGAACTGTGT AAGC	CACCGCAACAGAACTGTG TAAGC	ACCCGTTACACA GTTTCTGTTGC

9321	GCAAGAGAAATGCAG AGAAG	CACCGCAAGAGAAATGCAG AGAAG	ACCCCTTCTCTGC ATTTCTCTTGC
9301	GCAAGGTGGGGAGAC ATAGC	CACCGCAAGGTGGGGAGAC ATAGC	ACCCGCTATGTCT CCCCACCTTGC
9303	GCACAGATAAGAAAGT GAAC	CACCGCACAGATAAGAAAG TGAAC	ACCCGTTCACTTTC TTATCTGTGC
9314	GCACCCACAGTGGAGG AATA	CACCGCACCCACAGTGGAG GAATA	ACCCTATTCCTCC ACTGTGGGTGC
9348	GCACCCACAGAGGGG GGAG	CACCGCACCCACAGAGGG GGGAG	ACCCCTCCCCCCT CTGTGGGGTGC
9356	GCAGACACCAGGCTGG CTGC	CACCGCAGACACCAGGCTG GCTGC	ACCCGCAGCCAGC CTGGTGTCTGC
9316	GCAGCAGGCAGATTGT CTGA	CACCGCAGCAGGCAGATTG TCTGA	ACCCTCAGACAAT CTGCCTGCTGC
9302	GCAGCAGGCCCTTCCT TCCA	CACCGCAGCAGGCCCTTCCT TCCA	ACCCTGGAAGGAA GGGCCTGCTGC
9318	GCAGCTTGTTTCCCG AGTA	CACCGCAGCTTGTTTCCCG AGTA	ACCCTACTCGGGG AAACAAGCTGC
9353	GCAGTCCCGGGCCGC CGTG	CACCGCAGTCCCGGGCCG CCGTG	ACCCACGGCGGC CCGGGACCTGC
9342	GCCACAACACTGTGGTGG TTTT	CACCGCCACAACACTGTGGTGG TTTT	ACCCAAAACCACC ACAGTTGTGGC
9317	GCCAGCAGGGCACTCA ATGG	CACCGCCAGCAGGGCACTC AATGG	ACCCCATGAGT GCCCTGCTGGC
9346	GCCCAGGGTGCGGTGG GCCC	CACCGCCCAGGGTGCGGTG GGCCC	ACCCGGGCCACC GCACCCTGGGC
9306	GCCTTGAACACTCAGGGT TGGA	CACCGCCTTGAACACTCAGGGT TGGA	ACCCTCCAACCCT GAGTTCAAGGC
9354	GCGCGCGCGCGCGCGC ATGC	CACCGCGCGCGCGCGCGC CATGC	ACCCGCATGCGCG CGCGCGCGCGC
9343	GGAAGCTGTGGATGCT TCTG	CACCGGAAGCTGTGGATGCT TCTG	ACCCAGAAGCAT CCACAGCTCC
9308	GGAATAAACATAAAAT CTCA	CACCGGAATAAACATAAAA TCTCA	ACCCTGAGATTTT ATGTTTATTCC
9304	GGACAAAACAGGGAG CCTACT	CACCGGACAAAACAGGGAG CCTACT	ACCCAGTGGCTCC CTGTTTTGTCC
9311	GGAGACAGACTGTGCC ATTG	CACCGGAGACAGACTGTGC CATTG	ACCCCAATGGCAC AGTCTGTCTCC
9351	GGCCCTCAAGGCCAG GCTT	CACCGCCCTCAAGGCCA GGCTT	ACCCAAGCCTGGC CTGAGGGGCC
9340	GGCTGGAGGTAGGAAG GCCT	CACCGGCTGGAGGTAGGAA GGCCT	ACCCAGGCCTTCC TACCTCCAGCC
9323	GGGAGACATGTAAGAA GCAT	CACCGGGAGACATGTAAGA AGCAT	ACCCATGCTTCTT ACATGTCTCCC
9345	GGGCTGAGCCCAAGAA AGCC	CACCGGGCTGAGCCCAAGA AAGCC	ACCCGGCTTTCTT GGGCTCAGCCC
9309	GGTCAATGCTGTCAGG CAGG	CACCGGTCAATGCTGTCAGG CAGG	ACCCCTGCCTGA CAGCATTGACC
9315	GGTGCAGCCCGTTGT AAGC	CACCGGTGCAGCCCGTTGT AAGC	ACCCGCTTACAAC CGGGCTGCACC
9327	GTATAAAAATAGACAG AATT	CACCGTATAAAAATAGACA GAATT	ACCCAATTCTGTG TATTTTTATAC
9324	GTCAAGTGACAGGCCT GGAA	CACCGTCAAGTGACAGGCCT GGAA	ACCCTTCCAGGCC TGTCATTGAC
9310	GTCCCAAGCCAGCAGC CACT	CACCGTCCCAAGCCAGCAG CCTACT	ACCCAGTGGCTGC TGGCTTGGGAC

9326	GTCCCAGATGTGAAAA AGGT	CACCGTCCCAGATGTGAAA AAGGT	ACCCACCTTTTTCA CATCTGGGAC
9349	GTCTGGTTTCGAGACC AGCC	CACCGTCTGGTTTCGAGACC AGCC	ACCCGGCTGGTCT CGAAACCAGAC
9333	GTGAAGACCTTCACCA GGAG	CACCGTGAAGACCTTCACCA GGAG	ACCCCTCTGGTG AAGGTCTTAC
9322	GTGAGGAGGAAGTGG GGCTC	CACCGTGAGGAGGAAGTGG GGCTC	ACCCGAGCCCCAC TTCCTCCTAC
9320	GTGGTTCTAAAGTGGC TTTA	CACCGTGGTTCTAAAGTGGC TTTA	ACCCTAAAGCCAC TTTAGAACAC
9307	GTTAGTCATCATGACC TCCA	CACCGTTAGTCATCATGACC TCCA	ACCCTGGAGGTCA TGATGACTAAC

Primerek az stgRNS és kanonikus gRNS összehasonlításához

Kanonikus gRNS-ek klónozásához használt primerek

gRNS száma	Spacer szekvenciája	oligo1	oligo2
gRNS 1	GGAGAAAAAGTGAC AGAGCG	CACCGGAGAAAAAGTG ACAGAGCG	AAACCGCTCTGTCAC TTTTTCTCC
gRNS 2	GAGAGAAATGTGAC CGCGCC	CACCGAGAGAAATGTG ACCGCGCC	AAACGGCGCGGTCA CATTCTCTC
gRNS 3	GACCGCAAAGTAAC CATAGA	CACCGACCGCAAAGTA ACCATAGA	AAACTCTATGGTTAC TTTGCGGTC
gRNS 4	GCGCGCATAGTAAC AGAGCC	CACCGCGCGCATAGTA ACAGAGCC	AAACGGCTCTGTTAC TATGCGCGC
gRNS 5	GGGGCAATGTGGG CGCGAA	CACCGGGGCAATGTG GGCGCGAA	AAACTTCGCGCCCAC ATTGCCCCC
gRNS 6	GACCGGACTGTGTCC GGTAA	CACCGACCGGACTGTGT CCGGTAA	AAACTACCGGACA CAGTCCGGTC
gRNS 7	GAGTGAAAGAGGAC ACGGCC	CACCGAGTGAAAGAGG ACACGGCC	AAACGGCCGTGTCCT CTTTCACTC
gRNS 8	GACCGAAAAAAGC GGTAG	CACCGACCGAAAAAA GCGGGTAG	AAACCTACCGCTTT TTTTCGGTC
gRNS 9	GACCGTAACATTAGC AGGGA	CACCGACCGTAACATTA GCAGGGA	AAACTCCCTGTAAT GTTACGGTC
gRNS 10	GTGGGCCAGTTGTCC GGTTA	CACCGTGGGCCAGTTGT CCGGTTA	AAACTAACCGGACA ACTGGCCCAC
gRNS 11	GATATAGCCCGGCA ACCAGA	CACCGATATAGCCCGG CAACCAGA	AAACTCTGGTTGCCG GGCTATATC
gRNS 12	GTAGCGTAGCATGTA ACACG	CACCGTAGCGTAGCAT GTAACACG	AAACCGTGTTACATG CTACGCTAC
gRNS 13	GAAGTACGCCCAGA ACAAGG	CACCGAAGTACGCCCA GAACAAGG	AAACCCTTGTTCTGG GCGTACTTC
gRNS 14	GATCCCAGCCCAGA AACATG	CACCGATCCCAGCCCA GAAACATG	AAACCATGTTTCTGG GCTGGGATC
gRNS 15	GAAGCATTGCACCA AACACT	CACCGAAGCATTGCAC CAAACACT	AAACAGTGTTTGGTG CAATGCTTC
gRNS 16	GATATGTATCCCCAA TACGA	CACCGATATGTATCCCC AATACGA	AAACTCGTATTGGG GATACATATC
gRNS 17	GGAATGGGTCAAGA GACATC	CACCGGAATGGGTCAA GAGACATC	AAACGATGTCTCTTG ACCCATTCC
gRNS 18	GTTGCATTCCATCAA TCATA	CACCGTTGCATTCCATC AATCATA	AAACTATGATTGATG GAATGCAAC

gRNS 19	GGTACCTCGCACCAA CCAGT	CACCGGTACCTCGCACC AACCAGT	AAACACTGGTTGGT GCGAGGTACC
gRNS 20	GCTATGTGCCAAGTA ACACA	CACCGCTATGTGCCAAG TAACACA	AAACTGTGTTACTTG GCACATAGC

stgRNS-ek klónozásához használt primerek

stgRNS száma	Spacer szekvenciája	oligo1	oligo2
gRNS 1	GGAGAAAAAGTGAC AGAGCG	CACCGGAGAAAAAGTG ACAGAGCG	ACCCCGCTCTGTCAC TTTTTCTCCGGTG
gRNS 2	GAGAGAAATGTGAC CGCGCC	CACCGAGAGAAATGTG ACCGCGCC	ACCCGGCGCGGTCA CATTCTCTCGGTG
gRNS 3	GACCGCAAAGTAAC CATAGA	CACCGACGCAAAGTA ACCATAGA	ACCTCTATGGTTAC TTGCGGTCCGGTG
gRNS 4	GCGCGCATAGTAAC AGAGCC	CACCGCGCGCATAGTA ACAGAGCC	ACCCGCTCTGTTAC TATGCGCGCGGTG
gRNS 5	GGGGCAATGTGGG CGCGAA	CACCGGGGCAATGTG GGCGCGAA	ACCCTTCGCGCCCAC ATTGCCCCCGGTG
gRNS 6	GACCGACTGTGTCC GGTAA	CACCGACCGACTGTGT CCGGTAA	ACCCTTACCGGACAC AGTCCGGTCCGGTG
gRNS 7	GAGTGAAAGAGGAC ACGGCC	CACCGAGTGAAAGAGG ACACGGCC	ACCCGGCCGTGTCCT CTTCACTCGGTG
gRNS 8	GACCGAAAAAAGC GGGTAG	CACCGACGAAAAAAA GCGGGTAG	ACCCCTACCCGCTTT TTTTCGGTCCGGTG
gRNS 9	GACCGTAACATTAGC AGGGA	CACCGACCGTAACATTA GCAGGGA	ACCCTCCCTGCTAAT GTTACGGTCCGGTG
gRNS 10	GTGGGCCAGTTGTCC GGTTA	CACCGTGGGCCAGTTGT CCGGTTA	ACCCTAACCGGACA ACTGGCCACCGGTG
gRNS 11	GATATAGCCCGCA ACCAGA	CACCGATATAGCCCGG CAACCAGA	ACCCTCTGGTTGCCG GGCTATATCGGTG
gRNS 12	GTAGCGTAGCATGTA ACACG	CACCGTAGCGTAGCAT GTAACACG	ACCCCGTGTTACATG CTACGCTACGGTG
gRNS 13	GAAGTACGCCCAGA ACAAGG	CACCGAAGTACGCCCA GAACAAGG	ACCCCTTGTCTCTGG GCGTACTTCCGGTG
gRNS 14	GATCCCAGCCCAGA AACATG	CACCGATCCCAGCCCA GAAACATG	ACCCCATGTTTCTGG GCTGGGATCCGGTG
gRNS 15	GAAGCATTGCACCA AACACT	CACCGAAGCATTGCAC CAAACACT	ACCCAGTGTTTGGTG CAATGCTTCCGGTG
gRNS 16	GATATGTATCCCCAA TACGA	CACCGATATGTATCCCC AATACGA	ACCCTCGTATTGGGG ATACATATCCGGTG
gRNS 17	GGAATGGGTCAAGA GACATC	CACCGAATGGGTCAA GAGACATC	ACCCGATGTCTCTTG ACCCATTCCGGTG
gRNS 18	GTTGCATTCCATCAA TCATA	CACCGTTGCATTCCATC AATCATA	ACCCTATGATTGATG GAATGCAACGGTG
gRNS 19	GGTACCTCGCACCAA CCAGT	CACCGGTACCTCGCACC AACCAGT	ACCCACTGGTTGGTG CGAGGTACCGGTG
gRNS 20	GCTATGTGCCAAGTA ACACA	CACCGCTATGTGCCAAG TAACACA	ACCCTGTGTTACTTG GCACATAGCGGTG

GFxFP célplazmidok klónozásához használt primerek

gRNS száma	GFxFP oligo 1	GFxFP oligo 2
gRNS 1	GATCAACACCGAGAGAAAAAGTGACA GAGCGGGGTTAGA	AATTTCTAACCCGCTCTGTCAC TTTTCTCCGGTGTT
gRNS 2	GATCAACACCGAGAGAAATGTGACC GCGCCGGGTTAGA	AATTTCTAACCCGCGCGGTCA CATTCTCTCGGTGTT
gRNS 3	GATCAACACCGACCGCAAAGTAACC ATAGAGGGTTAGA	AATTTCTAACCCCTCTATGGTTAC TTTGCGGTCGGTGTT
gRNS 4	GATCAACACCGCGCGCATAGTAACA GAGCCGGGTTAGA	AATTTCTAACCCGCTCTGTTAC TATGCGCGCGGTGTT
gRNS 5	GATCAACACCGGGGGCAATGTGGGC GCCAAGGGTTAGA	AATTTCTAACCCCTTCGCGCCAC ATTGCCCCGGTGTT
gRNS 6	GATCAACACCGACCGGACTGTGTCC GGTAAGGGTTAGA	AATTTCTAACCCCTTACCGGACA CAGTCCGGTCGGTGTT
gRNS 7	GATCAACACCGAGTGAAAGAGGACA CGGCCGGGTTAGA	AATTTCTAACCCGGCCGTGTCTCT CTTCACTCGGTGTT
gRNS 8	GATCAACACCGACCGAAAAAAGCG GGTAGGGGTTAGA	AATTTCTAACCCCTACCCGCTTT TTTTCGGTTCGGTGTT
gRNS 9	GATCAACACCGACCGTAACATTAGC AGGGAGGGTTAGA	AATTTCTAACCCCTCCCTGCTAAT GTTACGGTCGGTGTT
gRNS 10	GATCAACACCGTGGGCCAGTTGTCCG GTTAGGGTTAGA	AATTTCTAACCCCTAACCGGACA ACTGGCCACGGTGTT
gRNS 11	GATCAACACCGATATAGCCCGCAA CCAGAGGGTTAGA	AATTTCTAACCCCTCTGGTTGCCG GGCTATATCGGTGTT
gRNS 12	GATCAACACCGTAGCGTAGCATGTA ACACGGGGTTAGA	AATTTCTAACCCCGTGTACATG CTACGCTACGGTGTT
gRNS 13	GATCAACACCGAAGTACGCCAGAA CAAGGGGGTTAGA	AATTTCTAACCCCTTGTCTGG GCGTACTTCGGTGTT
gRNS 14	GATCAACACCGATCCCAGCCAGAA ACATGGGGTTAGA	AATTTCTAACCCCATGTTTCTGG GCTGGGATCGGTGTT
gRNS 15	GATCAACACCGAAGCATTGCACCAA ACACTGGGTAGTA	AATTTCTAACCCAGTGTGTTGGTG CAATGCTTCGGTGTT
gRNS 16	GATCAACACCGATATGTATCCCAAT ACGAGGGTTAGA	AATTTCTAACCCCTCGTATTGGGG ATACATATCGGTGTT
gRNS 17	GATCAACACCGGAATGGGTCAAGAG ACATCGGGTTAGA	AATTTCTAACCCGATGTCTCTTG ACCCATTCCGGTGTT
gRNS 18	GATCAACACCGTTGCATTCCATCAAT CATAGGGTTAGA	AATTTCTAACCCCTATGATTGATG GAATGCAACGGTGTT
gRNS 19	GATCAACACCGGTACCTCGCACCAA CCAGTGGGTAGTA	AATTTCTAACCCACTGGTTGGTG CGAGGTACCGGTGTT
gRNS 20	GATCAACACCGCTATGTGCCAAGTA ACACAGGGTTAGA	AATTTCTAACCCCTGTGTTACTTG GCACATAGCGGTGTT